



MINISTÉRIO DA EDUCAÇÃO
Universidade Federal de Alfenas. Unifal-MG
Rua Gabriel Monteiro da Silva, 714 – Alfenas/MG – CEP 37130-000
Fone: (35) 3299-1000. Fax: (35) 3299-1063



ALBERT KATCHBORIAN NETO

**INTEGRATING METABOLOMICS AND COMPUTATIONAL TOOLS TO
DISCOVER ANTI-INFLAMMATORY BIOACTIVE AGENTS IN *Ocotea* SPECIES**

Alfenas/MG

2024

ALBERT KATCHBORIAN NETO

**INTEGRATING METABOLOMICS AND COMPUTATIONAL TOOLS TO
DISCOVER ANTI-INFLAMMATORY BIOACTIVE AGENTS IN *Ocotea* SPECIES**

This thesis is part of the requirements for obtaining a PhD degree in Organic Chemistry from the Federal University of Alfenas - Minas Gerais.

Area of concentration: Organic Chemistry (Chemistry of Natural Products and Metabolomics)

Advisor: PhD. Daniela A. Chagas de Paula

Co-advisor: PhD. Ana Cláudia C. P. Ladvoat

Alfenas/MG

2024

Sistema de Bibliotecas da Universidade Federal de Alfenas
Biblioteca Central

Katchborian, Albert Neto.

Integrating metabolomics and computational tools to discover anti-inflammatory bioactive agents in *Ocotea* species / Albert Neto Katchborian. - Alfenas, MG, 2024.
276 f. : il. -

Orientador(a): Daniela Aparecida Chagas de Paula.

Tese (Doutorado em Química) - Universidade Federal de Alfenas, Alfenas, MG, 2024.

Bibliografia.

1. Metabolômica. 2. Espectrometria de massas. 3. *Ocotea*. 4. Anti-inflamatório. 5. Produtos Naturais. I. Chagas de Paula, Daniela Aparecida, orient. II. Título.

Ficha gerada automaticamente com dados fornecidos pelo autor.

O presente trabalho foi realizado com apoio financeiro da “Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG)” e a da “Pró-reitoria de Pesquisa e Pós-graduação da UNIFAL-MG.

ALBERT KATCHBORIAN NETO

INTEGRATING METABOLOMICS AND COMPUTATIONAL TOOLS TO DISCOVER ANTI-INFLAMMATORY BIOACTIVE AGENTS IN *Ocotea* SPECIES

The President of the examining board below signs the approval of the Thesis presented as part of the requirements for obtaining the title of Doctor in Organic Chemistry from the Federal University of Alfenas. Concentration area: Chemistry of Natural Products

Aprovada em: 04 de Outubro de 2024

Prof. Dr. Alan César Pilon
Universidade Estadual Paulista – UNESP-SP

Profa. Dra. Denise Brentan da Silva
Universidade Federal do Mato Grosso do Sul – UFMT-MT

Prof. Dr. Fernando Batista da Costa
Universidade de São Paulo – USP/RP-SP

Prof. Dr. Hector Ferreira Koolen
Universidade do Estado do Amazonas – UEA-AM

Profa. Dra. Fernanda Maria Pinto Vilela
Universidade Federal de Juiz de Fora – UFJF-MG

Eu, Profa. Dra. Daniela Aparecida Chagas de Paula, Presidente da Comissão Examinadora, declaro que ata foi lida e aprovada por todos os membros da comissão examinadora, durante a sessão de defesa, e não havendo mais nada a registrar, assino a presente ata.



Documento assinado eletronicamente por **Daniela Aparecida Chagas de Paula, Professor do Magistério Superior**, em 07/10/2024, às 10:54, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.unifal-mg.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1356835** e o código CRC **0868C22F**.

ACKNOWLEDGEMENT

First and foremost, I want to express my deepest gratitude to what we might call God, whose manifestation throughout the energy in the universe has placed me in a position of privilege, allowing me to embark on this academic journey with plenitude.

To my parents, Francisca and João Katchborian, words cannot fully capture my gratitude. Your unconditional love, care, and sacrifices have been the foundation that allowed me to stand where I am today. Without your unwavering support, none of this would have been possible. And without you, none of that would have made sense. I also dedicate this work to the memory of my grandfather, whose strength and spirit live on in me. Though he is no longer with us, his legacy continues to push me forward, reminding me to persist even when the path is difficult. I know you will be proud whenever you are, we all will be together in our minds.

To the professors of the LFQMM, Dr. Marisi Gomes Soares and Dr. Danielle Ferreira Dias, thank you for being more than just mentors. Your guidance, help, knowledge, and friendship have shaped me and helped me to evolve both as a researcher and as a person. Thanks for allowing me to be part of this amazing family. I am also deeply grateful to Dr. Paula Bueno for all your support, transmitted knowledge, your attention, and for being such a true scientific inspiration for me in the last two years. Also, to all my colleagues at LFQMM for being there whenever I needed help. Special thanks to Dr. Karen Nicácio (UFMT), Dr. Mario Ferreira (UFES) and Dr. Tiago Branquinho (UFS) for the knowledge shared along this journey and for the collaborations that have enriched my scientific path.

To the Federal University of Alfenas (UNIFAL), thank you for nurturing me as a researcher, and to all the professors of the PPGQ, I am grateful for the knowledge and wisdom you have imparted. I also want to extend my heartfelt thanks to the dedicated employees of UNIFAL who work behind the scenes, ensuring that our spaces are clean, comfortable, and prepared to push the boundaries of science every day. A special thanks to the funding agencies—CAPES, CNPq, FAPEMIG, and FINEP—for their financial support. More specifically, the FAPEMIG APQ APQ02882-24, APQ-05218-23, APQ-00544-23 e BPD-00760-22 and the CNPQ 408115/2023-8,

316204/2021-8 and 406837/2021-0. In addition, my deepest gratitude to the CESJ Herbarium at the Federal University of Juiz de Fora (UFJF) and the OUPR Herbarium at the Federal University of Ouro Preto (UFOP), for providing access to the *Ocotea* species that were so crucial to my research.

I owe a great debt of gratitude to Dr. Michael Murgu from Waters Corporation for your incredible assistance with the UPLC/MS experiments as well Dr. João Lago from the Federal University of ABC (UFABC) for the incredible support with NMR experiments. These collaborations were fundamental to the success of this metabolomics research. To Dr. Ivo Caldas, thank you for your unwavering availability and support whenever needed. To my co-advisor, Dr. Ana Chagas, thank you for standing by me through every step of this journey, from my first publication to my PhD qualification stage. Your support, availability, and wisdom have been pivotal, and I am deeply grateful for your care and guidance.

A very special thanks to Professor RuAngelie from the University of Strathclyde – Glasgow / UK. Your kindness, brilliance, and beautiful passion for science have left an indelible mark on me. You've shown me what it means to pursue proper knowledge with heart and soul and of course showing that also comes with big responsibilities. The experiences I shared with you and my colleagues at SIBPS during this first semester of 2024 continue to inspire me every day to push further, to think bigger, and to follow my curiosity through science. Special thanks to Garry and Riggs for their kindness and friendship in the data processing workstation during my time as a visiting researcher at SIBPS. My wonderful flat mates Mary and Jobin for helping me live in the UK in the nicest way possible. I also need to thank Marion Bowles for her company and for such amazing shared moments, and of course my best-ever Scottish friend in the world Michael for all the Rock n' roll nights at Keg in Paisley town.

To my friends, PhD candidate Luiz Paulo Belchior, Dr. Miller Ferreira (UNIFAL), MS.c Gabriel Viana, and MSc. candidate Matheus Fernandes Alves, our bond goes beyond the walls of the lab, and I'm thankful for the laughter, the deep conversations, and the shared empathy we've cultivated over these years, along with all my other friends from LFQMM. Special thanks to Dr. Jonas Cruz (USP-RP), not only for your long friendship but for your indispensable support in different aspects of my research, as well as Dr.

Fausto Carnevale for all the gas-phase fragmentation reactions help, for the friendship and tips in the last part of this PhD journey. I am also thankful to the friends outside the laboratory and the 'academic world' who have been vital to my growth, not academically but as a human being, allowing me to fully live true natural human experiences. Citing names here would be easily unfair.

Finally, and most importantly, my deepest thanks to my advisor, Dr. Daniela Aparecida Chagas de Paula. Dani, you have been my guiding light through this entire process. Over the course of these seven years, from scientific initiation to PhD, you have been more than just an advisor. You have been a mentor, a friend, and an inspiration. You've taught me not only how to think critically and research with rigour but also how to live with passion, enthusiasm and happiness. Thank you for your captivating laughs, endless patience, and countless support, and for showing me that it is possible to build a future doing what you truly love. I've admired and learned from you since the first day you accepted me to be your student. I really do appreciate all the time you have given to improve my weaknesses and strengthen my qualities. I couldn't ask for better than that. Thank you, Dani! You are the best! To everyone mentioned, and to those whose names may not appear here but whose contributions were equally important—thank you, from the depths of my heart, for making these PhD years such an incredible journey to experience.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

"Life might just be a cosmic giggle, and a breath in the universe."

Kristle Cole, 2011

ABSTRACT

The quest for innovative anti-inflammatory agents with reduced side effects is a compelling area of interest in pharmaceutical research. The inhibition of prostaglandin E2 (PGE2) and leukotriene (LTB4) releasing are significant as they are key mediators in different inflammatory diseases. Natural Products (NP) offer a rich source of structurally diverse and functionally distinct specialised metabolites that can assist in the discovery of novel therapeutic agents. In this context, our research explored the *ex vivo* anti-inflammatory potential of 60 plant extracts from the *Ocotea* genus, most studied for the first time. Our objectives consist of applying metabolomics and computational tools to annotate biomarkers associated with anti-inflammatory activity. Machine learning models were also built to predict novel anti-inflammatory *Ocotea* extracts. The chemical composition of various endemic and threatened *Ocotea* species from different biomes in Brazil were analysed, leading to the annotation of multiple classes of specialized metabolites. Chapter I provides a classic review of all the identified metabolites within the *Ocotea* genus, culminating in the creation of a comprehensive *in-house* database named *OcoteaDB*. This literature review also explores the relevant biosynthetic pathways of the bioactive chemical scaffolds found in the genus. Chapter II is a research article of a metabolomics study that revealed the PGE2 inhibition release of *Ocotea* spp. can be majorly attributed to aporphine alkaloids. By employing ultra-performance liquid chromatography coupled with high-resolution mass spectrometry (UPLC-MS) combined with *OcoteaDB*, we achieved a rapid and reliable annotation process. Chapter III is a research article on the dual COX/LOX inhibitory biomarkers in promising *Ocotea* species using a concatenated UPLC-MS - Nuclear Magnetic Resonance (NMR) metabolomics approach and machine learning prediction models. This time, alkaloids, a glycosylated flavonoid and a sesquiterpenoid were correlated with the potential dual anti-inflammatory activity. Chapter IV solidifies the significance of the *Ocotea* genus as a producer of different classes of alkaloids, lignoids and glycosylated flavonoids using gas-phase fragmentation reactions and molecular networking in the annotation procedure. A UPLC-MS-DIA (*data-independent acquisition*) open pipeline was developed by integrating data conversion, processing, and metabolomics tools analysis. This thesis provided valuable insights into the metabolome and anti-inflammatory profile of several

unstudied *Ocotea* species, as well as a pipeline for future NP research computational analyses. The scientific knowledge generated can be applied in the search for new anti-inflammatory compounds and also contributes to highlighting the importance of conserving *Ocotea* species in Brazil, encouraging their preservation for future research.

Keywords: Metabolomics, Mass spectrometry, *Ocotea*, Anti-inflammatory, Natural Products, Chemometrics

RESUMO

A busca por agentes anti-inflamatórios inovadores com menos efeitos colaterais é uma área de grande interesse na pesquisa farmacêutica. A inibição da liberação de prostaglandina E2 (PGE2) e leucotrieno B4 (LTB4) é relevante pois estes são mediadores chave em diversas doenças inflamatórias. Produtos Naturais (PN) oferecem uma rica fonte de metabólitos especializados, estruturalmente diversos e funcionalmente distintos, que podem auxiliar na descoberta de novos agentes terapêuticos. Neste contexto, nossa pesquisa explorou o potencial anti-inflamatório *ex vivo* de 60 extratos de plantas do gênero *Ocotea*, muitas estudados pela primeira vez. O objetivo deste projeto consiste em aplicar ferramentas computacionais de metabolômica para anotar biomarcadores associados à atividade anti-inflamatória e modelos de aprendizado de máquina para prever novos extratos de *Ocotea* ativos. A composição química de várias espécies endêmicas e ameaçadas de *Ocotea* de diferentes biomas no Brasil foram analisadas, levando à anotação de múltiplas classes de metabólitos especializados. O Capítulo I apresenta uma revisão clássica de todos os metabólitos identificados dentro do gênero *Ocotea*, culminando na criação de um banco de dados abrangente chamado *OcoteaDB*. Esta revisão da literatura também explora as vias biossintéticas relevantes dos esqueletos químicos bioativos no gênero. O Capítulo II é um artigo de pesquisa de um estudo metabolômico que revelou que a inibição da liberação de PGE2 em espécies de *Ocotea* pode ser atribuída majoritariamente a alcaloides do tipo aporfina. Ao empregar cromatografia líquida de ultra performance acoplada à espectrometria de massas de alta resolução (CLAE-EM) combinada com o *OcoteaDB*, alcançamos um processo de anotação rápido e confiável. O Capítulo III é um artigo de pesquisa sobre biomarcadores inibidores duais das vias COX/LOX em espécies promissoras de *Ocotea*, utilizando uma abordagem metabolômica concatenada CLAE-EM - Ressonância Magnética Nuclear (RMN) e modelos de predição por aprendizado de máquina. Desta vez, os alcaloides, seguidos por um flavonoide glicosilado e um sesquiterpenoide foram correlacionados com a atividade anti-inflamatória dual. O Capítulo IV solidifica a importância do gênero *Ocotea* como um produtor de diferentes classes de alcaloides, lignoides e flavonoides glicosilados, utilizando reações de fragmentação em fase gasosa e redes moleculares no processo de anotação. Foi criado um pipeline aberto CLAE-EM-DIA (*aquisição de dados independente*) pela integração da conversão de dados, processamento e análise usando ferramentas metabolômicas. Esta tese gerou conhecimento sobre o metaboloma e o perfil anti-inflamatório de várias espécies de *Ocotea* ainda não estudadas, e um fluxo de

trabalho para futuras análises computacionais de PN. O conhecimento científico gerado pode ser aplicado na busca de novas substâncias ativas na inflamação e também contribuir para destacar a relevância da conservação de plantas do gênero *Ocotea* no Brasil, estimulando a preservação de suas espécies para futuras pesquisas.

Palavras-chave: Metabolômica, Espectrometria de massas, *Ocotea*, Anti-inflamatório, Produtos Naturais, Quimiometria.

FIGURE AND TABLE LIST

Figure 1 - Chemical structure of popular currently approved anti-inflammatory drugs.....	21
Figure 2 - Schematic representation of the connection between the lymphangiogenic growth factors (VEGF-C and VEGF-D), PGE2 pathway and the cancer metastasis through lymphatic vessels regulated by enzyme 15-hydroxyprostaglandin dehydrogenase (pgdh) inactivation and COX-2 overexpression.....	24
Figure 3 - Chemical structure of popular NP approved drugs.....	26
Figure 4 - The map represents the distribution of species of the <i>Ocotea</i> genus worldwide. The yellow dots show that the <i>Ocotea</i> species are attending mainly the American continent, and also present in the African continent, in areas nearest to Madagascar Island...	29
Figure 5 - Anti-inflammatory and cytotoxic NP isolated from <i>Ocotea</i> genus..	31
Table 1 - Other scientific collaborations	254

ABBREVIATIONS LIST

AA	- Arachidonic Acid
COX	- Cyclooxygenase
CYP80G2	- Cytochrome P-450 80G2
DEX	- Dexamethasone
DDA	- Data Dependent Acquisition
DIA	- Data Independent Acquisition
DNP	- Dictionary of Natural Products
EET	- Epoxyeicosatrienoic acids
EP4	- Prostaglandin E2 receptor 4
ESI	- Electrospray Ionisation
GC-MS	- Gas Chromatography – Mass Spectrometry
GNPS	- Global Natural Product Social Molecular Networking
MF	- Molecular Formula
MS	- Mass Spectrometry
NP	- Natural Products
HCA	- Hierarchical Cluster Analysis
HPLC	- High-Performance Liquid Chromatography
IL-1 β	- Interleukin 1 β
IND	- Indomethacin
LC-HRMS	- Liquid Chromatography hyphenated with High-Resolution Mass Spectrometry

LT	- Leukotriene
LFQMM	- <i>Laboratório de Fitoquímica, Química Medicinal e Metabolômica</i>
LPS	- Lipopolysaccharides
LOX	- Lipoxygenase
LTB4	- Leukotriene B4
LX	- Lipoxin
ML	- Machine Learning
MN	- Molecular Networking
MSA	- Multivariate Statistical Analyses
MS ^E	- Mass Spectrometry Fragment ion spectra - Data-independent analysis
mPGES	- prostaglandin E synthase-1
NMR	- Nuclear Magnetic Resonance
NO	- Nitric oxide
NP	- Natural Products
NSAIDs	- Non-Steroidal Anti-Inflammatory Drugs
OPLS-DA	- Orthogonal Partial Least Squares Regression – Discriminant Analyses
PCA	- Principal Component Analysis
PC	- Prostacyclin

PG	- Prostaglandin
PGE2	- Prostaglandin E2
PLS-DA	- Partial Least Squares Regression – Discriminant Analyses
QC	- Quality Control
QTOF	- Quadrupole-Time of Flight
SAIDs	- Steroidal Anti-Inflammatory Drugs
TX	- Thromboxanes
TME	- Tumor microenvironment
UHPLC	- Ultra-High-Performance Liquid Chromatography
VEGF-C	- Vascular Endothelial Growth Factor C

SUMMARY

1	INTRODUCTION.....	17
1.1	Inflammation.....	17
1.2	Natural Products.....	23
1.3	Lauraceae family and <i>Ocotea</i> genus.....	26
1.4	Metabolomics and computational analysis.....	30
2	OBJECTIVES.....	37
2.1	General objective.....	37
2.2	Specified objectives.....	37
3	CHAPTER I - (THEORETICAL REFERENCE) – BIOSYNTHETIC ASPECTS OF <i>Ocotea</i> METABOLOME.....	38
3.1	Review article: The metabolome of <i>Ocotea</i> genus (Lauraceae): from biosynthetic aspects to the widespread bioactive chemical scaffolds.....	(1-29)
4	CHAPTER II – PART A: ANTI-INFLAMMATORY METABOLOMICS STUDY OF <i>Ocotea</i> SPECIES.....	68
4.1	Research article: Bioprospecting-based untargeted metabolomics identifies alkaloids as potential anti-inflammatory bioactive markers of <i>Ocotea</i> species (Lauraceae).....	(1-28)
5	CHAPTER III – PART B: ANTI-INFLAMMATORY METABOLOMICS STUDY OF <i>Ocotea</i> SPECIES.....	97
5.1	Research article: Discovery of dual COX/LOX inhibitory bioactive markers in <i>Ocotea</i> (Lauraceae) using concatenated UPLC/MS-NMR metabolomics and machine learning.....	(1-75)
6	CHAPTER IV – DIA-MS DATA PROCESSING AND MOLECULAR NETWORKING PIPELINE.....	173
6.1	Research article: Integrative open workflow for confident annotation and molecular networking of metabolomics MS ^E /DIA data.....	(1-79)
7	CHAPTER V – OTHER SCIENTIFIC COLLABORATIONS.....	253
8	FINAL REMARKS.....	263
	REFERENCES.....	266

1 INTRODUCTION

1.1 Inflammation

Inflammation is a crucial defence mechanism of the immune system, triggered by various factors in response to harmful stimuli such as fungi, viral or bacterial infections, tissue damage, or the development of different diseases. Inflammatory pathways are essential for protecting the body and initiating healing processes. During the initial inflammatory phase, the body works to eliminate the cause of cell injury, remove damaged cells, and restore tissue homeostasis. However, when inflammation is exacerbated, prolonged, or chronic and becomes associated with pathological conditions, it can lead to significant tissue damage and contribute to the progression of a range of diseases such as psoriasis, gouty arthritis, diabetes, atherosclerosis, trauma, ischemia, cancer, and autoimmune diseases (Chen *et al.*, 2018; Landskron *et al.*, 2014; Meirer; Steinhilber; Proschak, 2014).

The inflammatory response can manifest acutely or chronically and may occur either systemically or locally, depending on the underlying pathology (Landskron *et al.*, 2014; Multhoff; Molls; Radons, 2012). In response to tissue injury, the body initiates chemical signalling cascades that stimulate inflammation at the affected sites. A key process in this response is the activation of arachidonic acid (AA) metabolism, which plays a crucial role in the progression of inflammation. This process involves the release of key chemical mediators, such as prostaglandins (PGs) and leukotrienes (LTs), which not only regulate vascular permeability but also control the recruitment of leukocytes during inflammation. Additionally, the release of various growth factors, chemokines, and cytokines further contributes to the regulation and progression of the inflammatory process (Chen *et al.*, 2018; Coussens; Werb, 2012; Fiorucci *et al.*, 2001; Landskron *et al.*, 2014).

The AA is a 20-carbon polyunsaturated fatty acid that is released from membrane phospholipids through physiological stimulation by phospholipases enzymes, predominantly by the type-IV cytosolic PLA2- α , after the initial damage to the body tissues (Broughton; Janis; Attinger, 2006; Meirer; Steinhilber; Proschak, 2014). After AA is released, it can undergo metabolism through various enzymatic pathways. One such pathway involves the action of cyclooxygenase (COX), which is responsible for the production of PGs and thromboxanes (TXs). Another pathway involves lipoxygenase (LOX), which produces LTs, and lipoxins (LXs). Additionally,

cytochrome P450 enzymes (CYP) play a crucial role in the production of epoxyeicosatrienoic acids (EETs). These enzymatic pathways contribute to the diverse array of lipid mediators involved in physiological and pathological processes (Coussens; Werb, 2012; Meirer; Steinhilber; Proschak, 2014).

These chemical inflammatory mediators produced, such as PGs, TXs and LTs are substances that act in an autocrine and paracrine manner on target cells after their extracellular release. The prostaglandin E2 (PGE2), is one of the key final mediators derived from the AA inflammatory cascade, mediated mainly by COX enzymes and also terminal synthases in the pathway such as microsomal prostaglandin E synthase-1 (mPGES) (Chagas-paula *et al.*, 2015a; Ding *et al.*, 2018; Fiorucci *et al.*, 2001; Mahesh; Kumar; Reddanna, 2021). Targeting the COX pathway has been a central focus of pharmaceutical interventions to effectively reduce acute inflammatory responses and mitigate inflammation-related diseases. Inhibiting PGE2 is particularly important due to its central role in driving the hallmark symptoms of inflammation—pain, redness, heat, swelling, and the resulting loss of cell function (Funk, 2001; Meirer; Steinhilber; Proschak, 2014).

In this context, steroidal anti-inflammatory drugs (SAIDs), such as glucocorticoids, e.g. dexamethasone (Figure 1), are well-known for their ability to inhibit phospholipase A2 (PLA2 α) at the very beginning of the inflammatory cascade, as well as nuclear factor- κ B (NF- κ B), a family of inducible transcription factors that regulate a wide array of genes involved in immune and inflammatory responses (Hwang *et al.*, 2013). This inhibition by SAIDs leads to a broad reduction in the production of inflammatory mediators including a decrease in PGE2, though it occurs earlier in the inflammatory pathway, and not directly via COX enzyme inhibition. Instead, non-steroidal anti-inflammatory drugs (NSAIDs), which are among the most commonly used anti-inflammatory drugs on the market, act as selective or non-selective inhibitors of the COX-1 and COX-2 isoforms, important examples include acetylsalicylic acid (aspirin), ibuprofen, diclofenac, naproxen, indomethacin and others (Figure 1). By directly inhibiting these enzymes, NSAIDs reduce the secretion of PGs, reducing inflammation, alleviating pain, and lowering fever, particularly through the more selective reduction of PGE2 levels at sites of inflammation, making them indispensable treatments for a variety of inflammatory conditions (Leslie, 2015; Meirer; Steinhilber; Proschak, 2014).

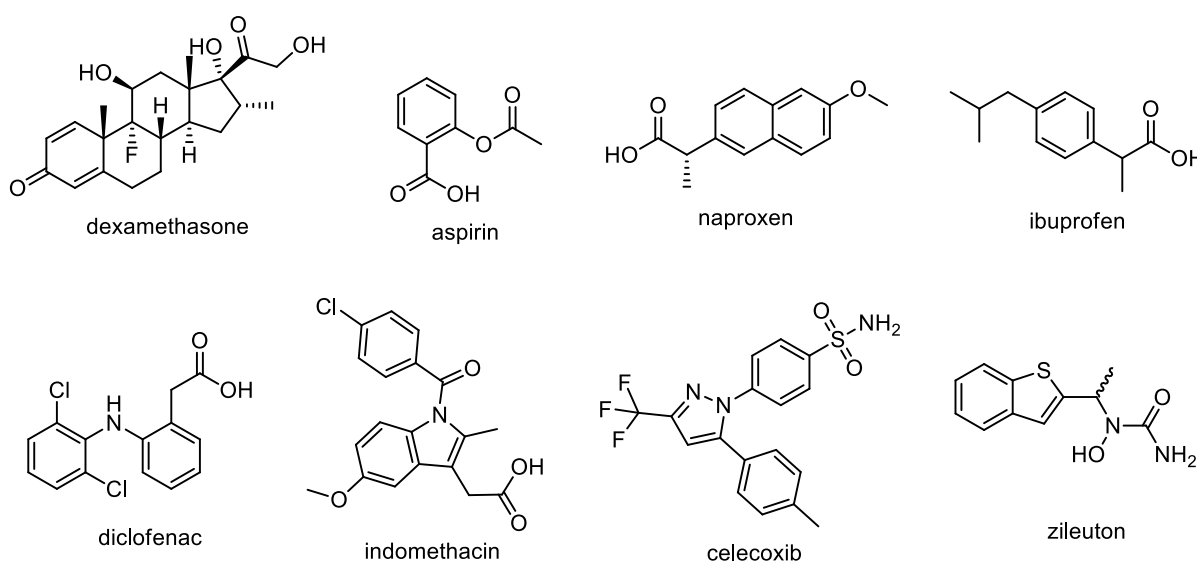
Similarly, leukotriene B₄ (LTB₄) is a potent lipid mediator involved in inflammatory responses, particularly known for its role in recruiting and activating leukocytes, chemokinetic and chemotactic responses which amplifies inflammation and contributes to tissue damage (Rang *et al.*, 2011). Unlike NSAIDs that target COX enzymes, the inhibition of LOX enzymes, particularly 5-LOX, such as with zileuton (Figure 1), is essential for reducing LTB₄ synthesis. By lowering LTB₄ levels, LOX inhibitors can potentially reduce the severity of inflammation by limiting leukocyte infiltration and activation. This targeted approach is especially effective in managing both chronic and acute inflammatory conditions and is primarily used in the treatment of asthma, where it helps reduce bronchoconstriction, inflammation, and mucus production associated with leukotrienes, which play a significant role in respiratory inflammatory responses (Leslie, 2015; Meirer; Steinhilber; Proschak, 2014; Rang *et al.*, 2011).

Moreover, it is recognized that anti-inflammatory drugs are among the most used drugs worldwide. However, the effectiveness of currently available drugs can significantly decrease when used in long-term treatments or chronic situations, leading to a lower effectiveness-to-side-effect ratio and a higher incidence of adverse reactions (Dugowson; Gnanashanmugam, 2006; Hwang *et al.*, 2013). Anti-inflammatory drugs on the market can induce adverse drug reactions, often due to a lack of specificity to inhibit PGE₂ release or subsequent imbalances in the AA metabolic pathways. These side effects are particularly problematic for individuals requiring long-term therapy, where millions of people suffer from pain, and the prolonged use of NSAIDs has become a common health concern (Ghosh; Alajbegovic; Gomes, 2015; Hwang *et al.*, 2013).

While SAIDs are known for their potency, they can cause a range of mild to severe side effects, especially when used at higher doses or for extended periods. These side effects include high blood pressure, weight gain, fluid retention, mood swings, and detrimental effects on bone and eye health. In contrast, NSAIDs are generally safer for short-term use and effective at reducing pain and inflammation. However, in long-term, they affect the renal and cardiovascular systems and the gastrointestinal tract, where they can cause ulcers and internal bleeding, especially when taken orally (Meirer; Steinhilber; Proschak, 2014; Yang *et al.*, 2007). More recently, COX-2 selective inhibitors, a subclass of NSAIDs such as celecoxib (Figure 1), were developed to selectively inhibit the COX-2 enzyme while sparing the

constitutive and more physiological COX-1, thereby reducing the risk of gastrointestinal side effects. However, their use has been associated with an increased risk of cardiovascular events and stroke, particularly when used at doses higher than recommended, leading to greater caution in their prescription and use (Hwang *et al.*, 2013; Mahesh; Kumar; Reddanna, 2021).

Figure 1 - Chemical structure of popular currently approved anti-inflammatory drugs.



Source: From author (2024)

Additionally, a common side effect of NSAIDs is their potential to unbalance the AA pathway, as the sole inhibition of COXs may lead to a diversion of the AA metabolism towards the LTs pathway, leading to an increased LT levels production. This overproduction can result in adverse effects, particularly for individuals with respiratory conditions, where it may exacerbate bronchoconstriction, or for those with gastric mucosa damage, worsening their condition. These effects could potentially be mitigated by administering 5-LOX inhibitors or LT antagonists. Such risks underscore the importance of careful use of these drugs in susceptible populations (Ghosh; Alajbegovic; Gomes, 2015; Rang *et al.*, 2011).

In this context, many currently available anti-inflammatory drugs can trigger mild to serious adverse reactions. These drawbacks highlight the urgent need for innovative therapies. The enzyme mPGES-1 has attracted considerable attention because it is the final step in the inflammatory cascade, offering a more targeted approach to PGE2 inhibition with potentially fewer adverse effects (Ding *et al.*, 2018; Meirer; Steinhilber;

Proschak, 2014). Also, dual COX/LOX inhibitors show promise as a novel strategy for the next generation of anti-inflammatory drugs. By simultaneously targeting both pathways, dual inhibitors might induce an enhanced anti-inflammatory effect without damaging the gastrointestinal mucosa. Therefore, the ongoing search for new, viable anti-inflammatory agents remains crucial, intending to develop substances that offer enhanced efficacy and reduced side effects (Chagas-paula *et al.*, 2015a; Cui *et al.*, 2018; Rudrapal *et al.*, 2023).

In addition to relieving pain, anti-inflammatory drugs, particularly NSAIDs, are effective anticancer agents due to their ability to reduce inflammation, a process that promotes tumor development. Inflammation attracts immune and inflammatory cells to the tumor microenvironment (TME), contributing to metastasis. The overexpression of COX enzymes leads to an accumulation of PGE2 in the TME, which is closely linked to cancer progression, especially in breast cancer. PGE2 plays a crucial role in promoting cancer cell growth and survival by modulating the TME, and also by stimulating pro-inflammatory cytokines and chemokines that recruit more immune cells, creating a pro-tumor environment. This supports cancer cell proliferation, migration, invasion, and inhibits apoptosis (Mahesh; Kumar; Reddanna, 2021; Nandi *et al.*, 2017; Timoshenko *et al.*, 2003; Yuan *et al.*, 2012).

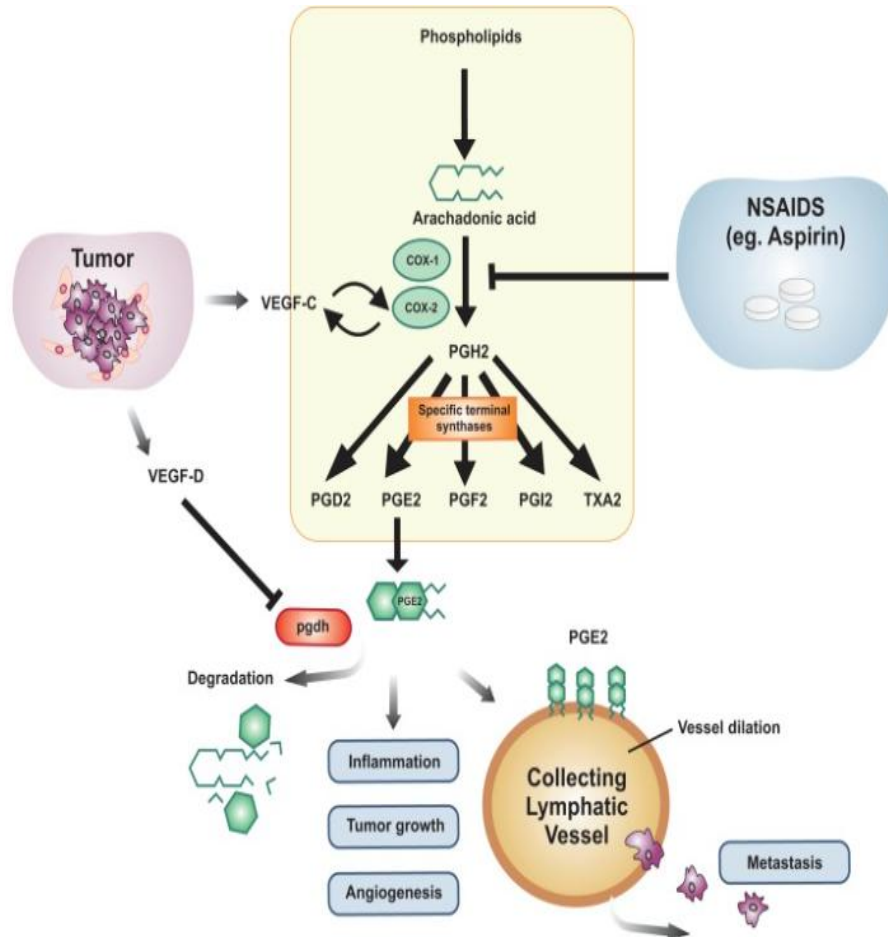
Cancer and inflammation are interconnected at various levels, with PGE2 inhibition emerging as a key therapeutic target, particularly when combined with chemotherapy. PGE2 plays a well-documented role in promoting lymphangiogenesis, a process crucial for cancer metastasis via lymphatic vessels. This is particularly evident in breast cancer, where COX-2 overexpression leads to elevated PGE2 levels. The presence of PGE2 in the TME, driven by increased COX-2 expression, contributes to cancer progression through mechanisms such as lymphangiogenesis (Karnezis *et al.*, 2012; Paduch, 2016). This negatively impacts patient health and it is a frequent manifestation in almost all epithelial cancers, including lungs, stomach, colon, breast, pharynx and larynx, uterine cervix, prostate and the ovary (Nandi *et al.*, 2017; Paduch, 2016).

The inflammatory conditions are responsible for one of the first cancer routes of spread, via bloodstream, and then later metastasize from the lymph nodes to other organs. The newly formed lymphatic capillaries work as conduits for the entrance and spread of cancer cells to lymph nodes (Karnezis *et al.*, 2012; Nandi *et al.*, 2017; Paduch, 2016). Depending on the tumour, metastasis already occur in the existing

lymphatic vessels, and lymphangiogenesis is triggered as a late process (Nandi *et al.*, 2017; Paduch, 2016). Thus, targeting inflammation, including molecules like PGE₂, could be an effective cancer treatment strategy, as PGE₂ is produced by various immune and cancer cells, and plays a crucial role in cancer pathology (Chini *et al.*, 2020; Mantovani, 2018). Studies show that COX-2 inhibitors can reduce tumour growth by inhibiting tumour-associated angiogenesis and lymphangiogenesis (Karnezis *et al.*, 2012; Lyons *et al.*, 2014; Wang; Honn; Nie, 2007).

Recent research unfolded the mechanism revealing the upregulation of lymphangiogenic growth factors VEGF-C or VEGF-D in tumour cells (Nandi *et al.*, 2017; Rozic; Chakraborty; Lala, 2001; Xu *et al.*, 2019). Physiologically, PGE₂ is metabolized by the enzyme 15-hydroxyprostaglandin dehydrogenase (pgdh). However, primary tumors can secrete VEGF-D, which downregulates the expression of the pgdh gene, inhibiting the degradation of PGE₂ in the extracellular environment. This inhibition of PGE₂ degradation by VEGF-D leads to lymphatic vessel dilation, promoting an increase in metastases and tumor growth (Karnezis *et al.*, 2012; Nandi *et al.*, 2017). Also, VEGF-C or VEGF-D secreted by tumour cells can increase vascular permeability or have important effects on the pressure of the tumour interstitial fluid, which can promote the entry of tumour cells into the lymph. In addition, VEGF-C increases COX-2 expression, further elevating PGE₂ levels (Karnezis *et al.*, 2012; Nandi *et al.*, 2017). Consequently, either COX or VEGF-C/D inhibitors have shown promise in reducing PGE₂ in the TME and preventing metastases (Figure 2) (Karnezis *et al.*, 2012; Majumder *et al.*, 2014; Nandi *et al.*, 2017).

Figure 2 - Schematic representation of the connection between the lymphangiogenic growth factors (VEGF-C and VEGF-D), PGE2 level and the cancer metastasis through lymphatic vessels regulated by enzyme 15-hydroxyprostaglandin dehydrogenase (pgdh) inactivation and COX-2 overexpression.



Source: Karnezis *et al.* (2012)

As a result, it is crucial to study anti-inflammatory drugs for their significant role in both cancer treatment and management of a range of inflammatory related-diseases. These drugs are vital in reducing inflammation, a common underlying factor in several of these chronic conditions. Novel therapeutic strategies can help manage symptoms, slow disease progression, and improve patient outcomes. Therefore, studying new therapies can lead to advancements in treating a wide spectrum of diseases where inflammation plays a critical role.

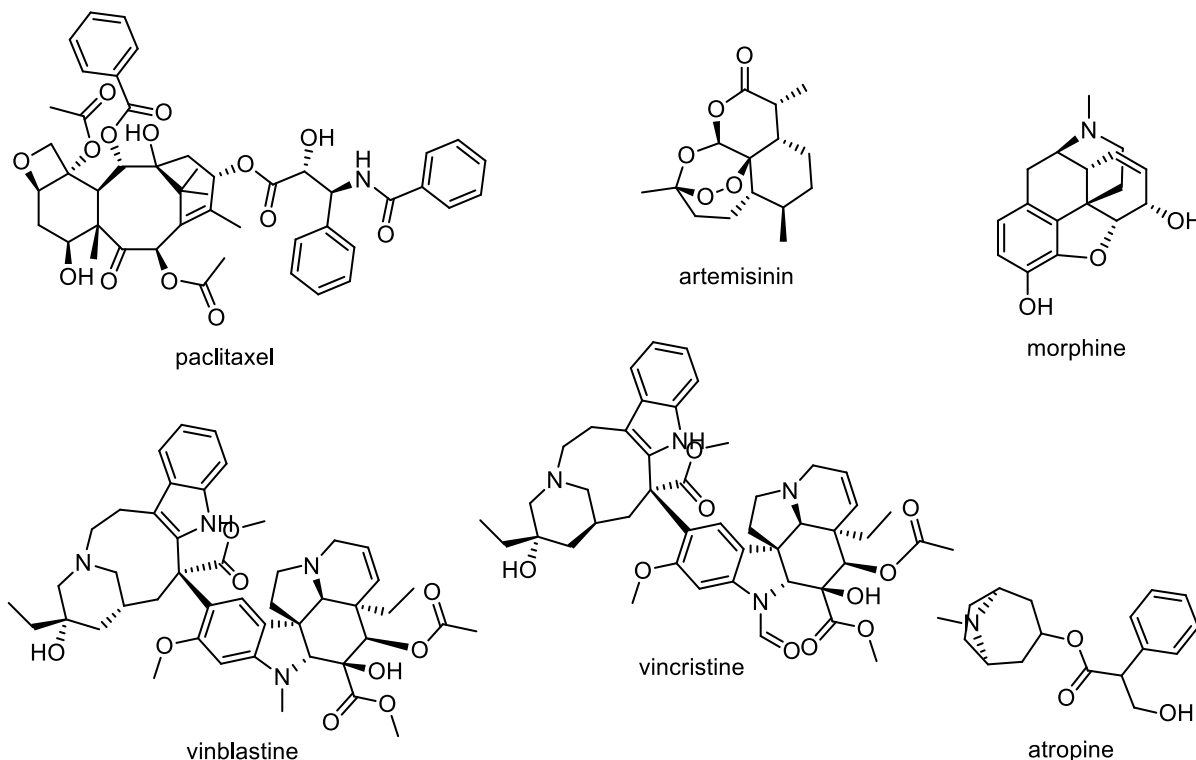
1.2 Natural Products

Natural products (NP) are chemical compounds of living organisms such as plants, animals, fungi, bacteria, and marine organisms, encompassing a vast array of

specialized metabolites with complex biosynthetic pathways, wide bioactivity profiles, and unique chemical structures (Dewick, 2009; Harvey; Edrada-ebel; Quinn, 2015). These compounds have played a pivotal role in the development of new therapies and drugs, offering a rich source of structurally diverse and functionally distinct specialised metabolites that serve as leads for the discovery of novel therapeutic agents (Gaudêncio; Pereira, 2015). Ethnobotanical and ethnopharmacological studies are key to identifying species with medicinal potential, as they rely on the traditional knowledge of local communities regarding the use of plants and other natural resources in their popular medicine (Mangisa *et al.*, 2021; Verpoorte; Choi; Kim, 2005). Chemotaxonomy-based plant selection studies provide another valuable approach, allowing researchers to trace new species and bioactive compounds based on previous chemical knowledge of the genus or family. This method complements traditional ethnobotanical approaches by providing a scientific framework for identifying and prioritizing species for further study (Gottlieb, 1972; Zidorn, 2019).

The research in NP comprehends multiple stages, from the experimental design, and collection of biological samples to the extraction, isolation, and structural characterization of compounds, followed by the evaluation of their biological activities. The traditional approach to discovering bioactive constituents from complex NP mixtures is known as "bioassay-guided fractionation". This method involves simplifying and separating extracts using chromatographic techniques, followed by screening the resulting fractions for biological activity. This iterative process continues until a single bioactive compound is isolated and identified (Chan *et al.*, 2006; Zidorn, 2019). While bioassay-guided fractionation has been the gold standard in NP discovery, leading to the identification of significant compounds like paclitaxel, artemisinin, and vincristine/vinblastine, morphine and atropine (Figure 3), it is not without challenges. One common issue is the loss of biological activity during the process or the failure to isolate a single active component from a complex mixture. Additionally, this method tends to favour the most abundant, which often results in the re-isolation of already-known compounds. To address these limitations, modern NP approaches have integrated metabolomics and chemometrics tools, to access the metabolite profile of natural sources and investigate a higher number of known compounds in less amount of time (Demarque *et al.*, 2020; Harvey; Edrada-ebel; Quinn, 2015; Patil; Patil; Maheshwari, 2016).

Figure 3 - Chemical structure of popular NP approved drugs.



Source: From author (2024)

Moreover, NP chemistry is not only of pharmacological importance but also plays a fundamental role in understanding the biochemical and ecological processes of living organisms and has implications for the functioning of ecosystems. The structural diversity of natural compounds reflects the evolutionary adaptation of species to survive and interact with their environment (Dyer *et al.*, 2018; Schmidt *et al.*, 2019). Specialised metabolites, such as terpenoids, flavonoids, and alkaloids, are produced by organisms to fulfil specific ecological functions, including defence against predators, the attraction of pollinators, and competition with other species (Dyer *et al.*, 2018; Findlay, 2016). Therefore, the study of NP chemistry not only contributes to drug development but also provides valuable insights into the chemical ecology of organisms.

Specialised metabolites such as alkaloids, terpenoids, lignoids, polyphenols, and glycosides often possess intricate chemical structures, featuring multiple chiral centres, conjugated systems, and fused rings (Conceição *et al.*, 2020; Macedo *et al.*, 2020). The elucidation of these structures requires the combined use of advanced techniques such as nuclear magnetic resonance (NMR) spectroscopy, which provides detailed information on the chemical environment and connection of atoms, while liquid

chromatography coupled to high-resolution mass spectrometry (LC-HRMS) allows separation, precise determination of molecular masses and the identification of characteristic fragments. Both techniques have been extensively used in traditional and modern approaches in NP research (Alarcon-barrera *et al.*, 2022; Zhang *et al.*, 2012). These analytical techniques corroborate each other to identify the chemical structures of natural compounds, a potential drug prototype and thus enable the possibility of structural analogues development with potentially enhanced biological properties in the medicinal chemistry field. Thus, NP research enriches our understanding of nature's chemical diversity and significantly contributes to the continuous advancement of drug discovery (Gobbo-neto; Lopes, 2007; Harvey; Edrada-ebel; Quinn, 2015).

1.3 Lauraceae family and *Ocotea* genus

Following the FLORA BRASIL 2024, to date, 46.732 species belonging to the Plantae kingdom are officially recognized for the Brazilian flora. However, there is an estimate that Brazil shows the potential to possess around 400.000 plant species in the several phytogeographical domains of the country, which corresponds to the most extensive plant genetic diversity in the world (Silva *et al.*, 2012). Moreover, Brazil has a significant record of ethnopharmacology use from a vast number of plant species, which were transmitted by popular knowledge through generations. Brazil's total land is extensive filled with many biomes and typical climates, as a result of different ecosystems of its natural abundant resources. Thus, in each part of the country, there are different vegetal species, with unique characteristics (Breitbach *et al.*, 2013). The existing biomes in Brazil are the Amazon Forest, Cerrado, Pampas, Mata Atlântica, Caatinga, Pantanal, the transition bands, and the Brazilian marine biome (Ibama, 2017). Thus, Brazil is one of the richest sources of pharmacologically active compounds worldwide, and most of them arise from the specialized metabolism of plants (Padilla *et al.*, 2018).

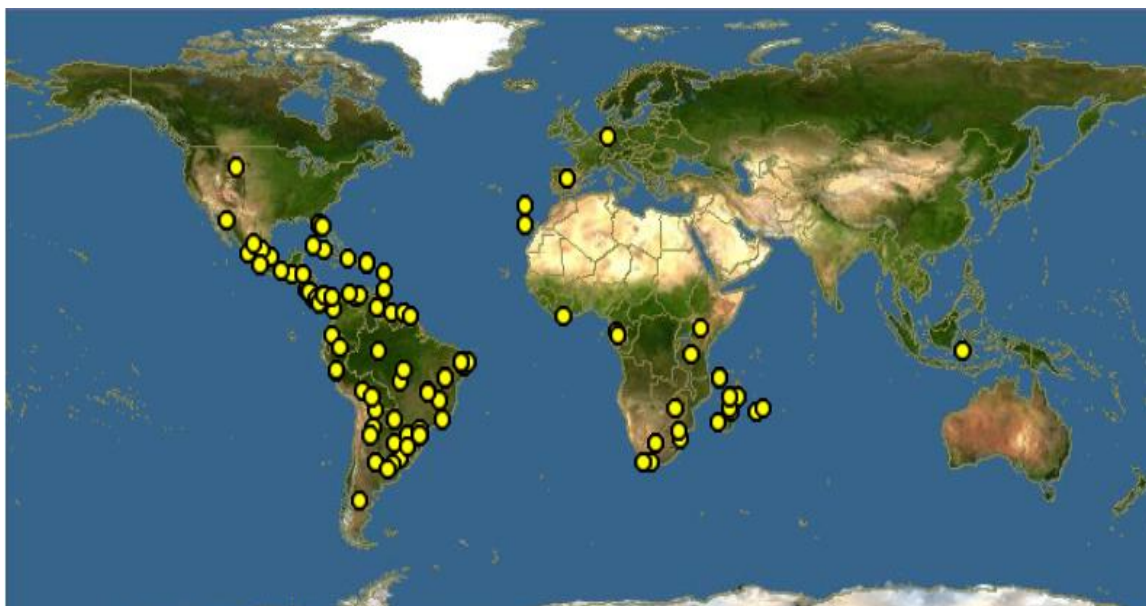
Regarding the botanical families of the Brazilian flora, the Lauraceae Jussieu is an important botanical family popularly called as laurel family. The Lauraceae has a natural distribution in the world's tropical and subtropical regions, found mainly in Southeast Africa and South America. This family gather more than 2850 species distributed along 67 genera (Veiga junior, 2014; Yamaguchi *et al.*, 2011). Antoine

Laurent de Jussieu named the Lauraceae in 1789 after years of botanical research. The Lauraceae belongs to the order of Laurales, which is considered one of the most primitive families of the Magnoliophyta division and to the Plantae kingdom, which is represented by the organisms named the flowering plants, scientifically known as the angiosperms (Chase *et al.*, 2016). Among the most known species in the family are the laurel (*Laurus nobilis* L.), the avocado (*Persea americana* Mill.), the cinnamon (*Cinnamomum verum* J. Presl.), and the camphor tree (*Cinnamomum camphor* Meisn.). The Lauraceae have a diversity of utility for humans, with broad economic input associated. Besides the use of these plant species in cooking, papermaking, carpentry and civil construction, several Lauraceae species are recognized in popular medicine. In addition, its essential oil is well explored by the pharmaceutical and cosmetic industry (Passos *et al.*, 2022).

Regarding Lauraceae, 27 native genera have been registered in Brazil, but not necessarily endemic to the country land, some also occur in other subtropical regions around the world. The Lauraceae genera found in Brazil are *Aiouea*, *Anaueria*, *Aniba*, *Beilschmiedia*, *Cassytha*, *Cinnamomum*, *Cryptocaria*, *Dicypellium*, *Endlicheria*, *Kubitzkia*, *Licaria*, *Mezilaurus*, *Misanteca*, *Nectandra*, *Ocotea*, *Paraia*, *Persea*, *Phoebe*, *Phyllostemonodaphne*, *Pleurothyrium*, *Rhodostemonodaphne*, *Sextonia*, *Systemonodaphne*, *Urbanodendron* and *Williamdendron*. Besides, the *Laurus* and *Litsea* genera are well established and grown in the country, although both were introduced to Brazil due to their economic importance (Chaverri; Cicció, 2005; Veiga Junior, 2014; Sacchetti *et al.*, 2006; Trofimov; De Moraes; Rohwer, 2019; Yamaguchi *et al.*, 2011). Particularly, the *Ocotea* genus is the largest genus of the Lauraceae family in the American continent and comprehends around 400 species widespread in world tropical and sub-tropical areas, such as Central and South America, Southern Africa and Madagascar. The *Ocotea* distribution worldwide is detailed in Figure 4.

There are 101 endemic *Ocotea* spp. in Brazil, with around 160 species in total. Most of these *Ocotea*, endemic or not, have never been chemically or biologically evaluated before. Several *Ocotea* plants species occur in the Atlantic Forest biome, and thus they cover diverse states of the country, especially in the south and the central east. The *Ocotea* spp. are mostly found in the states of Paraná, Santa Catarina, Minas Gerais, São Paulo, Rio de Janeiro, Mato Grosso/Mato Grosso do Sul, Bahia, and Alagoas (Yamaguchi *et al.*, 2011; Florabrasil, 2018).

Figure 4 - The map represents the distribution of species of the *Ocotea* genus worldwide. The yellow dots show that the *Ocotea* species are found mainly the American continent, and also present in the African continent, in areas nearest to Madagascar Island.



Source: The Polistes Corporation/ <https://www.discoverlife.org>. (2020)

The *Ocotea* genus arouses great economic interest in the logging industry, such as the *O. porosa* or popularly known 'imbuia', and the wood from other species, for example, the *O. puberula* and *O. bullata*. The genus is represented mainly by aromatic trees which are fundamentally essential oil producers, rich in camphor (Chaverri; Ciccío, 2005; Salleh; Ahmad, 2017). The essential oils of the popularly known medicinal plant sassafras (*O. odorifera*), and others produced by the *Ocotea* genus are used by the cosmetics and perfumery industry, e.g. *O. cymbarum*, *O. caudata*, *O. pretiosa*, *O. usambarensis*, *O. sassafras* and *O. brenessi*. Besides, *Ocotea* spp. is known to have the potential pharmacological activity for a diverse spectrum of diseases (Chaverri; Ciccío, 2005; Sacchetti *et al.*, 2006). In the popular medicine of native people from South America, there are several *Ocotea* species recognized for valued medicinal properties, such as analgesic, antioxidant and anti-inflammatory properties (Salleh; Ahmad, 2017).

For example, in Peru and Equator, there is a traditional use of the *O. quixos* as a cinnamon substitute to aromatise food. However, it has demonstrated local anaesthetic and anti-diarrheic properties (Salleh; Ahmad, 2017). Moreover, the essential oil of *O. quixos* showed a significant suppression effect on LPS-induced nitric

oxide (NO) release from macrophages at non-toxic concentrations and also inhibited LPS-induced PGE2 production in a significant manner. The western blotting analysis suggested *O. quixos* could suppress LPS-mediated iNOS and COX-2 elevation, and at higher concentrations revealed even better inhibition results. Also, the study showed that the inhibition ability of *O. quixos* was found to be comparable to that of curcumin, a well-known anti-inflammatory NP (Amilia destryana *et al.*, 2014).

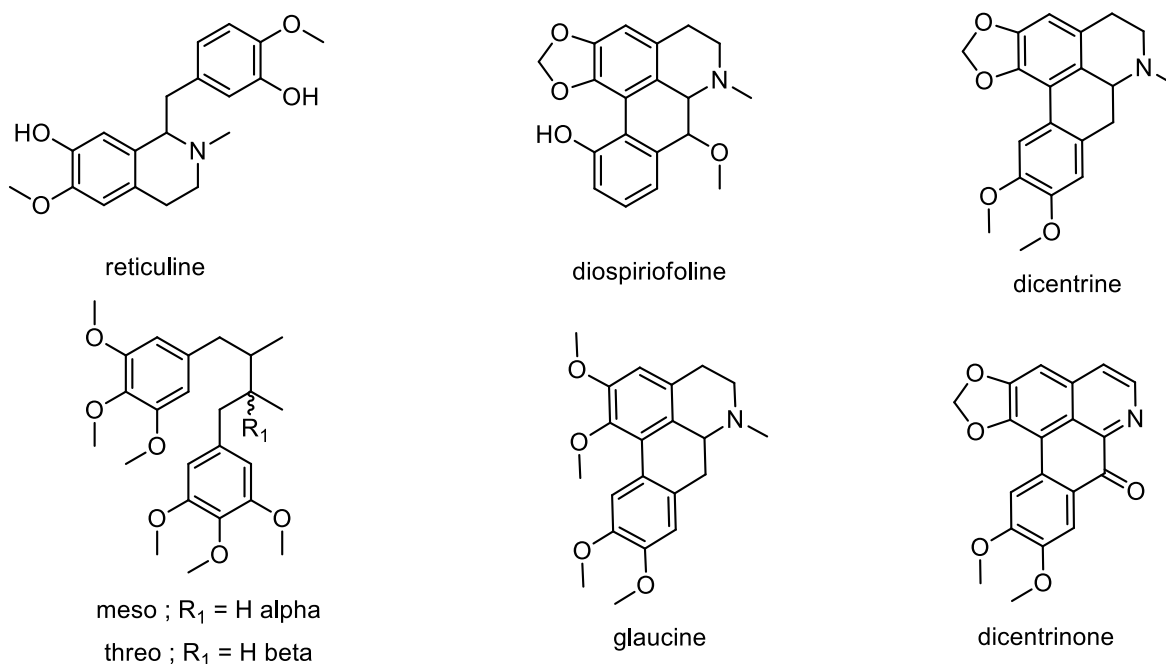
In Brazil, *O. lancifolia* is known for anti-rheumatic properties, and a series of new biologically active sesquiterpenes have been isolated together with aporphine alkaloids and other specialised metabolites (Palomino *et al.*, 1996; Raquel *et al.*, 2010). Also, *O. caparrapi* has demonstrated cytotoxic effects and commonly is used to heal bites, such as insects or even snakes; it also has been indicated for bronchitis (Palomino *et al.*, 1996). Likewise, *O. leucoxydon* has cytotoxic potential, as its crude extracts exhibit Topoisomerase I inhibition in anti-proliferative assays (Zhou *et al.*, 2000).

Recently, our research group have evidenced the anti-inflammatory properties of *O. odorifera* and *O. diospyrifolia* leaves, which demonstrated potent *in vivo* anti-inflammatory effects with dual inhibition of COX and LOX pathways. The process led to the isolation of a benzyloquinoline alkaloid, the reticuline, and also a new aporphine alkaloid named diospirofoline (Figure 5) (De alcântara *et al.*, 2021; Silva *et al.*, 2021). Moreover, diastereomeric lignans of *O. macrophylla* (Figure 5) were revealed to be a potent dual COX-2/5-LOX inhibitor as well as platelet-activating factor (PAF) antagonist in a screening of Lauraceae lignans (Coy; Cuca; Sefkow, 2009).

Chemically, the *Ocotea* genus was revealed to contain specialised metabolites from different chemical classes in different plant parts such as leaves and barks, containing mainly aporphine and benzyloquinoline alkaloids, lignans, neolignans, flavonoids, phenylpropanoids, and terpenes (Batista *et al.*, 2010; De Camargo *et al.*, 2013; Gottlieb, 1972; Ludy Cristina; Luis Enrique, 2010; Marques, 2001). Regarding the main chemical compounds in the genus, the alkaloid classes of benzyloquinoline and aporphine, which are common in the *Ocotea* genus, have been described with a wide range of different bioactivities and mechanisms of action. The aporphine alkaloids, besides anti-inflammatory activity (De alcântara *et al.*, 2021), are also recognized to have affinities by the dopaminergic, adrenergic and serotonergic receptors system, with potent agonist action (Kapadia; Harding, 2016).

Aporphine alkaloids from *Dactylicapnos scandens* (Papaveraceae) and semi-synthetic aporphine alkaloids demonstrated potent *in vivo* anti-inflammatory activity, and the bioassay results suggested that the mechanism involved possibly caused peripherally anti-inflammatory effects by inhibition of the expression of cytokines TNF- α and IL-1 β , and also the chemical inflammation mediator PGE2 (Wang *et al.*, 2020). The aporphine derivatives found in the *Ocotea* genus, dicentrine and glaucine (Figure 5) showed also *in vitro* cytotoxic activity. The possible mechanisms investigated revealed that these compounds can bind to DNA and act as intercalating agents. Also, dicentrine and dicentrinone (Figure 5) can interfere with the catalyst activity of topoisomerases, which are enzymes that are important for keeping the proper topology of the double-helical DNA structure (Hoet *et al.*, 2004; Salleh; Ahmad, 2017; Zhou *et al.*, 2000).

Figure 5 - Anti-inflammatory and cytotoxic NP isolated from *Ocotea* genus.



Source: From author (2024)

1.4 Metabolomics and computational analysis

Metabolites are the end products of cellular processes and reflect the combined influences of genes, environmental factors, and the organism's phenotype (Alarcon-barrera *et al.*, 2022; Aydođan *et al.*, 2020). In recent years, the integration of modern approaches, such as metabolomics, chemometrics tools and machine learning (ML)

techniques, has revolutionized the field of NP research and metabolite profile characterization (Chagas-paula *et al.*, 2015b; Ebbels *et al.*, 2023; Mannocho-russo *et al.*, 2023). Particularly, metabolomics comprehends the study of small molecules, or metabolites, within cells, tissues, or organisms. It aims to comprehensively profile and quantify these metabolites to understand biological processes and how they change under different conditions (Alseekh *et al.*, 2021; Canuto *et al.*, 2018). On the other hand, chemometric and computational tools are statistical and mathematical methods used to analyze chemical data, such as those obtained from metabolomics. Chemometric tools help in interpreting multidimensional data sets, identifying patterns, and extracting meaningful information (Boccard *et al.*, 2019; Yi *et al.*, 2016). Whereas, ML is a subset of artificial intelligence that involves training algorithms to learn from data and make predictions or decisions without being explicitly programmed. In NP research, ML models can predict the biological activity of compounds and extracts, identify patterns in the metabolomics data, and assist in the prediction of new bioactive samples (Alcântara *et al.*, 2023; Carpenter *et al.*, 2018; Chagas-paula *et al.*, 2015b)

Metabolomics currently plays a central role in several areas of plant sciences and offers new perspectives for the advancement of drug discovery, chemical ecology and taxonomy research fields. These tools allow for a more systematic and comprehensive analysis of complex biological matrices, facilitating the rapid identification of bioactive compounds and their biosynthetic pathways (Ebbels *et al.*, 2023; Kosmides *et al.*, 2013; Pilon *et al.*, 2020). The fusion of traditional NP research with metabolomics and computational tools can accelerate the discovery of new drugs and open new frontiers in understanding the ecological roles and evolutionary significance of living organisms (Harvey; Edrada-ebel; Quinn, 2015; Nephali *et al.*, 2022; Pilon *et al.*, 2020; Zhang *et al.*, 2020).

Metabolomics studies the metabolome, and the metabolome is defined as all metabolites of low-weight molecules of any biological sample of a living organism (Fiehn, 2001; Van Der Laan *et al.*, 2020; Vuckovic, 2012). While, we are still far from covering the whole metabolome with only one single analytical technique, especially for plants due to their widespread specialized metabolism, metabolomics strategies can significantly aid the identification and characterization of metabolites under a given condition, qualitatively and quantitatively (Pilon *et al.*, 2020; Vereyken *et al.*, 2019). Thus, metabolomics is currently perceived as a rapidly advancing multidisciplinary field that involves the qualitative and quantitative analysis of the metabolome in a range of

different biological samples. Metabolomics can still be defined as a multiparametric approach that involves the systematic and comprehensive analysis of present metabolites in any biological system (Theodoridis; Gika; Wilson, 2011; Worley; Powers, 2012).

Under metabolomics studies, the main chemical composition of an organism, a specific metabolite or metabolic via can be detected using different sets of hyphenated analytical techniques. LC-HRMS, GC-MS and NMR are the most commonly used techniques, which could be analysed separately or concatenated, and the generated data can be further analyzed in a range of different software and computational tools (Clendinen; Monge; Fernández, 2017; Ebbels *et al.*, 2023; Farag *et al.*, 2012; SPICER *et al.*, 2017). As the samples of a metabolomics study usually contain complex chemical content, of several different classes of molecules, the sample preparation is challenging and typically followed by careful organization of data acquisition, configuration and type of the analytical platform selected, extensive computational chemical analysis, which involves the normalization, deep data mining, multivariate statistical analyses (MSA), and sometimes the use of even other bioinformatics tools to perform adequate data interpretation (Alseekh *et al.*, 2021; Evans *et al.*, 2020; Long *et al.*, 2020).

However, metabolomics studies face several challenges, which include mainly the dynamic changing nature of the metabolism itself, including ongoing cellular interactions and environmental influences that complicate experimental standardization (Holmes; Wilson; Nicholson, 2008; Theodoridis; Gika; Wilson, 2011). Particularly, in plant metabolomics, the additional extensive chemical diversity of primary and secondary metabolites together with the wide concentration gradients of different substances, poses significant challenges in precisely understanding biological functions and characterizing structures, besides complicating experimental setups and data analysis (Pilon *et al.*, 2020). Consequently, metabolomics analyses generate a complex, large and unavoidable amount of raw data, and thus metabolomics researchers rely on a range of different computational software and tools to perform the data processing and analysis, including open ones such as MZmine, MS-DIAL and OpenMS or commercial such as Target Analyses, Masslynx, UNIFI, XCMS and others. In general, these software allows data processing and the use of free and commercial databases, as well as, *in house* ones, which facilitates the dereplication and targeting of new substances to isolate, or those with pharmacological interest (Demarque *et al.*,

2020; Tsugawa *et al.*, 2015; Vereyken *et al.*, 2019). Dereplication also commonly named annotation or putative identification, is the identification of compounds by comparison with data from known compounds with different levels of confidence (Spicer; Salek; Steinbeck, 2017; Sumner *et al.*, 2007).

Basically, there are two main approaches under metabolomics studies, the targeted and untargeted methodologies. The targeted consists of sample analyses that the chemical compounds that will be studied are already known, and thus the extraction demands to be selective for the quantitative monitoring of these analytes. Thus, the instrumental set-up is optimized for maximum sensitivity at the specific m/z related to the metabolites of interest (Braga; Adamec, 2018; Canuto *et al.*, 2018). In general, these techniques require the use of analytical patterns to uncover a potential mechanism of a disease or a primary study regarding genetic and epigenetic changes. Whereas, the untargeted aims at unknown components in a sample, and thus a more comprehensive extraction is necessary. The analytical instrument is optimized for a broad range of molecular masses (m/z), in typical untargeted metabolomics study ranging from 50–2200 m/z . An untargeted approach is commonly used to discover biomarkers, e.g. of diseases, pharmacological activity or toxicological property. For instance, biomarkers have been defined as a substance that indicates a normal biological or pathological process or pharmacological response (Chagas-paula; Oliveira; Faleiro, 2015; Chen *et al.*, 2021; De Vos *et al.*, 2007; Vinayavekhin; Saghatelian, 2010).

In particular, LC-HRMS-based untargeted metabolomics represents a modern approach in NP research, with great potential for field advancement by enhancing metabolite coverage (Chen *et al.*, 2021). This approach typically employs data-dependent acquisition (DDA) and data-independent acquisition (DIA) techniques, which are pivotal for acquiring both precursor (MS^1) and fragment (MS^2) ion data (Ebbels *et al.*, 2023). DDA selects precursor ions based on their measured MS^1 scan abundance, to acquire their corresponding MS^2 spectra (Davies *et al.*, 2021). In contrast, DIA indiscriminately fragments all detectable precursor ions within a larger mass range. Although DIA's methods, spectral libraries, and software are better established in the proteomics field (Lou; Shui, 2024), DIA development and applications in metabolomics are starting to rise (Alka *et al.*, 2022; Fenaille *et al.*, 2017; Katchborian-neto *et al.*, 2024). The classic DIA methods are based on alternation between low and high-energy channels, to acquire both MS^1 and MS^2 scans,

respectively. The terminology varies across different vendors' MS platforms. For example, in Waters quadrupole-time of flight (QTOF) instruments it is referred to as MS^E, while in Thermo Fisher Orbitrap™ instruments it is termed all-ion fragmentation (AIF) (Carnevale Neto *et al.*, 2022; Wang; Yin; Zhu, 2019).

Apart from the fact that DDA is the most popular method in metabolomics, DIA is an MS acquisition mode that offers exceptional sensitivity for characterizing complex metabolite samples, such as extracts of plants (Chagas-paula *et al.*, 2015a; Li *et al.*, 2020; Van Der Laan *et al.*, 2020). By systematically fragmenting precursor ions within a specific mass-to-charge ratio (m/z) range, DIA is particularly effective at identifying low-abundance metabolites, which are often overlooked by conventional DDA modes (Guo; Huan, 2020; Klont *et al.*, 2020; Li *et al.*, 2020). The analysis of processed data, independently of their modes of MS acquisition, usually is further followed by statistical analysis, which is suitable for the study of the correlation between variables of the study. The most common chemometrics methods used for metabolomics data are principal component analysis (PCA), hierarchical cluster analysis (HCA), partial least squares regression (PLS) and orthogonal partial least squares™ (OPLS). One of the main objectives of these multivariate analyses is to clean and reduce data dimensionality, in the case of the unsupervised analysis, and in the case of supervised methods to facilitate the visualization of the most discriminant compounds in the chemical composition of the investigated species (Jae-won; Heon, 2015; Katajamaa; Orešič, 2007; Worley; Powers, 2012; Yuliana *et al.*, 2011).

Furthermore, molecular networking (MN) is another powerful computational approach tool employed in NP and metabolomics research, especially from the last 10 to 15 years. It involves the clustering and visualization of LC-HRMS data based on the similarity of their MS/MS spectra (Aron *et al.*, 2020; Bonde *et al.*, 2021). This method enables an integrative approach to enhance data visualization and aid in the annotation of known compounds, as well as, the discovery of novel or structurally related compounds. It thus assist in the prioritization of unknown compounds of interest for further investigation. In MN, each mass spectrum from the dataset is represented as a node, and spectral similarities between nodes are calculated using algorithms such as the cosine similarity in the global natural product social molecular networking (GNPS). The GNPS is a data-driven online platform for the storage, analysis, consult of MS/MS spectra, and for sharing spectral data with the scientific community (Aron *et al.*, 2020; Bonde *et al.*, 2021; Demarque *et al.*, 2020; Wang *et al.*, 2016).

Thus, the use of MN under metabolomics studies allows the exploration of the chemical diversity within any complex dataset to elucidate the chemical structural relationships among the studied compounds, based on their fragmentation patterns. This approach has already proven to be particularly valuable in NP chemistry, where it has accelerated the discovery of new bioactive compounds from various sources, including plants, microbes, and marine organisms (Aron *et al.*, 2020; Nothias *et al.*, 2020). Thus, with the integration of MS spectral information with online chemical databases, modern dereplication strategies, and MN together enable efficient chemical characterization and metabolomics analysis. Additionally, metabolomics studies can focus on the prioritization of compounds with unique structural features or/and potential biological activities (Aron *et al.*, 2020; Bonde *et al.*, 2021; Wang *et al.*, 2016). Thus, metabolomics and computational approaches can be treated as an ally for the drug development process, and as a comprehensive strategy that allows profiling complex mixtures of myriad chemical components in crude NP extracts (Chagas-paula *et al.*, 2015a; Vinayavekhin; Saghatelian, 2010; Wolfender; Marti; Queiroz, 2010; Yuliana *et al.*, 2011).

Overall, metabolomics studies can aid faster new drug candidates by discovering active natural compounds for a high number of samples, through the integration of modern efficient analysis. The untargeted metabolomics approach is one of the latest omics technologies that has been successfully applied across a wide range of scientific research areas to analyze the global metabolome of organisms (Vinayavekhin; Saghatelian, 2010; Wolfender *et al.*, 2019). Untargeted approaches could include obtaining metabolic fingerprints, which provide a semi-quantitative chemical profile of a sample under specific biological conditions, representing the full range of metabolites produced by an organism (Sumner *et al.*, 2007; Van Der Kooy *et al.*, 2009). This approach in metabolomics has several practical applications, for example, the consistent quality control of herbal materials through comparative metabolic profiling (Rao Gajula; Nanjappan, 2021). In drug discovery, metabolomics is increasingly being used to analyze plant products, trace their metabolic profiles, and statistically identify metabolites correlated with pharmacological activity. This approach also helps in identifying key metabolic pathways that can be targeted for drug development (Chagas-paula *et al.*, 2015a; Vinayavekhin; Saghatelian, 2010; Wolfender; Marti; Queiroz, 2010).

Chapters II, III and IV describe the use of these computational techniques to study bioactive compounds of the metabolome of *Ocotea* plant species from Brazil. A chemical review of the current known biosynthetic aspects that led to the main chemical cores and isolated metabolites reported in the literature from *Ocotea* spp. is presented in Chapter I. The anti-inflammatory untargeted metabolomics study of *Ocotea* species is presented in Chapters II and III. In addition, the chemical diversity of these species was explored using MN and gas-phase fragmentation reactions, and it is presented in Chapter IV. The other scientific contributions developed during the time of this PhD is presented in Chapter V.

2 OBJECTIVES

2.1 General objective

To investigate the anti-inflammatory bioactive markers and the chemical diversity of *Ocotea* species using metabolomics and computational approaches.

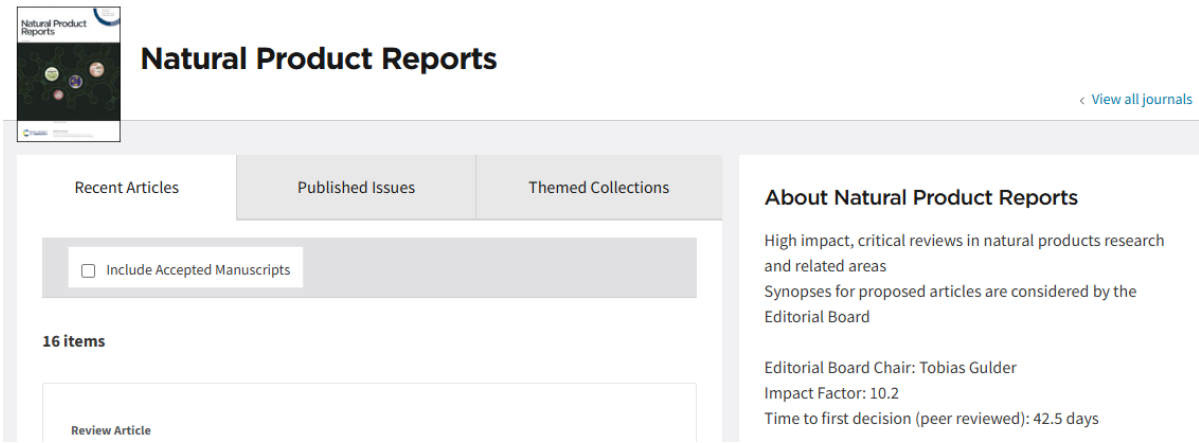
2.2 Specified objectives

- 1) Create a comprehensive database with all metabolites already identified in the *Ocotea* genus.
- 2) Perform crude extract preparation of 60 different *Ocotea* spp. to cover a significant amount of species of the genus.
- 3) Perform the acquisition of LC-HRMS–DIA data of each selected *Ocotea* (n=60) and the quality controls (QCs) for untargeted metabolomics analyses.
- 4) Perform *ex-vivo* anti-inflammatory assays for PGE2 and LTB4 determination.
- 5) Perform the acquisition of NMR data of promissory extracts to explore the chemical shifts most correlated with bioactivity for untargeted metabolomics analyses.
- 6) Perform MS-Data processing and NMR-Data processing prior to statistical analysis.
- 7) Perform Multivariate Statistical Analysis using computational tools to explore the bioactive markers.
- 8) Perform Molecular networking of LC-HRMS–DIA data using computational tools to explore the chemical profile of the species.
- 9) Perform characterization and annotation of the chemical profile and bioactive markers after Multivariate Statistical Analysis
- 10) Develop anti-inflammatory prediction models based on metabolomics data for future *Ocotea* species bioprospecting studies.

3 CHAPTER I - (THEORETICAL REFERENCE) – BIOSYNTHETIC ASPECTS OF *Ocotea* METABOLOME

The theoretical reference of this thesis is presented in a manuscript format. It includes the biosynthetic aspects and the chemical scaffolds of the *Ocotea* metabolome. Chapter I containing the review manuscript is separately provided as an attached material.

Targeted Academic Journal: Natural Products Reports (RSC) (not submitted yet)
Impact factor (2024): 10.2



The screenshot displays the homepage of the journal "Natural Product Reports". At the top left is the journal's logo, which includes the text "Natural Product Reports" and a graphic of four colored spheres. To the right of the logo is the journal title "Natural Product Reports" in a large, bold font. Further right is a link that says "< View all journals". Below the title is a navigation bar with three tabs: "Recent Articles", "Published Issues", and "Themed Collections". Under the "Recent Articles" tab, there is a checkbox labeled "Include Accepted Manuscripts" which is currently unchecked. Below this, it states "16 items". A single article is visible in the list, titled "Review Article". On the right side of the page, there is a section titled "About Natural Product Reports" which contains the following text: "High impact, critical reviews in natural products research and related areas", "Synopsis for proposed articles are considered by the Editorial Board", "Editorial Board Chair: Tobias Gulder", "Impact Factor: 10.2", and "Time to first decision (peer reviewed): 42.5 days".

The metabolome of *Ocotea* genus (Lauraceae): From biosynthetic aspects to widespread bioactive chemical scaffolds

Albert Katchborian-Neto^a, Karen J. Nicácio^b, Paula C. P. Bueno^c, Matheus Fernandes Alves^a, Miller Santos Ferreira^a, Wanderleya S. Toledo^d, Daniele de Oliveira Silva^a, João Henrique G. Lago^e, Daniele F. Dias^a, Marisi G. Soares^a, Ana C. C. Paula^d, RuAngelie Edrada-Ebel^f, Daniela A. Chagas-Paula^a

Received 00th January 20xx,

Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

The *Ocotea* genus has been reported as one of the main potential sources of bioactive agents within the Lauraceae, and yet it remains insufficiently explored in academic literature. There are species of *Ocotea* recognized as essential oil-producing trees and others that are considered valuable and diverse sources of specialised metabolites, e.g. the *O. duckei*, *O. gomezii*, *O. lanciofolia* and others. Additionally, several *Ocotea* species have a long history of traditional medicinal use, and recent scientific studies have reinforced their potential as promissory sources of bioactive agents. However, to date, accurately identifying this genus remains a challenge, as it is easily confused with other genera within the Lauraceae, and botanically recognized as part of the multifaceted 'Ocotea complex', a phylogenetically unresolved group. This review addressed the most current knowledge on the chemical steps involved in the biosynthesis of the *Ocotea* metabolome. It highlighted the metabolic pathways that are chief to lead to the bioactive scaffolds of alkaloids, lignoids, flavonoids, and terpenoids. This article aims to be a reliable and updated source of information to guide further research in the fields of Natural Products Research, chemophenetics, chemoinformatics, metabolomics, and medicinal chemistry regarding the *Ocotea* genus, its chemical scaffolds, and its specialised metabolites.

1. Introduction
 - 1.1 Biosynthesis, chemical aspects and bioactivities
2. Chemical database construction (*Ocotea*DB)
3. *Ocotea* spp. metabolites: Alkaloids
 - 3.1 *Ocotea* alkaloids: A biosynthetic approach
 - 3.2 *Ocotea* alkaloid profile
4. *Ocotea* spp. metabolites: Lignoids
 - 4.1 *Ocotea* lignoids: A biosynthetic approach
 - 4.2 *Ocotea* lignoid profile
5. *Ocotea* spp. metabolites: Flavonoids
 - 5.1 *Ocotea* flavonoids: A biosynthetic approach
 - 5.2 *Ocotea* flavonoid profile
6. *Ocotea* spp. metabolites: Terpenoids
 - 6.1 *Ocotea* terpenoids: A biosynthetic approach
 - 6.2 *Ocotea* terpenoid profile
 - 6.2.1 Monoterpenes
 - 6.2.2 Sesquiterpenes
- 6.3 *Ocotea* spp. essential oils
- 6.4 *Ocotea* spp. phenylpropanoids

- 6.5 Remarking considerations and recent advances
7. Conclusion remarks and future prospective
8. Author contributions
9. Conflicts of interest
10. Acknowledgments
11. References

1. Introduction

The *Ocotea* genus belongs to the Lauraceae Jussieu family, popularly known as the laurel family, which gathers more than 3000 species distributed amongst 55 genera worldwide.^{1–4} Lauraceae is amongst the top five largest in the world's tropical and sub-tropical forests, and it is classified into the Magnoliid clade, as part of the order Laurales.^{5–8} Several Lauraceae plants are traditionally used in culinary, e.g. laurel (*Laurus nobilis*), which is used as a flavouring agent to season food in native American cooking recipes, and cinnamon (*Cinnamomum verum*), a famous spice used globally as a condiment in various culinary traditions, such as in the Asian recipes. *Persea americana* is a tree that produces one of the most important tropical fruits in the world, popularly known as avocado.^{6–10} In general, Lauraceae species are aromatic trees, of which several are recognised as natural sources of essential oils. Example species include *Aniba rosaeodora*, known in Brazil as "pau rosa", and *Ocotea catharinensis*, known as "canela preta", both popular for their high linalool oil content and their pleasant scents. Linalool is a monoterpene highly requested to enhance woody fragrances in several commercial products.^{8,11–13} Other species of Lauraceae are documented as popular medicines, which are known for their significant antioxidant, analgesic, anti-inflammatory, and antitumor properties.^{6,7,14} For

^a Institute of Chemistry, Federal University of Alfenas-MG, 37130-001, Alfenas, Minas Gerais, Brazil.

^b Department of Chemistry, Federal University of Mato Grosso, 78060-900, Cuiaba, Mato Grosso, Brazil.

^c Leibniz Institute of Vegetable and Ornamental Crops (IGZ), Theodor-Echtermeyer-Weg, 14979, Großbeeren, Germany

^d Department of Pharmaceutical Sciences, Federal University of Juiz de Fora, 36036-900, Juiz de Fora, Minas Gerais, Brazil.

^e Center of Natural Sciences and Humanities, Federal University of ABC, 09210-180, Santo Andre, São Paulo, Brazil.

^f Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, G4 ORE, Glasgow, Scotland.

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

example, *Cinnamomum camphora* is a plant recognized by traditional medicine as a pain relief agent and is used to reduce local pain and itching. The plant is rich in the monoterpenoid camphor, which is the active main ingredient of commercially available topical analgesic ointments in the market.^{13,15–17}

Within the popular ethnomedicine of native South American people, traditional uses of several *Ocotea* species with their significant medicinal properties are of great value to local communities. *O. quixos*, a medicinal plant native to the Andes Mountains is used as an anaesthetic against pains in wounds and skin conditions as well as for the treatment of digestive problems and stomach aches.^{6,18} *O. lancifolia* and *O. cymbarum* were used against parasitic diseases caused by *Plasmodium*, *Leishmania* and *Trypanosoma* along with other medicinal properties, including antioxidant and anti-inflammatory effects.^{6,19,20} The metabolites responsible for these biological activities are mainly alkaloids, lignoids, and terpenoids.^{6,18–20}

Natural Products (NP) research plays a key role in the discovery of new bioactive compounds, as it is a consensus that nature is one of the main accessible sources of novel metabolites.^{21–23} Thus, biodiversity protection is a consequence of bioprospecting natural resources, such as the plants from the *Ocotea* genus, which have been a target of academic investigations not only due to their ethnomedicinal properties but also due to their ecological, botanical, and economic relevance.^{5–8} Most *Ocotea* species are woody trees of high-quality desired timber, such as those from *O. porosa*, *O. odorifera* and *O. catharinensis*, amongst others, which humans have been using in carpentry and civil constructions for decades in South America.^{8,24} Unfortunately, illegal logging activity has led to several *Ocotea* species being currently endangered in Brazil.^{24–26}

Ocotea Aubl. is the largest genus of the Lauraceae family, comprising more than 400 species.¹ To date, only 113 species (~28.5%) have been investigated for their chemical content either for classical or modern NP studies, including our recent *Ocotea* metabolic profiling study (Supplementary **Table S1**, **S1.1**, and **S1.2**, respectively, available to download at Zenodo's link- <https://doi.org/10.5281/zenodo.1067>).²⁷ The majority of the *Ocotea* species are neotropical species found mainly in Central and South America. However, *Ocotea* species are also found in Western and Southern Africa. It also occurs in Macaronesia, Madagascar, and Comoro islands.^{1,28,29} In South America, the genus *Ocotea* is the third most common in the Amazon biome, which offers the largest biodiversity hotspot in the world. In the main Brazilian biomes: Amazon rainforest, Atlantic forest, Cerrado, Caatinga and Pampas, a total of 176 species have been registered, of which 112 are considered endemic.³⁰ At least an estimate of 160 species remains yet to be found and taxonomically identified, and then phytochemically investigated.^{31–33}

Botanically, *Ocotea* is considered an intriguing genus that at present is classified as belonging to the Perseeae-Laureae clade. Phylogenetically and based on DNA sequencing markers and morphological features, *Ocotea* belongs to the *Ocotea* complex *sensu*, which currently includes about 950 species of distinct 17 genera.^{1,4,32,34,35} The *Ocotea* complex raises great interest to NP chemists due to the diverse biological activities such as anti-inflammatory, antibacterial, antiviral, or insecticidal that are associated with the chemical

composition of distinct plants. However, when DNA material is not available for comparison, based on only morphological characteristics, *Ocotea* species are a challenge to taxonomically identify and can be misclassified with other species from other genera such as *Nectandra*, *Persea*, *Licaria*, *Litsea* or many others from within the Lauraceae family. Also, phylogenetic studies on the genus still present gaps concerning their taxonomic classification, as some species can overlap morphologically with other genera from the *Ocotea* complex. Thus, *Ocotea* is considered one of the most challenging genera for species identification within the Lauraceae family.^{4,32,34,36,37} Synthetic systematics including chemistry and molecular biology can aid in the taxonomic classification of known *Ocotea* species, and despite the great economic interest in the genus, and relevant published works in the literature, there is still a lack of phytochemical and bioactivity studies.^{6,8,10,32}

1.1 Biosynthesis, chemical aspects and bioactivities

The first academic medical report on *Ocotea* was published in the medical journal, The Lancet, in 1830 reporting the therapeutic valuable oil of an *Ocotea* tree native to South America. Various parts of the tree were used to relieve spasmodic complaints, convulsions, and cramps. It was also reported to be effective against acute and chronic inflammation in the treatment of cutaneous eruption or rheumatic pain.³⁸ The first published data regarding the chemical aspects of *Ocotea* oil was from 1844. It was coined in the literature as the “oil of laurel turpentine”.³⁹ Likewise, the first non-volatile compounds reported for *Ocotea* were lignans, including sesamin (pseudocubebin), which was isolated in 1916 from the bark of *O. usambarensis*, a traditional medicinal plant used by African natives in the Kimboza Forest Reserve of Tanzania.^{40,41}

In the last seventy years, and especially in the current 21st century, an increased number of scientific papers in the literature have demonstrated that *Ocotea* is a promissory bioactive genus with relevant antimicrobial, larvicidal, anaesthetic, anti-inflammatory, and antitumoral properties.^{6,32,42–44} Our research group has confirmed the significant anti-inflammatory activity by dual inhibition of cyclooxygenase (COX) and lipoxygenase (LOX) pathways for reticuline, a benzylisoquinoline congener, which was isolated from the crude extract of *O. odorifera*, and also for the new aporphine, named diospirofoline isolated from the *O. diospyrifolia*.^{45,46} Therefore, *Ocotea* species stand as an interesting source of potential therapeutic compounds.^{28,29}

Concerning its phytochemistry, the genus *Ocotea* is perceived to be an extremely diverse genus, with a high degree of chemical diversity along with its specialised metabolites. Alkaloids are known as the most frequently occurring chemical class encountered in the genus. A series of different subclasses are observed such as a wide range of aporphines, and also benzylisoquinolines, and less common subclasses, such as the morphinans and phenanthrenes.^{6,32,47} Succeeding the alkaloids, the lignoids and flavonoids are the most commonly described NP classes in the *Ocotea* genus, which will be covered in this review. The metabolome of *Ocotea* also includes volatile aldehydes, terpenes, terpenoids and phenylpropanoids, and other specialised metabolites, which are less commonly found, such as coumarins, tannins,

polyketides, saponins, glycosides, benzopyrans and steroids (Supplementary Table S2-S7, available to download at Zenodo's link- <https://doi.org/10.5281/zenodo.1067>).^{6,31,32,48-50}

This considerable metabolic variability and diversity are corroborated by the fact that specialised metabolites in medicinal plants are biosynthesized by a combination of biosynthetic building blocks, which could be of the same type, or by a mixture of different building blocks. The chemical reactions that occur among these blocks are mediated through an abundant number of specific enzymes and cofactors that are responsible for the existence of diversified chemical pathways in plant biosynthesis.

As mentioned, the most shared alkaloid classes in the *Ocotea* genus consist of isoquinoline derivatives with aporphines, making up the majority. These classes have been considered some of the most important biomarkers of the genus and the main source of bioactivities, such as anti-inflammatory, antimicrobial, and antiparasitic.^{43,51,52} The combination of shikimate and pentose-erythrose-phosphate building blocks is responsible for the chemical core variations of these alkaloids (Fig. 1). The lignoids and essential oils can be also derived from these shared pathways.^{32,53,54}

With the aid of modern analytical molecular methods, specialised metabolites can be chemically discriminated in a given taxon. This area has been recently named chemophenetics. This approach can be seen as a modern extension of chemotaxonomy, which traditionally involves the classification of plants and other organisms based on their chemical composition, particularly specialised metabolites, that are unique to specific groups.^{32,55,56} Modern analytical instrumentation can support chemical studies of plant biosynthesis, and together with metabolomics investigations can provide a more accurate characterization of clades. The association with DNA sequencing and gene data can also help reach more integrative and robust data to solve phylogenetic issues. In addition, metabolic fingerprints can enable the identification of chemical markers within the *Ocotea* genus, e.g., metabolites from the classes of morphinans or *O*-glycosylated flavonoids can be used for metabolic fingerprint discrimination. Even the flavone:flavonol ratio can assist in the evolutionary lineage pattern studies, and once they are linked to the species specialization index, their chemical presence can be associated with a genetic modification during the evolution of these plants. The consequence is a more evolved metabolome of higher chemical complexity that can be differentiated using metabolomics strategies.^{32,55,57,58} A current application can include the use of the metabolic fingerprint content to avoid taxonomic errors in laboratories and industry due to morphologically similar botanical features.^{32,55,56} In particular, metabolomics investigation can support a faster discovery of new bioactive metabolites by high throughput screenings. Also, new potential bioactivities can come across for known or unknown metabolites.^{32,59-62}

Thus, a robust chemical database associated with the main biosynthetic pathways of the *Ocotea* metabolome can ease the recognition of new potential biomarkers and acquire more accurate and faster species differentiation. In addition, systematically describing respective chemical scaffolds can increase accessibility to an ample *Ocotea* chemical database.^{63,64} From our knowledge, this is the most comprehensive chemical review of the *Ocotea* genus, which

includes more than 900 chemical structures in a curated database (Available at Zenodo's link- <https://doi.org/10.5281/zenodo.1067>). This review aims to be an updated version of the current knowledge of the chemical and biosynthetic characteristics of the *Ocotea* metabolome to promote insights and to orientate further studies targeting *Ocotea* species. It is intended to be easy access to chemical cores and unique metabolites, integrated with the current biosynthetic knowledge. The different biosynthetic chemical classes were separately discussed in sections 3 (alkaloids), 4 (lignoids), 5 (flavonoids), and 6 (terpenoids).

2. Chemical database construction (*OcoteaDB*)

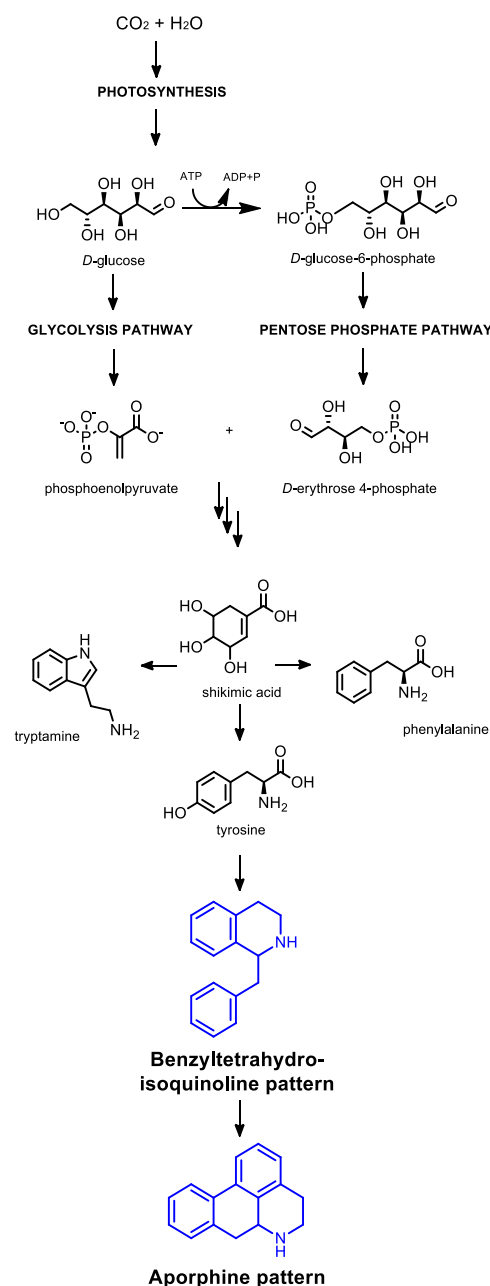


Fig. 1 The combination of the biosynthetic building blocks of the shikimate and the pentose-erythrose-phosphate pathway, derived from the aromatic amino acid tyrosine (*L*-Tyr), leading to the biosynthesis of the benzyltetraisoquinoline and aporphines alkaloids.

The *Ocotea* chemical database (*OcoteaDB*) was put together using 928 metabolites obtained from the literature and NP online databases, gathering them together according to their respective NP classes (Supplementary Table S2-S7, available to download at Zenodo's link-<https://doi.org/10.5281/zenodo.1067>). The bibliographic databases used for the research included: PubMed, Web of Science, ScienceDirect, and the Google Scholar platforms. Regarding chemical databases (DB), it included: DNP© (Dictionary of Natural Products), KNApSACK (A Comprehensive Species Metabolites Relationship Database), and NuBBE (Nuclei of Bioassays, Ecophysiology and Biosynthesis of Natural Products Database). The keywords “*Ocotea*”, “phytochemistry of *Ocotea*”, “metabolites of *Ocotea*”, “compounds of *Ocotea*”, “molecules of *Ocotea*”, “essential oils of *Ocotea*”, “secondary metabolism of *Ocotea*”, “chemosystematics of *Ocotea*”, and “metabolomics of *Ocotea*” were used to search for published articles between the years of 1830 and 2024.

OcoteaDB included chemical and/or trivial names, molecular formula, monoisotopic mass, metabolite core, biosynthetic class, SMILES, *Ocotea* species, and literature citation. The chemical structures were redrawn from original articles and downloaded in .mol format from ChempSpider®, Pubchem®, and J-GLOBAL® online databases when available. The structures were drawn using the software ChemDraw ultra 12.0 (Perkin Elmer Informatics, Cambridge, England) for chemical figures generation. Data visualization was performed using various Python libraries accessed through the Google Colab online cloud service platform (<https://colab.research.google.com/>). Pandas was employed for data manipulation, while NumPy supported numerical calculations and Matplotlib graph generation.

3. *Ocotea* spp. metabolites

Most studies on *Ocotea* species in the literature concern botanical, phylogenetics, phytochemistry, and biological activity investigations. From the chemical point of view, alkaloids, lignoids and terpenoids are the main representative classes, found in monoecious and dioecious species. Flavonoids are also found, although in a smaller number of species.^{6,10,27,32,47,54} However, chemical literature is scarce compared to the number of the already known *Ocotea* species.^{6,32} Even though notable discoveries were made by Otto Gottlieb and collaborators between 1960-1980, it is only recently that the chemical research interest has shifted back to the genus, with an increased number of relevant publications in the last 15 years.^{6,32,35,65}

3.1 *Ocotea* alkaloids: a biosynthetic approach

The main alkaloid class produced by *Ocotea* is the aporphines. These types of alkaloids are not exclusive to Lauraceae, as they also occur in other plant families such as Magnoliaceae, Menispermaceae, Papaveraceae, Ranunculaceae, Hernandiaceae, and Annonaceae.^{53,66} Aporphine alkaloids are one of the largest groups of isoquinoline alkaloid derivatives, with more than 500 known representatives in the literature. The biosynthetic pathway of aporphines is greatly dependent on nitrogen-based sources, especially the amino acid precursor tyrosine (*L*-Tyr). Despite

the chemical diversity and distribution of this chemical class, there is still a small amount of robust data regarding their metabolic biosynthetic pathways in the literature.^{53,54,67}

It has been established that in plant metabolism, *L*-Tyr undergoes modification through tyrosine hydroxylase enzymes (E1), resulting in the addition of a second hydroxyl group to the *ortho* position of the phenol group in the aromatic moiety. It allows the conversion of the amino acid into *L*-DOPA (**1**), which is the first step leading to the biosynthesis of the aporphine core. The next step includes dopamine (**2**) production through decarboxylation in the presence of DOPA decarboxylase (E2). In parallel, the *L*-Tyr also undergoes transamination by tyrosine aminotransferase (E3) and PLP coenzyme, giving rise to 4-hydroxyphenylpyruvic acid (**3**) that is also decarboxylated to 4-hydroxyphenyl-acetaldehyde (**4**). Subsequently, through a Pictet-Spengler mechanism, a mannich-like reaction takes place to join precursors **2** and **4** by amino alkylation to form (*S*)-norcoclaurine (**5**), the primary precursor for the benzyloisoquinoline class of compounds. Methylation of **5** yields (*S*)-coclaurine (**6**). This step is mediated *via* norcoclaurine 6-*O*-methyltransferase (E4) and SAM (*S*-adenosyl methionine), a common cosubstrate involved in the stereoselective mechanism reaction of methyl group transfers by SN₂-type nucleophilic substitution. Then, precursor **6** is catalysed to (*S*)-*N*-methylcoclaurine (**7**) by coclaurine *N*-methyltransferase (E5). Following the pathway, **7** can be hydroxylated at position 3' of the benzyl ring to form the tetrahydroxy substituted pattern in the presence of the enzyme *N*-methylcoclaurine 3'-hydroxylase (E6) producing the (*S*)-3'-hydroxy-*N*-methylcoclaurine (**8**). In addition, the hydroxyl group at position 4' can be methylated by 3'-hydroxy-*N*-methylcoclaurine-4'-*O*-methyltransferase (E7)/SAM to afford the alkaloid (*S*)-reticuline (**9**), which is one of the main benzyloisoquinoline alkaloids found in the *Ocotea* genus. For instance, **9** has been isolated from the crude extracts of several species, such as *O. lancifolia*, *O. caparrari*, *O. odorifera*, *O. caudata* and others.^{6,45,53} The biosynthetic route is summarized in Fig. 2A.

Although *Ocotea* does produce *R* isomers as well,^{45,46,53} herein, the dominant presence of *S* isomers in the *Ocotea* genus has been described for benzyloisoquinoline alkaloids. Thus, as E4, E5, E6 and E7 are stereoselective enzymes, it is expected that the biosynthesis of either *R* or *S* stereoisomers after the Mannich reaction to be dependent on the configuration of the precursor. As a consequence, derivatives **5**, **6**, **7**, **8** and **9** could assume either of both isomer forms in the *Ocotea* metabolism, considering they might retain the stereochemistry of their initial precursor throughout their biosynthetic pathway. This is an appropriate piece of information contributing to *Ocotea* species differentiation and phylogenetic studies. For instance, our research group isolated the (*6R*)-reticuline (**14**) from *O. diospyrifolia*, while (*6S*)-reticuline has also been described from *O. odorifera*.^{45,46} Different (*R*)-alkaloid isomers were also reported from other *Ocotea* species, such as those isolated from *O. caesia*, *O. lancifolia*, *O. velloziana* and others.^{20,68,69} Therefore, these findings corroborate the significance of chemophenetics and biosynthetic studies, which to date are poorly described in the literature and therefore, can be further explored by using specific enzymes and substances

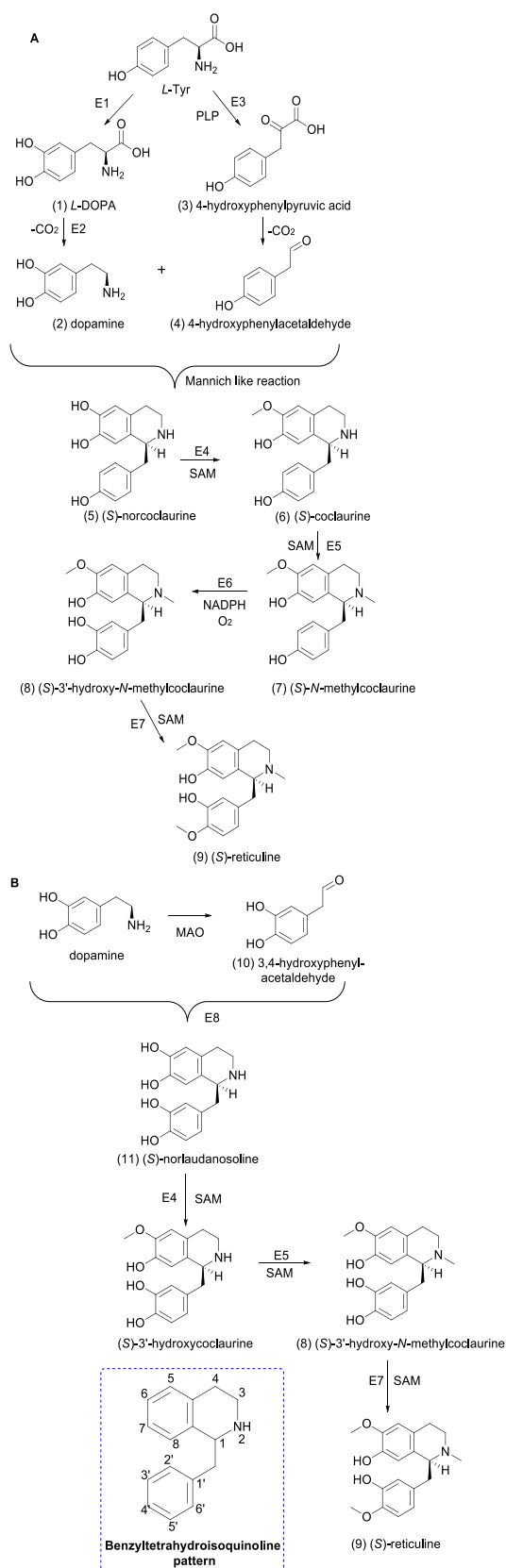


Fig. 2 Biosynthetic pathway of benzylisoquinoline and the alkaloid (*S*)-reticuline. A – (*S*)-reticuline produced via norcoclaurine. B- (*S*)-reticuline produced via norlaudanansoline.

with atom labelling, to increase and confirm the assumptions and observations highlighted in this review.^{46,68,70}

Recently, (*S*)-reticuline (**9**) was successfully obtained in a transgenic system by using genetic engineering to incorporate genes from *Coptis japonica* (Ranunculaceae) into *Escherichia coli* host cells. The bacterial encoding gene of monoamine oxidase enzymes (MAO) converted the supplied dopamine to the culture into 3,4-dihydroxyphenylacetaldehyde (**10**). Subsequently, consolidating dopamine with **10** by the action of norcoclaurine synthase (E8), (*S*)-norlaudanansoline (**11**) could be formed. The *E. coli* host system was inherently responsible for the presence of the methylating agent SAM while *C. japonica* genes provided the enzymes for the ongoing biosynthetic pathways (E4, E5, and E7).^{53,71} Therefore, this engineered system is an alternative pathway to obtain **9** via the formation of (*S*)-norlaudanansoline (**11**) (Fig. 2B).

Furthermore, (*S*)-reticuline (**9**) can undergo phenolic oxidative coupling, which plays a significant role in modifying the basic benzyltetrahydroisoquinoline skeleton to yield several other types of alkaloids, such as aporphines and morphinan derivatives. While the former is widespread in the *Ocotea* genus, the latter is also present, but only in fewer *Ocotea* species.^{47,53,72} The aporphines, (*S*)-corytuberine (**12**) and (*S*)-isoboldine (**13**) have been found in different *Ocotea* species ($n = 8$) in literature, such as *O. caesia*, *O. caudata*, and *O. lancifolia*, as well as in other Lauraceae genera.^{6,32,47,53} Thus, regarding the biosynthetic aspects, the first reaction step consists of the oxidation at the unprotected hydroxyl group of the aromatic rings to give the intermediate ketones (Fig. 3A). Further, electron oxidation can result in resonance-stabilized radicals that cause the oxidative coupling at position 8 and position 2' or 5' of the benzylisoquinoline aromatic ring, as the C-C bond can rotate freely. In addition, a cytochrome P-450-dependent enzyme is necessary (e.g. CYP80G2) to allow oxidative coupling catalysis.⁵³

Literature data supports the change in the absolute configuration of **9** to its *R* enantiomer (**14**) that could be achieved by an oxidation-reduction process through the formation of 1,2-dehydroreticulium cation intermediate (**15**). The *R* enantiomer is formed due to oxidation that turns it up in a planar iminium cation enabling the formation **14**. These steps allow the change in the stereochemistry through the action of two coordinated enzymes: 1,2-dehydroreticuline synthase (E9)/NADP⁺, which causes the first heterocyclic nitrogen oxidation, and 1,2-dehydroreticuline reductase (E10)/NADPH, reducing the ion to form the enantiomeric congener **14** (Fig. 3B).^{53,73}

As previously discussed, some *L*-Tyr-derived alkaloids can assume the *R* configuration, e.g. in the classes of opioids and morphinans that are predominantly biosynthetically derivatised from (*R*)-reticuline (**14**) stereoisomer.⁵³ As mentioned before, both *R* and *S* isomers could be found in the genus *Ocotea*, including benzyltetrahydroisoquinoline precursors such as **6**, **9**, and **14**, but also for several aporphines and morphinans, that are biosynthesized farther in the biosynthetic pathway.^{53,73} Thus, a systematic analysis of this biosynthetic pathway differentiation is a piece of useful information for further studies on lineage evolution.

Therefore, **14** is the main precursor for the morphinan class of alkaloids, which gives rise to important classic alkaloids such as codeine and morphine then later on salutaridine (**16**) and pallidine (**17**). Initially, one-electron

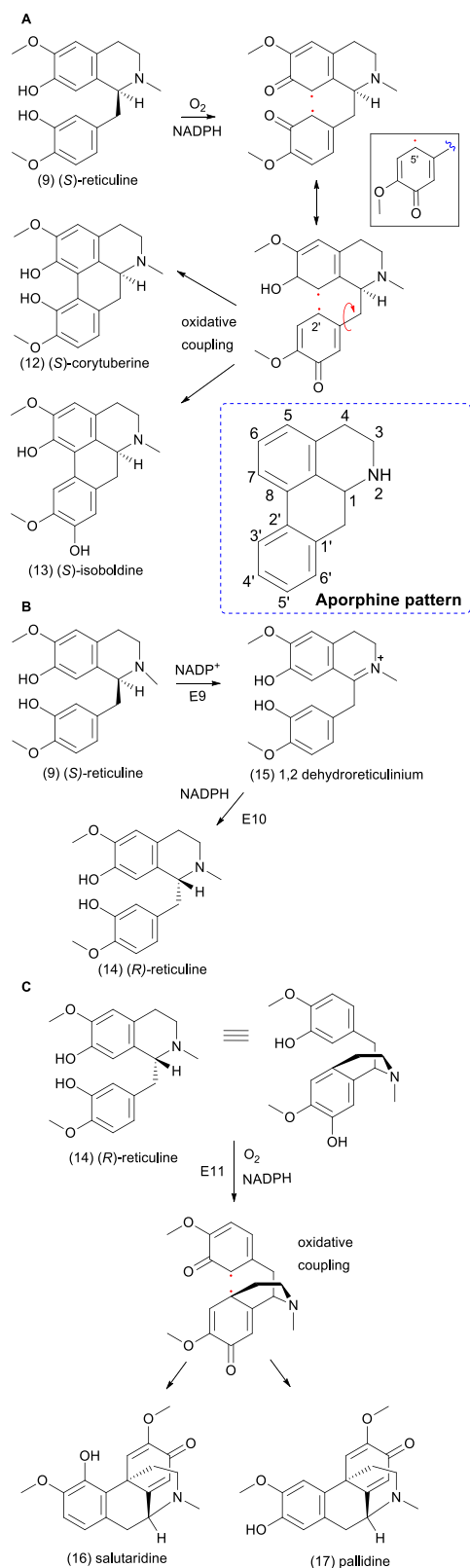


Fig. 3 A- Aporphine alkaloids (S)-isoboldine and (S)-corytuberine is produced by phenolic oxidative coupling of the (S)-reticuline precursor. B- The oxidation-reduction process through the intermediate 1,2-dehydroreticulinium cation to convert (S)-reticuline to (R)-reticuline. C- (R)-reticuline precursor yields salutaridine and pallidine morphinan scaffolds through oxidative coupling.

oxidation takes place in the two aromatic rings of a phenolic group forming ketone intermediates while the resonance-stabilized diradical allows an *ortho* coupling to the intermediate ketone in the isoquinoline moiety at position 7, and also *para* coupling to the ketone benzyl substituent at the 5' position. Biosynthetically these reactions are mediated via coupling enzyme synthases that are cytochrome P450-dependent monooxygenase, like salutaridine synthase (E11), responsible for giving rise to the dienones morphinan alkaloids **16** and **17** (Fig. 3C). Even though *Ocotea* species are not known to be an opioid producer, the morphinan alkaloid **17** has already been isolated from *O. acutangula*, *O. acutifolia*, and *O. lancifolia*, while **16** is commonly found in the Papaveraceae family.^{6,53,73}

Moreover, the CYP complex enzyme (Cyp80A1) can also catalyse the C-O intermolecular phenol-coupling reaction between benzylisoquinoline cores to afford a dimer class of alkaloids, the bisbenzylisoquinolines. Therefore, alkaloid **14** is also their main gateway precursor. Bisbenzylisoquinolines occur in plant families such as Lauraceae, Menispermaceae, Berberidaceae, and Ranunculaceae. This class of dimer alkaloids is not characteristic of the *Ocotea* genus and was only found in three *Ocotea* species, which include *O. rodiaei*, *O. rodiei* and *O. venenosa*.^{6,32}

Bisbenzylisoquinoline alkaloids consist of two benzyltetrahydroisoquinolines that are linked together through intramolecular bonds, which is a result of the phenolic oxidative coupling mechanism and the formation of resonance-stabilized radicals.^{53,74,75} Several different benzylisoquinoline precursors can undergo one-electron oxidations at the free phenol groups within each aromatic ring. Consequently, these oxidized groups can act as diradicals, coupling to initiate either an ether or a carbon-carbon (C-C) bridge between different precursors, thereby giving rise to a vast diversity of bisbenzylisoquinoline dimers. Bisbenzylisoquinolines, such as tetrandrine (**18**) and thalmine (**19**) can be found in the *Ocotea* genus (Fig. 4), and could be potential specific biomarkers for certain *Ocotea* species.

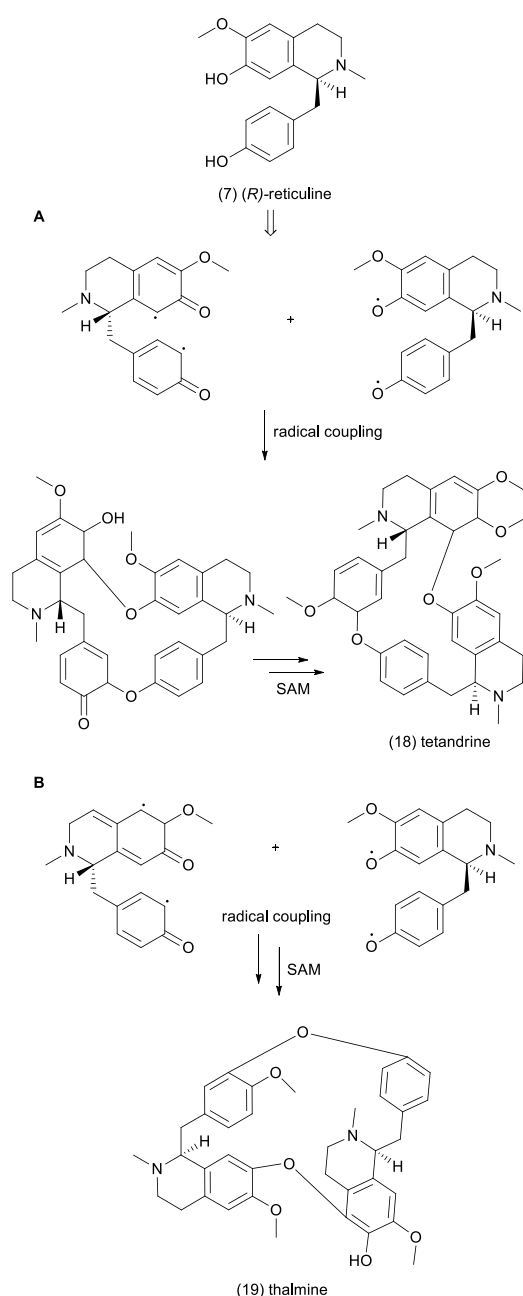


Fig. 5 The biosynthesis of bisbenzylisoquinolines via oxidative diradical formation and coupling of two benzylisoquinoline precursors to yield tetrandrine (A) and thalmine (B) alkaloids.

3.2 *Ocotea* spp. alkaloid profile

A comprehensive review with multivariate statistical analysis of the data by Antonio *et al.* (2020) claimed that approximately less than 10% of *Ocotea* species have been chemically evaluated for their alkaloidal content.³² Their study used 31 alkaloid-producing *Ocotea* species as extensively described earlier by Silva Teles *et al.* (2019).⁶⁸ Herein, the present review updated this number to 41 *Ocotea* species reported to date for those eliciting biosynthetic pathways for alkaloid production. Thus, the alkaloid class stand as the most prevalent class amongst the investigated *Ocotea* species with

174 unique reported metabolites. This number was increased to 60 *Ocotea* species by our recent publication applying modern analytical techniques, in which most of the *Ocotea* sp. were evaluated for the first time.²⁷ (Supplementary Table S2 and S2.1 respectively, available online at Zenodo's link-<https://doi.org/10.5281/zenodo.1067>).

Regarding only phytochemical and NP isolation studies, the aporphine alkaloids extend as the most common alkaloid source reported in *Ocotea* species with 107 representative structures in the literature (chemical cores in Fig 5). Of the 41 alkaloid-producing species, 35 (~85.4%) had the aporphine core described in at least one study (cores C2-C3). Moreover, beyond the aporphines, classes such as their precursors or derivatives that included benzylisoquinoline, bisbenzylisoquinoline, proaporphine, phenanthrene and morphinans, were reported in less number of *Ocotea* species. Thus, 26 (~63.4%) species have reported one of these latter cores in at least one phytochemical study, with a total of 34 unique metabolites (cores C1, C4-C6). Even less common chemical cores have also been reported, *e.g.* protoberberine (core C7) was described in *O. duckei*, while aminoaporphine (core C8) was found in *O. variabilis* and *O. glaziovii*. In addition, the indole core (C9) was reported in *O. minarum*, and the pyrrolidine core (C10) was found in *O. caudata*, whereas the isoquinoline core (C11) was isolated from *O. diospyrifolia*.^{6,46,76,77}

Aporphine is a relevant alkaloid class for evolutionary lineages in Lauraceae and is considered a biomarker for discriminating *Ocotea* as a basal genus from others in the family.³² For example, the various patterns of the substituted aporphines can aid in the differentiation of *Ocotea* from *Cinnamomum*, and even from other genera that could be morphologically similar and also phylogenetically placed at the unresolved *Ocotea* complex *sensu*.^{1,32} Thus, an alkaloid fingerprint profile including aporphine patterns can aid and support phylogenetic studies of Lauraceae, once gene sequence could be significant key information to better understand alkaloid expression.^{32,53} However, within the *Ocotea* genus, only specific substituted aporphines can be significant differential biomarkers, especially those with high levels of oxidation, *e.g.* 3-hydroxydicentrine or dehydroocoteine.³² This is expected as the majority of the *Ocotea* sp. are aporphine producers while sharing the same biosynthetic pathway. Thus, besides its potential for genera differentiation in the *Ocotea* complex, other metabolite classes could be utilised as a more efficient biomarker for a chemosystematic approach in the *Ocotea* genus, such as for some specialised lignans (item 4).

Phylogenetically and evolutionary classification states that the *Ocotea* species produces majoritarian aporphines as being part of the Old World, and thus it is assumed that they are more basal in the Lauraceae lineage.^{1,32} Instead, *Ocotea* species that produce more evolved chemical structures, such as the morphinan alkaloids could be considered from the New World, as they evolved from specialised biosynthetic pathways, such as in *O. acutangula*, from which six different morphinan alkaloids were isolated.^{1,6,32,54,76,78} Only five *Ocotea* species (*O. acutangula*, *O. acutifolia*, *O. brachybotra*, *O. caudata* and *O. lancifolia*) have been described in the literature as morphinan producers.

Furthermore, chemometrics and metabolomics studies can aid in the search for specific biomarkers and alkaloidal

fingerprints to facilitate species differentiation. For example, plants from *Licaria* and *Nectandra* genera might be also taxonomically misunderstood with the *Ocotea* genus due to their botanical similarity in terms of their leaf and flower morphologies.^{4,32,34} Some of these species can share the same stomatal surface shape, which makes morphological differentiation difficult. Thus, plants such as those from the *Ocotea* genus can be more easily differentiated by their chemical profiles and through biomarker comparison, when aided with chemometric analysis of their secondary metabolites. The chemical profile data together with respective morphological features along with phylogenetic studies can avoid inaccurate taxonomic assumptions and incorrect genus identifications.^{1,4,32,79}

Regarding the chemistry and bioactivity, as mentioned before, the classical chemical skeleton of the aporphine alkaloid class occurs when the C2' of the benzyloquinoline nucleus is attached to the C8 position. This chemical modification confers to this NP class substantial structural diversity and a broad variety of pharmacological activities. As a consequence, aporphine cores exhibit diverse chemical possibilities of -R substituents ranging from different positions as methoxy, methylenedioxy and hydroxy substituents attached to the aromatic ring.^{32,47,53,54} This review describes all substituted aporphine cores within *Ocotea* reported in the literature. It included derivative classes of oxoaporphine, proaporphine, *O*-aporphine, dihydroaporphine, dihydroaporphine, and phenanthrene.

To date, approximately 200 different alkaloids (174 distinct alkaloids of different subclasses found exclusively in *Ocotea* species) have been isolated from the *Ocotea* genus. Fig. 5-6 describes all known alkaloid chemical cores (C1-C11). The detailed identity of the chemical compounds including monoisotopic mass, MF, chemical name, SMILES-ID, and *Ocotea* species from which the metabolite was isolated, is presented in Supplementary Tables S2-S7, available at Zenodo's link- <https://doi.org/10.5281/zenodo.1067>.

Furthermore, the aporphine alkaloids (C2.1-2.6; C3.1-3.2), from different *Ocotea* species are reported with pronounced anti-inflammatory and antineoplastic properties. The aporphine alkaloid boldine (C2.1), for example, those found in *O. lancifolia*, *O. spixiana* and in other *Ocotea* and Lauraceae species, can induce antipyretic effects that can be due to the effect on the COX pathway, leading to inhibition of prostaglandins.^{80,81} Likewise, the aporphine alkaloid glaucine, which occurred in different *Ocotea* sp., such as the *O. vellosiana* and the *O. macrophylla*, exhibited interesting anticancer properties by suppressing nuclear factor kappa-activated B cells (NF-κB) activity. In addition, the glaucine congener has also demonstrated a reduction in metastatic breast cancer cell invasion.⁸²

The aporphine alkaloid dicentrine (C2.1), and the oxoaporphine dicentrinone (C3.1), demonstrated a broad spectrum of biological activities, e.g. inhibition of topoisomerase I and II enzymes, which are relevant therapeutic targets of current chemotherapy of protocols. In addition, more recent data has confirmed significant *in vivo* antinociceptive effects of dicentrine. The mechanism of action included attenuating mechanical and cold hypersensitivity in inflammatory conditions via activation of the transient receptor potential TRPA1.⁸³ Dicentrinone has also shown potent antiparasitic activity against trypanomastigote forms of

Trypanosoma cruzi, and significant changes in lipid biological surfaces with reduced mammalian cytotoxicity.^{52,84}

Regarding the benzyloquinoline alkaloids with the C1a core (Fig. 6), *R*-coclaurine demonstrated a highly pronounced anti-HIV activity with an EC₅₀ value of 0.8 μg/mL.^{85,86} In addition, promising butyrylcholinesterase inhibition activities were shown for other two benzyloquinoline alkaloids, reticuline with an IC₅₀ value of 33.6 ± 3.0 μM) and *N*-methylcoclaurine with an IC₅₀ value of 15.0 ± 1.4 μM.⁸⁷ Our research group has confirmed the anti-inflammatory activity of the *O. odorifera* leaf extracts validating the ethnopharmacological use of the plant decoction as a phytomedicine. The leaf decoction is rich in reticuline, which after alkaloid isolation and evaluation, exhibited a significant anti-oedematogenic effect together with neutrophil recruitment inhibition in a dose-dependent manner, suggesting that both COX and LOX inflammatory pathways were inhibited.^{45,88}

Benzyloquinolines and aporphinoids are the major classes isolated in *Ocotea* (n= 39 species, ~95.1%, 167 metabolites), and can be considered natural bioactive agents, encountered in several different *Ocotea* species, e.g. the *O. puberula*, *O. vellosiana*, *O. acutifolia*, *O. macropoda*, *O. leucoxydon*, *O. discolor*, *O. caesia*, *O. odorifera*, *O. diospyrifolia*, *O. lancifolia*, *O. brachybotra* and several other *Ocotea* sp. (Supplementary Table S2 and S2.1, available for download at Zenodo's link- <https://doi.org/10.5281/zenodo.1067>).

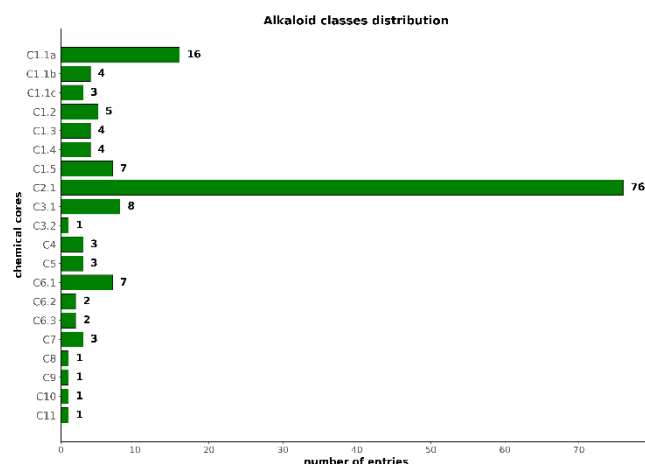


Fig. 5 Chemical diversity of alkaloids in the *Ocotea* genus expressed by the number of different chemical structures reported in the literature for each subclass.

