HELENA MANNOCHIO RUSSO

Metabolomic tools for chemosystematic studies and identification of bioactive natural products in Malpighiaceae species

Thesis submitted to the Institute of Chemistry, São Paulo University (UNESP), as part of the requirements for obtaining the title of Doctor in Chemistry.

Supervisor: Profª. Drª. Vanderlan da Silva Bolzani

Araraquara
2022

Russo, Helena Mannochio
R969m Metabolomic tools for chemosystematic studies and identification of bioactive natural products in Malpighiaceae species / Helena Mannochio Russo – Araraquara: [s.n.], 2022
285 p.: il.

Thesis (doctor) – Universidade Estadual Paulista, Instituto de Química
Advisor: Vanderlan da Silva Bolzani

1. Metabolites. 2. Mass spectrometry. 3. Chemotaxonomy. 4. Evolution. 5. Nitro compounds. I. Title.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Química, Araraquara. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada

HELENA MANNOCHIO RUSSO

Ferramentas metabolômicas para estudos quimiossistemáticos e identificação de produtos naturais bioativos em espécies de Malpighiaceae

Tese apresentada ao Instituto de Química, Universidade Estadual Paulista (UNESP), como parte dos requisitos para obtenção do título de Doutora em Química

Orientadora: Profª. Drª. Vanderlan da Silva Bolzani

Araraquara
2022

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Araraquara

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: "Ferramentas metabolômicas para estudos quimiossistemáticos e identificação de produtos naturais bioativos em espécies de Malpighiaceae"

AUTORA: HELENA MANNOCHIO RUSSO
ORIENTADORA: VANDERLAN DA SILVA BOLZANI

Aprovada como parte das exigências para obtenção do Título de Doutora em QUÍMICA, pela Comissão Examinadora:

Prof.ª Dr.ª VANDERLAN DA SILVA BOLZANI (Participaçao Virtual)
Departamento de Bioquimica e Quimica Organica / Instituto de Quimica - UNESP - Araraquara

Prof. Dr. CRISTIANO SOLEO DE FUNARI (Participaçao Virtual) p.
Departamento de Bioprocessos e Biotecnologia / Faculdade de Ciências Agronômicas - UNESP - Botucatu

Prof. Dr. EMERSON FERREIRA QUEIROZ (Participaçao Virtual) p.
Escola de Farmácia Genève-Lausanne / Universidade de Genebra - UNIGE - Genebra

Prof. Dr. RICARDO ROBERTO DA SILVA (Participaçao Virtual) p.
Departamento de Ciências BioMoleculares / Faculdade de Ciências Farmacêuticas de Ribeirão Preto - USP - Ribeirão Preto

Prof. Dr. IAN CASTRO GAMBOA (Participaçao Virtual)
Departamento de Bioquímica e Química Orgânica / Instituto de Química - UNESP - Araraquara

Araraquara, 26 de abril de 2022

Curriculum Vitae

1) Identification

Full name: Helena Mannochio Russo

Name in scientific citations: RUSSO, H. M.; RUSSO, HELENA MANNOCHIO; RUSSO, HELENA M.; MANNOCHIO RUSSO, HELENA; MANNOCHIO, RUSSO H; MANNOCHIO-RUSSO, H.; MANNOCHIO-RUSSO, HELENA.

Professional address: Nucleus of Bioassays, Biosynthesis, and Ecophysiology of Natural Products (NuBBE), Organic Chemistry Department, Institute of Chemistry, São Paulo State University (UNESP). Rua Professor Francisco Degni, 55 - CEP: 14800-900, Araraquara, SP, Brazil.

Link to CV Lattes: http://lattes.cnpq.br/9014119833113231

ORCID: 0000-0002-4961-2353

SCOPUS: https://www.scopus.com/authid/detail.uri?authorId=57208494456

Web of Science ResearcherID: ABD-5514-2020

E-mail: helenamrusso@gmail.com or helena.russo@unesp.br

2) Education

2018 – 2022 Ph.D. in Chemistry

Institute of Chemistry, São Paulo State University (UNESP), Araraquara, Brazil.

Title: Metabolomic tools for chemosystematic studies and identification of bioactive natural products in Malpighiaceae species.

Supervisor: Profª. Drª. Vanderlan da Silva Bolzani.

Scholarship: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Process number: 142014/2018-4.

2019 – 2020 Research Internship (8 months).

Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego (UCSD), California, United States.

Title: Molecular networking: a rational and innovative strategy for chemosystematic studies of Malpighiacae and identification of biologically active natural products.

Supervisor: Prof. Dr. Pieter Dorrestein.

Scholarship: Fulbright Program.

2015 – 2018 Master's in Chemistry

Institute of Chemistry, São Paulo State University (UNESP), Araraquara, Brazil.

Title: Evaluation of the chemical profile of the methanolic extract of *Niedenzuella multiglandulosa* leaves: isolation, characterization and identification of the bioactive compounds.

Supervisor: Profª. Drª. Vanderlan da Silva Bolzani (Co-supervisor: Prof. Dr. Emerson Ferreira Queiroz).

Scholarship: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Process number: 152431/2015-2.

Research Internship: School of Pharmacy Genève-Lausanne. University of Geneva (UNIGE), Geneva, Switzerland (5 months).

2010 – 2015 Bachelor in Chemistry

Institute of Chemistry, University of Campinas (UNICAMP), Campinas, Brazil.

Undergraduate research project: Application of continuous flow chemistry in the synthesis of natural products and derivatives with potential antitumor activity.

Supervisor: Prof. Dr. Julio Cezar Pastre.

Scholarship: Fundo de Apoio ao Ensino, à Pesquisa e à Extensão (FAEPEX).

2013 Undergraduate Research Internship

School of Pharmacy Genève-Lausanne. University of Geneva (UNIGE), Geneva, Switzerland (6 months).

Supervisor: Prof. Dr. Emerson Ferreira Queiroz.

Title: Phytochemical investigation of cashew (*Anacardium occidentale*).

3) Scientific awards

- **2022:** Selected to participate in the 71st Lindau Nobel Laureate Meeting (dedicated to Chemistry), in Lindau (Germany) – June 26th to July 1st, 2022.

- **2022:** Selected to attend the Baden-Würtemberg International Post Conference Programme of the 71st Lindau Nobel Laureate Meeting – July 1st to July 8th, 2022.
- **2022:** CAS Future Leaders in Chemistry, American Chemical Society.
- **2022:** Women in Science – UNESP. Category: Graduate student – Exact and Earth Sciences. Dean of Graduate Studies (PROPG/UNESP).
- **2019:** Fulbright Doctoral Dissertation Research Award (DDRA): 2019-2020. Fulbright Commission.
- **2019:** Panel awarded by the Natural Products Chemistry Division – 42nd Annual Meeting of the Brazilian Chemical Society.
- **2017:** Best presentation award in the VIII Symposium of Thermal Analysis (Brazil).

4) Publications in scientific journals

- BAUERMEISTER, ANELIZE; **MANNOCHIO-RUSSO, HELENA**; COSTA-LOTUFO, LETÍCIA V.; JARMUSCH, ALAN K.; DORRESTEIN, PIETER C. Mass spectrometry-based metabolomics in microbiome investigations. *Nature Reviews Microbiology*, v. 20, p. 143-160, **2022**.
- **MANNOCHIO-RUSSO, H.**; DE ALMEIDA, R.; NUNES, W. D.; BUENO, P. C. P.; CARABALLO-RODRÍGUEZ, A. M.; BAUERMEISTER, A.; DORRESTEIN, P.; BOLZANI, V. S. Untargeted metabolomics sheds light on the diversity of major classes of secondary metabolites in the Malpighiaceae botanical family. *Frontiers in Plant Science*, v. 13, p. 854842, **2022**.
- FERRARI, A. B. S.; OLIVEIRA, G. A.; **RUSSO, H. M.;** BERTOZO, L. C.; BOLZANI, V. S.; ZIED, D. C.; XIMENES, V. F.; ZERAIK, M. L. *Pleurotus ostreatus* and *Agaricus subrufescens*: investigation of chemical composition and antioxidant properties of these mushrooms cultivated with different handmade and commercial supplements. *International Journal Of Food Science And Technology*, v. 56, p. 452-460, **2021**.
- RIBEIRO, D. C.*; **RUSSO, H. M.*;** Fraige, K.; ZERAIK, M. L.; NOGUEIRA, C. R.; SILVA, P. B.; CODO, A. C.; CALIXTO, G. M. F.; MEDEIROS, A. I.; CHORILLI, M.; BOLZANI, V. S. Bioactive Bioflavonoids from *Platonia insignis* (Bacuri) Residues as Added Value Compounds. *Journal Of The Brazilian Chemical Society*, p. 786-799, **2021**.

- FERRARI, A. B. S.; MARCHEAFAVE, G. G.; **MANNOCHIO-RUSSO, H.;** BOLZANI, VANDERLAN S.; ZIED, D. C.; SCARMINIO, I. S.; ZERAIK, M. L. Chemical composition and chromatographic fingerprint of three strains of *Agaricus subrufescens* cultivated with handmade and commercial supplements. *Food Chemistry*, v. 363, p. 130227, **2021**.
- SCHMID, R. PETRAS, D. NOTHIAS, L-F. WANG, M. ARON, A. JAGELS, A. TSUGAWA, H. RAINER, J. GARCIA-ALOY, M. DUHRKOP, K. KORF, A. PLUSKAL, T. KAMENIK, Z. JARMUSCH, A. CARABALLO-RODRIGUEZ, A. WELDON, K. NOTHIAS-ESPOSITO, M. AKSENOV, A. BAUERMEISTER, A. ORIO, A. A. GRUNDMANN, C. VARGAS, F. KOESTER, I. GAUGLITZ, J. GENTRY, E. , HOVELMANN, Y. KALININA, S. PENDERGRAFT, M. PANITCHPAKDI, M. TEHAN, R. GOUELLEC, A. L. **MANNOCHIO RUSSO, HELENA** ARNDT, B. HUBNER, F. HAYEN, H. ZHI, H. RAFFATELLU, M. PRATHER, K. ALUWIHARE, L. BOCKER, S. MCPHAIL, K. HUMPF, H. KARST, U. DORRESTEIN, P. C. Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment. *Nature Communications*, v. 12, p. 3832, **2021**.
- GONÇALVES NUNES, WILHAN DONIZETE*; **MANNOCHIO RUSSO, HELENA***; DA SILVA BOLZANI, VANDERLAN; CAIRES, FLÁVIO JUNIOR. Thermal characterization and compounds identification of commercial *Stevia rebaudiana* Bertoni sweeteners and thermal degradation products at high temperatures by TG-DSC, IR and LC-MS/MS. *Journal of Thermal Analysis And Calorimetry*, v. 146, p. 1149-1155, **2021**.
- **RUSSO, H. M.**; QUEIROZ, E. F.; MARCOURT, L.; RUTZ, A.; ALLARD, P-M.; ALMEIDA, R. F.; CARVALHO, N. M.; WOLFENDER, J. L.; BOLZANI, V. S. Phytochemical analysis of the methanolic leaves extract of *Niedenzuella multiglandulosa* (Malpighiaceae), a plant species toxic to cattle in Brazil. *Phytochemistry Letters*, v. 37, p. 10-16, **2020**.
- SCUPINARI, T.*; **RUSSO, H. M.*;** FERRARI, A. B. S.; BOLZANI, V. S.; DIAS, W. P.; NUNES, E. O.; HOFFMANN-CAMPO, C. B.; ZERAIK, M. L. *Crotalaria spectabilis* as a source of pyrrolizidine alkaloids and phenolic compounds: HPLC-MS/MS dereplication and monocrotaline quantification of seed and leaf extracts. *Phytochemical Analysis*, v. 31, p. 747-755, **2020**.

- **MANNOCHIO-RUSSO, H.**; BUENO, P. C. P.; BAUERMEISTER, A.; ALMEIDA, R. F.; DORRESTEIN, P. C.; CAVALHEIRO, A. J.; BOLZANI, V. S. Can Statistical Evaluation Tools for Chromatographic Method Development Assist in the Natural Products Workflow? A Case Study on Selected Species of the Plant Family Malpighiaceae. *Journal Of Natural Products*, p. 3239-3249, **2020**.
- QUEIROZ, SUZANA APARECIDA S.; PINTO, MERI EMILI F.; BOBEY, ANTONIO F.; **RUSSO, HELENA M.;** BATISTA, ANDREA N.L.; BATISTA, JOAO M.; CODO, ANA C.; MEDEIROS, ALEXANDRA I.; BOLZANI, VANDERLAN S. Diterpenoids with inhibitory activity of nitrite production from *Croton floribundus*. *Journal Of Ethnopharmacology*, v. 249, p. 112320, **2020**.
- MATUTINO BASTOS, TANIRA; **MANNOCHIO RUSSO, HELENA**; SILVIO MORETTI, NILMAR; SCHENKMAN, SERGIO; MARCOURT, LAURENCE; GUPTA, MAHABIR; WOLFENDER, JEAN-LUC; FERREIRA QUEIROZ, EMERSON; BOTELHO PEREIRA SOARES, MILENA. Chemical Constituents of *Anacardium occidentale* as Inhibitors of *Trypanosoma cruzi* Sirtuins. *Molecules*, v. 24, p. 1299, **2019**.
- VALLI, MARILIA; **RUSSO, HELENA MANNOCHIO**; PILON, ALAN CESAR; PINTO, MERI EMILI FERREIRA; DIAS, NATHALIA B.; FREIRE, RAFAEL TEIXEIRA; CASTRO-GAMBOA, IAN; BOLZANI, VANDERLAN da SILVA. Computational methods for NMR and MS for structure elucidation II: database resources and advanced methods. *Physical Sciences Reviews*, v. 4, p. 20180167, **2019**.
- VALLI, MARILIA; **RUSSO, HELENA MANNOCHIO**; PILON, ALAN CESAR; PINTO, MERI EMILI FERREIRA; DIAS, NATHALIA B.; FREIRE, RAFAEL TEIXEIRA; CASTRO-GAMBOA, IAN; BOLZANI, VANDERLAN da SILVA. Computational methods for NMR and MS for structure elucidation I: software for basic NMR. *Physical Sciences Reviews*, v. 4, p. 20180108, **2019**.
- **RUSSO, HELENA MANNOCHIO***; NUNES, WILHAN DONIZETE GONÇALVES*; DA SILVA BOLZANI, VANDERLAN; IONASHIRO, MASSAO; CAIRES, FLÁVIO JUNIOR. Thermoanalytical and spectroscopic characteristics of young and old leaves powder and methanolic extracts of *Niedenzuella multiglandulosa*. *Journal of Thermal Analysis and Calorimetry*, v. 132, p. 771-776, **2018**.

- MEIRA, CÁSSIO S.; FILHO, JOSÉ MAURÍCIO DOS SANTOS; SOUSA, CAROLINE C.; ANJOS, PÂMELA S.; CERQUEIRA, JÉSSICA V.; NETO, HUMBERTO A. DIAS; DA SILVEIRA, RAFAEL G.; **RUSSO, HELENA M.;** WOLFENDER, JEAN-LUC; QUEIROZ, EMERSON F.; MOREIRA, DIOGO R.M.; SOARES, MILENA B.P. Structural design, synthesis and substituent effect of hydrazone-*N*-acylhydrazones reveal potent immunomodulatory agents. *Bioorganic & Medicinal Chemistry*, v. 26, p. 1971-1985, **2018**.
- MARILIA VALLI; **RUSSO, HELENA M.;** BOLZANI, V. S. The potential contribution of the natural products from Brazilian biodiversity to bioeconomy. *Anais Da Academia Brasileira De Ciências*, v. 90, p. 763-778, **2018**.

*Authors contributed equally.

5) Articles accepted for publication in scientific journals

- SILVA, G. L.; CAMPIDELLI, M.; FERRARI, A. B. S.; **MANNOCHIO-RUSSO, H.**; Fraige, K.; DAMETTO, A. C.; BOLZANI, V.; ZERAIK, M. L. In vitro antiglycation and antioxidant properties of *Eugenia pyriformis* leaves and fruits. *Natural Product Research*, **2022**. http://dx.doi.org/10.1080/14786419.2021.2005049.
- TAKAYAMA, K. S.; MONTEIRO, M. C.; SAITO, P.; PINTO, I. C.; NAKANO, C. T.; MARTINEZ, R. M.; THOMAZ, D. V.; VERRI JR, W. A.; BARACAT, M. M.; ARAKAWA, N. S.; **RUSSO, H. M.**; ZERAIK, M. L.; CASAGRANDE, R.; COUTO, R. O.; GEORGETTI, S. R. *Rosmarinus officinalis* extract-loaded emulgel prevents UVB irradiation damage to the skin. *Anais Da Academia Brasileira De Ciências*, **2022**.

6) Articles submitted in scientific journals

- SILVA, DULCE HELENA SIQUEIRA; **MANNOCHIO-RUSSO, HELENA**; LAGO, JOÃO HENRIQUE GHILARDI; BUENO, PAULA CAROLINA PIRES; MEDINA, REBECA PREVIATE; BOLZANI, VANDERLAN DA SILVA; VILEGAS, WAGNER, NUNES, WILHAN DONIZETE GONÇALVES. Bioprospecting as a strategy for conservation and sustainable use of the Brazilian Flora. *Biota Neotropica*, **2022**.
- ZANATTA, ANA C.; BORGES, MAIARA S.; **MANNOCHIO-RUSSO, HELENA**; HEREDIA-VIEIRA, SILVIA CRISTINA; SANTOS, LOURDES CAMPANER; RINALDO, DANIEL; VILEGAS, WAGNER. Green chromatography as a novel

alternative for the quality control of *Serjania marginata* Casar. leaves. *Microchemical Journal*, **2022**.

7) Book chapters

- MARILIA VALLI; **MANNOCHIO RUSSO, HELENA**; PILON, ALAN CESAR; PINTO, M. E. F.; DIAS, NATHALIA B.; FREIRE, RAFAEL TEIXEIRA; CASTRO-GAMBOA, IAN; BOLZANI, V. S. Computational methods for NMR and MS for structure elucidation II: database resources and advanced methods. In: Fidele Ntie-Kang. (Org.). Chemoinformatics of Natural Products - Volume 1 Fundamental Concepts. 1ed. Berlim: De Gruyter, **2020**, p. 205-228.
- MARILIA VALLI; **MANNOCHIO RUSSO, HELENA**; PILON, ALAN CESAR; PINTO, M. E. F.; DIAS, NATHALIA B.; FREIRE, RAFAEL TEIXEIRA; CASTRO-GAMBOA, IAN; BOLZANI, V. S. Computational methods for NMR and MS for structure elucidation I: software for basic NMR. In: Fidele Ntie-Kang. (Org.). Chemoinformatics of Natural Products - Volume 1 Fundamental Concepts. 1ed. Berlim: De Gruyter, **2020**, p. 177-204.

8) Participations in scientific events during the doctorate period

- **MANNOCHIO-RUSSO, H.**; BUENO, P. C. P.; BAUERMEISTER, A.; ALMEIDA, R. F.; DORRESTEIN, P. C.; CAVALHEIRO, A. J.; BOLZANI, V. S. Use of Statistical Tools and Automation to Accelerate the Natural Products Workflow – A Malpighiaceae Case Study. In: II Bio.Natural-Bioactive Natural Products Research Meeting, Lisbon, Portugal, **2021**. Invited talk (online).
- **MANNOCHIO-RUSSO, H.**; ALMEIDA, R. F.; NUNES, W. D. G.; ALBERNAZ, L.; ESPINDOLA, L. S.; BOLZANI, V. S. Screening of Malpighiaceae Plant Extracts Against *Aedes aegypti* Larvae and Metabolomic Investigation of the Most Active Species. In: 8th Brazilian Conference on Natural Products, Goiânia/GO, Brazil, **2021**. Poster presentation (online).
- RAMOS, A. V. G.; **MANNOCHIO-RUSSO, H.**; CABRAL, M. R. P.; SAHM, B.; LOTUFO, L.; CARMO, M.; SARRAGIOTTO, M. H.; BALDOQUI, D. C. Antitumoral activity of *Vernonanthura cuneifolia* and targeted isolation of sesquiterpene lactones based on Molecular Networking. In: 8th Brazilian Conference on Natural Products, Goiânia/GO, Brazil, **2021**. Poster presentation (online).

- **MANNOCHIO-RUSSO, H.**; ALMEIDA, R. F.; BUENO, P. C. P.; BAUERMEISTER, A.; CARABALLO-RODRIGUEZ, A. M.; DORRESTEIN, P. C.; BOLZANI, V. Untargeted metabolomics sheds light on the secondary metabolism of Malpighiaceae family. In: GA – 69th Annual Meeting, Bonn, Germany, **2021**. Poster presentation (online).
- **RUSSO, H. M.**; BUENO, P.; BAUERMEISTER, A.; ALMEIDA, R. F.; DORRESTEIN, P. C.; CAVALHEIRO, A. J.; BOLZANI, V. S. Experimental design applied in UPLC-DAD method development of Malpighiaceae species extracts and metabolite identification. In: American Chemical Society Spring 2020 National Meeting, Philadelphia, United States **2020**. Poster presentation (online).
- ZANATTA, A. C.; **RUSSO, H. M.**; MEDEIROS, R. D.; RINALDO, D.; VILEGAS, W. Perfil químico e screening de atividades biológicas para os extratos de *Byrsonima intermedia*, *Serjania marginata* e *Terminalia catappa*. In: XII Simpósio Brasileiro de Farmacognosia, Petrópolis/RJ, Brazil, **2019**. Poster presentation.
- **RUSSO, H. M.**; ALVES, N. F.; RIBEIRO, D. C.; ZOCOLO, G. J.; VALLI, M.; BOLZANI, V. S. Identificação de metabólitos secundários com atividade anti-glicante em extratos da espécie *Talisia esculenta*. In: XII Simpósio Brasileiro de Farmacognosia, Petrópolis/RJ, Brazil, **2019**. Poster presentation.
- **RUSSO, H. M.**; SCUPINARI, T.; FERRARI, A. B. S.; NUNES, E. O.; BOLZANI, V. S.; ZERAIK, M. L. Chemical profile of *Crotalaria spectabilis* extracts and evaluation of nematode control in *Heterodera glycines*. In: 42a Reunião Anual da Sociedade Brasileira de Química, Joinville/SC, Brazil, **2019**. Poster presentation.
- **RUSSO, H. M.**; ALMEIDA, R. F.; CAVALHEIRO, A. J.; BOLZANI, V. S. Experimental planning for UPLC method development of hydroalcoholic extracts from Malpighiaceae species. In: XVII Latin American Symposium on Chromatography and Related Techniques, Aracaju/SE, Brazil, **2019**. Poster presentation.
- FERRARI, A. B. S.; **RUSSO, H. M.**; ZIED, D. C.; BOLZANI, V. S.; ZERAIK, M. L. Dereplication by LC-MS/MS of three strains of *Agaricus subrufescens* grown with different supplements. In: XVII Latin American Symposium on Chromatography and Related Techniques, Aracaju/SE, Brazil, **2019**. Poster presentation.
- FERRARI, A. B. S.; SCUPINARI, T.; RUSSO, H. M.; DIAS, W. P.; NUNES, E. O.; ZERAIK, M. L. LC-MS/MS analysis of *Crotalaria spectabilis* leaves and seeds

extracts and its nematicidal and nematostatic effects. In: XVII Latin American Symposium on Chromatography and Related Techniques, Aracaju/SE, Brazil, **2019**. Poster and oral presentation.

- RUSSO, HELENA MANNOCHIO; RIBEIRO, D. C.; Fraige, K.; ZERAIK, M. L.; BOLZANI, V. S. Search for natural products of economic interest in fruit residues of *Platonia insignis*: bioactivity evaluation and structural characterization. In: XXV Encontro de Química da Região Sul (SBQSul), Londrina/PR, Brazil, **2018**. Oral presentation.

9) Short-term courses

- **2021:** Machine Learning in Chemometrics, 6 hours. Lecturer: Prof. Dr. Peter Harrington. V Escola de Inverno de Quimiometria. Online.
- **2021:** Multivariate regression: basic concepts and application examples, 8 hours. Lecturer: Prof. Dr. Dmitry Kirsanov. Escola de Inverno de Quimiometria. Online.
- **2018:** Morfologia e sistemática de Malpighiaceae Neotropical, 30 hours. Lecturer: Dr. Rafael Felipe de Almeida. Universidade Federal de Minas Gerais (UFMG).

10) Organization and/or presentations of workshops and courses

- **2021:** Lecturer of the short-term course “Basic concepts of mass spectrometry and use of the Global Natural Products Social Molecular Networking (GNPS) platform”. 10 hours. Online. Faculty of Pharmaceutical Sciences, Federal University of Alfenas (UNIFAL). 35 participants.
- **2021:** Workshop – “Molecular Networking with GC-MS data (in Portuguese)”. 4 hours. Online. Estimated audience: 80 people.
- **2020:** Workshop – “Feature-Based Molecular Networking (in Portuguese)”. 4 hours. Online. Estimated audience: 200 people
- **2020:** Workshop – “Molecular Networking (in Portuguese)”. 2 hours. Online. Estimated audience: 450 people.
- **2020:** Workshop – “Automated tools for developing chromatographic methods: Fusion QbD”. 2 hours. Online. Estimated audience: 90 people.
- **2019:** Workshop – “Young Researchers: future leaders in the academic and industrial sectors”. 4 hours. Workshop held in the 42a Reunião Anual da Sociedade Brasileira de Química, Joinville/SC, Brazil. Estimated audience: 80 people.

11) Scientific supervisions

- **2018-2019:** Anna Beatriz Sabino Ferrari (Department of Chemistry, State University of Londrina, UEL). Master research project, “Chemical analysis of *Agaricus subrufescens* mushroom grown with different handmade and commercial supplements”. Supervisor: Profª. Drª. Maria Luiza Zeraik.
- **2018-2019:** Giselle Lopes da Silva (Department of Chemistry, State University of Londrina, UEL). Master research project, “Phytochemical study and antiglycant and antioxidant evaluation of fruits *Psidium cattleianum* var. lucidum”. Supervisor: Profª. Drª. Maria Luiza Zeraik.
- **2019:** Natalia Felix Alves (Chemical Engineering, São Paulo State University, UNESP). Undergraduate research project, “Bioguided study of *Talisia esculenta* for the identification of secondary metabolites inhibitors of advanced glycation”. Supervisor: Profª. Drª. Vanderlan da Silva Bolzani.
- **2018:** Dayane Castro Ribeiro (Faculty of Pharmaceutical Sciences, São Paulo State University, UNESP). Undergraduate research project, “Evaluation of the potential of Brazilian fruits for the development of safe functional food of economic interest”. Supervisor: Profª. Drª. Vanderlan da Silva Bolzani.

12) Scientific affiliations

- Member of the Brazilian Chemical Society (SBQ, 2019 – Present) – Member of the SBQ Young Researchers Committee (JP-SBQ) since its creation.
- Member of the Metabolomics Society (2021 – Present).
- Member of the International Young Chemists Network (IYCN/IUPAC) – Brazilian delegate (2021 – Present).

13) Other relevant information

- Member of the editorial board of Revista Virtual de Química (RVSBQ) – Special Edition coordinated by JP-SBQ: “Science Developed by Young Brazilian Researchers / A Ciência Desenvolvida por Jovens Pesquisadores Brasileiros”. **2021-2022**.
- Seven interviews granted to news outlets about the doctoral work focused on the search for natural substances to combat *Aedes aegypti*, **2021**.

I dedicate this thesis to my dear parents, Elisa and Valter, for their unconditional support in my personal and professional life. My journey to this point is also a victory for them!

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation and thanks to (in Portuguese):

Inicialmente, gostaria de agradecer aos meus amados pais, Elisa e Valter, por sempre terem me apoiado e servido de exemplo não só de carreira, mas para a vida. Sem vocês eu não teria conseguido chegar tão longe.

Ao meu marido e companheiro, Wilhan D. G. Nunes, que acompanha todos os meus passos e conquistas desde o mestrado. Ter você ao meu lado fez com que essa trajetória tenha sido mais leve. Obrigada pela ajuda e apoio, tanto no âmbito pessoal quanto no profissional! Eu te amo!

À minha orientadora, professora Dra. Vanderlan, por ter me aceitado como aluna de doutorado, pela orientação, confiança, por todas as oportunidades que me concedeu, e por todos os ensinamentos transmitidos durante todos esses anos. Tê-la como orientadora é, sem sombra de dúvidas, uma grande inspiração!

Ao professor Dr. Pieter Dorrestein, por ter me recebido em seu laboratório durante meu estágio de doutorado-sanduíche, por todas as oportunidades que me concedeu e por todos os conhecimentos transmitidos. A minha estadia no Dorrestein Lab foi muito além do que eu poderia imaginar!

Um agradecimento especial ao Dr. Rafael Felipe de Almeida, pelos ensinamentos transmitidos, amizade e pela excelente colaboração. Grande parte desse trabalho só foi possível de ser realizada devido à sua grande ajuda!

Às professoras Dra. Lorena Albernaz e Dra. Laila Espíndola pelos ensaios realizados com o *Aedes aegypti*, pela excelente colaboração, atenção, e por toda a paciência.

Outro agradecimento especial à professora Dra. Maria Luiza Zeraik, pela grande amizade de longa data e por todas as oportunidades de colaboração que me concedeu durante o período de doutorado! Aprendi e aprendo muito com você, obrigada pela amizade e confiança.

Aos doutores Nivaldo Boralle, Juliana Rodrigues, João Bronzel pelo excelente apoio técnico no laboratório de RMN e nos laboratórios do NuBBE. Obrigada pelo tempo dedicado às análises e treinamentos, eu aprendi muito com vocês.

À professora Dra. Paula Bueno, pela amizade e ensinamentos transmitidos. Ter sua colaboração em grande parte desse trabalho foi essencial para produzirmos um trabalho de qualidade. É uma parceria que quero levar para a vida!

À Dra. Marilia Valli, pela amizade, por todas as oportunidades de colaboração, e por ter me chamado para participar da JP-SBQ!

Ao professor Dr. Alberto Cavalheiro pelos ensinamentos transmitidos, pela colaboração e amizade.

Também gostaria de agradecer ao professor Dr. Craig Nelson e ao M.Sc. Sean Swift, excelentes colaboradores da University of Hawai‘i, pelos ensinamentos transmitidos e pela oportunidade de colaborar ativamente nos projetos de Waimea Bay.

Aos meus colegas e amigos do Bolzani Team, NuBBE, e UNESP: Suzana Queiroz, Meri Emili Pinto, Antonio Bobey, Luiz Dutra, Ana Letícia Pires, Andrés Peña, Maria do Carmo, Rafael Vieira, Rafael Viana, Marcus Vinicius, Naira Anhesine, Karina Fontanin, Ana Zanatta e muitos outros. Também um agradecimento especial ao Marlon Oliveira, que mesmo já tendo defendido a alguns anos, mantemos uma amizade sincera.

Aos colegas e amigos do Dorrestein Lab: Alan Jarmusch, Alexander Aksenov, Allegra Aron, Amina Bouslimani, Anelize Bauermeister, Anupriya Tripathi, Carlismari Grundmann, Daniel Petras, Emily Gentry, Julia Gautlitz, Mabel González, Mauricio Caraballo, Mingxun Wang, Morgan Panitchpakdi, Tiago Leão. Thank you for the instructions, help, collaboration and friendship! My stay in San Diego would not be the same without you all.

Ao programa Fulbright pela bolsa concedida, tornando possível o sonho de realizar o doutorado-sanduíche em San Diego, Estados Unidos. Além da bolsa concedida, a

experiência Fulbright como um todo foi muito gratificante e me permitiu conhecer pessoas incríveis de diversos lugares do mundo. Sou muito feliz em fazer arte da família Fulbright!

Aos servidores do Instituto de Química, em especial à Seção Técnica de Pós-Graduação e à Biblioteca pelo apoio, excelente trabalho e competência.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de doutorado concedida (#142014/2018-4).

Aos órgãos de fomento à pesquisa, FAPESP e CAPES que financiaram direta ou indiretamente este trabalho.

Finalmente, a todos que contribuíram direta ou indiretamente para este trabalho.

“If I have seen further, it is by standing on the shoulders of giants”

– Isaac Newton

RESUMO

Historicamente, as plantas estão entre as fontes de produtos naturais mais investigadas, e contribuíram significativamente para o desenvolvimento do campo da química de produtos naturais. Metabólitos de plantas representam uma diversidade estrutural rica, com inúmeras atividades biológicas que inspiraram o desenvolvimento de diversos fármacos. O Brasil é considerado um *hotspot* de biodiversidade contendo milhares de espécies sem nenhuma informação química relatada até o momento, representando um vasto potencial para estudos químicos. O desenvolvimento contínuo de ferramentas e bancos de dados de bioinformática de produtos naturais pode ser de grande valia para auxiliar e acelerar estudos na área de produtos naturais no Brasil. Nesse contexto, a família botânica Malpighiaceae se destaca por apresentar uma grande diversidade de produtos naturais reportados na literatura, e ainda contém centenas de espécies quimicamente inexploradas. O presente trabalho tem como objetivo estudar plantas da família Malpighiaceae, com três principais objetivos: (1) desenvolver um método cromatográfico otimizado (em UHPLC-PDA) para análise de uma mistura de extratos de espécies da família Malpighiaceae (MIX); (2) realizar um estudo quimiossistemático baseado em análises de UHPLC-MS/MS de 137 espécies de Malpighiaceae baseado em metabolômica não direcionada, buscas espectrais, e em ferramentas de classificação *in silico* recentemente descritas; (3) buscar por compostos bioativos com atividade larvicida contra o *Aedes aegypti* usando uma abordagem metabolômica. Os resultados obtidos destacam o poder do uso da abordagem *Quality by Design* para desenvolver um método cromatográfico robusto de amostras muito complexas, além do uso da ferramenta *Molecular Networking*, que levou à anotação bem-sucedida de 61 compostos nos nove extratos usados para preparar o MIX. Além disso, as ferramentas bioinformáticas utilizadas nas investigações quimiotaxonômicas revelaram diversas classes químicas específicas de determinados clados ou gêneros, e foram usados para reconstruir a filogenia de Malpighiaceae. Esta análise forneceu classes de metabólitos identificados como homoplásticos (ou seja, não exclusivos) ou sinapomórficos (ou seja, exclusivos) para clados e gêneros, e permitiu um estudo químico evolutivo abrangente em Malpighiaceae. Finalmente, para a busca de compostos bioativos, uma triagem inicial revelou a espécie *Heteropterys umbellata* como a mais promissora, que foi selecionada para uma investigação química mais aprofundada. Foi possível determinar as diferenças nos perfis metabólicos de caules e folhas e de coletas realizadas em diferentes estados do bioma Cerrado (Minas Gerais e São Paulo). Um fracionamento bioguiado levou ao isolamento de nitropropanoil glicosídeos, uma classe incomum de metabólitos que apresentaram atividade larvicida de moderada a potente. Em conclusão, este estudo contribuiu para investigar a química de dezenas de espécies de Malpighiaceae que ainda não haviam sido estudadas até então. O estudo permitiu a análise das classes químicas em Malpighiaceae em um contexto quimioevolutivo, além de levar ao isolamento de compostos naturais incomuns por uma abordagem de base metabolômica. Esses resultados contribuíram para ampliar o conhecimento químico sobre Malpighiaceae, e os fluxos de trabalho seguidos podem ser reproduzidos em outros táxons para futuras investigações químicas.

Palavras-chave: *Quality by Design*; quimiotaxonomia; metabolômica de plantas; desreplicação; espectrometria de massas; evolução; nitrocompostos.

ABSTRACT

Historically, plants are among the most investigated sources of natural products, which substantially contributed to the development of the natural products field. Plant metabolites represent a rich structural diversity, with countless biological activities that have inspired the development of several drugs. Brazil is considered a megadiverse hotspot, containing thousands of species without any chemical information reported to date, representing a vast potential for chemical studies. The continuous development of natural product bioinformatics tools and databases can be of great value to assist and accelerate studies in the natural products field in Brazil. In this context, the Malpighiaceae botanical family stands out for presenting a wide diversity of natural products reported to date, but still containing hundreds of chemically underexplored species. This research aims to study plant species from the Malpighiaceae family, with three main objectives: (1) develop an optimized chromatographic method (UHPLC-PDA) for the analysis of a mixture of extracts from Malpighiaceae species (MIX); (2) perform a chemosystematic investigation based on UHPLC-MS/MS analysis of 137 Malpighiaceae species based on MS/MS untargeted metabolomics, spectral searches, and recently described *in silico* classification tools; (3) search for bioactive compounds with larvicidal activity against *Aedes aegypti* using an untargeted metabolomics approach. The results obtained highlight the power of using the Quality by Design approach to develop a robust chromatographic method of very complex samples, as well as of the Molecular Networking tool, which led to the successful annotation of 61 compounds in the nine extracts used to prepare the MIX. In addition, the bioinformatics tools used in the chemotaxonomic investigations revealed several clade or genera-specific chemical classes, which were successfully used to reconstruct the Malpighiaceae phylogeny. This analysis provided classes of metabolites identified as homoplastic (*i.e.*, non-exclusive) or synapomorphic (*i.e.*, exclusive) for clades and genera, and allowed a comprehensive evolutionary chemical study in Malpighiaceae. Finally, for the search of bioactive compounds, an initial screening revealed *Heteropterys umbellata* as the most promising species, which was selected for a deeper chemical investigation. It was possible to determine the differences in the metabolic profiles of stems and leaves, and of collections performed in different states in the Cerrado biome (Minas Gerais and São Paulo). A bio-guided fractionation led to the isolation of nitropropanoyl glucosides, an unusual class of metabolites that showed moderate to potent larvicidal activity. In conclusion, this study contributed to investigate the chemistry of dozens of Malpighiaceae species which had not been investigated to date. It allowed the inspection of the chemical classes in Malpighiaceae in a chemoevolutionary context, besides leading to the isolation of unusual natural compounds by a metabolomic-based approach. These results contributed to expand the chemical knowledge of Malpighiaceae, and the workflows followed can be reproduced in other taxa for further chemical investigations.

Keywords: Quality by Design; chemotaxonomy; plant metabolomics; dereplication; mass spectrometry; evolution; nitro compounds.

RESUMO EXPANDIDO

Os produtos naturais oriundos de plantas, microorganismos e organismos marinhos têm fascinado e inspirado a humanidade desde a antiguidade, principalmente devido aos efeitos terapêuticos observados para os mais diversos tipos estruturais. Estima-se que aproximadamente 50% dos fármacos comercializados são produtos naturais e seus derivados, ou inspirados em produtos naturais (NEWMAN; CRAGG, 2020). De fato, a busca por moléculas inéditas de origem natural foi foco de inúmeros grupos de pesquisa nas últimas décadas, e estima-se que aproximadamente 300.000 metabólitos secundários foram identificados até os dias de hoje, sendo que muitas destas substâncias foram avaliadas para somente um, ou mesmo nenhum alvo biológico (HUBERT; NUZILLARD; RENAULT, 2017). Apesar de centenas de milhares de substâncias naturais terem sido descritas, estima-se que somente uma pequena porcentagem da biodiversidade mundial tenha sido explorada quimicamente. Dessa forma, a área da química de produtos naturais ainda tem muito a ser explorada, principalmente em países considerados *hotspots* de biodiversidade, como o Brasil.

Nas últimas décadas, o desenvolvimento de técnicas hifenadas sensíveis tornou possível estudos mais abrangentes de extratos. Nesse contexto, as técnicas hifenadas ganham destaque, principalmente a cromatografia gasosa e cromatografia líquida acopladas a espectrômetro de massas (CG-EM e CL-EM). A possibilidade de se acoplar um detector extremamente sensível, como o espectrômetro de massas, a uma técnica de separação cromatográfica prévia possibilita explorar amostras complexas (como extratos) de forma mais global. Além disso, o desenvolvimento de bases de dados e ferramentas computacionais de processamento de dados e de anotação, causaram uma revolução na área de produtos naturais.

A fitoquímica clássica com foco de isolar substâncias bioativas foi muito utilizada nas últimas décadas, mas muitas vezes levava ao re-isolamento de substâncias já bastante exploradas. Com o uso das técnicas descritas acima, é possível determinar as substâncias conhecidas a partir de comparação com bases de dados espectrais e focar o estudo em substâncias potencialmente inéditas, ou mesmo em metabólitos conhecidos de interesse. Essa forma mais racional de busca por substâncias naturais tem sido empregada com sucesso na química de produtos naturais, e ferramentas computacionais ainda mais robustas vem sendo desenvolvidas para auxiliar no

processo de interpretação de grandes conjuntos de dados. Esses avanços permitem a realização de trabalhos cada vez mais abrangentes, com quantidades de amostras cada vez maiores.

A família botânica Malpighiaceae se destaca pela diversidade de produtos naturais descritos na literatura. Diversos estudos em Malpighiaceae dividem essa família em dez grandes grupos filogenéticos, compreendendo 74 gêneros e cerca de 1300 espécies distribuídas nas regiões tropicais e subtropicais do globo (DAVIS; ANDERSON, 2010). Estudos químicos prévios demonstraram que as espécies de Malpighiaceae acumulam uma ampla gama de metabólitos secundários, incluindo alcaloides β-carbonílicos e derivados de triptamina, isoprenoides (como fitoecdisteroides), e compostos fenólicos abrangendo diferentes subclasses de flavonoides a derivados de ácido quínico (FRAIGE *et al.*, 2018; MANNOCHIO RUSSO *et al.*, 2020; QUEIROZ *et al.*, 2014; SAMOYLENKO *et al.*, 2010). Além da complexidade química, as espécies de Malpighiaceae também se destacam pelo valor nutricional de seus frutos, usos etnofarmacológicos, e propriedades tóxicas. Apesar de diversos estudos químicos reportados, essa família ainda é considerada quimicamente pouco explorada, com centenas de espécies que ainda não foram estudadas com esse propósito. Dessa forma, um estudo químico abrangente em Malpighiaceae é de grande interesse para ampliar o conhecimento químico da família, e correlações quimiotaxonômicas (i.e., classes de substâncias específicas ou mais abundantes em determinados clados filogenéticos ou gêneros) podem auxiliar a guiar futuros estudos nessa família.

A produção de compostos tóxicos por espécies de Malpighiaceae despertam interesse na busca de substâncias com as mais diversas bioatividades. Um estudo prévio realizado com extratos de *Hiptage benghalensis* revelou que essa espécie apresentava importantes atividades inseticidas e repelentes contra mosquitos transmissores de doenças negligenciadas em países do sudeste asiático (LALROTLUANGA *et al.*, 2012). Dessa forma, essa família botânica pode ser promissora para a busca de substâncias tóxicas para o combate de mosquitos transmissores de doenças negligenciadas, como o *Aedes aegypti*. Esse mosquito é o vetor de diversas doenças que se espalham em território nacional todos os anos, como a dengue, zika e chikungunya. Vacinas e tratamento efetivos para essas doenças ainda não foram desenvolvidos, e a forma mais efetiva de combate a essas doenças é o controlar o vetor (SILVÉRIO *et al.*, 2020).

Dessa forma, o presente trabalho tem como principais objetivos realizar uma ampla investigação química de 137 espécies de Malpighiaceae, visando utilizar ferramentas computacionais recentemente descritas para encontrar correlações quimiossistemáticas e buscar compostos bioativos. Nesse contexto, dividimos esta tese em três capítulos para melhor compreensão de seu conteúdo:

1. Desenvolver um método cromatográfico UHPLC de uma mistura de extratos de Malpighiaceae usando a abordagem Quality by Design e automação instrumental, e anotar os metabólitos secundários com base no algoritmo de rede molecular e comparação com bibliotecas espectrais MS/MS.
2. Realizar uma investigação quimiotaxonômica usando metabolômica não direcionada, redes moleculares, buscas em bibliotecas espectrais e ferramentas *in silico* para classificação de metabólitos para correlacionar os resultados químicos com a filogenia atualmente aceita para a família.
3. Buscar por compostos com atividade larvicida contra o *Aedes aegypti* através de uma triagem inicial de extratos de Malpighiaceae, seguida de uma investigação química mais aprofundada da espécie mais ativa por meio de uma abordagem metabolômica.

O primeiro capítulo dessa tese descreve uma metodologia analítica para desenvolver um método cromatográfico robusto para análise de espécies de Malpighiaceae, além da anotação dos metabólitos secundários produzidos por estas. Para tal, uma mistura (MIX) em partes iguais de nove extratos hidroetanólicos de Malpighiaceae foi preparada, compreendendo uma espécie de cada grupo filogenético da família, sendo estas: *Byrsonima intermedia*, *Mcvaughia bahiana*, *Barnebya harleyi*, *Ptilochaeta densiflora*, *Bunchosia pallescens*, *Hiraea restingae*, *Niedenzuella multiglandulosa*, *Banisteriopsis laevifolia* e *Amorimia septentrionalis*. Os extratos de cada uma dessas espécies foram obtidos por micro-extração em ultrassom. O MIX foi, portanto, utilizado para desenvolvimento do método cromatográfico utilizando a abordagem de Quality by Design (QbD). Para tal, contamos com o auxílio do software Fusion Quality by Design® e de automação instrumental para podermos avaliar diversos parâmetros em um curto tempo.

O desenvolvimento do método em *Ultra-High Performance Liquid Chromatography- photodiode array* (UHPLC-PDA) contou com três principais etapas: triagem, otimização e simulação de robustez. Além disso, as respostas selecionadas

para análise estatística foram: número de picos totais, número de pares de picos com resolução ≥ 1.5, número de pares de picos com resolução ≥ 2.0, e número de picos com encaudamento ≤ 1.2. Na primeira etapa foram avaliados seis colunas, dois solventes orgânicos, três valores de pH (ácido, neutro e básico) e dois tempos de gradiente. Na segunda etapa, a temperatura, pH (três valores de acidez), % CH_3CN–MeOH em fase orgânica e % final de solvente orgânico foram variados. A etapa final consistiu no simulador de robustez, que proporcionou as melhores condições: coluna Acquity BEH Shield RP18 (2.1 x 100 mm, 1.7 µm), CH_3CN–MeOH (80/20):H_2O, pH 2.22, 275 nm, 45 °C, 5-55% B em 30 min. Devido à automação computacional e instrumental, foi possível chegar a esses resultados em apenas quatro dias.

Esse trabalho também incluiu a validação do método desenvolvido através do uso dos padrões (+)-catequina, (−)-epicatequina, rutina, e ecdisterona, baseado nas diretrizes do ICH (*International Conference on Harmonization*). O método atingiu todos os critérios estabelecidos, provando ser extremamente sensível até mesmo para diversas classes de substâncias. Além disso, a análise dos nove extratos individualmente também mostrou uma separação satisfatória dos picos de cada extrato, indicando que o uso de uma mistura de extratos para o desenvolvimento do método foi uma abordagem eficaz.

Finalmente, esses extratos individuais foram analisados por UHPLC-MS/MS em um método não direcionado. O uso da plataforma *Global Natural Product Social Molecular Networking* (GNPS) possibilitou a obtenção de redes moleculares e comparação com bibliotecas espectrais públicas. Essa abordagem levou à anotação de 61 compostos, incluindo flavonoides *O*-glicosilados, flavonoides *C*-glicosilados, derivados de ácido quínico, esteróis, e outros compostos fenólicos.

Dessa forma, os resultados desse capítulo mostraram a eficácia de se empregar uma mistura de extratos, do uso de automação e design experimental na obtenção de métodos cromatográficos sensíveis, e da utilização da plataforma GNPS como uma ferramenta importante para anotação de substâncias conhecidas em extratos. Esse trabalho pode servir de inspiração para acelerar o fluxo de trabalho em pesquisas na área de produtos naturais, uma vez que tempo e recursos são poupados. Além disso, foi possível ter as primeiras informações químicas acerca de três gêneros de Malpighiaceae até então ainda não explorados quimicamente: *Mcvaughia*, *Barnebya* e *Ptilochaeta*.

O segundo capítulo da tese teve como principal foco o estudo quimiossistemático da família Malpighiaceae. Para tal, uma amostragem compreendendo 39 gêneros e 137 espécies foi empregada, compreendendo todos os dez grandes clados filogenéticos da família. Extratos hidroetanólicos (80%) e acetato de etila foram preparados a partir de 197 amostras (em sua grande maioria, folhas) por micro extração em ultrassom. Essas amostras foram, portanto, analisadas por UHPLC-MS/MS em um método não directionado, tanto no modo positivo quanto negativo de ionização. Dessa forma, um total de 788 corridas foram realizadas. Os dados foram processados no software MZmine2 para análises subsequentes.

Inicialmente, uma avaliação do impacto do solvente extrator e modo de ionização foi realizada a partir de Análise de Coordenadas Principais (PCoA). A partir desses resultados, foi possível concluir que tanto o solvente extrator quanto o modo de ionização impactam significativamente o perfil químico dos extratos. Além disso, correlacionando os grupos gerados nessas análises multivariadas com a filogenia atualmente aceita para a família, foi possível verificar que, diferentemente dos extratos hidroetanólicos (80%), os grupos observados para os extratos acetato de etila não mostraram correlação com a filogenia. Dessa forma, foi possível inferir que os protocolos de extração podem impactar significativamente em estudos quimiossistemáticos.

A partir dos extratos hidroetanólicos (80%) adquiridos tanto em modo positivo quanto negativo, foram geradas redes moleculares e os espectros MS/MS foram comparados com as bibliotecas espectrais públicas da plataforma GNPS. Diversas famílias mostraram-se específicas de determinados clados filogenéticos. Dentre as principais famílias moleculares consideradas específicas de clados, destacam-se: flavonoides di-*C*-glicosilados (clado I), ácido quínico ligados a substituintes galloil (clados A e G), ácido quínico ligados a substituintes fenilpropanoil (clado H), alcaloides benzilisoquinolínicos (clado J), ecdisteroides (clado H) e iridoides (clado J).

De modo a expandir o conhecimento químico da família, e ter informações adicionais sobre as classes de compostos detectadas, utilizamos o *plug in* Qemistree combinado à ferramenta CANOPUS de classificação *in silico*. Dessa forma, foi possível realizar uma análise de hierarquia química e obter classificação *in silico* a nível de superclasse, classe e subclasse. Os resultados obtidos em modo positivo de ionização foram consideravelmente diferentes e complementares ao modo negativo

de ionização, principalmente pois as bases espectrais são mais populadas com dados em modo positivo.

Os resultados das anotações *in silico* foram utilizados para realizar análises de reconstrução de caráter ancestral, com o intuito de mapear a história evolutiva dessas classes usando os critérios de máxima verossimilhança na filogenia molecular de Malpighiaceae. A partir dessa análise, os dados das classificações obtidas em modo positivo e negativo foram combinados, e as classificações foram recuperadas como homoplasias (não-exclusivas) e sinapomorfias (exclusivas) dos principais clados filogenéticos da família.

Dessa forma, os resultados desse capítulo mostram uma nova abordagem para estudos quimiossistemáticos abrangentes utilizando a família Malpighiaceae como modelo. A partir de pouca quantidade de material vegetal e de análises extremamente sensíveis, foi possível obter uma grande quantidade de informações químicas a partir de comparação com bibliotecas espectrais públicas e uso de ferramentas *in silico* de classificação. A comparação dessas classes com a filogenia possibilitou a interpretação desses resultados em um contexto evolutivo da família, sendo extremamente relevante para estudos quimiotaxonômicos. Esses resultados devem ser confirmados em futuros estudos mais aprofundados em cada espécie ou gênero, e para tal, as ferramentas clássicas são de grande valia. A partir desse trabalho, dezenas de espécies de Malpighiaceae sem qualquer estudo químico prévio foram exploradas pela primeira vez, iluminando o metabolismo desse táxon ainda pouco estudado. O fluxo de trabalho seguido nesse capítulo utilizou ferramentas e bases de dados publicamente disponíveis, e pode ser seguido em estudos em outros táxons pouco explorados.

O terceiro capítulo dessa tese consistiu na busca de substâncias naturais com atividade larvicida contra o *Aedes aegypti* em Malpighiaceae, utilizando uma abordagem metabolômica. A mesma coleção de amostras utilizada no segundo capítulo foi empregada nesse trabalho para se fazer uma triagem inicial. Dessa forma, extratos hidroetanólicos (80%) e acetato de etila obtidos por micro-extração em ultrassom foram avaliados em ensaios larvicidas contra *Ae. aegypti*. Essa triagem inicial mostrou que as espécies *Tetrapterys phlomoides* e *Verrucularina piresii* apresentaram 70% e 65% de mortalidade após 72 h de incubação, enquanto a *Heteropterys umbellata* causou 100% de mortalidade após apenas 24 h. Nenhuma dessas espécies apresenta relatos de estudos químicos e biológicos aprofundados

até o momento. Dessa forma, maiores quantidades de materiais vegetais de *H. umbellata* foram coletadas para dois órgãos vegetais diferentes (folhas e caule) em dois estados diferentes (São Paulo e Minas Gerais), totalizando 14 amostras para investigação metabolômica.

As amostras coletadas foram submetidas a extração por maceração em solução hidroalcoólica (80%), e os extratos obtidos foram analisados por HPLC-MS/MS em um método não direcionado, em triplicata. Os dados foram processados no software MZmine2 e submetidos a análises nas plataformas MetaboAnalyst (para análise multivariada) e GNPS (para obtenção das redes moleculares e comparação com as bibliotecas espectrais). Na plataforma MetaboAnalyst, os dados foram submetidos a análise não supervisionada (*Principal Component Analysis*, PCA) e supervisionada (*Partial Least Squares Discriminant Analysis*, PLS-DA). Essas análises revelaram que os perfis químicos obtidos dos extratos das folhas são significativamente diferentes dos extratos do caule. Além disso, o estado de coleta também tem um impacto significativo no perfil metabólico quando somente os extratos de folhas ou de caules são considerados. Através da análise de PLS-DA também foi possível recuperar uma lista das substâncias que impactam mais significativamente da diferenciação dos grupos descritos. Dessa forma, foi possível direcionar a anotação dessas substâncias de maior interesse.

Os resultados das redes moleculares e busca nas bibliotecas espectrais publicamente disponíveis no GNPS mostraram três principais famílias moleculares com correspondências: flavonoides *O*-glicosilados, flavonoides *C*-glicosilados, e derivados de derivados de ácido cafeoilquínico. Essas duas últimas classes compreendiam substâncias relevantes recuperadas nas análises de PLS-DA para a diferenciação de órgãos da planta e estado de coleta. Além disso, outras duas substâncias abundantes não mostram correspondência alguma com as bibliotecas, sendo recuperadas como importantes na diferenciação dos pontos de coleta – tanto para folhas, quanto para caule. Essas substâncias poderiam ser substâncias potencialmente inéditas (ou incomuns), e o fato de serem menos polares também chamou atenção devido ao potencial da atividade buscada.

Portanto, esforços foram direcionados para o isolamento dessas substâncias, que foram caracterizadas por Espectrometria de Massas de Alta Resolução, Ressonância Magnética Nuclear, Espectrometria de Massas em Tandem, sendo caracterizadas como 1,2,6-tris-*O*-[3-nitropropanoyl]-beta-D-glucopyranose (karakin, **2**) e 1,2,3,6-

tetrakis-*O*-[3-nitropropanoyl]-beta-D-glucopyranose (**3**). A substância 3,5-di-*O*-caffeoylquinic acid (isochlorogenic acid A, **1**) também foi isolada por se tratar de uma substância recuperada como relevante em todas as análises de PLS-DA realizadas. Os nitrocompostos isolados são considerados incomuns na natureza, com poucos relatos em poucas famílias botânicas e já relatados para outras espécies de *Heteropterys*.

Essas substâncias isoladas foram subsequentemente utilizadas para realizar a validação do método de análise, utilizando um método direcionado para recuperar as principais transições MS/MS no modo *Multiple Reaction Monitoring* (MRM). O método se mostrou sensível para os compostos em questão, e consiste no primeiro método validado que permite a quantificação desses ésteres intactos. Essas três substâncias foram quantificadas nos 14 extratos inicialmente obtidos, e foi possível confirmar os resultados obtidos para as análises de PLS-DA, além de levantar hipóteses de possíveis funções ecológicas. Os compostos nitrogenados mostraram moderada atividade larvicida quando testados puros, mas a atividade foi potencializada quando foram testados em fração contendo outros isômeros com mesmo padrão de fragmentação, sugerindo um possível efeito sinergístico na atividade.

Os resultados do terceiro capítulo da tese demonstram que a família Malpighiaceae representa um táxon promissor para a busca de compostos larvicida para o combate de vetores de arboviroses. Através do uso de abordagens metabolômicas, foi possível realizar um estudo aprofundado da espécie mais ativa selecionada da triagem inicial. Essas ferramentas levaram à seleção de compostos nitrogenados incomuns na natureza e que ainda não haviam sido avaliadas para a bioatividade em questão. Esperamos que esse estudo possa guiar futuras pesquisas com o gênero *Heteropterys*, além de inspirar outras investigações na área da química de produtos naturais utilizando ferramentas metabolômicas.

LIST OF FIGURES

Figure 1 – Illustrative scheme of the Molecular Networking approach: a mixture of compounds is analyzed by LC-MS/MS and then organized into molecular families through computational methods. 46

Figure 2 – (A) Molecular networks obtained for the 292 extracts of Euphorbiaceae species, in which a different color was assigned to each genus; (B) Networks characterized by only one color, indicating specific chemical structures of certain genera; (C) Selection of chlorinated substances of interest and targeted isolation of these substances in *C. peltatum*. Node size is proportional to the MS^1 peak area of the crude extract.....47

Figure 3 – Major phylogenetic groups (clades) currently accepted for the Malpighiaceae botanical Family. 53

Figure 4 – General overview of the step-by-step in which the software communication is based on. The QbD software created and exported the methods needed for the method development directly to the UHPLC software. The chromatograms are processed, and the selected responses are imported back to the QbD software, which provides the statistical analyses for each of the responses selected (such as the Analysis of Variance, mathematical equations, and surface response).....64

Figure 5 – Workflow followed during the chromatographic method optimization using the QbD approach. (A) Three main steps were followed in the chromatographic method development: screening, optimization, and robustness simulator. The designs, variables chosen, and fixed parameters are described. (B) Chromatograms with the higher overall responses obtained for the MIX sample for each step are shown (275 nm), such as the conditions employed. 72

Figure 6 – Response surface plots for the responses number of total peaks, number of peaks with resolution ≥ 1.5, number of peaks with resolution ≥ 2.0, and number of peaks with tailing ≤ 1.2 for the screening step of the chromatographic method development.....73

Figure 7 – Response surface plots for the responses number of total peaks, number of peaks with resolution ≥ 1.5, number of peaks with resolution ≥ 2.0, and number of peaks with tailing ≤ 1.2 for the first optimization step of the chromatographic method development. Constant variables: final percentage of organic solvent (60 ˚C) and CH_3CN to MeOH ratio (89:11). 75

Figure 8 –Response surface plots for the responses number of total peaks, number of peaks with resolution ≥ 1.5, number of peaks with resolution ≥ 2.0, and number of peaks with tailing ≤ 1.2 for the first optimization step of the chromatographic method development. Constant variables: oven temperature (45 ˚C) and pH (2.26). 76

Figure 9 – Response surface plots for the responses number of total peaks, number of peaks with resolution ≥ 1.5, number of peaks with resolution ≥ 2.0, and number of peaks with tailing ≤ 1.2 for the second optimization step of the chromatographic method development.....77

Figure 10 – Chromatograms (275 nm) obtained by UHPLC-PDA from the MIX and the individual extracts from Malpighiaceae plant species. Sodium diclofenac (*) was used as the internal standard. 80

Figure 11 – Distribution of the standards used in method validation among the species used in the MIX preparation. 83

Figure 12 – Molecular families annotated using the classical molecular networking workflow: *O*-glycosylated flavonoids, *C*-glycosylated flavonoids, quinic/shikimic acid derivatives, sterols, and other phenols. These are level 2 matches according to the 2007 metabolomics standards initiative (SUMNER et al., 2007). Structures shown with a defined configuration were confirmed with standards isolated from *N. multiglandulosa*. 84

Figure 13 – Experimental workflow followed for the metabolomics and chemosystematics analyses of Malpighiaceae samples. (1) The samples were initially collected, (2) the extracts were prepared with different solvents [EtOH–H_2O (4:1, v/v) or EtOAc], and then (3) subjected to LC-ESI-MS/MS analysis in positive and negative ionization modes in an untargeted method. (4) The data acquired were processed for feature finding, and the exported data were used for multivariate analysis. The clustering groups observed were merged to the phylogeny using the Maximum Likelihood Estimation (MLE) for preliminary chemotaxonomic investigations. (5) The data was also used for Feature-Based Molecular Networking and library searches workflows to observe clade-specific molecular families. (6) A chemical hierarchy analysis and in silico classifications were obtained and finally (7) merged to the currently accepted Malpighiaceae phylogeny to determine the ubiquitous and the taxa-specific *in silico* classes. 104

Figure 14 – (A) Distribution map showing the collection sites of all samples within the American and African continents. A complete record of all collection sites (numbers on black circles) is listed in Supplementary Box 1. Photograph on the left representing a New World tropic species of Malpighiaceae (*Camarea ericoides* by R.F. Almeida). Photograph on the right representing an Old World tropic species of Malpighiaceae (*Acridocarpus excelsus* by T. Randrianarivony). (B) Ten major phylogenetic clades currently accepted in Malpighiaceae, based on plastid and nuclear genes, according to Davis and Anderson (2010). Major clades are shaded in different colors. Species = number of species sampled by each clade in our study. Clade A = Byrsonimoid clade; B = Acridocarpoid clade; C = Mcvaughioid clade; D = Barnebyoid clade; E = Ptilochaetoid clade; F = Bunchosioid clade; G = Hiraeoid clade; H = Tetrapteroid clade; I = Malpighioid clade; and J = Stigmaphylloid clade. 112

Figure 15 – Diversity of metabolic profiles obtained in different extraction protocols and ionization modes. (A) Venn diagrams for the different extraction protocols in positive and negative ionization modes. (B) Three-dimensional PCoA plots of the samples analyzed in different ionization modes (positive: left; negative: right) determined by Bray–Curtis distance metric. The percentage of variance explained by the principal coordinates is presented on each axis. 113

Figure 16 – Clustering trends observed in different ionization modes for the hydroethanolic extracts and correlation with phylogeny. (A) Three-dimensional PCoA plots of the hydroethanolic extracts analyzed in different ionization modes determined

by Canberra distance. The percentage of variance explained by the principal coordinates is presented on each axis. (B) Groups observed in the PCoA plots optimized using maximum likelihood criteria in the most recent molecular phylogeny of Malpighiaceae. Pie charts located on the tip of branches represent presence/absence of metabolomic profiles from groups A and B. Pie charts located on branch nodes of the tree represent the statistical results (presented in %) from the optimization analyzes. Colors represent all ten major phylogenetic clades (i.e., natural groups) currently recognized in Malpighiaceae: light yellow- Byrsonimoid clade (A); dark blue- Acridocarpoid clade (B); pink- Mcvaughioid clade (C); grey- Barnebyoid clade (D); light blue- Ptilochaetoid clade (E); dark green- Bunchosioid clade (F); purple- Hiraeoid clade (G); light green- Tetrapteroid clade (H); yellow- Malpighioid clade (I); and red- Stigmaphylloid clade (J). 115

Figure 17 – Clustering trends observed in different ionization modes for the ethyl acetate extracts and correlation with phylogeny. (A) Three-dimensional PCoA plots of the ethyl acetate extracts analyzed in positive ionization mode determined by Canberra distance. The percentage of variance explained by the principal coordinates is presented on each axis. (B) Groups observed in the PCoA plots optimized using maximum likelihood criteria in the most recent molecular phylogeny of Malpighiaceae. (C) Three-dimensional PCoA plots of the ethyl acetate extracts analyzed in negative ionization mode determined by Canberra, Bray–Curtis and Jaccard distances. The percentages of variance explained by the principal coordinates are presented on each axis. 116

Figure 18 – Molecular family composed of flavonoids containing two *C*-glycosylated portions (positive ionization mode). 118

Figure 19 – Molecular family composed of flavonoids containing one *C*-glycosylated portion (positive ionization mode). 119

Figure 20 – Molecular family composed of *O*-glycosylated flavonoids (positive ionization mode). 120

Figure 21 – Molecular family composed of *O*-glycosylated flavonoids containing phenylpropanoid potions (positive ionization mode). 121

Figure 22 – Molecular family composed of *O*-glycosylated flavonoids containing galloyl potions (positive ionization mode). 121

Figure 23 – Molecular family mainly composed of non-glycosylated flavonoids (positive ionization mode). 122

Figure 24 – Molecular family mainly composed of catechin, afzelechin and their derivatives (positive ionization mode). 123

Figure 25 – Molecular family composed of quinic acid bound to phenylpropanoids substituents (positive ionization mode). 124

Figure 26 – Molecular family composed of quinic acid bound to galloyl substituents (positive ionization mode). 125

Figure 27 – Molecular family composed of fatty acids and fatty esters (positive ionization mode) 125

Figure 28 – Molecular families composed of glycerophospholipids (positive ionization mode). 126

Figure 29 – Molecular family composed of jasmonic acid derivatives (positive ionization mode). 127

Figure 30 – Molecular families composed of beta-carboline alkaloids and other tryptophan derivatives (positive ionization mode). 128

Figure 31 – Molecular family composed of isoquinoline alkaloids (positive ionization mode). 129

Figure 32 – Molecular family composed of protoberberine alkaloids (positive ionization mode). 129

Figure 33 – Molecular family composed of benzylisoquinoline alkaloids (positive ionization mode). 130

Figure 34 – Molecular families composed of amides (positive ionization mode). 131

Figure 35 – Molecular family composed of polyamines (positive ionization mode). 131

Figure 36 – Molecular family composed of triterpenoids and precursors (positive ionization mode). 132

Figure 37 – Molecular family composed of ecdysteroids (positive ionization mode). 133

Figure 38 – Molecular families composed of iridoids (positive ionization mode). 134

Figure 39 – Molecular family composed of secoiridoids (positive ionization mode). 134

Figure 40 – Molecular families composed of neolignans and furofuranoid lignans (positive ionization mode). 135

Figure 41 – Molecular families composed of condensed tannins (negative ionization mode). 136

Figure 42 – Molecular families composed of lignans (negative ionization mode). ... 136

Figure 43 – *In silico* annotations obtained for the Malpighiaceae dataset from the Qemistree workflow combined with the CANOPUS classification tool. These are level 3 annotations according to the 2007 metabolomics standards initiative (SUMNER *et al.*, 2007). (A) Chemical hierarchies of the predicted molecular fingerprints from the Malpighiaceae plant samples analyzed in positive (left) and negative (right) ionization modes. The trees are pruned to keep fingerprints that were classified up to a superclass level in CANOPUS. The branch colors indicate the superclasses, while the bar plots of the outer ring indicate the relative abundance of a molecular fingerprint in each Malpighiaceae clade. (B) The ion features classified *in silico* are mapped based

on the CANOPUS superclass (same colormap described in A). The x and y axes indicate the retention time and *m/z* value, respectively. 138

Figure 44 – Compound classification distribution within Malpighiaceae species in positive and negative ionization modes at a (A) CANOPUS superclass level and (B) CANOPUS class level. 139

Figure 45 – Summary of the maximum likelihood ancestral state reconstruction for the *in silico* classifications obtained at a class level. Each chemical class was treated as a character (0 to 77), and character states were binary-coded for each genera (1: present; 0: absent). Black and red circles represent homoplasies and synapomorphies, respectively. Clades highlighted represent the Malpighiaceae major clades recognized by recent molecular phylogenetic studies (DAVIS; ANDERSON, 2010). 142

Figure 46 – General workflow for the identification of bioactive compounds with larvicidal activity. (A) Initially, a collection of 197 samples from Malpighiaceae species leaves were extracted with solvents of different polarity, and all extracts were subjected to a larvicidal screening against *Aedes aegypti*. (B) *Heteropterys umbellata* was selected for a deeper study for being the most active extract. Both stem and leaves were collected in different locations in São Paulo and Minas Gerais states in Brazil, yielding 14 extracts. These extracts were analyzed through LC-MS/MS for untargeted metabolomics, in which the data generated was processed and subjected to multivariate analysis, molecular networks and library searches. (C) Three features important for group differentiation and/or without library matches were isolated. These compounds were used to validate the analytical method and to quantify these compounds in the different collection points and plant organs. The isolated compounds were once again evaluated for biological activity.....151

Figure 47 – Base Peak Chromatograms (BPCs) obtained from the *Heteropterys umbellata* crude hydroethanolic extracts. BPCs from the leaves samples are shown on the left (green), and from the stem samples are shown on the right (brown). The collection state of each sample is depicted (MG = Minas Gerais state; SP = São Paulo state). Sodium diclofenac (*) was used as the internal standard. 162

Figure 48 – Statistical analyses for differentiating *H. umbellata* plant organs (A–C) and collection sites (D–I). Scores plot obtained for the PCA for plant organs (A) and collection sites (leaves: D; stem: G). Scores plot obtained for the PLS-DA for plant organs (B) and collection sites (leaves: E; stem: H). VIP scores indicating the top 15 metabolites with the highest VIP values (1st component of PLS-DA) for plant organs (C), and collection sites (leaves: F; stem: I).....164

Figure 49 – Molecular families obtained from the Feature-Based Molecular Networking workflow and annotated based on spectral matches within the public spectral libraries in the GNPS platform. Each node represents a tandem mass spectrometry spectrum (MS/MS), while the edges that connect them represent the MS/MS fragmentation similarity (cosine > 0.7). The pie charts indicate the relative abundance of an ion feature in each Heteropterys umbellata plant organ (stem and leaves). Node sizes are relative to the summed peak areas of the precursor ion in MS^1 scans. 166

Figure 50 – Base Peak Chromatograms (BPCs) obtained from the *Heteropterys umbellata* Fr80%, fraction 6 and fraction 9. Fraction 6 contains three isomers of *m/z*

501, while fraction 9 presented two isomers of *m/z* 602 and isomers of *m/z* 501 in low amounts. 169

Figure 51 – Compounds isolated from *H. umbellata* leaves extract. 170

Figure 52 – Proposed fragmentation pathway for compound 1 isolated from *H. umbellata*. 171

Figure 53 – Proposed fragmentation pathway for compound 2 isolated from *H. umbellata*. 171

Figure 54 – Proposed fragmentation pathway for compound 3 isolated from *H. umbellata*. 172

Figure 55 – Boxplots of the concentration levels of compounds 1, 2, and 3 obtained by Multiple Reaction Monitoring (MRM) mode by plant organ (A), and by collection site for leaves (B) and stem (C). Black dots represent all individual sample concentrations for each group. 176

LIST OF TABLES

Table 1 – Description of the columns selected for the screening procedure in the chromatographic method development. 65

Table 2 – Method Validation Items, Parameters, Results, and Acceptance Criteria (according to the ICH guidelines). 82

Table 3 – Metabolites annotated by UHPLC-ESI-MS^2 and Molecular Network in the extracts present used to elaborate the MIX sample – positive ionization mode. 86

Table 4 – Larvicidal results obtained for Fractions 6 and 9, and for the isolated compounds 2 and 3. 173

Table 5 – Method validation data of compounds 1, 2, and 3 for the *H. umbellata* analyses (according to the ICH guidelines). 175

LIST OF BOXES

Box 1 – List of plant species used to prepare the MIX of Malpighiaceae: collection sites, dates and voucher codes, biomes, and phylogenetic groups. 61

Box 2 – Characters retrieved from the ancestral characters reconstruction (clades) based on the classifications obtained *in silico* for Malpighiaceae samples. The ionization mode in which each classification was obtained is described (pos = positive ionization mode; neg = negative ionization mode; both = both ionization modes)…143

LIST OF ACRONYMS

®	Registered trademark
*g*COSY	Correlation Spectroscopy
d	doublet
dd	doublet of doublets
ddd	doublet of doublets of doublets
*g*DEPTQ	Distortionless Enhancement by Polarization Transfer with quaternary carbon detection
ESI	Electrospray Ionization
FBMN	Feature-Based Molecular Networking
GNPS	Global Natural Product Social Molecular Networking
*g*HMBC	Heteronuclear Multiple Bond Coherence
HPLC	High Performance Liquid Chromatography
HPLC-PDA	High Performance Liquid Chromatography - Photodiode Array Detector
HRMS	High-Resolution Mass Spectrometry
*g*HSQC	Heteronuclear Single Quantum Coherence
J	Coupling constant
LC-MS	Liquid Chromatography - Mass spectrometry
m	multiplet
m/z	Mass to charge ratio
MS	Mass Spectrometry
MS/MS	Tandem mass spectrometry
NMR	Nuclear Magnetic Resonance
*g*NOESY	Nuclear Overhauser Effect Spectroscopy
PCA	Principal Component Analysis
PLS-DA	Partial Least-Squares Discriminant Analysis
QToF	Quadrupole time-of-flight
Rt	Retention time
s	singlet
t	triplet
UHPLC	Ultra-High-Performance Liquid Chromatography

VIP score	Variable in Importance in Projection score
δ	Chemical shift
δC	Carbon chemical shift
δH	Hydrogen chemical shift

SUMMARY

1 INTRODUCTION 43

1.1 Metabolomics in natural products investigations 43

1.2 Molecular networking and the GNPS platform 45

1.3 *In silico* tools for metabolites annotation 47

1.4 Malpighiaceae botanical family 49

1.5 Chemosystematics studies of plants 51

1.6 Chemosystematics of Malpighiaceae botanical family 52

1.7 Bioactive natural products as candidates for *Aedes aegypti* mosquito control 53

2 OBJECTIVES 55

CHAPTER 1 – CHROMATOGRAPHIC METHOD DEVELOPMENT USING THE QUALITY BY DESIGN APPROACH AND ANNOTATION OF SECONDARY METABOLITES IN MALPIGHIACEAE 56

1 Introduction 58

2 Materials and Methods 60

2.1 General experimental procedures 60

2.2 Plant material 60

2.2.1 Collections 60

2.2.2 MIX preparation and extraction procedure 61

2.3 Ultra-high performance liquid chromatography – method development 62

2.4 UHPLC-MS/MS analyses 63

2.5 Method development and optimization 63

2.5.1 Software 63

2.5.2 Initial screening 64

2.5.3 Method optimization 65

2.5.4 Robustness simulator 66

2.6 Validation of the analytical method....66
2.6.1 Specificity....67
2.6.2 Precision (repeatability and intermediate precision)....67
2.6.3 Linearity....67
2.6.4 Limits of detection (LOD) and limits of quantification (LOQ)....68
2.6.5 Accuracy....68
2.6.6 Quantification of the standards in the individual extracts....68
2.7 MS/MS data processing and Molecular Networking analysis....69
3 Results and Discussion....70
3.1 Sample preparation....70
3.2 Chromatographic method development....70
3.2.1 Screening....70
3.2.2 Optimization....73
3.2.3 Robustness simulator....78
3.3 Analytical method validation....79
3.4 Annotation of the metabolites....83
4 Conclusions....96
CHAPTER 2 – CHEMOSYSTEMATICS STUDIES OF MALPIGHIACEAE BOTANICAL FAMILY THROUGH THE USE OF SPECTRAL LIBRARY SEARCHES AND *IN SILICO* TOOLS FOR METABOLITES ANNOTATION....99
1 Introduction....101
2 Materials and Methods....105
2.1 General information....105
2.2 Plant material....105
2.3 Extraction procedure....105
2.4 UHPLC-MS/MS analysis....106
2.5 Mass spectrometry data pre-processing....106
2.6 Feature-Based Molecular Networking (FBMN) workflow....107

2.7 Chemical hierarchy analysis using Qemistree 108
2.8 Statistical analysis 109
2.9 Phylogenetic analyses 109
2.10 Ancestral state reconstruction 110
3 Results and Discussion 111
3.1 Evaluation of extraction solvent and ionization mode in Malpighiaceae chemical diversity and chemosystematics 111
3.2 Metabolites annotation 117
3.3 Chemical hierarchy analysis 137
3.4 Ancestral character reconstructions for the classes of secondary metabolites annotated in Malpighiaceae 140
4 Conclusions 145
CHAPTER 3 – SEARCH FOR BIOACTIVE COMPOUNDS IN MALPIGHIACEAE SPECIES WITH TOXIC EFFECTS TO AEDES *AEGYPTI LARVAE* 147
1 Introduction 149
2 Materials and Methods 152
2.1 General experimental procedures 152
2.2 Plant material 152
2.3 Extraction procedures 153
2.4 Bioactivity evaluation 153
2.4.1 Larvicidal bioassay 153
2.4.2 Pupicidal bioassay 154
2.4.3 Adulticide bioassay 154
2.5 HPLC-MS/MS analyses 155
2.6 UHPLC-HRMS analysis 155
2.7 MS/MS data processing 156
2.8 Feature-Based Molecular Networking (FBMN) 156
2.9 MetaboAnalyst platform 157

2.10 Fractionation and isolation of *Heteropterys umbellata* secondary metabolites 158

2.11 Method validation 159

2.11.1 Specificity 159

2.11.2 Precision (repeatability and intermediate precision) 159

2.11.3 Linearity 159

2.11.4 Limits of detection (LOD) and limits of quantification (LOQ) 160

2.11.5 Accuracy 160

3 Results and discussion 160

3.1 Larvicidal screening 160

3.2 Metabolic profiling and multivariate analyses of *Heteropterys umbellata* extracts 161

3.3 Metabolite annotation 165

3.4 Isolation of the bioactive compounds 168

3.5 Method validation and quantification 173

4 Conclusions 177

3 FINAL CONSIDERATIONS: DISCUSSION AND CONCLUSION 179

REFERENCES 181

SUPPLEMENTARY MATERIAL A – CHAPTER 1 199

SUPPLEMENTARY MATERIAL B – CHAPTER 2 214

SUPPLEMENTARY MATERIAL C – CHAPTER 3 252

1 INTRODUCTION

The chemistry of natural products from plants, microorganisms, and marine organisms has fascinated and inspired humanity since ancient times, mainly due to the therapeutic effects reported for the most diverse structural types, being the molecular base of several drugs used in the treatment and prevention of diseases. These compounds of natural sources are biosynthesized through complex pathways, originating a significant number of pharmacophoric groups with high selectivity, being a rich source for the production of drugs, chemical pesticides, fragrances, cosmetics, food supplements, among others. The impressive number of compounds from natural origin identified to date are important sources to identify hits and leads, even for the most challenging targets (HARVEY; EDRADA-EBEL; QUINN, 2015).

For many decades, the search for novel molecules of natural origin has been the focus of numerous research groups on all continents, and it is estimated that approximately 300,000 secondary metabolites have been identified to date, with many of these substances being evaluated for only one, or even none, biological target (HUBERT; NUZILLARD; RENAULT, 2017). Traditionally, the active principles in a bioactive extract are identified using purification strategies, such as liquid-liquid extraction and chromatographic separation methods. This approach demands high effort and investment, and can lead to the isolation of well-known substances, or even without the expected activity (NOTHIAS *et al.*, 2018).

Considering the large number of known natural compounds, new strategies have been employed to save time, cost, and effort spent on research in the chemistry of natural products field. The development of increasingly sensitive hyphenated analytical techniques, such as GC-MS, HPLC-MS, HPLC-NMR, HPLC-NMR-MS, and HPLC-SPE-NMR, enables researchers to perform broader scope chemical investigations using metabolomic approaches.

1.1 Metabolomics in natural products investigations

Metabolomic studies allow a more comprehensive overview of the metabolites (usually in the range of 50 to 1500 Da) produced by a specific organism. Consequently, one can investigate if, and how, groups chemically differ as a response to biotic stimuli,

stress, environmental changes, among others. Basically, these studies can be classified as targeted or untargeted, in which the first one will focus on possible chemical changes in pre-selected compounds, while the second one will look at the whole data aiming to annotate and/or quantify as many compounds as possible (PILON *et al*., 2020).

Gas Chromatography-Mass Spectrometry (GC-MS) and Liquid Chromatography-Mass Spectrometry (LC-MS) are among the most widely used hyphenated techniques in metabolomic studies for their great sensitivity, and for the previous chromatographic separation. In fact, LC-MS has shown great acceptance by the scientific community due to the possibility of detecting a larger number of intact metabolites without the need for chemical derivatization (GOWDA; DJUKOVIC, 2014).

In metabolomic studies, several statistical analyses can be used to investigate a given dataset and how groups can chemically differentiate. For instance, supervised multivariate analyses, such as partial least squares discriminant analysis (PLS-DA), or non-supervised multivariate analyses, such as Principal Component Analysis (PCA), are widely employed. On the other hand, the univariate analyses are also largely used to evaluate the variables individually, and the relationships between them are not considered; in this case, statistical tests such as Analysis of Variance (ANOVA) and Student's t-test are commonly employed (PILON *et al*., 2020).

Besides these chemometric approaches, the annotation of known compounds (also called "dereplication") is of great interest for a biological interpretation of the hypothesis initially raised. In the case of LC-MS, the annotation of these compounds can be performed by comparing the experimental mass spectra with the ones reported in the literature. By identifying the known compounds in a complex mixture, it is possible to focus the research on potentially new compounds, or even isolate known substances for a biological reassessment (GAUDÊNCIO; PEREIRA, 2015; HUBERT; NUZILLARD; RENAULT, 2017).

Even though there has been an evident development of analytical techniques and chemometrics analyses, the annotation of compounds in complex matrices is still considered time-consuming when performed manually. This is particularly critical for large datasets, and nuances can be lost in manual comparisons between analyses.

It is estimated that, considering tandem mass spectra (MS/MS) data obtained from a biological sample, only about 1.8% of the spectra of an untargeted analysis can be determined by comparison with existing data (QUINN *et al*., 2017). With such

approaches, 98% of the spectra obtained in these analyses can be called the “dark matter” of metabolomics (SILVA; DORRESTEIN; QUINN, 2015). It consists of a vast amount of chemical information that, at first, does not help in identifying known compounds.

In recent years, increasingly sophisticated computational methods have been developed to aid and accelerate data processing and annotation. In this context, the molecular networking is an approach that has been successfully used in the natural products field, becoming a powerful and valuable computational strategy for the visualization and interpretation of MS/MS data acquired.

1.2 Molecular networking and the GNPS platform

The molecular networking approach has been widely used in recent studies on the chemistry of natural products. It has proven to be a valuable computational strategy for interpreting mass spectrometry data obtained from complex matrices. The use of molecular networking for the study of natural products led to the development of an open-access platform called GNPS (Global Natural Product Social Molecular Networking), in which it is possible to store, analyze and share information about MS/MS data. In addition, several reference spectral libraries are publicly available on the platform, which is a valuable asset to assist in annotating known compounds in a complex mixture or dataset (QUINN *et al*., 2017; WANG *et al*., 2016; YANG *et al*., 2013). On this platform, MS/MS data can be analyzed and, after similarity calculations of the fragmentation patterns of the substances present in the samples, it is possible to generate a map consisting of molecular families (or networks). These networks are made up of nodes and edges, where a node corresponds to an MS/MS spectrum of a given substance, while the edges connecting these nodes indicate the degree of similarity between the spectra by cosine similarity. In other words, the connection between the nodes represents structurally similar molecules sharing similar MS/MS fragmentation patterns. An illustrative scheme of this approach is shown in Figure 1.

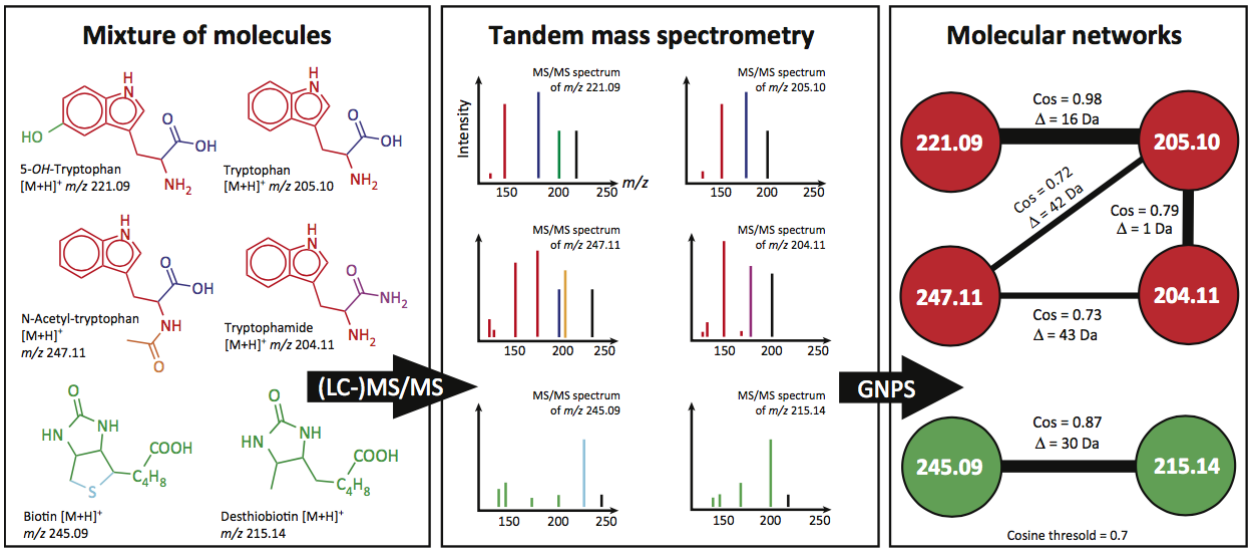


Figure 1 – Illustrative scheme of the Molecular Networking approach: a mixture of compounds is analyzed by LC-MS/MS and then organized into molecular families through computational methods.

Source: Quinn *et al.* (2017).

When a large number of samples is analyzed, the results obtained using the Molecular Networking approach make it possible to find correlations between classes of compounds with the presence or absence of genes that encode them, being particularly useful in studies regarding biosynthetic pathways and their evolution (QUINN *et al.*, 2017). The Molecular Networking approach has been widely used in articles published in literally thousands of high-impact indexed journals in recent years. Particularly in the natural products field, for instance, this tool can be applied to 1) reinvestigate a well-known species in order to identify new natural compounds (FOX RAMOS *et al.*, 2017); 2) study bioactive compounds by combining bioactivity data from natural products and MS/MS spectra (NOTHIAS *et al.*, 2018); 3) study libraries of extracts from a specific taxon (OLIVON *et al.*, 2018).

An article with this later focus consisted in investigating 292 extracts obtained from 107 species of the Euphorbiaceae family collected on New Caledonia. These extracts were analyzed by UHPLC-MS/MS, and the analytical data obtained were subjected to the Molecular Networking approach in the GNPS platform. In this case, the nodes were colored based on the genera each feature was detected genera, and genera-specific networks were obtained (OLIVON *et al.*, 2018), as depicted in Figure 2.

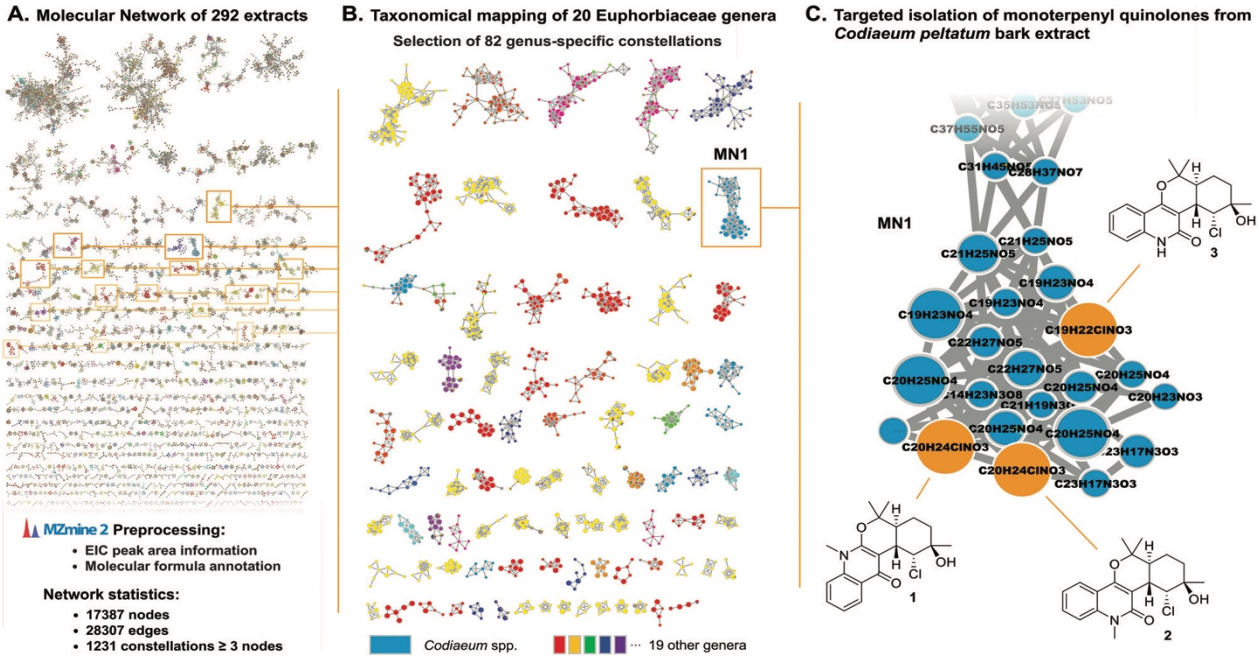


Figure 2 – (**A**) Molecular networks obtained for the 292 extracts of Euphorbiaceae species, in which a different color was assigned to each genus; (**B**) Networks characterized by only one color, indicating specific chemical structures of certain genera; (**C**) Selection of chlorinated substances of interest and targeted isolation of these substances in *C. peltatum*. Node size is proportional to the MS^1 peak area of the crude extract.

Source: Olivon *et al*. (2018).

These results show the importance of state-of-the-art computational tools in the chemistry of natural products from a more comprehensive point of view. It provides valuable support for the metabolomic studies of complex matrices and assists phylogeny studies of poorly studied taxa. In addition, the molecular networking approach combined with searches in spectral libraries also enables the direct search for potential new natural compounds with promising bioactivities. However, despite the evident boost in the annotation of known compounds due to the growing number of spectral databases, this step is still considered a bottleneck in the natural products field. Within this scenario, other computational tools can be used to assist in the annotation process.

1.3 In silico tools for metabolites annotation

From the estimated 2% average annotation rate estimated in 2017 for complex samples (QUINN *et al.*, 2017), recently published articles have described annotation up to more than 10% depending on the sample type (DEAN *et al.*, 2021). Part of this jump is a consequence of the scientific community's effort to populate public spectral libraries with experimental spectra of known compounds. However, there is still much to be chemically explored, and several *in silico* tools have been developed in the last decade to expand the annotation of chemical compounds in MS/MS datasets (BAUERMEISTER *et al*., 2022; MEDEMA, 2021; MISRA, 2021).

These *in silico* tools make use of structural databases and enable users to have more general information regarding the classes of the compounds detected. The information retrieved from such tools has lower confidence than the spectral library searches, and is relative to a level 3 according to the Metabolomics Standards Initiative (MSI) (SUMNER *et al*., 2007). Even though there is a lower confidence level in the*in silico* annotations, they are extremely useful for gathering broad chemical information and expanding chemical knowledge of large datasets with a very low annotation rate retrieved from spectral searches.

There are dozens of these tools reported in the literature making use of sophisticated algorithms to predict the chemical classes through the most diverse methods. For instance, SIRIUS uses a method based on the computation of fragmentation trees to select the molecular formula that best explains the MS data of a particular compound and, along with CSI:FingerID, it predicts a molecular fingerprint encoding substructural properties of the molecule (DÜHRKOP *et al*., 2015, 2019). On the other hand, MetFrag uses combinatorial fragmentation methods to explain MS/MS peaks based on substructures generated by disconnecting the bonds of known chemical compounds (WOLF *et al*., 2010).

In fact, *in silico* classification tools have also been combined with molecular networking, such as the Network Annotation Propagation (NAP). NAP is based on the MetFrag algorithm, and it makes it possible to compare the structural candidates of each node in a network and re-rank a particular structure when a connected node has a related structure in the network (SILVA *et al.*, 2018).

A few studies in the literature make use of specific *in silico* classifications tools in chemotaxonomy investigations, and it has proven to be a promising approach to explore large plant datasets in an evolutionary context (KANG *et al*., 2019). In this way,

the expanded chemical knowledge retrieved from such tools can be of great value to explore plant chemotaxonomy at a family level.

1.4 Malpighiaceae botanical family

The Malpighiaceae family belongs to the group of angiosperms and presents a great diversity of forms such as trees, shrubs, and vines, found in tropical, subtropical, and savannah regions. This family currently consists of 74 genera and approximately 1300 species (http://powo.science.kew.org/), located in the New and Old World. It is estimated that about 85% of plant species belong to the New World, while the other 15% are found in other tropical and subtropical regions of the globe. Furthermore, approximately 150 species belonging to 17 genera are found exclusively in the Old World (DAVIS *et al.*, 2002; DAVIS; ANDERSON, 2010).

Most Malpighiaceae species occur in South America, north of the Tropic of Capricorn, with Brazil being the South American country with the highest representation of this family (ANDERSON, 2004). Approximately 46 genera and 588 species are found in the entire national territory, predominantly in the Atlantic Forest, Cerrado, and Amazon biomes (ALMEIDA *et al.*, 2021).

The largest genera of this family are *Heteropterys* (151 species), *Byrsonima* (136), *Stigmaphyllon* (117), *Bunchosia* (79), *Hiraea* (74), *Banisteriopsis* (59), *Malpighia* (53), and *Tetrapterys* (50), and 16 genera are made up of only one species (http://powo.science.kew.org/). It should be noted that several taxonomic revisions of the species in this family have been described, and a species previously belonging to a genus can be reclassified to a different (or new) genus. For example, some *Tetrapterys* species (such as *T. multiglandulosa* and *T. acutifolia*) were reclassified to the genus *Niedenzuella* (*N. multiglandulosa* and *N. acutifolia*) (ANDERSON, 2006), while *T. cardiophylla* and *T. microphylla* were reclassified to the *Glicophyllum* genus (*G. cardiophyllum* and *G. microphyllum*) (ALMEIDA; VAN DEN BERG, 2021).

Byrsonima crassifolia and *Malpighia emarginata* are two of the economically important species of Malpighiaceae, whose fruits are popularly known as murici and acerola, respectively, and are widely consumed in Brazil. In addition to the nutritional properties, several species of this family are also ornamental, such as *Stygmaphyllon floribundum*, *Lophanthera lactescens*, and *Malpighia coccigera* (ANDERSON, 2004).

One of the most studied Malpighiaceae species is *Banisteriopsis caapi*, whose vine has been known since pre-Columbian times and used in the preparation of Ayahuasca tea. This tea is consumed to this day, not only in indigenous rituals but also in syncretic religious movements in Brazil, such as the União do Vegetal and Santo Daime. It is prepared by the decoction of the vine of *B. caapi* and the leaves of *Psychotria viridis* (Rubiaceae), and the substances present in this tea affect the central nervous system when ingested due to the toxic components and hallucinogens present (MCKENNA; CALLAWAY; GROB, 1998). Phytochemical studies carried out with *B. caapi* showed that this species accumulates, mainly in its vine, β-carboline alkaloids, which are generally very toxic (SAMOYLENKO *et al.*, 2010).

Another species of Malpighiaceae, morphologically very similar to *B. caapi*, is *Tetrapterys mucronata*. Due to the morphological similarity, *T. mucronata* can be erroneously used instead of *B. caapi* to prepare this tea. In studies carried out recently by the NuBBE research group, the presence of β-carboline alkaloids similar to the ones described for *B. caapi* was determined in *T. mucronata*. However, these are even more toxic than the ones present in *B. caapi* and can lead to death when ingested (QUEIROZ *et al.*, 2014, 2015).

In addition to species toxic to humans, outbreaks of acute and chronic cattle poisoning have also been reported in central Brazilian regions caused by species belonging to *Amorimia* (LEE *et al*., 2012) and *Niedenzuella* (mainly *N. multiglandulosa* and *N. acutifolia*) genera (CARVALHO *et al*., 2008; RIET-CORREA; MEDEIROS; SCHILD, 2012). Since these species are considered invasive of pasture fields, these plants can cause considerable economic losses since the Brazilian economy is highly dependent on agribusiness. In this way, livestock producers are committed to eliminating these species from pasture fields, and some of these species (mainly *Amorimia*) are becoming more difficult to find in nature. Therefore, these species should be studied for a better chemotaxonomic understanding of the family and the economic problems caused. With this concern, a recent phytochemical study for *Niedenzuella multiglandulosa* was performed in the NuBBE research group, in which several phytoecdysteroids were identified in the leaves of this species (MANNOCHIO RUSSO *et al.*, 2020).

Several other species of this family are reported to present biological activities, such as antifungal (*Heteropterys byrsonimifolia*), antidepressant, and anxiolytic (*Heteropterys brachiata*) (HUERTA-REYES; FONSECA JUÁREZ; AGUILAR-ROJAS,

2015). However, many species and genera still do not present any chemical study. Therefore, a more in-depth study of this family is of interest for a better chemical understanding of this taxon.

1.5 Chemosystematics studies of plants

Plant systematics can be defined as the science of the diversification of an organism, involving the discovery, description, and interpretation of the biological diversity existing on the planet today, considering its evolutionary history. For this, it is necessary to study all branches of an evolutionary tree in order to document the changes that occurred during the evolution of these branches, and to describe all species (the tips' branches) (JUDD *et al*., 2002; STUESSY, 2008).

There are several manners to carry out the taxonomic classification of living organisms. Particularly for plants, the approach can be carried out based on their most suitable habitat, observing the macromorphology of the species, among others. The macromorphological classification was one of the first to be used to infer evolutionary relationships (phylogeny) (JUDD *et al*., 2002). It is also worth mentioning that evolution depends on a series of internal and external factors such as mutation, recombination of genetic differences, and selection.

From the angiosperm evolution point of view, a generation of morphologically similar plants can occur in the evolutionary process without having a common ancestor. This phenomenon, known in biology studies, is called "convergence" or "parallel development". On the other hand, evolutionarily related plants can be morphologically different (in this case, the phenomenon is called "divergence") (ERDTMAN, 1963). Considering these phenomena, the taxonomic classification may, in some cases, be imprecise. Recent taxonomic studies use molecular tools (a more precise and robust approach) in taxonomic studies regarding the reclassification of many previously misclassified species (ANDERSON, 2006). In fact, molecular approaches in consonance with chemical data can assist taxonomic classification of species and assist phylogeny studies.

The classification of plants based on their secondary metabolites has emerged as another promising possibility. This classification, called chemosystematics (or chemotaxonomy), is valuable to assist studies in taxonomy and ecology. Chemosystematic studies are, in principle, little dependent on classical botanical

studies. In this way, the results obtained can help in the classification of a species and confirm whether the classifications and conclusions obtained previously (based on the morphological characters, for instance) also make sense from the point of view of chemical diversity (ERDTMAN, 1963).

In the last decades, most of the chemosystematics studies were conducted based on vast literature surveys of isolated and characterized compounds. Several important conclusions were withdrawn from such studies, correlating geological times with the origin of classes of metabolites and their oxidation levels (GOTTLIEB, 1989). These studies can provide valuable information about the biosynthetic pathways present in the species, which would make it possible to target enzymatic and biotechnological studies, which are very useful for plants (BHARGAVA; PATEL; DESAI, 2013).

Due to the advances in highly sensitive hyphenated techniques, data processing, and low amounts of plant material necessary for such study, comprehensive chemosystematic studies can be carried out through collaborative research between botanists and chemists. In this way, a more comprehensive understanding of the secondary metabolism at a botanical family level can be performed (STUESSY, 2008).

1.6 Systematics of the Malpighiaceae botanical family

Given the interest of Malpighiaceae species for the Brazilian economy, a more detailed study of this family is of interest, including mapping natural products that can help in the taxonomy and phylogeny of the family. Since the 19th century, Malpighiaceae species have been classified into subfamilies and tribes according to the characteristics of their fruits, classified as dry or fleshy fruits. For centuries this classification was accepted and increasingly detailed for differentiating genera into large groups. However, in 2001, the first studies on the molecular phylogeny of Malpighiaceae were published, and a new classification was proposed for the family (CAMERON *et al*., 2001; DAVIS; ANDERSON; DONOGHUE, 2001), based on the plastid *ndhF* , *matK* , and *rbcL* and nuclear *PHYC* (DAVIS; ANDERSON, 2010).

Currently, this family is divided into ten major phylogenetic groups (ALMEIDA *et al.*, 2017; DAVIS; ANDERSON, 2010), and it can be observed that similar morphologies in the fruits originated independently several times during the diversification of the family. In other words, species with fleshy or dry fruits do not have the same common

ancestor. The classification of the ten major phylogenetic groups, along with the distribution of fruit types in the family, is illustrated in Figure 3.

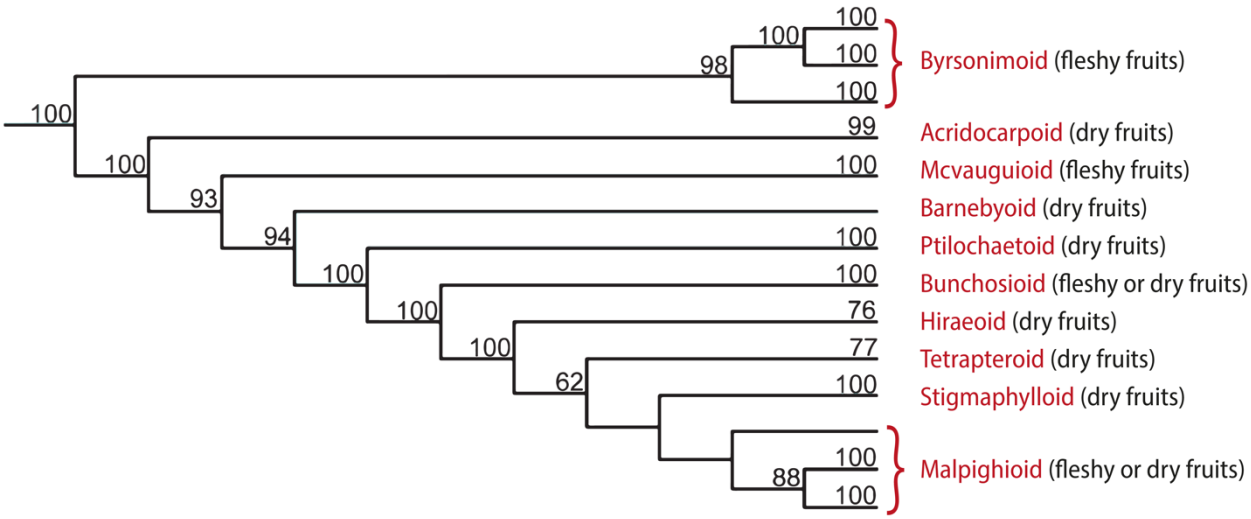


Figure 3 – Major phylogenetic groups (clades) currently accepted for the Malpighiaceae botanical family.

Source: Adapted from Davis and Anderson (2010).

Since these DNA-based studies, several authors have gradually proposed new genera and combinations to accommodate these newly identified genetic lineages, and many species still lack such kind of study. In addition, no morphological characters were ever recovered, circumscribed, or discussed for its major phylogenetic clades (ALMEIDA; VAN DEN BERG, 2021). In this way, chemical studies and metabolomic tools are of great value to assist in establishing a new classification system based on chemical compounds and morphology (ALMEIDA *et al.*, 2017).

1.7 Bioactive natural products as candidates for Aedes aegypti mosquito control

Mosquitoes are vectors of several human diseases (also called arboviruses), such as malaria, dengue, zika, filariasis, and yellow fever, among many others, being more frequent in tropical and subtropical regions of the world (GERIS *et al.*, 2012). The *Aedes aegypti* mosquito (Diptera: Culicidae) is the primary vector of dengue, zika, and chikungunya viruses, which afflicts thousands of Brazilians every year.

The prevention of diseases transmitted by the *Ae. aegypti* mosquito mainly consists of combating the vector since effective vaccines have not been developed to date. Some of the most efficient insecticides developed for this purpose are organophosphates and pyrethroids, which are very efficient synthetic molecules, but mosquitoes have shown increasing resistance (SILVÉRIO *et al*., 2020). In this way, an interest in more natural approaches to combating the mosquito arose, especially in countries considered biodiversity hotspots, such as Brazil.

Very recently, several Brazilian plants have shown promising results to fight *Ae. aegypti* (MAGALHAES *et al*., 2021; MELO *et al*., 2021; SILVA *et al*., 2021, 2022). In addition, A previous study revealed that *Hiptage benghalensis* (Malpighiaceae) extracts presented promising insecticidal and repellent activity against mosquito vectors (specifically, *Anopheles barbirostris*, *Culex quinquefasciatus*, and *Aedes albopictus*) in South East Asian countries (LALROTLUANGA *et al.*, 2012). Since the Malpighiaceae family has known toxic species, such as *Banisteriopsis caapi*, *Tetrapterys mucronata*, and *Niedenzuella multiglandulosa*, evaluating the toxicity of Malpighiaceae extracts towards *Ae. aegypti* can generate promising results for combating dengue and other neglected diseases that spread in the national territory. In this context, the search for bioactive natural compounds can also benefit from the metabolomic tools previously described, being possible to guide the study towards possible new bioactive compounds.

2 OBJECTIVES

The main goal of this work is to know about the chemistry of 137 Malpighiaceae species, by using recently described computational tools to find chemosystematics correlations and bioactive compounds. In this context, we divided this thesis into three chapters seeking to:

1. Develop a UHPLC chromatographic method of a mixture of Malpighiaceae extracts using the Quality by Design approach and instrumental automation, and annotate the secondary metabolites based on molecular networking algorithm and comparison with MS/MS spectral libraries.
2. Perform a chemotaxonomic investigation using untargeted metabolomics, molecular networking, searches in spectral libraries, and *in silico* tools for metabolites' classification to correlate the chemical results with the phylogeny currently accepted for the family.
3. Search for compounds with larvicidal activity against *Aedes aegypti* through an initial screening of Malpighiaceae extracts and deeper chemical investigation of the most active species through a metabolomic approach.

CHAPTER 1 – CHROMATOGRAPHIC METHOD DEVELOPMENT USING THE QUALITY BY DESIGN APPROACH AND ANNOTATION OF SECONDARY METABOLITES IN MALPIGHIACEAE

Article title: Can Statistical Evaluation Tools for Chromatographic Method Development Assist in the Natural Products Workflow? A Case Study on Selected Species of the Plant Family Malpighiaceae (doi: 10.1021/acs.jnatprod.0c00495).

Journal: Journal of Natural Products.

Status: Published.

Authors: Helena Mannochio-Russo[1,2], Paula Carolina P. Bueno[3,4], Anelize Bauermeister[2,5], Rafael Felipe de Almeida[6], Pieter C. Dorrestein[2], Alberto José Cavalheiro[1], and Vanderlan S. Bolzani[1]

Affiliations:

[1]NuBBE, Department of Biochemistry and Organic Chemistry, Institute of Chemistry, São Paulo State University (UNESP), Araraquara, SP, Brazil;

[2]Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, CA, USA;

[3]Faculty of Pharmaceutical Sciences of Ribeirão Preto, Department of Physics and Chemistry, University of São Paulo, Ribeirão Preto, SP, Brazil;

[4]Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany;

[5]Biomedical Sciences Institute, University of São Paulo, São Paulo, SP, Brazil;

[6]Lamol Lab, Feira de Santana State University (UEFS), Department of Biological Sciences, Feira de Santana, BA, Brazil.

Abstract

Malpighiaceae species present important biological activities, such as cytotoxic, hallucinogen, acetylcholinesterase inhibition, and antifungal; however, this family can still be considered chemically underexplored. Considering the increasing need and will of minimizing time and cost in many research projects, including in natural products workflows, the Quality by Design (QbD) approach stands out for reaching these needs. The employment of commercial software that automatically applies the QbD concepts for chromatographic method development of samples such as botanical extracts has been little explored. However, it has a great potential to assist the community of natural products in this crucial step. In this context, this study has investigated the chemical diversity of Malpighiaceae extracts by preparing a representative sample (called MIX) composed of nine species of this family. For this, the QbD principles for UPLC-PDA method development were applied, and their secondary metabolites were annotated by UPLC-PDA-HRMS. The employment of an automated tool for such a purpose allowed this step to be accomplished in only four days, comprising the screening, optimization, and robustness simulation steps. This approach resulted in a sensitive method with low limits of detection and quantification. The analyses of the individual extracts also highlighted the advantages of using a MIX for this analytical process. The molecular networking algorithm and library searches allowed the annotation of 61 compounds, including *O*-glycosylated flavonoids, *C*-glycosylated flavonoids, quinic/shikimic acid derivatives, steroids, and other phenols. These results showed that QbD application for studying botanical extracts could optimize the natural products workflow, besides being useful for future Malpighiaceae studies.

Keywords: Malpighiaceae; Quality by Design; molecular networking; metabolite profiling; method optimization.

1 Introduction

Developing an appropriate chromatographic method can be very challenging and time-consuming, particularly for complex matrices such as plant extracts. The employment of the Quality by Design (QbD) approach for chromatographic method development for natural crude extracts analysis is becoming more popular (BUENO *et al*., 2015; FUNARI *et al*., 2014; SHAO *et al*., 2018), although still underutilized for this purpose. The QbD principles have been extensively applied in analytical method development, including different techniques such as supercritical fluid chromatography (DISPAS *et al*., 2014), capillary electrophoresis (ORLANDINI *et al*., 2015), and even in cell therapy products manufacturing (LIPSITZ; TIMMINS; ZANDSTRA, 2016).

This approach provides innumerous advantages compared to the traditional method development that typically uses the One Factor at Time (OFAT) approach, which can be time-consuming (even taking weeks of work) and costly (TYTECA *et al*., 2016). The QbD uses Design of Experiment (DoE) methods, which provide combinations of parameters, requiring fewer experiments to be performed. Additionally, by using the QbD approach, the user can obtain the designed space of the multidimensional combination of variables where the responses are maximized, resulting in an optimal condition to analyze a certain sample (SCHMIDT; STANIC; MOLNÁR, 2014; SHAO *et al*., 2018).

The analytical method development in liquid chromatographic techniques (UHPLC/HPLC) usually aims to optimize the chromatographic conditions, such as the mobile and stationary phases, pH, oven temperature, and mobile phase gradient. With such optimization in chromatography, the number of total peaks and peaks' resolution can be maximized to get more precise quantitative and qualitative results. The OFAT approach can be enough for the analytical method optimization depending on the sample complexity. However, using the OFAT approach can be very challenging for complex matrices, such as botanical extracts, and the community of natural products often struggles with chromatographic method development.

Optimizing the chromatographic parameters are also useful to enhance the detection in hyphenated techniques, such as liquid chromatography coupled to mass spectrometry (LC-MS). In an MS detector, the coelution of chemical compounds can result in ion suppression or ion enhancement (FACCIN *et al*., 2016), leading to the loss of precise chemical information of a given sample. Thus, ensuring proper separation

of the compounds can also benefit the dereplication of known metabolites in a biological sample, especially since a higher quality of the MS data will be granted, enabling the searches in spectral databases.

Nowadays, several commercial software that automatically applies the QbD concepts on the chromatographic method development is available, such as Nexera Method Scouting Solutions (Shimadzu), ChromSwordAuto and ChromSword Scout (Thermo-Fisher Scientific), AutoChrom (ACS Labs), and Fusion Method Development (Waters) (D'ATRI *et al*., 2019). In fact, the use of commercial software to develop chromatographic methods applied for the analysis of botanical extracts can be very useful, especially to the studies aiming to explore the huge diversity of compounds and isomers present in plant species.

Species from the Malpighiaceae family are excellent examples regarding the high diversity of compounds biosynthesized by plant species. The majority of its plant species are found in the tropical and subtropical areas of the globe (DAVIS *et al*., 2002), being Brazil its most representative country. Among the compounds reported to date, some of them stand out for their biological and nutritional importance: (1) phenolic compounds, reported for the *Byrsonima* and *Malpighia* species (BELWAL *et al*., 2018; NERI-NUMA *et al*., 2018), two genera whose fruits are very consumed in some Brazilian regions; (2) alkaloids, reported for *Banisteriopsis caapi* and *Tetrapterys mucronata* species (QUEIROZ *et al.*, 2014; SAMOYLENKO *et al.*, 2010), species used in the Ayahuasca tea preparation (a hallucinogen tea consumed by indians from the Amazon region); (3) steroids, reported for *Niedenzuella multiglandulosa* (MANNOCHIO RUSSO *et al*., 2020), a species known for causing chronic cattle intoxication in Brazil. Despite the evident pharmacological, nutritional, and economic potentials of Malpighiaceae species, this family can still be considered chemically underexplored.

Therefore, the present study aimed to develop a robust and efficient chromatographic method to analyze complex samples, taking a mixture of extracts of nine Malpighiaceae species as a case study. These species were selected based on previous phylogenetic studies and included clades comprising genera without any chemical report in the literature (*Ptilochaeta*, *Mcvaughia*, and *Barnebya* genera). The method development was performed based on the QbD principles and consisted of three main steps: screening, optimization, and robustness simulation. The method was further validated based on the simultaneous quantification of (+)-catechin, (−)-

epicatechin, rutin, and ecdysterone. The individual extracts were analyzed using this chromatographic method in an untargeted mass spectrometry workflow (UHPLC-ESI-MS/MS). The data obtained was subjected to the Molecular Networking workflow in the GNPS platform, and the occurrence of the annotated metabolites was inspected in the *Dictionary of Natural Products*.

2 Materials and Methods

2.1 General experimental procedures

Acetonitrile (HPLC grade), methanol (HPLC grade), formic acid 88% (v/v) and ammonium hydroxide 28-30% (v/v) were purchased from J. T. Baker (USA). Ethanol 98% from Êxodo Científica was used in the extraction protocol. Integristerone A, podecdysone B, 5-hydroxypodecdysone B, and ecdysterone standards were isolated from *Niedenzuella multiglandulosa* in a previous study performed in our laboratory (MANNOCHIO RUSSO *et al*., 2020). For the chromatographic method validation procedure, (+)-catechin (≥98% purity), (−)-epicatechin (≥98% purity), rutin (≥94% purity), and ecdysterone (77.6% pure) were used, and sodium diclofenac was used as an internal standard. The catechins, rutin, and sodium diclofenac were all acquired from Sigma-Aldrich (Germany).

2.2 Plant material

2.2.1 Collections

Leaves from nine Malpighiaceae plant species [*Byrsonima intermedia* A. Juss., *Mcvaughia bahiana* W. R. Anderson, *Barnebya harleyi* W. R. Anderson & B. Gates, *Ptilochaeta densiflora* Nied., *Bunchosia pallescens* Skottsb., *Hiraea restingae* C. E. Anderson, *Niedenzuella multiglandulosa* (A. Juss.) W. R. Anderson, *Banisteriopsis laevifolia* (A. Juss.) B. Gates and *Amorimia septentrionalis* W. R. Anderson] were collected in different Brazilian regions between 2010 and 2018, and voucher specimens were deposited in herbaria, as shown in Box 1. Field photographs of each of the selected species are shown in Supplementary Figure 1. The authorization for conducting this study was conceded by the National System for Management of Genetic Heritage and Associated Traditional Knowledge (SISGEN), n˚ A6FDC2E. The

leaves were initially dried in a desiccator containing silica at room temperature. Subsequently, samples were frozen in liquid nitrogen, grounded in a ball mill, and stored in a freezer at -20 ˚C until the preparation of the extracts.

Box 1 – List of plant species used to prepare the MIX of Malpighiaceae: collection sites, dates and voucher codes, biomes, and phylogenetic groups.

Species	Collection sites	Collection dates	Codes	Herbarium*	Biome	Phylogenetic group
Byrsonima intermedia	Mogi Guaçu/SP	Jan/2014	IAC 55281	IAC	Cerrado	Byrsonimoid
Mcvaughia bahiana	Monte Santo/BA	Jan/2006	Guedes 12148	ALCB	Atlantic Forest	Mcvaughioid
Barnebya harleyi	Itatim/BA	Oct/2014	Melo 1518	HUEFS	Caatinga	Barnebyoid
Ptilochaeta densiflora	Corumbá/MS	Apr/2010	Carvalho 290	HUEFS	Pantanal	Ptilochaetoid
Bunchosia maritima	Rio de Janeiro/RJ	Sep/2018	I.R.C. 183	RBv	Atlantic Forest	Bunchosioid
Hiraea restingae	Sooretama/ES	Jan/2012	Almeida 518	SP	Atlantic Forest	Hiraeoid
Niedenzuella multiglandulosa	Campo Grande/MS	Nov/2015	HMS 5206	CGMS	Cerrado	Tetrapteroid
Banisteriopsis laevifolia	Rio de Janeiro/RJ	Sep/2018	Mattos 317	RBv	Atlantic Forest	Stigmaphylloid
Amorimia septentrionalis	Maruim/SE	Nov/2015	Almeida 800	HUEFS	Atlantic Forest	Malpighioid

***IAC:** Agronomic Institute of Campinas; **ALCB:** Herbarium Alexandre Leal Costa – Federal University of Bahia; **HUEFS:** Herbarium of State University of Feira de Santana; **RBv:** Arboretum of Rio de Janeiro Botanical Garden (living collection); **SP:** Institute of Botany of São Paulo; **CGMS:** Herbarium of Federal University of Mato Grosso do Sul.

Source: elaborated by the Author (2022).

2.2.2 MIX preparation and extraction procedure

The nine species selected for this study were selected based on previous Malpighiaceae phylogenetic studies, in which one species from each of the major phylogenetic groups was chosen to prepare a sample that could be representative for

the family, named as a MIX. This MIX could, therefore, provide a more general and comprehensive chemical information regarding the secondary metabolites produced by Malpighiaceae species. The species from each group were chosen considering the ones with previous chemical studies reported in literature, and for the groups that had not been studied to date, the species were chosen based only in the species with the higher plant material amount available considering the whole sampling of Malpighiaceae species present in our laboratory.

To prepare the MIX sample, 20 mg of each powdered plant material was transferred to a 1.5 mL Eppendorf tube along with 1 mL EtOH–H_2O (80:20). Samples were sonicated during 15 min, centrifuged for 5 min at 6200 *g*, and the supernatant was dried at Speed Vac (RVT4104 Thermo® equipped with a vacuum pump) to obtain the dried extracts. The MIX sample was prepared by transferring 0.5 mg of each dried extract to a 1.5 mL Eppendorf flask, and solubilized in 1 mL of MeOH–H_2O (1:1). This solution was filtered through a 0.22 µm Nylon Millipore filter and subjected to the chromatographic analyses. For the chromatographic method validation, the same procedure was followed, but the solution used to solubilize the sample contained 15 µg mL^{-1} of sodium diclofenac as internal standard.

The individual nine extracts were also individually solubilized at a concentration of 1 mg mL^{-1}, using MeOH–H_2O (1:1) for the UHPLC-PDA-QToF analyses, and MeOH–H_2O (1:1) containing 15 µg mL^{-1} of sodium diclofenac for the quantification of the standards in each individual samples by UHPLC-PDA analysis.

2.3 Ultra-high performance liquid chromatography – method development

The chromatographic method development was performed using a Waters Acquity UPLC H-Class (Waters, USA) equipped with a quaternary pump (A to D) in which one of the channels subdivides into six (D1–D6), degassing system, autosampler, three individual column manager ovens with capacity for two columns each, and a photodiode array (PDA) detector (200−400 nm). Acetonitrile (A) and methanol (B) were used as organic solvents, while H_2O with different modifiers was evaluated (D1–D3). During the screening step, an acidic (+0.1% formic acid, pH 2.54), a neutral (no modifier, pH 6.60) and an alkaline (+0.1% ammonium hydroxide, pH 10.45) aqueous phase were evaluated. For the optimization procedure, three acidic aqueous phases were evaluated: +0.4% formic acid, pH 2.26; +0.1% formic acid, 2.57; +0.013% formic

acid, pH 3.07. The flow rate and the injection volume were set to 0.35 µL min^{-1} and 2 µL, respectively. All chromatographic data was visualized and processed through Empower 3 software (Waters, USA), and the method development experiments were designed using the Fusion QbD® software (S-Matrix, USA).

2.4 UHPLC-MS/MS analyses

The high-resolution mass spectrometry (HRMS) analyses were performed on a Waters Acquity UPLC H-Class system equipped with a quaternary pump, degassing system, autosampler, column oven, PDA detector, and a Xevo G2-XS QToF mass spectrometer (Waters, USA). The mixture (MIX) and the nine extracts were analyzed, separately. The stationary phase consisted of an Acquity BEH Shield RP18 (ethylene-bridged hybrid C18, 1.7 µm, 100 x 2.1 mm), with oven temperature set to 45 ˚C. The injection volume for the MIX and for the individual extracts were 0.8 and 0.3 µL, respectively. The quaternary pump system consisted of acetonitrile (A), methanol (B), and H_2O (+0.4% formic acid; v/v, D), and a linear gradient was employed from 5% to 55% of organic solvent [constituted of CH_3CN−MeOH (80:20)] in 30 min, and the flow rate was set to 0.35 mL min^{-1}. The system was also equipped with a PDA detector, with detection range set to 200−400 nm. The mass spectra were acquired both in positive and negative mode, separately. Mass range was set to 50−1600 Da using Data-Dependeng Acquisition (DDA) mode. The parameters used for data acquisition were as follows: capillary voltage of 3.0 kV, cone and desolvation gas flow of 50 and 800 L h^{-1}, respectively, source temperature of 100 ˚C, and desolvation temperature of 650 ˚C with sampling cone and source offset at 40 and 80, respectively. A ramp collision energy (CE) was employed for MS/MS fragmentation: low CE from 6 to 9 eV and a high CE of 60 to 80 eV. The data were processed through MassLynx 4.1 software.

2.5 Method development and optimization

2.5.1 Software

The software Fusion Quality by Design® was employed to develop a chromatographic method by using Design of Experiments (DoE) approach, which

consisted in three main steps: screening (also named as “scouting”), optimization and robustness simulator. This software communicates directly with Waters Empower 3, allowing a simpler process. The responses selected for statistical analyses were the number of total peaks, number of peaks with resolution ≥ 1.5, number of peaks with resolution ≥ 2.0 and number of peaks with tailing ≤ 1.2. These responses are imported back to Fusion QbD®, enabling calculation and prediction of the best chromatographic answer for each step. In addition, several statistical responses are generated for each step.

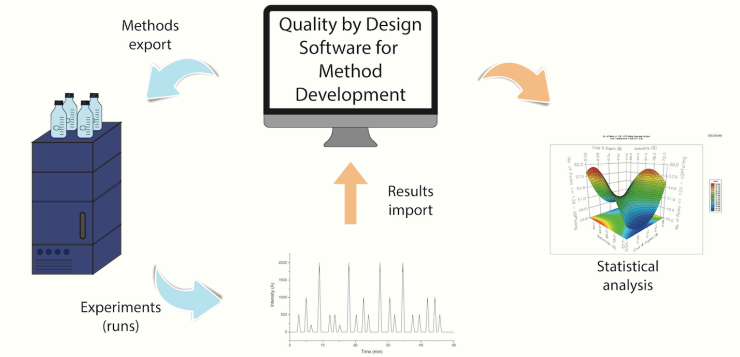


Figure 4 – General overview of the step-by-step in which the software communication is based on. The QbD software created and exported the methods needed for the method development directly to the UHPLC software. The chromatograms are processed, and the selected responses are imported back to the QbD software, which provides the statistical analyses for each of the responses selected (such as the Analysis of Variance, mathematical equations, and surface response).

Source: elaborated by the Author (2022).

2.5.2 Initial screening

For the screening procedure, more abrupt changes were performed in the parameters evaluated. Therefore, six different columns (Table 1), two organic solvents (CH_3CN and MeOH), three pH values (2.54, 6.60, 10.45) and two gradient times (15 and 30 min) were considered for the experimental planning. The statistical quadratic model A- and G- optimal was applied to determine the best conditions for this first step in the chromatographic method development. All the chromatographic conditions of these runs, along with the responses obtained, are shown in Supplementary Table 1.

During the screening, two parameters were set constant: temperature (40 ˚C) and flow rate (0.35 mL min^{-1}).

Table 1 – Description of the columns selected for the screening procedure in the chromatographic method development.

Column	Particle	Bonded phase	pH range	Max temperature (˚C)	Particle size (µm)
Cortecs	Solid Core	C8	2–8	45	1.6
Acquity HSS	High Strength Silica	T3 (C18)	2–8	45	1.8
Acquity BEH	Ethylene Bridged Hybrid	Shield RP 18	2–11	High pH: 45 Low pH: 50	1.7
Acquity HSS	High Strength Silica	PFP (pentafluorophenyl)	2–8	45	1.8
Cortecs	Solid Core	C18+	2–8	45	1.6
Acquity BEH	Ethylene Bridged Hybrid	C18	2–12	High pH: 60 Low pH: 80	1.7

Source: elaborated by the Author (2022).

After the chromatograms were processed and the results evaluated, the conditions in which all the responses are simultaneously maximized were as follows: Acquity BEH Shield RP18 column, gradient time of 30 min, acidic conditions and CH_3CN as organic solvent.

2.5.3 Method optimization

During the optimization step, slighter variations on the parameters were performed, in which the Acquity BEH Shield RP18 column, gradient time of 30 min and flow rate of 0.35 mL min^{-1} were set constant, as a result of the previous screening step. At this point, three acidic pH values (2.26, 2.54 and 3.07), column temperatures (35 and 45 ˚C), final percentage of organic solvent (60% and 80%) and organic mobile phase composition (CH_3CN to MeOH ratio, 70% and 100%) were evaluated. The central composite with face-centered axial point experimental design was used at this point, and the matrix containing the analyses conditions is shown at the Supplementary Table 2. The chromatograms were processed, and the best conditions (maximizing all the

responses) consisted in the lowest pH (2.26), higher column temperature (45 ˚C), 60% of final organic solvent and 89% CH_3CN in the organic phase.

Considering the column limitations and the possibility of still improving the peaks separation, the pH and temperature values were maintained constant at 2.26 and 45 ˚C, respectively, and a second optimization step was performed. At this time, only the final percentage of organic solvent (40% and 70%) and organic mobile phase composition (CH_3CN to MeOH ratio, 60% and 100%) were evaluated. The central composite with face-centered axial point experimental design was once again employed, and the matrix containing the analyses conditions and all the responses obtained is shown at the Supplementary Table 3. Finally, the best conditions after data processing consisted in 60% of final percentage of organic solvent and 64.8% CH_3CN/MeOH ratio.

2.5.4 Robustness simulator

Another functionality included in the Fusion Quality by Design® package software is the robustness simulator step, in which a series of simulations and calculations are performed based on the maximum variations the user set for each chromatographic parameter. At this final step, maximum variation of the final percentage of organic solvent and the CH_3CN percentage in organic solvent was set at 1%, and the maximum expected variations in all the responses was set at 1 (± 1Δ). In this way, the proposed robust method considering these variations would be with 55% of final organic solvent and 80% CH_3CN in the organic solvent. Therefore, the developed method consisted in the following final conditions: Acquity BEH Shield RP18 (2.1x100 mm, 1.7 µm), CH_3CN/MeOH (80/20)–H_2O, 5-55% B gradient in 30 min, pH 2.26, 275 nm, 45 ˚C, injection volume of 2 µL and flow rate of 0.35 mL min^{-1}, 275 nm.

2.6 Validation of the analytical method

The chromatographic validation was performed according to the ICH guidelines (ICH, 2005), in which the following parameters were evaluated: specificity, precision (repeatability and intermediate precision), linearity, limit of detection (LOD), limit of quantification (LOQ) and accuracy. The method was validated based on the area of

each peak, in which the area was divided by the area of the internal standard (diclofenac) to minimize variations and provide more precise results.

2.6.1 Specificity

The specificity of the developed method was determined through the injection of the solvent solution containing only the internal standard (blank), the standards solution and the MIX solution (n=3). The retention times and resolution of the chromatographic peaks of (+)-catechin, (–)-epicatechin, ecdysterone and rutin, present in the standard and MIX sample solution, were used to calculate the relative standard deviation (RSD). The high-resolution mass spectra of each compound present in both standard and MIX solutions confirmed the specificity and identity of the peaks.

2.6.2 Precision (repeatability and intermediate precision)

The precision was estimated by the analysis of six MIX solutions, each one injected once, while the repeatability (or intra-day precision) was calculated as the RSD of the standards' concentrations of injection (μg mL^{-1}). The intermediate precision consisted in the same procedure performed in a different day. The results were compared through F-test.

2.6.3 Linearity

The linearity of the method was calculated by the calibration curves of the four standards selected, in a concentration range expected for each compound in the MIX sample. Initially, A stock standard solution containing catechin (100 μg mL^{-1}), epicatechin (150 μg mL^{-1}), ecdysterone (776 μg mL^{-1}), and rutin (300 μg mL^{-1}) solubilized in MeOH–H_2O 1:1 was prepared for simultaneous acquisition of the analytical curves. From this stock solution, 11 concentration levels were prepared in (MeOH–H_2O 1:1) containing the internal standard sodium diclofenac at 15 μg mL^{-1}. The concentrations ranged from (i) 0.2 to 25 μg mL^{-1} for catechin, (ii) 0.3 to 37.5 μg mL^{-1} for epicatechin, (iii) 1.6 to 194 μg mL^{-1} for ecdysterone, and (iv) 0.6 to 75 μg mL^{-1} for rutin. Each calibration concentration sample was injected in triplicate. The ratio

between the areas each calibration point divided by the area of the internal standard ($R=A_{compound}/A_{IS}$), along with the nominal concentrations were used to build the analytical curves for each standard. The results were interpreted in function of the correlation coefficients (R) calculated for each compound.

2.6.4 Limits of detection (LOD) and limits of quantification (LOQ)

The LODs and LOQs were estimated from the standard deviation of the y-intercept (Sb) and the slope (a) of three calibration curves prepared in three low concentrations, prepared from the same stock solution used in the linearity determination. The solutions were also prepared in MeOH–H_2O 1:1, containing sodium diclofenac at 15 μg mL^{-1}. The three concentration levels ranged from: (i) 0.2 to 1 μg mL^{-1} for catechin, (ii) 0.3 to 1.5 μg mL^{-1} for epicatechin, (iii) 1.6 to 7.8 μg mL^{-1} for ecdysterone, and (iv) 0.6 to 3 μg mL^{-1} for rutin. The calculations were performed using the following equations: $LOD = 3.3Sb.a^{-1}$ and $LOQ = 10Sb.a^{-1}$, where *a* is the mean of slopes of the calibration curves and *Sb* is the SD of the y-intercept.

2.6.5 Accuracy

Accuracy was evaluated considering the specified range of the analytical procedure by a recovery study. Preanalyzed standard and MIX solutions were used to prepare three different concentrations levels by spiking known amounts of the stock standards solution into the MIX solution. Three concentration levels were spiked, named low, intermediate, and high (66.7, 100.0 and 166.7 % levels, respectively). In this way, solutions containing the expected concentrations of 2.99, 8.99 and 14.99 μg mL^{-1} of catechin, 4.18, 13.18 and 22.18 μg mL^{-1} of epicatechin, 17.99, 64.55 and 111.11 μg mL^{-1} of ecdysterone and 7.64, 25.64 and 43.64 μg mL^{-1} of rutin were prepared. The accuracy was calculated by the difference between the nominal and the experimentally calculated contents.

2.6.6 Quantification of the standards in the individual extracts

The standards used for the chromatographic method validation procedure [(+)-catechin, (–)-epicatechin, ecdysterone and rutin] were also quantified in the individual extracts considering the equations obtained for each compound in the linearity evaluation. Each individual extract was prepared in triplicate for RSD calculation.

2.7 MS/MS data processing and Molecular Networking analysis

Initially, both samples and blank analysis (CH_3CN–H_2O, 1:1) were converted to .mzXML format using the MSConvert from ProteoWizard software (ProteoWizard, Palo Alto, CA, USA) (KESSNER *et al*., 2008). These files were used to create molecular networks and perform library searches in the Global Natural Products Social Molecular Networking (GNPS) (WANG *et al*., 2016) platform, in which the positive and negative ionization mode data were considered separately. MS/MS peaks within ±17 Da of the precursor ion were removed, as well as only the 6 most intense fragment ions in a window of ±50 Da were selected to filter the data. MSCluster was used by setting both the parent mass tolerance and the MS/MS fragment ion tolerance in 0.02 Da to create consensus spectra. The molecular networks were created with a cosine score threshold above 0.6, and more than four matched fragment ions. The clustered spectra were also searched against the GNPS public spectral libraries, also considering the cosine threshold of 0.65, and four matched fragment ions. The visualization of the networks was performed in Cytoscape software (SHANNON *et al*., 2003) (version 3.7.2, Cytoscape consortium, San Diego, CA, USA). The blank analysis was also included in the analysis, and the nodes corresponding to features detected in the blank analysis were excluded from the networks. Node colors were determined based on the MS/MS data source files (a different color for each species), and the edge thickness was defined to reflect the similarities between the nodes: the thicker the lines, the more similar the spectra are.

The UHPLC-HRMS/MS data were deposited in the Mass Spectrometry Interactive Virtual Environment (MassIVE) repository and is publicly available. The molecular networking job on GNPS can be found at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=206279f229df48f69514c36cd7447cb5.

3 Results and Discussion

3.1 Sample preparation

A representative sample of nine Malpighiaceae species, named MIX, was initially prepared. which were selected considering their phylogenetic classification. By choosing one species from each phylogenetic group found in Brazil, it would be expected that the most evolutionary distant species could produce the most structurally different secondary metabolites (even though the environmental stimuli could also affect the metabolite production) (FANG; LUO, 2019). Therefore, this MIX was considered a robust representative sample for method optimization

3.2 Chromatographic method development

The chromatographic method development was performed using Fusion QbD® software, which consisted in three main steps: screening, optimization, and robustness simulator (in which the last one is completely based on simulations and calculations). This software combined with the automation of the chromatographic system automatic column change valves and the subdivision of channel D (D1−D6, which made possible testing different pHs in the same batch), enabled this process to be performed efficiently, without manual intervention.

For the screening and optimization steps, the software provides statistical analyses in a semi-automated fashion, such as the Analysis of Variance (ANOVA) table, the equation that best describes the model proposed, the Pareto chart, and the surface responses. This information gives a higher statistical basis in the analyses performed in each step.

3.2.1 Screening

During the screening procedure, usually more abrupt changes in chromatographic selectivity are performed. At this point, six columns, two organic solvents, gradient time, and large variations of pH were assessed Figure 5A. Considering the entered variables, the statistical quadratic model A- and G- optimal was applied, which is usually recommended when categorical variables is higher than 2 levels (such as the

six columns evaluated) (MONTGOMERY, 2012; S-MATRIX, [s.d.]). Due to the possibility of automatic column change, the quaternary pump and the subdivision of channel D, this step could be performed in only one day and a half. At this point, the wavelength that provided the higher number of peaks and a relatively low matrix effect intensity (baseline deviation due to tannins in the sample) was 275 nm, being this wavelength chosen for data processing in all steps. All chromatograms were processed, and the responses chosen for statistical analyses were the number of total peaks, number of peaks with resolution ≥ 1.5, number of peaks with resolution ≥ 2.0 and number of peaks with tailing ≤ 1.2. These responses were imported back to Fusion QbD® which provided the best chromatographic condition for the screening procedure in which all the responses should be optimized simultaneously: column Acquity UPLC BEH Shield RP18, acidic pH, 30 min of gradient time, and CH_3CN as organic solvent. The chromatogram obtained for this condition is shown in Figure 5B.

As previously stated, several statistical responses can be automatically retrieved from Fusion QbD®. The ANOVA table is shown in Supplementary Table 4, while the mathematical equations and Pareto charts obtained for the four responses in this step are shown in Supplementary Table 5. The R^2 value for the four responses ranged from 0.78 to 0.90, indicating that the models can explain 78% to 90% of the results obtained for each response. To test the model for statistical significance, two conditions were specified: mean square regression (MSR) ≥ MSR significance threshold and mean square of the lack-of-fit (MS-LOF) ≤ MS-LOF significance threshold. These models proved to be statistically significant for the number of total peaks and the number of peaks with high resolution; however, a lack of fit was observed for the number of peaks with low tailing.

Through the Pareto charts, it is possible to visualize the impact that each parameter has on the model. For instance, the number of total peaks and the number of peaks with low tailing is highly influenced by the pH and column type, while the number of peaks with high resolution is impacted by the gradient time and the column type. The response surfaces obtained for the screening step by setting the column as Acquity UPLC BEH Shield RP18 column and the organic solvent as CH_3CN (Figure 6), it is possible to observe that, overall, higher responses were obtained in lower pH values, and that the gradient time has little impact on the responses.

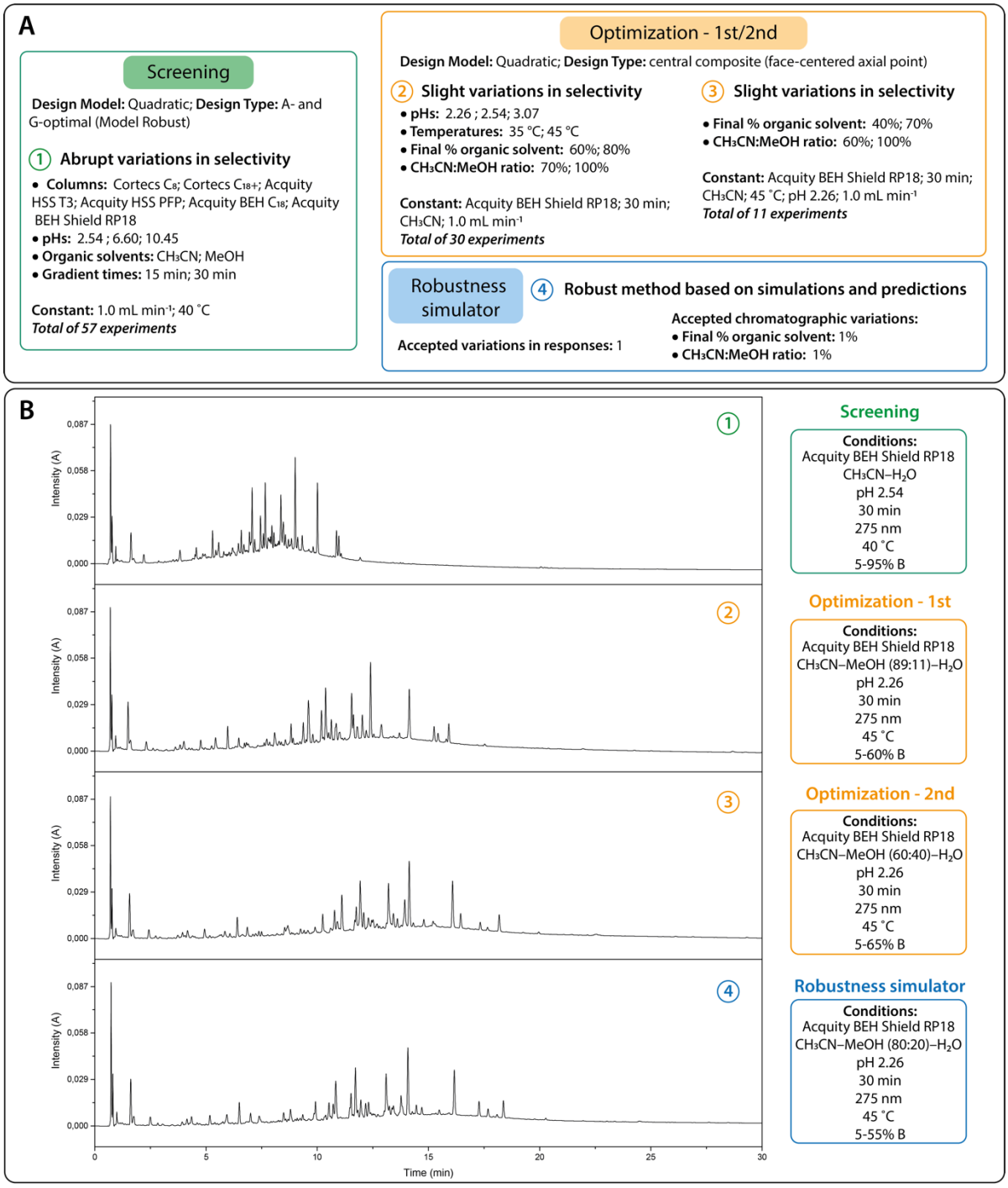
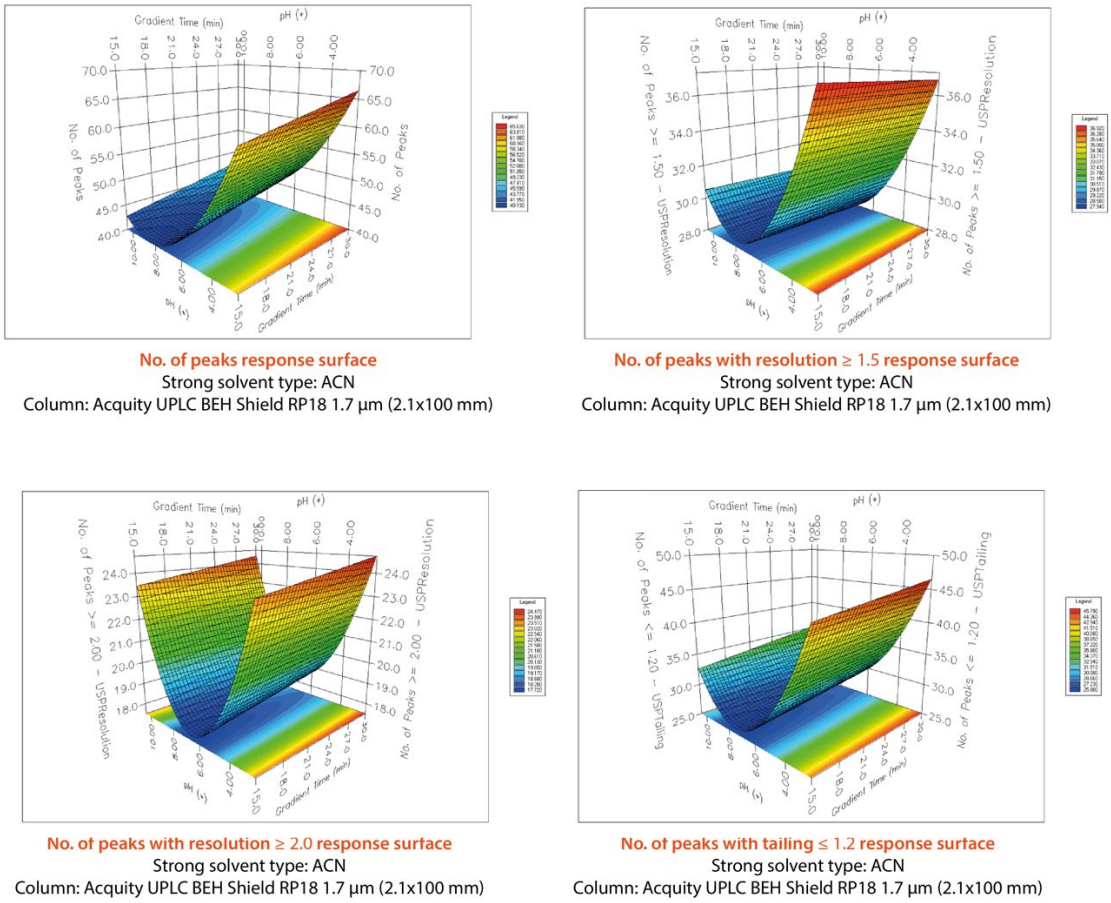


Figure 5 – Workflow followed during the chromatographic method optimization using the QbD approach. (**A**) Three main steps were followed in the chromatographic method development: screening, optimization, and robustness simulator. The designs, variables chosen, and fixed parameters are described. (**B**) Chromatograms with the higher overall responses obtained for the MIX sample for each step are shown (275 nm), such as the conditions employed.