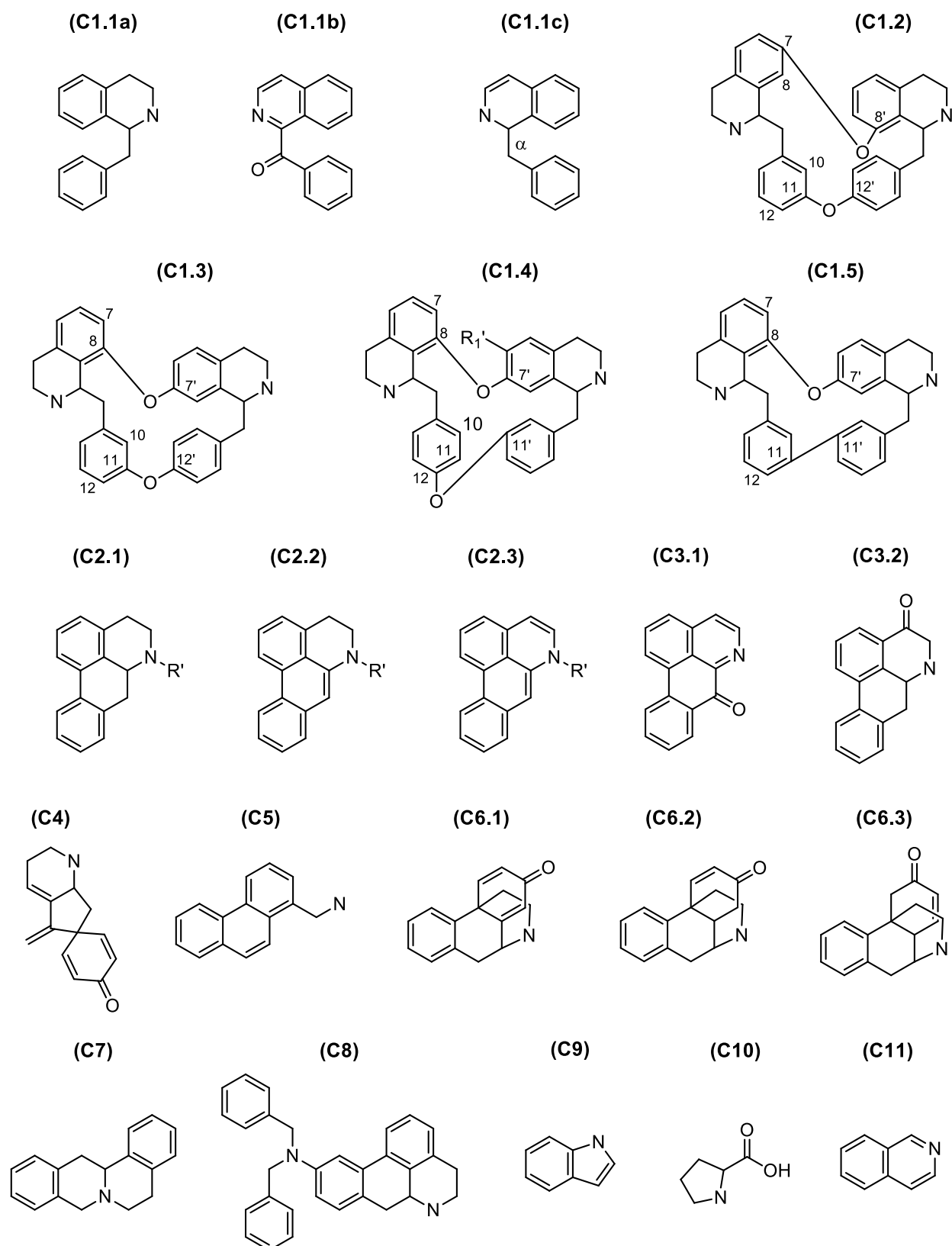


Fig. 6 Alkaloids chemical scaffolds found in *Ocotea* species: (C1.1a) benzylisoquinoline, (C1.1b) oxobenzylisoquinoline, (C1.1c) α -benzylisoquinoline, (C1.2) 8,7'-12,11'-bisbenzylisoquinoline I, (C1.3) 8,7'-12,11'-bisbenzylisoquinoline II, (C1.4) 8,7'-12,11'-bisbenzylisoquinoline III, (C1.5) 8,7'-11,11'-bisbenzylisoquinoline, (C2.1) aporphine, (C2.2) dihydroaporphine, (C2.3) didehydroaporphine, (C3.1) oxo-aporphine, (C3.2) didehydro-oxo-aporphine, (C4) pro-aporphine, (C5) phenanthrene, (C6.1) morphinan-7-one, (C6.2) dihydromorphinan-7-one, (C6.3) morphinan-6-one, (C7) protoberberine, (C8*) aminoaporphine, (C9*) indole, (C10*) pyrrolidine, and (C11*) isoquinoline alkaloids from *Ocotea* genus. *Only one study in literature have evidenced this chemical core for the respective *Ocotea* sp. evaluated.

4 *Ocotea* spp. metabolites: lignoids

4.1 *Ocotea* lignoids: a biosynthetic approach

Since the highly bioactive lignan, podophyllotoxin was identified back in 1933, the biosynthetic pathways of lignoids have been thoroughly elucidated, establishing their main bioactive scaffolds. Podophyllotoxin is an 8-8',2-7' aryltetralin lignan lactone widely described for its antiviral and especially for its anticancer properties. It was first isolated from the Himalayan rhizomes and shrub of *Podophyllum*, a member of the Berberidaceae family, in India. The anticancer properties of podophyllotoxin derive from its ability to inhibit DNA replication. To date, semi-synthetic derivatives of podophyllotoxin, such as teniposide and etoposide, have been approved for the pharmaceutical market. These drugs block cell division by inhibiting topoisomerase II enzymes and are favoured in several current chemotherapy protocols due to their fewer side effects compared to natural lignan.^{48,89,90} Consequently, the lignan derivatives of podophyllotoxin turned out to be a target of a wide range of academic investigations in the last few years. Additionally, the biosynthetic pathway that leads to this bioactive lactone lignan scaffold has been identified in other plant families, including Lauraceae, Cupressaceae, and Linaceae. The 8-8',2-7' aryltetralin lignan metabolite (–)-morelensin, isolated from *O. macrophylla*, comprises only two methoxy groups in its aromatic moiety, as opposed to three in podophyllotoxin, and thus originating from the same biosynthetic mechanisms and routes.^{53,90–92}

Regarding the chemical cores of lignoids, their general structure can be classified based on the character of the carbon-carbon (C-C) or carbon-oxygen-carbon (C-O-C) bond (oxyneolignans), which are responsible for joining the two phenylpropane units. Lignoid compounds are generally referred to as phenylpropanoid dimers, divided into three main classes: lignans, neolignans, and oxyneolignans.^{32,48,89} Among *Ocotea* species reported in the phytochemical studies in literature, ~58.8% produce lignans, while ~88.2% have at least one neolignan in their metabolome. Oxyneolignans are less common with only ~11.7% of the reported species being known for producing it, such as *O. cymosa* and *O. costulatum*.

Lignoids represent the second most common class of secondary metabolites in the *Ocotea* genus, characterized in 17 different *Ocotea* sp.^{32,93,94} Our recent publication, which employed modern analytical techniques, increased the number of known *Ocotea* species able to elicit lignoids biosynthesis for 26 (Supplementary Table S3 and S3.1, available at Zenodo's link-<https://doi.org/10.5281/zenodo.1067>).²⁷ As mentioned before, that recent study marked the first evaluation of several *Ocotea* species that have never been evaluated before.

Moreover, Fig. 7 illustrates Haworth's definition (1937) of a lignan, which is still recognized by the academic community today. This definition encompasses C6–C3 units connected at the 8-8' (β - β' linked) positions, with possible variations at the 7-7', 9-9', and 2-7' positions. Linkages differing from this pattern, such as 3-3', 7-1', 7-3', 8-1', 8-3'-linked, or other linkages like 8-O-4', are classified as neolignans.^{48,53,89} Later, Gottlieb (1974) suggested that the classification can be based on their respective precursors, leading to a better understanding of lignoid biosynthesis. Thus, the basic lignoid core is formed through an oxidative coupling of phenylpropanoids, yet the plant's metabolism shows a degree of biosynthetic independence, resulting in various congeners. The term 'lignans' should be retained for derivatives condensed by oxidative coupling of cinnamyl alcohol and/or cinnamic acid (20)

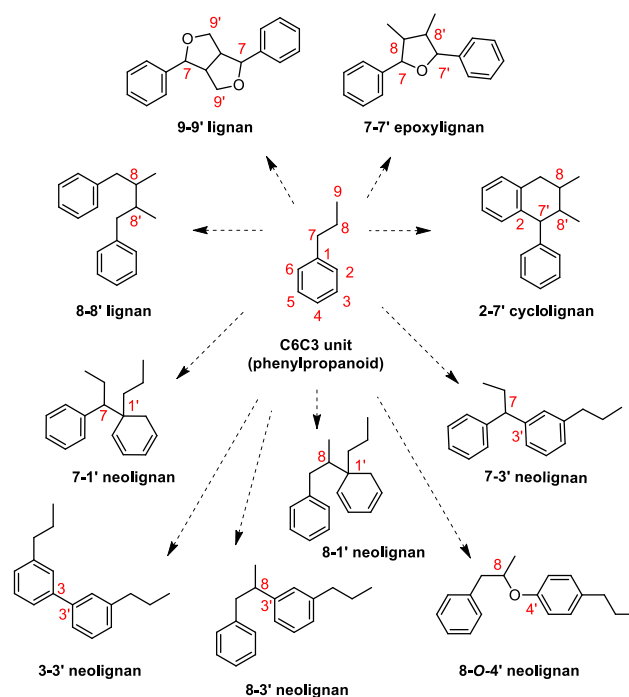


Fig. 7 C6C3 units of a phenylpropanoid core and example of the difference between lignans and neolignans based on the different potential linkages.

(8-8' linked). While neolignans are derived by condensation and oxidative coupling between propenylphenols and allylphenols (Fig. 8). Nevertheless, the recent lignan handbook published by Newman and collaborators (2022) suggested a rational need for an essential revision of phenylpropanoid (C6–C3) lignan and neolignan nomenclatures. They recommended that the original definition by Haworth could conveniently suit both lignan and neolignan chemical structures. Previously, Gottlieb (1984) had indicated a similar perspective as well, but at that time, the academic community, particularly organic synthetic chemists, were already employing the neolignoid nomenclature. Consequently, this current approach systematically expands the lignan nomenclature to encompass all potential structural lignoid inter-unit linkages, simplifying the current taxonomy while addressing relevant incongruences, such as those in C6–C3 trimers containing conventional "lignan" and "neolignan" linkages under the same lignoid core. Therefore, these lignoids had no clear classification before.^{95,96} In that context, the monomers that make up the lignoid core would rather not be the same with different linkage types, they do all come from the same biosynthetic pathway. The handbook strongly recommends discontinuing the use of the prefix 'neo-' in neolignans.^{95,96}

Despite agreeing with Newman's reclassification, this review retains the traditional lignan and neolignan division, as the general academic community in NP has not yet adopted the new (retro) nomenclature. Furthermore, most of the cited references in this review have cited and identified the compounds as either lignoids or neolignoids. Therefore, to prevent confusion by altering class names, we maintain the established nomenclature division. Nonetheless, we have chosen to cover the biosynthetic pathways of lignans and neolignans in a more integrative manner.

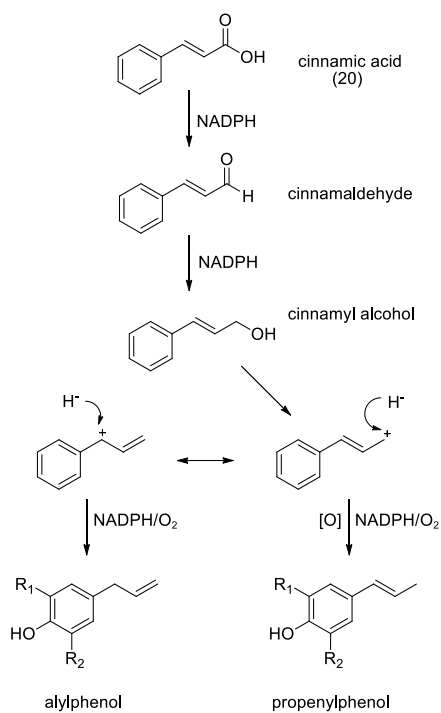


Fig. 8 C6-C3 units of neolignans formation via allylphenol and propenylphenol.

The study of these metabolites was extensive during the period from 1930 to 1980, leading to a well-established understanding of the main biosynthetic pathways for lignoids. Despite this foundation, gaps persist in the literature regarding the specific biosynthetic steps, particularly for the more complex neolignans, which exhibit a wide range of structural and stereochemical diversity. Lignoids are often optically active and display various substitution patterns, including different degrees of side-chain oxidation on their aromatic rings, attributable to their biosynthesis by stereoselective enzymes.^{48,53,89}

These compounds are synthesized in plants via the shikimate pathway, utilizing the amino acids *L*-Phenylalanine (*L*-Phe) and *L*-Tyr as precursors for the formation of phenylpropanoid units (C6-C3).^{32,53,89,91} The initial step in lignoid biogenesis involves a stereospecific, non-oxidative enzymatic reaction called deamination. This reaction removes ammonia from *L*-Phe to form the *E*-cinnamic acid (**20**), and from *L*-Tyr to the *p*-coumaric acid (**21**), mediated by the enzymes phenylalanine ammonia-lyase (PAL) and tyrosine ammonia-lyase (TAL) enzymes, respectively. Subsequently, the P450 enzyme cinnamate 4-hydroxylase enzyme (C4H), along with the cofactor NADPH and molecular oxygen, hydroxylates cinnamic acid by adding a hydroxyl group at the C4 (*para*) position of the aromatic ring. The resulting *p*-coumaric acid serves as a critical core in the phenylpropanoid pathway and has been isolated from *O. minarum*.^{53,89,97,98} Further biosynthetic steps yield catechols and organic acids, such as ferulic (**22**) and sinapic acids (**23**), and pave the way for the synthesis of more complex classes of compounds, including coumarins, flavonoids, and other lignoid cores (**Fig. 9**).

Following the formation of *p*-coumaric acid, the enzyme 4-coumaroyl CoA-ligase (4-CL) converts it into *p*-coumaroyl-CoA, initiating the cascade of biosynthetic reactions within the phenylpropanoid pathway. The primary precursors for lignoid biosynthesis are the monolignols, which include 4-hydroxycinnamyl alcohol (**24**) (*p*-coumaryl alcohol), the coniferyl

alcohol (**25**) and the sinapyl alcohol (**26**) (**Fig. 10**). An extensive range of different specific enzymes, many of which remain unidentified, catalyses the formation of these metabolites. This enzymatic arsenal includes *O*-methyltransferase (OMT), hydroxylases, lyases, oxidases, and synthases, among others. Additionally, isomerization reactions within different plant biosynthetic pathways can convert *E*-monolignols to *Z*-monolignols through the action of *E*→*Z* isomerases, further diversifying the structural possibilities of lignoids.^{53,89,99}

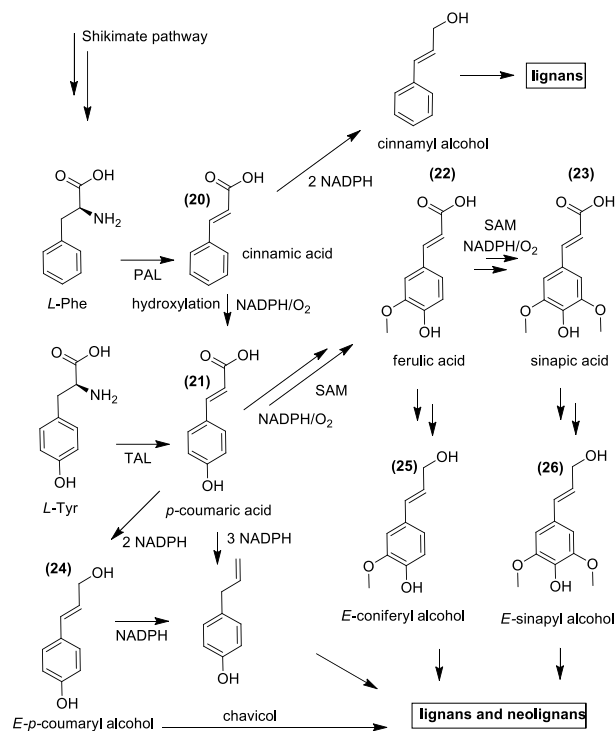


Fig. 9 The main monomers derived from shikimate pathway in the lignans and neolignans biosynthesis.

The biosynthesis of lignoids often results in enantiomerically pure metabolic products, attributed to the control exerted by stereoselective enzymes during the coupling reactions that form lignans and neolignans. Examples include the oxylignans (*S*)-virolongin B (**C7.2**), or the (*S*)-ococymosin (**C6.6**), both neolignans isolated from *O. cymosa*, showcasing the typical enantiomeric purity of these compounds. As a result, lignoids are predominantly found in optically active forms, similar to certain alkaloids. Regarding the lignan scaffolds, the 7-7' epoxy lignan (furan lignan – **C2**) and 9-9' diepoxy lignans (furofuran lignan – **C3**) are the most common lignan found in the *Ocotea* genus accounting together for 79.2% of all *Ocotea* isolated lignans reported in the literature.

The formation of these lignan structures depends critically on the synthesis of monomer **25** and their transformation through pathways that involve oxidases and dirigent proteins (DIR). The discovery of DIR, facilitated by kinetic studies at the beginning of this century, has been instrumental in unraveling the mechanisms behind the stereoselective biosynthesis of lignoids.^{53,90–92,100} The biosynthetic pathway of lignoids, following phenol oxidation, is initiated by the action of DIR, which serves as a stereoselective coupling enzymes. This process begins after the initial oxidation of monolignols, leading to the formation of radical species. A key intermediate in this pathway is the coniferyl alcohol radical (**26**) that gives rise to the (*8R,8R'*)-quinone methide (**27**), the secondary intermediate. This step is pivotal in the formation of complex

lignoid structures and exemplifies the importance of stereoselective enzymatic control in these biosynthetic processes.

Subsequently, DIR-mediated stereoselective intramolecular cyclisation leads to the formation of the furofuran lignan (+)-pinoselin (**28**), a 9-9' diepoxylignan. This reaction is catalyzed by the coupling process involving (+)-pinoselin synthase-DIR, which, depending on the specific enzyme variant, such as (-)-pinoselin synthase-DIR, can result in the product assuming an inverted configuration.^{53,89,91} To summarize, the biosynthetic reaction starts with one-electron oxidation at a phenol group of the *p*-coumaryl alcohols, leading to radical intermediates. At that point, phenol oxidation causes the delocalization of an unpaired electron by a resonance mechanism from one side-chain to another of the C6-C3 unit. So, four resonance structures can conjugate each other by radical pairing and then afford an array of possible dimeric or oligomeric reactive systems. Those are susceptible to nucleophilic attack from the intramolecular OH groups, or even from external nucleophiles, as represented in Fig. 10.^{53,97} Thus, monolignols can generate a variety of coupling radicals that are further modified and rearranged, originating a wide range of different lignoids and neolignoids, and thus a widespread chemical structural possibilities.^{53,89,97,101} For example, the formation of the furofuran syringaresinol (**29**) and yangambin (**30**), which are derived from precursor **28** through enzymatic hydroxylation and SAM/OMT methylation. The compounds **29** and **30** were described in the *O. duckei* and the *O. heterochroma*, respectively.

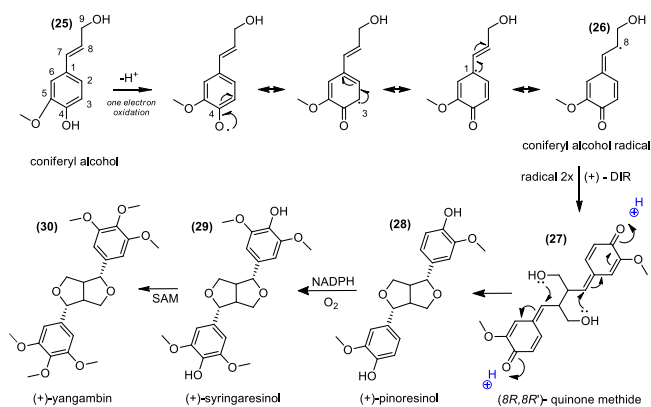


Fig. 10 Coniferyl alcohol radicals and DIR enantioselective formation of (+)-pinoselin and conversion to yangambin, via (+)-syringaresinol.

Moreover, the furan lignans, e.g. the 7-7' epoxyfuran lignans, are formed by the coupling of propenylphenols monomers units, which are coniferyl alcohol derivatives, such as the eugenol (**31**), isoeugenol (**32**) and any other potential methylated derivatives. The dimerization of the *E*-isoeugenol through radical coupling can give rise to the (+)-verrucosin (**33**). Similarly, its methoxylated monomers after radical coupling can generate the (+)-veraguensin (**34**) and feasibly the (+)-galgravin (**35**) furan lignans, which are stereoisomers (Fig. 11). Configurational aspects thus play a significant role in the lignoids biosynthesis, as different subclasses can incorporate more than four to five asymmetric carbons into the same scaffold, leading to a high level of complexity in the natural biosynthesis of these plants. Furan lignan cores **34** and **35** are the most commonly found in the *Ocotea* genus and have been isolated from *O. foetans*, *O. veraguensis*, *O. catharinensis*, and several other *Ocotea* species (Supplementary Table S3, available at Zenodo's link- <https://doi.org/10.5281/zenodo.1067>).

The biosynthesis of more oxygenated lignoids, though less common, also can be found in the *Ocotea* genus, and it is

potentially derived from the radicals of methoxylated monolignols precursors.^{97,101} Recent studies of *Ocimum basilicum* (sweet basil) metabolism have shed light on this process. These studies highlighted that NADPH-dependent reductases play a crucial role in catalysing the conversion of coniferyl acetate in chavicol and anol units of phenylpropene, such as eugenol and isoeugenol. The enzymes responsible for the catalysis are eugenol and isoeugenol synthases, respectively.^{97,102} (Fig. 11).

On the other hand, the neolignans scaffolds are abundant in

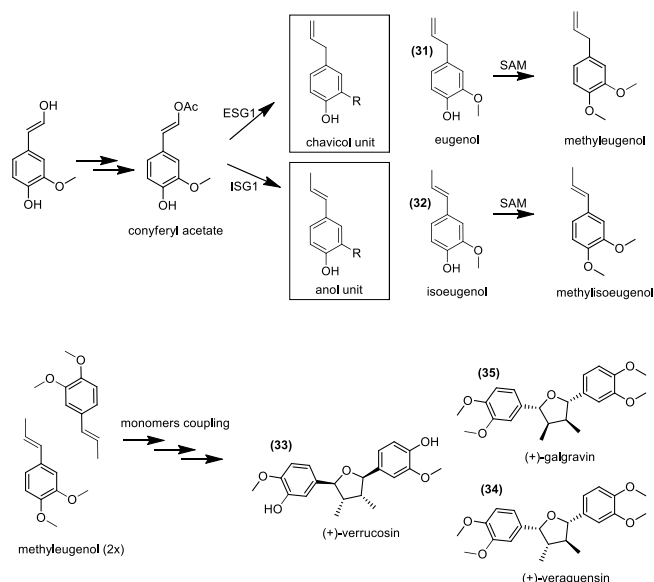


Fig. 11 The proposed galgravin and veraguensin biosynthesis via methyleugenol precursors.

Lauraceae, particularly in the *Ocotea* genus, where 82.2% of all lignoids are considered neolignoids (Supplementary Table S3). As already argued, radical coupling formation is the primary mechanism responsible for the oxidative dimeric formation of neolignans, offering a wide range of inter-unit linkage possibilities. This process depends on proton and electron abstraction from the phenylpropanoid units to generate propenylphenol, allylphenol, and monolignol radicals, which can be stabilized by their respective resonance effects. The chemical reactions are responsible for the different binding positions of the monomers, such as the 3-3', 7-1', 7-3', 8-1', 8-3', 8-*O*-4', 3-*O*-4', 2-*O*-3', and other possible linkages occur through the Michael addition reaction, via carbanion formation and resonance between the generated enolate ions. Thus, a carbon-carbon bond is created at the acceptor's carbon of the coupling oxidated radical monomer. This can occur via Diels-Alder's concerted mechanism for a range of other neolignoids scaffolds. Additionally, further combinations of secondary aromatic rings arise through the Claisen condensation mechanism coupled with Friedel-Crafts reactions for aromatic ring alkylation, allowing a wide variety of stereoselective and complex lignoid scaffolds.^{53,96,103}

However, even though lignoids display diversity, within the *Ocotea* genus, more than one-third (36.3%) are bicyclic neolignoids, making this subclass the most representative lignoid subclass in the *Ocotea* species. These compounds have been identified in different species such as *O. aciphylla*, *O. bullata*, *O. catharinensis*, *O. veraguensis*, and *O. porosa*, among others.^{26,104-108} Following closely, the benzofuran subclass accounts for 32.87% of the isolated lignoids, found in species such as *O. catharinensis*, *O. veraguensis*, *O. porosa*, *O. macrophylla* and other *Ocotea*

species (Supplementary Table S3, available at Zenodo's link-<https://doi.org/10.5281/zenodo.1067>). The high prevalence of these two subclasses within the *Ocotea* spp. is not merely coincidental. Back to the biosynthesis aspects, benzofuran neolignans, particularly those with 8-1' and 8-3' linkages, are considered key precursors in the biosynthetic pathway before rearrangement into the bicyclic [3.2.1] octane neolignoids. Fig. 12 illustrates the acid-catalyzed rearrangement of the 8-1' linkage to form the classic bicyclic neolignoids scaffold. Additionally, the reverse pathway back to the benzofuran core is also feasible, allowing for further biochemical modifications within the plant biosynthesis.^{96,109,110}

4.2 *Ocotea* spp. lignoids profile

Over the past century, lignoids have been extensively studied for their role as phytochemical biomarkers and through various

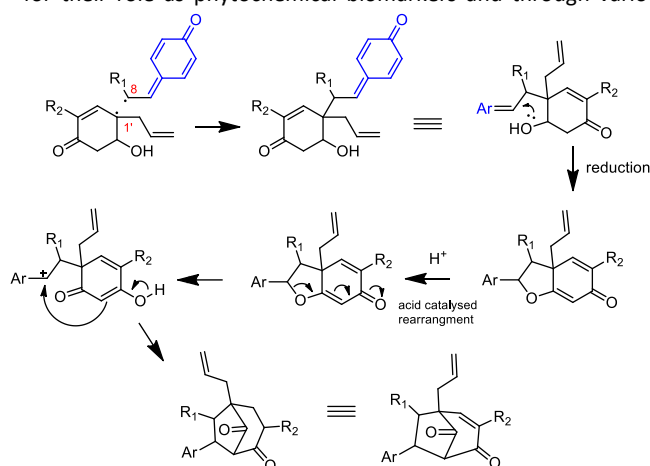


Fig. 12 The proposed bicyclic [3.2.1] octane neolignoids via 8-1' linkage radical monomers based on acid catalyzed rearrangement mechanism.

systematic approaches, prominently by Otto Gottlieb and collaborators from 1970 and 2000.^{10,104,110,111} More recently, a metabolomic study using only literature data by Antonio *et al.* (2020) shown a shift in academic interest towards alkaloids in the 21st century, with lignoids receiving less focus. Nevertheless, besides their research limitation regarding acquisition data for metabolomics models, and even the limited lignoids available data in the literature (small number of *Ocotea* sp. available), their multivariate statistical analysis corroborated that *Ocotea* lignoids are a valuable piece of information to the field of chemophenetics research. In addition, their work allowed the comprehension that Gottlieb's chemophenetic statement that *Ocotea* species exclusively biosynthesize either alkaloids or lignoids is no longer valid, as modern chemical analytical technics evidenced that different *Ocotea* sp. can elicit both alkaloids and lignoids, *e.g.* as the case of *O. macrophylla*, *O. duckei* and *O. minarum*.³²

In this context, the study of metabolomic features through untargeted metabolomics strategies has the potential enhance plant taxonomic issues in the genus. Species differentiation could be rapidly achieved using metabolic fingerprints obtained from state-of-the-art analytical tools like Ultra Performance Liquid Chromatography coupled with High-Resolution Mass Spectrometry (UPLC-HRMS). Antonio *et al.* (2020) also demonstrated that comparing lignoid content is an effective method for distinguishing similar *Ocotea* species due to the significant variation in their lignoid profiles. Diepoxylignans, such as compound **30**, have proven to be effective phylogenetic

biomarkers within the genus. Continued application of untargeted metabolomics may uncover new markers, aiding taxonomic classification, as these methods are already widely used in drug and phytomedicine quality control.^{32,112–114}

The increasing focus on the chemical data and bioactivity of lignoids over the last decade suggests that these compounds may soon regain broader interest in the academic community. Original articles accessing chemical data and the bioactivity of lignoids have been increasing in the last 10 years. Noteworthy, pharmacological effects have been reported in approximately 20% of the chemically characterized *Ocotea* species. To date, 172 different lignoid chemical structures have been isolated from the genus (Supplementary Table S3, available at Zenodo's link-<https://doi.org/10.5281/zenodo.1067>). Fig. 13-18 describe the various lignoid cores found among the different *Ocotea* lignoid producers reported in the literature.

The diverse lignoid profile of *Ocotea* results from complex enzymatic reactions within specialised biosynthetic pathways. Recent studies have shown that *Ocotea* species-producing lignoids have a high potential for treating neglected tropical diseases (NTDs) such as dengue fever and leishmaniasis, which affect over 8 million people in Latin America. Accordingly, potent larvicidal and antiprotozoal effects were reported for different lignoids isolated from plants of the *Ocotea* genus. For example, neolignans isolated from *O. cymosa* have shown high larvicidal potential, halting the *Aedes aegypti* life cycle at the larval stage with 100% mortality.^{99,115} The yangambin lignan (**C3**) and the licarin A neolignan (**C6.6**) from *O. macrophylla* have exhibited low IC₅₀ values for amastigote and promastigote forms of different *Leishmania* (L.) subspecies. In addition, the lignoid **30** (Fig. 9) and burchelin (**C6.3**) were also reported in the literature as highly cytotoxic agents against trypomastigote forms of *Trypanosoma cruzi*.^{101,116–119}

Moreover, lignoids are recognized as pertinent bioactive compounds with notable pharmacological properties, including pronounced antitumoral, antiviral, and anti-inflammatory effects.^{6,48,92,120,121} For instance, the antitumor activity of the furofuran lignan (**C3**), the aryltetralin lactone (**C4**), and the cyclolignan (**C5**) have been reported as active against a broad spectrum of cancer cell lines, including human lung adenocarcinoma (A-549), human colon adenocarcinoma (HT-29), human breast adenocarcinoma (MCF-7), and murine lymphocytic leukemia (P-388).⁹⁰ Regarding antiviral activities, different classes of lignoids have shown to be effective against hepatitis B virus (HBV), human cytomegalovirus (HCMV), human immunodeficiency virus (HIV), SARS-virus (SARS-CoV) and Zika virus (ZIKV).^{92,120} Additionally, diastereomeric lignans isolated from the *O. macrophylla* exhibited PAF-antagonism and potent dual inhibition of COX-2/5-LOX pathways.^{6,108} A study of our research group confirms potent *in vivo* anti-inflammatory activity by dual inhibition of the PGE2 production and neutrophil recruitment of the bicyclo [3.2.1] octaneneolignans (**C9**). Although, they were isolated from *Aniba firmula*, which also belongs to Lauraceae.^{121–123} The derivatives of the bicyclo [3.2.1] octaneneolignans represent 49.3% of the *Ocotea* known lignoids in the literature, which are encountered in 64.7% of the *Ocotea* sp. producer, *e.g.* the *O. macrophylla*, *O. catharinensis*, *O. porosa*, *O. cymosa*, and others.^{94,111,121} (Supplementary Table S3, available at Zenodo's link-<https://doi.org/10.5281/zenodo.1067>). Furthermore, the bicyclic neolignoid silyllenone (**C9.1b**), isolated from *O. bullata*, has shown promising anti-inflammatory activity through inhibition of 5-LOX enzymes.²⁶

Lastly, the comprehensive review of Teppone *et al.* (2016) reported interesting data on recent advances in the lignoids pattern biosynthesis, detailing a classification into ten lignan core subtypes and fifteen neolignan subtypes.⁸⁹ Our review highlights that the *Ocotea* genus encompasses 5 of these 10 lignan subtypes: 8-8' lignans (**C1**); 7-7' epoxy lignans (**C2**), 9-9' diepoxy lignans (**C3**), 8,8',2,7' aryltetralin lignan lactones (**C4**), and 2-7' cyclolignans (**C5**). Additionally, the genus hosts 9 different neolignan core subtypes, including benzofuran neolignans (**C6**), oxyneolignans (**C7**), 8-1'/8-3'/other linkage neolignoids (**C8**), along with bicycle [3.2.1] octaneneolignans (**C9**). Thus, alongside its alkaloids, *Ocotea* is also a remarkably diverse source of lignans and neolignans, offering significant variations in scaffold structures (Fig. 13-18).

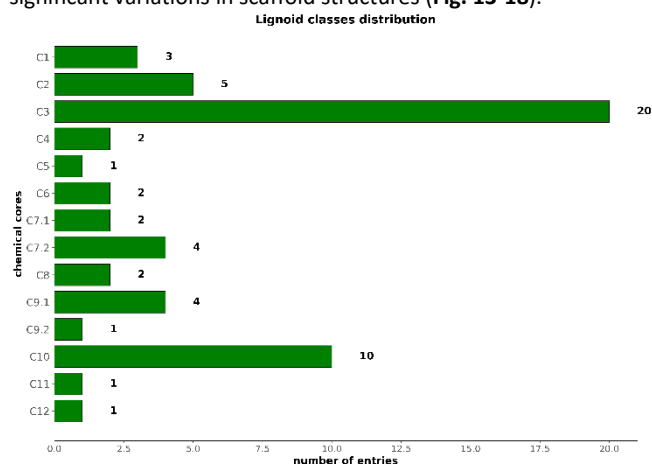


Fig. 13 Chemical diversity of lignoids in the *Ocotea* genus expressed by the number of different chemical structures reported in literature.

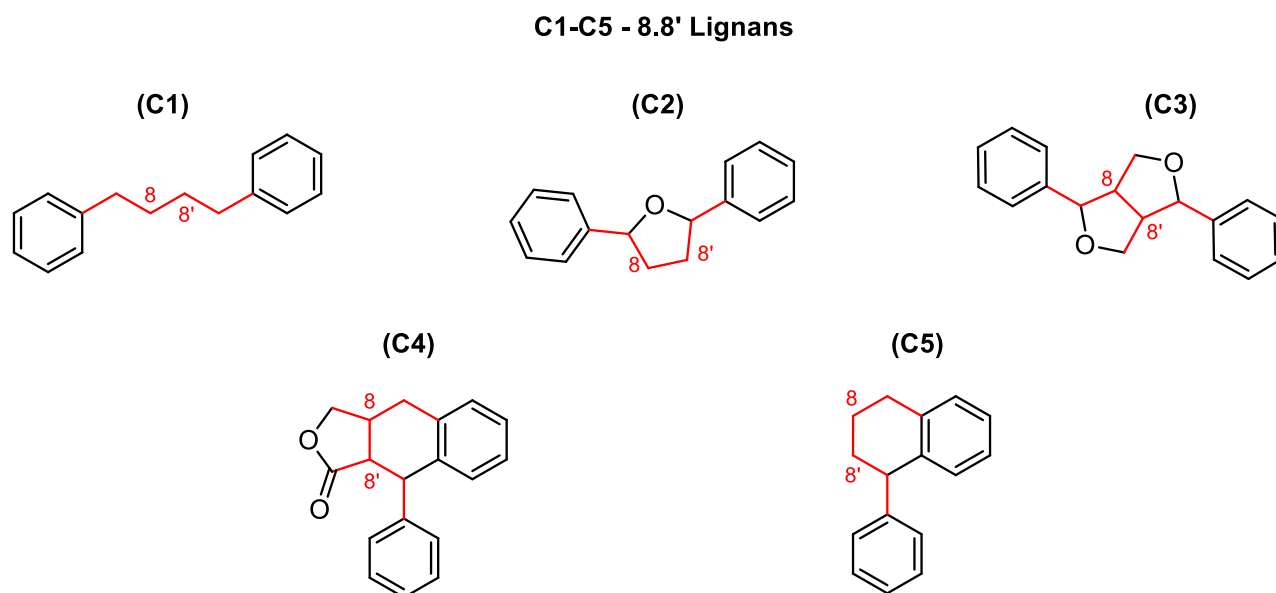


Fig. 14 Lignans chemical scaffolds found in *Ocotea* species: (C1) 8-8'-lignans, (C2) 7-7' epoxy lignans, (C3) 9-9' diepoxy lignans, (C4) 8,8',2,7' aryltetralin lignan lactones, and (C5) 2-7' cyclolignans. Molecular diversity of lignans in the genus *Ocotea* expressed by the number of different chemical structures reported in the literature for each subclass Lignans chemical scaffolds.

(C6) - 8.1' / 8.3' benzofuran neolignans

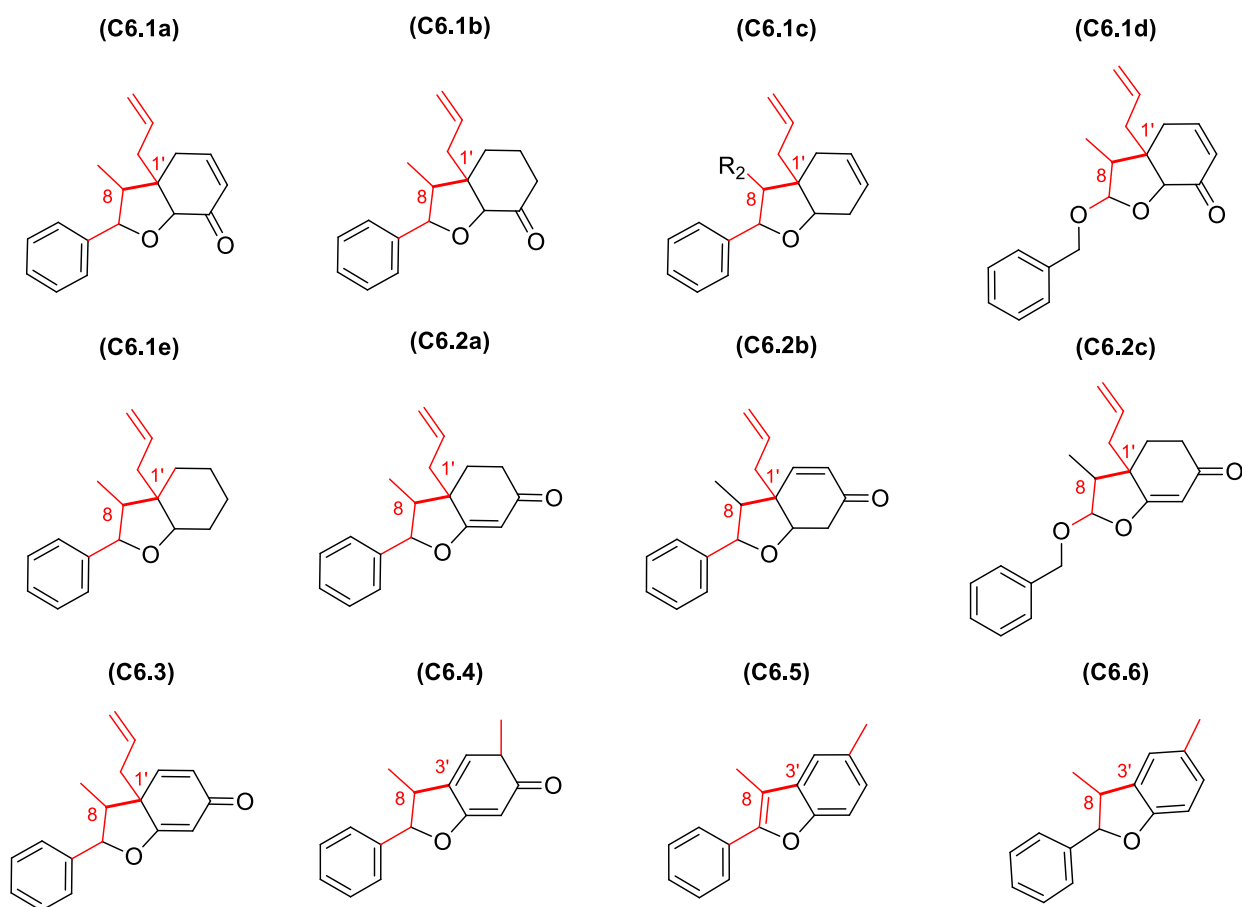


Fig. 15 Benzofuran neolignans chemical scaffolds found in *Ocotea* species: (6.1a) ferrearin type I, (6.1b) ferrearin dehydro type, (6.1c) ferrearin type II, (6.1d) ferrearin type III, (6.2a) porosin type I, (6.2b) porosin type II, (6.2c) porosin type III, (6.3) burchelin, (6.4) alpha-beta ketone type, (6.5) ocophyllal type, and (6.6) ocozymosin type.

(C7) - 7.3', 8.4' / 8.4' oxyneolignans



Fig. 16 Oxyneolignans chemical cores found in *Ocotea* species: (C7.1) dioxyneolignan and (C7.2) oxyneolignan.

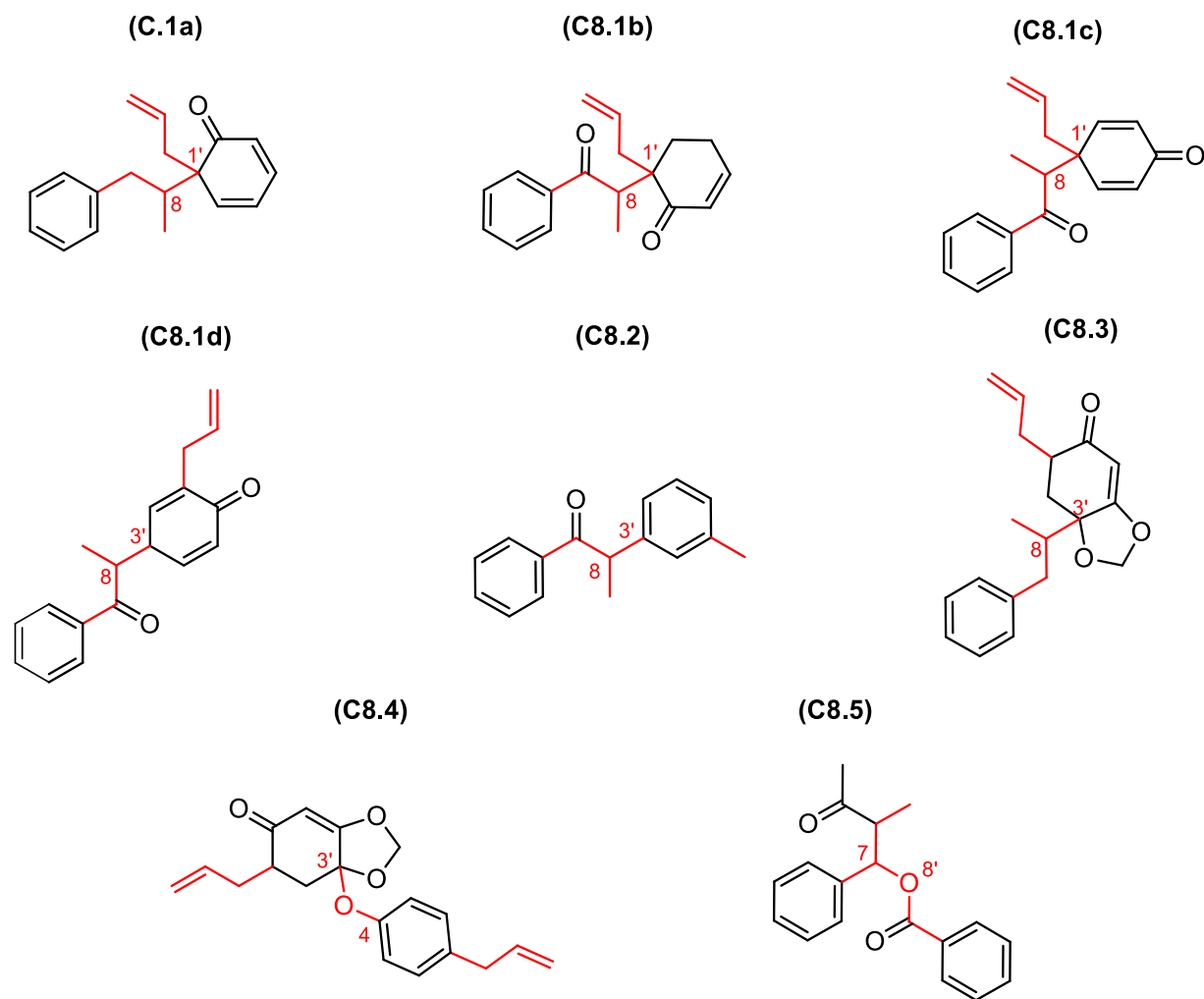
(C8) - 8.1' / 8.3' / 4.3' / 7.8 neolignans

Fig. 17 Lignoids chemical scaffolds found in *Ocotea* species: (C8.1a) 8.1' neolignan I, (C8.1b) 8.1' neolignan II, (C8.2) 8'-oxo-1' neolignan I, (C8.3) 8.3' neolignan I, (C8.4) 4.3'- didymochlaenone, and (C8.5) 7.8- secolignan.

(C9) bicyclo [3.2.1]octaneneolignans

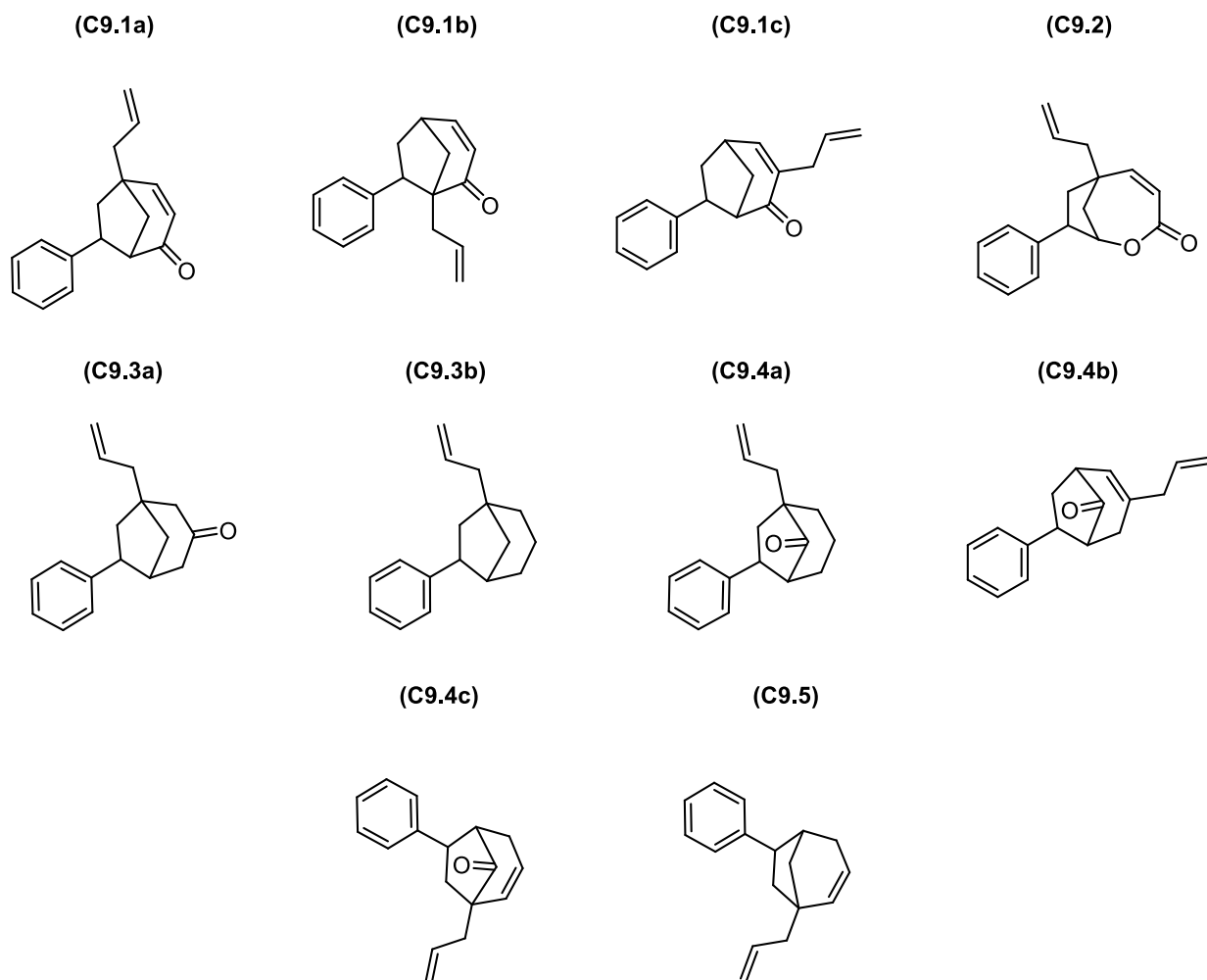


Fig. 18 Bicyclo neolignans chemical cores found in *Ocotea* species: (C9.1a) 7.3',8.1' cycloneolignan, (C9.1b) 7.1',8.3' cycloneolignan, (C9.1c) 7.3',8.5' cycloneolignan, (C9.2) 7.3',8.1' oxaguianin type, (C9.3a) 7.3',8.1' canellin type, (C9.3b) 7.3',8.1' canellin type II, (C9.4a) 7.3',8.1' cycloneolignan ketone type I, (C9.4b) 7.3',8.1' cycloneolignan ketone type II, (C9.4c) 7.3',8.1' cycloneolignan ketone type III, and (C9.5) 7.1',8.3' oxycycloneolignan.

5. *Ocotea* sp. metabolites: flavonoids

5.1 *Ocotea* sp. flavonoids biosynthetic approach

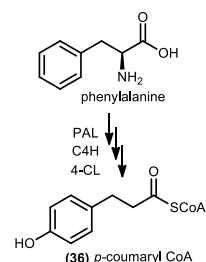
Flavonoids play a crucial role in plant ecology, acting as defense agents against several environmental and biotic stresses such as ultraviolet radiation, pests, and diseases. Likewise, they are integral in regulating plant growth and development by modulating different plant hormones, thus also playing a pivotal role in plant evolution.^{124–126} In addition, they are known for their beneficial effects on human health, including antioxidant, anti-inflammatory, anti-cancer, and anti-viral properties.^{126–129} These diverse groups of naturally occurring polyphenolic compounds are prevalent in the plant kingdom, including the Lauraceae and the *Ocotea* genus.^{6,32,124,130} In this review, flavonoids were identified as the third most commonly isolated metabolite class in the *Ocotea* genus reported in the literature, with 19 species currently recognized to produce these polyhydroxylated aromatic compounds. In our recent publication using modern annotation analytical techniques, this number was highly increased up to 59 *Ocotea* species able to biosynthesize flavonoids.²⁷ (Supplementary Table S4 and S4.1, available at Zenodo's link- <https://doi.org/10.5281/zenodo.1067>).

The study of flavonoid biosynthesis and their chemical diversity contributes significantly to the understanding of plant evolution. The diversified chemical structures of flavonoids across different plant species reflect changes in the genetic sequences of their biosynthetic enzymes, which may allow plants to adapt to varying environmental conditions.^{131,132} In particular, the analysis of flavonoid chemical cores can aid in taxonomic classification and fill potential phylogenetic gaps. By identifying specific flavonoid subclasses, investigations in the fields of chemophenetic, metabolomics, and transcriptomics can be implemented to elucidate the relationships between different plant genera.^{32,53,124,133} In the *Ocotea* genus, known to produce a wide variety of flavonoids, there is a potential for identifying specific biomarkers aligning with phytochemistry strategies guided through metabolomics studies.^{32,79,124,134}

The flavonoids are chemically characterized by a general structure of a 15-carbon skeleton, that consists of two phenyl rings (A and B) and a heterocyclic ring (C). Biosynthetically, the carbon chain is named as C6-C3-C6 unit. Significant efforts have been made by the scientific community to elucidate the biosynthetic pathways of these bioactive polyphenolic metabolites.^{53,124,127} Previous studies have shown that flavonoid biosynthesis begins with the shikimate pathway, although the mevalonate pathway also plays a crucial role in producing different flavonoid subclasses. Such molecules can exhibit chirality and interesting bioactivity, and thus, the study of enzymatic stereospecificity is closely associated with the chemical study of these specialised metabolites.^{53,124,127}

The first step of flavonoid biosynthesis is the same as lignoids and includes the conversion of the *L*-Phe / *L*-Tyr until the formation of the *p*-coumaroyl-CoA precursor (**36**). The next step of flavonoid biosynthesis consists of the catalysis by the chalcone synthase (CHS) enzymes and the combination with mevalonate pathway precursors. Thus, the *p*-coumaroyl-CoA is extended with three units of the malonyl-CoA precursors. The final poly- β -keto chain, is a tetrahydrochalcone, which undergoes Claisen cyclisation to produce the aromatic ring A, leading to the formation of the naringenin-chalcone precursor (**37**) (Fig. 19). The following cyclisation is responsible for the six-membered heterocyclic ring of flavonoids.^{53,124}

SHIKIMATE PATHWAY



ACETATE PATHWAY

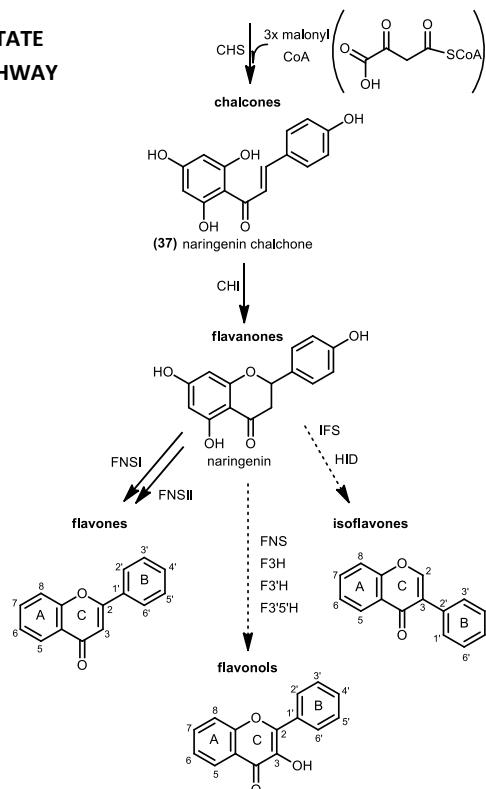


Fig. 19 The general flavones, isoflavones and flavonols biosynthesis.

The subsequent transformation involves the activity of flavanone isomerase (CHI) enzymes, catalyzing a path via intramolecular Michael-type nucleophilic attack. Specifically, a phenol group acts as the nucleophile, attacking the α,β -unsaturated ketone to form the heterocycle C ring and give rise to the naringenin precursor (**38**). The aromatic ring B originates from the shikimate pathway, while ring A is derived from a polyketide origin, and the heterocycle C is formed by nucleophilic addition. This union of rings A, B, and C forms the stereospecific flavanone core, which is the primary precursor of the flavones, isoflavones, and flavonol subclasses, with the latter being most prevalent in the *Ocotea* genus, exemplified by derivatives such as kaempferol and quercetin, and also less present myricetin derivatives.^{53,127,130,135} In addition to the core modifications, further diversification of flavonoid chemical diversity is driven by various enzymatic groups. Flavone synthases (FNS), isoflavone synthases (IFS), flavonol hydroxylases (F3H, F3'H, and F3'5'H), dehydratases (HID), along with other isomerases and reductases, facilitate the formation of a vast array of analogues including flavones, flavan-3-ols, anthocyanidins, catechins, and glycoside flavonoids. Moreover, other enzyme groups of glycosyltransferases (GTs), methyltransferases, and acyltransferases can perform additional

modifications on the flavonoid chemical scaffolds, leading to an assorted range of chemical structure derivatives.^{53,124,127} For example, flavonolignans represent a hybrid class of NP biosynthesized through the oxidative coupling of taxifolin (5,7,3',4'-flavan-on-ol) with a phenylpropanoid, typically coniferyl alcohol. Recently, a flavonolignan was isolated for the first time within the *Ocotea* genus, specifically, from *O. diospyrifolia*.⁴⁶

5.2 *Ocotea* spp. flavonoids profile

Flavonoids are specialised metabolites characterized by their structure of two benzene rings linked by a heterocyclic pyran ring, which can undergo a wide variety of chemical modifications. These include substitutions with hydroxyl, methoxy, or a range of different glycosyl groups, which significantly affect their physical and chemical properties, as well as their biological activity.^{127,135,136}

Furthermore, in the *Ocotea* genus, flavonoids are predominantly derivatives of catechin, epicatechin, quercetin, and kaempferol.^{32,130} Despite growing research interest, knowledge of *Ocotea* flavonoid profiles remains limited.¹³⁰ Still, 56 distinct flavonoids were reported, of which ~14.3% are non-substituted flavonoid cores (n= 8) and ~85.7% (n= 48) are glycosylated. For the glycosylated, ~91.6% (n= 44) are *O*- glycosylated, and ~8.4% (n= 4) are *C*-glycosylated derivatives. For example, *O*-alkyl flavonoids were found only in *O. porosa*, while *C*-glycosylated were already found in four other *Ocotea* sp., including *O. aciphylla*, *O. foentans*, *O. nutans* and *O. odorifera*. Additionally, a recent comprehensive UPLC-HRMS flavonoid profiling from our research indicated the presence of apigenin flavone backbone in other 6 *Ocotea* species including *O. diospyrifolia*, *O. guianensis*, *O. lancifolia*, *O. notata*, *O. odorifera* and *O. porosa*. The apigenin derivatives were found mainly in *C*-glycoside form, with *O. porosa* exhibiting the highest flavone : flavonol ratio.¹³⁰ Additionally, This distribution pattern corroborates that *Ocotea* could occupy either a basal or intermediate position in the evolutionary lineage of the Lauraceae, as previously pointed out by Antonio and collaborators (2020).³² This is supported by the low occurrence of *O*-alkylated flavonoids, which are more commonly found in evolutionarily advanced plant groups.³² However, as the evolutionary studies become more accurate and precise with a larger set of species evaluated, additional studies using DNA and morphological data could provide further evidence to support or refute that hypothesis.³⁷

Flavonoids are not only crucial from an evolutionary and phylogenetic perspective but also noted for their diverse bioactivities. One of the most common flavonoid cores found in *Ocotea* species is the 3-*O*-glycoside flavonol (**core 3**), with at stands for isoquercitrin, a representative example, which was already isolated from *O. notata*, *O. elegans*, *O. corymbosa*, and *O. caudata*. Isoquercitrin has been shown to possess antibacterial activity against *E. coli* cells, inducing apoptosis-like death and damaging membrane dynamics by inducing oxidative stress.¹³⁷ This underscores the potential of *Ocotea* flavonoid's bioactivity, which warrants further exploration of their biological properties. Flavonoids have also been increasingly recognized for their antioxidant and pro-oxidant properties, as well as for relevant antiviral and antiprotozoal effects.^{32,124,126,127,138} A series of flavonols can be used to seek more effective treatment for NTDs, such as leishmaniasis and trypanosomiasis.^{6,126} Also, *Ocotea* species, such as *O. minarum* and *O. odorifera*, have been recognized as potent natural antioxidants.^{98,136} Antiviral effects against herpes virus (HSV-1 and 2) were reported for flavonoids of the *O. notata* leaves extract.⁶ Recently, bacteriostatic effects were

evidenced for *O. minarum* flavonoid fractions against the *Salmonella*, the *Pseudomonas* and the *Proteus* genera.⁹⁸

Flavonoids also exhibit other interesting biological activities, and low toxicity for human cells, and thus are considered safe substances for consumption. Some flavonoids, such as biflavonoids from *Caesalpinia pluviosa* (Fabaceae), have shown notable cytotoxic activity against various tumor cell lines.^{126–129,139} More specifically, the biflavonoid caesalpinioflavone, extracted from the stem bark of *C. pluviosa*, was able to reduce the viability of tumor cell lines, including A549, MCF-7, Hst578T, and HTC. The mechanism of action involved cell cycle arrest at the G1/S transition at least in A549 and MCF-7 cells.¹³⁹ For instance, biflavonoids from *O. odorifera* and *O. caniculata* (**core C10**) might also hold promise for cytotoxic activity and might be worthy of further investigation.^{32,136}

This extensive chemical diversity of flavonoids underscores their potential contribution to our understanding of plant evolution and adaptive metabolism. Therefore, the development of flavonoid-based drug candidates is of great interest due to the broad range of biological activities they exhibit, such as antioxidant, anti-inflammatory, and antitumor, besides the low cytotoxicity effects for human cells. As chemical data accumulates, future studies could delineate specific flavonoids within *Ocotea* that might serve therapeutic purposes, potentially leading to novel drug prototypes. Moreover, they might also be used as bioactive markers for taxonomy together with the development of a more comprehensive understanding of their chemical diversity.

In this context, a broad diversity of specific flavonoid glycoside derivatives were isolated from several *Ocotea* species, such as *O. vellosiana*, *O. odorifera*, *O. porosa*, *O. notata*, *O. lancifolia* and others.^{6,32,127,135} Fig. 20–21 describes all the different flavonoid cores encountered in the genus. It includes a wide number of flavonoids subclasses, including, flavone (**core 1**), flavonol (3-hydroxy flavone, **core 2**), 3-*O*-glycoside flavonol (**core 3**), 7-*O*-glycoside flavone-flavonol (**core 4**), flavanonol (**core 5**). Also, the 3-*O*-glycoside flavanonol (**core 6**), flavan-3-ol (**core 7.1**), phenylpropanoid-substituted flavan-3-ol, (**core 7.2**), 7-*O*-glycoside flavanone (**core 8**), 8-glycoside flavone (**core 9.1**), 6-glycoside flavone (**core 9.2**), flavanonol dimer or biflavonoids (**core 10**), tetrahydrochalcones (**core 11**) and flavonolignan (**core 12**).

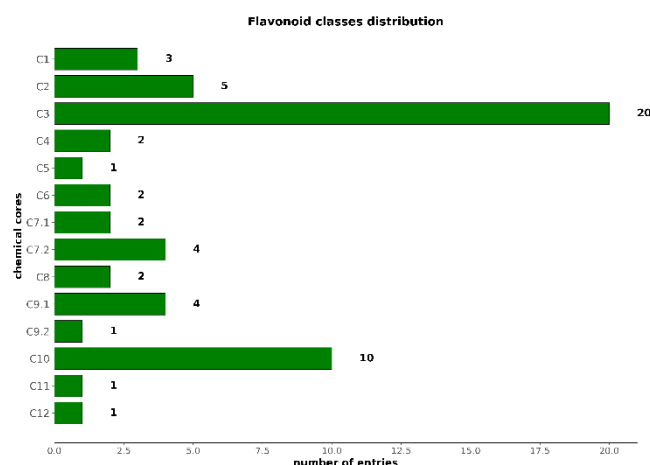


Fig. 20 Chemical diversity of flavonoids in the *Ocotea* genus expressed by the number of different chemical structures reported in literature.

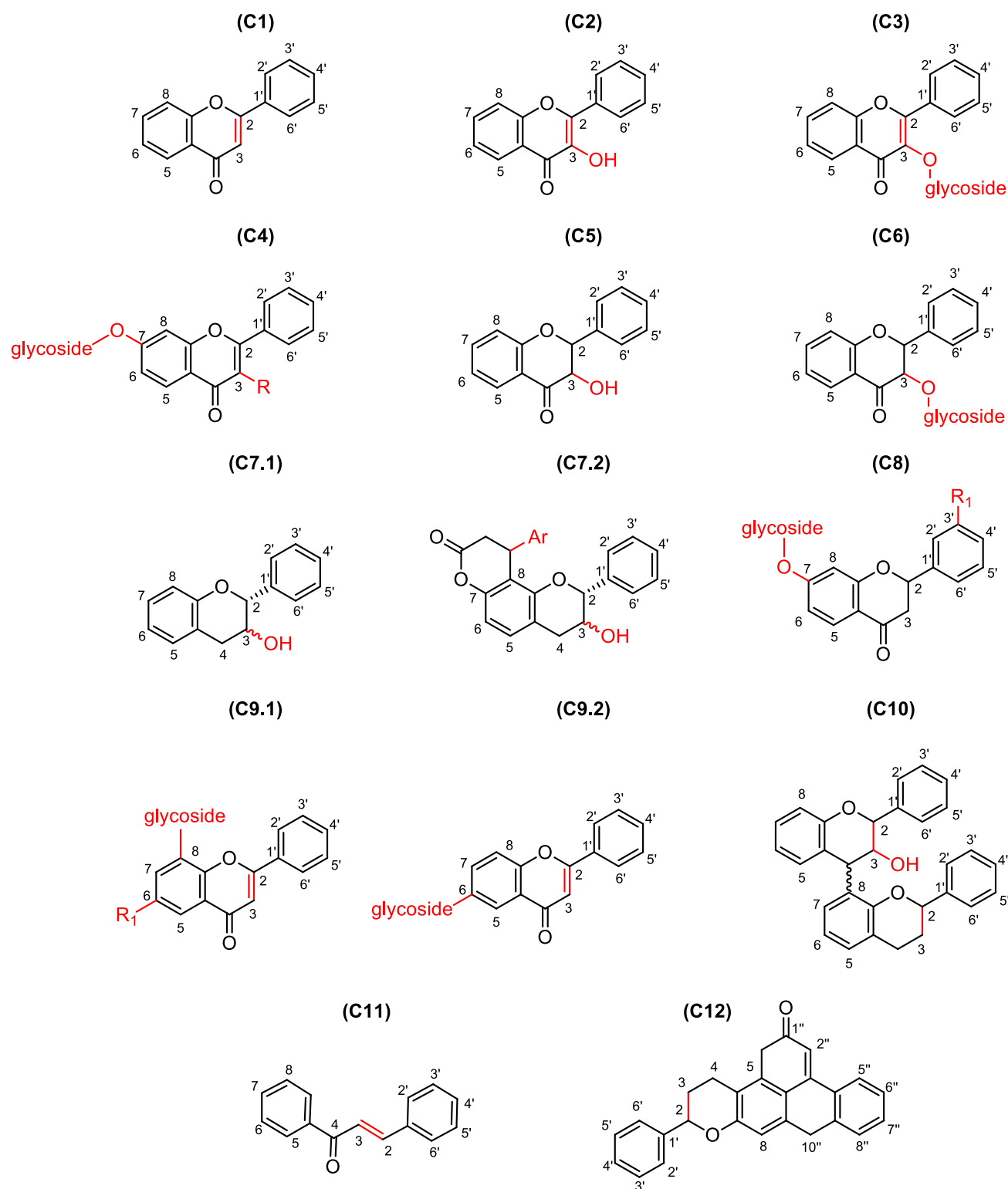


Fig. 21 Flavonoids cores found in *Ocotea* species: (C1) flavone, (C2) flavonol, (C3) 3-*O*-glycoside flavonol, (C4) 7-*O*-glycoside flavonol, (C5) flavanone, (C6) 3-*O*-glycoside flavanone, (C7.1) flavan-3-ol, (C7.2) phenylpropanoid-substituted flavan-3-ol, (8) 7-*O*-glycoside flavanone, (C9.1) 8-glycoside flavone (C9.2) 6-glycoside flavone, (C10) flavanone dimer or biflavonoid, (C11*) tetrahydroxychalcone and (C12#) flavonolignan *Flavonoid precursor; # Flavonoid mixed biosynthetic class.

6 *Ocotea* sp. metabolites: terpenoids

6.1 *Ocotea* spp. terpenoids biosynthetic approach

Terpenes and terpenoids represent the largest class of chemicals found in the essential oils of *Ocotea* species, followed by phenyl compounds such as phenylpropanoids and phenylpropene derivatives, along with aldehydes, alcohols, and esters. Indeed, volatile terpenes are the predominant constituents in the essential oils of a vast array of plants and flowers globally.^{140–142} These compounds are biosynthesized predominantly through two pathways: the mevalonic acid (MVA) (**39**) and the 2C-methyl-*D*-erythritol-4-phosphate (MEP) (**40**) pathways. The biosynthesis begins with the isoprene-like C₅ building block, which can be assembled in a head-to-tail fashion or through a central tail-to-tail linkage. These building blocks undergo further modifications such as cyclisation, polymerization, oxidation, and dimerization, allowing for extensive structural diversity.^{53,143}

The terpenes subclasses can be classified according to the number of isoprene units as hemiterpenes (C₅), monoterpenes (C₁₀), sesquiterpenes (C₁₅), diterpenes (C₂₀), sesterterpenes (C₂₅), triterpenes (C₃₀), and tetraterpenes (C₄₀). Even though the isoprene unit is naturally produced, essentially it is not involved in the terpenes formation, instead, the biochemically active isoprene units are the isopentenyl diphosphate (IPP) (**41**) and the dimethylallyl diphosphate (DMAPP) (**42**).^{53,144,145} In plants, the MVA and the MEP pathways operate in different cellular compartments within photosynthetic tissue. The cytosol is the site of activity for the MVA pathway, while the MEP pathway occurs within the plastids. Different classes of terpenoids are synthesized via each pathway: sesquiterpenes, sterols, and triterpenes typically arise from the MVA pathway, whereas monoterpenes and diterpenes are generally products of the MEP pathway.^{53,143,146} *Ocotea* species produce majoritarian sesquiterpenes (via MVA) and monoterpenes (via MEP), indicating both pathways are highly evolved and active within the *Ocotea* metabolome.

In detail, the MVA pathway begins with the enzyme acetoacetyl-CoA synthase (E1) combining two acetyl-CoA units via a Claisen condensation to form acetoacetyl-CoA. This precursor is then converted by hydroxymethylglutaryl-CoA synthase (E2), which adds a third acetyl-CoA molecule to produce 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA). The thioester bond of HMG-CoA is reduced by HMG-CoA reductase (E3), along with 2 NADPH + 2H⁺, transforming it into MVA. This acid is then sequentially phosphorylated by ATP-dependent mevalonate kinase (E4), phosphomevalonate kinase (E5), and mevalonate 5-diphosphate decarboxylase (E6) to yield IPP. Finally, IPP isomerase (E7) catalyzes the conversion of IPP to DMAPP (Fig. 22a).^{53,147}

The MEP pathway starts with the condensation of the glyceraldehyde-3-phosphate and the pyruvate catalyzed by DXP synthase (E8)/thiamine pyrophosphate cofactor (TPP). This reaction produces the 5-carbon intermediate, 1-deoxy-*D*-xylulose-5-phosphate (DXP) (**43**). The activation mechanism involves the decarboxylation of pyruvic acid to form an acetaldehyde-enamine intermediate. This intermediate allows the glyceraldehyde 3-phosphate unit to undergo a nucleophilic attack by the enamine. Subsequently, DXP is converted MEP through a pinacol-like rearrangement, followed by a NADPH-dependent reduction catalyzed by 1-deoxy-*D*-xylulose-5-phosphate isomerase (E9). At this point, MEP reacts with the cytidine triphosphate (CTP) followed by phosphorylation via the ATP-dependent kinase 4-diphosphocytidyl-2C-methyl-*D*-erythritol (E10) and cyclisation with

loss of CTP by 2C-methyl-*D*-erythritol-2,4-cyclodiphosphate synthase (E11), forming the the cMEPP intermediate. A two-electron reduction by 1-hydroxy-2-methyl-2-(*E*)-butenyl-4-diphosphate synthase (E12) produces 4-hydroxy-3-methyl-but-2-enyl diphosphate, which is subsequently converted into IPP and DMAPP by 4-hydroxy-3-methyl-2-(*E*)-butenyl-4-diphosphate reductase (E13) (Fig. 22b).^{53,147}

Thus, the IPP and DMAPP are the primary building blocks for

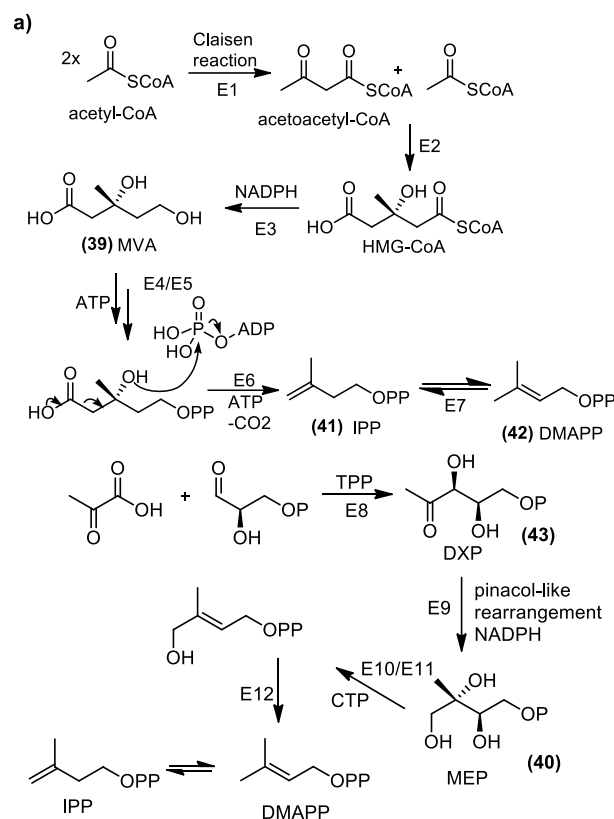


Fig. 22 Biosynthetic pathway of IPP and DMAPP via a) MVA and b) MEP formation.

terpenes formation, which can be further modified generating a broad and diverse number of structures with significant biological properties. The IPP represents a nucleophilic isoprene unit, while the DMAPP are electrophilic specie. Thus, these two groups commonly assemble in a head-to-tail fashion that is catalyzed by GPP synthase (E14). The geranyl pyrophosphate (GPP) (**44**) forms the backbone of most monoterpenoids (C₁₀), such as geraniol (**45**), citronellol (**46**), linalool (**47**), and cyclic terpenoids like camphor (**48**), borneol (**49**), limonene (**50**), terpineol (**51**), and 1,8-cineol (**52**), which all are widespread in the *Ocotea* genus.^{53,145,148} Monoterpenoid derivatives result from a series of enzymatic modifications, undergoing various oxidation and reduction reactions. For example, linalool synthase (LIS) acts on linalool pyrophosphate (LPP) (**53**), producing linalool (**47**) (Fig. 23).

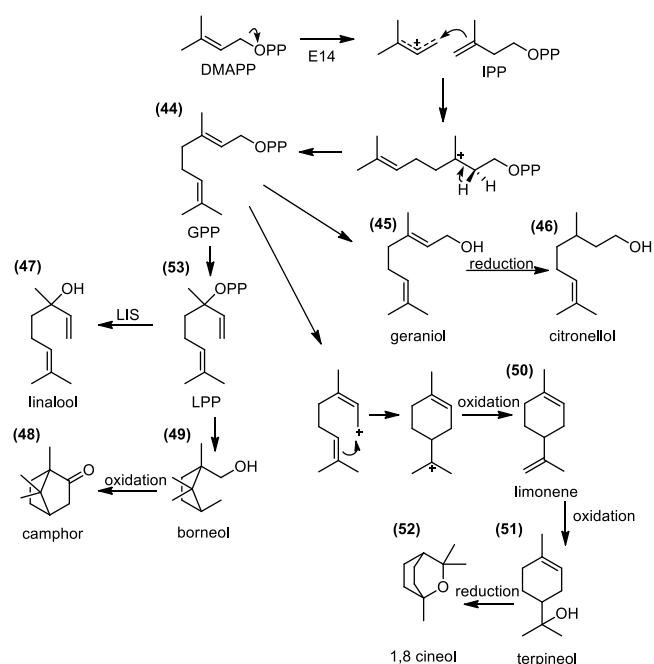


Fig. 23 Biosynthesis of GPP, the backbone for monoterpenes such as geraniol, citronellol, linalool and linalyl acetate.

The mevalonic acid (MVA) pathway predominantly leads to the biosynthesis of sesquiterpenes (C_{15}), a major class of terpenes formed by the union of three isoprene units. These units originate from the farnesyl diphosphate (FPP) precursor (**54**), which is synthesized by the condensation of an isopentenyl diphosphate (IPP) unit with geranyl pyrophosphate (GPP) through the action of GPP synthase (E14). The resulting GPP cation facilitates the addition of an IPP unit, culminating in the formation of the FPP cation. This process is followed by the loss of a proton in a reaction catalyzed by FPPS prenyltransferase (E15) (**Fig. 24**). The chain length can be even increased by further modifications through oxidative cyclisation reactions to give rise to a diversified number of metabolites. Thus, besides the great number of linear structures, a vast range of sesquiterpenes motifs such as mono-, bi-, and tricyclic structures can be biosynthesized.^{53,145} The sesquiterpenes such as the (+)-germacrene A (**55**) and the (+)-germacrene B (**56**) are relevant metabolites encountered in different *Ocotea* species, and also well present in their essential oils.^{19,149}

Likewise, the biosynthesis of diterpenes (C_{20}) occurs with the addition of another IPP unit to the FPP core through the action of geranylgeranyl diphosphate synthases. In this way, diterpenes are characterized by four isoprene units with additional cyclized modifications. Regardless of that, diterpenes are less common in essential oils, and this class has been found only in a few *Ocotea* species, for example, *O. floribunda*, *O. nutans* and *O. bicolor* (Supplementary Table S5-S6 and S5.1-S6.1, available at Zenodo's link- <https://doi.org/10.5281/zenodo.1067>). Due to their rarity, diterpenes may not serve as primary biomarkers for the *Ocotea* genus. Instead, they provide specific chemical profiles useful for differentiating species within the genus.

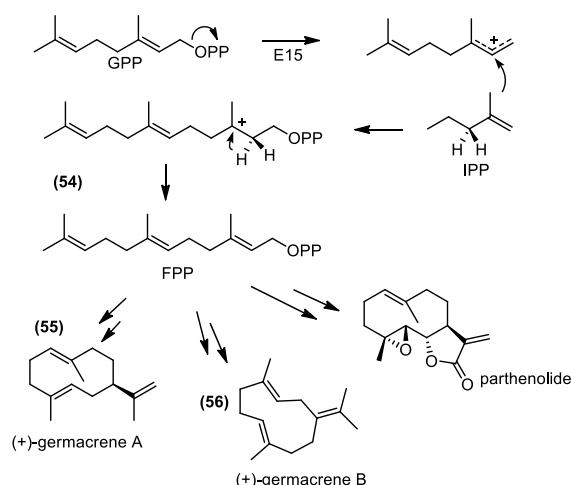


Fig. 24 Biosynthesis of FPP, the backbone for sesquiterpenes such as (+)-germacrene A and B and parthenolide.

6.2 *Ocotea* spp. terpenoids profiles

6.2.1 Monoterpenes

Ocotea species are known for producing a diverse array of monoterpenes, which include both oxygenated and non-oxygenated, as well as linear and cyclic compounds. These monoterpenes are distinguished by their varied chemical structures and unique biological activities, demonstrating a range of pharmacological properties such as antimicrobial, antifungal, antitumoral, and anti-inflammatory effects. Studies have suggested that the therapeutic potential of monoterpenes is due to their ability to interact with various biological targets, including enzymes, receptors, and ion channels. Such interactions can significantly affect cellular processes by modulating signaling pathways and influencing inflammation, apoptosis, and cell growth.^{11,143,150,151}

Among these, several components like **47**, **49**, **51**, **52**, limonene (**50**), β -myrcene (**69**), 4-terpineol (**70**), *cis*- β -ocimene (**71**), *trans*-ocimene (**72**), α -pinene (**73**), β -pinene (**74**), and camphene (**75**) are prevalent across various *Ocotea* species (**Fig. 25**). Specifically, **69** is a major constituent not only in several plants but also abundantly found in twenty-two different *Ocotea* species (Supplementary Table S6 and S6.1, available at Zenodo's link- <https://doi.org/10.5281/zenodo.1067>). This compound is known as a highly important raw material for the production of flavors, fragrances, cosmetics, and vitamins, and also plays an important role in the pharmaceutical field since **69** has exhibited significant biological properties, e.g. anxiolytic, antioxidant, anti-ageing, anti-inflammatory and analgesic.¹⁵² Similarly, **52** (also known as eucalyptol) is found in twenty-one *Ocotea* species and is renowned for its anti-inflammatory effects, primarily through the inhibition of interleukins and interference with TNF- α production. It also exhibits antinociceptive, cardiovascular, and vasorelaxant properties.^{153,154}

Moreover, **51** and **70**, which are commonly found in the Tea tree (*Melaleuca alternifolia*), were present in fifteen *Ocotea* species, in addition to the β -, γ - and δ -terpineol that were also reported. These compounds act as anti-inflammatory agents by inhibiting the production of LPS-induced mediators such as IL-1 β , IL-6, and IL-10.¹⁵⁵ The **51** is also utilized industrially as an ingredient in aromatic scents, perfumes, and cosmetics, given its lilac-like

odor. Additionally, it exhibits a wide range of biological properties, including antioxidant, anticancer, anticonvulsant, antiulcer, antihypertensive, antinociceptive, and insecticidal effects.¹⁵⁶ Fig. 25 describes the most commonly identified monoterpenes in the *Ocotea* genus.

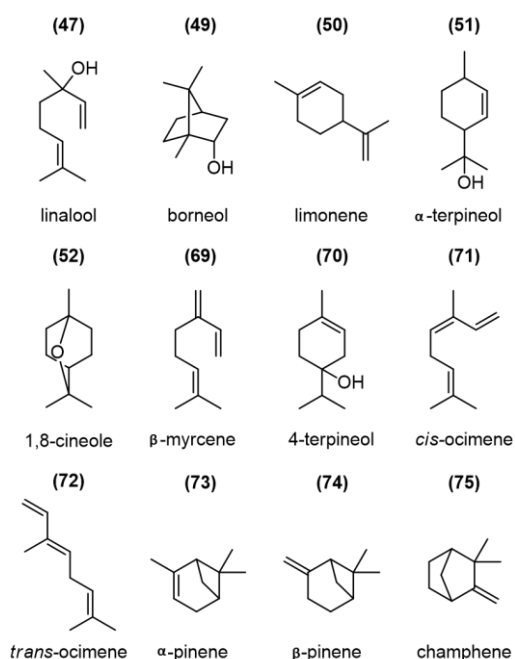


Fig. 25 Widespread components of *Ocotea* sp. essential oils from the monoterpene class.

The **47** was found in thirteen *Ocotea* species, and it is also recognised as a biologically active compound with commercial interest due to its beneficial aromatic properties. This compound has demonstrated repellent properties against various crop-destroying insects, and potential antioxidant, anti-inflammatory, anticancer, anxiolytic, analgesic, and sedative effects.^{157,158} Moreover, the **49** is a common component in plants essential oils, which is present in numerous medicinal plants, such as the *Valeriana officinalis*, the *Matricaria chamomilla*, the *Lavandula officinalis*, and it was also present in twelve *Ocotea* species. In addition, **49** can be used in fragrances and cosmetics, and also hold potential for medicinal applications, once there are reports of biological activities that include antimicrobial, anti-inflammatory, and antiviral effects.^{150,159,160}

Terpenoids from several *Ocotea* species have shown promising antitumor properties against a variety of cancer cell lines, including breast adenocarcinoma epithelial cells (MCF-7 and MDA-MB-231), cerebral glioblastoma (A-172, U-87MG), cerebral astrocytoma (CCF-STTG1), and hepatocellular carcinoma (Hep-G2) cells.^{57,160–164} Moreover, terpenoids **70** found in species like *O. quixos*, *O. bofo*, *O. opfera* and others, exhibit potent cytotoxic and antitumoral effects, mediated through various mechanisms including the induction of apoptosis, inhibition of cell migration, DNA fragmentation, and cell cycle arrest.¹⁶⁰

6.2.2 Sesquiterpenes

Sesquiterpenes and sesquiterpenoids constitute the most predominant terpenoid subclass in *Ocotea* species and are major components of the essential oils extracted from various *Ocotea* sp. In addition, several terpenoids could be also isolated from other

different parts of *Ocotea* plant species, including bark, leaves and flowers.^{6,149} Thus, the most prevalent compounds are described in Fig. 26 and Tables S.5 and S5.1, including the spathulenol (**57**), α -cadinol (**58**), elemol (**59**), bicyclogermacrene (**60**), caryophyllene oxide (**61**), globulol (**62**), α -humulene (**63**), β -caryophyllene (**64**), 1-*epi*-cubenol (**65**), viridiflorol (**66**), *trans*-nerolidol (**67**), and α -cadinene (**68**).

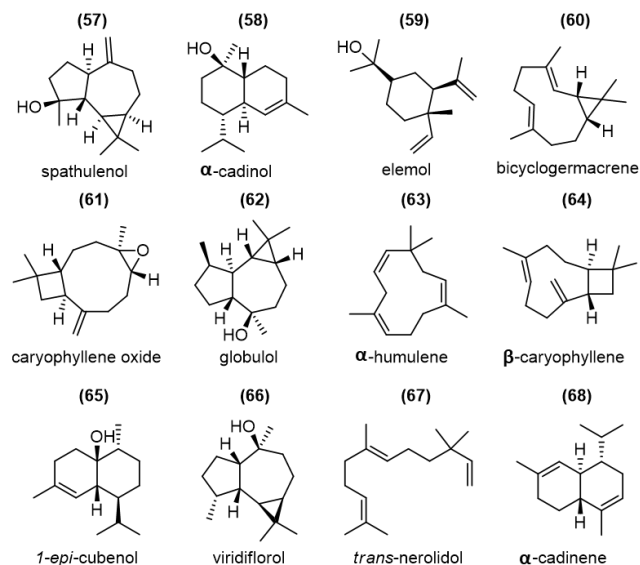


Fig. 26 Widespread components of *Ocotea* sp. essential oils from the sesquiterpene class.

The sesquiterpenoid **57** was identified in twenty-five different *Ocotea* sp., e.g. *O. diospyrifolia*, *O. lancifolia*, and *O. nutans*, while the isospathulenol was found only in *O. dispersa*. These compounds are tricyclic sesquiterpene alcohols, whose core skeleton is highly similar to the popular azulenes class. There are several aromatic medicinal plants in which the **57** is recognised as a major volatile constituent, for example, the essential oils of *Psidium guineense*, which is also a plant with ethnopharmacological significance, and popularly used for treating a range of inflammatory diseases, as the *O. odorifera*.⁴⁵ In addition to the anti-inflammatory properties, **57** is also reported with antioxidant, antiproliferative, immunomodulatory and antibacterial effects in different studies in the literature.^{165–167}

Likewise, **58** has been found in twenty-three different *Ocotea* species, alongside other less concentrated isomers. Also, the sesquiterpenoid **58** and its isomers can be encountered in the essential oils of other plant genera and are recognized for their antioxidant, cytotoxic, and antimicrobial activities.^{50,167,168} Additionally, the sesquiterpenes **64** and **67**, and sesquiterpenoids **59**, **61** and **62** were identified in different isomeric forms and belonging to the chemical composition of several *Ocotea* species (Supplementary Table S6 and S6.1, available at Zenodo's link-<https://doi.org/10.5281/zenodo.1067>). Similarly, sesquiterpenoid **66** was found in eleven *Ocotea* plant species, and it is noted for its anti-inflammatory action against leukocyte migration and also exhibits antioxidant and antibacterial activities.¹⁶⁹

6.3 *Ocotea* spp. essential oils

Essential oils are complex mixtures of volatile organic compounds primarily biosynthesized in specialised plant cells or produced as byproducts during the distillation process, which

stands as the most common method for essential oil extraction.^{29,143,170} Some components of essential oils are heat-derived artefacts, not originally present in the plant material but formed during distillation from chemical precursors in the plant's metabolome.¹⁴³ Sometimes these artefacts are present in the raw plant material, however, their concentration might increase during the distillation extraction. Therefore, although essential oils are made from natural ingredients, not all components of essential oils are indeed from a natural origin.¹⁴³ For instance, spathulenol, a sesquiterpene commonly found in *Ocotea* species, is produced from the bicyclogermacrene precursor, although it is naturally present in smaller quantities.^{143,171} These terpenoids can be found in varying proportions in *Ocotea* spp. essential oils.

The family of the Lauraceae, to which *Ocotea* belongs, is highly valued for the essential oils produced by its plants. *Ocotea* species are predominantly aromatic, making significant contributions to the Lauraceae's reputation for producing a variety of volatile compounds.^{17,172} Despite numerous studies, only 48 out of 113 chemically characterized species (~42.5%) have been examined for their essential oil composition. To date, 36 unique medium-weight terpenoids (e.g. sesquiterpenes and triterpenes) were isolated from *Ocotea* species. In addition, gas chromatography coupled with mass spectrometry (GC-MS) analysis has identified 456 different compounds as constituents of *Ocotea* essential oils, including a range of different monoterpenoids. (Supplementary **Table S6** and **S6.1** respectively, available at Zenodo's link- <https://doi.org/10.5281/zenodo.1067>).

Essential oils are utilized widely in conventional medicine and aromatherapy, and they are prevalent in cosmetics, dentistry, agriculture, food flavoring, cleaning products, and solvents, making them one of the industry's most successful commodities.^{140,143,158,173} Their popularity in modern society is justified by their association with various desirable biological activities, such as antimicrobial, antioxidant, anti-inflammatory, and cytotoxic, and supported by extensive scientific research.^{158,162,169,172,174,175}

Moreover, terpenes represent the most abundant components in *Ocotea* spp. essential oils, with sesquiterpenes (64.6%) being the dominant subclass, followed by monoterpenes (28.5%). The qualitative and quantitative variations in these components can be influenced by factors such as genetic specificity, developmental stage, environmental stress, and adaptation to local climatic and soil conditions, as well as interspecies interactions.¹⁷² The essential oils of various *Ocotea* species also contain minor components like aldehydes (1.8%), phenylpropene (1.4%), phenylpropanoids (1.0%), alcohols (0.7%), ketones (0.6%), diterpenes (0.5%), and even smaller percentages of aliphatic and aromatic hydrocarbons, carboxylic acids, and esters (Supplementary **Table S6** and **S6.1**, available at Zenodo's link- <https://doi.org/10.5281/zenodo.1067>).

6.4 *Ocotea* spp. phenylpropanoids profile

Phenylpropanoids are a diverse class of compounds commonly found in plant essential oils, primarily derived from the aromatic amino acid phenylalanine, which after deamination, typically catalyzed by PAL, forms cinnamic acid, which is the basic building block for the vast array of phenylpropanoid compounds. These compounds are also prevalent in various plant parts, extending beyond just essential oils. Furthermore, saffrole, a well-known phenylpropanoid, is notably prevalent across the Lauraceae family and was once suggested by Professor Otto R. Gottlieb as a chemotaxonomic marker for this family, due to its widespread occurrence in several species.¹⁰ Interestingly, to date, biosynthesis

of saffrole has only been confirmed in a few *Ocotea* sp., e.g. *O. pretiosa*, *O. odorifera*, *O. cymbarum*, *O. opfera* and *O. zahamenensis*.^{45,172,176} Notably, saffrole is the major component in *O. odorifera* essential oil, where its content can range dramatically from 30% to 90%. Despite its prevalence, based on the current data available since Gottlieb's initial studies, our review suggests that saffrole may not serve as a reliable chemotaxonomic marker within the *Ocotea* genus. Nonetheless, it may still hold validity as a marker for other genera within the Lauraceae family. This underscores the complexity and variability of phenylpropanoid biosynthesis and their potential utility in the chemotaxonomic classification of related plant groups.

6.5 Remarkings considerations and recent advances

Recent research has not only highlighted the anti-inflammatory, cytotoxic and antitumor potential of the metabolome of *Ocotea* species but also demonstrated the significant role that *Ocotea* essential oils play in enhancing the efficacy of antibiotics against multidrug-resistant bacteria.¹⁷⁷ The authors demonstrated that the essential oil extracted from *O. odorifera* and its primary component, saffrole, can enhance the activity of macrolides and aminoglycosides antibiotics, such as erythromycin and gentamicin, respectively. This is achieved through the direct inhibition of efflux pumps, which are crucial in modulating bacterial resistance. Consequently, the combination of *O. odorifera* essential oil and saffrole with antibiotics lead to improved antibacterial activity and clinically relevant effects against *Staphylococcus aureus*.¹⁷⁷

Furthermore, alongside their potential as therapeutic agents, recent applications of nanotechnology combined with NP present a promising alternative for ecological management. This approach serves as a potential substitute for synthetic anti-parasitic drugs and insecticides, which have been linked to significant environmental damage and adverse effects on non-target organisms, including the development of insect resistance.^{172,178–180} In this context, recently, the potential of a nanoemulsion containing essential oil from *O. pulchella* was evaluated as a control agent in the schistosomiasis cycle. This nanoemulsion presented molluscicidal, ovicidal, and cercaricidal activities against the schistosomiasis transmitter *Biomphalaria glabrata*, causing the death of adults and preventing oviposition. The major component found in the leaves was myristicin (~29.0%), followed by **73** (~17.2%) and **60** (~16.6%).¹⁷² Additionally, it showed relevant antiparasitic activity against *Schistosoma mansoni*, the infectious agent of schistosomiasis.¹⁷²

In terms of environmental management, in the search for safer and more sustainable insecticide formulations, nanotechnology associated with NP was also evaluated against the *Aedes aegypti* mosquito, which is a serious health problem in Brazil. An optimized nanoemulsion containing essential oil from *O. indecora* leaves was developed and tested for its larvicidal properties against *A. aegypti* larvae.^{172,178} The major constituent found in the essential oil was sesquirosefuran (~81.4%). The *in-silico* analysis suggested that the larvicide property is related to acetylcholinesterase enzyme inhibition. Notably, the nanoemulsion demonstrated no toxicity against the non-target organism *Apis mellifera*, ensuring safety for pollinator bees. Additionally, this formulation showed stability when stored at room temperature or refrigerated.

7. Conclusion remarks and future prospective

In summary, this bibliographic review sheds light on the metabolome of the *Ocotea* genus. We have meticulously detailed the principal metabolic pathways that give rise to the bioactive scaffolds of alkaloids, lignoids, flavonoids, and terpenoids. This comprehensive overview of the chemical and biosynthetic characteristics provides a valuable resource for gaining insights and guiding further research targeting *Ocotea* species. Our review synthesizes data from 113 different *Ocotea* species and 928 metabolites, including 174 alkaloids, 172 lignoids, 57 flavonoids, 527 terpenoids (encompassing monoterpenes, sesquiterpenes, diterpenes, and triterpenes isolated from crude extracts and identified in essential oils) and 42 metabolites from less frequent classes. To our knowledge, this constitutes the most thorough examination of the genus available in the literature to date. The presented chemical data can aid in species identification, differentiation and chemophenetic studies, providing support for the integration of the chemical information with taxonomical and phylogenetic investigations in further evolutionary studies of the *Ocotea* genus.

This review also discusses the specific stereochemistry differentiation of alkaloids and the presence of common and uncommon NP classes among the *Ocotea* species. Moreover, the structural patterns and metabolites explored here hold significant promise for the pharmaceutical and medicinal chemistry communities. Promissory bioactive scaffolds were detailed corroborating the *Ocotea* species as a valuable source for drug discovery. The compiled data in this review can support further computational, chemical, and bioactivity investigations, and could be utilized to add complementary information for innovative chemometric and metabolomic models. Future investigations are likely to reveal even more intricate biosynthetic pathways and novel compounds. Ultimately, this review offers a comprehensive understanding of the *Ocotea* metabolome, providing valuable insights and laying the groundwork for continued advancements in various NP research fields related to the *Ocotea* plant species.

8. Author contributions

Albert Katchborian-Neto and **Daniela A. Chagas-Paula** contributed to the conceptualization, methodology, and writing of the original draft, and editing the final manuscript version. Both **Albert Katchborian-Neto** and **Karen J. Nicácio** were involved in all sections including the methodology, writing, reviewing, and editing the main sections of the manuscript. **Matheus F. Alves**, **Paula C. P. Bueno**, **Miller S. Ferreira**, **Wanderleya S. Toledo** and **Danielle O. Silva** were responsible for chemical data curation, reviewing, and editing. **Paula C. P. Bueno**, **João H. G. Lago**, **Daniele F. Dias**, **Marisi G. Soares**, **Ana C. C. Paula**, **Daniela A. Chagas-Paula** and **RuAngelie Edrada-Ebel** provided resources, funding, and supervision, and contributed critical thinking and also to the writing, reviewing, and editing the final manuscript version.

9. Conflicts of interest

The authors declare no conflicts of interest.

10. Acknowledgements

The authors express their gratitude to the funding agencies: National Council for Science and Technology (CNPq), grant numbers 408115/2023-8, 316204/2021-8, and 406837/2021-0;

Coordination for the Improvement of Higher Education Personnel (CAPES), finance code 001; and Minas Gerais Research Funding Foundation (FAPEMIG), grant numbers APQ-05218-23, APQ-00544-23, APQ-02353-17, APQ-00207-18, and BPD-00760-22, for their financial support.

11. References

- 1 D. Trofimov and J. G. Rohwer, *Bot. J. Linn. Soc.*, 2020, **192**, 510–535.
- 2 D. L. Custódio and V. Florêncio da Veiga Junior, *RSC Adv.*, 2014, **4**, 21864–21890.
- 3 M. U. Yamaguchi, F. P. Garcia, D. A. G. Cortez, T. Ueda-Nakamura, B. P. D. Filho and C. V. Nakamura, *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.*, 2011, **99**, 507–514.
- 4 D. Trofimov, P. L. R. De Moraes and J. G. Rohwer, *Bot. J. Linn. Soc.*, 2019, **190**, 25–50.
- 5 C. S. Wu, T. J. Wang, Chia-Wenwu, Y. N. Wang and S. M. Chaw, *Genome Biol. Evol.*, 2017, **9**, 2604–2614.
- 6 W. M. N. H. W. Salleh and F. Ahmad, *J. Appl. Pharm. Sci.*, 2017, **7**, 204–218.
- 7 J. Wang, B. Su, H. Jiang, N. Cui, Z. Yu, Y. Yang and Y. Sun, *Fitoterapia*, 2020, **146**, 104675.
- 8 C. A. Marques, *Floresta e Ambient.*, 2001, **8**, 195–206.
- 9 M. E. Galindo-Tovar, A. M. Arzate-Fernández, N. Ogata-Aguilar and I. Landero-Torres, *Harvard Pap. Bot.*, 2007, **12**, 325–334.
- 10 O. R. Gottlieb, *Phytochemistry*, 1972, **11**, 1537–1570.
- 11 L. D. F. S. Sampaio, J. G. S. Maia, A. M. De Parijós, R. Z. De Souza and L. E. S. Barata, *Phyther. Res.*, 2012, **26**, 73–77.
- 12 R. P. Santos, P. Cristina, P. De Tarso, B. Sampaio, C. Quisen, Â. Maria, C. Leite and C. L. De Oliveira, *Acta Amaz.*, 2008, **38**, 459–466.
- 13 S. C. Joshi, R. C. Padalia, D. S. Bisht and C. S. Mathela, *Chem. Biodivers.*, 2009, **6**, 1364–1373.
- 14 B. G. V. de Alcântara, A. K. Neto, D. A. Garcia, R. Casoti, T. B. Oliveira, A. C. C. de P. Ladvoat, R. Edrada-Ebel, M. G. Soares, D. F. Dias and D. Chagas-Paula, *Chem. Biodivers.*, 2023, **20**, 1–10.
- 15 R. Hamidpour, S. Hamidpour, M. Hamidpour and M. Shahlari, *Int. J. Case Reports Images*, 2013, **4**, 86.
- 16 J. Pavlis and G. Yosipovitch, *Am. J. Clin. Dermatol.*, 2018, **19**, 319–332.
- 17 A. Simić, M. D. Soković, M. Ristić, S. Grujić-Jovanović, J. Vukojević and P. D. Marin, *Phyther. Res.*, 2004, **18**, 713–717.
- 18 R. Bruni, A. Medici, E. Andreotti, C. Fantin, M. Muzzoli, M. Dehesa, C. Romagnoli, G. Sacchetti, A. Guerrini, P. Noriega, A. Bianchi and R. Bruni, *Food Chem.*, 2004, **85**, 415–421.
- 19 M. J. De Camargo, M. L. D. Miranda, C. M. Kagamida, E. D. Rodrigues, F. R. Garcez and W. S. Garcez, *Quim.*

- Nova*, 2013, **36**, 1008–1013. 44
- 20 A. Fournet, M. E. Ferreira, A. Rojas de Arias, I. Guy, H. Guinaudeau and H. Heinzen, *Fitoterapia*, 2007, **78**, 382–384.
- 21 A. L. Harvey, R. Edrada-Ebel and R. J. Quinn, 2015, 3–19. 45
- 22 N. D. Yuliana, A. Khatib, Y. H. Choi and R. Verpoorte, *Phyther. Res.*, 2011, 25, 157–169.
- 23 G. M. Cragg and D. J. Newman, *Biochim. Biophys. Acta - Gen. Subj.*, 2013, **1830**, 3670–3695. 46
- 24 E. M. Martins, R. W. Lamont, G. Martinelli, C. F. Lira-Medeiros, A. Quinet and A. Shapcott, *Conserv. Genet.*, 2015, **16**, 1–14.
- 25 D. C. Gontijo, G. C. Brandão and P. C. Gontijo, 2017.
- 26 S. Zschocke, J. Van Staden, K. Paulus, R. Bauer, M. M. Horn, O. Q. Munro, N. J. Brown and S. E. Drewes, *Phytochemistry*, 2000, **54**, 591–595. 47
- 27 A. Katchborian-Neto, M. F. Alves, P. C. P. Bueno, K. De Jesus Nicácio, M. S. Ferreira, T. B. Oliveira, H. Barbosa, M. Murgu, A. C. C. De Paula Ladvoat, D. F. Dias, M. G. Soares, J. H. G. Lago and D. A. Chagas-Paula, *Brief. Bioinform.*, 2024, **25**, 1–18. 48
- 28 C. Chaverri and J. F. Cicció, *Rev. Biol. Trop.*, 2005, **53**, 431–436. 49
- 29 G. Sacchetti, A. Guerrini, P. Noriega, A. Bianchi and R. Bruni, *Flavour Fragr. J.*, 2006, **21**, 674–676.
- 30 Jardim Botânico do Rio de Janeiro, Ocotea in Flora e Funga do Brasil, <https://floradobrasil.jbrj.gov.br/FB8456>, (accessed 17 April 2024). 50
- 31 U. M. Yamaguchi, P. F. Garcia, A. D. Garcia and T. Ueda-Nakamura, *Antonie van Leeuwenhoek J. Microbiol.*, 2011, **99**, 507–514. 51
- 32 A. S. Antonio, V. F. Veiga-Junior and L. S. M. Wiedemann, *Phytochemistry*, 2020, **173**, 112314. 52
- 33 M. L. Brotto, A. C. Cervi and É. P. Dos Santos, *Rodriguesia*, 2013, **64**, 495–525. 53
- 34 A. S. Chanderbali, H. Van Der Werff and S. S. Renner, *Ann. Missouri Bot. Gard.*, 2001, **88**, 104–134. 54
- 35 D. Trofimov, D. Cadar, J. Schmidt-Chanasit, P. L. Rodrigues de Moraes and J. G. Rohwer, *Sci. Rep.*, 2022, **12**, 1–13. 55
- 36 D. Trofimov and J. G. Rohwer, *Perspect. Plant Ecol. Evol. Syst.*, 2018, **31**, 17–35. 56
- 37 J. C. Penagos Zuluaga, H. van der Werff, B. Park, D. A. R. Eaton, L. S. Comita, S. A. Queenborough and M. J. Donoghue, *Am. J. Bot.*, 2021, **108**, 664–679. 57
- 38 T. Hancock, *Lancet*, 1830, **361**, 688–690. 58
- 39 J. Stenhouse, *London, Edinburgh, Dublin Philos. Mag. J. Sci.*, 1844, **25**, 200–200. 59
- 40 Halberkann Josef, *Chem. life Sci.*, 1916, **9**, 246–255. 60
- 41 A. E. and K. D.P., *J. Ethnobiol. Ethnomed.*, 2012, **8**, 1. 61
- 42 M. J. De Camargo, M. Lazaro, D. Miranda, C. M. Kagamida, E. D. Rodrigues, F. Rodrigues, W. Silva, S. Of and L. Six, *Quim. Nov.*, 2013, **36**, 1008–1013. 62
- 43 P. Ludy Cristina and C. Luis Enrique, *Quim. Nova*, 2010, **33**, 875–879.
- A. N. L. Batista, J. M. B. Junior, S. N. López, M. Furlan, A. J. Cavalheiro, D. H. S. Silva, V. D. S. Bolzani, S. M. Nunomura and M. Yoshida, *Quim. Nova*, 2010, **33**, 321–323.
- B. G. V. de Alcântara, F. P. de Oliveira, A. Katchborian-Neto, R. Casoti, O. da S. Domingos, M. F. C. Santos, R. B. de Oliveira, A. C. C. de Paula, D. F. Dias, M. G. Soares and D. A. Chagas-Paula, *J. Ethnopharmacol.*, 2021, **264**, 133378.
- A. F. Silva, M. F. C. Santos, T. S. C. Maiolini, P. P. O. Salem, M. Murgu, A. C. C. Paula, E. O. Silva, K. J. Nicácio, A. G. Ferreira, D. F. Dias, M. G. Soares and D. A. Chagas-Paula, *Phytochem. Lett.*, 2021, **42**, 52–60.
- S. M. W. Zanin and A. L. L. Lordello, *Quim. Nova*, 2007, **30**, 92–98.
- Y. Li, S. Xie, J. Ying, W. Wei and K. Gao, *Molecules*, 2018, **23**, 1–18.
- J. R. d. A. Silva, D. F. M. do Carmo, E. M. Reis, G. M. C. Machado, L. L. Leon, B. O. da Silva, J. L. P. Ferreira and A. C. F. Amaral, *J. Braz. Chem. Soc.*, 2009, **20**, 1071–1076.
- F. C. M. Betim, C. F. de Oliveira, A. M. de Souza, E. M. Szabo, S. M. W. Zanin, O. G. Miguel, M. D. Miguel and J. de F. G. Dias, *Brazilian J. Pharm. Sci.*, 2019, **55**, 1–10.
- A. Katchborian-neto, K. De Jesus, J. C. Carolina, Cruz, Paula, P. Bueno, M. Murgu, D. F. Dias, M. G. Soares, A. C. C. Paula and D. A. Chagas-paula, *Phytomedicine*, 2023, **120**, 155060.
- H. Barbosa, R. L. C. G. da Silva, T. A. Costa-Silva, A. G. Tempone, G. M. Antar, J. H. G. Lago and L. Caseli, *Bioorg. Chem.*, 2020, **101**, 103978.
- P. M. Dewick, *Medicinal Natural: A Biosynthetic Approach*, 2009, vol. 3.
- M. Madalena, R. Silva, A. Angel and V. Pinheiro, in *Alkaloids of the Lauraceae*, 2019, vol. 82, pp. 1–32.
- C. Zidorn, *Phytochemistry*, 2019, **163**, 147–148.
- N. J. Sadgrove, I. R. H. Telford, G. F. Padilla-gonzález, B. W. Greatrex, J. J. Bruhl, R. Botanic, G. Kew and U. Kingdom, *Phytochem. Lett.*, 2020, **38**, 112–120.
- Q. Quentin Li, R. X. Lee, H. Liang and Y. Zhong, *Anticancer Res.*, 2013, **33**, 65–76.
- Y. Liu, G. Fan, J. Zhang, Y. Zhang, J. Li, C. Xiong, Q. Zhang, X. Li and X. Lai, *Sci. Rep.*, 2017, **7**, 1–11.
- S. M. U. P. Mawalagedera, D. L. Callahan and A. C. Gaskett, 2019, **7**, 1–11.
- D. A. Chagas-Paula, T. B. Oliveira, T. Zhang, R. Edrada-Ebel and F. B. Da Costa, *Planta Med.*, 2015, **81**, 450–458.
- L. Yi, N. Dong, Y. Yun, B. Deng, D. Ren, S. Liu and Y. Liang, *Anal. Chim. Acta*, 2016, **914**, 17–34.
- A. Katchborian-Neto, W. T. Santos, K. de Jesus Nicácio, J. O. A. Corrêa, M. Murgu, T. M. M. Martins, D. A. Gomes, A. M. Goes, M. G. Soares, D. F. Dias, D. A. Chagas-Paula and A. C. C. Paula, *J. Ethnopharmacol.*, 2020, **255**, 112743.

- 63 B. Xi, H. Gu, H. Baniasadi and D. Raftery, *Methods Mol Biol.*, 2015, 1–20.
- 64 J. Trygg, E. Holmes and T. Lundstedt, *J. Proteome Res.*, 2007, **6**, 469–479.
- 65 E. M. Martins, G. Martinelli, M. P. Arbetman, R. W. Lamont, J. L. Simões-Araújo, D. Powell, M. Ciampi-Guillard, C. Baldauf, A. Quinet, P. Galisa and A. Shapcott, *Genet. Mol. Res.*, 2014, **13**, 5138–5142.
- 66 E. V. Costa, P. E. O. da Cruz, M. L. B. Pinheiro, F. A. Marques, A. L. T. G. Ruiz, G. M. Marchetti, J. E. de Carvalho, A. Barisonn and B. H. L. N. S. Maia, *J. Braz. Chem. Soc. Braz. Chem. Soc.*, 2013, **24**, 788–796.
- 67 I. M. Menéndez-Perdomo and P. J. Facchini, *Molecules*, 2018, **23**, 1–17.
- 68 M. M. R. Silva Teles, A. A. Vieira Pinheiro, C. Da Silva Dias, J. Fachine Tavares, J. M. Barbosa Filho and E. V. Leitão Da Cunha, *Alkaloids Chem. Biol.*, 2019, **82**, 147–304.
- 69 J. H. Y. Vilegas and O. Gottlieb, *Phytochemistry*, 1989, **28**, 3577–3578.
- 70 T. Meelaph, K. Kobtrakul, N. N. Chansilpa, Y. Han, D. Rani, W. De-Eknamkul and S. Vimolmangkang, *ACS Omega*, 2018, **3**, 8794–8802.
- 71 M. G. Marinus and A. Løbner-Olesen, *EcoSal Plus*, 2014, **6**, 1–62.
- 72 F. Carnevale Neto, M. A. Andréo, D. Raftery, J. L. C. Lopes, N. P. Lopes, I. Castro-Gamboa, B. H. Lameiro de Noronha Sales Maia, E. V. Costa and R. Vessecchi, *Rapid Commun. Mass Spectrom.*, 2019, 0–2.
- 73 J. Ziegler, P. J. Facchini, R. Geißler, J. Schmidt, C. Ammer, R. Kramell, S. Voigtländer, A. Gesell, S. Pienkny and W. Brandt, *Phytochemistry*, 2009, **70**, 1696–1707.
- 74 C. Weber and T. Opatz, *Bisbenzylisoquinoline Alkaloids*, Elsevier Inc., 1st edn., 2019, vol. 81.
- 75 A. Gesell, M. Rolf, J. Ziegler, M. L. D. Chávez, F. C. Huang and T. M. Kutchan, *J. Biol. Chem.*, 2009, **284**, 24432–24442.
- 76 E. Gil Archila and L. E. Cuca Suárez, *Nat. Prod. Res.*, 2018, **32**, 195–201.
- 77 D. S. A. Cassiano, I. M. A. Reis, I. de O. Estrela, H. F. de Freitas, S. S. da R. Pita, J. M. David and A. Branco, *Comput. Biol. Chem.*, 2019, **83**, 107129.
- 78 F. R. Garcez, A. F. G. Da Silva, W. S. Garcez, G. Linck, M. D. F. C. Matos, E. C. S. Santos and L. M. M. Queiroz, *Planta Med.*, 2011, **77**, 383–387.
- 79 J. L. Wolfender, M. Litaudon, D. Touboul and E. F. Queiroz, *Nat. Prod. Rep.*, 2019, **36**, 855–868.
- 80 N. Backhouse, C. Delporte, M. Givernau, B. K. Cassels, A. Valenzuela and H. Speisky, *Agents Actions*, 1994, **42**, 114–117.
- 81 X. Yang, X. Gao, Y. Cao, Q. Guo, S. Li, Z. Zhu, Y. Zhao, P. Tu and X. Chai, *Planta Med.*, 2018, **84**, 20–25.
- 82 H. Kang, S. W. Jang, J. H. Pak and S. Shim, *Mol. Cell. Biochem.*, 2015, **403**, 85–94.
- 83 D. P. Montrucchio, M. M. Córdova and A. R. Soares Santos, *PLoS One*, 2013, **8**, 2–10.
- 84 J. W. Dong, L. Cai, Y. S. Fang, H. Xiao, Z. J. Li and Z. T. Ding, *Fitoterapia*, 2015, **104**, 102–107.
- 85 I. P. Singh and H. S. Bodiwala, *Nat. Prod. Rep.*, 2010, **27**, 1781–1800.
- 86 Y. Kashiwada, A. Aoshima, Y. Ikeshiro, Y. P. Chen, H. Furukawa, M. Itoigawa, T. Fujioka, K. Mihashi, L. M. Cosentino, S. L. Morris-Natschke and K. H. Lee, *Bioorganic Med. Chem.*, 2005, **13**, 443–448.
- 87 A. Hosalkova, L. Opletal, J. Kunes, Z. Novak, M. Hrabínova, J. Chlebek, L. Eegan and L. Cahlikova, *Nat. Prod. Commun.*, 2015, **10**, 577–580.
- 88 D. Chagas-Paula, T. Zhang, F. Da Costa and R. Edrada-Ebel, *Metabolites*, 2015, **5**, 404–430.
- 89 R. B. Teponno, S. Kusari and M. Spiteller, *Recent advances in research on lignans and neolignans*, 2016, vol. 33.
- 90 M. Gordaliza, M. Castro, J. Miguel del Corral and A. San Feliciano, *Curr. Pharm. Des.*, 2005, **6**, 1811–1839.
- 91 J. D. Ickovski, J. L. J. Pavlovic, M. N. Mitic, I. R. Palic, D. A. Kostic, G. M. Petrovic and G. S. Stojanovic, *J. Serbian Chem. Soc.*, 2020, **85**, 575–600.
- 92 Q. Cui, R. Du, M. Liu and L. Rong, *Molecules*, 2020, **25**, 1–17.
- 93 W. D. Da Silva, R. Braz-Filho and O. R. Gottlieb, *Phytochemistry*, 1989, **28**, 661–662.
- 94 M. Ishige, M. Motidome, M. Yoshida and O. R. Gottlieb, *Phytochemistry*, 1991, **23**, 2101–2104.
- 95 N. G. Lewis, L. B. Davin and A. D. R. V. Ranjit N. Munasinghe, *The lignan handbook*, 2022, vol. 1.
- 96 O. R. Gottlieb and M. Yoshida, *Quim. Nova*, 1984, **7**, 250–273.
- 97 S. Susuki and T. Umezawa, *J. Wood Sci.*, 2007, **9**, 273–284.
- 98 A. B. Rodrigues, A. A. De Almeida-Apolonio, T. M. Alfredo, F. G. Da Silva Dantas, J. F. Campos, C. A. L. Cardoso, K. De Picoli Souza and K. M. P. De Oliveira, *Oxid. Med. Cell. Longev.*, 2019, **2019**, 1–14.
- 99 L. H. Rakotondraibe, P. R. Graupner, Q. Xiong, M. Olson, J. D. Wiley, P. Krai, P. J. Brodie, M. W. Callmander, E. Rakotobe, F. Ratovoson, V. E. Rasamison, M. B. Cassera, D. R. Hahn, D. G. I. Kingston and S. Fotso, *J. Nat. Prod.*, 2015, **78**, 431–440.
- 100 D. R. Gang, M. A. Costa, M. Fujita, A. T. Dinkova-Kostova, H. Bin Wang, V. Burlat, W. Martin, S. Sarkanen, L. B. Davin and N. G. Lewis, *Chem. Biol.*, 1999, **6**, 143–151.
- 101 N. P. Lopes, M. Yoshida and M. J. Kato, *Brazilian J. Pharm. Sci.*, 2004, **40**, 1–5.
- 102 B. L. Jackson, C. M. Kish, T. Koeduka, E. Fridman, D. R. Gang, D. G. Vassa, I. Orlova, S. M. Spassova, N. G. Lewis, J. P. Noel, T. J. Baiga, N. Dudareva and E. Pichersky, *Proc. Natl. Acad. Sci.*, 2006, **103**, 10128–10133.
- 103 N. G. Lewis, L. B. Davin and S. Sarkanen, in *American Chemical Society*, 1998, vol. 1, pp. 1–27.