Source: elaborated by the Author (2022).



No. of peaks response surface
Strong solvent type: ACN
Column: Acquity UPLC BEH Shield RP18 1.7 µm (2.1x100 mm)

No. of peaks with resolution ≥ 1.5 response surface
Strong solvent type: ACN
Column: Acquity UPLC BEH Shield RP18 1.7 µm (2.1x100 mm)

No. of peaks with resolution ≥ 2.0 response surface
Strong solvent type: ACN
Column: Acquity UPLC BEH Shield RP18 1.7 µm (2.1x100 mm)

No. of peaks with tailing ≤ 1.2 response surface
Strong solvent type: ACN
Column: Acquity UPLC BEH Shield RP18 1.7 µm (2.1x100 mm)

Figure 6 – Response surface plots for the responses number of total peaks, number of peaks with resolution ≥ 1.5, number of peaks with resolution ≥ 2.0, and number of peaks with tailing ≤ 1.2 for the screening step of the chromatographic method development.

Source: elaborated by the Author (2022).

3.2.2 Optimization

The chromatogram representing the best condition of the screening step still presented poorly resolved peaks, and most of them eluting in the first 15 min of analysis. Therefore, during the optimization step, slight variations on the pH were evaluated, as well as oven temperature, the final percentage of organic solvent, and the organic phase composition (CH_3CN to MeOH ratio). The variations of the gradient and organic phase composition could provide a better peak distribution in the chromatogram, while slight variations in temperature and pH could provide better peak resolution and less peak tailing (SNYDER; KIRKLAND; DOLAN, 2009).

Considering these variables, the central composite with a face-centered axial point experimental design was employed at this point. This design is usually recommended in the optimization process, in which continuous variables are evaluated (in this case, the final percentage of organic solvent, temperature, and CH_3CN to MeOH ratio) since it is possible to include intermediate points between the minimum and maximum desired in the evaluation of each parameter (MONTGOMERY, 2012).

Similarly to the screening procedure, the responses obtained for the matrix of experiments were imported back to Fusion QbD®, and the combination of parameters that provided the best overall responses was: pH 2.26, oven temperature of 45 ˚C, gradient of 5−60% B and CH_3CN−MeOH (89:11)−H_2O. The chromatogram (275 nm) obtained for this step is shown in Figure 5B. The ANOVA table, equations and Pareto charts can be found in the Supplementary Material (Supplementary Table 6 and Supplementary Table 7). For all these responses, the two conditions specified (MSR ≥ MSR significance threshold and MS-LOF ≤ MS-LOF significance threshold) were achieved, indicating that the models were statistically significant.

The response surfaces obtained by setting the final percentage of organic solvent to 60%, and the organic solvent ratio to 89:11 (Figure 7) showed that, in general, all the responses evaluated were higher in lower pH values and higher temperatures. The response surfaces generated by setting pH to 2.26 and temperature to 45 ˚C (Figure 8) demonstrated that the number of total peaks was maximized at approximately 9:1 CH_3CN:MeOH in the organic phase composition, while the final percentage of organic solvent did not influence this response greatly. A maximum was observed in the response surface obtained for the number of peaks with resolution ≥ 1.5, at 67% as the final organic solvent and 80% CH_3CN in the organic phase. In contrast, the number of peaks with low tailing is maximized in two different regions: both in lower and higher percentages of final organic solvent.

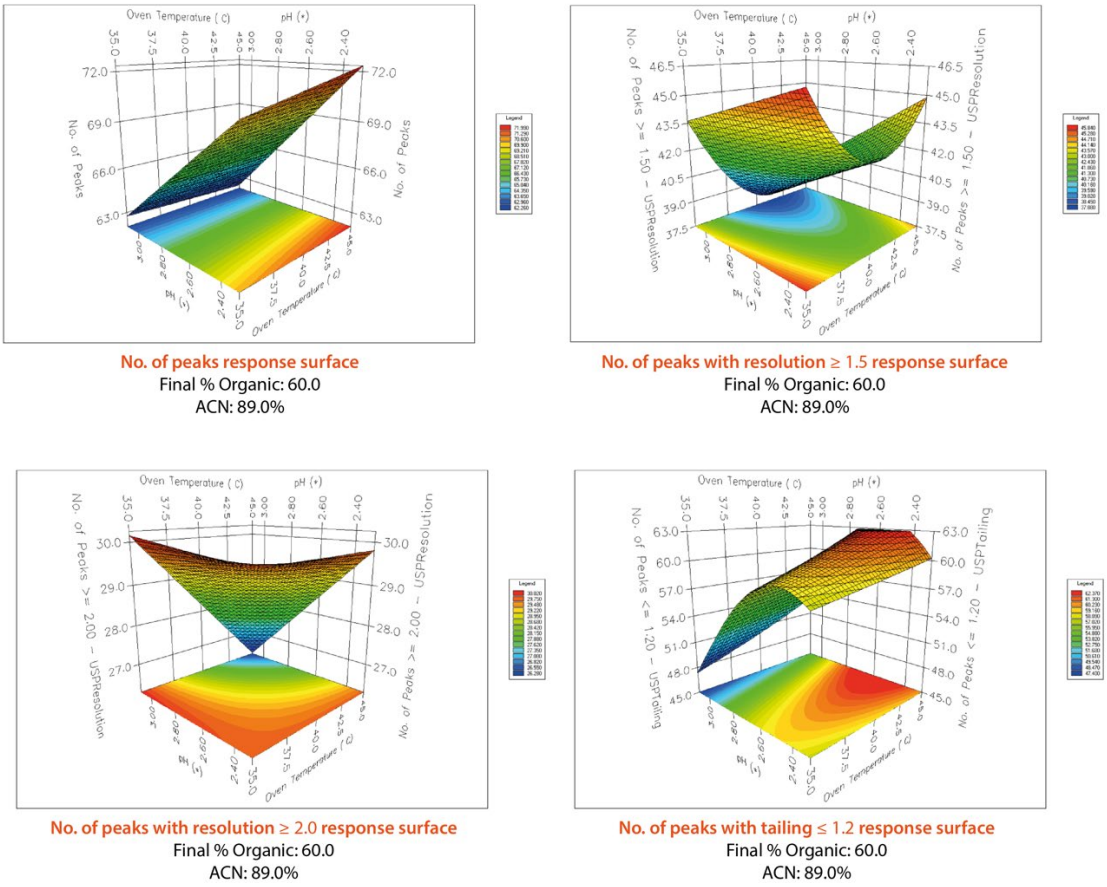
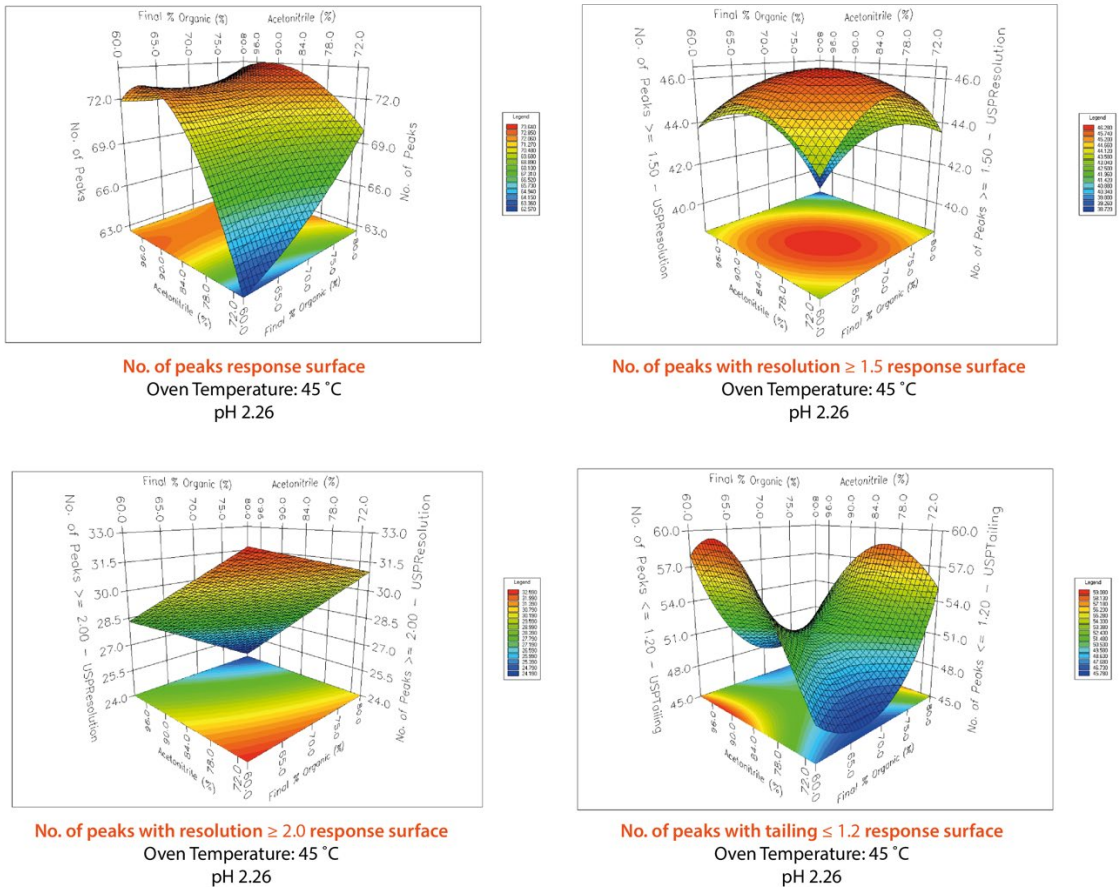


Figure 7 – Response surface plots for the responses number of total peaks, number of peaks with resolution ≥ 1.5, number of peaks with resolution ≥ 2.0, and number of peaks with tailing ≤ 1.2 for the first optimization step of the chromatographic method development. Constant variables: final percentage of organic solvent (60 ˚C) and CH_3CN to MeOH ratio (89:11).

Source: elaborated by the Author (2022).



No. of peaks response surface
Oven Temperature: 45 ˚C
pH 2.26

No. of peaks with resolution ≥ 1.5 response surface
Oven Temperature: 45 ˚C
pH 2.26

No. of peaks with resolution ≥ 2.0 response surface
Oven Temperature: 45 ˚C
pH 2.26

No. of peaks with tailing ≤ 1.2 response surface
Oven Temperature: 45 ˚C
pH 2.26

Figure 8 –Response surface plots for the responses number of total peaks, number of peaks with resolution ≥ 1.5, number of peaks with resolution ≥ 2.0, and number of peaks with tailing ≤ 1.2 for the first optimization step of the chromatographic method development. Constant variables: oven temperature (45 ˚C) and pH (2.26).

Source: elaborated by the Author (2022).

By comparing the chromatograms obtained for the best chromatographic conditions in all the runs performed in both the screening and optimization steps (Figure 5B), there was an evident chromatographic improvement, especially in terms of total peaks and their distribution in the chromatograms. However, the distribution of the peaks could still be improved, and possibly resolve more peaks that could be coeluting. At this point, pH and temperature would be limiting parameters to the column initially chosen (Table 1) and therefore were fixed in their optimal values retrieved in the optimization step: the lowest pH (2.26) and highest temperature (45 ˚C).

Therefore, a final experimental design (named as a second optimization step) was applied in order to re-evaluate only the final percentage of organic solvent and organic phase composition. Once again, the central composite with a face-centered axial point

design was applied. From the responses obtained in the last matrix of experiments, the best conditions for this second optimization step were as follows: CH_3CN−MeOH (60:40)−H_2O and a gradient of 5−65% (chromatogram shown in Figure 5B). The ANOVA tables, Pareto charts and equations are detailed in Supplementary Table 8 and Supplementary Table 9, while the response surfaces obtained are depicted in Figure 9.

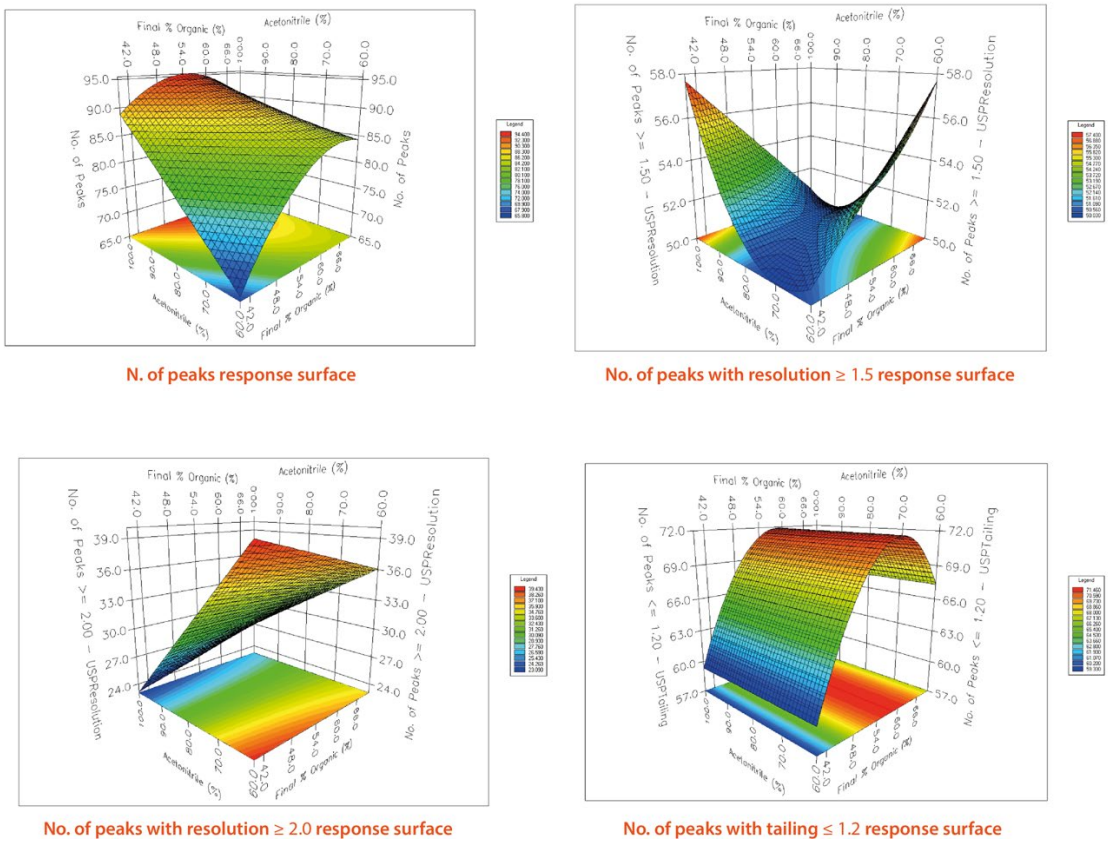


Figure 9 – Response surface plots for the responses number of total peaks, number of peaks with resolution ≥ 1.5, number of peaks with resolution ≥ 2.0, and number of peaks with tailing ≤ 1.2 for the second optimization step of the chromatographic method development.

Source: elaborated by the Author (2022).

These surfaces show a higher number of total peaks can be obtained when the final percentage of organic solvent approaches 54% and the organic phase composition approaches 100% CH_3CN. The number of peaks with a resolution of ≥1.5 could be observed at two extreme points (100% CH_3CN as solvent B and 40% of final solvent B/ CH_3CN−MeOH (60:40) as solvent B and 70% of final solvent B), and those

with a resolution of ≥ 2.0 were obtained with lower final percentages of organic solvent and CH_3CN percentages in the organic solvent. The number of peaks with tailing of ≤1.2 was higher when the final percentage of solvent B was close to 60%, and the composition of the organic phase had a small impact on this response. Even though the peaks were not distributed equally in the chromatogram obtained in this phase, it is important to emphasize that a higher number of peaks may not necessarily correspond to peaks with high resolution or reduced tailing.

3.2.3 Robustness simulator

The Fusion QbD® software also includes another tool called Robustness Simulation, which is an important final step to determine a robust chromatographic method. In other words, a robust method would still be able to provide acceptable results if slight variations in the parameters are made. The robustness simulation was performed in the second optimization step, in which the only variables evaluated were the final % organic solvent and the % CH_3CN in the organic mobile phase. After simulations and calculations, the software provides a robust method to optimize all the responses simultaneously. Therefore, the final conditions determined for the separation of the compounds of the MIX sample were as follows: Acquity BEH Shield RP18 (2.1 x 100 mm, 1.7 µm), CH_3CN−MeOH (80:20)−H_2O, pH 2.22, 275 nm, 45 ˚C, 5−55% B in 30 min (Figure 5B).

By comparing the chromatogram retrieved as the best one in the initial screening step with the last one obtained in the robustness simulation, there was an evident chromatographic improvement in all the responses selected for this study. The number of total peaks, the number of peaks with resolution ≥ 1.5, the number of peaks with resolution ≥ 2.0 and number of peaks with tailing of ≤1.2 increased by 54%, 21%, 19%, and 43%, respectively.

After developing this robust chromatographic method in only four days, the plant extracts were analyzed individually, and the chromatograms are shown in Figure 10. A very satisfactory overall peak separation was observed for all the individual extracts. These results highlight that the Quality by Design approach, the elaboration of a mixture of extracts, and instrumental automatization efficiently accelerate this crucial step in the natural products workflow, especially compared with the traditional OFAT approach.

3.3 Analytical method validation

In order to determine if the developed method would provide consistent, reliable, and accurate data, it was validated based on the ICH guidelines (ICH, 2005). The specificity, precision, linearity, limits of detection (LOD) and quantification (LOQ), and accuracy were determined based on the standards (+)-catechin, (−)-epicatechin, ecdysterone, and rutin.

Specificity is the ability of the method to unequivocally assess and differentiate the signals of the analyte from other compounds, such as impurities, other matrix components or degradants, among others. The detected peaks corresponded to (+)-catechin (4.88 min), (–)-epicatechin (6.25 min), ecdysterone (10.61 min), and rutin (11.56 min) in the MIX solution, showed good resolutions (5.38, 1.98, 7.71 and 1.25, respectively) and did not show any interferences when compared to the standard solution containing only the pure compounds. Moreover, the recorded mass spectra allowed the confirmation of each compound's identity.

The repeatability was determined on two days, based on the relative standard deviation (RSD) of six MIX solution replicates. The concentration of each compound during the injection (μg mL^{-1}) was taken into account for the calculations. For the first day, the average concentrations measured for (+)-catechin, (–)-epicatechin, ecdysterone, and rutin were 10.101, 13.644, 52.377, and 23.672 μg mL^{-1}, respectively. During the second day, the average concentrations measured for (+)-catechin, (–)-epicatechin, ecdysterone, and rutin were 10.446, 13.829, 52.545, and 23.605μg mL^{-1}, respectively. On both days, the RSDs determined for all compounds were lower than 1% (Table 2). The intermediate precision, calculated by the F-test between the two different days, was also very satisfactory since no significant difference at F= 0.05 (n=6-1) was detected.

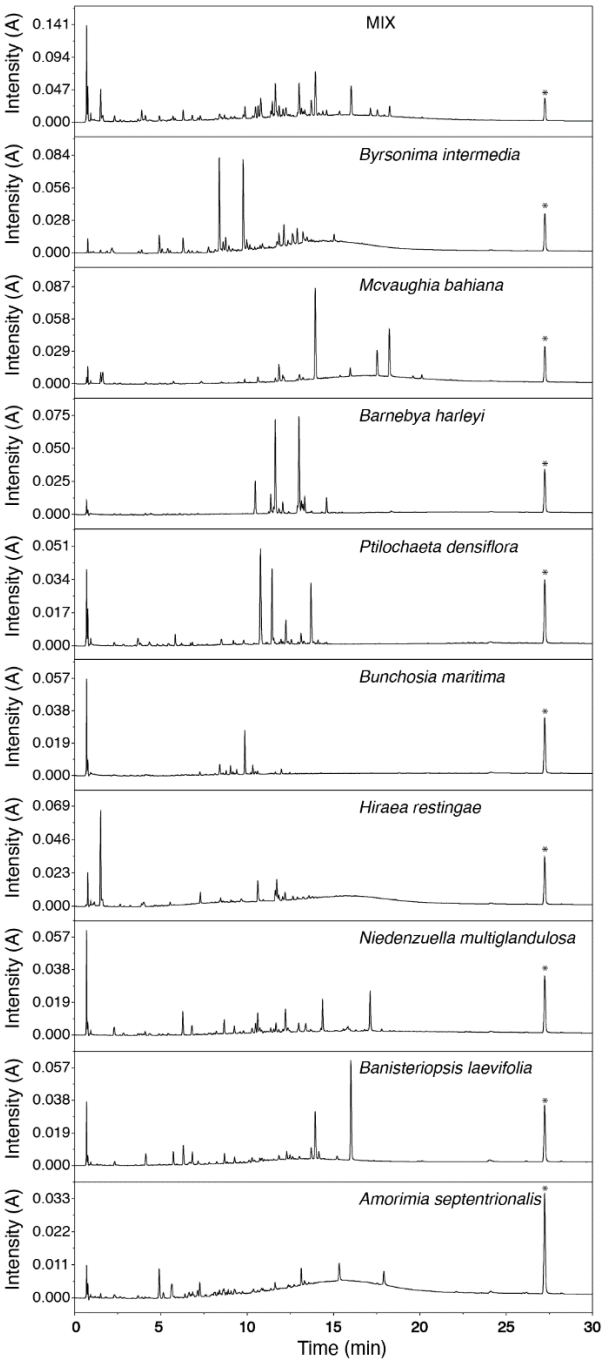


Figure 10 – Chromatograms (275 nm) obtained by UHPLC-PDA from the MIX and the individual extracts from Malpighiaceae plant species. Sodium diclofenac (*) was used as the internal standard.

Source: elaborated by the Author (2022).

The linearity was determined by calibration curves of each compound ranging from 2.2 to 278% of the working standards concentrations (n=3). The linear ranges of concentrations obtained were adequate for all compounds, and the obtained values of

R coefficients were for 0.9996, 0.9994, 0.9995, 0.9994 for catechin, epicatechin, ecdysterone, and rutin, respectively (Table 2).

Both LOD and LOQ were calculated through the elaboration and injection of three analytical curves, prepared at three low concentrations levels. The results obtained for the LOD and LOQ for catechin, epicatechin, ecdysterone, and rutin were 0.06, 0.02, 0.07, and 0.03 μg mL^{-1} and 0.18, 0.06, 0.22, and 0.11 μg mL^{-1}, respectively.

The accuracy results show an average of recovery corresponded to 86.72% (low level), 88.71% (intermediate level) and 93.73% (high level), taking into account all compounds in each level. Considering each compound separately, the RSD was lower than 5%.

This method reached all the acceptance criteria specified for each validation parameter and proved to be sensitive for the compounds used. By comparing the LOQ and LOD values of the compounds used in this validation with those reported in the literature [catechins: LOD of 0.10−0.41 μg mL^{-1}, LOQ of 0.30^{-1}.41 μg mL^{-1} (NALDI *et al*., 2014); ecdysterone: LOD of 2.63 μg mL^{-1}, LOQ of 7.97 μg mL^{-1} (ROSA *et al*., 2018); rutin: LOD of 0.18 μg mL^{-1}, LOQ of 0.60 μg mL^{-1} (MAGIERA; BARANOWSKA; LAUTENSZLEGER, 2015)], it was possible to determine that the approach followed proved to be effective for the development of a sensitive method, even for different classes of metabolites. Therefore, the approach followed in our study was also efficient in detecting compounds in very low concentrations, which could have been missed if coeluted with other major compounds.

These standards were also quantified in the individual extracts, and their distribution in the species is shown in Figure 11. The presence of these compounds was not detected in only two species: *P. densiflora* and *H. restingae*. *B. intermedia* contained both catechins in similar amounts, while *B. harleyi* presented the highest concentration of rutin among all the species. In addition, ecdysterone was only detected in *N. multiglandulosa*.

Table 2 – Method Validation Items, Parameters, Results, and Acceptance Criteria (according to the ICH guidelines).

Validation item	Parameter	Compounds				Acceptance criteria
		(+)-catechin	(–)-epicatechin	ecdysterone	rutin	
Linearity (n=3)	Linear dynamic range[a]	0.2 to 25.0	0.3 to 37.5	1.6 to 194.0	0.6 to 75.0	informative
	Equation	y = 0.017x – 0.0032	y = 0.0191x – 0.0049	y = 0.0054x + 0.0009	y = 0.0314x – 0.0169	informative
	R	0.9996	0.9994	0.9995	0.9994	R>0.9900
Precision (n=6)	Repeatability: Day 1	0.48	0.72	0.53	0.27	RSD<5%
	Repeatability: Day 2	0.41	0.39	0.61	0.26	RSD<5%
	Intermediate precision[b]	1.28	3.32	0.76	1.10	$F_{calc} < F_{tabl}$
Accuracy[c] (n=3)	Low level average	86.30	87.96	83.95	88.64	80-120%
	Intermediate level average	90.58	91.59	86.11	86.54	80-120%
	High level average	96.51	97.13	90.42	90.87	80-120%
LOD and LOQ (n=3)	Slope average (±SD)	0.015 (±0.0004)	0.016 (±0.0001)	0.005 (±0.0001)	0.027 (±0.0003)	–
	y-intercept average (±SD)	0.001 (± 0.0003)	0.001 (± 0.0001)	0.000 (± 0.0001)	0.003 (± 0.0003)	–
	LOD[a]	0.06	0.02	0.07	0.03	informative
	LOQ[a]	0.18	0.06	0.22	0.11	informative

[a]Concentration in μg mL^{-1}.

[b]F = 0.05; n = 6−1.

[c]Percentage of recovery.

Source: elaborated by the Author (2022).

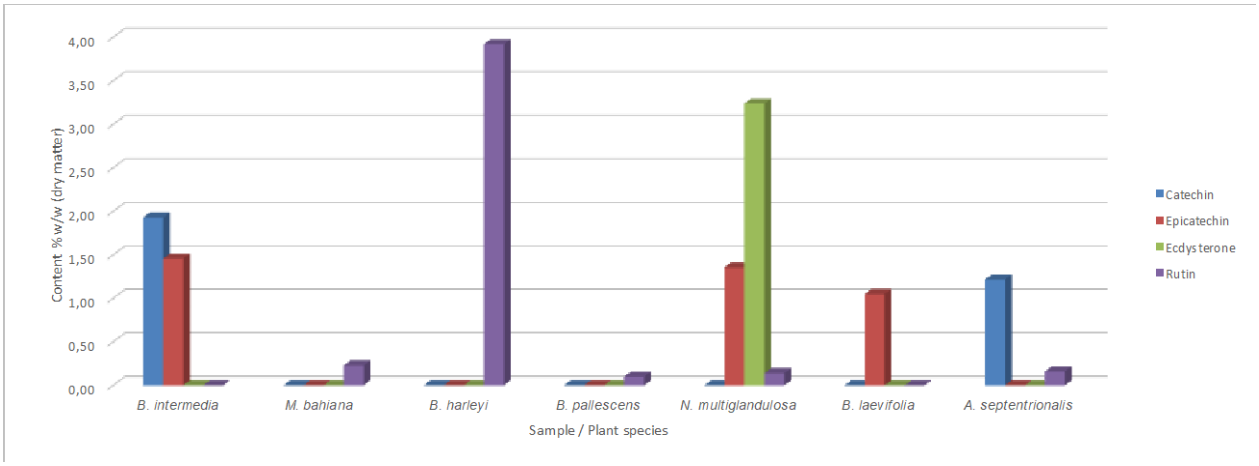


Figure 11 – Distribution of the standards used in method validation among the species used in the MIX preparation.

Source: elaborated by the Author (2022).

3.4 Annotation of the metabolites

The possibility of coupling liquid chromatography to mass spectrometers is valuable for compounds annotation in extracts, especially tandem mass spectrometry (MS/MS) in high-resolution spectrometers, such as the quadrupole time of flight (QToF) (AKSENOV *et al*., 2017). With the increasing number of experimental MS/MS databases and spectral libraries, such as MassBank (HORAI *et al*., 2010), METLIN (GUIJAS *et al*., 2018), and ReSpect (SAWADA *et al*., 2012), the employment of bioinformatics tools for MS/MS data processing and comparison with databases has become more popular. In this context, the Global Natural Product Social Molecular Networking (GNPS) (WANG *et al*., 2016) stands out for the public spectral libraries available and the possibility to visually inspect similarities among the MS/MS data through networks indicating the molecular families present in the dataset.

The individual extracts were then analyzed in the optimized chromatographic method by UHPLC-HRMS/MS aiming to annotate the secondary metabolites present in the samples. Three genera included in our sampling did not present any chemical study to date (*Mcvaughia*, *Barnebya*, and *Ptilochaeta*); therefore, these analyses would provide the first chemical information regarding their metabolites. The MS/MS data were searched against the GNPS spectral reference library, and a molecular network was constructed for data visualization. In addition, all the MS/MS spectra were also manually compared with spectra previously reported in the literature to verify the

annotation made with the GNPS platform. Additionally, the ions not annotated in GNPS were searched in the *Dictionary of Natural Products*.

By this approach, it was possible to annotate 61 compounds and to observe a wide diversity of compounds present in the Malpighiaceae family, such as *O*-glycosylated flavonoids, *C*-glycosylated flavonoids, quinic/shikimic acid derivatives, steroids, and polyphenols, highlighted in Figure 12. The molecular formulas and fragmentation of the annotated compounds were also confirmed to result in a level 2 or 3 annotation (SCHRIMPE-RUTLEDGE *et al.*, 2016) and can be found in Table 3. These results show that the optimized method allowed the annotation of a large number of compounds, even in plants that have not been previously investigated.

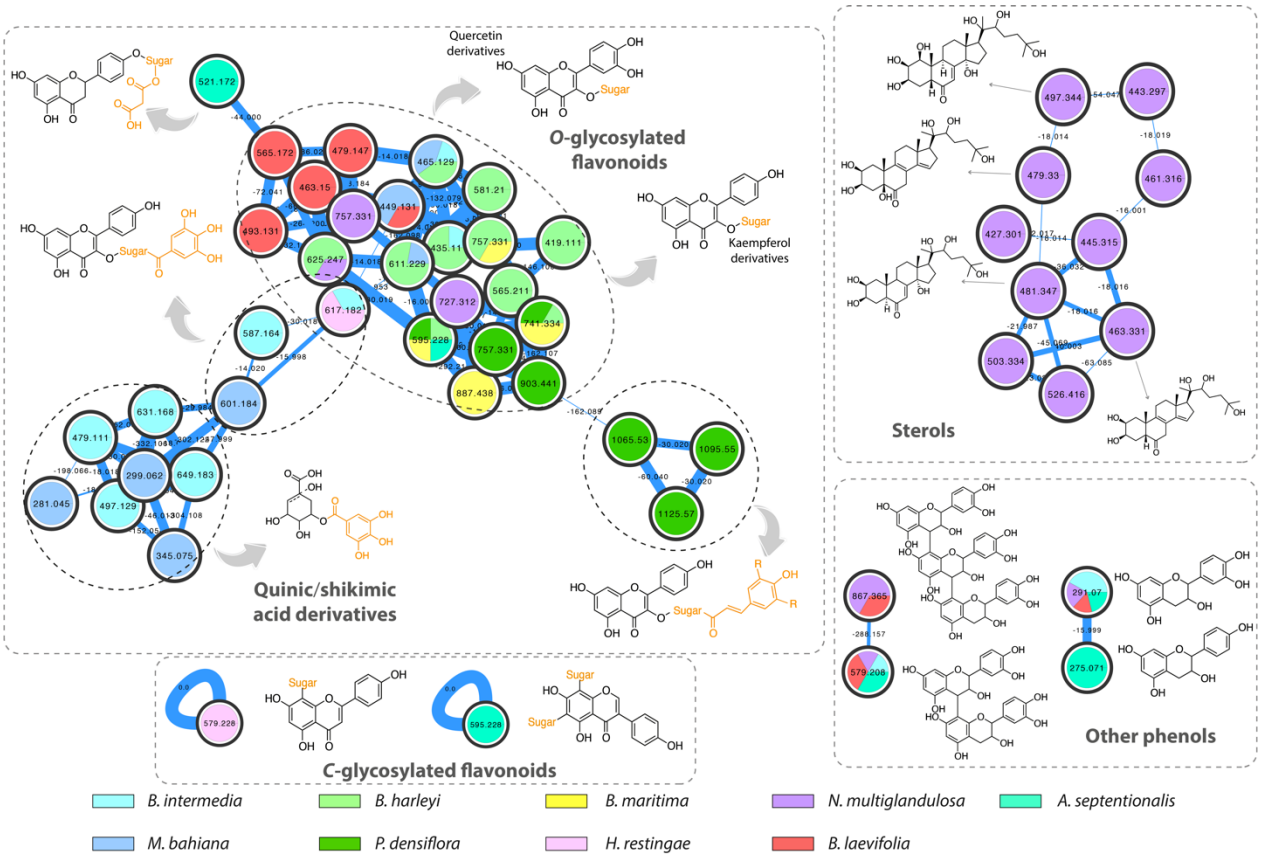


Figure 12 – Molecular families annotated using the classical molecular networking workflow: *O*-glycosylated flavonoids, *C*-glycosylated flavonoids, quinic/shikimic acid derivatives, sterols, and other phenols. These are level 2 matches according to the 2007 metabolomics standards initiative (SUMNER *et al.*, 2007). Structures shown with a defined configuration were confirmed with standards isolated from *N. multiglandulosa*.

Source: elaborated by the Author (2022).

On the molecular family characterized by phenolic compounds, quercetin and kaempferol derivatives are slightly separated. Quercetin derivatives (aglycone characterized by a fragment at *m/z* 303) can be found in *B. intermedia*, *M. bahiana*, *B. harleyi*, *N. multiglandulosa*, and *B. laevifolia* species, while kaempferol derivatives (aglycone at *m/z* 287) were observed in *B. harleyi*, *B. pallescens,* and *N. multiglandulosa* extracts. The substituents observed for the flavonoids were, mainly, hexoside (162 Da), deoxyhexoside (146 Da), and pentoside (132 Da). Some of these flavonoids were observed both as a protonated ion ($[M+H]^+$) and as a sodiated adduct ($[M+Na]^+$). Two points are worth highlighting at this moment: (1) the number of nodes does not reflect the number of metabolites present in the sample, as the same compound can be retrieved as different adducts; (2) in the classical molecular networking workflow, isomers presenting the same fragmentation pattern are clustered in the same node by MScluster. Thus, the inspection for isomers must be performed carefully.

Table 3 – Metabolites annotated by UHPLC-ESI-MS^2 and Molecular Network in the extracts present used to elaborate the MIX sample – positive ionization mode.

(to be continued)

#	Rt (min)	Species	Molecular Formula	$[M+H]^+$ observed	$[M+H]^+$ calculated	Error (ppm)	MS/MS fragments (% abundance)	Metabolite
1	1.81	*M. bahiana*	$C_{14}H_{17}O_{10}$	345.0832	345.0822	2.9	153.0202 (100)	Galloyl quinic acid
2	5.25	*B. intermedia*	$C_{15}H_{15}O_6$	291.0877	291.0869	2.7	207.0675 (20), 165.0560 (15), 147.0459 (20), 139.0403 (100), 123.0459 (50)	Catechin
2	5.25	*A. septentrionalis*	$C_{15}H_{15}O_6$	291.0874	291.0869	1.7	207.0667 (15), 165.0555 (15), 147.0462 (20), 139.0411 (100), 123.0459 (35)	Catechin
3	5.55	*A. septentrionalis*	$C_{30}H_{27}O_{12}$	579.1505	579.1503	0.3	409.0920 (30), 287.0572 (80), 275.0572 (60), 247.0611 (50), 233.0461 (30), 163.0416 (50), 139.0410 (70), 127.0390 (100), 123.0457 (40)	Proanthocyanidin dimer
4	5.77	*B. intermedia*	$C_{30}H_{27}O_{12}$	579.1501	579.1503	-0.3	409.0939 (40), 301.0712 (15), 287.0581 (85), 271.0616 (45), 247.0648 (40), 233.0464 (15), 191.0378 (15), 163.0419 (40), 139.0420 (60), 127.0412 (100), 123.0467 (30)	Proanthocyanidin dimer
5	6.04	*A. septentrionalis*	$C_{30}H_{27}O_{12}$	579.1508	579.1503	0.9	409.0963 (70), 301.0708 (20), 287.0569 (100), 275.0569 (60), 247.0605 (40), 233.0446 (20), 163.0405 (45), 139.0428 (60), 127.0405 (55), 123.0458 (35)	Proanthocyanidin dimer
6	6.65	*B. intermedia*	$C_{15}H_{15}O_6$	291.0877	291.0869	2.7	207.0670 (15), 147.0459 (25), 139.0408 (100), 123.0462 (40)	Epicatechin
6	6.65	*B. laevifolia*	$C_{15}H_{15}O_6$	291.0870	291.0869	0.3	207.0670 (10), 165.0558 (100), 147.0466 (20), 139.0408 (100), 123.0465 (40)	Epicatechin
6	6.66	*N. multiglandulosa*	$C_{15}H_{15}O_6$	291.0882	291.0869	4.5	207.0660 (15), 165.0559 (15), 147.0463 (20), 139.0408 (100), 123.0457 (35)	Epicatechin

Table 3 – Metabolites annotated by UHPLC-ESI-MS^2 and Molecular Network in the extracts present used to elaborate the MIX sample – positive ionization mode.

(continuation)

#	Rt (min)	Species	Molecular Formula	$[M+H]^+$ observed	$[M+H]^+$ calculated	Error (ppm)	MS/MS fragments (% abundance)	Metabolite
7	7.23	*N. multiglandulosa*	$C_{30}H_{27}O_{12}$	579.1499	579.1503	-0.7	409.0945 (45), 301.0731 (35), 287.0569 (90), 275.0583 (70), 247.0613 (35), 233.0454 (20), 163.0408 (70), 139.0403 (70), 127.0413 (100), 123.0779 (35)	Proanthocyanidin dimer
7	7.23	*B. laevifolia*	$C_{30}H_{27}O_{12}$	579.1505	579.1503	0.3	409.0950 (30), 287.0583 (70), 271.0630 (40), 247.0613 (30), 233.0455 (20), 163.0404 (40), 139.0422 (60), 127.0408 (100), 123.0458 (30)	Proanthocyanidin dimer
8	7.25	*A. septentrionalis*	$C_{15}H_{15}O_5$	275.0924	275.0919	1.8	201.0483 (10), 191.0726 (15), 149.0616 (20), 139.0405 (100), 107.0509 (25)	Afzelechin
9	7.70	*A. septentrionalis*	$C_{27}H_{31}O_{15}$	595.1657	595.1663	-1.0	541.1246 (20), 523.1274 (20), 481.1144 (40), 457.1113 (45), 427.1031 (65), 409.0929 (75), 379.0817 (90), 337.0736 (55), 325.0726 (100), 307.0597 (45), 295.0616 (30)	Genistein-di-*C*-hexoside
10	8.81	*B. intermedia*	$C_{21}H_{19}O_{13}$	479.0831	479.0826	1.0	309.0630 (15), 153.0200 (100)	Digalloyl shikimic acid
11	8.84	*B. intermedia*	$C_{21}H_{21}O_{14}$	497.0938	497.0931	1.4	309.0636 (10), 153.0199 (100)	Digalloyl quinic acid
12	8.87	*B. maritima*	$C_{39}H_{51}O_{23}$	887.2811	887.2821	-1.1	287.0568 (100)	Kaempferol-*O*-hexoside-deoxyhexoside-deoxyhexoside-deoxyhexoside
13	8.88	*B. intermedia*	$C_{28}H_{25}O_{18}$	649.1038	649.1041	-0.5	309.0635 (10), 153.0201 (100)	Trigalloyl quinic acid
14	9.11	*B. laevifolia*	$C_{45}H_{39}O_{18}$	867.2119	867.2136	-2.9	545.1077 (10), 527.1068 (15), 419.0790 (30), 409.0927 (60), 407.0764 (55), 393.0641 (30), 301.072 (30), 289.0740 (40), 287.0574 (60), 275.0552 (95), 247.0612 (100), 245.0452 (90),	Proanthocyanidin trimer

Table 3 – Metabolites annotated by UHPLC-ESI-MS^2 and Molecular Network in the extracts present used to elaborate the MIX sample – positive ionization mode.

(continuation)

#	Rt (min)	Species	Molecular Formula	$[M+H]^+$ observed	$[M+H]^+$ calculated	Error (ppm)	MS/MS fragments (% abundance)	Metabolite
							163.0412 (80), 139.0418 (80), 127.0418 (70), 123.0496 (40)	
14	9.12	*N. multiglandulosa*	$C_{45}H_{39}O_{18}$	867.2115	867.2136	-2.4	545.1047 (15), 527.0963 (25), 409.0902 (65), 407.0797 (40), 393.0627 (20), 301.0681 (25), 289.0714 (55), 287.0561 (70), 247.0615 (100), 245.0453 (90), 163.0425 (75), 127.0416 (70), 123.0459 (45)	Proanthocyanidin trimer
15	9.46	*B. maritima*	$C_{33}H_{41}O_{20}$	757.2178	757.2191	-1.7	303.1518 (100)	Quercetin-*O*-hexoside-deoxyhexoside-deoxyhexoside
16	10.11	*H. restingae*	$C_{27}H_{31}O_{14}$	579.1713	579.1714	-0.2	433.1136 (100), 415.1046 (55), 397.0939 (40), 379.0836 (10), 367.0835 (20), 337.0737 (15), 313.0727 (60), 283.0600 (10), 271.0587 (5)	Apigenin-*C*-hexoside-deoxyhexoside
17	10.20	*B. intermedia*	$C_{28}H_{23}O_{17}$	631.0937	631.0935	0.3	153.0201 (100)	Trigalloyl shikimic acid
18	10.31	*B. maritima*	$C_{33}H_{41}O_{19}$	741.2253	741.2242	1.5	287.0567 (100)	Kaempferol-*O*-hexoside-deoxyhexoside-deoxyhexoside
19	10.74	*P. densiflora*	$C_{27}H_{31}O_{15}$	595.1664	595.1663	0.2	287.0567 (100)	Kaempferol-*O*-hexoside-deoxyhexoside
19	10.74	*B. maritima*	$C_{27}H_{31}O_{15}$	595.1670	595.1663	1.2	287.0566 (100)	Kaempferol-*O*-hexoside-deoxyhexoside
20	10.76	*B. maritima*	$C_{33}H_{41}O_{19}$	741.2241	741.2242	-0.1	287.0565 (100)	Kaempferol-*O*-hexoside-deoxyhexoside-deoxyhexoside
21	10.86	*B. harleyi*	$C_{33}H_{41}O_{20}$	757.2187	757.2191	-0.5	303.0515 (100)	Quercetin-*O*-hexoside-deoxyhexoside-deoxyhexoside

Table 3 – Metabolites annotated by UHPLC-ESI-MS^2 and Molecular Network in the extracts present used to elaborate the MIX sample – positive ionization mode.

(continuation)

#	Rt (min)	Species	Molecular Formula	$[M+H]^+$ observed	$[M+H]^+$ calculated	Error (ppm)	MS/MS fragments (% abundance)	Metabolite
22	10.90	*B. harleyi*	$C_{27}H_{31}O_{16}$	611.1614	611.1612	0.3	303.0517 (100)	Quercetin-*O*-hexoside-deoxyhexoside
23	10.93	*N. multiglandulosa*	$C_{27}H_{43}O_7$	479.3013	479.3009	0.8	479.3017 (10), 461.2912 (30), 443.2816 (70), 425.2692 (100), 407.2585 (50), 389.2487 (10), 383.2599 (25), 351.1992 (10), 343.2313 (25), 311.2036 (25), 281.566 (30)	5-hydroxypodecdysone B
24	10.96	*N. multiglandulosa*	$C_{27}H_{45}O_8$	497.3119	497.3114	1.0	497.3134 (40), 479.3018 (10), 461.2907 (20), 443.2819 (70), 425.2698 (100), 407.2598 (45), 387.219 (20), 369.2086 (45), 351.1957 (65), 343.2288 (20), 311.2013 (25), 309.1983 (20), 281.1558 (30), 269.1534 (20)	Integristerone A
25	11.04	*N. multiglandulosa*	$C_{27}H_{43}O_6$	463.3063	463.3060	0.6	463.3047 (10), 445.2959 (100), 427.2848 (40), 409.2748 (30), 391.2643 (10), 371.2236 (15), 353.2130 (10), 329.1170 (20), 301.1812 (80), 283.1714 (20), 165.287 (50)	Podecdysone B
26	11.06	*H. restingae*	$C_{27}H_{31}O_{14}$	579.1715	579.1714	0.3	433.1161 (20), 415.1041 (15), 397.0915 (30), 379.0824 (45), 367.0836 (45), 349.0751 (15), 337.0729 (90), 313.0727 (100), 283.0616 (85)	Apigenin-*C*-hexoside-deoxyhexoside
27	11.09	*N. multiglandulosa*	$C_{27}H_{45}O_7$	481.3162	481.3165	-0.6	481.1698 (10), 445.2972 (100), 427.2838 (55), 409.2762 (40), 371.2232 (70), 162.1285 (80)	Ecdysterone
28	11.19	*P. densiflora*	$C_{39}H_{51}O_{24}$	903.2742	903.2770	-3.1	287.0557 (100)	Kaempferol-*O*-hexoside-hexoside-deoxyhexoside-deoxyhexoside

Table 3 – Metabolites annotated by UHPLC-ESI-MS^2 and Molecular Network in the extracts present used to elaborate the MIX sample – positive ionization mode.

(continuation)

#	Rt (min)	Species	Molecular Formula	$[M+H]^+$ observed	$[M+H]^+$ calculated	Error (ppm)	MS/MS fragments (% abundance)	Metabolite
29	11.22	*P. densiflora*	$C_{33}H_{41}O_{20}$	757.2161	757.2191	-4.0	287.0558 (100)	Kaempferol-*O*-hexoside-hexoside-deoxyhexoside
30	11.77	*B. harleyi*	$C_{27}H_{31}O_{16}$	611.1612	611.1612	0.0	303.0515 (100)	Quercetin-*O*-hexoside-deoxyhexoside
31	11.92	*P. densiflora*	$C_{33}H_{41}O_{19}$	741.2241	741.2242	-0.1	287.0565 (100)	Kaempferol-*O*-hexoside-deoxyhexoside-deoxyhexoside
31	11.93	*B. harleyi*	$C_{33}H_{41}O_{19}$	741.2224	741.2242	-2.4	287.0559 (100)	Kaempferol-*O*-hexoside-deoxyhexoside-deoxyhexoside
32	12.01	*B. harleyi*	$C_{21}H_{21}O_{12}$	465.1036	465.1033	0.6	303.0513 (100)	Quercetin-*O*-hexoside
33	12.03	*M. bahiana*	$C_{27}H_{31}O_{16}$	611.1608	611.1612	-0.7	303.0517 (100)	Quercetin-*O*-hexoside-deoxyhexoside
34	12.04	*N. multiglandulosa*	$C_{32}H_{39}O_{19}$	727.2078	727.2086	-0.8	287.0565 (100)	Kaempferol-*O*-hexoside-deoxyhexoside-pentoside
35	12.05	*H. restingae*	$C_{28}H_{25}O_{16}$	617.1136	617.1143	-1.1	303.0513 (90), 297.0614 (15), 171.0312 (10), 153.0197 (100)	Quercetin-*O*-galloyl-hexoside
36	12.06	*B. harleyi*	$C_{27}H_{31}O_{16}$	611.1617	611.1612	0.8	303.0515 (100)	Quercetin-*O*-hexoside-deoxyhexoside
37	12.16	*H. restingae*	$C_{28}H_{25}O_{16}$	617.1126	617.1143	-2.8	303.0510 (100), 153,0196 (95)	Quercetin-*O*-galloyl-hexoside
38	12.27	*B. intermedia*	$C_{21}H_{21}O_{12}$	465.1026	465.1033	-1.5	303.0508 (100)	Quercetin-*O*-hexoside
38	12.27	*M. bahiana*	$C_{21}H_{21}O_{12}$	465.1038	465.1033	1.1	303.0514 (100)	Quercetin-*O*-hexoside
39	12.61	*N. multiglandulosa*	$C_{33}H_{41}O_{20}$	757.2170	757.2191	-2.8	317.0662 (100)	Methoxy-quercetin-*O*-hexoside-deoxyhexoside-pentoside

Table 3 – Metabolites annotated by UHPLC-ESI-MS^2 and Molecular Network in the extracts present used to elaborate the MIX sample – positive ionization mode.

(continuation)

#	Rt (min)	Species	Molecular Formula	$[M+H]^+$ observed	$[M+H]^+$ calculated	Error (ppm)	MS/MS fragments (% abundance)	Metabolite
40	12.66	*P. densiflora*	$C_{50}H_{61}O_{29}$	1125.3295	1125.3299	-0.4	287.0564 (50), 207.0675 (100), 175.0408 (20)	Kaempferol-*O*-hexoside-hexoside-hexoside-deoxyhexoside-dimethoxyferulic acid
41	13.02	*B. intermedia*	$C_{28}H_{25}O_{16}$	617.1155	617.1143	1.9	303.0524 (30), 233.0470 (25), 205.0526 (20), 153.0203 (100)	Quercetin-*O*-galloyl-hexoside
42	13.41	*N. multiglandulosa*	$C_{28}H_{33}O_{16}$	625.1757	625.1769	-1.9	317.0670 (100)	Methoxy-quercetin-*O*-hexoside-deoxyhexoside
43	13.43	*B. harleyi*	$C_{26}H_{29}O_{15}$	581.1502	581.1506	-0.7	303.0516 (1000	Quercetin-*O*-deoxyhexoside-pentoside
44	13.44	*B. harleyi*	$C_{20}H_{19}O_{11}$	435.0935	435.0927	1.8	303.0506 (100)	Quercetin-*O*-pentoside
45	13.52	*B. harleyi*	$C_{27}H_{31}O_{15}$	595.1666	595.1663	0.5	287.0568 (100)	Kaempferol-*O*-hexoside-deoxyhexoside
46	13.52	*P. densiflora*	$C_{49}H_{59}O_{28}$	1095.3199	1095.3193	0.5	287,0564 (50), 177.0569 (100), 145.0292 (10)	Kaempferol-*O*-hexoside-hexoside-hexoside-deoxyhexoside-methoxycaffeic acid
47	13.53	*A. septentrionalis*	$C_{27}H_{31}O_{15}$	595.1670	595.1663	1.2	287.0562 (100)	Kaempferol-*O*-hexoside-deoxyhexoside
48	13.65	*B. intermedia*	$C_{20}H_{19}O_{11}$	435.0933	435.0927	1.4	303.0513 (100)	Quercetin-*O*-pentoside
49	13.70	*B. harleyi*	$C_{28}H_{33}O_{16}$	625.1766	625.1769	-0.5	317.0670 (100)	Methoxy-quercetin-*O*-hexoside-deoxyhexoside
50	14.06	*P. densiflora*	$C_{48}H_{57}O_{27}$	1065.3075	1065.3087	-1.1	287.0571 (80), 147.0458 (100)	Kaempferol-*O*-hexoside-hexoside-hexoside-deoxyhexoside-coumaric acid

Table 3 – Metabolites annotated by UHPLC-ESI-MS^2 and Molecular Network in the extracts present used to elaborate the MIX sample – positive ionization mode.

(continuation)

#	Rt (min)	Species	Molecular Formula	$[M+H]^+$ observed	$[M+H]^+$ calculated	Error (ppm)	MS/MS fragments (% abundance)	Metabolite
51	14.10	*B. laevifolia*	$C_{22}H_{23}O_{12}$	479.1192	479.1190	0.4	317.0664 (100)	Methoxy-quercetin-*O*-hexoside
52	14.32	*B. laevifolia*	$C_{21}H_{21}O_{11}$	449.1085	449.1084	0.2	303.0512 (100)	Quercetin-*O*-deoxyhexoside
52	14.32	*M. bahiana*	$C_{21}H_{21}O_{11}$	449.1088	449.1084	0.9	303.0513 (100)	Quercetin-*O*-deoxyhexoside
53	14.54	*B. laevifolia*	$C_{22}H_{21}O_{13}$	493.0980	493.0982	-0.4	317.0673 (100)	Methoxy-quercetin-*O*-glucuronic acid
54	14.96	*B. harleyi*	$C_{20}H_{19}O_{10}$	419.0980	419.0978	0.5	287.0564 (100)	Kaempferol-*O*-pentoside
55	14.99	*B. harleyi*	$C_{26}H_{29}O_{14}$	565.1560	565.1557	0.5	287.0568 (100)	Kaempferol-*O*-deoxyhexoside-pentoside
56	15.40	*B. intermedia*	$C_{27}H_{23}O_{15}$	587.1040	587.1037	0.5	303.0490 (10), 267.0518 (35), 249.0437 (20), 231.0308 (20), 207.0311 (45), 205.0518 (35), 154.0200 (100)	Quercetin-*O*-galloyl-pentoside
57	15.60	*B. laevifolia*	$C_{25}H_{25}O_{15}$	565.1213	565.1193	3.5	317.0667 (100)	Methoxy-quercetin-*O*-malonyl-hexoside
58	16.33	*B. laevifolia*	$C_{15}H_{11}O_{17}$	463.1250	462.9996	0.9	317.0670 (100)	Methoxy-quercetin-*O*-hexoside
59	17.88	*M. bahiana*	$C_{28}H_{25}O_{15}$	601.1186	601.1193	-1.2	303.0513 (10), 153.0201 (100)	Quercetin-*O*-galloyl-deoxyhexoside
60	18.30	*A. septentrionalis*	$C_{24}H_{25}O_{13}$	521.1296	521.1295	0.2	273.0774 (100)	Naringenin-*O*-malonyl-hexoside
61	18.58	*M. bahiana*	$C_{28}H_{25}O_{15}$	601.1190	601.1193	-0.5	303.0508 (10), 153.0191 (100)	Quercetin-*O*-galloyl-deoxyhexoside
62	18.58	*M. bahiana*	$C_{13}H_{15}O_8$	299.0771	299.0767	1.3	153.0200 (100)	Not identified

Table 3 – Metabolites annotated by UHPLC-ESI-MS^2 and Molecular Network in the extracts present used to elaborate the MIX sample – positive ionization mode.

(conclusion)

#	Rt (min)	Species	Molecular Formula	$[M+H]^+$ observed	$[M+H]^+$ calculated	Error (ppm)	MS/MS fragments (% abundance)	Metabolite
63	18.58	*M. bahiana*	$C_{13}H_{13}O_7$	281.0666	281.0661	1.8	280.0967 (50), 153.0203 (100)	Not identified

Source: elaborated by the Author (2022).

Specifically for *P. densiflora*, a small cluster into the phenolics network characterized by long chains of sugars connected with coumaric acid (**50**, *m/z* 1065.3075 [M+H]+), methoxycaffeic acid (**46**, *m/z* 1095.3199 [M+H]+) and dimethoxyferulic acid (**40**, *m/z* 1125.3295 [M+H]+) was observed. This is the first report of this subclass of compounds for the Malpighiaceae family, besides being the first chemical study of *Ptilochaeta* genus. These subclasses of compounds containing *p*-coumaric acid and methoxycaffeic acid substituents were previously reported in species from Brassicaceae, Theaceae, and Amaranthaceae families (ABD EL-LATIF *et al*., 2014; CORRADI *et al*., 2016; LEE; LEE; LEE, 2012; YOSHIKAWA *et al*., 2008), however, a dimethoxyferulic acid as substituent has not been described to date. Considering that these three nodes cluster together in the network and present the same MS/MS spectra (an intense ion corresponding to the aglycone – kaempferol), we propose compound **40** for the first time, since the mass shift of 30 Da between **40** and **46** can indicate the presence of an additional methoxy group in the substituent (Supplementary Figure 2). Further studies are necessary to unequivocally identify this compound.

Malonylated flavonoid glycosides were also observed in the phenolic cluster, specifically present in *B. laevifolia* (**57**, *m/z* 565.1213 $[M+H]^+$) and *A. septentrionalis* (**60**, *m/z* 521.1296 $[M+H]^+$), in which both present a neutral loss of 248 Da corresponding to a hexose (162 Da) bonded to a malonyl (86 Da) group (YANG *et al*., 2016). Compound **57** has been previously described for *Pyrus* spp. (LIN; HARNLY, 2008; WALD *et al*., 1989), berries plants (TIAN *et al*., 2017, 2019), *Ambrosia artemisiifolia* (BOŽIČEVIĆ *et al*., 2017), *Fagonia cretica* (NAZIR *et al*., 2017), and *Phyteuma orbiculare* (ABBET *et al*., 2013). To the best of our knowledge, **60** has not been reported to date. Considering that this compound presents the same neutral loss as **57**, we propose that the hexose bonded to the malonyl group are also present in **60**, being naringenin the most likely aglycone due to the MS/MS fragment at *m/z* 273.0774. This is the first report of this subclass of compounds in the Malpighiaceae family.

Another small cluster into the phenolics network was observed, characterized by quinic and shikimic acid derivatives containing galloyl as substituents, observed for *B. intermedia* and *M. bahiana* species. These compounds could be determined based on the presence of *m/z* 153.0201 fragments, corresponding to the galloyl substituent. Quinic acid derivatives have been previously reported for other *Byrsonima* species

(FRAIGE *et al.*, 2018). On the other hand, chemical studies of *Mcvaughia* species are scarce, being this the first chemical report of this genus. Three nodes also stand out for representing compounds annotated as quercetin derivatives bonded to galloyl as one of the substituents (which could be inferred by the presence of fragments at *m/z* 303.0515 and *m/z* 153.0201 of compounds **35**, **37**, **56**, and **59**), which were detected in *B. intermedia*, *M. bahiana*, and *H. restingae.*

C-glycosylated flavonoids were observed as single nodes retrieved from *H. restingae* (**16**) and *A. septentrionalis* (**9**) species. Differently from the *O*-glycosylated flavonoids (which usually present a neutral loss of the whole glycosidic portion), the fragmentation pattern of *C*-glycosylated flavonoids presents much more fragments between the protonated compound and the aglycone. For instance, compound **16** (*m/z* 579.1713 $[M+H]^+$) presented fragment *m/z* 433.1136 $[M+H-146]^+$, indicating the neutral loss of a deoxyhexoside; however, the loss of the hexoside portion was observed as many fragments to reach the aglycone ion (*m/z* 271.0587), including consecutive losses of water molecules and 120 Da, characteristic of *C*-glycosylated flavonoids. On the other hand, compound **9** (*m/z* 595.1663 $[M+H]^+$) presented two hexosides connected to the genistein in different positions by a C-C bound, which has been reported in *Dalbergia* spp. (NUNES; HAAG; BESTMANN, 1989; VAN HEERDEN; BRANDT; ROUX, 1980) and *Linus reflexus* (STOBIECKI *et al*., 2010). *C*-glycosylated flavonoids have been previously described for species of the Malpighiaceae family, including *Hiraea reclinata* species (HUSSEIN *et al*., 2003).

Other polyphenols were also observed, such as dimers (for *B. intermedia*, *N. multiglandulosa*, *B. laevifolia*, and *A. septentrionalis*) and trimers (for *N. multiglandulosa* and *B. laevifolia*) of proanthocyanidins, also named as condensed tannins. In this class of compounds, it is possible to clearly observe fragments resulting from the retro Diels-Alder (RDA) reaction, yielding neutral losses of 152 Da ($C_8H_8O_3$) (AABY; EKEBERG; SKREDE, 2007; FRAIGE *et al.*, 2018; RUE; RUSH; VAN BREEMEN, 2018). The fragments observed for the dimers consisted mainly of ions at *m/z* 409 and *m/z* 275, corresponding to the neutral loss of a water molecule with one RDA, and two RDA reactions, respectively. Similarly, the fragments observed for the trimers, *m/z* 545 and *m/z* 409, corresponded to the loss of water along with two RDA reactions, and to the loss of a catechin portion, along with a water molecule and RDA reaction, respectively.

Additionally, monomers, such as catechin (detected in *B. intermedia*, *N. multiglandulosa*, *B. laevifolia*, and *A. septentrionalis*) and afzelechin (only observed in *A. septentrionalis*) were detected, and the MS/MS spectra of both compounds presented a base peak at *m/z* 139, also corresponding to RDA reactions (SPÁČIL; NOVÁKOVÁ; SOLICH, 2010). These compounds are largely found in plants, including the Malpighiaceae family (FRAIGE *et al.*, 2018; HUERTA-REYES; FONSECA JUÁREZ; AGUILAR-ROJAS, 2015).

A molecular family characterized only by nodes representing *N. multiglandulosa* species was also retrieved, and corresponds to the presence of phytoecdysteroids that were confirmed by comparison with standards previously isolated from this species in our laboratory (MANNOCHIO RUSSO *et al.*, 2020): 5-hydroxypodecdysone B (**23**), integristerone A (**24**), podecdysone B (**25**), and ecdysterone (**27**). This class of compounds has only been recently described for the *Niedenzuella* genus in the Malpighiaceae botanical family.

It should be mentioned that some ions observed in this study were not identified by GNPS or by search on DNP, highlighting the need for future studies with a greater number of species to verify the possibility of new molecular families from Malpighiaceae species. In this context, the chromatographic method developed in this work could be of great use for future investigations.

4 Conclusions

In conclusion, a chromatographic method was developed with the assistance of commercial automated statistical evaluation tools and a mixture of samples, which allowed it to be performed in only four days. A significant increase in the responses selected was obtained, such as an increase of 57% in the number of total peaks, which is of extreme importance in natural products research. Besides, the development of an appropriate method for the chromatographic analysis also enabled the detection of minor compounds, which could be challenging if traditional procedures had been used. Therefore, it can be considered a promising approach for chromatographic method development in natural products workflows. The case study performed with Malpighiaceae samples enabled the separation and detection of different classes of compounds, such as *O*-glycosylated flavonoids, *C*-glycosylated flavonoids, quinic acid derivatives, steroids, among others. Additionally, many compounds were described for the first time in this family, which can contribute to a better understanding of these

compounds’ distribution. We hope that this study will encourage the use of such tools for chromatographic method development even more, and that the developed method can assist in future Malpighiaceae studies.

Adapted with permission from (MANNOCHIO-RUSSO *et al.* 2020). Copyright 2020 American Chemical Society.

03/05/2022 11:14 Rightslink® by Copyright Clearance Center

RightsLink®

Home | Help | Live Chat | Sign in | Create Account

ACS Publications
Most Trusted. Most Cited. Most Read.

Can Statistical Evaluation Tools for Chromatographic Method Development Assist in the Natural Products Workflow? A Case Study on Selected Species of the Plant Family Malpighiaceae

Author:
Helena Mannochio-Russo, Paula Carolina P. Bueno, Anelize Bauermeister, et al

Publication: Journal of Natural Products

Publisher: American Chemical Society

Date: Nov 1, 2020

Copyright © 2020, American Chemical Society

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

BACK CLOSE WINDOW

© 2022 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

https://s100.copyright.com/AppDispatchServlet#formTop 1/1

CHAPTER 2 – CHEMOSYSTEMATICS STUDIES OF MALPIGHIACEAE BOTANICAL FAMILY THROUGH THE USE OF SPECTRAL LIBRARY SEARCHES AND *IN SILICO* TOOLS FOR METABOLITES ANNOTATION

Article title: Untargeted metabolomics sheds light on the diversity of major classes of secondary metabolites in the Malpighiaceae botanical family (doi: 10.3389/fpls.2022.854842)

Journal: Frontiers in Plant Science

Status: Published

Authors: Helena Mannochio-Russo[1,2], Rafael F. De Almeida[3,4], Wilhan D. Gonçalves Nunes[5], Paula C. Pires Bueno[6,7], Andrés M. Caraballo-Rodríguez[2], Anelize Bauermeister[2], Pieter Dorrestein[2], Vanderlan Da Silva Bolzani[1]

Affiliations:

[1]NuBBE, Department of Biochemistry and Organic Chemistry, Institute of Chemistry, São Paulo State University (UNESP), 14800-901 Araraquara, SP, Brazil.

[2]Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California 92093, United States.

[3]Royal Botanical Gardens Kew, Science, Ecosystem Stewardship, Diversity and Livelihoods, Richmond, Surrey, TW9 3AE, United Kingdom.

[4]Department of Biological Sciences, Lamol Lab, Feira de Santana State University (UEFS), Feira de Santana, BA 44036-900, Brazil.

[5]Federal Institute of Education, Science and Technology of Rondônia (IFRO), Ji-Paraná campus, Ji-Paraná, RO, Brazil.

[6]Max Planck Institute of Molecular Plant Physiology, 14476, Potsdam-Golm, Germany.

[7]Institute of Chemistry, Federal University of Alfenas (UNIFAL), 37130-001, Alfenas, MG, Brazil.

Abstract

Plant metabolites are among the most studied natural products, with thousands of reports describing new chemical structures and valuable biological activities. However, the study of phylogenetic diversification and distribution patterns of plant secondary metabolites is still in its early steps, and several botanical families have not been studied in this context. Highly sensitive hyphenated detection techniques, such as Liquid Chromatography-Mass Spectrometry (LC-MS), along with bioinformatics tools for the classification of metabolites, have proven to be of great value in chemosystematics investigations. In this work, our goals were to perform a chemosystematic study of the Malpighiaceae botanical family using MS/MS untargeted metabolomics, spectral searches, and recently described *in silico* classification tools. For such, a collection of 137 species of Malpighiaceae comprising all the major phylogenetic clades currently accepted for the family was employed. Our results show that different ionization modes and extraction protocols significantly impacted the chemical profiles, influencing the chemotaxonomic results. Molecular Networking and spectral searches within public databases revealed several clades or genera-specific molecular families, being potential chemical markers for these taxa. The *in silico* classification tools expanded the Malpighiaceae chemical space, and the classes putatively annotated were used for ancestral character reconstructions, which provided classes of metabolites identified as homoplastic (i.e., non-exclusive) or synapomorphic (i.e., exclusive) for clades and genera. The workflow described herein combines several recently developed tools to perform a comprehensive chemotaxonomic study, which will be useful for future studies in Malpighiaceae and other chemoevolutionary studies in other taxa.

Keywords: ancestral character reconstruction; chemotaxonomy; evolution; mass spectrometry; metabolite annotation; metabolomics.

1 Introduction

Secondary metabolites, also known as natural products, are compounds produced by bacteria, fungi, and plants, which can be extremely complex due to the large number of biosynthetic pathways involved in their production. The chemical diversity of plants has been widely explored from the 1930s until nowadays, mainly aiming to study plants with ethnopharmacological value, bioactive extracts and compounds, and new chemical structures (ATANASOV *et al*., 2021; NEWMAN; CRAGG, 2020). Due to the development of highly sensitive detection techniques, such as mass spectrometry (MS), it is now possible to explore plant extracts as a whole and have a more comprehensive overview of the metabolites biosynthesized by plant species (DE VOS *et al*., 2007; GEMPERLINE; KELLER; LI, 2016; LI; GAQUEREL, 2021).

In this context, chemosystematics studies employing MS and bioinformatic tools for exploring big plant datasets have been successfully performed, raising valuable insights regarding the biosynthetic pathways involved in different phylogenetic groups (KANG *et al*., 2019). It is now possible to support phylogenetic studies at a molecular level and assist in elaborating evolutionary hypotheses based on metabolomics analyses. In fact, the rapid development of natural product bioinformatics tools and databases (MEDEMA, 2021) can be of great value to assist and accelerate more comprehensive chemosystematics studies.

Even though evolutionary studies at a metabolite level are of great interest, certain factors can greatly impact the detection of metabolites and, consequently, the conclusions drawn. For instance, the extraction solvent used and the ionization mode in MS can prioritize one class of compounds over another (CREYDT; FISCHER, 2017; FLOROS *et al*., 2016), leading to biased results. However, the literature still lacks investigations of these factors in chemosystematics studies. Therefore, to evaluate these variables, a plant family with a broad diversity of compounds classes identified, in addition to extensive previous DNA-based phylogenetic studies, would be ideal for a deeper chemotaxonomic investigation.

The Malpighiaceae plant family is an excellent example for a deeper chemical investigation due to the high diversity of secondary metabolites produced by these plant species, besides being a taxon of flowering plants with all its genera sampled in molecular phylogenetic studies (DAVIS; ANDERSON, 2010). Malpighiaceae is one of the 36 families of flowering plants placed in Malpighiales, one of the most important

and diverse orders of angiosperms in tropical forests (CAI *et al*., 2016; XI *et al*., 2012). This family currently comprises 70 genera and ca. 1,300 species mostly confined to the American tropics, being Brazil its most representative country, with few genera and species reaching the tropics of Africa, Asia, and Oceania (ALMEIDA *et al.*, 2021; ALMEIDA; VAN DEN BERG, 2021; DAVIS; ANDERSON, 2010). Some Amazonian species of Malpighiaceae are traditionally known for their psychedelic or aphrodisiac properties, with several studies focusing on the chemical characterization of these species (QUEIROZ *et al*., 2014, 2015; SAMOYLENKO *et al.*, 2010). On the other hand, several extra-Amazonian Malpighiaceae species are long reported as toxic to cattle (such as *Amorimia*, *Niedenzuella,* and *Tetrapterys spp.*) (LEE *et al*., 2014; RIET-CORREA; MEDEIROS; SCHILD, 2012), causing significant economic losses in the growing cattle industry of Brazil. Additionally, *Malpighia* and *Byrsonima* spp. also stand out for the nutritional value of their fruits (BELWAL *et al*., 2018; NERI-NUMA *et al*., 2018).

In the past two decades, Malpighiaceae has gone through unparalleled changes in its traditional classification as a direct result of several molecular phylogenetic studies (ALMEIDA *et al.*, 2017; ALMEIDA; AMORIM; VAN DEN BERG, 2018; CAMERON *et al*., 2001; DAVIS; ANDERSON, 2010; DAVIS; ANDERSON; DONOGHUE, 2001). Key morphological characters of its traditional classification system (*i.e.*, fruit types) were recovered as highly homoplastic (CAMERON *et al*., 2001; DAVIS; ANDERSON; DONOGHUE, 2001). The inevitable recognition of unforeseen relationships within Malpighiaceae brought to light deep taxonomic problems regarding the monophyly of several genera [*e.g.*, *Banisteriopsis* C.R.Rob., *Mascagnia* (Bertero ex DC.) Bertero, *Stigmaphyllon* A.Juss., and *Tetrapterys* Cav.], tribes (*e.g.*, just Gaudichaudieae Horan. was recovered as monophyletic), and all its subfamilies (*e.g.*, Byrsonimoideae W.R.Anderson, Banisterioideae A.Juss. and Malpighioideae A.Juss.) (ALMEIDA *et al.*, 2017; CAMERON *et al.*, 2001; DAVIS; ANDERSON, 2010; DAVIS; ANDERSON; DONOGHUE, 2001). Since then, different authors have gradually proposed new genera and combinations to accommodate these newly identified lineages (ALMEIDA; VAN DEN BERG, 2021; ANDERSON, 2011, 2006; ANDERSON; DAVIS, 2006; DAVIS; MARINHO; AMORIM, 2020). Even though morphological characters were used to reconstruct the last generic phylogeny of Malpighiaceae, no morphological synapomorphies were ever recovered or discussed for all its major clades (ALMEIDA; VAN DEN BERG, 2021). As a result, its traditional classification was rejected, and

several informal clades, without any morphological circumscription, were recognized in the most recent generic phylogeny for Malpighiaceae (DAVIS; ANDERSON, 2010).