- 104 J. D. Felício, M. Motidome, M. Yoshida and O. R. Gottlieb, *Phytochemistry*, 1986, **25**, 1707–1710.
- 105 A. L. L. Lordello and M. Yoshida, *Phytochemistry*, 1997, **46**, 6–9.
- 106 C. D. Dodson, F. R. Stermitz, O. Castro and D. H. Janzen, *Phytochemistry*, 1987, **26**, 2037–2040.
- 107 D. A. Dias, M. Yoshida and O. R. Gottlieb, *Phytochemistry*, 1986, **25**, 2613–2616.
- 108 E. D. Coy-Barrera, L. E. Cuca-Suárez and M. Sefkow, *Phytochemistry*, 2009, **70**, 1309–1314.
- 109 M. A. De Alvarenga, O. Castro C., A. M. Giesbrecht and O. R. Gottlieb, *Phytochemistry*, 1977, **16**, 1801–1804.
- 110 M. C. C. P. Gomes, M. Yoshida and R. Gottlieb, *Phytochemistry*, 1983, **22**, 269–273.
- 111 J. M. David, M. Yoshida and O. R. Gottlieb, *Phytochemistry*, 1994, **36**, 491–499.
- 112 A. Roux, D. Lison, C. Junot and J. F. Heilier, *Clin. Biochem.*, 2011, **44**, 119–135.
- 113 A. K. Kosmides, K. Kamisoglu, S. E. Calvano, S. A. Corbett and I. P. Androulakis, *Crit. Rev. Biomed. Eng.*, 2013, **41**, 205–221.
- 114 F. Van Der Kooy, F. Maltese, H. C. Young, K. K. Hye and R. Verpoorte, *Planta Med.*, 2009, **75**, 763–775.
- 115 J. O. A. Narciso, R. O. De Araújo Soares, J. Reis Dos Santos Mallet, A. É. Guimarães, M. C. De Oliveira Chaves, J. M. Barbosa-Filho and M. Maleck, *Parasites and Vectors*, 2014, **7**, 1–10.
- 116 D. T. da Silva, R. Herrera, B. F. Batista, B. M. Heinzmann and J. Labidi, *Int. Biodeterior. Biodegrad.*, 2017, **117**, 158–170.
- 117 P. L. N. Nêris, J. P. A. Caldas, Y. K. S. Rodrigues, F. M. Amorim, J. A. Leite, S. Rodrigues-Mascarenhas, J. M. Barbosa-Filho, L. C. Rodrigues and M. R. Oliveira, *Exp. Parasitol.*, 2013, **135**, 307–313.
- 118 J. Sánchez-Suárez, E. Coy-Barrera, L. E. Cuca and G. Delgado, *Nat. Prod. Commun.*, 2011, **6**, 231–234.
- 119 M. M. O. Cabral, J. M. Barbosa-Filho, G. L. A. Maia, M. C. O. Chaves, M. V. Braga, W. De Souza and R. O. A. Soares, *Exp. Parasitol.*, 2010, **124**, 319–324.
- 120 M. C. Marcotullio, A. Pelosi and M. Curini, *Molecules*, 2014, **19**, 14862–14878.
- 121 M. F. C. Santos, B. G. V. Alcântara, C. dos R. Feliciano, A. F. Silva, T. C. S. Maiolini, A. K. Neto, M. Murgu, D. A. C. de Paula and M. G. Soares, *Phytochem. Lett.*, 2019, **30**, 31–37.
- 122 M. F. C. Santos, C. dos R. Feliciano, A. K. Neto, D. A. C. de Paula and M. G. Soares, *Nat. Prod. Res.*, 2021, **0**, 1–4.
- 123 M. F. C. Santos, K. de J. Nicácio, A. Katchborian-Neto, M. S. Ferreira, D. de O. Miranda, J. V. Andrade, H. de A. Pereira, E. G. de Jesus, T. B. S. Souza, R. P. Moraes-Urano, D. F. Dias, D. A. Chagas-Paula and M. G. Soares, *Nat. Prod. Res.*, 2024, **38**, 393–401.
- 124 K. Yonekura-Sakakibara, Y. Higashi and R. Nakabayashi, *Front. Plant Sci.*, 2019, **10**, 1–16.
- 125 W. P. P. André, G. S. Cavalcante, W. L. C. Ribeiro, J. M. L. Dos Santos, I. T. F. Macedo, H. C. B. De Paula, S. M. De Moraes, J. V. De Melo and C. M. L. Bevilaqua, *Rev. Bras. Parasitol. Vet.*, 2017, **26**, 323–330.
- 126 J. L. Baldim, B. G. V. De Alcântara, O. D. S. Domingos, M. G. Soares, I. S. Caldas, R. D. Novaes, T. B. Oliveira, J. H. G. Lago and D. A. Chagas-Paula, *Oxid. Med. Cell. Longev.*, 2017, **2017**, 1–12.
- 127 E. L. Santos, B. H. S. maia L.N., A. P. Ferriani and S. D. Teixeira, in *Intech*, 2017, p. 13.
- 128 C. Rodríguez-García, C. Sánchez-Quesada, J. J. Gaforio and J. J. Gaforio, *Antioxidants*, 2019, **8**, 1–23.
- 129 A. García-Lafuente, E. Guillamón, A. Villares, M. A. Rostagno and J. A. Martínez, *Inflamm. Res.*, 2009, **58**, 537–552.
- 130 M. F. Alves, A. Katchborian-Neto, P. C. P. Bueno, F. Carnevale-Neto, R. Casoti, M. S. Ferreira, M. Murgu, A. C. C. de Paula, D. F. Dias, M. G. Soares and D. A. Chagas-Paula, *RSC Adv.*, 2024, **14**, 10481–10498.
- 131 G. F. Padilla-González, M. Frey, J. Gómez-Zeledón, F. B. Da Costa and O. Spring, *Sci. Rep.*, 2019, **9**, 1–15.
- 132 R. Koes, W. Verweij and F. Quattrocchio, *Trends Plant Sci.*, 2005, **10**, 236–242.
- 133 D. B. Silva, I. C. C. Turatti, D. R. Gouveia, M. Ernst, S. P. Teixeira and N. P. Lopes, *Sci. Rep.*, 2014, **4**, 1–8.
- 134 C. S. Bonde, L. Bornancin, Y. Lu, H. T. Simonsen, M. Martínez-Valladares, M. Peña-Espinoza, H. Mejer, A. R. Williams and S. M. Thamsborg, *Front. Pharmacol.*, 2021, **12**, 1–12.
- 135 A. Crozier, I. B. Jaganath and M. N. Clifford, *Plant Second. Metab. Occur. Struct. Role Hum. Diet*, 2007, 1–24.
- 136 D. C. Gontijo, G. C. Brandão, P. C. Gontijo, A. B. de Oliveira, M. A. N. Diaz, L. G. Fietto and J. P. V. Leite, *Food Chem.*, 2017, **230**, 618–626.
- 137 J. E. Yun, E. R. Woo and D. G. Lee, *Biochim. Biophys. Acta - Biomembr.*, 2018, **1860**, 357–363.
- 138 R. Gomes, T. C. Almeida, A. C. Cardoso, S. Augusto, V. Filho, G. C. Brandão, G. Nicioli, H. C. De Sousa, V. L. De Almeida, J. C. Dias, G. Henrique, B. De Souza, S. Augusto, V. Filho, G. C. Brandão, G. Nicioli, C. De Sousa, V. L. De Almeida, J. César, D. Lopes and G. H. Bianco, *Nat. Prod. Res.*, 2020, **0**, 1–6.
- 139 J. L. B. Zanin, M. Massoni, M. H. Dos Santos, G. C. De Freitas, E. L. O. Niero, R. R. Schefer, J. H. G. Lago, M. Ionta and M. G. Soares, *J. Braz. Chem. Soc.*, 2015, **26**, 804–809.
- 140 F. Abbas, Y. Ke, R. Yu, Y. Yue, S. Amanullah, M. M. Jahangir and Y. Fan, *Planta*, 2017, **246**, 803–816.
- 141 D. M. Martin, J. Gershenzon and J. Bohlmann, *Plant Physiol.*, 2003, **132**, 1586–1599.
- 142 E. Oldfield and F. Lin, *Angew Chem Int Ed Engl.*, 2012, **51**, 1124–1137.
- 143 N. J. Sadgrove, G. F. Padilla-González and M. Phumthum, *Plants*, 2022, **11**, 1–35.
- 144 E. J. N. Helfrich, G. M. Lin, C. A. Voigt and J. Clardy, *Beilstein J. Org. Chem.*, 2019, **15**, 2889–2906.

- 145 A. L. Zografos, *From Biosynth. to Total Synth. Strateg. Tactics Nat. Prod.*, 2016, 1–569.
- 146 M. E. Bergman, B. Davis and M. A. Phillips, *Molecules*, 2019, 3961.
- 147 M. E. Pyne, L. Narcross and V. J. J. Martin, *Plant Physiol.*, 2019, **179**, 844–861.
- 148 C. Terreaux, M. Maillard, K. Hostettmann, C. Universita, V. Horwri and E. Hakizamungu, *Phytochem. anal*, 1994, **5**, 233–238.
- 149 C. R. Nogueira, L. H. Carbonezi, C. T. F. de Oliveira, W. S. Garcez and F. R. Garcez, *Phytochem. Lett.*, 2021, **42**, 8–14.
- 150 M. Zielinska-Błajet and J. Feder-Kubis, *Int. J. Mol. Sci.*, 2020, **21**, 1–38.
- 151 G. Farré-Armengol, I. Filella, J. Llusà and J. Peñuelas, *Molecules*, 2017, **22**, 1–6.
- 152 K. Połec, M. Broniatowski, P. Wydro and K. Hąc-Wydro, *J. Mol. Liq.*, 2020, **308**, 1–6.
- 153 F. A. Santos and V. S. N. Rao, *Phyther. Res.*, 2000, **14**, 240–244.
- 154 M. R. V. Santos, F. V. Moreira, B. P. Fraga, D. P. de Sousa, L. R. Bonjardim and L. J. Quintans, *Rev. Bras. Farmacogn.*, 2011, **21**, 764–771.
- 155 M. N. M. Nogueira, S. G. Aquino, C. Rossa and D. M. P. Spolidorio, *Inflamm. Res.*, 2014, **63**, 769–778.
- 156 C. Khaleel, N. Tabanca and G. Buchbauer, *Open Chem*, 2018, **16**, 349–361.
- 157 A. C. Aprotosoiaie, M. Hăncianu, I. I. Costache and A. Miron, *Flavour Fragr. J.*, 2014, **29**, 193–219.
- 158 G. P. P. Kamatou and A. M. Viljoen, *Nat. Prod. Commun.*, 2008, **3**, 1183–1192.
- 159 R. M. C. Vasconcelos, F. C. Leite, J. A. Leite, S. Rodrigues Mascarenhas, L. C. Rodrigues and M. R. Piuvezam, *Immunopharmacol. Immunotoxicol.*, 2012, **34**, 1028–1038.
- 160 B. Shi, D. Luan, S. Wang, L. Zhao, L. Tao, Q. Yuan and X. Wang, *RSC Adv.*, 2015, **5**, 51947–51952.
- 161 K. Fidy, A. Fiedorowicz, L. Strządała and A. Szumny, *Cancer Med.*, 2016, **5**, 3007–3017.
- 162 V. Hanusova, H. Svobodová, M. Ambro, A. Skarka, N. Murínová, P. Tom and L. Skálová, *Biomed. Pharmacother.*, 2017, **95**, 828–836.
- 163 S. Qu, G. Wang, W. Duan, S. Yao, J. Zuo and C. Tan, *Bioorg. Med. Chem.*, 2011, **19**, 3120–3127.
- 164 J. J. Lu, Y. Y. Dang, M. Huang, W. S. Xu, X. P. Chen and Y. T. Wang, *J. Ethnopharmacol.*, 2012, **143**, 406–411.
- 165 A. V. Del Pacciaroni, E. Mongelli, L. A. Espinar, A. Romano, G. Ciccia and G. L. Silva, *Planta Med.*, 2000, **66**, 720–723.
- 166 A. Ziaei, M. Ramezani, L. Wright, C. Paetz, B. Schneider and Z. Amirghofran, *Phyther. Res.*, 2011, **25**, 557–562.
- 167 K. F. do Nascimento, F. M. F. Moreira, J. Alencar Santos, C. A. L. Kassuya, J. H. R. Croda, C. A. L. Cardoso, M. do C. Vieira, A. L. T. Góis Ruiz, M. Ann Foglio, J. E. de Carvalho and A. S. N. Formagio, *J. Ethnopharmacol.*, 2018, **210**, 351–358.
- 168 Y. C. Su, K. P. Hsu, E. I. C. Wang and C. L. Ho, *Nat. Prod. Commun.*, 2015, **10**, 1311–1314.
- 169 L. N. F. Trevizan, K. F. do Nascimento, J. A. Santos, C. A. L. Kassuya, C. A. L. Cardoso, M. do C. Vieira, F. M. F. Moreira, J. Croda and A. S. N. Formagio, *J. Ethnopharmacol.*, 2016, **192**, 510–515.
- 170 F. Bakkali, S. Averbeck, D. Averbeck and M. Idaomar, *Food Chem. Toxicol.*, 2008, **46**, 446–475.
- 171 N. J. Sadgrove, G. F. Padilla-González, A. Green, M. K. Langat, E. Mas-Claret, D. Lyddiard, J. Klepp, S. V. A. M. Legendre, B. W. Greatrex, G. L. Jones, I. M. Ramli, O. Leuner and E. Fernandez-Cusimamani, *Plants*, 2021, **10**, 1–41.
- 172 B. G. Passos, R. D. D. G. de Albuquerque, A. Muñoz-Acevedo, J. Echeverria, A. M. Llaure-Mora, M. L. Ganoza-Yupanqui and L. Rocha, *Fitoterapia*, 2022, **156**, 10565.
- 173 B. S. Jugreet, S. Suroowan, R. R. K. Rengasamy and M. F. Mahomoodally, *Trends Food Sci. Technol.*, 2020, **101**, 89–105.
- 174 M. Govindarajan and G. Benelli, *Ecotoxicol. Environ. Saf.*, 2016, **133**, 395–402.
- 175 S. Sharma, J. Gupta, P. K. Prabhakar, P. Gupta, P. Solanki and A. Rajput, *Assay Drug Dev. Technol.*, 2019, **17**, 339–351.
- 176 Razafiarimanga Zara Nomentsoa, Randriamampianina Lovarintsoa Judicael, Randrianarivo Hanitra Ranjàna, Ralitera Andrianirina Manampisoa, Rakoto Danielle Aurore Doll and Jeannoda Victor Louis, *GSC Biol. Pharm. Sci.*, 2021, **16**, 115–125.
- 177 R. S. de Almeida, J. Ribeiro-Filho, P. R. Freitas, A. C. J. de Araújo, E. L. dos Santos, S. R. Tintino, T. F. Moura, V. A. Ferreira, B. A. Ferreira, V. Juno Alencar Fonseca, P. I. P. Leite, A. C. Albuquerque da Silva, L. Everson da Silva, W. do Amaral, C. Deschamps, A. Siyadatpanah, P. Wilairatana and H. D. M. Coutinho, *J. Infect. Public Health*, 2022, **15**, 373–377.
- 178 F. P. Machado, D. Folly, J. J. Salas Enriquez, C. B. Mello, R. Esteves, R. S. Araújo, P. F. S. Toledo, J. G. Mantilla-Afanador, M. G. Santos, E. E. Oliveira, E. Ricci-Junior and L. Rocha, *Ind. Crops Prod.*, 2023, **192**, 116031.
- 179 V. M. Pathak, V. K. Verma, B. S. Rawat, B. Kaur, N. Babu, A. Sharma, S. Dewali, M. Yadav, R. Kumari, S. Singh, A. Mohapatra, V. Pandey, N. Rana and J. M. Cunill, *Front. Microbiol.*, 2022, **13**, 1–29.
- 180 E. V. R. Campos, P. L. F. Proença, J. L. Oliveira, M. Bakshi, P. C. Abhilash and L. F. Fraceto, *Ecol. Indic.*, 2019, **105**, 483–495.

4 CHAPTER II – PART A: ANTI-INFLAMMATORY METABOLOMICS STUDY OF *Ocotea* SPECIES

Chapter II containing the published research is separately provided as an attached material of this thesis.

Academic Journal: Phytomedicine (Elsevier) (published)

Impact factor (2024): 6.7

DOI: <https://doi.org/10.1016/j.phymed.2023.155060>

Phytomedicine 120 (2023) 155060



Contents lists available at ScienceDirect

Phytomedicine

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



Original Article

Bioprospecting-based untargeted metabolomics identifies alkaloids as potential anti-inflammatory bioactive markers of *Ocotea* species (Lauraceae)

Albert Katchborian-Neto^a, Karen de Jesus Nicácio^b, Jonas C. Cruz^c,
Paula Carolina Pires Bueno^{a,d}, Michael Murgu^e, Danielle F. Dias^a, Marisi G. Soares^a,
Ana C.C. Paula^f, Daniela A. Chagas-Paula^{a,*}

^a Institute of Chemistry, Federal University of Alfenas (UNIFAL), 37130-001, Alfenas, Minas Gerais, Brazil

^b Department of Chemistry, Federal University of Mato Grosso (UFMT), 78060-900, Cuiabá, Mato Grosso, Brazil

^c Department of Chemistry, University of São Paulo (USP), 14040-901, Ribeirão Preto, São Paulo, Brazil

^d Leibniz Institute of Vegetable and Ornamental Crops (IGZ), Theodor-Echtermeyer-Weg 1, 14979 Großbeeren, Germany

^e Waters Corporation, Alameda Tocantins 125, 27th floor, Alphaville, 06455-020, Barueri, São Paulo, Brazil

^f Department of Pharmaceutical Sciences, Federal University of Juiz de Fora (UFJF), 36036-900, Juiz de Fora, Minas Gerais, Brazil



ARTICLE INFO

Keywords:

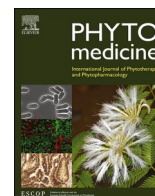
Prostaglandin
UPLC-ESI-QTOF
Metabolite annotation
Ocotea
Alkaloids
Anti-inflammatory

ABSTRACT

Background: Species within the *Ocotea* genus (Lauraceae), have demonstrated an interesting profile of bio-activities. Renowned for their diverse morphology and intricate specialized metabolite composition, *Ocotea* species have re-emerged as compelling candidates for bioprospecting in drug discovery research. However, it is a genus insufficiently studied, particularly regarding anti-inflammatory activity.

Purpose: To investigate the anti-inflammatory activity of *Ocotea* spp. extracts and determine the major markers in this genus.

Methods: Extracts of 60 different *Ocotea* spp. were analysed by an *ex vivo* anti-inflammatory assay in human whole blood. The experiment estimates the prostaglandin E2 levels, which is one of the main mediators of the inflammatory cascade, responsible for the classical symptoms of fever, pain, and other common effects of the inflammatory process. Untargeted metabolomics analysis through liquid chromatography coupled with high-resolution mass spectrometry was performed, along with statistical analysis, to investigate which *Ocotea* me-



Original Article

Bioprospecting-based untargeted metabolomics identifies alkaloids as potential anti-inflammatory bioactive markers of *Ocotea* species (Lauraceae)

Albert Katchborian-Neto^a, Karen de Jesus Nicácio^b, Jonas C. Cruz^c,
Paula Carolina Pires Bueno^{a,d}, Michael Murgu^e, Danielle F. Dias^a, Marisi G. Soares^a,
Ana C.C. Paula^f, Daniela A. Chagas-Paula^{a,*}

^a Institute of Chemistry, Federal University of Alfenas (UNIFAL), 37130-001, Alfenas, Minas Gerais, Brazil

^b Department of Chemistry, Federal University of Mato Grosso (UFMT), 78060-900, Cuiabá, Mato Grosso, Brazil

^c Department of Chemistry, University of São Paulo (USP), 14040-901, Ribeirão Preto, São Paulo, Brazil

^d Leibniz Institute of Vegetable and Ornamental Crops (IGZ), Theodor-Echtermeyer-Weg 1, 14979 Großbeeren, Germany

^e Waters Corporation, Alameda Tocantins 125, 27th floor, Alphaville, 06455-020, Barueri, São Paulo, Brazil

^f Department of Pharmaceutical Sciences, Federal University of Juiz de Fora (UFJF), 36036-900, Juiz de Fora, Minas Gerais, Brazil



ARTICLE INFO

Keywords:

Prostaglandin
UPLC-ESI-QTOF
Metabolite annotation
Ocotea
Alkaloids
Anti-inflammatory

ABSTRACT

Background: Species within the *Ocotea* genus (Lauraceae), have demonstrated an interesting profile of bio-activities. Renowned for their diverse morphology and intricate specialized metabolite composition, *Ocotea* species have re-emerged as compelling candidates for bioprospecting in drug discovery research. However, it is a genus insufficiently studied, particularly regarding anti-inflammatory activity.

Purpose: To investigate the anti-inflammatory activity of *Ocotea* spp. extracts and determine the major markers in this genus.

Methods: Extracts of 60 different *Ocotea* spp. were analysed by an *ex vivo* anti-inflammatory assay in human whole blood. The experiment estimates the prostaglandin E2 levels, which is one of the main mediators of the inflammatory cascade, responsible for the classical symptoms of fever, pain, and other common effects of the inflammatory process. Untargeted metabolomics analysis through liquid chromatography coupled with high-resolution mass spectrometry was performed, along with statistical analysis, to investigate which *Ocotea* metabolites are correlated with their anti-inflammatory activity.

Results: The anti-inflammatory screening indicated that 49 out of 60 *Ocotea* spp. extracts exhibited significant inhibition of PGE2 release compared to the vehicle ($p < 0.05$). Furthermore, 10 of these extracts showed statistical similarity to the reference drugs. The bioactive markers were accurately identified using multivariate statistics combined with a fold change (> 1.5) and adjusted false discovery rate analysis as unknown compounds and alkaloids, with a majority of aporphine and benzylisoquinolines. These alkaloids were annotated with an increased level of confidence since MS^E spectra were compared with comprehensive databases.

Conclusion: This study represents the first bioprospecting report revealing the anti-inflammatory potential of several *Ocotea* spp. The determination of their anti-inflammatory markers could contribute to drug discovery and the chemical knowledge of the *Ocotea* genus.

Abbreviations: ANOVA, Variance analyses; CAP, Chloramphenicol; CE, Crude extracts; CorrCoef, Correlation coefficients values; COX, Cyclooxygenase enzyme; DIA (MS^E), Data independent acquisition; DEX, Dexamethasone; DNP, Dictionary of natural products©; FDR, False discovery rate; GNPS, Global natural product social molecular networking; HPLC, High-performance liquid chromatography; HRMS, High-resolution mass spectrometry; ID, Identification number; IND, Indomethacin; LPS, Lipopolysaccharides; MS, Mass spectrometry; MSA, Multivariate statistical analyses; MF, Molecular formula; NSAIDs, Non-steroidal anti-inflammatory drugs; NP, Natural products; PBS, Phosphate-buffered saline; PCA, Principal component analysis; PGE2, Prostaglandin E2; PLS-DA, Partial least squares regression – discriminant analyses; PTFE, Polytetrafluoroethylene; RT, Retention time; SPE, Solid-phase extraction; SRM, Single reaction monitoring; UPLC-ESI-QTOF/MS, Ultra-performance liquid chromatography coupled to electrospray ionization source, and quadrupole time-of-flight/mass spectrometry; UVar, Unit variance; VIP, Variable important in projection.

* Corresponding author at: Institute of Chemistry, Federal University of Alfenas, Rua Gabriel Monteiro da Silva, 700, Alfenas-MG, 37130-001, Brazil.

E-mail address: daniela.chagas@unifal-mg.edu.br (D.A. Chagas-Paula).

<https://doi.org/10.1016/j.phymed.2023.155060>

Received 7 June 2023; Received in revised form 8 August 2023; Accepted 29 August 2023

Available online 3 September 2023

0944-7113/© 2023 Elsevier GmbH. All rights reserved.

Introduction

Natural products (NP) have been a valuable source of biologically active compounds all over the drug discovery history. The Lauraceae is a plant family that includes genera of high commercial and pharmaceutical importance worldwide, such as the *Cinnamomum*, *Persea*, *Laurus*, *Ocotea*, and others (Antonio et al., 2020; Gaudêncio and Pereira, 2015; Gottlieb, 1972; Marques, 2001). Renowned for their diverse morphology and intricate specialized metabolite composition, *Ocotea* species have reemerged as compelling candidates for bioprospecting in drug discovery research. With approximately 400 species, the *Ocotea* genus contains the potential to find therapeutic agents as demonstrated by important research in the literature (Brotto et al., 2013; de Alcântara et al., 2021; Marques, 2001; Salleh and Ahmad, 2017; Trofimov and Rohwer, 2020). Even though, the genus is considered few chemically and pharmacologically studied. The reasons for that knowledge gap can be attributed to the complex taxonomic issues in accurately identifying *Ocotea* species and the limited accessibility to these species, particularly in regions where they are naturally found (Antonio et al., 2020; Penagos Zuluaga et al., 2021).

The traditional people of Peru and Equator have used an extract of *O. quixos* as a local anesthetic and natural healing agent. Indeed, scientific investigations revealed anti-bacterial and anti-inflammatory properties for extracts of this specie (Bruni et al., 2004; Costa et al., 2015). In Brazilian traditional medicine, the decoction of *O. odorifera* leaves has been traditionally employed for treating inflammatory-related conditions, which has shown a dual mechanism of action that effectively inhibits both edema formation and neutrophil recruitment in vivo experiments (de Alcântara et al., 2021). Moreover, *O. diospyrifolia*, *O. bullata*, *O. notata*, and *O. cymbarum* have also exhibited anti-inflammatory effects or their isolated metabolites. These findings highlight the anti-inflammatory potential within the *Ocotea* genus, albeit limited to a small number of species when considering the overall size of this genus. Thus, it underscores the importance of further bioprospecting efforts to explore the therapeutic possibilities of *Ocotea* spp, particularly the anti-inflammatory potential (Costa et al., 2015; Silva et al., 2021; Zschocke et al., 2000; Ferreira et al., 2023).

Compounds inhibiting the arachidonic acid (AA) inflammatory signaling pathway, including non-steroidal anti-inflammatory drugs (NSAIDs), have been extensively studied in pharmaceutical research. NSAIDs exert their effects by targeting cyclooxygenase enzymes (COX), reducing the release of prostaglandins (PGs), such as prostaglandin E2 (PGE2), which play a crucial role in inflammatory processes (Leslie, 2015; Meirer et al., 2014; Parente, 2001). Despite progress in the last two decades, there remains a need for novel anti-inflammatory agents with enhanced efficacy, different mechanisms of action, and reduced risk of adverse reactions, given the limited long-term effectiveness of current treatments for numerous inflammatory disorders. Thus, ongoing investigations in this field aim to expand therapeutic options, addressing unmet medical needs and providing potential solutions for managing inflammation-related conditions (Chagas-Paula et al., 2015a; Chagas-Paula et al., 2015b; Ding et al., 2018; Funk, 2001; Meirer et al., 2014).

Metabolomics studies have emerged as a systematic approach to analysing endogenous metabolites in various samples, offering comprehensive insights into physiological, pathological, and pharmacological aspects (Alarcon-Barrera et al., 2022; Zhang et al., 2020). In metabolomics, a combination of analytical techniques, including liquid chromatography coupled with high-resolution mass spectrometry (LC–HRMS), are commonly used to investigate the metabolome. Which, by the way, encompasses all low-weight molecules found in samples (Fenaille et al., 2017; Vinayavekhin and Saghatelian, 2010; Zhang et al., 2020). One notable advancement in LC–HRMS is the application of data-independent acquisition (DIA), which enables unbiased and comprehensive analysis of metabolites in a single chromatographic run (Fenaille et al., 2017; Tsugawa et al., 2015).

The DIA-MS approaches offers improved data coverage, allowing for

the search of a wider range of metabolites in metabolomics studies, including a few abundant compounds from complex samples, such as crude plant extracts (Azmi et al., 2021; Rao Gajula and Nanjappan, 2021). Integrating metabolomics approaches such as data treatment and multivariate statistical analysis (MSA) it is possible to explore bioactive markers, uncover novel chemical compounds, and gain valuable biological insights into the metabolome (Xi et al., 2015; Zanatta et al., 2021). The continuous advancements in LC-MS instrumentation contribute to improved sensibility, resolution, and as a consequence also complexity, albeit expanding the possibilities for discovering novel compounds, and new bioactivities for those already known in a lesser amount of time (Ebbels et al., 2023; Perez de Souza et al., 2021; Spicer et al., 2017). Herein, we employed LC–HRMS-based metabolomics with the DIA (MS^E type) approach to explore bioactive markers of *Ocotea* species, with a focus on their ability to inhibit PGE2 release. This study is the first bioprospecting report on the chemical and bioactivity of several *Ocotea* species.

Material and methods

Plant materials, drugs, and reagents

Considering the taxonomic issues, complicated access to endemic *Ocotea* species, in addition to ecological aspects of threatened species in Brazil (Antonio et al., 2020; Martins et al., 2014; Penagos Zuluaga et al., 2021), 1 – 3 leaves were donated from 60 vouchers specimen deposited at Brazilians herbariums of Ouro Preto (OUPR, Federal University of Ouro Preto - UFOP) and Leopoldo Krieger - Centro de Ensino Superior de Juiz de Fora (CESJ, Federal University of Juiz de For a - UFJF). The research was registered on the National System for Governance of Genetic Heritage and Associated Traditional Knowledge (SisGen # A5A8F67). The *Ocotea* samples received an identification code (ID) according to their species names and herbarium of origin. The ID names and additional information are shown in Supplementary Table S1.

Dexamethasone and indomethacin were acquired from Eurofarma® (São Paulo, SP, Brazil). Chloramphenicol (CAP) and PGE2 standards, together with *E. coli* O26:B6 lipopolysaccharides (LPS) were purchased from Sigma Aldrich® (St Louis, MO, USA). All solvents used were high-performance liquid chromatography (HPLC) grade, including hexane, methanol, ethanol, and acetonitrile Sigma Aldrich® (St Louis, MO, USA). The formic acid was supplied by Sigma-Aldrich® (St Louis, MO, USA). Ultrapure water was purified using a Millipore Milli-Q® water purification system (Millipore, Bedford, MA, USA). The liquid nitrogen was obtained from Linde® (Pullach, Munique, Germany).

Crude extracts preparation

The 60 different *Ocotea* ssp. samples were crushed using pistil and liquid nitrogen until pulverization. To the powdered material (20 mg), 1.7 ml of ethanol/water, 7:3 (v/v) was added to extract the most polar and semi-polar compounds. The samples were placed in a warm ultrasound bath (35 °C) for 15 min (170 W, 50 kHz, L100 Schuster), and then centrifuged at 22 °C and 112 rcf (G-force). To remove fatty substances, the supernatants were partitioned with hexane (2 × 300 µl). After, the extracted samples were filtered through polytetrafluoroethylene (PTFE – pore size 22 µm) syringe filters and dried using vacuum equipment for 3 h at 40 °C. The samples were kept in a freezer (–20 °C) until the moment of the analyses. The dried extracts were weighted and the respective yields were calculated (Supplementary Table S2).

Ex-vivo anti-inflammatory evaluation

The *ex-vivo* anti-inflammatory experiment in human blood was performed according to (Rosa et al., 2021; Nicácio et al., 2022; Santos et al., 2022; Silva et al., 2023). The experiment was approved by the research ethics committee of the Federal University of Alfenas (89,325,

Table 1

The single reaction monitoring (SRM) parameters in the negative mode for PGE2 and the internal standard chloramphenicol (CAP). Q1, CE, and Q3 in eV.

Compound	Transition (<i>m/z</i>)	Q1	CE	Q3
CAP	320.90 → 152.10	16	20	29
PGE2	351.10 → 271.30	13	18	28

818.1.0000.5142, approved on 15th August 2022) and all the donors provided written informed consent before blood collection. The positive controls dexamethasone (DEX) and indomethacin (IND) were tested at 1 µg/ml (final well concentration). The *Ocotea* extracts were assayed at 10 µg/ml (final well concentration). A phosphate-buffered saline 1x (PBS - pH 7.2; 0.15 M chloride; 0.01 M phosphate) was used for dilutions. The LPS at 100 µg/ml (final well concentration) was used as the inflammation inducer agent. Negative control (PBS 1x) was performed to estimate the maximum amount of PGE2 produced in the inflammatory process. The plating sequence consisted of 25 µl of each *Ocotea* extract, 200 µl of blood in all wells and, 25 µl of LPS solution. Following, the 96-well plates were incubated for 24 h in a 5% CO₂ atmosphere at 37 °C. After incubation time, the plates were centrifuged for 5 min at 157 rcf and 4 °C. Then, 100 µl of plasma was removed from the blood of each well and then frozen until posterior analyses.

Prostaglandin E2 (PGE2) quantification by UPLC-MS/MS

The PGE2 quantification was done according to the developed and validated method (Rosa et al., 2021). 100 µl of each plasma sample was spiked with 500 µl of the precipitating agent (ACN/MeOH, 1:1, v/v) and centrifuged (4025 rcf, 4 °C) for 10 min. The supernatant was transferred to a polypropylene tube containing 4.5 ml of ultrapure water. The samples were loaded on an LC-18 cartridge of solid-phase extraction (SPE, 500 mg, #57,012 Supelco) after conditioning with 2 ml of MeOH, followed by 2 ml of acidified ultrapure water containing 0.1% of acetic acid. The cartridges were washed with 2 ml of aqueous solution with 0.1% of acetic acid, the PGE2 analyte was eluted into polypropylene tubes using a 1.8 ml methanolic solution with 0.1% acetic acid, and later the eluted methanolic fraction was evaporated to dryness. Chloramphenicol at the concentration of 25 ng/ml was used as an internal standard (IS).

The quantification was done by ultra-performance LC coupled to tandem mass spectrometry (UPLC-MS/MS) analyses carried out using system model 8030 (Shimadzu®, Kyoto, Japan) equipped with a triple-quadrupole mass analyser operating in the negative mode. The controls and samples were injected (20 µl) in the chromatographic system containing a column (1.7 µm, 2.1 × 100 mm, Kinetex® C18) maintained at 30 °C. The mobile phase consisted of (A) ultrapure water acidified with formic acid at 0.1% and (B) acetonitrile (100%) at a constant flow rate of 300 µl/min. The gradient elution was established as follows: 40% of B and 60% of A until 100% of B (0 – 3 min), which remained in this configuration for up to 4 min. After, the method returned to the initial mobile phase (40% of B) for 0.5 min, followed by 1 min of re-equilibration. The overall run time was 8.5 min. The source and MS parameters were as follows: nebulizing gas nitrogen at flow 2 l/min, drying gas nitrogen at flow 15 l/min, interface voltage 3.5 kV, DL temperature 250 °C, oven temperature 35 °C, detector voltage 2.44 kV and collision gas argon at 230 kPa. The transitions for the PGE2 and the IS are shown (Table 1).

Data acquisition was performed using LabSolutions® software. The linear equation of the calibration curve ($y = 0.04341 \cdot x + 0.07256$) obtained a correlation coefficient (R^2) = 0.9996, where y is the relative area of PGE2/IS and x is the concentration of PGE2 measured in ng/ml. The PGE2 concentration in ng/ml and the percentage of inhibition were calculated for each evaluated *Ocotea* crude extract sample. The results of PGE2 concentration were also expressed as the mean ± standard error (SD) and statistically analysed via *one-way* analysis of variance

(ANOVA) followed by Dunnett's multiple comparison tests on GraphPad Prism® 9.1.2 (GraphPad software©, La Jolla, CA, USA).

Metabolomic analysis – LC–HRMS

The analysis of plant extracts was performed on an ultra-performance liquid chromatography coupled to an electrospray ionization source and a quadrupole time-of-flight mass analyser (UPLC-ESI-QTOF, Xevo-qTOF/MS, Waters Corp., Milford, USA) instrument. The analysis was set for DIA, more specifically, the mass spectra of MS^E type from all precursor ions. DIA is an MS data acquisition approach that involves systematically fragmenting all detected ions in a sample without relying on predefined precursor ion selection. DIA provides comprehensive coverage of the metabolome by acquiring fragment ion spectra for all ions in the sample (Rosnack et al., 2016; Tsugawa et al., 2015; van der Laan et al., 2020).

The analysis was performed using Masslynx™ MS Software (Waters Corp., Milford, USA). An aliquot of 5 µl of each *Ocotea* extract sample was injected, and the separation occurred in a reversed-phase column (C18, 1.8 µm, 100 × 2.1 mm, ACQUITY UPLC® HSS T3) maintained at 40 °C. The mobile phases consisted of (A) acidified water with 0.1% formic acid, and (B) pure acetonitrile, delivered at a 0.5 ml/min flow rate. The chromatographic run was as follows: 1% of B in 0.1 min, 15% of B in 7.5 min, 80% of B in 8.5 min, 99% of B in 8.6 min, and 1% of B until 10 min.

The ESI operated in the positive and negative ionization modes. The mass spectrometer parameters were executed with alternative high and low-energy scans. The low CE was set at 3 eV and the high CE was ramped from 25 to 40 eV. Cone gas flow, 30 l/h; desolvation temperature, 300 °; source temperature, 120 °; and desolvation gas flow, 600 l/h. The mass scan range was set at m/z 50 to m/z 1000 for functions 1 and 2. The MS data were collected in profile mode, using the lock spray for calibration, to guarantee accuracy and reproducibility. Leucine-enkephalin was used as a lock mass, identified by the m/z 554.2622 (ESI⁻) and m/z 556.2768 (ESI⁺), which was acquired every 10 s.

For the chromatographic method development, an analytical quality control (QC) sample was prepared by gathering together 10 µl of each *Ocotea* sample extract (1 mg/ml). Samples were analysed randomly, with one replicate, one blank, and the QC at the beginning, middle, and end of the chromatographic batch.

Data treatment and chemical annotation, considering MS^E spectra for untargeted metabolomics, were performed using the UNIFI scientific information system software 1.8.1 (Waters Corp., Milford, USA), which combines data acquisition, processing, visualization, and compliance of the raw data (Rosnack et al., 2016; Wang et al., 2019). The raw data treatment using UNIFI software included peak detection, alignment processing algorithms for chromatographic peak corrections, deconvolution, deisotope, and baseline correction. A peak intensity threshold was set as 250 counts for high-energy fragment detection and 500 counts for low-energy. The noise was established as 10.000 peaks per channel and the error mass tolerance of 10 ppm. RT tolerance was defined as 0.3 min. For adduct search, the positive mode included Na⁺ and K⁺, and the negative mode Cl⁻, HCO₂⁻. The area was normalised regarding the total ion count, to generate a data matrix with the m/z value, retention time (RT) and the normalized peak area. The raw data of positive (ESI⁺) and negative (ESI⁻) modes of ionization were treated separately.

Univariate and multivariate statistical analyses (MSA)

The data containing peak area, m/z and RT processed by UNIFI software was exported in .xlsx Excel reading format. The spreadsheet was imported through MetaboAnalyst 5.0 software (Montreal, QC, Canada, <https://www.metaboanalyst.ca/>) for the MSA. The data were quantile normalized and mean-centered scaled before the unsupervised and supervised multivariate analysis. The metabolic profile was initially analysed by unsupervised statistical analyses using initially principal

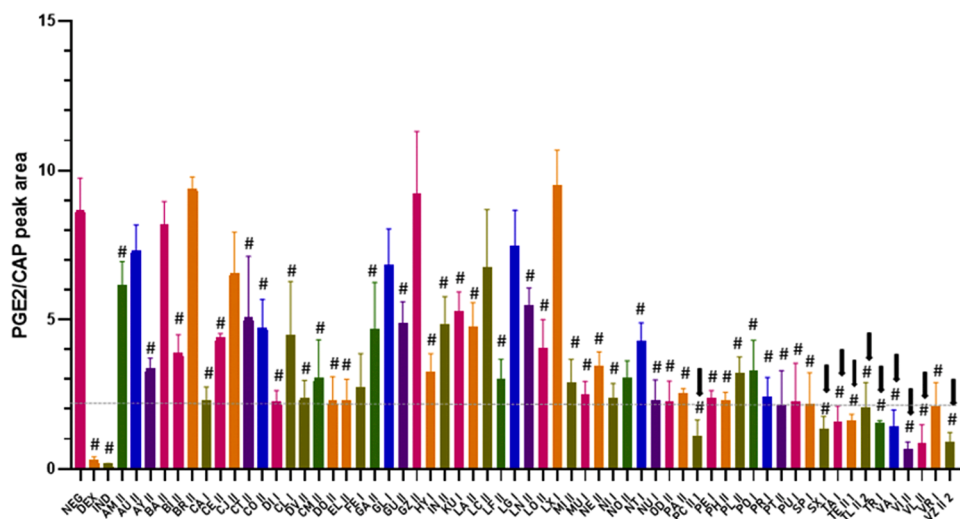


Fig. 1. Comparison of 60 *Ocotea* extracts and their *ex vivo* anti-inflammatory activity through PGE2 inhibition release screening levels ($n = 3$ replicates for samples and 6 for controls), in which 49 extracts were statistically different to the negative control (NEG, PBS + blood + LPS). Note: The results were analysed by *one-way* ANOVA, followed by Dunnett's multiple comparison test. Bars represent mean \pm SD. # indicates statistical difference to (NEG) control, where $p \leq 0.05$. Arrows indicate the 10 most active samples without a significant difference when compared to the positive control, the reference drugs: dexamethasone (DEX) and indomethacin (IND), $p > 0.05$.

component analyses (PCA), hierarchical cluster analyses (HCA), and heatmaps. Subsequently, data were supervised by the anti-inflammatory activity using the partial least squares regression - discriminant analysis (PLS-DA).

Variables important in projection (VIPs) with values >1 and positive correlation coefficients (CorrCoef) from PLS-DA were used to evaluate the difference between active and inactive samples (Chagas-Paula et al., 2015a; Cragg and Pezzuto, 2016; Roux et al., 2011). Permutation tests ($n = 100$) were additionally implemented to ensure the reliability of the metabolomics model. Moreover, for avoiding misleading potential metabolites from possible outlier samples in the MSA models, a univariate statistical method of classical volcano plot (CVP) was also applied. CVP is based on both p values from a t -test and fold-change (FC) values. Thus, for the CVP analysis, a metabolite was only considered statistically significant if the p -value < 0.05 and $\log_2 FC > 1.5$ (Kumar et al., 2018). In addition, the false discovery rate (FDR) method was applied to rank the obtained p -values from low to high, multiplying each p -value by the number of variables after the cut-off (p -value < 0.05 and $\log_2 FC > 1.5$, $n = 55$ variables for positive and $n = 32$ for negative), and dividing by their order of rank according to the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995; Zanatta et al., 2021)

Overall, the RT- m/z pairs referent the metabolites that acquired VIP > 1 , $\log_2 FC$ values > 1.5 , p -value < 0.05 , positive CorrCoef, and FDR (q -value ≤ 0.05) were considered potentially valid bioactive markers of this untargeted metabolomics study (Chagas-Paula et al., 2015a; Kumar et al., 2018; Wu et al., 2018; Zanatta et al., 2021).

Chemical annotation

Thus, through univariate and multivariate statistics, biological experimentation and metabolomic data, we were able to annotate potential bioactive markers. The use of MS^E for data acquisition was crucial in allowing us to detect and analyze all MSⁿ fragments from the precursor ions (MS¹) of the metabolites of interest. The UNIFI standard database library and the dictionary of natural products (DNP) were employed to assist in the annotation of the compounds (Rosnack et al., 2016; Zanatta et al., 2021). To enhance confidence in the annotation of these markers, their m/z values and fragmentation patterns were manually compared to those in the global natural product social molecular Networking (GNPS) spectral library as well as published MS/MS data in the literature. In addition, the strategies used for molecular formula (MF) determination were based on the scientific tool i-FIT (Norm), which is a core methodology for the analyses of small molecular weight metabolites obtained through TOF instrumentation. The list of the proposed elemental compositions was according to the highest

agreement of theoretical isotopic pattern and lower error of mass in mDa. The low i-FIT Norm score is better (Iglesias, 2013; Katchborian-Neto et al., 2020). In cases where alkaloids did not have matches in the GNPS database, their identification was based on a comprehensive review of electrospray ionization (ESI) fragmentation patterns as investigated by (Demarque et al., 2016; Qing et al., 2020) supported by spectral literature data.

Following the guidelines set by the metabolomics standards initiative (MSI), the identification level in this study was categorized as level two for the annotated alkaloids and three for a few other compounds. Level two identification requires the presence of spectral matching with an authentic reference standard, and thus the fragmentation patterns consistent with the proposed compound structure. This approach ensures increased chemical information regarding the annotated metabolites (Creek et al., 2014; Spicer et al., 2017; Sumner et al., 2007). On the other hand, level three identification relies on the match of mono-isotopic mass and chemical formula, providing precise chemical class determination, while not yielding detailed structural information. By employing both level two and level three identification approaches, this study ensures a comprehensive analysis of the metabolome.

Results and discussion

Anti-inflammatory *ex vivo* evaluation

The PGE2 release is one of the main targets of anti-inflammatory drugs with a main role in a wide range of inflammatory diseases that rely on the COX activation cascade pathway, and the release of PGE2 inhibition is a valid therapeutic strategy for drug development (Chini et al., 2020; Majumder et al., 2014; Salehifar and Hosseinimehr, 2016). The positive controls DEX (4.61 ± 2.57 ng/ml of PGE2) and IND (1.78 ± 0.59 ng/ml of PGE2) were considered statistically similar to each other and different from the negative control (NEG) (197.98 ± 22.07 ng/ml of PGE2) ($p < 0.0001$), giving confidence to the screening (Fig. 1). According to the *one-way* ANOVA and Dunnett's multiple comparisons post-test, 49 out of the 60 *Ocotea* extracts (81.7%) evaluated were statistically different from the negative control ($p < 0.0001$; Fig. 1 and Table 2), and thus able to inhibit of PGE2 release, thereby inhibiting the cyclooxygenase (COX) pathway, directly or indirectly. Thus, these results themselves support and corroborate the anti-inflammatory activity described by the genus. Furthermore, a subsequent post-test indicated 10 out of the 49 active *Ocotea* species showed no statistical difference compared to the reference anti-inflammatory drugs (DEX and IND; $p > 0.05$; Fig. 1).

Thus, these 10 *Ocotea* spp. were considered the most promissory anti-

Table 2
Percentage of PGE2 releasing inhibition for the 60 *Ocotea* species.

Samples IDs	% PGE2 inhibition	[PGE2] ng/ml	Samples IDs	%PGE2 inhibition	[PGE2] ng/ml
NEG	0	197.98 ± 22.07	LN II	36.9	124.91 ± 10.44
DEX	97.7	4.61 ± 2.58	LO II	53.6	91.78 ± 17.60
IND	99.1	1.78 ± 0.59	LX I	0	258.94 ± 61.10
AM II	29.5	139.54 ± 15.22	MI II	67.3	64.70 ± 14.81
AU II	15.8	166.57 ± 16.21	MU I	71.6	56.18 ± 7.77
AY II	61.8	75.57 ± 6.49	NE II	60.8	77.55 ± 8.95
BA II	5.5	186.99 ± 14.29	NI I	73.0	53.38 ± 8.84
BI II	57.0	84.99 ± 13.05	NO II	65.5	68.38 ± 10.87
BR II	0	214.52 ± 6.33	NT I	51.1	96.80 ± 11.45
CA I	74.1	51.33 ± 10.71	NU I	74.0	51.41 ± 12.65
CE II	49.9	99.22 ± 2.82	OD II	78.8	41.94 ± 15.74
CJ II	30.4	137.81 ± 21.72	PA II	71.4	56.68 ± 2.92
CL I	51.9	95.15 ± 38.87	PC II 1	87.9	23.89 ± 9.87
CM II	70.7	57.92 ± 16.28	PE I	73.2	53.07 ± 4.58
CO II	43.9	111.05 ± 19.61	pH II	74.3	50.91 ± 5.23
CT II	42.0	114.82 ± 38.67	PL II	63.6	72.04 ± 10.26
DI I	76.4	46.75 ± 5.76	PO I	62.5	74.27 ± 18.94
DO II	74.5	50.46 ± 15.36	PR I	72.8	53.84 ± 12.24
DV II	73.2	53.00 ± 10.99	PT II	76.2	47.16 ± 23.32
EL II	74.1	51.31 ± 13.00	PU I	74.8	49.85 ± 24.48
FE I	68.9	61.44 ± 20.97	SP I	75.5	48.49 ± 20.77
GA II	40.4	117.89 ± 45.33	SX I	85.2	29.20 ± 7.65
GL I	28.7	141.23 ± 26.33	TA I	82.2	35.11 ± 9.43
GU II	5.0	188.07 ± 109.60	TE II 1	81.9	35.72 ± 3.76
GZ II	13.3	171.56 ± 62.51	TL II 2	80.1	39.41 ± 13.07
HY I	63.6	72.09 ± 14.02	TR I	83.1	33.50 ± 1.52
IN II	44.3	110.28 ± 17.12	VA I	84.5	30.65 ± 12.44
KU I	39.5	119.73 ± 12.23	VI II	92.8	14.09 ± 5.09
LA II	45.4	108.06 ± 15.11	VL II	90.7	18.37 ± 12.36
LC II	22.1	154.21 ± 36.03	VR I	76.4	46.70 ± 16.32
LF II	65.9	67.36 ± 12.66	VZ II 2	90.2	19.37 ± 6.18
LG I	24.1	150.21 ± 32.54			

inflammatory species ($p > 0.05$ compared to current anti-inflammatory drugs): PC II 1 – *O. pulchella*, SX I – *O. spixiana*, TA I – *O. tabacifolia*, TE II 1 – *tenuiflora*, TL II – *O. teleiandra*, TR I – *O. tristis*, VA I – *O. vaccinioides*, VI II – *O. villosa*, VL II – *O. velutina* and VZ II 2 – *O. velloziana*. To date, this is the first anti-inflammatory evidence for those species, besides the other 35 active *Ocotea* spp, except for the *O. odorifera* (78.8% PGE2 inhibition), *O. pretiosa* (76.2% PGE2

inhibition), *O. diospyrifolia* (74.5% PGE2 inhibition), and *O. notata* (65.5% PGE2 inhibition), which had previous scientific evidence of anti-inflammatory activity (Costa et al., 2015; de Alcântara et al., 2021; Silva et al., 2021), and thus was also corroborated by our results. Therefore, by using an *ex vivo* anti-inflammatory activity we have demonstrated for the first time in literature the anti-inflammatory potential of other 45 different *Ocotea* spp.

Principal component analyses (PCA)

Normalization is a critical step in metabolomics analysis that aids to mitigate systematic bias, such as the peak areas from different analytical measurements before the statistical analysis. By removing these variations, the data matrix is transformed into a more Gaussian-type distribution, allowing more accurate comparisons of metabolites among the different samples of the dataset (Bartel et al., 2013; Katajamaa and Orešić, 2007). Thus, after quantile normalization and mean centering the data, PCA was appropriately applied to visualize the general samples clustering trends of the dataset and to verify the metabolomic characteristics of the extracts, independently of their class. In the score plot, the samples near each other have similar metabolite content and thus are clustered together, whereas observations far away from each other are dissimilar (Trygg et al., 2007; Xi et al., 2015). The PCA can assist in reducing the dimensionalities of complex datasets and can provide discrimination of the score observations, by evidencing groupings, and outliers and by ensuring reproducibility (Chagas-Paula et al., 2015a; Yuliana et al., 2011).

The PCA score plot analysis demonstrated that the BR II and CL I samples are outliers in our model, and thus they were excluded from the subsequent multivariate analysis. Moreover, the reproducibility of data in metabolomics experiments is crucial to ensure reliable results. To check that, after filtering the data (mean intensity values), the 3D PCA scatter plots were analysed. It indicated the analytical replicates (QC controls and VI II) clustered at near spots to each other in both negative and positive ionization modes, respectively (Fig. 2). The QC samples clustered close to the plot center for all PC components indicated the satisfactory fitness of the unsupervised metabolomics exploratory analysis since is as a mix of samples. Thus, these results taken together demonstrated that the analyses were reproducible and the data treatment was accurate (Fig. 2). The PCA, with 5 components, obtained $R^2 = 0.546$ and 0.526 (ESI^+ and ESI^-) indicating that the models were well-fitted since R^2 was > 0.5 (Chagas-Paula et al., 2015a; Yuliana et al., 2011; Zanatta et al., 2021).

Partial least square – discriminant analyses (PLS-DA)

In this study, the MSA approach was employed to visualize the chemical differences among the samples and identify potential markers for differentiation into two groups: active and inactive samples. Specifically, PLS-DA was applied as it is a powerful dual matrix classification model that establishes fundamental relationships between multiple variables using linear regression models. In our case between the metabolomic data (X data, m/z -RT pair and peak area) and the Y variables (active and inactive samples; Fig. 3). As active samples were considered the 10 most active extracts, which showed significant anti-inflammatory effect compared to the negative group ($p < 0.05$, Fig. 1) and similar effect compared to reference anti-inflammatory drugs ($p > 0.05$, Fig. 1). As inactive samples were considered those samples without any anti-inflammatory activity, which did not show significant difference from negative control ($p > 0.05$, Fig. 1). This method enabled us to effectively to identify the key discriminant features in metabolomic profiles according to the biological response of interest (Chagas-Paula et al., 2015a; Xi et al., 2015; Yuliana et al., 2011).

The PLS-DA was applied and the three-dimensional data (m/z / RT / peak area) were supervised by the anti-inflammatory profile of the samples from *ex vivo* results. The developed model with 4 components

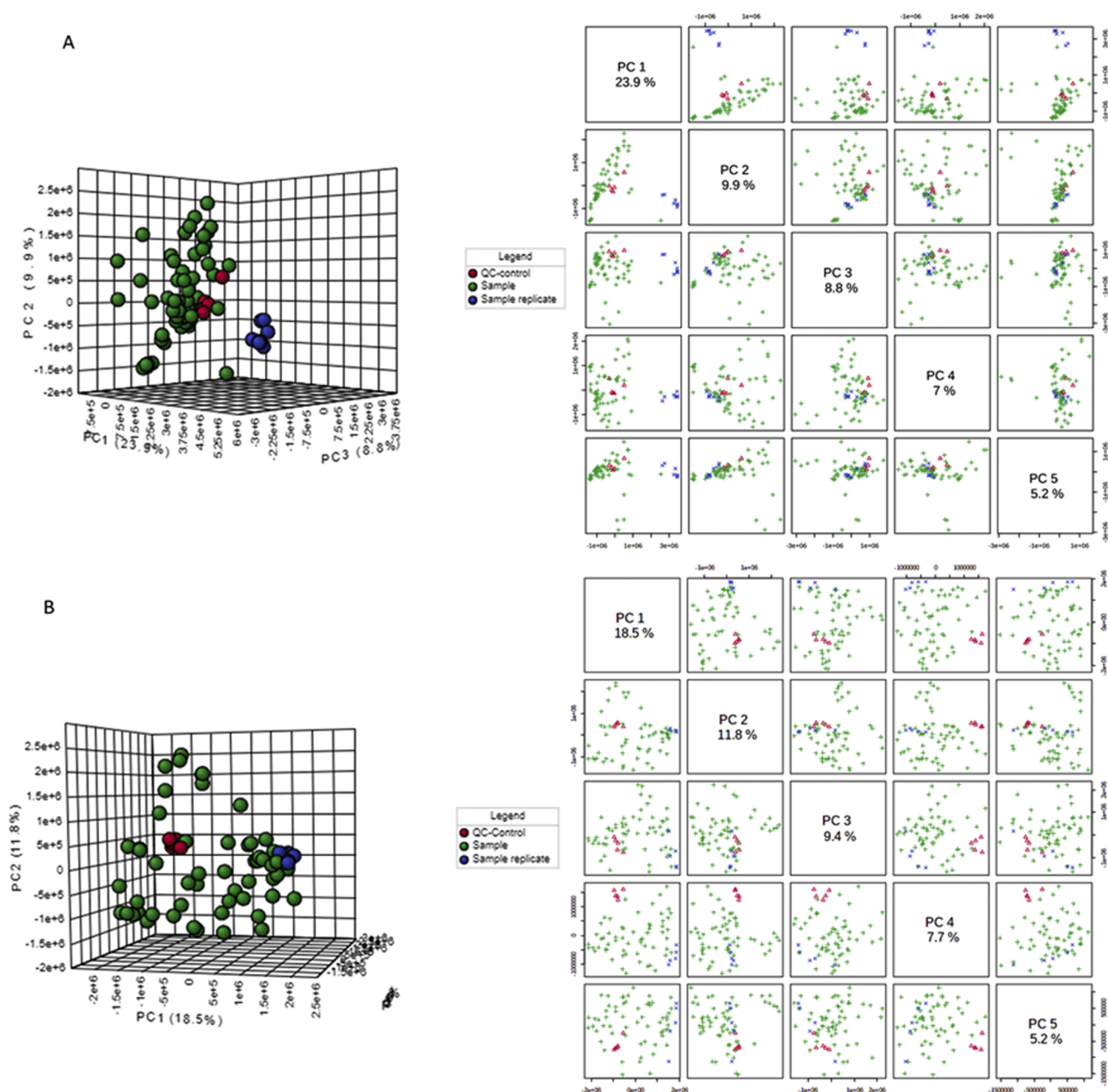


Fig. 2. Score plot of principal component analyses (PCA) of metabolomic data of UPLC–HRMS analysis in ESI⁺ (A) and ESI⁻ (B), hotelling ellipse = 95%, 5 components, $R^2 = 0.546$ and $R^2 = 0.526$, respectively. QC – Quality control; VI II – *O. villosa* replicates.

and 2 components were well fitted for ESI⁺ for ESI⁻ data, respectively, since obtained $R^2 = 0.968$ and 0.940 . Also, the models showed to be robust to prediction since obtained $Q^2 = 0.581$ and 0.509 , respectively for ESI⁺ for ESI⁻ data. For metabolomics studies, considering their complex and large dataset, values of $Q^2 > 0.5$ indicate robustness (Chagas-Paula et al., 2015a; Yuliana et al., 2011).

In parallel, Hotelling's T2 ellipse in a PLS-DA score plot allowed the identification of two active outliers in the negative mode (TE II 1 and TL II 2) and (VA I and VL II), in the positive ionization mode. Thus, these samples were excluded from MSA (Fig. 3). The further 100-permutation tests performed allowed the estimation of adjusted significance levels tests ($p = 0.04$) independently of the metabolomic variables and ensured that the model validity and predictability is not over-fitted (Peluso et al., 2021). Additionally, it was possible to determine the VIP values, which are the m/z -RT pair of the dataset. Those VIP values > 1 are estimated as the most important variables to differentiate the samples according to

their respective class, in our case, active and inactive (Chagas-Paula et al., 2015a; Katchborian-Neto et al., 2020; Yuliana et al., 2011).

Univariate statistical analysis – volcano plot

Metabolomics studies commonly generate a substantial amount of complex data that requires computational mining for proper interpretation. After data acquisition through hyphenated analytical techniques, the processing through analytical software is mandatory, which generally requires MSA for evaluation of the high metabolomic coverage. These steps aim to interpret the proper correlation of the biological response with minimal possible bias (Chagas-Paula et al., 2015a; Kosmides et al., 2013; Peluso et al., 2021).

The volcano univariate analysis is a robust statistical approach that can be integrated into MSA to reach an appropriate threshold cut-off for the annotated bioactive markers and can be used to find potential

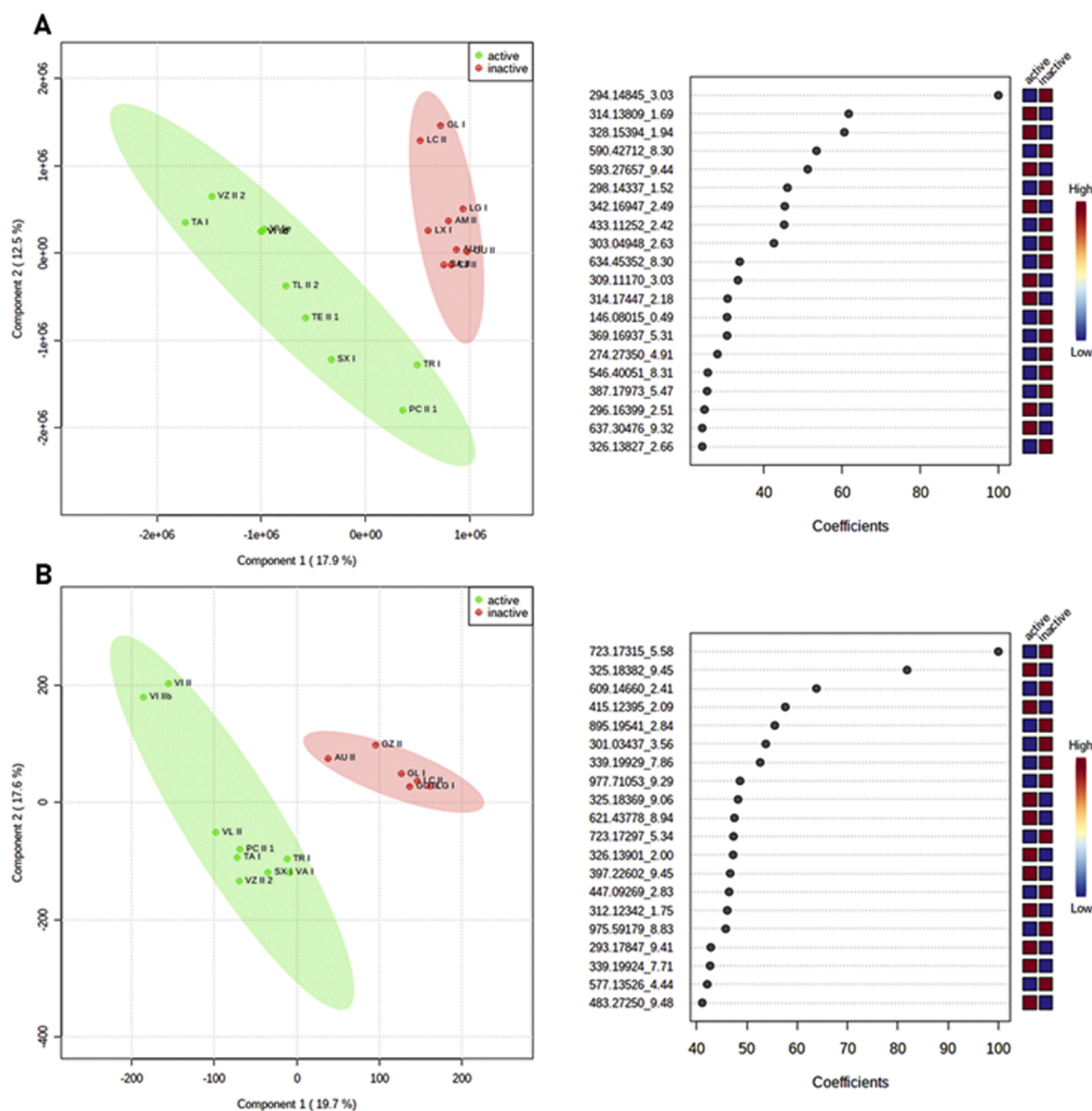


Fig. 3. Score plot of PLS-DA model indicating the discrimination of the active (green) and inactive sample groups (red), quartile normalized. (A) Positive mode: $R^2 = 0.968$ and $Q^2 = 0.581$. (B) For negative mode $R^2 = 0.940$ and $Q^2 = 0.509$. The highest fifteen correlation scores are displayed for m/z / RT pairs among the two groups. Note: QC – Quality control; VI II – *O. villosa* replicates.

outliers (Peluso et al., 2021). The classical volcano plot performed corroborated the bioactive markers pointed out by MSA and indicated 3 additional metabolites, two in the positive (m/z 386.39861_RT 7.43 min; m/z 189.06885_RT 3.04 min) mode and one in the negative (m/z 395.15510_RT 2.31) that acquired p -value < 0.05 and \log_2 FC values > 1.5 . Although, the ions in the positive mode at m/z 189 and m/z 386 scored VIP values < 1 . Thus, that explains why they have not figured out among the MSA positively correlated bioactive markers, besides their positive correlation with the anti-inflammatory activity. In addition, no hits on the searched databases were found for the ions at m/z 386 ($M + H$)⁺ and at m/z 395 ($M-H$)⁻, indicating potential unknown compounds (Fig. 4; Table 4).

Additionally, a heatmap and FDR calculations were performed for all significant metabolites from univariate and multivariate statistical analysis combined ($n = 17$, $FDR < 0.05$) between the active and the inactive anti-inflammatory groups of *Ocotea* spp. (Fig. 5, Tables 3 and 4). A clear separation in metabolomic profiles was observed, as indicated by the red and green colours (Fig. 5), representing higher and lower-peak areas. In addition, the chemical scaffolds of the bioactive markers

level annotated 2 by MSA are shown.

Furthermore, Fig. 6 demonstrates the overlapped LC–HRMS metabolic fingerprint from the 10 most active *Ocotea* species. The fingerprints are displayed as total ion chromatograms (TIC) in the positive mode, as based peak ions (BPI). The classes of the main annotated bioactive markers annotated at confidence level 2 according to MSI are demonstrated, including three alkaloid subclasses: aporphines (dehydronuciferine, laurelliptine, boldine, dicentrine and caaverine), benzyloquinolines (armepavine and laudanine), and the phenanthrenes (argentinine). Details of the biomarkers including m/z , MF, mass error, VIP values, and adjusted q -values are shown in Table 3 (Level 2 annotation) and Table 4 (Level 3 annotation + unknown).

The literature is filled with significant *in vitro* and *in vivo* studies evidencing the anti-inflammatory and antipyretic effects of boldine, as well as other benzyloquinoline and aporphine alkaloids (Backhouse et al., 1994; Peng et al., 2019; Yang et al., 2018). The aporphine isomers of boldine and laurelliptine aporphines were annotated in both ionization modes for different *Ocotea* species and were indicated by MSA as bioactive markers. Our findings support a recent study highlighting the

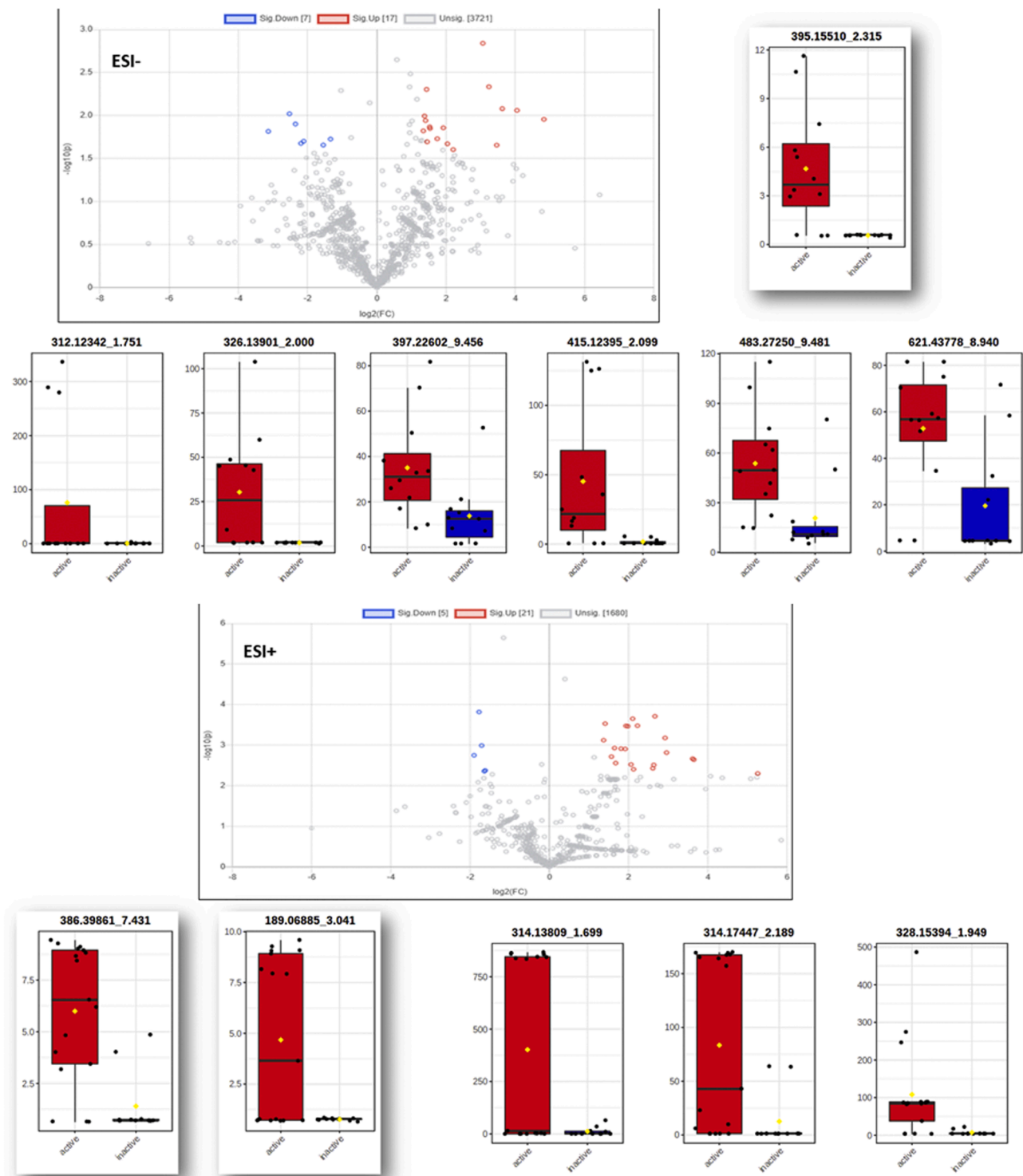
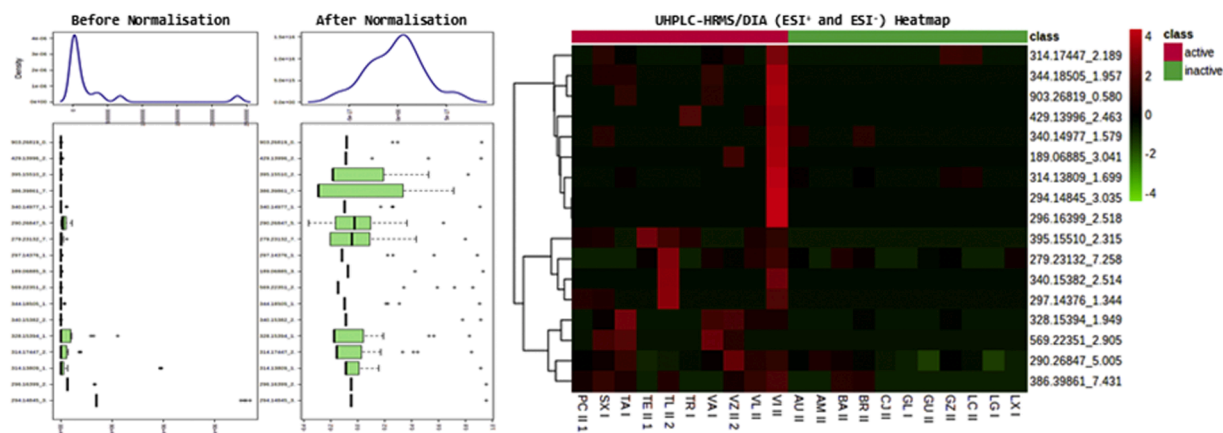


Fig. 4. Classical volcano plots were generated for both positive and negative modes using the original PLS-DA dataset obtained from UPLC—HRMS data processing. The univariate statistical analysis revealed three new markers, one in the negative mode and two in the positive mode (395.1551, m/z 386.3986, and m/z 189.0685, respectively). Additionally, the other 14 previously identified bioactive markers by MSA were confirmed. The statistical cut-off for considering metabolites as significantly different in the volcano plot was defined as fold change > 1.5 and p -value < 0.05 .

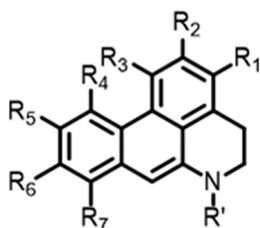
anti-inflammatory effect of reticuline, a benzyloisoquinoline alkaloid, in combination with boldine. The study demonstrated that this combination reduced paw edema induced by carrageenan and inhibited PGE2 in a time and dose-dependent manner. Notably, the combined administration of boldine and reticuline exhibited a synergistic effect, leading to enhanced efficacy compared to their individual use of the alkaloids (Backhouse et al., 1994; Peng et al., 2019; Yang et al., 2018).

In this study, boldine exhibited a positive correlation with anti-inflammatory activity and VIP values > 1 . On the other hand, reticuline showed a positive correlation but had a lower VIP value of 0.56 in the PLS-DA model of the positive mode, and a p -value > 0.05 in the univariate analysis. This suggests that the presence of reticuline alone does not strongly differentiate the active and inactive groups, and its concentration varied significantly among samples. While reticuline



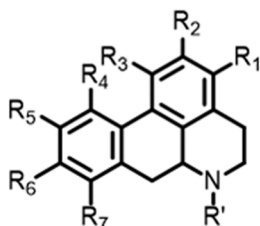
ANTI-INFLAMMATORY ANNOTATED BIOACTIVE MARKERS

APORPHINES



1 - Dehydronuciferine - m/z 294.1484 / RT 3.03 min - M+H

R_1 -H; R_2 -OCH₃; R_3 -OCH₃; R_4 -H; R_5 -H; R_6 -H; R_7 -H; R' -CH₃



3 - Laurelptine - m/z 314.1381 / RT 1.69 min - M+H

R_1 -H; R_2 -OCH₃; R_3 -OH; R_4 -H; R_5 -OCH₃; R_6 -OH; R_7 -H; R' -H

5 - Boldine - m/z 328.1539 / RT 1.94 min - M+H

R_1 -H; R_2 -OH; R_3 -OCH₃; R_4 -H; R_5 -OCH₃; R_6 -OH; R_7 -H; R' -CH₃

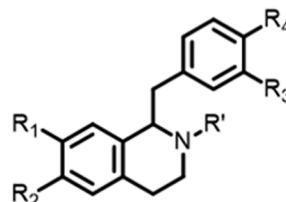
6 - Dicentrine - m/z 340.1538 / RT 2.51 min - M+H

R_1 -H; R_2/R_3 -OCH₂O; R_4 -H; R_5 -OCH₃; R_6 -OCH₃; R_7 -H; R' -CH₃

8 - Caaverine - m/z 266.1098 / RT .2.90 min - M-H

R_1 -H; R_2 -OCH₃; R_3 -OH; R_4 -H; R_5 -H; R_6 -H; R_7 -H; R' -H

BENZYLISOQUINOLINES



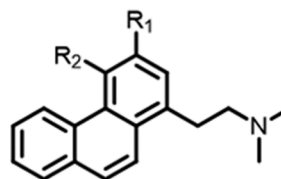
4 - Armepevine - m/z 314.1744 / RT 2.18 min - M+H

R_1 -OCH₃; R_2 -OCH₃; R_3 -H; R_4 -OH; R' -CH₃

7 - Laudanine - m/z 344.1850 / RT 1.96 min - M+H

R_1 -OCH₃; R_2 -OCH₃; R_3 -OH; R_4 -OCH₃; R' -CH₃

PHENANTRENES



2 - Argentinine - m/z 296.1639 / RT 2.50 min - M+H

R_1 -OH; R_2 -OCH₃

Fig. 5. Heatmap plot for the 17 differential markers of *Ocotea* species (red indicates high expression and green indicates low expression) using normalised peak area (square root transformation and range scaling by mean-centered and divided by the range of each variable). Representation of each chemical scaffold annotated is shown: aporphine, benzylisoquinoline and phenanthrene alkaloids, indicated by both univariate and multivariate analysis.

showed a positive correlation with anti-inflammatory activity, it is not one of the main compounds strongly associated with these effects. On the other hand, the chemically similar benzylisoquinolines, armepevine and laudanine, demonstrated VIP values greater than 1, a positive correlation, and a significant q -value below 0.05 (FDR). In addition,

regarding the chemical composition of the extracts, it is becoming clear that isoquinoline derivatives, especially benzylisoquinoline and aporphines represent an interesting and potentially useful group of anti-inflammatory potential scaffold (Fig. 5 and Table 3). It is worth noting that the anti-inflammatory activity of *Ocotea* species may be attributed

Table 3

The list of potential bioactive markers annotated at a Level 2 confidence level positively correlated with anti-inflammatory activity by inhibiting the release of PGE2. The 'Sample' column indicates the *Ocotea* extract with the highest detector counts associated with each respective bioactive marker.

ID	Sample [#]	Observed <i>m/z</i>	RT ^a	DC ^b	VIP ^c	FDR ^d (<i>q</i> -value)	MF ^e	Error (mDa)	Adduct	I-FIT	Compound Class	Proposed Annotation (Level 2)	Observed product-ions (<i>m/z</i>)	Source
1	VI II	294.1484	3.03	7.2 × 10 ⁶	15.45	8.410E-03	C ₁₉ H ₁₉ NO ₂	0.6	<i>M + H</i>	0.2	Aporphine alkaloids	Dehydronuciferine	249.0907; 219.0806; 191.0837	Proposed
2	VI II	296.1639	2.50	3.5 × 10 ⁴	4.23	1.22E-02	C ₁₉ H ₂₁ NO ₂	0.4	<i>M + H</i>	0.4	Phenanthrene alkaloids	Argentinine	251.1056; 236.0832; 219.0806; 208.0856; 191.0837	GNPS
3	VI II	314.1381	1.69	2.8 × 10 ⁴	10.54	9.36E-03	C ₁₈ H ₂₀ NO ₄	0.5	<i>M + H</i>	0.2	Aporphine alkaloids	Laurelliptine	298.1082; 283.0838; 255;0879	GNPS
4	VZ II 2	314.1744	2.18	2.5 × 10 ⁴	4.70	4.84E-02	C ₁₉ H ₂₃ NO ₃	0.6	<i>M + H</i>	0.2	Benzylisoquinoline alkaloids	Armepavine	298.1070; 271.1328; 269.1160; 107.0495; 58.0675	GNPS
5	PC II 1	328.1539	1.94	3.4 × 10 ⁴	10.81	1.29E-02	C ₁₉ H ₂₃ NO ₄	0.3	<i>M + H</i>	0.0	Aporphine alkaloids	Boldine	297.1028; 265.0842; 237.0899; 205.0641	GNPS
6	TL II 2	340.1538	2.51	1.2 × 10 ⁴	1.71	9.63E-03	C ₂₀ H ₂₁ NO ₄	0.2	<i>M + H</i>	0.3	Aporphine alkaloids	Dicentrine	309.1108; 279.1030; 264.0762; 251.1056	GNPS
7	VI II	344.1850	1.96	1.6 × 10 ⁵	1.03	2.08E-02	C ₂₀ H ₂₅ NO ₄	0.2	<i>M + H</i>	0.4	Benzylisoquinoline alkaloids	Laudanine	298.1082; 206.0727; 189.0678; 174.0905; 137.0596	Proposed
8	VA I	569.2235	2.90	2.6 × 10 ³	1.35	4.99E-02	C ₃₄ H ₃₄ N ₂ O ₄	1.6	2M+Cl	0.9	Aporphine alkaloids	Caaverine	266.1098; 234.0754	Proposed

^a Retention Time,.

^b Detector counts,.

^c Variant Important in Projection,.

^d False Discovery Rate,.

^e Molecular formula. For those alkaloids of unknown available spectra, we have followed [Qing et al., 2020](#) and [Demarque et al., 2016](#) fragmentation mechanism elucidation to propose the fragment ions (metabolites 1, 7 and 8). Spectra of all annotated biomarkers are attached to Supplementary data including MS^E spectra of low and high-energy scans (Fig. S1-S8).

Note:.

[#] The table displays only the *Ocotea* sample in which the bioactive marker was found with the highest peak area.

Table 4

The list of potential bioactive markers annotated at a *Level 3 confidence level positively correlated with anti-inflammatory activity by inhibiting the release of PGE2. The 'Sample' column indicates the *Ocotea* extract with the highest detector counts associated with each respective bioactive marker. The unknown compounds also indicated by MSA are discriminated.

Sample ^f	Observed m/z	RT ^a	DC ^b	VIP ^c	FDR ^d (q-value)	MF ^e	Error (mDa)	Adduct	Hits (DNP ^f)	Proposed Compound class	Proposed annotation (*Level 3)	Isolated in <i>Ocotea</i> sp.?
VI II	189.0688	3.04	4.4 × 10 ⁴	*0.31	1.45E-02	C ₁₀ H ₈ N ₂ O ₂	2.4	M + H	15	Quinazolinones	e.g. *8-Methylquinazoline-4-carboxylic acid	No [#]
VI II	297.1437	1.34	3.0 × 10 ⁴	1.67	9.30E-03	C ₁₅ H ₂₄ O ₃	0.2	M + 2Na	298	Sesquiterpenoids	* (rel)-4β,5β,7β Eremophil-1(10)-en-2-oxo-12-oic acid	Yes. <i>O. insularis</i> , <i>O. holdrigeana</i>
TL II 2	279.2313	7.26	2.4 × 10 ⁵	1.19	2.43E-02	C ₁₈ H ₃₀ O ₂	0.8	M + H	70	Fatty acids	-	No
VZ II 2	290.2684	5.00	4.5 × 10 ⁵	1.99	2.01E-03	C ₁₆ H ₃₅ NO ₃	0.7	M + H	1	Unknown	Unknown	No
VI II	340.1498	1.58	2.3 × 10 ⁴	1.43	4.73E-02	C ₁₇ H ₂₃ O ₇	1.9	M + H	0	Unknown	Unknown	No
VL II	386.3986	7.43	2.8 × 10 ⁴	*0.41	6.425E-03	C ₂₄ H ₅₁ NO ₂	0.8	M + H	15	Unknown	Unknown	No
TL II 2	395.1551	2.31	1.2 × 10 ⁴	1.52	4.62E-02	C ₁₆ H ₂₈ O ₁₁	0.6	M-H	3	Sugar derivatives	-	No
TR I	429.1399	2.46	3.2 × 10 ⁴	2.27	4.991E-02	C ₁₉ H ₂₆ O ₁₁	0.2	M-H	16	Iridoid glycosides	-	No
VI II	903.2682	0.58	2.6 × 10 ⁴	1.34	4.660E-02	C ₅₆ H ₄₂ NO ₁₁	0.2	M-H	0	Unknown	Unknown	No

Source: From the author.

^a Retention Time,.

^b Detector counts,.

^c Variable Important in Projection,.

^d False Discovery Rate,.

^e Molecular formula,.

^f Dictionary of Natural Products.

Note:

[#] The table displays only the sample in which the bioactive marker was found with the highest peak area. Other active samples also have these bioactive markers.

* Metabolites that acquired VIP < 1, although were statistically different by volcano plot analysis. [#] Not a reliable hit, * Previously described in the literature in *Ocotea* genus: *O. lancifolia* (De Camargo et al., 2013).

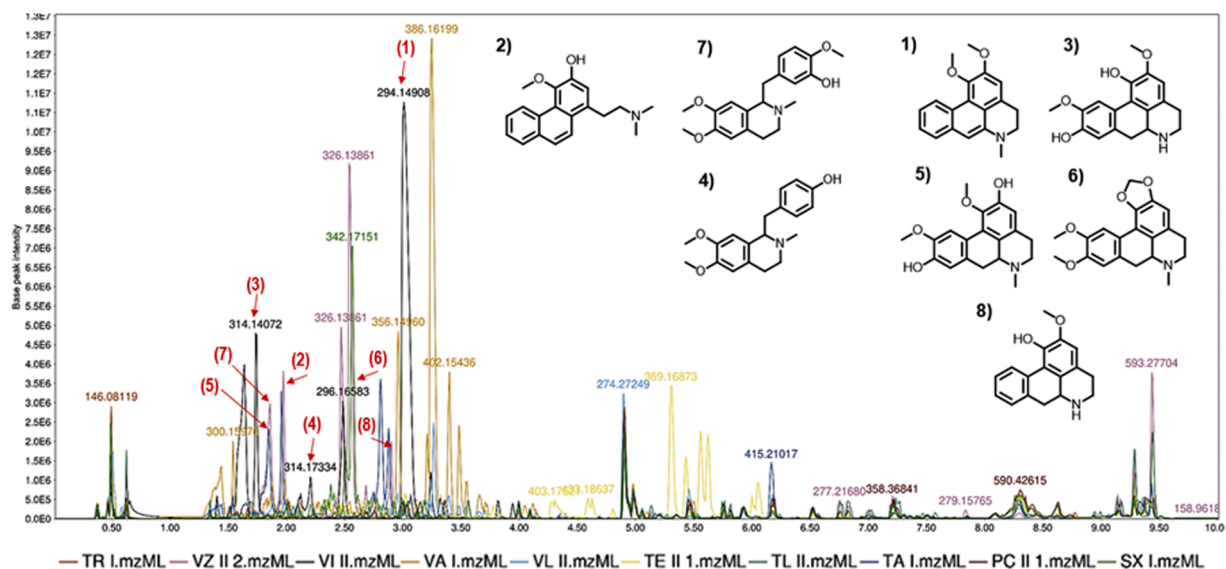


Fig. 6. LC-MS/MS-ESI-DIA metabolic fingerprints of total ion chromatograms (TIC) in the positive mode showing the overlapped metabolic fingerprint of the 10 most active *Ocotea* species, highlighting the presence of 8 alkaloids (1-dehydronuciferine, 2-argentinine, 3-laurelliptine, 4-armepavine, 5-boldine, 6-dicentrine, 7-laudanine and 8-caaverine).

to more than one bioactive marker, indicating that compounds could be acting synergistically to produce pronounced anti-inflammatory effects on the crude extract of these plants. Therefore, further studies are needed to explore the possible synergistic effects of these metabolites.

Our study corroborates the *Ocotea* species for the search of anti-inflammatory bioactive compounds, and also to the understanding of the anti-inflammatory properties of *Ocotea* species. Statistical analysis identified 10 extracts, including *O. pulchella*, *O. teleiandra*, *O. tenuiflora*, *O. spixiana*, *O. tabacifolia*, *O. tristis*, *O. vaccinioides*, *O. villosa*, *O. velutina*, and *O. velloziana*, with high PGE2 inhibition release. These extracts hold significant potential as novel sources for developing plant-based anti-inflammatory treatments. With many inflammatory diseases still poorly managed, NP offers untapped resources for discovering new metabolites and biological activities with anti-inflammatory properties (Peng et al., 2019; Yuliana et al., 2011). Our study underscores the importance of exploring natural chemical biodiversity to expand the repertoire of anti-inflammatory drugs.

Regarding the metabolomics results, the R² and Q² values obtained suggest that the metabolomics model is robust and reliable enough for identifying anti-inflammatory bioactive markers. The additional permutation tests performed ensured the absence of overfitting. According to the MSA and univariate statistical results mainly aporphine and benzylisoquinoline isomers are the main correlated metabolites of *Ocotea* sp. with the PGE2 inhibition release. Thus, as a final point, under a holistic approach in terms of NP, our untargeted metabolomic reach to candidates of anti-inflammatory activity. The study was performed with a non-time-consuming isolation step, which is generally mandatory for classic phytochemical analyses (Wolfender et al., 2019; Yuliana et al., 2011). In addition, this work is the first research regarding the chemical and biological activity of several endemic and endangered *Ocotea* species.

Conclusions

In this research, we have successfully addressed a critical gap in the existing literature concerning *Ocotea* genus chemical investigations by conducting a comprehensive metabolomics screening. We have laid the foundation for future phytochemical and biological explorations of *Ocotea* species with groundbreaking insights into the chemical content and bioactivity of several endemic *Ocotea* spp., previously unexplored. The bioactive markers revealed have the potential for inhibiting PGE2 release, a critical inflammatory mediator in several inflammatory diseases. This discovery opens new avenues for research, as further investigations into their molecular scaffolds could lead to the development of new potential therapeutic agents, and increase the current repertoire of anti-inflammatory treatments. This study contributes to the broader understanding of NP alkaloid chemistry within the context of *Ocotea* plant species. By highlighting the untapped potential of these endemic plants for anti-inflammatory bioprospecting, we also underscore the importance of biodiversity protection of such valuable natural resources.

Author contributions

A.K., A.P., D.C., M.S. conceived the idea. A.K., J.C., and K.N. performed the anti-inflammatory assay and UPLC-MS/MS analysis. M.M. performed the High-resolution MS experiments. A.K. performed metabolomics data treatment and M.M., A.P., K.N., P.B., and D.C. participated in data analysis interpretation. A.K. drafted the manuscript, all other authors co-wrote the manuscript. A.P., M.S., D.F., and D.C. provided the funding. All authors reviewed and approved the final manuscript. All data were generated *in-house*, and no paper mill was used. All authors agree to be accountable for all aspects of work ensuring integrity and accuracy.

Funding

The Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil (CNPq; grant numbers 406837/2021-0, 316204/2021-8), and Fundação de Amparo à Pesquisa do Estado de Minas Gerais - Brazil (FAPEMIG; grant numbers APQ-02353-17, APQ-00207-18, BPD-00760-22).

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

The authors extend their gratitude to the José Badini OUPR herbarium (UFOP- Federal University of Ouro Preto, Minas Gerais- MG, Brazil), and Il- Leopoldo Krieger CESJ herbarium (UFJF- Federal University of Juiz de Fora - Minas Gerais, MG, Brazil), for generously providing the *Ocotea* spp. material. Their contributions were invaluable to the success of this research.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.phymed.2023.155060](https://doi.org/10.1016/j.phymed.2023.155060).

References

- Alarcon-Barrera, J.C., Kostidis, S., Ondo-Mendez, A., Giera, M., 2022. Recent advances in metabolomics analysis for early drug development. *Drug Discov. Today* 27, 1763–1773. <https://doi.org/10.1016/j.drudis.2022.02.018>.
- Antonio, A.S., Veiga-Junior, V.F., Wiedemann, L.S.M., 2020. *Ocotea* complex: A metabolomic analysis of a lauraceae genus. *Phytochemistry* 173, 112314. <https://doi.org/10.1016/j.phytochem.2020.112314>.
- Azmi, L., Srivastava, A., Shukla, I., Gautam, A., 2021. Multivariate Analysis of Herbal Drugs With Diverse Pharmacological Activities: Metabolomics study, Medicinal and Aromatic Plants. Elsevier Inc. <https://doi.org/10.1016/b978-0-12-819590-1.00004-5>.
- Backhouse, N., Delporte, C., Givernau, M., Cassels, B.K., Valenzuela, A., Speisky, H., 1994. Anti-inflammatory and antipyretic effects of boldine. *Agents Actions* 42, 114–117. <https://doi.org/10.1007/BF01983475>.
- Bartel, J., Krumsiek, J., Theis, F.J., 2013. Statistical methods for the analysis of high-throughput metabolomics data. *Comput. Struct. Biotechnol. J.* 4, e201301009. <https://doi.org/10.5936/CSBJ.201301009>.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- Brotto, M.L., Cervi, A.C., Dos Santos, É.P., 2013. O gênero *Ocotea* (Lauraceae) no estado do Paraná. *Brasil. Rodriguesia* 64, 495–525. <https://doi.org/10.1590/S2175-78602013000300004>.
- Bruni, R., Medici, A., Andreotti, E., Fantin, C., Muzzoli, M., Dehesa, M., Romagnoli, C., Sacchetti, G., 2004. Chemical composition and biological activities of Ishpingo essential oil, a traditional Ecuadorian spice from *Ocotea quixos* (Lam.) Kosterm. (Lauraceae) flower calices. *Food Chem.* <https://doi.org/10.1016/j.foodchem.2003.07.019>.
- Chagas-Paula, D., Zhang, T., Da Costa, F., Edrada-Ebel, R., 2015a. A Metabolomic Approach to Target Compounds from the Asteraceae Family for Dual COX and LOX Inhibition. *Metabolites* 5, 404–430. <https://doi.org/10.3390/metabo5030404>.
- Chagas-Paula, D.A., Oliveira, T.B., Zhang, T., Edrada-Ebel, R., Da Costa, F.B., 2015b. Prediction of anti-inflammatory plants and discovery of their biomarkers by machine learning algorithms and metabolomic studies. *Planta Med.* 81, 450–458. <https://doi.org/10.1055/s-0034-1396206>.
- Chini, M.G., Giordano, A., Potenza, M., Terracciano, S., Fischer, K., Vaccaro, M.C., Colarusso, E., Bruno, I., Riccio, R., Koeberle, A., Werz, O., Bifulco, G., 2020. Targeting mPGES-1 by a Combinatorial Approach: identification of the Aminobenzothiazole Scaffold to Suppress PGE2 Levels. *ACS Med. Chem. Lett.* 11, 783–789. <https://doi.org/10.1021/acsmchemlett.9b00618>.
- Costa, I.F.B., Calixto, S.D., Heggdorne De Araujo, M., Konno, T.U.P., Tinoco, L.W., Guimarães, D.O., Lasunskaja, E.B., Leal, I.R.C., Muzitano, M.F., 2015. Antimicrobial and nitric oxide production inhibitory activities of *ocotea notata* from Brazilian restinga. *Sci. World J.* 2015 <https://doi.org/10.1155/2015/947248>.
- Cragg, G.M., Pezzuto, J.M., 2016. Natural Products as a Vital Source for the Discovery of Cancer Chemotherapeutic and Chemopreventive Agents. *Med. Princ. Pract.* 25, 41–59. <https://doi.org/10.1159/000443404>.
- Creek, D.J., Dunn, W.B., Fiehn, O., Griffin, J.L., Hall, R.D., Lei, Z., Mistrik, R., Neumann, S., Schymanski, E.L., Sumner, L.W., Trengove, R., Wolfender, J.L., 2014.

- JAK2/STAT3 and NF- κ B Signaling Pathways. *Planta Med.* 84, 20–25. <https://doi.org/10.1055/s-0043-113447>.
- Yuliana, N.D., Khatib, A., Choi, Y.H., Verpoorte, R., 2011. Metabolomics for bioactivity assessment of natural products. *Phytother. Res.* 25, 157–169. <https://doi.org/10.1002/ptr.3258>.
- Zanatta, A.C., Vilegas, W., Edrada-Ebel, R.A., 2021. UHPLC-(ESI)-HRMS and NMR-Based Metabolomics Approach to Access the Seasonality of *Byrsonima intermedia* and *Serjania marginata* From Brazilian Cerrado Flora Diversity. *Front. Chem.* 9, 1–17. <https://doi.org/10.3389/fchem.2021.710025>.
- Zhang, X.W., Li, Q.H., Xu, Z.Di, Dou, J.J., 2020. Mass spectrometry-based metabolomics in health and medical science: a systematic review. *RSC Adv.* 10, 3092–3104. <https://doi.org/10.1039/c9ra08985c>.
- Zschocke, S., Van Staden, J., Paulus, K., Bauer, R., Horn, M.M., Munro, O.Q., Brown, N.J., Drewes, S.E., 2000. Stereostructure and anti-inflammatory activity of three diastereomers of ocobullenone from *Ocotea bullata*. *Phytochemistry* 54, 591–595. [https://doi.org/10.1016/S0031-9422\(00\)00163-1](https://doi.org/10.1016/S0031-9422(00)00163-1).

Bioprospecting-based untargeted metabolomics identifies alkaloids as potential anti-inflammatory bioactive markers of *Ocotea* species (Lauraceae)

Albert Katchborian-Neto^a, Karen de Jesus Nicácio^b, Jonas C. Cruz^c, Michael Murgu^e, Danielle F. Dias^a, Paula Carolina Pires Bueno^{a,f}, Marisi G. Soares^a, Ana C. C. Paula^d, Daniela A. Chagas-Paula^{a*}

^a Institute of Chemistry, Federal University of Alfenas, 37130-001, Alfenas, Minas Gerais, Brazil;

^b Department of Chemistry, Federal University of Mato Grosso, 78060-900, Cuiabá, Mato Grosso, Brazil;

^c Department of Chemistry, University of São Paulo, 14040-901, Ribeirão Preto, São Paulo, Brazil;

^d Department of Pharmaceutical Sciences, Federal University of Juiz de Fora, 36036-900, Juiz de Fora, Minas Gerais, Brazil;

^e Waters Corporation, Alameda Tocantins 125, 27th floor, Alphaville, 06455-020, São Paulo, São Paulo, Brazil;

^f Leibniz Institute of Vegetable and Ornamental Crops (IGZ), Theodor-Echtermeyer-Weg 1, 14979 Großbeeren, Germany

***Corresponding author:**

Prof. Dr. Daniela A. Chagas-Paula

daniela.chagas@unifal-mg.edu.br

Institute of Chemistry, Federal University of Alfenas, Rua Gabriel Monteiro da Silva, 700, Alfenas-MG, 37130-001, Brazil.

Appendix A. Supplementary data

Table S1. Vouchers from UFJF and UFOP and the *Ocotea* spp. details I.

Number	Code	Popular name	Specie Name	Endemic (E)		Geographical location
				Non-endemic (N)		
1	AY II	canela-amarela	<i>Ocotea aciphylla</i> (Nees & Mart.) Mez	N		13°32'14.0"S 41°54'14.0" W
2	AU II	canela-branca	<i>Ocotea acutifolia</i> (Nees) Mez	N		72°28'18.0"S 58°08'21.0" W
3	AM II	unknown	<i>Ocotea amazonica</i> (Meiss) Mez	N		-
4	BI II	canela-preta	<i>Ocotea bicolor</i> Vattimo-Gil	N		-
5	BA II	louro-verdadeiro	<i>Ocotea brachybotrya</i> (Meisn.) Mez	E		19°35'28.0"S 42°34'07.0" W
6	BR II	unknown	<i>Ocotea bragai</i> Coe-Teix.	E		-
7	CL I	unknown	<i>Ocotea calliscypha</i> L.C.S.Assis & Mello-Silva	E		20°17'15.0"S 43°30'19.1" W
8	CA I	unknown	<i>Ocotea caesia</i> Mez	E		-
9	CT II	canela-coqueiro	<i>Ocotea catharinensis</i> Mez	N		-
10	CE II	moena negra	<i>Ocotea cernua</i> (Nees) Mez	N		-
11	CM II	Unknown	<i>Ocotea complicata</i> (Meisn.) Mez	N		-
12	CO II	canela-fedida	<i>Ocotea corymbosa</i> (Meisn.)Mez	N		-
13	CJ II	Cuchumari	<i>Ocotea cujumary</i> Mart.	N		-
14	DO II	canela-louro	<i>Ocotea diospyrifolia</i> (Meisn.) Mez	N		-
15	DI I	canela-sassafrás	<i>Ocotea dispersa</i> (Nees & Mart.) Mez	E		-
16	DV II	canela-segueira	<i>Ocotea divaricata</i> (Nees) Mez	E		-
17	EL II	canela-broto	<i>Ocotea elegans</i> Mez/ <i>Ocotea indecora</i> (Schott) Mez	E		-
18	FE I	Unknown	<i>Ocotea felix</i> Coe-Teix.	E		-
19	GL I	louro	<i>Ocotea glauca</i> (Nees & Mart.) Mez	E		20°22'40.0"S 43°24'57.9" W

20	GU II	unknown	<i>Ocotea glaucina</i> (Meisn.) Mez	E	16°35'47.0"S 42°54'05.0" W
21	GZ II	canela-amarela	<i>Ocotea glaziovii</i> Mez	E	-
22	GA II	canela-seda	<i>Ocotea guianensis</i> Aubl.	N	-
23	HY I	unknown	<i>Ocotea hypoglauca</i> (Nees & Mart.) Mez	E	-
24	IN II	canela	<i>Ocotea indecora</i> (Schott) Mez	E	-
25	KU II	canela-burra	<i>Ocotea kuhlmannii</i> Vattimo-Gi/ <i>Ocotea nectandrifolia</i> Mez	E	-
26	LA II	unknown	<i>Ocotea lanata</i> (Nees & Mart.) Mez	E	27°37'49.0"S 49°02'58.0" W
27	LN II	canela-pilosa	<i>Ocotea lanceolata</i> (Nees) Nees/ <i>Ocotea lancifolia</i> (Schott)	N	25°32'52.0"S 54°35'17.1" W
28	LC II	canela-sabão	<i>Ocotea lancifolia</i> (Schott) Mez	N	18°06'54.0"S 43°20'28.0" W
29	LG I	unknown	<i>Ocotea langsdorffii</i> (Meisn.) Mez	E	-
30	LX I	canela-pimenta	<i>Ocotea laxa</i> (Nees) Mez	E	20°17'15.0"S 43°30'19.0" W
31	LO II	unknown	<i>Ocotea lobbii</i> (Meisn.) Rohwer	E	22°05'21.1"S 43°49'40.0" W
32	LF II	louro-ingá	<i>Ocotea longifolia</i> Kunth	N	-
33	MI II	canela-vassoura	<i>Ocotea minarum</i> (Nees & Mart.) Mez	E	-
34	NT I	unknown	<i>Ocotea nitidula</i> (Nees et Mart. ex Ness)	E	-
35	NE II	canela-burra	<i>Ocotea nectandrifolia</i> Mez	E	26°54'36.0"S 50°13'13.0" W
36	NI II	louro	<i>Ocotea nitida</i> (Meisn.) Rohwer	E	19°50'03.0"S 42°33'07.0" W
37	NO II	louro-pipoca	<i>Ocotea notata</i> (Nees & Mart.) Mez / <i>Ocotea glaucina</i>	E	10°43'58.0"S 41°19'35.0" W
38	MU I	canelinha	<i>Ocotea nummularia</i> / <i>Ocotea tristis</i>	E	19°52'47.9"S 43°40'10.9" W
39	NU I	unknown	<i>Ocotea nutans</i> (Nees) Mez	E	-
40	OD II	canela-sassafrás	<i>Ocotea odorifera</i> Vell. Rohwer	E	-
41	PA II	unknown	<i>Ocotea paranaenses</i> Brotto, Baitello, Cervi & E.P.Santos	E	25°52'58.0"S 48°34'28.9" W
42	PE I	Unknown	<i>Ocotea percoriacea</i> Kosterm.	E	20°17'15.0"S 43°30'19.0" W
43	PO I	Canela	<i>Ocotea pomaderroides</i> (Meisn.) Mez	E	-

44	PR II	imbuia	<i>Ocotea porosa</i> (Nees & Mart.) Barroso	N	-
45	PT II	canela-sassafrás	<i>Ocotea pretiosa</i> (Nees) Mez/ <i>Ocotea odorifera</i> (Vell.)	E	-
46	PU I	canela-babosa	<i>Ocotea puberula</i> (Rich.) Nees	N	20°17'15.0"S 43°30'19.1" W
47	PL II	unknown	<i>Ocotea pulchea</i> Vattimo-Gil	E	-
48	PC II 1	canela-lageana	<i>Ocotea pulchella</i> (Nees & Mart.) Mez	N	21°55'24.9"S 46°23'09.9" W
49	PH II	Canela	<i>Ocotea pulchra</i> Vattimo-Gil	E	27°21'38.0"S 49°08'13.0" W
50	SP I	canela-baraúna	<i>Ocotea spectabilis</i> (Meisn.) Mez	E	20°22'40.0"S 43°24'57.9" W
51	SX I	Canelão	<i>Ocotea spixiana</i> (Nees) Mez	E	20°17'15.0"S 43°30'19.1" W
52	TA I	unknown	<i>Ocotea tabacifolia</i> (Meisn.) Rohwer	E	-
53	TL II 1	canela-limão	<i>Ocotea teleiandra</i> (Meisn.) Mez	E	-
54	TE II 1	unknown	<i>Ocotea tenuiflora</i> (Nees) Mez	E	-
55	TR I	canelinha	<i>Ocotea tristis</i> (Nees & Mart.) Mez	E	20°17'15.0"S 43°30'29.1" W
56	VA I	unknown	<i>Ocotea vaccinioides</i> (Meisn.) Mez/ <i>Ocotea daphnifolia</i>	E	-
57	VR I	canela-pilosa	<i>Ocotea variabilis</i> Mart./ <i>Ocotea lancifolia</i> (Schott) Mez	N	-
58	VZ II 2	canela-verde	<i>Ocotea velloziana</i> (Meisn.) Mez	E	-
59	VL II	canelão-amarelo	<i>Ocotea velutina</i> (Nees) Rohwer	E	-
60	VI II	unknown	<i>Ocotea villosa</i> Kosterm.	E	-

* I- OUPR herbarium (UFOP- Federal University of Ouro Preto, Minas Gerais- MG) and II- CESJ herbarium (UFJF- Federal University of Juiz de Fora - Minas Gerais, MG), e.g., the sample “VL II” – *Ocotea velutina* – from CESJ.

Table S2. The *Ocotea* spp. extracts yields.

Code	Yield (%)	Code	Yield (%)	Code	Yield (%)
AY II	20,0	GA II	8,5	PO I	16,0
AU II	18,5	HY I	14,0	PR II	16,5
AM II	10,5	IN II	10,0	PT II	10,5
BA II	10,5	KU II	7,5	PU I	18,0
BR II	16,0	LA II	9,5	PL II	9,5
BI II	13,0	LN II	16,0	PC II 1	29,0
CA I	27,0	LC II	9,0	PH II	10,5
CL I	9,5	LG I	17,5	SP I	5,5
CT II	13,5	LX I	23,5	SX I	14,0
CJ II	12,5	LO II	10,0	TA I	21,5
CE II	15,5	LF II	11,0	TE II 2	14,0
CO II	15,0	MI II	7,0	TL II 2	8,0
CM II	6,0	MU I	23,0	TR I	13,0
DI I	14,5	NE II	8,0	VA I	13,5
DO II	15,5	NI II	21,5	VR I	11,0
DV II	13,5	NT I	7,5	VZ II 2	14,5
EL II	19,0	NO II	20,5	VL II	9,0
FE I	14,0	NU I	14,5	VI II	20,5
GL I	11,0	OD II	9,5		
GU II	8,5	PA II	13,5		
GZ II	13,5	PE I	16,0		

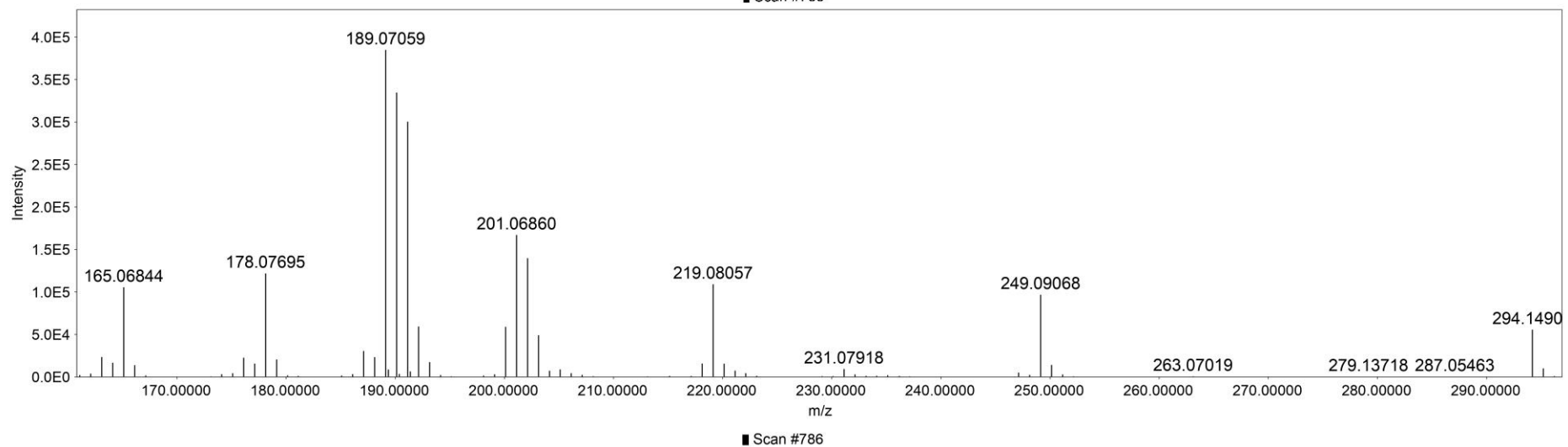
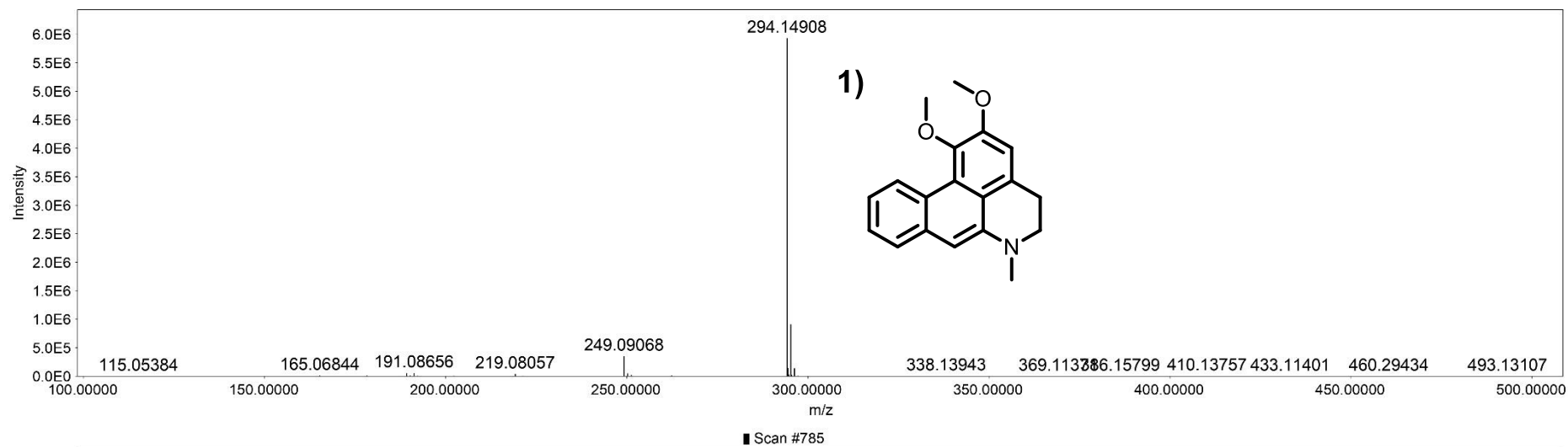


Fig. S1. MS^E spectra of low and high-energy channels for the proposed annotation of the dehydronuciferine (M+H).

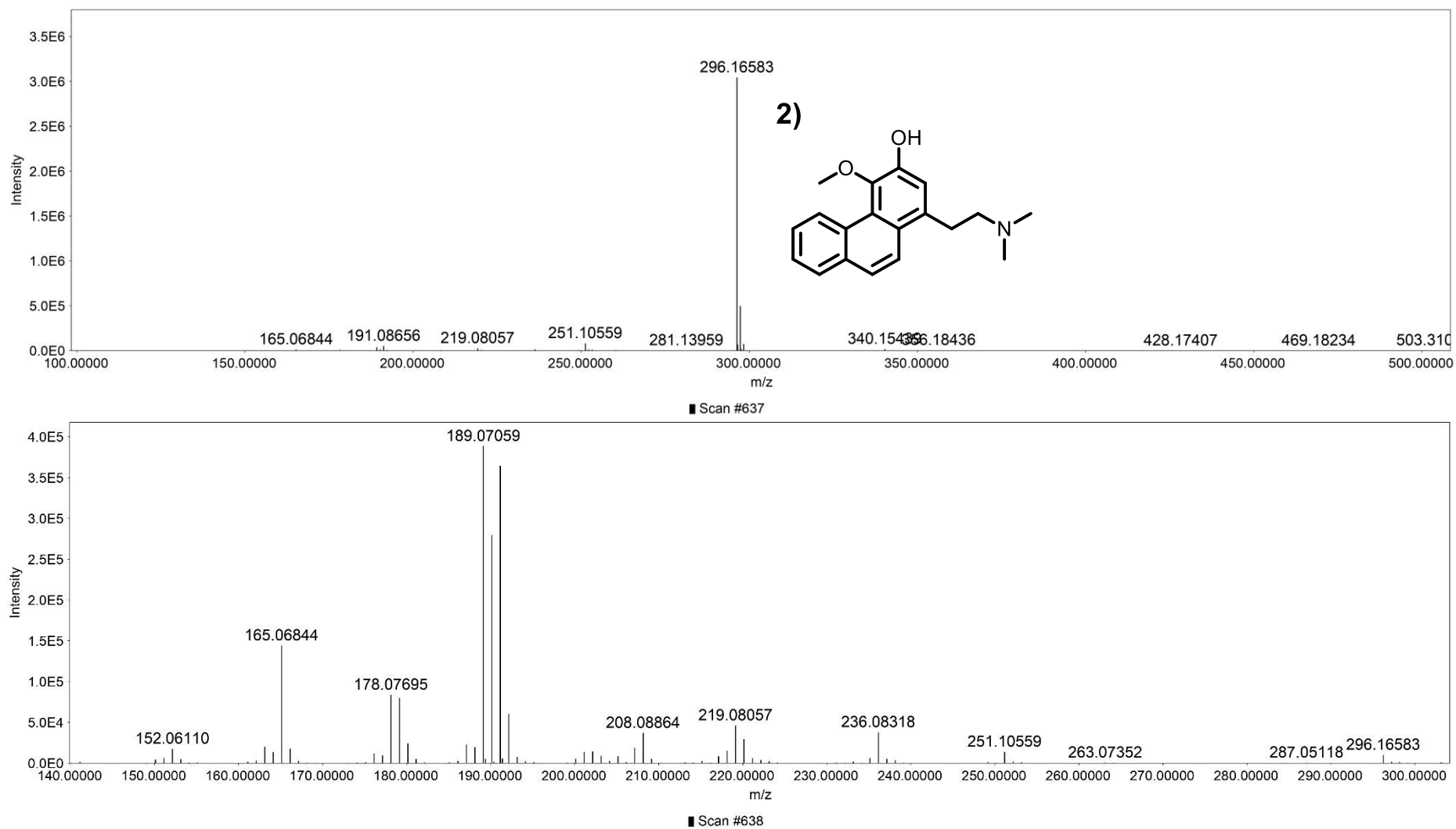


Fig. S2. MS^E spectra of low and high-energy channels for the proposed annotation of the argentineine.

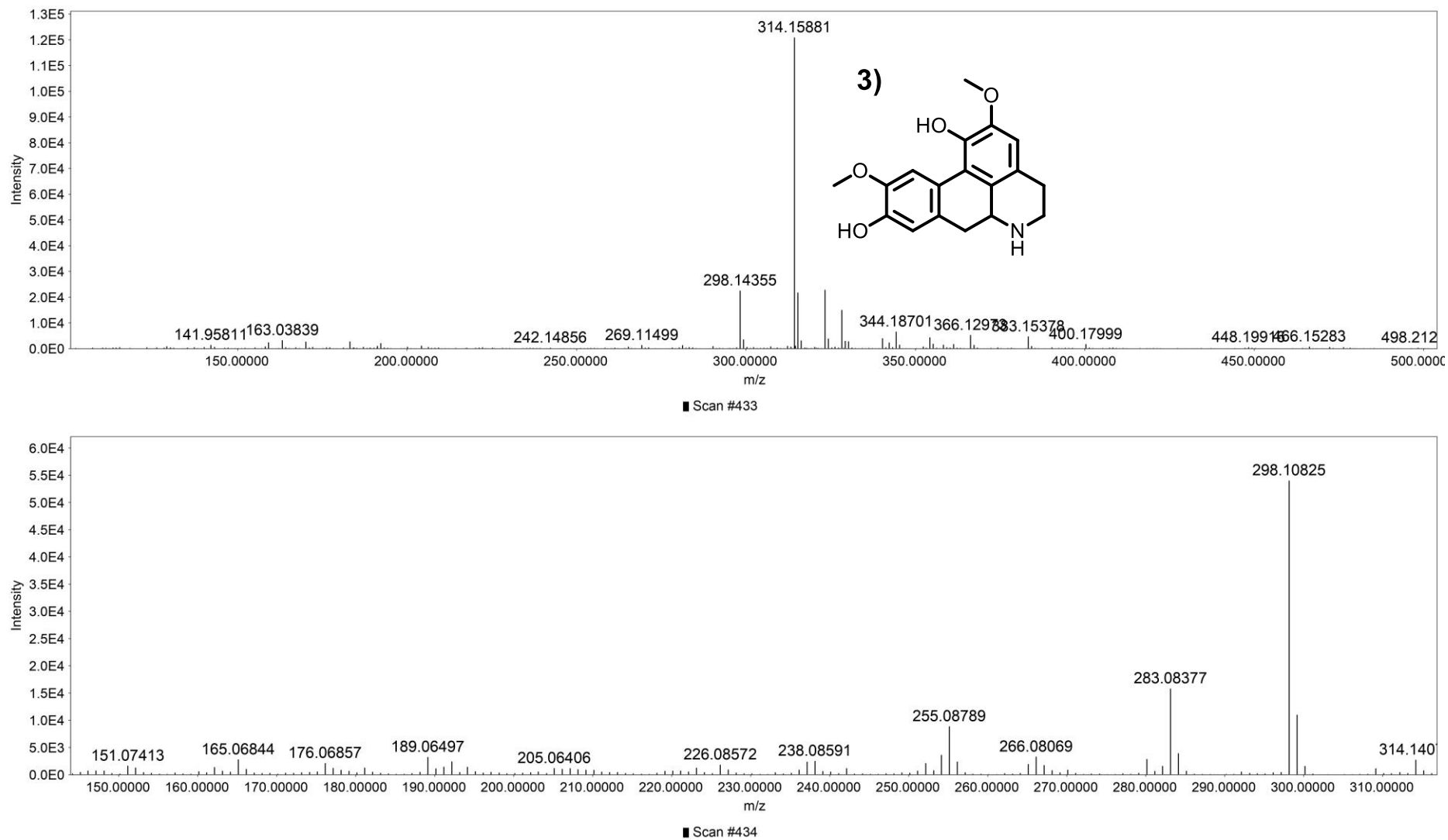


Fig. S3. MS^E spectra of low and high-energy channels for the proposed annotation of the laurelliptine (M+H).

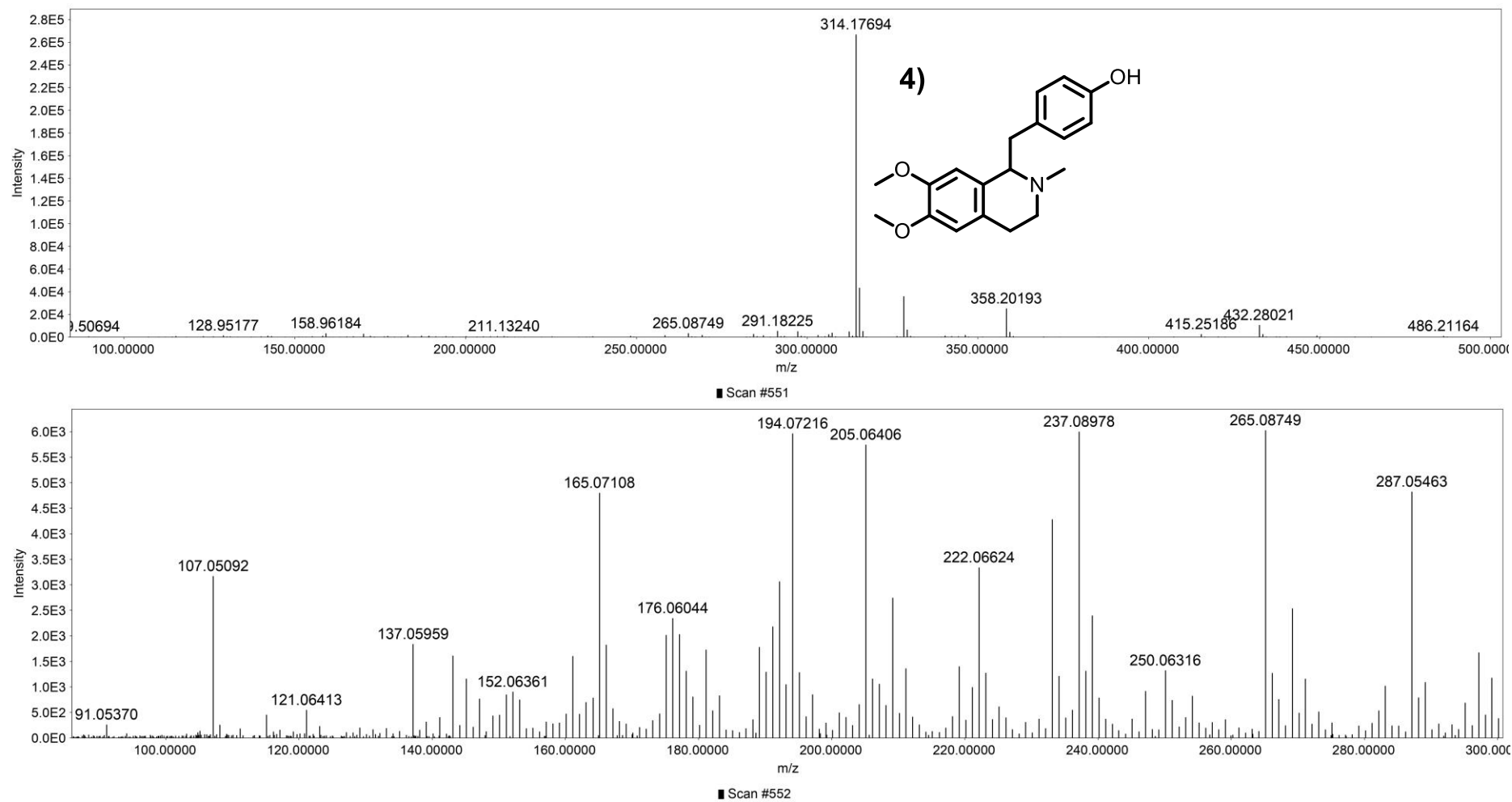


Fig. S4. MS^E spectra of low and high-energy channels for the proposed annotation of the artemepavine (M+H).

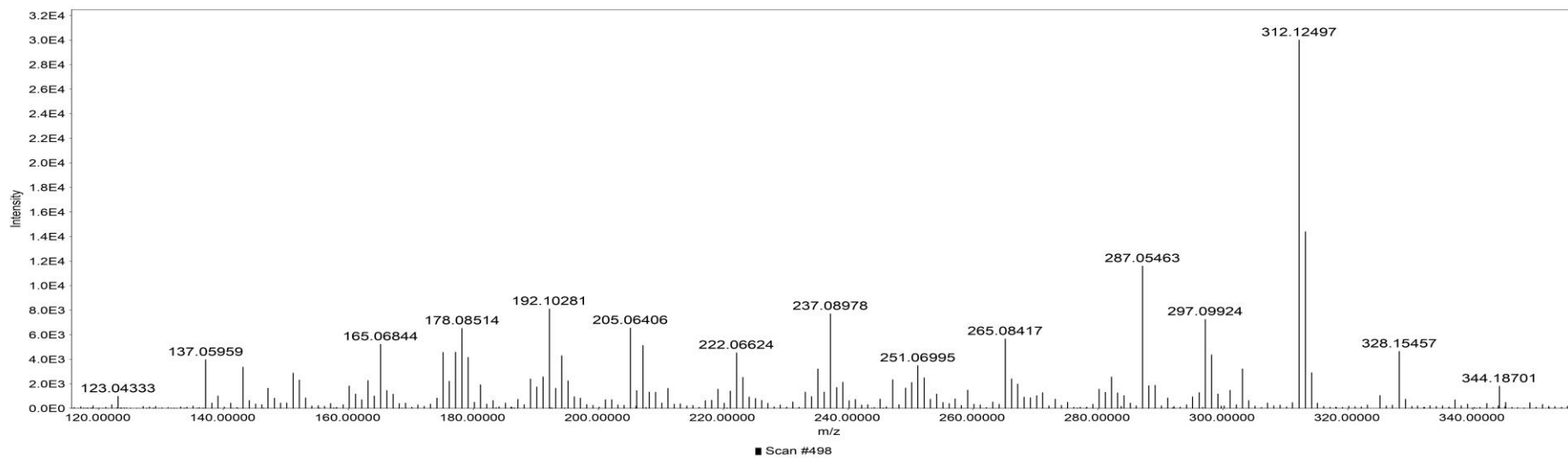
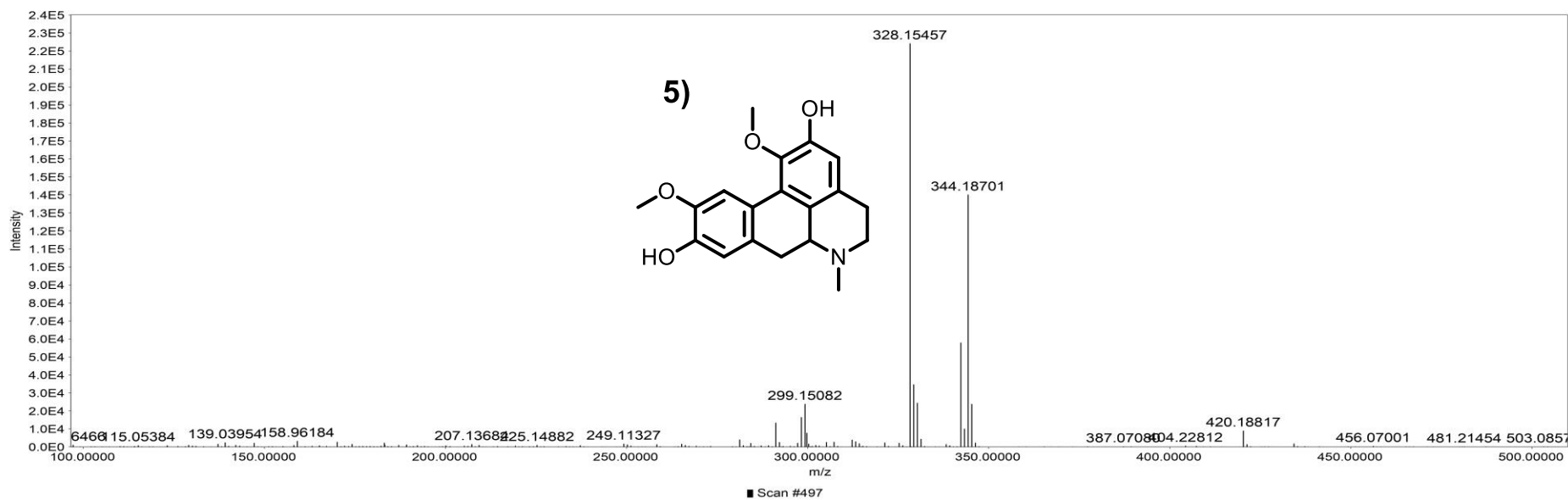


Fig. S5. MS^E spectra of low and high-energy channels for the proposed annotation of the boldine (M+H).