In this study, we propose a methodology for chemosystematics studies, in which we performed metabolomics analysis in combination with *in silico* fragmentation predictors, chemical hierarchy analysis, and ancestral character reconstructions to compare the clustering patterns observed in multivariate analysis and classifications retrieved (Figure 13). For this, we evaluated a unique collection of Malpighiaceae samples, comprising 39 genera (out of 74) and 139 species from each of the major phylogenetic groups currently accepted for the family. Different extraction solvents and ionization modes were also assessed. By this approach, we were able to provide insights regarding the classes of compounds detected in distinct extraction protocols and different ionization modes, give the first chemical information of several Malpighiaceae genera, besides drawing conclusions regarding the evolution of biosynthetic pathways in the Malpighiaceae family.

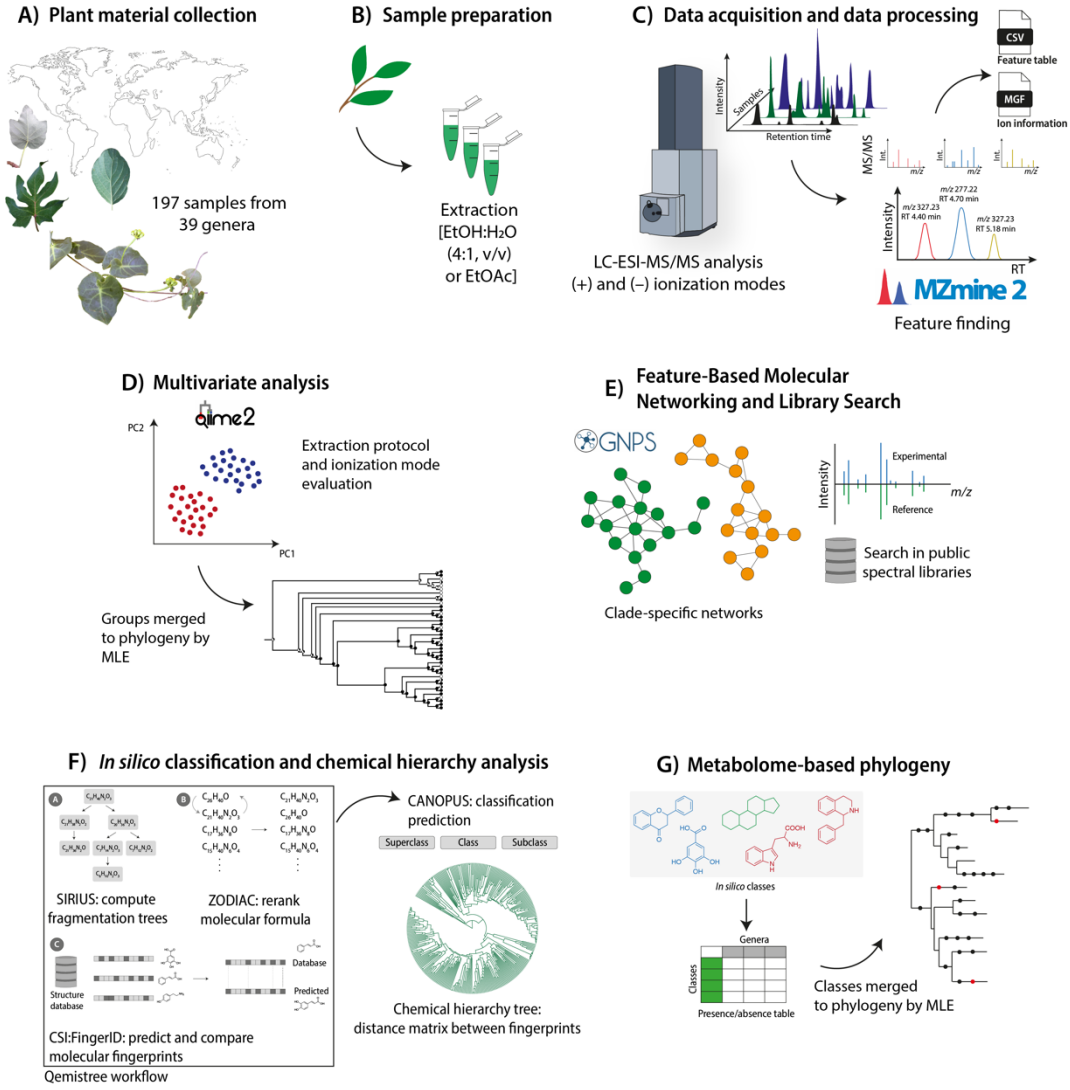


Figure 13 – Experimental workflow followed for the metabolomics and chemosystematics analyses of Malpighiaceae samples. (1) The samples were initially collected, (2) the extracts were prepared with different solvents [EtOH–H_2O (4:1, v/v) or EtOAc], and then (3) subjected to LC-ESI-MS/MS analysis in positive and negative ionization modes in an untargeted method. (4) The data acquired were processed for feature finding, and the exported data were used for multivariate analysis. The clustering groups observed were merged to the phylogeny using the Maximum Likelihood Estimation (MLE) for preliminary chemotaxonomic investigations. (5) The data was also used for Feature-Based Molecular Networking and library searches workflows to observe clade-specific molecular families. (6) A chemical hierarchy analysis and in silico classifications were obtained and finally (7) merged to the currently accepted Malpighiaceae phylogeny to determine the ubiquitous and the taxa-specific *in silico* classes.

Source: elaborated by the Author (2022).

2 Materials and Methods

2.1 General information

The ethanol and ethyl acetate used for the extraction procedures were obtained from Sigma-Aldrich (St. Louis, USA) and J.T. Baker (J.T. Baker-Avantor, Radnor, USA), respectively. Acetonitrile and water, both LC-MS grade, were obtained from Fisher Scientific (Fair Lawn, NJ, USA).

2.2 Plant material

Most plant samples were collected by R.F. Almeida on field expeditions throughout Brazil from 2013 to 2017 or were retrieved from discarded fragmented samples used for DNA extraction in molecular studies (ALMEIDA *et al*., 2017; ALMEIDA; AMORIM; VAN DEN BERG, 2018; ALMEIDA; VAN DEN BERG, 2020, 2021). All the information regorging the samples can be found in Supplementary Box 1. The plant materials were dried in a desiccator containing silica at room temperature. Samples were then frozen in liquid nitrogen and grounded in a ball mill. The samples were stored in a freezer at -20 ˚C until the preparation of the extracts.

The authorization for conducting this study was conceded by the National System for Management of Genetic Heritage and Associated Traditional Knowledge (SISGEN), n˚ A6FDC2E.

2.3 Extraction procedure

The extracts were prepared in a proportion of 20 mg of plant material to 1 mL of extraction solvent [EtOH–H_2O (4:1, v/v) or EtOAc (100%)]. The samples were homogenized in a Qiagen TissueLyzer II (Qiagen, Hilden, Germany) for 5 min at 25 MHz and extracted at room temperature for an additional 30 min. The samples were centrifuged (5,000 *g*) for 15 min, and 300 µL of the supernatants were transferred to a 96-deep-well plate. The solvent was dried in a Labconco CentriVap (USA), and the plates were sealed and stored at –80 ˚C prior to analyses.

2.4 UHPLC-MS/MS analysis

The extracts were initially resuspended in 200 µL MeOH–H_2O (4:1) containing sulfachloropyridazine (2 µM) as internal standard, and sonicated for 15 min. The plates were centrifuged for 10 min at 1,300 *g*, and the supernatants were then transferred to a new 96-well plate for metabolomics analyses.

The analyses were carried out with a Thermo Scientific UltiMate 3000 UHPLC system coupled to a Maxis Impact QTOF mass spectrometer (Bruker Daltonics, Germany), controlled by the Otof Control and Hystar software packages, and equipped with an ESI (electrospray ionization) source. The extracts were analyzed using a Kinetex 1.7 µm C18 reversed-phase UHPLC column (50 X 2.1 mm) (Phenomenex, Torrance, CA, USA), at 40 ˚C, and an injection volume of 5 µL. The pump system consisted of water (A) and acetonitrile (B), both acidified with formic acid (0.1%, v/v), and the flow rate was set at 0.5 mL/min. The metabolites separation was achieved with 5% solvent B for 1 min, followed by a linear gradient from 5% to 100% in 5 min. The column was washed at 100% solvent B for 2 min, then returned to the initial 5% in 1 min, and the equilibration of the column was achieved for 1 min at 5% solvent B. The mass spectra were acquired in both positive and negative ionization modes, separately, in a 50–2000 Da mass range in the data-dependent acquisition (DDA) mode. The parameters used for data acquisition were set as follows: nitrogen used as nebulizer gas with pressure at 2 Bar, a capillary voltage of 4200 V, ion source temperature of 200 °C, dry gas flow at 9 L/min, and spectra rate acquisition of 3 spectra/s. The 5 most intense selected ions per spectrum were fragmented (MS/MS) using ramped collision-induced dissociation energy, ranging from 22 to 50 eV. MS/MS active exclusion was set after 5 spectra and released after 30 s.

The UHPLC-MS/MS data were deposited in the MassIVE Public GNPS dataset (http://massive.ucsd.edu, MSV000085119) and are publicly available.

2.5 Mass spectrometry data pre-processing

The raw data files (.d) were converted to .mzXML format using DataAnalysis® software (Bruker) after lock mass correction using hexakis (1H, 1H, 2H-difluoroethoxy) phosphazene (Synquest Laboratories, Alachua, FL), with *m/z* 622.029509. The quality

of the analyses was evaluated considering the retention time and the *m/z* of a standard solution containing a mixture of six standards, which was analyzed every 11 injections.

The .mzXML files were processed in MZmine2 (version 2.37.corr17.7_kai_merge2) in the positive and negative ionization modes, separately. The parameters used for feature finding were as follows:

- mass detection (centroid, 1.0E3 and 1.0E1 for MS1 and MS2, respectively);
- chromatogram builder (minimum time span of 0.01 min, minimum height of 3.0E3, and *m/z* tolerance of 20 ppm);
- chromatogram deconvolution (baseline cut-off algorithm: minimum peak height: 1.0E3, peak duration range: 0.01 to 3 min; baseline level: 1.0E3) with median *m/z* center calculation, *m/z* range for MS2 scan pairing of 0.02 Da, and retention time [RT] range for MS2 scan pairing of 0.1 min;
- isotope peaks grouper (*m/z* tolerance set at 20 ppm, RT tolerance of 0.1 min, maximum charge of 3, and representative isotope set to most intense), join alignment *(m/z* tolerance of 20 ppm, weight for m/z and RT of 75 and 25, respectively, and RT tolerance of 0.1 min).

A filter was applied to keep only the features with MS/MS spectra. This feature list was exported as a feature quantification table (.csv), as an MS2 spectral summary (.mgf), and with the SIRIUS export module (.mgf) for downstream analyses. The exporter files can also be found in the MassIVE repository (MSV000085119).

2.6 Feature-Based Molecular Networking (FBMN) workflow

The processed LC-MS/MS data (.mgf and .csv) were used to create a Feature-Based Molecular Network (NOTHIAS *et al*., 2020) on the GNPS platform (WANG *et al*., 2016) with input files containing only the features detected in the hydroethanolic extracts. The data were filtered by removing all MS/MS fragment ions within +/- 17 Da of the precursor ion. Additionally, MS/MS spectra were window filtered to select only the top six fragment ions in the +/- 50 Da window throughout the spectrum. The precursor ion and the MS/MS fragment ion tolerance were set to 0.02 Da. A molecular network was created, in which the edges were filtered to have a cosine score above 0.7 and at least four matched peaks. Similarly, the parameters for the library search (for comparison between the experimental and library spectra) were set to have a

score above 0.7 and at least four matched peaks to assist in the metabolites' annotation - level three according to the metabolomics standards initiative (SUMNER *et al.*, 2007). The FBMN jobs on GNPS can be found at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=2c5f11403ac847a298e4d7866a491143 (positive mode) and https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=501c16500476451f978311057266fbdf (negative mode).

The molecular network visualization was performed in Cytoscape (version 3.7.2, Cytoscape consortium, San Diego, CA, USA) (SHANNON *et al.*, 2003), in which the nodes correspond to ion features, while the edges between the nodes represent the MS/MS cosine scores calculated. Nodes corresponding to features present in the blank analyses were excluded to avoid misinterpretations due to contaminants in the analyses. Sample type information was added to color the nodes as pie charts representing the relative abundance of the features across the samples (colors based on the phylogenetic clades A–J). Node size was scaled relative to the sum of the peak areas obtained in the samples where the feature was detected. Compounds with identical MS/MS spectra but different retention times were represented as separate nodes, indicating isomers. The Cytoscape visualization files can be accessed in MSV000085119 (MassIVE).

2.7 Chemical hierarchy analysis using Qemistree

A chemical hierarchy analysis (Qemistree) (TRIPATHI *et al.*, 2021) was performed with the metabolites detected in the hydroethanolic extraction protocol. For this, the q2-qemistree qiime2 plugin was used (https://github.com/biocore/q2-qemistree), in which the feature quantification table (.csv) and the file obtained from the SIRIUS (DÜHRKOP *et al.*, 2019) (export module (.mgf) exported from MZmine were used as input. Briefly, the Qemistree workflow applies SIRIUS (version 4.8.2) to the .mgf file (containing ion information), generating predicted molecular formulas for each feature. The predicted molecular formulas were reranked using ZODIAC (LUDWIG *et al.*, 2020), and the predicted molecular fingerprints were subsequently generated using fragmentation trees via CSI:FingerID (DÜHRKOP *et al.*, 2015). The chemical taxonomy of the predicted metabolite structures was obtained by CANOPUS (superclass, class, and subclass) (DÜHRKOP *et al.*, 2021). The Euclidean pairwise

distances between the molecular fingerprints were calculated, and the fingerprint vectors were hierarchically clustered to generate a tree representing the structural chemical relationships of this dataset. The tree was then pruned in order to keep only the fingerprints classified up to a superclass level. The trees were visualized interactively in EMPRESS (CANTRELL *et al.*, 2021) for data exploration, in which clade information was added as relative abundance stacked bar charts to each feature. The dendrogram obtained can be interactively visualized with the .qzv files found in the Github repository (https://github.com/helenamrusso/Malpighiaceae_supplementary). The files generated in the Qemistree workflow in Qiime2 are also available in MSV000085119

2.8 Statistical analysis

The feature table exported from MZmine was used to perform unsupervised analysis using Qiime 2 (version 2020.2) (BOLYEN *et al.*, 2019) bioinformatics pipeline within a Jupyter notebook. The metabolomic profiles of the different extraction protocols were compared using the Bray–Curtis distance metric, while the Canberra distance metric was used to investigate the four subsets individually (two extraction solvents and two ionization modes). The Principal Coordinates Analysis (PCoA) plots showing the top three principal coordinates were visualized using EMPeror (VÁZQUEZ-BAEZA *et al.*, 2013). Permutational multivariate analysis of variance (PERMANOVA) (ANDERSON, 2001) was also performed in Qiime 2 on metabolite distance matrices to test for clustering significance (with 999 permutations), and the *F* statistic was reported as a measure of effect size.

2.9 Phylogenetic analyses

In total, 39 genera (out of 74) and 137 species (out of 1,300) of Malpighiaceae were sampled, representing all of its ten major clades recognized by recent molecular phylogenetic studies Supplementary Box 1. Since our sampling focused on the diversity of genera and clades within the Malpighiaceae family, genera not sampled on the chemical analyses were not included in the molecular phylogenetic analysis. Sequences for the genes *matK*, *ndhF*, PHYC, and *rbcL* were retrieved from GenBank

(https://www.ncbi.nlm.nih.gov/genbank/), edited using Geneious (KEARSE *et al*., 2012), and aligned using Muscle (EDGAR, 2004), with subsequent adjustments in the preliminary matrices to ensure that the nitrogenous bases were correctly aligned. The complete data matrices are available at TreeBase (https://www.treebase.org/treebase-web/search/studySearch.html, accession number S11008).

All trees were rooted in the Byrsonimoid clade, which is considered the sister-group of the other Malpighiaceae clades (DAVIS; ANDERSON, 2010). Combined analysis of plastid + nuclear regions was performed using Bayesian inference and Maximum Likelihood criteria to reconstruct our phylogenetic hypotheses. Both model-based methods were conducted with a mixed substitution model (GTR+G+I) and unlinked parameters, using MrBayes 3.1.2 (RONQUIST; HUELSENBECK, 2003) and raxmlGUI2 (EDLER *et al*., 2021). For the Bayesian inference, the Markov Chain Monte Carlo (MCMC) was run using two simultaneous independent runs with four chains each (one cold and three heated), saving one tree every 1,000 generations, for a total of ten million generations. We excluded 20% of retained trees as 'burn in', and checked for a stationary phase of likelihood, checking for ESS values higher than 200 for all parameters on Tracer 1.6 (RAMBAUT *et al*., 2014). The posterior probabilities (PP) of clades were based on the majority rule consensus, using the stored trees, and calculated with MrBayes 3.1.2 (RONQUIST; HUELSENBECK, 2003). Support values are presented below branches, with posterior probabilities shown before bootstrap values.

2.10 Ancestral state reconstruction

Chemical profiles were selected based on the clustering trends of the PCoA analysis for EtOH80% and EtOAc100% extracts, in both positive and negative ionization modes (four subsets). The subsets in which samples clustered into two groups (A and B) represented different chemical profiles. Character coding followed the recommendations of (SERENO, 2007) for morphological phylogenies. Primary homology hypotheses (PINNA, 1991) were proposed for different chemical profiles, with a total of two subgroups of classes scored (i.e., presence and absence). All characters were optimized on the concatenated tree from the Bayesian inference using the maximum likelihood function with Mesquite 2.73 (MADDISON; MADDISON, 2006) using the mk1 algorithm.

After annotating the main classes of compounds present in the hydroethanolic extracts, we transformed the obtained quantitative matrices for the positive and negative ionization modes into qualitative matrices considering only the presence and absence of all annotated classes. Since we do not have all the analyzed species of this study sampled in the latest generic molecular phylogeny of Malpighiaceae, we summarized our results at the generic level using an arithmetic mean equation. After summarizing both matrices (retrieved from the positive and negative ionization modes), we compared all classes of secondary metabolites identified in each of them, merging their binary coding into a single column. By doing so, we obtained a total of 78 classes of secondary metabolites that were optimized into a summarized version of the phylogenetic tree presented by Davis and Anderson (2010), including only the genera sampled in this study. Ancestral character reconstructions were performed using Maximum Likelihood approaches into the software Mesquite 2.73 (MADDISON; MADDISON, 2006) and visualized in the software Winclada (NIXON, 1999) using the fast optimization method, which favors homoplasies in the analysis.

3 Results and Discussion

3.1 Evaluation of extraction solvent and ionization mode in Malpighiaceae chemical diversity and chemosystematics

The distribution of the collection points of Malpighiaceae plant species in study is shown in Figure 14. This unique collection of Malpighiaceae samples comprises 39 genera and 137 species from each of the major phylogenetic groups currently accepted for the family. A total of 197 samples were collected, in which it was possible to obtain the same species found in different biomes, as described in Supplementary Box 1.

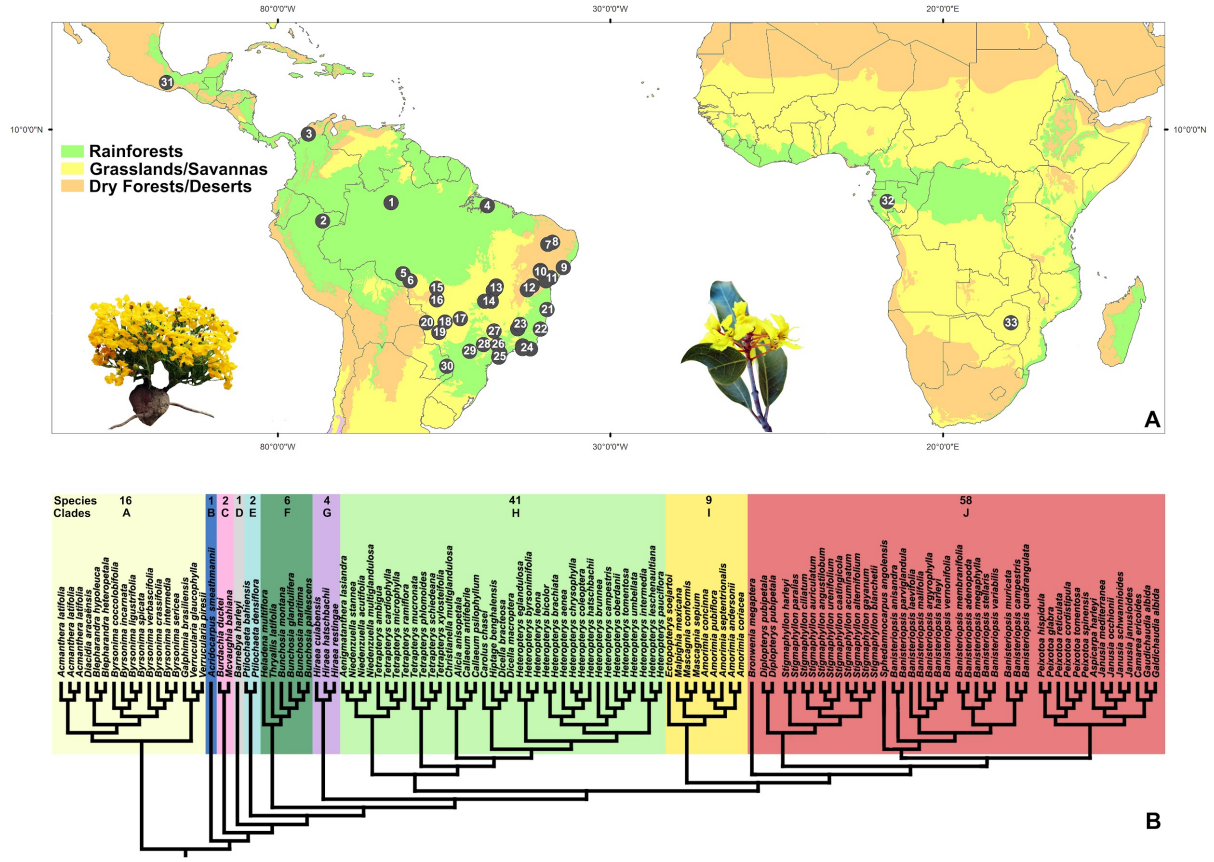


Figure 14 – (**A**) Distribution map showing the collection sites of all samples within the American and African continents. A complete record of all collection sites (numbers on black circles) is listed in Supplementary Box 1. Photograph on the left representing a New World tropic species of Malpighiaceae (*Camarea ericoides* by R.F. Almeida). Photograph on the right representing an Old World tropic species of Malpighiaceae (*Acridocarpus excelsus* by T. Randrianarivony). (**B**) Ten major phylogenetic clades currently accepted in Malpighiaceae, based on plastid and nuclear genes, according to Davis and Anderson (2010). Major clades are shaded in different colors. Species = number of species sampled by each clade in our study. Clade A = Byrsonimoid clade; B = Acridocarpoid clade; C = Mcvaughioid clade; D = Barnebyoid clade; E = Ptilochaetoid clade; F = Bunchosioid clade; G = Hiraeoid clade; H = Tetrapteroid clade; I = Malpighioid clade; and J = Stigmaphylloid clade.

Source: elaborated by the Author (2022).

The extracts (EtOH–H_2O, 4:1, v/v; EtOAc) obtained from the Malpighiaceae samples were analyzed by LC-MS/MS in the positive and negative ionization modes and processed in MZmine2. The feature finding step resulted in 24,440 and 20,150 features detected in the positive and negative ionization modes, respectively. The

Venn diagrams obtained for the different extraction protocols in positive and negative ionization modes (Figure 15A) showed that only about 5% of the features were detected exclusively with ethyl acetate as extraction solvent regardless of the ionization mode used. In the positive ionization mode, 74% of the features were shared between the two extraction protocols, while in the negative ionization mode, this number dropped to 59%, with the number of metabolites exclusively detected in the hydroethanolic extraction rising from 20% to 36%. These results are in accordance with the previously reported since even though there is a common core metabolome, solvent-specific metabolites can be observed (CRÜSEMANN *et al*., 2017). In this way, in order to obtain a broader chemical diversity, different extraction solvents can be used. Additionally, the diversity of metabolites in common and solvent-specific will also vary depending on the ionization mode employed.

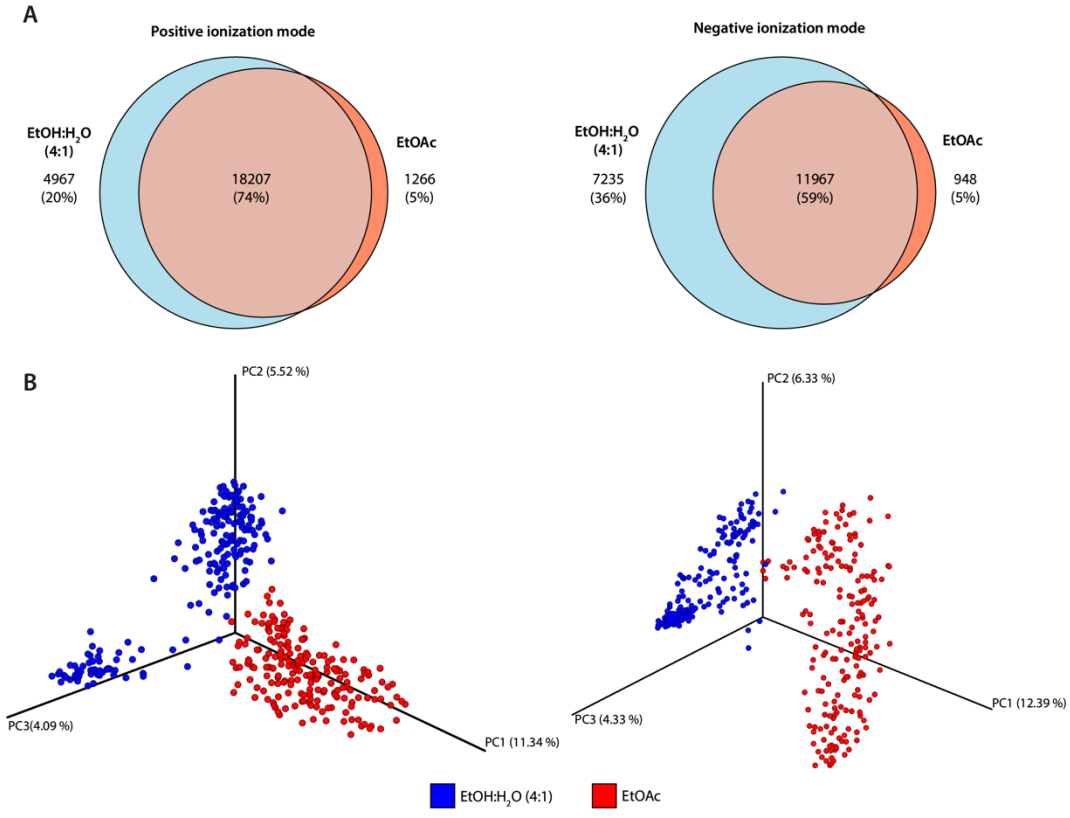


Figure 15 – Diversity of metabolic profiles obtained in different extraction protocols and ionization modes. (**A**) Venn diagrams for the different extraction protocols in positive and negative ionization modes. (**B**) Three-dimensional PCoA plots of the samples analyzed in different ionization modes (positive: left; negative: right) determined by Bray–Curtis distance metric. The percentage of variance explained by the principal coordinates is presented on each axis.

Source: elaborated by the Author (2022).

To observe the metabolomic profiles obtained by the different extraction protocols, Principal Coordinates Analysis (PCoA) of the positive and negative subsets were obtained (Figure 15B). The PCoA plots showed that the two extraction solvents resulted in very different metabolomic profiles on both ionization modes (for positive ionization mode: PERMANOVA F = 40.39, p = 0.001; for negative ionization mode: PERMANOVA F = 38.37, p = 0.001). Even though most of the features detected are shared between the two extraction protocols, the compounds' relative abundance may vary significantly to result in different metabolomic profiles.

In order to determine whether there would be clustering trends for each of the extracts obtained and in both ionization modes, PCoAs for these four subsets were also obtained. For the hydroethanolic extract, two main groups were formed in both ionization modes (Figure 16A) (for positive ionization mode: PERMANOVA F = 28.59, p = 0.001; for negative ionization mode: PERMANOVA F = 43.92, p = 0.001), while for the ethyl acetate extract, a separation into two groups was obtained in the positive ionization mode (PERMANOVA F = 18.71, p = 0.001; Figure 17A), and no group separation was observed in the negative ionization mode in any of the distance matrices tested (Figure 17C), reinforcing that the extraction solvent and ionization mode can influence data acquisition.

These clustering trends were optimized using maximum likelihood criteria into the most recent molecular phylogenetic tree of Malpighiaceae obtained from (DAVIS; ANDERSON, 2010), resulting in the cladograms shown in Figure 16B. These trees are graphical representations of how these different chemical profiles obtained from hydroethanolic extracts originated in specific clades (*i.e.*, groups) through time (*i.e.*, geological time). For the metabolites detected in the positive ionization mode, we can infer that the production of metabolites from group B originated at least three separate times (firstly in part of the Byrsonimoid clade, secondly in the Mcvaughioid clade, and thirdly in the most recent common ancestor of the Ptilochaetoid clade shared with all remaining clades of Malpighiaceae (*i.e.*, Ptilochaetoid + Bunchosioid + Hiraeoid + Malpighioid + Tetrapteroid + Stigmaphylloid). In other words, the samples from group B are, in general, evolutionary more recent than the ones from group A. Some specific genera (such as *Amorimia*, *Alicia*, and *Banisteriopsis*), which are more recent in terms of their origins in past geological times, also cluster together with group A. A possible

explanation for this tendency is that these genera are mostly found in dry environments, just like some of the early-diverging genera of Malpighiaceae (ALMEIDA *et al*., 2021; DAVIS *et al*., 2014).

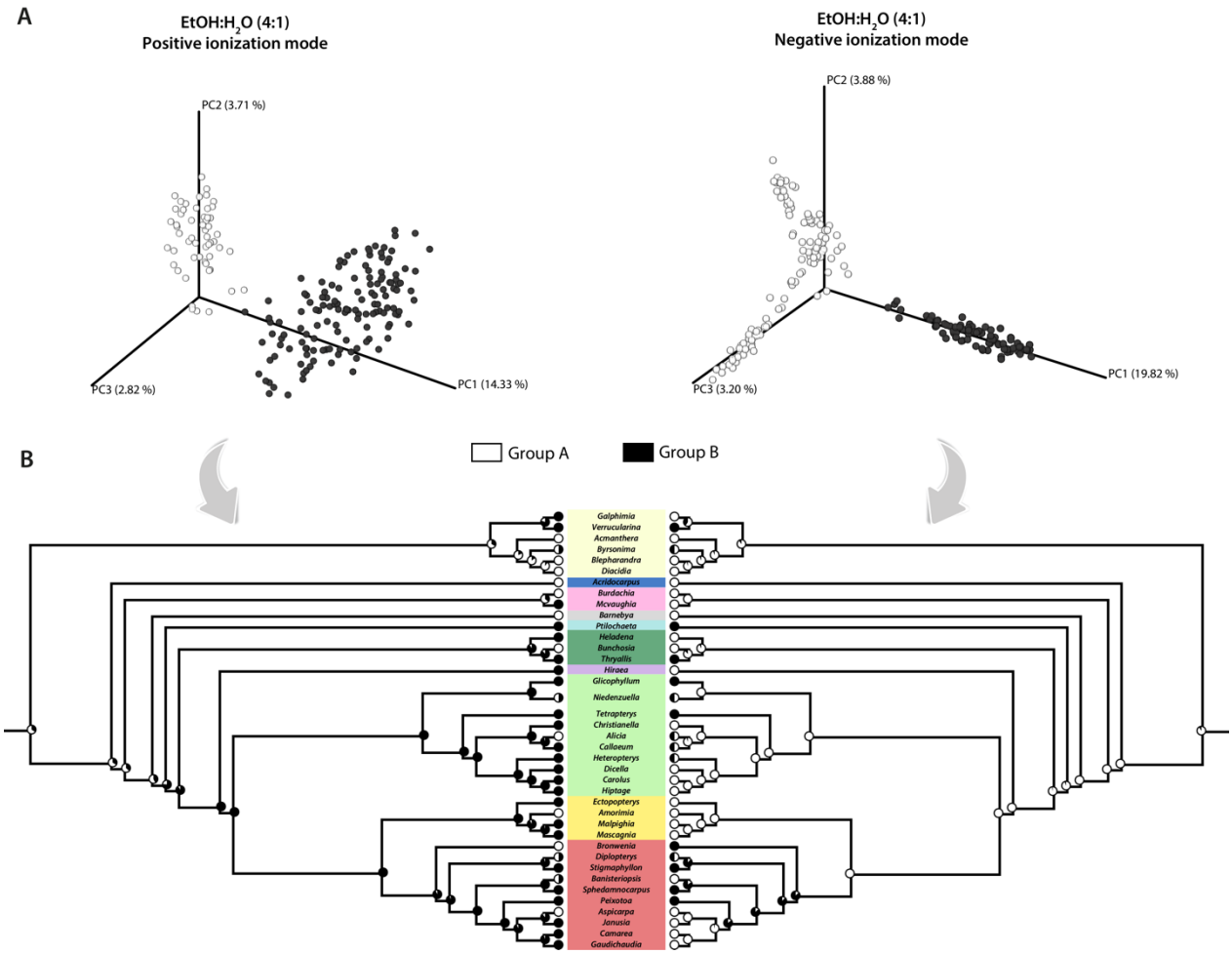


Figure 16 – Clustering trends observed in different ionization modes for the hydroethanolic extracts and correlation with phylogeny. (**A**) Three-dimensional PCoA plots of the hydroethanolic extracts analyzed in different ionization modes determined by Canberra distance. The percentage of variance explained by the principal coordinates is presented on each axis. (**B**) Groups observed in the PCoA plots optimized using maximum likelihood criteria in the most recent molecular phylogeny of Malpighiaceae. Pie charts located on the tip of branches represent presence/absence of metabolomic profiles from groups A and B. Pie charts located on branch nodes of the tree represent the statistical results (presented in %) from the optimization analyzes. Colors represent all ten major phylogenetic clades (i.e., natural groups) currently recognized in Malpighiaceae: light yellow- Byrsonimoid clade (A); dark blue- Acridocarpoid clade (B); pink- Mcvaughioid clade (C); grey- Barnebyoid clade (D); light

blue- Ptilochaetoid clade (E); dark green- Bunchosioid clade (F); purple- Hiraeoid clade (G); light green- Tetrapteroid clade (H); yellow- Malpighioid clade (I); and red- Stigmaphylloid clade (J).

Source: elaborated by the Author (2022).

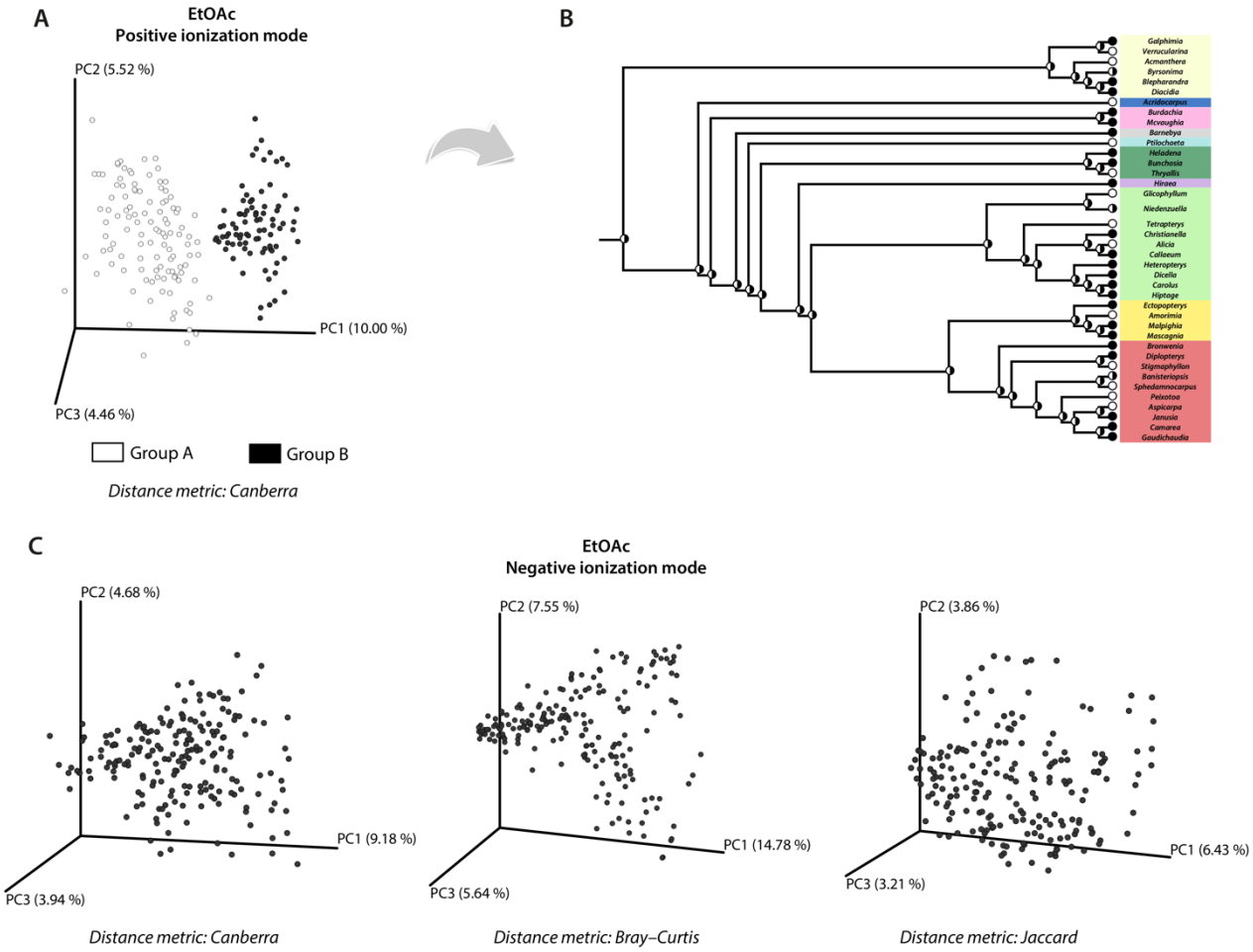


Figure 17 – Clustering trends observed in different ionization modes for the ethyl acetate extracts and correlation with phylogeny. (**A**) Three-dimensional PCoA plots of the ethyl acetate extracts analyzed in positive ionization mode determined by Canberra distance. The percentage of variance explained by the principal coordinates is presented on each axis. (**B**) Groups observed in the PCoA plots optimized using maximum likelihood criteria in the most recent molecular phylogeny of Malpighiaceae. (**C**) Three-dimensional PCoA plots of the ethyl acetate extracts analyzed in negative ionization mode determined by Canberra, Bray–Curtis and Jaccard distances. The percentages of variance explained by the principal coordinates are presented on each axis.

Source: elaborated by the Author (2022).

On the other hand, the metabolomic profile obtained in the negative ionization mode analyses for the hydroethanolic extracts showed a different tendency. The different chemical profiles obtained from group B in the negative ionization mode were only recovered as synapomorphic to the Stigmaphylloid clade (i.e., *Bronwenia*, *Diplopterys*, *Stigmaphyllon*, *Sphedamnocarpus*, and *Peixotoa*), besides also independently originated in some genera from the Tetrapteroid, Bunchosioid, Ptilochaetoid, and Byrsonimoid clades.

We also evaluated the metabolomic profile obtained for the ethyl acetate extracts in positive ionization mode (since the negative ionization mode did not show clustering trends). Contrary to what we previously observed, there was no clear correlation of the metabolite clustering trends observed with the phylogeny of the Malpighiaceae family (Figure 17B). These differences in both extraction protocols could be explained due to the different polarity ranges covered by these two solvents. From these results, it is important to emphasize that different extraction protocols and ionization modes in MS can significantly impact the results from multivariate analysis and chemosystematics investigations.

3.2 Metabolites annotation

Since the chemical profiles obtained for the hydroethanolic extracts showed a promising correlation with the evolution of Malpighiaceae, we made an in-depth investigation of the data obtained in both positive and negative ionization modes. The MS/MS library search performed in GNPS resulted in 1,070 hits for the positive, and 1,025 hits for the negative ionization modes, resulting in 4.6% and 5.3% of the detected chemical spaces, respectively. These matches were carefully evaluated and compared with the literature, resulting in level 2 or 3 annotations according to the metabolomics standards initiative (MSI) (SUMNER *et al*., 2007). In addition, these compounds were also searched in the *Dictionary of Natural Products*, the *naturaL prOducTs occUrrences databaSe* (LOTUS) (RUTZ *et al*., 2021), and Scifinder databases to inspect for previous reports in Malpighiaceae.

The matches from the positive ionization mode showed the presence of a high diversity of classes of compounds, including *C*-glycosylated and *O*-glycosylated flavonoids, lipids, alkaloids, quinic acid derivatives, amides, triterpenes, iridoids, and

lignans. Some of these classes of compounds were widely detected in all clades, while others were more specific to particular clades or even genera.

Flavonoids with two *C*-glycosylated portions were mainly detected in samples from clade I, especially in *Amorimia* and *Mascagnia* genera (Figure 18). Library matches from this molecular family included apigenin-di-*C*-hexoside-pentoside and luteolin-di-*C*-hexoside, being in accordance with the previous reports for *Amorimia* genera (MANNOCHIO-RUSSO *et al*., 2020). On the other hand, spectral matches to flavonoids containing only one *C*-glycosylated portion (Figure 19) were less clade-specific, being detected in samples from clades A, G, H, I, and J.

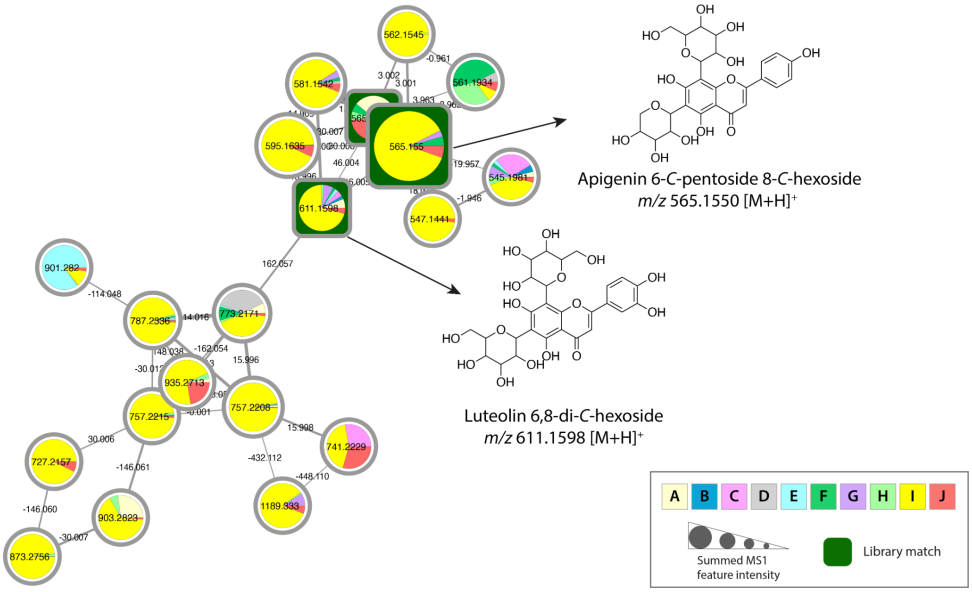


Figure 18 – Molecular family composed of flavonoids containing two *C*-glycosylated portions (positive ionization mode).

Source: elaborated by the Author (2022).

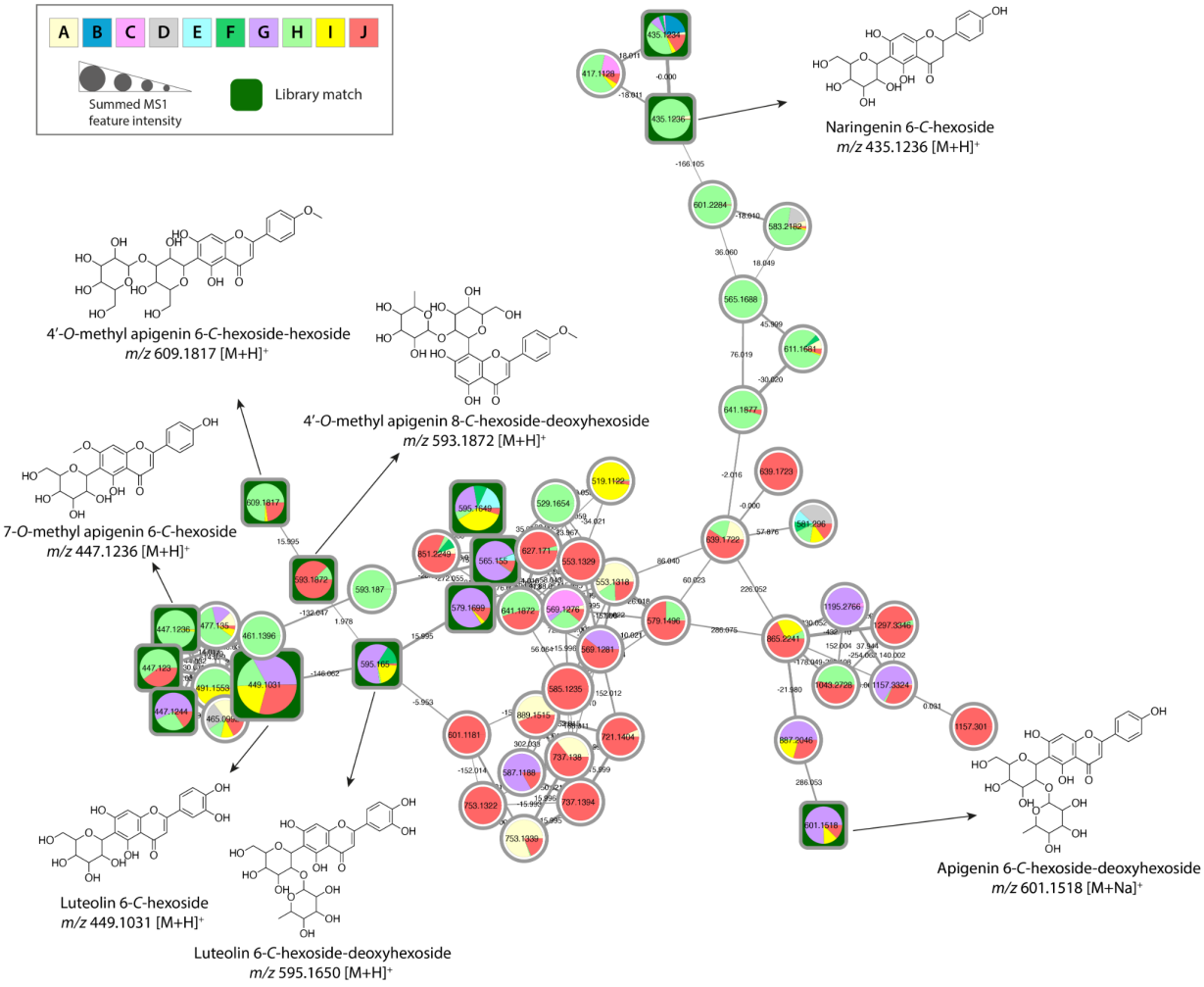


Figure 19 – Molecular family composed of flavonoids containing one *C*-glycosylated portion (positive ionization mode).

Source: elaborated by the Author (2022).

Several matches to *O*-glycosylated flavonoids were observed in all Malpighiaceae clades (Figure 20), which is expected since this class of compounds is considered ubiquitous to plant species. This network primarily comprises quercetin and kaempferol derivatives bounded to hexose, deoxyhexose, and pentose sugars, presenting neutral losses of 162 Da, 146 Da, and 132 Da, respectively. A small cluster in this network highlighted compounds mostly present in clade J, specifically from *Stigmaphyllon* species. These compounds were annotated as flavonoids with glucuronide, malonyl hexoside, and acetylated and malonylated hexoside substituents, which have not been reported in the literature for the *Stigmaphyllon* genus to date.

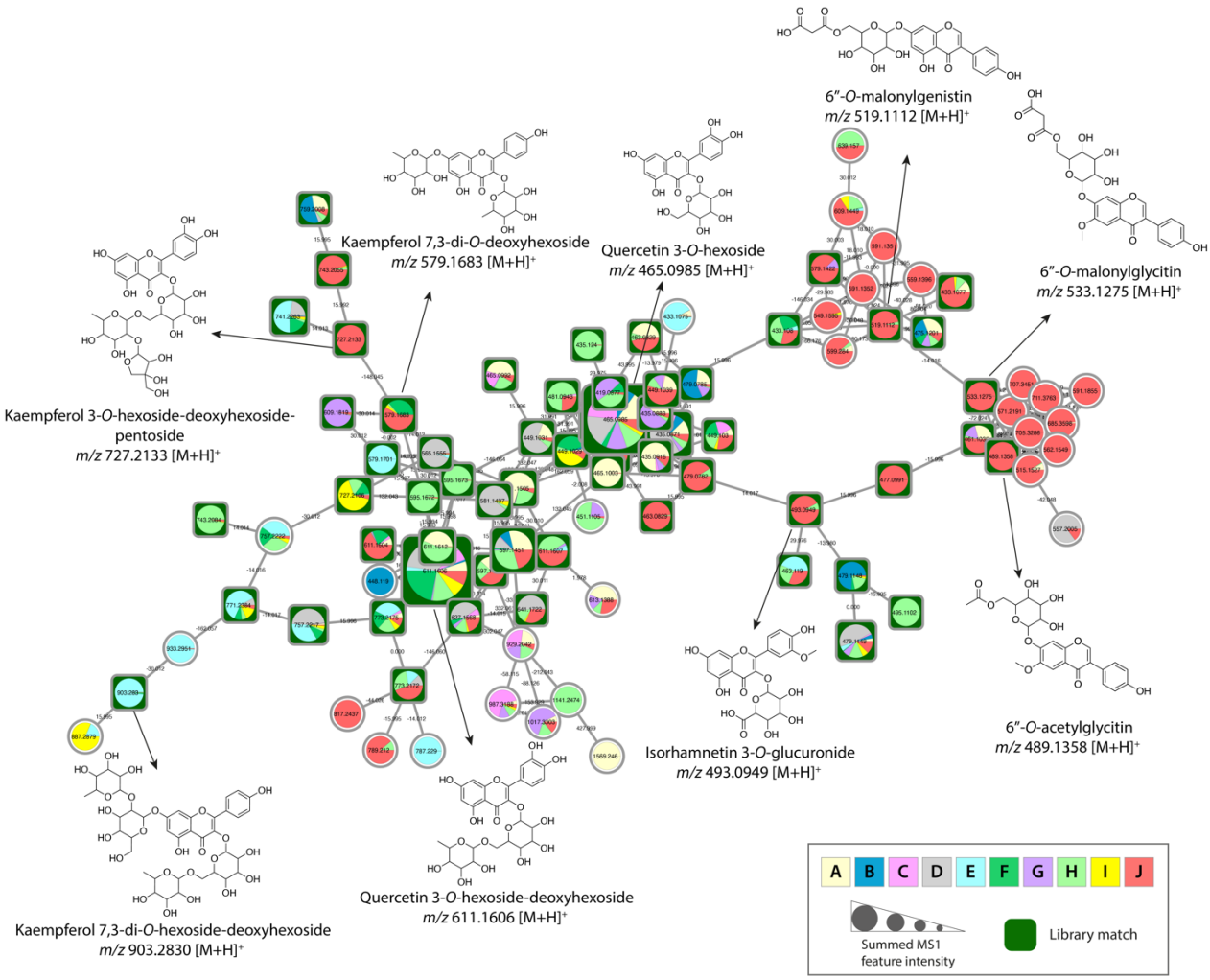


Figure 20 – Molecular family composed of *O*-glycosylated flavonoids (positive ionization mode).

Source: elaborated by the Author (2022).

Other networks containing *O*-glycosylated flavonoids were also observed. Substituents containing sugars and phenylpropanoid portions (Figure 21) were mainly present in the *Ptilochaeta* genus (clade E), which is in accordance with previous reports (MANNOCHIO-RUSSO *et al*., 2020). In addition, sugars bounded to galloyl portions (Figure 22) were also detected as flavonoid substituents and widely found in all Malpighiaceae clades.

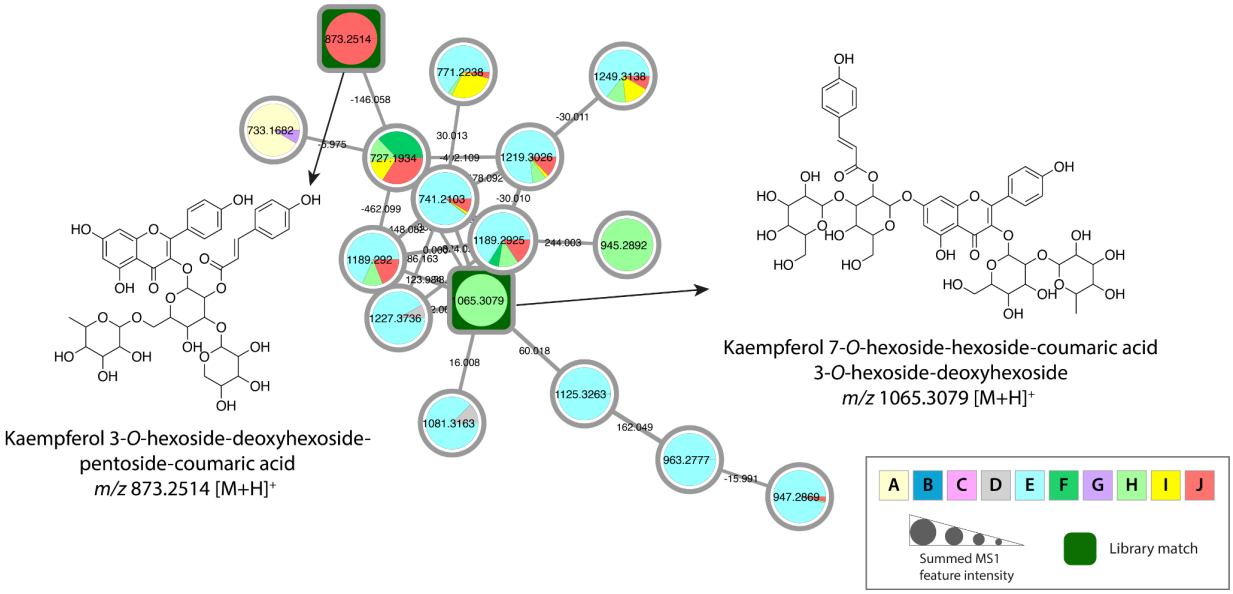


Figure 21 – Molecular family composed of *O*-glycosylated flavonoids containing phenylpropanoid potions (positive ionization mode).

Source: elaborated by the Author (2022).

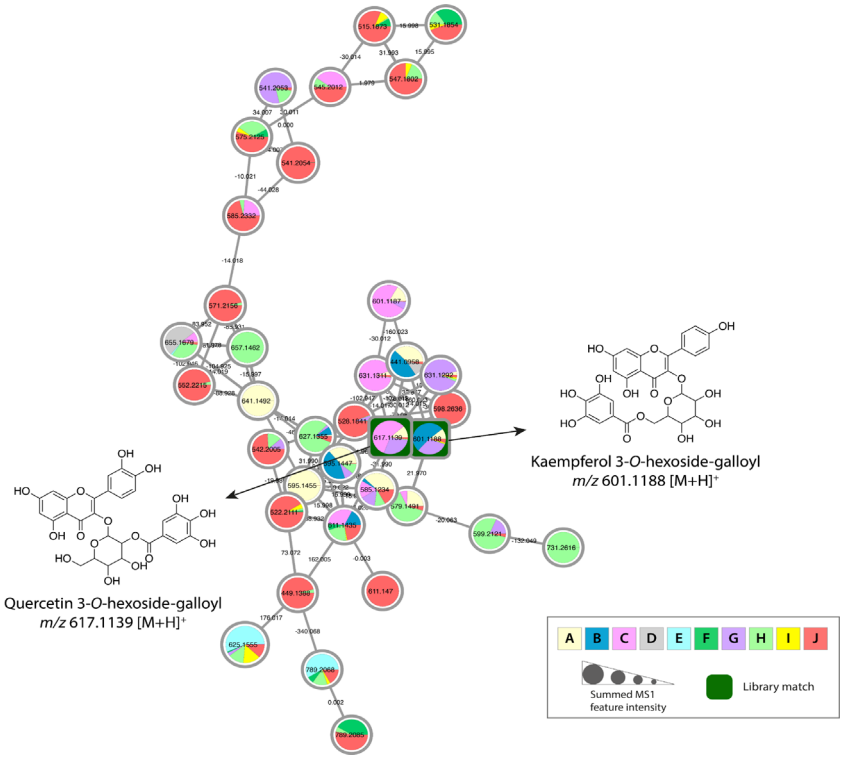


Figure 22 – Molecular family composed of *O*-glycosylated flavonoids containing galloyl potions (positive ionization mode).

Source: elaborated by the Author (2022).

A network containing mainly non-glycosylated flavonoids (Figure 23), widely distributed in all clades, was also observed. These compounds can represent aglycones, or even in-source fragments of the *O*-glycosylated flavonoids. Additionally, a molecular family with spectral matches to catechin and afzelechin and their derivatives (Figure 24) was also formed, containing mainly methoxylated portions and galloyl and sugar substituents.

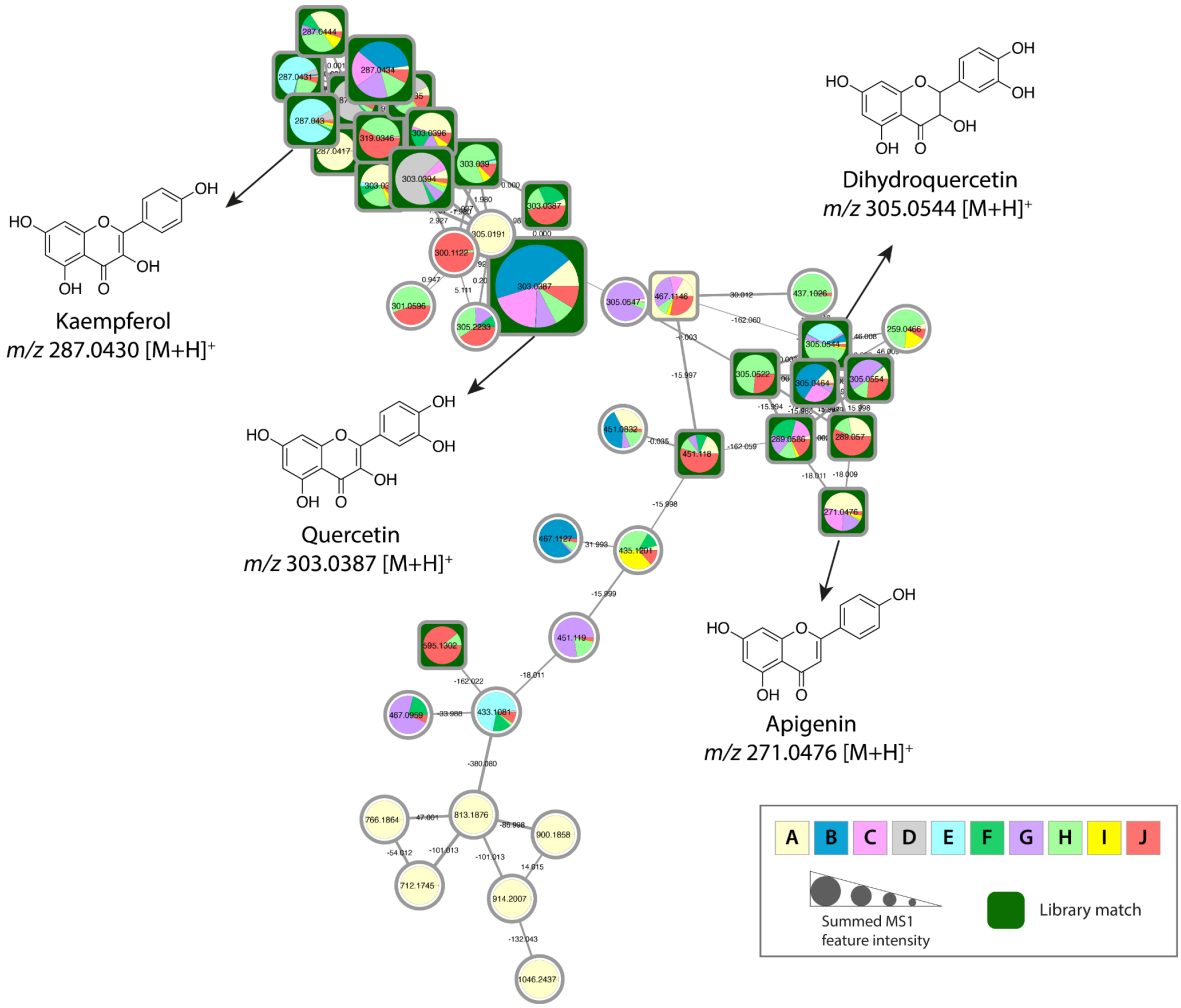


Figure 23 – Molecular family mainly composed of non-glycosylated flavonoids (positive ionization mode).

Source: elaborated by the Author (2022).

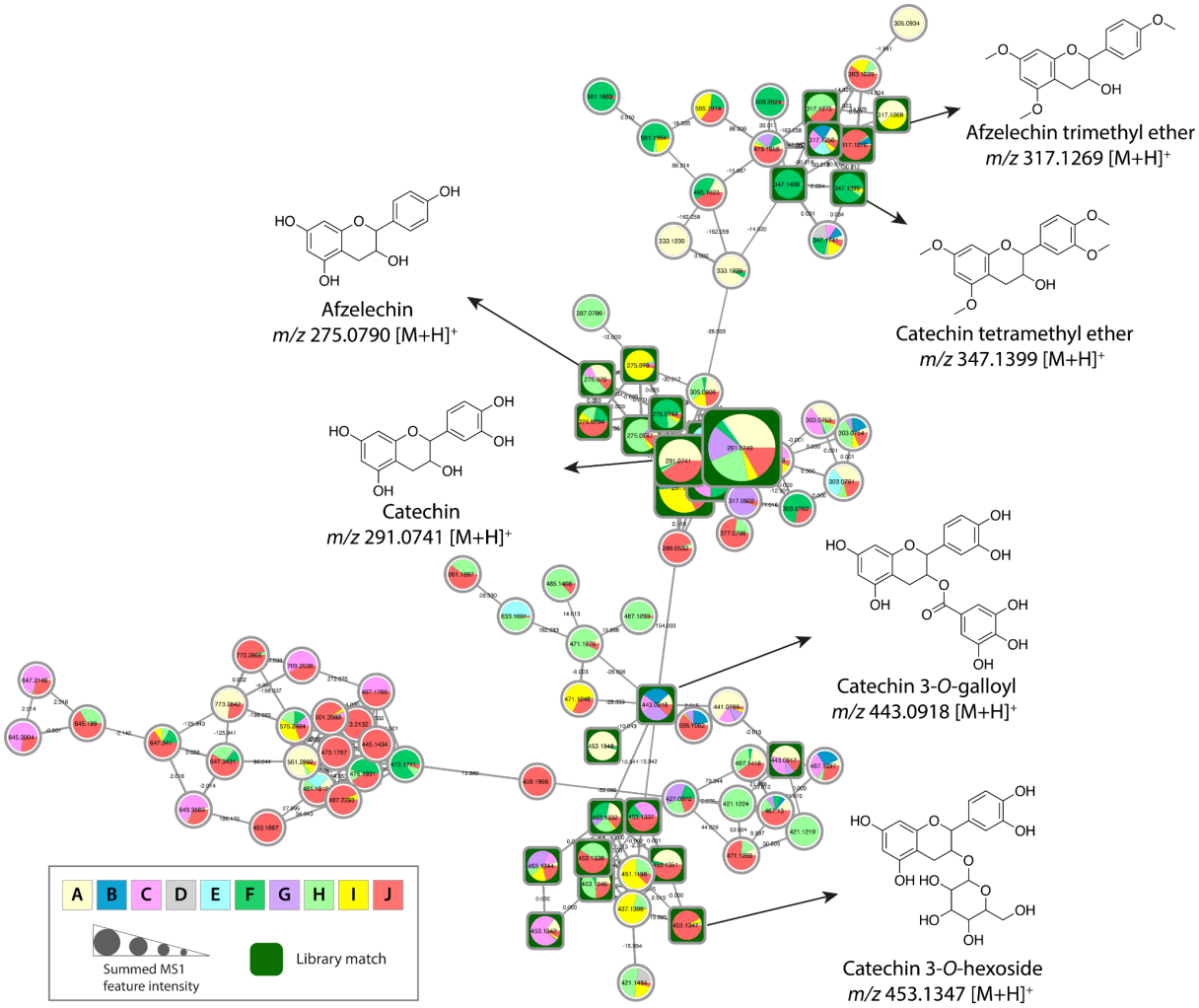


Figure 24 – Molecular family mainly composed of catechin, afzelechin and their derivatives (positive ionization mode).

Source: elaborated by the Author (2022).

Quinic acid derivatives were detected in higher amounts in clades H and J for phenylpropanoid substituents, and in clades A, B, and G for galloyl substituents (Figure 25, Figure 26). Previous reports described the presence of galloylquinic acids in *Byrsonima* (clade A) species (FRAIGE *et al.*, 2018; MANNOCHIO-RUSSO *et al.*, 2020), while quinic acids containing phenylpropanoids substituents have been described for *Heteropterys* (clade H) genus (HUERTA-REYES *et al.*, 2013; PAULA-FREIRE *et al.*, 2013).

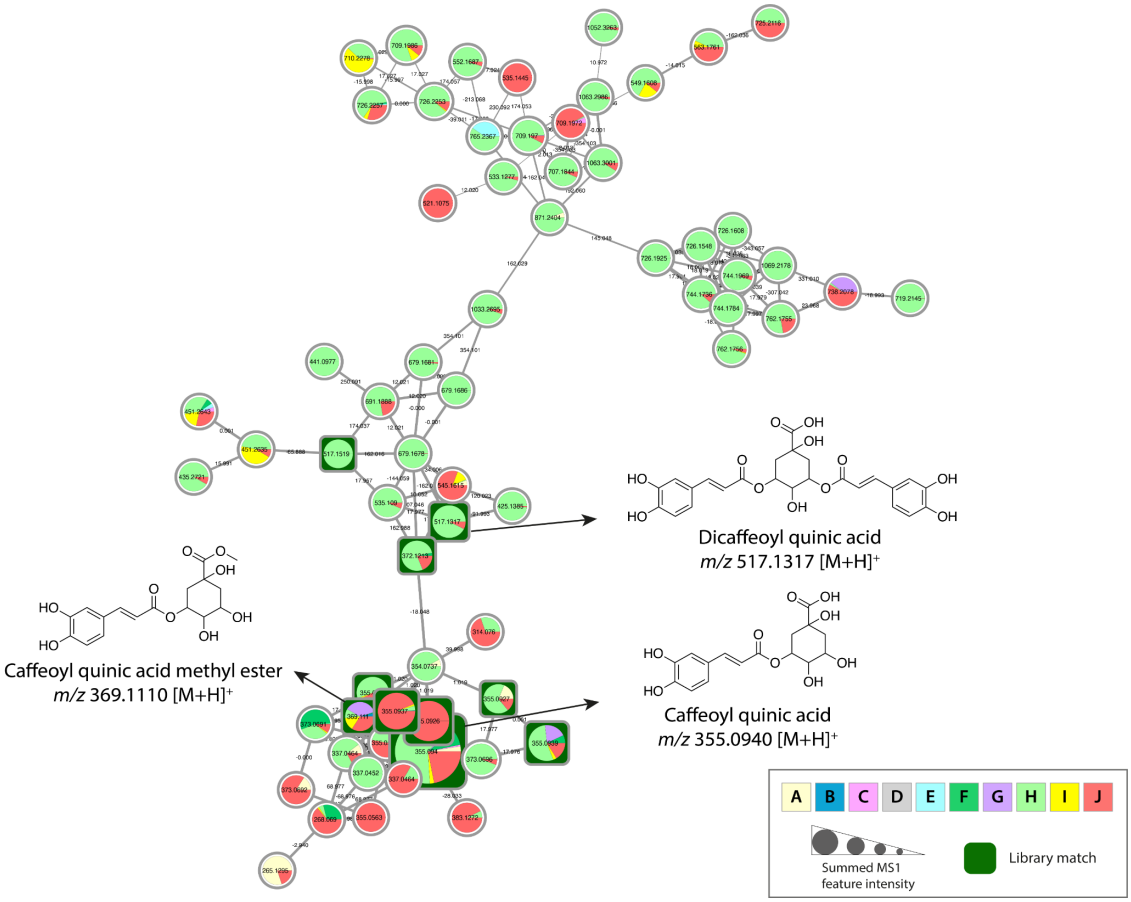


Figure 25 – Molecular family composed of quinic acid bound to phenylpropanoids substituents (positive ionization mode).

Source: elaborated by the Author (2022).

Many lipid-like molecules presented library matches, mainly corresponding to the fatty acids, and fatty esters (Figure 27), and glycerophospholipids (Figure 28) classes, besides jasmonic acid derivatives (Figure 29), which were largely distributed in all Malpighiaceae clades. Lipids represent an important class of compounds, widely found in plants, with key roles in multiple signaling processes (MAMODE CASSIM *et al*., 2019). Lipids, along with sugars, are among the main constituents of the oil glands present in Malpighiaceae leaves and flowers (POSSOBOM; GUIMARÃES; MACHADO, 2010). These glands play a crucial ecological role in Malpighiaceae, a botanical family mainly pollinated by oil collecting bees, and one of the oldest families characterized by oil-bee pollination (ANDERSON, 1990; DAVIS *et al*., 2014; RENNER; SCHAEFER, 2010).

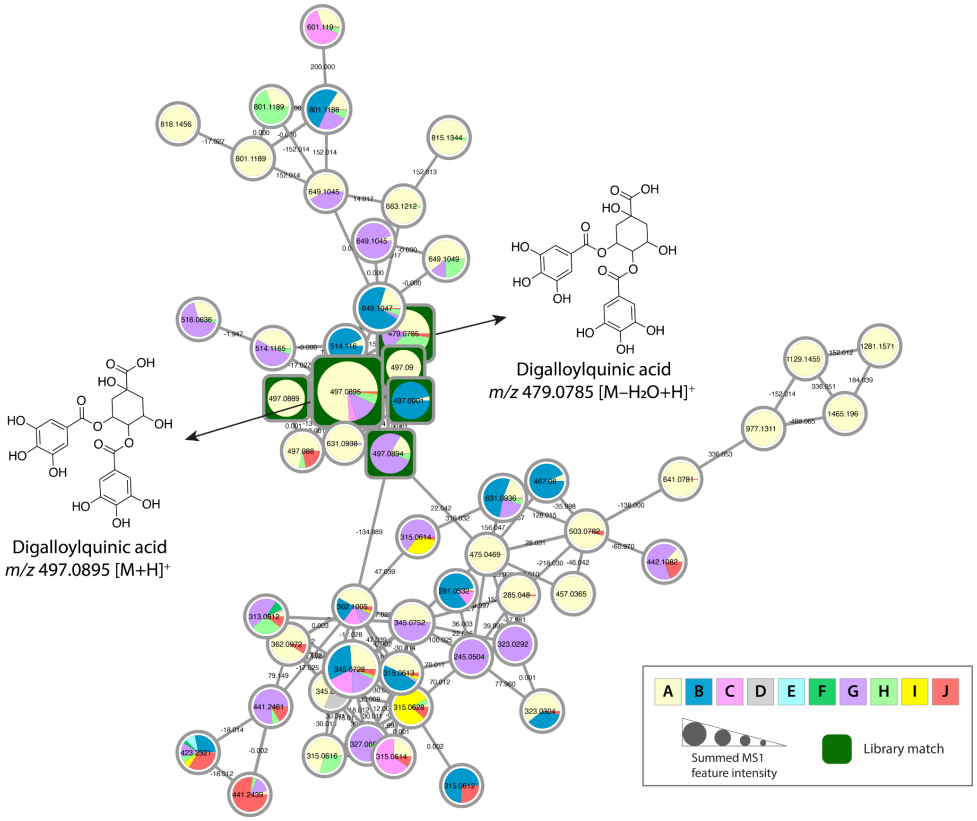


Figure 26 – Molecular family composed of quinic acid bound to galloyl substituents (positive ionization mode).

Source: elaborated by the Author (2022).

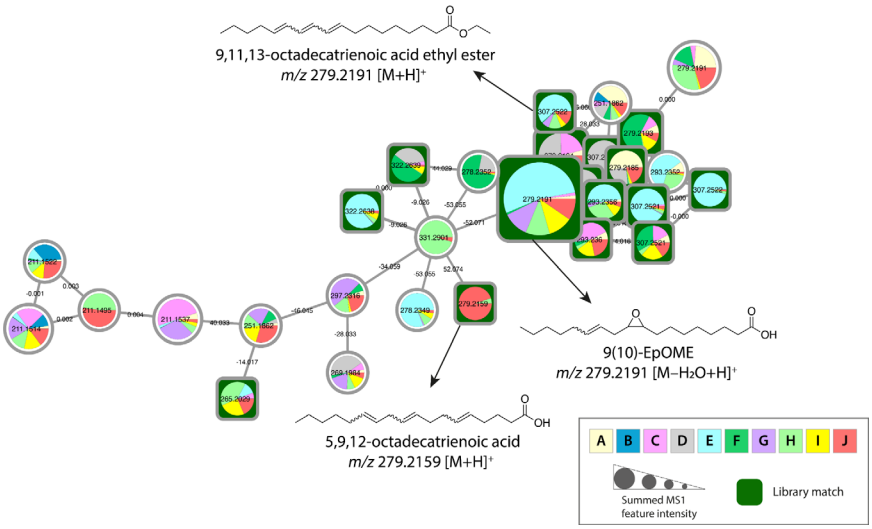


Figure 27 – Molecular family composed of fatty acids and fatty esters (positive ionization mode)

Source: elaborated by the Author (2022).

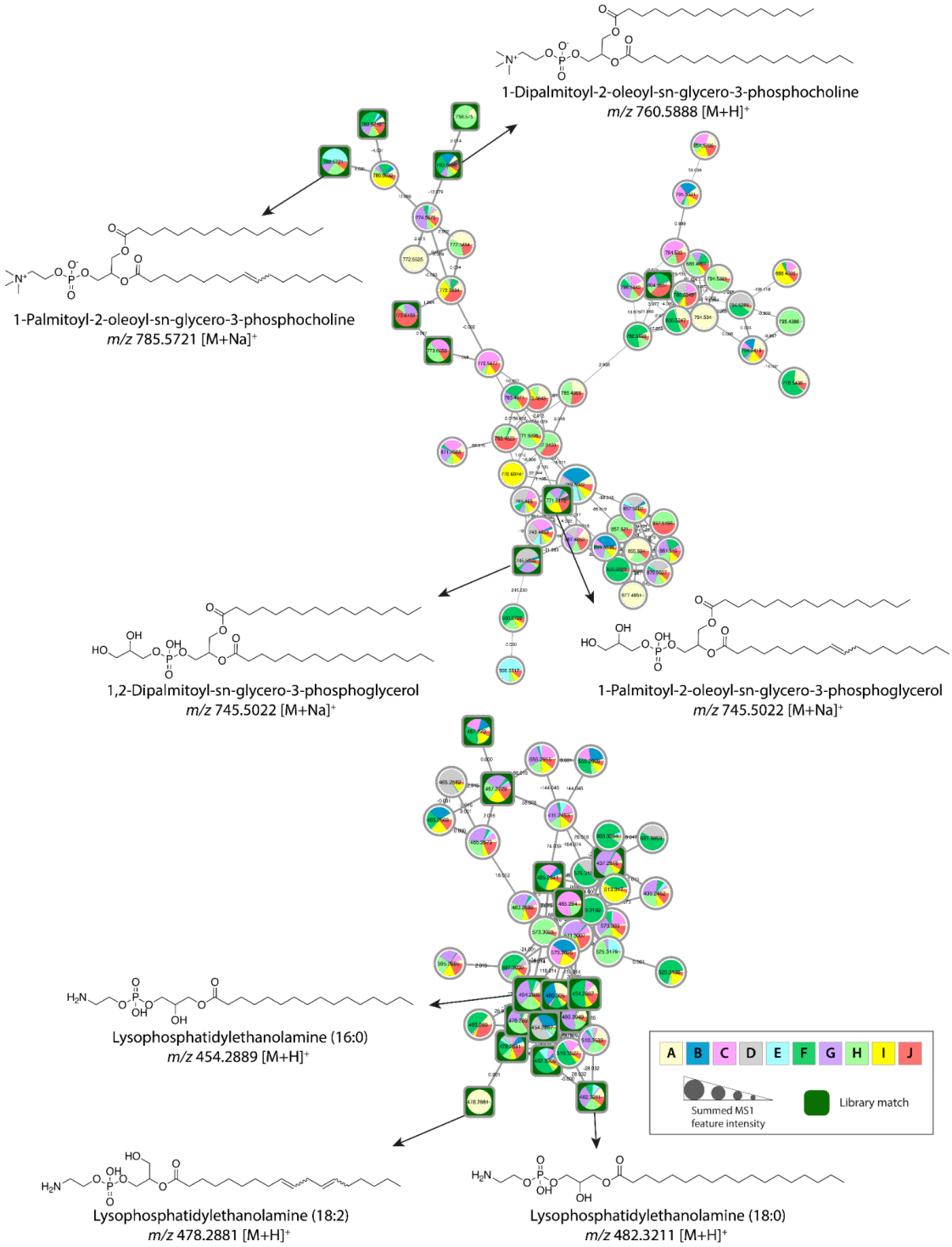


Figure 28 – Molecular families composed of glycerophospholipids (positive ionization mode).

Source: elaborated by the Author (2022).

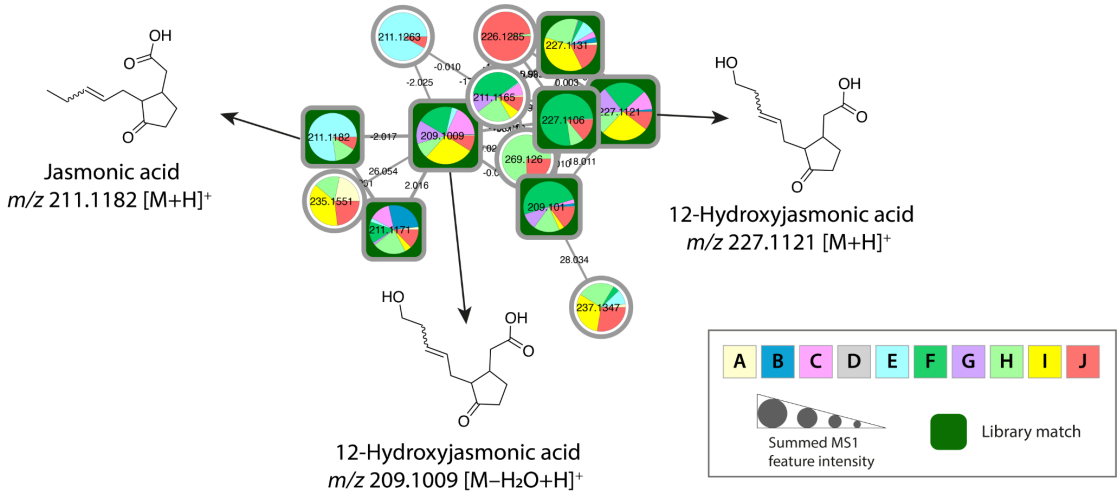


Figure 29 – Molecular family composed of jasmonic acid derivatives (positive ionization mode).

Source: elaborated by the Author (2022).

Spectral matches to β-carboline alkaloids and other tryptophan derivatives were observed in samples from many clades (Figure 30), including *Tetrapterys* and *Banisteriopsis* species, corroborating previous reports (QUEIROZ *et al.*, 2014; SAMOYLENKO *et al.*, 2010). A small cluster into the network from Figure 30 (top) showed the presence of compounds without library matches with higher *m/z* in samples from clade A and J, mainly in *Janusia*, *Banisteriopsis*, and *Byrsonima* species, which can indicate the presence of glycosylated portions due to the mass differences observed between the nodes.

Networks containing library matches corresponding to isoquinoline, protoberberine, and benzylisoquinoline alkaloids were also observed (Figure 31, Figure 32, and Figure 33, respectively). Isoquinoline alkaloids were mainly observed in samples from clades G, H, and J. The protoberberine alkaloid network showed nodes corresponding to compounds mainly detected in clades H and J, particularly in *Stigmaphyllon* and *Alicia* species. Similarly, benzylisoquinoline alkaloids were also mainly detected in samples from clades H and J, especially in *Stigmaphyllon* species. These classes of compounds have not been described to date for Malpighiaceae species and represent important classes for chemosystematics studies in Malpighiaceae due to their specificity to the clades mentioned above, especially protoberberine and benzylisoquinoline alkaloids.

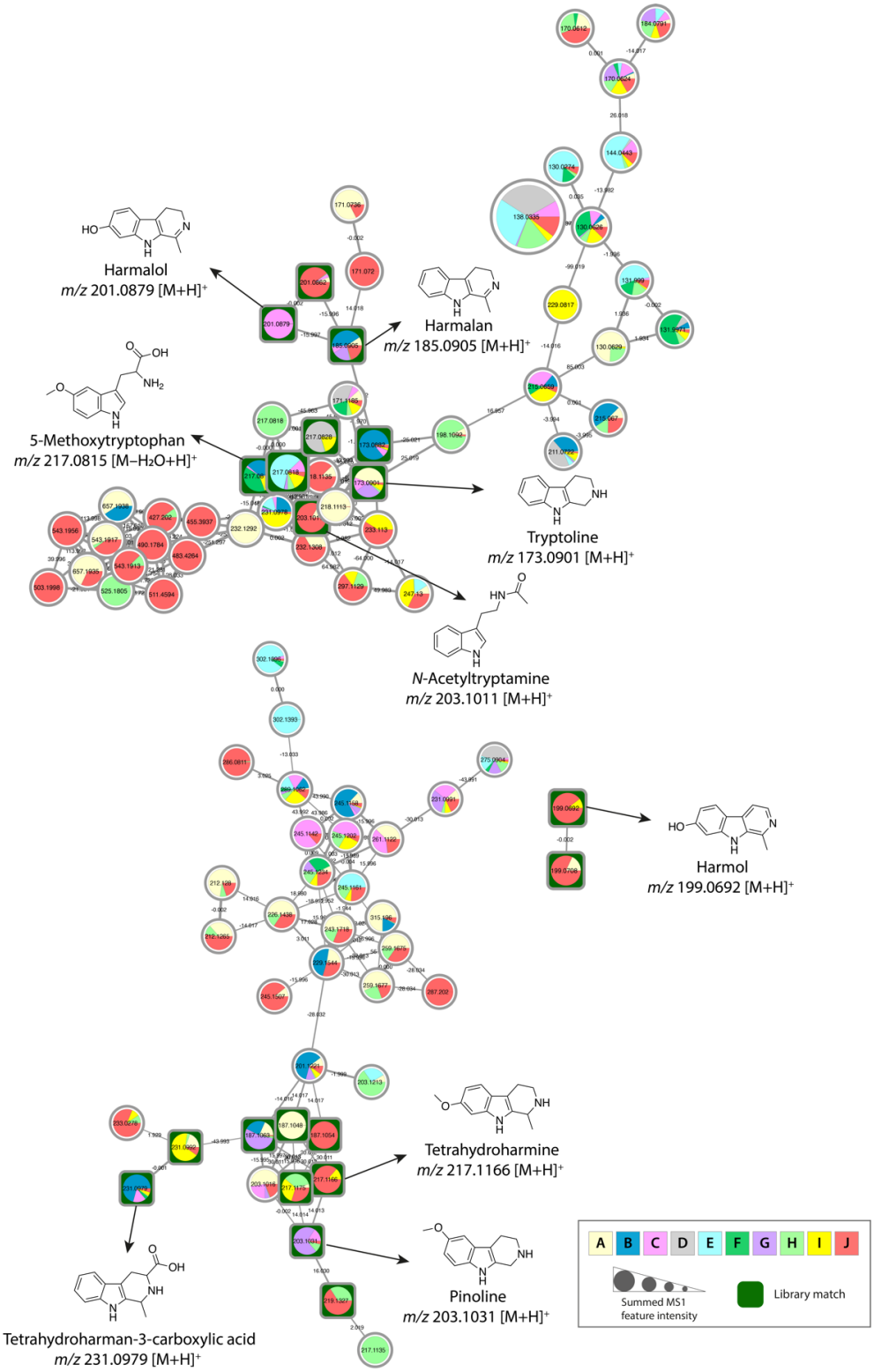


Figure 30 – Molecular families composed of beta-carboline alkaloids and other tryptophan derivatives (positive ionization mode).

Source: elaborated by the Author (2022).

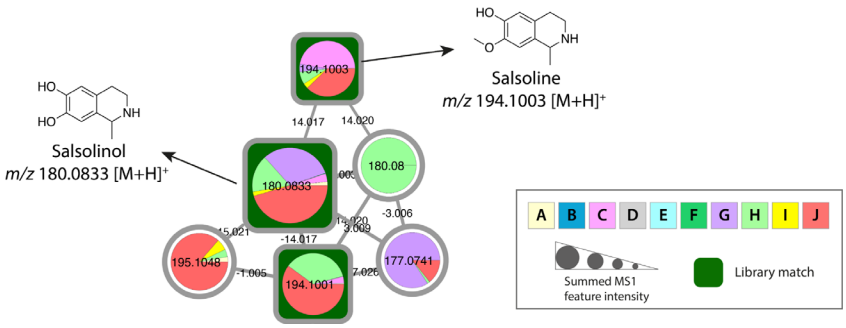


Figure 31 – Molecular family composed of isoquinoline alkaloids (positive ionization mode).

Source: elaborated by the Author (2022).

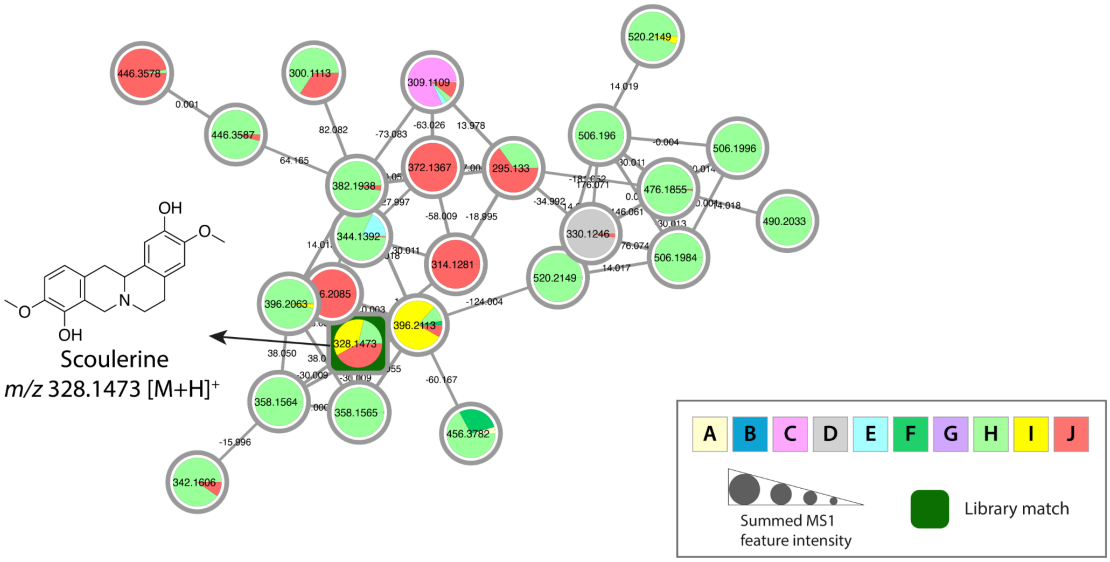


Figure 32 – Molecular family composed of protoberberine alkaloids (positive ionization mode).

Source: elaborated by the Author (2022).

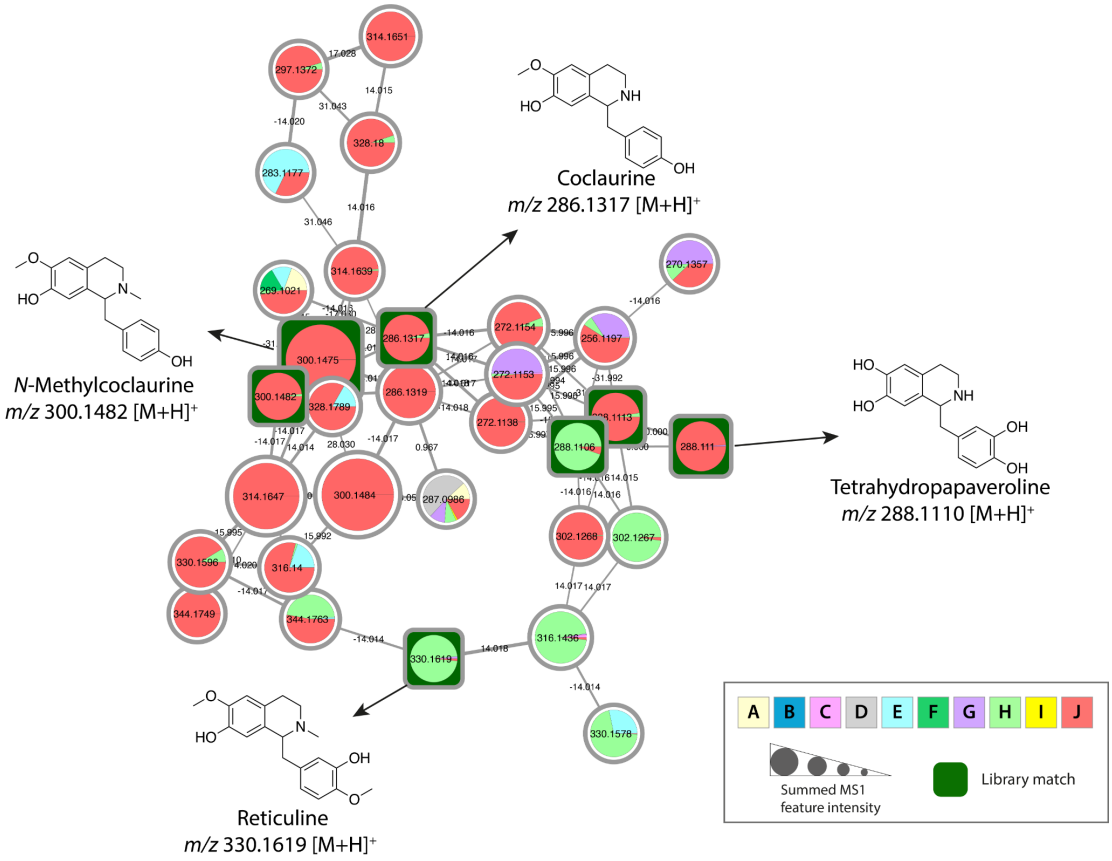


Figure 33 – Molecular family composed of benzylisoquinoline alkaloids (positive ionization mode).

Source: elaborated by the Author (2022).

Spectral matches corresponding to amides and polyamines were also observed in the networks shown in Figure 34 and Figure 35, respectively. Amides were mainly detected in the *Ptilochaeta* genus (clade E) as fatty amides and small dipeptides, while the polyamines were primarily observed in samples from clades E, H, I, J, being this the first report of these classes of compounds in Malpighiaceae botanical family.

Several molecular families related to terpenoids were also observed. One of the largest networks obtained (Figure 36) showed library matches to triterpenes and their precursors. These compounds were also broadly distributed among all Malpighiaceae clades, being in accordance with previous reports. In fact, triterpenoids are the most described class of compounds from Malpighiaceae species to date, with numerous reports in the literature for *Acridocarpus*, *Byrsonima*, and *Galphimia* genera (AGUIAR; DAVID; DAVID, 2005; CAO *et al*., 2004; CARDOSO TAKETA *et al*., 2004).

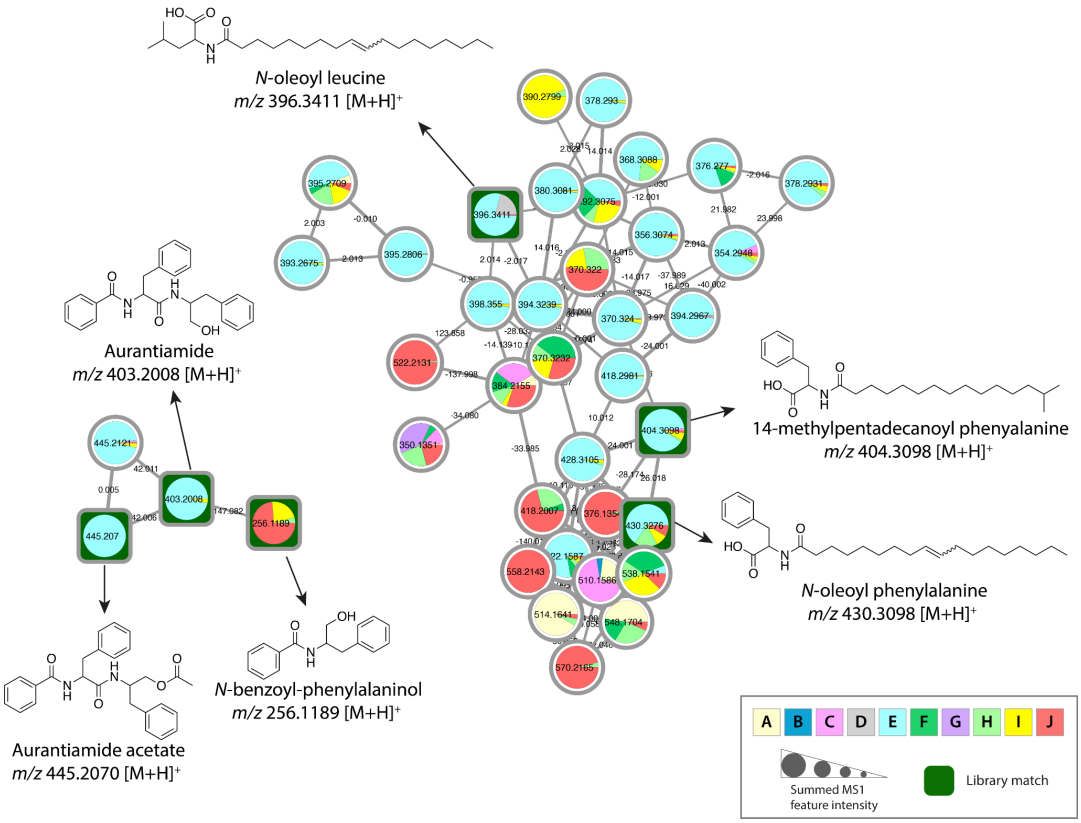


Figure 34 – Molecular families composed of amides (positive ionization mode).

Source: elaborated by the Author (2022).

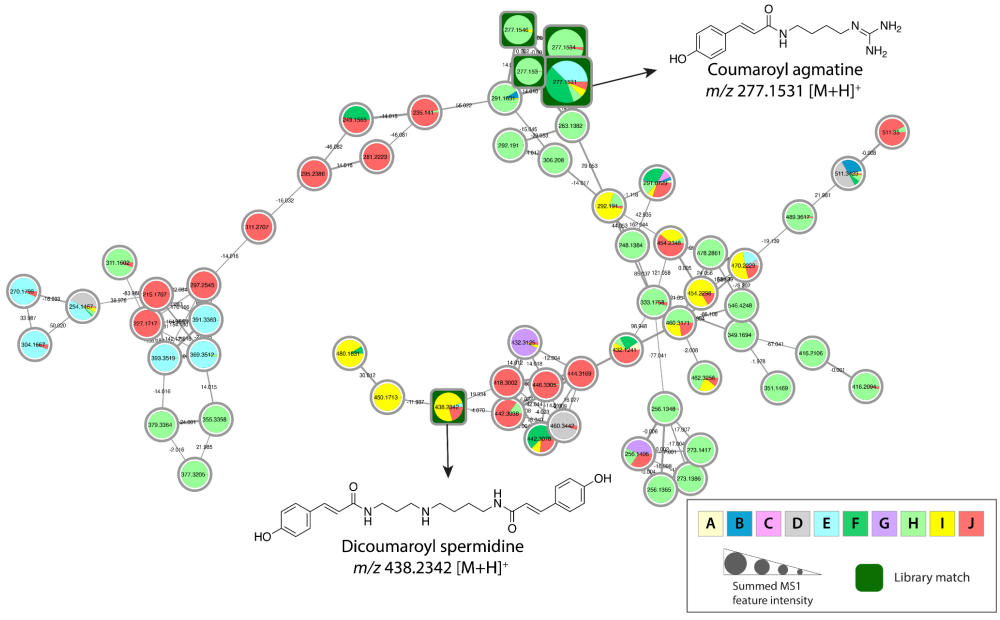


Figure 35 – Molecular family composed of polyamines (positive ionization mode).

Source: elaborated by the Author (2022).

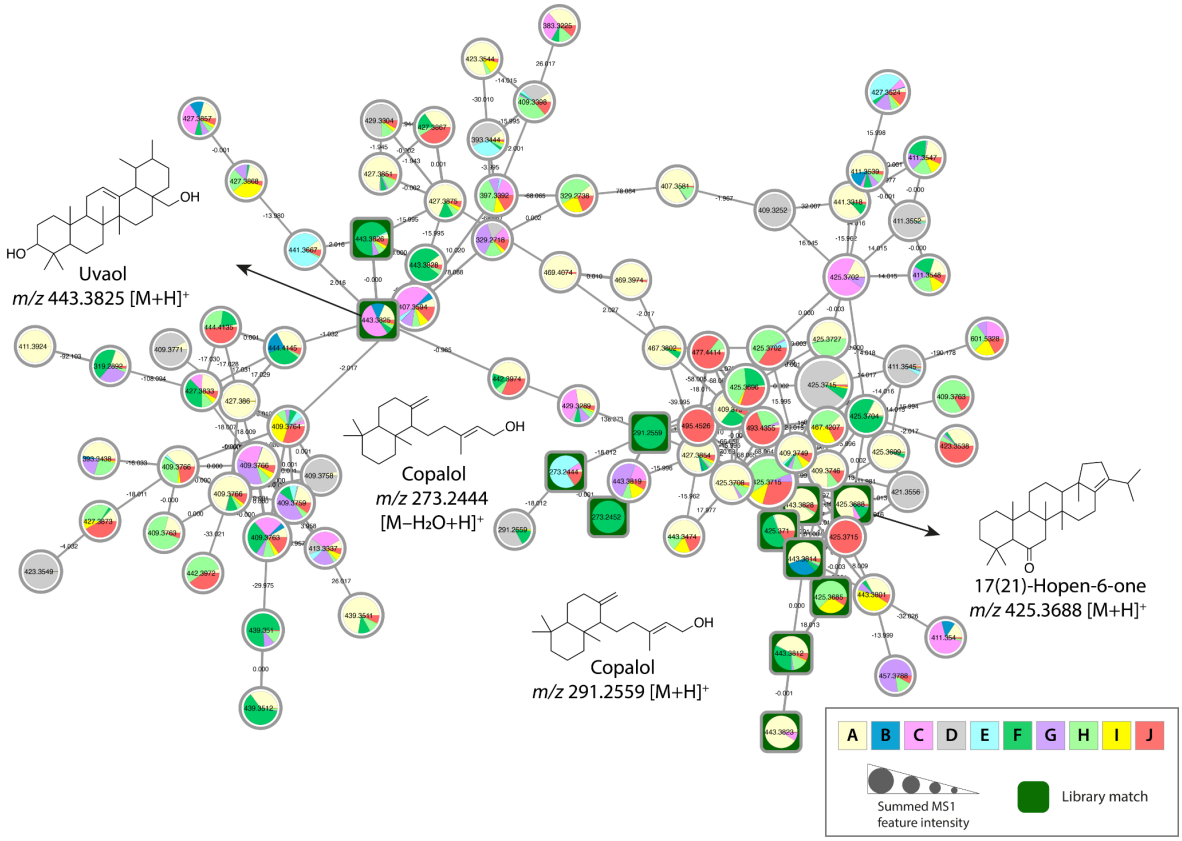


Figure 36 – Molecular family composed of triterpenoids and precursors (positive ionization mode).

Source: elaborated by the Author (2022).

Ecdysteroids were also observed (Figure 37), detected mainly in samples from clade H, particularly *Niedenzuella* species. The presence of ecdysteroids in *Niedenzuella multiglandulosa* species has been recently described in the literature (MANNOCHIO-RUSSO *et al*., 2020; MANNOCHIO RUSSO *et al*., 2020), being the only reports of this class of compounds in the Malpighiaceae family. The analyses showed a wide diversity of ecdysteroids is present in all six *Niedenzuella* species sampled, being possible chemosystematics markers for the *Niedenzuella* genus. The sampled *Niedenzuella* species represents about 40% of all the species reported to date for this genus, and future studies with other *Niedenzuella* species should be performed to corroborate this hypothesis. Moreover, these steroids were also detected in *Hiraea*, *Tetrapterys* (including the recently segregated *Glicophyllum* species, evolutionary close to *Niedenzuella*), and *Peixotoa* genera (clades G, H, and J, respectively).

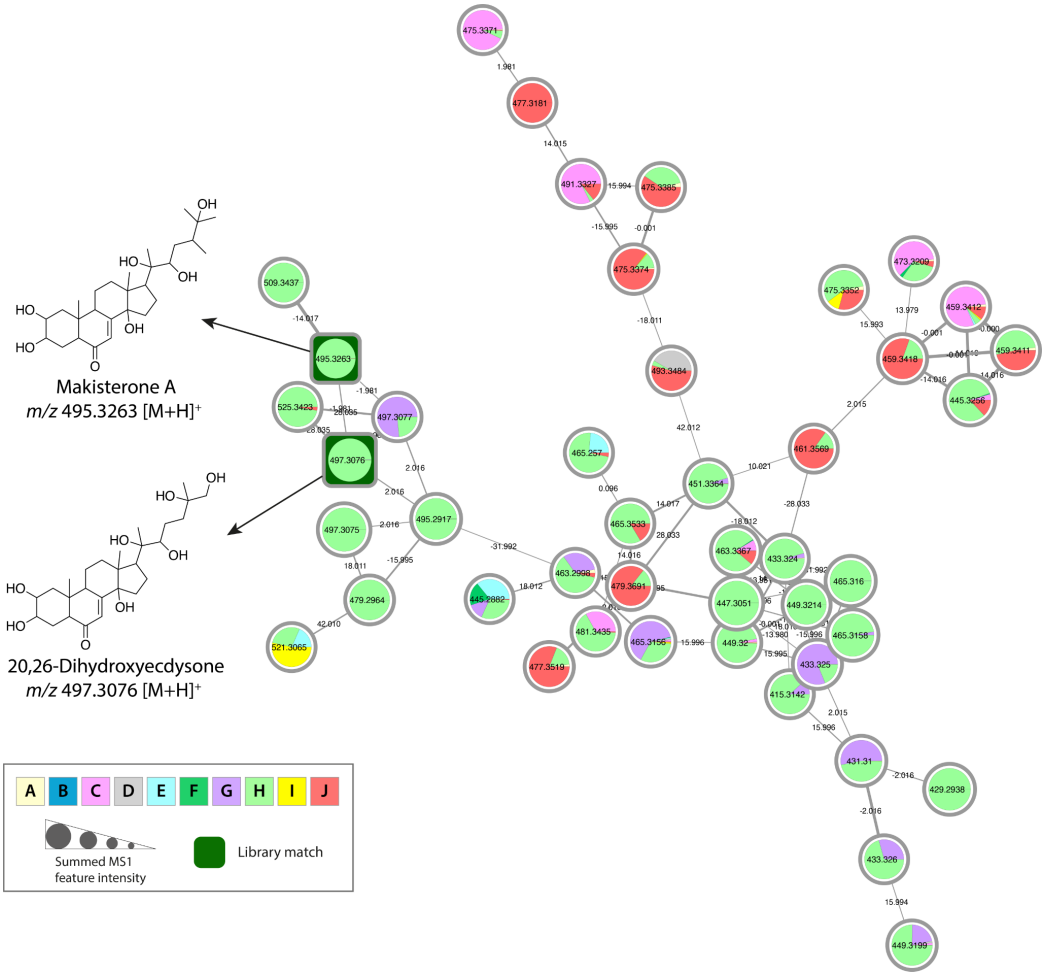


Figure 37 – Molecular family composed of ecdysteroids (positive ionization mode).
Source: elaborated by the Author (2022).

This dataset also presented library matches to iridoids, another relevant class of terpenoids, shown in Figure 38. These networks revealed that these iridoids are mainly present in samples from clade J, especially in *Stigmaphyllon* species, being in accordance with previous reports (DAVIOUD *et al*., 1985; SAINTY *et al*., 1981). To date, iridoids have only been reported for *Stigmaphyllon* genus in Malpighiaceae, and in this study, we observe that these compounds are also present in samples from other clades, in particular *Heteropterys oberdanii* species (clade H). Furthermore, a molecular family with a spectral match to a secoiridoid (Figure 39) was also observed, with part of the compounds widely distributed among all the clades, and part of them mainly detected in *Stigmaphyllon* species. Finally, molecular families relative to neolignans and furofuranoid lignans (Figure 40) were observed, widely distributed in Malpighiaceae clades, being this the first report of these classes of compounds in this family.

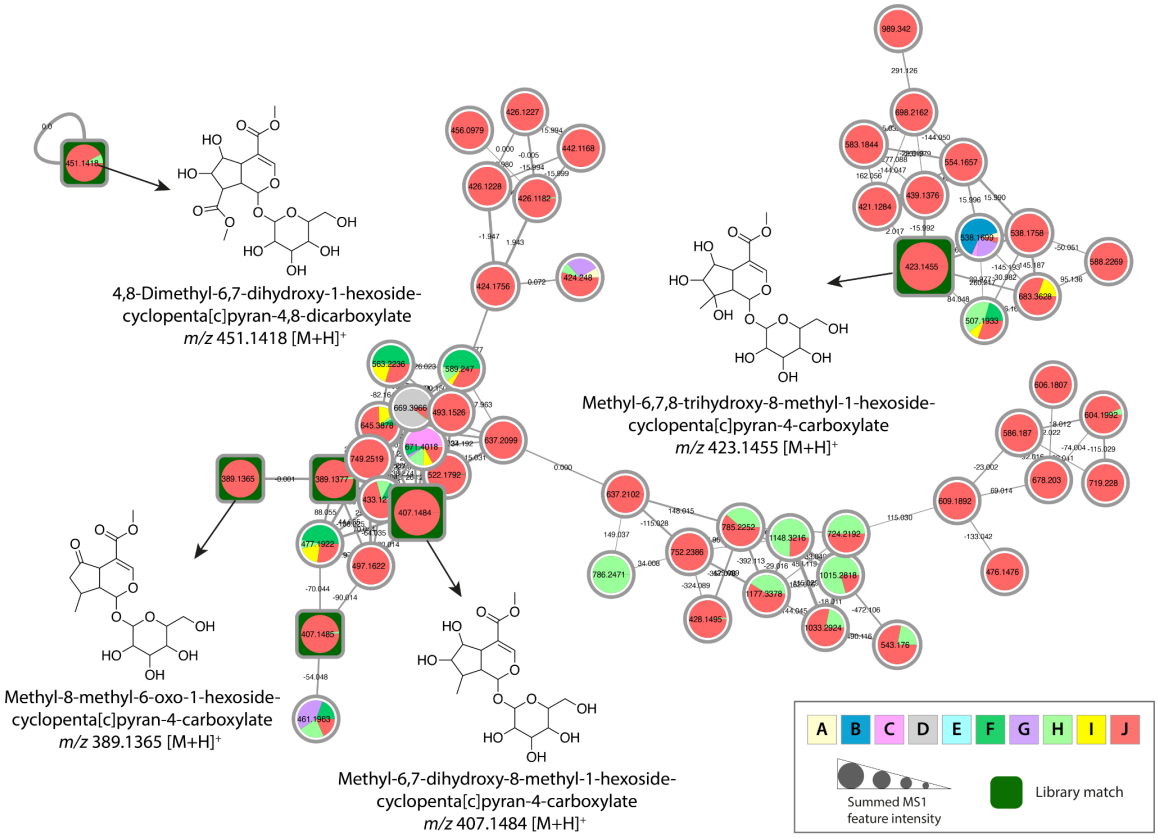


Figure 38 – Molecular families composed of iridoids (positive ionization mode).

Source: elaborated by the Author (2022).

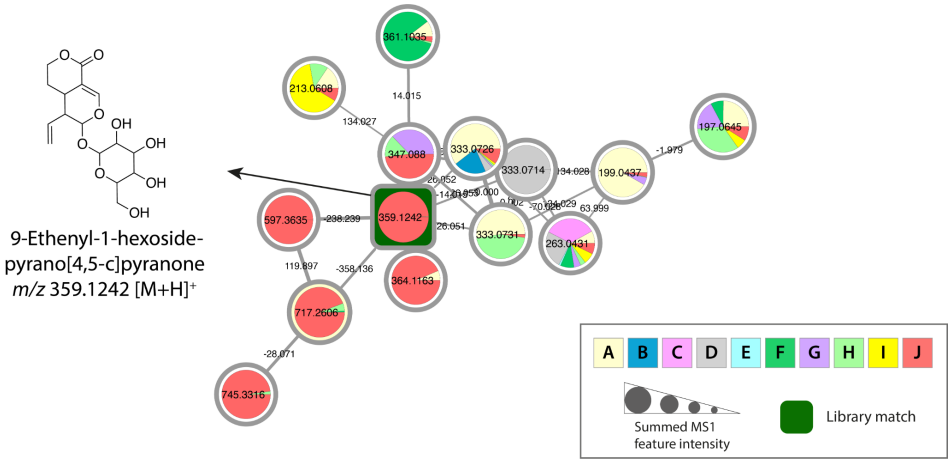


Figure 39 – Molecular family composed of secoiridoids (positive ionization mode).

Source: elaborated by the Author (2022).

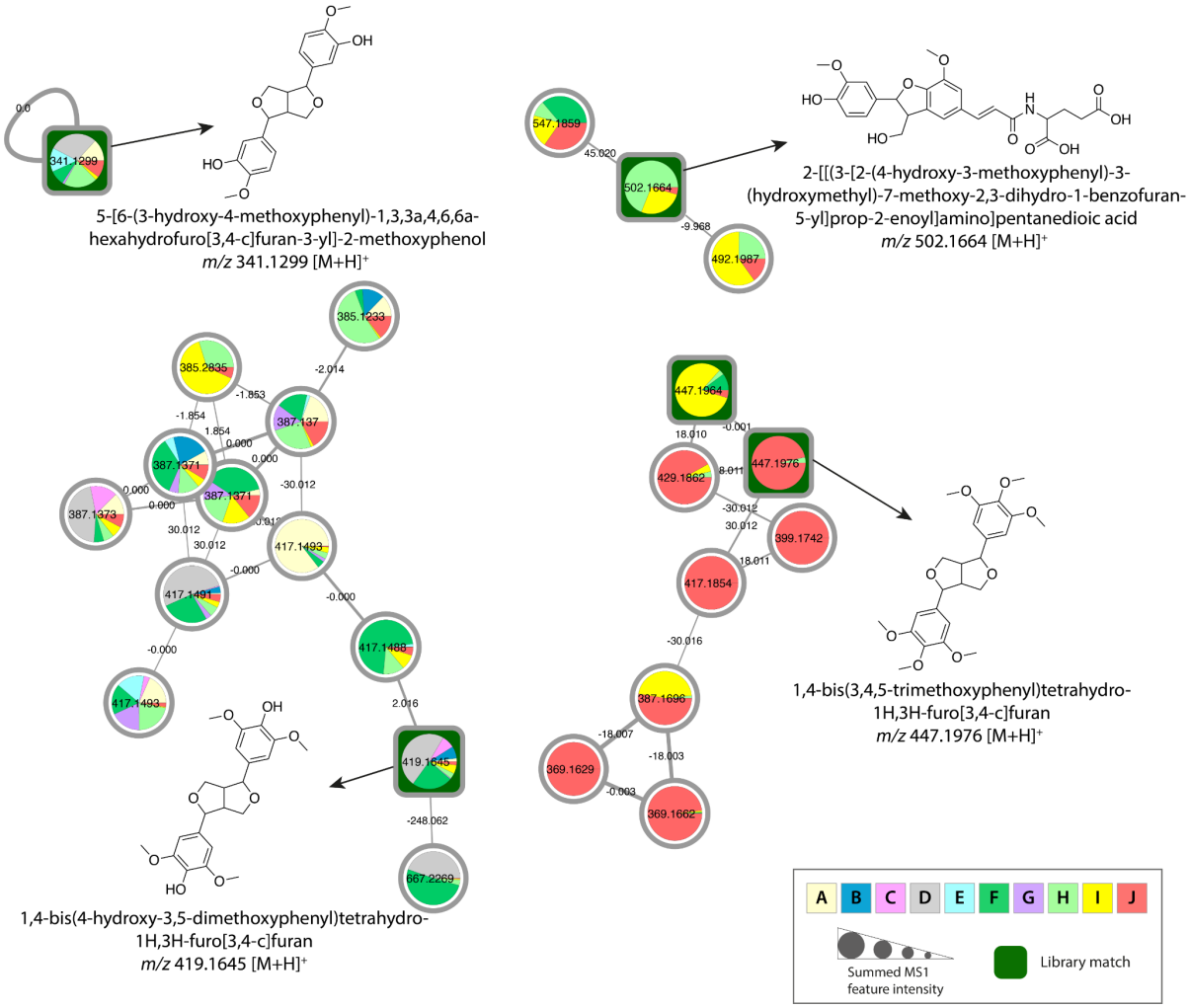


Figure 40 – Molecular families composed of neolignans and furofuranoid lignans (positive ionization mode).

Source: elaborated by the Author (2022).

The library matches obtained for the negative ionization mode and the molecular networks formed were also inspected, and these analyses showed mainly some of the same classes of compounds described for the positive ionization mode. Two networks stood out for presenting library matches to classes of compounds not observed for the positive ionization mode: proanthocyanins and lignans (Figure 41 and Figure 42). Proanthocyanins dimers, trimers, and tetramers were observed in all Malpighiaceae clades, corroborating previous reports *(FRAIGE et al., 2018; MANNOCHIO-RUSSO et al.*, 2020). Lignans have already been reported for *Tetrapterys mucronata* (QUEIROZ *et al*., 2014), and here we observe that, in addition to *T. mucronata* species (clade H), they are also present in other clades.

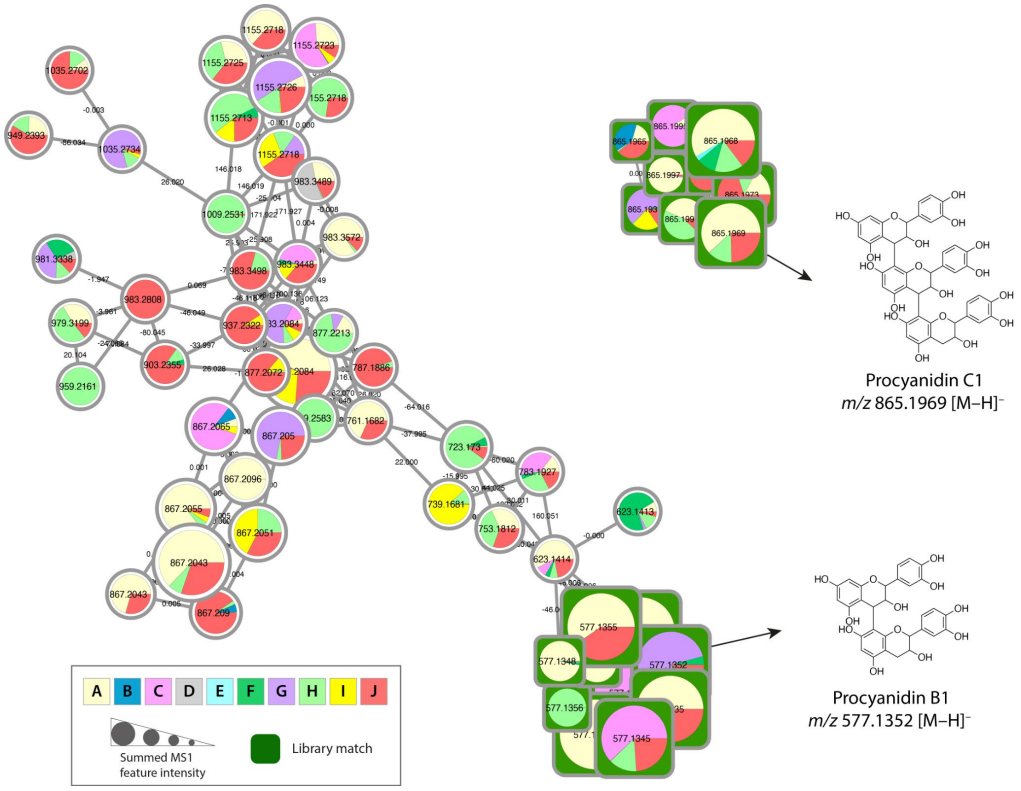


Figure 41 – Molecular families composed of condensed tannins (negative ionization mode).

Source: elaborated by the Author (2022).

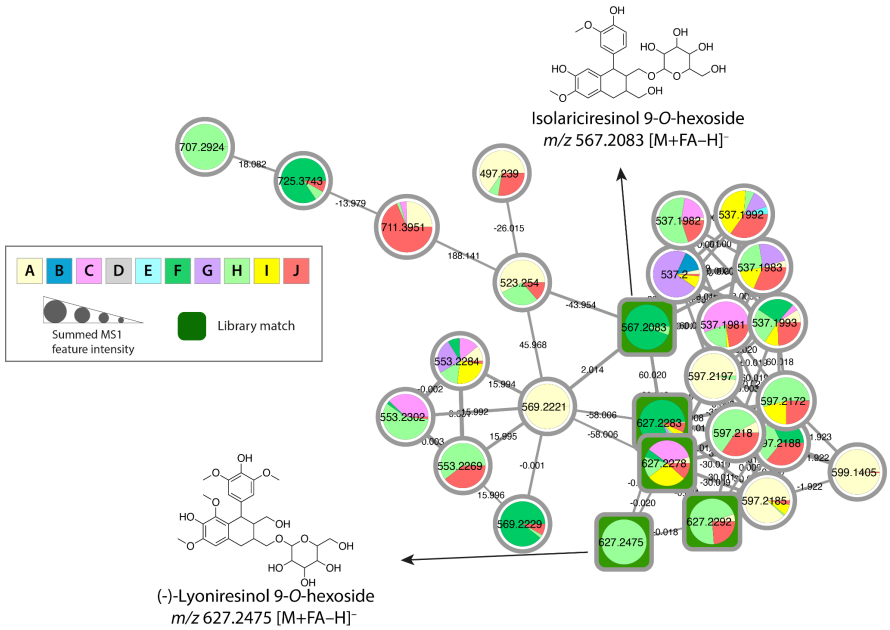


Figure 42 – Molecular families composed of lignans (negative ionization mode).

Source: elaborated by the Author (2022).

It is important to emphasize that the compounds discussed above represent only a part of the compounds detected in this study. In fact, many networks important for chemosystematics studies (*i.e.*, molecular families composed mainly or exclusively by nodes representing a specific clade/genera – Supplementary Figure 5 and Supplementary Figure 6), did not show any spectral match, indicating even potential new natural products. On the other hand, even considering only the spectral matches, it was possible to obtain important chemical information regarding the classes of compounds produced by these plant species. Another crucial point to be considered is that the sampling in this study comprises a larger number of samples from clades A, H, I, and J, and that these clades presented most of the spectral matches observed. It is possible that if a greater number of samples from other clades are included, more spectral matches can be retrieved to give additional insights on other clades as well.

3.3 Chemical hierarchy analysis

In order to amplify the chemical space from the Malpighiaceae dataset and have additional information about the classes of compounds detected, we used the Qemistree workflow combined with the CANOPUS classification tool for systematic annotation of the compounds classes (DÜHRKOP *et al.*, 2021; TRIPATHI *et al.*, 2021). For more information about the Qemistree workflow, see section 2.7. In this way, it was possible to construct a chemical hierarchy tree based on molecular fingerprints from MS/MS spectra and *in silico* classification tools. A total of 7,489 and 3,773 fingerprints were generated and classified at a superclass level for the positive and negative ionization modes, respectively. In this way, two chemical hierarchy trees were obtained, as shown in Figure 43A.

From the results obtained, it is evident that the ionization mode employed greatly influences the classifications obtained. At a superclass level, the ones most retrieved in the positive ionization modes were the "organic acids and derivatives", followed by "benzenoids" and "organoheterocyclic compounds". In contrast, most of the superclasses retrieved in the negative ionization mode consisted of "lipids and lipid-like molecules", "organic acids and derivatives", and "benzenoids" (Figure 44). At a CANOPUS class level, the "carboxylic acids and derivatives" was the main recovered class in both ionization modes. Ion features classified as "benzene and substituted derivatives", and "azoles" were also observed several times for the positive ionization

mode, while for the negative ionization mode, "organooxygen compounds" and "fatty acyls" were the second and third most abundant classifications.

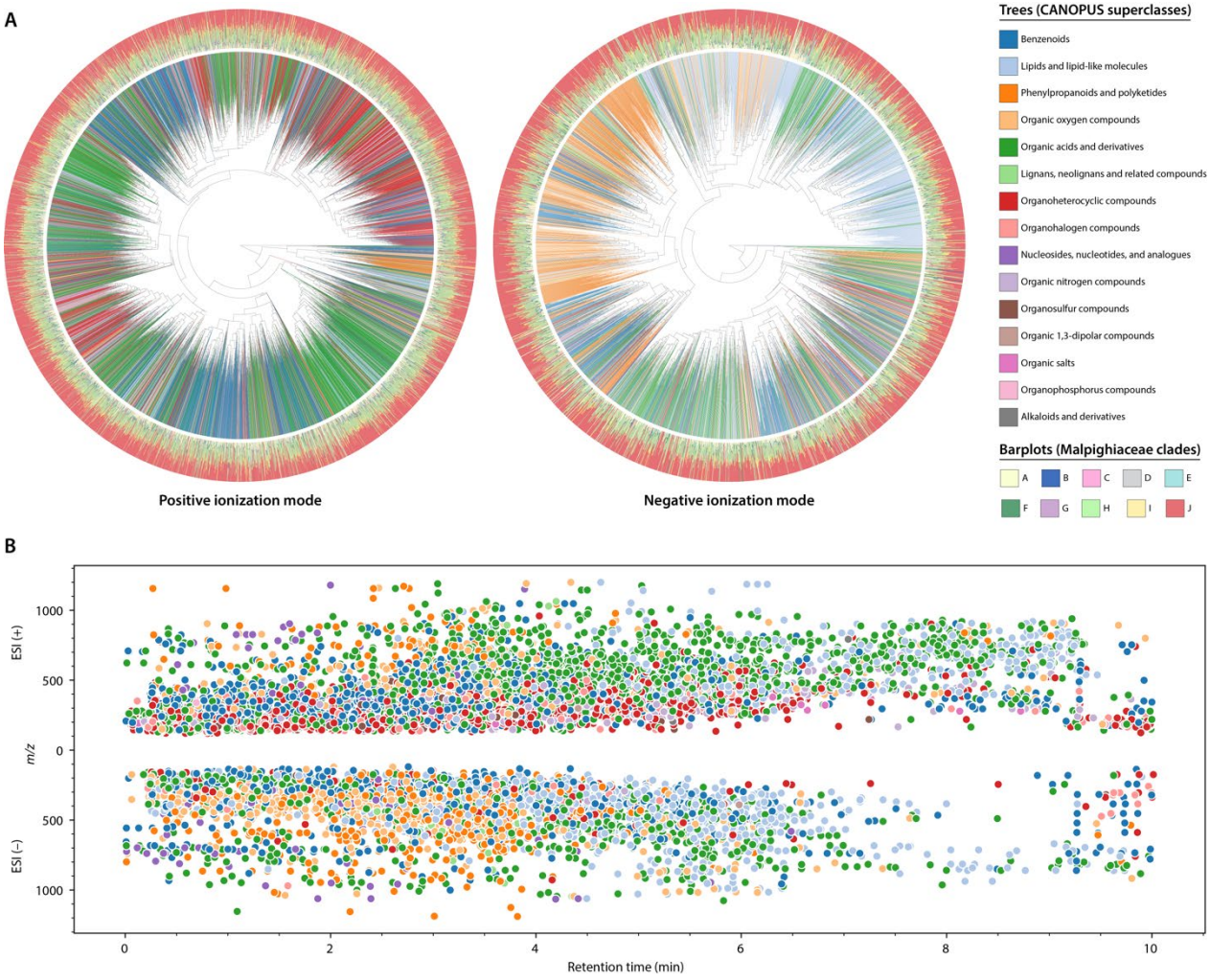


Figure 43 – *In silico* annotations obtained for the Malpighiaceae dataset from the Qemistree workflow combined with the CANOPUS classification tool. These are level 3 annotations according to the 2007 metabolomics standards initiative (SUMNER *et al.*, 2007). (**A**) Chemical hierarchies of the predicted molecular fingerprints from the Malpighiaceae plant samples analyzed in positive (left) and negative (right) ionization modes. The trees are pruned to keep fingerprints that were classified up to a superclass level in CANOPUS. The branch colors indicate the superclasses, while the bar plots of the outer ring indicate the relative abundance of a molecular fingerprint in each Malpighiaceae clade. (**B**) The ion features classified *in silico* are mapped based on the CANOPUS superclass (same colormap described in A). The x and y axes indicate the retention time and *m/z* value, respectively.

Source: elaborated by the Author (2022).

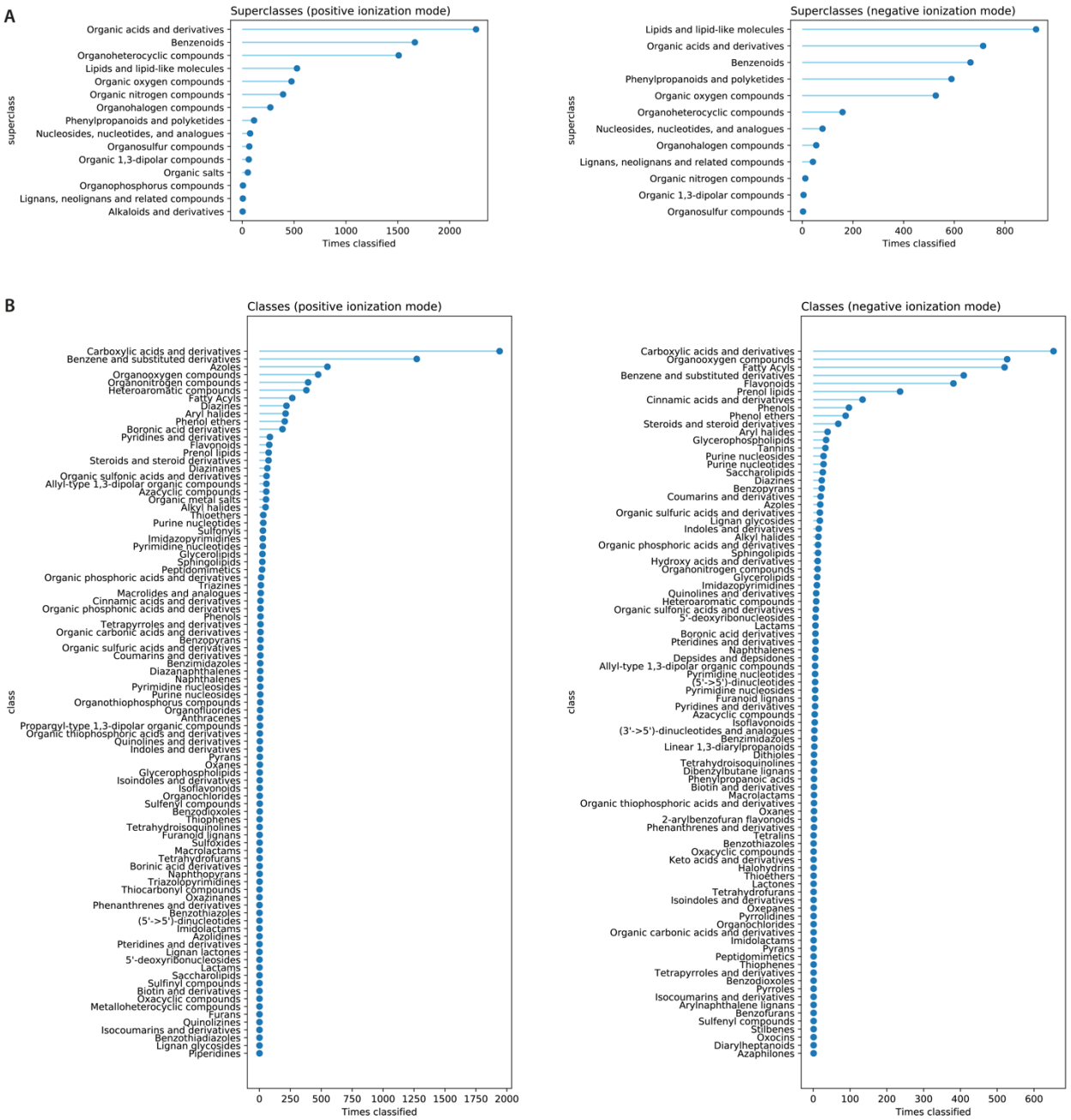


Figure 44 – Compound classification distribution within Malpighiaceae species in positive and negative ionization modes at a (**A**) CANOPUS superclass level and (**B**) CANOPUS class level.

Source: elaborated by the Author (2022).

These results confirm some of the conclusions drawn from the molecular networks, such as the presence of lipids and lipid-like molecules in several Malpighiaceae clades. The molecular networks and library searches in spectral libraries, combined with the *in silico* approaches based on structural databases allowed us to expand the

Malpighiaceae chemical space. Several hypotheses raised from the molecular networks were corroborated with the *in silico* classifications, giving higher confidence in these results.

The classifications obtained in the different ionization modes are also shown distributed in the chromatographic run (Figure 43B). In addition to being possible to observe differences in classifications between ionization modes, the retention time ranges also vary for specific superclasses. For instance, in the negative ionization mode, organic oxygen compounds, and phenylpropanoids and polyketides elute from near the dead volume to approximately 4 min, while the ion features classified as lipids and lipid-like molecules present higher retention times.

These results show that both ionization modes result in complementary chemical classifications, which is crucial for comprehensive chemotaxonomic investigations. In fact, the *in silico* tools used in this study relies on public spectral and structural databases, which are known to be more populated with data on the positive ionization mode (TRIPATHI *et al*., 2021). In this way, the results obtained for the negative ionization mode are less extensive compared to the positive ionization mode.

These results show that both ionization modes result in complementary chemical classifications, which is crucial for comprehensive chemotaxonomic investigations. In fact, the *in silico* tools used in this study relies on public spectral and structural databases, which are known to be more populated with data on the positive ionization mode (TRIPATHI *et al*., 2021). In this way, the results obtained for the negative ionization mode are less extensive than the positive ionization mode.

3.4 Ancestral character reconstructions for the classes of secondary metabolites annotated in Malpighiaceae

Ancestral character reconstruction analyzes were recently applied to a metabolomic study comprising a specific genus and alkaloid class (CHEN *et al*., 2020), being a promising approach for exploring the evolution of chemical compounds in a specific genus or family. This analysis statistically tests and recovers which characters, such as chemical classes, characterizes a given taxonomic group within the group studied (major clades recovered by DNA-based phylogenetic studies in Malpighiaceae). In the present study, the *in silico* classifications obtained were used to map the evolutionary history of the classes retrieved in our analyses using the

maximum likelihood criteria in the Malpighiaceae molecular phylogeny (Figure 45). The classifications obtained in the positive and negative ionization modes were combined to provide a general overview of the total number of chemical classes obtained for the family. Classes retrieved as homoplasies or synapomorphies for the major Malpighiaceae clades (A–J) are described in Box 2, while the ones retrieved for all Malpighiaceae clades and genera are extensively presented in Supplementary Box 2. From the 113 *in silico* classes retrieved, 35 were present in all genera, such as fatty acyls, flavonoids, glycerolipids, phenols, and purine nucleosides, which are compounds widely distributed in plants, with a variety of ecological roles in these organisms.

Regarding the ten major clades of Malpighiaceae, all of them were recovered with at least one homoplastic or synapomorphic class of metabolites supporting them (Box 2, Supplementary Box 2). The ten major clades in Malpighiaceae were characterized by the presence and absence of 22 and 23 classes of metabolites (i.e., homoplasies), respectively, with the absence of glycerophospholipids being recovered as a synapomorphy of Clade D (Barnebyoid). It is worth mentioning that even though most classes of secondary metabolites circumscribing all the ten major clades of Malpighiaceae were recovered as homoplasies, their presence or absence was recorded exclusively for each clade, with reversions (*i.e.*, parallelisms) only being recorded within a few distantly related subclades or genera. The results retrieved from this analysis corroborate some of the conclusions obtained from the molecular networks and from the *in silico* classifications. For instance, glycerophospholipids, fatty acids, fatty esters, furofuranoid lignans, and prenol lipids classes were widely distributed in our molecular networking analyses for all phylogenetic clades, and these classes were also extensively observed in the ancestral character reconstructions (i.e., appeared in different geological times). In addition, it also clearly shows which classes were important to circumscribe specific clades/genera.

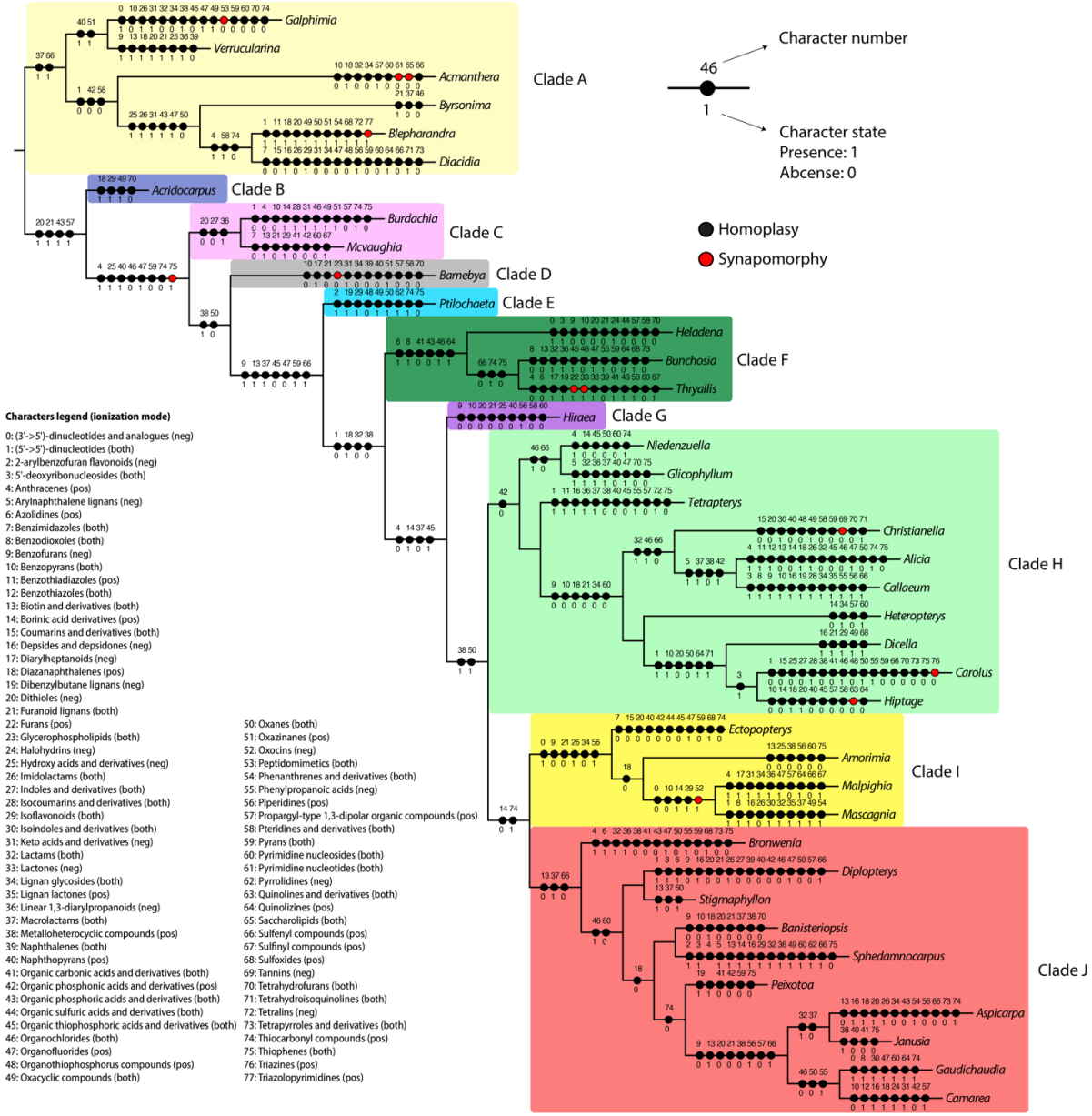


Figure 45 – Summary of the maximum likelihood ancestral state reconstruction for the *in silico* classifications obtained at a class level. Each chemical class was treated as a character (0 to 77), and character states were binary-coded for each genera (1: present; 0: absent). Black and red circles represent homoplasies and synapomorphies, respectively. Clades highlighted represent the Malpighiaceae major clades recognized by recent molecular phylogenetic studies (DAVIS; ANDERSON, 2010).

Source: elaborated by the Author (2022).

Box 2 – Characters retrieved from the ancestral characters reconstruction (clades) based on the classifications obtained *in silico* for Malpighiaceae samples. The ionization mode in which each classification was obtained is described (pos = positive ionization mode; neg = negative ionization mode; both = both ionization modes).

	Classes present†	**Classes absent**
Clade A (Byrsonimoid clade)	Macrolactams (both); Sulfenyl compounds (pos)	–
Clade B (Acridocarpoid clade)	Diazanaphthalenes (pos); Isoflavonoids (both); Oxacyclic compounds (both)	Tetrahydrofurans (both)
Clade C (Mcvaughioid clade)	Linear 1,3-diarylpropanoids (neg)	Dithioles (neg); Indoles and derivatives (both)
Clade D (Barnebyoid clade)	Diarylheptanoids (neg); Keto acids and derivatives (neg); Oxazinanes (pos)	Benzopyrans (both); Furanoid lignans (both); Glycerophospholipids (both)§, Lignan glycosides (both); Naphthalenes (both); Naphthopyrans (pos); Propargyl-type 1,3-dipolar organic compounds (pos); Pteridines and derivatives (both); Tetrahydrofurans (both)
Clade E (Ptilochaetoid clade)	2-arylbenzofuran flavonoids (neg); Dibenzylbutane lignans (neg); Isoflavonoids (both); Oxacyclic compounds (both); Oxanes (both); Pyrrolidines (neg); Thiocarbonyl compounds (pos)	Organothiophosphorus compounds (pos); Thiophenes (both)
Clade F (Bunchosioid clade)	Azolidines (pos); Benzodioxoles (both); Organochlorides (both); Quinolizines (pos)	Organic carbonic acids and derivatives (both); Organic phosphoric acids and derivatives (both)
Clade G (Hiraeoid clade)	Piperidines (pos)	Benzofurans (neg); Benzopyrans (both); Dithioles (neg); Furanoid lignans (both); Hydroxy acids and derivatives (neg); Naphthopyrans (pos); Pteridines and derivatives (both); Pyrimidine nucleosides (both)
Clade H (Tetrapteroid clade)	–	Organic phosphonic acids and derivatives (pos)
Clade I (Malpighioid clade)	(3'→5')-dinucleotides and analogues (neg); Imidolactams (both); Piperidines (pos)	Benzofurans (neg); Furanoid lignans (both); Lignan glycosides (both)
Clade J (Stigmaphylloid clade)	Macrolactams (both)	Biotin and derivatives (both); Sulfenyl compounds (pos)

†**Classes retrieved as present in all Malpighiaceae clades:** Alkyl halides (both); Allyl-type 1,3-dipolar organic compounds (both); Aryl halides (both); Azacyclic compounds (both); Azoles (both); Benzene and substituted derivatives (both); Boronic acid derivatives (both); Carboxylic acids and derivatives (both); Cinnamic acids and derivatives (both); Diazinanes (pos); Diazines (both); Fatty Acyls (both); Flavonoids (both); Glycerolipids (both); Heteroaromatic compounds (both); Imidazopyrimidines (both);

Macrolides and analogues (pos); Organic metal salts (pos); Organic sulfonic acids and derivatives (both); Organonitrogen compounds (both); Organooxygen compounds (both); Phenol ethers (both); Phenols (both); Prenol lipids (both); Purine nucleosides (both); Purine nucleotides (both); Pyridines and derivatives (both); Sphingolipids (both); Steroids and steroid derivatives (both); Sulfonyls (pos); Thioethers (both).

[§]Synapomorphy.

Source: elaborated by the Author (2022).

Our chemotaxonomic approach based on MS/MS analyses of Malpighiaceae plant samples allowed us to obtain a comprehensive overview of the classes of secondary metabolites produced by this taxon. Plant secondary metabolites are known to show patterns of occurrence in certain taxa (WINK, 2003). However, it is important to emphasize that many factors influence chemical diversity, such as genetic variation and environmental influences (soil nutrients, humidity, herbivory, ecological interactions, among others) (ISAH, 2019). Our study provides a starting point for follow-up and systematic evaluation of such factors, and in-depth studies must be conducted to confirm and expand these chemotaxonomic conclusions. In addition, our efforts were directed to obtain chemical information at a major phylogenetic clade and genus level. There might be discrepancies if the ancestral character state reconstructions are obtained at different taxonomic levels, such as intrageneric and species levels.