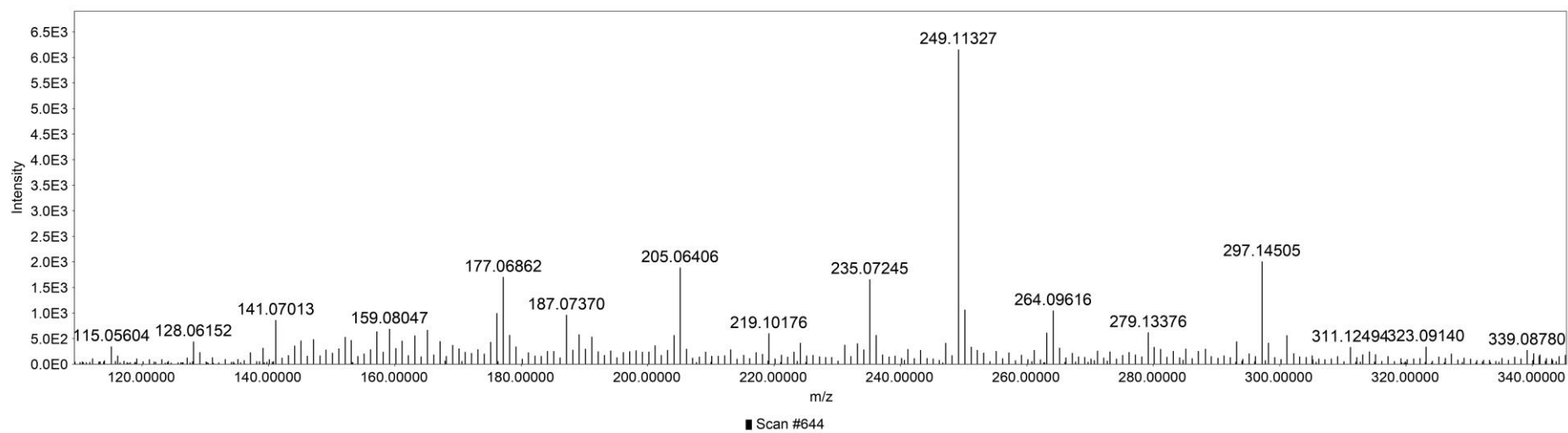
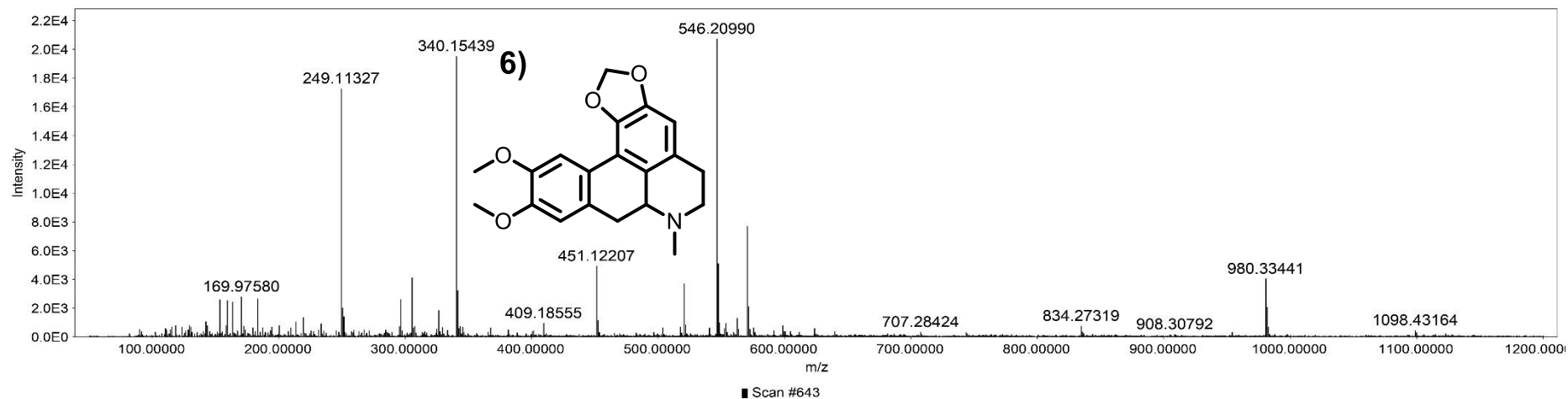


Fig. S6. MS^E spectra of low and high-energy channels for the proposed annotation of the dicentrine.

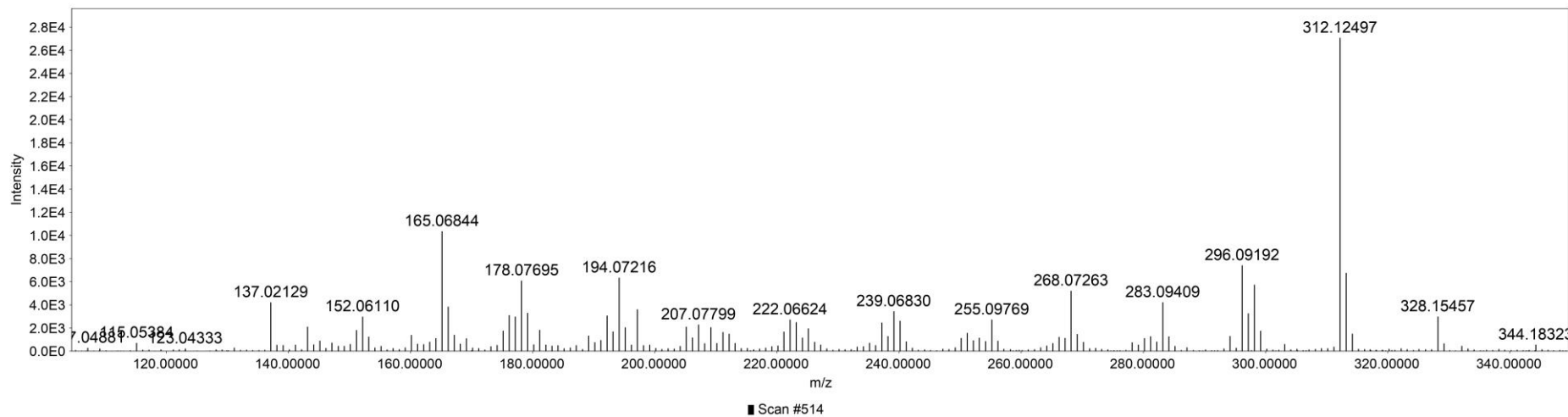
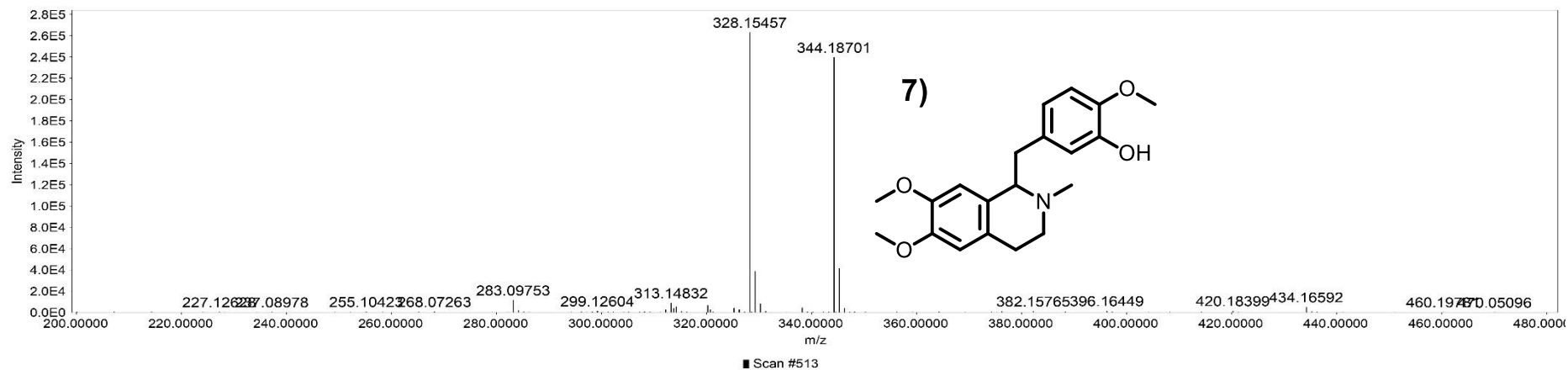
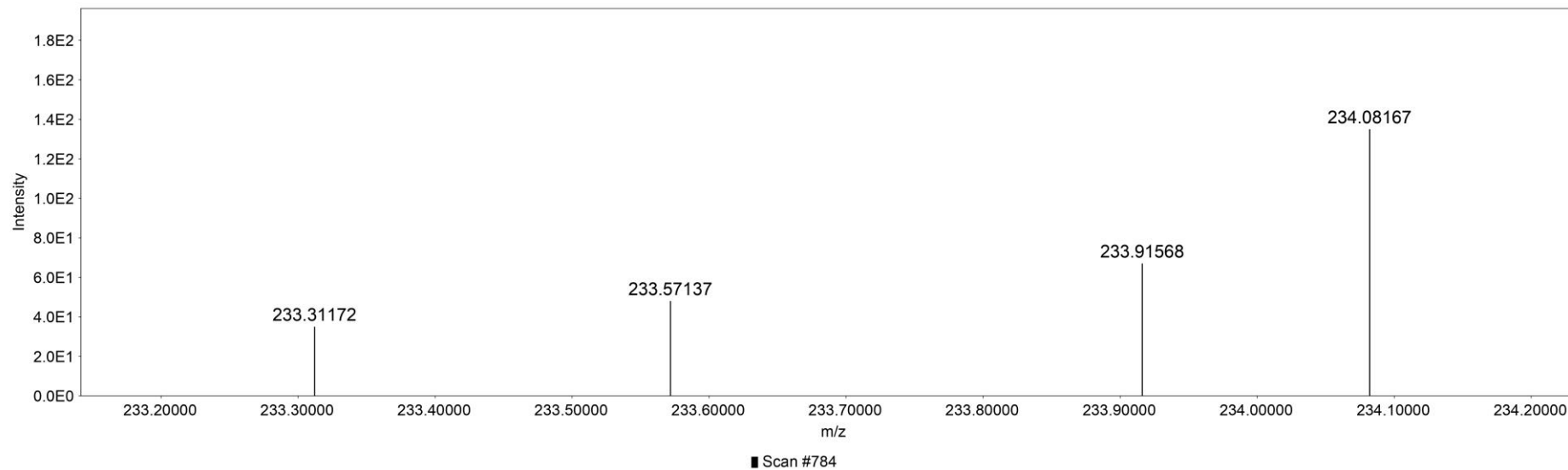
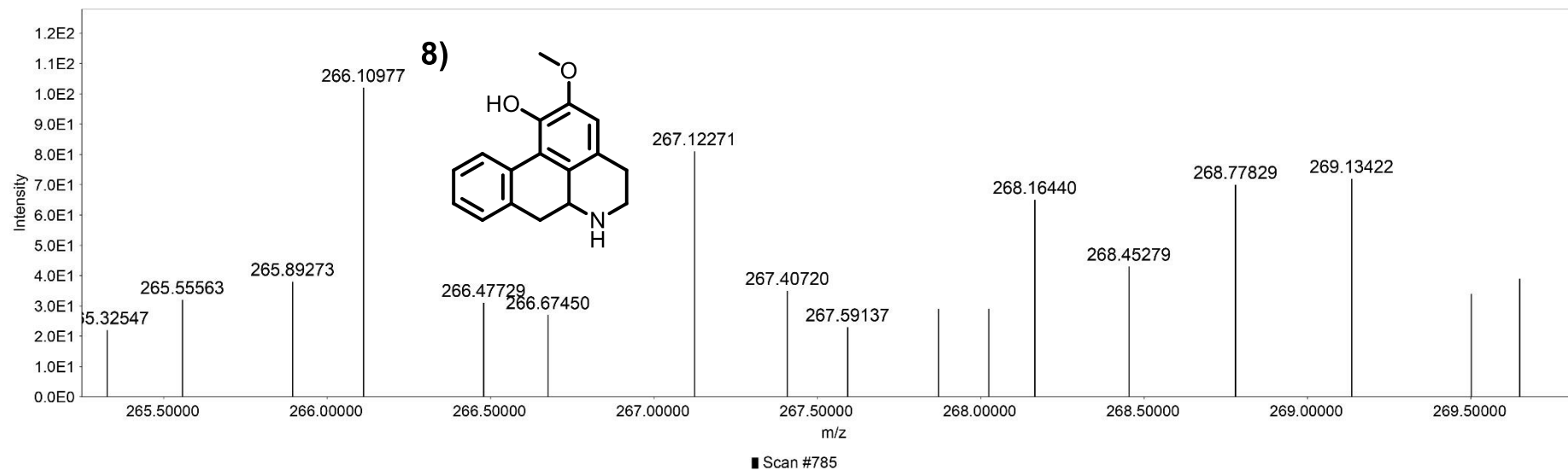


Fig. S7. MS^E spectra of low and high-energy channels for the proposed annotation of the laudanine (M+H).



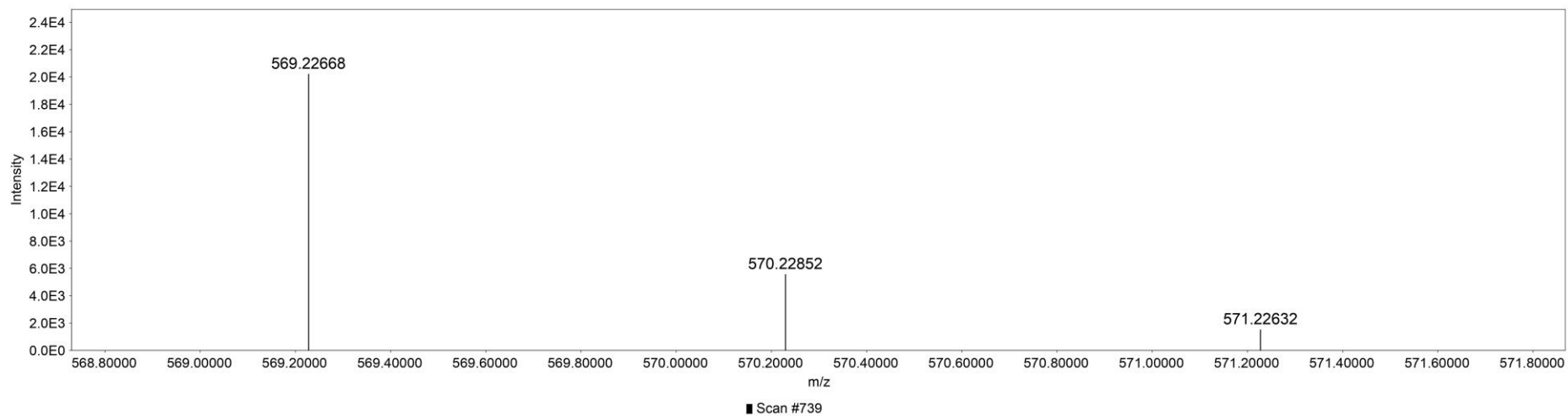
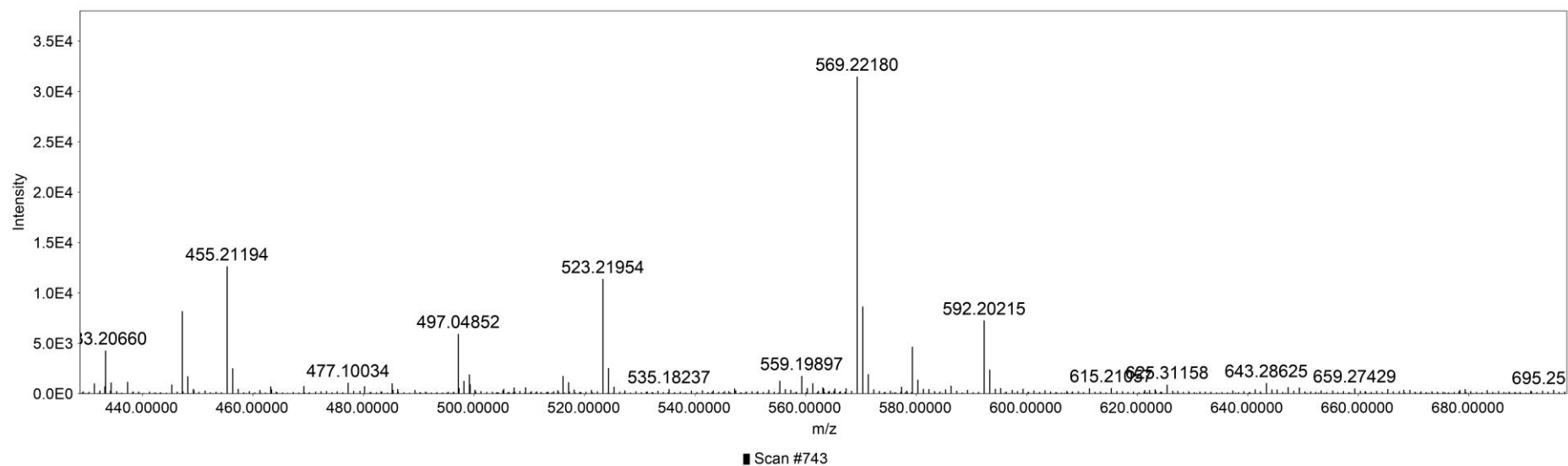


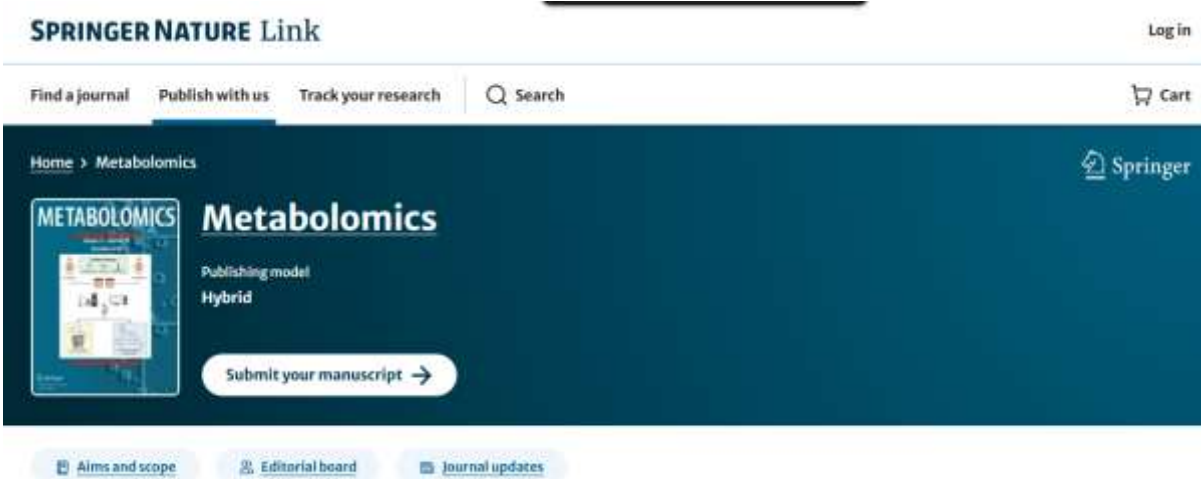
Fig. S8. MS^E spectra of low and high-energy channels for the proposed annotation of the caaverine (M+H and 2M+Cl).

5 CHAPTER III – PART B: ANTI-INFLAMMATORY METABOLOMICS STUDY OF *Ocotea* SPECIES

Chapter III containing the research article in a manuscript format is separately provided as an attached material of this thesis.

Academic Journal: Metabolomics (not submitted yet)

Impact factor (2024): 3.5



The screenshot displays the Springer Nature website interface for the journal *Metabolomics*. At the top, the "SPRINGER NATURE Link" logo is visible on the left, and a "Log in" link is on the right. Below the logo, there are navigation links: "Find a journal", "Publish with us", and "Track your research", followed by a search bar labeled "Search" and a "Cart" icon. The main content area features a dark blue header with "Home > Metabolomics" on the left and the Springer logo on the right. The central focus is the journal's title "Metabolomics" in large white text, with a "Publishing model Hybrid" label below it. To the left of the title is a thumbnail image of the journal cover. A prominent white button with a right-pointing arrow says "Submit your manuscript". At the bottom of the page, there are three light blue buttons: "Aims and scope", "Editorial board", and "Journal updates".

Anti-inflammatory markers of *Ocotea* (Lauraceae) uncovered through concatenated UPLC/MS-NMR metabolomics approach– dual inhibition of COX and LOX pathways

Albert Katchborian-Neto^{a,c}, Gabriel Viana^a, Miller Santos Ferreira^a, Matheus Fernandes Alves^a, Michael Murgu^b, Marina de Monroe Gonçalves^c, Ana Cláudia. C. Paula^d, Daniele Ferreira Dias^a, Marisi Gomes Soares^a, João Henrique G. Lago^c, RuAngelie Edrada-Ebel^e,
Daniela A. Chagas-Paula^{a*}

The Lauraceae family, particularly the genus *Ocotea*, is recognized for its wide chemical diversity and medicinal properties, such as anti-inflammatory. Inflammation, driven by pathways such as cyclooxygenase (COX) and lipoxygenase (LOX), plays a central role in several human pathologies. These pathways are essential for the production of key pro-inflammatory mediators, such as prostaglandin E2 (PGE2) and leukotriene B4 (LTB4). This study aimed to prospect dual COX/LOX inhibitory biomarkers in *Ocotea* species using a concatenated ultra-performance liquid chromatography coupled to high-resolution mass spectrometry (UPLC/MS) - nuclear magnetic resonance (NMR) metabolomics approach and machine learning (ML) prediction models. Sixteen *Ocotea* species were analysed, with PGE2 and LTB4 levels determined by using an *ex-vivo* anti-inflammatory assay. The metabolomic data were subjected to block scaling and concatenated, with UPLC/MS and ¹H-NMR data analyzed as distinct blocks. Both unsupervised and supervised multivariate statistical methods were performed. Additional statistical total correlation spectroscopy (STOCSY) was performed to identify and correlate chemical shifts in the NMR data. In addition, gas-phase fragmentation reactions were proposed to increase metabolite confidence annotation in the UPLC/MS data. ML models were integrated to predict novel promissory extracts based on the discriminant bioactive markers of the metabolomics data. Nine *Ocotea* species showed dual inhibition, annotating two aporphine and two benzylisoquinoline alkaloids, one sesquiterpenoid and one glycosylated flavonoid as key potential anti-inflammatory compounds. This metabolomics approach also reinforces the anti-inflammatory potential of *Ocotea* species as a source for the discovery of novel anti-inflammatory agents with innovative mechanisms of action through dual inhibition of COX and LOX pathways.

Keywords: Natural Products, Biochemometrics, Chemical profiling, Spectroscopy and Spectrometric Techniques, Inflammation.

Introduction

The Lauraceae family has garnered significant attention for its chemical diversity, traditional medicinal value, and widespread biological activities.¹⁻⁴ Particularly, the genus *Ocotea*, one of the largest in the family has demonstrated significantly anti-inflammatory activity.⁵⁻⁸ Previous studies have reported the anti-inflammatory effects of various *Ocotea* and also other species in the Lauraceae plants, attributing these effects mainly to the presence of alkaloids.⁷⁻⁹ Inflammation, a protective biological response to harmful stimuli, such as from infections, tissue damage, or the development of different diseases, involves complex biochemical cascades, including the activation of cyclooxygenase (COX) and lipoxygenase (LOX) pathways. They are critical in the biosynthesis of relevant pro-inflammatory mediators such as prostaglandin E2 (PGE2) and leukotriene B4 (LTB4), respectively.^{10,11} The concurrent inhibition of these pathways is viewed as a valuable approach to drug discovery, given that most currently available drugs target them individually and inflammation is involved in a range of different pathological conditions.^{1,4,5} As drugs that demonstrate simultaneous activity in these pathways are scarce, discovering compounds with dual activity is highly desirable in to quest for novel effective anti-inflammatory therapies.¹¹⁻¹³

The field of natural products (NP) research, particularly in drug discovery, is undergoing a significant transformation with the increased implementation of modern metabolomics techniques.¹⁴⁻¹⁶ These methods offer an approach to exploring promising compounds early in the research process, reducing and rationalizing repetitive and labour-intensive steps associated with traditional bioassay-guided fractionation.^{17,18} Contemporary research strategies increasingly utilize comprehensive analysis of raw extracts which could also include preliminary fractions. This shift is allowing metabolomics research to streamline faster-targeted metabolite identification of potential bioactive markers.^{19,20} Such strategies can aid in prioritizing which extracts to explore for further fractionation, avoiding already known compound characterisation, and thus isolating new NPs or targeting those of interest more efficiently.^{21,22}

Central to these methodologies are advanced analytical techniques such as ultra-performance liquid chromatography coupled to high-resolution mass spectrometry (UPLC/MS) and nuclear magnetic resonance (NMR) spectroscopy. These techniques are chosen for their sensitivity and selectivity, and their complementary capabilities are essential for the reliable annotation and potential identification of compounds. Concatenated metabolomics represents a sophisticated modern approach that integrates data from UPLC/MS and NMR to provide comprehensive metabolomic profiling.²³⁻²⁵ This

approach pools the strengths of both methods to better characterise specialised metabolites from complex natural sources. This comprehensive coverage allows the detection of a broader range of metabolites, including those that might be missed by either technique alone. Additionally, results from each technique can validate and corroborate findings from the other, enhancing confidence in the bioactive markers annotation process.^{23,25,26}

In addition to the advanced analytical techniques, the integration of computational tools and machine learning (ML) models into NP research is also contributing to a more effective way to prospect bioactive compounds.^{14,15} ML algorithms can process and aid in analysing complex datasets generated by metabolomics studies, enabling the identification of patterns and correlations that might be overlooked using traditional methods. These models are particularly valuable in predicting chemical and biological aspects, including patterns of co-occurring fragments and neutral losses acquired in tandem mass spectrometry (MS/MS) or even the bioactivity of compounds based on their metabolomics data, thus streamlining the identification of potential bioactive agents.^{8,15,27}

This study aimed to find dual biomarkers in *Ocotea* species responsible for inhibiting both COX and LOX pathways using a combined UPLC/MS-NMR metabolomics approach. Even though the literature supports the anti-inflammatory potential of *Ocotea* species, the dual inhibition of COX and LOX pathways by the selected *Ocotea* plant species has not been yet investigated. By integrating UPLC/MS and NMR techniques, this study sought to comprehensively profile the active compounds in *Ocotea* extracts and gain detailed insights into the metabolites responsible for anti-inflammatory properties, and their potential dual activity. An *ex vivo* anti-inflammatory assay was performed to assess the blood inhibition levels of PGE2 and LTB4 using UPLC-MS/MS. Later, ML was employed to validate the anti-inflammatory biomarkers of dual inhibition of COX and LOX pathways by building prediction models based on their discriminant metabolomics data.

Results and discussion

Anti-inflammatory activity. The results of *ex-vivo* anti-inflammatory activity for 16 *Ocotea* extracts were assessed by one-way ANOVA statistical analysis (**Figure 1**). To identify which groups differ, Dunnett's multiple comparison test (**Table S1-2**) was applied post-ANOVA.²⁸ Dunnett's post-test is indicated for statistical analysis of several groups against the control,²⁹ as in the case of the anti-inflammatory screening evaluation. The evaluated reference drugs, dexamethasone (DEX) and indomethacin (IND) exhibited substantial and comparable anti-inflammatory activity, especially for the PGE2 experiment, both statistically different from negative (NEG) controls ($p < 0.0001$).

Notably, the NEG controls showed high levels of PGE2 and LTB4, which are key pro-inflammatory mediators targeted by current anti-inflammatory therapies,^{30–32} thus validating the efficacy of the anti-inflammatory screening protocol employed.

The *Ocotea* crude extracts statistically different from the NEG controls were categorized as active samples ($p < 0.05$). The PGE2 and LTB4 inhibition release results demonstrated relevant insights into the dual anti-inflammatory potential of nine evaluated *Ocotea* extracts, specifically *O. pulchella* (PC), *O. teleiandra* (TE), *O. tenuiflora* (TL), *O. spixiana* (SX), *O. tabacifolia* (TA), *O. tristis* (TR), *O. vaccinioides* (VA), *O. villosa* (VI), and *O. veloziana* (VZ) (**Table 1**). However, similarly to previous screening of PGE2 inhibition release *O. villosa* stands with the best results, for the present assay with approximately 72% of inhibition for PGE2. It has inhibited by 68% the LTB4 release, similar to the DEX positive control (62% and 68% respectively), which is the reference drug well known to inhibit both COX and LOX pathways.¹² Of notice, *O. glaucina* extract was inactive, besides pro-inflammatory, particularly for LTB4 levels.

The *Ocotea* species investigated in this research have been previously evaluated only regarding their PGE2 inhibition activity, but not LTB4.⁹ Thus, the obtained results externally validated the previous one regarding their anti-inflammatory activity in the COX pathway. On the subject of the LOX pathway, previous works found that the leaf extract of *O. diospyrifolia* and *O. odorifera* and other Lauraceae species such as from the *Aniba* genus could inhibit myeloperoxidase (MPO) levels, an enzyme primarily found in neutrophils that indicates their recruitment to inflamed tissue. Given that LTB4 is a key mediator in neutrophil recruitment, it is also possible that *Ocotea* species might also inhibit the LOX pathway.^{1,5,6,33} The results presented here corroborate the literature and provide additional findings suggesting the anti-inflammatory activity of several *Ocotea* extracts can be via COX and LOX pathways, by dual inhibition of both PGE2 and LTB4 release, simultaneously. These findings further substantiate the therapeutic promise of *Ocotea* plant species in treating inflammation-driven conditions, potentially by a mechanism of action different from current anti-inflammatory drugs in the market, supporting the claims of anti-inflammatory activity of plants from this genus.^{5–7,9} However, based on this *ex vivo* assay results, using human blood, is not possible to determine which portion of the COX and LOX inflammatory pathways are inhibited. To acquire such information, further investigation using specific enzymatic investigations would be necessary.

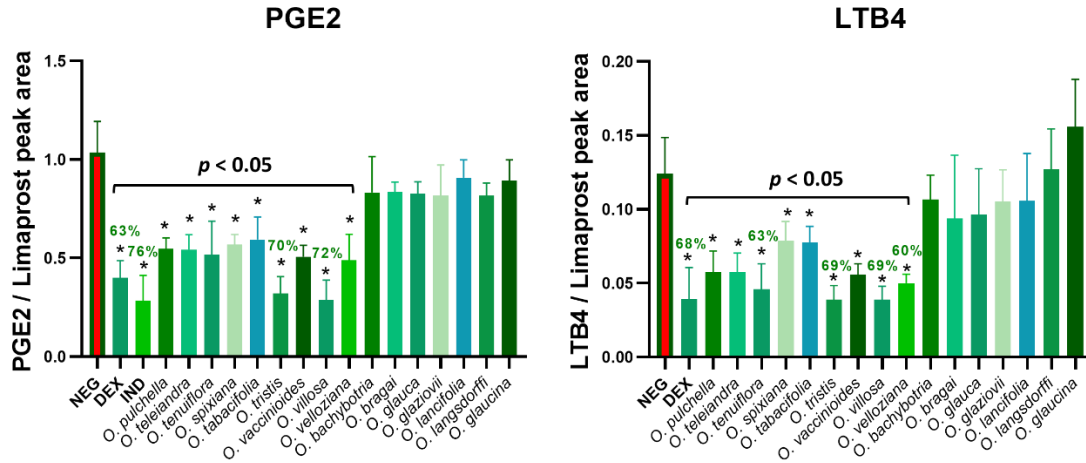


Figure 1. *Ex-vivo* anti-inflammatory activity of 16 *Ocotea* crude extracts and their inhibitory effects on PGE2 and LTB4 levels. The data is represented as the mean \pm standard deviation (SD) of four replicates per sample and six replicates for the controls. Statistical analysis was conducted using one-way ANOVA followed by Dunnett's multiple comparison test. Asterisks (*) denote a statistically significant difference from the negative control (NEG), with p -values ≤ 0.05 . Reference anti-inflammatory drugs are dexamethasone (DEX) and indomethacin (IND).

Table 1. The percentage list of PGE2 and LTB4 releasing inhibition for the 16 evaluated *Ocotea* species.

Specie	rrPGE2 (SD)	PI _{PGE2}	rrLTB4 (SD)	PI _{LTB4}
<i>O. pulchella</i>	0.526 (0.096)	47.4 %	0.465 (0.147)	53.5 %
<i>O. teleiandra</i>	0.524 (0.107)	47.6 %	0.464 (0.138)	53.6 %
<i>O. tenuiflora</i>	0.501 (0.178)	49.9 %	0.370 (0.157)	63.0 %
<i>O. spixiana</i>	0.548 (0.095)	45.2 %	0.608 (0.162)	39.2 %
<i>O. tabacifolia</i>	0.572 (0.141)	42.8 %	0.626 (0.152)	37.4 %
<i>O. tristis</i>	0.308 (0.095)	69.2 %	0.312 (0.100)	68.8 %
<i>O. vaccinioides</i>	0.489 (0.092)	51.1 %	0.451 (0.106)	54.9 %
<i>O. villosa</i>	0.277 (0.105)	72.3 %	0.314 (0.095)	68.6 %
<i>O. velloziana</i>	0.472 (0.145)	52.8 %	0.403 (0.094)	59.7 %
<i>O. bachybotria</i>	0.803 (0.213)	19.7 %	0.860 (0.217)	14.0 %
<i>O. bragai</i>	0.807 (0.131)	19.3 %	0.757 (0.376)	24.3 %
<i>O. glauca</i>	0.797 (0.134)	20.3 %	0.778 (0.293)	22.2 %
<i>O. glaziovii</i>	0.786 (0.192)	21.4 %	0.849 (0.242)	15.1 %
<i>O. lancifolia</i>	0.876 (0.159)	12.4 %	0.854 (0.308)	14.6 %
<i>O. langsdorffi</i>	0.788 (0.134)	21.2 %	0.926 (0.217)	7.4 %
<i>O. glaucina</i>	0.862 (0.165)	13.8 %	1.257 (0.359)	0 %
DEX (+ control)	0.368 (0.09)	63.1 %	0.316 (0.183)	68.4 %
IND (+ control)	0.238 (0.104)	76.2 %	-	-
NEG (- control)	0	0 %	0	0 %

Note: Relative release of PGE2 related to the NEG control = $rrPGE2 = (\text{PGE2 sample signal} / \text{NEG control signal})$. $rrLTB4 = (\text{LTB4 sample signal} / \text{NEG control signal})$. SD: standard deviation obtained by the propagation of error from sample signal uncertainty and NEG control signal uncertainty. PI: Percentage inhibition = $(1 - rr) * 100$.

UPLC/MS and NMR Data Processing. After the exclusion of high-intensity features from the blank in both UPLC/MS and NMR datasets, for UPLC/MS data in the positive ionization mode, 1290 distinct features were detected (based on m/z and retention time - RT), while 1043 features were recorded in the negative mode. The m/z data contained features ranging from 150 to 1250 m/z and RT from 0.2 - 9.8 min from both ionization modes were processed separately before being combined to create a comprehensive data matrix of 2333 features of m/z - RT pairs (**Table Z1**- Available at <https://doi.org/10.5281/zenodo.13826734>). UPLC/MS data of the 16 *Ocotea* species in positive ionization mode exhibited a visually higher number of peaks between 1.5 to 3.5 minutes. For the negative ionization mode, peaks were distributed throughout the entire chromatogram regions, especially from 1.5 - 5.5 min and 7.0 - 9.5 min (Supplementary **Figure S1-2**). For NMR data processing, 274 chemical shifts remained in the final data matrix after deleting blank signals (**Table Z2**- Available at <https://doi.org/10.5281/zenodo.13826734>). The NMR stacked spectral profile of the *Ocotea* extracts obtained exhibited signals in the regions of 0.6–1.4 ppm (methyl proton shifts), 1.5–2.8 ppm (unusual methyl proton or methylene shifts), 3.0–4.0 ppm (methoxy proton shifts), 4.2–5.5 ppm (anomeric proton shifts), and 5.0–8.0 ppm (conjugated non-aromatic double bonds or aromatic proton shifts) (Supplementary **Figure S3**).

Data fusion. The processed datasets of UPLC/MS (positive/negative modes = 2285 features; 48 features below 150 m/z were excluded) and NMR data (274 features) were gathered together generating a final concatenated dataset of 2559 features (**Table Z3**- Available at <https://doi.org/10.5281/zenodo.13826734>). This disparity of almost ten times in the number of features highlighted the variable nature of detection across the different techniques. The UPLC/MS data generated an extensive higher number of features, as this technique has far more sensibility. NMR data besides fewer features, has higher reproducibility and thus these techniques provide complementary information under metabolomics studies.^{25,34,35} Thus, to handle this different number of variables in each block of the analytical data, a block-wise scaling tool using SIMCA-P was employed to manage this natural variability effectively. By considering the appropriate variance and standard deviation of each block, it ensured that no single block of data overshadowed the other. Also, treating each data block (block 1 – UPLC/MS and block 2 – NMR) as a cohesive unit and scaling them appropriately, can lead to more balanced data adequate to proper multivariate statistical analysis (MSA).^{36,37} The use of block scaling is especially

beneficial in metabolomics studies, where the diversity of metabolite concentrations and their biological relevance can greatly affect the interpretability of MSA models.^{25,38}

Multivariate Statistical Analysis. The NMR, UPLC/MS and fused datasets were separately investigated. For our metabolomics study, Principal Component Analysis (PCA) was the method of choice for an initial exploration of data structure due to its ability to facilitate the visualization of sample distributions across the dataset, and to evaluate the quality of the analytical acquired data, as well as for observation of potential outliers.^{39,40} PCA is an unsupervised method that reduces original data to a few principal components to describe maximum variation within the data.^{21,39} Thus, PCA is a useful multivariate analysis that was employed to explore the complexity of the metabolic characteristics of *Ocotea* plant extracts, irrespective of their predefined classifications under the *ex vivo* anti-inflammatory experiments.

In our analysis, after applying a log transformation to minimize data skewness and Pareto scaling to make metabolite levels comparable, PCA was used to visualize general clustering trends within the datasets.^{41,42} For both UPLC/MS and NMR data, PCA achieved satisfactory goodness-of-fit and explained variances with R^2 values of 0.68 and 0.64, respectively, using 5 components. Higher R^2 values indicate a greater ability of the model to explain variance in the response data, thus it is a statistical measure that shows how well the data fit the regression model. Often for metabolomics studies R^2 values greater than 0.5-0.6 are considered satisfactory.^{13,43,44} The UPLC/MS scatter plot analysis suggested a potential trending for the separation of active and inactive samples, while for NMR data, this trending pattern was less discriminative (**Figure 2 A and B**).

For the fused dataset (**Figure 2 C**), PCA could effectively summarize the variation between groups with 5 components, with a well-explained variance and R^2 value of 0.63. The separation trend was also clear, with inactive samples clustering together in the upper right quadrant of Hotelling's ellipse (95%), and active samples being more widespread (**Figure 2 C**). Therefore, the fusion of datasets was beneficial because of enhanced clustering discrimination patterns. Still, by the analysis of the PCA score plots (**Figure 2**) it is evident that any data has achieved perfect clustering regarding the anti-inflammatory profile, demonstrating that other variables such as the potential inter-specie chemical composition variation are present also play a role. As shown in the loading plot (**Figure 2**), there are leveraged loading features across different active *Ocotea* extracts (left side) that might be able to contribute to the presence of dual anti-inflammatory effects.

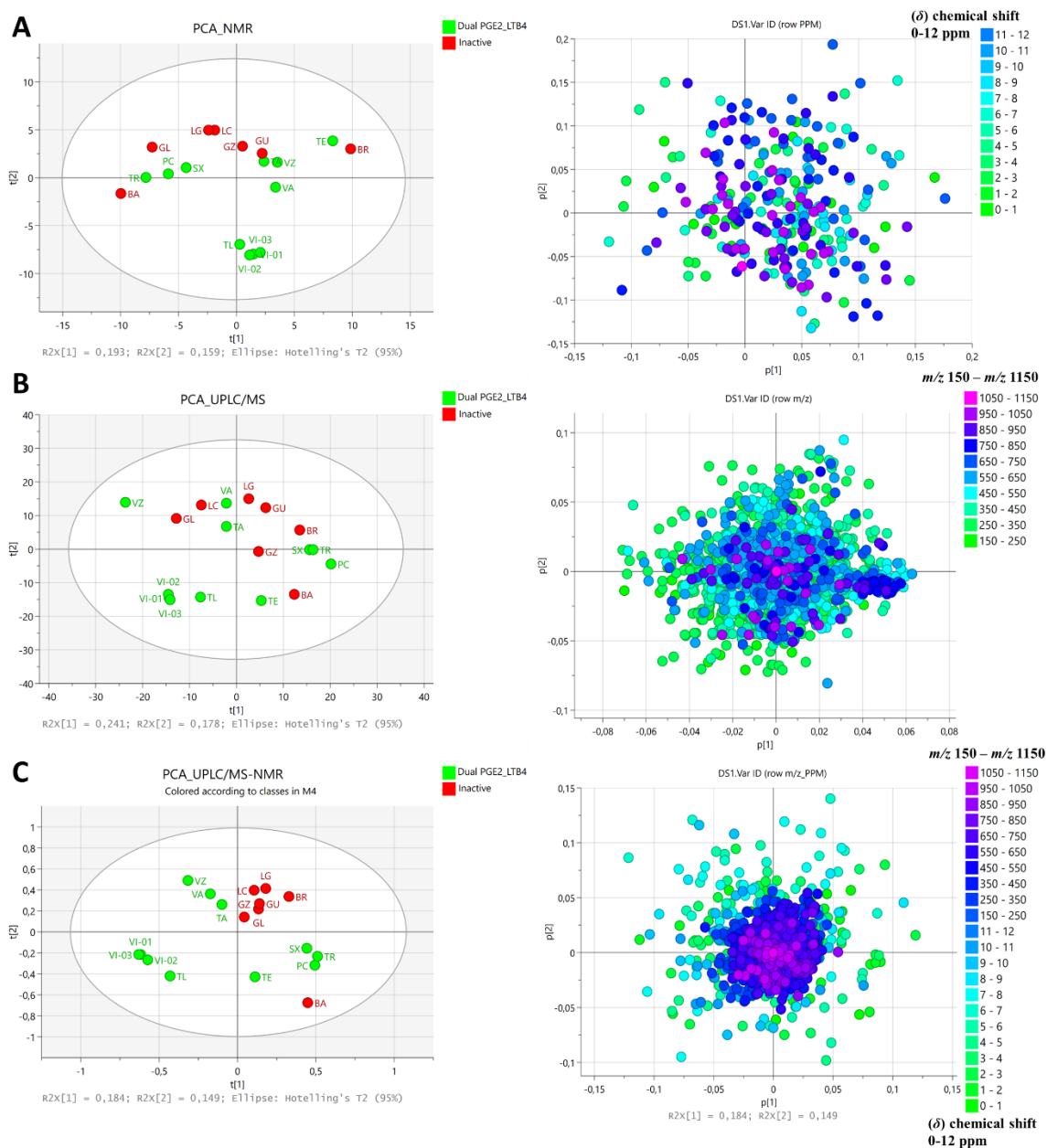


Figure 2. Score (left side) and loading (right side) plots of principal component analysis (PCA) of *Ocotea* crude extracts using PC1 vs. PC2. (A) NMR data, (B) UPLC/MS data and (C) UPLC/MS-NMR fused data. PC - *O. pulchella*, TE - *O. teleiandra*, TL - *O. tenuiflora*, SX - *O. spixiana*, TA - *O. tabacifolia*, TR - *O. tristis*, VA - *O. vaccinioides*, VI - *O. villosa*, VZ - *O. veloziana*, BA - *O. bachybotria*, BR - *O. bragai*, GL - *O. glauca*, GZ - *O. glaziovii*, LC - *O. lancifolia*, LG - *O. langsdorffi*, and GU - *O. glaucina*.

Supervised analyses of Orthogonal Projections to Latent Structures - Discriminant Analysis (OPLS-DA) evaluated the impact of some features on the anti-inflammatory activity of the samples, using the UPLC/MS, NMR and concatenated datasets. The OPLS is an extension of the partial least squares regression (PLS) method which integrates an orthogonal signal correction filter to enhance model interpretability and accuracy, making it particularly useful in complex datasets to distinguish subtle differences between two groups.^{39,45} OPLS-DA is a robust statistical model that separates the predictive variability

(the variability that helps differentiate between active and inactive) from the orthogonal variability, which is the variability that is present in the data but does not help distinguish the groups in the model construction. OPLS-DA separate variations that are orthogonal to the prediction, focusing on the variation that is directly related to the classification of the different classes or Y variables, generating models of easier interpretation.⁴⁶

The metrics of the concatenated model were similar to UPLC/MS and superior to the NMR individual's metrics (**Figure 3; Figure S4-5**). For the fused dataset, after Pareto the scaling, the different blocks were supervised by their anti-inflammatory profiles. After GZ sample (inactive) exclusion (**Table S4**), an enhanced clustering of the inactive samples was observed. The OPLS-DA model, generated with 3 components, yielded an excellent fit, demonstrated by R^2 of 0.99, Q^2 of 0.87, with an R^2Q^2 difference of only 0.12, indicating a robust classification and predictive capacity. Different from R^2 which explains the variance, Q^2 predicts the variance, providing information about model predictability. Thus, statistical models with a difference higher than 0.3 between R^2 and Q^2 are considered overfitted models.^{36,47} Thus, the developed OPLS-DA were well fitted and can be reliable for the investigation of bioactive markers in the samples and potentially useful for creating prediction models.

To ensure the validity of the obtained results for the concatenated OPLS-DA model, the possibility of model overfitness was double-checked, and additional 100-permutation tests were conducted. Permuted UPLC/MS-NMR concatenated dataset ($Q^2 = -0.08$), and NMR dataset alone acquired negative intercepted values ($Q^2 = -0.61$), thus corroborating the reliability of the developed model (**Figure 3; Figure S7**). Permuted UPLC/MS dataset acquired Q^2 values near zero (**Figure S7**). Permutation tests certify that the model is valid and its predictability was not compromised. The models that are not overfitted show in the permutation test a lower Y-intercept of R^2 compared to the non-permuted model and a Y-intercept of Q^2 less than 0.05 or negative.⁴⁸ In addition, the cross-validated residuals (CV-ANOVA) acquired a p -value of 0.036 and a Root Mean Squared Error (RMSE) of 0.2 for the first component and 0.1 for the second and third, respectively, thus providing additional satisfactory validation metrics for the model. The RMSE estimates the mean error expected for the predictions of the Y-variables, while the CV-ANOVA provides an estimation of the statistical significance of the model, indicating whether the observed separation between groups is likely due to genuine differences rather than random chance.⁴⁹

For data interpretation S-plot was investigated, as it provides a visual tool to interpret the results of OPLS-DA models, aiding in visualizing the relationship between the magnitude and the reliability of the features in distinguishing between active and

inactive groups. The bottom left quadrant of the S-plot from OPLS-DA indicates the presence of discriminant features, including NMR chemical shifts (green) and m/z -RT pairs (blue) (**Figure 3**). Features of chemical shifts (δ) 2.89, 3.05 and 4.89 ppm were clearly highlighted with high magnitude. Similarly, in the UPLC/MS data, such as the m/z 314.1870 at RT 1.97 min and m/z 328.1538 at RT 1.96 min were visually evident as a discriminant of less magnitude in the model but of high reliability. Additionally, the loadings plots (**Figure S6**) have corroborated the S-plots observations. Loadings represent the weights or coefficients that indicate the contribution of each original variable to the latent components, demonstrating how strongly each variable is correlated and differentiating between the Y variable in the model, thus helping to identify the most important and distinguishing features.^{25,45,50}

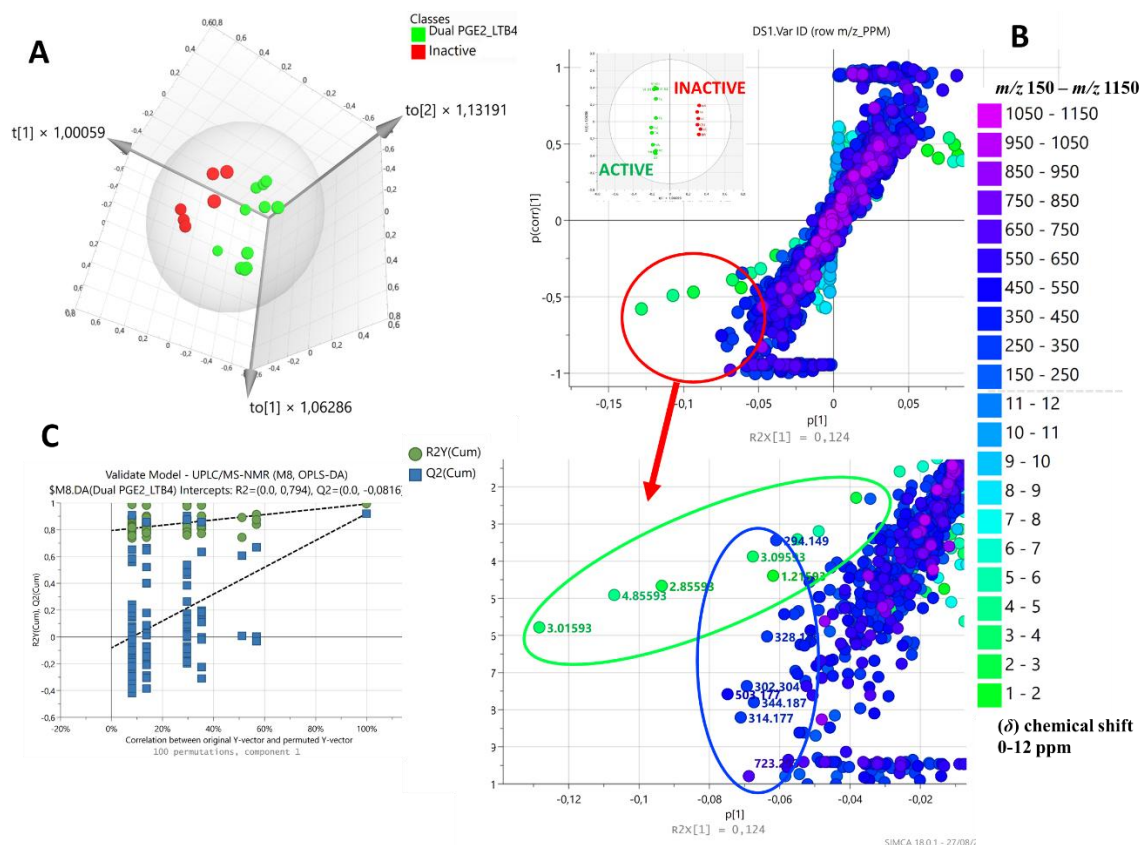


Figure 3. OPLS-DA model of *Ocotea* crude extracts for metabolite discrimination supervised by the dual anti-inflammatory activity (R^2 : 0.99, Q^2 : 0.87). (A) 3D score scatter plot of UPLC/MS-NMR fused data divided into two clusters: active (green) and inactive (red). (B) S-plots of UPLC/MS-NMR fused data highlighting discriminant features. (C) Model validation by 100x permutations test (intercepts, lower R^2 : 0.79 and negative Q^2 : -0.08).

Moreover, biplots based on the OPLS model of the fused UPLC/MS-NMR model were performed to simultaneously examine the correlation between the chemical variables and the active and inactive *Ocotea* samples. From the NMR data, a notable shift was observed in active groups within the chemical shift regions of (δ) 2-3 and (δ) 3-4 ppm.

Equally, in the UPLC/MS data, the m/z range of 250-550 was shifted in the direction of the active *Ocotea* species (Figure 4). Also, the OPLS model was used to define the key variables important for projection, identified by Variable Importance in Projection (VIP) scores (Figure 4). Variables with VIP scores > 1 are considered essential for separating samples based on their supervised activity.^{16,25} VIP scores higher than 1 and positively correlated with the active *Ocotea* species group were identified as important discriminant features, which could be responsible for the changes in the dual anti-inflammatory activity profile, particularly regarding the inhibition of PGE2 and LTB4 release.

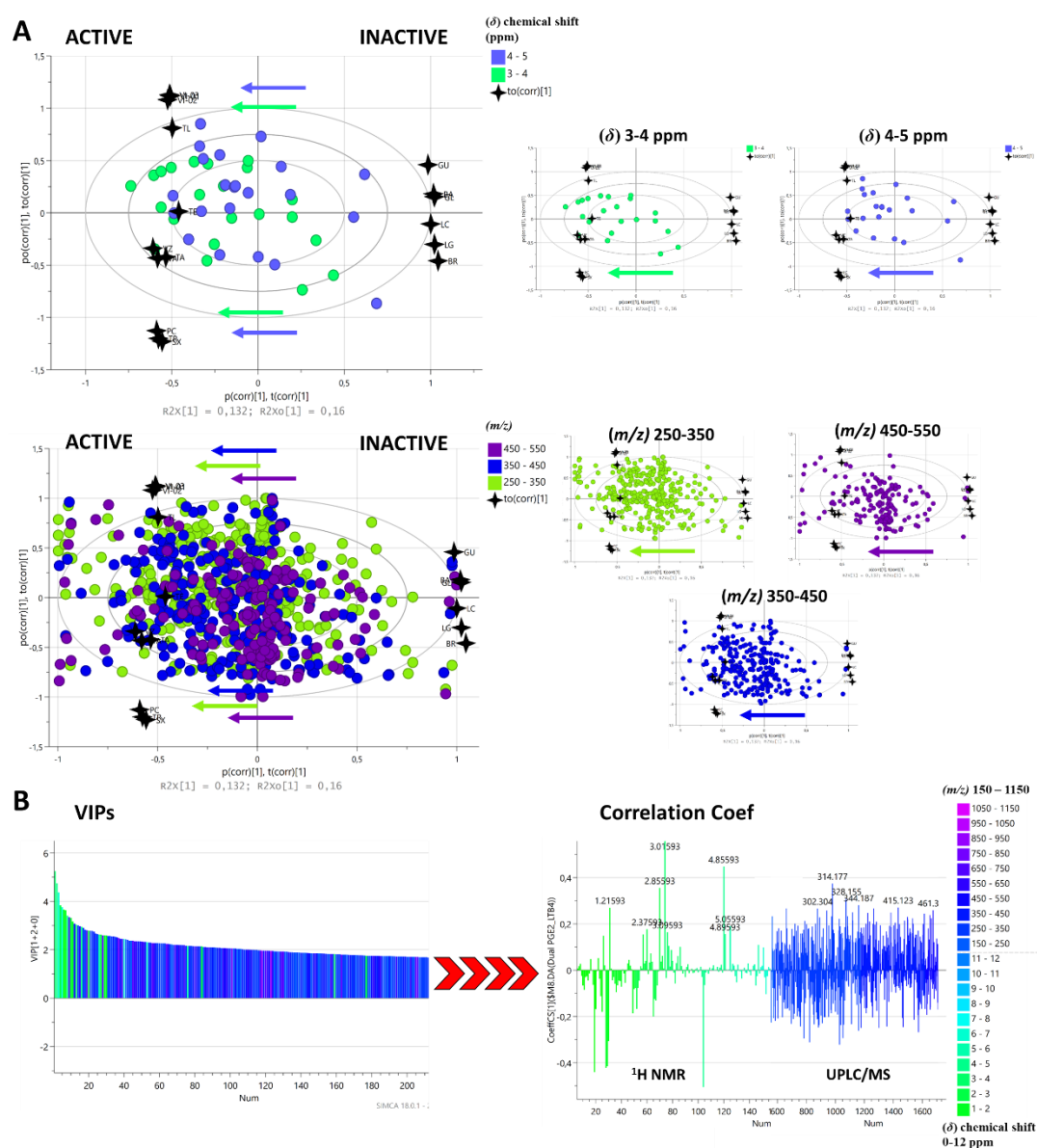


Figure 4. (A) OPLS-DA model biplots (loadings p and scores t) describing the simultaneous correlation between the chemical variables (^1H NMR chemical shifts and m/z features) and the *Ocotea* extract samples. NMR data highlighting chemical shift regions of (δ) 2-3 and (δ) 3-4 ppm. UPLC/MS data showing the m/z range of 250-550. (B) Variable Importance in Projection (VIP) scores > 1 (in green to blue) and positive Correlation Coefficients (CorrCoef) discriminate features in the concatenated UPLC/MS- ^1H NMR dataset.

Thus, the positive correlated discriminant features were ranked according to their significance to the model by the VIP score, leading to 311 features of VIP score from 1 to 5.24 (26 chemical shifts and 285 m/z -RT pairs). Those not positively correlated with the dual activity were excluded leading to 216 features (16 chemical shifts and 200 m/z -RT pairs). Lastly, only those features with p -value < 0.05 were considered for the final list, leading to 50 features, 4 chemical shifts and 41 m/z -RT pairs, which all might be the main ones related to the compounds that are worthy of further consideration. The false discovery rate (FDR) was calculated by adjusting the p -values to Benjamin-hock q -values (**Table 2; Table Z4**-Available at <https://doi.org/10.5281/zenodo.13826734>) increasing the confidence of the discriminating features. In addition, satisfactory area under the curve (AUC) values > 0.7 were also obtained for 4 NMR shifts and 15 from UPLC/MS m/z –RT pairs (**Table 2; Table S3**). These results corroborate the confidence in discriminating the most significant metabolites associated with the dual COX and LOX inhibition pathways.

Metabolite Annotation. Adhering to the criteria set forth by the Metabolomics Standards Initiative (MSI), the annotation of the compounds correlated with anti-inflammatory activity, by inhibiting the release of PGE2 and LTB4, was at levels level 2 and 3 of confidence.^{51–53} This approach ensures a standard annotation of metabolites, enhancing the reliability and depth of our results. Initially, level 3 automated annotation with the *OcoteaDB in-house* database was achieved by matching the precursor's (MS^1) monoisotopic masses.^{9,54,55} This strategy gives directions to chemotaxonomy-related compounds, once the database is constituted of NP previously isolated from the genus, avoiding initial misleading annotations.^{54,56}

The annotated hits had their fragmentation spectra manually inspected using MZmine 3 for comparison with available spectral data in online databases. This process enabled the level 2 confidence annotations of four bioactive markers (**Table 2**). For the other two, laudanine and 4-oxo-lanceolic acid, no public mass spectral data was available, thus their fragmentation patterns were proposed based on possible gas-phase fragmentation reactions. Consequently, these compounds were assigned at level 2b of confidence, according to the classification system proposed by Schymanski et al. (2014).⁵⁷ Other seven metabolites were annotated as level 3 according to the MSI confidence annotation guidelines related to hits from the Dictionary of Natural Products (DNP), which included caffeoyl-glycerol, sinapyl alcohol and stigmasterol glycoside derivatives (**Table S3**). In addition, intense unidentified features with VIP > 1.5 were subjected to molecular formula proposal using MZmine algorithms and following fundamental MS rules (*e.g.*, nitrogen rule and isotopic pattern). This strategy resulted in, two metabolites annotated as level 4,

they had zero hits in the DNP, offering insights into the future isolation of potential new bioactive compounds (**Table S3**).

Table 2. The list of potential bioactive markers annotated at a Level 2 confidence level positively correlated with anti-inflammatory activity by inhibiting the release of PGE2 and LTB4.

Observed m/z	RT ^a	VIP ^b	FDR ^c (q -value) / AUC	MF ^d	Error (ppm)	Adduct	Proposed Annotation (Level 2)	Observed product-ions (m/z)
314.1769	1.71	2.9	4.37E-02 / 0.53	C ₁₉ H ₂₃ NO ₃	4.7	M+H	Armepavine	103.056; (107.049); 115.054; 131.049; 151.077; 165.069; 176.063 ; 181.067; 194.072; 237.090; 268.099 (<i>b</i>) 137.059 ; 176.063 ; 189.068; 298.119 ; 313.127 ; 329.155 (<i>c</i>)
344.1870	1.97	2.8	3.08E-02 / 0.76	C ₂₀ H ₂₅ NO ₄	3.8	M+H	Laudanine	178.077; 194.072; 191.208; 205.064 ; 222.066 ; 237.090 ; 251.096 ; 265.084 ; 281.069 ; 296.093 ; (311.118) (<i>b</i>) 151.054; 165.072; 177.069 ; 197.072; 205.064 ; 222.066 ; 237.089 ; 265.084 ; 282.0868 ; (297.099) (<i>a</i>) (<i>b</i>)
328.1538	1.96	2.4	4.56E-02 / 0.77	C ₁₉ H ₂₃ NO ₄	0.7	M+H	Laurotetanine or Norisocorydine	151.004; 152.010; 180.004; 275.054; 284.033; (285.038); 297.098; 303.050 ; 431.135 ; (449.1080) (<i>a</i>)
328.1539	1.94	2.7	4.72E-02 / 0.76	C ₁₉ H ₂₃ NO ₄	0.7	M+H	Isoboldine or Boldine	107.0482 ; 165.0621 ; (203.142) 219.142 (<i>c</i>)
449.1080	2.62	2.2	4.51E-02 / 0.68	C ₂₁ H ₂₂ O ₁₁	-2.2	M-H	Astilbin	
247.1330	5.57	2.1	1.39E-02 / 0.70	C ₁₅ H ₂₀ O ₃	-3.8	M-H	4-oxo-lanceolic acid	

^a Retention Time, ^b Variant Important in Projection, ^c False Discovery Rate, ^d Molecular formula. MS/MS database comparison: (a) – MoNA, (b) – GNPS. For compounds of unknown available spectra (*c*), annotation followed by Demarque et al., 2016 and Carnevale Neto et al., 2020^{58,59}, annotated also according to Schymanski et al. (2014). MS^E of low and high-energy scans are shown in **Figure S15-S20**). Note: The highest abundance fragment ion is shown in parenthesis. Product-ions are highlighted in bold and had their fragmentation mechanism proposed.

The OPLS-DA pointed ions at m/z between 250 and 550 correlated to the dual anti-inflammatory bioactivity profile (**Figure 4**). These features presented FDR (q -values) lower than 0.05, representing the true positive annotation that distinguishes the potential discriminating metabolites associated with anti-inflammatory activity (**Table 2**). Gas-phase fragmentation reactions aided in uncovering distinct patterns for the annotated bioactive markers. Manual analysis of the high-energy positive mode MS^E (fragment spectra) showed the ions at m/z 328 (RT 1.94 and 1.96 min) presented neutral losses related to aporphine alkaloids, while the ions at m/z 314 (RT 1.71 min) and 344 (RT 1.97 min) present neutral losses commonly observed in benzylisoquinoline alkaloids (**Figure 5-6**; **Figure S8**). The fragmentation pathways of these compounds can be highly influenced by the *N*-substitution pattern and the presence of adjacent methoxyl and

hydroxyl groups at the A and C aromatic rings,^{54,60,61} thus different fragmentation mechanistic pathways were proposed (**Figure 5-6; Figure S8**). The displayed m/z of the drawn fragments was calculated based on theoretical mass values (**Figures 5-8**).

The fragmentation patterns and diagnostic ions observed for the protonated alkaloids at m/z 328 have been previously described in the *Ocotea* genus,^{9,62} and match aporphines, such as the isomers isoboldine and boldine, as well as noraporphines, such as laurotetanine and norisocorydine. Due to the complexity of the employed data-independent acquisition (DIA) method and the non-targeted approach applied, it was not possible to precisely differentiate either of these pairs solely on matching MS/MS spectra, as they display several common fragment ions. The noraporphines, laurotetanine or norisocorydine, besides displaying the same precursor ion of the aporphines boldine and isoboldine at m/z 328, exhibit a different neutral loss of 17 Da (m/z 328 \rightarrow m/z 311; M+H-17). This pathway is related to the simple inductive α -cleavage releasing ammonia -NH₃ instead of 31 Da (m/z 328 \rightarrow m/z 297; M+H-31) in the neutral loss of methylamine -NH₂CH₃, which is common to aporphines (**Figure 5**). This change is attributed to the replacement of the methyl group by hydrogen in their *N*-substitution pattern, which allows the differentiation of these compounds by their fragmentation pattern.⁵⁹ The mechanistic pathway can also be explained by the opening of the B ring (**Figure S8**). Further losses in (nor)aporphine rings of CH₃OH (32 Da) were also observed (m/z 297 \rightarrow m/z 265 and 311 \rightarrow 279; M+H-32), followed by the loss of CO (28 Da) (m/z 265 \rightarrow m/z 237 and 279 \rightarrow 251; M+H-28) (**Figure 5; Figure S8**), which are expected neutral losses for these alkaloids.^{54,59,63}

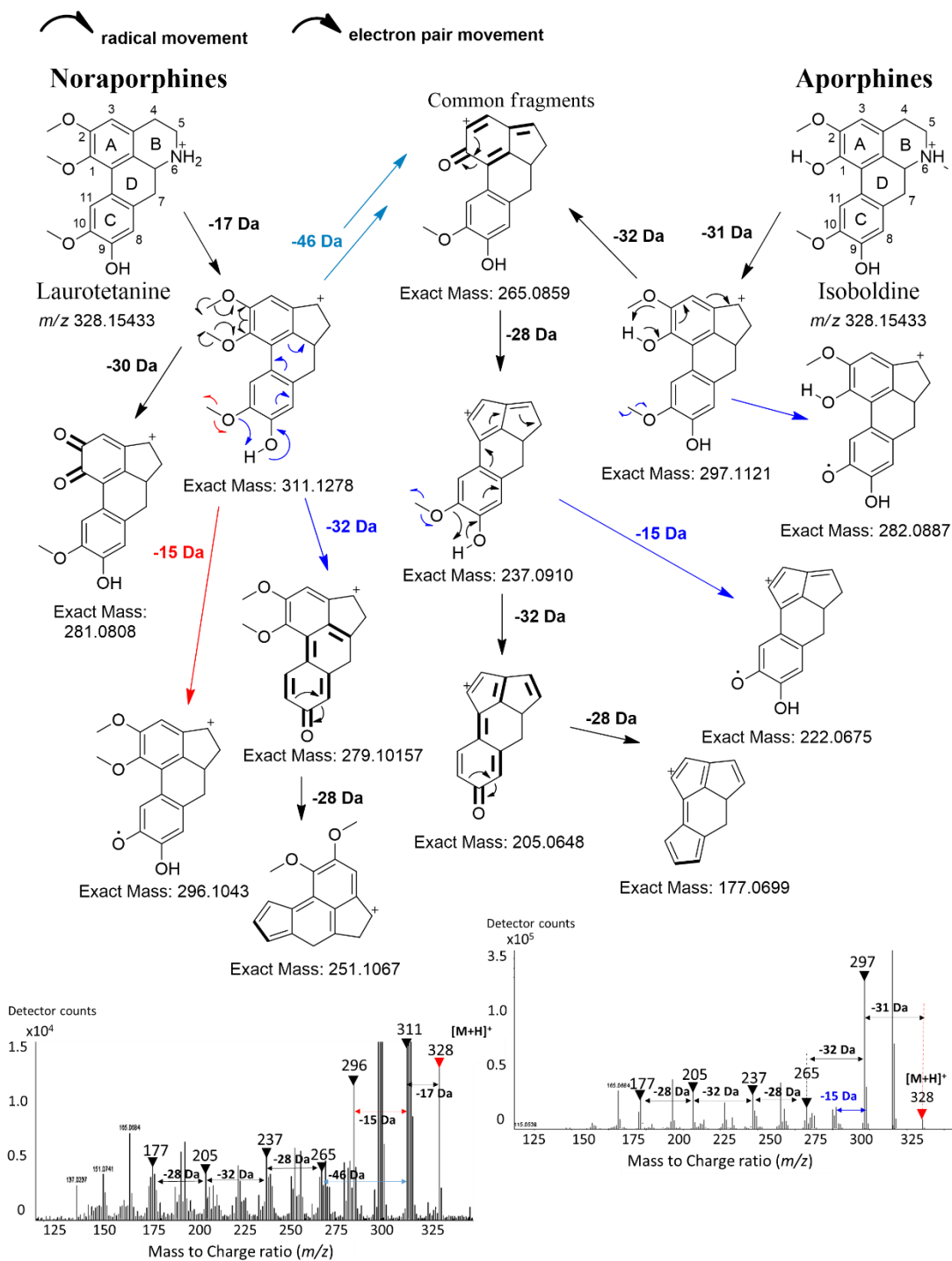


Figure 5. Proposed fragmentation pathway of the noraporphine laurotetanine (left) and the aporphine isoboldine (right). Similar fragment ions and common neutral losses are highlighted (15 Da – CH₃; 17 Da – NH₃; 28 Da – CO; 31 Da- NH₂CH₃; 32 Da – CH₃OH).

For the annotated benzyloisoquinoline armepavine besides matching important fragments to the reference spectra, the high-intensity diagnostic fragment ion at *m/z* 107 suggested the presence of benzyloisoquinolines mono-hydroxylated in their C ring corroborating the annotation of armepavine. This ion at *m/z* 107 is a diagnostic ion proposed to be produced by direct elimination of the C ring, resulting in the neutral loss of the isoquinoline ring (-207 Da; C₁₂H₁₇NO₂). This pathway might be favoured due to the formation of a benzylic carbocation (C₆H₅CH₂⁺), a more stable fragment, where the

positive charge can be stabilized by the equilibrium with the aromatic π system, and in this case, even extended due to the presence of the *p*-hydroxyl substituent (**Figure 6**). Following the same mechanistic pathway, for laudanine, the diagnostic fragment ion at m/z 137 indicates the presence of an additional methoxyl group in the C ring, which for laudanine is at position 5' (**Figure 6**). Moreover, the observed 30 Da loss is likely due to a homolytic cleavage at the methoxyl groups ($-\text{OCH}_3$), between C-O bonds, leading to the loss of two methyl radicals ($-\text{CH}_3$, 15 Da each), and the subsequent formation of carbonyl groups ($\text{C}=\text{O}$). Also, other losses of 15 Da ($-\text{CH}_3$) (m/z 313 \rightarrow 298; m/z 283 \rightarrow 268 - **Figure S8**) were also observed for the benzyloquinoline class.

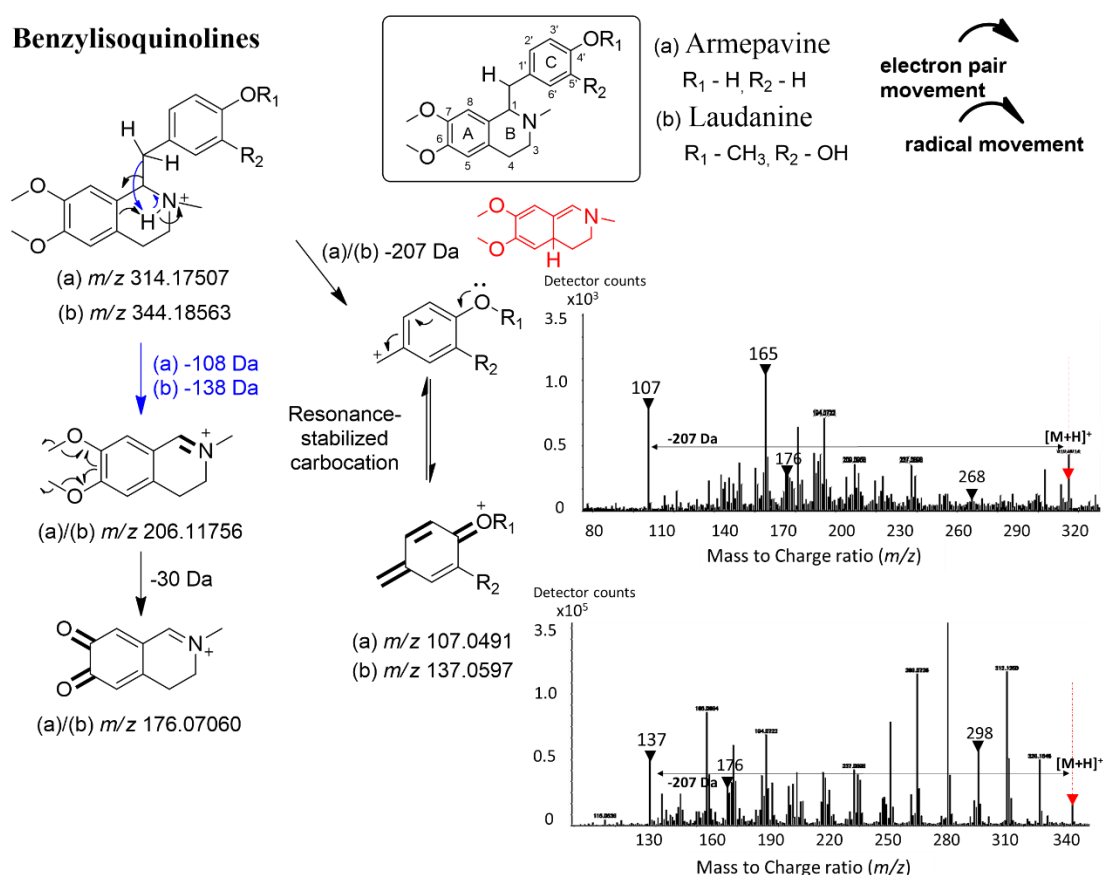


Figure 6. Proposed fragmentation pathway of the benzyloquinolines armepavine and laudanine. Similar fragment ions and common neutral losses are shown (108 Da $-\text{C}_7\text{H}_8\text{O}$; 138 Da $-\text{C}_8\text{H}_{10}\text{O}_2$; 207 Da $-\text{C}_{12}\text{H}_{16}\text{NO}_2$).

Similarly, the high-energy negative mode MS^E spectra of the ion at m/z 449, showed neutral losses commonly observed in the fragmentation of glycosylated flavonoids due to the observation of a neutral loss of the attached sugars.^{25,55} This allowed proposing a mechanistic pathway for deprotonated flavonoids. The observation of high abundance fragments at m/z 303 $[\text{M}-\text{H}-146]^-$ and at m/z 285 $[\text{M}-\text{H}-164]^-$ suggested the loss of a deoxyhexose unit attached by an *O*-glycosidic link, and the presence of the

taxifolin flavanonol (or 2,3-dihydroflavonol) aglycone. Deoxyhexose units (164.07 Da) are six-carbon tetra-substituted saccharides that compared with common hexoses, such as glucose, the deoxyhexoses do not present hydroxyl groups at C6 (deoxy), which is replaced by a hydrogen, forming a methyl group in C5 (**Figure 7**).

Accordingly, the fragment ion at m/z 303 represents the intact taxifolin aglycone, which is the dihydro form of quercetin, thus retaining the hydroxyl at C3 after a heterolytic cleavage of the *O*-glycosidic bond. The observed fragment at m/z 285, suggests also the heterolytic mechanism but directed to the flavanonol ring instead of the sugar unit, with loss of the hydroxyl at C3 and formation of a double bond between C2 and C3. While the observation of m/z 285 is typically associated with flavonol kaempferol aglycone⁶⁴, the concomitant presence of m/z 303 suggests the taxifolin flavanonol aglycone⁶⁵, once the loss of the sugar unit with the elimination of the hydroxyl group is not common in flavonols due to the presence of already a double bond between C2 and C3.^{55,64} In addition, the dehydration pathway is also likely to occur due to the observation of a less abundant neutral loss of 18 Da, potentially related to one water molecule elimination (m/z 449 \rightarrow m/z 431; $[M-H-18]^-$). Besides, the fragment at m/z 303 matches the taxifolin deprotonated molecule in online MS/MS databases (**Figure 7**; **Table 2**).

Glycosylated flavanonol

Taxifolin *O*-deoxyhexose (astilbin)

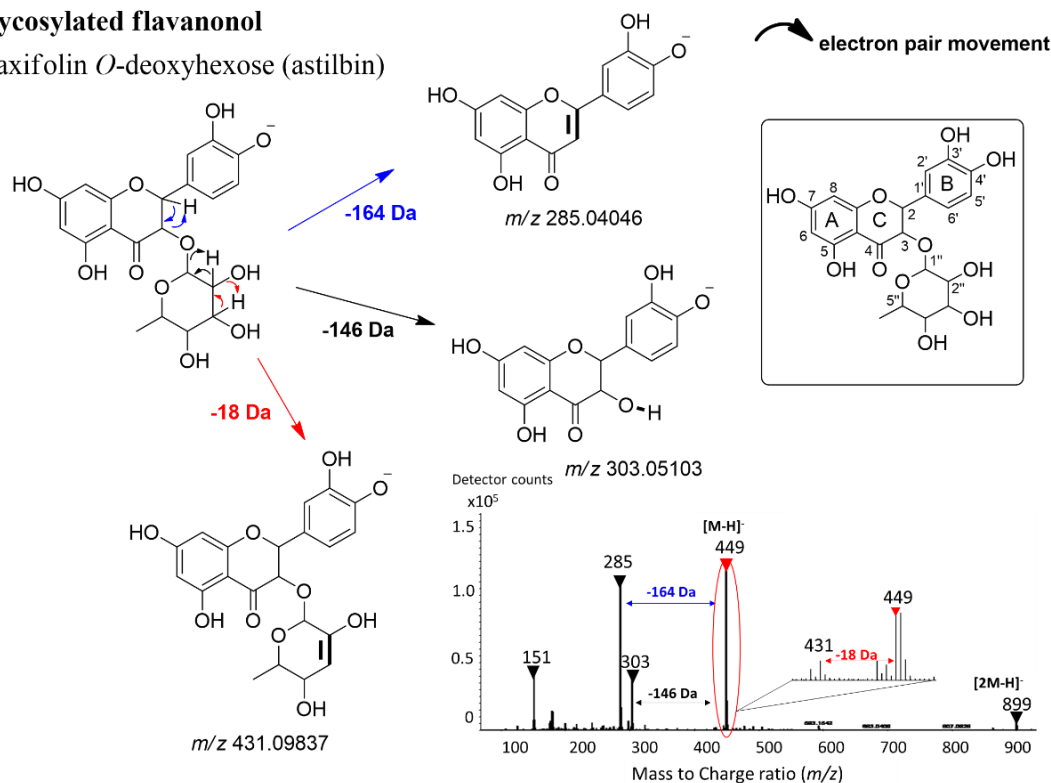


Figure 7. Proposed fragmentation pathway of astilbin (taxifolin 3-*O*-deoxyhexose). Three pathways are highlighted: loss of rhamnosyl (-146 and -164 Da neutral loss) at position 3 via glycosidic bond cleavage, and loss of water (18 Da neutral loss).

Thus, these observed neutral losses and fragments helped identify the substitution pattern as an *O*-glycosylation of a deoxyhexose sugar in a taxifolin aglycone. The ion at

m/z 449 and RT 2.62 min matched the 3'-*O*-hydroxy-glycosylated flavanonol astilbin (taxifolin 3-*O*-rhamnoside), from the *Ocotea*DB⁵¹. Thus, despite limitations regarding establishing the exact position of the deoxyhexose moiety and the specific deoxyhexose sugar, chemotaxonomy and MS data supported the annotation of astilbin (**Figure 7**).

Moreover, the high-energy negative mode MS^E spectra of the ion at m/z 247 allowed the annotation of the sesquiterpenoid 4-oxo-lanceolic acid. It displayed an evident main neutral loss of 44 Da resulting from a potential decarboxylation reaction through α -cleavage (m/z 247 \rightarrow m/z 203; [M-H-44]⁻). Thus, that allowed proposing a mechanistic pathway for deprotonated compound, where the formed fragment at m/z 203 can further undergo a γ -cleavage to form a stable *O*-methylphenolate anion (m/z 203 \rightarrow m/z 107; [M-H-44-96]⁻). Another possible fragmentation pathway is a Retro Diels Alder (RDA) reaction originating a fragment ion at m/z 165 (m/z 247 \rightarrow m/z 165; [M-H-82]⁻). In addition, the deprotonated molecular ion can undergo a CO loss to form a fragment ion at m/z 219 (m/z 247 \rightarrow m/z 219; [M-H-28]⁻) as well as another γ -cleavage to also form the *O*-methylphenolate anion (m/z 247 \rightarrow m/z 107; [M-H-140]⁻) (**Figure 8; Table 2**).

Furthermore, this annotated sesquiterpenoid has been previously isolated in the leaves of *O. minarum*,⁶⁶ along with other β -sesquiphellandrene sesquiterpenoid derivatives. Although, there are no available MS/MS spectra for direct data comparison, based on MS^E analysis it matched expected precursor and fragment ions and has an associated *Ocotea* biosynthetic background, which further increases the confidence in this metabolite's annotation.

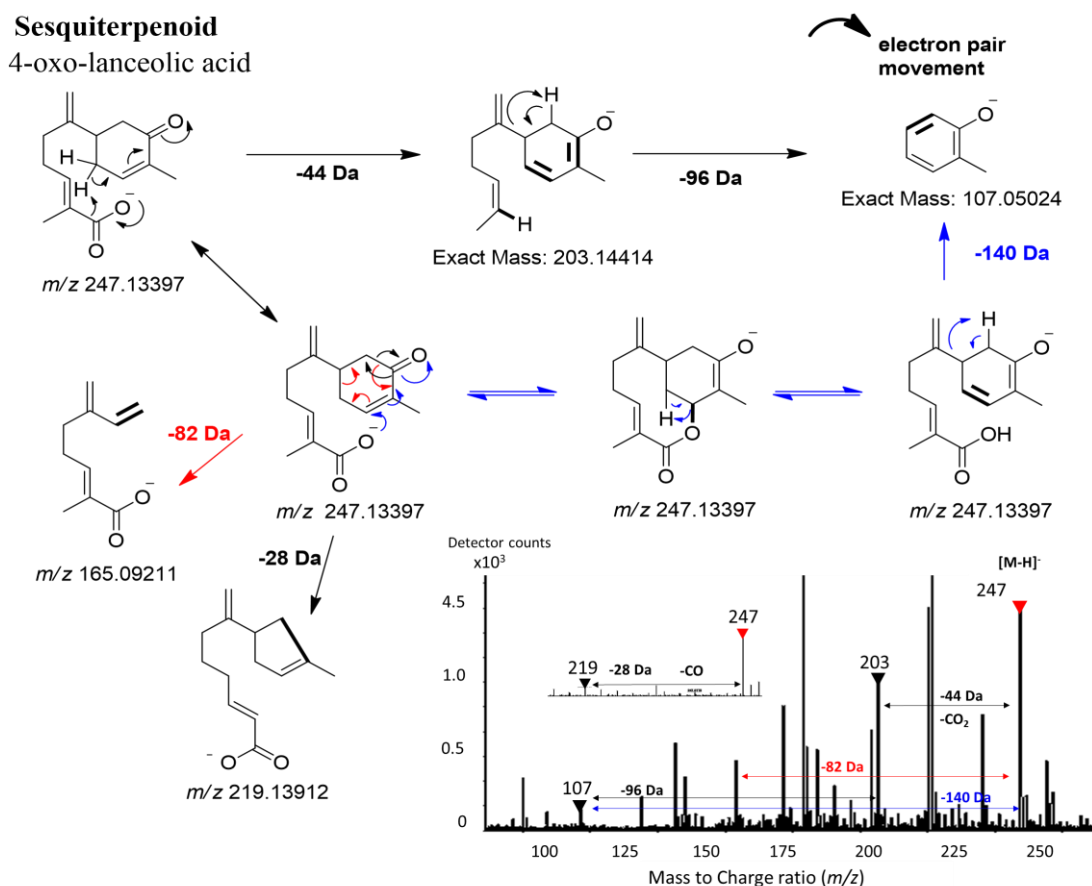


Figure 8. Proposed fragmentation reactions of 4-oxo-lanceolic acid. Decarboxylation reactions of α -cleavage lead to the loss of 44 Da (CO₂) (m/z 247 \rightarrow m/z 203) followed by to form an *O*-methylphenolate anion (m/z 203 \rightarrow m/z 107; [M-H-44-96]⁻). Other alternative pathway include 3 mechanistic pathways: Retro Diels Alder (RDA) reaction originating a fragment ion at m/z 165 (m/z 247 \rightarrow m/z 165; [M-H-82]⁻); CO loss to form a fragment ion at m/z 219 (m/z 247 \rightarrow m/z 219; [M-H-28]⁻); γ -cleavage to also form a *O*-methylphenolate anion (m/z 247 \rightarrow m/z 107; [M-H-140]⁻).

Regarding the NMR data, there are intense signals in the ¹H NMR spectrum (2.0–4.0 ppm) that were observed as singlets or broad singlets. Based on existing chemical knowledge of the genus and the annotation results obtained from UPLC/MS, the signals observed are reasonable with the presence of benzyloisoquinoline and aporphine-substituted alkaloids, as well as glycosylated flavonoids.

In the case of the alkaloids, the more unshielded ones (5.0–4.0 ppm) could be related to the presence of potential methoxyl groups in the A and C aromatic rings of these alkaloids, and the more shielded (2.0–3.0 ppm) potentially *N*-methyl or *N*-H groups in the B ring.^{36,67,68} However, these regions in the range of 3.00–5.0 ppm also align with the proton environments of glycosylated compounds, such as saccharides moieties (Figure 7; Figure S3).^{36,69} In addition, the presence of overlapping signals in this region is indicative of several protons from carbohydrates, such as hexoses, deoxyhexoses, pentoses and other potential saccharide moieties commonly found attached to flavonoid units.^{55,69,70} Due to the intensity of this multiplexed region, it was perceived that the concentration of the flavonoids in these crude *Ocotea* leaf extract samples might be higher

than the alkaloid content (**Figure S9**). Moreover, the signals in the ^1H NMR spectrum (6.00 - 8.00 ppm) are characteristic of the aromatic protons, which are present in both classes of alkaloids and flavonoids NP classes. Thus, the observation of these unshielded chemical shifts could be attributed to the aromatic protons of the aglycone portion of flavonoids (rings A and B), as well as to the aromatic rings of aporphines and benzyloquinolines (rings A and C) (**Figure S10**).

The loadings plot and S-plot from OPLS-DA, based on the UPLC/MS-NMR concatenated datasets (**Figure 3** and **Figure S6**), highlighted chemical shift signals in the more shielded proton region (2.80–4.90 ppm) that show a higher correlation with the presence of anti-inflammatory bioactivity (**Figure 4**). The chemical shift signals at 2.89, 3.05, 3.25, and 4.89 ppm, with false discovery rate (FDR) q -values below 0.05 represent true positive annotations and ^1H NMR discriminant signals associated with dual anti-inflammatory activity (**Table 3**). The less shielded signals at 4.89 and 3.25 ppm were associated with the presence of glycosylated flavonoids. The signal at 4.89 ppm is likely attributed to the methyne hydrogen of anomeric carbons (H-1'). Anomeric protons generally resonate between 4.0 and 5.5 ppm due to the deshielding effect of the oxygen of the glycosylic linkage attached to the anomeric carbon.⁶⁹ Whereas, the signal at 3.25 ppm could possibly come from the other sugar protons (adjacent carbinol groups), such as H-2', H-3', H-4', and H-5' protons. The more shielded methyne groups of the carbohydrates resonate between δH 3.0 and 5.0 ppm, varying with the specific sugar, glycosidic linkage, and substitution pattern on the flavonoid backbone.^{36,69} The signal at 2.89 ppm could be indicative of the presence of protons from adjacent methylene groups or unusual methyl proton shifts due to the potential presence of close electronegative atoms.⁷⁰ While the signal at 3.05 ppm likely corresponds to protons derived from methylene groups in the aliphatic portion of aporphine and benzyloquinoline alkaloids (**Table 3; Figure 5-6**). Aliphatic methylene protons typically resonate between 2.5 - 3.5 ppm, depending on the degree of substitution and proximity to heteroatoms or aromatic rings that can influence their shielding.⁷⁰

The signal at 4.89 ppm was assigned to a proton likely associated with the anomeric carbon of glycosylated flavonoids. The observed multiplicity (doublet) is likely derived from the coupling of H1' (the proton on the carbon in the glycosidic linkage) with H2', the adjacent carbinol in the saccharide moiety, resulting in a characteristic deshielded doublet, commonly observed in the NMR spectra of glycosylated flavonoids. This assignment is supported by the consistent appearance of this doublet in the ^1H NMR spectra of the active *Ocotea* species (**Figure 9A**), corroborating the presence of the

annotated taxifolin-3-*O*-rhamnoside (astilbin) or even other similar glycosylated flavonoids. The anomeric configurations of the sugar units were inferred from their measured $^3J_{H-1/H-2}$ coupling constants. Since this doublet resonates at 4.9 ppm with a coupling constant of 3.55 Hz, it is inferred that the sugar moiety is likely in the α -configuration ($J < 5.0$ Hz).^{36,69} The highest peak areas of this signal were observed mostly in the spectral data of the active samples, particularly *O. teleiandra* and *O. tabacifolia* (Table Z3, Figure 9A). Interestingly, reports of isolated astilbin in the literature also indicate the α -configuration, such as in *O. canaliculata* and *O. elegans*, further supporting this assignment through chemosystematic and phytochemical studies.^{71,72}

Table 3. The list of NMR discriminant features from the MS-NMR fused data that contribute to the differentiation of the dual anti-inflammatory activity model of *Ocotea* crude sample extracts.

ID	δH (ppm) ^a	δH , Mult. ^b (<i>J</i> in Hz)	Proposed type of proton	STOCSY ^c δH (<i>J</i>) – corr.	VIP ^c	AUC ^d	FDR ^e (<i>q</i> - value)	Proposed annotation
nmr_205	2.85- 2.89	2.89, <i>bt</i>	Alkyl (methyne), -CH	3.53, <i>m</i> – 0.95	3.82	0.78	4.53E- 02	Saccharide unit (Flavonoids)
nmr_201	3.01- 3.05	3.05, <i>m</i>	Alkyl (methylene), -CH ₂	3.09, <i>m</i> – 0.94; 6.85 <i>m</i> – 0.87	5.24	0.82	4.45E- 02	Alkaloids
nmr_196	3.21- 3.25	3.25, <i>bs</i>	Alkyl (methyne), -CH	3.14, <i>m</i> – 0.92; 0.82 <i>m</i> – 0.88	1.14	0.88	4.62E- 02	Saccharide unit (Flavonoids)
nmr_155	4.85- 4.89	4.89, <i>d</i> , (≈ 3.55)	Anomeric – CH-OH	7.29, <i>d</i> , (≈ 2.7) – 0.95; 5.26, <i>bs</i> – 0.90; 4.26, <i>bd</i> (≈ 7.55) – 0.87	4.78	0.70	4.95- 02	Saccharide unit (Flavonoids)

^a Chemical shift (δH – ppm) range in the 1D projection spectra, ^b Multiplicity and coupling constant in hertz (Hz). *b* – broad signal of compromised multiplicity visualization. ^c Statistical Total Correlation Spectroscopy – corr – correlation value, ^e Variant Important in Projection, ^d Area Under the Curve ^e False Discovery Rate. NA – not annotated. # Note: Only the most significant correlations with the driver signal of STOCSY were reported (> 0.85).

Statistical Total Correlation Spectroscopy (STOCSY) was utilized to explore the variance and the potential metabolite correlations within the ¹H NMR dataset to further assist in enhancing the confidence in the bioactive marker annotations.^{23,73,74} STOCSY employs covariance (combined variability among different variables) and correlation calculations between peaks to identify features from the same molecular structure or similar pathways. The usage of STOCSY calculations has increased in metabolomics studies in the last ten years, in order to facilitate the biomarker identification stage.^{23,73–75} Thus, the STOCSY analysis method was applied and has successfully aided in identifying

significant correlations among the chemical shifts across the ^1H NMR spectra. This method starts by selecting a “driver peak” and then measures its covariance and correlation with all intensity peaks across samples, assuming that multicollinearity indicates a similar molecular origin.^{23,75}

STOCSY successfully aided in identifying the multicollinearity of chemical shift intensities across the analysed active dataset. By using the discriminant signals (FDR < 0.05) as the signal drivers (**Table 3**), most of the correlated signals were found to correspond to the chemical shifts in the aromatic (Ar-H), anomeric (CH-OH), and carbohydrate proton regions, which are characteristic of glycosylated flavonoids (**Figure 9B**). The most deshielded aromatic signals in the range of 7-8 ppm, typically observed as doublets with $J < 5$ Hz, are often associated with B-ring protons of a flavonoid aglycone with meta-coupling, such as at positions 2' and 6'. In contrast, higher coupling constants ($J > 7$ are observed for ortho-coupling at positions 5' and 6' of the B-ring.^{69,70} Interestingly, STOCSY revealed a high correlation between the driver signal at 4.89 ppm and the aromatic signal at 7.29 ppm, suggesting that these discriminant signals could indeed correspond to glycosylated flavonoid derivatives (**Table 3, Figure 9B**). The aromatic signal at 7.29 ppm appeared as an unresolved doublet or broad singlet (**Figure S11**). This may be due to the lower intensity of the signal, as well as the natural farther coupling distance of meta protons, potentially leading to the observed unresolved coupling. Nonetheless, a coupling constant ($J \approx 2.6$ Hz), could be measured, supporting the correlation with a potential proton at a meta position on the aromatic ring. Also, high correlations at 5.26 (*bs*) and 4.26 (*bd*, $J=7.55$ Hz) ppm were also observed. The unshielded signal at 5.26 ppm likely arises from methyne groups directly attached to oxygen atoms, at positions 2 or 3 of the C-ring in taxifolin-derived glycosylated flavonoid cores (**Figure 8B**), as the oxygens pull electron density and typically deshields these protons. The presence of these high correlations further supports the designation of the signal at δH 4.89 as an alkyl methyne group of anomeric carbon of flavonoids.

Furthermore, the methylene proton region at δH 3.05 (*m*) ppm could be indicative of aliphatic methylene groups in the B-ring of either aporphines or benzyloquinolines. STOCSY also showed strong correlations in the same multiplet region at δH 3.09, as well as in the aromatic region at 6.85 ppm (**Figure S12**).

The signal in the shielded region may be attributed to methylene groups within the core structure, such as the adjacent methylene groups of the aporphine (positions H4 and H5) and benzyloquinoline B-ring (positions H3 and H4) or the methylene groups connecting the isoquinoline B-ring to the benzyl C-ring in benzyloquinolines and

related alkaloids, which also could include aporphines unsubstituted at the D-ring (position H7) (**Figure S12**). These correlated chemical shifts by STOCSY corroborate that these discriminant chemical shifts could indeed be derived from aliphatic methylene groups likely originating from the presence of benzyloquinolines and aporphine alkaloids. Additionally, still relevant correlations (>0.65) were observed for potential methoxy group signals at δ H 3.65, δ H 3.73, and δ H 3.77, which is a well-known characteristic chemical shift range of aromatic methoxyl groups in these alkaloid classes.^{5,67,68,76} (**Figure S12**).

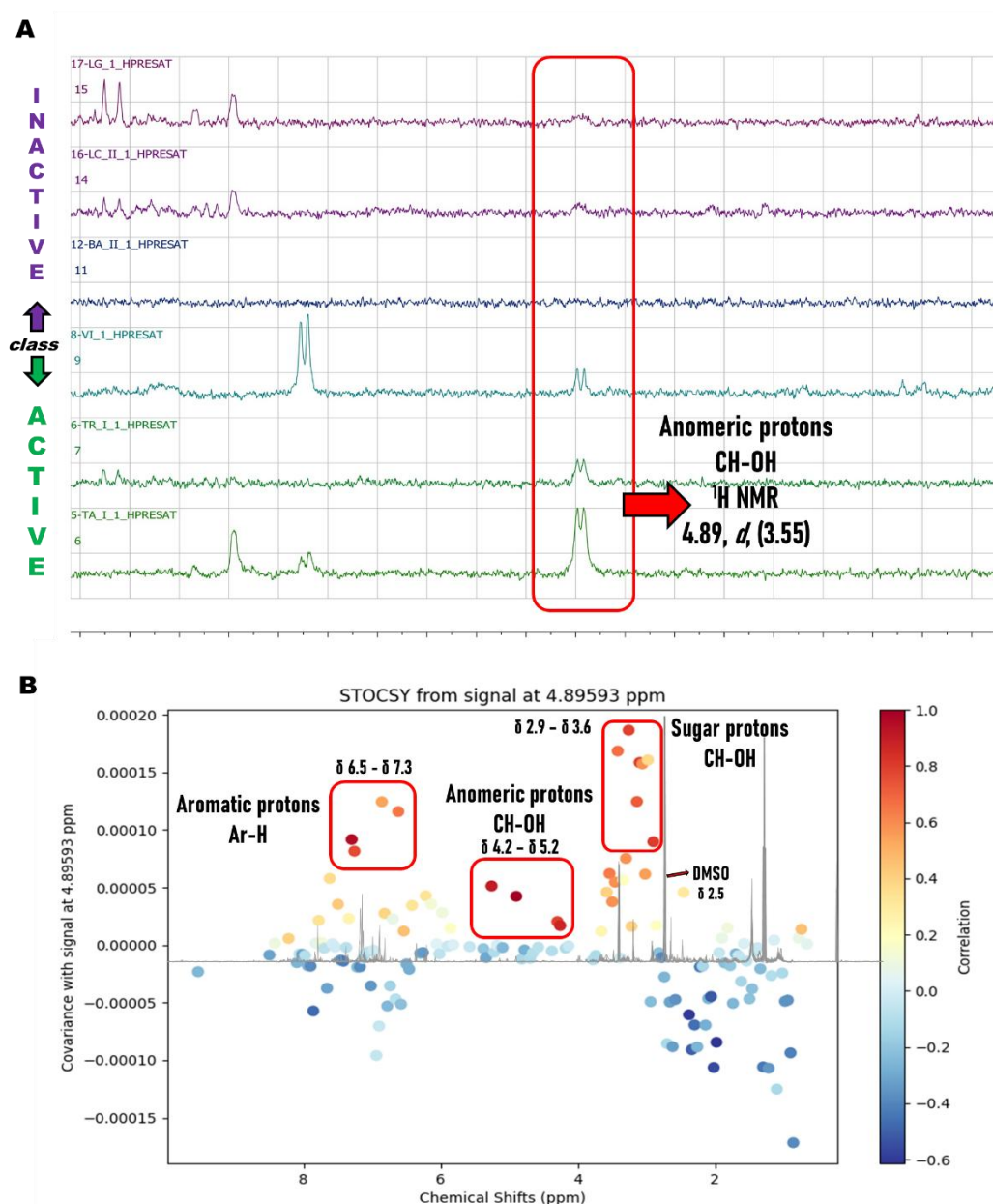


Figure 9. (A)- Hydrogen of anomeric carbon that appears predominantly in the ^1H NMR spectra of the active *Ocotea* species. (B)- STOCSY correlation with driver signal at 4.89 ppm. The gradient color of the STOCSY scatter plot represents the degree of correlation, with warmer colors (yellow/red) indicating a stronger positive correlation, and cooler colors (blue/purple) indicating negative or weak correlations.

In this context, OPLS-DA and STOCSY aided in exploring the variance and finding the most pronounced correlations among the several signals throughout the NMR spectral dataset, enhancing the ability to discern and interpret complex metabolic interactions. STOCSY aided in identifying signals from the same feature based on the variance in its detected levels within the different analysed spectra. Of notice, it is important to mention that two or more metabolites involved in the same biosynthetic pathway can also present significant intermolecular correlations in STOCSY because of natural biological covariance.²³ However, even though signals may come not from the same molecule, these signals covariate and this type of information could be relevant and sometimes detrimental for robust bioactive markers investigation and analysis, as they can corroborate the existence of similar chemical structures that can also discriminate active from inactive samples.^{23,75}

Furthermore, as previously mentioned, aporphines and benzyloquinoline alkaloids are a class of NP known for their significant biological activities, including anti-inflammatory.^{6,7,9} These alkaloids are primarily found in various plant species, where they have been traditionally used for their therapeutic properties.^{6,77} Previous studies in the literature have demonstrated that aporphines and benzyloquinoline alkaloids can modulate inflammatory responses by inhibiting key enzymes and signaling pathways involved in multiple inflammation targets, including the COX and LOX enzymes.^{6,8,9,63,78,79} The ability of aporphines and benzyloquinoline alkaloids to target important inflammatory pathways makes them promising candidates for developing new anti-inflammatory drugs. The results from multivariate analysis indicated the aporphines boldine/isoboldine, the noraporphines laurotetanine / norisocorydine as significant discriminant metabolites as well as the benzyloquinolines armapevine and laudanine.

Regarding the glycosylated flavonoids, this class of metabolites is increasingly recognized for their extensive profile of antioxidant and anti-inflammatory properties. These naturally occurring substances are found in a wide range of fruits, vegetables, and medicinal plants.⁸⁰⁻⁸² Particularly, astilbin is commonly found in various plants, such as those belonging to the *Astilbe* species, with associated popular use in traditional Chinese medicine, although also known to be produced by Lauraceae plants, and has been already isolated in the *Ocotea* genus.^{71,72,83} Astilbin was reported in the literature to promote *in vivo* immunomodulation and inhibit inflammatory pathways by reduction in the production of pro-inflammatory cytokines and mediators, such as PGE₂, TNF- α , IL-1 β , IL-10 and nitric oxide. It was proposed that the anti-inflammatory effect was also due to the inhibition of the NF- κ B pathway.⁸⁴ Thus, the literature suggests that astilbin could be

an effective NP anti-inflammatory agent, corroborating the findings of this metabolomics investigation. Herein, it is demonstrated the potential of astilbin as a bioactive anti-inflammatory marker in the dual COX and LOX pathways, which are important enzymes for the release of several pro-inflammatory mediators, such as PGE2 and LTB4.

Furthermore, β -sesquiphellandrene is a sesquiterpene found in several essential oils, such as those from ginger, turmeric, and citrus species and also found in *Ocotea* plant species.^{66,85–87} It is known for its aromatic properties and biological activities, including anti-inflammatory and antioxidant effects. Studies suggest that essential oils⁸⁸ and plant extracts⁸⁹ containing β -sesquiphellandrene can modulate important inflammatory signaling pathways, such as NF- κ B, and inhibit the production of pro-inflammatory mediators prostaglandins (such as PGE2), and cytokines (such as IL-1 β , and IL-6), which are typically elevated in inflammatory conditions and play key roles in the inflammatory response.^{88–90} Due to similar chemical structures, the 4-oxo-lanceolic acid is a potential derivative of β -sesquiphellandrene, featuring additional carboxylic acid and ketone groups, thus, is likely to retain or even enhance the anti-inflammatory properties. As this metabolite was associated as a potential marker in this metabolomics study, it is expected that the 4-oxo-lanceolic acid could improve its interaction with enzymes or receptors involved in inflammatory pathways, possibly through additional hydrogen bonding within active sites of enzymes due to the presence of the carbonyl and carboxyl groups. However, there is no research on the anti-inflammatory effects of 4-oxo-lanceolic acid specifically, so its potential anti-inflammatory effects need to be confirmed through further experimental studies. Although the anti-inflammatory potential of β -sesquiphellandrene is supported by the literature, and together with the previous isolation of the 4-oxo-lanceolic acid in the genus, these shreds of evidence support the annotation of this sesquiterpenoid as a discriminant bioactive metabolite, which herein demonstrated the potential to be active in both dual COX and LOX inflammatory pathways.

Prediction Models. Our study employed two ML methods, LogisticBase (Logistic regression) and Artificial Neural Networks (ANN) to examine and forecast the dual anti-inflammatory activity of *Ocotea* extracts based on their annotated significant discriminant features of the concatenated UPLC/MS-NMR dataset. These models revealed high accuracy in the external test sets (**Table 4**). This indicates a satisfactory performance with an equitable balance of internal accuracy and external evaluation, which helps to avoid building overfitted models. However, the internal validation (CV-10 folds) for the simple logistic regression assumed only a modest performance of 67%. Regarding the more sophisticated ANN method, satisfactory internal validation of 77%

was achieved. Metrics such as sensitivity and specificity that accurately detect true positives (active samples) and true negatives (inactive samples) were satisfactory for the models. Other additional validation metrics were also assessed and showed high robustness, especially for ANN (**Table 4**). Still even applying simple regression models, such as LogisticBase, the discriminant features of the OPLS-DA analysis seem to elicit satisfactory predictions. Even though the dataset is relatively small, it successfully predicts new active samples with adequate performance using simple logistic regression and a more sophisticated ANN algorithm.

Table 4. ML algorithms and performance metrics for the discriminant features of the concatenated dataset. Describes accuracy, Cohen’s kappa, sensitivity, specificity, precision, recall and F-measure.

ML algorithms		Accuracy (%)	Cohen's Kappa	Sensitive	Specificity	Precision	Recall	F-measure
LogisticBase	Training	100	1	1	1	1	1	1
	External Validation	80	0.6	0.7	1	1	0.7	0.8
	Internal Validation	67	0.3	0.6	0.7	0.8	0.6	0.7
	Validation							
ANN	Training	100	1	1	1	1	1	1
	External Validation	100	1	0.9	1	1	0.9	1
	Internal Validation	77	0.5	0.8	0.7	0.9	0.8	0.8
	Validation							

Furthermore, additional scrambling tests were also performed to ensure data reliability, and when comparing the validation metrics of the real models with the scrambled data, a significant decrease in model performance was observed. Notably, the ANN model experienced a considerable decrease in external validation performance when using the scrambled dataset, with accuracy falling from 100% to 50% and Cohen’s Kappa from 1 to -0.3, LogisticBase following the same negative Cohen’s Kappa pattern for external validation (**Table S6**). These discrepancies highlight that scrambled data leads to highly inferior and unreliable model performance compared to real data. Thus, the prediction models demonstrated overall high satisfactory validation metrics, especially for the ANN model (**Table 4**). Additionally, even applying simple regression models, like LogisticBase, the discriminant features of the OPLS-DA analysis seem to elicit adequate predictions. Thus, based on the quality of the built models without the use of additional hyperparameter tuning or boosting algorithms, demonstrates that this approach offers hopeful directions for computational drug discovery strategies based on discriminant features of untargeted metabolomics data.

In summary, the application of standard values in developing anti-inflammatory prediction models has yielded highly satisfactory metrics, demonstrating that validated results from multivariate analysis can be ground-breaking for identifying biomarkers in crude plant extracts, particularly in the context of modern metabolomics studies. This strategy of MSA and ML association reinforces the presence of discriminant metabolites in the evaluated *Ocotea* extracts, which the anti-inflammatory profile could be explained by differences in their chemical profile. It also demonstrates the reliability of bioinformatics strategies to predict new active *Ocotea* crude extract samples using computational methods, thus supporting the development of faster natural anti-inflammatory agents using modern metabolomics and ML strategies.

Final Remarks. Sixteen species from the *Ocotea* genus were investigated, and nine of them demonstrated the ability to inhibit both the COX and LOX pathways, suggesting the presence of bioactive markers as potential dual-action anti-inflammatory agents. The obtained results reproduced literature regarding the inhibition of PGE2 production for some of the previously evaluated *Ocotea* species and provided novel findings concerning the inhibition of LTB4 release to all of them. This work presents highly promising results for the advancement of novel anti-inflammatory agents with mechanisms of action different from drugs in the market and also fulfils a gap in the literature for effectively screening plant species for dual anti-inflammatory activity investigation in the same experiment using a small volume of blood for evaluation.

OPLS-DA successfully aided in identifying key discriminant metabolites, while STOCSY correlation analysis complemented the OPLS-DA findings by indicating correlated NMR chemical shifts for these metabolites. Four alkaloids, one glycosylated flavonoid and one sesquiterpenoid were revealed to be significant potential bioactive markers of the dual COX and LOX inhibition pathways. In addition, the anti-inflammatory effects of β -sesquiphellandrene provide a solid rationale for considering the potential anti-inflammatory properties of its derivatives, such as the 4-oxo-lanceolic acid. Further experimental validation through *in vitro*, *ex vivo* and *in vivo* studies would be necessary to confirm this first evidence. Of notice, only UPLC/MS data pointed out the oxo-lanceolic acid as a bioactive marker, likely due to their concentration in the *Ocotea* extracts, as UPLC/MS analysis suggested being a minor compound. Nonetheless, only MS due to its higher sensibility was able to detect this metabolite, demonstrating the importance of using both NMR and UPLC/MS techniques, if available, for enhanced metabolomics coverage and confidence in bioactive markers annotation.

The prediction models generated from discriminant metabolomics data using OPLS-DA demonstrated high accuracy, sensitivity, and precision, indicating robust performance across the two different models. The accuracy and performance of the ANN model (100% accuracy on the external set) were highly satisfactory, considering only the standard parameters. These approaches offer a strategic method for identifying bioactive compounds and compound classes in NP extracts, such as the evaluated *Ocotea* extracts, and reliably predicting the activity of novel crude extract samples. In the context of this research, with potential dual-anti-inflammatory action and applicability domain to the *Ocotea* species. This strategy would allow researchers to prioritize extracts that have been predicted as potentially active before conducting extensive experimental plant bioactivity assays. Thus, this approach enhances the efficiency of bioprospecting NP under untargeted metabolomics studies.

However, despite the advantages of concatenated UPLC/MS-NMR metabolomics, it faces challenges, such as the complexity of data integration and ensuring that sample preparation is suitable for both techniques. Combining and interpreting data from two different techniques also required sophisticated computational tools. However, the benefits of this approach include enhanced metabolite coverage, improved accuracy through cross-verification between UPLC/MS and NMR data, and therefore increased confidence in the biological interpretations and findings. Overall, concatenated metabolomics is a powerful approach for comprehensive and accurate metabolomic analysis^{25,37,38}, leveraging the strengths of both analytical platforms of UPL/MS and NMR to provide detailed insights into the chemistry of specialised metabolites in NP research.

The combined use of *ex-vivo* anti-inflammatory assays, concatenation metabolomics techniques, and chemometric strategies, complemented with gas-phase fragmentation reactions proposals and spectroscopic correlation analysis, facilitated the confidence level 2 of metabolite annotation of the key alkaloids, a glycosylated flavonoid and a sesquiterpenoid derivative, as potential dual anti-inflammatory discriminant markers in the crude extract of the investigated *Ocotea* species. This research underscores the efficacy of *Ocotea* species as a source of natural anti-inflammatory compounds with the potential to inhibit both COX and LOX pathways, thus contributing to the growing body of evidence supporting the medicinal value of the *Ocotea* genus. The application of concatenated UPLC/MS-NMR and robust multivariate statistical methods provides an effective approach for biomarker investigation in modern drug discovery. It showcases the successful combinations in NP research of advanced analytical techniques with

computational methods and bioinformatics tools as powerful strategies for biomarker discovery and prediction model constructions based on metabolomics data.

Experimental Section

Drugs and Reagents. Dexamethasone, indomethacin, limaprost, PGE2 authentic standard, and LTB4 authentic standard, *E. coli* O26:B6 lipopolysaccharides (LPS) and calcium ionophore (CI) were acquired from Sigma Aldrich® (St Louis, MO, USA). The deuterated dimethylsulfoxide (DMSO-*d*₆) was sourced from Sigma-Aldrich (Dorset, SWE, UK). All solvents, including hexane, methanol, ethanol, acetonitrile and formic acid, were of high-performance liquid chromatography (HPLC) grade and also supplied by Sigma Aldrich® (St Louis, MO, USA). Ultrapure water was produced using a Millipore Milli-Q® water purification system (Bedford, MA, USA). Liquid nitrogen was sourced from Lindet® (Pullach, MU, DE).

Plant Materials. To address the ecological concerns regarding threatened *Ocotea* species in Brazil,^{91–93} we collected 1-3 leaves from 16 different voucher specimens. These specimens have been previously deposited in the herbariums of the Federal University of Ouro Preto (OUPR, UFOP) and the Federal University of Juiz de Fora (Leopoldo Krieger - Centro de Ensino Superior de Juiz de Fora, CESJ, UFJF). Each *Ocotea* sample was assigned an identification code (ID) detailed along with the geographical location of the plant collection, together with the respective deposit voucher numbers (**Table S6**). This research was registered with the National System for Governance of Genetic Heritage and Associated Traditional Knowledge (SisGen) under the code #A5A8F67.

Metabolite Extraction and Sample Preparation. Each of the 16 *Ocotea* samples was weighed (50 mg) and crushed using a pestle and liquid nitrogen until pulverized. To extract the most polar and semi-polar compounds, 1.7 mL of ethanol:water (7:3, v/v) was added to the powdered material. The samples were placed in a warm ultrasound bath at 35 °C for 15 minutes (170 W, 50 kHz, L100 Schuster) and then centrifuged at 22 °C and 90 rcf (G-force). To remove fatty substances, the supernatants were partitioned with hexane (2 x 300 µL). The extracted samples were then filtered through polytetrafluoroethylene (PTFE, Millex®) filters (22 µm) and dried using a Speed Vacuum Eppendorf at 40 °C for 4 hours. The samples were stored in a freezer at -20 °C until analysis.⁹ The dried extracts were weighed, and the yield was calculated (**Table S7**). For the NMR analyses, each dried extract was prepared by dissolving it in 650 µL of DMSO-*d*₆ to achieve a concentration of 5 mg/mL. The solutions were then transferred into 5 mm 7" NMR tubes, ensuring they were ready for analysis. In parallel, for the UPLC-MS analysis, the same extracts were prepared by suspending them in methanol to achieve a concentration of 1.0 mg/mL. This mixture was then filtered through a PTFE filter with a pore size of 0.22 µm to prevent any particulate contamination during the LC-MS analysis process.

Ex-vivo Anti-inflammatory Assays. Experiments were conducted using human whole blood from volunteers (Research Ethics Committee of the Federal University of Alfenas 60344622.1.0000.5142). The detailed anti-inflammatory assay protocol is available in our latest publications^{1,9,94,95}. However, for this time the LTB4 release quantitative assessment was also included together with additional sample preparation improvements. Basically, pure substances such as dexamethasone (DEX) and indomethacin

(IND) were tested at 1 µg/mL as positive controls. The *Ocotea* extracts were assayed at 10 µg/mL. Phosphate-buffered saline 1x (PBS 1x, pH 7.2; 0.15 mol/L chloride; 0.01 mol/L phosphate) was used for dilution of the inducer solutions, including lipopolysaccharides (LPS) at 10 µg/mL and calcium ionophore (CI) at 20 µmol/L, to activate both PGE2 and LTB4 mediators release, respectively. PBS 1x solution was also used as a negative (NEG) control to estimate the average maximum amount of PGE2 and LTB4 produced during the inflammatory process. The plating sequence was as follows: 35 µL of each *Ocotea* extract, 280 µL of blood in all wells, and 35 µL of LPS + CI solution. The samples in sterile 96-well plates were incubated for 24 hours in a 5% CO₂ atmosphere at 37 °C. After incubation, samples were manipulated out of the light exposure. The plates were centrifuged for 5 minutes at 1207 rcf and 4 °C. Subsequently, 180 µL of plasma was removed from each well and frozen for later lyophilization. The lyophilized plasma was spiked with 280 µL of a precipitating solution (ACN:MeOH 1:1 v/v) containing 30 ng/mL limaprost as internal standard (IS). Samples were then centrifuged (8400 rcf, 4 °C) for 10 minutes. The supernatant (200 µL) was transferred to a polypropylene tube containing 1.8 mL of ultrapure water. The samples were then loaded into a Supelco® Solid-Phase Extraction cartridge (Supelclean® LC-18 100 mg, 1 mL tubes, #504270) after conditioning with 2 mL of MeOH followed by 2 mL of acidified ultrapure water (0.1% HAc). The cartridges were washed with 1 mL of ultrapure water (0.1% HAc), and the PGE2 and LTB4 analytes were eluted into polypropylene tubes using 1 mL of MeOH (0.1% HAc) and then evaporated to dryness. Samples were then resuspended with 50 µL of ACN before injection into the UPLC-MS/MS instrument system.

PGE2 and LTB4 Determination by UPLC-MS/MS. The UPLC-MS/MS analyses were performed using a Shimadzu® (Kyoto, Japan) UPLC-MS/MS system with a triple-quadrupole mass analyzer operating in negative mode. Multiple Reaction Monitoring (MRM) experiments were performed for each analyte (PGE2, LTB4 and Limaprost – internal standard), according to their specific *m/z* transitions (**Table S6**). The samples and controls were injected (30 µL) into a chromatographic system with a Poroshell 120 EC-C₁₈ (2.7 µm, 4.6 x 150 mm) column maintained at 30 °C. The mobile phase consisted of (A) ultrapure water acidified with 0.1% formic acid and (B) acetonitrile (100%) at a constant flow rate of 300 µL/min. Gradient elution was established as follows: 40% B until 100% B (0-5 min), maintained for 2 minutes, followed by reversion to the initial configuration for 1 minute, and 3 minutes of re-equilibration. The total run time was 11 minutes. The source and MS parameters included: nebulizing gas nitrogen at 2 L/min, drying gas nitrogen at 15 L/min, interface voltage 3.5 kV, DL temperature 250 °C, oven temperature 35 °C, detector voltage 2.16 kV, and collision gas argon at 230 kPa. Data acquisition was done using LabSolutions® software.

Statistical Analysis of Anti-inflammatory Data. The PGE2 and LTB4 signals obtained from the anti-inflammatory assay were normalized by Limaprost signal (IS) and statistically analysed via one-way analysis of variance (ANOVA) followed by Dunnett's test ($p < 0.05$) using GraphPad Prism® v. 8.0.1 (La Jolla, CA, USA). The analytical signals from the NEG samples were assumed as the control in Dunnett's multiple comparisons test. The anti-inflammatory activities were expressed as the relative release of PGE2 (rrPGE2) and LTB4 (rrLTB4), consisting of the ratio of the analytical signal of the treated samples (evaluated *Ocotea* extracts) related to NEG control samples. The uncertainties of these ratios were calculated by propagation of error and expressed as standard deviation (SD) (Supplemental material –

Scheme 1). The percentage inhibition (PI) of PGE2 and LTB4 consists in complementary percentage of rrPGE2 and rrLTB4 respectively, thus, $PI_{PGE2} = 1 - rrPGE2$, and $PI_{LTB4} = 1 - rrLTB4$.

UPLC-HRMS/MS Experiments and Data Processing. The UPLC-MS analyses were performed using a Waters Xevo G2® (Milford, USA) instrument (Quadrupole Time-of-Flight - QTOF mass analyzer) set for profile data-independent acquisition (DIA, MS^E mode). Analyses were conducted using Masslynx® MS Software (Waters Corp., Milford, USA). An aliquot of 5 µL of each *Ocotea* sample was injected, and separation occurred in a C18 (ACQUITY UPLC®HSS T3) reversed-phase column (1.8 µm, 100 x 2.1 mm) maintained at 40 °C. The mobile phases consisted of (A) 1% acidified water with formic acid, and (B) pure acetonitrile, delivered at a 0.5 mL/min flow rate. The chromatographic run was: 1% B and 99% A for 0.1 min, 85% A and 15% B for 7.5 min, 20% A and 80% B for 8.5 min, and 1% A and 99% B until 10 min. The electrospray (ESI) operated in both positive and negative ionization modes. The mass spectrometer parameters included alternating high and low-energy scans: low CE (collision energy) at 3 eV and high CE ramped from 10 to 40 eV, cone gas flow at 30 L/h, desolvation temperature at 300 °C, source temperature at 120 °C, and desolvation gas flow at 600 L/h. The mass scan range was set at m/z 50 to m/z 1000 for functions 1 and 2. MS data were collected in profile mode using the lock spray for calibration to ensure accuracy and reproducibility. Leucine-enkephalin was used as a lock mass, with m/z 554.2622 (ESI⁻) and m/z 556.2768 (ESI⁺) at a frequency of 10 s.

Before data processing the raw Waters MS^E data was converted using the Waters2mzML tool to the universal mzML data format. Waters2mzML is freely online available (URL:<https://github.com/AnP311/Waters2mzML/releases/tag/v1.2.0>). The *Ocotea* extracts were processed using MZmine 3.8.1 (<https://mzmine.github.io/>) (MZmine Development Team) with the following parameters: Mass detection of MS¹ scans used scan filters at MS¹ level = 1 and the centroid algorithm, with noise levels of 1000 (positive and negative). Feature detection utilized the ADAP Chromatogram Builder module with a scan filter at MS¹ level = 1, a minimum of 5 consecutive scans, minimum intensity for consecutive scans at five times the noise level, minimum absolute height also at five times the noise level, and an m/z tolerance of 0.005 m/z or 10 ppm. Chromatogram resolution used the Local Minimum Feature Resolver with a retention time dimension, chromatographic threshold of 5%, minimum search range RT/mobility of 0.15, minimum absolute height equal to ADAP, peak top/edge ratio of 3, peak duration range of 0.2-1.0 absolute, and a minimum of 5 data points. Isotope filtering employed the ¹³C isotope filter module with an m/z tolerance of 0.003 m/z or 5.0 ppm, retention time tolerance of 0.1 absolute, monotonic shape set to false, maximum charge of 2, and the most intense isotope as representative. Alignment was done using the Join aligner with an m/z tolerance of 5 ppm, RT relative tolerance of 5%, weight for m/z . 20, and weight for RT 20. Gap filling used the Peak Finder algorithm with an intensity tolerance of 0.2, m/z tolerance of 0.005 m/z or 10 ppm, RT tolerance of 0.1 absolute, and a minimum of 4 data points. Duplicate feature list rows were filtered using a new average filter mode with an m/z tolerance of 0.005 m/z or 10 ppm, and an RT tolerance of 0.07 absolute. All other unspecified parameters were set to default. Processed data containing peak area, m/z -Rt pair was exported as a .csv table format (**Table Z1** - <https://doi.org/10.5281/zenodo.13826734>).

NMR Experiments and Data Processing. All *Ocotea* sp. crude extracts (n= 18) were prepared at a concentration of 4.5 mg per 600 µL of DMSO-*d*₆ in 5 mm, 7-inch NMR tubes. The ¹H-NMR (500 MHz) spectra were recorded for 1D correlation spectroscopy NMR using a Bruker-Biospin Ascend 500

instrument, Avance III series®, (Rheinstetten, Germany), operating at 11.7 T. Each spectrum was accumulated with 64 scans, a delay time of 4 seconds, and an acquisition time of 3.41 minutes with 32 Kb. A pre-saturation pulse sequence was applied to achieve one-dimensional proton spectra with effective water suppression (3.3 ppm). All chemical shifts were expressed in ppm relative to the deuterium solvent or TMS, and data were acquired and processed using TOPSPIN v.3.2 (Bruker-Biospin, Rheinstetten, Germany). Before MSA, the data obtained were processed with MestReNova (Mnova v.14.2.3) software (Mestrelab Research, Santiago de Compostela, Spain) (<https://mestrelab.com/download/mnova/>). The ¹H-NMR spectra were processed, including Whittaker Smoother, automatically followed by manual phase correction for still non-phased spectra. Binning was applied at an integral region of 0.04 ppm using the Sum method. Subsequently, the spectral data were stacked and the peak intensities Excel® (Microsoft Windows) (.xlsx) as a text file and then converted to a .csv format file. Chemical shift (δ) values were established between 0.0 and 12.0 ppm (Table Z2 - <https://doi.org/10.5281/zenodo.13826734>).

UPLC/MS-NMR Concatenation. The output data from NMR and UPLC/MS data processing were then concatenated to expedite the annotation of specialized bioactive metabolite markers associated with the dual anti-inflammatory activity. The data was pre-combined within an Excel® spreadsheet and then saved in .csv format (Table Z3 - <https://doi.org/10.5281/zenodo.13826734>). To address the challenge of integrating disparate data types (UPLC-HRMS and ¹H-NMR), block-wise scaling was employed using SIMCA-P (Umetrics©) software. This method ensures that each data type (block) contributes equally to the analysis, equalizing the total variance of each data block, thus preventing the dominance of one data type over another due to differences in magnitude, scale or variance. The UPLC/MS data was defined as block 1 and ¹H-NMR data as block 2. Each block was scaled according to the variance within that block. Block scaling was implemented with the weights set according to Equation (1) as follows:

(1) $(1/\sqrt{Kblock})$, where K block is the number of variables in the block. This setting aims for the block to contribute a total variance of 1.

Multivariate Statistical Analysis (MSA) and Data Visualisation. To analyse the features of the fused dataset, of the UPLC/MS and NMR data, the datasets were analysed also using SIMCA-P tools for the MSA. All datasets underwent previous Log data transformation and Pareto scaling prior to analysis, in order to completely eradicate remaining data skewness. Initially, Principal Component Analysis (PCA) was employed to reduce the dimensionality of the data and identify underlying patterns. Following PCA, Pareto-scaled, Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA) was conducted to supervise data, maximize the separation between the two groups (active and inactive) and aid find potential discriminant bioactive markers. The metabolomics models were validated based on the coefficient of determination (R^2) and cross-validation (Q^2) metrics > 0.5 .⁹⁶ To corroborate the validation of the models, the Root Mean Square Error of Prediction (RMSE) and CV-ANOVA were assessed, and permutation tests ($n=100$) were implemented to ensure the model's robustness and prevent overfitting. Additionally, additional validation metrics were applied: the Benjamini-Hochberg procedure that was applied for the generation of q -values of the discriminant metabolites to control the false discovery rate (FDR), ensuring the statistical significance of the findings.^{9,25,97} Variables Important in Projection (VIP) scores were calculated to pinpoint the most influential features contributing to the model.^{25,92} Correlation coefficients (Coef) of the variable with the dual activity were computed to explore and determine the discriminant

features. The area under the curve-receiver operating characteristic (AUC-ROC) was also calculated to corroborate validation metrics, as the graphical plots illustrate the performance of a binary classifier model and prediction variables (**Figure S13**).

Metabolite Annotation. According to the Metabolomics Standards Initiative (MSI),^{98,99} confidence level 3 annotations were performed for the discriminant features using the custom database search module in MZmine, by cross-matching the precursor mass with the *Ocotea*DB,⁵⁴ which contains substances isolated specifically for the *Ocotea* genus. The set parameters included m/z tolerance of 0.005 m/z or 10 ppm, maximum charge of 1, maximum of 2 molecules per cluster, and adduct usage true, for positive mode ($[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[M+NH_4]^+$) and negative mode ($[M-H]^-$, $[M+Cl]^-$, $[M+Br]^-$, $[M+FA]^-$). The feature tables were exported containing information on aligned peak areas, exact mass, retention time, molecular formula and chemical annotation for all samples (**Table Z4** - <https://doi.org/10.5281/zenodo.13826734>). To assist in a broader identification of the compounds, we utilized the Dictionary of Natural Products (DNP). Additionally, to enhance the confidence of the annotations manual literature spectral comparisons were conducted using the Mass Bank of North America (MoNA). The high-energy MS^E spectra were checked using the MZmine 3 raw data inspection visualisation modules. For those bioactive markers that did not match any entries in the MoNA database, the MS^E product-ions were proposed based on a chemical knowledge of ESI fragmentation patterns in the literature.^{58,60} Chemical structures relevant to the study were precisely drawn using ChemDraw Ultra 12.0 (Perkin Elmer InformaticsTM, Cambridge, England) for inclusion in the manuscript figures.

STOCSY Correlation Analysis. Additional spectroscopic data analysis was conducted using the statistical total correlation spectroscopy (STOCSY) algorithm. This analysis aimed to correlate the important spectroscopic features generated from MSA results to identify significant correlated NMR shifts, thus also supporting the validation of annotated biomarkers.^{23,73} The processed NMR spectra segmented into equal-sized bins of 0.04 ppm were utilised for 1D STOCSY analyses. The input driver peaks were the discriminant signals of ¹H NMR with FDR < 0.05. Scatter plots were built to individually visualise the correlations. Scripts utilized were adapted from DAF-discovery²³ using the Google Colab online platform (<https://colab.research.google.com>) (script available in the <https://doi.org/10.5281/zenodo.13826734>), which several libraries were used including pandas for data structuring and manipulation, NumPy for numerical data handling, matplotlib for visual representation of the data, and Plotly for creating interactive .html graph plots.

Prediction Models and Validation. For the construction of the input metabolomic database, concatenated data obtained from UPLC/MS and NMR were utilised, where variable X represented the discriminant UPLC/MS data (m/z , RT and peak area) and the NMR chemical shifts derived from the OPLS-DA model, whereas the Y variable was obtained from the *ex vivo* anti-inflammatory assay (**Table Z5-6** - <https://doi.org/10.5281/zenodo.13826734>). Binary classification models were applied to categorize samples into active and inactive groups based on the ANOVA and post-test statistical analysis. Logistic regression and ML prediction models were developed utilizing the open-source data analytics platform KNIME (version 5.1, <https://www.knime.com>). The modeling process began with the "CSV Reader" node to import the dataset, which represented dual anti-inflammatory activity as a binary categorical response. The dataset, stored in .csv files, was subsequently normalized to a range of 0 to 1 using the "Normalization"

node. Following this, the "Partitioning" node was applied to partition the data through stratified sampling of the classes, allocating 75% of the data for training and 25% for testing. To ensure robust validation of the predictive model, the training dataset was partitioned using the "X-Partitioner" node in a 10-fold cross-validation loop. This node, employing the leave-one-out parameter, iteratively trained the model by excluding one data point in each iteration. The excluded point was used for testing, while the remaining data constituted the training set. This process was repeated until all data points had been used for testing once, thus enhancing the model's reliability across multiple data subsets. The "X-Aggregator" node was utilized to compile the validation data, and the "Scorer" node was used to present the statistical results (**Figure S14**).

To develop the metabolomic prediction models, two different algorithms were employed. Firstly, the LogisticBase 3.7 algorithm, a logistic regression-based method, was configured with standard parameters (-I 500 -WTB 0.0) (**Table S7**) to perform classification by modeling the probability of an instance belonging to a specific class. This algorithm enhances robustness by estimating parameters that maximize the likelihood of the observed data.¹⁰⁰ Subsequently, an artificial neural network (ANN), known as Multilayer Perceptron, was implemented and configured with the following optimised parameters: -L 0.3 -M 0.2 -N 500 -S 100 -H 46 (**Table S7**). The ANN is a supervised learning method that commonly applies backpropagation to train a multilayer perceptron before classifying and predicting instances from the dataset.^{101,102} Accuracy, precision, F-measure, mean absolute error (MAE), and root mean squared error (RMSE) were calculated to assess the overall performance and validation of the built models. Additionally, the scramble test was implemented, with the dual activity classes randomly shuffled and X features randomly selected, to provide additional validation for the developed metabolomic variable prediction models.

Associated Content

Additional supplementary data is online available at <https://doi.org/10.5281/zenodo.13826734>

Author information

Corresponding Authors

Daniela A. Chagas-Paula - ^a*Laboratory of Phytochemistry, Medicinal Chemistry and Metabolomics, Institute of Chemistry, Federal University of Alfenas, R. Gabriel Monteiro da Silva, 700, Alfenas/MG, 37130-001, Brazil. E-mail: daniela.chagas@unifal-mg.edu.br. orcid.org - 0000-0003-2274-4919*

Albert Katchborian-Neto - ^a*Laboratory of Phytochemistry, Medicinal Chemistry and Metabolomics, Institute of Chemistry, Federal University of Alfenas, R. Gabriel Monteiro da Silva, 700, Alfenas/MG, 37130-001, Brazil.* ^c*Center of Natural Sciences and Humanities, Federal University of ABC, Santo Andre, São Paulo, Brazil. E-mail: albert_katchborian@hotmail.com. orcid.org - 0000-0003-2711-2881*

Authors

Gabriel Viana - ^a*Institute of Chemistry, Federal University of Alfenas, Alfenas, Minas Gerais, Brazil.* orcid.org – 0000-0002-5450-9860

Miller Santos Ferreira - ^a*Institute of Chemistry, Federal University of Alfenas, Alfenas, Minas Gerais, Brazil.* orcid.org - 0000-0002-9723-0528

Matheus Fernandes Alves - ^a*Institute of Chemistry, Federal University of Alfenas, Alfenas, Minas Gerais, Brazil.* orcid.org - 0000-0003-2980-0370

Michael Murgu - *Waters Corporation, Alameda Tocantins 125, 27th floor, Alphaville, Barueri, Sao Paulo, Brazil.* orcid.org -0000-0002-7489-7637

Marina de Monroe Gonçalves - ^c*Center of Natural Sciences and Humanities, Federal University of ABC, Santo Andre, São Paulo, Brazil.* orcid.org – 0000-0003-0962-7224

Ana Cláudia C. Paula – ^d*Department of Pharmaceutical Sciences, Federal University of Juiz de Fora, Juiz de Fora, Minas Gerais, Brazil.* orcid.org – 0000-0001-7998-0950

Daniele Ferreira Dias - ^a*Institute of Chemistry, Federal University of Alfenas, Alfenas, Minas Gerais, Brazil.* orcid.org - 0000-0001-9129-4734

Marisi Gomes Soares - ^a*Institute of Chemistry, Federal University of Alfenas, Alfenas, Minas Gerais, Brazil.* orcid.org - 0000-0001-9221-9867

João Henrique G. Lago - ^c*Center of Natural Sciences and Humanities, Federal University of ABC, Santo Andre, São Paulo, Brazil.* orcid.org – 0000-0002-1193-8374

RuAngelie Edrada-Ebel - ^e*Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, G4 0RE, Glasgow, Scotland.* orcid.org - 0000-0003-2420-1117

Disclosure statement

The authors declare no conflict of interest.

Acknowledgements

The Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil (CNPq) (grant number 406837/2021-0, 316204/2021-8, 408115/2023-8), and Fundação de Amparo à Pesquisa do Estado de Minas Gerais - Brazil (FAPEMIG) (grant numbers APQ-02353-17, APQ-00207-18, APQ-05218-23, APQ-00544-23, BPD-00760-22). The authors acknowledge the UFJF and UFOP herbarium for providing the *Ocotea* sp. material. The São Paulo Research Foundation (FAPESP), Brazil, financed this study, in part. Process Number (#:24/04606-5).

References

- (1) Santos, M. F. C.; Nicácio, K. de J.; Katchborian-Neto, A.; Ferreira, M. S.; Miranda, D. de O.; Andrade, J. V.; Pereira, H. de A.; Jesus, E. G. de; B, T.; Souza, S.; Morais-Urano, R. P.; Dias, D. F.; Chagas-Paula, D. A.; Soares, M. G. Ex Vivo Inhibition of PGE2 Formation in Human Blood by Four Bicyclic [3.2.1] Octane Neolignans Isolated from Aniba Firmula Bark, Two with Unusual Structural Pattern. *Nat. Prod. Res.* **2024**, *38* (3), 393–401. <https://doi.org/10.1080/14786419.2022.2124248>.
- (2) Simić, A.; Soković, M. D.; Ristić, M.; Grujić-Jovanović, S.; Vukojević, J.; Marin, P. D. The Chemical Composition of Some Lauraceae Essential Oils and Their Antifungal Activities. *Phyther. Res.* **2004**, *18* (9), 713–717. <https://doi.org/10.1002/ptr.1516>.
- (3) Wang, J.; Su, B.; Jiang, H.; Cui, N.; Yu, Z.; Yang, Y.; Sun, Y. Traditional Uses, Phytochemistry and Pharmacological Activities of the Genus *Cinnamomum* (Lauraceae): A Review. *Fitoterapia* **2020**, *146*, 104675. <https://doi.org/10.1016/j.fitote.2020.104675>.
- (4) Gottlieb, O. R. Chemosystematics of the Lauraceae. *Phytochemistry* **1972**, *11* (5), 1537–1570. [https://doi.org/10.1016/0031-9422\(72\)85001-5](https://doi.org/10.1016/0031-9422(72)85001-5).
- (5) Silva, A. F.; Santos, M. F. C.; Maiolini, T. S. C.; Salem, P. P. O.; Murgu, M.; Paula, A. C. C.; Silva, E. O.; Nicácio, K. J.; Ferreira, A. G.; Dias, D. F.; Soares, M. G.; Chagas-Paula, D. A. Chemistry of Leaves, Bark, and Essential Oils from *Ocotea Diospyrifolia* and Anti-Inflammatory Activity – Dual Inhibition of Edema and Neutrophil Recruitment. *Phytochem. Lett.* **2021**, *42* (July 2020), 52–60. <https://doi.org/10.1016/j.phytol.2021.02.002>.
- (6) de Alcântara, B. G. V.; Oliveira, F. P. de; Katchborian-Neto, A.; Casoti, R.; Domingos, O. da S.; Santos, M. F. C.; Oliveira, R. B. de; Paula, A. C. C. de; Dias, D. F.; Soares, M. G.; Chagas-Paula, D. A. Confirmation of Ethnopharmacological Anti-Inflammatory Properties of *Ocotea Odorifera* and Determination of Its Main Active Compounds. *J. Ethnopharmacol.* **2021**, *264* (August 2020), 133378. <https://doi.org/10.1016/j.jep.2020.113378>.
- (7) Salleh, W. M. N. H. W.; Ahmad, F. Phytochemistry and Biological Activities of the Genus *Ocotea* (Lauraceae): A Review on Recent Research Results (2000–2016). *J. Appl. Pharm. Sci.* **2017**, *7* (5), 204–218. <https://doi.org/10.7324/JAPS.2017.70534>.
- (8) Alcântara, B. G. V. de; Neto, A. K.; Garcia, D. A.; Casoti, R.; Oliveira, T. B.; Ladvoat, A. C. C. de P.; Edrada-Ebel, R.; Soares, M. G.; Dias, D. F.; Chagas-Paula, D. Anti-Inflammatory Activity of Lauraceae Plant. Chemistry and Biodiversity.Pdf. *Chem. Biodivers.* **2023**, *20* (9), 1–10. <https://doi.org/doi.org/10.1002/cbdv.202300650>.
- (9) Katchborian-neto, A.; Jesus, K. De; Carolina, Cruz, J. C.; Paula; Bueno, P.; Murgu, M.; Dias, D. F.; Soares, M. G.; Paula, A. C. C.; Chagas-paula, D. A. Bioprospecting-Based Untargeted Metabolomics Identifies Alkaloids as Potential Anti-Inflammatory Bioactive Markers of *Ocotea* Species (Lauraceae). *Phytomedicine* **2023**, *120* (June), 155060.

<https://doi.org/10.1016/j.phymed.2023.155060>.

- (10) Mahesh, G.; Kumar, K. A.; Reddanna, P. Overview on the Discovery and Development of Anti-Inflammatory Drugs: Should the Focus Be on Synthesis or Degradation of Pge2? *J. Inflamm. Res.* **2021**, *14*, 253–263. <https://doi.org/10.2147/JIR.S278514>.
- (11) Hwang, S. H.; Wecksler, A. T.; Wagner, K.; Hammock, B. D. Rationally Designed Multitarget Agents Against Inflammation and Pain. *Curr. Med. Chem.* **2013**, *20* (13), 1783–1799. <https://doi.org/10.2174/0929867311320130013>.
- (12) Domingos, O. D. S.; Alcântara, B. G. V.; Santos, M. F. C.; Maiolini, T. C. S.; Dias, D. F.; Baldim, J. L.; Lago, J. H. G.; Soares, M. G.; Chagas-Paula, D. A. Anti-Inflammatory Derivatives with Dual Mechanism of Action from the Metabolomic Screening of *Poincianella Pluviosa*. *Molecules* **2019**, *24* (23), 1–16. <https://doi.org/10.3390/molecules24234375>.
- (13) Chagas-Paula, D.; Zhang, T.; Da Costa, F.; Edrada-Ebel, R. A Metabolomic Approach to Target Compounds from the Asteraceae Family for Dual COX and LOX Inhibition. *Metabolites* **2015**, *5* (3), 404–430. <https://doi.org/10.3390/metabo5030404>.
- (14) Chagas-Paula, D. A.; Oliveira, T. B.; Zhang, T.; Edrada-Ebel, R.; Da Costa, F. B. Prediction of Anti-Inflammatory Plants and Discovery of Their Biomarkers by Machine Learning Algorithms and Metabolomic Studies. *Planta Med.* **2015**, *81* (6), 450–458. <https://doi.org/10.1055/s-0034-1396206>.
- (15) Ebbels, T. M. D.; van der Hoof, J. J. J.; Chatelaine, H.; Broeckling, C.; Zamboni, N.; Hassoun, S.; Mathé, E. A. Recent Advances in Mass Spectrometry-Based Computational Metabolomics. *Curr. Opin. Chem. Biol.* **2023**, *74*, 102288. <https://doi.org/10.1016/j.cbpa.2023.102288>.
- (16) Yuliana, N. D.; Khatib, A.; Choi, Y. H.; Verpoorte, R. Metabolomics for Bioactivity Assessment of Natural Products. *Phyther. Res.* **2011**, *25* (2), 157–169. <https://doi.org/10.1002/ptr.3258>.
- (17) Caesar, L. K.; Kellogg, J. J.; Kvalheim, O. M.; Cech, N. B. Opportunities and Limitations for Untargeted Mass Spectrometry Metabolomics to Identify Biologically Active Constituents in Complex Natural Product Mixtures. *J. Nat. Prod.* **2019**, *82* (3), 469–484. <https://doi.org/10.1021/acs.jnatprod.9b00176>.
- (18) Alarcon-Barrera, J. C.; Kostidis, S.; Ondo-Mendez, A.; Giera, M. Recent Advances in Metabolomics Analysis for Early Drug Development. *Drug Discov. Today* **2022**, *27* (6), 1763–1773. <https://doi.org/10.1016/j.drudis.2022.02.018>.
- (19) Mannocho-Russo, H.; Nunes, W. D. G.; Almeida, R. F.; Albernaz, L. C.; Espindola, L. S.; Bolzani, V. S. Old Meets New: Mass Spectrometry-Based Untargeted Metabolomics Reveals Unusual Larvicidal Nitropropanoyl Glycosides from the Leaves of *Heteropterys Umbellata*. *J. Nat. Prod.* **2023**, *86* (3), 621–632. <https://doi.org/10.1021/acs.jnatprod.2c00788>.
- (20) Demarque, D. P.; Dusi, R. G.; de Sousa, F. D. M.; Grossi, S. M.; Silvério, M. R. S.; Lopes, N. P.;

- Espindola, L. S. Mass Spectrometry-Based Metabolomics Approach in the Isolation of Bioactive Natural Products. *Sci. Rep.* **2020**, *10* (1), 1–9. <https://doi.org/10.1038/s41598-020-58046-y>.
- (21) Pilon, A. C.; Selegato, D. M.; Fernandes, R. P.; Bueno, P. C. P.; Pinho, D. R.; Neto, F. C.; Freire, R. T.; Castro-Gamboa, I.; Bolzani, V. S.; Lopes, N. P. Plant Metabolomics: Methods and Challenges. *Quim. Nova* **2020**, *43* (3), 329–354. <https://doi.org/10.21577/0100-4042.20170499>.
- (22) Wang, X. J.; Ren, J. L.; Zhang, A. H.; Sun, H.; Yan, G. L.; Han, Y.; Liu, L. Novel Applications of Mass Spectrometry-Based Metabolomics in Herbal Medicines and Its Active Ingredients: Current Evidence. *Mass Spectrom. Rev.* **2019**, *38* (4–5), 380–402. <https://doi.org/10.1002/mas.21589>.
- (23) Borges, R. M. F. das N. C.; Fernanda A. Magno Teixeira; Yoon, J.; Márcio Barczyszyn Weiss2 | Camila Manoel Crnkovic2 | Alan Cesar Pilon3 | Bruno C. Garrido4 | Luz Adriana Betancur5 | Abel M. Forero6, 7 | Leonardo Castellanos6 | Freddy A. Ramos6 | Mônica T. Pupo3 | Stefan Kuhn. Data Fusion-Based Discovery (DAFdiscovery) Pipeline to Aid Compound Annotation and Bioactive Compound Discovery across Diverse Spectral Data. *Phytochem. anal* **2022**, *34*, 48–55.
- (24) Farag, M. A.; Porzel, A.; Schmidt, J.; Wessjohann, L. A. Metabolite Profiling and Fingerprinting of Commercial Cultivars of *Humulus Lupulus* L. (Hop): A Comparison of MS and NMR Methods in Metabolomics. *Metabolomics* **2012**, *8* (3), 492–507. <https://doi.org/10.1007/s11306-011-0335-y>.
- (25) Zanatta, A. C.; Vilegas, W.; Edrada-Ebel, R. A. UHPLC-(ESI)-HRMS and NMR-Based Metabolomics Approach to Access the Seasonality of *Byrsonima Intermedia* and *Serjania Marginata* From Brazilian Cerrado Flora Diversity. *Front. Chem.* **2021**, *9* (July), 1–17. <https://doi.org/10.3389/fchem.2021.710025>.
- (26) Tawfik, N. F.; Tawfike, A. F.; Abdou, R.; Abbott, G.; Abdelmohsen, U. R.; Edrada-Ebel, R.; Haggag, E. G. Metabolomics and Bioactivity Guided Isolation of Secondary Metabolites from the Endophytic Fungus *Chaetomium* Sp. **2017**, *1* (1), 2357–2547.
- (27) Chagas-Paula, D. A.; Oliveira, T. B.; Zhang, T.; Edrada-Ebel, R.; Da Costa, F. B. Prediction of Anti-Inflammatory Plants and Discovery of Their Biomarkers by Machine Learning Algorithms and Metabolomic Studies. *Planta Med.* **2015**, *81* (6), 450–458. <https://doi.org/10.1055/s-0034-1396206>.
- (28) Lee, S.; Lee, D. K. What Is the Proper Way to Apply the Multiple Comparison Test? *Korean J. Anesthesiol.* **2018**, *71* (5), 353–360. <https://doi.org/10.4097/kja.d.18.00242>.
- (29) Sawyer, S. F. Analysis of Variance: The Fundamental Concepts. *J. Man. Manip. Ther.* **2009**, *17* (2), 27E–38E. <https://doi.org/10.1179/jmt.2009.17.2.27e>.
- (30) Majumder, M.; Xin, X.; Liu, L.; Girish, G. V.; Lala, P. K. Prostaglandin E2 Receptor EP4 as the Common Target on Cancer Cells and Macrophages to Abolish Angiogenesis, Lymphangiogenesis, Metastasis, and Stem-like Cell Functions. *Cancer Sci.* **2014**, *105* (9), 1142–1151. <https://doi.org/10.1111/cas.12475>.

- (31) Salehifar, E.; Hosseinimehr, S. J. The Use of Cyclooxygenase-2 Inhibitors for Improvement of Efficacy of Radiotherapy in Cancers. *Drug Discov. Today* **2016**, *21* (4), 654–662. <https://doi.org/10.1016/j.drudis.2016.02.019>.
- (32) Fiorucci, S.; Meli, R.; Bucci, M.; Cirino, G. Dual Inhibitors of Cyclooxygenase and 5-Lipoxygenase. A New Avenue in Anti-Inflammatory Therapy? *Biochem. Pharmacol.* **2001**, *62* (1), 1433–1438.
- (33) Santos, M. F. C.; Alcântara, B. G. V.; dos R. Feliciano, C.; Silva, A. F.; Maiolini, T. C. S.; Neto, A. K.; Murgu, M.; de Paula, D. A. C.; Soares, M. G. New Bicyclic [3.2.1] Octane Neolignans Derivatives from Aniba Firmula with Potent in Vivo Anti-Inflammatory Activity on Account of Dual Inhibition of PGE₂ Production and Cell Recruitment. *Phytochem. Lett.* **2019**, *30* (3), 31–37. <https://doi.org/10.1016/j.phytol.2019.01.014>.
- (34) Roux, A.; Lison, D.; Junot, C.; Heilier, J. F. Applications of Liquid Chromatography Coupled to Mass Spectrometry-Based Metabolomics in Clinical Chemistry and Toxicology: A Review. *Clin. Biochem.* **2011**, *44* (1), 119–135. <https://doi.org/10.1016/j.clinbiochem.2010.08.016>.
- (35) Kirwan, J. A.; Gika, H.; Beger, R. D.; Bearden, D.; Dunn, W. B.; Goodacre, R.; Theodoridis, G.; Witting, M.; Yu, L. R.; Wilson, I. D. Quality Assurance and Quality Control Reporting in Untargeted Metabolic Phenotyping: MQACC Recommendations for Analytical Quality Management. *Metabolomics* **2022**, *18* (9), 1–16. <https://doi.org/10.1007/s11306-022-01926-3>.
- (36) Zanatta, A. C.; Vieira, N. C.; Dantas-Medeiros, R.; Vilegas, W.; Edrada-Ebel, R. A. Understanding the Seasonal Effect of Metabolite Production in Terminalia Catappa L. Leaves through a Concatenated MS- and NMR-Based Metabolomics Approach. *Metabolites* **2023**, *13* (3). <https://doi.org/10.3390/metabo13030349>.
- (37) 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. *Sci. Rep.* **2016**, *6* (1), 29265. <https://doi.org/10.1038/srep29265>.
- (38) Yusoff, Y. M.; Abbott, G.; Young, L.; Edrada-Ebel, R. Metabolomic Profiling of Malaysian and New Zealand Honey Using Concatenated NMR and HRMS Datasets. *Metabolites* **2022**, *12* (1). <https://doi.org/10.3390/metabo12010085>.
- (39) Worley, B.; Powers, R. Multivariate Analysis in Metabolomics. *Curr. Metabolomics* **2012**, *1* (1), 92–107. <https://doi.org/10.2174/2213235x130108>.
- (40) Van Der Kooy, F.; Maltese, F.; Young, H. C.; Hye, K. K.; Verpoorte, R. Quality Control of Herbal Material and Phytopharmaceuticals with MS and NMR Based Metabolic Fingerprinting. *Planta Med.* **2009**, *75* (7), 763–775. <https://doi.org/10.1055/s-0029-1185450>.
- (41) Yi, L.; Dong, N.; Yun, Y.; Deng, B.; Ren, D.; Liu, S.; Liang, Y. Chemometric Methods in Data Processing of Mass Spectrometry-Based Metabolomics: A Review. *Anal. Chim. Acta* **2016**, *914*, 17–34. <https://doi.org/10.1016/j.aca.2016.02.001>.

- (42) Rudaz, S. Identification and Data Processing Methods in Metabolomics. *Future science*. 2015, pp 1–184. <https://doi.org/10.4155/9781910420287>.
- (43) Ren, S.; Hinzman, A. A.; Kang, E. L.; Szczesniak, R. D.; Lu, L. J. Computational and Statistical Analysis of Metabolomics Data. *Metabolomics* **2015**, *11* (6), 1492–1513. <https://doi.org/10.1007/s11306-015-0823-6>.
- (44) Boccard, J.; González-Ruiz, V.; Codesido, S.; Rudaz, S. Mass Spectrometry Metabolomic Data Handling for Biomarker Discovery. *Proteomic Metabolomic Approaches to Biomark. Discov.* **2019**, 369–388. <https://doi.org/10.1016/B978-0-12-818607-7.00021-9>.
- (45) Trygg, J.; Holmes, E.; Lundstedt, T. Chemometrics in Metabonomics. *J. Proteome Res.* **2007**, *6* (2), 469–479. <https://doi.org/10.1021/pr060594q>.
- (46) Bylesjö, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J. K.; Holmes, E.; Trygg, J. OPLS Discriminant Analysis: Combining the Strengths of PLS-DA and SIMCA Classification. *J. Chemom.* **2006**, *20* (8–10), 341–351. <https://doi.org/10.1002/cem.1006>.
- (47) Pratim Roy, P.; Paul, S.; Mitra, I.; Roy, K. On Two Novel Parameters for Validation of Predictive QSAR Models. *Molecules* **2009**, *14* (5), 1660–1701. <https://doi.org/10.3390/molecules14051660>.
- (48) Peluso, A.; Glen, R.; Ebbels, T. M. D. Multiple-Testing Correction in Metabolome-Wide Association Studies. *BMC Bioinformatics* **2021**, *22* (1), 1–18. <https://doi.org/10.1186/s12859-021-03975-2>.
- (49) Eriksson, L.; Trygg, J.; Wold, S. CV-ANOVA for Significance Testing of PLS and OPLS® Models. *J. Chemom.* **2008**, *22* (11–12), 594–600. <https://doi.org/10.1002/cem.1187>.
- (50) Kemsley, E. K.; Tapp, H. S. OPLS Filtered Data Can Be Obtained Directly from Non-Orthogonalized PLS1. *J. Chemom.* **2009**, *23* (5), 263–264. <https://doi.org/10.1002/cem.1217>.
- (51) Creek, D. J.; Dunn, W. B.; Fiehn, O.; Griffin, J. L.; Hall, R. D.; Lei, Z.; Mistrik, R.; Neumann, S.; Schymanski, E. L.; Sumner, L. W.; Trengove, R.; Wolfender, J. L. Metabolite Identification: Are You Sure? And How Do Your Peers Gauge Your Confidence? *Metabolomics* **2014**, *10* (3), 350–353. <https://doi.org/10.1007/s11306-014-0656-8>.
- (52) Spicer, R. A.; Salek, R.; Steinbeck, C. Comment: A Decade after the Metabolomics Standards Initiative It's Time for a Revision. *Sci. Data* **2017**, *4*, 2–4. <https://doi.org/10.1038/sdata.2017.138>.
- (53) 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: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **2007**, *3* (3), 211–221. <https://doi.org/10.1007/s11306-007-0082-2>.

- (54) Katchborian-Neto, A.; Alves, M. F.; Bueno, P. C. P.; De Jesus Nicácio, K.; Ferreira, M. S.; Oliveira, T. B.; Barbosa, H.; Murgu, M.; De Paula Ladvoat, A. C. C.; Dias, D. F.; Soares, M. G.; Lago, J. H. G.; Chagas-Paula, D. A. Integrative Open Workflow for Confident Annotation and Molecular Networking of Metabolomics MSE/DIA Data. *Brief. Bioinform.* **2024**, *25* (2), 1–18. <https://doi.org/10.1093/bib/bbae013>.
- (55) Alves, M. F.; Katchborian-Neto, A.; Bueno, P. C. P.; Carnevale-Neto, F.; Casoti, R.; Ferreira, M. S.; Murgu, M.; de Paula, A. C. C.; Dias, D. F.; Soares, M. G.; Chagas-Paula, D. A. LC-MS/DIA-Based Strategy for Comprehensive Flavonoid Profiling: An *Ocotea* Spp. Applicability Case. *RSC Adv.* **2024**, *14* (15), 10481–10498. <https://doi.org/10.1039/d4ra01384k>.
- (56) Theodoridis, G.; Gika, H. G.; Wilson, I. D. Mass Spectrometry-based Holistic Analytical Approaches for Metabolite Profiling in Systems Biology Studies. *Mass Spectrom. Rev.* **2011**, *30* (5), 884–906. <https://doi.org/10.1002/mas.20306>.
- (57) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **2014**, *48* (4), 2097–2098. <https://doi.org/10.1021/es5002105>.
- (58) Demarque, D. P.; Crotti, A. E. M.; Vessecchi, R.; Lopes, J. L. C.; Lopes, N. P. Fragmentation Reactions Using Electrospray Ionization Mass Spectrometry: An Important Tool for the Structural Elucidation and Characterization of Synthetic and Natural Products. *Nat. Prod. Rep.* **2016**, *33* (3), 432–455. <https://doi.org/10.1039/c5np00073d>.
- (59) Carnevale Neto, F.; Andréo, M. A.; Raftery, D.; Lopes, J. L. C.; Lopes, N. P.; Castro-Gamboa, I.; Lameiro de Noronha Sales Maia, B. H.; Costa, E. V.; Vessecchi, R. Characterization of Aporphine Alkaloids by Electrospray Ionization Tandem Mass Spectrometry and Density Functional Theory Calculations. *Rapid Commun. Mass Spectrom.* **2020**, *34* (S3). <https://doi.org/10.1002/rcm.8533>.
- (60) Qing, Z.; Xu, Y.; Yu, L.; Liu, J.; Huang, X.; Tang, Z.; Cheng, P.; Zeng, J. Investigation of Fragmentation Behaviours of Isoquinoline Alkaloids by Mass Spectrometry Combined with Computational Chemistry. *Sci. Rep.* **2020**, *10* (1), 1–13. <https://doi.org/10.1038/s41598-019-57406-7>.
- (61) de Lima, B. R.; da Silva, F. M. A.; Soares, E. R.; de Almeida, R. A.; da Silva-Filho, F. A.; Barison, A.; Costa, E. V.; Koolen, H. H. F.; de Souza, A. D. L.; Pinheiro, M. L. B. Integrative Approach Based on Leaf Spray Mass Spectrometry, HPLC-DAD-MS/MS, and NMR for Comprehensive Characterization of Isoquinoline-derived Alkaloids in Leaves of *Onychopetalum Amazonicum* R. E. Fr. *J. Braz. Chem. Soc.* **2020**, *31* (1), 79–89. <https://doi.org/10.21577/0103-5053.20190125>.
- (62) Conceição, R. S.; Reis, I. M. A.; Cerqueira, A. P. M.; Perez, C. J.; Junior, M. C. do. S.; Branco, A.; Ifa, D. R.; Botura, M. B. Rapid Structural Characterisation of Benzylisoquinoline and Aporphine Alkaloids from *Ocotea Spixiana* Acaricide Extract by HPTLC-DESI-MSn. *Phytochem. Anal.* **2020**, *31* (6), 711–721. <https://doi.org/10.1002/pca.2935>.

- (63) Zhang, S.; Zhang, Q.; Guo, Q.; Zhao, Y.; Gao, X.; Chai, X.; Tu, P. Characterization and Simultaneous Quantification of Biological Aporphine Alkaloids in *Litsea Cubeba* by HPLC with Hybrid Ion Trap Time-of-Flight Mass Spectrometry and HPLC with Diode Array Detection. *J. Sep. Sci.* **2015**, *38* (15), 2614–2624. <https://doi.org/10.1002/jssc.201500286>.
- (64) Pilon, A. C.; Gu, H.; Raftery, D.; Bolzani, V. da S.; Lopes, N. P.; Castro-Gamboa, I.; Neto, F. C. Mass Spectral Similarity Networking and Gas-Phase Fragmentation Reactions in the Structural Analysis of Flavonoid Glycoconjugates. *Anal. Chem.* **2019**, *91* (16), 10413–10423. <https://doi.org/10.1021/acs.analchem.8b05479>.
- (65) Stenger Moura, F. C.; dos Santos Machado, C. L.; Reisdorfer Paula, F.; Garcia Couto, A.; Ricci, M.; Cechinel-Filho, V.; Bonomini, T. J.; Sandjo, L. P.; Bellé Bresolin, T. M. Taxifolin Stability: In Silico Prediction and in Vitro Degradation with HPLC-UV/UPLC–ESI-MS Monitoring. *J. Pharm. Anal.* **2021**, *11* (2), 232–240. <https://doi.org/10.1016/j.jpha.2020.06.008>.
- (66) Nogueira, C. R.; Carbonezi, L. H.; de Oliveira, C. T. F.; Garcez, W. S.; Garcez, F. R. Sesquiterpene Derivatives from *Ocotea Minarum* Leaves. *Phytochem. Lett.* **2021**, *42* (January), 8–14. <https://doi.org/10.1016/j.phytol.2021.01.004>.
- (67) Costa, E. V.; Cruz, P. E. O. da; Pinheiro, M. L. B.; Marques, F. A.; Ruiz, A. L. T. G.; Marchetti, G. M.; Carvalho, J. E. de; Barisonn, A.; Maia, B. H. L. N. S. Aporphine Alkaloids from the Leaves of *Guatteria Friesiana* (Annonaceae) and Their Cytotoxic Activities. *J. Braz. Chem. Soc. Braz. Chem. Soc.* **2013**, *24* (5), 788–796. <https://doi.org/10.3390/molecules18088994>.
- (68) Vilegas, J. H. Y.; Gottlieb, O. Aporphine Alkaloids from *Ocotea Caesia*. *Phytochemistry* **1989**, *28* (12), 3577–3578.
- (69) Agrawal, P. K. NMR Spectroscopy in the Structural Elucidation of Oligosaccharides and Glycosides. *Phytochemistry* **1992**, *31* (10), 3307–3330. [https://doi.org/10.1016/0031-9422\(92\)83678-R](https://doi.org/10.1016/0031-9422(92)83678-R).
- (70) Silverstein, R. Spectrometric Identification of Organic Compounds 6th Edition (R. M. Silverstein & F. X. Webster).Pdf. *Journal of Molecular Structure.* 1976, pp 424–425. <http://linkinghub.elsevier.com/retrieve/pii/002228607687024X>.
- (71) Batista, A. N. L.; Junior, J. M. B.; López, S. N.; Furlan, M.; Cavalheiro, A. J.; Silva, D. H. S.; Bolzani, V. D. S.; Nunomura, S. M.; Yoshida, M. Aromatic Compounds from Three Brazilian Lauraceae Species. *Quim. Nova* **2010**, *33* (2), 321–323. <https://doi.org/10.1590/S0100-40422010000200017>.
- (72) Antonio, S.; Ramalho, G.; Pereira, M. G.; Veiga-junior, V. F. Chemophenetic Study of *Ocotea Canaliculata* (Lauraceae) by UHPLC – HRMS and GNPS. *Nat. Prod. Res.* **2020**, *0* (0), 1–5. <https://doi.org/10.1080/14786419.2020.1837823>.
- (73) Garcia-Perez, I.; Poma, J. M.; Serrano-Contreras, J. I.; Boulangé, C. L.; Chan, Q.; Frost, G.; Stamler, J.; Elliott, P.; Lindon, J. C.; Holmes, E.; Nicholson, J. K. Identifying Unknown Metabolites

- Using NMR-Based Metabolic Profiling Techniques. *Nat. Protoc.* **2020**, *15* (8), 2538–2567. <https://doi.org/10.1038/s41596-020-0343-3>.
- (74) Lemus Ringele, G. B.; Beteinakis, S.; Papachristodoulou, A.; Axiotis, E.; Mikros, E.; Halabalaki, M. NMR Metabolite Profiling in the Quality and Authentication Assessment of Greek Honey—Exploitation of STOCSY for Markers Identification. *Foods* **2022**, *11* (18). <https://doi.org/10.3390/foods11182853>.
- (75) Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. Statistical Total Correlation Spectroscopy: An Exploratory Approach for Latent Biomarker Identification from Metabolic 1H NMR Data Sets. *Anal. Chem.* **2005**, *77* (5), 1282–1289. <https://doi.org/10.1021/ac048630x>.
- (76) Garcez, W. S.; Yoshida, M.; Gottlieb, O. R. Benzylisoquinoline Alkaloids and Flavonols From *Ocotea Velloziana*. *Phytochemistry* **1995**, *39* (4), 815–816. [https://doi.org/https://doi.org/10.1016/0031-9422\(94\)00961-R](https://doi.org/https://doi.org/10.1016/0031-9422(94)00961-R).
- (77) Amilia Destryana, R.; Gary Young, D.; Woolley, C. L.; Huang, T. C.; Wu, H. Y.; Shih, W. L. Antioxidant and Anti-Inflammation Activities of *Ocotea*, *Copaiba* and *Blue Cypress* Essential Oils in Vitro and in Vivo. *JAOCS, J. Am. Oil Chem. Soc.* **2014**, *91* (9), 1531–1542. <https://doi.org/10.1007/s11746-014-2504-4>.
- (78) Coy, E. D.; Cuca, L. E.; Sefkow, M. COX, LOX and Platelet Aggregation Inhibitory Properties of *Lauraceae Neolignans*. *Bioorganic Med. Chem. Lett.* **2009**, *19* (24), 6922–6925. <https://doi.org/10.1016/j.bmcl.2009.10.069>.
- (79) Ferreira, B. A.; de Moura, F. B. R.; Gomes, K. S.; da Silva Souza, D. C.; Lago, J. H. G.; Araújo, F. de A. Biseugenol from *Ocotea Cymbarum* (Lauraceae) Attenuates Inflammation, Angiogenesis and Collagen Deposition of Sponge-Induced Fibrovascular Tissue in Mice. *Inflammopharmacology* **2023**, No. Libby 2008. <https://doi.org/10.1007/s10787-023-01210-3>.
- (80) da Silva, R. G.; Almeida, T. C.; Reis, A. C. C.; Filho, S. A. V.; Brandão, G. C.; da Silva, G. N.; de Sousa, H. C.; de Almeida, V. L.; Lopes, J. C. D.; de Souza, G. H. B. In Silico Pharmacological Prediction and Cytotoxicity of Flavonoids Glycosides Identified by UPLC-DAD-ESI-MS/MS in Extracts of *Humulus Lupulus* Leaves Cultivated in Brazil. *Nat. Prod. Res.* **2020**, *0* (0), 1–6. <https://doi.org/10.1080/14786419.2020.1803308>.
- (81) García-Lafuente, A.; Guillamón, E.; Villares, A.; Rostagno, M. A.; Martínez, J. A. Flavonoids as Anti-Inflammatory Agents: Implications in Cancer and Cardiovascular Disease. *Inflamm. Res.* **2009**, *58* (9), 537–552. <https://doi.org/10.1007/s00011-009-0037-3>.
- (82) Yonekura-Sakakibara, K.; Higashi, Y.; Nakabayashi, R. The Origin and Evolution of Plant Flavonoid Metabolism. *Front. Plant Sci.* **2019**, *10* (August), 1–16. <https://doi.org/10.3389/fpls.2019.00943>.
- (83) Kimura, Y.; Sumiyoshi, M.; Sakanaka, M. Effects of *Astilbe Thunbergii* Rhizomes on Wound

- Healing. *J. Ethnopharmacol.* **2007**, *109* (1), 72–77. <https://doi.org/10.1016/j.jep.2006.07.007>.
- (84) Sharma, A.; Gupta, S.; Chauhan, S.; Nair, A.; Sharma, P. Astilbin: a promising unexplored compound with multidimensional medicinal and health benefits. *Pharmacol. Res.* **2020**, *158*, 104894. <https://doi.org/10.1016/j.phrs.2020.104894>.
- (85) Tyagi, A. K.; Prasad, S.; Yuan, W.; Li, S.; Aggarwal, B. B. Identification of a Novel Compound (β -Sesquiphellandrene) from Turmeric (*Curcuma Longa*) with Anticancer Potential: Comparison with Curcumin. *Invest. New Drugs* **2015**, *33* (6), 1175–1186. <https://doi.org/10.1007/s10637-015-0296-5>.
- (86) Kamal, G. M.; Nazi, N.; Sabir, A.; Saqib, M.; Zhang, X.; Jiang, B.; Khan, J.; Noreen, A.; Uddin, J.; Murtaza, S. Yield and Chemical Composition of Ginger Essential Oils as Affected by Inter-Varietal Variation and Drying Treatments of Rhizome. *Separations* **2023**, *10* (3), 186. <https://doi.org/10.3390/separations10030186>.
- (87) Damasceno, C. S. B.; Oliveira, L. F. de; Szabo, E. M.; Souza, Â. M.; Dias, J. F. G.; Miguel, M. D.; Miguel, O. G. Chemical Composition, Antioxidant and Biological Activity of *Ocotea Bicolor* Vattimo-Gil (LAURACEAE) Essential Oil. *Brazilian J. Pharm. Sci.* **2018**, *53* (4). <https://doi.org/10.1590/s2175-97902017000417298>.
- (88) Chiba, N.; Aiuchi, T.; Suzuki, T.; Mori, T.; Shibasaki, M.; Kawahito, Y.; Shioda, S. Comparison of Anti-Nociceptive and Anti-Inflammatory/Analgesic Effects of Essential Oils in Experimental Animal Models. *Jpn. J. Pharm. Palliat. Care Sci.* **2014**, *7*, 63–70.
- (89) Edo, G. I.; Onoharigho, F. O.; Kasar, K. A.; Ainyanbhor, I. E.; Jikah, A. N. Evaluation of the Anti-Inflammatory Potential of *Zingiber Officinale* on Adjuvant-Induced Arthritis. *Adv. Tradit. Med.* **2024**. <https://doi.org/10.1007/s13596-024-00779-6>.
- (90) Keyel, P. A. How Is Inflammation Initiated? Individual Influences of IL-1, IL-18 and HMGB1. *Cytokine* **2014**, *69* (1), 136–145. <https://doi.org/10.1016/j.cyto.2014.03.007>.
- (91) Antonio, A. S.; Veiga-Junior, V. F.; Wiedemann, L. S. M. *Ocotea* Complex: A Metabolomic Analysis of a Lauraceae Genus. *Phytochemistry* **2020**, *173* (January), 112314. <https://doi.org/10.1016/j.phytochem.2020.112314>.
- (92) Katchborian-Neto, A.; Santos, W. T.; de Jesus Nicácio, K.; Corrêa, J. O. A.; Murgu, M.; Martins, T. M. M.; Gomes, D. A.; Goes, A. M.; Soares, M. G.; Dias, D. F.; Chagas-Paula, D. A.; Paula, A. C. C. Neuroprotective Potential of Ayahuasca and Untargeted Metabolomics Analyses: Applicability to Parkinson's Disease. *J. Ethnopharmacol.* **2020**, *255* (March), 112743. <https://doi.org/10.1016/j.jep.2020.112743>.
- (93) Martins, E. M.; Lamont, R. W.; Martinelli, G.; Lira-Medeiros, C. F.; Quinet, A.; Shapcott, A. Genetic Diversity and Population Genetic Structure in Three Threatened *Ocotea* Species (Lauraceae) from Brazil's Atlantic Rainforest and Implications for Their Conservation. *Conserv. Genet.* **2015**, *16* (1), 1–14. <https://doi.org/10.1007/s10592-014-0635-7>.

- (94) Rosa, W.; da Silva Domingos, O.; de Oliveira Salem, P. P.; Caldas, I. S.; Murgu, M.; Lago, J. H. G.; Sartorelli, P.; Dias, D. F.; Chagas-Paula, D. A.; Soares, M. G. In Vivo Anti-Inflammatory Activity of Fabaceae Species Extracts Screened by a New Ex Vivo Assay Using Human Whole Blood. *Phytochem. Anal.* **2021**, No. December 2020, 1–25. <https://doi.org/10.1002/pca.3031>.
- (95) Nicácio, K. de J.; Ferreira, M. S.; Katchborian-Neto, A.; Costa, M. L.; Murgu, M.; Dias, D. F.; Soares, M. G.; Chagas-Paula, D. A. Anti-Inflammatory Markers of Hops Cultivars (*Humulus Lupulus* L.) Evaluated by Untargeted Metabolomics Strategy. *Chem. Biodivers.* **2022**, *19* (4). <https://doi.org/10.1002/cbdv.202100966>.
- (96) Worley, B.; Powers, R. PCA as a Practical Indicator of OPLS-DA Model Reliability. *Curr. Metabolomics* **2016**, *4* (2), 97–103. <https://doi.org/10.2174/2213235x04666160613122429>.
- (97) Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **1995**, *57* (1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- (98) Fiehn, O.; Robertson, D.; Griffin, J.; vab der Werf, M.; Nikolau, B.; Morrison, N.; Sumner, L. W.; Goodacre, R.; Hardy, N. W.; Taylor, C.; Fostel, J.; Kristal, B.; Kaddurah-Daouk, R.; Mendes, P.; van Ommen, B.; Lindon, J. C.; Sansone, S. A. The Metabolomics Standards Initiative (MSI). *Metabolomics* **2007**, *3* (3), 175–178. <https://doi.org/10.1007/s11306-007-0070-6>.
- (99) Spicer, R. A.; Salek, R.; Steinbeck, C. Comment: A Decade after the Metabolomics Standards Initiative It's Time for a Revision. *Sci. Data* **2017**, *4*, 2–4. <https://doi.org/10.1038/sdata.2017.138>.
- (100) Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. H. The WEKA Data Mining Software: An Update. *ACM SIGKDD Explor. Newsl.* **2009**, *11* (1), 10–18.
- (101) Baldim, J. L.; Alcantara, B. G. V. De; Domingos, O. D. S.; Soares, M. G.; Caldas, I. S.; Novaes, R. D.; Oliveira, T. B.; Lago, J. H. G.; Chagas-Paula, D. A. The Correlation between Chemical Structures and Antioxidant, Prooxidant, and Antitrypanosomatid Properties of Flavonoids. *Oxid. Med. Cell. Longev.* **2017**, *2017* (3789856), 1–12. <https://doi.org/10.1155/2017/3789856>.
- (102) Frank, E., Hall, M. A. *Data Mining: Practical Machine Learning Tools and Techniques*; 2016.

Anti-inflammatory markers of *Ocotea* (Lauraceae) uncovered through concatenated UPLC/MS-NMR metabolomics approach – dual inhibition of COX and LOX pathways

Albert Katchborian-Neto^{a,c}, Gabriel Viana^a, Miller Santos Ferreira^a, Matheus Fernandes Alves^a, Michael Murgu^b, Marina de Monroe Gonçalves^c, Ana Cláudia. C. Paula^d, Daniele Ferreira Dias^a, Marisi Gomes Soares^a, João Henrique G. Lago^c, RuAngelie Edrada-Ebel^e,
Daniela A. Chagas-Paula^{a*}

^aInstitute of Chemistry, Federal University of Alfenas, Alfenas, Minas Gerais, Brazil

^bWaters Corporation, Barueri, São Paulo, Brazil

^cCenter of Natural Sciences and Humanities, Federal University of ABC, Santo Andre, São Paulo, Brazil.

^dFaculty of Pharmacy, Federal University of Juiz de Fora, Juiz de Fora, Minas Gerais, Brazil

^eStrathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, Scotland

*Corresponding Authors

E-mail addresses: albert_katchborian@hotmail.com and daniela.chagas@unifal-mg.edu.br

Supplementary material

Tables

Table S1. Statistical analysis was conducted using one-way ANOVA followed by Dunnett's multiple comparison test for PGE2 inhibition levels.

Table S2. Statistical analysis was conducted using one-way ANOVA followed by Dunnett's multiple comparison test for LTB4 inhibition levels.

Table S3. The list of potential new bioactive markers annotated as confidence level 4 positively correlated with anti-inflammatory activity by inhibiting the release of PGE2 and LTB4.

Table S4. *Ocotea* sample was assigned an identification code (ID) detailed along with the geographical location of the plant collection, together with the respective deposit voucher numbers.

Table S5. Multiple Reaction Monitoring (MRM) experiments were performed for each analyte (PGE2, LTB4 and Limaprost), according to their specific m/z transitions.

Table S6. Performance evaluation of LogisticBase, ANN and Random Forest models for scramble data.

Table S7. Parameters description of the prediction models: logistic regression and ANN.

Figures

Figure S1. Stacked UPLC/MS data of 16 *Ocotea* species of positive ionization mode exhibiting a visually higher number of peaks between RT 1.5 to 3.5 min.

Figure S2. Stacked UPLC-MS chromatograms of 16 *Ocotea* species analyzed in negative ionization mode, displaying peaks distributed throughout the entire chromatographic profiles.

Figure S3. Representative stacked NMR spectral profile of 16 *Ocotea* species extracts. Signals in the regions of δ H 0.6–1.4 (methyl proton shifts), δ H 1.5–2.8 (unusual methyl proton or methylene shifts), δ H 3.0–4.0 (methoxy proton shifts), δ H 4.2–5.5 (anomeric proton shifts), and δ H 6.0–8.0 (aromatic proton shifts).

Figure S4. Multivariate statistical analyses of the *Ocotea* crude extract species for discrimination of anti-inflammatory metabolites. Orthogonal Partial Least Square-Discriminant Analysis (OPLS-DA) score scatter plot (left) of UPLC/MS data and loading plot (right) with samples categorized according to the *ex-vivo* anti-inflammatory results (Dual PGE2_LTB4 and Inactive).

Figure S5. Multivariate statistical analyses of the *Ocotea* crude extract species for discrimination of anti-inflammatory metabolites. Orthogonal Partial Least Square-Discriminant Analysis (OPLS-DA) score scatter plot (left) of 1 H NMR data and loading plot (right) with samples categorized according to the *ex-vivo* anti-inflammatory results (Dual PGE2_LTB4 and Inactive).

Figure S6. Orthogonal Partial Least Square- Discriminant Analysis (OPLS-DA) of the concatenated UPLC/MS - NMR fused data. (A) 3D- scatter plot of the *ex-vivo* anti-inflammatory results (Dual PGE2_LTB4 and Inactive). (B) Loadings plot with highlighted mass to charge ratios (m/z) and chemical shifts (δ). (C) Zoom of the highlighted discriminating features of high loading values.

Figure S7. Permutation tests for OPLS-DA models. Permutation tests (100 permutations) for UPLC/MS (left) and 1 H NMR (right) models. Green circles represent cumulative R^2Y values (model fit), and blue squares represent cumulative Q^2 values (predictive ability). Dashed lines indicate trends for permuted data. For UPLC/MS, $Q^2 = 0.00768$ and NMR $Q^2 = -0.611$, indicating poor model predictability.

Figure S8. Additional fragmentation mechanistic pathways proposals for the aporphine, noraporphines and benzyloquinolines.

Figure S9. All stacked 1 H NMR spectra of *Ocotea* leaf extract samples replicates ranging from δ H 2.55 to 3.55. This region of the spectra shows a multiplexed signal intensity in a common area to protons of carbohydrates.

Figure S10. Stacked 1 H NMR spectra (δ H 6.00–8.00) of *Ocotea* leaf extract samples. The region shown highlights signals characteristic of aromatic protons, common to both alkaloids and flavonoids.

Figure S11. Representative active *Ocotea* species stacked ^1H NMR spectra highlighting the STOCSY correlation of driver peak at δH 4.89 related to aromatic signal at δH 7.29 that appeared as a unresolved doublet.

Figure S12. (A) STOCSY analysis of the driver peak signal at δH 3.05 revealed strong correlations in this region (δH 3.09) and in the aromatic region (δH 6.85) suggestive of adjacent CH_2 groups. (B) STOCSY analysis of the driver peak at δH 3.25 revealed high correlations in this region (δH 3.0-3.5) suggestive of other carbohydrate protons.

Figure S13. Area under the curve - receiver operating characteristic (AUC-ROC) graphical plots illustrating the performance of a binary classifier model and their discriminant predictive variables.

Figure S14. KNIME workflow for machine learning model training and validation. The workflow shows the process of reading data, filtering columns, normalizing features, and partitioning into training and test sets, including internal and external validation using cross-validation (10x) and accuracy scoring. A: Multilayer Perceptron - ANN, B: Logistic Regression

Figure S15-S20. Low and high energy MS^E spectra of level 2 annotated bioactive markers as potential dual anti-inflammatory agents.

Scheme 1. Uncertainty of a quotient (Z) - propagation error equation.

Table S1. Statistical analysis was conducted using one-way ANOVA followed by Dunnett's multiple comparison test for PGE2 inhibition levels.

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted <i>p</i> -value
NEG vs. DEX	0.6367	0.4320 to 0.8413	Yes	****	<0.0001
NEG vs. IND	0.7526	0.5480 to 0.9573	Yes	****	<0.0001
NEG vs. <i>O. pulchella</i>	0.4906	0.2638 to 0.7174	Yes	****	<0.0001
NEG vs. <i>O. teleiandra</i>	0.4928	0.2460 to 0.7397	Yes	****	<0.0001
NEG vs. <i>O. tenuiflora</i>	0.5174	0.2906 to 0.7441	Yes	****	<0.0001
NEG vs. <i>O. spixiana</i>	0.4683	0.2215 to 0.7152	Yes	****	<0.0001
NEG vs. <i>O. tabacifolia</i>	0.4436	0.2168 to 0.6703	Yes	****	<0.0001
NEG vs. <i>O. tristis</i>	0.7161	0.4894 to 0.9429	Yes	****	<0.0001
NEG vs. <i>O. vaccinioides</i>	0.5286	0.3019 to 0.7554	Yes	****	<0.0001
NEG vs. <i>O. villosa</i>	0.7489	0.5021 to 0.9958	Yes	****	<0.0001
NEG vs. <i>O. velloziana</i>	0.5469	0.3000 to 0.7938	Yes	****	<0.0001
NEG vs. <i>O. bachybotria</i>	0.2041	-0.04281 to 0.4509	No	ns	0.1705
NEG vs. <i>O. bragai</i>	0.2001	-0.04673 to 0.4470	No	ns	0.1885
NEG vs. <i>O. glauca</i>	0.2104	-0.03644 to 0.4573	No	ns	0.1441
NEG vs. <i>O. glaziovii</i>	0.2210	-0.02589 to 0.4678	No	ns	0.1079
NEG vs. <i>O. lancifolia</i>	0.1287	-0.1182 to 0.3756	No	ns	0.7458
NEG vs. <i>O. langsdorffi</i>	0.2194	-0.02746 to 0.4663	No	ns	0.1127
NEG vs. <i>O. glaucina</i>	0.1433	-0.1035 to 0.3902	No	ns	0.6113

Note: ns – non significant. *Statistical difference.

Table S2. Statistical analysis was conducted using one-way ANOVA followed by Dunnett's multiple comparison test for LTB4 inhibition levels.

Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted <i>p</i> -value
NEG vs. DEX	0.08476	0.04064 to 0.1289	Yes	****	<0.0001
NEG vs. <i>O. pulchella</i>	0.06627	0.02215 to 0.1104	Yes	***	0.0006
NEG vs. <i>O. teleiandra</i>	0.06645	0.02232 to 0.1106	Yes	***	0.0005
NEG vs. <i>O. tenuiflora</i>	0.07799	0.02995 to 0.1260	Yes	***	0.0002
NEG vs. <i>O. spixiana</i>	0.04512	0.0009965 to 0.08925	Yes	*	0.0420
NEG vs. <i>O. tabacifolia</i>	0.04633	0.002198 to 0.09045	Yes	*	0.0338
NEG vs. <i>O. tristis</i>	0.08521	0.04108 to 0.1293	Yes	****	<0.0001
NEG vs. <i>O. vaccinioides</i>	0.06798	0.02385 to 0.1121	Yes	***	0.0004
NEG vs. <i>O. villosa</i>	0.08499	0.04086 to 0.1291	Yes	****	<0.0001
NEG vs. <i>O. velloziana</i>	0.07404	0.02600 to 0.1221	Yes	***	0.0004
NEG vs. <i>O. bachybotria</i>	0.01731	-0.02681 to 0.06144	No	ns	0.9513
NEG vs. <i>O. bragai</i>	0.03014	-0.01398 to 0.07427	No	ns	0.3848
NEG vs. <i>O. glauca</i>	0.02745	-0.01668 to 0.07158	No	ns	0.5115
NEG vs. <i>O. glaziovii</i>	0.01865	-0.02548 to 0.06278	No	ns	0.9178
NEG vs. <i>O. lancifolia</i>	0.01813	-0.02600 to 0.06225	No	ns	0.9323
NEG vs. <i>O. langsdorffi</i>	-0.003171	-0.04730 to 0.04096	No	ns	0.9997
NEG vs. <i>O. glaucina</i>	-0.03186	-0.07990 to 0.01618	No	ns	0.4243

Note: ns – non significant. *Statistical difference.

Table S3. The list of potential bioactive markers annotated as confidence level 4 positively correlated with anti-inflammatory activity by inhibiting the release of PGE2 and LTB4.

ID	RT ^a	Observed <i>m/z</i>	M8.VIP ^b	CorrCoef ^b	<i>q</i> -value (FDR)	AUC	MF	Error (ppm)	Sample	Hits DNP
786_P	4.9	290.269	1.5586	0.158639	4.39E-02	0.80	C ₁₆ H ₃₅ NO ₃	-1.4	<i>O. villosa</i>	1 hit- plants - phytosphingosine derivatives
938_P	5.7	302.304	3.00399	0.266747	4.98E-02	0.64	C ₁₈ H ₃₉ NO ₂	-5.9	<i>O. villosa</i>	1 hit- animals - fatty acids
1104_P	7.2	374.362	1.83946	0.209747	3.26E-02	0.89	C ₂₂ H ₄₇ NO ₃	-2.7	<i>O. tabacifolia</i>	0 hits
199_N	2.1	415.123	2.54787	0.271011	4.20E-02	0.83	C ₁₈ H ₂₄ O ₁₁	-3.8	<i>O. villosa</i>	15 hits - plants - Caffeoylglycerol glycoside derivatives
287_N	2.5	429.14	2.31824	0.230898	4.23E-02	0.71	C ₁₉ H ₂₆ O ₁₁	0.5	<i>O. villosa</i> / <i>O. tristis</i>	16 hits - plants - Caffeoylglycerol glycoside derivatives
1259_N	9.5	461.364	1.90427	0.259117	4.56E-02	0.88	C ₂₉ H ₅₀ O ₄	0.6	<i>O. vaccinoides</i>	27 hits - plants - Stigmasterol derivatives
499_N	2.9	503.177	3.08045	0.367132	4.77E-02	0.59	C ₂₂ H ₃₂ O ₁₃	-0.4	<i>O. spixiana</i>	14 hits - plants -Sinapyl alcohol glycoside derivates
229_N	2.2	565.25	2.32456	0.186682	4.87E-02	0.53	C ₂₅ H ₄₂ O ₁₄	0.1	<i>O. villosa</i>	0 hits
361_N	2.6	597.22	1.64886	0.194322	4.09E-02	0.61	C ₂₈ H ₃₈ O ₁₄	1.7	<i>O. tabacifolia</i>	15 hits - plants - epoxy neolignans glucopyranosides

^a Retention Time, ^b Variant Important in Projection, ^c Correlation coefficients Y variable, ^d False Discovery Rate, ^e Molecular formula. # The 'Sample' column indicates the *Ocotea* extract with the highest peak area associated with each respective bioactive marker.

Table S4. *Ocotea* sample was assigned an identification code (ID) detailed along with the geographical location of the plant collection, deposit voucher numbers and respective dried extract and yield.

<i>ID</i>	<i>Popular name</i>	<i>Scientific name</i>	<i>Geographical location</i>	<i>Voucher</i>	<i>Dried extract</i>	<i>Yield</i>
BA II	louro-verdadeiro	<i>Ocotea brachybotrya</i> (Meisn.) Mez	19°35'28.0"S 42°34'07.0" W	CESJ 45562	4.9 mg	9.8 %
BR II	unknown	<i>Ocotea bragai</i> Coe-Teix.	x	CESJ 43604	12.0 mg	24.0 %
GL I	louro	<i>Ocotea glauca</i> (Nees & Mart.) Mez	20°22'40.0"S 43°24'57.9" W	OUPR 5467	5.9 mg	11.8 %
GZ II	canela-amarela	<i>Ocotea glaziovii</i> Mez	x	CESJ 50895	7.7 mg	15.4 %
LC II	canela-sabão	<i>Ocotea lancifolia</i> (Schott) Mez	18°06'54.0"S 43°20'28.0" W	CESJ 45567	10.9 mg	21.8 %
LG I	unknown	<i>Ocotea langsdorffii</i> (Meisn.) Mez	x	CESJ 16362	7.6 mg	15.2 %
GU II	unknown	<i>Ocotea glaucina</i> (Meisn.) Mez	16°35'47.0"S 42°54'05.0" W	CESJ 50402	5.2 mg	10.4 %
PC II	canela-lageana	<i>Ocotea pulchella</i> (Nees & Mart.) Mez	21°55'24.9"S 46°23'09.9" W	CESJ 49900	8.0 mg	16.0 %
TA I	unknown	<i>Ocotea tabacifolia</i> (Meisn.) Rohwer	x	OUPR 45565	10.1 mg	20.2 %
TL II	canela-limão	<i>Ocotea teleiandra</i> (Meisn.) Mez	x	CESJ 34581	6.1 mg	12.2 %
TE II	unknown	<i>Ocotea tenuiflora</i> (Nees) Mez	x	CESJ 33596	8.7 mg	17.4 %
TR I	canelinha	<i>Ocotea tristis</i> (Nees & Mart.) Mez	20°17'15.0"S 43°30'29.1" W	OUPR 6504	6.4 mg	12.8 %
VA I	unknown	<i>Ocotea vaccinioides</i> (Meisn.) Mez/ <i>Ocotea daphnifolia</i>	x	OUPR 18269	9.0 mg	18.0 %
SX I	canelão	<i>Ocotea spixiana</i> (Nees) Mez	20°17'15.0"S 43°30'19.1" W	OUPR 1390	7.5 mg	15.0 %
VZ II	canela-verde	<i>Ocotea velloziana</i> (Meisn.) Mez	x	CESJ 20721	7.1 mg	14.2 %
VI II	unknown	<i>Ocotea villosa</i> Kosterm.	x	CESJ 50002	11.3 mg	22.6 %

Note: CESJ herbarium – Federal University of Juiz de Fora (UFJF) and OUPR herbarium from Federal University of Ouro Preto (UFOP) in Brazil

Table S5. Multiple Reaction Monitoring (MRM) experiments were performed for each analyte (PGE2, LTB4 and Limaprost – internal standard), according to their specific *m/z* transitions.

<i>Analyte</i>	<i>ESI mode</i>	<i>Precursor (m/z)</i>	<i>Product (m/z)</i>	<i>Dwell time (msec)</i>	<i>Q1 Pre Bias (V)</i>	<i>CE</i>	<i>Q3 Pre Bias (V)</i>
----------------	-----------------	------------------------	----------------------	--------------------------	------------------------	-----------	------------------------

LTB4	-	335.2	195.2	100	11	18	19
LTB4	-	335.2	59.1	100	14	25	20
PGE2	-	351.2	271.2	100	14	20	17
PGE2	-	351.2	315.25	100	14	15	21
PGE2	-	351.2	333.35	100	11	15	23
Limaprost	-	379.2	343.45	100	11	13	24
Limaprost	-	379.2	299.3	100	14	21	21

Table S6. Performance evaluation of LogisticBase and ANN models for scramble data.

<i>ML algorithms</i>		<i>Accuracy (%)</i>	<i>Cohen's Kappa</i>
LogisticBase	Training	67	0
	External Validation	60	0
	Internal Validation	58	-0.1
ANN	Training	100	1
	External Validation	50	-0.3
	Internal Validation	54	0.1

Table S7. Parameters description of prediction models: logistic regression and ANN.

<i>Method</i>	<i>Parameter</i>	<i>Values</i>	<i>Description</i>
	H_hiddenLayers	46	The hidden layers are to be created for the network.

ANN	L_learningRate	0.3	Learning Rate for the backpropagation algorithm.
	M_momentum	0.2	Momentum Rate for the backpropagation algorithm
	N_trainingTime	500	Number of epochs to train through.
	S_seed	100	The value used to seed the random number Generator
	Decay	False	Learning rate decay will occur. (Set this to cause the learning rate to decay).
	Reset	True	Resetting the network will NOT be allowed. (Set this to not allow the network to reset)
	Nominaltobinaryfilter	True	A NominalToBinary filter will NOT automatically be used.
LogisticBase	Debug	False	Outputs detailed information about the model's training process
	maxIterations	500	Maximum number of iterations that the Logistic Regression algorithm can perform during the optimization process
	useAIC	False	To evaluate model quality by balancing goodness of fit and model complexity, penalizing models with excessive parameters
	weighttrimBeta	0.0	It defines a threshold to trim instances with very small weights during the training process.

Figure S1. Stacked UPLC/MS data of 16 *Ocotea* species of positive ionization mode exhibiting visually higher number of peaks between RT 1.5 to 3.5 min.

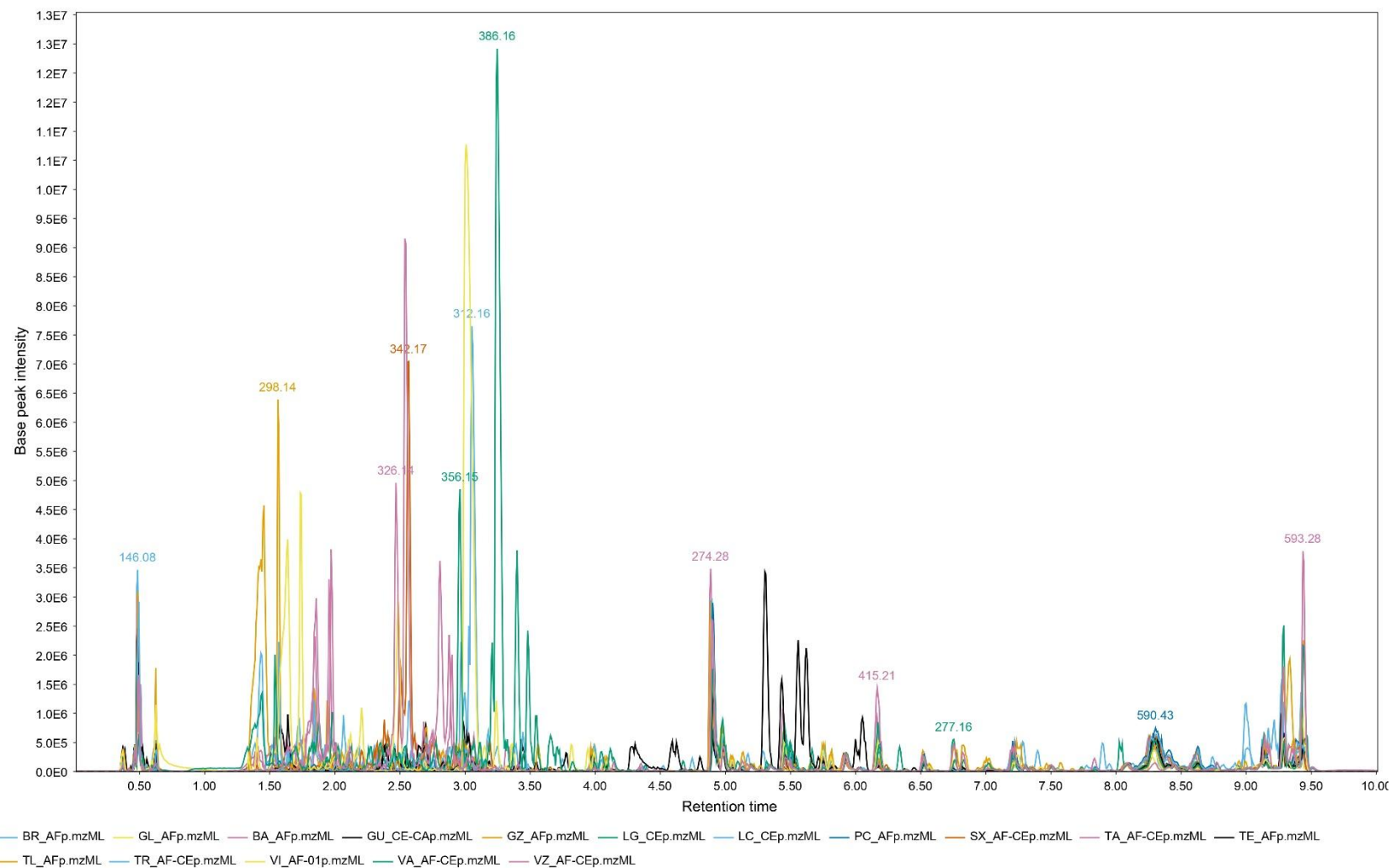


Figure S2. Stacked UPLC-MS chromatograms of 16 *Ocotoca* species analyzed in negative ionization mode, displaying peaks distributed throughout the entire chromatographic profiles.

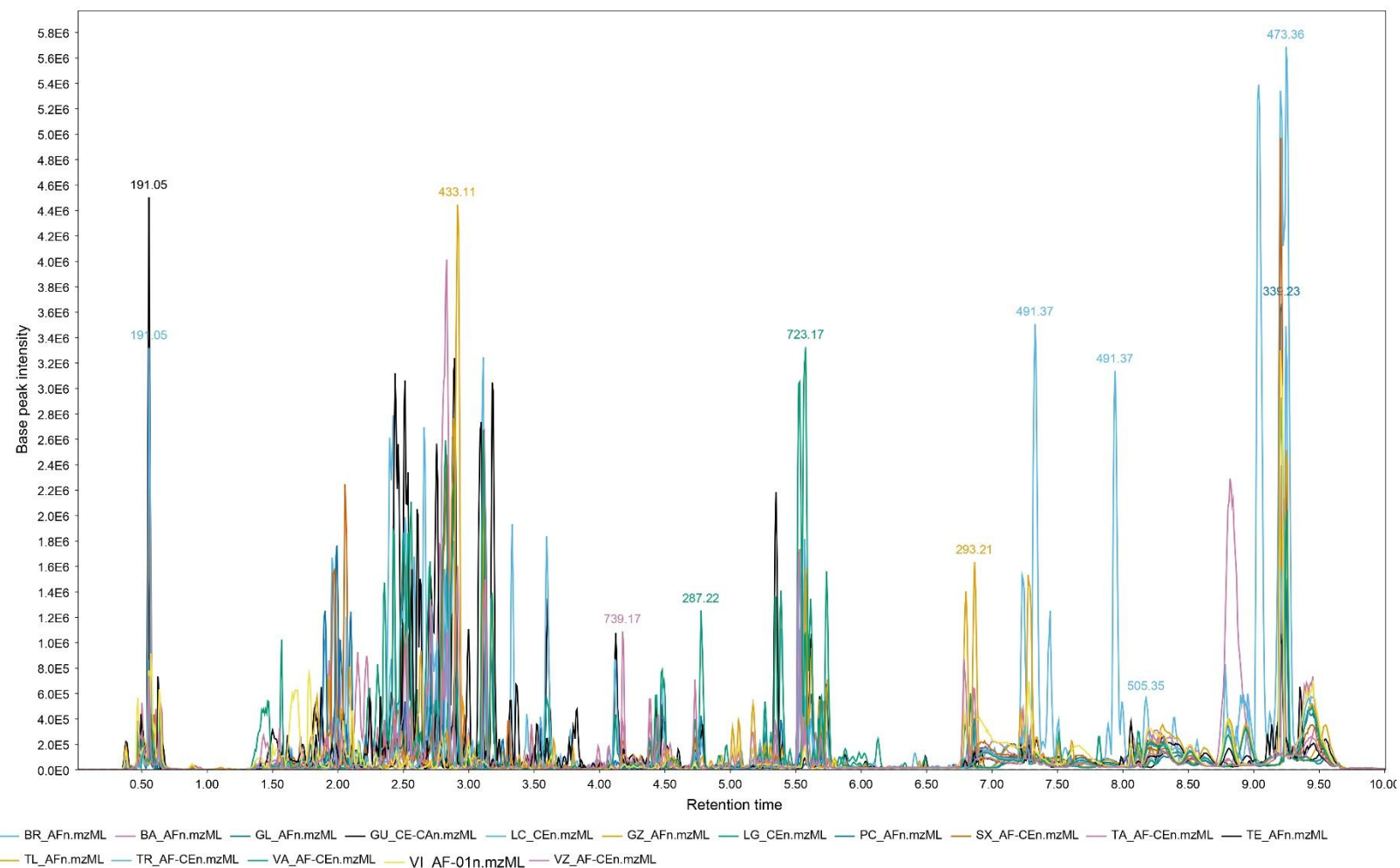


Figure S3. Representative stacked NMR spectral profile of 16 *Ocotea* species extracts. Signals in the regions of δ H 0.6–1.4 (methyl proton shifts), δ H 1.5–2.8 (unusual methyl proton or methylene shifts), δ H 3.0–4.0 (methoxy proton shifts), δ H 4.2–5.5 (anomeric proton shifts), and δ H 6.0–8.0 (aromatic proton shifts).

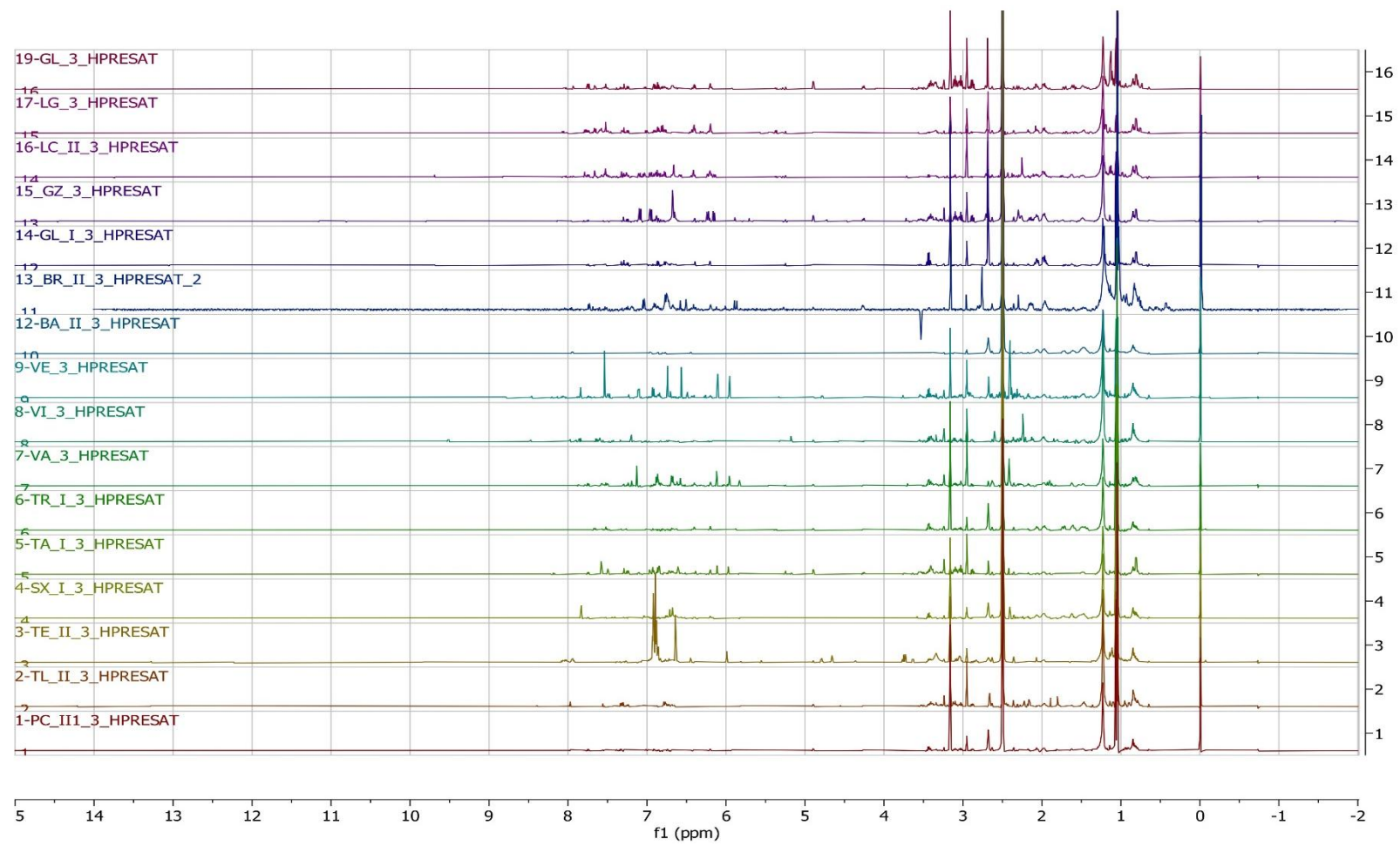


Figure S4. Multivariate statistical analyses of the *Ocotea* crude extract species for discrimination of anti-inflammatory metabolites. Orthogonal Partial Least Square- Discriminant Analysis (OPLS-DA) score scatter plot (left) of UPLC/MS data and loading plot (right) with samples categorized according to the *ex-vivo* anti-inflammatory results (Dual PGE2_LTB4 and Inactive).

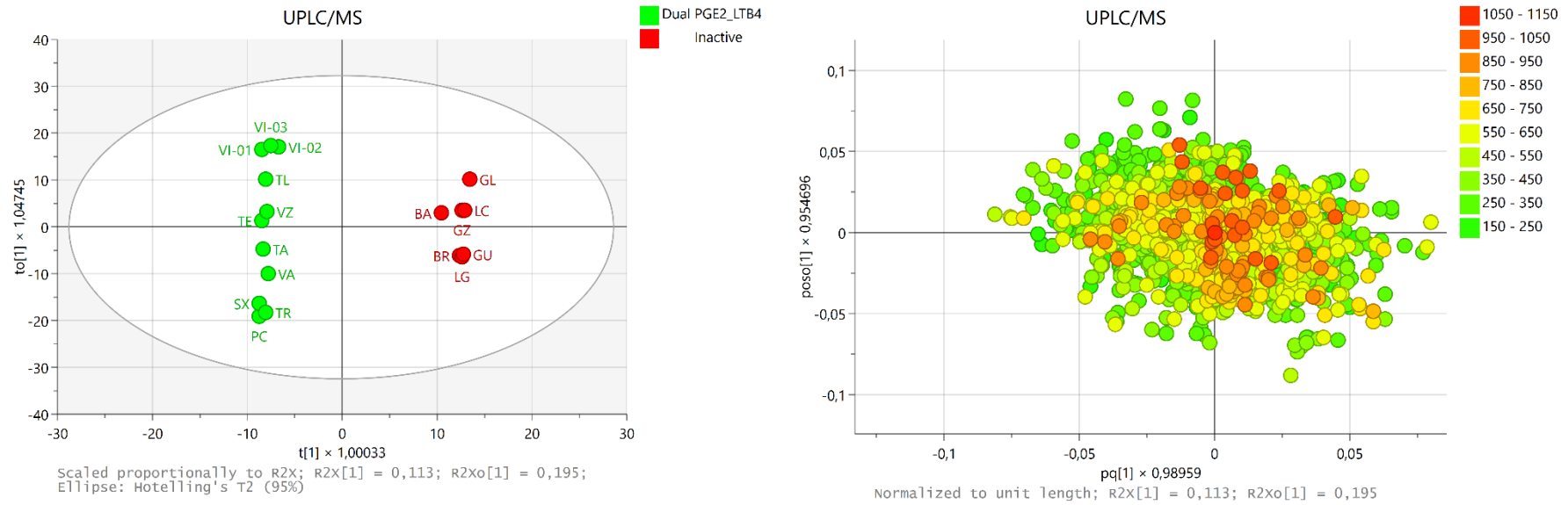


Figure S5. Multivariate statistical analyses of the *Ocotea* crude extract species for discrimination of anti-inflammatory metabolites. Orthogonal Partial Least Square- Discriminant Analysis (OPLS-DA) score scatter plot (left) of ^1H NMR data and loading plot (right) with samples categorized according to the *ex-vivo* anti-inflammatory results (Dual PGE₂_LTB₄ and Inactive).

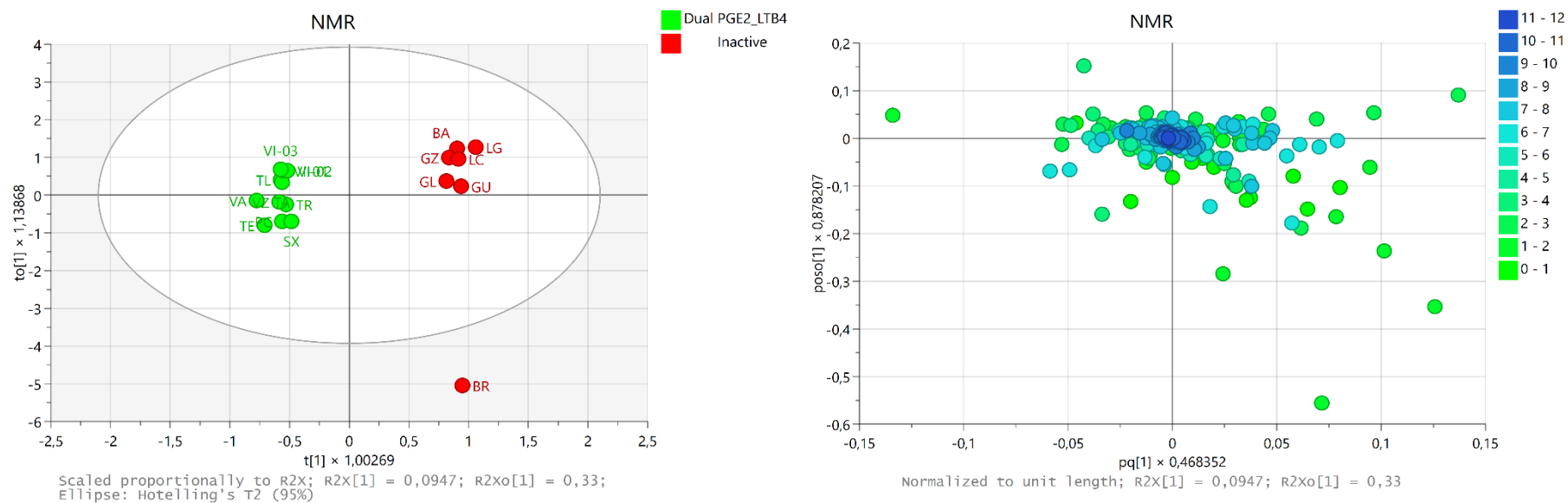


Figure S6. Orthogonal Partial Least Square- Discriminant Analysis (OPLS-DA) of the concatenated UPLC/MS - NMR fused data. (A) 3D- scatter plot of the *ex-vivo* anti-inflammatory results (Dual PGE2_LTB4 and Inactive). (B) Loadings plot with highlighted mass to charge ratios (m/z) and chemical shifts (δ). (C) Zoom of the highlighted discriminating features of high loading values.

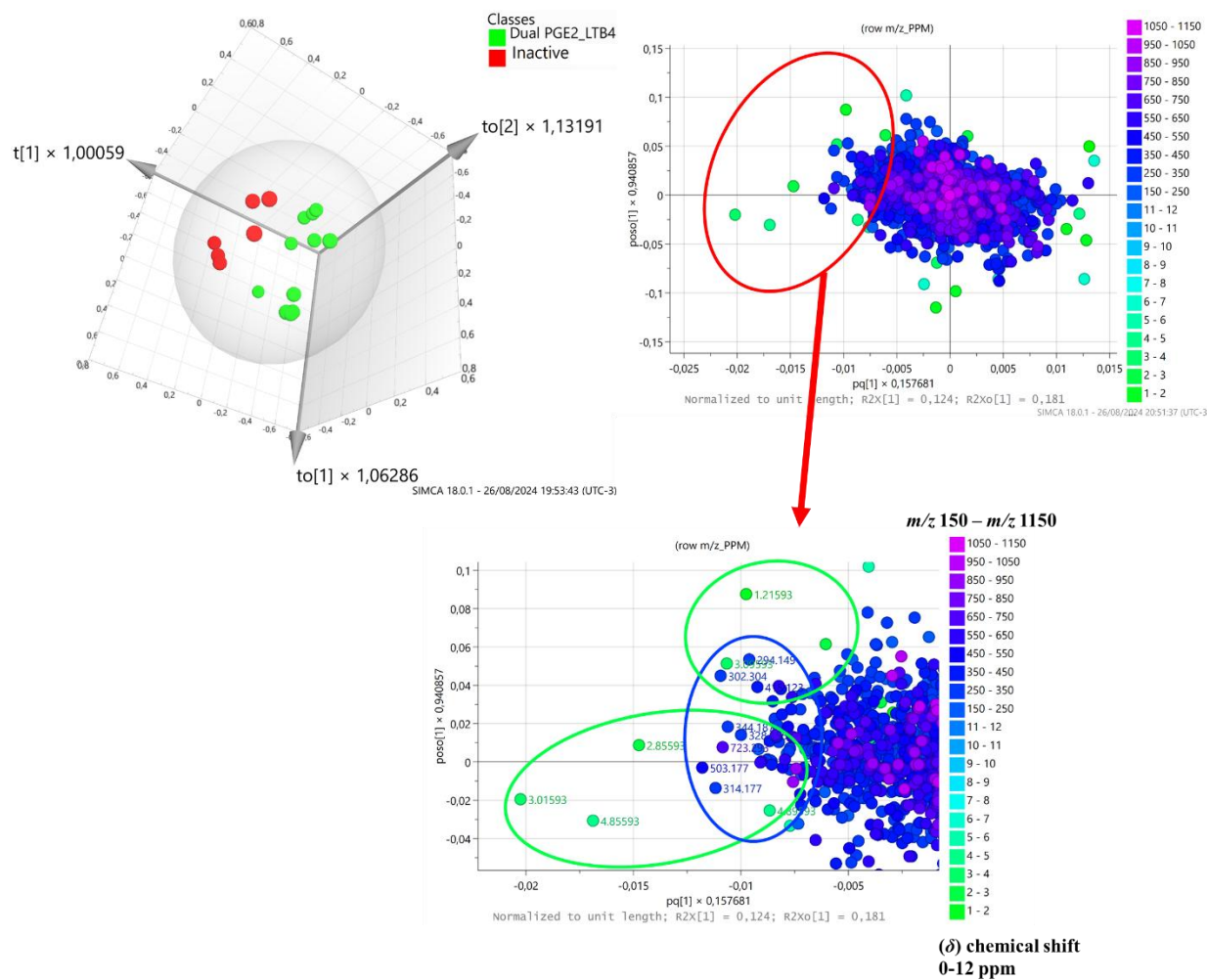
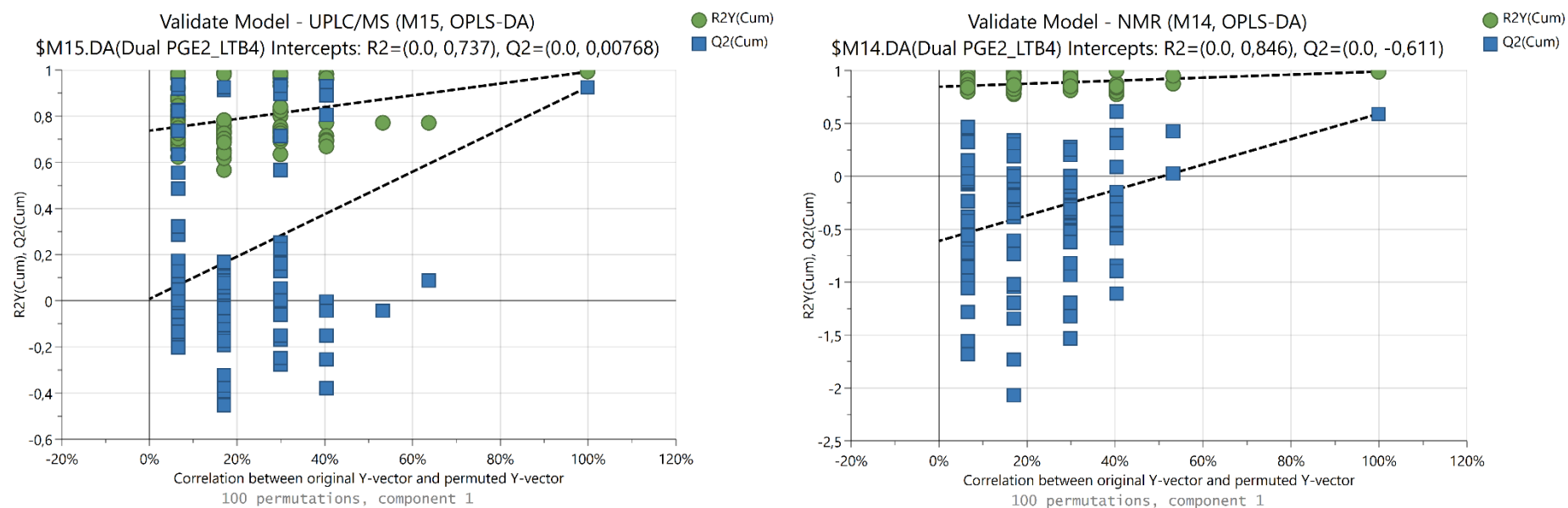
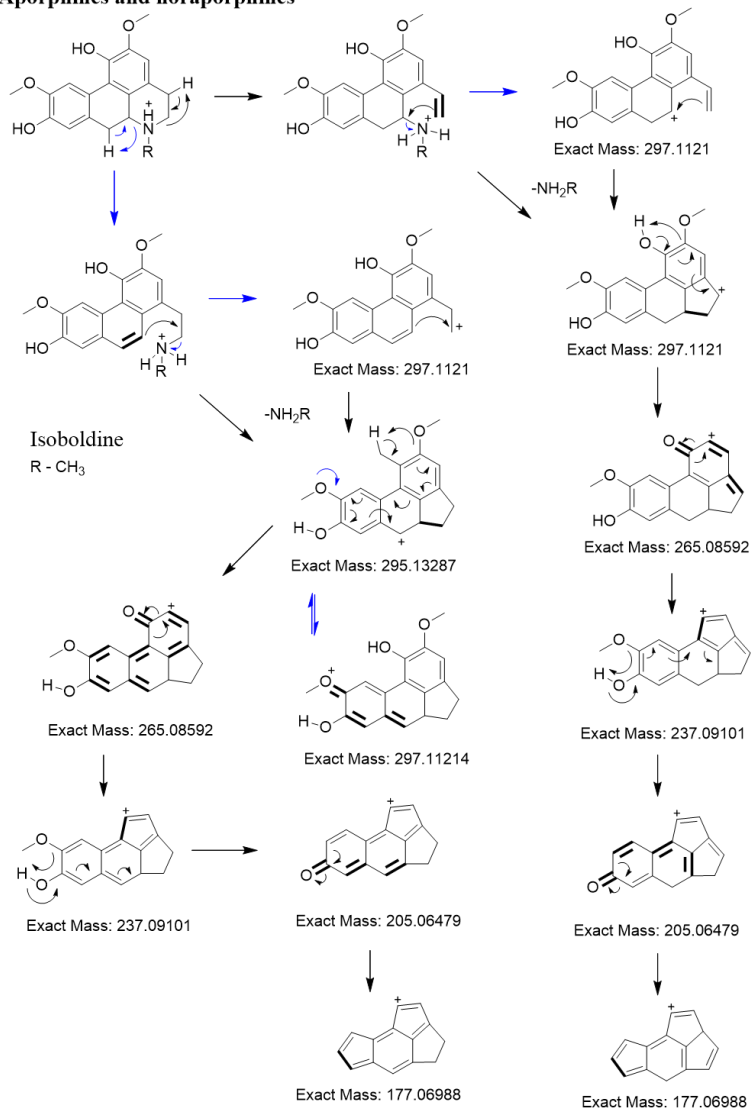


Figure S7. Permutation tests for OPLS-DA models. Permutation tests (100 permutations) for UPLC/MS (left) and ^1H NMR (right) models. Green circles represent cumulative R^2Y values (model fit), and blue squares represent cumulative Q^2 values (predictive ability). Dashed lines indicate trends for permuted data. For UPLC/MS, $Q^2 = 0.00768$ and NMR $Q^2 = -0.611$, indicating poor model predictability.



Figures S8. Additional fragmentation mechanistic pathways proposals for the aporphine, noraporphines and benzyloquinolines.

Aporphines and noraporphines



Benzylisoquinolines

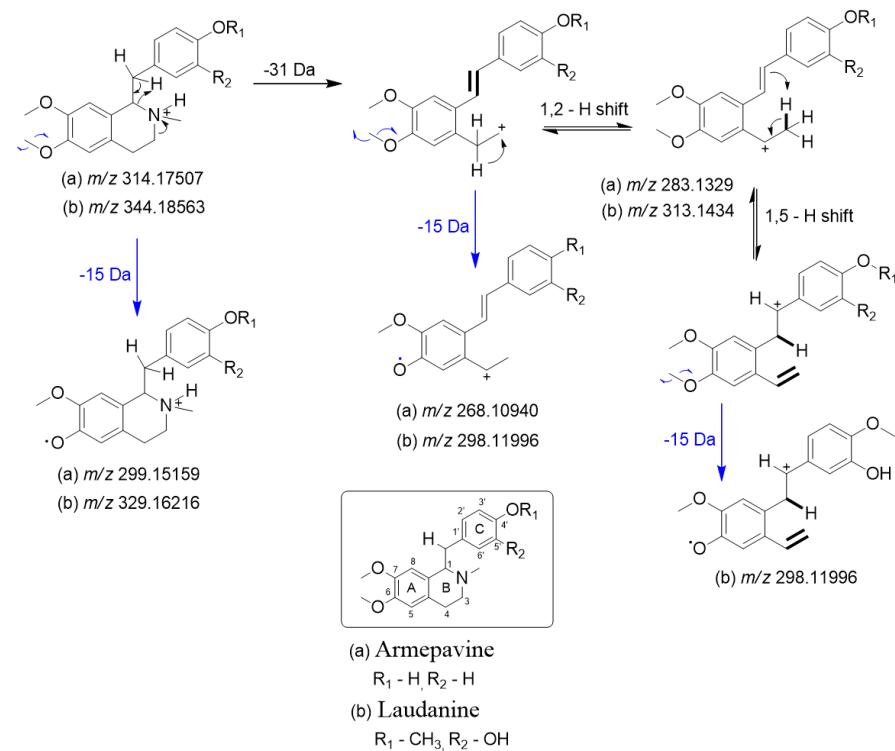


Figure S9. All stacked ^1H NMR spectra of *Ocotea* leaf extract samples replicates ranging from δH 2.55 to 3.55. This region of the spectra shows a multiplexed signal intensity in a common area to protons of carbohydrates.

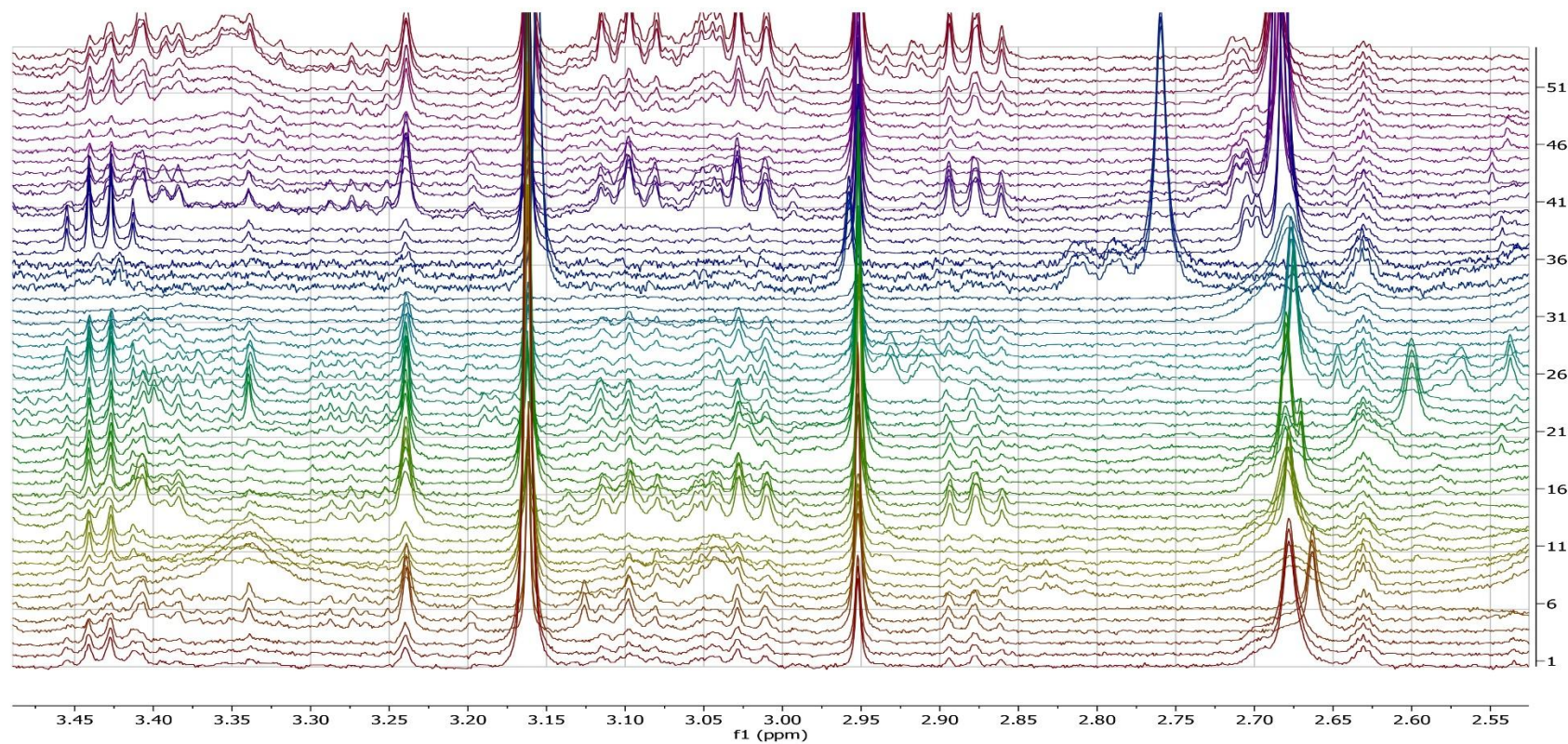


Figure S10. All stacked ^1H NMR spectra (δH 6.00–8.00) of *Ocotea* leaf extract samples replicates. The region shown highlights signals characteristic of aromatic protons, common to both alkaloids and flavonoids.

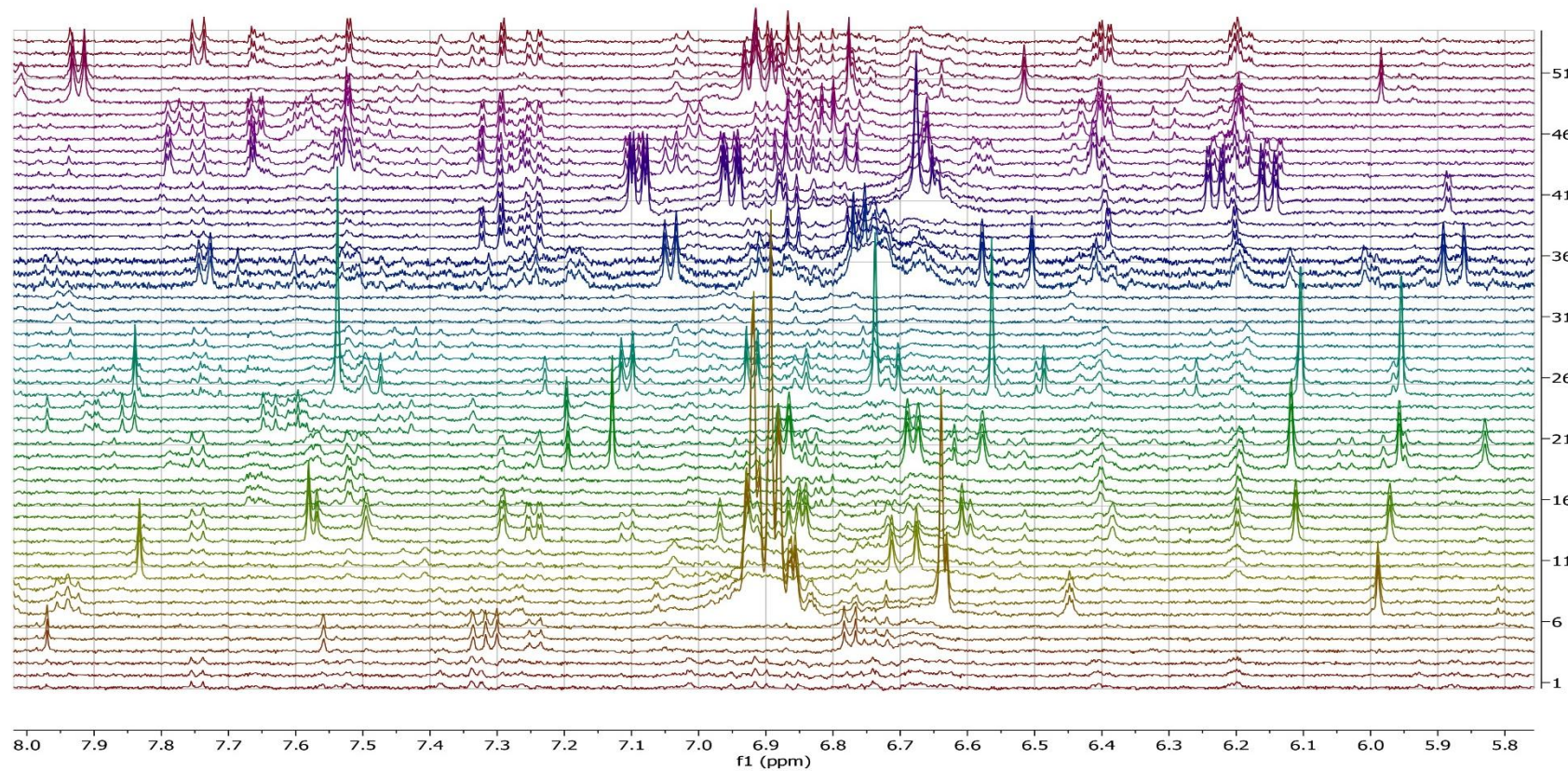


Figure S11. Representative active *Ocotea* species stacked ^1H NMR spectra highlighting the STOCSY correlation of driver peak at δH 4.89 related to aromatic signal at δH 7.29 that appeared as an unresolved doublet.

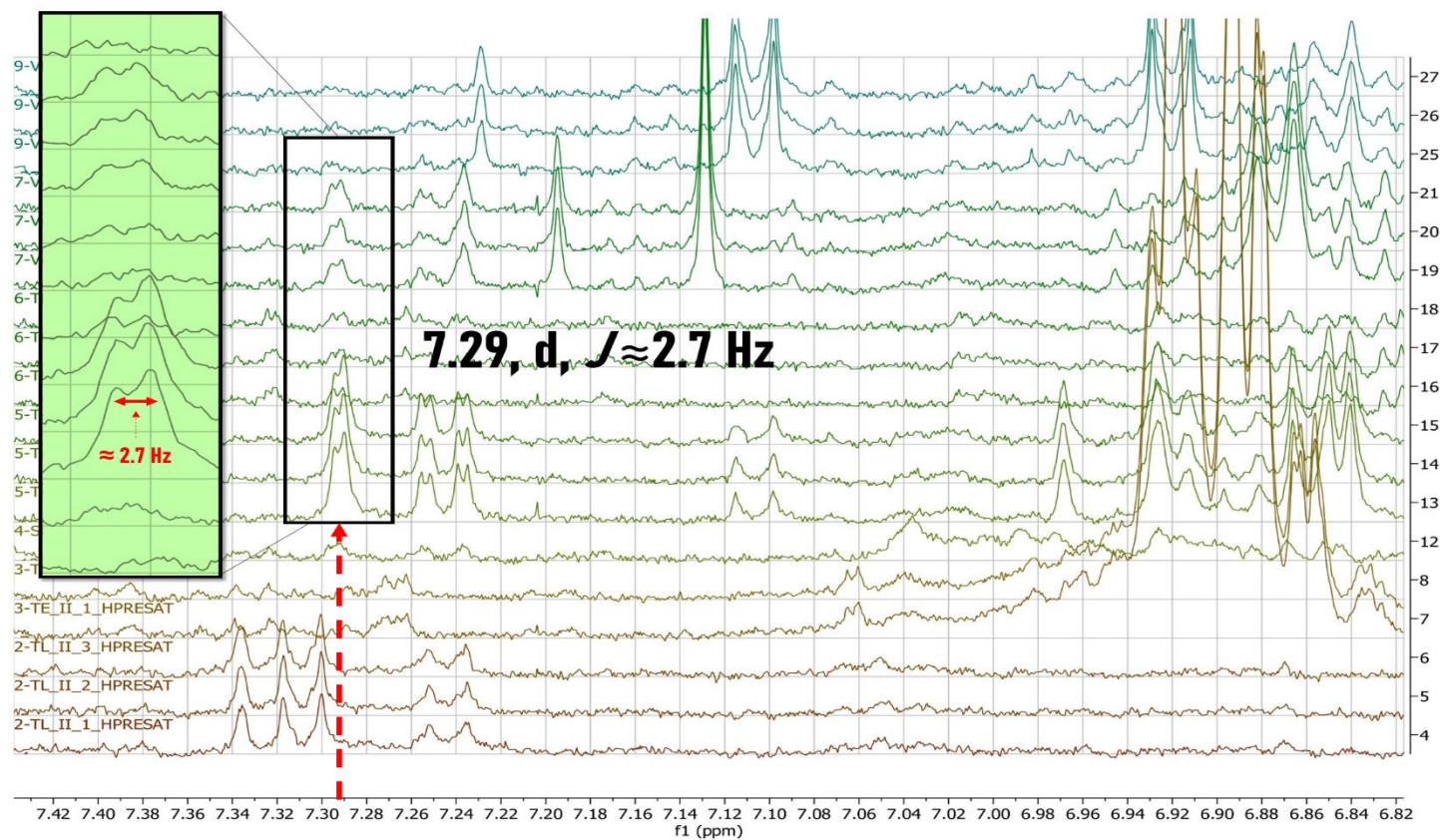


Figure S12. (A) STOCSY analysis of the driver peak signal at δH 3.05 revealed strong correlations in this region (δH 3.09) and in the aromatic region (δH 6.85) suggestive of adjacent CH_2 groups. (B) STOCSY analysis of the driver peak at δH 3.25 revealed high correlations in this region (δH 3.0-3.5) suggestive of other carbohydrate protons.

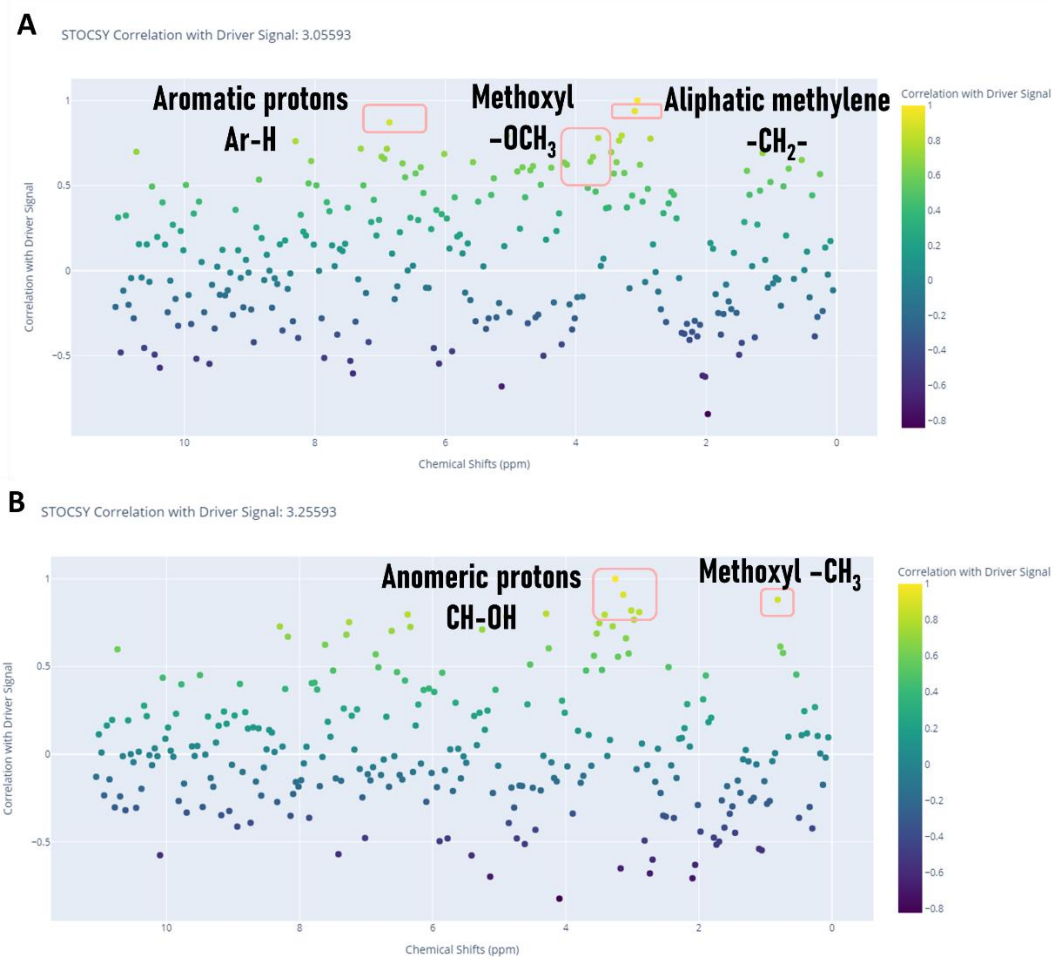


Figure S13. Area under the curve - receiver operating characteristic (AUC-ROC) graphical plots illustrating the performance of a binary classifier model and their discriminant predictive variables.

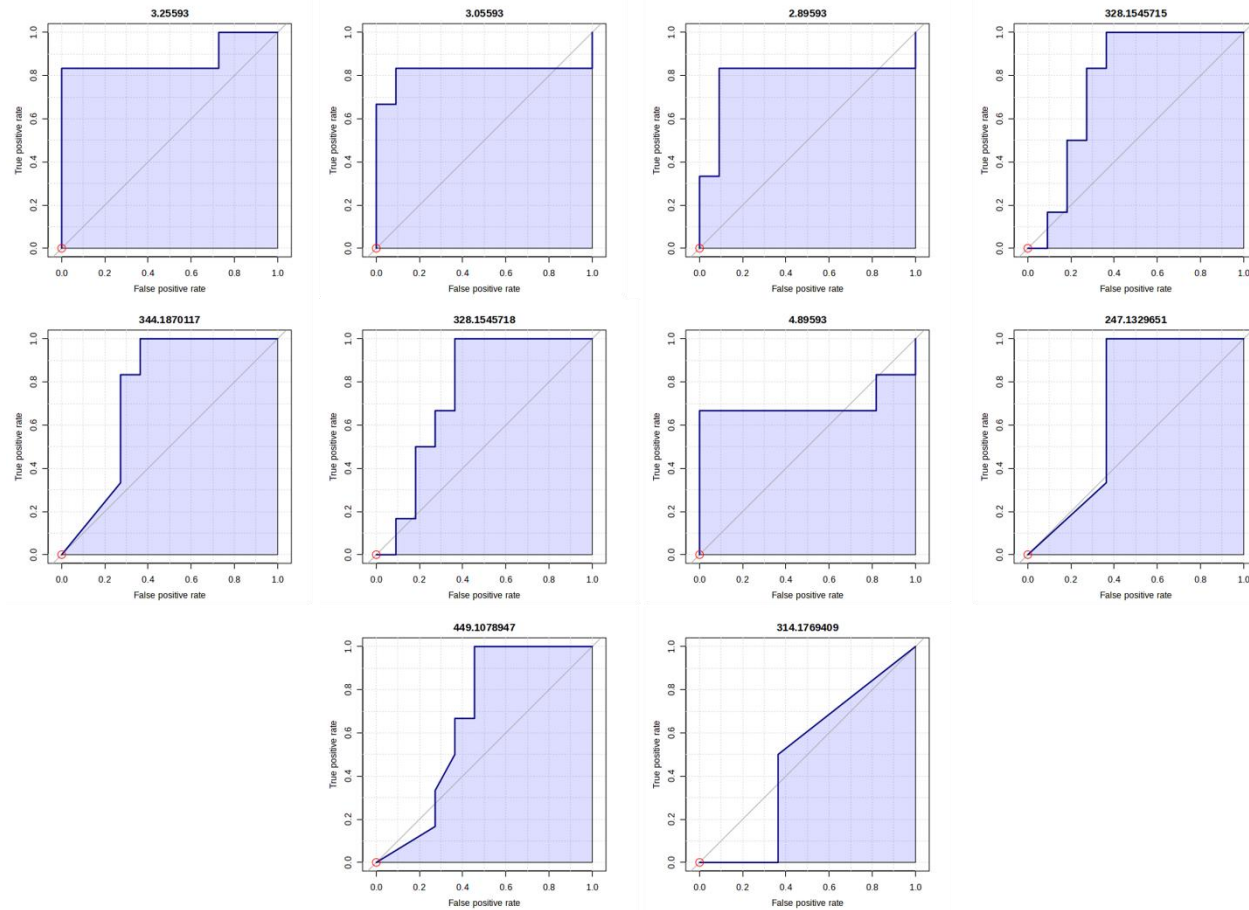


Figure S14. KNIME workflow for machine learning model training and validation. The workflow shows the process of reading data, filtering columns, normalizing features, and partitioning into training and test sets, including internal and external validation using cross-validation (10x) and accuracy scoring. A: Multilayer Perceptron - ANN, B: Logistic Regression

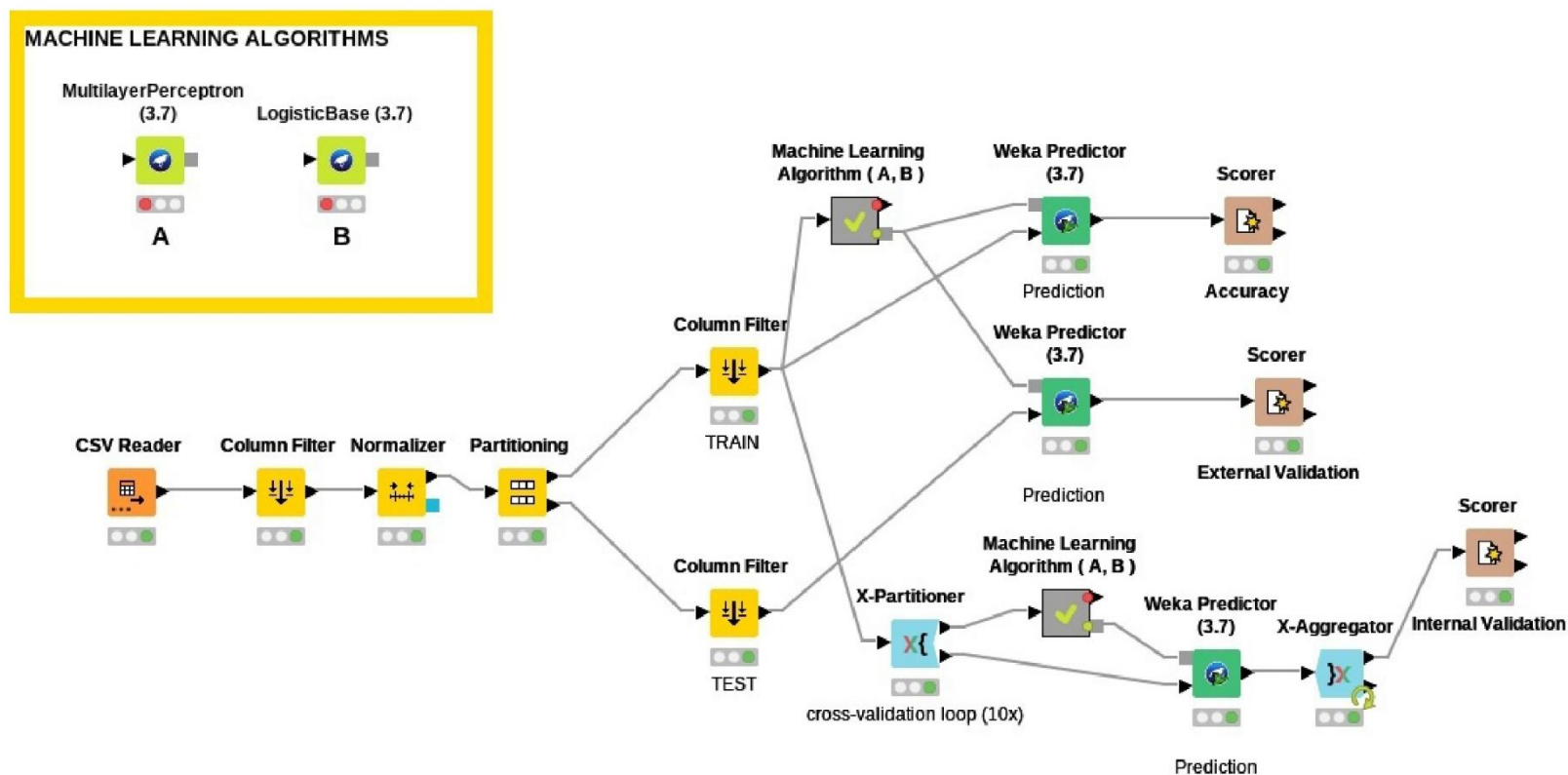


Figure S15. Low and high energy MS^E spectra of the ion at m/z 314.1769 / RT 1.71 annotated as armepavine (M+H; sample: VA).