The relevance of our evolutionary approach to the study of secondary metabolites can be evidenced if we consider three relevant classes of plant secondary metabolites as examples, such as furanoid lignans, isoflavonoids, and piperidines. The absence of the furanoid lignans in clades D (Barnebyoid), G (Hiraeoid), and I (Malpighioid) is regarded as an informative homoplasy in our analyzes since this information can be used to point out which clades in the family one must focus on in future studies to search for this class of metabolites. This information can also be used in chemosystematic studies of Malpighiaceae to chemically characterize these clades and help to establish a new classification system based on chemical compounds and morphology (ALMEIDA *et al*., 2017). On the other hand, the presence of isoflavonoids in clades B (Acridocarpoid) and E (Ptilochaetoid), and piperidines in clades G (Hiraeoid) and I (Malpighioid) are also informative homoplasies that can be interpreted in the same light as the furanoid lignans.

The chemical characterization of all Malpighiaceae clades is the first step towards enabling a new research line on the evolution of secondary metabolites in this plant

family since this family already has a dated and calibrated molecular phylogeny available in the literature (DAVIS *et al.*, 2014; WINK, 2003). Merging both analyses would allow us to infer the geologic time in which all classes of secondary metabolites have arisen in Malpighiaceae and correlate these dates with past biogeographic events, such as the colonization of different biomes by the most recent common ancestor of all analyzed clades (ALMEIDA; AMORIM; VAN DEN BERG, 2018). Additionally, it would also be possible to evaluate which Malpighiaceae lineages experienced a higher diversification throughout the geological time and identify which classes of secondary metabolites are correlated with these diversification events (XI *et al*., 2012).

4 Conclusions

Metabolomics analyses based on tandem mass spectrometry and bioinformatics tools have enabled a more comprehensive investigation of the metabolites produced by organisms, and have been increasingly used for this purpose. In fact, due to the low amount of material necessary, it is possible to investigate entire families for chemoevolutionary studies based on the remaining samples from molecular phylogenies used for total DNA extraction. Even though there was a significant advance in this field, some tendencies can be obtained and considered. The ionization mode and extraction protocols must be carefully evaluated since these factors influence the results, especially when aiming at chemotaxonomic investigations. Our results showed that positive and negative ionization modes lead to complementary results both in library searches and *in silico* classification tools. However, as the public libraries and structural databases are more populated with data acquired in the positive ionization mode, less extensive information can be retrieved from analyses performed in the negative ionization mode. More complete and precise results will undoubtedly be obtained for chemotaxonomic studies as these databases get more populated and new bioinformatic tools are developed. Similarly, conclusions obtained from *in silico* approaches must be confirmed with complementary techniques, and the classical methodologies are of great value for deeper investigations.

Our study explored several Malpighiaceae plant species, genera, and clades for the first time, which significantly improved the chemical knowledge of this botanical family. We hope that our findings guide future studies in Malpighiaceae since we reported evidence of specific classes of compounds that most likely occur in specific

clades or genera. Therefore, if a particular chemical class is of interest for presenting specific biological activities, future studies can focus their search on specific taxonomic groups. All the software and libraries used in this study are publicly available, making this workflow accessible to be reproduced in other taxa. In addition, we expect that the workflow followed in this study will be used for future studies in several fields, such as chemotaxonomy, metabolomics, chemical ecology, and the discovery of new natural products.

Adapted with permission from (MANNOCHIO-RUSSO *et al.* 2022). Copyright 2022 Frontiers.

CHAPTER 3 – SEARCH FOR BIOACTIVE COMPOUNDS IN MALPIGHIACEAE SPECIES WITH TOXIC EFFECTS TO AEDES *AEGYPTI LARVAE*

Manuscript title: Mass spectrometry-based untargeted metabolomics reveals unusual larvicidal nitropropanoyl glycosides from *Heteropterys umbellata* (Malpighiaceae).

Target Journals: Journal of Natural Products; Phytochemistry; Scientific Reports

Status: Unpublished.

Authors: Helena Mannochio-Russo[1], Wilhan D. G. Nunes[2], Rafael F. Almeida[3], Lorena Albernaz[4], Laila S. Espindola[4], Vanderlan S. Bolzani[1]

Affiliations:

[1]NuBBE, Department of Biochemistry and Organic Chemistry, Institute of Chemistry, São Paulo State University (UNESP), SP, 14800-901 Araraquara, SP, Brazil.

[2]Federal Institute of Education, Science and Technology of São Paulo (IFSP), Ilha Solteira, SP, 15385-000, Brazil.

[3]Royal Botanical Gardens Kew, Science, Ecosystem Stewardship, Diversity and Livelihoods, Richmond, Surrey, TW9 3AE, United Kingdom.

[4]Laboratório de Farmacognosia, Universidade de Brasília (UnB), Campus Universitário Darcy Ribeiro, Brasília, DF, 70910-900, Brazil.

Abstract

The *Aedes aegypti* (Diptera: Culicidae) mosquito is the vector of several arboviruses in the tropical and subtropical areas of the globe, and the search for alternatives to fight this vector is considered an urgent task. Several studies searching for natural products to fight *Ae. aegypti* have been reported to date; however, synthetic pesticides are still the most widely used for this purpose. This study describes the investigation for larvicidal compounds from the Malpighiaceae botanical family, which has been reported to present toxic compounds in several species. Our workflow initially consisted of an extensive larvicidal screening, which led to the selection of *Heteropterys umbellata* for deeper chemical and biological investigation. By employing mass spectrometry-based metabolomics tools and multivariate analyses, we determined that the metabolic profile in stems and leaves, and the collections performed in Minas Gerais (MG) *vs.* São Paulo (SP) states, differed significantly. This metabolomic-guided approach led to the isolation of nitropropanoyl glycosides with larvicidal activity, in addition to the indication of possible synergistic effects of their isomers. Additionally, quantifying these compounds in different extracts raised several insights regarding their production in this plant. These results support the metabolomic-guided approach to search for bioactive compounds in the natural products workflow.

Keywords: molecular networking; metabolite annotation; nitropropanoyl esters of D-glucopyranose; nitro compounds; targeted isolation; method validation; LC-MS/MS.

1 Introduction

Natural products represent an important source of bioactive compounds, with thousands of reports in literature from the most diverse organisms with a wide variety of biological activities. In fact, most of the commercialized drugs in the market today are derived or inspired from natural compounds (ATANASOV *et al*., 2021). Plants are among the most studied natural organisms for this purpose, with reports of medicinal use since ancient times. For several decades, the phytochemical investigations relied on the so-called classical methodologies, aiming at the extraction, isolation, and bioactivity evaluation of natural or semisynthetic compounds. This kind of approach undoubtedly led to important discoveries, which significantly improved the knowledge in the field. However, this workflow can oftentimes be costly, time-consuming, and have poor environmental impacts, even leading to the re-isolation of well-known compounds.

There was a paradigm shift in natural products investigations in the last decade, mainly due to the development of increasingly sensitive detection techniques (such as Mass Spectrometry – MS) and computational tools to assist in data exploration prior to the isolation step. Therefore, with metabolomics analyses, it is possible to gather a significant amount of information regarding the dataset (or even of a specific molecule) to guide the study towards answering the hypothesis raised (BAUERMEISTER *et al*., 2022). This approach is particularly useful for exploring unprecedented natural products or reinvestigating a particular compound to a new biological activity. In addition to guiding the isolation of natural compounds, metabolomic tools also present a crucial point in understanding the possible metabolic and ecological differences in distinct groups even of a single species. For instance, indicating the metabolites potentially responsible for differentiation plant organs or collection sites can raise several hypotheses of possible ecological roles (LI; GAQUEREL, 2021), besides determining in which particular group a compound of interest is produced in higher amounts.

Arboviruses are a concern in several tropical and subtropical countries around the globe, and the search for natural products as an alternative to control their vectors is increasing significantly (DEMARQUE; ESPINDOLA, 2021; SILVÉRIO *et al*., 2020; VALLI *et al*., 2021). *Aedes aegypti* (Diptera: Culicidae) mosquito, for instance, is the vector of several arboviruses such as dengue, Zika, chikungunya, and yellow fever

(BENELLI; JEFFRIES; WALKER, 2016; SILVÉRIO *et al*., 2020) in Brazil, and so far, the most effective way to control these diseases is to combat the vector. Synthetic pesticides, such as neonicotinoids, oxadiazines, organophosphates, pyrethroids, and carbamates, are still the most widely used for this purpose (REZENDE-TEIXEIRA *et al*., 2022; SILVÉRIO *et al*., 2020), but natural products are emerging as a favorable alternative. Several studies regarding plants from the Brazilian biodiversity have shown promising results against *Ae. aegypti* (MAGALHAES *et al*., 2021; MELO *et al*., 2021; SILVA *et al*., 2021, 2022). In this way, the yet considered underexplored Brazilian biodiversity represents a valuable source of still-to-be-discovered compounds for this target.

The Malpighiaceae botanical family stands out among the several chemically underexplored plant families with a wide distribution in Brazilian territory. Even though Malpighiaceae species present some well-studied species, with a meaningful chemical diversity of compounds reported, a significant percentage of this family has not been deeply explored regarding their metabolites and biological activities. Several Malpighiaceae species are known for presenting toxic constituents, such as *Banisteriopsis caapi* (SAMOYLENKO *et al*., 2010), *Tetrapterys mutiglandulosa* (QUEIROZ et al., 2014), *Niedenzuella multiglandulosa* (MANNOCHIO RUSSO *et al*., 2020) and *Amorimia* species (LEE *et al*., 2012). A previous study with *Hiptage benghalensis* extracts showed that this species presented promising insecticidal and repellent activity against mosquito vectors (specifically, *Anopheles barbirostris*, *Culex quinquefasciatus*, and *Aedes albopictus*) in South East Asian countries (LALROTLUANGA *et al.*, 2012). Therefore, neotropical Malpighiaceae species constitutes of an interesting source in searching for natural products toxic to *Ae. aegypti*.

Therefore, the main objective of this investigation was to search for bioactive natural products from the Malpighiaceae botanical family toxic to *Ae. aegypti* larvae. In our workflow (Figure 46), an initial screening of about 400 extracts was initially performed to determine the more promising species for a metabolomic-guided approach. This step led to the more profound evaluation of *Heteropterys umbellata*, a species without any deeper chemical information. Different plant organs and collection sites were investigated using an untargeted metabolomics approach and multivariate analysis, aiming for a deeper understanding of these factors in the metabolic profile of *H. umbellata*. With this approach, we were able to isolate unusual secondary

metabolites, determine their bioactivity, validate the analysis method, quantify these compounds in different groups, and infer possible ecological roles for these compounds.

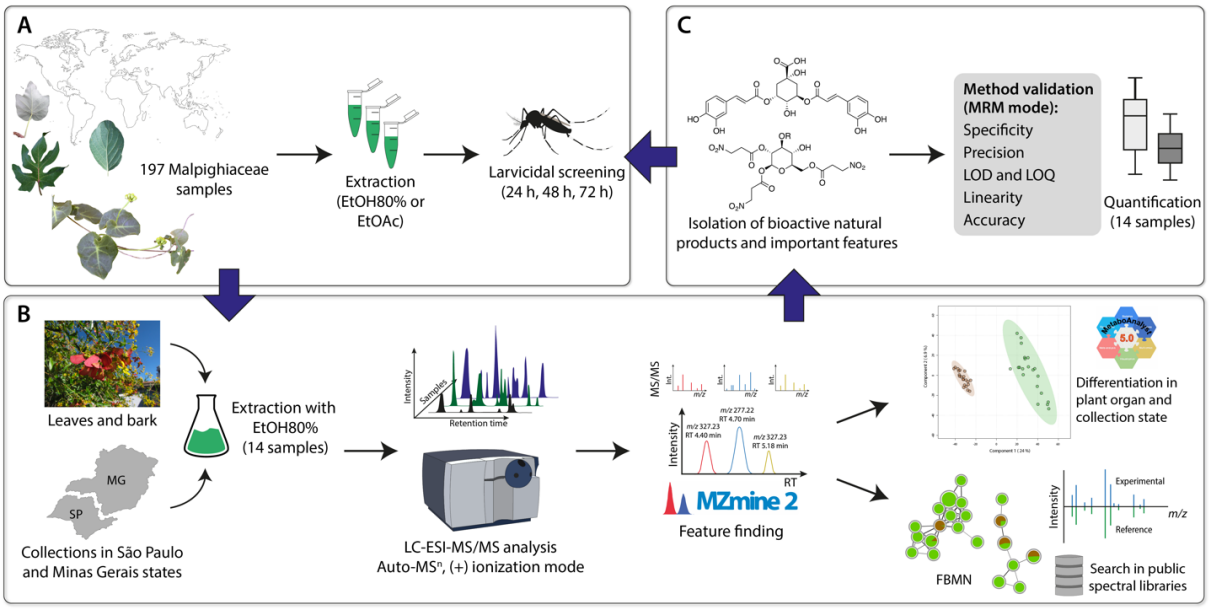


Figure 46 – General workflow for the identification of bioactive compounds with larvicidal activity. (**A**) Initially, a collection of 197 samples from Malpighiaceae species leaves were extracted with solvents of different polarity, and all extracts were subjected to a larvicidal screening against *Aedes aegypti*. (**B**) *Heteropterys umbellata* was selected for a deeper study for being the most active extract. Both stem and leaves were collected in different locations in São Paulo and Minas Gerais states in Brazil, yielding 14 extracts. These extracts were analyzed through LC-MS/MS for untargeted metabolomics, in which the data generated was processed and subjected to multivariate analysis, molecular networks and library searches. (**C**) Three features important for group differentiation and/or without library matches were isolated. These compounds were used to validate the analytical method and to quantify these compounds in the different collection points and plant organs. The isolated compounds were once again evaluated for biological activity.

Source: elaborated by the Author (2022).

2 Materials and Methods

2.1 General experimental procedures

All solvents used for LC-MS/MS analysis were HPLC grade. Acetonitrile and formic acid 88% (v/v) were purchased from J.T. Baker (J.T. Baker-Avantor, Radnor, USA), while ultrapure water was obtained using a Millipore water purification system (Millipore, USA). For extraction purposes, analytical grade ethanol (98%) was obtained from Êxodo Científica, and ethyl acetate was acquired from Synth and further destilated. The NMR spectra were recorded in a Bruker Avance III HD 600 MHz NMR spectrometer equipped with a Triple Inverse TCI 5 mm Cryo-probehead and automated sample changer SampleXpress. The structural elucidation was obtained through 1D and 2D-NMR experiments (^{1}H, ^{13}C, 1D-TOCSY, 1D-NOESY, COSY, HSQC, and HMBC). For compound **1**, spectra were obtained in DMSO-d_6 (calibrated considering the residual solvent residual signals δH 2.50 and δC 39.5), while for **2** and **3**, spectra were obtained in CD_3CN-d_3 (calibrated considering the residual solvent residual signals δH 1.94 and δC 1.3). DMSO-d_6 was acquired from Sigma-Aldrich (St. Louis, USA), and CD_3CN-d_3 was acquired from Cambridge Isotope Laboratories, Inc (Tewksbury, USA).

2.2 Plant material

Leaves from 194 samples of Malpighiaceae species were collected between 2010 and 2018. Collection sites, dates, and voucher codes are listed in Supplementary Box 1. The leaves were dried in a desiccator containing silica gel at room temperature, frozen in liquid nitrogen, and grounded in an analytical mill.

Heteropterys umbellata stems and leaves were collected in seven different locations (three in Minas Gerais state, and four in São Paulo state) in January 2021. The plant material was dried in an oven at 40 ˚C, frozen in liquid nitrogen, and grounded in an analytical mill. The material was identified by Dr. R. F. Almeida and voucher specimens (H. Mannochio-Russo1 – H. Mannochio-Russo7) were deposited in the herbarium of the Federal University of São Carlos (SPSC). More information regarding the collections can be found in Supplementary Box 3, and the morphological details of this species are shown in Supplementary Figure 7. The authorization for

conducting this study was granted by SisGen (no. A6FDC2E) – Brazilian Ministry of Environment.

2.3 Extraction procedures

During the initial screening for larvicidal bioactivity, the extraction procedure consisted of transferring 20 mg of each powdered plant material to 2.0 mL Eppendorf tubes along with 1 mL of extraction solvent [EtOH–H_2O (80:20, v/v) or EtOAc (100%)], totalizing 388 extracts. The samples were sonicated for 15 min, centrifuged for 15 min at 5000 *g*, and supernatants were transferred to clean sample tubes and dried in N_2 gaseous flux.

For the *H. umbellata* study, the leaves and stem samples regarding the seven collection points were processed individually, resulting in 14 samples. The dried and pulverized plant materials were extracted with EtOH–H_2O (80:20, v/v) in a proportion of 30 g of dried plant material to 200 mL of extraction solvent. The extraction was performed by sonication for 1 h, in triplicate. Each extract was concentrated under reduced pressure by rotatory evaporation and further lyophilized. This procedure resulted in an average yield range of 24% for leaves crude extracts, and 8% for stem extracts – for a full list of extraction yields, see Supplementary Table 10.

2.4 Bioactivity evaluation

All bioassays were performed with the *Aedes aegypti* Rockefeller strain. Third instar larvae (L3), pupae and female adults were obtained from infection-free colonies mantained in the insectary of the Pharmacognosy Laboratory, at the University of Brasília (UnB) (SILVA *et al*., 2020). Larvae were obtained from non-infected colonies maintained by the Pharmacognosy Laboratory. The maintenance of the colony is in accordance with the guidelines of the World Health Organization (WHO).

2.4.1 Larvicidal bioassay

The initial screening for larvicidal activity in the 194 Malpighiaceae samples (394 extracts) was performed as previously described (SILVA *et al*., 2020). Briefly, samples

were tested in triplicate in 12-well plates containing 10 L3 stage larvae, 3 mL of water, and 50 µL of the sample or negative control (<2% dimethyl sulfoxide). Temephos was used as a positive control (0.35 µg mL^{-1}). The hydroethanolic extracts were tested at 250 µg mL^{-1}, while the ethyl acetate extracts were tested at 150 µg mL^{-1}. Larvae mortality was determined after 24, 48, and 72 h of incubation at room temperature. Extracts that resulted in >80% of larval mortality were considered active and were tested one more time for bioactivity confirmation.

The larvicidal tests performed with the *H. umbellata* fraction were tested at a concentration of 125 µg mL^{-1}, while the pure compounds were tested at 22.5, 45, 90 and 150 µg mL^{-1}.

2.4.2 Pupicidal bioassay

The pupicidal assays were performed in quadruplicate in 12-well plates containing 10 pupae, 3 mL of tap water, and 50 µL of the sample or negative control (<2% dimethyl sulfoxide) (SILVA *et al*., 2020). The isolated compounds were tested at 150 µg mL^{-1}, and the mortality was evaluated after 24 h and 48 h of incubation at room temperature. The samples with mortality higher than 80% were considered active.

2.4.3 Adulticide bioassay

The adulticide bioassay was conducted using Ae. aegypti based on the CDC bottle assay (Centers for Disease Control and Prevention, 2019). The isolated compounds were tested in quadruplicate in Scott bottles (250 mL). The bottles used for this test were internally coated with 1 mL of each sample solution solubilized in acetone (250 µg/bottle). Each bottle was impregnated with the sample diluted in acetone and pipetted enough to achieve 250 µg per bottle. After 2 h in a hood and complete solvent drying, 10 to 20 females were added in each bottle. The readings were performed after 0, 15, 30, 45, 60, 75, 90, and 120 min. For each replicate, the number of dead females and the number of females that did not fly were recorded. Acetone was used as negative control, and malathion (50 µg/bottle) as positive control.

2.5 HPLC-MS/MS analyses

The extracts were analyzed in an HPLC Prominence (Shimadzu®, Japan) coupled to an ion trap mass spectrometer amaZon SL Bruker Daltonics® (Bruker, Billerica, MA, USA), equipped with electrospray ionization (ESI) for dereplication purposes. The HPLC system consisted of an LC-20AD pump, DGU 20-A3R online degassing unit, SPD-M20A diode array detector (200-800 nm), CTO-20A column oven, SIL-20AHT automatic injector, and CBM-20 communication module. A Phenomenex® C18-Luna (250 mm × 4.6 mm, 5.0 µm) column was employed as stationary phase and the solvent system consisted of water (A) and acetonitrile (CH_3CN, B), both acidified with 0.1% v/v formic acid, a 1.0 mL/min flow rate, 35˚C column temperature, and 2 µL of injection volume. The extracts were solubilized in CH_3CN–H_2O (1:1, v/v) at the concentration of 1.0 mg mL^{-1}, then centrifuged for 5 min at 4500 × *g* and the supernatants were analyzed. The gradient employed for the analyses consisted of a gradient of 5 to 83% B in 33 min. The mass spectra were obtained in the positive ionization mode considering a mass range of 50–1200 Da and applying auto-MS^n mode (n = 3) for exploratory analysis. The analyses regarding the method validation were acquired with the same mass range using the Multiple Reaction Monitoring (MRM) mode. In the MRM analyses, the transitions 499.10 → 319.10, 501.10 → 365.10, 602.20 → 466.10, and 296.10 → 278.00 were chosen to monitor compounds **1**, **2**, **3**, and diclofenac, respectively. The mass spectrometer source parameters were set as follows: capillary voltage at 4.5 kV, nitrogen used as the nebulizing and drying gas (50 psi, 10 L/min, 300˚C). The data were processed through Bruker Compass Data Analysis 4.3® software.

2.6 UHPLC-HRMS analysis

The fractions were solubilized in CH_3CN–H_2O (1:1) at a concentration of 1 mg mL^{-1} and further filtered in a 0.22-µm filter (hydrophobic PTFE). The samples were analyzed in an UHPLC-MS/MS system (Bruker Daltonics, Elute pump UHPLC, Elute autosampler UHPLC, Elute DAD) coupled to a QToF mass spectrometer (Compact QToF, Bruker Daltonics) equipped with electrospray (ESI) ionization source, and controlled by Bruker oToF control. The mobile phase consisted of CH_3CN (A) and H_2O (B), both acidified with 0.1% formic acid, while the column Phenomenex® C18(2)-Luna

column (150 × 4.6 mm i.d.; 5 µm), was chosen as stationary phase. The oven temperature, the injection volume, and the flow rate were set to 40 ˚C, 5 µL, and 0.5 mL min^{-1}, respectively. A linear gradient from 5 to 100% B in 8 min was employed. The mass spectra were obtained in the positive ionization mode considering a mass range of 50–1300 Da. The source parameters were set as follows: capillary voltage of 4.5 kV, with nitrogen used as the nebulizing and drying gas (1.8 bar, 9.0 L min^{-1}, 220 ˚C). The quadrupole cone voltage was set to 4.0 eV, and the collision energy to 7.0 eV. The three most abundant ions in each scan were selected for MS/MS, and excluded after 3 subsequent acquisitions. The data were processed through Bruker Compass Data Analysis 5.1® software.

2.7 MS/MS data processing

The raw data files (.d) obtained from the ion trap analyzes were converted to .mzML format using DataAnalysis 4.3® software (Bruker). The .mzML files were processed in MZmine2 (version 2.53) for positive ionization mode. The parameters used for feature finding were as follows: mass detection (centroid, 1.0E4 and 1.0E3 for MS1 and MS2, respectively); ADAP chromatogram builder (minimum group size in number of scans of 2, group intensity threshold and minimum highest intensity set to 5.0E4, and tolerance of 0.5 *m/z*); chromatogram deconvolution (baseline cut-off algorithm: minimum peak height: 1.0E5, peak duration range: 0.01 to 2 min; baseline level: 5.0E5) with median *m/z* center calculation, and *m/z* range for MS2 scan pairing of 0.5 Da; isotope peaks grouper (tolerance set at 0.5 *m/z*, RT tolerance of 0.1 min, maximum charge of 2, and representative isotope set to most intense), join alignment (tolerance set at 0.5 *m/z*, weight for *m/z* and RT of 50, and RT tolerance of 0.1 min). A filter was applied in order to keep only the features present in at least 3 samples and containing MS/MS spectra. This feature list was exported as a feature quantification table (.csv), and as an MS^2 spectral summary (.mgf).

2.8 Feature-Based Molecular Networking (FBMN)

To investigate the metabolic profile of the dataset, the processed LC-MS/MS data (.mgf and .csv) were used to create a Feature-Based Molecular Network (NOTHIAS *et*

al., 2020) on the GNPS platform (WANG *et al*., 2016). The data was filtered by removing all MS/MS fragment ions within +/- 17 Da of the precursor ion. Additionally, MS/MS spectra were window filtered to select only the top 6 fragment ions in the +/- 50 Da window throughout the spectrum. Both the precursor ion and the MS/MS fragment ion tolerance were set to 0.5 Da. A molecular network was created, in which the edges were filtered to have a cosine score above 0.7 and at least 4 matched peaks. Similarly, the parameters for the library search (for comparison between the experimental and library spectra) were set to have a score above 0.7 and at least 4 matched peaks to assist in the metabolites annotation - level three according to the metabolomics standards initiative (SUMNER *et al*., 2007). The FBMN jobs on GNPS can be found at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=acc800910da64f60ac9910f689fdaceb.

Subsequently, MASST searches were conducted to search for the presence of features 9 and 325 in other public datasets available in the workflow (https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=37cd3d87e49945fd9735ea798592ba8b for feature 9; https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=a631bc8a037a4d828438f1c0459cf3fe for feature 325).

The mass spectrometry data is publicly available on the *Mass spectrometry Interactive Virtual Environment* (MassIVE) as the dataset MSV000089124. The Cytoscape visualization file, the feature table (.csv) and ion information (.mgf) file exported from MZmine2 are also available at this address.

2.9 MetaboAnalyst platform

The quantification table (.csv) file exported from MZmine2 was further modified to include the plant organ relative to each sample. In addition, this table was also split into two other tables containing only the samples relative to each plant organ in order to explore the impact of the collection site. In this way, three separate .csv tables were obtained: one containing all samples and plant organ information, another one relative only to the leaves extracts and collection state information, and another one with only the stem samples and collection state information. Each .csv file was uploaded to the MetaboAnalyst 5.0 platform (www.metaboanalyst.ca) (PANG *et al*., 2021) for statistical

analysis. The data filtering was performed by the median intensity value, the samples were normalized by median, and the data was scaled using Pareto scaling. PCA and PLS-DA analyses were performed, and the important features for discriminating two groups (plant organs or collection states) were identified based on the VIP scores (component 1).

2.10 Fractionation and isolation of Heteropterys umbellata secondary metabolites

The EtOH–H_2O (80:20, v/v) of the leaves extract collected in location 1 was selected for the fractionation and isolation procedure. A total of 2.86 g of crude extract was initially subjected to an initial fractionation on a solid phase extraction column containing C18 as stationary phase. The mobile phases were H_2O (A) and CH_3CN (B), and the eluents were CH_3CN 20%, CH_3CN 80%, and CH_3CN 100%. The fractions were concentrated under reduced pressure and further lyophilized, resulting in 1.52 g, 1.75 g, and 69.6 mg of fractions Fr20, Fr80, and Fr100, respectively.

Fraction Fr80 was selected to perform a subfractionation in preparative HPLC-DAD using a Phenomenex® C18-Luna column (150 × 21.2 mm i.d.; 5 µm), with an acidic (0.1 % formic acid) H_2O (A) and CH_3CN (B) as the mobile phase, gradient from 30 to 65 %B in 20 min, 8 mL min^{-1} flow rate, and the UV absorbance was detected at 220 nm and 254 nm. This step resulted in fractions f1 (0 to 3.0 min), f2 (3.0 to 8.8 min,), f3 (8.8 to 11.5 min), f4 (11.5 to 13.3 min), f5 (13.3 to 14.4 min), f6 (14.4 to 16.0 min), f7 (16.0 to 16.5 min), f8 (16.5 to 20.5 min), and f9 (20.5 to 21.5 min). Fractions f3, f6, and f9 were further purified with a Phenomenex® C18-Luna semi preparative column (250 × 10 mm i.d.; 5 µm), also using acidic H_2O and CH_3CN as mobile phases. Fraction f3 was purified by an isocratic elution (20% CH_3CN in 30 min), leading to the isolation of compound **I** (12.5 mg, Rt of 22 min, yellow oil). An isocratic elution (30% CH_3CN in 25 min) was used to purify fraction f6, affording compound **II** (25.7 mg, Rt of 19 min). Finally, fraction f9 was also purified by an isocratic elution (40% CH_3CN in 30 min), affording compound **III** (16.6 mg, Rt of 25 min). The structures of the isolated compounds were characterized by NMR, ESI-MS/MS, and ESI-HRMS.

2.11 Method validation

2.11.1 Specificity

The specificity consists of the ability of the method to detect and differentiate the signals of the analytes from possible impurities, other compounds present in the natural matrix, among others. In this study, the specificity was determined based on the injection of a blank solution (solvent containing diclofenac as internal standard), and the solution of the standards **1**, **2**, and **3** (n = 3). The relative standard deviation (RSD) was calculated based on each peak retention times in the sample solution **HU1F**. Additionally, the MS^1 and MS^2 mass spectra of each compound was used to confirm the specificity and identity of the peaks.

2.11.2 Precision (repeatability and intermediate precision)

The precision of the method was determined by the analysis of six sample solutions **HU1F**, each one infected one time (n=6), while the repeatability (intraday precision) was estimated as the relative standard deviation (RSD) of the three standards **1**, **2**, and **3** concentrations (μg mL^{-1}) measured in two consecutive days. On the first day, the concentrations measured for **1**, **2**, and **3** were 36.46 (±1.63), 32.75 (±1.45), and 29.08 (±1.02), respectively. On the second day, the concentration values were 39.36 (±1.91), 30.33 (±0.70), and 24.41 (±1.06), respectively. The intermediate precision was measured by the comparison of the results of the two days through an F-test (n = 6–1).

2.11.3 Linearity

The linearity of the method at the MS detector was determined by calibration curves in concentration ranges comprising each compound at the HU1Leaves sample solution. Initially, a stock solution containing **1** (500 μg mL^{-1}), **2** (150 μg mL^{-1}), and **3** (150 μg mL^{-1}) in $CH_3CN–H_2O$ (1:1) was prepared to acquire the analytical curves simultaneously. From this solution, seven concentration levels were prepared for compound **1**, and five concentration levels for **2** and **3**, all in $CH_3CN–H_2O$ (1:1). These levels ranged from 5.0 to 150.0 μg mL^{-1} for **1**, and from 7.5 to 45.0 μg mL^{-1} for both **2**

and **3**. Each concentration level was injected in triplicate. Therefore, the linearity was determined by the analytical curves built based on the nominal concentrations and the average of the rations between the areas of each compound and the internal standard ($R=A_{compound}/A_{IS}$). The correlation coefficients (R) were calculated for each compound.

2.11.4 Limits of detection (LOD) and limits of quantification (LOQ)

The LODs and LOQs were estimated by the mean of the slopes (a) and the standard deviation of the y-intercept (Sb) on three calibration curves in the three lower concentrations for each compound (from 5 to 25 μg mL^{-1} for **1**, and from 7.5 to 22.5 μg mL^{-1} for **2** and **3**). These limits were calculated by the following equations: LOD = $3.3Sb.a^{-1}$ and LOQ = $10Sb.a^{-1}$.

2.11.5 Accuracy

The accuracy of the method was determined by recovery studies, in which known amounts of the isolated standards were spiked to the HU1F solution in three different concentrations (low, intermediate, and high) considering the predetermined linearity concentration range. Three replicates for each level were prepared to obtain solutions containing 25.22, 50.22, and 75.22 μg mL^{-1} of **1**, 11.09, 18.59, and 26.09 μg mL^{-1} for **2**, and 11.89, 19.39, and 26.89 μg mL^{-1} for **3**. The percentage of recovery was analyzed using the validated method and the accuracy was determined by the difference between the nominal and the experimental values.

3 Results and discussion

3.1 Larvicidal screening

Initially, a larvicidal screening was performed with 394 extracts obtained from 134 Malpighiaceae plant species primarily collected in Brazil. Solely three species showed high to moderate activity against *Aedes aegypti* larvae: *Heteropterys umbellata* (EtOH80% and EtOAc extracts, both with 100% mortality in 24 h), *Tetrapterys phlomoides* (EtOAc extract, 0%, 50%, and 70% mortality in 24 h, 48 h, and 72 h,

respectively), and *Verrucularina piresii* (EtOAc extract, 0%, 40%, and 65% mortality in 24 h, 48 h, and 72 h, respectively).

Only one hydroethanolic extract was active, while the others were obtained with ethyl acetate extraction solvent. This result indicates that, possibly, compounds with lower polarity may be more likely to be toxic against these larvae, as previously reported (SILVÉRIO *et al*., 2020). None of the active species (*H. umbellata*, *T. phlomoides*, and *V. piresii*) were studied to date regarding any biological activity or deeper chemical characterization. These species were only used in a large chemotaxonomic study of the Malpighiaceae family recently performed in our laboratory, described in the Chapter 2 of this thesis. Therefore, *Heteropterys umbellata* was chosen for a deeper chemical investigation in order to determine its bioactive compounds.

3.2 Metabolic profiling and multivariate analyses of Heteropterys umbellata extracts

Heteropterys umbellata, the most active species, was selected for a deeper chemical investigation, aiming to explore how the metabolic profile differs based on different plant organs and collection sites, in addition to determining its bioactive compounds. Stem and leaves samples were collected from seven (named HU1 to HU7) collection sites in São Paulo (SP) and Minas Gerais (MG) states, Brazil, resulting in a total of 14 samples. These samples were subjected to the extraction procedure with hydroethanolic solution (extraction yields in Supplementary Table 10). The crude extracts obtained were analyzed by LC-MS/MS in an untargeted method, whose metabolic profile is shown in Figure 47.

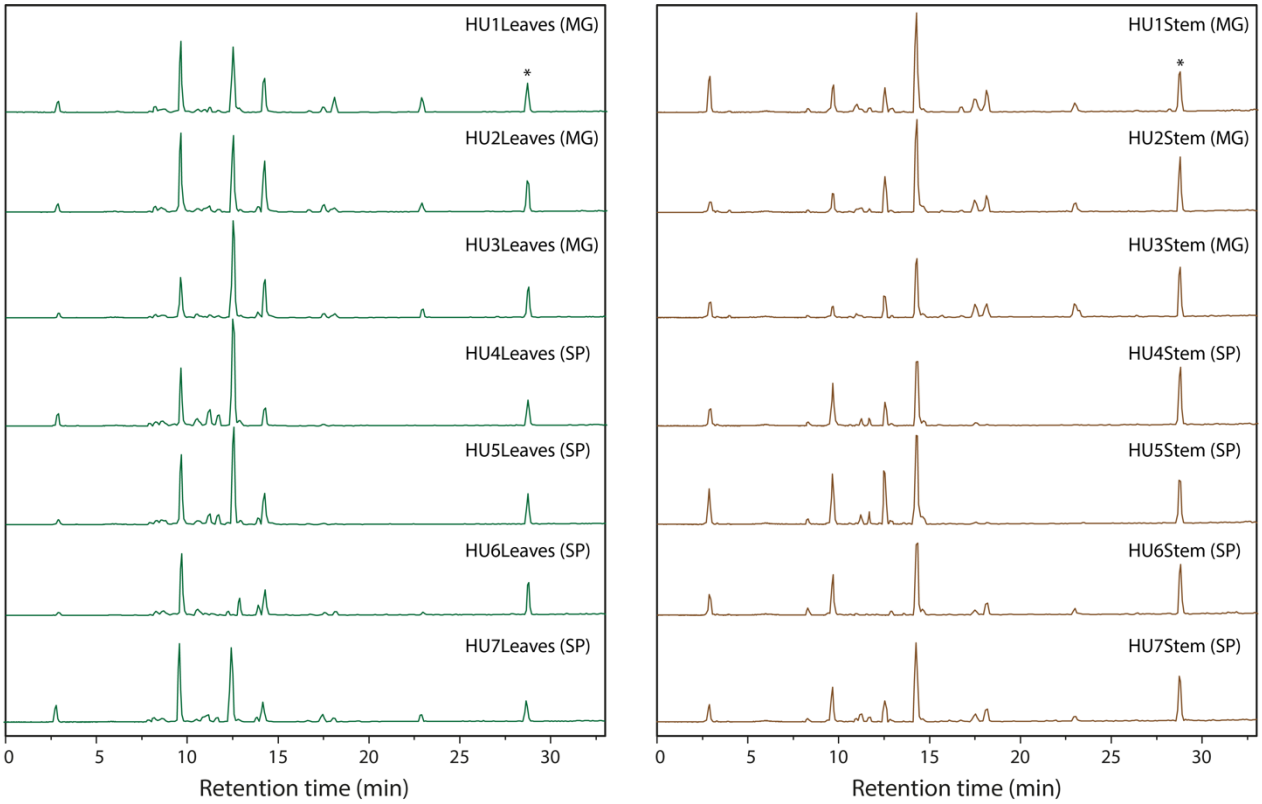


Figure 47 – Base Peak Chromatograms (BPCs) obtained from the *Heteropterys umbellata* crude hydroethanolic extracts. BPCs from the leaves samples are shown on the left (green), and from the stem samples are shown on the right (brown). The collection state of each sample is depicted (MG = Minas Gerais state; SP = São Paulo state). Sodium diclofenac (*) was used as the internal standard.

Source: elaborated by the Author (2022).

A visual comparison between the chromatograms obtained indicates that, in general, all extracts present the same major metabolites, with substantial variations in their relative abundances. In addition, compounds with higher retention times (from 15 to 25 min) presented evident variations, being even absent in some extracts such as HU4 and HU5 samples. Therefore, to explore the chemical differentiations between plant organs and collection sites and have more accurate information regarding these metabolic trends, the LC-MS/MS data were processed and used for multivariate analyses in the Metaboanalyst platform (PANG *et al*., 2021). Multivariate analyses have been successfully used to guide the search for larvicidal compounds (DEMARQUE *et al*., 2020), in addition to exploring tendencies based on plant organs and collection sites (BUENO *et al*., 2021; CIECHOMSKA *et al*., 2019). Therefore, it consists of an invaluable approach to guide the search for bioactive natural

compounds, better understand their possible ecological roles, and investigate site impact.

Principal component analysis (PCA), an unsupervised method, was initially applied to the analyses of all 14 extracts (Figure 48A), showing that, indeed, the metabolites detected in stem and leaves were statistically different. In addition, considering only the leaves and the stem samples, separately, the collection site also impacted the metabolic profiles, causing statistically significant metabolic differences, even though part of these groups overlap in the bidimensional plots (Figure 48D,G). In addition, a supervised method, Partial Least Squares Discriminant Analysis (PLS-DA), was employed to investigate the features relevant for the groups' differentiation (Figure 48B,E,H). The PLS-DA model had acceptable predictive ability under LOOCV cross-validation in all three analyses (Acc. = 1.00, R2 = 0.94, Q2 = 0.91 for plant organ; Acc. = 1.00, R2 = 0.99, Q2 = 0.71 for leaves collection state; and Acc. = 0.90, R2 = 0.83, Q2 = 0.57 for stem collection state).

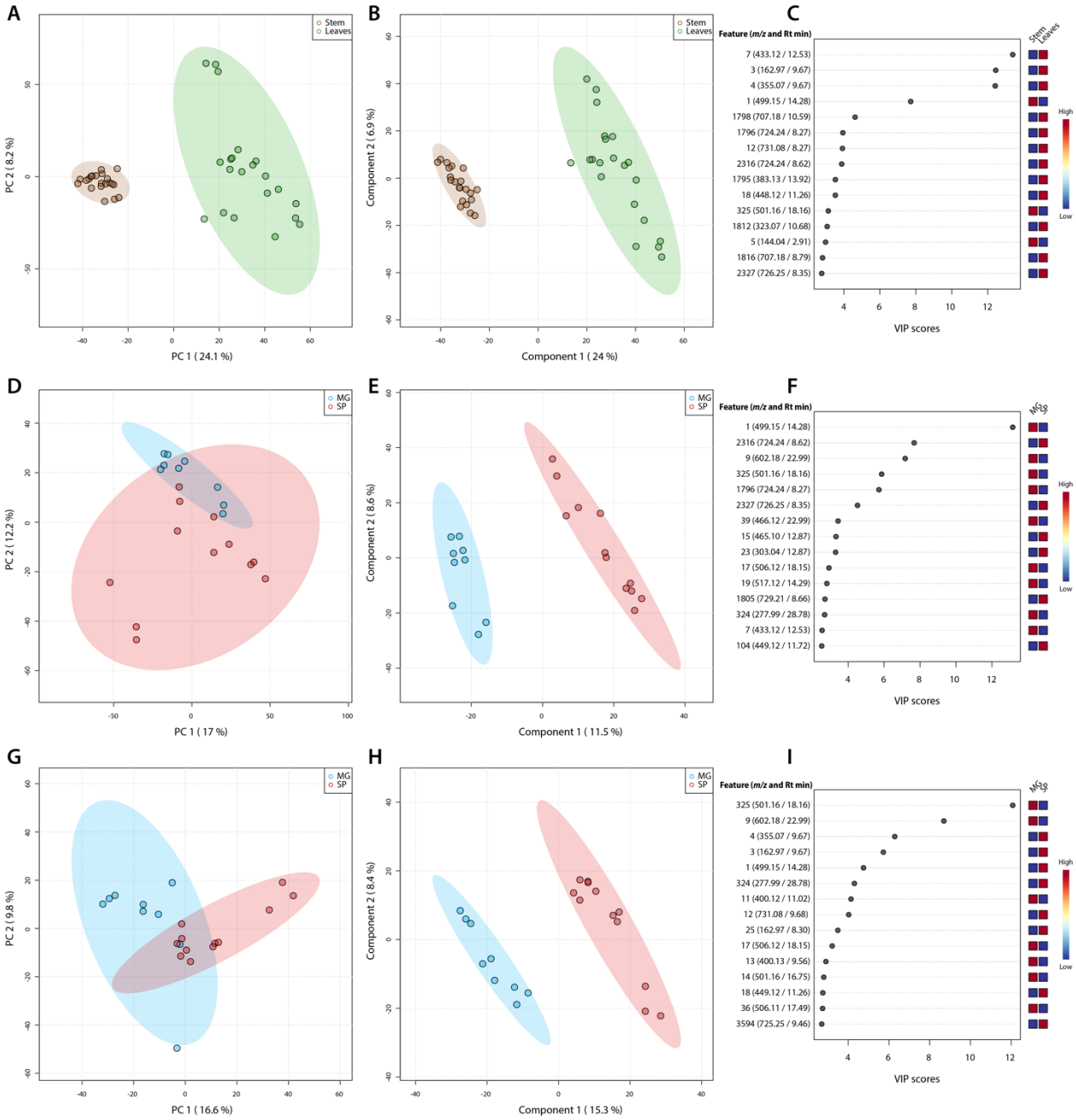


Figure 48 – Statistical analyses for differentiating *H. umbellata* plant organs (**A**–**C**) and collection sites (**D**–**I**). Scores plot obtained for the PCA for plant organs (**A**) and collection sites (leaves: **D**; stem: **G**). Scores plot obtained for the PLS-DA for plant organs (**B**) and collection sites (leaves: **E**; stem: **H**). VIP scores indicating the top 15 metabolites with the highest VIP values (1st component of PLS-DA) for plant organs (**C**), and collection sites (leaves: **F**; stem: **I**).

Source: elaborated by the Author (2022).

The 15 features with the highest Variable in Importance in Projection (VIP) scores along component 1 were evaluated as a measure of the importance of each feature to

discern the plant organs and collection sites (Figure 48C,F,I). In addition, the colored boxes on the right of the VIP scores plot indicate the relative concentrations of the corresponding metabolite in each of the studied groups. For the differentiation of plant organs, features 7 (*m/z* 433.12, Rt 12.53 min), 3 (*m/z* 162.97, Rt 9.67 min), 4 (*m/z* 355.07, Rt 9.67 min), and 1 (*m/z* 499.15, Rt 14.28 min) presented VIP scores higher than 7.0 (Figure 48C). On the leaves subset, the influence of the collection site revealed that features 1, 2316 (*m/z* 724.24, Rt 8.62 min), 9 (*m/z* 602.18, Rt 22.99 min), 325 (*m/z* 501.16, Rt 18.16 min), and 1796 (*m/z* 724.24, Rt 8.27 min) were the top five features for such differentiation (Figure 48F). Finally, on the stem subset, features 325, 9, 4, 3, and 1 were retrieved as the most significant features for collection site distinction (Figure 48I).

In summary, the LC-MS untargeted metabolomics approach used along with multivariate analyses successfully identified tendencies in the relative abundance of specific metabolites in the samples collected. In addition, the possibility to determine the most important features to differentiate groups (by the VIP scores calculation) provides valuable information to guide the study to understand how certain metabolites vary in each group.

3.3 Metabolite annotation

After determining the relevant features in plant organ and collection site differentiations, the processed LC-MS/MS data were used for constructing molecular networks to assist in data visualization, and searched against the GNPS public spectral reference libraries (WANG *et al*., 2016). All the matches were manually inspected to determine their reliability. Three networks stood out for presenting a significant number of features with library matches (Figure 49), being caffeoylquinic acid derivatives, *C*-glycosylated flavonoids, and *O*-glycosylated flavonoids. The annotations of these compounds are described in Supplementary Table 11, and consist of level 2 or 3 annotations according to the Metabolomics Standards Initiative (MSI) (SUMNER *et al*., 2007).

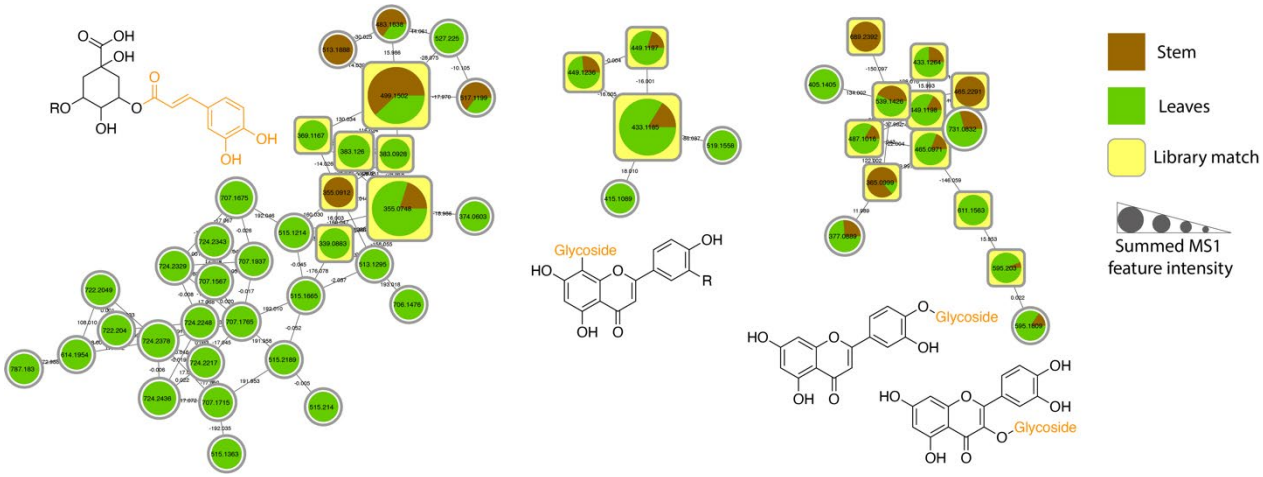


Figure 49 – Molecular families obtained from the Feature-Based Molecular Networking workflow and annotated based on spectral matches within the public spectral libraries in the GNPS platform. Each node represents a tandem mass spectrometry spectrum (MS/MS), while the edges that connect them represent the MS/MS fragmentation similarity (cosine > 0.7). The pie charts indicate the relative abundance of an ion feature in each Heteropterys umbellata plant organ (stem and leaves). Node sizes are relative to the summed peak areas of the precursor ion in MS^1 scans.

Source: elaborated by the Author (2022).

The features highlighted in VIP scores obtained from the PLS-DA analyses presented library matches to *C*-glycosylated flavonoids and caffeoylquinic acids. Feature 7 (*m/z* 433.12 $[M+H]^+$, Rt 12.53 min) was attributed to apigenin-6-*C*-hexoside and was retrieved as in higher concentration in leaves, and in leaves collected in MG state Figure 48C,F. Feature 18 (*m/z* 449.12 $[M+H]^+$, Rt 11.26 min), putatively annotated as luteolin-8-*C*-hexoside, was also recovered as in higher concentration in leaves samples, in addition to higher concentration in stems collected in SP (Figure 48C,I). Another feature also annotated as a *C*-glycosylated flavonoid, feature 104 (*m/z* 449.12 $[M+H]^+$, Rt 11.72 min), showed to be an isomer of feature 18, putatively annotated as luteolin-6-*C*-hexoside. This feature was retrieved only in higher concentrations in leaves collections performed in SP state. All these features presented MS/MS patterns characteristic of *C*-glycosylated flavonoids, with several fragments detected between the protonated compound and the aglycone, such as the consecutive neutral losses of water molecules and the loss of 120 Da.

Two features with high VIP scores in the PLS-DA analyses were annotated as quinic acid derivatives, with the main MS/MS fragments relative to the loss of a caffeoyl

unit, and the fragment at *m/z* 163 representing the loss of the quinic acid moiety. Feature 4 (*m/z* 355.07 $[M+H]^+$, Rt 9.67 min) presented a library match to a caffeoylquinic acid, while feature 3 (*m/z* 162.97, Rt 9.67 min) could be attributed to feature 4 in-source fragmentation due to its MS/MS spectrum and to the same retention time these two features present. Both features 3 and 4 were recovered as in higher concentration in leaves samples, and in stem samples collected in SP. In addition, feature 1 (*m/z* 499.15 $[M–H_2O+H]^+$, Rt 14.28 min) was annotated as a dicaffeoylquinic acid, being retrieved as an important feature in all PLS-DA analyses performed: highly present in stem samples, MG collection sites for leaves, and in SP state for stem samples.

These classes of compounds are widely distributed in plant species, including Malpighiaceae (MANNOCHIO-RUSSO *et al*., 2020; MANNOCHIO RUSSO *et al*., 2020). *O*-glycosylated flavonoids and caffeoylquinic acid derivatives have been described for other *Heteropterys* species, such as *H. tomentosa* (BEZERRA *et al*., 2016; PAULA-FREIRE *et al*., 2013), *H. brachiata* (HUERTA-REYES *et al*., 2013), *H. chrysophylla* (BOBACH *et al*., 2014), and *H. byrsonimifolia* (SANTOS JÚNIOR *et al*., 2014); however, the presence of *C*-glycosylated flavonoids in *Heteropterys* genus has not been reported to date.

The chemical differentiation between different plant organs is expected due to the different ecological roles of each metabolite. Flavonoids are known for being produced in all plant organs, with a series of ecological roles and bioactivities reported in the literature (MIERZIAK; KOSTYN; KULMA, 2014). In *H. umbellata*, the metabolites annotated as apigenin-6-*C*-hexoside and luteolin-8-*C*-hexoside were retrieved as highly accumulated in the leaves, which may indicate a possible role in ultraviolet protection, as reported for other *C*-glycosylated flavonoids (SILVA *et al*., 2015).

Another interesting difference retrieved between leaves and stem samples was the indication that caffeoylquinic acid was present in higher relative concentrations in leaves, while dicaffeoylquinic acid was present in higher amounts in the stem. Both mono and dicaffeoylquinic acids have been attributed to plants' defense against insects, and chlorogenic acid was described as a metabolite involved in abiotic stresses (MONDOLOT *et al*., 2006). Flavonoids and caffeoylquinic acids are biosynthesized by the shikimate pathway, in which the phenylpropanoids are their precursors (FALCONE FERREYRA; RIUS; CASATI, 2012). In *H. umbellata* extracts, our analyses show an inverse correlation between dicaffeoylquinic acid and apigenin-

6-*C*-hexoside contents, and a direct correlation between the mono caffeoylquinic acid and this flavonoid. This may indicate a higher necessity of the phenylpropanoid portion (such as caffeoyl) to produce the flavonoid skeleton in the leaves.

In addition, two other features without any library match presented high VIP scores for collection site differentiation in both plant organs: features 9 and 325. These features presented similar properties such as higher retention times in the chromatographic analyses, similar MS/MS fragmentation patterns (for instance, the initial neutral loss of 136 Da) as confirmed by cosine similarity (0.74), both only ionized in the positive ionization mode, and both presented λ_{max} 220 nm – indicating the absence of aromatic groups. We performed a MASST search of those features, and no datasets were retrieved, indicating that these features are not present in the public datasets included in the search and could be unusual compounds or even new molecules.

3.4 Isolation of the bioactive compounds

Due to the statistical importance of features 9 and 325, the absence of matches within public libraries, and the fact that compounds with larvicidal activity usually present less polar properties, the guided isolation of these compounds was performed. In addition, feature 1 was also aimed due to its statistical importance in the top VIP scores in all PLS-DA analyses performed. Using this metabolomics-guided approach combined with the classical phytochemical methodologies, an initial fractionation was performed in preparative HPLC, yielding nine fractions (f1-f9). All fractions obtained were evaluated for larvicidal activity at 125 μg mL^{-1} (Table 4). Two fractions showed promising activity: fraction 6 presented a mortality rate of 85% in 72 h, while fraction 9 led to 100% mortality in only 24 h.

These fractions were analyzed by LC-MS/MS to explore their chromatographic profiles (Figure 50). The analysis of fraction 6 revealed the presence of three peaks at *m/z* 501 with the same fragmentation pattern, indicating the presence of isomers, in which one of them was the feature 325 retrieved in the PLS-DA analyses. Fraction 9 presented two isomers at *m/z* 602, including feature 9 previously retrieved.

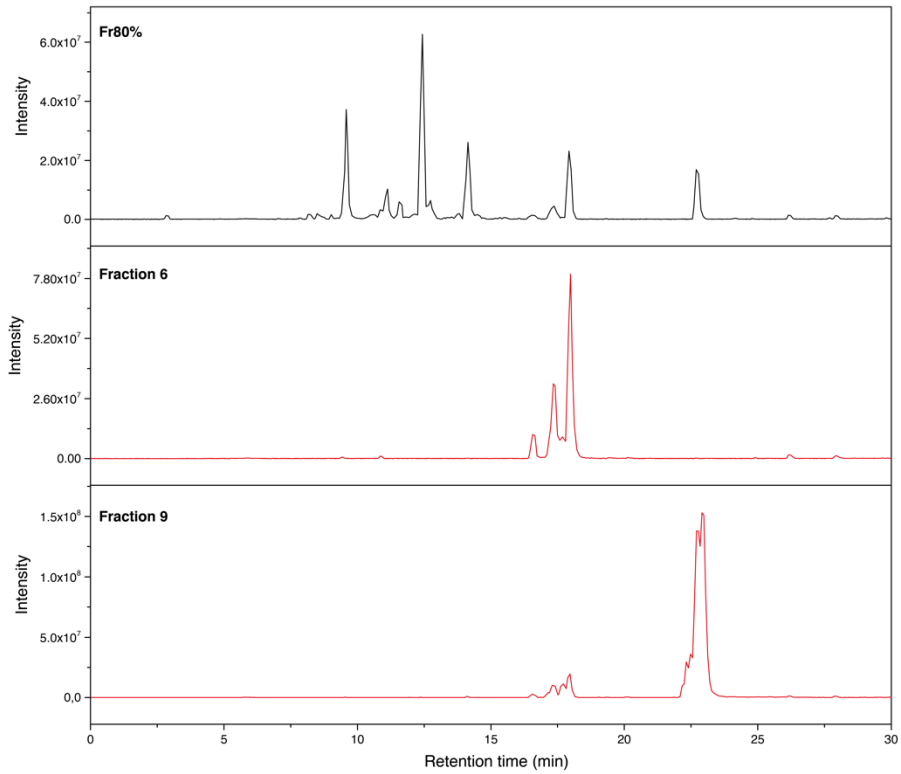


Figure 50 – Base Peak Chromatograms (BPCs) obtained from the *Heteropterys umbellata* Fr80%, fraction 6 and fraction 9. Fraction 6 contains three isomers of *m/z* 501, while fraction 9 presented two isomers of *m/z* 602 and isomers of *m/z* 501 in low amounts.

Source: elaborated by the Author (2022).

Subsequently, compounds **1, 2**, and **3** were isolated by semi-preparative HPLC from fraction 3, fraction 6, and fraction 9, respectively. The structure elucidation of the isolated compounds was performed based on 1D and 2D NMR spectroscopic data (Supplementary Figure 8 to Supplementary Figure 30), HRMS analysis (Supplementary Figure 31 to Supplementary Figure 33), tandem MS analysis (Supplementary Figure 34 to Supplementary Figure 36), and comparison with data on the literature. The compounds were characterized as 3,5-di-*O*-caffeoylquinic acid (isochlorogenic acid A) (**1**) (ALKHATIB *et al*., 2009), 1,2,6-tris-*O*-[3-nitropropanoyl]-beta-glucopyranose (karakin) (**2**) (MAJAK *et al*., 1992), and 1,2,3,6-tetrakis-*O*-[3-nitropropanoyl]-beta-glucopyranose (**3**) (MAJAK; BENN, 1994), as shown in Figure 51.

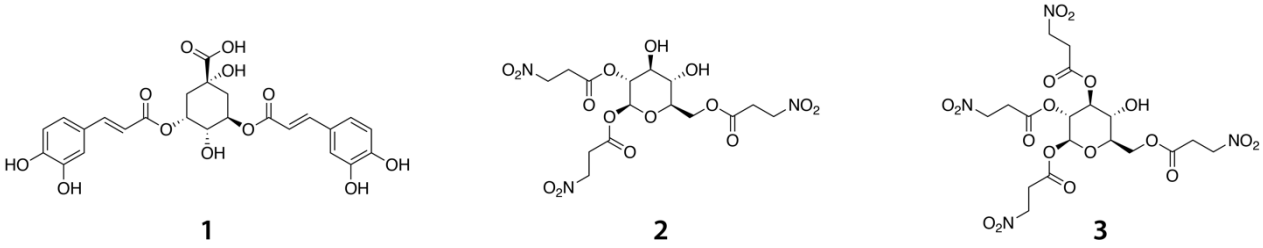


Figure 51 – Compounds isolated from *H. umbellata* leaves extract.

Source: elaborated by the Author (2022).

The ^{1}H and ^{13}C NMR chemical shifts for compounds **1**, **2**, and **3** are described in Supplementary Table 12, Supplementary Table 13, and Supplementary Table 14, while the MS/MS fragmentation proposals are depicted in Figure 52, Figure 53, and Figure 54, respectively. The fragmentation pathways of the nitropropanoyl glycosides (**2** and **3**) consisted primarily of the neutral losses of water (18 Da), HNO_2 (47 Da), and nitropropanoyl substituents (101 Da). Additionally, HRMS analyses were performed to calculate the mass errors, resulting in *m/z* 499.1229 $[M–H_2O+H]^+$ ($C_{25}H_{23}O_{11}$, *m/z* 499.1240, Δ ppm = –2.2), *m/z* 501.1309 $[M+NH_4]^+$ ($C_{15}H_{25}N_4O_{15}$, *m/z* 501.1316, Δ ppm = –1.4), and 602.1413 $[M+NH_4]^+$ ($C_{18}H_{28}N_5O_{18}$, *m/z* 602.1429, Δ ppm = –2.7) for compounds **1**, **2**, and **3**, respectively.

Nitropropanoyl glycosides have already been reported for the Malpighiaceae family and isolated from *Heteropterys aphrodisiaca* (MELO *et al*., 2008; ROMAN JÚNIOR *et al*., 2005), *Heteropterys angustifolia* (STERMITZ *et al*., 1975), and *Hiptage medablota* (GORTER, 1920). Even though these aliphatic nitro compounds have been known for over a century, only a few reports can be found in the literature regarding their isolation and biological activity. They have also been isolated from Fabaceae and Corynocarpaceae, particularly *Astragalus* species (BENN; BAI; MAJAK, 1995; HARLOW; STERMITZ; THOMAS, 1975; NA *et al*., 2015; STERMITZ *et al*., 1972; WILLIAMS; STERMITZ; THOMAS, 1975), *Coronilla varia* (MOYER *et al*., 1977; SIENTZOFF *et al*., 2015), *Corynocarpus laevigatus* (MAJAK; BENN, 1994; MOYER *et al*., 1979), *Hippocrepis comosa* (SALEM *et al*., 1995), *Indigofera* species (FINNEGAN; MUELLER, 1965; MAJAK *et al*., 1992; SU *et al*., 2005) and *Lotus pedunculatus* (GNANASUNDERAM; SUTHERLAND, 1986). In addition, previous reports also indicate their presence in insect (Coleoptera: Chrysomelidae) secretions (SUGENO; MATSUDA, 2002). Their occurrence in nature is considered limited, and two extensive

screenings performed with more than one hundred Argentinean and European plant species were performed decades ago (SALEM *et al*., 1995; STERMITZ *et al*., 1975).

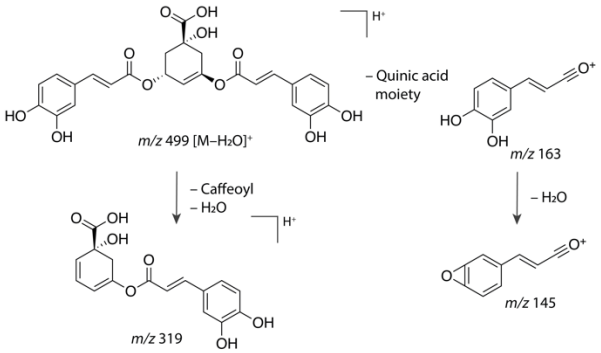


Figure 52 – Proposed fragmentation pathway for compound **1** isolated from *H. umbellata*.

Source: elaborated by the Author (2022).

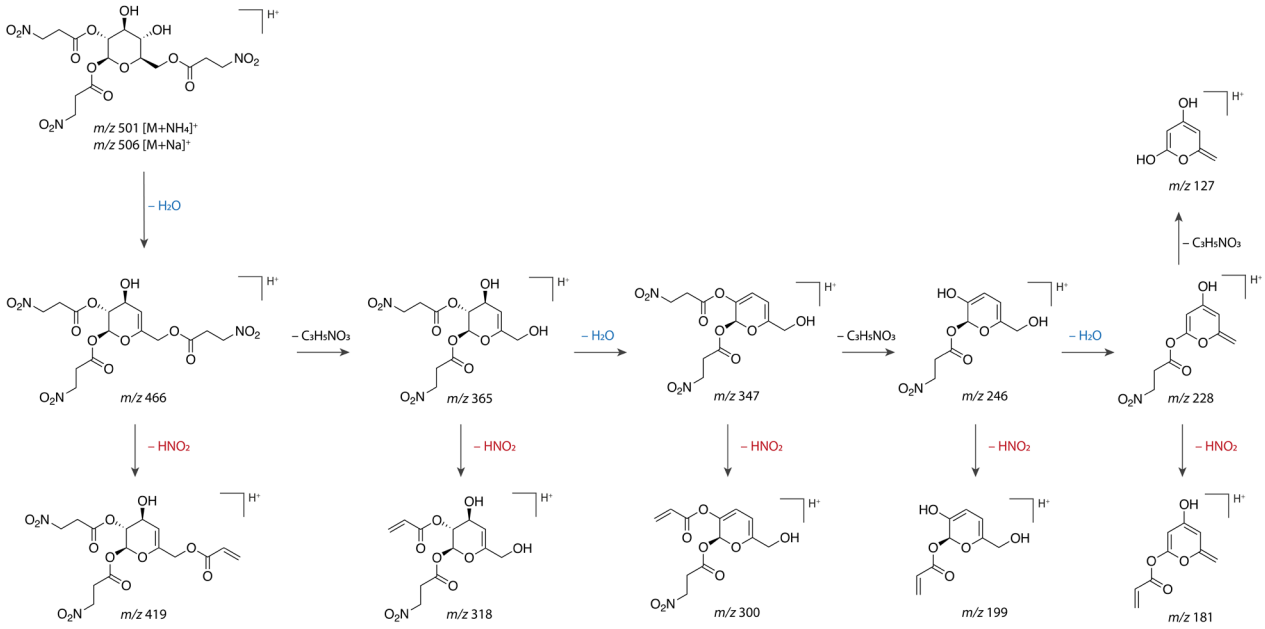


Figure 53 – Proposed fragmentation pathway for compound **2** isolated from *H. umbellata*.

Source: elaborated by the Author (2022).

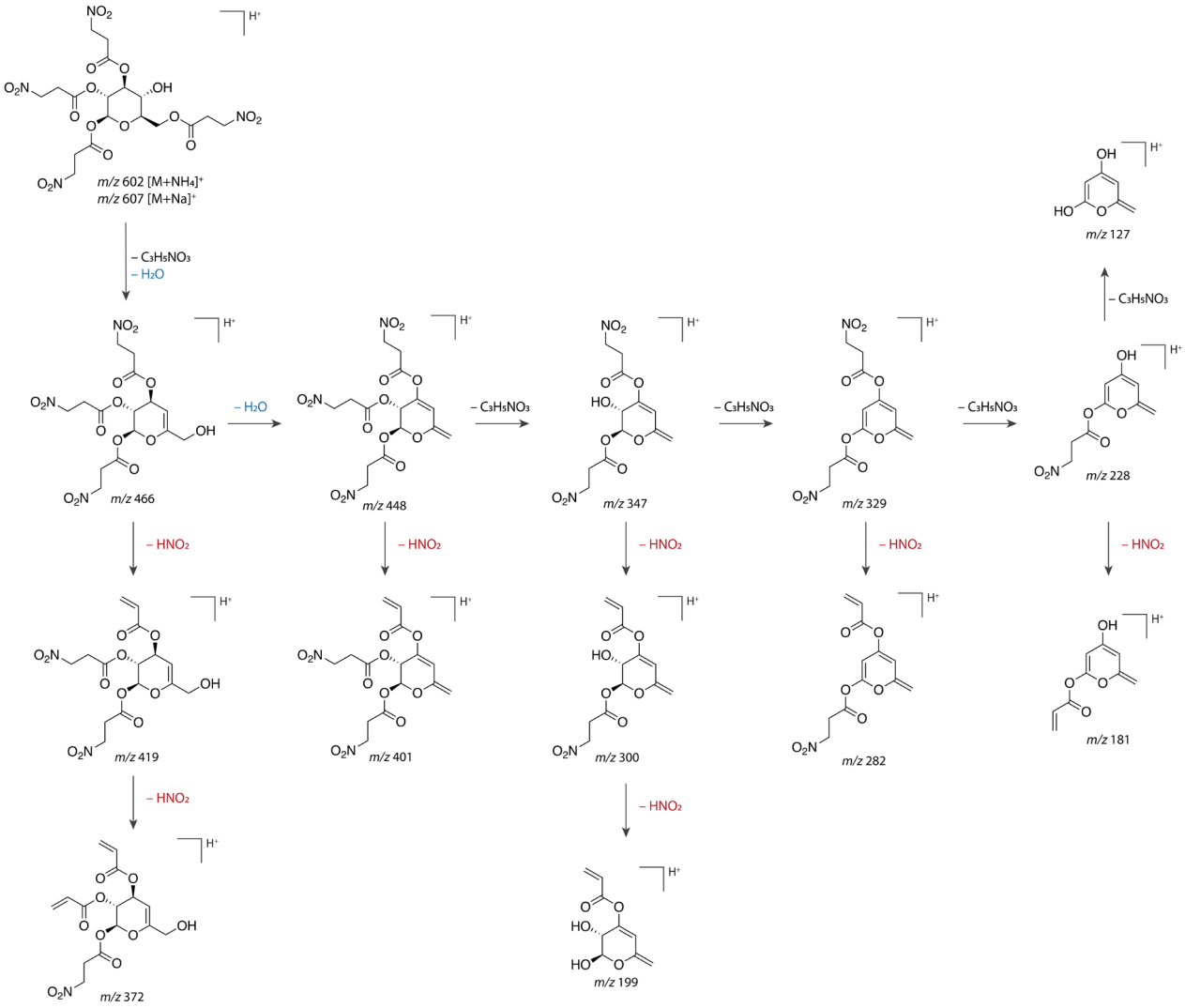


Figure 54 – Proposed fragmentation pathway for compound **3** isolated from *H. umbellata*.

Source: elaborated by the Author (2022).

These glycosidic esters are known for their toxicity, most likely produced as a defense mechanism against insect herbivory (BYERS *et al*., 1986; BYERS; GUSTINE; MOYER, 1977). In addition, studies show that they are usually not toxic to ruminants, being hydrolyzed by their microbiota, and the nitropropionic acid released is metabolized to nontoxic products (GUSTINE *et al*., 1977). Their biosynthetic precursor, 3-nitropropanoate (3-NPA), is also found in fungi (FRANCIS *et al*., 2013; WEI *et al*., 1994), also raising the question of whether the 3-NPA found in plants could actually be produced by their endophytic fungi (CHOMCHEON *et al*., 2005). 3-NPA is known for interfering in ATP synthesis and causing oxidative stress (SHAW; BILLING, 2006).

Pure compounds **2** and **3** were evaluated for larvicidal assay in concentrations ranging from 22.5 to 150 μg mL^{-1} (Table 4), with moderate mortality rates reaching up to 60% and 70% in 72 h, respectively. It was not possible to calculate the EC_{50} values since the mortality of all larvae was not achieved in the higher concentration tested in this protocol. In addition, these pure compounds were also tested in pupes and adult mosquitoes, but did not present toxicity in other stages of *Aedes aegypti* development.

Aliphatic nitro compounds have been reported for their antiviral, antimicrobial, and antifungal activities (MELO *et al*., 2008; ROMAN JÚNIOR *et al*., 2005), and their evaluation against *Aedes aegypti* larvae had not been explored to date. The higher activity of the fractions tested compared to the pure compounds shows that a possible synergistic effect between isomers plays an important role in potentializing the larvicidal activity. In addition, the number of substituents also seems to have a significant impact on the larvicidal activity, since the fraction containing isomers with four propanoyl substituents was more active than the other fraction containing compounds with three substituents. The activity of these compounds only in the larvae show that it is selective to the immature form of *Aedes aegypti*.

Table 4 – Larvicidal results obtained for Fractions 6 and 9, and for the isolated compounds **2** and **3**.

Sample	Concentration (μg mL^{-1})	Results – Mortality rate (%)		
		24 h	48 h	72 h
Fraction 6	125.0	0	2.5	85
Fraction 9	125.0	100	100	100
Compound **2**	90.0	0	0	20
	150.0	0	20	60
Compound **3**	90.0	0	0	40
	150.0	0	10	70

Source: elaborated by the Author (2022).

3.5 Method validation and quantification

In order to more precisely estimate the amount of the isolated compounds in the different extracts obtained for *H. umbellata*, the isolated compounds **1**–**3** were used to validate the LC-MS/MS targeted method in the Multiple Reaction Monitoring (MRM) mode. For this validation, sample HU1F was selected for containing a significant

amount of all three isolated compounds. In this way, it would be possible to determine if the analysis method could provide consistent, reliable, and accurate data. The validation was based on the International Conference on Harmonization (ICH) guidelines (ICH, 2005), in which the specificity, linearity, precision, accuracy, and limits of detection (LOD) and quantification (LOQ) were determined based on the peak areas. The method employed reached the acceptance criteria specified for each parameter, as described in Table 5.

Table 5 – Method validation data of compounds **1**, **2**, and **3** for the *H. umbellata* analyses (according to the ICH guidelines).