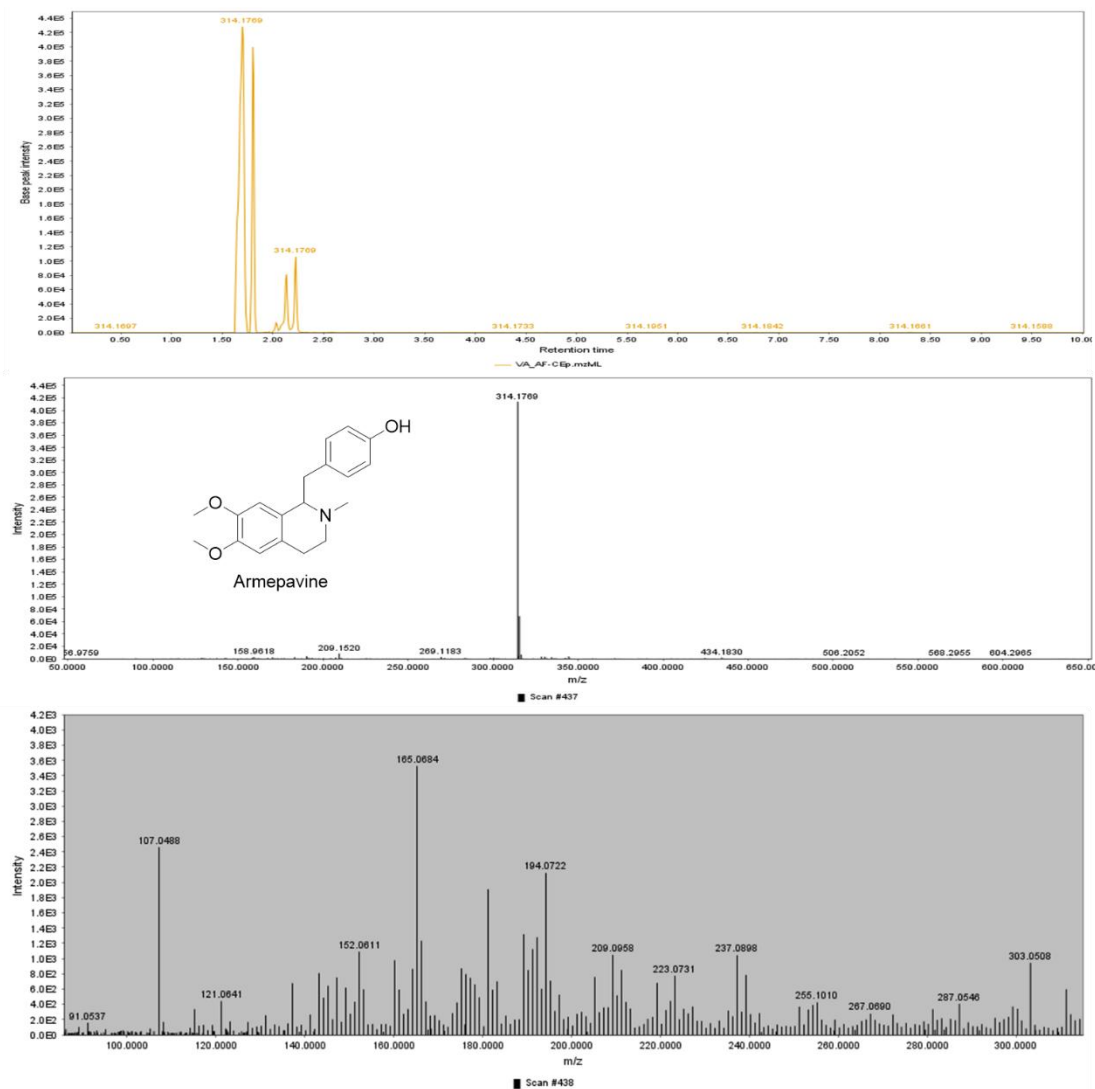


Figure S16. Low and high energy MS^E spectra of the ion at m/z 344.1870 / RT 1.97 annotated as laudanine (M+H; sample: VI).

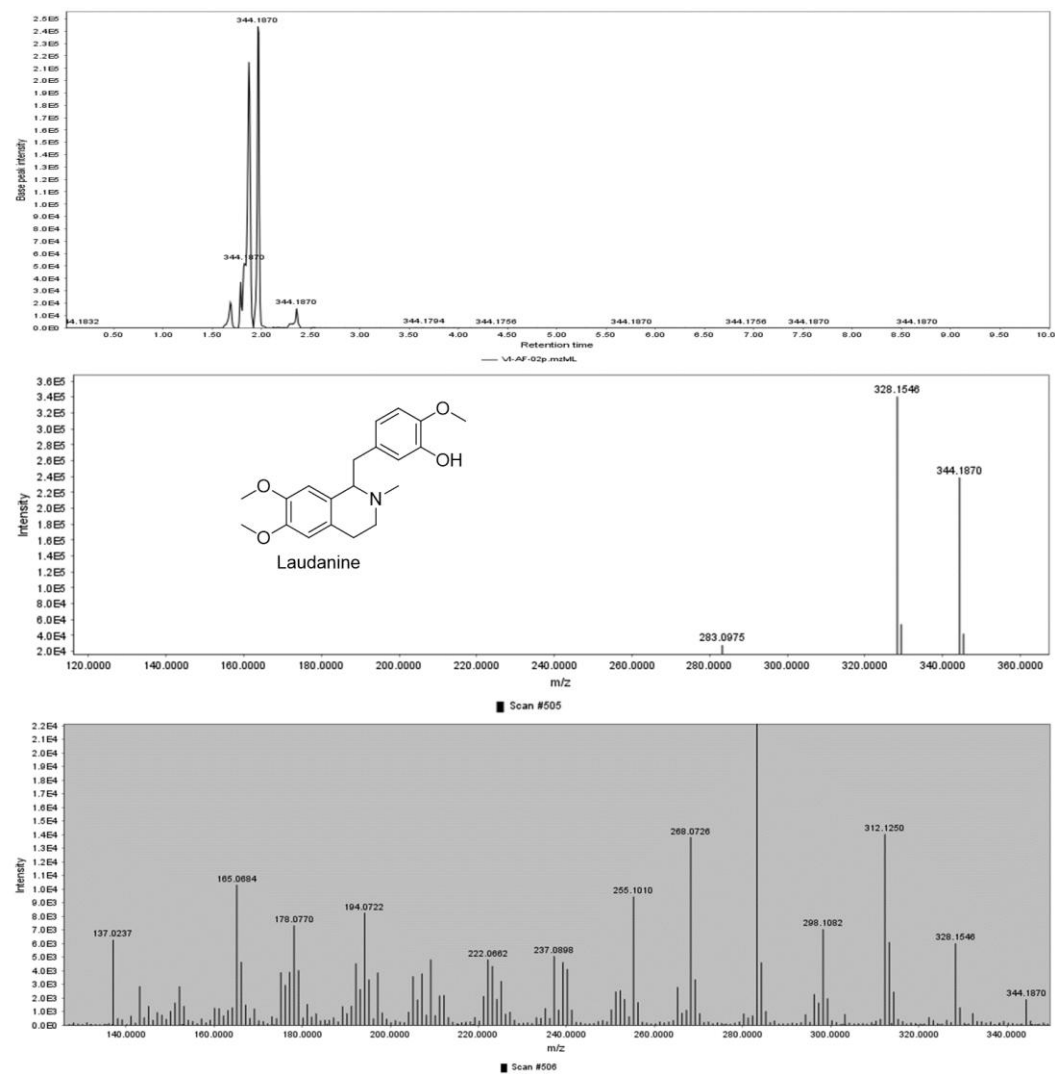


Figure S17. Low and high energy MS^E spectra at m/z 328.1538 / RT 1.94 annotated as laurotetanine / norisocorydine (M+H; sample: VZ).

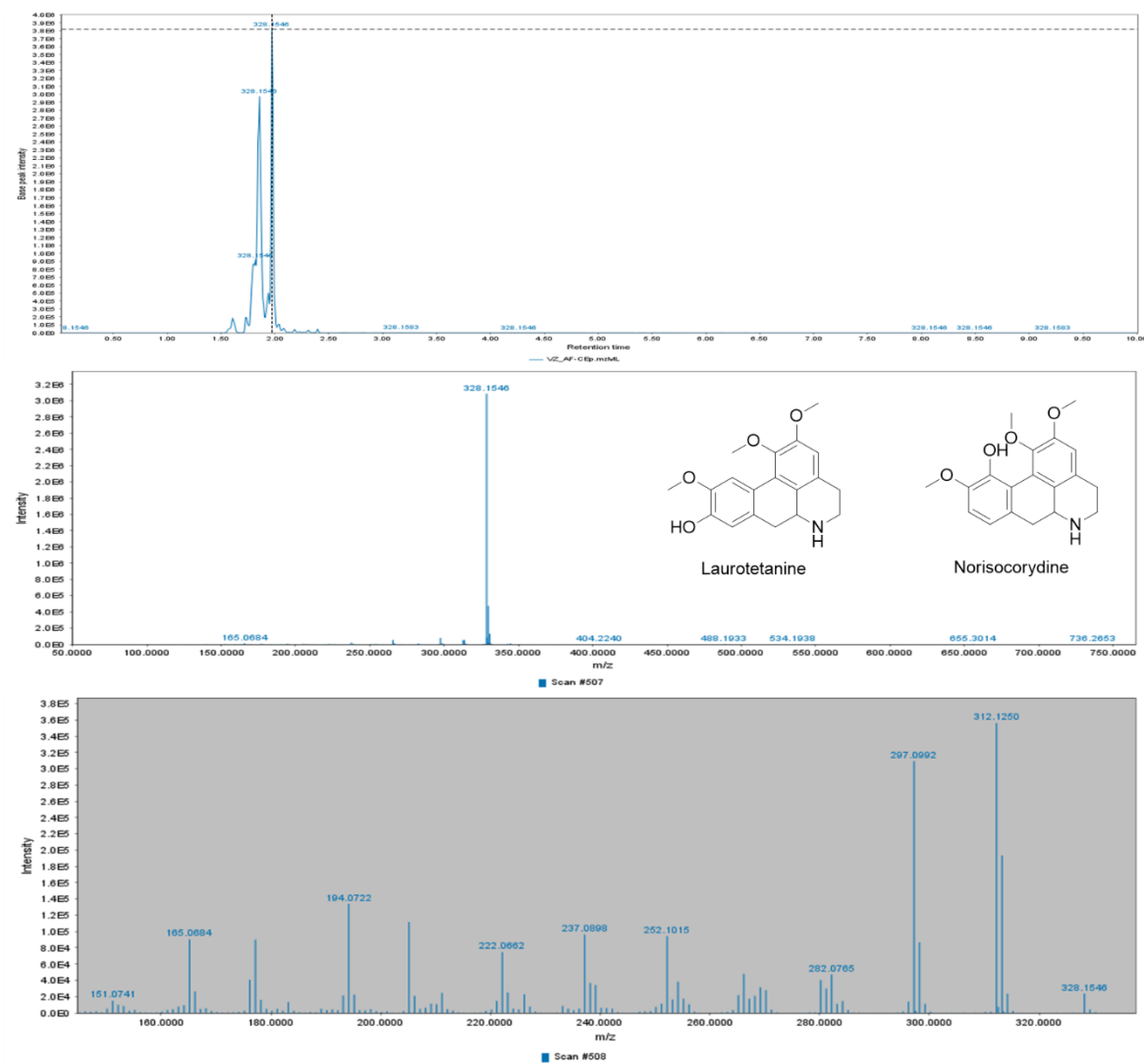


Figure S18. Low and high energy MS^E spectra of the ion at m/z 328.1538 / RT 1.97 annotated as isoboldine or boldine (M+H; sample: VI).

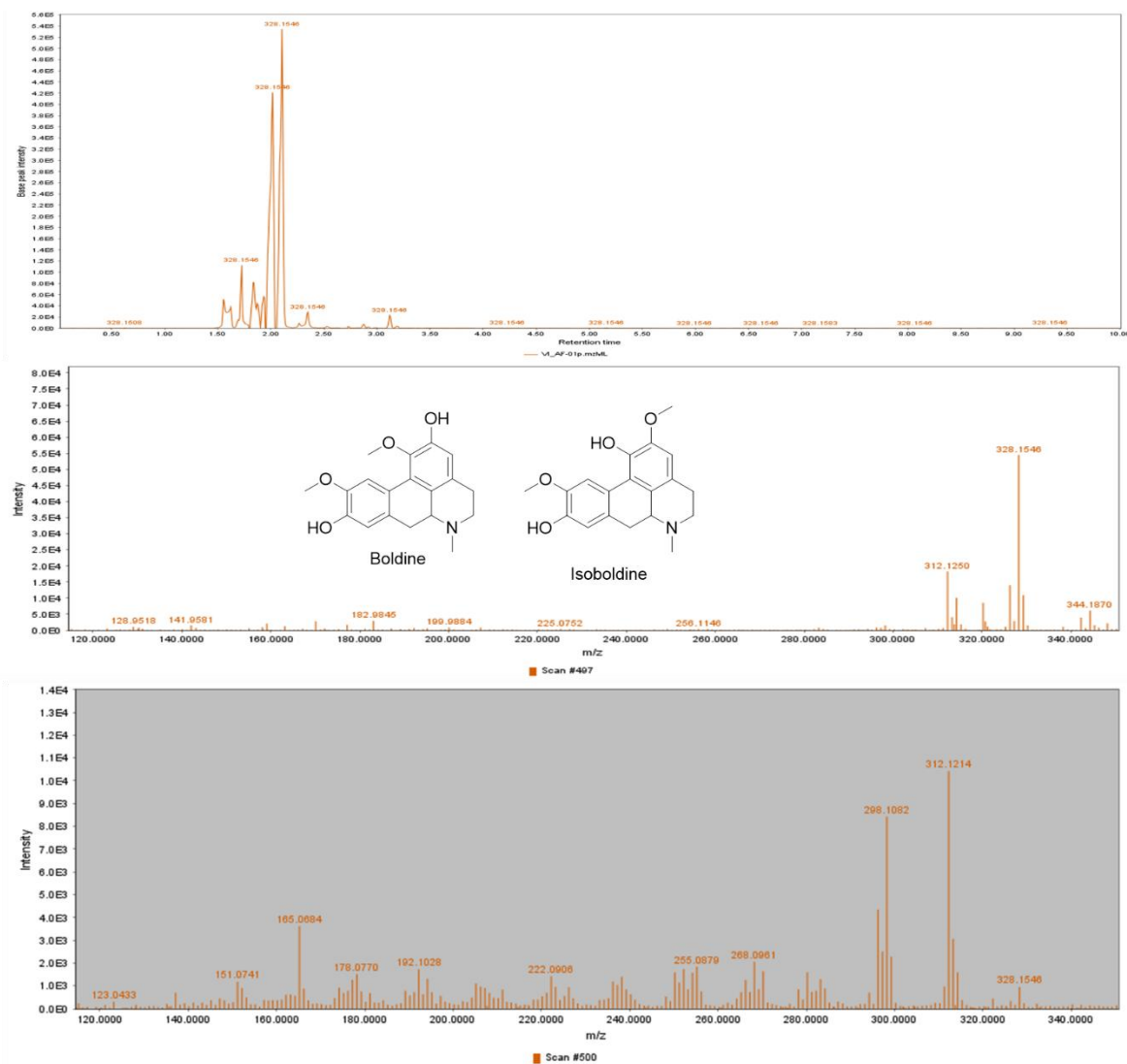


Figure S19. Low and high energy MS^E spectra of the ion at m/z 449.1080 / RT 2.62 annotated as astilbin (M-H; sample: TL).

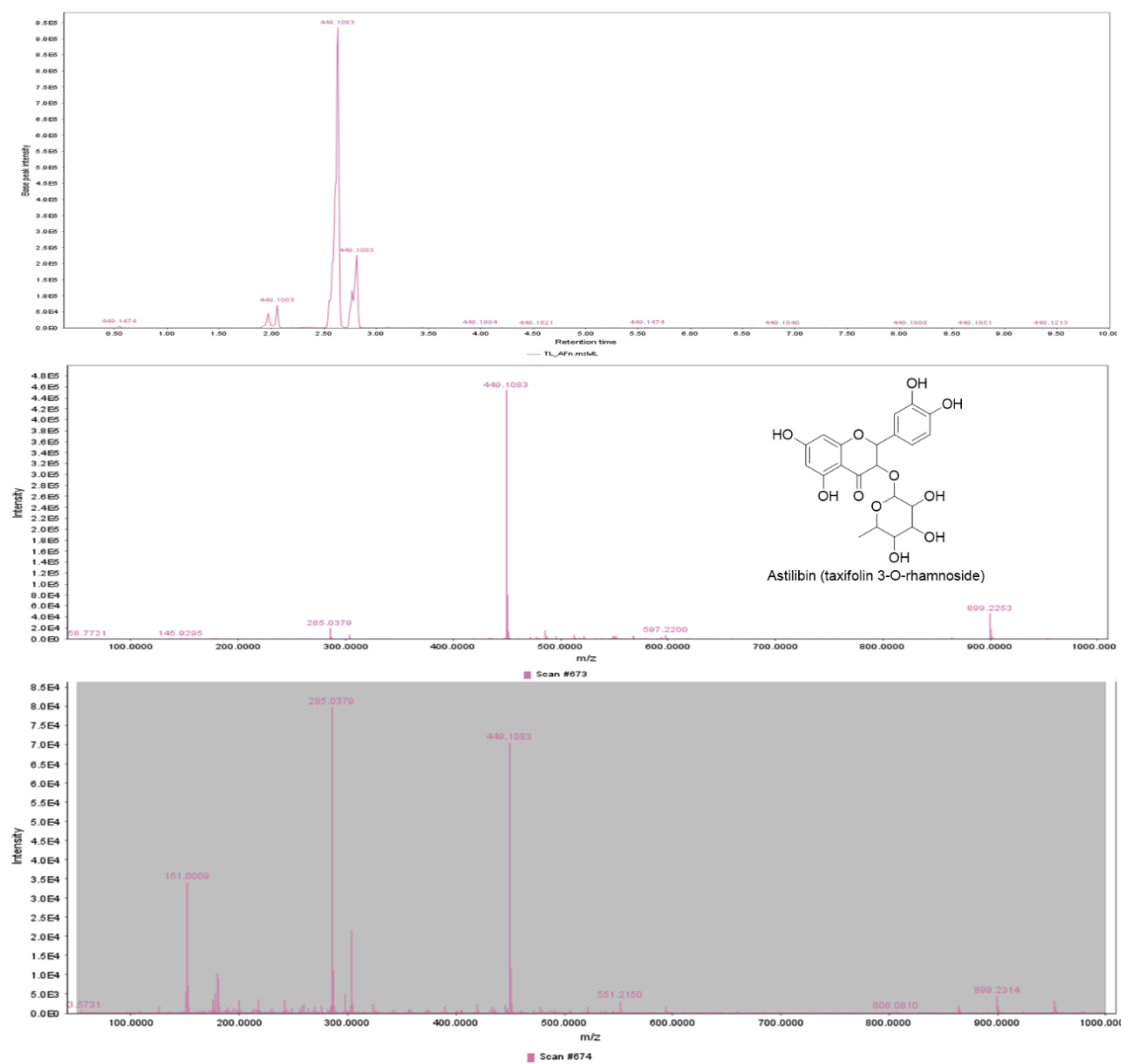
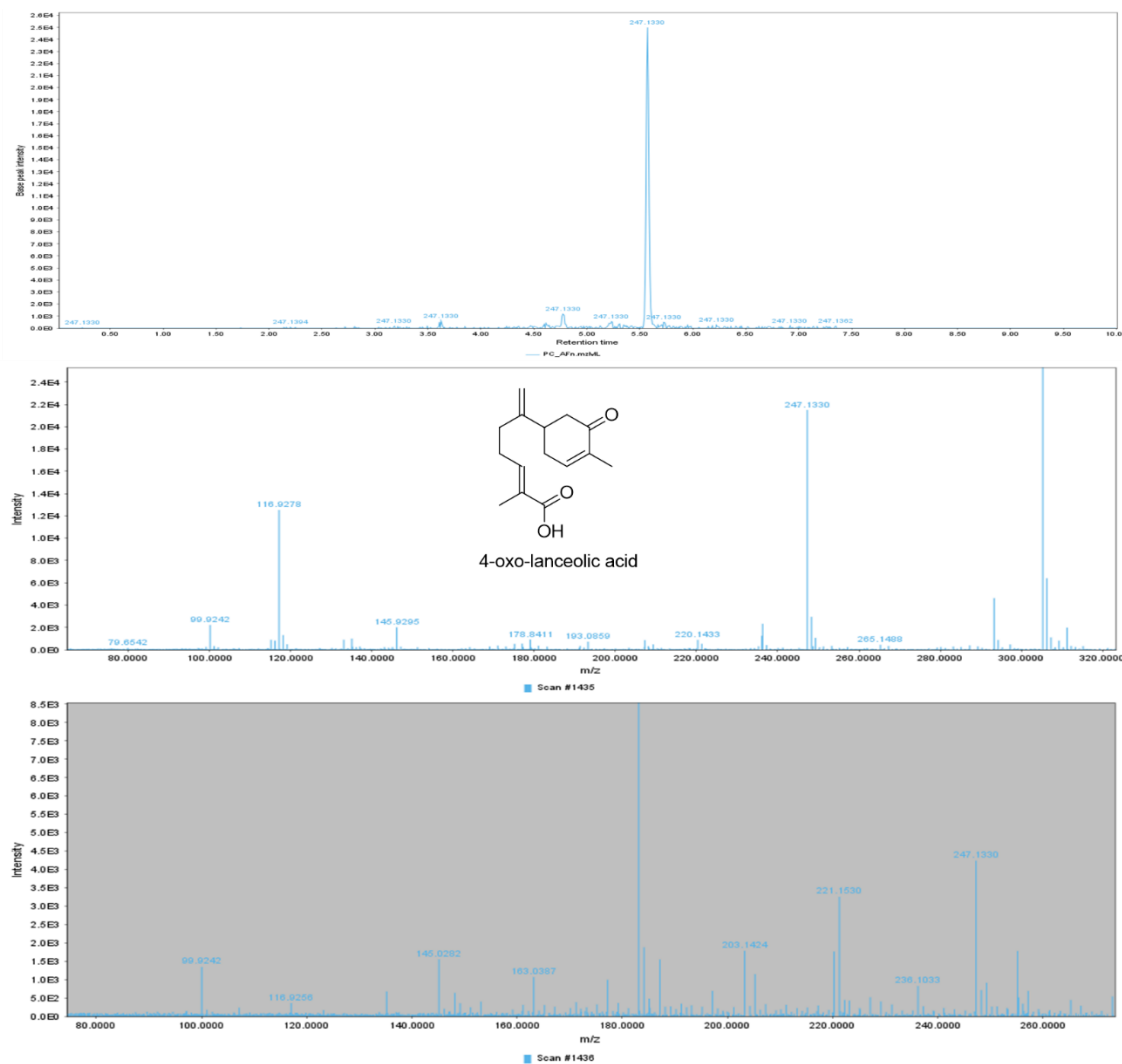


Figure S20. Low and high energy MS^E spectra of the ion at m/z 247.1330 / RT 5.57 annotated as 4-oxo-lanceolic acid (M-H; sample: PC).



Scheme 1 – uncertainty of a quotient (Z) propagation error equation

Calculating the uncertainty (σ_z) of a quotient (Z) according to propagation of error. In this case, X and Y are experimental (average) measurements and σ_x and σ_y are their respective uncertainties, expressed by standard deviation (SD).

If:

$$Z = \frac{X}{Y} \quad (\text{Equation 1})$$

So:

$$\sigma_z = Z * \sqrt{\left(\frac{\sigma_x}{X}\right)^2 + \left(\frac{\sigma_y}{Y}\right)^2} \quad (\text{Equation 2})$$

Reference: Daniel T. Holmes & Kevin A. Buhr. Error propagation in calculated ratios. *Clinical biochemistry*. (2007) 40.9-10: 728-734. DOI: 10.1016/j.clinbiochem.2006.12.014

6 CHAPTER IV – DIA-MS DATA PROCESSING AND MOLECULAR NETWORKING PIPELINE

Chapter IV containing the published research is separately provided as an attached material of this thesis.

Academic Journal: Briefings in Bioinformatics (Oxford Press) (published)

Impact factor (2024): 6.8

DOI: <https://doi.org/10.1093/bib/bbae013>
















Briefings in Bioinformatics, 2024, 25(2), 1–18

<https://doi.org/10.1093/bib/bbae013>

Problem Solving Protocol

Integrative open workflow for confident annotation and molecular networking of metabolomics MS^E/DIA data

Albert Katchborian-Neto , Matheus F. Alves , Paula C.P. Bueno , Karen de Jesus Nicácio , Miller S. Ferreira ,
Tiago B. Oliveira , Henrique Barbosa , Michael Murgu , Ana C.C. de Paula Ladvoat , Danielle F. Dias , Marisi G. Soares ,
João H.G. Lago  and Daniela A. Chagas-Paula 














Corresponding author: Daniela A. Chagas-Paula, Laboratory of Phytochemistry, Medicinal Chemistry and Metabolomics, Institute of Chemistry, Federal University of Alfenas, R. Gabriel Monteiro da Silva, 700, Alfenas, Minas Gerais 37130-001, Brazil. Tel.: +55 35 988817108; E-mail: daniela.chagas@unifal-mg.edu.br

Abstract

Liquid chromatography coupled with high-resolution mass spectrometry data-independent acquisition (LC-HRMS/DIA), including MS^E, enable comprehensive metabolomics analyses though they pose challenges for data processing with automatic annotation and molecular networking (MN) implementation. This motivated the present proposal, in which we introduce DIA-IntOpenStream, a new integrated workflow combining open-source software to streamline MS^E data handling. It provides 'in-house' custom database construction, allows the conversion of raw MS^E data to a universal format (.mzML) and leverages open software (MZmine 3 and MS-DIAL) all advantages for confident annotation and effective MN data interpretation. This pipeline significantly enhances the accessibility, reliability and reproducibility of complex MS^E/DIA studies, overcoming previous limitations of proprietary software and non-universal MS data formats that restricted integrative analysis. We demonstrate the utility of DIA-IntOpenStream with two independent datasets: dataset 1 consists of new data from 60 plant extracts from the *Ocotea* genus; dataset 2 is a publicly available actinobacterial extract spiked with authentic standard for detailed comparative analysis with existing methods. This user-friendly pipeline enables broader adoption of cutting-edge MS tools and provides value to the scientific community. Overall, it holds promise for speeding up metabolite discoveries toward a more collaborative and open environment for research.

Keywords: chemical annotation; data-independent acquisition; open software; *Ocotea*; mass spectrometry

Integrative open workflow for confident annotation and molecular networking of metabolomics MS^E/DIA data

Albert Katchborian-Neto , Matheus F. Alves , Paula C.P. Bueno , Karen de Jesus Nicácio , Miller S. Ferreira ,
Tiago B. Oliveira , Henrique Barbosa , Michael Murgu , Ana C.C. de Paula Ladvocat , Danielle F. Dias , Marisi G. Soares ,
João H.G. Lago  and Daniela A. Chagas-Paula 

Corresponding author: Daniela A. Chagas-Paula, Laboratory of Phytochemistry, Medicinal Chemistry and Metabolomics, Institute of Chemistry, Federal University of Alfenas, R. Gabriel Monteiro da Silva, 700, Alfenas, Minas Gerais 37130-001, Brazil. Tel.: +55 35 988817108; E-mail: daniela.chagas@unifal-mg.edu.br

Abstract

Liquid chromatography coupled with high-resolution mass spectrometry data-independent acquisition (LC-HRMS/DIA), including MS^E, enable comprehensive metabolomics analyses though they pose challenges for data processing with automatic annotation and molecular networking (MN) implementation. This motivated the present proposal, in which we introduce DIA-IntOpenStream, a new integrated workflow combining open-source software to streamline MS^E data handling. It provides 'in-house' custom database construction, allows the conversion of raw MS^E data to a universal format (.mzML) and leverages open software (MZmine 3 and MS-DIAL) all advantages for confident annotation and effective MN data interpretation. This pipeline significantly enhances the accessibility, reliability and reproducibility of complex MS^E/DIA studies, overcoming previous limitations of proprietary software and non-universal MS data formats that restricted integrative analysis. We demonstrate the utility of DIA-IntOpenStream with two independent datasets: dataset 1 consists of new data from 60 plant extracts from the *Ocotea* genus; dataset 2 is a publicly available actinobacterial extract spiked with authentic standard for detailed comparative analysis with existing methods. This user-friendly pipeline enables broader adoption of cutting-edge MS tools and provides value to the scientific community. Overall, it holds promise for speeding up metabolite discoveries toward a more collaborative and open environment for research.

Keywords: chemical annotation; data-independent acquisition; open software; *Ocotea*; mass spectrometry

INTRODUCTION

Classical purification/isolation procedures for chemical characterization in the field of natural products (NPs) are known for their

laborious nature, involving multiple chromatographic steps and frequently afford well-known compounds. To solve this problem, more recently, chemical annotation using liquid chromatography coupled with high-resolution mass spectrometry (LC-HRMS) has

Albert Katchborian-Neto, a doctoral candidate at the Institute of Chemistry of the Federal University of Alfenas, Brazil, focuses his research on natural products by applying mass spectrometry and machine learning to metabolomics for exploring pharmacological applications with particular interest in the psychedelics science.

Matheus F. Alves, a master's student at the Institute of Chemistry of the Federal University of Alfenas, Brazil, focuses his research on mass spectrometry-based metabolomics, exploring psychedelic natural products and biodiversity bioprospecting studies.

Paula C.P. Bueno is a scientist at the Leibniz Institute of Vegetable and Ornamental Crops (IGZ), Germany, and a guest researcher at the Chemistry Institute of the Federal University of Alfenas, Brazil. Her research line is focused on plant metabolomics, ecophysiology and chemical ecology.

Karen de Jesus Nicácio, a professor at the Chemistry Department at the Federal University of Mato Grosso, Brazil, researches natural products by isolating and identifying metabolites from plants and microorganisms using classical and modern metabolomics approaches.

Miller S. Ferreira is a post-doctoral researcher at the Institute of Chemistry of the Federal University of Alfenas, Brazil. His research interests include chemometrics and machine learning techniques applied to the chemistry of natural products.

Tiago B. Oliveira is a professor at the Department of Pharmacy of the Federal University of Sergipe, Brazil. His research interests are green chemistry methods including metabolomics, biochemometrics and virtual screenings to obtain food safety.

Henrique Barbosa is a doctoral student at the Center of Natural Sciences and Humanities at the Federal University of ABC, Brazil, focusing his research on bioprospecting natural resources using computational methods.

Michael Murgu is a senior scientist at Waters Corporation, São Paulo. He has expertise in metabolite characterization, tandem mass spectrometry and high-performance liquid chromatography, to drive innovation in mass spectrometry technologies.

Ana C.C. de Paula Ladvocat is a professor at the Department of Pharmacy at the Federal University of Juiz de Fora. Her interests are mainly in stem cells for regenerative medicine, modeling diseases and assessing pharmacological properties.

Danielle F. Dias, a professor at the Institute of Chemistry of the Federal University of Alfenas, Brazil, works with the synthesis and characterization of novel bioactive compounds with analgesic, anti-inflammatory, antimicrobial and antiparasitic activities.

Marisi G. Soares, a professor at the Institute of Chemistry of the Federal University of Alfenas, Brazil, researches natural products using chromatography, nuclear magnetic resonance and mass spectrometry for structural characterization and identification.

João H.G. Lago is a professor in the Center of Natural Sciences and Humanities at the Federal University of ABC, Brazil. His research interests involve the discovery of bioactive natural products, especially those with antiparasitic, anti-inflammatory or anticancer activity.

Daniela A. Chagas-Paula, a professor at the Institute of Chemistry of the Federal University of Alfenas, Brazil, focuses her research on natural products chemistry and implementation of metabolomics, biochemometric modeling and pharmacological evaluations, with additional interests in the biochemical basis of Parkinson's disease.

Received: October 2, 2023. **Revised:** December 20, 2023. **Accepted:** January 9, 2024

© The Author(s) 2024. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

become the gold standard in the pursuit of a more rapid and efficient metabolite content assessment for either known compounds, as well as the isolation of the unknown ones [1–4].

Data-independent acquisition (DIA) is a mass spectrometry (MS) acquisition mode that systematically fragments precursor ions within a specific mass-to-charge ratio (m/z) range. It has the advantage of detecting low-abundance metabolites, which are often overlooked by conventional data-dependent acquisition (DDA) methods, due to their unavoidable loss of MS data coverage [5, 6]. MS^E, developed by Waters™ for Quadrupole Time of Flight (Q-TOF) MS analyzers, is a DIA method that fragments all precursor ions within the entire acquisition window by alternating between low- and high-collision energies, thereby obtaining consecutive scans of precursors and their fragments. This unbiased tandem MS approach is therefore considered DIA due to its unbiased fragmentation of precursors, irrespective of their abundance [7, 8]. The terms MS^{all} and all-ion fragmentation (AIF) have been also employed for a similar type of fragmentation with Orbitrap analyzer-based instruments from Thermo Fisher™ [5, 7, 9].

Despite the advancement in MS techniques, challenges persist, especially in software availability for processing data obtained through DIA methods. While there are robust options for proprietary-specific software, e.g. UNIFI (Waters™), open software options are limited. In this scenario, the MS-DIAL is an extensively used option for users of a wide variety of mass spectrometers. Other more recent approaches for processing and annotation include DIAMetAlyzer, DecoID and MetaboMSDIA, each of which has particular advantages and limitations [10–13].

In molecular networking (MN), each processed mass spectrum is represented as a node, and spectral similarities between nodes can be calculated using different algorithms such as the cosine similarity of the Global Natural Product Social Molecular Networking (GNPS) [14, 15]. Despite its potential, processing only DIA-MS data for automated annotation and MNs in metabolomics remains challenging compared to the well-established DDA workflows [14, 16]. Therefore, these challenges have motivated the DIA-IntOpenStream pipeline to be built. The present study brings novelty by offering a comprehensive pipeline for processing LC-HRMS-DIA/MS^E data that automates the generation of custom databases using free commercial software. We additionally showcased the pipeline's advantages with the successful utilization of the universal .mzML MS data format to process, annotate and generate functional MN. Finally, we focus on the current challenges in LC-HRMS-DIA/MS^E data analysis, offering strategies to mitigate these drawbacks and providing critical insights for future advancements.

This integrative approach enhances confidence in the annotation of known compounds and facilitates the discovery of novel and/or structurally related compounds. Therefore, it also enables the prioritization of unknown metabolites of interest for further investigation. Our study validates the DIA-IntOpenStream pipeline with two independent datasets. Dataset 1 consists of LC-HRMS/DIA data from 60 *Ocotea* plant extracts, showcasing the pipeline's applicability in plant metabolomics. Dataset 2 is a publicly available actinobacterial extract dataset enriched with a diverse pool of chemical authentic standards, encompassing a range of antimicrobial and naturally occurring compounds. The inclusion of known standards allows evaluation of the pipeline's annotation accuracy and efficiency. Thus, this dataset provides a solid foundation for a detailed comparative study with the original well-designed research that has performed the study using non-open software [17]. A key advantage of DIA-IntOpenStream is that it relies exclusively on open-source software. Thus, it is a

cost-effective alternative to achieve equivalent and even complementary results to the standard approaches, thereby also enabling high-quality metabolomics analysis based on LC-MS/DIA data.

RESULTS

General pipeline workflow

LC-HRMS/DIA techniques such as MS^E generate highly complex datasets that require specialized software for processing and annotation. Until recently, MN generation required vendor software or the use of non-universal MS data format (e.g. ABF from MS-DIAL), limiting the execution of integrated and fully MS^E data analyses. In contrast, DIA-IntOpenStream uses a standardized MS data format (.mzML) and open software tools for MS^E data processing and annotation. The integrated workflow provides enhanced confidence for general, automated annotation strategies and provides an accessible way to increase the reliability of metabolome annotation coverage using DIA data. Indeed, the pipeline is adaptable for any MS^E or AIF LC-MS analysis. Step 1 starts with the raw MS^E acquisition data. In step 2, MS^E raw data are converted into standard .mzML format using a Waters2mzML (<https://github.com/AnP311/Waters2mzML>), first published to GitHub in late 2022; however it is still under development and limited to Microsoft Windows operating systems. Waters2mzML implements a Python-based wrapper for ProteoWizard msConvert (<https://proteowizard.sourceforge.io/>), the most used open MS converter software. Step 3 is the Konstanz Information Miner (KNIME) workflow that can be rapidly executed, and the result is the generation of custom 'in-house' databases (DBs). These DBs are imported in the following steps 4 (MZmine 3 processing) and 5 (MS-DIAL processing). Step 3 is important for automatic enhanced annotation with level 3 of confidence during the data processing steps, leveraging the quality, processing power and annotation of both MZmine 3 and MS-DIAL 4.9 software [18, 19]. Of note, the generated KNIME database output is exported as a .csv file and subsequently imported into MZmine 3, while a .txt file is used for MS-DIAL. Furthermore, from MS-DIAL, DIA data are exported in .mgf spectra format together with the GNPS feature table (.csv). In step 6, these two files and the additional metadata are submitted using WinSCP remote server software for feature-based molecular networking (FBMN) in the GNPS platform. Step 7 applies FBMN analysis and automated annotation with level 2 of confidence. Step 8 consists of semi-automated strategies in Cytoscape software that are employed for data, inspection, visualization and integration of FBMN and in-house DB annotations.

The integration of the processed data with online MS spectral libraries allowed for the automated annotation of metabolites. Supplementation with data gathered from customized 'in-house' annotations has bolstered confidence in the annotation of the molecular families generated. The construction of tailored in-house DBs with metabolites of interest is critical to increase annotation reliability, as it provides matches with metabolites specific to a given taxon under investigation. For example, the utilization of the *Ocotea*DB and *Actinomarine*DB built with KNIME dramatically enhanced the reliability of our metabolite annotations by reducing the likelihood of potential false positives, thereby increasing true hits. Overall, strategic integration of automatic custom, automated 'in-house' annotations with online libraries and optimized GNPS parameters described in this pipeline enables MN with robust metabolite annotation of DIA/MS^E data, as schematically demonstrated in Figure 1.

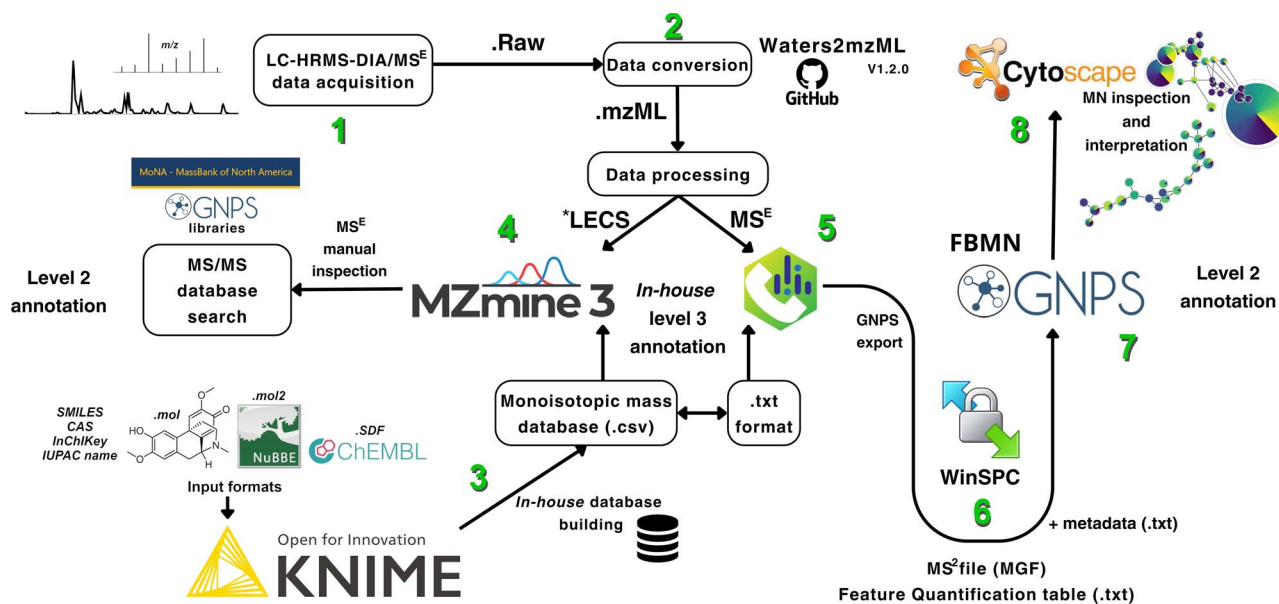


Figure 1. Schematic representation of the integrated process for generating an ‘in-house’ automatic database and performing LC-HRMS/DIA data processing to create molecular networking. After data acquisition (step 1), LC-HRMS/DIA data are converted to the standardized .mzML universal format using Waters2mzML V1.2.0 (step 2). Custom ‘in-house’ DB specific to the research is automatically prepared in the KNIME platform (step 3) using drawn chemical structures or can be downloaded from online libraries in various standard formats (.mol, .mol2, .sdf). Table formats (.csv) containing SMILES, CAS number, InChIKey, or IUPAC chemical name can also be utilized. The converted .mzML data are then imported into MZmine 3 (step 4) and MS-DIAL (step 5) for processing, where the output ‘in-house’ DB generated in KNIME is integrated to enable automatic annotation (level 3 of confidence). MS-DIAL exports the align results as a GNPS feature table (.txt) and MS² file (.mgf) that along with the custom metadata (.txt) are submitted via WinSCP remote server software to the GNPS environment (step 6). Feature-Based Molecular Networking (FBMN) analysis with automated level 2 of confidence annotation is performed on GNPS (step 7). Semi-automated strategies in Cytoscape software are employed for the data visualization and integration of FBMN and in-house DB annotations at levels 2 and 3 of confidence, respectively (step 8). *Low energy channel scans (LECS)/MS¹.

In-house database and KNIME workflow

In general, this workflow accepts the four most common types of chemical input data, namely, .mol, .mol2, .sdf and .csv. The table input files could be formatted as Simplified Molecular Input Line Entry System (SMILES), International Chemical Identifier (InChIKey), Chemical Abstracts Service (CAS) number or International Union of Pure and Applied Chemistry (IUPAC) names. The output is a .csv file with three columns: chemical structure name, calculated molecular formula and calculated monoisotopic mass. To generate the *Ocotea*DB dataset, the KNIME workflow was run with 492 molecular structures from *Ocotea* spp. in .mol format, drawn from online databases. The total runtime of the workflow was 84 s (see [Methods](#) section for desktop configuration check) and was used for later ‘in-house’ annotation during data processing. The *Actinomarine*DB dataset was generated with 6481 NPs sourced from the online npatlas database (<https://www.npatlas.org/>) in .csv format and comprised of the genera of *Actinomyces*, *Streptomyces*, *Salinospora*, *Micromonospora*, *Nocardia*, *Actinomadura* and *Rhodococcus*, running for 63 s. The *Ocotea*DB and the *Actinomarine*DB .csv files were successively uploaded into MZmine 3 for annotation. Additionally, .txt export versions can be imported into MS-DIAL for annotation. [Figure 2](#) illustrates the reader, converter and writer nodes.

LC-HRMS data conversion and processing

Dataset 1 along with quality controls (QCs) and blanks in the Waters™ .raw format was effectively converted with Waters2mzML to generate functional centroided .mzML files. The same step was performed for dataset 2. These conversions took ~36 and 1.5 h, respectively, on our computer configuration (detailed in the [Methods](#) section). The files were then processed

using the MZmine 3 and MS-DIAL 4.9. For MZmine 3, despite the large cohort of dataset 1, final batch processing required only ~7 min per ionization mode, while dataset 2 took only 1 min. While actual processing is quite fast (a few minutes), software parameter optimization is time-demanding, although empowers robust data processing for complex samples. Dataset 1 yielded 18 805 aligned features in the positive mode, including 3983 annotation hits from *Ocotea*DB with all potential adducts identified. Similarly, the negative mode yielded 23 304 features with 3216 database annotation hits.

In contrast, the positive mode analysis with MS-DIAL 4.9 required ~2.11 h for dataset 1, resulting in the annotation of 22 572 features, while the analysis from the negative mode analysis took around 1.58 h, yielding 21 838 features. For dataset 2, MS-DIAL 4.9 processing required only 4 min. The dataset acquisition parameters and data size have a great influence on processing time, especially aligning an elevated number of samples, as in the case of dataset 1. In addition, despite inherent variations in parameters and algorithms employed by the two programs, the results generated were comparable. More specifically, MS-DIAL exhibited a longer processing time as it can appropriately process and assign MS² fragment ions to MS¹ precursor ions in DIA data. This step is the slowest during data processing and is particularly mandatory for MN implementation. Details of the dataset 2 processing results are provided at SM-4 and 5.

Although DIA processing algorithms are present in MZmine 3, they were not employed in this pipeline as they remain in an experimental phase without publicly available guides or tutorials to standardize parameter values, different from DDA data processing, which is very well established. As such, we used MZmine only to perform MS¹ data processing, which explains

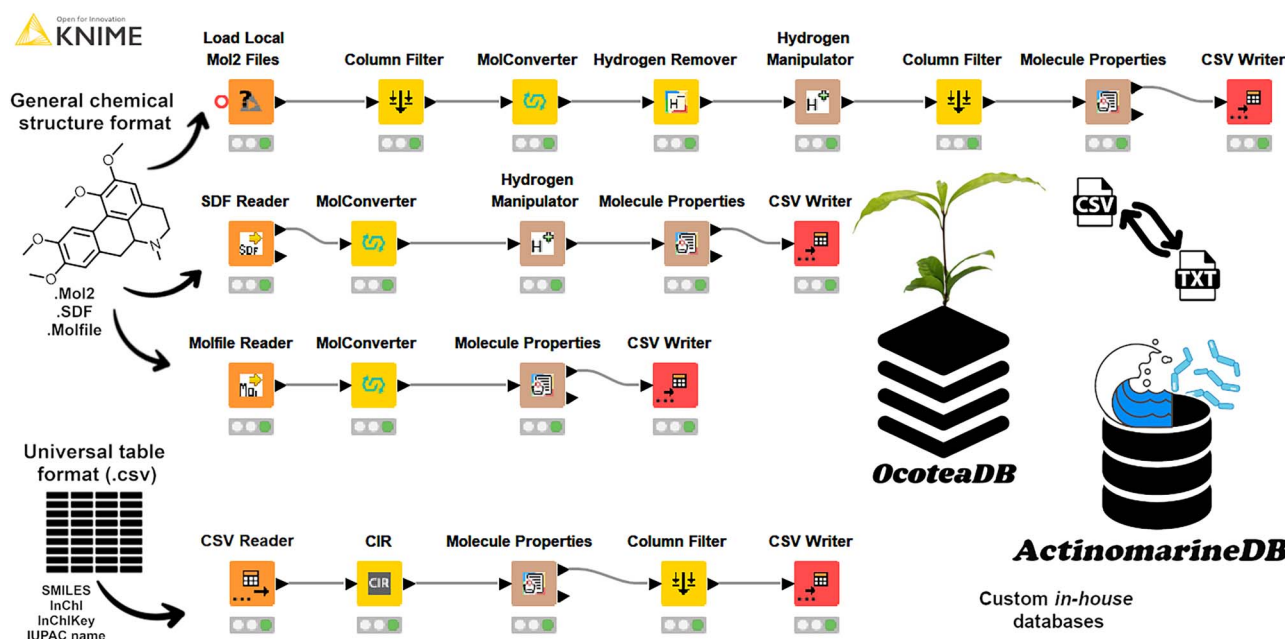


Figure 2. KNIME workflow for high confidence in-house automatic database assembly. The workflow includes .mol, .mol2 and .sdf chemical structures as input data for the respective node readers. The last node is a .csv table reader that accepts tables containing either SMILES, CAS numbers, InChIKey or IUPAC names as input data. The final output is a .csv (or .txt) file with three columns: chemical name, calculated molecular formula and calculated monoisotopic mass. The in-house DB specific for the aiming samples can now be integrated into the data processing step to increase confidence in the metabolite annotation.

the increased speed of data processing results when compared to MS-DIAL. Nevertheless, despite this limitation, we highlight that the MZmine 3 demonstrates remarkable transparency and guidance in data processing and annotation. Further insights and considerations are further provided in the Discussion section.

Chemical annotation

Following established guidelines for untargeted metabolomics, QC samples were prepared for dataset 1. After data analyses, consistent peak distributions and reproducible metabolic fingerprints across the QC injections were observed (Figures 3 and S1). Using QC samples to ensure that predefined quality thresholds are met is a critical step, thereby validating that the analytical system can acquire high-quality metabolomics data from the experimental samples [20, 21]. The developed LC-HRMS/MS^E method effectively separated and detected major and minor metabolite components in *Ocotea* spp. The comprehensive list of annotations with level 2 of confidence for the acquired metabolic fingerprint is shown in Table 1. Even with the complexity of our metabolomics data, careful and rigorous data processing led to reliable and clear reproducible results, as shown by the superimposed chromatograms of experimental replicates (Figure S2). The superimposed chromatograms of extract samples and QCs from both positive and negative modes are illustrated in Figures 3 and S3–S5, evidencing the high chemical complexity of data. Details of dataset 2 are provided in Figures S16–S18 and Tables S3 and S4.

For dataset 1, alkaloids, lignoids and flavonoids were the main annotated classes in the *Ocotea* spp. samples (Figure 3). In the positive mode, high-intensity levels were observed for an amino acid derivative 4-hydroxy-*N*-methylproline (**1**, m/z 146.0812— R_t 0.49 min), the morphinan alkaloid isomers of flavinantine (**9**, m/z 328.1546— R_t 1.75 min), the aporphine alkaloid isoboldine (**11**, m/z 328.1546— R_t 1.96 min), also some lignoids, including the ocophyllol B (**43**, m/z 359.1856— R_t 5.38 min) and flavonoids such as apigenin-7-*O*-rutinoside (**61**, m/z 577.1569— R_t 2.75 min)

and the kaempferol 3-(2'',4''-di-(*E*)-*p*-coumaryl)-rhamnoside (**66**, m/z 723.1728— R_t 5.46 min). Alkaloids and lignoids were ionized mainly in the positive mode, while flavonoids were annotated in the negative ionization mode.

Aporphines and benzyloquinoline alkaloids were annotated as major compounds, all of which have known fragmentation patterns and were typically distinguished from each other by neutral losses (Figure 4). Fragmentation patterns of the morphinandienones and phenanthrenes alkaloids, as well as the NP subclasses of lignoids and flavonoids, were also demonstrated (Figures 5 and 6). Automated integration of 'in-house' DB by matching MS¹ monoisotopic masses enabled manual examination of MS^E spectra using MZmine 3, leading to level 2 confidence identification of 66 specialized metabolites. No public mass spectral data were available for 15 annotated metabolites, so fragmentation was proposed based on relevant literature and chemical knowledge. Thus, for these metabolites, level 2 of confidence was given based on the classification by Schymanski et al. [22]. Diagnostic ion fragments were searched on the Mass Bank of North America (MoNA) and GNPS libraries. In parallel, automated annotation of MS^E spectra via MS-DIAL-GNPS-FBMN identified 155 NPs. A comparison between the manual and automated annotations revealed just 18 shared annotations (Table 1). The complete observed product ions and GNPS annotations are detailed in Table S1 and online (<https://zenodo.org/records/10383866>), respectively.

Gas-phase fragmentation reactions

The key distinction observed for alkaloids was based on the *N*-substitution pattern. Norapomorphines showed a 14.01 Da mass reduction versus aporphines due to the presence of a radical hydrogen instead of a methyl *N*-substitution. This allowed a distinction between the two alkaloid subclasses. Prevalent neutral losses were 17.03 Da (NH₃) and 31.01 Da (CH₃NH₂) from isoquinoline ring opening (Figure 4). Additionally, losses of CH₃OH

Table 1: ESI-MS^E positive and negative modes including annotation with level 2 of confidence of *Ocotea* metabolites from based on the in-house database. This table includes 41 alkaloids (pyrrolidine, proaporphine, noraporphine, aporphine, benzyloquinoline, morphinandienone and protoberberine subclasses), 6 lignoids (1 lignan and 5 neolignans), 18 flavonoids (glycosylated quercetin, kaempferol and apigenin derivative subclasses) and a cyclic polyol. The last two columns indicate the MS² spectral source used for manual annotation and if spectra were matched automatically on the GNPS platform, respectively

ID	R _t (min)	Putative metabolite	Metabolite class	MF	Ocotea DB	Observed m/z	Adduct	Error (ppm)	Spectral references	Automatic GNPS annotation
1	0.49	4-Hydroxy-N-methylproline	Pyrrolidine alkaloid	C ₆ H ₁₁ NO ₃	Yes	146.08119	[M + H] ⁺	0.14	Proposed	No
2	1.45	Crotsparine	Proaporphine alkaloid	C ₁₇ H ₁₇ NO ₃	Yes	284.12770	[M + H] ⁺	-1.48	GNPS	Yes
3	1.48	Glaziovine	Proaporphine alkaloid	C ₁₈ H ₁₉ NO ₃	Yes	298.14359	[M + H] ⁺	-0.60	GNPS	No
4	1.52	N-Methylcoclaurine	Benzyloquinoline alkaloid	C ₁₈ H ₂₁ NO ₃	Yes	300.15932	[M + H] ⁺	-0.33	GNPS	No
5	1.60	3-Hydroxynormuciferine	Noraporphine alkaloid	C ₁₈ H ₁₉ NO ₃	Yes	298.14355	[M + H] ⁺	-0.74	GNPS	Yes
6	1.61	Laurelliptine	Noraporphine alkaloid	C ₁₈ H ₁₉ NO ₄	Yes	314.13850	[M + H] ⁺	-0.57	Proposed	No
7	1.74	Lauroitsine	Noraporphine alkaloid	C ₁₈ H ₁₉ NO ₄	Yes	314.13853	[M + H] ⁺	-0.48	GNPS	No
8	1.63	Pallidine	Morphinandienone alkaloid	C ₁₉ H ₂₁ NO ₄	Yes	328.15444	[M + H] ⁺	0.34	GNPS	No
9	1.81	Flavinantine	Morphinandienone alkaloid	C ₁₉ H ₂₁ NO ₄	Yes	328.15448	[M + H] ⁺	0.46	Proposed	No
10	1.87	Boldine	Aporphine alkaloid	C ₁₉ H ₂₁ NO ₄	Yes	328.15447	[M + H] ⁺	0.43	GNPS	No
11	1.96	Isoboldine	Aporphine alkaloid	C ₁₉ H ₂₁ NO ₄	Yes	328.15440	[M + H] ⁺	0.21	GNPS	Yes
12	2.05	Corytuberine	Aporphine alkaloid	C ₁₉ H ₂₁ NO ₄	Yes	328.15438	[M + H] ⁺	0.15	GNPS	No
13	2.00	Lauroschoztzine	Aporphine alkaloid	C ₂₀ H ₂₃ NO ₄	Yes	342.17014	[M + H] ⁺	0.47	GNPS	Yes
14	2.07	Reticuline	Benzyloquinoline alkaloid	C ₁₉ H ₂₃ NO ₄	Yes	330.16918	[M + H] ⁺	-2.42	MoNA	Yes
15	2.19	Armepavine	Benzyloquinoline alkaloid	C ₁₉ H ₂₃ NO ₃	Yes	314.17505	[M + H] ⁺	-0.06	GNPS	No
16	1.93	Zenkerine	Noraporphine alkaloid	C ₁₈ H ₁₉ NO ₃	Yes	298.14367	[M + H] ⁺	-0.34	GNPS	No
17	2.08	Tuduranine	Noraporphine alkaloid	C ₁₈ H ₁₉ NO ₃	No	298.14343	[M + H] ⁺	-1.14	Proposed	No
18	2.08	Diospirifoline	Aporphine alkaloid	C ₁₉ H ₁₉ NO ₄	Yes	326.13859	[M + H] ⁺	-0.28	Proposed	No
19	2.27	Thaliporphine	Aporphine alkaloid	C ₂₀ H ₂₃ NO ₄	Yes	342.16985	[M + H] ⁺	-0.38	GNPS	No
20	2.43	Predicentrine	Aporphine alkaloid	C ₂₀ H ₂₃ NO ₄	Yes	342.17007	[M + H] ⁺	0.26	MoNA	No
21	2.47	Nuciferine	Aporphine alkaloid	C ₁₉ H ₂₁ NO ₂	No	296.16469	[M + H] ⁺	0.61	GNPS	No
22	2.54	Corydine	Aporphine alkaloid	C ₂₀ H ₂₃ NO ₄	Yes	342.17042	[M + H] ⁺	1.29	GNPS	Yes
23	2.50	Domesticine	Aporphine alkaloid	C ₁₉ H ₁₉ NO ₄	Yes	326.13834	[M + H] ⁺	-1.04	GNPS	No
24	2.98	Dehydrodicentrine	Aporphine alkaloid	C ₂₀ H ₂₁ NO ₄	Yes	338.13891	[M + H] ⁺	0.68	Proposed	No
25	2.46	Norisocorydine	Noraporphine alkaloid	C ₁₉ H ₂₁ NO ₄	Yes	328.15421	[M + H] ⁺	-0.37	GNPS	No
26	2.56	Lauroretanine	Noraporphine alkaloid	C ₁₉ H ₂₁ NO ₄	Yes	328.15429	[M + H] ⁺	-0.12	GNPS	No
27	2.79	Nordicentrine	Noraporphine alkaloid	C ₁₉ H ₂₁ NO ₄	Yes	326.13862	[M + H] ⁺	-0.18	GNPS	No
28	3.01	Nornantenine	Noraporphine alkaloid	C ₁₉ H ₁₉ NO ₄	Yes	326.13864	[M + H] ⁺	-0.12	GNPS	Yes
29	3.08	Nornuciferine	Noraporphine alkaloid	C ₁₈ H ₁₉ NO ₂	Yes	282.14870	[M + H] ⁺	-0.57	GNPS	No
30	2.69	Lirinidine	Aporphine alkaloid	C ₁₈ H ₁₉ NO ₂	Yes	282.14850	[M + H] ⁺	-1.28	GNPS	Yes
31	2.92	Glaucine	Aporphine alkaloid	C ₂₁ H ₂₅ NO ₄	Yes	356.18513	[M + H] ⁺	-1.40	GNPS	Yes
32	3.07	Roemerine	Aporphine alkaloid	C ₁₈ H ₁₇ NO ₂	Yes	280.13274	[M + H] ⁺	-1.68	GNPS	Yes
33	2.87	Nantenine	Aporphine alkaloid	C ₂₀ H ₂₁ NO ₄	Yes	340.15406	[M + H] ⁺	-0.79	MoNA	No
34	3.07	Dicentrine	Aporphine alkaloid	C ₂₀ H ₂₁ NO ₄	Yes	340.15426	[M + H] ⁺	-0.21	GNPS	No
35	3.15	Dehydronuciferine	Aporphine alkaloid	C ₁₉ H ₁₉ NO ₂	No	294.14903	[M + H] ⁺	0.58	Proposed	No
36	3.31	Dicentrinone	Oxoaporphine alkaloid	C ₁₉ H ₁₉ NO ₅	Yes	336.08626	[M + H] ⁺	-0.86	GNPS	No
37	3.52	Leucoxyllonine	Aporphine alkaloid	C ₂₂ H ₂₅ NO ₆	Yes	400.17563	[M + H] ⁺	0.42	Proposed	No
38	3.05	Stephananthrine	Phenanthrene alkaloid	C ₁₉ H ₁₉ NO ₂	No	294.14881	[M + H] ⁺	-0.17	Proposed	No
39	3.19	Argentinine	Phenanthrene alkaloid	C ₁₉ H ₂₁ NO ₂	Yes	296.16515	[M + H] ⁺	2.16	GNPS	Yes
40	3.66	Thalictuberine	Phenanthrene alkaloid	C ₂₁ H ₂₃ NO ₄	Yes	354.16966	[M + H] ⁺	-0.90	GNPS	No
41	3.86	Discretamine	Protoberberine alkaloids	C ₁₉ H ₂₁ NO ₄	Yes	328.15342	[M + H] ⁺	-2.77	GNPS	No
42	3.27	Sesamin	Lignan	C ₂₀ H ₁₈ O ₆	Yes	355.11784	[M + H] ⁺	0.65	MoNA	No
43	5.38	Ocophyllol B	Neolignan	C ₂₁ H ₂₆ O ₅	Yes	359.18565	[M + H] ⁺	0.97	Proposed	No
44	5.45	Eusiderin	Neolignan	C ₂₂ H ₂₆ O ₆	Yes	387.17956	[M + H] ⁺	-1.70	Proposed	No
45	5.99	Licarin B	Neolignan	C ₂₀ H ₂₀ O ₄	Yes	325.14306	[M + H] ⁺	-1.17	Proposed	No
46	6.30	Licarin A	Neolignan	C ₂₀ H ₂₂ O ₄	Yes	327.15865	[M + H] ⁺	-1.34	MoNA	No
47	6.34	Armenin B	Neolignan	C ₂₁ H ₂₄ O ₆	Yes	373.16390	[M + H] ⁺	-1.77	Proposed	No
48	0.56	Quinic acid	Cyclic polyol	C ₇ H ₁₂ O ₆	Yes	191.05433	[M-H] ⁻	-9.32	MoNA	No
49	1.75	Taxifolin	Flavanonol	C ₁₅ H ₁₂ O ₇	Yes	303.04983	[M-H] ⁻	-3.96	GNPS	No
50	2.07	Catechin/ Epicatechin	Flavonol	C ₁₅ H ₁₄ O ₆	Yes	289.07091	[M-H] ⁻	-2.94	MoNA	No
51	2.40	Isoquercitrin	Glycosylated flavone	C ₂₁ H ₂₀ O ₁₂	Yes	463.08689	[M-H] ⁻	-2.83	MoNA	Yes
52	2.42	Vitexin-2'-O-rhamnoside	Glycosylated flavone	C ₂₇ H ₃₀ O ₁₄	Yes	577.15701	[M-H] ⁻	1.26	GNPS	No
53	2.46	Rutin	Glycosylated flavone	C ₂₇ H ₃₀ O ₁₆	No	609.14699	[M-H] ⁻	1.44	GNPS	Yes

(Continued)

Table 1: Continued

ID	R _t (min)	Putative metabolite	Metabolite class	MF	Ocotea DB	Observed m/z	Adduct	Error (ppm)	Spectral references	Automatic GNPS annotation
54	2.54	Quercimeritrin	Glycosylated flavone	C ₂₁ H ₂₀ O ₁₂	Yes	463.08762	[M-H] ⁻	-1.25	Proposed	No
55	2.54	Vitexin	Glycosylated flavone	C ₂₁ H ₂₀ O ₁₀	Yes	431.09731	[M-H] ⁻	-2.46	GNPS	No
56	2.66	Quercitrin	Glycosylated flavone	C ₂₁ H ₂₀ O ₁₁	Yes	447.09299	[M-H] ⁻	-0.67	GNPS	Yes
57	2.73	Reynoutrin	Glycosylated flavone	C ₂₀ H ₁₈ O ₁₁	Yes	433.07686	[M-H] ⁻	-1.71	MoNA	Yes
58	2.74	Astragalin	Glycosylated flavone	C ₂₁ H ₂₀ O ₁₁	Yes	447.09298	[M-H] ⁻	-0.69	MoNA	No
59	2.74	Apigenin-7-O-rutinoside	Glycosylated flavone	C ₂₇ H ₃₀ O ₁₄	Yes	577.15675	[M-H] ⁻	0.81	MoNA	Yes
60	2.85	Schaftoside/ Isoschaftoside	Glycosylated flavone	C ₂₆ H ₂₈ O ₁₄	Yes	563.14164	[M-H] ⁻	1.79	GNPS	Yes
61	3.15	Afzelin	Glycosylated flavone	C ₂₁ H ₂₀ O ₁₀	Yes	431.09706	[M-H] ⁻	-3.04	MoNA	No
62	4.46	Kaempferol 3-4''-p-coumaryl- rhamnoside	Glycosylated flavone	C ₃₀ H ₂₆ O ₁₃	Yes	577.13474	[M-H] ⁻	-0.71	Proposed	No
63	3.57	Quercetin	Flavonol	C ₁₅ H ₁₀ O ₇	Yes	301.03286	[M-H] ⁻	-8.37	MoNA	No
64	4.04	Apigenin	Glycosylated flavone	C ₁₅ H ₁₀ O ₅	Yes	269.04407	[M-H] ⁻	-5.32	GNPS	No
65	4.10	Kaempferol	Flavonol	C ₁₅ H ₁₀ O ₆	Yes	285.03873	[M-H] ⁻	-6.07	GNPS	No
66	5.46	Kaempferol 3-(2'',4''- di-(E)-p-coumaryl- rhamnoside)/ Kaempferol 3-(3'',4''- di-(E)-p-coumaryl- rhamnoside)	Glycosylated flavone	C ₃₉ H ₃₂ O ₁₄	Yes	723.17282	[M-H] ⁻	1.23	GNPS	Yes

The IDs with the respective ion fragments observed on MS^E spectra are detailed in Table S1. Chemical structures are provided in Figures S8–S12. Spectral matching was done manually for all IDs using online MS reference spectra libraries (MoNA and GNPS). The proposed fragmentations were based on the literature with diagnostic ions for the annotated metabolite classes [23–27].

(32.03 Da) occurred due to adjacent hydroxyl and methoxy groups in aporphine rings followed by neutral CO loss (27.99 Da). Fragmentation patterns of some of the less common alkaloids found in *Ocotea* spp., including benzylisoquinolines, morphinandienones and phenanthrenes, are also depicted in Figure 5. This reveals some shared and distinctive fragmentation patterns among the diversity of the annotated alkaloids, as explained in ST-1. Fragmentation proposals were based on chemical knowledge and supported by the literature [23–27].

Flavonoids and lignoids displayed characteristic neutral losses and fragment ions as well (Figure 6). The fragmentation of lignoids was evidenced by neutral losses of methyl (14.01 Da), methoxy (32.03 Da), retro-Diels-Alder reactions and aromatic ring cleavages. These fragmentations formed diagnostic ions that allowed the differentiation of bicyclo neolignans and benzofuran lignoids. For flavonoids, fragmentation predominantly involved glycosidic bond cleavages and losses of saccharide units. These included losses of pentoses (132.04 Da), deoxyhexoses (146.05 Da), hexoses (162.05 Da), glucuronic acids (176.03 Da) and rutinoses (308.09 Da), giving characteristic product ions. These neutral losses provided clues to the types of glycosylation present on the flavonoid scaffolds. Key diagnostic ions for flavonoid aglycones allowed differentiation between subclasses such as apigenin, quercetin and kaempferol. Overall, these characteristic fragmentation patterns allowed differentiation between the main flavonoid subclasses present in *Ocotea* species (ST-1).

FBMN

Regarding the FBMN jobs with GNPS, positive mode analysis required ~5 h, whereas the negative mode took ~6 h. These analyses resulted in the generation of highly complex metabolic networks (Figures S6 and S7). Besides overall complexity, MN revealed intricate cluster families in the metabolome of *Ocotea* spp., which could be individually analyzed to get deeper

information. In addition, to extract more nuanced insights, LC-HRMS-DIA/MS^E data were reprocessed with higher amplitude cutoffs (e.g. 50 000 counts). The FBMN jobs from both datasets required less than 15 min to finish. The simplified MNs generated from the reprocessed data aided in the visualization and identification of key *Ocotea* spp. molecular families (Figure 8) and actinobacterial MNs (Figure 9).

Spectra were queried against GNPS libraries related to our dataset (e.g. IQAMDB and NIH NPs for positive and negative mode, respectively) and a complete list of matches is listed in Tables 1 and S1 and online at the Zenodo open digital library (<https://zenodo.org/records/10383866>). The FBMN analysis revealed distinct families of aporphine and benzylisoquinoline alkaloids (positive mode), alongside predominately O-glycosylated flavonoids (negative mode) across the *Ocotea* spp. (Figure 8). The pie charts illustrate relative metabolite abundance across 60 *Ocotea* species based on MS¹ precursor ion areas. Visual inspection of a positive mode alkaloid cluster shows 3-hydroxynornuciferine as highly abundant but specific to only a few *Ocotea* species, while glaucine to only a few others. Reticuline appears conserved across most *Ocotea* spp., suggesting a potential genus chemomarker. N-methylcoclaurine also arises broadly present, but with reduced abundance in this cluster family.

All of the spectral matches represented were thoroughly inspected to check their annotation and spectral similarity accuracy. Regarding the MN in negative mode, the highlighted family of glycosylated flavonoids is demonstrated, with the main metabolites across the highlighted cluster family including quercetin-3-O-rhamnoside (quercitrin), quercetin-3-O-galactoside (quercimetrin), quercetin-3-D-xyloside (reynoutrin), kaempferol-3-O-rhamnoside (afzelin), kaempferol-3-O-glucoside (astragalin) and kaempferol-3-O-arabinoside. The array of quercetin and kaempferol glycosides shows widespread distribution across the *Ocotea* spp. dataset. The broad interspecies

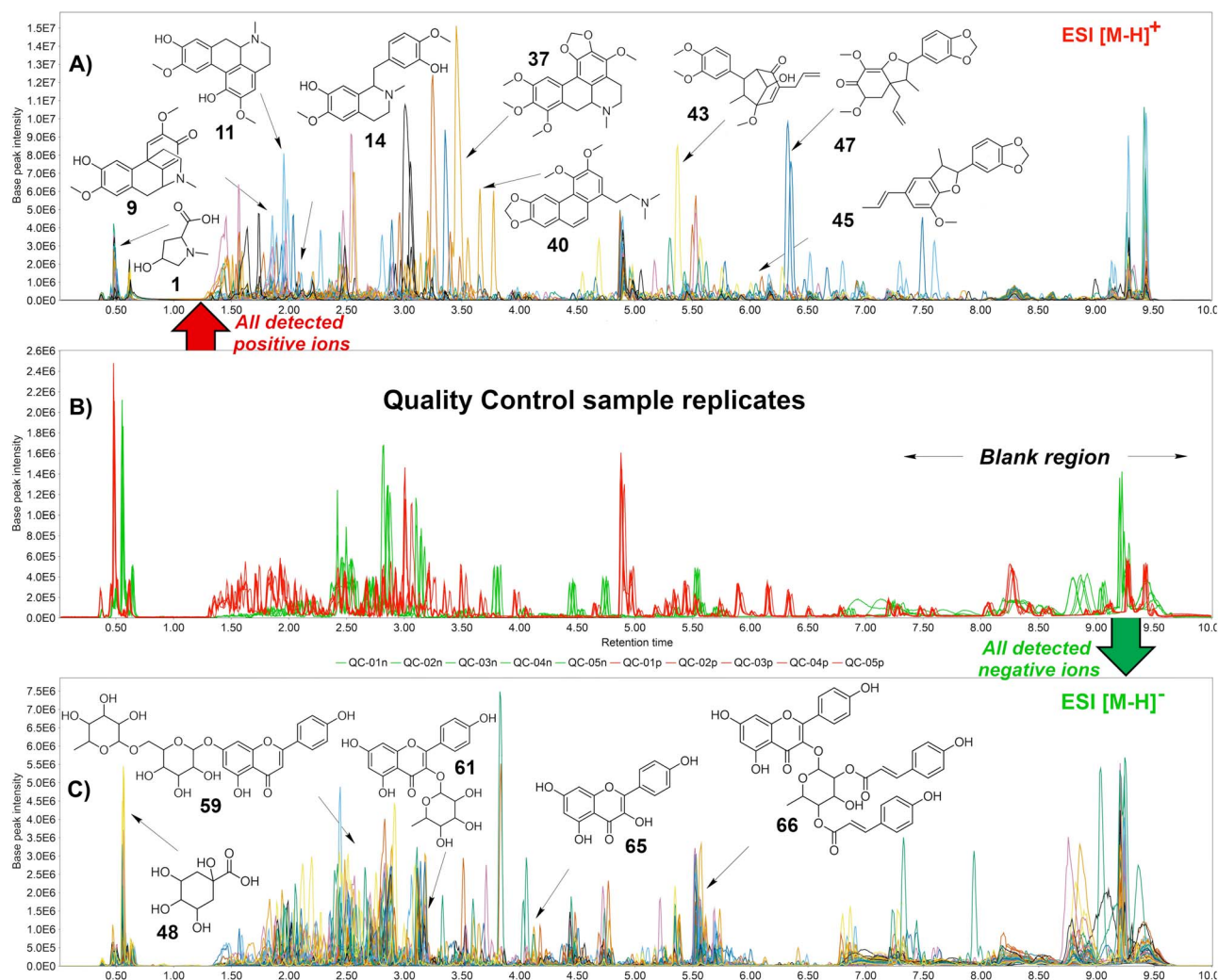


Figure 3. LC-HRMS-DIA/MS^E metabolic fingerprints displayed as base peak ions (BPIs) showing the overlapped QC replicates (pool of 60 *Ocotea* spp. extracts) and overlapped individual chromatograms. **(A)** The overlapped metabolic fingerprint of *Ocotea* species in positive ionization mode, including the amino acid derivative 4-hydroxy-*N*-methylproline (1), the alkaloids, flavinantine (9), isoboldine (11), leucoxylophine (37), thalictuberine (40), as well as the lignoids ocophyllol B (43), licarin B (45) and armenin B (47). **(B)** Pooled, overlapped QC samples, highlighting the blank region. Positive mode is displayed in red and negative in green. Peaks that are found in common among blanks are located after 7 min. **(C)** The overlapped metabolic fingerprint of *Ocotea* species in negative ionization mode, and representative annotation of some compounds, such as quinic acid (48) and the flavonoids apigenin-7-*O*-rutinoside (59), afzelin (61) and kaempferol 3-(2',4'-di-(*E*)-*p*-coumaryl-rhamnoside (66).

conservation of these flavonoid metabolites suggests the presence of essential, shared biosynthetic pathways within the *Ocotea* genus.

DISCUSSION

A core challenge in omics fields, including metabolomics, is the conversion and processing of DIA data (e.g. MS^E) compared to traditional DDA workflows. In DDA, preselected precursors are fragmented, enabling straightforward data conversion and processing by most open tools. However, DDA induces significant losses in spectral data coverage, varying with the sensitivity of the instrument and, in relation, the cycle time of the method, because it selects only ions above a certain cut-off area or intensity for fragmentation [5, 28]. In contrast, MS^E methods fragment all ions without any previous precursor ion selection, generating complex but unbiased spectra with a more complete metabolomic data coverage [7]. The lack of predefined precursors in MS^E means that fragment-precursor relationships must be reconstructed post-acquisition through computational

deconvolution approaches, which can be performed using different algorithms [5–7, 29]. This is more challenging compared to the inherent precursor-to-fragment associations made with DDA methods. Herein, we present an integrative strategy to leverage MS^E data in the community standard .mzML format using only freely available software and platforms. This study demonstrates the power of MS^E with accessible tools while highlighting current, ongoing challenges in data conversion, processing and interpretation.

This integrated pipeline provides an efficient and customizable solution for extracting the most biological information from MS^E data. To date, this is the first mention of the Waters2mzML in an applicability case, which is a recently introduced tool specifically designed to address the challenges associated with converting and centroiding Waters MS^E data, without the need to use vendor software such as UNIFI, Symphony or Progenesis QI. Waters2mzML is a simple tool that ensures compatibility and offers independence to convert raw Waters MS^E data into a more widely used format. This open tool can correctly reassign MS²-level data to MS^E MS/MS scans. Therefore, it is now possible to

ALKALOIDS

APORPHINE

NORAPORPHINE

PROAPORPHINE

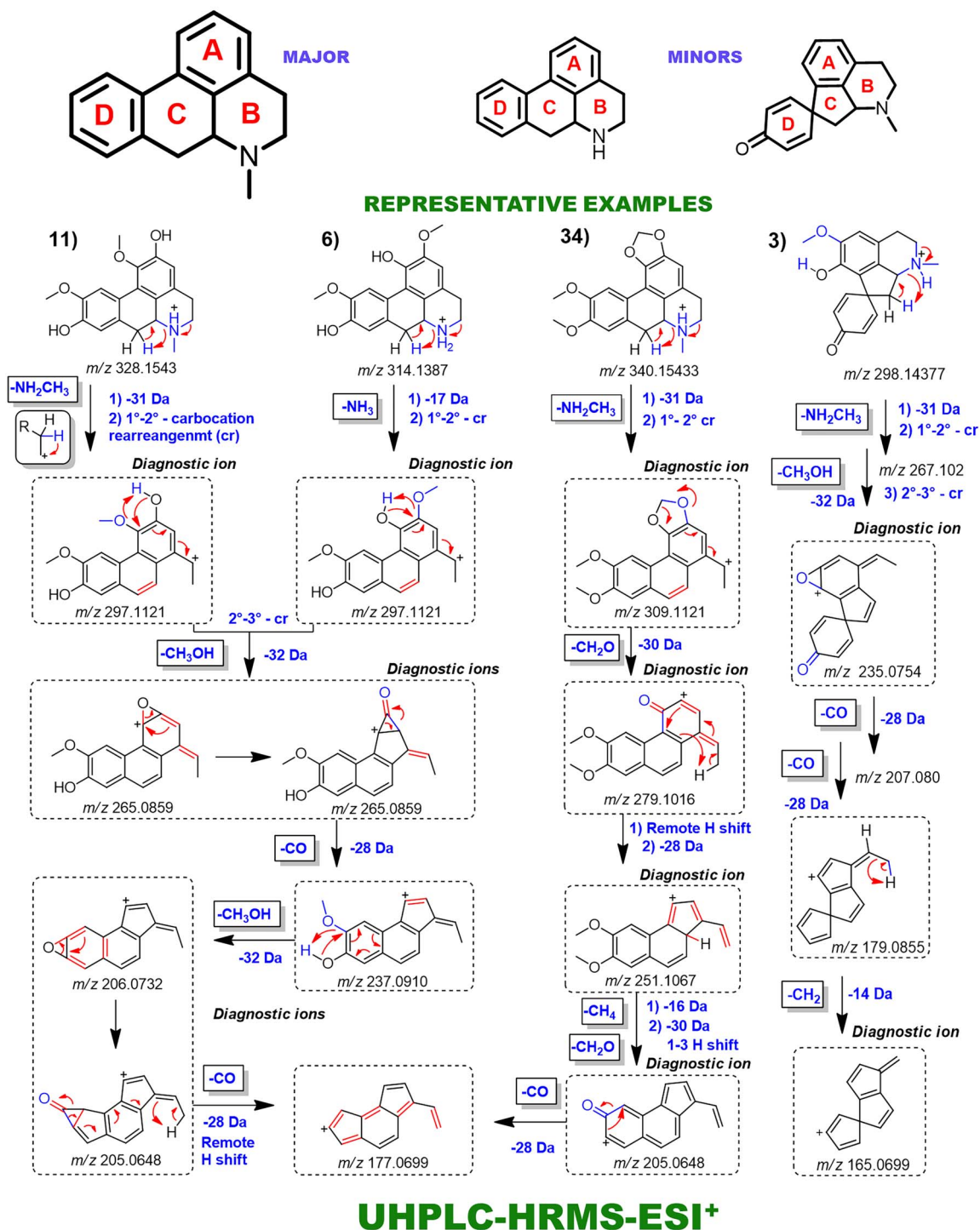


Figure 4. Main fragmentation pattern of proaporphine, noraporphine and aporphine alkaloids, along with their key diagnostic ions. Representative examples are demonstrated: proaporphines: 3—glaziiovine. Norapomorphines: 6—laurelliptine and aporphines: 10—boldine.

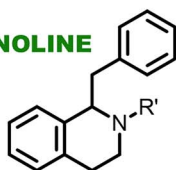
freely generate standard and functional .mzML spectra from MS^E data for further integrated downstream metabolomics analyses.

Moreover, this pipeline includes the ability to generate a custom, 'in-house' database through a KNIME workflow, enhancing its utility. This database integration into LC-HRMS- DIA/MS^E data

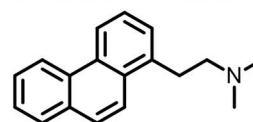
processing significantly improves metabolite annotation of MNs when using MZmine and MS-DIAL. In our case studies, our focus was to enhance the annotation of metabolites specifically from *Ocotea* plants and marine actinobacteria datasets. The creation and incorporation of *Ocotea*DB and *Actinomarine*DB provided a

ALKALOIDS

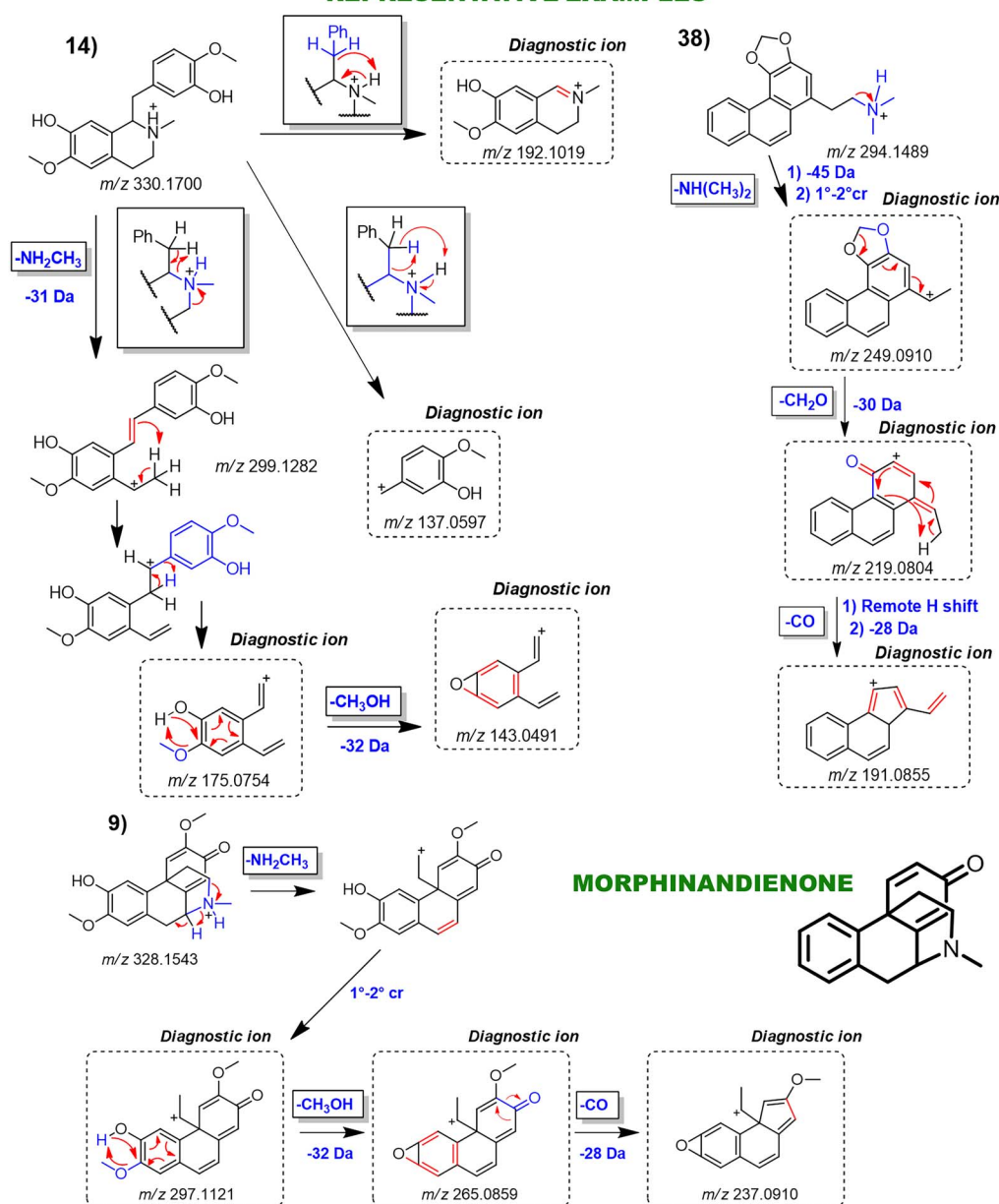
BENZYLISOQUINOLINE



PHENANTHRENE



REPRESENTATIVE EXAMPLES



UHPLC-HRMS-ESI⁺

Figure 5. Main fragmentation pattern of benzylisoquinoline, phenanthrene and morphinan alkaloids, along with their key diagnostic ions. Representative examples are demonstrated: Benzylisoquinoline: 14—reticuline. Morphinandiенone: 9—flavinantine. Phenanthrene: 38—stephananthrine.

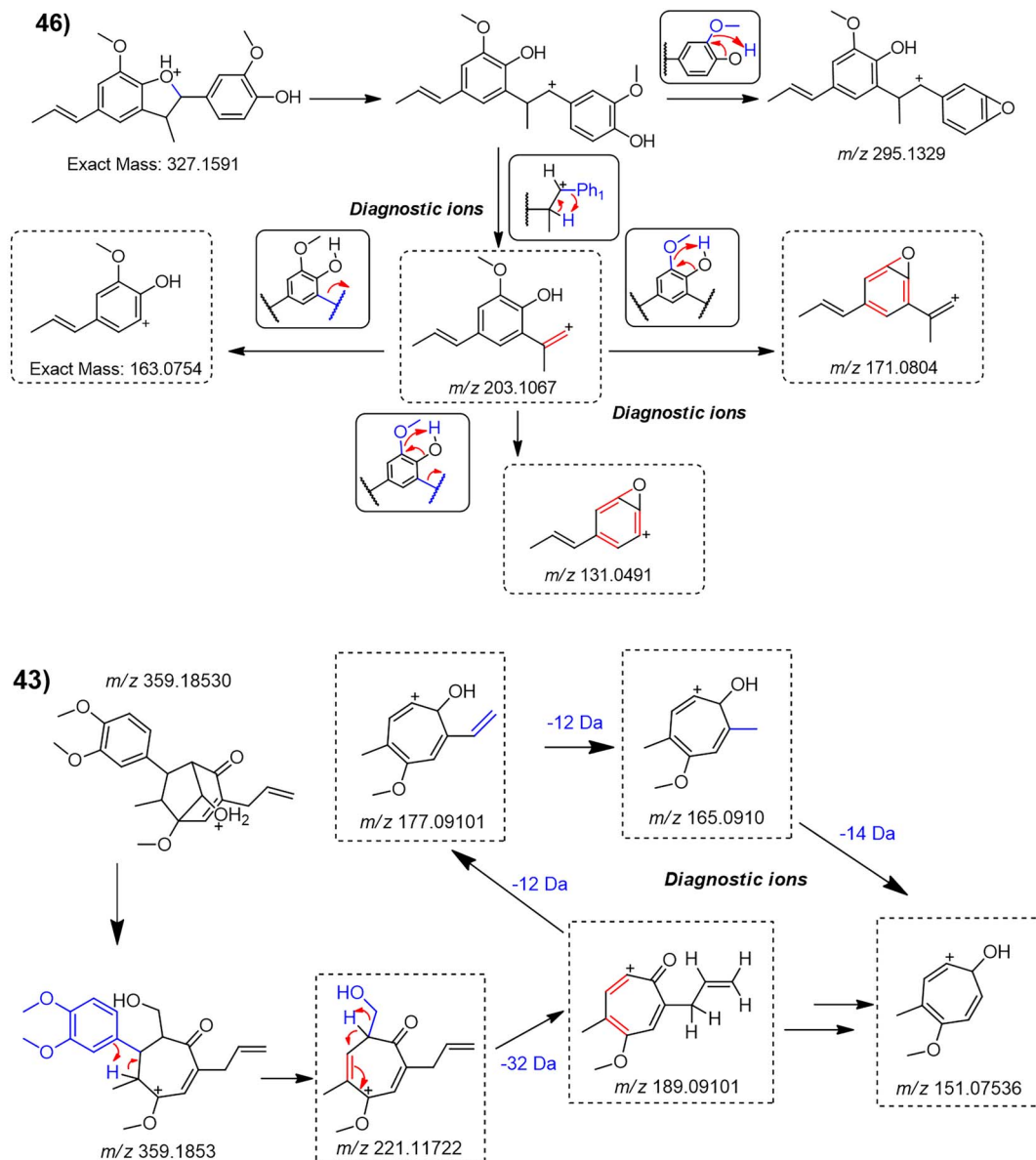
tailored metabolite reference library to complement the capabilities of the open-source software tools. This pipeline has demonstrated efficacy and accuracy, enabling streamlined annotations at confidence levels 2 and 3 according to Metabolomics Standards Initiative (MSI) guidelines [30, 31].

Regarding data processing, the export functionality of MS-DIAL allows DIA data to be cosine matched with high-quality spectral libraries on the GNPS platform, also allowing integration with any GNPS tools for enhanced data interpretation, including FBMN, which was used in this workflow. It also allows other GNPS

LIGNOIDS



REPRESENTATIVE EXAMPLES



UHPLC-HRMS-ESI⁺

Figure 6. Main fragmentation pattern of lignoids and their key diagnostic ions. Representative examples are demonstrated: Lignoids: 43—ocophyllol B and 46—licarin A.

analyses, such as MS2LDA, Network Annotation Propagation (NAP) and MolNetEnhancer, which are already well-implemented for DDA data in the GNPS platform [32–34]. MS-DIAL software therefore enables the full processing of MS^E spectra with correct

GNPS export in a generic file format (.mgf). Importantly, processed data from MS-DIAL can be employed for spectral similarity searches through a range of different algorithms and tools. The strength of MS-DIAL lies in its robust algorithm, MS²dec,

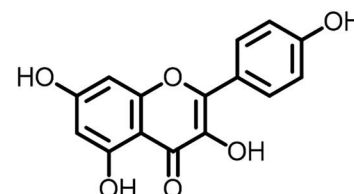
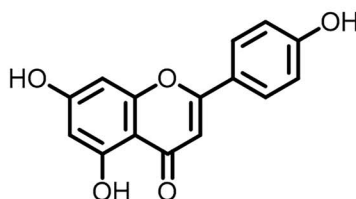
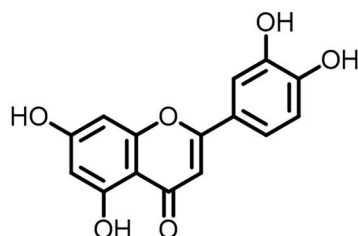
FLAVONOIDS

MAIN BACKBONES

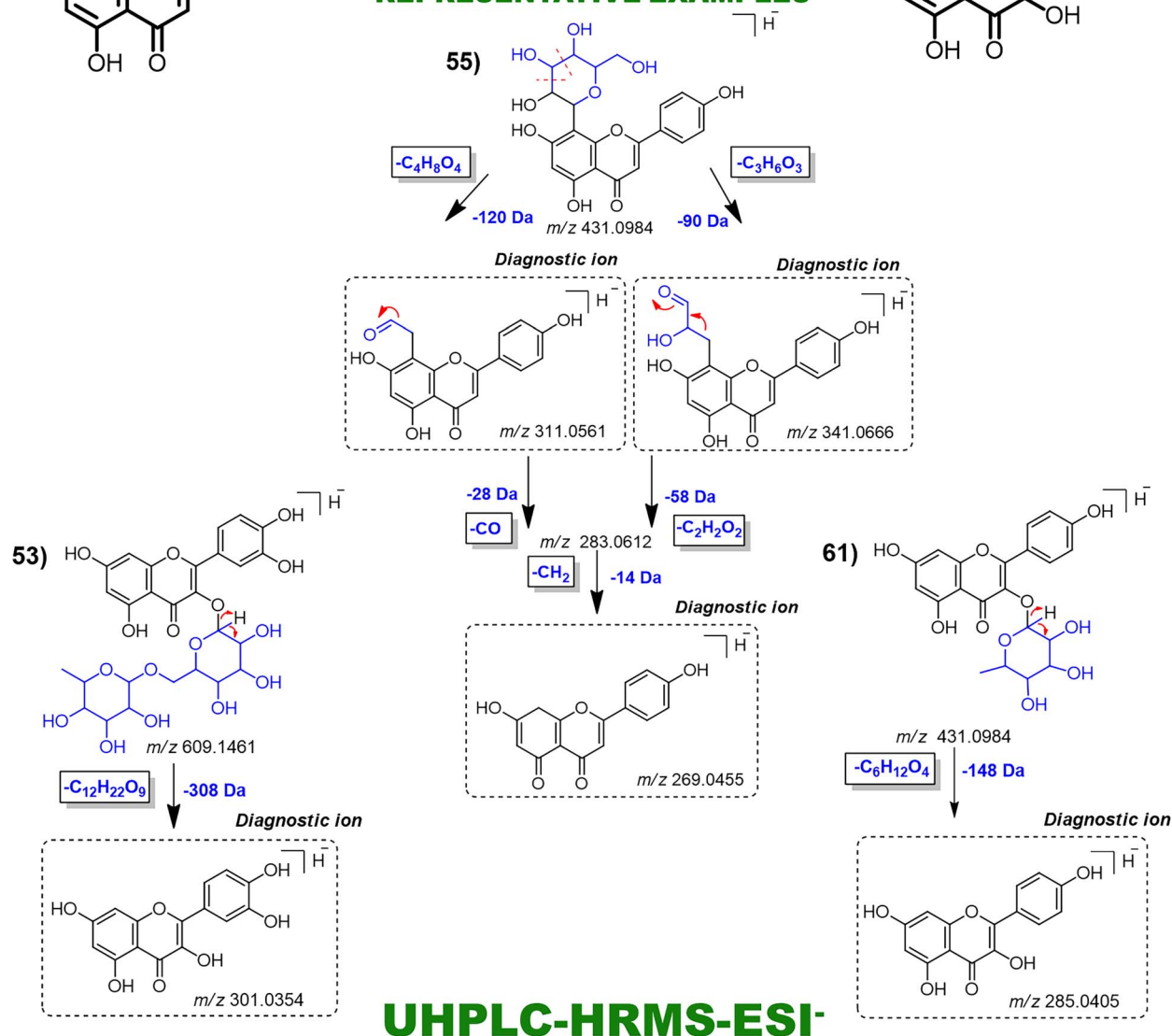
APIGENIN

KAEMPFEROL

QUERCETIN



REPRESENTATIVE EXAMPLES



UHPLC-HRMS-ESI-

Figure 7. Main fragmentation pattern of flavonoids and their key diagnostic ions. Representative examples are demonstrated: Glycosylated apigenin: 51—vitexin. Glycosylated quercetin: 54—rutin. Glycosylated kaempferol: 64—afzelin.

which successfully deconvolutes precursor ions and reassociates precursor-fragment links and whose effectiveness has been widely proven [14, 19, 32, 33].

In contrast, MZmine 3 is a powerful software for processing and analyzing DDA data, while effectively handling DIA data

and integrating with GNPS are still ongoing challenges. The dissociation of MS¹–MS² scans in DIA data remains a significant impediment for current MZmine 3. Full DIA-enabled algorithms within MZmine 3 are still in active development; nonetheless, the latest MZmine 3 can visualize DIA scans chronologically, enabling

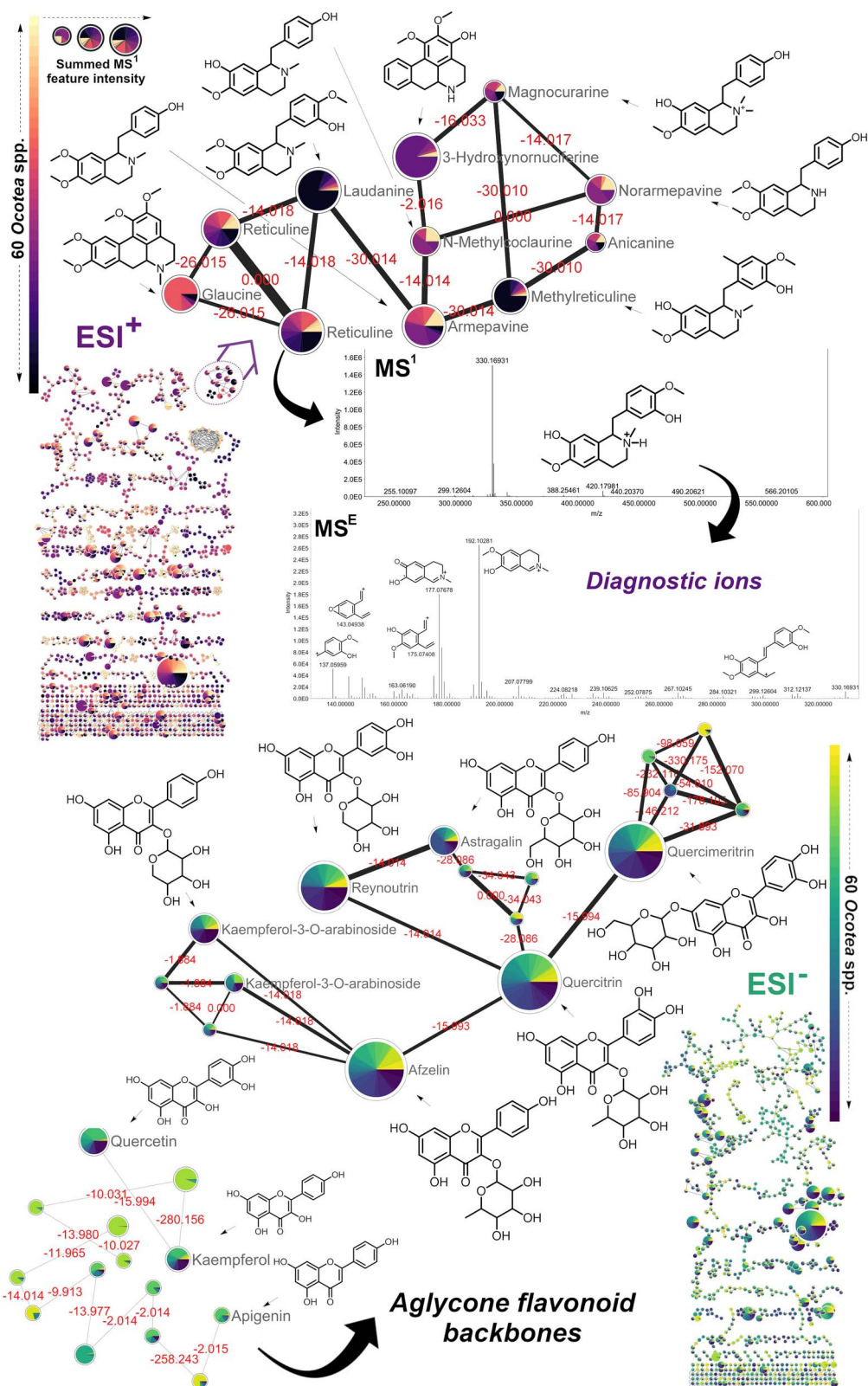


Figure 8. Molecular families of aporphine and benzyloquinoline alkaloids as well as the glycosylated flavonoid cluster families derived from the FBMN. Different alkaloids and flavonoids were annotated with levels 2 and 3 of confidence using GNPS and MoNA spectral matches, and the 'in-house' OcotecaDB. ESI⁺ demonstrates representative reticuline alkaloid MS^E spectra and fragmentation product ions. ESI⁻—Clustering of predominantly O-glycosylated flavonoids identified across Ocoteca spp. and respective aglycones. Each node represents an MS^E-acquired mass spectrum, and the edges connecting them show MS/MS fragmentation similarity (cosine > 0.6). The pie charts show the relative abundance of each Ocoteca plant species ($n = 60$). In MS¹ scans, node diameters are related to the sum of peak regions of the precursor ion in both positive (upper) and negative (lower) modes of ionization.

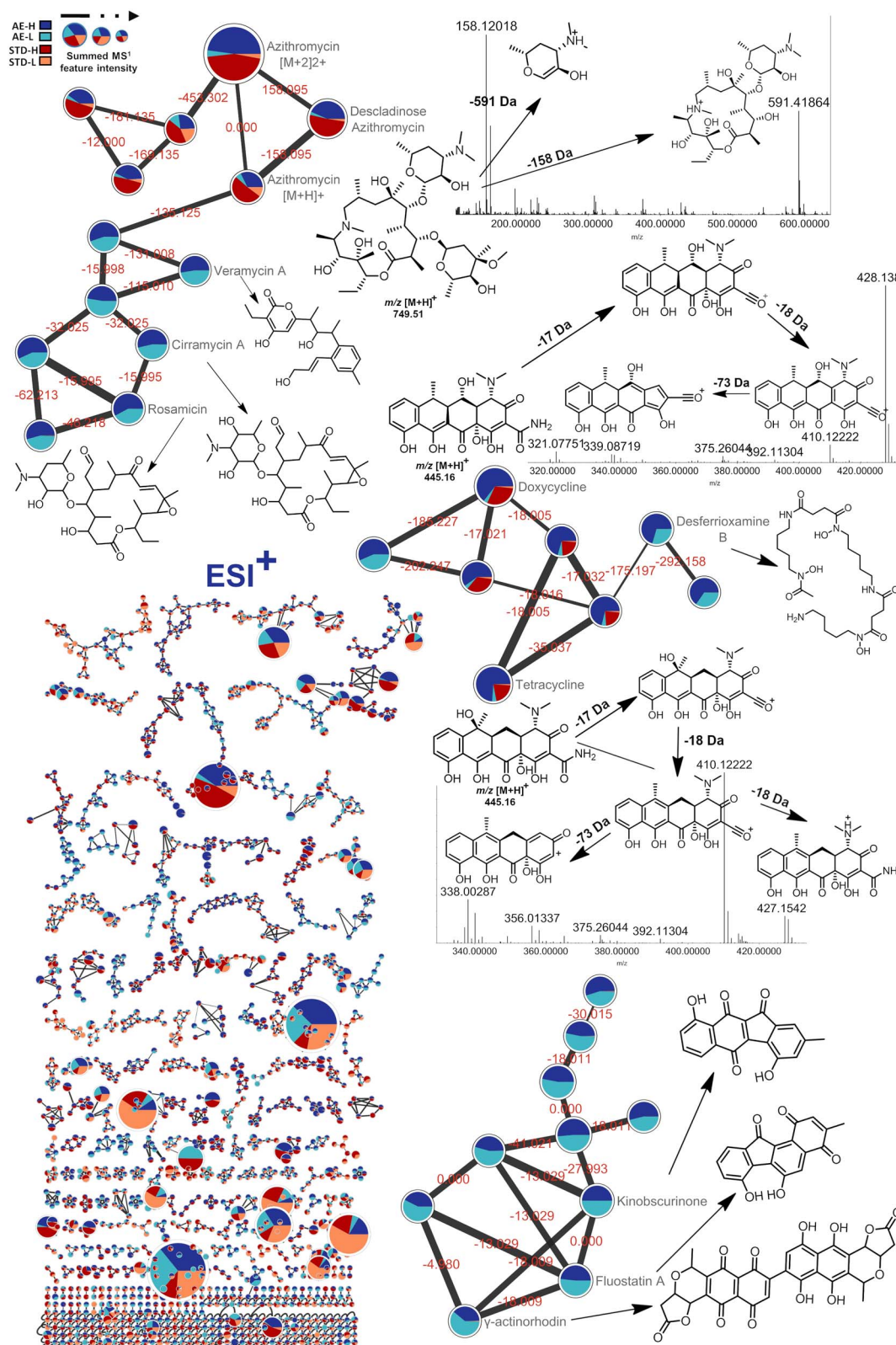


Figure 9. Molecular families of the actinobacterial extract derived from the FBMN at high and low concentrations of spiked chemical authentic standards. Different chemical standards were annotated with level 2 confidence using GNPS spectral matches, including azithromycin, tetracycline and doxycycline. Each node represents an MS^E-acquired mass spectrum, and the edges connecting them show MS/MS fragmentation similarity (cosine > 0.6). The pie charts show the relative abundance of each sample (AE-H—actinobacterial extract spiked with a high concentration of standards, AE-L—actinobacterial extract spiked with a low concentration of standards, STD-H—chemical standards at high concentration and STD-L—chemical standards at low concentration). In MS¹ scans, node diameters are related to the sum of the peak regions of the precursor ion in positive ionization mode.

manual inspection of raw chromatograms and spectra, which was previously only possible with vendor-provided software. While the integration of DIA algorithms within MZmine 3 remains a work in progress, it has emerged as a well-known ecosystem for open MS data processing [11, 18].

Even though MZmine 3 could not export processed DIA data to GNPS, it was utilized as the software platform in this pipeline to visualize the metabolic fingerprints of the pooled QC and crude extract samples, as well as to perform accurate level 3 annotation (Figure 3), providing an integrative MS^E data analysis. The flexibility in setting up parameters is particularly beneficial for the sample alignment step, allowing researchers to make modifications and visualize new results without the need to reprocess previous steps, as required in MS-DIAL. The utilization of MZmine 3 therefore enabled us to generate precise metabolic fingerprint images of the crude extracts from both dataset cases. Manual annotation of metabolites matching the 'in-house' DBs and online spectral libraries, such as the MoNA spectra database, was also incorporated for increased confidence in the annotation of matched metabolites.

The great advantage of this integrated pipeline is the ability to reliably process the entirety of MS^E data, providing optimal visualization of MS¹ and MS² raw and processed data within a user-friendly and open software pipeline environment. Even though both programs used here present limitations, we tried to benefit from their advantages to help overcome the bottlenecks of MS^E data analysis with MN implementation. The extensive development of MZmine 3 is evident through its active GitHub community and frequent updates, showcasing its commitment to continuous improvement. This dynamic environment fosters innovation, and DIA implementation tools seem to be on the horizon. In contrast, MS-DIAL has seen slower recent development, less frequent updates and fewer data processing features and parameters, indicating the need for further improvements. However, the limitations of MS-DIAL do not diminish its effectiveness in performing the necessary tasks for DIA data analysis.

This workflow offers guidance to the community for handling LC-HRMS-DIA/MS^E (and AIF), for which standardized protocols were previously lacking. Future software developments (e.g. DIA algorithms in MZmine 3) will build upon, rather than invalidate, the core foundations established here, as we have delineated key data handling steps for DIA workflow implementation, from data conversion to parameter tuning (see Supplementary Section). Overall, this provides an open-source framework to empower DIA-AIF/MS^E users with customizable workflows for enhanced metabolomics analyses.

Specific parameter adjustments were performed to ensure reliable results for our MS^E data during FBMN jobs, considering that most MN examples available are based on DDA data. Given the larger size of the dataset and the complexity of MS^E, we carefully modified search parameters such as cosine score, number of matched fragment ions and network organization parameters including TopK and maximum connected component size. It is crucial to fine-tune these parameters due to the lack of MS² specificity in MS^E data, where fragment ions originate from all co-eluted precursors (see Methods section). Thus, the TopK value directly influences the number of edges retained in the network and should be considered, as it influences the connections between nodes and the overall structure of the MN. We highlight the importance of considering appropriate values for TopK in the investigation of molecular families of any DIA or MS^E data. In addition, DDA is generally less effective compared to DIA for low-abundance compounds [35]. Lowering cosine parameters

is also applicable to DIA data since the search criteria need to be less restrictive for matches to occur. The association of these strategies provides a solid foundation for future improvements in metabolite identification and cluster analysis of DIA, AIF or MS^E data.

Furthermore, we recommend using more specific metabolite libraries in GNPS—like IQAMDB (IsoQuinoline and Annonaceous Metabolites Database) and NIH natural products—for broad metabolite coverage. DB-based annotation was consolidated with FBMN through feature metrics. For optimal annotation accuracy, curated, phylogeny-relevant libraries are preferable over comprehensive public counterparts. Targeted matching of detected metabolites to expected biosynthetic origins avoids erroneous assignments. Overall, harnessing biosynthetic knowledge through tailored libraries boosts reliability by connecting metabolites specifically to validated biological sources [36].

In this study, we rigorously demonstrate the utility and robustness of the DIA-IntOpenStream pipeline through its application to two distinct and carefully selected datasets, each chosen to showcase different aspects and capabilities of the workflow. In addition, the gas-phase fragmentation reactions were proposed for the different NP classes. The selection of dataset 1 was driven by its potential impact and applicability. Although, *Ocotea* spp. hold significant ethnobotanical importance, display promising medicinal potential and face taxonomic and ecological challenges. In addition, only a limited number of species within the genus have been chemically characterized. Given the high therapeutic potential of the *Ocotea* genus for drug discovery, there is an urgent need for NP chemical studies to support the bioprospecting use of *Ocotea* species, particularly those endangered in Brazil. Research topics focusing on the *Ocotea* genus have importance by themselves, and thus, this dataset also adds value and purpose to our study.

Dataset 2, an actinobacterial extract of MS public data spiked in high and low concentrations with 20 different chemical standards, was specifically chosen for a detailed comparative analysis with existing methods. The addition of known standards allows robust validation of the pipeline's annotation accuracy and efficiency. By using an actinobacterial extract, we also demonstrate the workflow's applicability to microbial metabolomics, an area of significant interest due to the role of microorganisms in environmental processes and human health. A comparative analysis with existing methods that have used this dataset highlights the advancements and improvements that DIA-IntOpenStream offers in terms of data processing efficiency, annotation accuracy and the ability to handle complex NP matrices. Several high-confidence annotations for both datasets were achieved. The results include a significant number of chemical annotations with level 2 confidence according to MSI guidelines [30, 31, 37, 38], with spectra having matched comprehensive spectral libraries of standard compounds (GNPS and MoNA).

Dataset 2, previously examined in a high-quality study [17], involved an advanced LC-HRMS analysis of complex NP mixtures. Among the strategies explored was the use of MS data acquired by DIA, specifically MS^E. In that study, vendor software was used for data analysis, which is an extremely commonly used approach at the time of this study. We re-analyzed their DIA data with the IntOpenStream pipeline, and the obtained results reinforced the pipeline's effectiveness. It successfully allowed the annotation of many authentic chemical standards in the complex NP microbial sources, a key indicator of its reliability. The initial study successfully identified 18 high-abundance and 16 low-abundance standards. In comparison, our pipeline yielded similar results, with a minimal difference of only two and three standards

fewer at each concentration level, respectively (Figure S15 and Table S3).

On the other hand, our pipeline demonstrated enhanced efficacy regarding 5 main aspects. (i) FBMN matched seven authentic standards with annotation confidence level 2 (Table S3), surpassing the two standard matches in the original study that used classical MN. (ii) We obtained an additional seven annotations with confidence level 2 for the original actinobacterial extract by utilizing the built-in ActinomarineDB and manual data inspection (Figures S16–S19 and Table S4). (iii) FBMN analysis revealed various GNPS-matched standards and related compounds in the actinobacterial extract, including the annotation of rosamicin in the azithromycin standard cluster family. (iv) We uniquely detected important genus-specific compounds such as fluostatin A, kinobscurinone and γ -actinorhodin in the actinobacterial extract (Figures 9 and S14). (v) Lastly, our pipeline annotated 450 features at confidence level 3 (details published online at <https://zenodo.org/records/10383866>). As such, our approach not only aligns with existing literature data but also provides complementary insights. The robust comparison against a dataset of established benchmarks highlights the reliability and validity of DIA-IntOpenStream. Our commitment is to offer a freely accessible, robust tool for the metabolomics community to provide independence from the advantages of proprietary software.

As a final comment, the metabolite annotation of constitutional isomers, as observed in the *Ocotea* dataset, can be facilitated using standard compounds for combined MS/MS experiments. However, the stereochemistry of compounds with a high degree of structural similarity demands additional characterization to confirm chemical identity, as in the case of the aporphines boldine and isoboldine, which exhibit the same parent ion and product ions (Figure 4 and Table 1). The ratios and proportions of formed ion fragments differ and might help to elucidate isomers and epimers, at standardized MS conditions, for reliable spectral matching. Implementation of our pipeline enabled us to state the chemical diversity of the studied *Ocotea* species as mainly alkaloid producers. Multiple aporphine alkaloids bearing various substituent patterns were annotated with level 2 of confidence. A range of different glycoside flavonoids were annotated as well. Lastly, a wide variety of lignoids were annotated with level 3 confidence (available online at <https://zenodo.org/records/10383866>). It is worth mentioning that the first report on the evaluation of the chemical composition of several of these endemic *Ocotea* spp. in Brazil was published in 2023 [39].

Several metabolites not previously reported in the *Ocotea* genus were annotated at level 2 confidence. For instance, dehydronuciferine (24) annotated here in some *Ocotea* extracts has only been documented in other plants like the Nymphaeaceae family, encountered in the sacred lotus *Nelumbo nucifera*. NP research on the *N. nucifera* allowed authors to isolate the dehydronuciferine together with other aporphines such as the nuciferine (21) and nor-nuciferine (29), which are common compounds found in the *Ocotea* genus and also reported by us in the present investigation. Also, the alkaloid leucoxyllone (37) is reported in the literature as produced only by two species of the *Ocotea* genus, including *Ocotea leucoxyllone* and *Ocotea minarum* [40, 41]. In this work, it was successfully annotated in other species with high-intensity peak areas, for the first time, in the VZ, VA and PU *Ocotea* extracts (Table S2). In this manner, the present research is also filling this gap and might contribute with chemical characterization data to further taxonomic classification studies associating chemosystematics strategies.

Using a custom database of metabolites previously isolated from the same biological source greatly aids annotation confidence. Large databases can complement this approach but require careful analysis to avoid improbable assignments. Our 'in-house' DBs built in KNIME enabled high-confidence annotations since matched compounds were previously isolated in the targeted genera of the studies. For study cases of biological samples such as urine and blood, a range of other databases are available in HMDB (<https://hmdb.ca/>) as well as other online repositories. In addition, automated annotation with a higher level of confidence can be also performed in MS-DIAL with metabolomics MSP spectral kits or by directly exporting data to the GNPS platform and selecting available spectral libraries. Critically important to perform quality analyses, manual and automated annotations were largely complementary. For example, dataset 1 contained only 18 common level 2 annotations, indicating both strategies are relevant and that combining them can be highly effective.

In conclusion, all these ongoing challenges around LC-HRMS/DIA analysis have motivated us to build this pipeline. We believe it represents an advancement in the field, providing an accessible and efficient workflow for handling complex MS^E data and conducting MN analyses. It can globally aid bioprospecting NP, as we did by unlocking the chemical diversity of plants and bacterial marine extracts. Also, the inclusion of known standards in dataset 2 allowed robust validation of the pipeline's annotation accuracy and efficiency. The use of both datasets highlighted DIA-IntOpenStream's versatility and potential in diverse metabolomics studies. By prioritizing accessibility and transparency, our pipeline ensures that all aspects of data analysis, including processing steps, parameters, software versions and computational setup, are precisely documented and available to the scientific community. This commitment to reproducibility fosters scientific progress and collaboration. Future works may integrate other valuable open data pre-processing, MS/MS annotation and *in silico* fragmentation tools into this pipeline, such as the TidyMS python library, SIRIUS software and MS-FINDER, respectively [42–44]. Overall, this pipeline embodies scientific rigor, and its implementation holds promise for speeding up chemical discoveries, ultimately guiding researchers toward a more collaborative and open environment for research.

METHODS

Solvents, plant material and crude extract preparation

Details regarding the solvents used and sample preparation methods are provided in the Supplementary Material SM-1. Information on solvent sources, purity levels, vegetal material, maceration extraction conditions and sample-handling procedures are all included.

Data acquisition and sample analysis

Chromatographic analysis was performed on an ultra-performance liquid chromatography–quadrupole time-of-flight tandem MS instrument (Xevo qTOF MS, Waters Corp., Milford, USA). Details concerning the QC preparation, chromatographic column, method details and mobile phase system information are described in Supplementary Material SM-2. The electrospray ionization (ESI) source operated in both positive and negative ion modes to capture a comprehensive range of analytes. MS^E, a type of DIA analysis, was conducted using MassLynx™ (v4.2; Waters

Corp., Milford, USA). The mass spectrometer and MS^F acquisition parameters are fully detailed in Supplementary Material SM-2.

Public samples dataset

We validated our pipeline using a publicly available LC-HRMS-DIA/MS^F dataset of a marine actinobacteria extract. This dataset, as described in the publication by Carnevale *et al.*, was enriched with a pool of 20 authentic standards [17]. The chosen standards encompass a wide range of antimicrobial and chemotherapeutic agents, along with naturally occurring compounds, thereby providing a diverse chemical profile suitable for comprehensive analysis. The dataset was obtained from the Massive repository (MSV000088316) and is accessible through the Global Natural Products Social Molecular Networking (GNPS) platform.

Data processing and annotation workflow

The KNIME workflow and subsequent open software in the pipeline were executed on a Windows 11 desktop computer with a 12-core (8 used) 64-bit Intel Core i7-12700—2.10 GHz processor with 32 GB of RAM. The GPU consisted of an NVIDIA T1000 8GB.

KNIME in-house database workflow

To perform the experiments, we have developed a robust workflow to establish an integrated 'in-house' database within Mzmine and MS-DIAL data processing software using the KNIME (University of Konstanz, Zurich, Switzerland, version 4.6.5). KNIME (www.knime.org) is an open-source workflow system with a graphical user interface built on a set of nodes known as 'extensions' that process data and transmit it via connections between those nodes. Thus KNIME provides a simple visual workbench that allows scientists to build and visualize complex workflows [45, 46]. The workflow is online and is publicly available to use (<https://hub.knime.com/-/spaces/-/~8bZEbbknV8tVpTea/current-state/>). Details regarding our custom 'in-house' DB are provided in the SM-3. The 'in-house' database allows level 3 annotation following the guidelines of the MSI [30, 31]. However, it holds more confidence because the 'in-house' DB supports fast annotation of previous metabolites previously isolated in the family or genus of the study.

MS data conversion

To ensure compatibility, accessibility and comparability, the raw Waters MS^F data from both independent datasets were converted to the widely used .mzML format using the recently developed open-source tool Waters2mzML 1.2.0, available on GitHub (<https://github.com/AnP311/Waters2mzML>) (SM-3). The generated .mzML files can be readily processed using Mzmine 3 and MS-DIAL software for further analyses and interpretation.

Mzmine 3 data processing and analysis

The raw data containing peak area and R_t - m/z pairs of 71 *Ocotea* samples (60 *Ocotea* spp. sample extracts, four QCs, four blanks and three VI sample extracts [replicates]) and 17 samples from actinobacterial extract (replicates and blanks), previously converted to .mzML format, were imported into Mzmine 3.4.27 (<https://mzmine.github.io/>; Mzmine Development Team). One QC and one blank sample replicate were excluded from processing due to higher shifts in the R_t compared to other replicates. The detected peaks were deconvoluted, isotopes were eliminated, identical peaks in the different chromatograms were aligned, the remaining gaps were filled, duplicated features were filtered and the blank chromatograms were subtracted. Then, the features were annotated according to their monoisotopic masses. Data

from each ionization mode were processed separately. The data processing parameters are fully detailed in the Supplementary material SM-4.

The treated MS data was then exported in .xlsx format. The Mass Bank of North America (MoNA) (<https://mona.fiehnlab.ucdavis.edu/>) was used for manual spectral comparisons and fragment MS data matches. These manual annotations were listed as level 2 according to the current standards initiative [31, 37, 38, 47]. The chosen modules and algorithms of processing were the standard ones, although this software offers an array of different modern tools that can be used to improve data processing results [6, 11, 18].