Validation item	Parameter	Compound			Acceptance criteria
		1	2	3	
Linearity (n = 3)	Linear dynamic range (μg mL^{-1})	5.0 to 150.0	7.5 to 45	7.5 to 45	Informative
	Equation	y = 0.0185x + 0.0957	y = 0.0193 – 0.0798	y = 0.0165 – 0.1128	Informative
	R	0.9972	0.9966	0.9983	>0.9900
Precision (n = 6)	Repeatability: day 1	4.48	4.41	3.50	RSD < 5%
	Repeatability: day 2	4.87	2.31	4.35	RSD < 5%
	Intermediate precision (F = 0.05; n = 6–1)	0.73	4.26	0.92	$F_{calc} < F_{tabl}$
Accuracy (n = 3)	Low level average	117.61	119.80	118.80	80–120%
	Intermediate level average	118.80	117.21	112.97	80–120%
	High level average	118.58	108.08	105.68	80–120%
Limit of Detection and Limit of Quantification (n = 3)	Slope average (±SD)	0.0257 (±0.0006)	0.0231 (±0.0005)	0.0171 (±0.0002)	
	y-intercept average (±SD)	-0.0288 (±0.0051)	-0.1306 (±0.0025)	-0.1194 (±0.0034)	
	LOD (μg mL^{-1})	0.6561	0.3618	0.6638	Informative
	LOQ (μg mL^{-1})	1.9883	1.0963	2.0114	Informative

Source: elaborated by the Author (2022).

All 14 extracts were analyzed in triplicate in this validated targeted method, and the quantification of compounds **1**–**3** in these extracts was performed (Figure 55, Supplementary Table 15). These results confirmed the PLS-DA analyses performed initially. Compound **1** was retrieved in higher amounts in stem samples, while compounds **2** and **3** were not statistically different comparing the two plant organs – even though these compounds were below the LOD for samples HU4F and HU5F (Figure 55A). For the leaves subset, compounds **1**–**3** were in higher concentration in all collections performed in MG state (Figure 55B). Finally, in the stem subset,

compound **1** was present in higher amounts in collections from SP state, while compounds **2** and **3**, similarly to the leaves subset, were also present in higher amounts MG (Figure 55C).

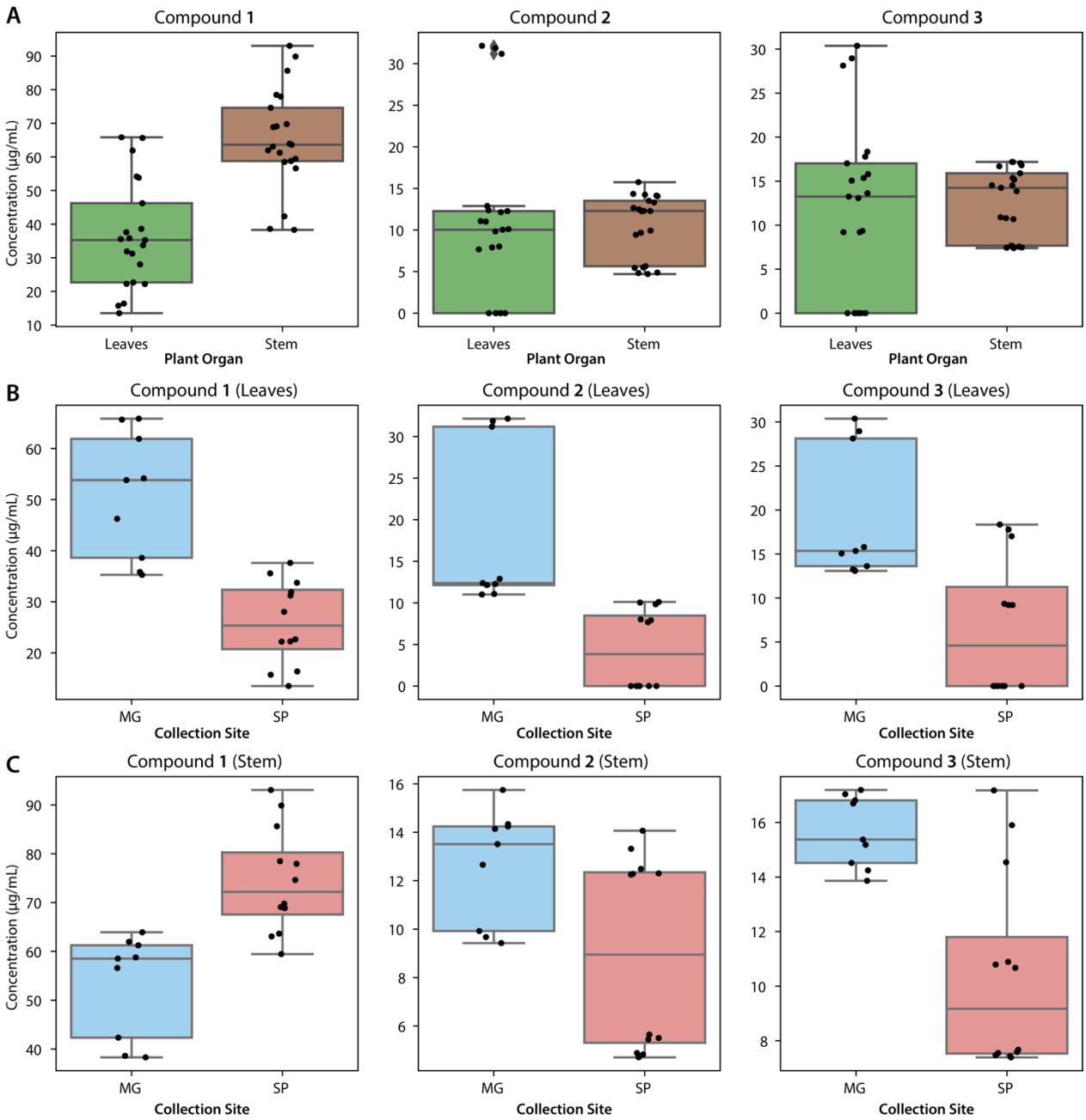


Figure 55 – Boxplots of the concentration levels of compounds **1, 2**, and **3** obtained by Multiple Reaction Monitoring (MRM) mode by plant organ (**A**), and by collection site for leaves (**B**) and stem (**C**). Black dots represent all individual sample concentrations for each group.

Source: elaborated by the Author (2022).

The validation of the LC-MS/MS method allowed us to determine the concentration of the isolated compounds in different groups, and proved to be sensitive since low LOD and LOQ values were obtained. A previous study describes the development of an HPLC method to indirectly quantify nitropropanoyl esters of D-glucopyranose as 3-NPA in karaka nuts (3-NPA LOQ of 2.98 μg mL^{-1}) (MACASKILL; MANLEY-HARRIS; FIELD, 2015); however, the direct quantification of these esters has not yet been reported to date. By comparing both methods, the one described in this study proved to be slightly more sensitive regarding direct detection and quantification of these esters.

A significantly higher amount of these nitropropanoyl glycosides (compounds **2** and **3**) was observed in MG collections compared to SP for both plant organs. Previous reports have correlated a lower production of 3-NPA to lower temperatures throughout the year in *Coronilla varia* (BYERS *et al*., 1986; FAIX; GUSTINE; WRIGHT, 1978). We evaluated the average temperature in a 30-day period prior to the collection of the samples, and we observed that the SP locality was about 2 ˚C hotter than MG localities (28.6 ˚C and 31.0 ˚C in MG and SP, respectively). Therefore, we believe that the higher concentrations of these esters in MG collection points are due to the MG specimens being located in preserved areas, being more susceptible to herbivory, which may impact the production of compounds **2** and **3**. Further seasonality studies should be performed to understand their production in *H. umbellata.*

4 Conclusions

The present study demonstrated that the Malpighiaceae family represents a promising taxon to further explore larvicidal compounds against the vectors of arboviruses. The most active species in our unique collection of extracts was our aim to search for bioactive compounds, and the metabolomic tools used were of great value to guide our search to isolate them. In addition, the quantification of the isolated compounds also corroborated the results we obtained in the multivariate analyses. The nitropropanoyl glycosides isolated are considered unusual, with limited distribution in nature, and we showed strong evidence of synergistic effects that potentialize the larvicidal results. Our study provided the first chemical exploration of *Heteropterys umbellata*, and we expect that our study can guide future research in the *Heteropterys*

genus, in addition to the search for bioactive compounds using a metabolomic-based approach.

3 FINAL CONSIDERATIONS: DISCUSSION AND CONCLUSION

The chemistry of natural products has been extensively investigated for decades, and plant secondary metabolites are among the most widely explored. Even though thousands of secondary metabolites have been described for plants, there is still much to be explored, and the sensitive hyphenated techniques and computational tools have proven to be of great value for more comprehensive studies. In fact, studies at a botanical family level are becoming possible, and chemotaxonomic investigations can benefit from these novel approaches.

This study took the Malpighiaceae botanical family as an example of a broad chemical investigation. The investigation began with the development of an analytical method, moved on to chemosystematic studies, and finally, to the search for bioactive compounds. All these investigations included the use of cutting-edge computational tools to guide the studies and answer our initial hypotheses.

The robust chromatographic method developed can be of great use in future studies in Malpighiaceae, and the fast workflow followed to get to these results can inspire future studies to save time and resources in natural products research. One more time, the Quality by Design approach proved to be effective for the analysis of extremely complex samples, such as mixtures of extracts. The use of the molecular networking algorithm and library searches in GNPS allowed us to have a first contact with the tool, and to realize the potential of its use in chemotaxonomy if all the 197 samples were analyzed.

The chemotaxonomic investigations performed in this study made use of cutting-edge computational tools to study Malpighiaceae at a family level. Most of our samples were retrieved from discarded fragmented samples used for DNA extraction in previous molecular studies, and the small amount of sample necessary for our study allowed us to use this material for further chemical analyses. The tools used for MS/MS data processing and analyses expanded the chemical information of Malpighiaceae in the literature, and all the MS/MS datasets generated in this study are publicly available. In fact, the use of open-access software and databases is another plus since our methodology can be reproduced with other taxa (at a family, clade, or genus level).

Correlating the *in silico* classifications with the phylogeny currently accepted for Malpighiaceae enabled a greater evolutionary understanding of the family. The results obtained can guide future studies in Malpighiaceae since we provided evidence of

chemical classes that can be enriched in particular clades. For instance, if future studies want to search for iridoids or benzylisoquinoline alkaloids in Malpighiaceae, they should probably focus on species from clade J (particularly, *Stigmaphyllon* genus). On the other hand, ecdysteroids should be explored in clade H (especially *Niedenzuella* genus), and so on. Further studies should be performed in the following decades to confirm the results obtained with our approach, and the classical methodologies are of great value for such. However, relatively quickly, we were able to carry out this study and propose several chemical classes.

The search for bioactive compounds was also accelerated by using metabolomics tools. The initial bioactivity screening allowed us to select the species with higher larvicidal activity against *Aedes aegypti*, and *Heteropterys umbellata* became the focus of our study. The metabolomic approach effectively directed the isolation of unusual natural nitro compounds, making it possible to infer possible synergistic effects in larvicidal activity and raise hypotheses regarding their ecological roles.

In conclusion, the combination of metabolomic strategies with state-of-the-art analytical techniques, along with cutting-edge computational tools for metabolites annotation, were crucial to leverage Malpighiaceae chemical space and determine a few bioactive compounds. The collaborative research among several research groups enabled this study to be conducted interdisciplinary, making it possible to withdraw more comprehensive conclusions.

REFERENCES

AABY, K.; EKEBERG, D.; SKREDE, G. Characterization of phenolic compounds in strawberry (*Fragaria* × *ananassa*) fruits by different HPLC detectors and contribution of individual compounds to total antioxidant capacity. **Journal of Agricultural and Food Chemistry**, v. 55, n. 11, p. 4395–4406, 2007.

ABBET, C.; SLACANIN, I.; HAMBURGER, M.; POTTERAT, O. Comprehensive analysis of *Phyteuma orbiculare* L., a wild Alpine food plant. **Food Chemistry**, v. 136, n. 2, p. 595–603, 2013.

ABD EL-LATIF, R. R.; MANSOUR, R. M. A.; SHARAF, M.; FARAG, A. Three new flavonol glycosides from *Suaeda maritima*. **Journal of Asian Natural Products Research**, v. 16, n. 5, p. 434–439, 2014.

AGUIAR, R. M.; DAVID, J. P.; DAVID, J. M. Unusual naphthoquinones, catechin and triterpene from *Byrsonima microphylla*. **Phytochemistry**, v. 66, n. 19, p. 2388–2392, 2005.

AKSENOV, A. A.; DA SILVA, R.; KNIGHT, R.; LOPES, N. P.; DORRESTEIN, P. C. Global chemical analysis of biology by mass spectrometry. **Nature Reviews Chemistry**, v. 1, n. 7, p. 0054, 2017.

ALKHATIB, R.; HENNEBELLE, T.; ROUMY, V.; SAHPAZ, S.; SÜZGEÇ, S.; AKALIN, E.; MERIÇLI, A. H.; BAILLEUL, F. Coumarins, caffeoyl derivatives and a monoterpenoid glycoside from *Ferulago asparagifolia*. **Biochemical Systematics and Ecology**, v. 37, n. 3, p. 230–233, 2009.

ALMEIDA, R. F. .; FRANCENER, A. .; PESSOA, C. .; SEBASTIANI, R. .; OLIVEIRA, Y. R. .; AMORIM, A. M. A. .; MAMEDE, M. C. H. **Malpighiaceae in Flora do Brasil 2020 (em construção).** Jardim Botânico do Rio de Janeiro, 2021.

ALMEIDA, R. F.; AMORIM, A. M. A.; VAN DEN BERG, C. Timing the origin and past connections between Andean and Atlantic Seasonally Dry Tropical Forests in South America: Insights from the biogeographical history of Amorimia (Malpighiaceae). **Taxon**, v. 67, n. 4, p. 739–751, 2018.

ALMEIDA, R. F.; AMORIM, A. M.; CORRÊA, A. M. da S.; VAN DEN BERG, C. A new infrageneric classification for Amorimia (Malpighiaceae) based on morphological, phytochemical and molecular evidence. **Phytotaxa**, v. 313, n. 3, p. 231, 2017.

ALMEIDA, R. F.; VAN DEN BERG, C. Biogeography of Stigmaphyllon (Malpighiaceae) and a Meta-Analysis of Vascular Plant Lineages Diversified in the Brazilian Atlantic Rainforests Point to the Late Eocene Origins of This Megadiverse Biome. **Plants**, v. 9, n. 11, p. 1569, 2020.

ALMEIDA, R. F.; VAN DEN BERG, C. Molecular phylogeny and character mapping support generic adjustments in the Tetrapteroid clade (Malpighiaceae). **Nordic**

Journal of Botany, v. 39, n. 1, p. e02876, 2021.

ANDERSON, C. Revision of *Ryssopterys* and transfer to *Stigmaphyllon* (Malpighiaceae). **Blumea - Biodiversity, Evolution and Biogeography of Plants**, v. 56, n. 1, p. 73–104, 2011.

ANDERSON, M. J. A new method for non-parametric multivariate analysis of variance. **Austral Ecology**, v. 26, n. 1, p. 32–46, 2001.

ANDERSON, W. R. The origin of the Malpighiaceae - the evidence from morphology. **Memoirs of the New York Botanical Garden**, v. 64, p. 210–224, 1990.

ANDERSON, W. R. Malpighiaceae (Malpighia Family). In: SMITH, N.; MORI, S. A.; HENDERSON, A.; STEVENSON, D. W.; HEALD, S. V. (Eds.). **Flowering plants of the neotropics**. [*S.l.*] : Princeton University Press; New York Botanical Garden, 2004. p. 229–232.

ANDERSON, W. R. Eight segregates from the neotropical genus *Mascagnia* (Malpighiaceae). **Novon**, v. 16, n. 2, p. 168–204, 2006.

ANDERSON, W. R.; DAVIS, C. C. Expansion of *Diplopterys* at the expense of *Banisteriopsis* (Malpighiaceae). **Harvard Papers in Botany**, v. 11, n. 1, p. 1–16, 2006.

ATANASOV, A. G.; ZOTCHEV, S. B.; DIRSCH, V. M.; SUPURAN, C. T. Natural products in drug discovery: advances and opportunities. **Nature Reviews Drug Discovery**, 2021.

BAUERMEISTER, A.; MANNOCHIO-RUSSO, H.; COSTA-LOTUFO, L. V.; JARMUSCH, A. K.; DORRESTEIN, P. C. Mass spectrometry-based metabolomics in microbiome investigations. **Nature Reviews Microbiology**, v. 20, n. 3, p. 143–160, 2022.

BELWAL, T.; DEVKOTA, H. P.; HASSAN, H. A.; AHLUWALIA, S.; RAMADAN, M. F.; MOCAN, A.; ATANASOV, A. G. Phytopharmacology of Acerola (*Malpighia* spp.) and its potential as functional food. **Trends in Food Science & Technology**, v. 74, p. 99–106, 2018.

BENELLI, G.; JEFFRIES, C.; WALKER, T. Biological control of mosquito vectors: past, present, and future. **Insects**, v. 7, n. 4, p. 52, 2016.

BENN, M.; BAI, Y.; MAJAK, W. Aliphatic nitro-compounds in *Astragalus canadensis*. **Phytochemistry**, [s. l.], v. 40, n. 6, p. 1629–1631, 1995.

BEZERRA, A. G.; NEGRI, G.; DUARTE-ALMEIDA, J. M.; SMAILI, S. S.; CARLINI, E. A. Phytochemical analysis of hydroethanolic extracts from powdered roots of *Panax ginseng* C. A. Meyer and *Heteropterys tomentosa* A. Juss and evaluation of their effects on astricyte cell death. **Química Nova**, v. 39, n. 5, p. 581–587, 2016.

BHARGAVA, V. V.; PATEL, S. C.; DESAI, K. S. Importance of Terpenoids and Essential Oils in Chemotaxonomic Approach. **International Journal of Herbal Medicine**, v. 1, n. 2, p. 14–21, 2013.

BOBACH, C.; SCHURWANZ, J.; FRANKE, K.; DENKERT, A.; SUNG, T. Van; KUSTER, R.; MUTISO, P. C.; SELIGER, B.; WESSJOHANN, L. A. Multiple readout assay for hormonal (androgenic and antiandrogenic) and cytotoxic activity of plant and fungal extracts based on differential prostate cancer cell line behavior. **Journal of Ethnopharmacology**, v. 155, n. 1, p. 721–730, 2014.

BOLYEN, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. **Nature Biotechnology**, v. 37, n. 8, p. 852–857, 2019.

BOŽIČEVIĆ, A.; MIERI, M.; NASSENSTEIN, C.; WIEGAND, S.; HAMBURGER, M. Secondary Metabolites in Allergic Plant Pollen Samples Modulate Afferent Neurons and Murine Tracheal Rings. **Journal of Natural Products**, v. 80, n. 11, p. 2953–2961, 2017.

BUENO, P. C. P.; ABARCA, L. F. S.; ANHESINE, N. B.; GIFFONI, M. S.; PEREIRA, F. M. V.; TORRES, R. B.; DE, R. W. R.; FERREIRA, P. M. P.; PESSOA, C.; CAVALHEIRO, A. J. Infraspecific chemical variability and biological activity of *Casearia sylvestris* from different Brazilian biomes. **Planta Medica**, v. 87, n. 01/02, p. 148–159, 2021.

BUENO, P. C. P.; PEREIRA, F. M. V.; TORRES, R. B.; CAVALHEIRO, A. J. Development of a comprehensive method for analyzing clerodane-type diterpenes and phenolic compounds from *Casearia sylvestris* Swartz (Salicaceae) based on ultra high performance liquid chromatography combined with chemometric tools. **Journal of Separation Science**, v. 38, n. 10, p. 1649–1656, 2015.

BYERS, R. A.; GUSTINE, D. L.; MOYER, B. G. Toxicity of β-Nitropropionic Acid to *Trichoplusia ni*. **Environmental Entomology**, v. 6, n. 2, p. 229–232, 1977.

BYERS, R. A.; GUSTINE, D. L.; MOYER, B. G.; BIERLEIN, D. L. 3-Nitropropionate in Crownvetch: A Natural Deterrent to Insects? In: GREEN, M. B.; HEDIN, P. A. (Eds.). *In:* **Natural Resistance of Plants to Pests**. [*S.l.*] : American Chemical Society, 1986. p. 95–105.

CAI, L.; XI, Z.; PETERSON, K.; RUSHWORTH, C.; BEAULIEU, J.; DAVIS, C. C. Phylogeny of Elatinaceae and the Tropical Gondwanan Origin of the Centroplacaceae (Malpighiaceae, Elatinaceae) Clade. **PLOS ONE**, v. 11, n. 9, p. e0161881, 2016.

CAMERON, K. M.; CHASE, M. W.; ANDERSON, W. R.; HILLS, H. G. Molecular systematics of Malpighiaceae: evidence from plastid *rbcL* and *matK* sequences. **American Journal of Botany**, v. 88, n. 10, p. 1847–1862, 2001.

CANTRELL, K.; FEDARKO, M. W.; RAHMAN, G.; MCDONALD, D.; YANG, Y.; ZAW, T.; GONZALEZ, A.; JANSSEN, S.; ESTAKI, M.; HAIMINEN, N.; BECK, K. L.; ZHU,

Q.; SAYYARI, E.; MORTON, J. T.; ARMSTRONG, G.; TRIPATHI, A.; GAUGLITZ, J. M.; MAROTZ, C.; MATTESON, N. L.; MARTINO, C.; SANDERS, J. G.; CARRIERI, A. P.; SONG, S. J.; SWAFFORD, A. D.; DORRESTEIN, P. C.; ANDERSEN, K. G.; PARIDA, L.; KIM, H.-C.; VÁZQUEZ-BAEZA, Y.; KNIGHT, R. EMPress enables tree-guided, interactive, and exploratory analyses of multi-omic data sets. **mSystems**, v. 6, n. 2, p. 1–10, 2021.

CAO, S.; GUZA, R. C.; MILLER, J. S.; ANDRIANTSIFERANA, R.; RASAMISON, V. E.; KINGSTON, D. G. I. Cytotoxic triterpenoids from *Acridocarpus vivy* from the Madagascar Rain Forest. **Journal of Natural Products**, v. 67, n. 6, p. 986–989, 2004.

CARDOSO TAKETA, A. T.; LOZADA-LECHUGA, J.; FRAGOSO-SERRANO, M.; VILLARREAL, M. L.; PEREDA-MIRANDA, R. Isolation of Nor-secofriedelanes from the Sedative Extracts of *Galphimia glauca*. **Journal of Natural Products**, v. 67, n. 4, p. 644–649, 2004.

CARVALHO, G. D.; NUNES, L. C.; BRAGANÇA, H. B. N.; PORFÍRIO, L. C. Principais plantas tóxicas causadoras de morte súbita em bovinos no estado do Espírito Santo - Brasil. **Archivos de Zootecnia**, v. 58, n. 224, p. 87–98, 2008.

CHEN, Y.; MULDER, P. P. J.; SCHAAP, O.; MEMELINK, J.; KLINKHAMER, P. G. L.; VRIELING, K. The evolution of pyrrolizidine alkaloid diversity among and within *Jacobaea* species. **Journal of Systematics and Evolution**, v. 60, n. 2, p. 361–376, 2020.

CHOMCHEON, P.; WIYAKRUTTA, S.; SRIUBOLMAS, N.; NGAMROJANAVANICH, N.; ISARANGKUL, D.; KITTAKOOP, P. 3-Nitropropionic Acid (3-NPA), a potent antimycobacterial agent from endophytic fungi: is 3-NPA in some plants Produced by endophytes? **Journal of Natural Products**, v. 68, n. 7, p. 1103–1105, 2005.

CIECHOMSKA, M.; WOŹNIAKIEWICZ, M.; MACHLOWSKA, K.; KLEPACKI, P.; KOŚCIELNIAK, P. Differentiation of Solanaceae psychoactive plants based on GC-MS analysis supported by chemometric tools. **Microchemical Journal**, v. 150, p. 104098, Apr., 2019.

CORRADI, E.; DE MIERI, M.; CADISCH, L.; ABBET, C.; HAMBURGER, M.; POTTERAT, O. New acylated flavonol glycosides and a phenolic profile of *Pritzelago alpina*, a forgotten edible Alpine plant. **Chemistry & Biodiversity**, v. 13, n. 2, p. 188–197, 2016.

CREYDT, M.; FISCHER, M. Plant metabolomics: maximizing metabolome coverage by optimizing mobile phase additives for nontargeted mass spectrometry in positive and negative electrospray ionization mode. **Analytical Chemistry**, v. 89, n. 19, p. 10474–10486, 2017.

CRÜSEMANN, M.; O'NEILL, E. C.; LARSON, C. B.; MELNIK, A. V.; FLOROS, D. J.; DA SILVA, R. R.; JENSEN, P. R.; DORRESTEIN, P. C.; MOORE, B. S. Prioritizing natural product diversity in a collection of 146 bacterial strains based on growth and

extraction protocols. **Journal of Natural Products**, v. 80, n. 3, p. 588–597, 2017.

D'ATRI, V.; FEKETE, S.; CLARKE, A.; VEUTHEY, J.-L.; GUILLARME, D. Recent advances in chromatography for pharmaceutical analysis. **Analytical Chemistry**, v. 91, n. 1, p. 210–239, 2019.

DAVIOUD, E.; BAILLEUL, F.; DELAVEAU, P.; JACQUEMIN, H. Iridoids of Guyanese species of *Stigmaphyllon*. **Planta Medica**, v. 51, n. 01, p. 78–78, 1985.

DAVIS, C. C.; ANDERSON, W. R. A complete generic phylogeny of Malpighiaceae inferred from nucleotide sequence data and morphology. **American Journal of Botany**, v. 97, n. 12, p. 2031–2048, 2010.

DAVIS, C. C.; ANDERSON, W. R.; DONOGHUE, M. J. Phylogeny of Malpighiaceae: evidence from chloroplast *ndhF* and *trnl-F* nucleotide sequences. **American Journal of Botany**, v. 88, n. 10, p. 1830–1846, 2001.

DAVIS, C. C.; BELL, C. D.; MATHEWS, S.; DONOGHUE, M. J. Laurasian migration explains Gondwanan disjunctions: Evidence from Malpighiaceae. **Proceedings of the National Academy of Sciences**, v. 99, n. 10, p. 6833–6837, 2002.

DAVIS, C. C.; MARINHO, L. C.; AMORIM, A. M. *Andersoniodoxa*, a replacement name for *Andersoniella* (Malpighiaceae). **Phytotaxa**, v. 470, n. 1, p. 121–122, 2020.

DAVIS, C. C.; SCHAEFER, H.; XI, Z.; BAUM, D. A.; DONOGHUE, M. J.; HARMON, L. J. Long-term morphological stasis maintained by a plant–pollinator mutualism. **Proceedings of the National Academy of Sciences**, v. 111, n. 16, p. 5914–5919, 2014.

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

DEAN, D. A.; GAUTHAM, G.; SIQUEIRA-NETO, J. L.; MCKERROW, J. H.; DORRESTEIN, P. C.; MCCALL, L.-I. Spatial metabolomics identifies localized chemical changes in heart tissue during chronic cardiac Chagas Disease. **PLOS Neglected Tropical Diseases**, v. 15, n. 10, p. e0009819, 2021.

DEMARQUE, D. P.; DUSI, R. G.; 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. **Scientific Reports**, v. 10, n. 1, p. 1051, 2020.

DEMARQUE, D. P.; ESPINDOLA, L. S. Challenges, advances and opportunities in exploring natural products to control arboviral disease vectors. **Frontiers in Chemistry**, v. 9, Nov, 2021.

DISPAS, A.; LEBRUN, P.; ANDRI, B.; ROZET, E.; HUBERT, P. Robust method optimization strategy—A useful tool for method transfer: The case of SFC. **Journal**

of Pharmaceutical and Biomedical Analysis, v. 88, p. 519–524, 2014.

DÜHRKOP, K.; FLEISCHAUER, M.; LUDWIG, M.; AKSENOV, A. A.; MELNIK, A. V.; MEUSEL, M.; DORRESTEIN, P. C.; ROUSU, J.; BÖCKER, S. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. **Nature Methods**, v. 16, n. 4, p. 299–302, 2019.

DÜHRKOP, K.; NOTHIAS, L.-F.; FLEISCHAUER, M.; REHER, R.; LUDWIG, M.; HOFFMANN, M. A.; PETRAS, D.; GERWICK, W. H.; ROUSU, J.; DORRESTEIN, P. C.; BÖCKER, S. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. **Nature Biotechnology**, v. 39, n. 4, p. 462–471, 2021.

DÜHRKOP, K.; SHEN, H.; MEUSEL, M.; ROUSU, J.; BÖCKER, S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. **Proceedings of the National Academy of Sciences**, v. 112, n. 41, p. 12580–12585, 2015.

EDGAR, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. **BMC Bioinformatics**, v. 5, n. 1, p. 113, 2004.

EDLER, D.; KLEIN, J.; ANTONELLI, A.; SILVESTRO, D. raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. **Methods in Ecology and Evolution**, v. 12, n. 2, p. 373–377, 2021.

ERDTMAN, H. Some aspects of chemotaxonomy. **Pure and Applied Chemistry**, v. 6, n. 4, p. 679–708, 1963.

FACCIN, H.; VIANA, C.; NASCIMENTO, P. C.; BOHRER, D.; CARVALHO, L. M. Study of ion suppression for phenolic compounds in medicinal plant extracts using liquid chromatography–electrospray tandem mass spectrometry. **Journal of Chromatography A**, v. 1427, p. 111–124, 2016.

FAIX, J. J.; GUSTINE, D. L.; WRIGHT, M. J. Beta-nitropropionic acid concentration of crownvetch plant parts as affected by growth-temperature-maturation variables 1. **Agronomy Journal**, v. 70, n. 4, p. 689–691, 1978.

FALCONE FERREYRA, M. L.; RIUS, S. P.; CASATI, P. Flavonoids: biosynthesis, biological functions, and biotechnological applications. **Frontiers in Plant Science**, v. 3, Sep., 2012.

FANG, C.; LUO, J. Metabolic GWAS-based dissection of genetic bases underlying the diversity of plant metabolism. **The Plant Journal**, v. 97, n. 1, p. 91–100, 2019.

FINNEGAN, R. A.; MUELLER, W. H. Chemical examination of a toxic extract of *Indigofera endecaphylla*. The endecaphyllins. **Journal of Pharmaceutical Sciences**, v. 54, n. 8, p. 1136–1144, 1965.

FLOROS, D. J.; JENSEN, P. R.; DORRESTEIN, P. C.; KOYAMA, N. A metabolomics

guided exploration of marine natural product chemical space. **Metabolomics**, v. 12, n. 9, p. 145, 2016.

FOX RAMOS, A. E.; ALCOVER, C.; EVANNO, L.; MACIUK, A.; LITAUDON, M.; DUPLAIS, C.; BERNADAT, G.; GALLARD, J.-F.; JULLIAN, J.-C.; MOURAY, E.; GRELLIER, P.; LOISEAU, P. M.; POMEL, S.; POUPON, E.; CHAMPY, P.; BENIDDIR, M. A. Revisiting previously investigated plants: A molecular networking-based study of *Geissospermum laeve*. **Journal of Natural Products**, v. 80, n. 4, p. 1007–1014, 2017.

FRAIGE, K.; DAMETTO, A. C.; ZERAIK, M. L.; DE FREITAS, L.; SARAIVA, A. C.; MEDEIROS, A. I.; CASTRO-GAMBOA, I.; CAVALHEIRO, A. J.; SILVA, D. H. S.; LOPES, N. P.; BOLZANI, V. S. Dereplication by HPLC-DAD-ESI-MS/MS and screening for biological activities of *Byrsonima* Species (Malpighiaceae). **Phytochemical Analysis**, v. 29, n. 2, p. 196–204, 2018.

FRANCIS, K.; SMITHERMAN, C.; NISHINO, S. F.; SPAIN, J. C.; GADDA, G. The biochemistry of the metabolic poison propionate 3-nitronate and its conjugate acid, 3-nitropropionate. **IUBMB Life**, v. 65, n. 9, p. 759–768, 2013.

FUNARI, C. S.; CARNEIRO, R. L.; ANDRADE, A. M.; HILDER, E. F.; CAVALHEIRO, A. J. Green chromatographic fingerprinting: An environmentally friendly approach for the development of separation methods for fingerprinting complex matrices. **Journal of Separation Science**, v. 37, n. 1–2, p. 37–44, 2014.

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.

GEMPERLINE, E.; KELLER, C.; LI, L. Mass spectrometry in plant-omics. **Analytical Chemistry**, v. 88, n. 7, p. 3422–3434, 2016.

GERIS, R.; RIBEIRO, P. R.; DA SILVA BRANDÃO, M.; DA SILVA, H. H. G.; DA SILVA, I. G. Bioactive Natural Products as Potential Candidates to Control *Aedes aegypti*, the Vector of Dengue. *In:* **Studies in Natural Products Chemistry**. [*S.l.*]: Elsevier, 2012. v. 37, p. 277–376. Chapter 10.

GNANASUNDERAM, C.; SUTHERLAND, O. R. W. Hiptagin and other aliphatic nitro esters in *Lotus pedunculatus*. **Phytochemistry**, v. 25, n. 2, p. 409–410, 1986.

GORTER, K. Hiptagin, a new glucoside from *Hiptage madablota*. **Gaertu Bull Jordin Bot Buitenzorg**, v. 2, p. 187–202, 1920.

GOTTLIEB, O. R. Evolution of Natural Products. In: ROWE, J. W. (Ed.). **Natural Products of Woody Plants I: chemicals extraneous to the lignocellulosic cell wall**. Berlin, Heidelberg: Springer, 1989. p. 125–153. Chapter 3.

GOWDA, G. A. N.; DJUKOVIC, D. Overview of Mass Spectrometry-Based Metabolomics: Opportunities and Challenges. In: RAFTERY, D. (Ed.). **Mass**

Spectrometry in Metabolomics. Methods in Molecular Biology (Methods and Protocols). New York, NY: Humana Press, 2014. v. 1198, p. 3–12.

GUIJAS, C.; MONTENEGRO-BURKE, J. R.; DOMINGO-ALMENARA, X.; PALERMO, A.; WARTH, B.; HERMANN, G.; KOELLENSPERGER, G.; HUAN, T.; URITBOONTHAI, W.; AISPORNA, A. E.; WOLAN, D. W.; SPILKER, M. E.; BENTON, H. P.; SIUZDAK, G. METLIN: A Technology platform for identifying knowns and unknowns. **Analytical Chemistry**, v. 90, n. 5, p. 3156–3164, 2018.

GUSTINE, D. L.; MOYER, B. G.; WANGSNESS, P. J.; SHENK, J. S. Ruminal metabolism of 3-nitropropanoyl-D-glucopyranoses from Crownvetch. **Journal of Animal Science**, v. 44, n. 6, p. 1107–1111, 1977.

HARLOW, M. C.; STERMITZ, F. R.; THOMAS, R. D. Isolation of nitro compounds from *Astragalus species*. **Phytochemistry**, v. 14, n. 5–6, p. 1421–1423, 1975.

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, 2015.

HORAI, H.; ARITA, M.; KANAYA, S.; NIHEI, Y.; IKEDA, T.; SUWA, K.; OJIMA, Y.; TANAKA, K.; TANAKA, S.; AOSHIMA, K.; ODA, Y.; KAKAZU, Y.; KUSANO, M.; TOHGE, T.; MATSUDA, F.; SAWADA, Y.; HIRAI, M. Y.; NAKANISHI, H.; IKEDA, K.; AKIMOTO, N.; MAOKA, T.; TAKAHASHI, H.; ARA, T.; SAKURAI, N.; SUZUKI, H.; SHIBATA, D.; NEUMANN, S.; IIDA, T.; TANAKA, K.; FUNATSU, K.; MATSUURA, F.; SOGA, T.; TAGUCHI, R.; SAITO, K.; NISHIOKA, T. MassBank: a public repository for sharing mass spectral data for life sciences. **Journal of Mass Spectrometry**, v. 45, n. 7, p. 703–714, 2010.

HUBERT, J.; NUZILLARD, J.-M.; RENAULT, J.-H. Dereplication strategies in natural product research: How many tools and methodologies behind the same concept? **Phytochemistry Reviews**, v. 16, n. 1, p. 55–95, 2017.

HUERTA-REYES, M.; FONSECA JUÁREZ, R. M.; AGUILAR-ROJAS, A. *Heteropterys* genus: A review of its phytochemistry and pharmacology. **International Journal of Pharmacology**, v. 11, n. 6, p. 523–531, 2015.

HUERTA-REYES, M.; HERRERA-RUIZ, M.; GONZÁLEZ-CORTAZAR, M.; ZAMILPA, A.; LEÓN, E.; REYES-CHILPA, R.; AGUILAR-ROJAS, A.; TORTORIELLO, J. Neuropharmacological in vivo effects and phytochemical profile of the extract from the aerial parts of *Heteropterys brachiata* (L.) DC. (Malpighiaceae). **Journal of Ethnopharmacology**, v. 146, n. 1, p. 311–317, 2013.

HUSSEIN, A. A.; GÓMEZ, B.; RAMOS, M.; HELLER, M.; COLEY, P. D.; SOLÍS, P. N.; GUPTA, M. P. Constituents of *Hiraea reclinata* and their anti-HIV activity. **Revista Latinoamericana de Química**, v. 31, n. 2, p. 5–8, 2003.

ICH. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. **ICH - Q2(R1) Validation of**

analytical procedures: text and methodology, 2005.

ISAH, T. Stress and defense responses in plant secondary metabolites production. **Biological Research**, v. 52, n. 1, p. 39, 2019.

JUDD, W. S.; CAMPBELL, C. S.; KELLOGG, E. A.; STEVENS, P. F.; DONOGHUE, M. J. **Plant Systematics: A Phylogenetic Approach**. 2nd. ed. Sunderland, MA: Sinauer Associates, Inc., 2002.

KANG, K. B.; ERNST, M.; HOOFT, J. J. J.; SILVA, R. R.; PARK, J.; MEDEMA, M. H.; SUNG, S. H.; DORRESTEIN, P. C. Comprehensive mass spectrometry-guided phenotyping of plant specialized metabolites reveals metabolic diversity in the cosmopolitan plant family Rhamnaceae. **The Plant Journal**, v. 98, n. 6, p. tpj.14292, 2019.

KEARSE, M.; MOIR, R.; WILSON, A.; STONES-HAVAS, S.; CHEUNG, M.; STURROCK, S.; BUXTON, S.; COOPER, A.; MARKOWITZ, S.; DURAN, C.; THIERER, T.; ASHTON, B.; MEINTJES, P.; DRUMMOND, A. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. **Bioinformatics**, v. 28, n. 12, p. 1647–1649, 2012.

KESSNER, D.; CHAMBERS, M.; BURKE, R.; AGUS, D.; MALLICK, P. ProteoWizard: open source software for rapid proteomics tools development. **Bioinformatics**, v. 24, n. 21, p. 2534–2536, 2008.

LALROTLUANGA; NGENTE, L.; NACHIMUTHU, S. K.; GURUSWAMI, G. Insecticidal and repellent activity of *Hiptage benghalensis* L. Kruz (Malpighiaceae) against mosquito vectors. **Parasitology Research**, v. 111, n. 3, p. 1007–1017, 2012.

LEE, S. T.; COOK, D.; PFISTER, J. A.; ALLEN, J. G.; COLEGATE, S. M.; RIET-CORREA, F.; TAYLOR, C. M. Monofluoroacetate-containing plants that are potentially toxic to livestock. **Journal of Agricultural and Food Chemistry**, v. 62, n. 30, p. 7345–7354, 2014.

LEE, S. T.; COOK, D.; RIET-CORREA, F.; PFISTER, J. A.; ANDERSON, W. R.; LIMA, F. G.; GARDNER, D. R. Detection of monofluoroacetate in *Palicourea* and *Amorimia* species. **Toxicon**, v. 60, n. 5, p. 791–796, 2012.

LEE, S. Y.; LEE, I. K.; LEE, K. R. Flavonol glycosides from *Cardamine komarovii*. **Canadian Journal of Chemistry**, v. 90, n. 9, p. 753–757, 2012.

LI, D.; GAQUEREL, E. Next-Generation Mass Spectrometry Metabolomics Revives the Functional Analysis of Plant Metabolic Diversity. **Annual Review of Plant Biology**, v. 72, n. 1, p. 867–891, 2021.

LIN, L.-Z.; HARNLY, J. M. Phenolic compounds and chromatographic profiles of pear skins (*Pyrus* spp.). **Journal of Agricultural and Food Chemistry**, v. 56, n. 19, p. 9094–9101, 2008.

LIPSITZ, Y. Y.; TIMMINS, N. E.; ZANDSTRA, P. W. Quality cell therapy manufacturing by design. **Nature Biotechnology**, v. 34, n. 4, p. 393–400, 2016.

LUDWIG, M.; NOTHIAS, L.-F.; DÜHRKOP, K.; KOESTER, I.; FLEISCHAUER, M.; HOFFMANN, M. A.; PETRAS, D.; VARGAS, F.; MORSY, M.; ALUWIHARE, L.; DORRESTEIN, P. C.; BÖCKER, S. Database-independent molecular formula annotation using Gibbs sampling through ZODIAC. **Nature Machine Intelligence**, v. 2, n. 10, p. 629–641, 2020.

MACASKILL, J. J.; MANLEY-HARRIS, M.; FIELD, R. J. Quantification of nitropropanoyl glucosides in karaka nuts before and after treatment. **Food Chemistry**, v. 175, p. 543–548, 2015.

MADDISON, W. P.; MADDISON, D. R. **Mesquite:** a modular system for evolutionary analysis. Version 3.61. [*S.l.*] : Mesquite Project, 2006. Avaliable in: https://www.mesquiteproject.org/. Access in 25 Mar., 2022.

MAGALHAES, N. M. G.; SOUSA, J. P. B.; DEMARQUE, D. P.; SALVADOR, C. E. de M.; ALBERNAZ, L. C.; ACHEE, N. L.; GRIECO, J. P.; ESPINDOLA, L. S. Combining chemometric and phytochemical tools to isolate and characterize activity of *Vismia gracilis* compounds against *Aedes aegypti*. **Natural Product Research**, 2021. doi: 10.1080/14786419.2021.1912747

MAGIERA, S.; BARANOWSKA, I.; LAUTENSZLEGER, A. UHPLC-UV method for the determination of flavonoids in dietary supplements and for evaluation of their antioxidant activities. **Journal of Pharmaceutical and Biomedical Analysis**, v. 102, p. 468–475, 2015.

MAJAK, W.; BENN, M. Additional esters of 3-nitropropanoic acid and glucose from fruit of the New Zealand karaka tree, *Corynocarpus laevigatus*. **Phytochemistry**, v. 35, n. 4, p. 901–903, 1994.

MAJAK, W.; BENN, M.; MCEWAN, D.; PASS, M. A. Three nitropropanoyl esters of glucose from *Indigofera linnaei*. **Phytochemistry**, v. 31, n. 7, p. 2393–2395, 1992.

MAMODE CASSIM, A.; GOUGUET, P.; GRONNIER, J.; LAURENT, N.; GERMAIN, V.; GRISON, M.; BOUTTÉ, Y.; GERBEAU-PISSOT, P.; SIMON-PLAS, F.; MONGRAND, S. Plant lipids: Key players of plasma membrane organization and function. **Progress in Lipid Research**, v. 73, p. 1–27, 2019.

MANNOCHIO-RUSSO, H.; BUENO, P. C. P.; BAUERMEISTER, A.; DE ALMEIDA, R. F.; DORRESTEIN, P. C.; CAVALHEIRO, A. J.; BOLZANI, V. S. Can statistical evaluation tools for chromatographic method development assist in the natural products workflow? A case study on selected species of the plant family Malpighiaceae. **Journal of Natural Products**, v. 83, n. 11, p. 3239–3249, 2020.

MANNOCHIO RUSSO, H.; FERREIRA QUEIROZ, E.; MARCOURT, L.; RUTZ, A.; ALLARD, P.; DE ALMEIDA, R. F.; MARQUES CARVALHO, N.; WOLFENDER, J.; DA SILVA BOLZANI, V. Phytochemical analysis of the methanolic leaves extract of

Niedenzuella multiglandulosa (Malpighiaceae), a plant species toxic to cattle in Brazil. **Phytochemistry Letters**, v. 37, p. 10–16, Jun., 2020.

MCKENNA, D. J.; CALLAWAY, J. C.; GROB, C. S. The scientific investigation of ayahuasca: A review of past and current research. **The Heffer Review of Psychedelic Research**, v. 1, p. 65–76, Sep., 1998.

MEDEMA, M. H. The year 2020 in natural product bioinformatics: an overview of the latest tools and databases. **Natural Product Reports**, v. 38, n. 2, p. 301–306, 2021.

MELO, F. L.; BENATI, F. J.; JUNIOR, W. A. R.; DE MELLO, J. C. P.; NOZAWA, C.; LINHARES, R. E. C. The *in vitro* antiviral activity of an aliphatic nitro compound from *Heteropteris aphrodisiaca*. **Microbiological Research**, v. 163, n. 2, p. 136–139, 2008.

MELO, S. J.; SOUSA, J. P. B.; SÁ, M. G.; MORAIS, L. S.; MAGALHÃES, N. M. G.; GOUVEIA, F. N.; ALBERNAZ, L. C.; ESPINDOLA, L. S. *Machaerium acutifolium* compounds with larvicidal activity against *Aedes aegypti*. **Pest Management Science**, v. 77, n. 3, p. 1444–1451, 2021.

MIERZIAK, J.; KOSTYN, K.; KULMA, A. Flavonoids as omportant molecules of plant interactions with the environment. **Molecules**, v. 19, n. 10, p. 16240–16265, 2014.

MISRA, B. B. New software tools, databases, and resources in metabolomics: updates from 2020. **Metabolomics**, v. 17, n. 5, p. 49, 2021.

MONDOLOT, L.; LA FISCA, P.; BUATOIS, B.; TALANSIER, E.; DE KOCHKO, A.; CAMPA, C. Evolution in caffeoylquinic acid content and histolocalization during *Coffea canephora* leaf development. **Annals of Botany**, v. 98, n. 1, p. 33–40, 2006.

MONTGOMERY, D. C. **Design and Analysis of Experiments**. 8th. ed. [*S.l.*] : John Wiley & Sons, Inc., 2012.

MOYER, B. G.; PFEFFER, P. E.; MONIOT, J. L.; SHAMMA, M.; GUSTINE, D. L. Corollin, coronillin and coronarian: Three new 3-nitropropanoyl-d-glucopyranoses from *Coronilla varia*. **Phytochemistry**, v. 16, n. 3, p. 375–377, 1977.

MOYER, B. G.; PFEFFER, P. E.; VALENTINE, K. M.; GUSTINE, D. L. 3-Nitropropanoyl-d-glucopyranoses of *Corynocarpus laevigatus*. **Phytochemistry**, v. 18, n. 1, p. 111–113, 1979.

NA, C. S.; LEE, Y. H.; KIM, T. W.; MURAI, Y.; HONG, S. H. Aliphatic nitro compounds from roots of *Astragalus sikokianus* (Fabaceae) in Korea. **Biochemical Systematics and Ecology**, v. 60, p. 120–122, 2015.

NALDI, M.; FIORI, J.; GOTTI, R.; PÉRIAT, A.; VEUTHEY, J.-L.; GUILLARME, D.; ANDRISANO, V. UHPLC determination of catechins for the quality control of green tea. **Journal of Pharmaceutical and Biomedical Analysis**, v. 88, p. 307–314, 2014.

NAZIR, I.; UR RAHMAN, N.; ALVI, Z.; HAFIZUR RAHMAN, M.; SENDKER, J.; ZHANG, T.; FRANKISH, N.; SHERIDAN, H. Antidiabetic Activities of an LC/MS Fingerprinted Aqueous Extract of *Fagonia cretica* L. in Preclinical Models. **Planta Medica**, v. 83, n. 14/15, p. 1141–1148, 2017.

NERI-NUMA, I. A.; SORIANO SANCHO, R. A.; PEREIRA, A. P. A.; PASTORE, G. M. Small Brazilian wild fruits: Nutrients, bioactive compounds, health-promotion properties and commercial interest. **Food Research International**, v. 103, p. 345–360, 2018.

NEWMAN, D. J.; CRAGG, G. M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. **Journal of Natural Products**, v. 83, n. 3, p. 770–803, 2020.

NIXON, K. **Winlada.** Version 1.0, 1999. Available in: http://www.diversityoflife.org/winclada/. Access in 25 Mar., 2022.

NOTHIAS, L.F. F.; NOTHIAS-ESPOSITO, M.; DA SILVA, R.; WANG, M.; PROTSYUK, I.; ZHANG, Z.; SARVEPALLI, A.; LEYSSEN, P.; TOUBOUL, D.; COSTA, J.; PAOLINI, J.; ALEXANDROV, T.; LITAUDON, M.; DORRESTEIN, P. C. Bioactivity-based molecular networking for the discovery of drug leads in natural product bioassay-guided fractionation. **Journal of Natural Products**, v. 81, n. 4, p. 758–767, 2018.

NOTHIAS, L. et al. Feature-based molecular networking in the GNPS analysis environment. **Nature Methods**, v. 17, n. 9, p. 905–908, 2020.

NUNES, D. S.; HAAG, A.; BESTMANN, H. J. Inhaltsstoffe der Rinde von *Dalbergia monetaria* L. Drei neue Isoflavon-C-glucoside. **Liebigs Annalen der Chemie**, v. 1989, n. 4, p. 331–335, 1989.

OLIVON, F.; APEL, C.; RETAILLEAU, P.; ALLARD, P. M.; WOLFENDER, J. L.; TOUBOUL, D.; ROUSSI, F.; LITAUDON, M.; DESRAT, S. Searching for original natural products by molecular networking: detection, isolation and total synthesis of chloroaustralasines. **Organic Chemistry Frontiers**, v. 5, n. 14, p. 2171–2178, 2018.

ORLANDINI, S.; PASQUINI, B.; CAPRINI, C.; DEL BUBBA, M.; PINZAUTI, S.; FURLANETTO, S. Analytical Quality by Design in pharmaceutical quality assurance: Development of a capillary electrophoresis method for the analysis of zolmitriptan and its impurities. **Electrophoresis**, v. 36, n. 21–22, p. 2642–2649, 2015.

PANG, Z.; CHONG, J.; ZHOU, G.; LIMA MORAIS, D. A.; CHANG, L.; BARRETTE, M.; GAUTHIER, C.; JACQUES, P.É.; LI, S.; XIA, J. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. **Nucleic Acids Research**, v. 49, n. W1, p. W388–W396, 2021.

PAULA-FREIRE, L. I. G.; MENDES, F. R.; MOLSKA, G. R.; DUARTE-ALMEIDA, J. M.; CARLINI, E. A. Comparison of the chemical composition and biological effects of the roots, branches and leaves of *Heteropterys tomentosa* A. Juss. **Journal of**

Ethnopharmacology, v. 145, n. 2, p. 647–652, 2013.

PILON, A.; SELEGATO, D.; FERNANDES, R.; BUENO, P.; PINHO, D.; CARNEVALE NETO, F.; FREIRE, R.; CASTRO-GAMBOA, I.; BOLZANI, V.; LOPES, N. Metabolômica de plantas: métodos e desafios. **Química Nova**, v. 43, n. 3, p. 329–354, 2020.

PINNA, M. C. C. Concepts and tests of homology in the cladistic paradigm. **Cladistics**, v. 7, n. 4, p. 367–394, 1991.

POSSOBOM, C. C. F.; GUIMARÃES, E.; MACHADO, S. R. Leaf glands act as nectaries in *Diplopterys pubipetala* (Malpighiaceae). **Plant Biology**, v. 12, n. 6, p. 863–870, 2010.

QUEIROZ, M. M. F.; MARTI, G.; QUEIROZ, E. F.; MARCOURT, L.; CASTRO-GAMBOA, I.; BOLZANI, V. S.; WOLFENDER, J. L. LC-MS/MS quantitative determination of *Tetrapterys mucronata* alkaloids, a plant occasionally used in Ayahuasca preparation. **Phytochemical Analysis**, v. 26, n. 3, p. 183–188, 2015.

QUEIROZ, M. M. F.; QUEIROZ, E. F.; ZERAIK, M. L.; EBRAHIMI, S. N.; MARCOURT, L.; CUENDET, M.; CASTRO-GAMBOA, I.; HAMBURGER, M.; SILVA BOLZANI, V.; WOLFENDER, J. L. Chemical composition of the bark of *Tetrapterys mucronata* and identification of acetylcholinesterase inhibitory constituents. **Journal of Natural Products**, v. 77, n. 3, p. 650–656, 2014.

QUINN, R. A.; NOTHIAS, L. F.; VINING, O.; MEEHAN, M.; ESQUENAZI, E.; DORRESTEIN, P. C. Molecular networking as a drug discovery, drug metabolism, and precision medicine strategy. **Trends in Pharmacological Sciences**, v. 38, n. 2, p. 143–154, 2017.

RAMBAUT, A.; SUCHARD, M. A.; XIE, D.; DRUMMOND, A. J. **Tracer.** Version v1.6, 2014. Available in: http://beast.bio.ed.ac.uk/Tracer. 25 Mar., 2022.

RENNER, S. S.; SCHAEFER, H. The evolution and loss of oil-offering flowers: new insights from dated phylogenies for angiosperms and bees. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 365, n. 1539, p. 423–435, 2010.

REZENDE-TEIXEIRA, P.; DUSI, R. G.; JIMENEZ, P. C.; ESPINDOLA, L. S.; COSTA-LOTUFO, L. V. What can we learn from commercial insecticides? Efficacy, toxicity, environmental impacts, and future developments. **Environmental Pollution**, v. 300, n. February, p. 118983, 2022.

RIET-CORREA, F.; MEDEIROS, R. M. T.; SCHILD, A. L. A review of poisonous plants that cause reproductive failure and malformations in the ruminants of Brazil. **Journal of Applied Toxicology**, v. 32, n. 4, p. 245–254, 2012.

ROMAN JÚNIOR, W. A.; CARDOSO, M. L. C.; VILEGAS, W.; NAKAMURA, C. V.; DIAS FILHO, B. P.; MELLO, J. C. P. De. 2,3,4,6-Tetra-O-(3-nitropropanoyl)-O-β-D-

glucopyranoside, a new antimicrobial from the roots of *Heteropteris aphrodisiaca*. **Acta Farmaceutica Bonaerense**, v. 24, n. 4, p. 543–545, 2005.

RONQUIST, F.; HUELSENBECK, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. **Bioinformatics**, v. 19, n. 12, p. 1572–1574, 2003.

ROSA, H. S.; KOETZ, M.; SANTOS, M. C.; JANDREY, E. H. F.; FOLMER, V.; HENRIQUES, A. T.; MENDEZ, A. S. L. Extraction optimization and UHPLC method development for determination of the 20-hydroxyecdysone in *Sida tuberculata* leaves. **Steroids**, v. 132, n. January, p. 33–39, 2018.

RUE, E. A.; RUSH, M. D.; VAN BREEMEN, R. B. Procyanidins: a comprehensive review encompassing structure elucidation via mass spectrometry. **Phytochemistry Reviews**, v. 17, n. 1, p. 1–16, 2018.

RUTZ, A.; SOROKINA, M.; GALGONEK, J.; MIETCHEN, D.; WILLIGHAGEN, E.; GRAHAM, J.; STEPHAN, R.; PAGE, R.; VONDRÁŠEK, J.; STEINBECK, C.; PAULI, G. F.; WOLFENDER, J.-L.; BISSON, J.; ALLARD, P.-M. Open natural products research: curation and dissemination of biological occurrences of chemical structures through Wikidata. **bioRxiv**, p. 2021.02.28.433265, 2021.

S-MATRIX. Fusion QbD Software: **LC Method Development - User's Guide**, [201-].

SAINTY, D.; BAILLEUL, F.; DELAVEAU, P.; JACQUEMIN, H. Malpighiacees: nouvelle famille a iridoides etude du *Stigmaphyllon sagittatum*. **Journal of Natural Products**, v. 44, n. 5, p. 576–578, 1981.

SALEM, M. A.; MICHAEL WILLIAMS, J.; WAINWRIGHT, S. J.; HIPKIN, C. R. Nitroaliphatic compounds in *Hippocrepis comosa* and other legumes in the European flora. **Phytochemistry**, v. 40, n. 1, p. 89–91, 1995.

SAMOYLENKO, V.; RAHMAN, M. M.; TEKWANI, B. L.; TRIPATHI, L. M.; WANG, Y.-H.; KHAN, S. I.; KHAN, I. A.; MILLER, L. S.; JOSHI, V. C.; MUHAMMAD, I. *Banisteriopsis caapi*, a unique combination of MAO inhibitory and antioxidative constituents for the activities relevant to neurodegenerative disorders and Parkinson's disease. **Journal of Ethnopharmacology**, v. 127, n. 2, p. 357–367, 2010.

SANTOS JÚNIOR, H. M.; CAMPOS, V. A. C.; ALVES, D. S.; CAVALHEIRO, A. J.; SOUZA, L. P.; BOTELHO, D. M. S.; CHALFOUN, S. M.; OLIVEIRA, D. F. Antifungal activity of flavonoids from *Heteropterys byrsonimifolia* and a commercial source against Aspergillus ochraceus: *In silico* interactions of these compounds with a protein kinase. **Crop Protection**, v. 62, p. 107–114, 2014.

SAWADA, Y.; NAKABAYASHI, R.; YAMADA, Y.; SUZUKI, M.; SATO, M.; SAKATA, A.; AKIYAMA, K.; SAKURAI, T.; MATSUDA, F.; AOKI, T.; HIRAI, M. Y.; SAITO, K. RIKEN tandem mass spectral database (ReSpect) for phytochemicals: A plant-specific MS/MS-based data resource and database. **Phytochemistry**, v. 82, p. 38–45, 2012.

SCHMIDT, A. H.; STANIC, M.; MOLNÁR, I. *In silico* robustness testing of a compendial HPLC purity method by using of a multidimensional design space build by chromatography modeling—Case study pramipexole. **Journal of Pharmaceutical and Biomedical Analysis**, v. 91, p. 97–107, 2014.

SCHRIMPE-RUTLEDGE, A. C.; CODREANU, S. G.; SHERROD, S. D.; MCLEAN, J. A. Untargeted metabolomics strategies—Challenges and emerging directions. **Journal of the American Society for Mass Spectrometry**, v. 27, n. 12, p. 1897–1905, 2016.

SERENO, P. C. Logical basis for morphological characters in phylogenetics. **Cladistics**, v. 24, n. 3, p. 565–587, 2007.

SHANNON, P.; MARKIEL, A.; OZIER, O.; BALIGA, N. S.; WANG, J. T.; RAMAGE, D.; AMIN, N.; SCHWIKOWSKI, B.; IDEKER, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. **Genome Research**, v. 13, n. 11, p. 2498–2504, 2003.

SHAO, J.; CAO, W.; QU, H.; PAN, J.; GONG, X. A novel quality by design approach for developing an HPLC method to analyze herbal extracts: A case study of sugar content analysis. **PLOS ONE**, v. 13, n. 6, p. e0198515, 2018.

SHAW, S. D.; BILLING, T. Karaka (*Corynocarpus laevigatus*) toxicosis in north island brown kiwi (*Apteryx mantelli*). **Veterinary Clinics of North America: Exotic Animal Practice**, v. 9, n. 3, p. 545–549, 2006.

SIENTZOFF, P.; HUBERT, J.; JANIN, C.; VOUTQUENNE-NAZABADIOKO, L.; RENAULT, J.-H.; NUZILLARD, J.-M.; HARAKAT, D.; MAGID, A. Fast identification of radical scavengers from *Securigera varia* by combining ^{13}C-NMR-based dereplication to bioactivity-guided fractionation. **Molecules**, v. 20, n. 8, p. 14970–14984, 2015.

SILVA, D. B.; TURATTI, I. C. C.; GOUVEIA, D. R.; ERNST, M.; TEIXEIRA, S. P.; LOPES, N. P. Mass spectrometry of flavonoid vicenin-2, based sunlight barriers in *Lychnophora* species. **Scientific Reports**, v. 4, n. 1, p. 4309, 2015.

SILVA, M. V.; SILVA, S. A.; TEIXERA, T. L.; DE OLIVEIRA, A.; MORAIS, S. A. L.; DA SILVA, C. V.; ESPINDOLA, L. S.; SOUSA, R. M. F. Essential oil from leaves of *Eugenia calycina* Cambes: Natural larvicidal against *Aedes aegypti*. **Journal of the Science of Food and Agriculture**, v. 101, n. 3, p. 1202–1208, 2021.

SILVA, R. L.; DEMARQUE, D. P.; DUSI, R. G.; SOUSA, J. P. B.; ALBERNAZ, L. C.; ESPINDOLA, L. S. Residual larvicidal activity of quinones against *Aedes aegypti*. **Molecules**, v. 25, n. 17, p. 3978, 2020.

SILVA, R. L.; MELLO, T. R. B.; SOUSA, J. P. B.; ALBERNAZ, L. C.; MAGALHÃES, N. M. G.; MORAIS, L. S.; FRANCISCO, L. R.; LEAL, W. S.; ESPINDOLA, L. S. Brazilian Cerrado biome essential oils to control the arbovirus vectors *Aedes aegypti* and *Culex quinquefasciatus*. **Industrial Crops and Products**, v. 178, p. 114568, Apr., 2022.

SILVA, R. R.; DORRESTEIN, P. C.; QUINN, R. A. Illuminating the dark matter in metabolomics. **Proceedings of the National Academy of Sciences**, v. 112, n. 41, p. 12549–12550, 2015.

SILVA, R. R.; WANG, M.; NOTHIAS, L.-F.; VAN DER HOOFT, J. J. J.; CARABALLO-RODRÍGUEZ, A. M.; FOX, E.; BALUNAS, M. J.; KLASSEN, J. L.; LOPES, N. P.; DORRESTEIN, P. C. Propagating annotations of molecular networks using *in silico* fragmentation. **PLOS Computational Biology**, v. 14, n. 4, p. e1006089, 2018.

SILVÉRIO, M. R. S.; ESPINDOLA, L. S.; LOPES, N. P.; VIEIRA, P. C. Plant natural products for the control of *Aedes aegypti*: The main vector of important arboviruses. **Molecules**, v. 25, n. 15, p. 3484, 2020.

SNYDER, L. R.; KIRKLAND, J. J.; DOLAN, J. W. **Introduction to Modern Liquid Chromatography**. 3rd. ed. [*S.l.*] : John Wiley & Sons, Inc., 2009.

SPÁČIL, Z.; NOVÁKOVÁ, L.; SOLICH, P. Comparison of positive and negative ion detection of tea catechins using tandem mass spectrometry and ultra high performance liquid chromatography. **Food Chemistry**, v. 123, n. 2, p. 535–541, 2010.

STERMITZ, F. R.; HNATYSZYN, O.; BANDONI, A. L.; RONDINA, R. V.D.; COUSSIO, J. D. Screening of Argentine plants for aliphatic nitro compounds: Hiptagin from *Heteropteris angustifolia*. **Phytochemistry**, v. 14, n. 5–6, p. 1341–1345, 1975.

STERMITZ, F. R.; LOWRY, W. T.; UBBEN, E.; SHARIFI, I. 1,6-Di-3-nitropropanoyl-β-D-glucopyranoside from *Astragalus cibarius*. **Phytochemistry**, v. 11, n. 12, p. 3525–3527, 1972.

STOBIECKI, M.; STASZKÓW, A.; PIASECKA, A.; GARCIA-LOPEZ, P. M.; ZAMORA-NATERA, F.; KACHLICKI, P. LC-MSMS Profiling of flavonoid conjugates in wild Mexican lupine, *Lupinus reflexus*. **Journal of Natural Products**, v. 73, n. 7, p. 1254–1260, 2010.

STUESSY, T. F. **Plant taxonomy: the systematic evaluation of comparative data**. 2nd. ed. New York: Columbia University Press, 2008.

SU, Y.; LI, C.; GAO, Y.; DI, L.; ZHANG, X.; GUO, D. Acryloylated glucose 3-nitropropanoates from *Indigofera kirilowii*. **Journal of Natural Products**, v. 68, n. 12, p. 1785–1786, 2005.

SUGENO, W.; MATSUDA, K. Adult secretions of four Japanese Chrysomelinae (Coleoptera: Chrysomelidae). **Applied Entomology and Zoology**, v. 37, n. 1, p. 191–197, 2002.

SUMNER, L. W.; AMBERG, A.; BARRETT, D.; BEALE, M. H.; BEGER, R.; DAYKIN, C. A.; FAN, T. W.-M.; FIEHN, O.; GOODACRE, R.; GRIFFIN, J. L.; HANKEMEIER, T.; HARDY, N.; HARNLY, J.; HIGASHI, R.; KOPKA, J.; LANE, A. N.; LINDON, J. C.;

MARRIOTT, P.; NICHOLLS, A. W.; REILY, M. D.; THADEN, J. J.; VIANT, M. R. Proposed minimum reporting standards for chemical analysis. **Metabolomics**, v. 3, n. 3, p. 211–221, 2007.

TIAN, Y.; LAAKSONEN, O.; HAIKONEN, H.; VANAG, A.; EJAZ, H.; LINDERBORG, K.; KARHU, S.; YANG, B. Compositional diversity among blackcurrant (*Ribes nigrum*) cultivars originating from European countries. **Journal of Agricultural and Food Chemistry**, v. 67, n. 19, p. 5621–5633, 2019.

TIAN, Y.; LIIMATAINEN, J.; ALANNE, A.-L.; LINDSTEDT, A.; LIU, P.; SINKKONEN, J.; KALLIO, H.; YANG, B. Phenolic compounds extracted by acidic aqueous ethanol from berries and leaves of different berry plants. **Food Chemistry**, v. 220, p. 266–281, 2017.

TRIPATHI, A.; VÁZQUEZ-BAEZA, Y.; GAUGLITZ, J. M.; WANG, M.; DÜHRKOP, K.; NOTHIAS-ESPOSITO, M.; ACHARYA, D. D.; ERNST, M.; VAN DER HOOFT, J. J. J.; ZHU, Q.; MCDONALD, D.; BREJNROD, A. D.; GONZALEZ, A.; HANDELSMAN, J.; FLEISCHAUER, M.; LUDWIG, M.; BÖCKER, S.; NOTHIAS, L.-F.; KNIGHT, R.; DORRESTEIN, P. C. Chemically informed analyses of metabolomics mass spectrometry data with Qemistree. **Nature Chemical Biology**, v. 17, n. 2, p. 146–151, 2021.

TYTECA, E.; VEUTHEY, J.-L.; DESMET, G.; GUILLARME, D.; FEKETE, S. Computer assisted liquid chromatographic method development for the separation of therapeutic proteins. **The Analyst**, v. 141, n. 19, p. 5488–5501, 2016.

VALLI, M.; ATANÁZIO, L. C. V.; MONTEIRO, G. C.; COELHO, R. R.; DEMARQUE, D. P.; ANDRICOPULO, A. D.; ESPINDOLA, L. S.; BOLZANI, V. D. S. The potential of biologically active Brazilian plant species as a strategy to search for molecular models for mosquito control. **Planta Medica**, v. 87, n. 01/02, p. 6–23, 2021.

VAN HEERDEN, F. R.; BRANDT, E. V.; ROUX, D. G. Synthesis of the pyranoisoflavonoid, heminitidulan. Isoflavanoid and rotenoid glycosides from the bark of *Dalbergia nitidula* Welw. ex Bak. **Journal of the Chemical Society, Perkin Transactions 1**, p. 2463–2469, 1980.

VÁZQUEZ-BAEZA, Y.; PIRRUNG, M.; GONZALEZ, A.; KNIGHT, R. EMPeror: a tool for visualizing high-throughput microbial community data. **GigaScience**, v. 2, n. 1, p. 16, 2013.

WALD, B.; WRAY, V.; GALENSA, R.; HERRMANN, K. Malonated flavonol glycosides and 3,5-dicaffeoylquinic acid from pears. **Phytochemistry**, v. 28, n. 2, p. 663–664, 1989.

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.

WEI, D.L.; CHANG, S. C.; LIN, S. C.; DOONG, M. L.; JONG, S. C. Production of 3-

nitropropionic acid by *Arthrinium* species. **Current Microbiology**, v. 28, n. 1, p. 1–5, 1994.

WILLIAMS, M. C.; STERMITZ, F. R.; THOMAS, R. D. Nitro compounds in *Astragalus* species. **Phytochemistry**, v. 14, n. 10, p. 2306–2308, 1975.

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

WOLF, S.; SCHMIDT, S.; MÜLLER-HANNEMANN, M.; NEUMANN, S. *In silico* fragmentation for computer assisted identification of metabolite mass spectra. **BMC Bioinformatics**, v. 11, n. 1, p. 148, 2010.

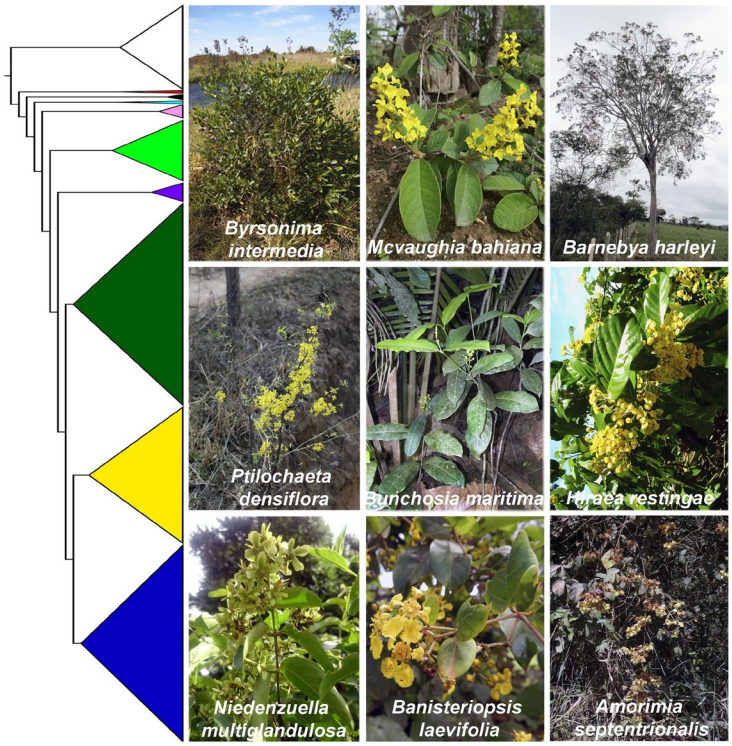
XI, Z.; RUHFEL, B. R.; SCHAEFER, H.; AMORIM, A. M.; SUGUMARAN, M.; WURDACK, K. J.; ENDRESS, P. K.; MATTHEWS, M. L.; STEVENS, P. F.; MATHEWS, S.; DAVIS, C. C. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. **Proceedings of the National Academy of Sciences**, v. 109, n. 43, p. 17519–17524, 2012.

YANG, J. Y.; SANCHEZ, L. M.; RATH, C. M.; LIU, X.; BOUDREAU, P. D.; BRUNS, N.; GLUKHOV, E.; WODTKE, A.; DE FELICIO, R.; FENNER, A.; WONG, W. R.; LININGTON, R. G.; ZHANG, L.; DEBONSI, H. M.; GERWICK, W. H.; DORRESTEIN, P. C. Molecular networking as a dereplication strategy. **Journal of Natural Products**, v. 76, n. 9, p. 1686–1699, 2013.