MS-DIAL data processing and analysis

The 'Analysis Base File' (.abf) format, generated using Reifys Abf converter software (<https://www.reifys.com/AbfConverter/>), is a traditional data format for MS-DIAL MS^F data processing aimed at MN implementation. However, to ensure maximum compatibility with other software, we chose a more universal approach by converting the data into .mzML format, enabling simultaneous MZmine and MS-DIAL usage and MS data comparison.

The converted .mzML data were successfully loaded into MS-DIAL version 4.9.2 (<http://prime.psc.riken.jp/comprms/msdial/main.html>) for data processing, following a procedure similar to the one used with MZmine 3. In MS-DIAL, we configured the project settings according to our specific data requirements: ionization mode (soft ionization; chromatography; conventional DIA-all-ions method-AIF), the experiment file (available in the Supplementary Material) and data type (i.e. centroid MS¹ and MS/MS data). It is also necessary to process positive and negative ion modes data separately. For data processing, the parameters are fully detailed in the Supplementary Material (SM-5). Subsequently, the data were uploaded to the GNPS server using the open FTP tool named WinSCP (<https://winscp.net/eng/download.php>). All the other existing parameters not mentioned were left at software default.

Molecular networking and metabolite annotation analysis

Metabolite annotation in our study involved a combination of automated and manual approaches (detailed in SM-6). Post-data processing is performed by exporting the results from MS-DIAL, i.e. 'MS2 File' (.mgf), 'Feature Quantification Table' (.txt) and the metadata to the GNPS (<https://gnps.ucsd.edu/>) environment. The metadata file was built in .txt format with the filenames and the respective attributes of species, sample type, region/state of plant collection and endemic occurrence in Brazil. These files are available online on the Zenodo platform (<https://zenodo.org/records/10383866>). FBMN was generated using the respective workflow in the GNPS ecosystem [48] using FBMN parameters described in SM-6. Most of the metabolites were annotated at levels 2 and 3 according to MSI levels. All combined FBMN jobs with level 3 annotated metabolites are listed in Tables S3 and S4 and can be found on the Zenodo platform (<https://zenodo.org/records/10383866>).

Molecular networking visualization and interpretation

The generated networks from GNPS were downloaded and visualized using Cytoscape network software (version 3.8.2). The metadata-rich GNPS table, when opened in Cytoscape, can be exported as a .csv file. This facilitates semi-automated integration with the 'in-house' annotation. Subsequently, the annotated table

can be reimported into the software to perform MN investigation and analysis.

Key Points

- An open, integrated workflow is presented that leverages both universal data formats (mzML) and open-source software tools (KNIME, MZmine, MS-DIAL and GNPS) for enhanced DIA-MS^F data handling.
- The workflow demonstrated its applicability by characterizing *Ocotea* crude plant and marine actinobacterial extract, revealing the chemical diversity of different natural product classes.
- By promoting open science, the pipeline provides a framework to advance DIA-MS^F data handling, transparency, reproducibility and analysis through integrative approaches, overcoming the limitations of commercial solutions.
- We aim to propel the field forward, empowering researchers to achieve more accessible MS^F data processing, with a reliable annotation process, leveraging the potential of DIA-MS to drive the community toward further improvements.

SUPPLEMENTARY DATA

Supplementary data are available online at <http://bib.oxfordjournals.org/>.

ACKNOWLEDGEMENTS

The authors extend their gratitude to the José Badini OUPR herbarium (UFOP—Federal University of Ouro Preto, Minas Gerais—MG, Brazil) and II-Leopoldo Krieger CESJ herbarium (UFJF—Federal University of Juiz de Fora—Minas Gerais, MG, Brazil), for generously providing the *Ocotea* spp. material. Their contributions were invaluable to the applicability case of this research.

AUTHOR CONTRIBUTIONS

A.K.N. and D.A.C.P. designed the pipeline. A.K.N., M.F.A. and D.A.C.P. wrote the draft paper. M.M. performed the high-resolution DIA-MS experiments. A.K.N. and M.F.A. performed MS data conversion. A.K.N., M.S.F. and T.B.O. developed the KNIME workflow. A.K.N., M.F.A., M.S.F. and K.J.N. performed *Ocotea* data curation. A.K.N., K.J.N., M.F.A., M.S.F. and P.C.P.B. performed data processing. A.K.N. and M.F.A. performed the FBMN experiments. A.K.N., J.H.G.L., K.J.N., M.F.A., M.G.S., H.B. and D.A.C.P. aided in gas-phase fragmentation reactions elucidation. A.C.C.P.L., A.K.N., M.F.A., M.S.F., P.C.P.B., T.B.O., H.B., M.G.S. and D.F.D. aided in the construction of the manuscript tables, preparation of the figures and report of annotation standards report. A.K.N., A.C.C.P.L., K.J.N., J.H.G.L., D.F.D., M.F.A., M.G.S., H.B., M.M., M.S.F., P.C.P.B., T.B.O. and D.A.C.P. participated in data analysis, interpretation, discussion and revision. All authors have read the manuscript.

FUNDING

This work received financial support from the Coordination for the Improvement of Higher Education Personnel (CAPES Foundation, Brazil), finance code 001; Minas Gerais State Research Foundation (FAPEMIG, Brazil), finances codes: APQ-05218-23,

APQ-00544-23, APQ-02353-17, APQ-00207-18 and BPD-00760-22; National Council for Scientific and Technological Development (CNPq, Brazil) finances codes 408115/2023-8, 316204/2021-8 and 406837/2021-0 for financial support; São Paulo State Research Foundation (FAPESP, Brazil), finances codes: 2021/02789-7 e 2023/12447-1.

DATA AVAILABILITY

The *Ocotea* dataset of the current study was deposited to the Mass Spectrometry Interactive Virtual Environment (MassIVE) repository. The data that support the findings of this study are available at <ftp://MSV000093006@massive.ucsd.edu> under the registry code MSV000093006. Guidelines for files submitted to MassIVE for public access can be found online (<https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp>).

ADDITIONAL INFORMATION

Accompanying information supplements this paper online at Zenodo open digital library: <https://zenodo.org/records/10383866>.

REFERENCES

1. Gaudêncio SP, Pereira F. Dereplication: racing to speed up the natural products discovery process. *Nat Prod Rep* 2015;**32**: 779–810.
2. Vereyken L, Dillen L, Vreeken RJ, et al. High-resolution mass spectrometry quantification: impact of differences in data processing of centroid and continuum data. *J Am Soc Mass Spectrom* 2019;**30**:203–12.
3. Yuliana ND, Khatib A, Choi YH, Verpoorte R. Metabolomics for bioactivity assessment of natural products. *Phytother Res* 2011;**25**:157–69.
4. K de J N, Ferreira MS, Katchborian-Neto A, et al. Anti-inflammatory markers of hops cultivars (*Humulus lupulus* L.) evaluated by untargeted metabolomics strategy. *Chem Biodivers* 2022;**19**(4):e202100966.
5. Valmori M, Marie V, Fenaille F, et al. Recent methodological developments in data-dependent analysis and data-independent analysis workflows for exhaustive lipidome coverage. *Front Anal Sci* 2023;**3**:1–9.
6. Alka O, Shanthamoorthy P, Witting M, et al. DIAMetAlyzer allows automated false-discovery rate-controlled analysis for data-independent acquisition in metabolomics. *Nat Commun* 2022;**13**: 1–9.
7. Wang R, Yin Y, Zhu ZJ. Advancing untargeted metabolomics using data-independent acquisition mass spectrometry technology. *Anal Bioanal Chem* 2019;**411**:4349–57.
8. Rosnack KJ, Reid MJ, Ladak A, et al. Screening solution using the software platform UNIFI: an integrated workflow by waters. *ACS Symp Ser* 2016;**1242**:155–72.
9. Naz S, Gallart-Ayala H, Reinke SN, et al. Development of a liquid chromatography-high resolution mass spectrometry metabolomics method with high specificity for metabolite identification using all ion fragmentation acquisition. *Anal Chem* 2017;**89**:7933–42.
10. Rakusanova S, Fiehn O, Cajka T. Toward building mass spectrometry-based metabolomics and lipidomics atlases for biological and clinical research. *Trends Anal Chem* 2023;**158**: 116825.
11. Ebbels TMD, Van Der HJ, Chatelaine H, et al. Recent advances in mass spectrometry-based computational metabolomics. *Curr Opin Chem Biol* 2023;**74**:102288.

12. Ledesma-Escobar CA, Priego-Capote F, Calderón-Santiago M. MetaboMSDIA: a tool for implementing data-independent acquisition in metabolomic-based mass spectrometry analysis. *Anal Chim Acta* 2023;**1266**:341308.
13. Stancliffe E, Schwaiger-Haber M, Sindelar M, Patti GJ. DecoID improves identification rates in metabolomics through database-assisted MS/MS deconvolution. *Nat Methods* 2021;**18**:779–87.
14. Aron AT, Gentry EC, McPhail KL, et al. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nat Protoc* 2020;**15**:1954–91.
15. Wang M, Carver JJ, Phelan VV, et al. Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nat Biotechnol* 2016;**34**:828–37.
16. Li C, Gao M, Yang W, et al. Diamond: a multi-modal DIA mass spectrometry data processing pipeline. *Bioinformatics* 2021;**37**:265–7.
17. Carnevale Neto F, Clark TN, Lopes NP, Linington RG. Evaluation of ion mobility spectrometry for improving constitutional assignment in natural product mixtures. *J Nat Prod* 2022;**85**:519–29.
18. Schmid R, Heuckeroth S, Korf A, et al. Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nat Biotechnol* 2023;**41**:447–9.
19. Tsugawa H, Cajka T, Kind T, et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods* 2015;**12**:523–6.
20. Evans AM, O'Donovan C, Playdon M, et al. Dissemination and analysis of the quality assurance (QA) and quality control (QC) practices of LC–MS based untargeted metabolomics practitioners. *Metabolomics* 2020;**16**:1–16.
21. Kirwan JA, Gika H, Beger RD, et al. Quality assurance and quality control reporting in untargeted metabolic phenotyping: mQACC recommendations for analytical quality management. *Metabolomics* 2022;**18**:1–16.
22. Schymanski EL, Jeon J, Gulde R, et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ Sci Technol* 2014;**48**:2097–8.
23. Demarque DP, Crotti AEM, Vessecchi R, et al. Fragmentation reactions using electrospray ionization mass spectrometry: an important tool for the structural elucidation and characterization of synthetic and natural products. *Nat Prod Rep* 2016;**33**:432–55.
24. Qing Z, Xu Y, Yu L, et al. Investigation of fragmentation behaviours of isoquinoline alkaloids by mass spectrometry combined with computational chemistry. *Sci Rep* 2020;**10**:1–13.
25. Stévigny C, Jiwan JLH, Rozenberg R, et al. Key fragmentation patterns of aporphine alkaloids by electrospray ionization with multistage mass spectrometry. *Rapid Commun Mass Spectrom* 2004;**18**:523–8.
26. Carnevale Neto F, Andréo MA, Raftery D, et al. Characterization of aporphine alkaloids by electrospray ionization tandem mass spectrometry and density functional theory calculations. *Rapid Commun Mass Spectrom* 2020;**34**(S3):e8533.
27. Conceição RS, Reis IMA, Cerqueira APM, et al. Rapid structural characterisation of benzyloquinoline and aporphine alkaloids from *Ocotea spixiana* acaricide extract by HPTLC-DESI-MSn. *Phytochem Anal* 2020;**31**:711–21.
28. Cho K, Schwaiger-Haber M, Naser FJ, et al. Targeting unique biological signals on the fly to improve MS/MS coverage and identification efficiency in metabolomics. *Anal Chim Acta* 2021;**1149**:338210.
29. Yang X, Gao X, Cao Y, et al. Anti-inflammatory effects of boldine and reticuline isolated from *Litsea cubeba* through JAK2/STAT3 and NF- κ B signaling pathways. *Planta Med* 2018;**84**:20–5.
30. Sumner LW, Samuel T, Noble R, et al. Proposed minimum reporting standards for chemical analysis chemical analysis working group (CAWG) metabolomics standards initiative (MSI). *Metabolomics* 2007;**3**:211–21.
31. Creek DJ, Dunn WB, Fiehn O, et al. Metabolite identification: are you sure? And how do your peers gauge your confidence? *Metabolomics* 2014;**10**:350–3.
32. Nephali L, Steenkamp P, Burgess K, et al. Mass spectral molecular networking to profile the metabolome of biostimulant bacillus strains. *Front Plant Sci* 2022;**13**:1–14.
33. Tinte MM, Masike K, Steenkamp PA, et al. Computational metabolomics tools reveal metabolic reconfigurations underlying the effects of biostimulant seaweed extracts on maize plants under drought stress conditions. *Metabolites* 2022;**12**:1–24.
34. Bonde CS, Bormancin L, Lu Y, et al. Bio-guided fractionation and molecular networking reveal fatty acids to be principal anti-parasitic compounds in Nordic seaweeds. *Front Pharmacol* 2021;**12**:1–12.
35. Li KW, Gonzalez-Lozano MA, Koopmans F, et al. Recent developments in data independent acquisition (DIA) mass spectrometry: application of quantitative analysis of the brain proteome. *Front Mol Neurosci* 2020;**13**:1–8.
36. Theodoridis G, Gika H, Raftery D, et al. Ensuring fact-based metabolite identification in liquid chromatography-mass spectrometry-based metabolomics. *Anal Chem* 2023;**95**:3909–16.
37. Fiehn O, Robertson D, Griffin J, et al. The metabolomics standards initiative (MSI). *Metabolomics* 2007;**3**:175–8.
38. Spicer RA, Salek R, Steinbeck C. Comment: a decade after the metabolomics standards initiative it's time for a revision. *Sci Data* 2017;**4**:2–4.
39. Katchborian-neto A, De JK, Carolina CJC, et al. Bioprospecting-based untargeted metabolomics identifies alkaloids as potential anti-inflammatory bioactive markers of *Ocotea* species (Lauraceae). *Phytomedicine* 2023;**120**:155060.
40. Vecchiatti V, Casagrande C, Ferrari G, et al. New aporphine alkaloids of *Ocotea minarum*. *Farm Ed Sci* 1979;**34**:829–40.
41. Zhou BN, Johnson RK, Mattern MR, et al. Isolation and biochemical characterization of a new topoisomerase I inhibitor from *Ocotea leucoxylo*. *J Nat Prod* 2000;**63**:217–21.
42. Riquelme G, Zabalegui N, Marchi P, et al. A python-based pipeline for preprocessing LC–MS data for untargeted metabolomics workflows. *Metabolites* 2020;**10**:1–14.
43. Dührkop K, Fleischauer M, Ludwig M, et al. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat Methods* 2019;**16**:299–302.
44. Mallmann LP, O Rios A, Rodrigues E. MS-FINDER and SIRIUS for phenolic compound identification from high-resolution mass spectrometry data. *Food Res Int* 2023;**163**:112315.
45. Khalid B, Ghorab H, Benkhemissa A. Automated QSPR modeling and data curation of physicochemical properties using KNIME platform: prediction of partition coefficients. *J Indian Chem Soc* 2022;**99**:100672.
46. Warr WA. Scientific workflow systems: pipeline pilot and KNIME. *J Comput Aided Mol Des* 2012;**26**:801–4.
47. Sumner LW, Lei Z, Nikolau BJ, Saito K. Modern plant metabolomics: advanced natural product gene discoveries, improved technologies, and future prospects. *Nat Prod Rep* 2015;**32**(2):212–29.
48. Nothias LF, Petras D, Schmid R, et al. Feature-based molecular networking in the GNPS analysis environment. *Nat Methods* 2020;**17**:905–8.

Integrative open workflow for confident annotation and molecular networking of metabolomics MS^E/DIA data

Albert Katchborian-Neto¹, Matheus Fernandes Alves¹, Paula Carolina Pires Bueno^{1,2}, Karen de Jesus Nicácio³, Miller Santos Ferreira¹, Tiago Branquinho Oliveira⁴, Henrique Barbosa⁵, Michael Murgu⁶, Ana Cláudia Chagas de Paula Ladvocat⁷, Danielle Ferreira Dias¹, Marisi Gomes Soares¹, João Henrique Ghilardi Lago⁵, Daniela Aparecida Chagas-Paula^{1*}

¹Chemistry Institute, Federal University of Alfenas, 37130-001, Alfenas, Minas Gerais, Brazil

²Leibniz Institute of Vegetable and Ornamental Crops (IGZ), Theodor-Echtermeyer-Weg 1, 14979, Großbeeren, Germany

³Department of Chemistry, Federal University of Mato Grosso, 14040-901, Cuiabá, Mato Grosso, Brazil

⁴Department of Pharmacy, Federal University of Sergipe, 49100-000, São Cristóvão, Sergipe, Brazil

⁵Center of Natural Sciences and Humanities, Federal University of ABC, 09210-180, Santo Andre, São Paulo, Brazil.

⁶Waters Corporation, Alameda Tocantins 125, 27th floor, Alphaville, 06455-020, São Paulo, São Paulo, Brazil

⁷Department of Pharmaceutical Sciences, Federal University of Juiz de Fora, 36036-900, Juiz de Fora, Minas Gerais, Brazil

Supplementary material

Supplementary Figures	1
Supplementary Tables	20
Supplementary Texts	39
Supplementary Protocol	46
Supplementary Material and Methods	54
Supplementary FMBN jobs and Zenodo links	60

Supplementary Figures

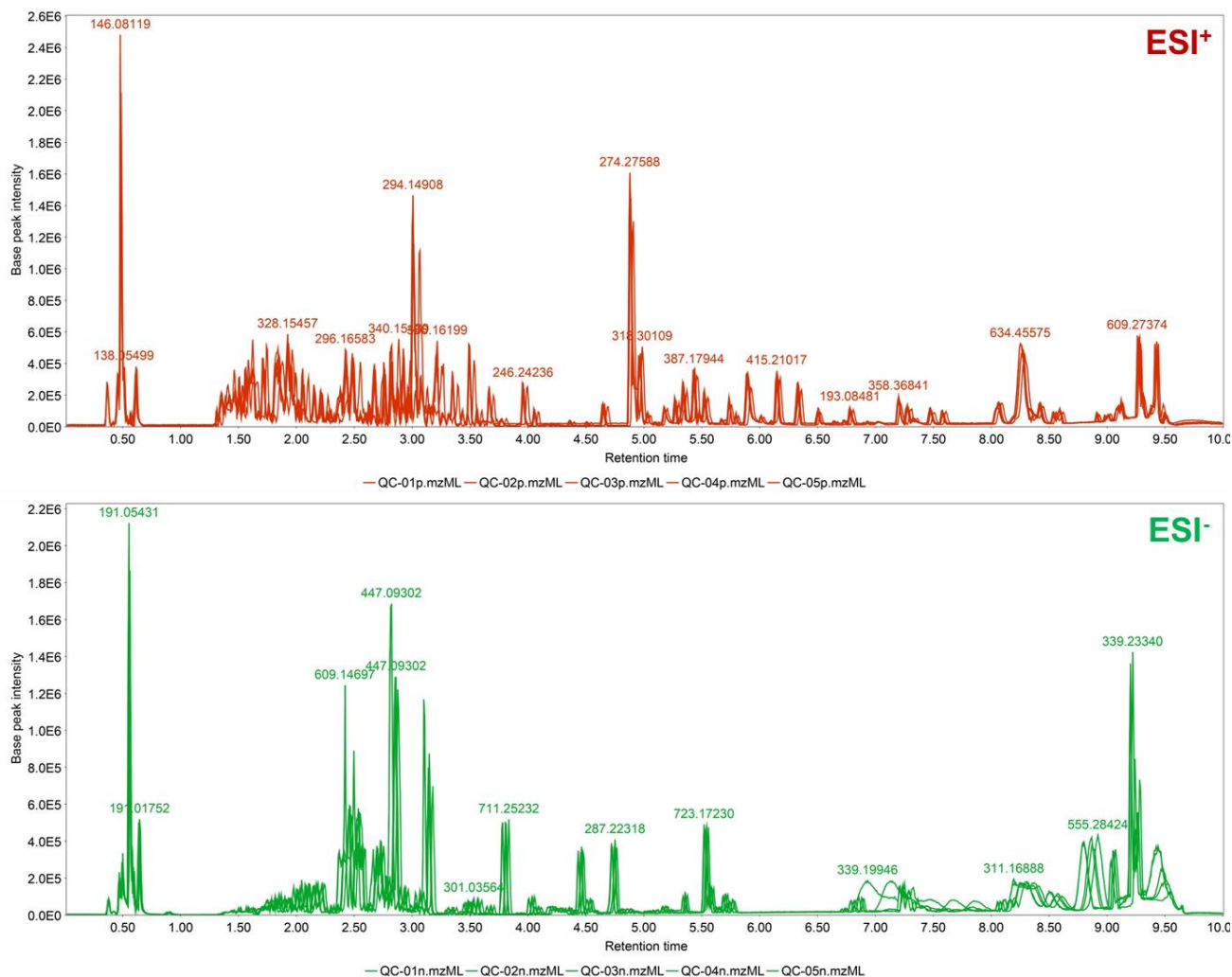


Fig. S1 | LC-HRMS-MS^E/DIA metabolic fingerprints shown as Base Peak Ion (BPI) chromatograms displaying the overlapped quality control (QC) replicates (pooled from 60 *Ocotea* spp. leaf extracts). The QC replicates were acquired in both positive (red) and negative (green) electrospray ionization modes under similar conditions. The overlay of the BPI traces for 5 replicate injections demonstrates highly reproducible chromatographic and mass spectrometric performance essential for robust comparative metabolomics analysis.

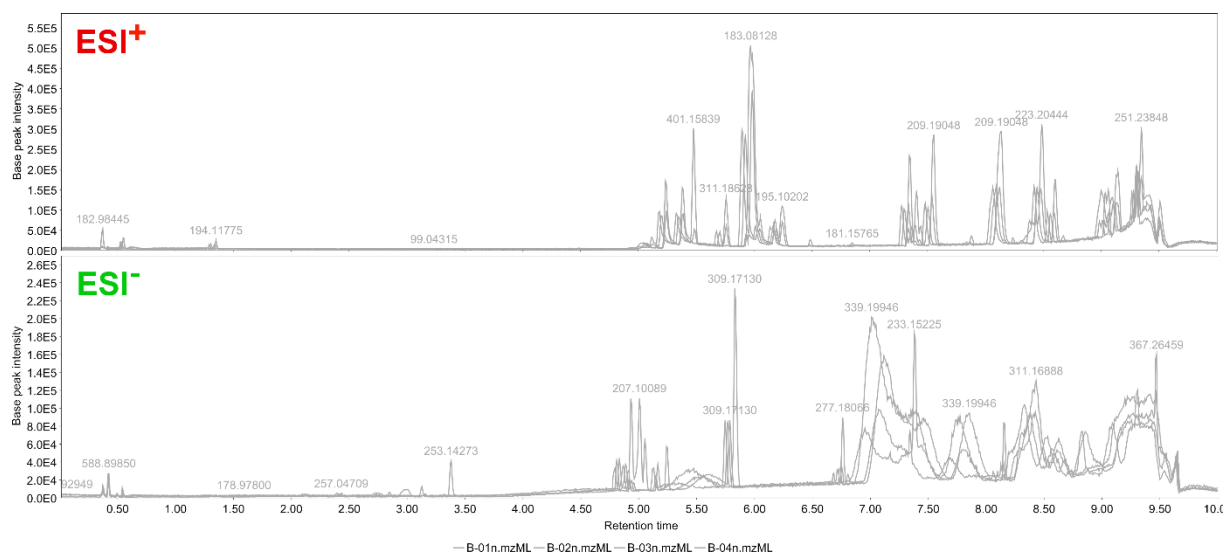


Fig. S2 | LC-HRMS-MS^E/DIA fingerprints shown as Base Peak Ion (BPI) chromatograms displaying the overlapped blank samples from the *Ocotea* dataset 1. The blank replicates were acquired in both positive (red) and negative (green) electrospray ionization modes under identical conditions. The overlay of the BPI traces for 5 replicate injections demonstrates highly reproducible chromatographic analysis.

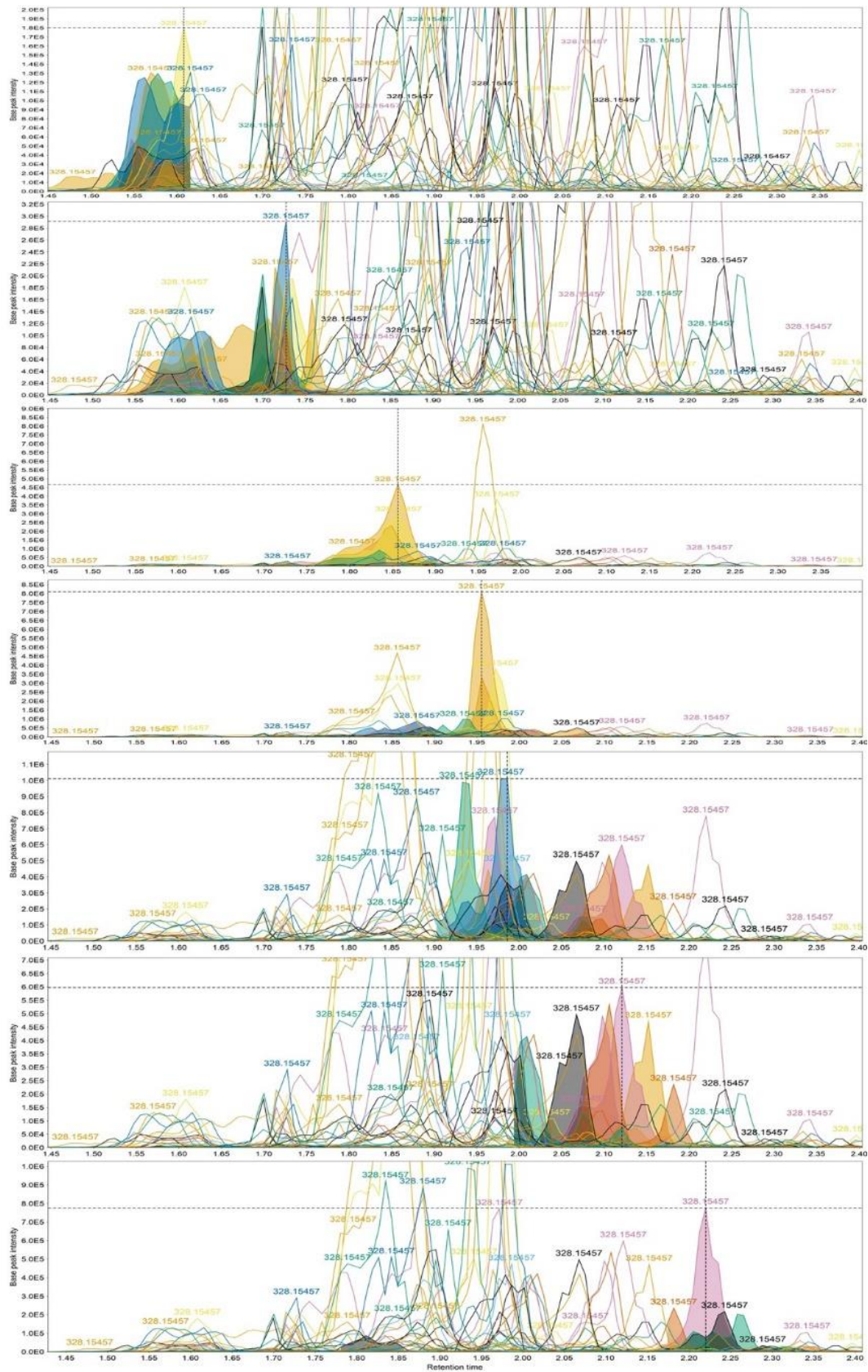


Fig. S3 | Aligned chromatograms from *Ocolea* spp. samples, with zoom on the features with m/z 328.154 on a 1.45-2.40 min retention time (R_t) window. The overlaid chromatograms exhibit precise retention time alignment of the isomeric compounds and high resolution of the close features.

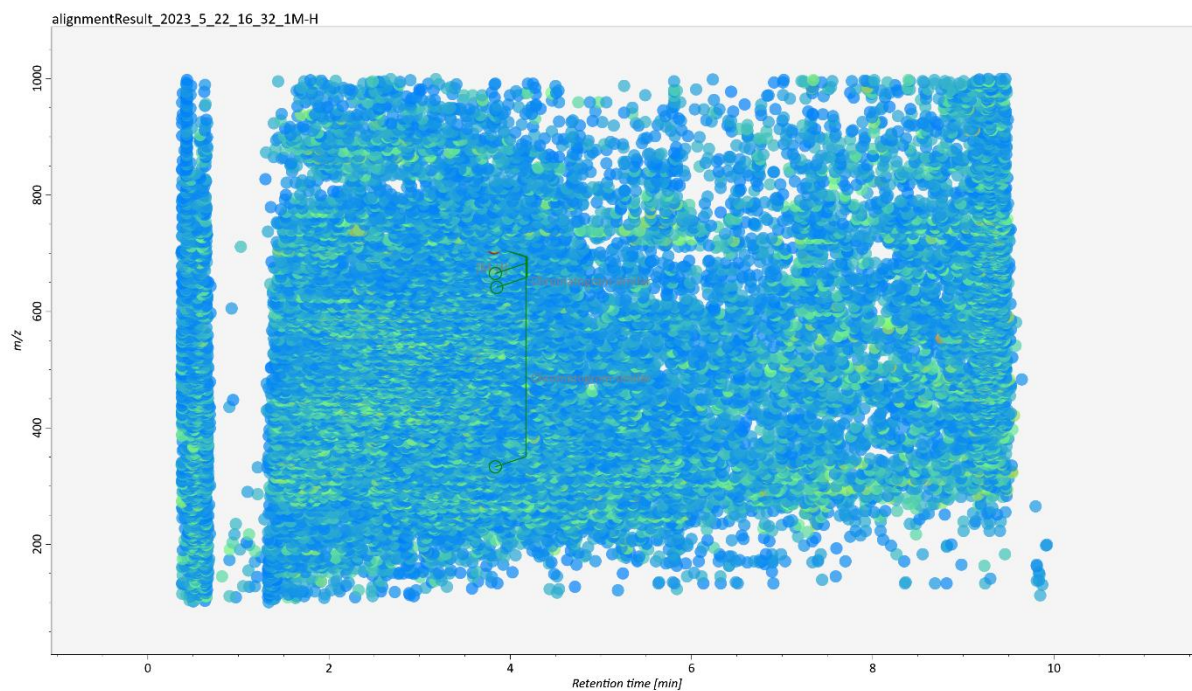


Fig. S4 | MS-DIAL alignment result of negative ionization mode from *Ocotea* dataset. Each dot (circle format) represents a particular feature of the processing with the colour pallet from blue to green varying with the intensity, where the green ones represent the most intense ions. In detail, the ions at m/z 711.2523 and R_t 3.82.

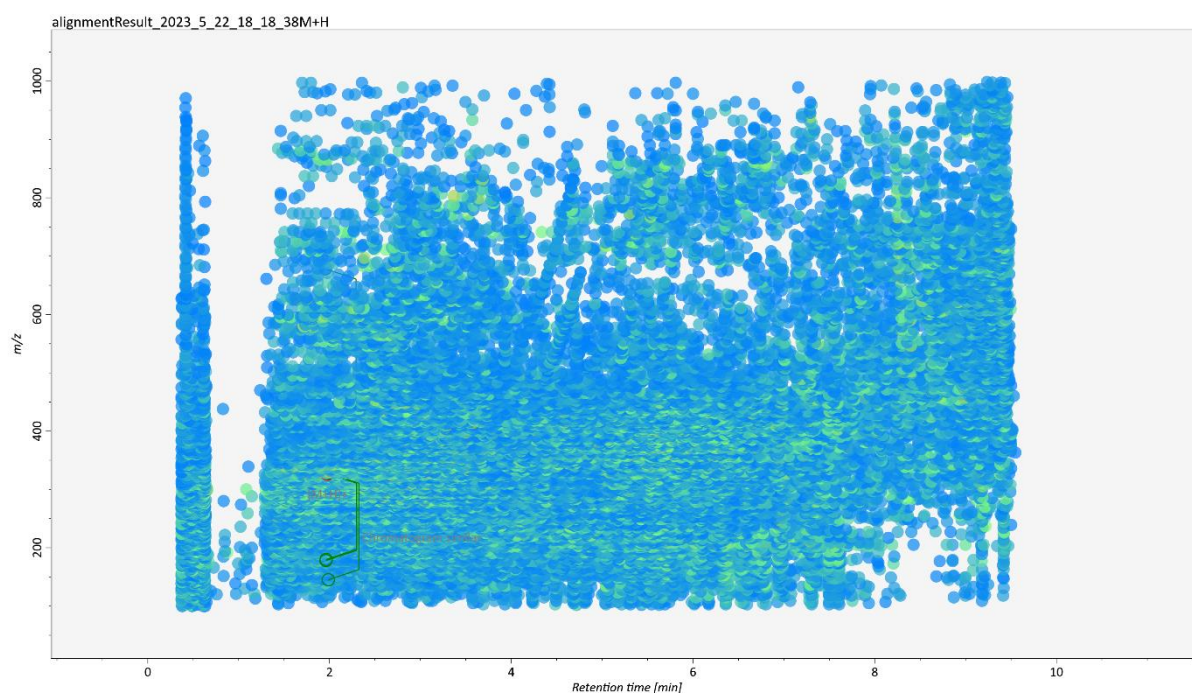


Fig. S5 | MS-DIAL alignment result of positive ionization mode from *Ocotea* dataset. Each dot (circle format) represents a particular feature of the processing with the same variation on the colour pallet. In detail, the ions at m/z 328.1546 and R_t 1.98.

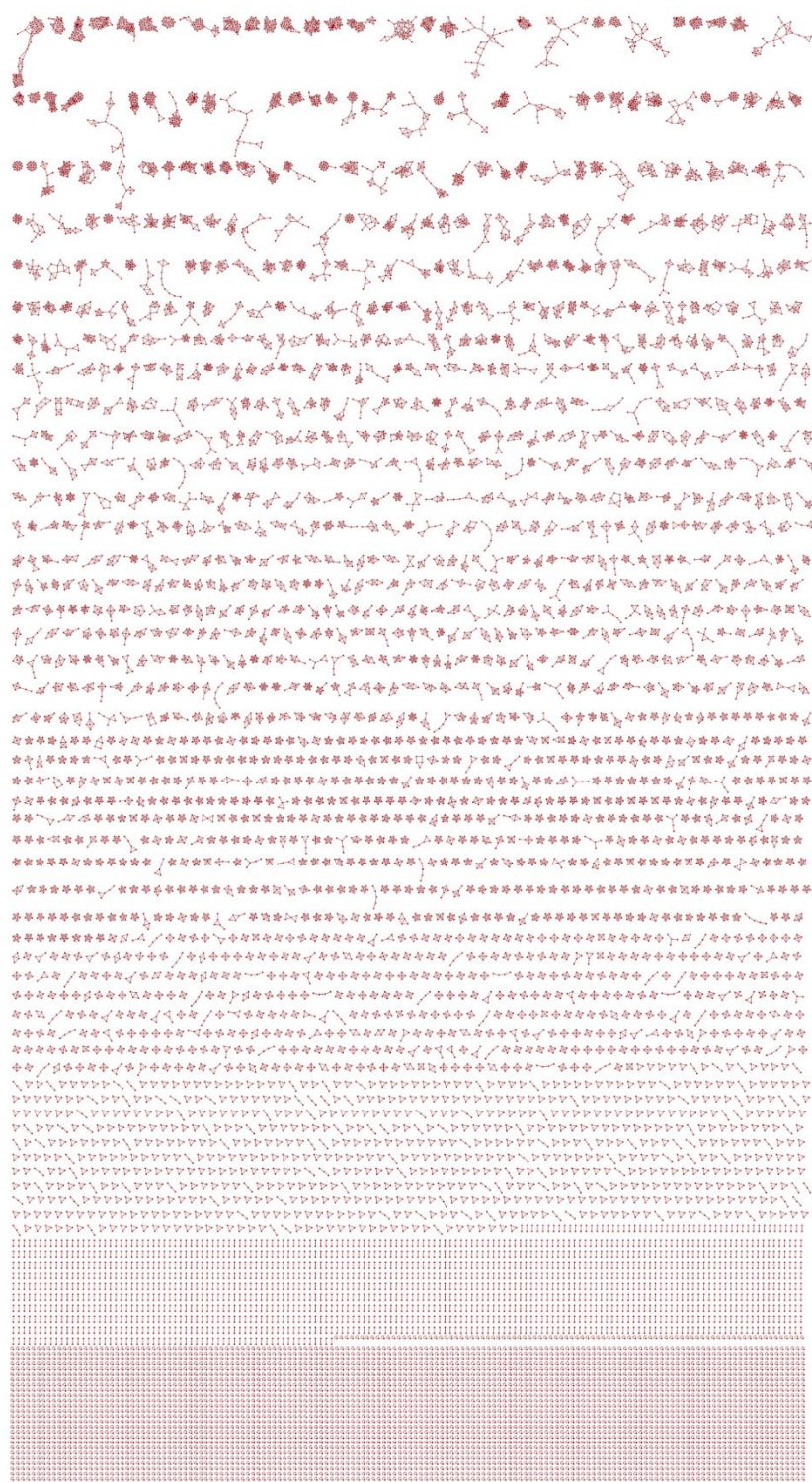


Fig. S6 | FBMN of positive (red) mode of ionization from *Ocotea* dataset. Each node on networks represents a feature (m/z / R_i) from MS-DIAL processing connected with a basis on MS/MS spectral similarity. 22,572 nodes are shown on positive modes. The amplitude used in the MS-DIAL data processing for visualization of these MN was set at 1000. The complexity of our data is illustrated by these networks, showing that complex matrices as plant extracts aligned in big datasets ($n=60$) demand sophisticated tools for reliable processing, visualization and interpretation of the results.



Fig. S7 | FBMN of negative (green) mode of ionization from *Ocotea* dataset. Each node on networks represents a feature (m/z / R_c) from MS-DIAL processing connected with a basis on MS/MS spectral similarity. 21.838 nodes are shown in negative mode. The amplitude used in the MS-DIAL data processing for visualization of these MN was set at 1000. The complexity of our data is illustrated by these networks, showing that complex matrices as plant extracts aligned in big datasets ($n=60$) demand sophisticated tools for reliable processing, visualization and interpretation of the results.

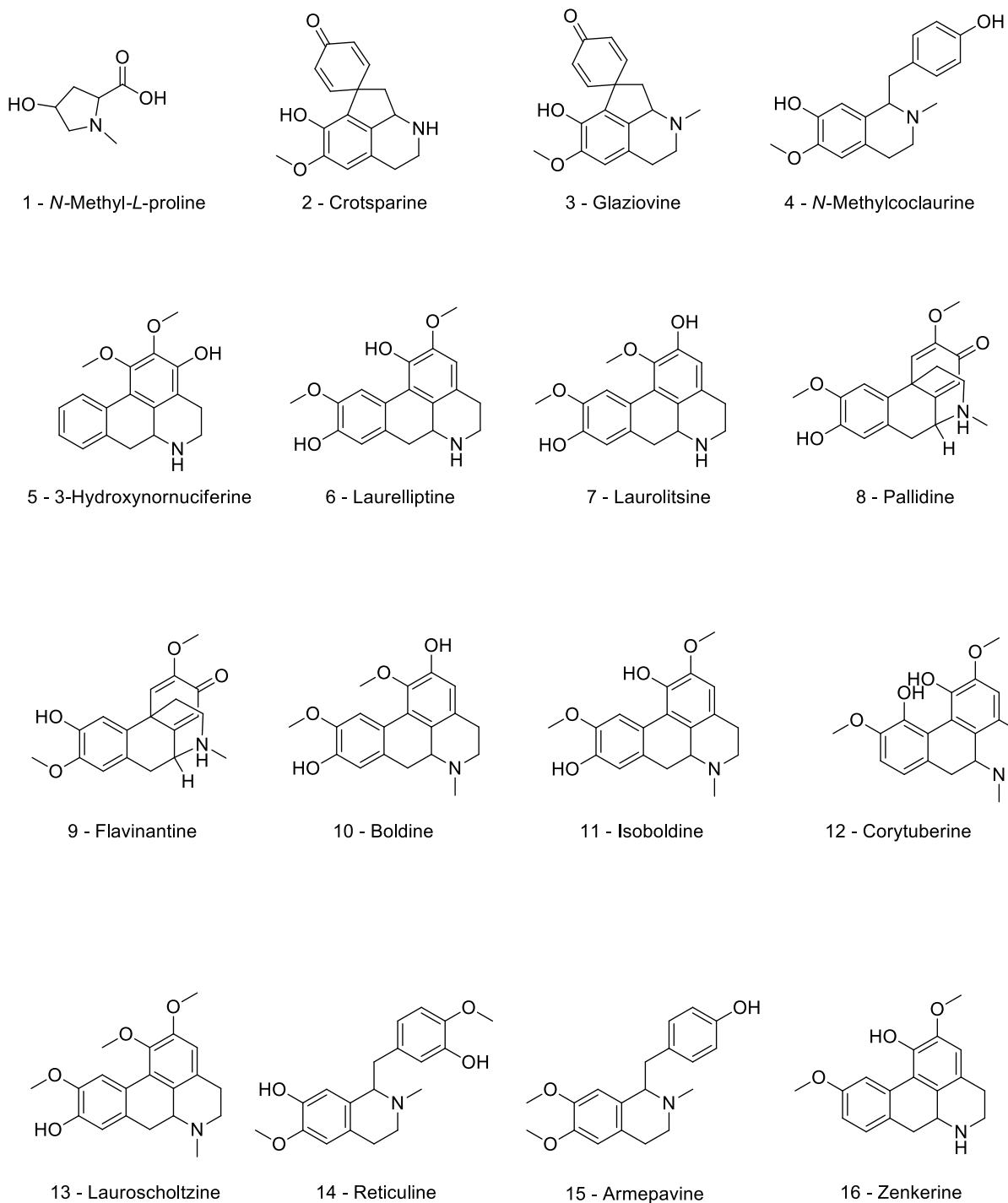
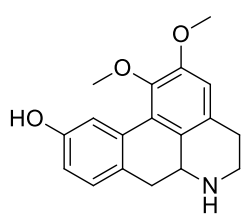
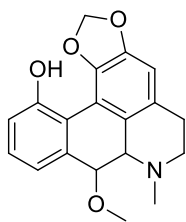


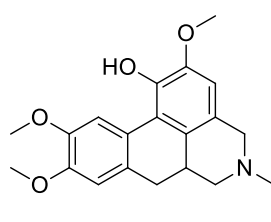
Fig. S8 | Chemical structures annotated for the *Ocotea* dataset with level 2 of confidence (ID 1-16).



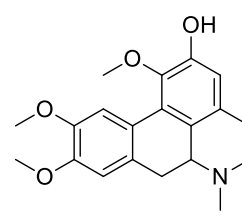
17 - Tuduranine



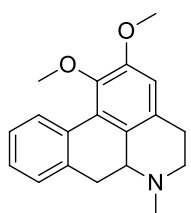
18 - Diospirifoline



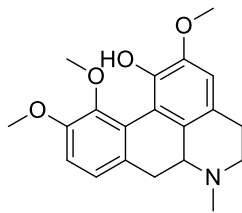
19 - Thaliporphine



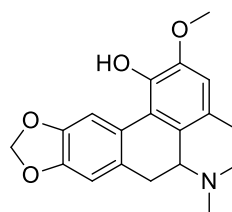
20 - Predicentrine



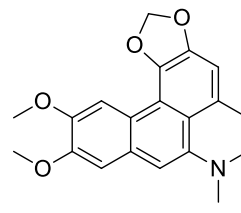
21 - Nuciferine



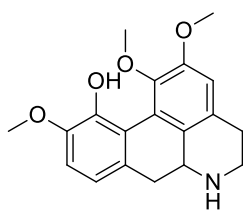
22 - Corydine



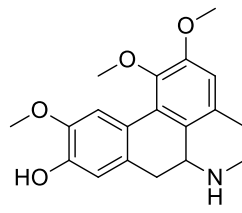
23 - Domesticine



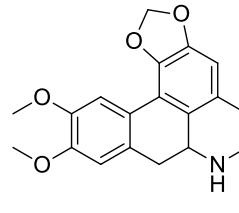
24 - Dehydrodicentrine



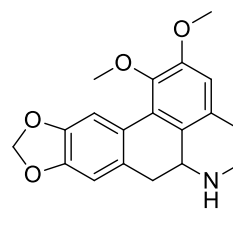
25 - Norisocorydine



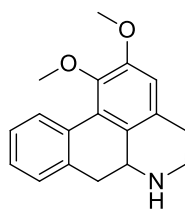
26 - Laurotetanine



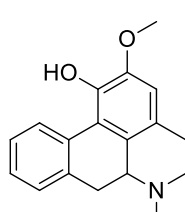
27 - Nordicentrine



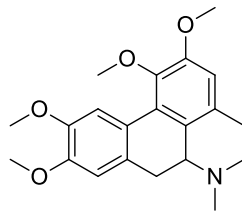
28 - Nornantenine



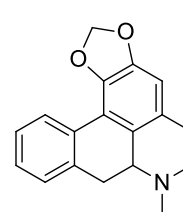
29 - Nornuciferine



30 - Lirinidine



31 - Glaucine



32 - Roemerine

Fig. S9 | Chemical structures annotated for the *Ocotea* dataset with level 2 of confidence (ID 17-32).

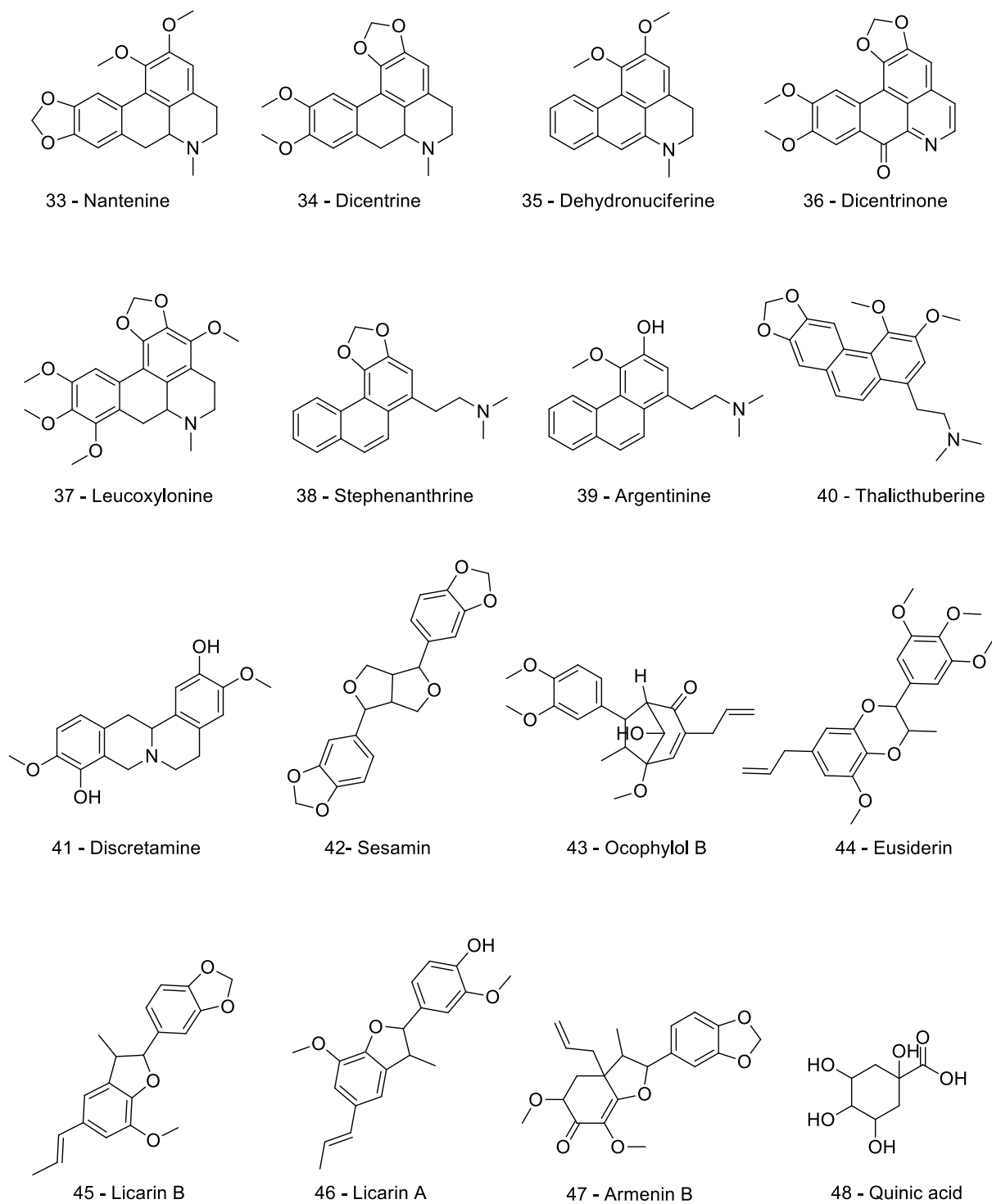


Fig. S10 | Chemical structures annotated for the *Ocotea* dataset with level 2 of confidence (ID 33-48).

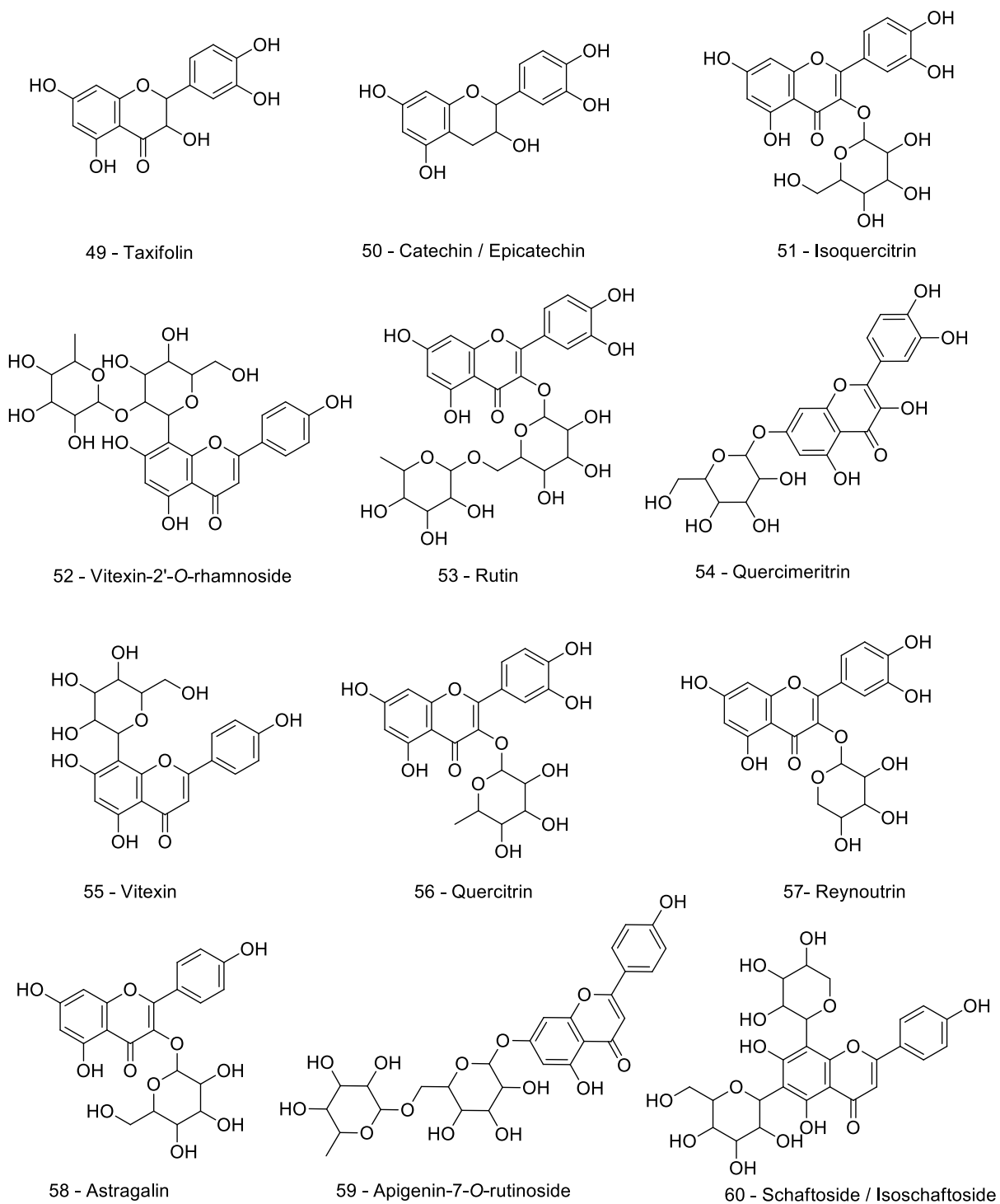
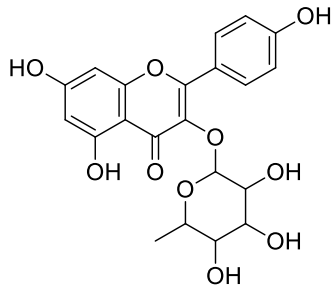
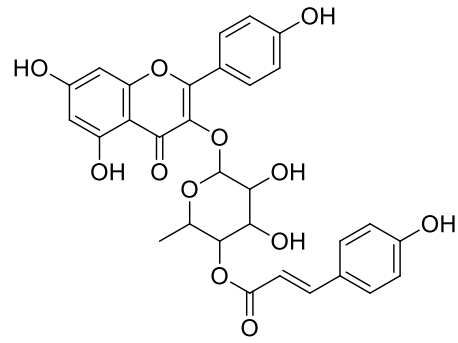


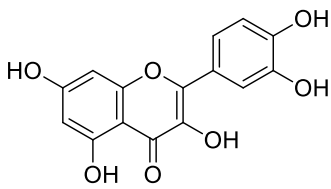
Fig. S11 | Chemical structures annotated for the *Ocotea* dataset with level 2 of confidence (ID 49-60).



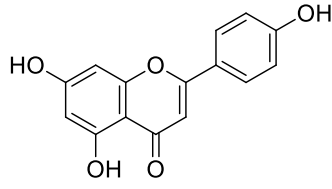
61 - Afzelin



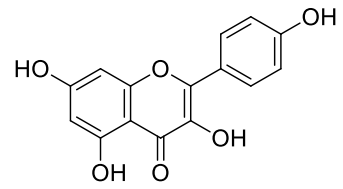
62 - Kaempferol 3-4''-*p*-coumarylrhamnoside



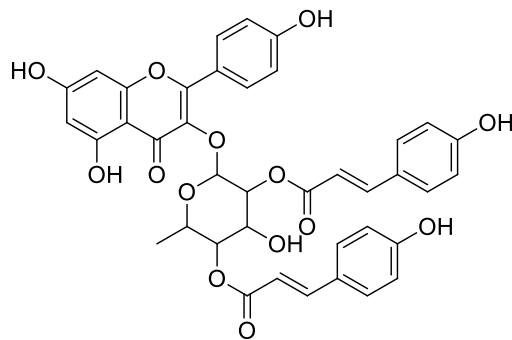
63 - Quercetin



64 - Apigenin



65 - Kaempferol



66- Kaempferol 3-(2'',4''-
di-(*E*)-*p*-coumarylrhamnoside) / Kaempferol 3-(3'',4''-
di-(*E*)-*p*-coumarylrhamnoside)

Fig. S12 | Chemical structures annotated for the *Ocotea* dataset with level 2 of confidence (ID 61-66).

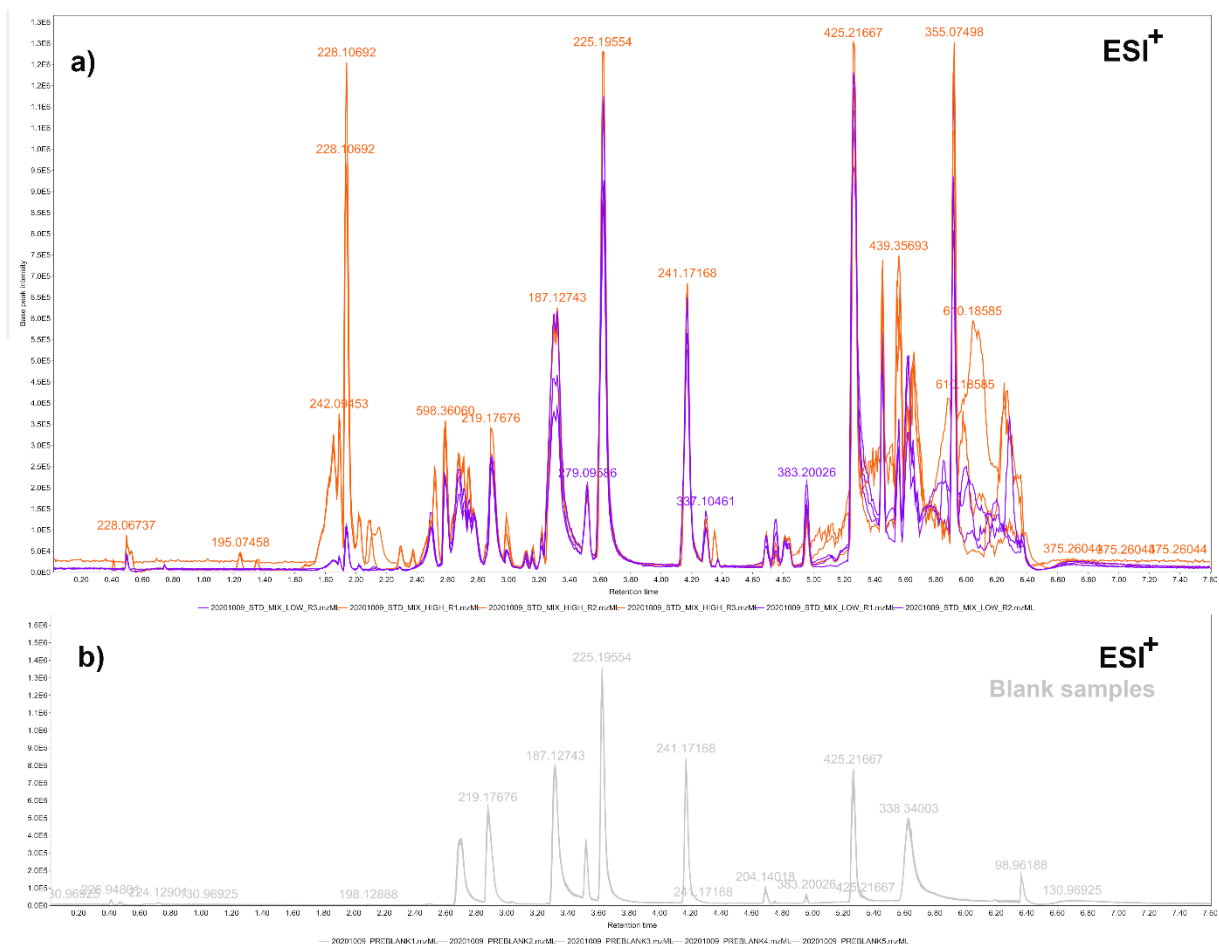


Fig. S13 | LC-HRMS-MS^E/DIA metabolic fingerprints shown as Base Peak Ion (BPI) chromatograms displaying in purple the overlapped actinobacterial extract spiked with chemical authentic standards in high (purple) and low (orange) concentration. (b) the overlapped blank replicates in the grey colour. The actinobacterial extract and blank replicates were acquired only in positive electrospray ionization. The overlay of the BPI traces for 6 samples and 5 blank replicate injections demonstrates highly reproducible chromatograms.

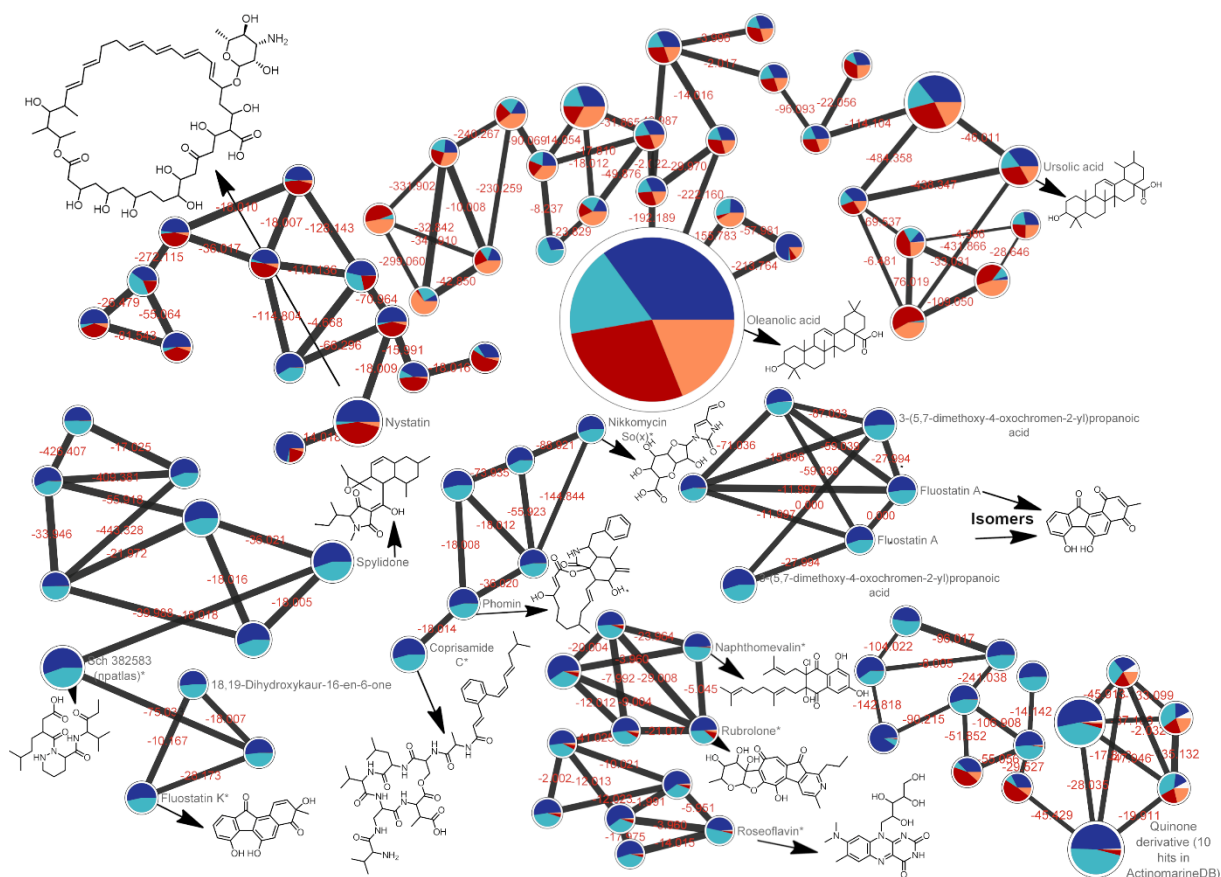
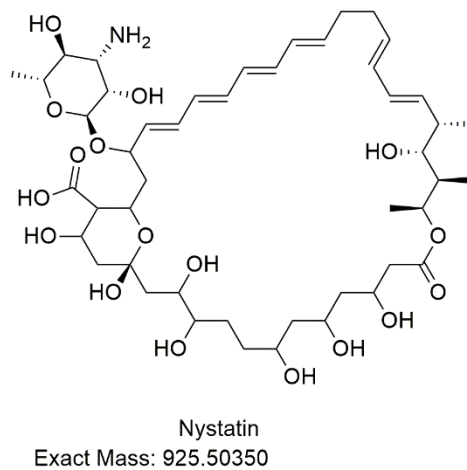
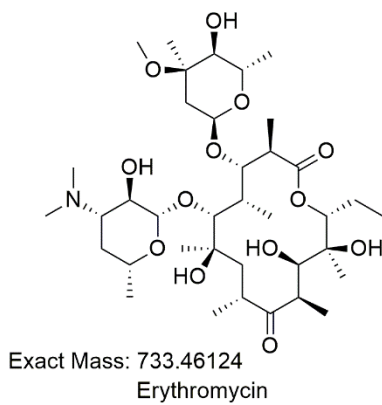
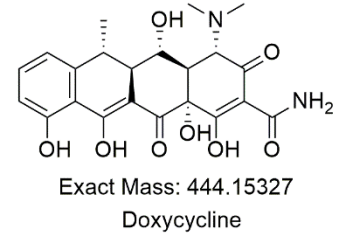
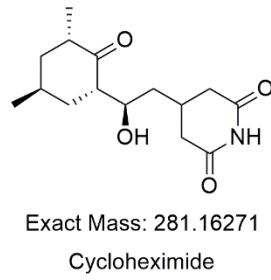
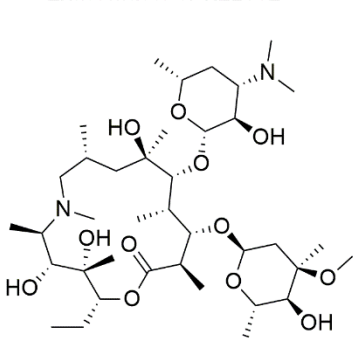
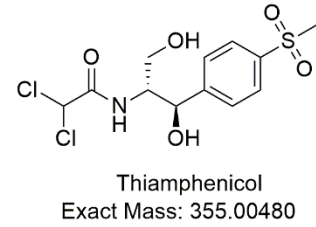
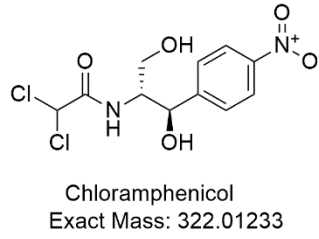
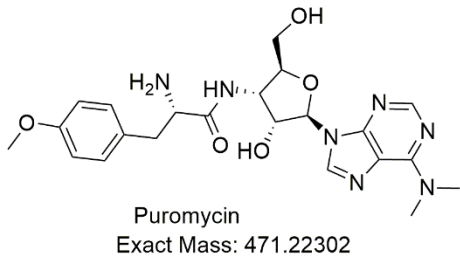
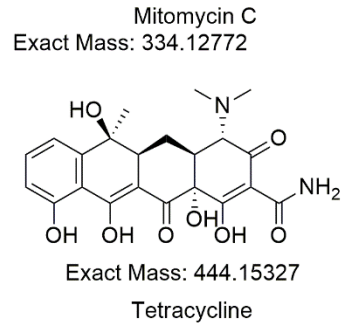
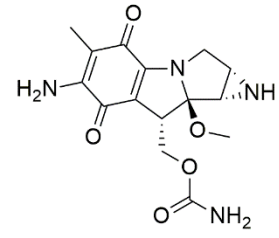
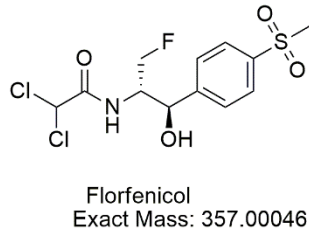
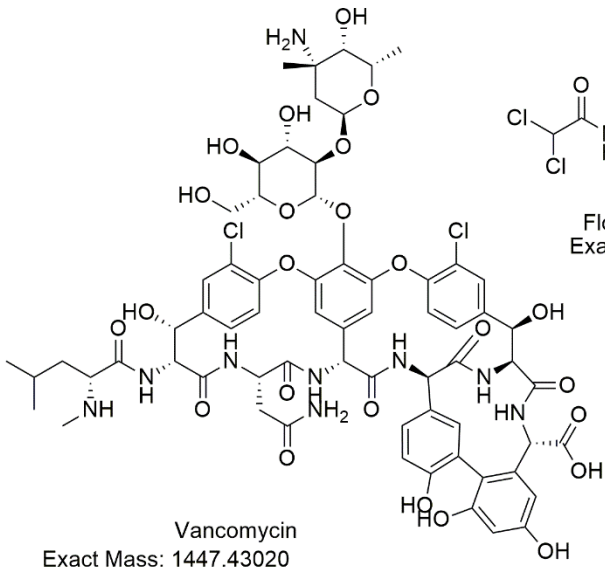
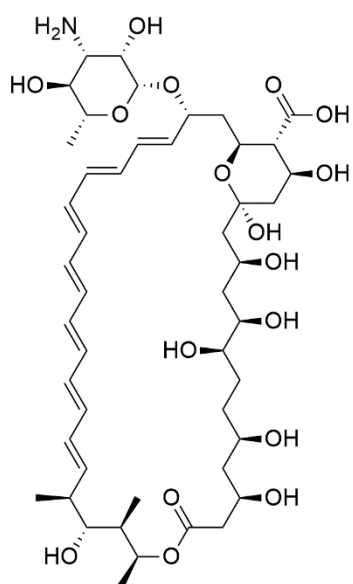
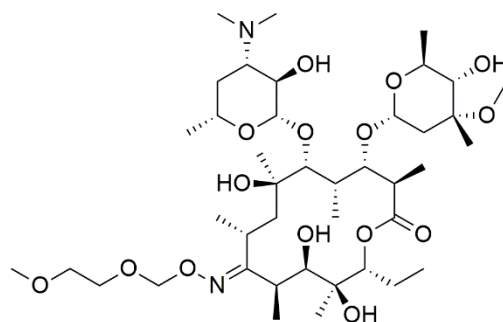


Fig. S14 | Molecular families derived from the Feature-Based Molecular Networking (FBMN) for the Actinobacterial dataset. Several marine NP were annotated with levels 2 and 3 of confidence using GNPS and MoNA spectral matches, and *in-house ActinomarineDB*. Besides, nystatin, oleanolic acid and ursolic acid were annotated as chemical standards spiked in the extracts. Pie chart colours refer to node distribution in extracts and standard samples. Blue colours represent actinobacterial extracts spiked with high (blue dark) and low (light blue) standard concentrations. Red and orange colours represent the pooled chemical standards samples, in high (red) and low (orange) concentrations, respectively. Fluostatin A isomers were clustered and automatically annotated as fluostatin A from GNPS libraries. Whereas, the coprisamide C, phomin and nikkomyacin So(x) were clustered and annotated at level 3 of confidence by the monoisotopic mass match with *ActinomarineDB*. As well fluostatin K and spylidone, alongside the roseoflavin, rubrolone, and naphthomevalin MN clusters. The bottom right cluster family presented a node related to ten level three annotations from *ActinomarineDB*, which were annotated as quinone derivatives, e.g 8-*O*-methyltetragomycin, fujianmycin B, hatomarugibin A and B, zumbericin A, brasiliquinone B and homo-dehydrabelomycin E, rubiginone M, and two others NP not named from our *in-houseDB*. *Level 3 of confidence annotations.

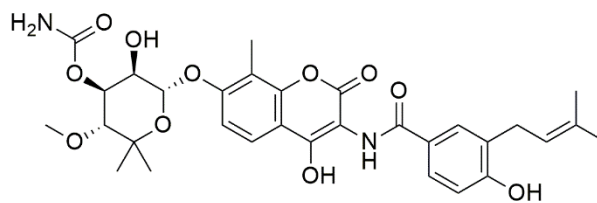




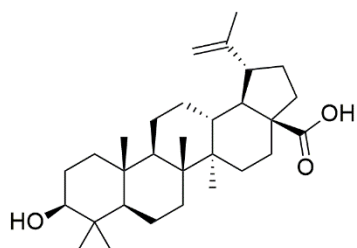
Exact Mass: 923.48785
Amphotericin B



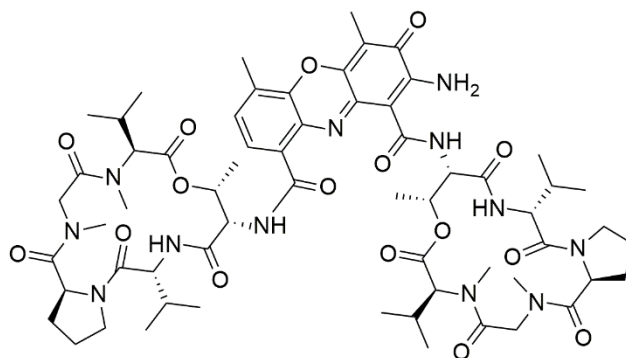
Roxithromycin
Exact Mass: 836.52457



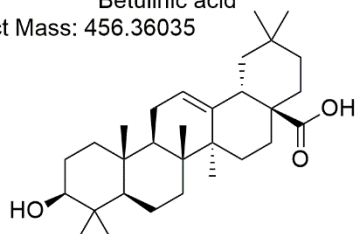
Novobiocin
Exact Mass: 612.23191



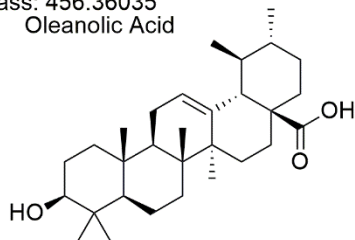
Betulinic acid
Exact Mass: 456.36035



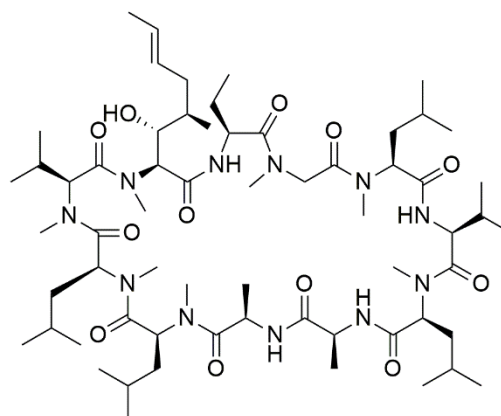
Actinomycin D
Exact Mass: 1254.62847



Exact Mass: 456.36035
Oleanolic Acid

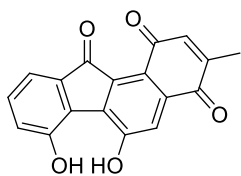


Ursolic acid
Exact Mass: 456.36035



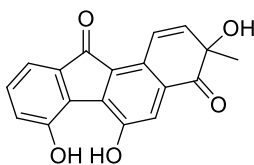
Cyclosporin A
Exact Mass: 1201.84137

Fig. S15 | All chemical structures of standards annotated from the dataset 2.



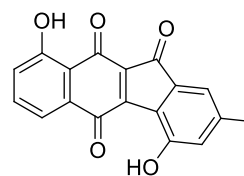
Exact Mass: 306.05282

Fluostatin A



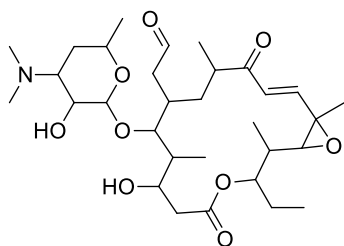
Exact Mass: 308.06847

Fluostatin K



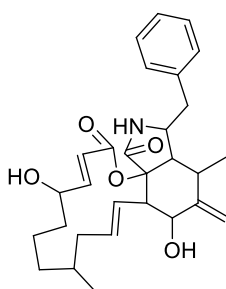
Exact Mass: 306.05282

Kinobscurinone



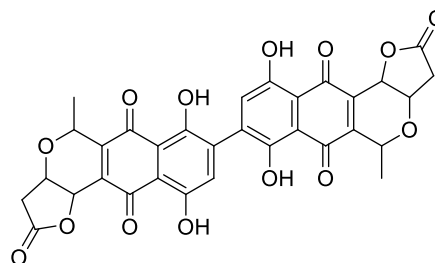
Exact Mass: 581.35638

Rosamicin



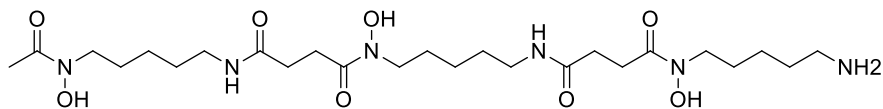
Exact Mass: 479.26717

Phomin



Exact Mass: 630.10096

gamma-actinorhodin



Desferrioxamine B

Exact Mass: 560.35336

Fig. S16 | Chemical structures of metabolites annotated with level 2 of confidence based on *ActinomarineDB*.

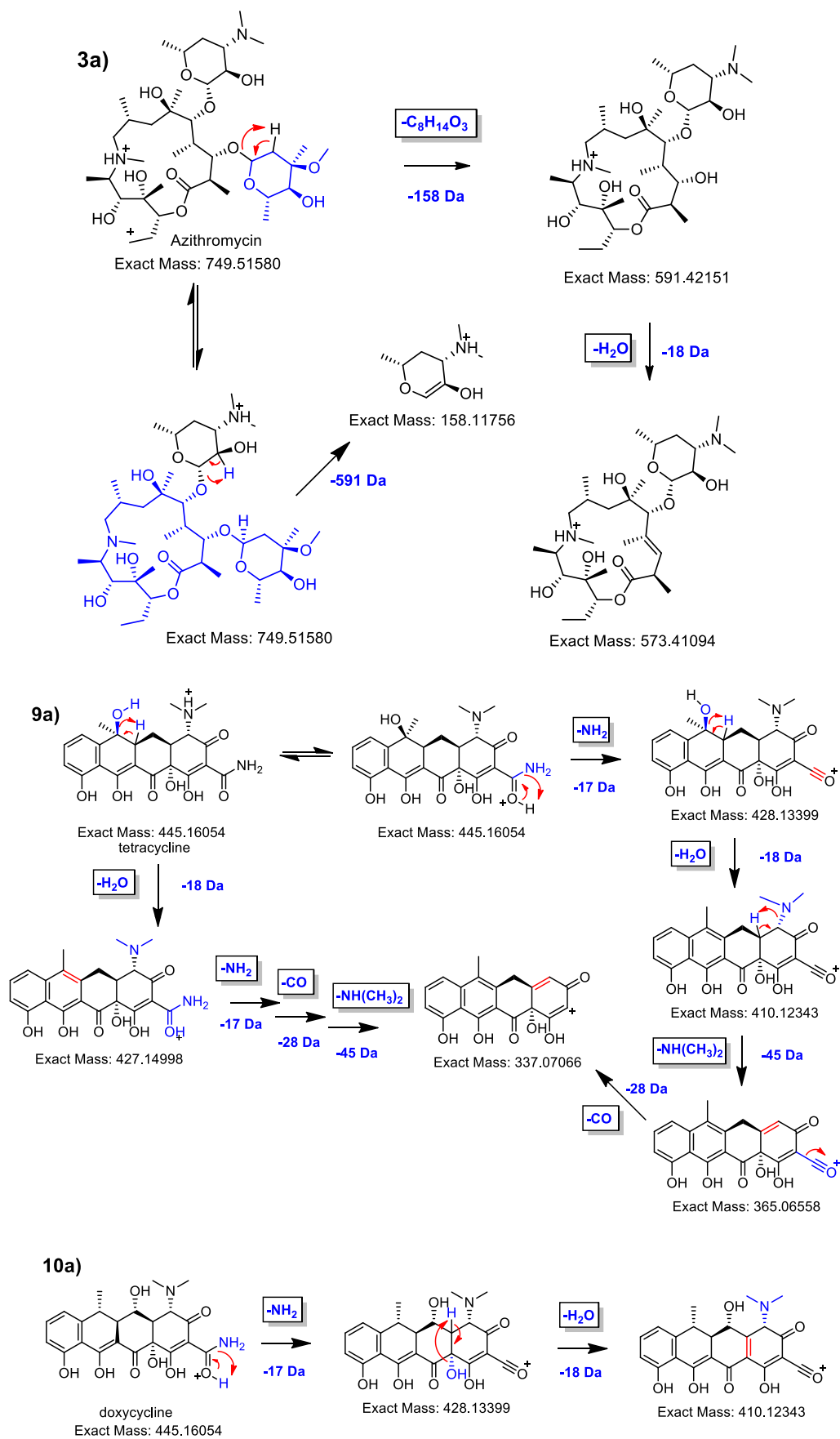


Fig. S17 | Main proposed fragmentation pattern of some annotated chemical standards used for dataset 2.

3a) Azithromycin. 9a) Tetracycline. 10a) Doxycycline.

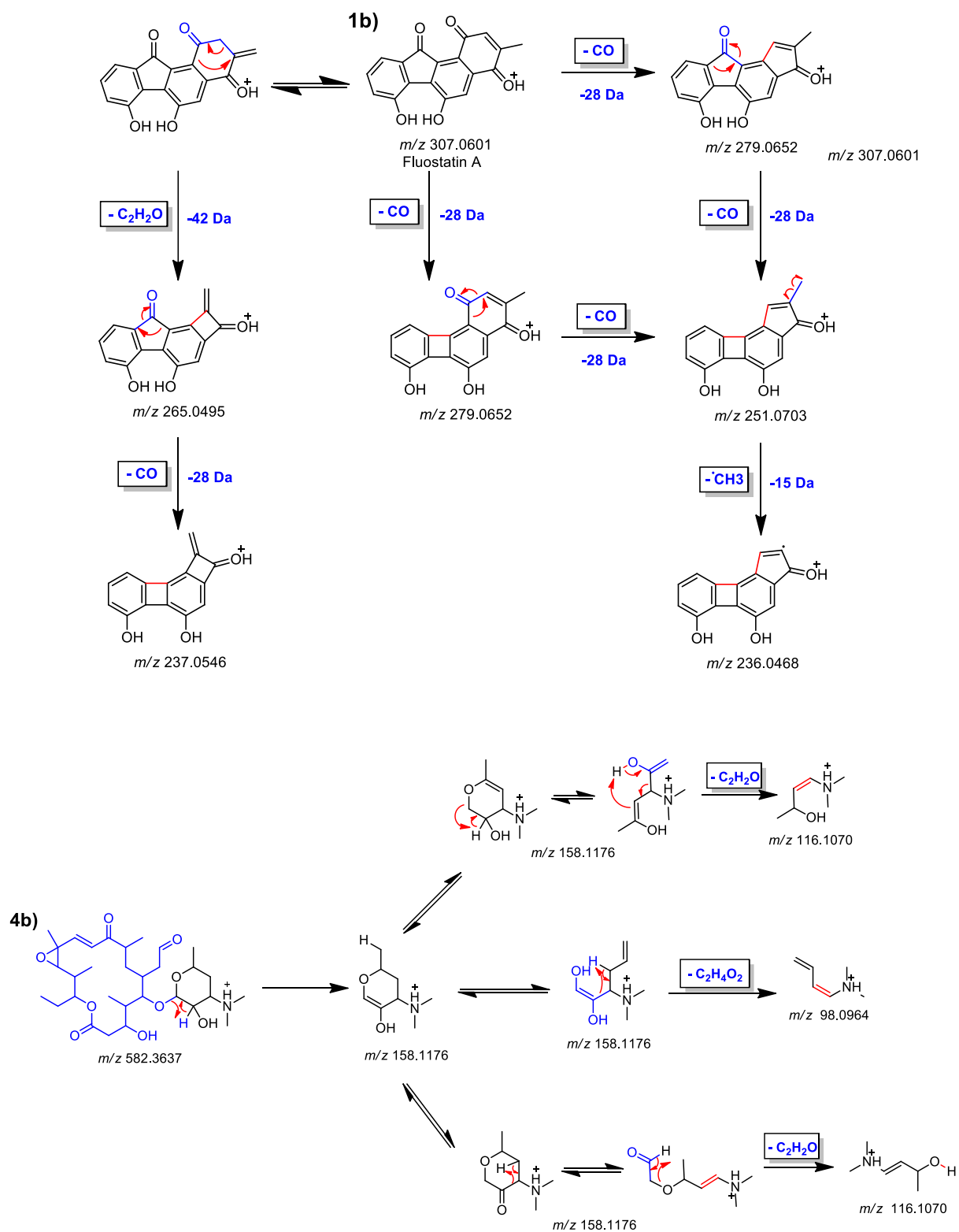


Fig. S18 | Main proposed fragmentation pattern of some level 2 annotated metabolites from the ActinomarineDB used for dataset 2. 1b) Fluostatin A. 4b) Rosimicin.

Table S1 | Details of metabolite annotation with level 2 of confidence for the *Ocotea* dataset. Putative annotations of metabolites identified by matching accurate mass, isotopic pattern, and fragmentation spectra to public databases literature sources and chemical knowledge. Putative name, cLogP, molecular formula, observed m/z , adducts, main fragments, database sources, and supporting international identifier (InChIKey) are provided.

ID	Putative metabolite name	Observed m/z	Adduct	MS/MS fragments	Reference	InChIKey
1	4-hydroxy- <i>N</i> -methylproline	146.08119	[M+H] ₊	82.06643; 100.07568	Proposed	FMIPNAUMSPFTHK-UHNVWZDZSA-N
2	Crotsparine	284.12770	[M+H] ₊	118.06763; 146.06143; 152.06110; 165.07108; 178.07968; 235.0754; 267.1027;	GNPS	LCAZZISCNMBVKG-LBPRGKRZSA-N
3	Glaziovine	298.14359	[M+H] ₊	141.07013; 152.06110; 165.06844; 167.08461; 178.07695; 194.07216; 235.0754; 267.1027;	GNPS	PNJUPRNTSJWJWAX-UHFFFAOYSA-N
4	<i>N</i> -methylcoclaurine	300.15932	[M+H] ₊	103.05380; 107.04881 115.05604; 165.07108; 194.07501	GNPS	BOKVLBSSPUTWLV-INIZCTEOSA-N
5	3-hydroxynornuciferine	298.14355	[M+H] ₊	152.06110; 165.07108; 177.06862; 178.07695; 189.06778; 195.07916; 205.06406; 218.07373; (281.14987)	GNPS	AOGVVFDNSYRXJL-CYBMUJFWSA-N
6	Laurelliptine	314.13850	[M+H] ₊	151.07413; 237.07719; 238.08591; 255.08789; 265.07416; 266.08069; 283.08377; 297.09570; 298.10825	Proposed	HORZNQYQXBFWNZ-UHFFFAOYSA-N
7	Laurolitsine	314.13853	[M+H] ₊	165.06844; 177.06862; 194.07216; 205.06406; 211.07594; 222.06624	GNPS	KYVJVURXKAZJRK-LBPRGKRZSA-N
8	Pallidine	328.15444	[M+H] ₊	153.06898; 163.08540; 165.06844; 168.05618; 177.06862; 178.08514; 205.06406; 211.07594; 227.07082; 239.07147; 251.06995; (297.11334)	GNPS	FBCNBECEGOCMPI-LIRRHJNSA-N
9	Flavinantine	328.15448	[M+H] ₊	134.05945; 163.06190; 178.08514; 205.06113; 297.14505	Proposed	GSNZKNRMDZYEAI-AUUYWEPGSA-N
10	Boldine	328.15447	[M+H] ₊	165.07108; 177.06862; 197.07216; 205.06406; 222.06624; 237.08978; (297.09924)	GNPS	LZJRNLRASBVRRX-ZDUSSCGKSA-N

11	Isoboldine	328.15440	[M+H] +	165.06844; 177.06862; 194.07216; 205.06406; 222.06624; 237.08978; (297.09924)	GNPS	LINHZVMHXABQLB-ZDUSSCGKSA-N
12	Corytuberine	328.15438	[M+H] +	165.06844; 178.07695; 189.07059; 191.08656; 193.06491; 194.07216; 205.06406; 222.06624; 239.07147	GNPS	WHFUDAOCYRYAKQ-LBPRGKRZSA-N
13	Lauroschooltzine	342.17014	[M+H] +	165.07108; 179.08540; 194.07216; 207.08092; 210.06694; 222.06624; 237.09293; 265.08417	GNPS	ZFLRVRLYWHNAEC-AWEZLNQCLSA-N
14	Reticuline	330.16918	[M+H] +	115.05604; 137.05959; 143.04938; 175.07408; 177.07678; 192.10281; 299.12604	MoNA	BHLYRWXGMIUIHG-HNNXBMFYSA-N
15	Arnepavine	314.17505	[M+H] +	103.05588; 107.04881; 115.05384; 121.06413; 131.04800; 151.07916; 237.09293	GNPS	ZBKFZIUXTWQTP-QGZVFWFLSA-N
16	Zenkerine	298.14367	[M+H] +	152.06110; 165.06844; 178.07695; 189.07059; 207.08092; 251.06995; (281.14987)	GNPS	RELZHBBKERFUJ-CQSZACIVSA-N
17	Tuduranine	298.14343	[M+H] +	152.06110; 165.06844; 178.07695; 189.07059; 207.06621; 235.05991; 281.11902	Proposed	KUECJOPWMRHEX-CQSZACIVSA-N
18	Diospirofoline	326.13859	[M+H] +	221.0586; 237.0898; 249.0552; 265.0875; 279.0586; 295.0969	Proposed	* CO[C@@H]1[C@@]2([H])N(CCc3c2c (c4c1cccc4O)c5c(OCO5)c3)C
19	Thaliporphine	342.16985	[M+H] +	152.06110; 165.07108; 178.07695; 189.07059; 193.06491; 205.06406; 207.08092; 221.05864; 251.10883	GNPS	SAERKXUSZPTMCQ-UHFFFAOYSA-N
20	Predicentrine	342.17007	[M+H] +	220.08670; 251.10559; 279.10300; 296.10599; 311.12857	MoNA	OUTYMWDDJORZOH-AWEZLNQCLSA-N

21	Nuciferine	296.16469	[M+H] +	165.06844; 178.07695; 179.08540; 189.07059; 190.07584; 191.08656; 207.08864; 219.08057; (265.12079)	GNPS	ORJVQPIHKOARKV-OAHLLOKOSA-N
22	Corydine	342.17042	[M+H] +	165.06844; 179.08540; 193.06491; 207.08092; 219.08057; 221.06168; 238.06065; 247.07423; 253.08632; 265.08417	GNPS	IDQUPXZJURZAGF-ZDUSSCGKSA-N
23	Domesticine	326.13834	[M+H] +	151.05402; 165.07108; 177.06862; 189.07059; 205.06406; 233.06061; 235.07558	GNPS	ZMNSHBTYBQNBPV-ZDUSSCGKSA-N
24	Dehydrodicentrine	338.13891	[M+H] +	165.07108; 177.06862; 205.06406; 235.07558; 263.07019; 295.09686;	Proposed	* COc1cc2cc3N(C)CCc4cc5OCOc5c(c2cc1OC)c34
25	Norisocorydine	328.15421	[M+H] +	152.06110; 165.06844; 178.07695; 189.06778; 191.08656; 207.07799; 219.08057; 221.05864; 236.08318; 238.06381; 247.07423; 253.08632; 265.08417; (311.12857)	GNPS	OHDQLTAYHMLRBA-LBPRGKRZSA-N
26	Laurotetanine	328.15429	[M+H] +	165.06844; 182.06990; 194.07216; 205.06406; 207.07799; 222.06624; 225.09052; 237.08978; 253.08632; 265.08417; 281.08130; (311.12857)	GNPS	GVVXPMORGFYVOO-ZDUSSCGKSA-N
27	Nordicentrine	326.13862	[M+H] +	152.06110; 165.06844; 179.08267; 189.07059; 205.06406; 207.08092; 221.05864; 235.07558; 263.07019	GNPS	YNWJEUJZYKLCJG-ZDUSSCGKSA-N
28	Nornantenine	326.13864	[M+H] +	165.06844; 177.06862; 189.07059; 205.06406; 223.07613; 235.07558; 251.06995; 263.07019; 279.06543	GNPS	JWXKBCGJLCEZTJ-UHFFFAOYSA-N

29	Nornuciferine	282.14870	[M+H] +	152.06110; 165.06844; 178.07695; 179.08540; 189.07059; 191.08656; 202.07614; 207.08092; 235.07558; (265.12411)	GNPS	QQKAHDMMPBQDAC-AWEZLNQCLSA-N
30	Lirinidine	282.14850	[M+H] +	141.06770; 152.06110; 165.06844; 178.07695; 179.08540; 189.07059; 190.07584; 191.08656; 207.08092; 218.07373; (251.10559)	GNPS	YXVXMURDCBMPRH-AWEZLNQCLSA-N
31	Glaucine	356.18513	[M+H] +	165.06844; 178.07695; 191.08374; 208.08569; 220.05029; 236.08318; 251.06670; 267.10580; 279.09958; 295.09686	GNPS	RUZIUYOSRDWYQF-HNNXBMFYSA-N
32	Roemerine	280.13274	[M+H] +	152.06110; 165.07108; 178.07695; 189.07059; 190.07584; 191.08656; 201.07150; 218.07071	GNPS	JCTYWRARKVGOBK-CQSZACIVSA-N
33	Nantenine	340.15406	[M+H] +	165.06844; 177.06862; 205.06406; 223.07307; 235.07558; 263.07019; 278.09259; 294.08945; 309.11176	MoNA	WSVWKHTVFGTTKJ-AWEZLNQCLSA-N
34	Dicentrine	340.15426	[M+H] +	251.06670; 278.09601; 279.06543; 280.06732; 309.11176	GNPS	YJWBWQWUHVXPNC-AWEZLNQCLSA-N
35	Dehydronuciferine	294.14903	[M+H] +	165.06844; 193.09901; 207.07799; 218.07071; 219.07755; 235.07558; 249.09068; 263.07019	Proposed	JBGSWIBJAGBOP-UHFFFAOYSA-N
36	Dicentrinone	336.08626	[M+H] +	206.05804; 234.05450; 263.05695; 275.05881; 292.06232; 320.05484	GNPS	NEQVOBXBOFZEMR-UHFFFAOYSA-N
37	Leucoxylinone	400.17563	[M+H] +	237.08978; 265.08749; 293.08026; 323.09140; 338.11688; 369.13336	Proposed	SFHWHWVEDBDXLV-AWEZLNQCLSA-N

38	Stephenanthrine	294.14881	[M+H] +	165.06844; 178.07695; 189.07059; 191.08374; 201.06860; 219.08057; 249.09068	Proposed	FXTBDJZGDJJCQU-UHFFFAOYSA-N
39	Argentinine	296.16515	[M+H] +	152.06110; 165.06844; 179.08540; 189.07059; 190.07584; 191.08656; 208.08864; 217.06319	GNPS	HCXNUWJYBNHDAE-UHFFFAOYSA-N
40	Thalichthuberine	354.16966	[M+H] +	165.07108; 177.05501; 189.06778; 205.06406; 223.07307; 233.06061; 235.07558; 251.06995; 278.09259; 279.06543; 294.08945	GNPS	DDCILWXYWBKXKC-UHFFFAOYSA-N
41	Discretamine	328.15342	[M+H] +	135.05959; 146.05649; 151.07413; 163.06190; 178.08514; 207.06621	GNPS	KNWVMRVOBAFFMH-UHFFFAOYSA-N
42	Sesamin	355.11784	[M+H] +	91.05564; 197.06081; 247.07423; 355.11987	MoNA	PEYUIKBAABKQKQ-AFHBHXEDSA-N
43	Ocophylol B	359.18565	[M+H] +	123.04333; 151.07413; 165.08946; 177.09039; 189.09027; 229.12169	Proposed	* COc1cc(cc(OC)c1OC)c2oc3c(OC)cc(C=O)cc 3c2C
44	Eusiderin	387.17956	[M+H] +	151.07413; 165.05530; 177.05501; 189.08746;	Proposed	* COc1c(OC)c(OC)cc([C@H]2Oc3c (O[C@@H]2C)c(OC)cc(CC=C)c3)c1
45	Licarin B	325.14306	[M+H] +	107.04881; 135.04391; 143.08607; 151.07413; 162.06853; 163.07495; 203.10658	Proposed	DMMQXURQRMNSBM-YZAYTREXSA-N
46	Licarin A	327.15865	[M+H] +	137.05719; 143.08607; 151.07413; 165.06844; 178.07695; 171.08144; 188.04839; 193.09901; 203.06870; 221.09512; 295.13553	MoNA	ITDOFWOJEDZPCF-FNINDUDTSA-N
47	Armenin B	373.16390	[M+H] +	153.09177; 165.05855; 179.03342; 203.07161; 229.08763; 241.08679; 257.08096; 341.13953; 379.17303	Proposed	SSPDVRMNHFFRCE-JDEKTIPCSA-N

48	Quinic acid	191.05433	[M-H]	85.02814; 93.03156; 127.03786; 137.02254; 171.02654; 173.04324; 191.05431	MoNA	AAWZDTNXLSGCEK-LNVDRNJUSA-N
49	Taxifolin	303.04983	[M-H]	125.02214; 137.02254; 153.01779; 175.03928; 217.04663; 241.04916; 259.05661; 275.05774; 285.03790	GNPS	CXQWRCVTCMQVQX-LSDHHAIUSA-N
50	Catechin / Epicatechin	289.07091	[M-H]	93.0156; 109.02729; 123.04298; 137.02254; 159.04279; 187.03732	MoNA	PFTAWBLQPZVEMU-DZGCQCFKSA-N / PFTAWBLQPZVEMU-UKRRQHHQSA-N
51	Isoquercitrin	463.08689	[M-H]	151.00093; 243.02718; 255.02904; 271.02179; 300.02765; 463.0877	MoNA	OVSQVDMCBVZWGM-QSOFNFLRSA-N
52	Vitexin-2'-O-rhamnoside	577.15701	[M-H]	293.04349; 311.05334; 341.06476; 413.08524; 457.11850	GNPS	LYGPBZVKGHHTIE-HUBYJIGHSA-N
53	Rutin	609.14699	[M-H]	151.00345; 178.99718; 243.02718; 255.02904; 271.02179; 300.02408; 301.03207	GNPS	IKGXIBQEEMLURG-NVPNHPEKSA-N
54	Quercimeritrin	463.08762	[M-H]	199.04004; 227.03430; 243.02718; 255.02904; 271.02515; 289.07208; 300.02765; 301.03207; 463.08667	Proposed	BBFYUPYFXSSMNV-HMGRVEAOSA-N
55	Vitexin	431.09731	[M-H]	269.04352; 283.06070; 311.05695; 323.05396; 341.06476	GNPS	SGEWCQFRYRRZDC-VPRICQMDSA-N
56	Quercitrin	447.09299	[M-H]	151.00345; 227.03430; 243.02718; 255.02904; 271.02515; 300.02765; 447.09302	GNPS	OXGUCUVFOIWWQJ-HQBVPOQASA-N
57	Reynoutrin	433.07686	[M-H]	151.00093; 199.04004; 227.03122; 227.03122; 243.02718; 255.02904; 271.02179; 300.02408	MoNA	PZZRDJXEMZMZFDBWYUNELBSA-N

58	Astragalin	447.09298	[M-H] ⁺	183.04088; 227.03430; 255.02904; 284.03290	MoNA	JPUKWEQWGBDDQB-QSOFNFLRSA-N
59	Apigenin-7-O-rutinoside	577.15675	[M-H] ⁺	117.03406; 268.03702; 269.0475	MoNA	FKIYLTVPDLUDL-SLNHTJRHSA-N
60	Schaftoside / Isoschaftoside	563.14164	[M-H] ⁺	353.06760; 383.08264; 413.08524; 443.09760; 473.10629; 563.14203	GNPS	MMDUKUSNQNWVET-VYUBKLCTSA-N / OVMFOVNOXASTPA-VYUBKLCTSA-N
61	Afzelin	431.09706	[M-H] ⁺	107.01432; 183.04364; 211.03810; 227.03430; 255.02904; 284.02945; 285.0238	MoNA	SOSLMHZOJATCCP-AEIZVZFYSA-N
62	Kaempferol 3-4"-p-coumarylrhamnoside	593.12992	[M-H] ⁺	163.0387; 285.0379; 291.0841;	Proposed	RFTKNPGPPJOBI-NSPOHDESEA-N
63	Quercetin	301.03286	[M-H] ⁺	107.01220; 121.02779; 151.00093; 227.03430; 243.02718; 271.02179; 301.03207	MoNA	REFJWTPEDVJJIY-UHFFFAOYSA-N
64	Apigenin	269.04407	[M-H] ⁺	117.03185; 149.02243; 151.00093; 201.05260; 225.05504; 227.03430; 269.04352	GNPS	KZNIHFPLKGYRTM-UHFFFAOYSA-N
65	Kaempferol	285.03873	[M-H] ⁺	143.04767; 159.04279; 171.04260; 187.03732; 201.05550; 211.03810; 227.03430; 239.05389; 285.03790	GNPS	IYRMWMYZSQPKC-UHFFFAOYSA-N
66	Kaempferol 3-(2",4"-di-(E)-p-coumarylrhamnoside) / Kaempferol 3-(3",4"-di-(E)-p-coumarylrhamnoside)	723.17282	[M-H] ⁺	145.02815; 187.03732; 229.05013; 255.02904; 284.03290; 285.03790; 437.12308; 577.13794; 723.17230	GNPS	KMOHJUXDKSMQOG-OLHCXIDTSA-N

#Fragment ions between brackets were not found in the reference spectra but were proposed based on chemical knowledge of MS fragmentation and observed MS^E spectra. *SMILES is provided, once referred InChkey is not available.

Table S2 | Metadata for *Ocotea* plant samples. Including collection details, species information, location, geographical coordinates, related activity, and references. Provides contextual information about analyzed samples to enable biological interpretation of metabolomics data.

Code	Date	Specie	Popular name	Synonym	Location	Geographical location	Related activity	Reference
AM	1935	<i>Ocotea amazonica</i> (Meiss) Mez	unknown	X	Fazenda da cachoeira, Tombos, Minas Gerais, Brazil (Jardim Botânico de Belo Horizonte)	-	No reported activity	x
AU	2013	<i>Ocotea acutifolia</i> (Nees) Mez	canela-branca	X	Santa Rosa, Ruta Nacional 118, Km 72, Concepción, Corrientes, Argentina	72°28'18.0"S 58°08'21.0" W	Cytotoxic activity against human cancer cell lines and mutagenic and genotoxic effects on wing cells of <i>Drosophila melanogaster</i>	Garcez <i>et al.</i> (2011) and Guterres <i>et al.</i> (2013)
AY	2005	<i>Ocotea aciphylla</i> (Nees & Mart.) Mez	canela-amarela	X	Serra das Almas - Inácio Pinto, Sítio Gaia da Mata, Rio de Contas, Bahia, Brazil	13°32'14.0"S 41°54'14.0" W	<i>In vitro</i> inhibitory activity over AChE and acaricidal activity against <i>Rhipicephalus (Boophilus) microplus</i>	Carneiro <i>et al.</i> (2018) and Conceição <i>et al.</i> (2017)
BA	2004	<i>Ocotea brachybotrya</i> (Meisn.) Mez	louro-verdadeiro	x	Paque estadual do vale do rio doce, Timóteo, Minas Gerais, Brazil (ICB-UFGM)	19°35'28.0"S 42°34'07.0" W	No reported activity	x
BI	2005	<i>Ocotea bicolor</i> Vattimo-Gil	canela-preta	x	Parque Estadual de Ibitipoca, Floresta no acero do parque, Lima Duarte, Minas Gerais, Brazil	-	<i>In vitro</i> antioxidant activity	Damasceno <i>et al.</i> (2017)

BR	2001	<i>Ocotea bragai</i> Coe-Teix.	unknown	x	Parque Estadual da Cantareira, Região das águas, Mariporã, São Paulo, Brazil	-	No reported activity	x
CA	2007	<i>Ocotea caesia</i> Mez	unknown	x	Serra de Antônio Pereira, Samarco. Alegria 7, Ouro Preto, Minas Gerais, Brazil	-	No reported activity	x
CE	2003	<i>Ocotea cernua</i> (Nees) Mez	moena negra	x	Morro do Imperador, Juiz de Fora, Brazil	-	No reported activity	x
CJ	2003	<i>Ocotea cujumary</i> Mart.	cuchumari	x	APA logoa Silvana, Caratinga, Minas Gerais, Brazil	-	<i>In vitro</i> antibacterial and cytotoxic activities against <i>E. coli</i> and MCF-7 cells, respectively	Da Silva <i>et al.</i> (2017)
CL	1974	<i>Ocotea calliscypha</i> L.C.S.Assis & Mello-Silva	unknown	x	Serra do Frazão, Ouro Preto, Minas Gerais, Brazil	20°17'15.0"S 43°30'19.1" W	No reported activity	x
CM	2003	<i>Ocotea complicata</i> (Meisn.) Mez	unknown	x	Rodovia BA-001, Una, Bahia, Brazil (Prefeitura de Curitiba)	-	<i>In vitro</i> leishmanicidal activity	Rebouças-Silva <i>et al.</i> (2023)
CO	2002	<i>Ocotea corymbosa</i> (Meisn.)Mez	canela-fedida	x	Eldorado, Mato Grosso do Sul, Brazil	-	No reported activity	x
CT	2002	<i>Ocotea catharinensis</i> Mez	canela-coqueiro	x	mata atlântica	-	No reported activity	x
DI	1993	<i>Ocotea dispersa</i> (Nees & Mart.) Mez	canela-sassafrás	x	Serra do Itacolomi, Ouro Preto, Minas Gerais, Brazil	-	<i>In vitro</i> antileishmanial activity	Alcoba <i>et al.</i> (2017)
DO	2004	<i>Ocotea diospyrifolia</i> (Meisn.) Mez	canela-louro	x	Mata do Baú, Barroso, Minas Gerais, Brazil	-	<i>In vivo</i> anti-inflammatory activity and <i>in vitro</i> antibacterial activity against <i>Salmonella</i> spp.	Silva <i>et al.</i> (2021) and Weber <i>et al.</i> (2018)
DV	2003	<i>Ocotea divaricata</i> (Nees) Mez	canela-segueira	x	Reserva Biológica da Represa do Grama, Descoberto, Minas Gerais, Brazil	-	No reported activity	x
EL	2001	<i>Ocotea elegans</i> Mez	canela-broto	<i>Ocotea indecora</i> (Schott) Mez	Estrada para São Mateus, Camanducaia, Minas Gerais, Brazil (ICB-UFG)	-	<i>In vitro</i> acaricidal and repellent	Figueiredo <i>et al.</i> (2018)

							activity on <i>Rhipicephalus (Boophilus) microplus</i> and insecticidal activity against <i>Dysdercus peruvianus</i> with AChE inhibition	and Nascimento <i>et al.</i> (2020)
FE	1999	<i>Ocotea felix</i> Coe-Teix.	unknown	x	Parque Estadual do Itacolomi, Ouro Preto, Minas Gerais, Brazil	-	No reported activity	x
GA	1997	<i>Ocotea guianensis</i> Aubl.	canela-seda	x	Margem do Rio Xingú, São Jose do Xingú, Mato Grosso, Brazil	-	No reported activity	x
GL	1971	<i>Ocotea glauca</i> (Nees & Mart.) Mez	louro	x	Santa Rita Durão, Mariana, Minas Gerais, Brazil	20°22'40.0"S 43°24'57.9" W	No reported activity	x
GU	2001	<i>Ocotea glaucina</i> (Meisn.) Mez	unknown	<i>Ocotea notata</i> (Nees & Mart.) Mez	Estrada Grão Mongol-Cristália KM 6, Grão Mongol, Minas Gerais, Brazil (USP)	16°35'47.0"S 42°54'05.0" W	<i>In vitro</i> antiviral activity against <i>Herpes simplex</i> virus types 1 and 2, antioxidant activity, <i>in vitro</i> anti-mycobacterial and immunomodulatory activities and <i>in vitro</i> antifungal activity against <i>Sporothrix brasiliensis</i>	Garrett <i>et al.</i> (2012), Pereira <i>et al.</i> (2019), Costa <i>et al.</i> (2015 and 2021) and de Souza <i>et al.</i> (2023)
GZ	2006	<i>Ocotea glaziovii</i> Mez	canela-amarela*	x	Mananciais da Serra, Represa do Carvalinho, FOM/FODM, Piraquara, Paraná, Brazil	-	Anxiolytic and antiviral	x
HY	1993	<i>Ocotea hypoglauca</i> (Nees & Mart.) Mez	unknown	x	Parque Estadual do Itacolomi, Estrada para a Fazenda do Manso, Ouro Preto, Minas Gerais, Brazil	-	No reported activity	x
IN	2001	<i>Ocotea indecora</i> (Schott) Mez	canela	<i>Ocotea elegans</i> Mez	Reserva Biológica da Represa do Grama, Descoberto, Minas Gerais, Brazil	-	<i>In vitro</i> larvicidal effect against	Machado <i>et al.</i> (2023)

							<i>Aedes aegypti</i> larvae	
KU	1979	<i>Ocotea kuhlmannii</i> Vattimo-Gi/ <i>Ocotea nectandrifolia</i> Mez	canela-burra	<i>Ocotea nectandrifolia</i> Mez	Ilha de Santa Catarina - Lagoa do Peri, Altitude: 300m., Florianópolis, Santa Catarina, Brazil	-	Antifungal- <i>Candida</i> sp	x
LA	2010	<i>Ocotea lanata</i> (Nees & Mart.) Mez	unknown	x	Rancho das Tábuas, Angelina, Paraná, Brazil (FURB)	27°37'49.0"S 49°02'58.0" W	No reported activity	x
LC	2003	<i>Ocotea lancifolia</i> (Schott) Mez	canela-sabão	<i>Ocotea lanceolata</i> (Nees) Nees; <i>Ocotea variabilis</i> Mart.	Parque do São Gonçalo do Rio Preto, São Gonçalo do Rio Preto, Minas Gerais, Brazil (ICB-UFMG)	18°06'54.0"S 43°20'28.0" W	<i>In vitro</i> acaricidal activity on <i>Rhipicephalus (Boophilus) microplus</i> , <i>in vitro</i> antifungal activity against <i>Fusarium moniliforme</i> (a), <i>Trametes versicolor</i> and <i>Gloeophyllum trabeum</i> (b), <i>in vitro</i> antioxidant activity and <i>in vitro</i> antiprotozoal activities against <i>Leishmania</i> spp. and <i>Trypanosoma cruzi</i>	Barbosa <i>et al.</i> (2013), da Silva <i>et al.</i> (2018) (a) and (2017) (b), and Fournet <i>et al.</i> (2007)
LF	1992	<i>Ocotea longifolia</i> Kunth	louro-ingá	x	Alto do Galo, Domingos Martins, Espírito Santo, Brazil	-	<i>In vitro</i> cytotoxic activities against cancer cell lines HepG2 and HL60 and insecticide activity against <i>Sitophilus zeamais</i>	da Silva <i>et al.</i> (2016) and Prieto <i>et al.</i> (2010)
LG	1938	<i>Ocotea langsdorffii</i> (Meisn.) Mez	unknown	x	Serra do Cipó Minas Gerais, Brazil	19°15'30.0"S 43°33'04.0" W	No reported activity	x

LN	2009	<i>Ocotea lanceolata</i> (Nees) Nees	canela-pilosa	<i>Ocotea lancifolia</i> (Schott) Mez	Parque Nacional do Iguaçu, Foz do Iguaçu, Paraná, Brazil (Prefeitura Municipal de Curitiba)	25°32'52.0"S 54°35'17.1" W	<i>In vitro</i> acaricidal activity on <i>Rhipicephalus (Boophilus) microplus</i> , <i>in vitro</i> antifungal activity against <i>Fusarium moniliforme</i> (a), <i>Trametes versicolor</i> and <i>Gloeophyllum trabeum</i> (b), <i>in vitro</i> antioxidant activity and <i>in vitro</i> antiprotozoal activities against <i>Leishmania spp.</i> and <i>Trypanosoma cruzi</i>	Barbosa <i>et al.</i> (2013), da Silva <i>et al.</i> (2018) (a)and (2017) (b), and Fournet <i>et al.</i> (2007)
LO	2003	<i>Ocotea lobbii</i> (Meisn.) Rohwer	unknown	x	Canavieiras, Bahia, Brazil (Prefeitura Municipal de Curitiba)	22°05'21.1"S 43°49'40.0" W	No reported activity	x
LX	1997	<i>Ocotea laxa</i> (Nees) Mez	canela-pimenta	x	Parque Estadual do Itacolomi, Córrego do Belchior, , Ouro Preto, Minas Gerais, Brazil	20°17'15.0"S 43°30'19.0" W	No reported activity	x
MI	2003	<i>Ocotea minarum</i> (Nees & Mart.) Mez	canela-vassoura	x	Fazenda Renascença, Bonito, Mato grosso do Sul, Brazil		<i>In vitro</i> antioxidant activity	Rodrigues <i>et al.</i> (2019)
MU	2007	<i>Ocotea nummularia</i>	canelinha	x	Serra da Piedade, Caeté, Minas Gerais, Brazil	19°52'47.9"S 43°40'10.9" W	No reported activity	x
NE	2010	<i>Ocotea nectandrifolia</i> Mez	canela-burra	x	Anta Branca (antigo Alto Rio do Oeste), Rio do Campo, Santa Catarina, Brazil (FURB)	26°54'36.0"S 50°13'13.0" W	No reported activity	x
NI	2003	<i>Ocotea nitida</i> (Meisn.) Rohwer	louro*	x	Morro do Gavião, Dionísio, Minas Gerais, Brazil	19°50'03.0"S 42°33'07.0" W	No reported activity	x

NO	2010	<i>Ocotea notata</i> (Nees & Mart.) Mez	louro-pipoca	<i>Ocotea glaucina</i> (Meisn.) Mez	Fazenda Lucuri, Serra do Curral Frio, Umburanas, Bahia, Brazil	10°43'58.0"S 41°19'35.0" W	<i>In vitro</i> antiviral activity against <i>Herpes simplex</i> virus types 1 and 2, antioxidant activity, <i>in vitro</i> anti-mycobacterial and immunomodulatory activities and <i>in vitro</i> antifungal activity against <i>Sporothrix brasiliensis</i>	Garrett <i>et al.</i> (2012), Pereira <i>et al.</i> (2019), Costa <i>et al.</i> (2015 and 2021) and de Souza <i>et al.</i> (2023)
NT	2008	<i>Ocotea nitidula</i> (Nees et Mart. ex Ness)	unknown	x	Parque Estadual do Itacolomi, Ouro Preto, Minas Gerais,	-	No reported activity	x
NU	2009	<i>Ocotea nutans</i> (Nees) Mez	unknown	x	Brazil, Minas Gerais, Mariana, Parque Estadual do Itacolomi	20°17'15.0"S 43°30'19.0" W	<i>In vitro</i> activity against <i>Aedes aegypti</i> larvae and antioxidant activity	Betim <i>et al.</i> (2019) and (2021)
OD	2004	<i>Ocotea odorifera</i> Vell. Rohwer	canela-sassafrás	<i>Ocotea pretiosa</i> (Nees) Mez	Reserva Biológica da Represa do Grama, Descoberto, Minas Gerais, Brazil	-	<i>In vivo</i> anti-inflammatory activity, <i>in vitro</i> antibacterial activity against <i>Staphylococcus aureus</i> (a, b and c), <i>in vitro</i> antioxidant and antimutagenic activities (a), <i>in vitro</i> antifungal activity against <i>Candida parapsilosis</i> , <i>in vitro</i> antileishmanial activity and insecticidal and repellent	Alcantara <i>et al.</i> (2021), Gontijo <i>et al.</i> (2017) (a), de Almeida <i>et al.</i> (2020) (b) and (2022) (c), Yamaguchi <i>et al.</i> (2010), Alcoba <i>et al.</i> (2017) and Mossi <i>et al.</i> (2013)

							activities against <i>Sitophilus zeamais</i>	
PA	2011	<i>Ocotea paranaenses</i> Brotto, Baitello, Cervi & E.P.Santos	unknown	x	Morro dos Perdidos, Serra da Araçatuba, Guaratuba, Paraná, Brazil	25°52'58.0"S 48°34'28.9" W	<i>In vitro</i> antibacterial activity against <i>Staphylococcus aureus</i> and antioxidant activity	Gribner <i>et al.</i> (2022)
PC	2008	<i>Ocotea pulchella</i> (Nees & Mart.) Mez	canela-lageana	x	Pedra Branca, Pocinhos do Rio Verde, Caldas, Minas Gerais, Brazil	21°55'24.9"S 46°23'09.9" W	<i>In vitro</i> antifungal activity against <i>Sporothrix brasiliensis</i> , <i>in vitro</i> antioxidant activity, <i>in vitro</i> antiviral activity against SuHV-1, and molluscicidal and antiparasitic effects on <i>Biomphalaria glabrata</i> and <i>Schistosoma mansoni</i> , respectively	de Souza <i>et al.</i> (2023), Reboucas <i>et al.</i> (2015), Padilla <i>et al.</i> (2018) and Passos <i>et al.</i> (2020)
PE	1998	<i>Ocotea percoriacea</i> Kosterm.	unknown	x	Cachoeira das Androinhas, Estrada à esquerda, Ouro Preto, Minas Gerais, Brazil	20°17'15.0"S 43°30'19.1" W	<i>In vitro</i> inhibitory activity over AChE	Cassiano <i>et al.</i> (2019)
PH	2010	<i>Ocotea pulchra</i> Vattimo-Gil	canela	x	Rio Veado, Nova Trento, Santa Catarina, Brazil	27°21'38.0"S 49°08'13.0" W	No reported activity	x
PL	2006	<i>Ocotea pulchea</i> Vattimo-Gil	unknown	x	Reserva Biológica Municipal Santa Cândida, Juiz de Fora, Minas Gerais, Brazil	-	No reported activity	x

PO	1906	<i>Ocotea pomaderroides</i> (Meisn.) Mez	canela	x	Miguel Burneir, Ouro Preto, Minas Gerais, Brazil	-	In vitro inhibitory activity over AChE	Reis <i>et al.</i> (2022)
PR	1995	<i>Ocotea porosa</i> (Nees & Mart.) Barroso	imbuia	x	Buraco do Padre, Ponta Grossa, Paraná (Prefeitura Municipal de Curitiba)	-	In vitro antibacterial activity against <i>Staphylococcus aureus</i> and in vitro cytotoxic effects on McCoy, B16F10 and MCF7 cell lines	Brito (2009) and Brutulim <i>et al.</i> (2020)
PT	1994	<i>Ocotea pretiosa</i> (Nees) Mez	canela-sassafrás	<i>Ocotea odorifera</i> Vell. Rohwer	Mata do Morro Redentor, Juiz de Fora, Minas Gerais, Brazil	-	In vivo anti-inflammatory activity, in vitro antibacterial activity against <i>Staphylococcus aureus</i> (a, b and c), in vitro antioxidant and antimutagenic activities (a), in vitro antifungal activity against <i>Candida parapsilosis</i> , in vitro antileishmanial activity and insecticidal and repellent activities against <i>Sitophilus zeamais</i>	Alcântara <i>et al.</i> (2021), Gontijo <i>et al.</i> (2017) (a), de Almeida <i>et al.</i> (2020) (b) and (2022) (c), Yamaguchi <i>et al.</i> (2010), Alcoba <i>et al.</i> (2017) and Mossi <i>et al.</i> (2013)
PU	2008	<i>Ocotea puberula</i> (Rich.) Nees	canela-babosa	x	Parque Estadual do Itacolomi, Ouro Preto, Minas Gerais,	20°17'15.0"S 43°30'19.1" W	In vitro activity against <i>Trypanosoma cruzi</i> , in vivo activities of wound healing	Barbosa <i>et al.</i> (2020) and (2021), Arcaro <i>et al.</i> (2023) and

								and antinociceptive activity in murine models	Montrucchio <i>et al.</i> (2012)
SP	1994	<i>Ocotea spectabilis (Meisn.) Mez</i>	canela-baraúna	x	Mariana, Minas Gerais, Brazil	20°22'40.0"S 43°24'57.9" W	No reported activity	x	
SX	1994	<i>Ocotea spixiana (Nees) Mez</i>	canelão	x	Parque Estadual do Itacolomi, Estrada para a Fazenda do Manso, Ouro Preto, Minas Gerais, Brazil	20°17'15.0"S 43°30'19.1" W	<i>In vitro</i> acaricidal activity against <i>Rhipicephalus</i> (<i>Boophilus</i>) <i>microplus</i>	Conceição <i>et al.</i> (2020)	
TA	2007	<i>Ocotea tabacifolia (Meisn.) Rohwer</i>	unknown	x	Serra de Antônio Pereira, Samarco. Alegria 7, Ouro Preto, Minas Gerais, Brazil	-	No reported activity	x	
TE	2001	<i>Ocotea tenuiflora (Nees) Mez</i>	unknown	x	Brazil, Minas Gerais, Descoberto, Reserva Biológica da Represa do grama	-	No reported activity		
TL	2001	<i>Ocotea teleiandra (Meisn.) Mez</i>	canela-limão	x	Reserva Biológica da Represa do Grama, Descoberto, Minas Gerais, Brazil	-	No reported activity	x	
TR	1997	<i>Ocotea tristis (Nees & Mart.) Mez</i>	canelinha	x	Antônio Pereira, Ouro Preto, Minas Gerais, Brazil	20°17'15.0"S 43°30'29.1" W	No reported activity	x	
VA	1977	<i>Ocotea vaccinioides (Meisn.) Mez</i>	unknown		<i>Ocotea daphnifolia</i> (<i>Meisn.</i>) <i>Mez</i>	Rancharia, Ouro Preto, Minas Gerais, Brazil	-	No reported activity	x
VI	2008	<i>Ocotea villosa Kosterm.</i>	unknown	x	Sítio Malícia, Mata do Krambeck, Juiz de Fora, Minas Gerais, Brazil	-	No reported activity	x	
VL	1994	<i>Ocotea velutina (Nees) Rohwer</i>	canelão-amarelo	x	Parque Estadual das Lauraceaes, Bocaiúva do sul, Paraná, Brazil	-	No reported activity	x	
VR	1967	<i>Ocotea variabilis Mart</i>	canela-pilosa		<i>Ocotea lancifolia</i> (<i>Schott</i>) <i>Mez</i>	Estrada de Campo Alegre para Araguaí, Campo Alegre, Minas Gerais, Brazil	-	<i>In vitro</i> acaricidal activity on <i>Rhipicephalus</i> (<i>Boophilus</i>) <i>microplus</i> , <i>in</i> <i>vitro</i> antifungal activity against <i>Fusarium</i> <i>moniliforme(a)</i> , <i>Trametes</i> <i>versicolor</i> and	Barbosa <i>et al.</i> (2013), da Silva <i>et al.</i> (2018) (a) and (2017) (b), and Fournet <i>et al.</i> (2007)

*Gloeophyllum
trabeum (b), in
vitro antioxidant
/ antiprotozoal
activities against
Leishmania spp.
and T. cruzi*

In vitro larvicidal
activity against
Aedes aegypti
larvae

Garcez *et al.*
(2009)

VZ I 2000 *Ocotea velloziana (Meisn.) Mez* canela-verde x

Camarinhas, Ouro Preto, Minas
Gerais, Brazil

-

Table S3 | Manual annotation chemical standards from Dataset 2 with level 2 of confidence on the final data processing list. Chemical standards were detected and identified by matching accurate mass, isotopic pattern, and fragmentation spectra to public databases literature sources, and chemical knowledge. Name, exact mass, observed m/z , RT, adducts, detected in high or low concentration in processed .mzML data, main MS/MS fragments, and supporting international identifier (InChIKey) are provided.

ID	Name	Exact mass	Observed m/z	RT	Adduct	High (H) or low (L) concentration	MS/MS fragments	InChIKey
1a	Vancomycin	1447.4302	724.72009	0.50	[M+2H] ²⁺	H	100.07672; 144.10208; 329.07394; 800.10370; 1087.28760; 1115.28979; 1143.27893	MYPYJXKWCTUITO- LYRMYLQWSA-N
2a	Nystatin	925.50350	926.51166	2.89	[M+H] ⁺	H / L	161.13705; 279.12503; 297.13632; 339.14346; 655.36560; 673.36810; 691.38489; 727.40198; 745.41968; 890.48761; 908.48932	VQOXZBDYSJBXMA- RKEBNKJGSA-N
3a	Azithromycin	748.50853	749.51324	2.10	[M+H] ⁺	H	72.0827; 98.09959; 116.10944; 158.12018; 186.14938; 257.08594; 398.29184; 434.31146; 573.41479; 591.41864	MQTOSJVFKKJCRP- BICOPXKESA-N
4a	Novobiocin	612.23191	613.24133	4.35	[M+H] ⁺	H / L	97.06465; 133.03056; 186.07791; 189.09122; 218.10526; 345.01031; 396.14288	YJQPYGGHQPGBLI- KGSXXDOSSA-N

5a	Puromycin	471.22302	472.23154	1.93	[M+H] ⁺	H / L	150.09512; 164.09294; 264.11874; 292.11609; 309.14447; 371. ;	RXWNC PJZOCPEPQ- NVWDDTSBSA-N
6a	Betulinic acid	456.36035	457.37149	5.55	[M+H] ⁺	H	95.08782; 163.14926; 191.17993; 231.20955; 249.05836; 297.08365; 393.35013; 411.36081; 439.35933	QGJZLNKBHJESQX- FZFNLFKSA-N
7a	Oleanolic Acid	456.36035	457.37149	5.67	[M+H] ⁺	H / L	95.08782; 163.14926; 191.17993; 221.08742; 231.20955; 269.98254; 297.08365; 393.35013; 411.36081; 439.35933	MIJYXULNPSFWEK- GTOFXWBISA-N
8a	Ursolic acid	456.36035	457.37149	5.68	[M+H] ⁺	H / L	95.08782; 163.14926; 191.17993; 221.08742; 231.20955; 297.08365; 393.35013; 411.36081; 439.35933	WCGUUGGRBIKTOS- GPOJBZKASA-N
9a	Tetracycline	444.15327	445.15884	1.95	[M+H] ⁺	H / L	86.02639; 98.06068; 126.05542; 154.05173; 241.00900; 269.07956; 337.07254; 365.06461; 410.12222; 427.1499; 428.13821	NWXMGUDVXFXRIG- WESIUV DSSA-N
10a	Doxycycline	444.15327	445.15884	2.29	[M+H] ⁺	H	84.04588 ; 98.09527; 126.05542; 154.05179; 201.05211; 267.06400; 321.07751; 339.08719; 410.12222; 428.13821	SGKRLCUYIXIAHR- AKNGSSGZSA-N
11 ^a	Florfenicol	357.00046	358.00943	2.52	[M+H] ⁺	H / L	132.06059; 170.06255; 206.04065; 208.03178; 210.05780; 241.00900; 243.00554; 319.99149; 339.00684; 341.00305	AYIRNRDRBQJXIF- NXEZZACHSA-N
12a	Thiamphenicol	355.0048	356.01337	2.02	[M+H] ⁺	H / L	146.06334; 198.06124; 228.10692; 230.00967; 240.00029; 242.01303; 307.12915 ; 309.14447 ; 311.14328	OTVAEFIXJLOWRX- NXEZZACHSA-N
13a	Mitomycin C	334.12772	335.13580	1.89	[M+H] ⁺	H / L	77.03879; 104.05007; 131.06418; 132.07062; 169.07913; 171.09450; 215.08574; 242.09453	NWIBSHFKIJFRCO- WUDYKRTC SA-N
14a	Chloramphenicol	322.01233	323.02032	2.63	[M+H] ⁺	H / L	132.06059; 165.06738; 206.03439; 241.00900; 243.00554; 257.99667; 259.99185; 274.99786; 276.9948; 305.0064; 307.00674	WIIZWVCIJKGZOK- RKDXNWHRSA-N
15a	Cycloheximide	281.16271	282.16724	2.70	[M+H] ⁺	H / L	107.08649; 159. 12080; 187.10951; 219.13799 229.12338; 246.15276; 264.16132	YPHMISFOHDHNIV- FSZOTQKASA-N

Table S4 Annotation with an enhanced level of confidence (level 2) of automated annotation with the level of confidence 3 using the ActinomarineDB. Putative metabolites were annotated by matching accurate mass, isotopic pattern, and fragmentation spectra to public databases literature sources and chemical knowledge. Putative name, exact mass, observed m/z , RT, adducts, main MS/MS fragments, *in-house*DB, and supporting international identifier (InChIKey) are provided.

ID	Putative metabolite name	Exact mass	Observed m/z	RT	Adduct	MS/MS fragments	Reference	Actinomarine DB hit	InChIKey
1b	Fluostatin A	306.0528	307.0603	3.16	[M+H] ⁺	236.04697; 251.06715; 263.06979; 279.06668; 281.0886	GNPS	Yes	ISHOMJGAOPXCEF- HFFFAOYSA-N
2b	Fluostatin K	308.0684	309.0754	3.22	[M+H] ⁺	263.06979; 281.07571; 291.06593	Proposed	Yes	CKTLJJYDUPSLCZ- SFHVURJKSA-N
3b	Kinobscurinone	306.0528	613.1116	4.32	[2M+H] ⁺	223.279.0666; 265.0489; 289.0502; 307.0603;	Proposed	Yes	CC1=CC(=C2C(=C1)C(=O)C3=C2C(=O)C4=C(C3=O)C(=CC=C4)O)O
4b	Rosamicin	581.3564	582.3661	2.27	[M+H] ⁺	98.0976; 113.05816; 116.10474; 158.12018	Proposed	Yes	IUPCWCLVECYZRV- ZMZINANSA-N
5b	γ -Actinorhodin	630.1009	631.1201	3.88	[M+H] ⁺	499.10809; 543.10413; 571.09656; 585.11890; 613.11163	GNPS	No	* CC1OC(CC(O7)=O)C7C(C(C3=C2C(O)=C(C4=CC(O)=C(C(C(C(O8)C(CC8=O)OC6C)=C6C5=O)=O)C5=C4O)C=C3O)=O)=C1C2=O
6b	Phomin	479.2671	480.2815	3.11	[M+H] ⁺	81.07118; 91.05359; 105.06780; 107.08649; 119.08701; 145.01472; 444.25668	GNPS	Yes	GBOGMAARMMDZGR- TYHYBEHESA-N
7b	Desferrioxamine B	560.3533	585.3195	1.36	[M-2H+Al] ⁺	70.06599; 241.11066; 300.10849; 303.19546; 368.17395; 385.20248; 467.2079	GNPS	No	UBQYURCVBFRUQT- UHFFFAOYSA-N

*SMILES are provided when InChIKey is not available.

Supplementary Text (ST-1)

Alkaloids

Forty-one alkaloids were level 2 annotated in the LC-HRMS QC ESI⁺ metabolic fingerprint. In the range of retention time at 0.5-1.5 min, the pyrrolidine alkaloid 4-hydroxy-*N*-methylproline (**1**, R_t 0.49 min) was annotated in high intensity with the parent ion at m/z 146.081. Subsequently, from 1.5 to 2.5 min, two proaporphines, a variety of noraporphines and aporphines and two benzyloquinoline alkaloids were annotated. More specifically, from 1.4 to 1.54 min, the proaporphines were annotated as crotsparine (**2**, R_t 1.42 min) and glaziovine (**3**, R_t 1.49 min), at m/z 284.127 and m/z 298.144, respectively. The *N*-methylcoclaurine (**4**, R_t 1.54 min), a benzyloquinoline, was annotated with parent ion at m/z 300.160, while the ion at m/z 298.144 was annotated as the aporphine 3-hydroxynornuciferine (**5**, R_t 1.60 min) (**Fig. 4; Table 1; Table S1**). Additionally, other two other noraporphine were annotated with parent ion at m/z 314.139, as the laurelliptine (**6**, R_t 1.66 min), and the laurilitine (norboldine) (**7**, R_t 1.74 min). Furthermore, five main alkaloid isomers eluted at R_t 1.61 – 1.96 min displayed the same observed parent ion at m/z = 328.155. Throughout MS^E fragment analyses, compounds were putatively annotated as pallidine (**8**, R_t 1.61 min), flavinantine (**9**, R_t 1.75 min), which belongs to the morphinandienone alkaloid class that derived from the benzyltetrahydroisoquinoline core. In addition, the aporphine isomers boldine (**10**, R_t 1.85 min), isoboldine (**11**, R_t 1.96 min), and corytuberine (**12**, R_t 2.06 min) could be differentiated by their respective observed MS^E fragment ions. Moreover, the aporphine lauroschoztine (*N*-methylaurotetanine) was also annotated at m/z 342.170 (**13**, R_t 2.02 min) (**Fig. 4-5; Table 1; Table S1**).

Regarding the gas-phase fragmentation reactions involved in high energy channels on MS^E/DIA mode to these isoquinoline alkaloids, several exhibited similar product ions. For example, for the noraporphines **6** and **7** and aporphines **10**, **11** and **12** the main difference between them lies in their respective spectra abundances and the first neutral losses they undergo, which are NH₃ and NH₂CH₃, respectively. Consequently, these compounds share a common diagnostic ion at m/z 297.112. Successively, the presence of the fragment ion at m/z 265.085 (C₁₇H₁₃O₃⁺) was attributed to the loss of CH₃OH (32.03 Da). The next observed fragment ion at m/z 237.092 (C₁₆H₁₃O₂⁺) is a result of a CO (27.99 Da) neutral loss from the m/z 265.085.

Moreover, the fragmentation pathway of proaporphine alkaloids, such as **2** and **3** is similar to the aporphines, although with two successive CO neutral losses. The fragment ion at m/z 267.102 corresponds to heterocycle ring opening for both alkaloids. However, for alkaloid

2 this fragment ion indicates an NH₃ loss, while for alkaloid **3** it indicates a CH₃NH₂ loss. The subsequent CH₃OH elimination yielded the ion at *m/z* 235.075. Consecutive loss of two CO units led to the ion at *m/z* 207.080 with contraction of the isoquinoline A ring, subsequently, the ion at *m/z* 179.086 was formed due to the contraction of ring D. A CH₂ (14.01 Da) elimination led to the diagnostic ion at *m/z* 165.068, which is the most intense fragment for these alkaloids (**Fig. 4; Table 1; Table S1**). Regarding the fragmentation of the morphinandienone alkaloids, the isomers **8** and **9** could not be accurately differentiated. They show the same fragmentation pattern of aporphines. The fragment ion at *m/z* 297.112 was observed due to the CH₃NH₂ neutral loss. Further, the ion at *m/z* 265.086 was formed due to a CH₃OH loss followed by CO, evidenced by the presence of the ion at *m/z* 237.091 (**Fig. 6; Table 1; Table S1**).

Following the annotations, in the R_t range of 2.0-2.33 min, with parent ion at *m/z* 330.170 and parent ion at *m/z* 314.175, the benzyloisoquinolines reticuline (**14**, R_t 2.07 min) and armapevine (**15**, R_t 2.19 min) were respectively annotated. Together with **4** (*m/z* 300.160, R_t 1.54 min), these benzyloisoquinoline alkaloids show a similar fragmentation. The common tertiary amine loss [M + H - CH₃NH₂]⁺ of 31 Da in the high energy spectrum, showed low-intensity fragment ions at *m/z* 269.117, 299.128 and 283.133 to **4**, **14** and **15**, respectively. The observed fragment ions at *m/z* 192.102, *m/z* 175.074, *m/z* 143.049 and *m/z* 137.057 from MS^E spectra are diagnostic ions for **14** (**Fig. 5; Table S1**). The ions at *m/z* 175 and 143 are products of subsequent fragmentations of *m/z* 299, first with a loss of the substituted phenyl group (124.06 Da) and then with a common CH₃OH neutral loss of 32.03 Da with epoxide formation. However, the most intense ions were at *m/z* 192 and *m/z* 137, which are a direct consequence of the fragmentation of the C ring (benzyl), forming either the charged isoquinoline or the broken C charged rings. The *m/z* 192 is also diagnostic for the benzyloisoquinoline alkaloid **4**, in which the rings A and B are identical to the **14** (**Fig. 5**). To alkaloid **15**, the respective fragment ion was observed at *m/z* 206.117, which indicates one more methyl group in the isoquinoline ring A. In addition, the ion observed at *m/z* 107.049 is common to **4** and **15** and is related to the charged benzyl fragment and equivalent to the mentioned *m/z* 137, with the difference of 30 Da due to an additional methoxy group attached to benzyl moiety of **14**. (**Fig. 5; Table 1; Table S1**).

In the next, two isomers of noraporphine alkaloids were annotated with parent ion at *m/z* 298.144, as the zenkerine (**16**, R_t 2.11 min) and tuduranine (**17**, R_t 2.16 min), exhibiting the same fragmentation pattern observed for the previously discussed aporphines. In the MS^E high energy spectra were observed for the product ions at *m/z* 281.150, which is related to the

isoquinoline opening ring and elimination of NH_3 . Also, the fragment ions at m/z 235.076 were attributed to the loss of $\text{CH}_3\text{CH}_2\text{OH}$ (46.04 Da), and at m/z 207.078, due to the common neutral loss of CO (**Fig. 4; Table 1; Table S1**). In addition, at m/z 326.139, the aporphines diospirifoline (**18**, R_t 2.13 min), and other two aporphine isomers were annotated with parent ion at m/z 342.167 as taliporphine (**19**, R_t 2.24 min) and predicentrine (**20**, R_t 2.35 min) (**Fig. 4-5; Table 1; Table S1**). Moreover, in the range of 2.4-2.5 min, several aporphine alkaloids were also annotated, as the nuciferine at m/z 296.165 (**21**, R_t 2.40 min), corydine (**22**, R_t 2.41 min) at m/z 342.171, domesticine (**23**, R_t 2.47 min) at m/z 326.139, and dehydrodicentrine (**24**, R_t 2.61 min) at m/z 340.154. In addition, two noraporphine isomers were also annotated at m/z 328.155 as norisocorydine (**25**, R_t 2.48 min) and the laurotetanine (**26**, R_t 2.52 min). Other three noraporphine were observed, which two are annotated isomers at m/z 326.139, as nordicentrine (**27**, R_t 2.81 min), and nornantenine (**28**, R_t 2.99 min), in addition to the ion at m/z 282.149 annotated as nornuciferine (**29**, R_t 3.12 min) (**Fig. 4-5; Table 1; Table S1**). Lastly, more six aporphines were observed in the range of 2.5-3.10 min, lirinidine (**30**, R_t 2.70 min) at m/z 282.149, R_t glaucine (**31**, R_t 2.90 min) at m/z 356.185, and R_t roemerine (**32**, R_t 3.07 min) at m/z 280.133. Also, two aporphine isomers with observed parent ion at m/z 340.154 were annotated as nantenine (**33**, R_t 2.89 min) and dicentrine (**34**, R_t 3.09 min), as well and dehydronuciferine annotated at m/z 294.149 (**35**, R_t 3.15 min) (**Fig. 4-5; Table 1; Table S1**).

The isomers **29** and **30** could be differentiated by the NH_3 and CH_3NH_2 neutral losses as well. The former generates the fragment ion at m/z 265.122 resulting from the characteristic elimination of NH_3 . Then, the product ion at m/z 234.104 is formed due to the OCH_3 loss. The fragment ion at m/z 250.099 ($\text{C}_{17}\text{H}_{14}\text{O}_2^+$) was observed because of the parallel loss of CH_3 . While, for the **30**, the neutral loss of CH_3NH_2 has resulted in the ion at m/z 251.107, in addition to the consequent neutral loss of CH_3OH and CO, which has yielded fragment ions at m/z 219.081 and m/z 191.086, respectively (**Fig. 4; Table 1; Table S1**). The same fragmentation pattern occurs for aporphines **21** and **35**. The fragmentation pathway for the aporphine **32** yielded a fragment ion at m/z 249.091 due to the characteristic elimination of CH_3NH_2 and CH_3OH , which formed the fragment ion at m/z 219.081. The known consequent neutral loss of CO generated a fragment ion at m/z 191.086 (**Fig. 5; Table 1; Table S1**). The same fragmentation pattern was also observed for the other six different aporphines: **13**, **20**, **22**, **24**, **33**, and **34** (**Fig. 5; Table 1; Table S1**).

Moreover, in the range of 3.3-3.9 min, fewer alkaloids were detected. Only one major oxo-noraporphine known as dicentrinone (**36**, R_t 3.33 min) was annotated at m/z 336.086. The last high-intensity aporphine found in the positive metabolic fingerprint included leucoxylophine

(**37**, R_t 3.52 min), annotated with precursor ion at m/z 400.176. Other high-intensity observed metabolites were phenanthrene alkaloids annotated as stephenanthrine (**38**, R_t 3.06 min) at m/z 294.149, argentinine (**39**, R_t 3.20 min) at m/z 296.165, and thalictuberine (**40**, R_t 3.67 min) at m/z 354.169 (**Fig. 4-5; Table 1**). At last, the only member of the tetrahydroprotoberberine alkaloid class was annotated as discretamine (**41**, R_t 3.88 min) at m/z 328.160 (**Fig. SX; Table 1**). The fragmentation pattern of the phenanthrene alkaloids was perceived for **38**, **39** and **40**. The fragment ion at m/z 251.107 was observed for **39**, while at m/z 249.091 it was attributed to **38**. A subsequent CH_3OH neutral loss followed by the CO led to the fragment ions at m/z 219.081 and m/z 191.086, respectively (**Fig. 5-6; Table 1; Table S1**).

Lignoids

To the lignoids class, a relatively smaller subset of compounds was successfully annotated with level 2 of confidence, compared to the alkaloid class. Nonetheless, our investigation revealed the presence of numerous lignoids in the *Ocotea* spp. samples. These lignoids were level 3 annotated according to the MSI, mainly in the LC-HRMS QC ESI⁺ metabolic fingerprint by multiple annotation hits from the *Ocotea*DB (<https://doi.org/10.5281/zenodo.8303382>). More specifically, 89 hits from *Ocotea*DB were observed, and thus, the presence of these compounds substantiates the notion that *Ocotea* species are indeed natural sources of lignoids as well. The limited availability of publicly accessible spectra, combined with the high diversity of possible isomers and complex scaffolds, posed challenges in annotating a larger number of lignoids with level 2 of confidence, where we have annotated specifically one lignan and five neolignans.

We were able to assign the parent ion at m/z 355.118 as the sesamin lignan (**42**, R_t 3.21 min), exhibiting characteristic fragments. Additionally, we successfully annotated five neolignoids at high-intensity levels, including the bicycloneolignan ocophylol B (**43**, R_t 5.38 min) of m/z 359.185 with its respective characteristics ions as well (see Supplementary **Table S1**). And the other four neolignans were annotated as eusiderin (**44**, R_t 5.46 min) with an ion at m/z 387, licarin B (**45**, R_t 6.01 min) with parent ion at m/z 325.143, licarin A (**46**, R_t 6.31 min) with an ion at m/z 327.159, and armenin B (**47**, R_t 6.35 min) with an ion at m/z 373.165. The fragmentation pattern of the last three, which are benzofuran is demonstrated in **Fig. 6** and observed fragments in **Table S1**.

Regarding the **46** a loss of water (18.01 Da) led to a fragment ion at m/z 309.147. This fragmentation event is suggestive of the presence of hydroxyl groups. The fragment at m/z 295.135 arises from a five-membered ring opening and the concurrent loss of a methoxy group

(32.03 Da). This fragmentation pathway suggests the formation of an epoxide and structural rearrangements. Whereas, the fragment ion at m/z 203.069 can also be formed after the opening of a five-membered ring, followed by a hydrogen rearrangement that leads to the breakage of phenylpropanoid units. This rearrangement results in a neutral loss of 124.05 Da, which is the diagnostic ion for different benzofurans, including **45** and **47**. Moreover, the **44** has a little different fragmentation pattern because there are no hydroxyl groups as substituents. Instead, there are only methoxy groups, and the five-membered ring is a heterocycle six-membered ring containing 2 oxygens, classified as an oxycneolignan. The characteristic ions are demonstrated together with the fragmentation of **42** and **43** in (Supplementary **Table S2**).

While these findings provide valuable insights into the composition of the investigated lignoids, it is important to acknowledge the limitations imposed by the unavailability of online spectra as well as the absence of deep investigations of fragmentation pathways for most of those compounds previously described in the *Ocotea* genus. Further efforts are warranted to obtain and incorporate additional spectrometric data, facilitating more comprehensive characterization and accurate annotation of lignoid compounds in future studies.

Flavonoids

The majority of the high-intensity ionized metabolites in the LC-HRMS QC ESI⁺ metabolic fingerprint were assigned as flavonoids, majorly the glycosylated flavonoids of kaempferol, quercetin and apigenin. Besides the flavonoids, a small phenolic metabolite was annotated as the cyclic polyol quinic acid (**48**, R_t 0.56 min) in high-intensity levels at m/z 191.054. The non-glycosylated flavonoid backbones of taxifolin (**49**, R_t 1.96 min) at m/z 303.050, catechin/epicatechin (**50**, R_t 2.07 min) at m/z 289.071, quercetin (**63**, R_t 3.57 min) at m/z 301.033, apigenin (**64**, R_t 2.89 min) at m/z 269.044, and kaempferol (**65**, R_t 4.10 min) at m/z 285.039 were annotated (**Table 1**).

High-intensity annotated glycosylated flavonoids include the class apigenin-based flavonoid patterns. Among them, vitexin (**55**, R_t 2.54 min) was annotated at m/z 431.097. Compound **55** corresponds to apigenin-8-*C*-glucoside, thus a β -*D*-glucosyl residue attached at position C-8 of the apigenin structure. Another notable glycosylated flavonoid is vitexin-2'-*O*-rhamnoside (**52**, R_t 2.42 min) annotated at m/z 577.157, a derivative of vitexin with an additional α -*L*-rhamnosyl residue attached at position C-2' of the flavonoid. Additionally, apigenin-7-*O*-rutinoside (**59**, R_t 2.74 min) at m/z 577.157 was identified as a rutinoside derivative of apigenin with a rutinose moiety attached at position C-7 of the apigenin structure. Lastly, the annotated isomers apigenin 6-*C*-glucoside-8-*C*-arabinoside (schaftoside) or apigenin

8-*C*-glucoside-6-*C*-arabinoside (isoschaftoside) (**60**, R_t 2.85 min) at m/z 563.142 was observed. The former exhibits a β -*D*-glucosyl residue attached at position C-6 and α -*L*-arabinosyl residue attached at position C-8 of the apigenin structure, while **60** is the opposite. Several apigenin-based flavonoid characteristic fragments could be observed. For compound **55**, the fragments correspond to the aglycone carbonyl-apigenin backbone at m/z 311.052 and a neutral loss of 120.04 Da related to the C-C bond cleavage product, a modified β -*D*-glucosyl residue (C₄H₈O₄) at position C-8. In the case of flavonoid **52**, the fragments also correspond to the carbonyl-apigenin backbone, but with the neutral loss of 266.10 Da related to the additional α -*L*-rhamnosyl residue attached to the oxygen at position 2' of **55**. To compound **59** the fragments also correspond to the classic apigenin backbone at m/z 269.047 with a loss of a deoxygenated rutinose moiety (C₁₂H₂₂O₉, 308.11 Da) at position C-7. Finally, the fragment ions observed for compound **60** also repeat the apigenin backbone structure, with losses corresponding to the modified β -*D*-glucosyl residue (120.04 Da) at position C-6, giving a fragment ion at m/z 443.09, and the α -*L*-arabinosyl residue (90.03 Da) at position C-8, giving a fragment ion at m/z 473.106 (**Fig. 6, Table 1; Table S1**).

A tetrahydroxyflavone compound exhibiting a pattern similar to quercetin, with a parent ion at m/z 609.147 (C₂₁H₂₀O₁₂), was detected and annotated as rutin (**53**, R_t 2.46 min). Other flavonoids with parent ions at m/z 463.087 (C₂₁H₂₀O₁₂) and m/z 447.093 (C₂₁H₂₀O₁₁) were annotated as the respective isoquercitrin (**51**, R_t 2.40 min) and quercitrin (**56**, R_t 2.66 min). These compounds are also tetrahydroxyflavone *O*-glycosides, where the quercetin is substituted by α -*L*-glucosyl and rhamnosyl moieties, respectively, at position C-3 via glycosidic linkage. Another tetrahydroxyflavone, with a parent fragment ion at m/z 463.088 (C₂₁H₂₀O₁₂), was identified as quercimeritrin (**54**, R_t 2.54 min). Flavonoid **51** corresponds to a quercetin *O*-glucoside, characterized by the presence of a β -*D*-glucosyl residue attached to position C-7 of the quercetin structure. Furthermore, the parent ion at m/z 433.077 (C₂₀H₁₈O₁₁), was annotated as reynoutrin (**57**, R_t 2.73 min), corresponding to quercetin-3-*O*-xylopyranoside. In the MS^E high-energy spectra corresponding to these flavonoids, several characteristic fragments of the quercetin backbone could be observed. For **53**, the fragments correspond to the aglycone quercetin backbone at m/z 301.035, with potential losses of the rutinose sugar (308.11 Da) moiety at position 3. The **54** also exhibits fragments corresponding to the quercetin backbone, with a loss of the β -*D*-glucosyl residue (162.053 Da) at position C-7. Flavonoid **57** shows fragments corresponding to the quercetin backbone as well, with a loss of the xylopyranosyl residue (132.042 Da) at position C-3. Finally, compounds, **56** and **51** also exhibit fragments

corresponding to the quercetin backbone, with a loss of the α -L-rhamnosyl residue (146.06 Da) at position C-3 and a glucosyl residue (162.05 Da), respectively (**Fig. 6, Table 1; Table S1**).

The kaempferol-based glycosylated flavonoids were also observed in high-intensity levels, including the ion at m/z 447.093 that was annotated as astragalin (**58**, R_t 2.74 min). Another parent ion at m/z 431.098 ($C_{21}H_{20}O_{10}$) was annotated as afzelin (**61**, R_t 3.15 min). Compounds **58** and **61** are glycosyl flavones in which the kaempferol aglycone is attached to α -L-glycosyl and rhamnosyl residues, respectively, via a 3-*O*-glycosidic bond. This is evidenced by the presence of the deprotonated aglycone ion at m/z 285.039 in the high-energy channel spectra, reflecting the loss of a rhamnose sugar moiety (146.06 Da) for **58** and a glucose moiety (162.05 Da) for **61**. Additionally, kaempferol 3-4"-*p*-coumarylrhamnoside (**62**, R_t 3.52 min) and isomers kaempferol 3-(2",4"-di-(*E*)-*p*-coumarylrhamnoside) and kaempferol 3-(3",4"-di-(*E*)-*p*-coumarylrhamnoside) (**66**, R_t 5.56 min) were also annotated. The former is also a glycosyl flavone with a parent ion at m/z 577.136 ($C_{30}H_{26}O_{12}$). The latter with observed parent ion at m/z 723.174 ($C_{39}H_{32}O_{14}$) (**Table 1**). **62** exhibited a single loss of the coumarylrhamnoside moiety (292.10 Da) in high-energy channel spectra, while the loss of a di-coumarylrhamnoside unit (438.132 Da) in the same channels indicated the presence of the isomers **66** (**Fig. 6, Table 1; Table S1**). The MS^E spectrum further confirmed the aglycone's identity with characteristic product ions at m/z 255.029 and 227.034, matching the literature and corroborating it as a kaempferol derivative.

Moreover, fragmentations not included in the manuscript, such as the fragmentation of **1** was proposed based on a 4-hydroxyproline available spectrum in the MoNA database. The difference is 14 Da more is related to the presence of a tertiary amine (*N*-methylated proline). In the MS^E high energy scan was observed the main fragment ions at m/z 100.0756, which was related to the carboxylic acid neutral loss (46.01 Da). In addition to the ions at m/z 82.0651 that are explained by a water loss (18.01 Da), and at m/z 72.0808 related to the opened pyrrolidine ring. Besides, the **1** was a hit from the *Ocotea*DB, and it was also putatively identified with high peak areas among the QCs and *Ocotea* sp. extract samples (**Table 1 and S1**).

Discussion regarding Supplementary Fig S3-S7.

The high complexity of our dataset, arising from the diverse matrix of plant extracts and large sample size ($n=60$), is evident in Supplementary **Fig. S3-S7**. Supplementary **Fig. S3** displays the overlaid chromatograms of features at m/z 328.145 within the R_t 1.45-2.40 min window from the final aligned feature list generated through MZmine 3 processing. These features represent 7 isomers eluting nearby that could be adequately resolved using the local

minimum peak resolver algorithm, although some peaks remained outside the aligned area, suggesting potential for even further refinement. Effective resolution of complex chromatographic regions enables reliable differentiation and semi-quantification of compounds for comparative metabolomics. Furthermore, Supplementary **Figs. S4** and **S5**, showing the final MS-DIAL alignment plots of m/z versus R_t , and Supplementary **Figs. S6** and **S7**, depicting the full MN of both ionization modes, demonstrate the substantial data complexity. These figures illustrate that intricate matrices in large datasets demand sophisticated analytics for robust processing, visualization, and interpretation.

Supplementary Protocol

Step-by-step protocol for the DIA-IntOpenStream pipeline

Software download

- Begin by downloading all necessary software tools for the pipeline. The list of required software includes Waters2mzML, Msconvert, MZmine 3, MS-DIAL, KNIME, WinSPC, and Cytoscape;
- Ensure that each software tool is compatible with your computer's operating system. Check the minimum system requirements on the respective websites of these tools to confirm that your machine has the necessary computational capacity to run them efficiently;
- Follow the detailed instructions provided on each software's website for proper download and installation. These guidelines are found on their respective websites;
- Waters2mzML: <https://github.com/AnP311/Waters2mzML>;
- MSconvert: <https://proteowizard.sourceforge.io/download.html>;
- MZmine 3: <http://mzmine.github.io/download.html>;
- MS-DIAL: <http://prime.psc.riken.jp/compms/msdial/main.html>;
- KNIME: <https://www.knime.com/downloads>;
- WinSPC: <https://winscp.net/eng/download.php>;
- Cytoscape: <https://cytoscape.org/download.html>;
- By carefully downloading and installing these software tools, you will establish a solid foundation for the successful execution of the pipeline;

Step 1) Sample preparation and data acquisition

- Begin by preparing your samples according to standard protocols. Collect LC-HRMS/DIA raw data (AIF or MS^E), ensuring consistency in data acquisition parameters. It could be either profile or centroided data;

Step 2) Data conversion to .mzML

- For MS^E Waters.RAW data, utilize Waters2mzML for conversion. Place your Waters.RAW files in the “raw_files folder” and run “Waters2mzML-1.2.0.exe”. Decide on centroiding and wait for the conversion completion. The processed .mzML files will be located in the “mzML_files folder”. For more details access the GitHub web page;
- In the case of other types of DIA data (e.g, AIF from Thermo Fischer Orbitrap), conventional conversion on Proteowizard’s msconvert is suitable as well described in MZmine and GNPS documentation, online available at https://mzmine.github.io/mzmine_documentation/data_conversion.html and <https://ccms-ucsd.github.io/GNPSDocumentation/fileconversion/>, respectively;

Step 3) Custom in-house database preparation

- Begin by selecting a suitable database containing metabolites pertinent to your study. For plant and microbial natural products, databases such as KNApSAcK and the Natural Product Atlas are open-recommended databases. For human metabolites, the Human Metabolome Database (HMDB) is a suitable choice.-;
- Download information related to the metabolites' structures (formats like .mol, .mol2, and .sdf) or simplified text identifiers (such as SMILES, CAS number, InChIKey, and IUPAC name).-;
- Download the KNIME workflow (<https://hub.knime.com/-/spaces/~8bZEbbknV8tVptea/current-state/>) for the creation of your *in-house*DB, which is compatible with four common types of chemical input data: .mol, .mol2, .sdf, and .csv (universal table format). As an alternative option, the information can be also added as table input files, which can include SMILES, InChIKey, CAS number, or IUPAC names;
- The outcome of this workflow is a .csv file containing three columns: chemical structure name, calculated molecular formula, and calculated monoisotopic mass. This file will

be instrumental for subsequent steps in the pipeline, particularly for enhancing the confidence level of the annotation to level 3;

- This step is crucial for ensuring that your workflow is tailored to the specific metabolites relevant to your research, thereby enhancing the accuracy and relevance of your analysis;

Step 4) MZmine 3 data processing

- Import the .mzML converted data into MZmine 3 either by dragging and dropping the files or using the “Import Data Module” under “Raw Data Methods”;
- Follow the basic processing sequence, utilizing separate modules for each step;
- Mass Detection: Identifying ions from the mass spectrometry data;
- ADAP Chromatogram Builder: Constructing chromatograms for detected ions;
- Chromatogram Deconvolution: Resolving overlapping signals in chromatograms;
- C¹³ Isotope Filter: Filtering out carbon-13 isotopes to reduce data complexity;
- Alignment: Using either “Join Aligner” or “RANSAC” depending on dataset size and complexity. The former is the standard, although if your data is highly complex more robust algorithms are sought such as the latter;
- Gap Filling: Filling in missing data points in the aligned feature list;
- Optional steps, such as “Duplicate Feature Filter” and “Feature List Blank Subtraction”, can be applied;
- The final step involves annotation using the *in-house* custom database in the .csv format in the “Annotation” Module. Proper mapping of .csv column names to MZmine features is crucial, as shown in **Figure S19**;

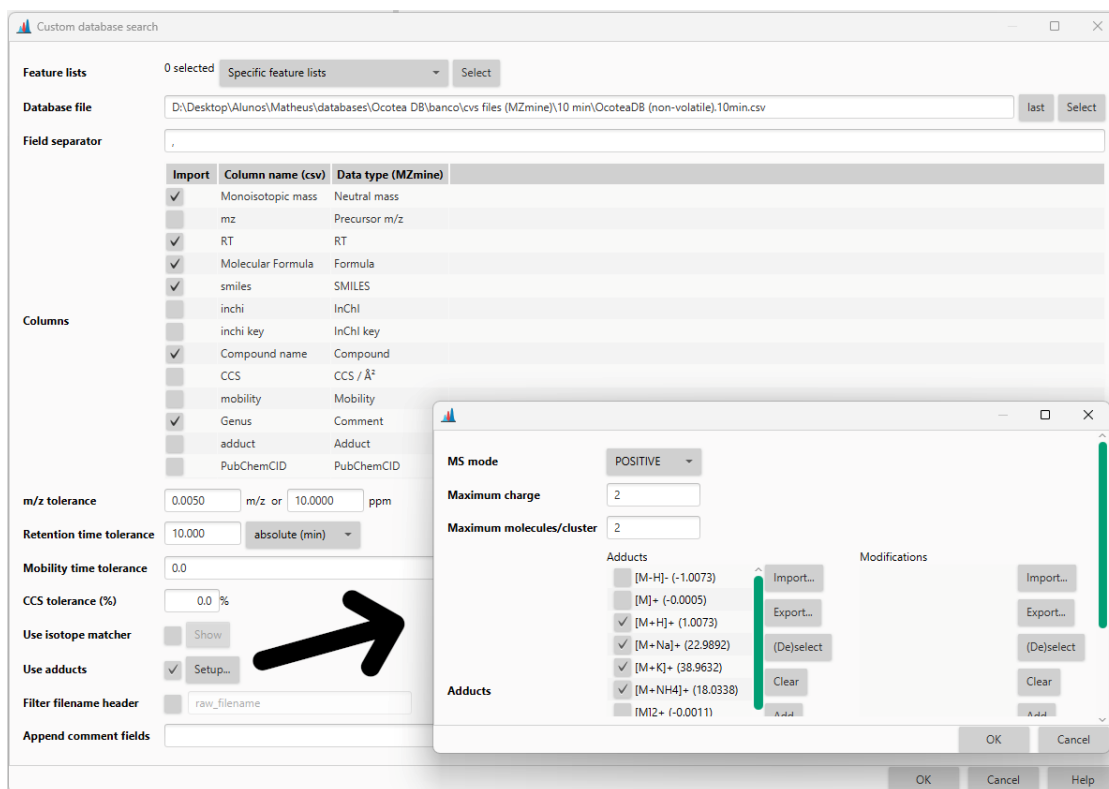


Fig. S19 | Custom database search settings of the “Annotation” module from MZmine 3.

- Set the “Use Adducts” option based on your sample and experimental requirements. For instance, in **Fig. S19**, different adducts were selected including $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[M+NH_4]^+$;
- For detailed explanations of each step and parameter, visit the MZmine Documentation, including tutorial videos https://mzmine.github.io/mzmine_documentation/index.html;

Step 5) Step MS-DIAL data processing

- Start a new project in MS-DIAL and ensure the directory path is set to where the .mzML files are stored;
- Configure your settings based on your experimental method of LC-MS/DIA, including soft ionization, chromatographic separation, all ions MS method, centroid data, the ionization mode, and metabolomics as target omics);
- Upload the experiment file as described in the MS-DIAL documentation. As an example, we have provided our experiment file at the bottom of this supplementary material;

- Define the sample types (blank, sample, QC) and classes as required;
- Processing in MS-DIAL is an integrated step encompassing peak picking, deconvolution, compound identification, and peak alignment;
- Upload a .txt version of your *in-house* database for custom database annotation during the identification step;
- Export the aligned results to your computer using the “GNPS export” option in MGF or MSP format. In the same page window of alignment parameter settings, the option to filter based on ion abundances from blank samples can also be set (**Figure S20**);

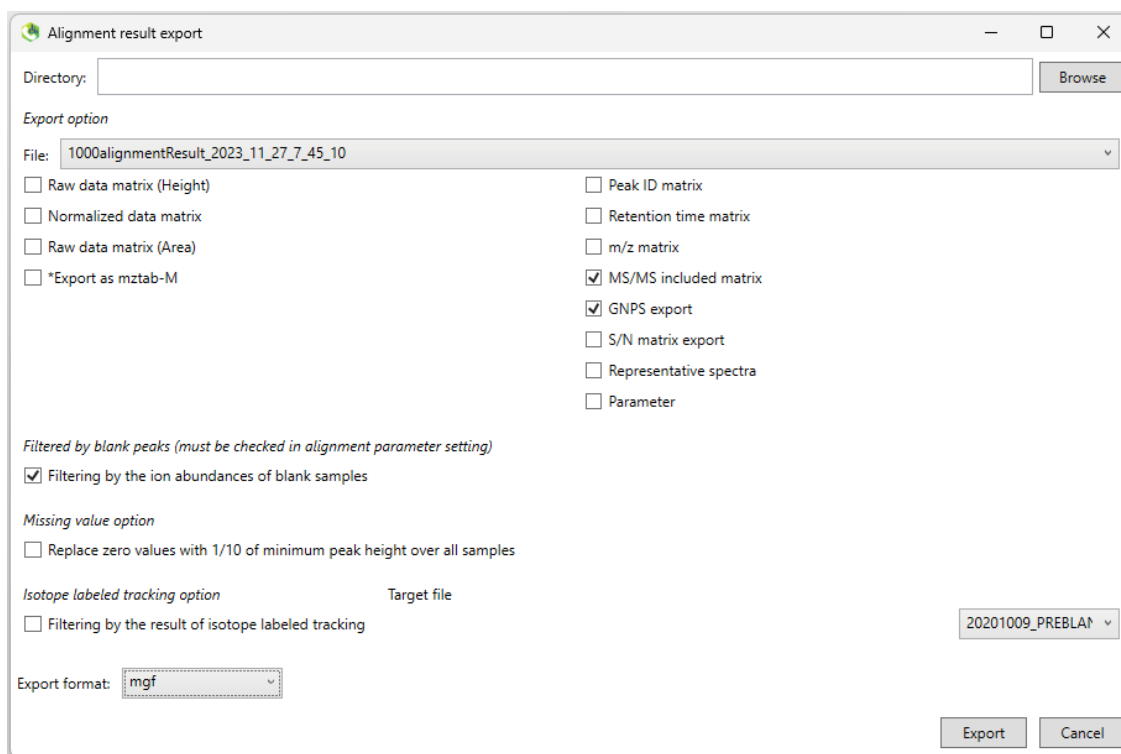


Fig. S20 | Alignment result export settings from MS-DIAL.

- Detailed information about processing steps and parameters is described online at <https://mtbinfo-team.github.io/mtbinfo.github.io/MS-DIAL/tutorial>;

Step 6) Data upload to GNPS server (WinSPC)

- For uploading files to GNPS, use an FTP client such as WinSPC. If you don't have a GNPS account, create one at <https://gnps.ucsd.edu/ProteoSAFe/user/register.jsp>;

- Connect to GNPS via WinSCP using FTP with no encryption, entering massive.ucsd.edu as the server (port 21), and using your GNPS credentials, as represented in **Figure S21**.

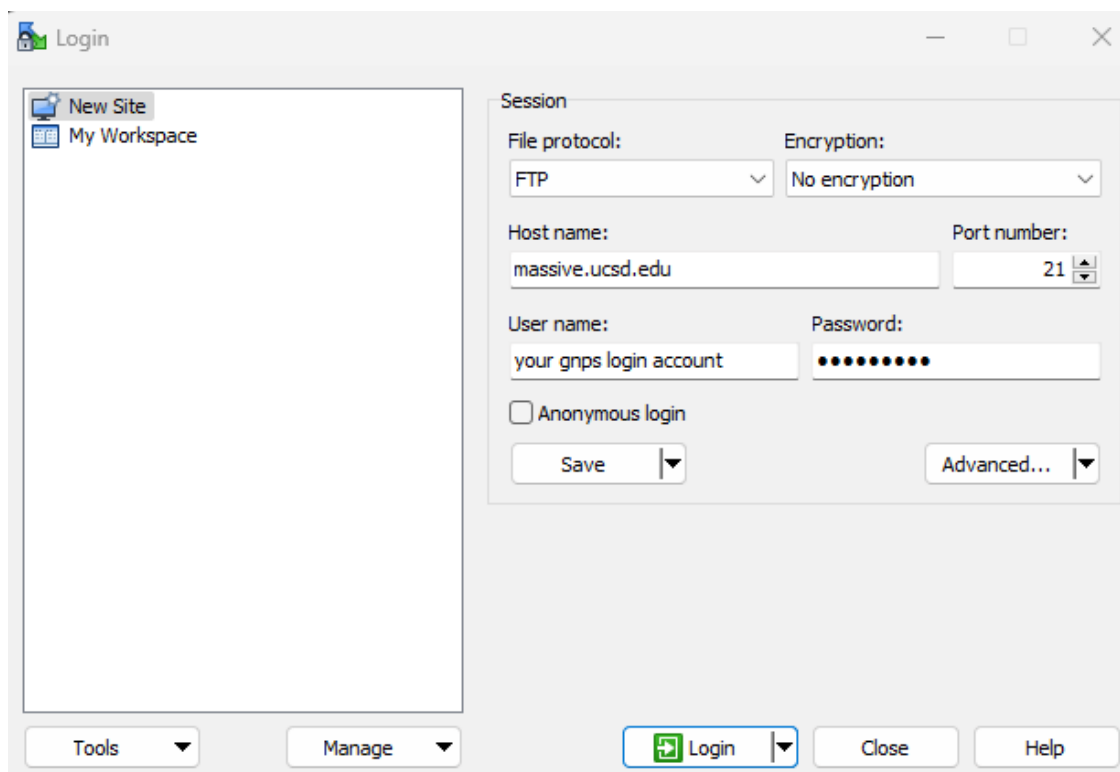


Fig. S21 | WinSCP login settings to connect to GNPS server.

- From there, you can upload the files related to the MS-DIAL alignment result, located in the respective directory on your computer (in the left panel) and then drag them over to the GNPS server (in the right panel);
- More information about this process can be found online at <https://ccms-ucsd.github.io/GNPSDocumentation/fileupload/> and <https://winscp.net/eng/docs/guides>;

Step 7) Feature-based Molecular Networking (FBMN) on GNPS

- On the GNPS platform homepage (<https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp>), after login in, select the option “Feature Networking” in the “Advanced Analysis tools” section;

- Essential for FBMN are two files: the feature table (.txt or .csv format) and the MS/MS spectral file (.mgf format), typically named GNPSStable.txt and GNPSmgf.mgf in MS-DIAL exports, respectively;
- Optionally, upload a metadata table via WinSPC for enhanced data analysis. This tab-separated text file can enhance dataset flexibility during data analysis and visualization. It is a text file (Tab-separated) that users must create these file themselves using a text editor (e.g. Microsoft Excel, Notepad++ for Windows, Gedit for Linux, TextWrangler for Mac OS). Using a Metadata table can greatly ease the visualization and analysis of data within Cytoscape before analysis;
- MS/MS libraries can also be uploaded to the GNPS server or selected among the GNPS libraries, to perform an automated annotation;
- In the FBMN interface (**Figure S22**), set parameters such as mass tolerance, cosine similarity, topK, and the number of fragments, as detailed in the main text. In addition, on the FBMN interface, by placing the cursor above some parameter an explanation about it appears;
- Comprehensive guidance on FBMN is online available at <https://lfnothias.github.io/GNPSDocumentation/featurebasedmolecularnetworking/>;

File Selection

MS2 File MGF/MSP(Progenesis QI)/mzML(MzTab-M): 1 file and 0 folders are selected

Feature Quantification Table: 1 file and 0 folders are selected

Original mzML Files:

Google Sheets Metadata URL (Experimental):

Sample Metadata Table: 1 file and 0 folders are selected

Basic Options

Quantification Table Source:

Precursor Ion Mass Tolerance: Da

Fragment Ion Mass Tolerance: Da

Advanced Network Options

Min Pairs Cos: Minimum Matched Fragment Ions: Maximum shift between precursors: Da

Network TopK: Maximum Connected Component Size (Beta):

Advanced Library Search Options

Spectral Library: 1 file and 0 folders are selected

[To import libraries for search click here](#)

Library Search Min Matched Peaks: Score Threshold:

Search Analogs: Maximum Analog Search Mass Difference: Da

Top results to report per query:

Advanced Filtering Options

Advanced Quantification Options

Advanced Multivariate Statistics Options

Advanced Univariate Statistics Options

Advanced External Tools

Advanced Extras

Fig. S22 | FBMN interface with File selection and other parameter settings.

Step 8) Data integration, interpretation and visualization

Part 1: MS/MS database search

- Inspect prominent annotations from the custom *in-house* database (annotation with confidence level 3). Once the biosynthetic-related hits were matched, to enhance the confidence level of these annotations we manually inspect the FBMN and the MS² spectra of these compounds;
- Pseudo MS² spectra from MS-DIAL (automatic) or raw MS² spectra on MZmine 3 (manual) can be compared with MS² spectra available in public databases to search for key fragments. We have applied the public repositories MassBank of North America (<https://mona.fiehnlab.ucdavis.edu/>) and Global Natural Products Social Molecular Networking (<https://gnps.ucsd.edu/ProteoSAFe/libraries.jsp>);

- For compounds without available spectra, propose fragmentation pathways using gas phase fragmentation reactions to bolster confidence in the annotations;

Part 2: Cytoscape visualization

- Upon completing FBMN, export the network for visualization in Cytoscape using the “Export/Download Network Files (Download Cytoscape Data)” or “Advanced Views - External Visualization (Direct Cytoscape Preview/Download)”;
- Information about how to customize your network on Cytoscape and improve the visualization can be found online at <https://cytoscape.org/cytoscape-tutorials/contents/index.html#/>;
- Integrate results from the custom database and GNPS annotations by adding the *in-house* data as a new column in the node table, as shown in **Figure S23**.

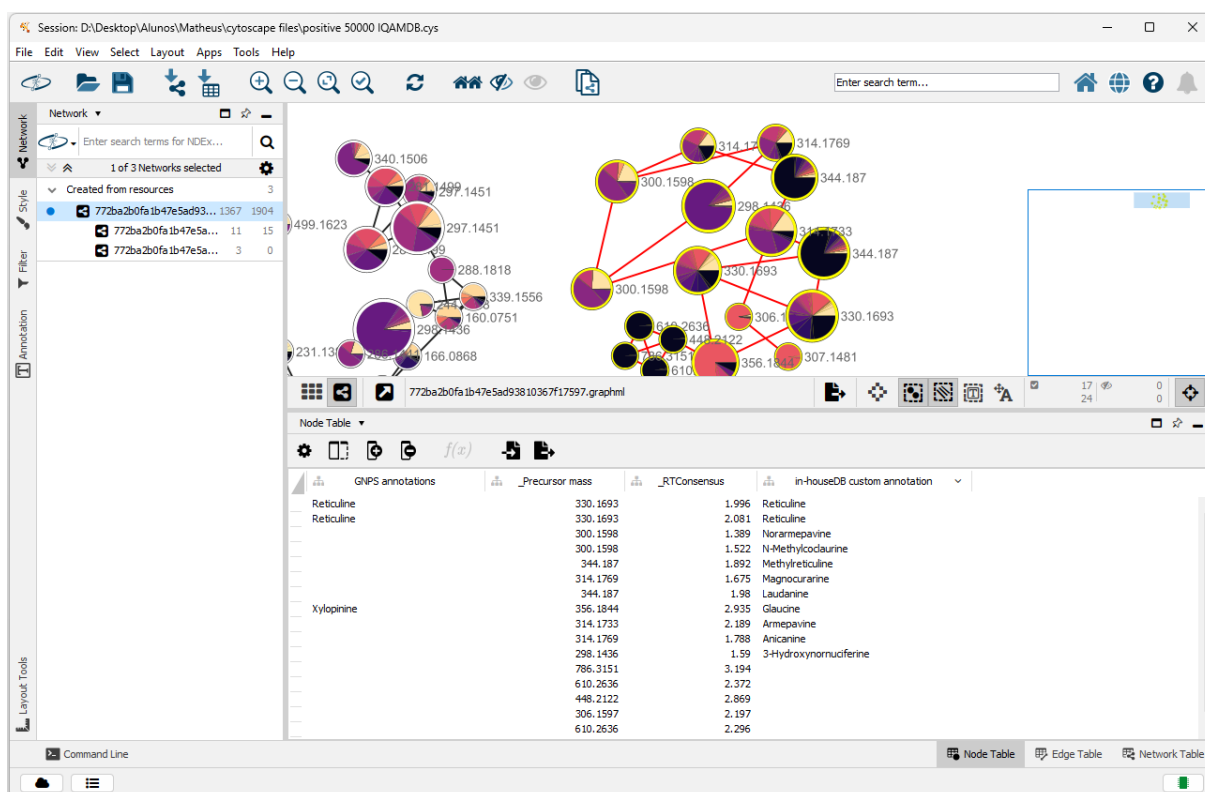


Fig. S23 | Customized molecular network on Cytoscape interface. In detail, the node table with GNPS automatic annotations for precursor mass and retention time, and *in-house* custom database annotations (monoisotopic mass match) integrated with visual resources.

Concluding remarks: Dive deep into your data! This exploratory phase is crucial in turning raw data into valuable knowledge, potentially guiding future research directions, informing decision-making processes, or contributing to scientific advancements.

Supplementary material and methods

Solvents, plant material and crude extracts preparation (SM-1)

All solvents used were LC-MS (acetonitrile) and HPLC grade, including hexane, methanol, and ethanol all were acquired from Sigma Aldrich® (St Louis, MO, USA). Formic acid was also supplied by Sigma-Aldrich® (St Louis, MO, USA). Ultrapure water was purified using a Millipore Milli-Q® water purification system (Millipore, Bedford, MA, USA). The liquid nitrogen was purchased from Linde® (Pullach, Munique, Germany). The OUPR herbarium (Federal University of Ouro Preto - UFOP) and CESJ herbarium (Federal University of Juiz de Fora - UFJF) supplied 1-3 leaves of 60 *Ocotea* sp. vouchers to this study. **Table S2** includes information about the geographical location of the plant collection, together with the deposit voucher numbers. The present research was registered on the National System for Governance of Genetic Heritage and Associated Traditional Knowledge (SisGen # A5A8F67). The *Ocotea* species received an identification code (ID) according to their specie scientific plant names, e.g. the *O. odorifera* (OD) (**Table S2**).

The 60 different *Ocotea* species' leaves vegetal material were weighed (20 mg each) and then pulverized using pistil and liquid nitrogen. Subsequently, 1.7 mL of ethanol: water 7:3 (v/v) was added to the powdered material for extraction. The extracts were placed into a warm ultrasound bath for 15 minutes at 35 °C (170 W, 50 kHz, L100 Schuster) and then centrifuged at 22 °C and 112 rcf (G-force). The collected supernatants (~1.6 mL) were partitioned with hexane (2 x 200 µL) for the removal of fatty material. Next, the extracts were filtered through polytetrafluoroethylene syringe filters (PTFE) of 0.22 µm and dried using a speed vacuum apparatus at 40 °C for 3 h. The extracts were maintained in a freezer at the temperature of -20 °C before UPLC-MS experiments.

Quality control, data acquisition and analysis (SM-2)

As part of our internal management system, to fulfil current quality requirement practices for data acquisition and metabolic fingerprint analysis, an analytical quality control (QC) sample was prepared by pooling together 10 µL from each *Ocotea* sample extract at a concentration of 1 mg/mL (600 µL total volume). Thus, the QC consisted of a composite of all *Ocotea* sp. crude extracts prepared at 1 mg/mL concentration. The QC was spread into three vials, kept at 4° in the instrument rack during the whole batch analysis, and injected at the beginning, middle and final chromatographic batch run. To monitor the system's background noise and reproducibility, one replicate (ID: VI), pooled QC, and one blank (water: acetonitrile,

1:1 v/v) were randomly injected every 15 injections, covering the start, middle, and end of the chromatographic batch data acquisition. The injection volume of each sample was 5 μ L.

Chromatographic separation was achieved using a high-quality C18 (ACQUITY UPLC®HSS T3) reversed-phase column (1.8 μ m, 100 x 2.1 mm) at 40 °C. The mobile phase consisted of two components: (A) 1% acidified water with formic acid, and (B) pure acetonitrile. A flow rate of 0.5 mL/min was maintained throughout the analysis. The chromatographic gradient began with an initial composition of 1% B and, was followed by a transition to 15% B at 0.1 min. Further changes in solvent composition occurred at 7.5 min (80% B), 8.5 min (99% B), and 8.6 min (1% B) until 10 min.

The mass spectrometer operated in MS^E acquisition mode with alternating high and low-energy scans. The collision energy was set at 3 eV for low-energy scans and 25-40 eV for high-energy scans. Instrument parameters, including cone voltage (40 V), capillary voltage (3.0 kV), cone gas flow (30 L/h), desolvation temperature (300 °C), source temperature (120 °C), and desolvation gas flow (600 L/h), were carefully optimized. High-purity nitrogen was employed for desolvation, collision, and cone gas. The full mass scan range was set from 50 to 1000 m/z for functions 1 and 2. To ensure accuracy and reproducibility, a solution of leucine-enkephalin was used as a lock mass with m/z 554.2622 (ESI⁻) and m/z 556.2768 (ESI⁺) for identification. MS data were continuously collected, and lock spray calibration was performed every 10 seconds.

Waters2mzML development and custom *in-house*DB (SM-3)

The converter was developed by Anja Miriam Prisching (email: anja.prisching@uni-oldenburg.de). Affiliation: Institute for Chemistry and Biology of the Marine Environment (ICBM), Carl-von-Ossietzky University, Oldenburg, Germany.

URL: <https://github.com/AnP311/Waters2mzML/releases/tag/v1.2.0>.

The proposed KNIME workflow streamlined and facilitated confident level 3 annotation by leveraging chemical structures from diverse sources. We utilize popular formats such as .mol, .mol2, or .sdf files, redrawn from original articles or downloaded from online chemical databases, e.g. ChEMBL (<https://www.ebi.ac.uk/chembl/>) as input files to the KNIME node readers. Our custom *in-house* database, *Ocotea*DB, contains 492 carefully curated chemical structures in .mol format. We have utilized the Nuclei of Bioassays, Ecophysiology and Biosynthesis of Natural Products Database (NuBBE) (<https://nubbe.iq.unesp.br/portal/nubbe-search.html>) on the construction of *Ocotea*DB. NUBBE (<https://nubbe.iq.unesp.br/portal/nubbe-search.html>) stores chemical structures from

Brazilian plant species with simple access to downloadable data, allowing the construction of particular plant species, genera, and family databases that can be automatically used in this workflow. For the *ActinomarineDB*, it comprised 6481 NPs related to the genera of *Actinomyces*, *Streptomyces*, *Salinospora*, *Micromonospora*, *Nocardia*, *Actinomadura* and *Rhodococcus*, downloaded from the npatlas database (<https://www.npatlas.org/>) in .csv format.

Also, the KNapSAck (http://www.knapsackfamily.com/knapsack_core/top.php) or any other NP product database can be used for the assembling of a SMILES or InChIKey table list (.csv or .xlsx). The KNIME output of this workflow is an automatic .csv table containing name, molecular formula and monoisotopic mass. This file can be directly uploaded to MZmine 3 or MS-DIAL. For MS-DIAL, a simple conversion to .txt format is necessary, with the table formatted to include columns for name, monoisotopic mass, and retention time (Metabolite name or ID, m/z , and RT, respectively). In both software, the retention time column of the .csv table should be null and ignored (typically set to the full run time for all rows).

MZmine 3 data processing and analysis (SM-4)

The following parameters were set of *Ocotea* extracts: mass detection of MS¹ scans (Scan filters, MS¹ level=1) using the centroid algorithm (Noise level, 500 and 350 for positive and negative, respectively), feature detection using the ADAP Chromatogram Builder module (Scan filter, MS¹ level = 1; Minimum consecutive scans, 5; Minimum intensity for consecutive scans, 2 times the noise level; Minimum absolute height, 4 times the noise level; m/z tolerance, 0.005 m/z or 10 ppm). Chromatogram resolving using Local minimum feature resolver (Dimension, retention time; Chromatographic threshold, 0.85; Minimum search range RT/mobility, 0.04; Minimum absolute height, equal to ADAP; Minimum ratio of peak top/edge, 1.7; Peak duration range, 0.0-1.0 absolute; Minimum # of data points, 5). Isotope filtering using ¹³C isotope filter module (m/z tolerance, 0.003 m/z or 5.0 ppm; Retention time tolerance, 0.05 absolute; Monotonic shape, true; Maximum charge, 1; Representative isotope, most intense). Alignment was performed using the RANSAC aligner (m/z tolerance, 0.007 or 12 ppm; RT tolerance, 0.15 absolute; RT tolerance after correction, 0.1 absolute; Minimum number of data points, 0.2; Threshold value, 0.1) and then gap filling using the Peak finder algorithm (Intensity tolerance, 0.2; m/z tolerance, 0.005 m/z or 10 ppm; RT tolerance, 0.1 absolute; Minimum data points, 4). Duplicate feature list rows filter (Filter mode, new average; m/z tolerance, 0.005 m/z or 10 ppm; R_t tolerance, 0.07 absolute) and Feature list blank subtraction (Select blanks raw files; Minimum # of detection in blanks, 3; Quantification, area; Ratio type, Average; Fold change increase, true, 3; Keep or remove features below fold change, remove), were applied to

achieve the final aligned feature list. All the other existing parameters not mentioned were set to default. The chemical level 3 annotation of the features was performed by applying the Search precursor mass using the local compound database (CSV) search of *Ocotea*DB (m/z tolerance: 0.005 m/z or 10 ppm; Retention time tolerance: 10 absolute; Use adducts, true, Setup: MS mode, respective to the data; Maximum charge, 1; Maximum molecules/cluster, 2; Adducts, $[M+H]^+$, $[M+Na]^+$; $[M+K]^+$ and $[M+NH_4]^+$ to positive mode; $[M-H]^-$, $[M+Cl]^-$, $[M+Br]^-$ and $[M+FA]^-$ to negative mode. The Feature list summaries of individual and aligned features with all parameters are online available (<https://zenodo.org/records/10383866>).

To optimize the processing parameters, QC samples were used to achieve better deconvolution, resolution and alignment of the features. The first processing with only the QC and blanks was performed looking for the best result and then a second processing with QC, samples and blanks were performed with the same parameters. To validate the data, samples and QC alignment were compared using the mainly known features as markers. Regarding data processing of the publicly available dataset of actinobacterial extracts, some minor alterations were made compared to the *Ocotea* dataset processing. For mass detection, the noise level was set to 200. Chromatogram building and resolving continued in the same configuration. To deisotoping, in addition to ^{13}C isotope filter (same configuration of parameters), the Isotope pattern finder module was utilised (Chemical elements, Br, Cl; m/z tolerance, 0.004 m/z or 7 ppm; maximum charge of isotope m/z , 1; search in scans, Single most intense). Alignment and gap filling were kept unaltered. Duplicate feature list rows filter was not necessary in this case due to the lower number of samples. Feature list blank subtraction was used in the same parameter configuration. The Search precursor mass using the local compound database (.csv) was also used to level 3 chemical annotation of the features but through *Actinomarine*DB (m/z tolerance: 0.005 m/z or 10 ppm; Retention time tolerance: 7.6 absolute; Use adducts, true, Setup: MS mode, respective to the data; Maximum charge, 2; Maximum molecules/cluster, 2; Adducts, $[M+H]^+$, $[M+2H]^{2+}$; $[2M+H]^+$).

MS-DIAL data processing and analysis (SM-5)

For data processing the parameters were set to *Ocotea* extracts as Data collection (MS^1 and MS^2 tolerance, 0.01 and 0.05 Da, respectively; Retention time begin/end, 0/10 minutes; MS^1 and MS/MS mass range begin/end 100/1000 Da; Maximum charged number, 2; Number of threads, 4), Peak detection (Maximum minimum peak height, 1000 and 50000 amplitude; Mass slice width of 0.1 Da for peak detection; Smoothing method, Linear weighted moving average; Smoothing level, 3 scans; Minimum peak width, 5 scans), MS^2 Dec data deconvolution

(Sigma window value, 0.5; MS/MS abundance cut-off, 10), Adduct search ($[M+H]^+$ and $[M-H]^-$, for each mode, respectively) and Alignment (Reference file, B-01p (blank sample); Retention time tolerance, 0.01 minute; MS^1 tolerance, 0.015; Remove features based on blank information, true; Sample max/blank average, 3 fold change; Gap filling by compulsion, true). Concerning the second data processing (publicly available dataset of actinobacterial extracts), the steps were highly similar to the *Ocotea* dataset, however, some minor alterations were performed to adequate the matrix differences. In Data collection, RT begin/end was set to 0/7 minutes and MS^1 and MS/MS mass range begin/end to 0/2000 Da. In addition, the option for considering Cl and Br elements was set. For peak detection, an amplitude of 1000 was applied to the Maximum minimum peak height. To adduct search, in addition to $[M+H]^+$, $[2M+H]^+$ and $[M+2H]^{2+}$ were also sought. For alignment, the option of removing features based on blank information was not set due to several blank chromatogram peaks being present in the same region of the peaks from the actinobacterial samples. The other parameters were not changed. Following data processing the alignment lists were exported to GNPS required file format (Export, Alignment results, GNPS export, true; MS/MS included matrix, true; Filtering by the ion abundances of blank samples, true; Export format, .mgf), with feature quantification table (.txt) and MS/MS spectral summary (.mgf).

Molecular networking and metabolite annotation analysis (SM-6)

The set parameters used for generating FBMN were the same for both datasets. Precursor ion mass tolerance of 0.02 Da and fragment ion mass tolerance of 0.05 Da. Advanced network options were set accordingly to our MS^E data type (Min pairs Cos, 0.6; Network TopK, 4; Minimum Matched Fragment ions, 4; Maximum Connected Component Size (Beta), 0; Maximum shift between precursors, 500 Da), as well the advanced library search options (Library Search Min Matched Peaks, 4; Search Analogs, Don't; Top results to report per query, 1; Score Threshold, 0.5; Maximum Analog Search Mass Difference, 100 Da). Other advanced options were not set or set as default. All the GNPS job links are provided together with Zenodo uploaded data.

The initial automated annotation was conducted by searching against GNPS libraries using FBMN. This automated annotation can also be performed within MS-DIAL using metabolomics MSP spectral kits. We utilized the isoquinoline-specific bank IQAMDB from GNPS to the positive mode for the *Ocotea* dataset as various alkaloids of *Ocotea* species belong to this class. Additionally, we used the NIH Natural Products Round 2 libraries in negative ion mode to access the flavonoid content in these samples. As no specific libraries to

actinobacterial-derived compounds NP were available on GNPS spectral libraries, we have preconized the main GNPS NP libraries to access the biggest number of metabolites in our samples. In this way, we have applied the following GNPS libraries: Sumner, Respect, NIH Natural Products Library, NIH Natural Products Library Round 2 in positive mode, Birmingham-UHPLC-MS-Pos, Berkeley Lab, GNPS Library, Tuebingen Natural Product Collection and MoNA libraries. It is worth noting that NP class and organism (family or genera) specific libraries are highly beneficial in enhancing the annotation process, as they help avoid irrelevant and non-real hits, and currently, it is not available in GNPS for use for annotation. Regarding the manual spectra checking, structural confirmation through careful inspection of fragmentation patterns by examining the MS¹ and MS^E spectra of the selected metabolite candidate was performed for each metabolite present in **Table 1** using the MZmine 3.

Supplementary archives (Zenodo)

Complete additional information regarding the applicability case of *Ocotea* spp. metabolomics data is available in the Zenodo link. The following data is available online: Level 3 annotation in both positive and negative mode; Metadata for GNPS jobs; Level 2 annotation from GNPS IQAMDB and NIH libraries; Feature tables generated from both MZmine 3 and MS-DIAL 4.9. The *Ocotea*DB and the *Actinomarine*DB were generated using the KNIME workflow (<https://zenodo.org/records/10383866>).

FBMN job links

Negative mode – NIH Natural Products (50000 amplitude cut-off)

<https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=79db88981e6844a7b9f1315dcebb587a>

Positive mode – IQAMDB (50000 amplitude cut-off)

<https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=6e3ef1b75b7341c786b1b740e844d970>

Negative mode – NIH Natural Products (1000 amplitude cut-off)

<https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=92dda20a10d84beab0dc6c5bcb6659a2>

Positive mode – IQAMDB (1000 amplitude cut-off)

<https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=cb02eaf164ac476195a8255a0af6a682>

Positive mode – Actinobacterial samples – NP GNPS libraries (1000 amplitude cut-off)

<https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=1148592a538c4a63b55f05a883cf2127>

MS-DIAL experiment file (.txt format file)

	A	B	C	D
1	Experiment	MS Type	Min m/z	Max m/z
2		0 SCAN	100	1000
3		1 MSMS	100	1000

7 CHAPTER V – OTHER SCIENTIFIC COLLABORATIONS

This chapter provides scientific collaborations in our research group and collaborators published in the literature during the time of this PhD. The knowledge, techniques and team-work learned during these four years of PhD were also built upon on these papers. Briefly, it consist of academic journal, year of publication, related main topic, and research title including publication DOI (Table 1).

Table 1. Other scientific collaborations.

(Continue)

N°	Academic Journal	Year	Topic	Title (DOI hyperlink)
1	Journal of Ethnopharmacology	2021	Ocotea and Lauracea plants	Confirmation of ethnopharmacological anti-inflammatory properties of <i>Ocotea odorifera</i> and determination of its main active compounds
2	Revista Brasileira de Farmacognosia	2021	Anti-inflammatory activity	Systematic review of anti-inflammatory agents from <i>Aspergillus Species</i>
3	The Lancet Child & Adolescent Health	2021	Data analysis	Risk factors for COVID-19 mortality in hospitalised children and adolescents in Brazil
4	Chemistry & Biodiversity	2022*	Other NP and bioactivities	Immunological modulation and control of parasitaemia by Ayahuasca compounds: therapeutic potential for Chagas's disease
5	Chemistry & Biodiversity	2022	Anti-inflammatory activity / Metabolomics	Anti-inflammatory markers of hops cultivars (<i>Humulus lupulus</i> L.) evaluated by untargeted metabolomics strategy
6	Revista Virtual de Química	2022	Other NP and bioactivities	Preparative HPLC chromatographic approach for the rapid isolation of phytotoxins from the fungus <i>Curvularia lunata</i> of <i>Spigelia antherlmia</i> leaves
7	Chemistry & Biodiversity	2023	Ocotea and Lauracea plants / Metabolomics	Anti-inflammatory activity of Lauraceae plant species and prediction models based on their metabolomics profiling data
8	Phytochemistry Letters	2023	Lauracea plants / Anti-inflammatory	New bicyclic [3.2.1] octane neolignan derivative from <i>Aniba firmula</i> with potent <i>in vivo</i> anti-inflammatory activity on account of dual inhibition
9	Natural Product Research	2023	Anti-inflammatory activity	In vivo evaluation of analgesic and anti-inflammatory activity of hydroalcoholic extracts from <i>Handroanthus impetiginosus</i> and their chemical composition
10	Journal of Toxicology and Environmental Health, Part A	2023	Data analysis	Age, gender, and 11 comorbidities as risk factors associated with COVID-19 mortality: A retrospective cohort including 1.8 million individuals
11	Brazilian Journal of Biology	2023	Other NP and bioactivities	Atividade schistosomicida do extrato etanólico dos galhos, folhas, flores e frutos de plantas de <i>Handroanthus impetiginosus</i> e caracterização do perfil metabólico por análise de UPLC-ESI-QTOF
12	Natural Product Research	2023	Other NP and bioactivities	Insecticidal activity of extracts of <i>Handroanthus impetiginosus</i> on <i>Plutella xylostella</i> (Lepidoptera: plutellidae) larvae

(Conclusion)

N°	Academic Journal	Year	Topic	Title (DOI hyperlink)
13	Chemistry & Biodiversity	2023	Other NP and bioactivities	Chemical composition, anthelmintic activity, and mechanism of action of <i>Lippia dominguensis</i> mold. essential oil on <i>Haemonchus contortus</i>
14	RSC Advances	2024	Ocotea and Lauracea plants	LC-MS/DIA-based strategy for comprehensive flavonoid profiling: an <i>Ocotea</i> spp. applicability case
15	Chemistry & Biodiversity	2024	Ocotea and Lauracea plants	<i>(submitted)</i> - UHPLC-HRMS/DIA metabolic profiling of <i>Ocotea diospyrifolia</i> (Meisn.) Mez
16	Chemico-Biological Interactions	2024	Ocotea and Lauracea plants	<i>(submitted)</i> - Cytotoxic activity of <i>Ocotea</i> species on breast cancer cells: insights into selectivity and cell cycle modulation
17	Natural Product Research	2024	Lauracea plants / Anti-inflammatory activity	Evaluation of the anti-inflammatory activity of <i>Acacia polyphylla</i> and identification of a new apigenin-3-C-glycosylated type flavonoid
18	Natural Product Research	2024	Lauracea plants / Anti-inflammatory activity	Ex vivo inhibition of PGE2 formation in human blood by four bicyclic [3.2.1] octane neolignans isolated from <i>Aniba firmula</i> bark, two with unusual structural pattern
19	Natural Product Research	2024	Anti-inflammatory activity	Phytochemical investigation of <i>Nigrospora zimmermanii</i> isolated from <i>Poincianella pluviosa</i> (Sibipiruna): metabolites characterisation and screening for anti-inflammatory activity
20	Chemical Biology and Drug Design	2024	Anti-inflammatory activity	<i>(in press)</i> - Novel synthesized benzophenone thiazole hybrids exhibited <i>ex vivo</i> and <i>in silico</i> anti-inflammatory activity
21	Food Research International	2024	Data analysis	Effects of geographical origin and post-harvesting processing on the bioactive compounds and sensory quality of Brazilian specialty coffee beans
22	ACS Chemical Neuroscience	2024	Metabolomics	Metabolomics unveils disrupted pathways in Parkinson's disease: towards biomarker-based diagnosis
23	Phytotherapy Research	2024	Other NP and bioactivities	<i>(submitted)</i> - Can Plant-Derived Sesquiterpene Lactones Be Promissory Anti-Melanoma Agents? A Systematic Review
24	Journal of Molecular Modeling	2024	Anti-inflammatory activity	<i>(submitted)</i> - Anti-inflammatory boosted-QSAR models for predicting Nitric Oxide inhibition by bioactive metabolites from <i>Aspergillus</i> species

Source: From author (2024)

Note: Each publication can be accessed by ctrl clicking the DOI hyperlink in each title.

*Scientific publication which I am the first author.

For the scientific publications listed in Table 1, the topics related to *Ocotea*/Lauraceae plants, anti-inflammatory activity, and metabolomics are highlighted in bold. The scientific papers from these topics were selected for further discrimination as they are related to the primary focus of this PhD project. Accordingly, brief summaries and the specific contributions of the candidate to these manuscripts have been highlighted below.

N°1 - Research article - (2021): Confirmation of ethnopharmacological anti-inflammatory properties of *Ocotea odorifera* and determination of its main active compounds

Ocotea odorifera has been traditionally used in southern Brazil to treat inflammatory conditions like rheumatism, though lacking scientific validation. This study investigated its anti-inflammatory potential and identified active compounds using phytochemical and metabolomic approaches. *In vivo* tests on extracts, fractions, and essential oils showed significant anti-inflammatory effects, particularly in reducing edema and neutrophil recruitment. S-(+)-reticuline was identified as a key important feature by multivariate analysis, which was later isolated and confirmed to be active. Another feature positively correlated with the anti-inflammatory activity is likely to be a new compound since zero hit on the comprehensive mass database were encountered. Essential oils also showed significant anti-inflammatory activity. Thus, this study supports the traditional use of *O. odorifera* and highlights it as a promising source of anti-inflammatory compounds.

Main contributions: To aid in drafting the manuscript, to assistance in the execution of anti-inflammatory assays, to assistance in the construction of metabolomics models and subsequent metabolite isolation, to aid in the elaboration of the final manuscript version.

N°2 - Review article - (2021): Systematic review of anti-inflammatory agents from *Aspergillus Species*

Fungi are known for causing various diseases but are also a promising source of bioactive metabolites. The chemical diversity of fungal metabolites offers potential for new anti-inflammatory therapies. Inflammation is linked to diseases like gout, autoimmune disorders, neurodegenerative diseases, and cancer. This systematic review identified 231 anti-inflammatory metabolites from *Aspergillus* fungi species, with four showing strong multi-target inhibition potential. *Aspergillus* species are highlighted as rich sources of anti-inflammatory compounds like alkaloids, terpenoids, and polyketides. However, more *in vivo* studies are needed to understand their mechanisms of action and translate it to clinical settings.

Main contributions: Assistance to conceive the idea and to draft the manuscript, assistance to data collection and analysis, and in the elaboration of the final manuscript version.

N°5 - Research article (2022): Anti-inflammatory markers of hops cultivars (*Humulus lupulus* L.) evaluated by untargeted metabolomics strategy

Hops (*Humulus lupulus* L.) are edible flowers commonly used to add flavour and aroma to beer, besides they have rich chemical diversity and medicinal potential. In this work, an *ex vivo* anti-inflammatory assay via the LPS-induced signalling pathway and metabolomics approaches were performed to evaluate the ability of hops to inhibit the production of prostaglandin E2 (PGE2) inflammatory mediator and analyze which metabolites produced by the nine different hop cultivars are potential anti-inflammatory markers. Columbus, Chinook and Hallertau Mittelfrüh hop cultivars yielded extracts with PGE2 release inhibition rates of 86.7, 92.5 and 73.5 %, respectively. According to the multivariate statistical analysis, the majority of the metabolites correlated with the activity were prenylated phloroglucinol and phenolic homologs. These results suggest promissory anti-inflammatory hop metabolites.

Main contributions: To aid in drafting the manuscript, to assistance in building the metabolomics models, and assistance in the execution of anti-inflammatory assays as well as statistical analysis of the results, to aid in gas phase fragmentation reactions interpretation, to aid in the elaboration of the final manuscript version.

N°7 - Research article - (2023): Anti-inflammatory activity of Lauraceae plant species and prediction models based on their metabolomics profiling data

The Lauraceae is a botanical family known for its anti-inflammatory potential. However, several species have not yet been studied. Thus, this work aimed to screen the anti-inflammatory activity of this plant family and to build statistical prediction models. The methodology was based on the statistical analysis of high-resolution liquid chromatography coupled with mass spectrometry data and the *ex vivo* anti-inflammatory activity of plant extracts. The results demonstrated significant anti-

inflammatory activity for several of these plants for the first time. Partial least square, artificial neural network, and stochastic gradient descent showed adequate fitting and predictive performance. Key anti-inflammatory markers, such as aporphine and benzyloisoquinoline alkaloids were annotated with confidence level 2. Additionally, the validated prediction models proved to be useful for predicting active extracts using metabolomics data and studying their most bioactive metabolites.

Main contributions: assistance to conceive the idea, to draft the manuscript, perform metabolite annotation, to build the metabolomics data prediction models, to aid in data analysis, to aid in the elaboration of the final manuscript version.

N°8 - Research article – (2023): New bicyclic [3.2.1] octane neolignan derivative from *Aniba firmula* with potent *in vivo* anti-inflammatory activity on account of dual inhibition

Two new bicyclic octane neolignans, along with a known one, were isolated from *Aniba firmula* (Lauraceae) through HPLC fractionation. Their structures were identified using spectroscopic analysis. Lauraceae is recognized for its anti-inflammatory potential. *In vivo* assays in mice showed that the two new neolignans significantly inhibited croton oil-induced ear edema, with one also inhibiting neutrophil recruitment. The findings suggest that the inhibition of edema occurred via the COX pathway, while the LOX pathway was not involved. These results indicate that these neolignans have a unique anti-inflammatory mechanism, making them promising leads for developing new effective anti-inflammatory drugs.

Main contributions: To aid in drafting the manuscript, to assistance plant material extraction and HPLC fractionation, and assistance in the execution of anti-inflammatory assays and analysis of the results, to aid in the elaboration of the final manuscript version

N°9 - Research article – (2023): *In vivo* evaluation of analgesic and anti-inflammatory activity of hydroalcoholic extracts from *Handroanthus impetiginosus* and their chemical composition by UPLC/MS analysis

This study demonstrates *in vivo* analgesic and anti-inflammatory properties of hydroalcoholic extracts of leaves, bark and flowers from the *Handroanthus impetiginosus* (Bignoniaceae) plant, recognized as 'Ipê roxo' in Brazil. The extracts were evaluated in male Swiss albino mice via oral administration. The *in vivo* paw edema test induced by carrageenan revealed that extracts of leaves and bark displayed relevant anti-inflammatory activity potential at the dosage of 100 mg/kg, 300 mg/kg, and 500 mg/kg. Likewise, the results obtained for leaves and flowers extracts suggested potent analgesic action in the conventional hot plate test. UPLC/MS analysis of the hydroalcoholic extracts samples identified metabolites belonging to several classes, mainly naphthoquinones and iridoids derivatives as well as flavonoids. The obtained results indicate that the extracts of *H. impetiginosus* plant parts could be considered as a complementary herbal medicine for the treatment of pain and inflammation disorders.

Main contributions: assistance to draft the manuscript, to perform metabolite annotation, to aid data analysis and the elaboration of the final manuscript version.

N°14 - Research article – (2024): LC-MS/DIA-based strategy for comprehensive flavonoid profiling: an *Ocotea* spp. applicability case

We introduce a liquid chromatography – mass spectrometry with data-independent acquisition (LC-MS/DIA)-based strategy, specifically tailored to achieve comprehensive and reliable glycosylated flavonoid profiling. It was applied to a dataset of six *Ocotea* plant species. This framework suggested 49 flavonoids potentially newly described for these plant species, alongside 45 known features within the genus. Flavonols kaempferol and quercetin, both exhibiting *O*-glycosylation patterns, were particularly prevalent. For the first time, the apigenin flavone backbone was also annotated in most of the examined *Ocotea* species. Apigenin derivatives were found mainly in the *C*-glycoside form, with *O. porosa* displaying the highest flavone : flavonol ratio. This work has underscored the untapped potential of LC-MS/DIA data for broad and reliable flavonoid profiling. Our study annotated more than 50 flavonoid backbones in each species, surpassing the current literature.

Main contributions: assistance to conceive the idea, to draft the manuscript, to perform metabolite annotation and characterization, to aid in data analysis, to aid in the elaboration of the final manuscript version.

N°15 - Research article – (2024): UHPLC-HRMS/DIA metabolic profiling of *Ocotea diospyrifolia* (Meisn.) Mez

This study presents the comprehensive metabolic profiling of *Ocotea diospyrifolia* (Meisn.) Mez using Ultra-High Performance Liquid Chromatography coupled with High-Resolution Data-Independent Acquisition Mass Spectrometry (UHPLC-HRMS/DIA). 43 distinct compounds were annotated across Metabolomics Standards Initiative (MSI) confidence levels 2, 3 and 4, including mainly alkaloids, glycosylated flavonoids, as well as lignoids, sesquiterpenoids and organic acids. This research highlights the advanced analytical capabilities for comprehensive metabolite analysis from a minimal amount sample, crucial for studying rare and endangered plants.

Main contributions: assistance to conceive the idea, to draft the manuscript, to perform data processing, metabolite annotation and characterization, to aid in the elaboration of the final manuscript version.

N°16 - Research article – (2024): Cytotoxic activity of *Ocotea* species on breast cancer cells: insights into selectivity and cell cycle modulation

Breast cancer is the most common cancer in women and a leading cause of death, prompting the search for new treatments with fewer side effects. This study investigates the cytotoxic activity of extracts from 60 *Ocotea* species, known for their pharmacological properties, against breast cancer cells (MCF-7). Using the MTT assay, the IC₅₀ values were calculated to assess cytotoxicity and selectivity towards cancer cells. *Ocotea villosa* extract showed promising results, with an IC₅₀ of 100 µg/mL, inhibiting colony formation, cell migration, and inducing a G1/S cell cycle arrest. These findings suggest *O. villosa*'s potential as a promising therapeutic candidate for more effective and selective breast cancer treatment with fewer adverse effects.

Main contributions: assistance to conceive the idea, to draft the manuscript, to perform plant material extraction, metabolite annotation, and to aid in the elaboration of the final manuscript version.

N°17 - Research article – (2024): Evaluation of the anti-inflammatory activity of *Acacia polyphylla* and identification of a new apigenin-3-C- glycosylated type flavonoid

Due to the harmful side effects of current anti-inflammatory drugs, this study focused on identifying new alternative substances by analyzing *A. polyphylla*. Various fractions of the *A. polyphylla* extract were tested in an *ex vivo* anti-inflammatory assay. The BH fraction showed the highest PGE2 inhibition (74.8%), outperforming reference drugs like dexamethasone and indomethacin, demonstrating relevant anti-inflammatory activity. Astragalin, a known 3-O-glucoside of kaempferol, was also isolated from *A. polyphylla* for the first time and exhibited moderate PGE2 inhibition (48.3%), while a newly identified compound, an apigenin-3-C-glycosylated flavonoid derivative, showed no anti-inflammatory activity. Besides, this study confirms the anti-inflammatory potential of *A. polyphylla*.

Main contributions: To aid in drafting the manuscript, to assistance HPLC fractionation, and assistance in the execution of anti-inflammatory assays as well as statistical analysis of the results, to aid in the elaboration of the final manuscript version.

N°18 - Research article – (2024): *Ex vivo* inhibition of PGE2 formation in human blood by four bicyclic [3.2.1] octane neolignans isolated from *Aniba firmula* bark, two with unusual structural pattern

Phytochemical investigation of the stem bark extract of *Aniba firmula* (Lauraceae) led to the isolation of two new bicyclic [3.2.1] octane neolignans, along with two known ones. These compounds were tested for their anti-inflammatory potential using an *ex vivo* model, focusing on the inhibition of the prostaglandin E2 (PGE2) inflammatory mediator as *A. firmula* stands out in the Lauraceae family as a source of potentially bioactive compounds. 2 and 3 exhibited significant anti-inflammatory activity by inhibiting the production of PGE2 in plasma samples, thus by interference with the

cyclooxygenase (COX) inflammatory pathway. These findings suggest that the bicyclic octane neolignan class holds promise for developing new anti-inflammatory agents.

Main contributions: To aid in drafting the manuscript, to assistance in sample preparation including extraction and fractionation, to assistance in the execution of anti-inflammatory assays and statistical analysis of the results, to aid in the elaboration of the final manuscript version.

N°19 - Research article – (2024): Phytochemical investigation of *Nigrospora zimmermanii* isolated from *Poincianella pluviosa* (Sibipiruna): metabolites characterisation and screening for anti-inflammatory activity

Endophytic fungi within plant tissues are promising sources of bioactive natural products. This study investigated the anti-inflammatory potential of an endophytic fungus isolated from the Brazilian medicinal plant *Poincianella pluviosa*. The fungus, identified as *Nigrospora zimmermanii* (FPD13), showed significant anti-inflammatory effects by inhibiting prostaglandin E2 (PGE2) release by 75.22% in an ex vivo assay. Phytochemical analysis led to the identification of three compounds: Nigrosporolide (a macrolide), Tyrosol (a phenyl-propanol), and Decarestrictine A (a terpene). These findings highlight the chemical diversity and anti-inflammatory potential of *P. pluviosa* endophytes, suggesting the need for further research into their bioactive metabolites.

Main contributions: To aid in drafting the manuscript, to assistance HPLC fractionation and metabolite characterization, and assistance in the execution of anti-inflammatory assays as well as statistical analysis of the results, to aid in the elaboration of the final manuscript version.

N°20 - Research article – (2024): Novel synthesized benzophenone thiazole hybrids exhibited ex vivo and in silico anti-inflammatory activity

Novel benzophenone–thiazole hybrids with different substituents were synthesized and evaluated for anti-inflammatory activity using an ex vivo human whole-blood assay. All hybrids (3c and 5a–h) showed significant anti-inflammatory activity via prostaglandin E2 (PGE2) release inhibition. Moreover, 5c (82.8% of PGE2 inhibition),

5e (83.1% of PGE2 inhibition), and 5h (82.1% of PGE2 inhibition) were comparable to the reference drugs. Molecular docking revealed potential preferable binding to the active sites of cyclooxygenase 2 (COX-2) and microsomal prostaglandin E synthase-1 (mPGES-1) enzymes. This study provides the first evidence that benzophenone-thiazole hybrids may also dock in mPGES-1, a new attractive anti-inflammatory drug target, besides providing promising *ex vivo* anti-inflammatory activity. Thus, the novel hybrids are promising anti-inflammatory lead compounds and highlight the significance of optimal substituent selection in the design of potent PGE2 inhibitors

Main contributions: Assistance to conceive the idea, to aid in drafting the manuscript, to assistance in the execution of anti-inflammatory assays and statistical analysis of the results, to aid in the elaboration of the final manuscript version.

N°22 - Research article – (2024): Metabolomics unveils disrupted pathways in Parkinson's disease: towards biomarker-based diagnosis

Parkinson's disease (PD) is a neurodegenerative disorder with complex symptoms, making accurate diagnosis challenging. This study used untargeted metabolomics, coupled with machine learning, to identify novel serum biomarkers for PD in a Brazilian cohort. Analyzing samples from 39 PD patients and 15 healthy controls, 15 metabolites were significantly associated with PD, with 11 being potential biomarkers identified for the first time. Disrupted metabolic pathways include caffeine metabolism, arachidonic acid metabolism, and primary bile acid biosynthesis. The machine learning model achieved a high accuracy of 94.1% in distinguishing PD patients from controls, surpassing the 80% accuracy of traditional clinical evaluations. These findings could improve the detection and monitoring of PD, paving the way for more precise diagnostics and therapeutic interventions. Our research emphasizes the role of metabolomics and machine learning in advancing our understanding of the chemical profile of neurodegenerative diseases.

Main contributions: To aid in drafting the manuscript, to assistance in data processing and interpretation, including metabolite annotation, pathway identification, and machine learning model prediction, to aid in the elaboration of the final manuscript version.

8. FINAL REMARKS

Metabolomics has overcome several of the practical challenges inherent to NP chemistry research, such as the speed of metabolite characterization when studying complex matrices. As a result, nowadays it represents a vital bridge between the intrinsic properties of compounds and the rediscovery of NP for drug discovery. The advent of more standardized databases, along with new chemometric tools and computational chemistry now allow for the rapid annotation and potential identification of previously reported compounds, while simultaneously pointing to unknown and novel bioactive compounds with therapeutic potential.

The cumulative findings of this project underscore the *Ocotea* genus as a rich source of bioactive compounds, with several species being acknowledged for significant pharmacological potential. This thesis presents a comprehensive investigation into the chemical diversity of the *Ocotea* genus, expanding the understanding of its metabolome and the practical implications of its bioactive compounds, particularly for bioprospecting anti-inflammatory compounds. Among these, *Ocotea villosa* stands out as a particularly promising species, exhibiting great anti-inflammatory effects, potentially with mechanisms of action different from current inflammatory drugs in the market.

Chapter I laid the groundwork by consolidating all known metabolites of *Ocotea* into the *OcoteaDB*, an *in-house* database that can be used as a valuable tool for the characterization of *Ocotea* species in further studies in the literature. Also a foundation for NP studies targeting bioactive compounds in the genus. This chapter also provided detailed insights into the biosynthetic pathways that give rise to bioactive scaffolds in the genus. By tracing these specialized biosynthetic routes, this work bridges the gap between NP chemistry and the broader pharmacological potential of *Ocotea* species.

Chapter II employed an untargeted metabolomics approach to chemically profile 60 different *Ocotea* species, many of which had never been chemically studied before. The exploratory analysis identified significant biomarkers correlated with anti-inflammatory activity, specifically through the inhibition of PGE₂, one of the crucial mediators in the inflammatory cascade. By correlating chemical profiles with pharmacological data, this work revealed critical alkaloids, such as aporphines and benzyloquinolines, which part of them were already reported to possess notable anti-

inflammatory properties. The discovery of these biomarkers from underexplored species highlights the bioprospecting potential of *Ocotea* for drug discovery.

Chapter III advanced the exploration of the genus by performing an untargeted study in promising species focusing on dual COX/LOX inhibitory biomarkers by concatenating data coming from UPLC-MS and NMR techniques. Multivariate statistics supplemented by machine learning models and STOCY annotated key markers, including alkaloids of both aporphine and benzilisoquinoline subclasses, one glycosylated flavonoid, and a sesquiterpenoid. These compounds were correlated to the dual anti-inflammatory activity, through simultaneous inhibition of PGE2 and LTB4 releasing. This chapter corroborates the anti-inflammatory relevance of previously correlated metabolites and points out novel potential anti-inflammatory agents. This, research emphasizes the potential of UPLC/MS – NMR data concatenation strategies for untargeted plant metabolomics studies.

Chapter IV presents the DIA-IntOpenStream pipeline, an open-source workflow designed to facilitate and enhance the annotation of complex metabolomics MS^E/DIA data using open software and tools. By putting together the advantages of MZmine and MS-DIAL software, and generating fast custom *in-house* databases using KNIME, this workflow addresses challenges in data processing and molecular networking of DIA-MS data using the GNPS platform. Its application to the *Ocotea* plant dataset demonstrated its effectiveness in annotating known compounds in complex matrices and offering a scalable and accessible solution for NP research. The inclusion of a publicly available actinobacterial extract spiked with authentic standards allowed detailed comparative analysis with existing methods. The pipeline holds promise for speeding up metabolite discoveries using DIA data towards more collaborative NP-based metabolomics research.

Therefore, the main scientific contributions of this thesis are twofold. First, it offers a detailed chemical and anti-inflammatory profile of several *Ocotea* species, many of which were chemically or pharmacologically investigated for the first time. Second, it provides innovative methodologies and practical tools for future bioprospecting, particularly under untargeted metabolomics and NP drug discovery. The discrimination of bioactive markers in *Ocotea* species underscores the urgent need to conserve these species, as many face the risk of extinction within different Brazil's biomes. Protecting biodiversity is not only crucial for maintaining ecological balance, but also for preserving the vast, untapped medicinal potential of these plant

species. By integrating classical phytochemistry with cutting-edge techniques such as UPLC/MS and NMR, alongside multivariate statistical analysis, STOCSY, molecular networking, and machine learning methods, this work establishes a solid foundation to prospect novel anti-inflammatory candidates using a minimal amount of plant sample extract and human blood volume to perform highly scalable *ex-vivo* screening anti-inflammatory experiments. These results corroborate future drug discovery efforts rooted in modern bioprospecting strategies, which simultaneously can alert and raise efforts for biodiversity protection.

Ultimately, NPs have long been central to drug discovery, as they offer an unparalleled arsenal of molecules shaped over millions of years through evolution on this planet. This thesis reinforces their continued relevance as sources of therapeutic innovation, particularly in the research of innovative anti-inflammatory treatments. Looking ahead, the isolation, semi-synthesis or synthesis of these metabolites, evaluation of their individual and synergic anti-inflammatory effects, and the further integration of metabolomics with other omics studies could be crucial to fully map all the metabolic pathways, targeted proteins and cells, and the mechanisms behind the anti-inflammatory activity of *Ocotea* species in a range of different pathological conditions. Herein, the obtained results support untargeted metabolomics for bioprospecting bioactive markers in the *Ocotea* genus, but also highlight the need for biodiversity protection in safeguarding future drug discovery.

REFERENCES

- ALARCON-BARRERA, J. C. *et al.* Recent advances in metabolomics analysis for early drug development. **Drug Discovery Today**, v. 27, n. 6, p. 1763–1773, 2022.
- ALCÂNTARA, B. G. V. *et al.* Anti-Inflammatory activity of Lauraceae plant. **Chemistry & Biodiversity**, v. 20, n. 9, p. 1–10, 2023.
- ALCÂNTARA, B. G. V. *et al.* Confirmation of ethnopharmacological anti-inflammatory properties of *Ocotea odorifera* and determination of its main active compounds. **Journal of Ethnopharmacology**, v. 264, p. 133378, Jan. 2021.
- ALKA, O. *et al.* DIAMetAlyzer allows automated false-discovery rate-controlled analysis for data-independent acquisition in metabolomics. **Nature Communications**, v. 13, n. 1, p. 1–9, 2022.
- ALSEEKH, S. *et al.* Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. **Nature Methods**, v. 18, n. 7, p. 747–756, 2021.
- AMILIA DESTRYANA, R. *et al.* Antioxidant and anti-inflammation activities of ocotea, copaiba and blue cypress essential oils *in vitro* and *in vivo*. **JAOCS, Journal of the American Oil Chemists' Society**, v. 91, n. 9, p. 1531–1542, 2014.
- ARON, A. T. *et al.* Reproducible molecular networking of untargeted mass spectrometry data using GNPS. **Nature Protocols**, v. 15, n. 6, p. 1954–1991, 2020.
- AYDOĞAN, C. *et al.* Miniaturized LC in molecular omics. **Analytical Chemistry**, v. 92, n. 17, p. 11485–11497, 2020.
- BATISTA, A. N. L. *et al.* Aromatic compounds from three brazilian Lauraceae species. **Quimica Nova**, v. 33, n. 2, p. 321–323, 2010.
- BOCCARD, J. *et al.* Mass spectrometry metabolomic data handling for biomarker discovery. **Proteomic and Metabolomic Approaches to Biomarker Discovery**, p. 369–388, 2019.
- BONDE, C. S. *et al.* Bio-guided fractionation and molecular networking reveal fatty acids to be principal anti-parasitic compounds in nordic seaweeds. **Frontiers in Pharmacology**, v. 12, p. 1–12, June 2021.
- BRAGA, C. P.; ADAMEC, J. Metabolome analysis. *In*: **ROUX, K. H.** (Ed.). Encyclopedia of bioinformatics and computational biology: ABC of Bioinformatics. [s.l.]. Academic press. v. 3, p. 463–475, 2019.

- BREITBACH, U. B. *et al.* Amazonian Brazilian medicinal plants described by C.F.P. von Martius in the 19th century. **Journal of Ethnopharmacology**, v. 147, n. 1, p. 180–189, 2013.
- BROUGHTON, G.; JANIS, J. E.; ATTINGER, C. E. The basic science of wound healing. **Plastic and Reconstructive Surgery**, v. 117, n. 7S, p. 12–34, 2006.
- CANUTO, G. A. B. *et al.* Metabolomics: Definitions, state-of-the-art and representative applications. **Química Nova**, v. 41, n. 1, p. 75–91, 2018.
- CARNEVALE NETO, F. *et al.* Evaluation of ion mobility spectrometry for improving constitutional assignment in natural product mixtures. **Journal of Natural Products**, v. 85, n. 3, p. 519–529, 2022.
- CARPENTER, K. A. *et al.* Deep learning and virtual drug screening. **Future Medicinal Chemistry**, v. 10, n. 21, p. 2557–2567, 2018.
- CHAGAS-PAULA, D. *et al.* A metabolomic approach to target compounds from the asteraceae family for dual COX and LOX inhibition. **Metabolites**, v. 5, n. 3, p. 404–430, 2015a.
- CHAGAS-PAULA, D. A. *et al.* Prediction of anti-inflammatory plants and discovery of their biomarkers by machine learning algorithms and metabolomic studies. **Planta Medica**, v. 81, n. 6, p. 450–458, 2015b.
- CHAGAS-PAULA, D.; OLIVEIRA, T.; FALEIRO, D. Outstanding anti-inflammatory potential of selected Asteraceae species through the potent dual inhibition of Cyclooxygenase-1 and 5-Lipoxygenase. **Planta Medica**, v. 81, p. 1296–1307, 2015.
- CHAN, J. Y. W. *et al.* Pheophorbide a, a major antitumor component purified from *Scutellaria barbata*, induces apoptosis in human hepatocellular carcinoma cells. **Planta Medica**, v. 72, n. 1, p. 28–33, 2006.
- CHASE, M. W. *et al.* An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. **Botanical Journal of the Linnean Society**, v. 181, n. 1, p. 1–20, 2016.
- CHAVERRI, C.; CICCIO, J. F. Essential oil of trees of the genus *Ocotea* (Lauraceae) in Costa Rica. I. *Ocotea brenesii*. **Revista de Biología Tropical**, v. 53, n. 3–4, p. 431–436, 2005.
- CHEN, L. *et al.* Inflammatory responses and inflammation-associated diseases in organs. **Oncotarget**, v. 9, n. 6, p. 7204–7218, 2018.

- CHEN, L. *et al.* Metabolite discovery through global annotation of untargeted metabolomics data. **Nature Methods**, v. 18, n. 11, p. 1377–1385, 2021.
- CHINI, M. G. *et al.* Targeting mPGES-1 by a combinatorial approach: identification of the aminobenzothiazole scaffold to suppress PGE2 levels. **ACS Medicinal Chemistry Letters**, v. 11, n. 5, p. 783–789, 2020.
- CLENDINEN, C. S.; MONGE, M. E.; FERNÁNDEZ, F. M. **Ambient Mass Spectrometry in Metabolomics**. [s.l.: s.n.]. v. 142, 2017.
- CONCEIÇÃO, R. S. *et al.* Rapid structural characterisation of benzyloquinoline and aporphine alkaloids from *Ocotea spixiana* acaricide extract by HPTLC-DESI-MSn. **Phytochemical Analysis**, v. 31, n. 6, p. 711–721, 2020.
- COUSSENS, L. M.; WERB, Z. Inflammation and cancer. **Nature**, v. 90, n. 1, p. 58–73, 2012.
- COY, E. D.; CUCA, L. E.; SEFKOW, M. COX, LOX and platelet aggregation inhibitory properties of Lauraceae neolignans. **Bioorganic and Medicinal Chemistry Letters**, v. 19, n. 24, p. 6922–6925, 2009.
- CUI, L. *et al.* Apoptosis induction by alantolactone in breast cancer MDA-MB-231 cells through reactive oxygen species-mediated mitochondrion-dependent pathway. **Archives of Pharmacal Research**, v. 41, n. 3, p. 299–313, 2018.
- CUSTÓDIO, D. L.; FLORÊNCIO DA VEIGA JUNIOR, V. Lauraceae alkaloids. **RSC Advances**, v. 4, n. 42, p. 21864–21890, 2014.
- DAVIES, V. *et al.* Rapid development of improved data-dependent acquisition strategies. **Analytical Chemistry**, v. 93, n. 14, p. 5676–5683, 13 apr. 2021.
- DE CAMARGO, M. J. *et al.* Sesquiterpenos de *Ocotea lancifolia* (Lauraceae). **Quimica Nova**, v. 36, n. 7, p. 1008–1013, 2013.
- DE VOS, R. C. H. *et al.* Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. **Nature Protocols**, v. 2, n. 4, p. 778–791, 2007.
- DEMARQUE, D. P. *et al.* Mass spectrometry-based metabolomics approach in the isolation of bioactive natural products. **Scientific Reports**, v. 10, n. 1, p. 1–9, 2020.
- DEWICK, P. M. **Medicinal natural**: a biosynthetic approach. [s.l.: s.n.]. v. 3, 2009.

- DING, K. *et al.* Structure-based discovery of mPGES-1 inhibitors suitable for preclinical testing in wild-type mice as a new generation of anti-inflammatory drugs. **Scientific Reports**, v. 8, n. 1, p. 1–9, 2018.
- DUGOWSON, C. E.; GNANASHANMUGAM, P. Nonsteroidal anti-inflammatory drugs. **Physical Medicine and Rehabilitation Clinics of North America**, v. 17, n. 2, p. 347–354, May 2006.
- DYER, L. A. *et al.* Modern approaches to study plant–insect interactions in chemical ecology. **Nature Reviews Chemistry**, v. 2, n. 6, p. 50–64, May 2018.
- EBBELS, T. M. D. *et al.* Recent advances in mass spectrometry-based computational metabolomics. **Current Opinion in Chemical Biology**, v. 74, p. 102288, June 2023.
- EVANS, A. M. *et al.* Dissemination and analysis of the quality assurance (QA) and quality control (QC) practices of LC–MS based untargeted metabolomics practitioners. **Metabolomics**, v. 16, n. 10, p. 1–16, 2020.
- FARAG, M. A. *et al.* Metabolite profiling and fingerprinting of commercial cultivars of *Humulus lupulus* L. (hop): A comparison of MS and NMR methods in metabolomics. **Metabolomics**, v. 8, n. 3, p. 492–507, 2012.
- FENAILLE, F. *et al.* Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: Where do we stand? **Journal of Chromatography A**, v. 1526, p. 1–12, Dec. 2017.
- FIEHN, O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. **Comparative and Functional Genomics**, v. 2, n. 3, p. 155–168, 2001.
- FINDLAY, B. L. The chemical ecology of predatory soil bacteria. **ACS Chemical Biology**, v. 11, n. 6, p. 1502–1510, June 2016.
- FIORUCCI, S. *et al.* Dual inhibitors of cyclooxygenase and 5-lipoxygenase. A new avenue in anti-inflammatory therapy? **Biochemical Pharmacology**, v. 62, n. 11, p. 1433–1438, 2001.
- FUNK, C. D. Prostaglandins and leukotrienes: Advances in eicosanoid biology. **Science**, v. 294, n. 5548, p. 1871–1875, 2001.
- GAUDÊNCIO, S. P.; PEREIRA, F. Dereplication: racing to speed up the natural products discovery process. **Natural Product Reports**, v. 32, n. 6, p. 779–810, 2015.

- GHOSH, R.; ALAJBEGOVIC, A.; GOMES, A. V. NSAIDs and cardiovascular diseases: role of reactive oxygen species. **Oxidative Medicine and Cellular Longevity**, v. 2015, p. 1–25, 2015.
- GOBBO-NETO, L.; LOPES, N. P. Plantas medicinais: fatores de influência no conteúdo de metabólitos. **Quimica Nova**, v. 30, n. 2, p. 374–381, 2007.
- GOTTLIEB, O. R. Chemosystematics of the Lauraceae. **Phytochemistry**, v. 11, n. 5, p. 1537–1570, 1972.
- GUO, J.; HUAN, T. Evaluation of significant features discovered from different data acquisition modes in mass spectrometry-based untargeted metabolomics. **Analytica Chimica Acta**, v. 1137, p. 37–46, 2020.
- HARVEY, A. L.; EDRADA-EBEL, R.; QUINN, R. J. The re-emergence of natural products for drug discovery in the genomics era. **Nature Reviews Drug Discovery**, v. 14, n. 2, p. 111–129, Feb. 2015.
- HOET, S. *et al.* Alkaloids from *Cassutha filiformis* and related aporphines: antitrypanosomal activity, cytotoxicity, and interaction with DNA and topoisomerases. **Planta Medica**, v. 70, n. 5, p. 407–413, 2004.
- HOLMES, E.; WILSON, I. D.; NICHOLSON, J. K. Metabolic phenotyping in health and disease. **Cell**, v. 134, n. 5, p. 714–717, Sept. 2008.
- HWANG, S. H. *et al.* Rationally designed multitarget agents against inflammation and pain. **Current Medicinal Chemistry**, v. 20, n. 13, p. 1783–1799, Mar. 2013.
- JAE-WON, L.; HEON, J. S. Metabolomics Based on UPLC-QTOF/MS Applied for the discrimination of *Cynanchum wilfordii* and *Cynanchum auriculatum*. **Journal of Postgenomics Drug & Biomarker Development**, v. 05, n. 04, 2015.
- KAPADIA, N.; HARDING, W. Aporphine alkaloids as ligands for serotonin receptors. **Medicinal Chemistry**, v. 06, n. 04, p. 241–249, 2016.
- KARNEZIS, T. *et al.* The connection between lymphangiogenic signalling and prostaglandin biology: A missing link in the metastatic pathway. **Oncotarget**, v. 3, n. 8, p. 890–903, 2012.
- KATAJAMAA, M.; OREŠIČ, M. Data processing for mass spectrometry-based metabolomics. **Journal of Chromatography A**, v. 1158, n. 1–2, p. 318–328, 2007.
- KATCHBORIAN-NETO, A. *et al.* Integrative open workflow for confident annotation and molecular networking of metabolomics MSE/DIA data. **Briefings in Bioinformatics**, v. 25, n. 2, p. 1–18, 2024.

- KLONT, F. *et al.* SWATH data independent acquisition mass spectrometry for screening of xenobiotics in biological fluids: opportunities and challenges for data processing. **Talanta**, v. 211, p. 120747, May. 2020.
- KOSMIDES, A. K. *et al.* Metabolomic fingerprinting: challenges and opportunities. **Critical Reviews in Biomedical Engineering**, v. 41, n. 3, p. 205–221, 2013.
- LANDSKRON, G. *et al.* Chronic inflammation and cytokines in the tumor microenvironment. **Journal of Immunology Research**, v. 2014, 2014.
- LESLIE, C. C. Cytosolic phospholipase A2: Physiological function and role in disease. **Journal of Lipid Research**, v. 56, n. 8, p. 1386–1402, 2015.
- LI, K. W. *et al.* Recent developments in data independent acquisition (DIA) mass spectrometry: application of quantitative analysis of the brain proteome. **Frontiers in Molecular Neuroscience**, v. 13, p. 1–8, Dec. 2020.
- LONG, N. P. *et al.* Toward a standardized strategy of clinical metabolomics for the advancement of precision medicine. **Metabolites**, v. 10, n. 2, p. 1–28, 2020.
- LOU, R.; SHUI, W. Acquisition and analysis of DIA-based proteomic data: a comprehensive survey in 2023. **Molecular & Cellular Proteomics**, v. 23, n. 2, p. 100712, Jan. 2024.
- LUDY CRISTINA, P.; LUIS ENRIQUE, C. Aporphine alkaloids from *Ocotea Macrophylla* (Lauraceae). **Quimica Nova**, v. 33, n. 4, p. 875–879, 2010.
- LYONS, T. R. *et al.* Cyclooxygenase-2-dependent lymphangiogenesis promotes nodal metastasis of postpartum breast cancer. **Journal of Clinical Investigation**, v. 124, n. 9, p. 3901–3912, 2014.
- MACEDO, A. L. *et al.* Absolute configuration of (–)-cubebin, a classical lignan with pharmacological potential, defined by means of chiroptical spectroscopy. **Journal of the Brazilian Chemical Society**, v. 31, n. 10, p. 2030–2037, 2020.
- MAHESH, G.; KUMAR, K. A.; REDDANNA, P. Overview on the discovery and development of anti-inflammatory drugs: Should the focus be on synthesis or degradation of pge2? **Journal of Inflammation Research**, v. 14, p. 253–263, 2021.
- MAJUMDER, M. *et al.* Prostaglandin E2 receptor EP4 as the common target on cancer cells and macrophages to abolish angiogenesis, lymphangiogenesis, metastasis, and stem-like cell functions. **Cancer Science**, v. 105, n. 9, p. 1142–1151, 2014.

- MANGISA, M. *et al.* Ethnomedicinal and phytochemical properties of sesquiterpene lactones from *Dicoma* (Asteraceae) and their anticancer pharmacological activities: A review. **Scientific African**, v. 13, p. e00919, 2021.
- MANNOCHIO-RUSSO, H. *et al.* Old meets new: mass spectrometry-based untargeted metabolomics reveals unusual larvicidal nitropropanoyl glycosides from the leaves of *Heteropterys umbellata*. **Journal of Natural Products**, v. 86, n. 3, p. 621–632, 2023.
- MANTOVANI, A. The inflammation – cancer connection. **FEBS Journal**, v. 285, n. 4, p. 638–640, 2018.
- MARQUES, C. A. Importância econômica da família Lauraceae Lindl. **Floresta e Ambiente**, v. 8, n. 1, p. 195–206, 2001.
- MD IDRIS, M. H. *et al.* Flavonoids as dual inhibitors of cyclooxygenase-2 (COX-2) and 5-lipoxygenase (5-LOX): molecular docking and in vitro studies. **Beni-Suef University Journal of Basic And Applied Sciences**, v. 11, n. 1, 2022.
- MEIRER, K.; STEINHILBER, D.; PROSCHAK, E. Inhibitors of the arachidonic acid cascade: Interfering with multiple pathways. **Basic and Clinical Pharmacology and Toxicology**, v. 114, p. 83–91, 2014.
- MULTHOFF, G.; MOLLS, M.; RADONS, J. Chronic inflammation in cancer development. **Frontiers in Immunology**, v. 2, p. 1–17, Jan. 2012.
- NANDI, P. *et al.* PGE2 promotes breast cancer-associated lymphangiogenesis by activation of EP4 receptor on lymphatic endothelial cells. **BMC Cancer**, v. 17, n. 1, p. 1–17, 2017.
- NEPHALI, L. *et al.* Mass spectral molecular networking to profile the metabolome of biostimulant bacillus strains. **Frontiers in Plant Science**, v. 13, p. 1–14, June. 2022.
- NOTHIAS, L. F. *et al.* Feature-based molecular networking in the GNPS analysis environment. **Nature Methods**, v. 17, n. 9, p. 905–908, 2020.
- PADUCH, R. The role of lymphangiogenesis and angiogenesis in tumor metastasis. **Cellular Oncology**, v. 39, n. 5, p. 397–410, 2016.
- PALOMINO, E. *et al.* Caparratriene, an active sesquiterpene hydrocarbon from *Ocotea caparrapi*. **Journal of Natural Products**, v. 59, n. 1, p. 77–79, 1996.
- PASSOS, B. G. *et al.* Essential oils from *Ocotea* species: chemical variety, biological activities and geographic availability. **Fitoterapia**, v. 156, p. 10565, Oct. 2022.

- PATIL, R. H.; PATIL, M. P.; MAHESHWARI, V. L. Bioactive secondary metabolites from endophytic fungi: a review of biotechnological production and their potential applications. **Studies in Natural Products Chemistry**, v. 49, p. 189–205, 2016.
- PILON, A. C. *et al.* Plant metabolomics: methods and challenges. **Quimica Nova**, v. 43, n. 3, p. 329–354, 2020.
- RANG, H. P. *et al.* **Farmacologia**. 7. ed. [s.l.: s.n.], 2011.
- RAO GAJULA, S. N.; NANJAPPAN, S. Metabolomics: a recent advanced omics technology in herbal medicine research. *In: Medicinal and aromatic plants: expanding their horizons through omics.* [s.l.]: Elsevier, p. 97–117, 2021.
- RAQUEL, J. *et al.* Sesquiterpenes of *Ocotea lancifolia* (Lauraceae). **Quimica Nova**, v. 33, n. 9, p. 1980–1986, 2010.
- ROZIC, J. G.; CHAKRABORTY, C.; LALA, P. K. Cyclooxygenase inhibitors retard murine mammary tumor progression by reducing tumor cell migration, invasiveness and angiogenesis. **International Journal of Cancer**, v. 93, n. 4, p. 497–506, 2001.
- RUDRAPAL, M. *et al.* Dual synergistic inhibition of COX and LOX by potential chemicals from Indian daily spices investigated through detailed computational studies. **Scientific Reports**, v. 13, n. 1, p. 1–27, 2023.
- SACCHETTI, G. *et al.* Essential oil of wild *Ocotea quixos* (Lam.) Kosterm. (Lauraceae) leaves from Amazonian Ecuador. **Flavour and Fragrance Journal**, v. 21, n. 4, p. 674–676, 2006.
- SALLEH, W. M. N. H. W.; AHMAD, F. Phytochemistry and biological activities of the genus *Ocotea* (Lauraceae): a review on recent research results (2000-2016). **Journal of Applied Pharmaceutical Science**, v. 7, n. 5, p. 204–218, 2017.
- SCHMIDT, R. *et al.* Microbe-driven chemical ecology: past, present and future. **The ISME Journal**, v. 13, n. 11, p. 2656–2663, Nov. 2019.
- SILVA, A. F. *et al.* Chemistry of leaves, bark, and essential oils from *Ocotea diospyrifolia* and anti-inflammatory activity – dual inhibition of edema and neutrophil recruitment. **Phytochemistry Letters**, v. 42, p. 52–60, July 2021.
- SPICER, R. *et al.* Navigating freely-available software tools for metabolomics analysis. **Metabolomics**, v. 13, n. 9, p. 1–16, 2017.
- SPICER, R. A.; SALEK, R.; STEINBECK, C. Comment: A decade after the metabolomics standards initiative it's time for a revision. **Scientific Data**, v. 4, p. 2–4, 2017.

- SUMNER, L. W. *et al.* Proposed minimum reporting standards for chemical analysis. Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). **Metabolomics**, v. 3, n. 3, p. 211–221, 2007.
- THEODORIDIS, G.; GIKA, H. G.; WILSON, I. D. Mass spectrometry-based holistic analytical approaches for metabolite profiling in systems biology studies. **Mass Spectrometry Reviews**, v. 30, n. 5, p. 884–906, Sept. 2011.
- TIMOSHENKO, A. V. *et al.* Role of prostaglandin E2 receptors in migration of murine and human breast cancer cells. **Experimental Cell Research**, v. 289, n. 2, p. 265–274, 2003.
- TROFIMOV, D.; DE MORAES, P. L. R.; ROHWER, J. G. Towards a phylogenetic classification of the *Ocotea* complex (Lauraceae): Classification principles and reinstatement of *Mespilodaphne*. **Botanical Journal of the Linnean Society**, v. 190, n. 1, p. 25–50, 2019.
- TSUGAWA, H. *et al.* MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. **Nature Methods**, v. 12, n. 6, p. 523–526, 2015.
- VAN DER KOOY, F. *et al.* Quality control of herbal material and phytopharmaceuticals with MS and NMR based metabolic fingerprinting. **Planta Medica**, v. 75, n. 7, p. 763–775, 2009.
- VAN DER LAAN, T. *et al.* Data-independent acquisition for the quantification and identification of metabolites in plasma. **Metabolites**, v. 10, n. 12, p. 1–14, 2020.
- VEREYKEN, L. *et al.* High-resolution mass spectrometry quantification: impact of differences in data processing of centroid and continuum data. **Journal of the American Society for Mass Spectrometry**, v. 30, n. 2, p. 203–212, 2019.
- VERPOORTE, R.; CHOI, Y. H.; KIM, H. K. Ethnopharmacology and systems biology: A perfect holistic match. **Journal of Ethnopharmacology**, v. 100, n. 1–2, p. 53–56, 2005.
- VINAYAVEKHIN, N.; SAGHATELIAN, A. Untargeted metabolomics. **Current Protocols in Molecular Biology**, v. 30.1.1, p. 1–24, 2010.
- VUCKOVIC, D. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. **Analytical and Bioanalytical Chemistry**, v. 403, n. 6, p. 1523–1548, 2012.

- WANG, B. *et al.* Exploring aporphine as anti-inflammatory and analgesic lead from *Dactylicapnos scandens*. **Organic Letters**, v. 22, n. 1, p. 257–260, 2020.
- WANG, M. *et al.* 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.
- WANG, M. T.; HONN, K. V.; NIE, D. Cyclooxygenases, prostanoids, and tumor progression. **Cancer and Metastasis Reviews**, v. 26, n. 3–4, p. 525–534, 2007.
- WANG, R.; YIN, Y.; ZHU, Z. J. Advancing untargeted metabolomics using data-independent acquisition mass spectrometry technology. **Analytical and Bioanalytical Chemistry**, v. 411, n. 19, p. 4349–4357, 2019.
- WOLFENDER, J. L. *et al.* Innovative omics-based approaches for prioritisation and targeted isolation of natural products-new strategies for drug discovery. **Natural Product Reports**, v. 36, n. 6, p. 855–868, 2019.
- WOLFENDER, J.; MARTI, G.; QUEIROZ, E. F. Advances in techniques for profiling crude extracts and for the rapid identification of natural products : dereplication , quality control and advances in techniques for profiling crude extracts and for the rapid identification of natural products : Derepl. **Current Organic Chemistry**, v. 14, n. 16, p. 1808–1832, 2010.
- WORLEY, B.; POWERS, R. Multivariate analysis in metabolomics. **Current Metabolomics**, v. 1, n. 1, p. 92–107, 2012.
- XU, X. *et al.* VEGF Induce vasculogenic mimicry of choroidal melanoma through the PI3k signal pathway. **BioMed Research International**, 2019.
- YAMAGUCHI, M. U. *et al.* Antifungal effects of ellagitannin isolated from leaves of *Ocotea odorifera* (Lauraceae). **Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology**, v. 99, n. 3, p. 507–514, 2011.
- YANG, K. *et al.* Dynamic simulations on the arachidonic acid metabolic network. **PLoS Computational Biology**, v. 3, n. 3, p. 0523–0530, 2007.
- YI, L. *et al.* Chemometric methods in data processing of mass spectrometry-based metabolomics: A review. **Analytica Chimica Acta**, v. 914, p. 17–34, 2016.
- YUAN, L. *et al.* 1,25-dihydroxyvitamin D3 inhibits growth of the breast cancer cell line MCF-7 and downregulates cytochrome P4501B1 through the COX-2/PGE2 pathway. **Oncology Reports**, v. 28, n. 6, p. 2131–2137, 2012.

- YULIANA, N. D. *et al.* Metabolomics for bioactivity assessment of natural products. **Phytotherapy Research**, v. 25, n. 2, p. 157–169, 2011.
- ZHANG, A. *et al.* Modern analytical techniques in metabolomics analysis. **Analyst**, v. 137, n. 2, p. 293–300, 2012.
- ZHANG, X. W. *et al.* Mass spectrometry-based metabolomics in health and medical science: A systematic review. **RSC Advances**, v. 10, n. 6, p. 3092–3104, 2020.
- ZHOU, B. N. *et al.* Isolation and biochemical characterization of a new topoisomerase I inhibitor from *Ocotea leucoxylo*n. **Journal of Natural Products**, v. 63, n. 2, p. 217–221, 2000.
- ZIDORN, C. Phytochemistry Plant chemophenetics – A new term for plant chemosystematics / plant chemotaxonomy in the macro-molecular era. **Phytochemistry**, v. 163, p. 147–148, Mar. 2019.