YANG, M.; ZHOU, Z.; YAO, S.; LI, S.; YANG, W.; JIANG, B.; LIU, X.; WU, W.; QV, H.; GUO, D. Neutral loss ion mapping experiment combined with precursor mass list and dynamic exclusion for screening unstable malonyl glucoside conjugates. **Journal of the American Society for Mass Spectrometry**, v. 27, n. 1, p. 99–107, 2016.

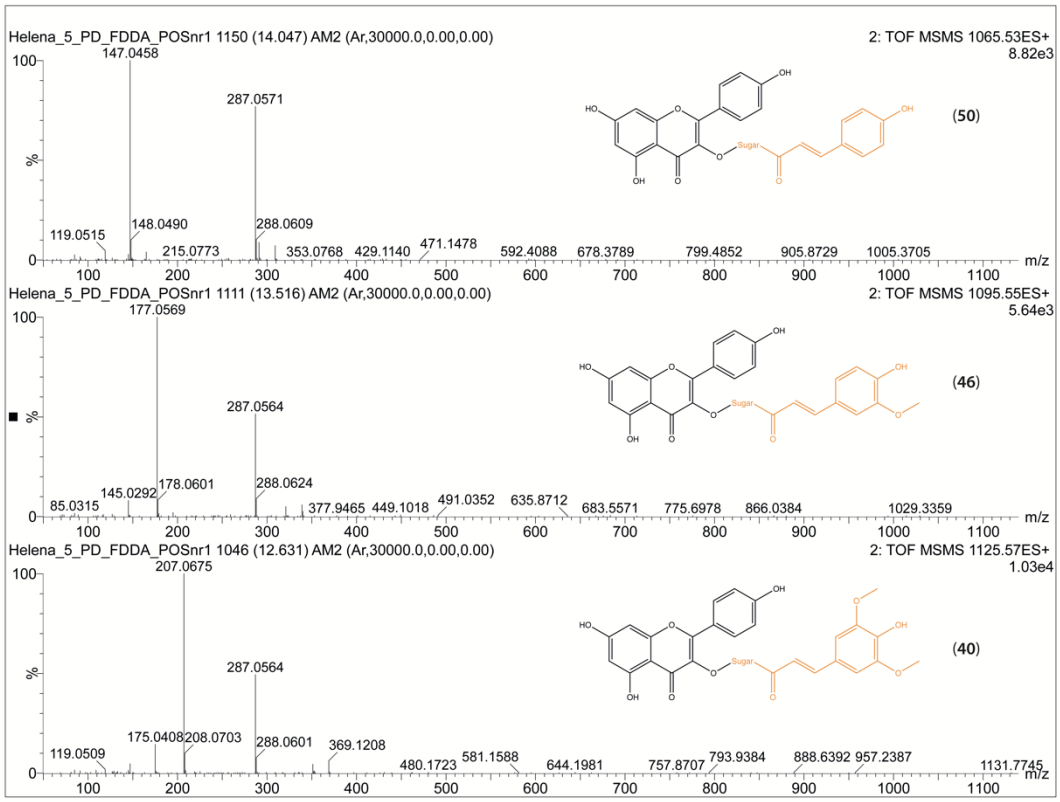
YOSHIKAWA, M.; SUGIMOTO, S.; NAKAMURA, S.; MATSUDA, H. Medicinal Flowers. XXII. Structures of chakasaponins V and VI, chakanoside I, and chakaflavonoside A from flower buds of chinese tea plant (*Camellia sinensis*). **Chemical & Pharmaceutical Bulletin**, v. 56, n. 9, p. 1297–1303, 2008.

SUPPLEMENTARY MATERIAL A – CHAPTER 1



Supplementary Figure 1 – Field photographs of one species from each phylogenetic group from Malpighiaceae family. White - Byrsonimoid clade represented by the species *Byrsonima intermedia* (photograph by R.F. Almeida), Red - Acridocarpoid clade (not sampled in this study), Black - Mcvaughioid clade represented by the species *Mcvaughia bahiana* (photograph by I.R. Guesdon), Light Blue - Barnebyoid clade represented by the species *Barnebya harleyi* (photograph by F. Flores), Pink - Ptilochaetoid clade represented by the species *Ptilochaeta densiflora* (photograph by R.F. Almeida), Light Green - Bunchosioid clade represented by the species *Bunchosia maritima* (photograph by J.M.Braga), Purple - Hiraeoid clade represented by the species *Hiraea restingae* (photograph by R.F. Almeida), Dark Green - Tetrapteroid clade represented by the species *Niedenzuella multiglandulosa* (photograph by N. Carvalho, Yellow - Stigmaphylloid clade represented by the species *Banisteriopsis laevifolia* (photograph by C.F. Hall), and Dark Blue - Malpighioid clade represented by the species *Amorimia septentrionalis* (photograph by M.O.O. Pellegrini) (phylogenetic classification according to Davis and Anderson, 2010).

Source: elaborated by the Author (2022).



Supplementary Figure 2 – MS/MS spectra comparison of compounds **40**, **46** and **50**.

Source: elaborated by the Author (2022).

Supplementary Table 1 – Factors levels, coded values and results of the experimental design for the screening step (quadratic model A- and G- optimal). Constant parameters: temperature (40 ˚C), injection volume (2 µL), flow rate (0.35 mL min^{-1}).

(to be continued)

Run	Variables				Responses			
	Organic Solvent	Gradient time (min)	pH	Column	Total number of peaks	Number of peaks with resolution ≥ 1.5	Number of peaks with resolution ≥ 2.0	Number of peaks with tailing ≤ 1.2
1	CH_3CN	18.8	2.54	Acquity HSS T3	58	32	18	35
2	CH_3CN	26.6	2.54	Cortecs C18+	64	40	26	43
3	CH_3CN	30.0	2.54	Acquity BEH Shield RP18	69	42	26	53
4	CH_3CN	15.0	2.54	Acquity BEH Shield RP18	61	39	25	47
5	CH_3CN	30.0	2.54	Acquity BEH C18	64	33	20	49
6	CH_3CN	15.0	2.54	Acquity BEH C18	59	27	18	45
7	CH_3CN	22.5	2.54	Acquity HSS PFP	47	30	22	28
8	CH_3CN	15.0	2.54	Acquity BEH Shield RP18	61	37	23	42
9	CH_3CN	22.5	6.60	Cortecs C8	47	28	18	29
10	CH_3CN	22.5	6.60	Acquity HSS T3	54	31	18	37
11	CH_3CN	22.5	6.60	Acquity BEH Shield RP18	48	30	20	30
12	CH_3CN	22.5	6.60	Acquity HSS PFP	35	24	14	23
13	CH_3CN	22.5	6.60	Cortecs C18+	46	27	17	28
14	CH_3CN	22.5	6.60	Acquity BEH C18	45	31	16	32
15	CH_3CN	30.0	6.60	Cortecs C8	49	31	23	34
16	CH_3CN	15.0	6.60	Cortecs C8	39	23	14	27
17	CH_3CN	30.0	6.60	Acquity HSS T3	58	32	21	40
18	CH_3CN	15.0	6.60	Acquity HSS T3	49	25	13	36

Supplementary Table 1 – Factors levels, coded values and results of the experimental design for the screening step (quadratic model A- and G- optimal). Constant parameters: temperature (40 ˚C), injection volume (2 µL), flow rate (0.35 mL min^{-1}).

(continuation)

Run	Variables				Responses			
	Organic Solvent	Gradient time (min)	pH	Column	Total number of peaks	Number of peaks with resolution ≥ 1.5	Number of peaks with resolution ≥ 2.0	Number of peaks with tailing ≤ 1.2
19	CH_3CN	30.0	6.60	Acquity HSS PFP	40	25	17	28
20	CH_3CN	15.0	6.60	Acquity HSS PFP	41	20	14	26
21	CH_3CN	30.0	6.60	Cortecs C18+	45	33	25	26
22	CH_3CN	15.0	6.60	Cortecs C18+	43	20	17	29
23	CH_3CN	22.5	6.60	Cortecs C8	45	27	20	30
24	CH_3CN	22.5	6.60	Acquity HSS T3	57	34	18	36
25	CH_3CN	22.5	6.60	Acquity BEH Shield RP18	43	27	20	29
26	CH_3CN	22.5	6.60	Acquity HSS PFP	38	19	17	20
27	CH_3CN	22.5	6.60	Cortecs C18+	44	28	16	23
28	CH_3CN	22.5	6.60	Acquity BEH C18	49	30	18	36
29	CH_3CN	18.8	10.45	Acquity BEH Shield RP18	47	32	25	36
30	CH_3CN	30.0	10.45	Acquity BEH Shield RP18	43	26	24	37
31	CH_3CN	22.5	10.45	Acquity BEH C18	43	29	22	33
32	MeOH	26.3	2.54	Acquity BEH Shield RP18	57	33	22	45
33	MeOH	30.0	2.54	Acquity HSS T3	55	34	24	45
34	MeOH	15.0	2.54	Acquity HSS T3	56	27	15	42
35	MeOH	30.0	2.54	Acquity HSS PFP	45	31	24	35
36	MeOH	15.0	2.54	Acquity HSS PFP	51	24	15	39

Supplementary Table 1 – Factors levels, coded values and results of the experimental design for the screening step (quadratic model A- and G- optimal). Constant parameters: temperature (40 ˚C), injection volume (2 µL), flow rate (0.35 mL min^{-1}).

(continuation)

Run	Variables				Responses			
	Organic Solvent	Gradient time (min)	pH	Column	Total number of peaks	Number of peaks with resolution ≥ 1.5	Number of peaks with resolution ≥ 2.0	Number of peaks with tailing ≤ 1.2
37	MeOH	30.0	2.54	Cortecs C18+	53	40	31	36
38	MeOH	15.0	2.54	Cortecs C18+	56	30	17	37
39	MeOH	22.5	2.54	Cortecs C8	57	40	24	45
40	MeOH	30.0	2.54	Acquity HSS T3	55	37	25	45
41	MeOH	15.0	2.54	Cortecs C18+	58	34	21	36
42	MeOH	22.5	6.60	Cortecs C8	44	30	16	32
43	MeOH	22.5	6.60	Acquity HSS T3	47	24	18	33
44	MeOH	22.5	6.60	Acquity BEH Shield RP18	37	24	16	26
45	MeOH	22.5	6.60	Acquity HSS PFP	39	24	19	30
46	MeOH	22.5	6.60	Cortecs C18+	43	28	22	33
47	MeOH	22.5	6.60	Acquity BEH C18	42	22	15	32
48	MeOH	22.5	6.60	Cortecs C8	40	27	18	30
49	MeOH	22.5	6.60	Acquity HSS T3	50	28	19	37
50	MeOH	22.5	6.60	Acquity BEH Shield RP18	32	22	16	21
51	MeOH	22.5	6.60	Acquity HSS PFP	35	20	16	25
52	MeOH	22.5	6.60	Cortecs C18+	39	28	19	29
53	MeOH	22.5	6.60	Acquity BEH C18	40	24	17	32
54	MeOH	30.0	10.45	Acquity BEH Shield RP18	40	29	21	32

Supplementary Table 1 – Factors levels, coded values and results of the experimental design for the screening step (quadratic model A- and G- optimal). Constant parameters: temperature (40 ˚C), injection volume (2 µL), flow rate (0.35 mL min^{-1}).

(conclusion)

Run	Variables				Responses			
	Organic Solvent	Gradient time (min)	pH	Column	Total number of peaks	Number of peaks with resolution ≥ 1.5	Number of peaks with resolution ≥ 2.0	Number of peaks with tailing ≤ 1.2
55	MeOH	15.0	10.45	Acquity BEH Shield RP18	43	27	19	29
56	MeOH	30.0	10.45	Acquity BEH C18	46	29	20	39
57	MeOH	15.0	10.45	Acquity BEH C18	50	26	19	42

Source: elaborated by the Author (2022).

Supplementary Table 2 – Factor levels, coded values and results of the experimental design for the first optimization step (central composite with face-centered axial point). Constant parameters: column (Acquity BEH Shield RP18 – Ethylene Bridged Hybrid C18, 1.7 µm, 100 x 2.1 mm), gradient time (30 min), injection volume (2 µL), flow rate (0.35 mL min^{-1}).

(to be continued)

Run	Variables				Responses			
	Final % organic solvent	CH_3CN % organic solvent	Oven temperature (˚C)	pH	Total number of peaks	Number of peaks with resolution ≥ 1.5	Number of peaks with resolution ≥ 2.0	Number of peaks with tailing ≤ 1.2
1	80	100	35	2.26	64	39	21	51
2	80	70	35	2.26	63	42	31	49
3	60	100	35	2.26	71	49	32	57
4	60	70	35	2.26	58	43	32	44
5	60	70	35	2.26	56	41	31	43
6	80	70	35	2.26	59	42	29	49
7	60	100	35	2.26	74	46	28	61
8	70	85	35	2.57	67	46	29	55
9	60	70	35	3.07	51	38	30	42
10	80	70	35	3.07	60	38	30	46
11	60	100	35	3.07	64	44	30	48
12	80	100	35	3.07	68	40	26	48
13	70	85	40	2.26	68	41	29	50
14	60	85	40	2.57	70	40	28	58
15	80	85	40	2.57	66	41	26	50
16	70	70	40	2.57	60	44	29	48
17	70	100	40	2.57	66	38	27	53

Supplementary Table 2 – Factor levels, coded values and results of the experimental design for the first optimization step (central composite with face-centered axial point). Constant parameters: column (Acquity BEH Shield RP18 – Ethylene Bridged Hybrid C18, 1.7 µm, 100 x 2.1 mm), gradient time (30 min), injection volume (2 µL), flow rate (0.35 mL min^{-1}).

(conclusion)

Run	Variables				Responses			
	Final % organic solvent	CH_3CN % organic solvent	Oven temperature (˚C)	pH	Total number of peaks	Number of peaks with resolution ≥ 1.5	Number of peaks with resolution ≥ 2.0	Number of peaks with tailing ≤ 1.2
18	70	85	40	2.57	68	44	27	59
19	70	85	40	2.57	67	39	30	56
20	70	85	40	2.57	67	39	29	56
21	70	85	40	3.07	63	41	27	51
22	60	70	45	2.26	63	43	33	48
23	80	70	45	2.26	71	42	30	57
24	60	100	45	2.26	71	44	27	57
25	80	100	45	2.26	74	40	26	51
26	70	85	45	2.57	66	46	29	56
27	60	70	45	3.07	53	37	31	45
28	80	70	45	3.07	55	38	27	44
29	60	100	45	3.07	61	37	23	53
30	80	100	45	3.07	63	33	21	48

Source: elaborated by the Author (2022).

Supplementary Table 3 – Factor levels, coded values and results of the experimental design for the second optimization step (central composite with face-centered axial point). Constant parameters: column (Acquity BEH Shield RP18 – Ethylene Bridged Hybrid C18, 1.7 µm, 100 x 2.1 mm), gradient time (30 min), pH (2.26), temperature (45 ˚C), injection volume (2 µL), flow rate (0.35 mL min^{-1}).

Run	Variables		Responses			
	Final % organic solvent	CH_3CN % organic solvent	Total number of peaks	Number of peaks with resolution ≥ 1.5	Number of peaks with resolution ≥ 2.0	Number of peaks with tailing ≤ 1.2
1	55	100	92	51	25	69
2	55	80	91	51	31	76
3	40	80	78	54	32	60
4	70	100	90	51	28	69
5	70	80	86	53	32	66
6	40	60	65	50	39	53
7	55	60	80	58	40	69
8	55	80	90	48	30	72
9	55	80	89	49	31	71
10	70	60	84	57	36	67
11	40	100	89	58	34	65

Source: elaborated by the Author (2022).

Supplementary Table 4 – Regression ANOVA statistics obtained for the **screening step** of the chromatographic method development for the responses selected.

Response (Screening)	Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Ratio	P-Value
Total Number of Peaks	*Regression*	3,504.9995	10	350.4999	41.8169	<0.0001
R^2: 0.9009	*Residual*	385.5619	46	8.3818		
*MSR: 0.9009; $MSR_{threshold}$: 0.0440	*Lack-of-fit*	307.0619	31	9.9052	1.8927	0.0956
**MS-LOF: 0.0255; $MS\text{-}LOF_{threshold}$: 0.0301	*Pure error*	78.5000	15	5.2333		
	Total	3,890.5614	56			
Number of peaks with resolution ≥ 1.5	*Regression*	1,281.7256	8	160.2157	21.4063	<0.0001
R^2: 0.7811	*Residual*	359.2568	48	7.4845		
*MSR: 0.7811; $MSR_{threshold}$: 0.0780	*Lack-of-fit*	297.2568	33	9.0078	2.1793	0.0550
**MS-LOF: 0.0439; $MS\text{-}LOF_{threshold}$: 0.0450	*Pure error*	62.000	15	4.1333		
	Total	1,640.9825	56			
Number of peaks with resolution ≥ 2.0	*Regression*	693.8530	14	49.5609	19.5713	<0.0001
R^2: 0.8671	*Residual*	106.3575	42	2.5323		
*MSR: 0.8671; $MSR_{threshold}$: 0.0857	*Lack-of-fit*	73.3575	27	2.7169	1.2350	0.3412
**MS-LOF: 0.0475; $MS\text{-}LOF_{threshold}$: 0.0872	*Pure error*	33.0000	15	2.2000		
	Total	800.2105	56			
Number of peaks with tailing ≤ 1.2	*Regression*	2,462.2860	10	246.2286	20.1102	<0.0001
R^2: 0.8138	*Residual*	563.2227	46	12.2440		
*MSR: 0.8138; $MSR_{threshold}$: 0.0827	*Lack-of-fit*	480.7227	31	15.5072	2.8195	0.0184
**MS-LOF: 0.0513; $MS\text{-}LOF_{threshold}$: 0.0407$^{\phi}$	*Pure error*	82.5000	15	5.5000		
	Total	3,025.5088	56			

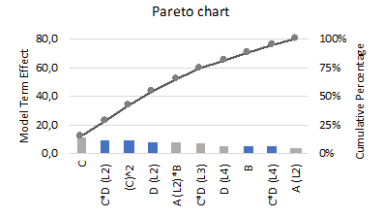
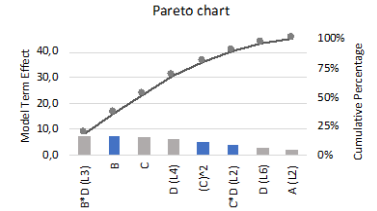
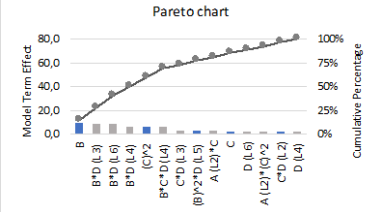
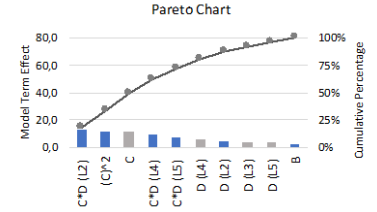
*Modeling goal: MSR ≥ $MSR_{threshold}$.

**Modeling goal: MS-LOF ≤ $MS\text{-}LOF_{threshold}$.

$^{\phi}$LOF is statistically significant (P-value < 0.0500).

Source: elaborated by the Author (2022).

Supplementary Table 5 – Equations and Pareto chart obtained for the responses selected for the **screening step** during the chromatographic method development.

Response (Screening)	Equation*	Model Term Ranking Pareto Chart**
Total Number of Peaks	$Y = 44.818 - 4.125 * A(L2) + 2.421 * B - 5.842 * C - 8.950 * D(L2) - 5.418 * D(L4) + 9.843 * C^2 - 4.121 * A(L2) * B + 9.635 * C * D(L3) - 3.615 * C * D(L3) + 4.667 * C * D(L4)$	
Number of peaks with resolution ≥ 1.5	$Y = 28.696 - 1.793 * A(L2) + 3.653 * B - 3.353 * C - 6.516 * D(L4) - 2.378 * D(L6) + 5.001 * C^2 - 3.845 * B * D(L3) + 3.703 * C * D(L2)$	
Number of peaks with resolution ≥ 2.0	$Y = 18.201 + 4.759 * B + 1.250 * C - 1.992 * D(L4) - 2.474 * D(L6) + 5.884 * C^2 - 1.340 * A(L2) * C - 4.277 * B * D(L3) - 3.181 * B * D(L4) - 4.009 * B * D(L6) + 2.196 * C * D(L2) - 1.397 * C * D(L3) + 2.737 * B^2 * D(L5) - 2.304 * A(L2) * C^2 - 2.922 * B * C * D(L4)$	
Number of peaks with tailing ≤ 1.2	$Y = 31.097 + 1.276 * B - 5.663 * C + 5.221 * D(L2) - 3.342 * D(L3) - 5.858 * D(L4) - 3.148 * D(L5) + 11.812 * C^2 + 12.204 * C * D(L2) + 8.718 * C * D(L4) + 7.266 * C * D(L5)$	

*__A__: strong solvent type; A (L2): methanol; **B**: gradient time; **C**: pH; **D**: column type; D (L2): Acquity HSS T3; D (L3): Acquity BEH Shield RP18; D (L4): Acquity HSS PFP; D (L5): Cortecs C_{18}^+; D (L6): Acquity BEH C_{18}. **__Blue:__ positive effects; **Grey:** negative effects.

Source: elaborated by the Author (2022).

Supplementary Table 6 – Regression ANOVA statistics obtained for the **first optimization step** of the chromatographic method development for the responses selected.

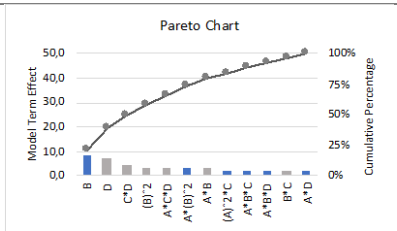
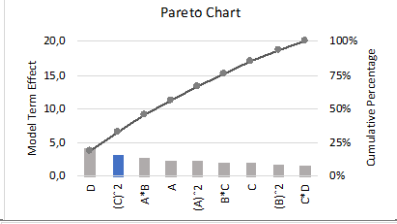
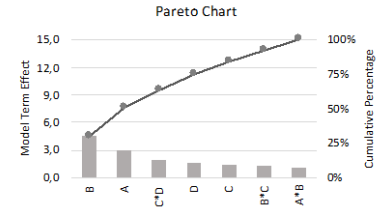
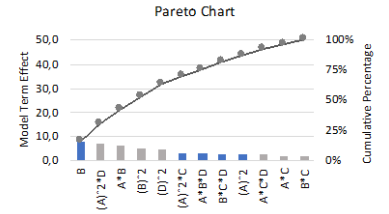
Response (1st optimization)	Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Ratio	P-Value
Total Number of Peaks	*Regression*	968.3290	12	80.6941	33.4278	<0.0001
R^2: 0.9593	*Residual*	41.0377	17	2.4140		
*MSR: 0.9593; $MSR_{threshold}$: 0.0683	*Lack-of-fit*	25.8710	12	2.1559	0.7107	0.7102
**MS-LOF: 0.0256; $MS\text{-}LOF_{threshold}$: 0.1687	*Pure error*	15.1667	5	3.0333		
	Total	1,009.3667	29			
Number of peaks with resolution ≥ 1.5	*Regression*	247.4675	9	27.4964	6.4927	0.0003
R^2: 0.7450	*Residual*	84.6991	20	4.2350		
*MSR: 0.7450; $MSR_{threshold}$: 0.2746	*Lack-of-fit*	61.5325	15	4.1022	0.8854	0.6130
**MS-LOF: 0.1111; $MS\text{-}LOF_{threshold}$: 0.5798	*Pure error*	23.1667	5	4.6333		
	Total	332.1667	29			
Number of peaks with resolution ≥ 2.0	*Regression*	189.6903	7	27.0986	10.2476	<0.0001
R^2: 0.7653	*Residual*	58.1764	22	2.6444		
*MSR: 0.7653; $MSR_{threshold}$: 0.1840	*Lack-of-fit*	43.0097	17	2.5300	0.8341	0.6488
**MS-LOF: 0.0714; $MS\text{-}LOF_{threshold}$: 0.3932	*Pure error*	15.1667	5	3.0333		
	Total	247.8667	29			
Number of peaks with tailing ≤ 1.2	*Regression*	0.0000	12	0.0000	54.1882	<0.0001
R^2: 0.9760	*Residual*	<0.0001	16	<0.0001		
*MSR: 0.9760; $MSR_{threshold}$: 0.0437	*Lack-of-fit*	0.0000	11	0.0000	0.7459	0.6826
**MS-LOF: 0.0163; $MS\text{-}LOF_{threshold}$: 0.1027	*Pure error*	0.0000	5	0.0000		
	Total	0.0000	28			

*Modeling goal: MSR ≥ $MSR_{threshold}$.

**Modeling goal: MS-LOF ≤ $MS\text{-}LOF_{threshold}$.

Source: elaborated by the Author (2022).

Supplementary Table 7 – Equations and Pareto chart obtained for the responses selected for the **first optimization step** during the chromatographic method development.

Response (1st optimization)	Equation*	Model Term Ranking Pareto Chart**
Total Number of Peaks	$Y = 66.239 + 4.020 * B - 3.564 * D - 3.266 * B^2 - 1.409 * A * B + 0.741 * A * D - 0.773 * B * C - 2.147 * C * D + 0.801 * A^2 * C + 1.438 * A * B^2 + 0.784 * A * B * C + 0.784 * A * B * D - 1.562 * A * C * D$	
Number of peaks with resolution ≥ 1.5	$Y = 41.301 - 1.164 * A - 1.028 * C - 2.041 * D - 2.265 * A^2 - 1.765 * B^2 + 3.235 * C^2 - 1.331 * A * B - 1.069 * B * C = 0.766 * C * D$	
Number of peaks with resolution ≥ 2.0	$Y = 28.031 - 1.506 * A - 2.251 * B - 0.722 * C - 0.808 * B - 0.557 * A * B - 0.593 * B * C - 0.981 * C * D$	
Number of peaks with tailing ≤ 1.2	$Y = -4.3E^{-5} * B - 2.8E^{-5} * A^2 + 6.7E^{-5} * B^2 + 6.8E^{-5} * D^2 + 3.5E^{-5} * A * B + 1.1E^{-5} * A * C + 1.0E^{-5} * B * C - 1.9E^{-5} * A^2 * C + 4E^{-5} * A^2 * D - 1.7E^{-5} * A * B * D + 1.3E^{-5} * A * C * D - 1.6E^{-5} * B * C * D$	

*__A__: final percentage of organic solvent; __B__: CH_3CN/MeOH ratio; __C__: oven temperature; __D__: pH.

__Blue:__ positive effects; **__Grey:__** negative effects.

Source: elaborated by the Author (2022).

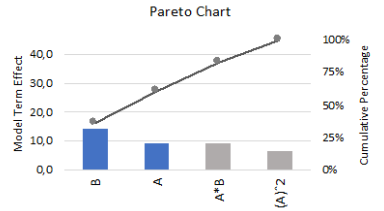
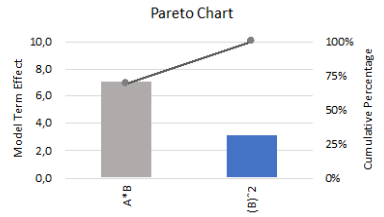
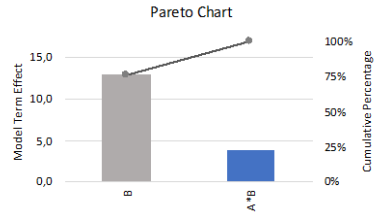
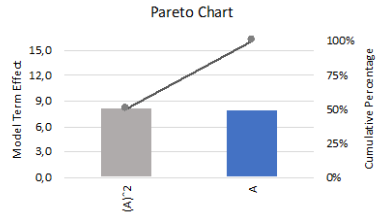
Supplementary Table 8 – Regression ANOVA statistics obtained for the **second optimization step** of the chromatographic method development for the responses selected.

Response (2nd optimization)	Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F-Ratio	P-Value
Total Number of Peaks	*Regression*	617.3758	4	154.3439	36.2688	0.0002
	Residual	25.5333	6	4.2556		
R^2: 0.9603	*Lack-of-fit*	23.5333	4	5.8833	5.8833	0.1505
*MSR: 0.9603; $MSR_{threshold}$: 0.1200	*Pure error*	2.0000	2	1.0000		
	Total	642.9091	10			
Number of peaks with resolution ≥ 1.5	*Regression*	76.3485	2	38.1742	5.8918	0.0267
	Residual	51.8333	8	6.4792		
R^2: 0.5956	*Lack-of-fit*	47.1667	6	7.8611	3.3690	0.2465
*MSR: 0.5956; $MSR_{threshold}$: 0.4508	*Pure error*	4.6667	2	2.333		
	Total	128.1818	10			
Number of peaks with resolution ≥ 2.0	*Regression*	189.1769	2	94.5885	71.7894	<0.0001
	Residual	9.2231	7	1.3176		
R^2: 0.9535	*Lack-of-fit*	8.5564	5	1.7113	5.1338	0.1710
*MSR: 0.9535; $MSR_{threshold}$: 0.0629	*Pure error*	0.6667	2	0.333		
	Total	198.4000	9			
Number of peaks with tailing ≤ 1.2	*Regression*	273.4667	2	136.7333	9.8963	0.0069
	Residual	110.5333	8	13.8167		
R^2: 0.7122	*Lack-of-fit*	96.5333	6	16.0889	2.2984	0.3339
*MSR: 0.7122; $MSR_{threshold}$: 0.3209	*Pure error*	14.0000	2	7.0000		
	Total	384.0000	10			

*Modeling goal: MSR ≥ $MSR_{threshold}$.

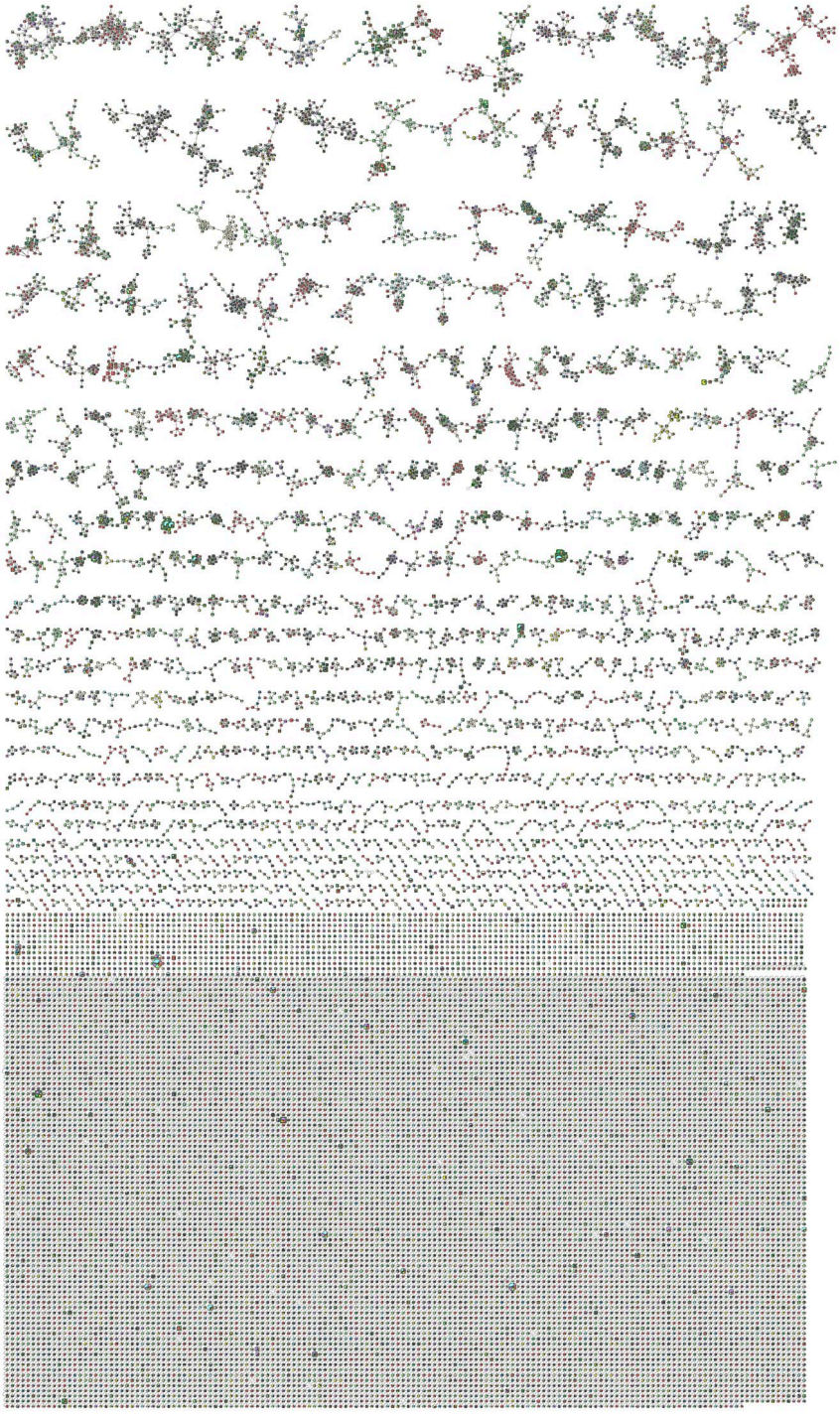
Source: elaborated by the Author (2022).

Supplementary Table 9 – Equations and Pareto chart obtained for the responses selected for the **second optimization step** in the chromatographic method development.

Response (2nd optimization)	Equation*	Model Term Ranking Pareto Chart**
Total Number of Peaks	$Y = 88.400 + 4.667 * A + 7.000 * B - 6.400 * A^2 - 4.5 * A * B$	
Number of peaks with resolution ≥ 1.5	$Y = 51 + 3.167 * B^2 - 3.500 * A * B$	
Number of peaks with resolution ≥ 2.0	$Y = 31.554 - 6.485 * B + 1.977 * A * B$	
Number of peaks with tailing ≤ 1.2	$Y = 71.400 + 4.000 * A - 8.067 * A^2$	

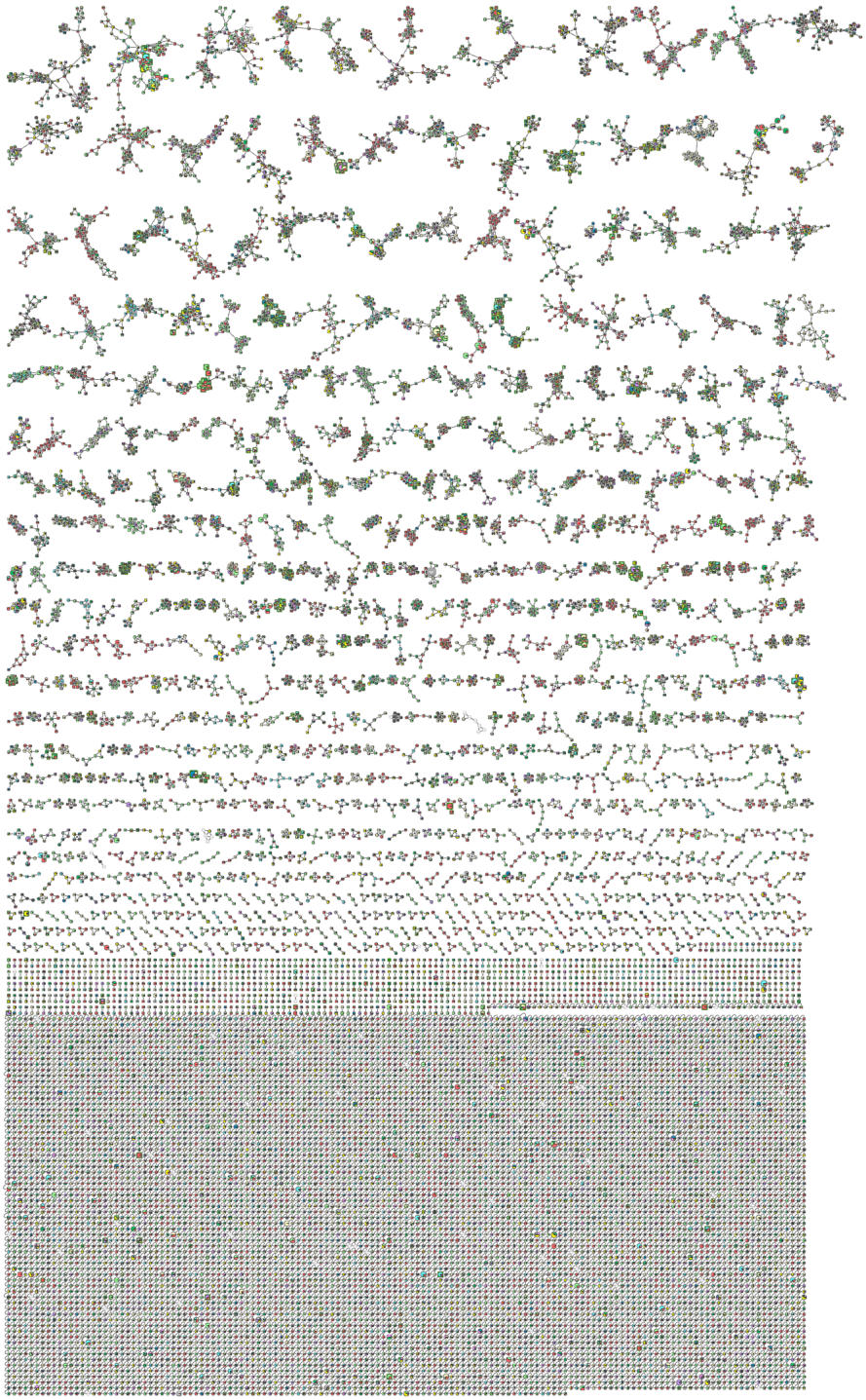
*__A__: final percentage of organic solvent; __B__: CH_3CN/MeOH ratio.
**__Blue:__ positive effects; __Grey:__ negative effects.
Source: elaborated by the Author (2022).

SUPPLEMENTARY MATERIAL B – CHAPTER 2



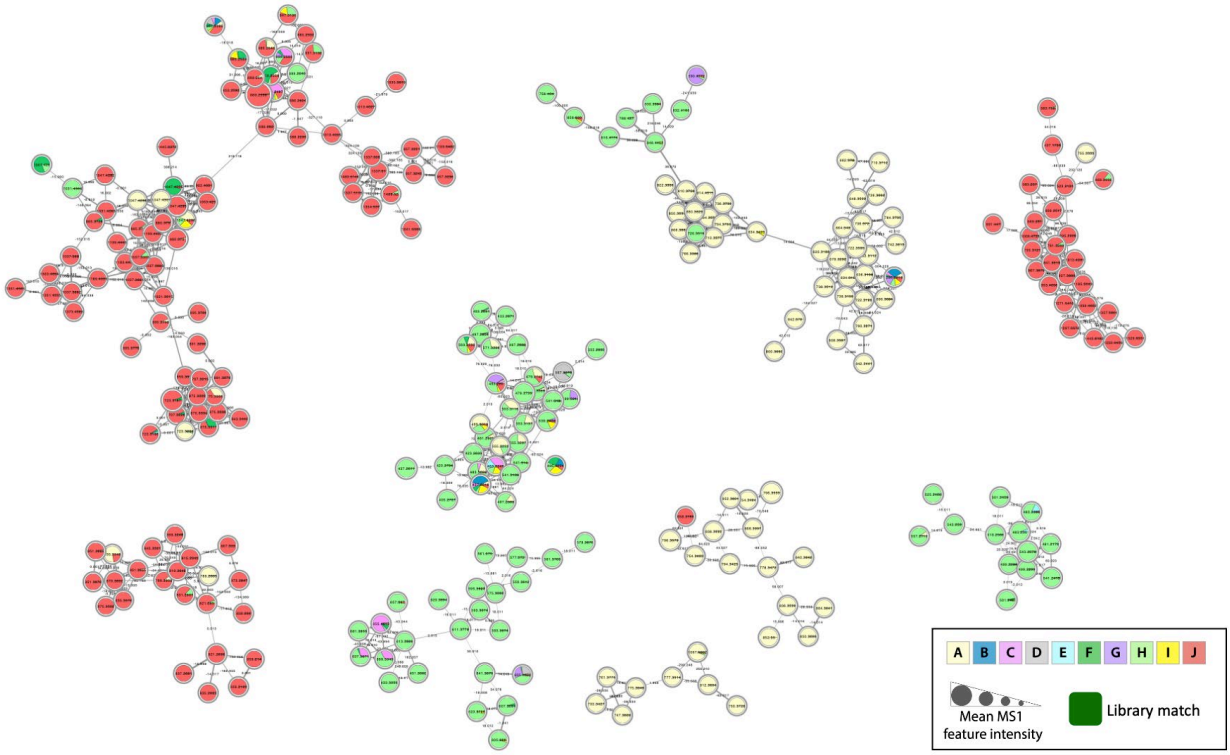
Supplementary Figure 3 – Molecular network obtained for Malpighiaceae samples in the positive ionization mode (EtOH80% extracts).

Source: elaborated by the Author (2022).



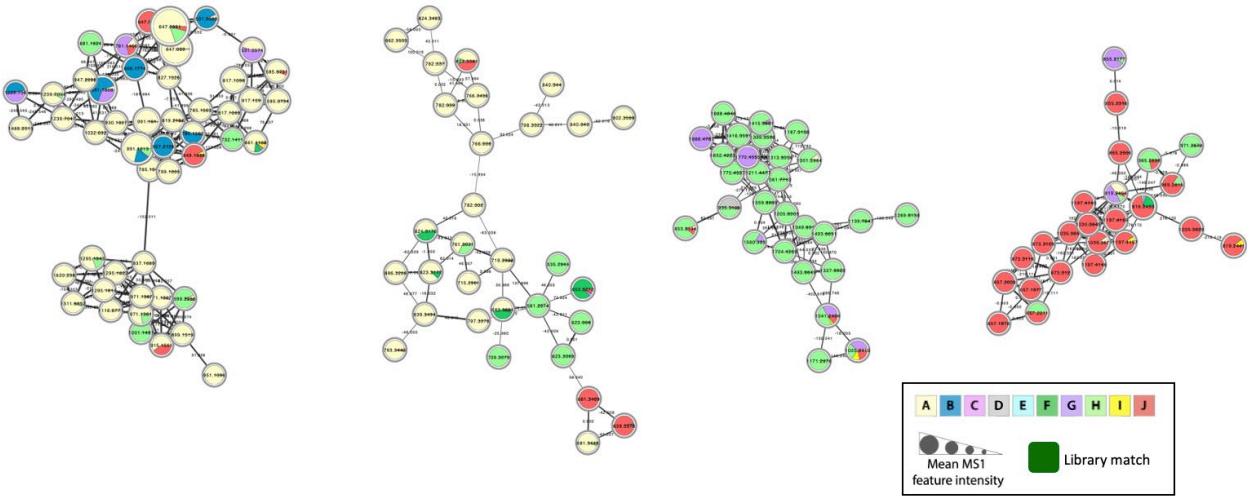
Supplementary Figure 4 – Molecular network obtained for Malpighiaceae samples in the negative ionization mode (EtOH80% extracts).

Source: elaborated by the Author (2022).



Supplementary Figure 5 – Molecular networks specific to Malpighiaceae clades without library matches (positive ionization mode).

Source: elaborated by the Author (2022).



Supplementary Figure 6 – Molecular networks specific to Malpighiaceae clades without library matches (negative ionization mode).

Source: elaborated by the Author (2022).

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(to be continued)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
1	*Acmanthera*	*latifolia*	Amorim 8575 (CEPEC)	Barcelos/AM; 15/04/2014	1	Amazon rainforest	Byrsonimoid (A)
2	*Acmanthera*	*latifolia*	Amorim 8580 (CEPEC)	Barcelos/AM; 15/04/2014	1	Amazon rainforest	Byrsonimoid (A)
3	*Acmanthera*	*latifolia*	Amorim 8649 (CEPEC)	Barcelos/AM; 25/04/2014	1	Amazon rainforest	Byrsonimoid (A)
4	*Acridocarpus*	*smeathmannii*	Letouzey 11760 (US)	Djoum/Camarões; 01/01/1973	32	African rainforest	Acridocarpoid (B)
5	*Alicia*	*anisopetala*	Rossetto 430 (HUEFS)	Londrina/PR; 15/07/2014	29	Atlantic forest	Tetrapteroid (H)
6	*Alicia*	*anisopetala*	Pace 200 (SPF)	Corumbá/MS; 26/06/2012	20	Pantanal	Tetrapteroid (H)
7	*Alicia*	*anisopetala*	Almeida 890 (RB)	Bauru/SP; 15/07/2018	28	Atlantic forest/Cerrado	Tetrapteroid (H)
8	*Amorimia*	*andersonii*	Marinho 654 (CEPEC)	Itamaraju/BA; 16/02/2014	21	Atlantic forest	Malpighioid (I)
9	*Amorimia*	*concinna*	Romero Castaneda 9252 (US)	Sincelejo/Colombia; 20/04/1963	3	Amazon rainforest	Malpighioid (I)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
10	*Amorimia*	*coriacea*	Almeida 615 (HUEFS)	Cabo Frio/RJ; 14/01/2016	24	Atlantic forest	Malpighioid (I)
11	*Amorimia*	*pubiflora*	Moleta S/N (HUEFS)	Cuiabá/MT; 01/06/2011	16	Cerrado	Malpighioid (I)
12	*Amorimia*	*septentrionalis*	Costa Lima 1163 (HUEFS)	Cachoeira dos índios/PB; 12/05/2014	8	Caatinga	Malpighioid (I)
13	*Amorimia*	*septentrionalis*	Almeida 800 (HUEFS)	Maruim/SE; 11/11/2015	9	Atlantic forest	Malpighioid (I)
14	*Aspicarpa*	*harleyi*	Almeida 874 (RB)	Caetité/BA; 02/11/2017	12	Caatinga/Cerrado	Stigmaphylloid (J)
15	*Banisteriopsis*	*adenopoda*	Almeida 813 (HUEFS)	São Paulo/SP; 10/04/2012	25	Atlantic forest	Stigmaphylloid (J)
16	*Banisteriopsis*	*adenopoda*	Almeida 813 (HUEFS)	São Paulo/SP; 10/04/2012	25	Atlantic forest	Stigmaphylloid (J)
17	*Banisteriopsis*	*argyrophylla*	Almeida 808 (HUEFS)	Mogi Guaçu/SP; 12/08/2012	26	Atlantic forest/Cerrado	Stigmaphylloid (J)
18	*Banisteriopsis*	*argyrophylla*	Almeida 811 (HUEFS)	Mogi Guaçu/SP; 12/08/2012	26	Atlantic forestCerrado	Stigmaphylloid (J)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
19	*Banisteriopsis*	*anisandra*	Francener 1138 (SP)	Pirenópolis/GO; 08/08/2012	14	Cerrado	Stigmaphylloid (J)
20	*Banisteriopsis*	*anisandra*	Almeida 611 (HUEFS)	Rio de Contas/BA; 30/07/13	12	Cerrado/Caatinga	Stigmaphylloid (J)
21	*Banisteriopsis*	*campestris*	Machado 20 (BHCB)	Serra do cipó/MG; 09/05/2018	23	Cerrado	Stigmaphylloid (J)
22	*Banisteriopsis*	*harleyi*	Almeida 635 (HUEFS)	Piatã/BA; 10/12/2013	10	Cerrado/Caatinga	Stigmaphylloid (J)
23	*Banisteriopsis*	*laevifolia*	Francener 1275 (SP)	Pedregulho/SP; 19/02/2013	27	Cerrado	Stigmaphylloid (J)
24	*Banisteriopsis*	*laevifolia*	Francener 1359 (SP)	Vila Bela de Santíssima Trindade/MT; 21/03/2013	6	Atlantic forest/Cerrado/ Pantanal	Stigmaphylloid (J)
25	*Banisteriopsis*	*laevifolia*	RBv 3814 (RB)	Rio de Janeiro Botanical Garden; 21/09/2018	24	Atlantic forest	Stigmaphylloid (J)
26	*Banisteriopsis*	*laevifolia*	Almeida 763 (HUEFS)	Aquidauana/MS; 26/01/2014	20	Cerrado	Stigmaphylloid (J)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
27	*Banisteriopsis*	*malifolia*	R1934 (IAC 54704)	Chapada do Araripe/CE; Jul/2013	7	Caatinga	Stigmaphylloid (J)
28	*Banisteriopsis*	*malifolia*	Francener 1344 (SP)	Vilhena/RO; 13/03/2013	5	Cerrado/Amazon rainforest	Stigmaphylloid (J)
29	*Banisteriopsis*	*malifolia*	Francener 1344 (SP)	Vilhena/RO; 13/03/2013	5	Cerrado/Amazon rainforest	Stigmaphylloid (J)
30	*Banisteriopsis*	*malifolia*	Oliveira 25 (BHCB)	Serra do cipó/MG; 09/05/2018	23	Cerrado	Stigmaphylloid (J)
31	*Banisteriopsis*	*megaphylla*	Francener 1131 (SP)	Brasília/DF; 07/08/2012	14	Cerrado	Stigmaphylloid (J)
32	*Banisteriopsis*	*membranifolia*	Pellegrini 494 (RB)	Petrópolis/RJ; 04/04/2016	24	Atlantic forest	Stigmaphylloid (J)
33	*Banisteriopsis*	*muricata*	Francener 1349 (SP)	Pimenta Bueno/RO; 20/03/2013	5	Amazon rainforest	Stigmaphylloid (J)
34	*Banisteriopsis*	*muricata*	Almeida 545 (HUEFS)	Foz do Iguaçu/PR; 08/06/2013	30	Atlantic forest	Stigmaphylloid (J)
35	*Banisteriopsis*	*parviglandula*	Costa-Lima 2180 (HUEFS)	Santa Teresa/ES; May/2015	22	Atlantic forest	Stigmaphylloid (J)
36	*Banisteriopsis*	*quadrangulata*	Almeida 581 (HUEFS)	Jacobina/BA; 13/08/2013	10	Caatinga	Stigmaphylloid (J)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
37	*Banisteriopsis*	*stellaris*	R1935 (IAC 54705)	Chapada do Araripe/CE; Jul/2013	7	Caatinga	Stigmaphylloid (J)
38	*Banisteriopsis*	*stellaris*	Francener 1123 (SP)	Nova Mutum/ MT; 02/05/2012	15	Cerrado	Stigmaphylloid (J)
39	*Banisteriopsis*	*stellaris*	RBv 3821 (RB)	Rio de Janeiro Botanical Garden; 21/09/2018	24	Atlantic forest	Stigmaphylloid (J)
40	*Banisteriopsis*	*stellaris*	Almeida 814 (HUEFS)	Mogi Guaçu/SP; 12/08/2012	26	Atlantic forest/Cerrado	Stigmaphylloid (J)
41	*Banisteriopsis*	*variabilis*	Almeida 664 (HUEFS)	Chapadão do Céu/GO; 30/01/2014	17	Cerrado	Stigmaphylloid (J)
42	*Banisteriopsis*	*variabilis*	Almeida 815 (HUEFS)	Mogi Guaçu/SP; 12/08/2012	26	Atlantic forest/Cerrado	Stigmaphylloid (J)
43	*Banisteriopsis*	*vernonifolia*	Almeida 660 (HUEFS)	Alto Paraíso/GO; 26/01/2014	13	Cerrado	Stigmaphylloid (J)
44	*Barnebya*	*harleyi*	Melo 1518 (HUEFS)	Itatim/BA; 30/10/2014	11	Caatinga	Barnebyoid (D)
45	*Blepharandra*	*heteropetala*	Amorim 8608 (CEPEC)	Barcelos/AM; 16/04/2014	1	Amazon rainforest	Byrsonimoid (A)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
46	*Blepharandra*	*hypoleuca*	Amorim 8618 (CEPEC)	Barcelos/AM; 20/04/2014	1	Amazon rainforest	Byrsonimoid (A)
47	*Bronwenia*	*megaptera*	Almeida 782 (HUEFS)	Itatim/BA; 20/10/2014	11	Caatinga	Stigmaphylloid (J)
48	*Bunchosia*	*glandulirifera*	RBv 3789 (RB)	Rio de Janeiro Botanical Garden; 21/09/2018	24	Atlantic forest	Bunchosioid (F)
49	*Bunchosia*	*montana*	Pace 418 (SPF)	Oaxaca/Mexico; 01/01/2013	32	Seasonally dry forest	Bunchosioid (F)
50	*Bunchosia*	*pallescens*	RBv 3808 (RB)	Rio de Janeiro Botanical Garden; 21/09/2018	24	Atlantic forest	Bunchosioid (F)
51	*Bunchosia*	*pallescens*	RBv 3808 (RB)	Rio de Janeiro Botanical Garden; 21/09/2018	24	Atlantic forest	Bunchosioid (F)
52	*Burdachia*	*duckei*	Amorim 8647 (CEPEC)	Barcelos/AM; 25/04/2014	1	Amazon rainforest	Mcvaughioid (C)
53	*Byrsonima*	*coccolobifolia*	M2 R1943 (IAC 55280)	Mogi Guaçu/SP; Jan/2014	26	Atlantic forest/Cerrado	Byrsonimoid (A)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
54	*Byrsonima*	*coccolobifolia*	M3 R1943 (IAC 55280)	Mogi Guaçu/SP; Jan/2014	26	Atlantic forest/Cerrado	Byrsonimoid (A)
55	*Byrsonima*	*coccolobifolia*	M6 R1944 (IAC 55278)	Mogi Guaçu/SP; Jan/2014	26	Atlantic forest/Cerrado	Byrsonimoid (A)
56	*Byrsonima*	*crassifolia*	Rapini 1553 (HUEFS)	Serra do cipó/MG; 09/05/2018	23	Cerrado	Byrsonimoid (A)
57	*Byrsonima*	*incarnata*	Almeida 901 (RB)	Ananindeua/PA; 18/04/2015	4	Amazon rainforest	Byrsonimoid (A)
58	*Byrsonima*	*intermedia*	M1 R1942 (IAC 55281)	Mogi Guaçu/SP; Jan/2014	26	Atlantic forest/Cerrado	Byrsonimoid (A)
59	*Byrsonima*	*intermedia*	M4 R1942 (IAC 55281)	Mogi Guaçu/SP; Jan/2014	26	Atlantic forest/Cerrado	Byrsonimoid (A)
60	*Byrsonima*	*ligustrifolia*	Almeida 805 (HUEFS)	Itanhaém/SP; 07/12/2015	25	Atlantic forest	Byrsonimoid (A)
61	*Byrsonima*	*sericea*	R1933 (IAC 54703)	Chapada do Araripe/CE; Jul/2013	7	Caatinga	Byrsonimoid (A)
62	*Byrsonima*	*sericea*	Trovó 656 (RB)	Rio de Janeiro/RJ; Abril/2015	24	Atlantic forest	Byrsonimoid (A)
63	*Byrsonima*	*spicata*	Almeida 902 (RB)	Ananindeua/PA; 18/04/2015	4	Amazon rainforest	Byrsonimoid (A)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
64	*Byrsonima*	*verbascifolia*	M7 (IAC 55279)	Mogi Guaçu/SP; Jan/2014	26	Atlantic forest/Cerrado	Byrsonimoid (A)
65	*Byrsonima*	*verbascifolia*§	M7 (IAC 55279)	Mogi Guaçu/SP; Jan/2014	26	Atlantic forest/Cerrado	Byrsonimoid (A)
66	*Callaeum*	*antifebrile*	Tamaio 110 (RB)	Rio de Janeiro Botanical Garden; 21/09/2018	24	Atlantic forest	Tetrapteroid (H)
67	*Callaeum*	*psilophyllum*	Almeida 734 (HUEFS)	Feira de Santana/BA; 10/08/2013	11	Caatinga	Tetrapteroid (H)
68	*Camarea*	*affinis X hirsuta*	Francener 1255 (SP)	Teresina de Goiás/GO; 10/01/2013	13	Cerrado	Stigmaphylloid (J)
69	*Camarea*	*ericoides*	Lovo 453 (SPF)	Serra do Cipó/MG; 2013	23	Cerrado	Stigmaphylloid (J)
70	*Carolus*	*chasei*	Almeida 585 (HUEFS)	Miguel Calmon/BA; 13/08/2013	10	Caatinga	Tetrapteroid (H)
71	*Christianella*	*multiglandulosa*	Francener 1173 (SP)	Corguinho/MS; 19/10/2012	18	Cerrado	Tetrapteroid (H)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
72	*Diacidia*	*aracaensis*	Amorim 8617 (CEPEC)	Barcelos/AM; 20/04/2014	1	Amazon rainforest	Byrsonimoid (A)
73	*Dicella*	*bracteosa*	Lombardi 1048 (BHCB)	Belo Horizonte/MG; 09/05/2018	23	Atlantic forest	Tetrapteroid (H)
74	*Dicella*	*macroptera*	Francener 1254 (SP)	Taipas do Tocantins/TO; 14/01/2013	13	Cerrado	Tetrapteroid (H)
75	*Diplopterys*	*pubipetala*	Francener 1126 (SP)	Nova Mutum/ MT; 02/05/2012	15	Cerrado	Stigmaphylloid (J)
76	*Diplopterys*	*pubipetala*	Francener 1334 (SP)	União do Sul/MT; 16/03/2013	16	Cerrado	Stigmaphylloid (J)
77	*Diplopterys*	*pubipetala*	Rando 1030 (BHCB)	Serra do cipó/MG; 09/05/2018	23	Cerrado	Stigmaphylloid (J)
78	*Ectopopterys*	*soejartoi*	Wurdack 2356 (US)	Loreto/Peru; 21/10/1962	2	Amazon rainforest	Malpighioid (I)
79	*Galphimia*	*brasiliensis*	Felix 372 (BHCB)	Belo Horizonte/MG; 09/05/2018	23	Atlantic forest	Byrsonimoid (A)
80	*Gaudichaudia*	*albida*	Pace 419 (SPF)	Oaxaca/México; 21/08/2013	32	Seasonally dry forest	Stigmaphylloid (J)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
81	*Gaudichaudia*	*albida*	Pace 423 (SPF)	Veracruz/México; 01/01/2013	31	Seasonally dry forest	Stigmaphylloid (J)
82	*Glicophyllum*	*ambiguum*	Francener 1165 (SP)	Araguainha/MT; 13/08/2012	14	Cerrado/Caatinga	Tetrapteroid (H)
83	*Glicophyllum*	*cardiophyllum*	Almeida 641 (HUEFS)	Rio de Contas/BA; 09/12/2013	12	Cerrado/Caatinga	Tetrapteroid (H)
84	*Glicophyllum*	*cardiophyllum*	Almeida 793 (HUEFS)	Rio de Contas/BA; 30/10/2014	12	Cerrado/Caatinga	Tetrapteroid (H)
85	*Glicophyllum*	*microphyllum*	Francener 1264 (SP)	Alto Paraíso/GO; 15/01/2013	13	Cerrado	Tetrapteroid (H)
86	*Glicophyllum*	*microphyllum*	Almeida 835b (HUEFS)	Serra do cipó/MG; 09/05/2018	23	Cerrado	Tetrapteroid (H)
87	*Glicophyllum*	*ramiflorum*	Francener 1170 (SP)	Corguinho/MS; 19/10/2012	18	Cerrado	Tetrapteroid (H)
88	*Heladena*	*multiflora*	Francener 1181 (SP)	Dourados/MS; 22/10/2012	19	Atlantic forest/Cerrado	Bunchosioid (F)
89	*Heteropterys*	*aenea*	Almeida 798 (HUEFS)	Itapema/SC; 25/11/2015	31	Atlantic forest	Tetrapteroid (H)
90	*Heteropterys*	*bicolor*	Barros 5309 (HUEFS)	Niterói/RJ; 11/11/2016	24	Atlantic forest	Tetrapteroid (H)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
91	*Heteropterys*	*brachiata*	Pace 406 (SPF)	Veracruz/Mexico; 01/01/2013	31	Seasonally dry forest	Tetrapteroid (H)
92	*Heteropterys*	*brunnea*	Almeida 684 (HUEFS)	Jequitinhonha/MG; 10/12/2014	12	Atlantic forest	Tetrapteroid (H)
93	*Heteropterys*	*byrsonimifolia*	Almeida 723 (HUEFS)	Seabra/BA; 15/11/2014	10	Caatinga	Tetrapteroid (H)
94	*Heteropterys*	*campestris*	Almeida 657 (HUEFS)	Alto Paraíso/GO; 26/01/2014	13	Cerrado	Tetrapteroid (H)
95	*Heteropterys*	*chrysophylla*	Trovó 658 (RB)	Rio de Janeiro/RJ; Apr/2015	24	Atlantic forest	Tetrapteroid (H)
96	*Heteropterys*	*coleoptera*	RBv 8909 (RB)	Rio de Janeiro Botanical Garden; 21/09/2018	24	Atlantic forest	Tetrapteroid (H)
97	*Heteropterys*	*coleoptera*	RBv 8909 (RB)	Rio de Janeiro Botanical Garden; 14/07/2018	24	Atlantic forest	Tetrapteroid (H)
98	*Heteropterys*	*eglandulosa*	Francener 1141 (SP)	Pirenópolis/GO; 08/08/2012	14	Cerrado	Tetrapteroid (H)
99	*Heteropterys*	*eglandulosa*	Almeida 726 (HUEFS)	Correntina/BA; 15/11/2014	12	Cerrado	Tetrapteroid (H)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
100	*Heteropterys*	*hatschbachii*	Francener 1184 (SP)	Rio Verde de Mato Grosso/MS; 24/10/2012	18	Cerrado	Tetrapteroid (H)
101	*Heteropterys*	*intermedia*	Pellegrini 491 (RB)	Petrópolis/RJ; 22/09/2018	24	Atlantic forest	Tetrapteroid (H)
102	*Heteropterys*	*intermedia*	Almeida 504 (SP)	Dores do Rio Preto/ES; 18/11/2011	24	Atlantic forest	Tetrapteroid (H)
103	*Heteropterys*	*intermedia*	Almeida 505 (SP)	Dores do Rio Preto/ES; 18/11/2011	24	Atlantic forest	Tetrapteroid (H)
104	*Heteropterys*	*leona*	Amorim 8651 (BHCB)	Barcelos/AM; 25/04/2014	1	Amazon rainforest	Tetrapteroid (H)
105	*Heteropterys*	*leona*	Amorim 8577 (BHCB)	Barcelos/AM; 15/04/2014	1	Amazon rainforest	Tetrapteroid (H)
106	*Heteropterys*	*leona*	Amorim 8597 (BHCB)	Barcelos/AM; 16/04/2014	1	Amazon rainforest	Tetrapteroid (H)
107	*Heteropterys*	*leschinaultiana*	Trovó 660 (RB)	Itatiaia/RJ; Sep/2015	24	Atlantic forest	Tetrapteroid (H)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
108	*Heteropterys*	*leschinaultiana*	Trovó 668 (RB)	Nova Friburgo/RJ; Sep/2015	24	Atlantic forest	Tetrapteroid (H)
109	*Heteropterys*	*oberdanii*	Almeida 521 (SP)	Sooretama/ES; 20/01/2012	22	Atlantic forest	Tetrapteroid (H)
110	*Heteropterys*	*pauciflora*	Almeida 804 (HUEFS)	Vacaria/RS; 04/12/2015	31	Atlantic forest	Tetrapteroid (H)
111	*Heteropterys*	*tomentosa*	Francener 1198 (SP)	Costa Rica/MS; 26/10/2012	17	Cerrado	Tetrapteroid (H)
112	*Heteropterys*	*umbellata*	Almeida 838b (RB)	Serra do cipó/MG; 09/05/2018	23	Cerrado	Tetrapteroid (H)
113	*Heteropterys*	*umbellata*	Almeida 838b (RB)	Serra do cipó/MG; 09/05/2018	23	Cerrado	Tetrapteroid (H)
114	*Hiptage*	*benghalensis*	RBv 3801 (RB)	Rio de Janeiro Botanical Garden; 21/09/2018	24	Atlantic forest	Tetrapteroid (H)
115	*Hiraea*	*cuiabensis*	Francener 1218 (SP)	Araguainha/MT; 29/12/2012	14	Atlantic forest	Hiraeoid (G)
116	*Hiraea*	*hatschbachii*	Almeida 548 (HUEFS)	Foz do Iguaçu/PR; 08/06/2013	30	Atlantic forest	Hiraeoid (G)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
117	*Hiraea*	*reclinata*	Pace 518 (SPF)	Loreto/Peru; 18/09/2014	2	Amazon rainforest	Hiraeoid (G)
118	*Hiraea*	*restingae*	Almeida 518 (SP)	Soretano/ES; 20/01/2012	22	Atlantic forest	Hiraeoid (G)
119	*Hiraea*	*restingae*	Almeida 542 (SP)	Guarapari/ES; 30/01/2012	22	Atlantic forest	Hiraeoid (G)
120	*Janusia*	*janusioides*	Francener 1117 (SP)	Nova Mutum/MT; 01/05/2004	15	Amazon rainforest/Cerrado	Stigmaphylloid (J)
121	*Janusia*	*mediterranea*	RBv 4767 (RB)	Rio de Janeiro Botanical Garden; 21/09/2018	24	Atlantic forest	Stigmaphylloid (J)
122	*Janusia*	*mediterranea*	Fontella 517 (RB)	Rio de Janeiro Botanical Garden, 11/08/2016	24	Atlantic forest	Stigmaphylloid (J)
123	*Janusia*	*occhionii*	Francener 1361 (SP)	Vila Bela de Santíssima Trindade/MT; 21/03/2013	6	Amazon rainforest/Cerrado/Pantanal	Stigmaphylloid (J)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
124	*Janusia*	*schwannioides*	Almeida 578 (HUEFS)	Boa Nova/BA; 01/07/2013	12	Atlantic forest	Stigmaphylloid (J)
125	*Malpighia*	*mexicana*	Pace 413 (SPF)	Oaxaca/Mexico; 21/08/2013	32	Seasonally dry forest	Malpighioid (I)
126	*Malpighia*	*mexicana*	Pace 417 (SPF)	Oaxaca/Mexico; 01/01/2013	32	Seasonally dry forest	Malpighioid (I)
127	*Mascagnia*	*conformis*	Amorim 8660 (CEPEC)	Barcelos/AM; 25/04/2014	1	Amazon rainforest	Malpighioid (I)
128	*Mascagnia*	*sepium*	Almeida 822 (HUEFS)	Marilândia/ES; 10/10/2015	22	Atlantic forest	Malpighioid (I)
129	*Mcvaughia*	*bahiana*	Guedes 12148 (ALCB)	Monte Santo/BA; 12/01/2006	10	Atlantic forest/Caatinga	Mcvaughioid (C)
130	*Niedenzuella*	*acutifolia*	Barros 5136 (RB)	Cachoeiras de Macacu/RJ; 21/09/2018	24	Atlantic forest	Tetrapteroid (H)
131	*Niedenzuella*	*acutifolia*§	Barros 5136 (RB)	Cachoeiras de Macacu/RJ; 21/09/2018	24	Atlantic forest	Tetrapteroid (H)
132	*Niedenzuella*	*acutifolia*	Almeida 519 (SP)	Sooretama/ES; 20/01/2012	22	Atlantic forest	Tetrapteroid (H)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
133	*Niedenzuella*	*lasiandra*	Francener 1182 (SP)	Rio verde de Mato Grosso/MS; 24/10/2012	18	Cerrado	Tetrapteroid (H)
134	*Niedenzuella*	*lucida*	Almeida 626 (HUEFS)	Barro Preto/BA; 09/12/2013	21	Atlantic forest	Tetrapteroid (H)
135	*Niedenzuella*	*multiglandulosa*	Francener 1199 (SP)	Costa Rica, MS; 26/10/2012	17	Cerrado	Tetrapteroid (H)
136	*Niedenzuella*	*multiglandulosa*	HMS 5206 (CGMS)	Campo Grande, MS; 22/11/2015	18	Cerrado	Tetrapteroid (H)
137	*Niedenzuella*	*multiglandulosa*	Almeida 523 (SP)	Sooretama/ES; 21/01/2012	22	Atlantic forest	Tetrapteroid (H)
138	*Niedenzuella*	*multiglandulosa*	Almeida 639 (HUEFS)	Rio de Contas/BA; 09/12/2013	12	Cerrado/Caatinga	Tetrapteroid (H)
139	*Niedenzuella*	*poeppigiana*	Pellegrini 490 (RB)	Petrópolis/RJ; 21/02/2016	24	Atlantic forest	Tetrapteroid (H)
140	*Niedenzuella*	*poeppigiana*	Pellegrini 490 (RB)	Petrópolis/RJ; 20/09/2018	24	Atlantic forest	Tetrapteroid (H)
141	*Niedenzuella*	*poeppigiana*§	Pellegrini 490 (RB)	Petrópolis/RJ; 20/09/2018	24	Atlantic forest	Tetrapteroid (H)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
142	*Niedenzuella*	*poeppigiana*	Almeida 508 (SP)	Santa Teresa/ES; 22/11/2011	22	Atlantic forest	Tetrapteroid (H)
143	*Niedenzuella*	*sericea*	Almeida 823 (HUEFS)	Rio de Contas/BA; 09/12/2013	12	Cerrado/Caatinga	Tetrapteroid (H)
144	*Niedenzuella*	*stannea*	Francener 1233 (SP)	Lagoa da Confusão/TO; 12/01/2013	13	Cerrado	Tetrapteroid (H)
145	*Peixotoa*	*cordistipula*	Almeida 666 (HUEFS)	Campo Grande/MS; 30/01/2014	18	Cerrado	Stigmaphylloid (J)
146	*Peixotoa*	*glabra*	Almeida 843 (RB)	Milho Verde/MG; 01/11/2017	23	Cerrado	Stigmaphylloid (J)
147	*Peixotoa*	*hispidula*	Almeida 510 (SP)	Guarapari/ES; 23/11/2011	22	Atlantic forest	Stigmaphylloid (J)
148	*Peixotoa*	*reticulata*	Almeida 812 (HUEFS)	Mogi Guaçu/SP; 12/08/2012	26	Atlantic forest/Cerrado	Stigmaphylloid (J)
149	*Peixotoa*	*spinensis*	Almeida 609 (HUEFS)	Mucugê/BA; 29/07/2013	10	Cerrado/Caatinga	Stigmaphylloid (J)
150	*Peixotoa*	*tomentosa*	Almeida 833 (HUEFS)	Ouro Branco/MG; 30/10/2017	23	Atlantic forest	Stigmaphylloid (J)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
151	*Peixotoa*	*tomentosa*	Sebastiani 299 (BHCB)	Serra do cipó/MG; 09/05/2018	23	Cerrado	Stigmaphylloid (J)
152	*Ptilochaeta*	*bahiensis*	Almeida 860 (RB)	Porteirinha/MG; 05/11/2017	12	Cerrado/Caatinga	Ptilochaetoid (E)
153	*Ptilochaeta*	*densiflora*	Carvalho 290 (HUEFS)	Corumbá/MS; 07/04/2010	20	Pantanal	Ptilochaetoid (E)
154	*Sphedamnocarpus*	*angolensis*	Humbert 15422 (US)	Matopo/Zimbabwe; 15/04/1934	33	African savanna	Stigmaphylloid (J)
155	*Stigmaphyllon*	*acuminatum*	Almeida 821 (HUEFS)	Mimoso do Sul/ES; 20/10/2015	24	Atlantic forest	Stigmaphylloid (J)
156	*Stigmaphyllon*	*alternifolium*	Almeida 511 (SP)	Divino de São Lourenço/ES; 17/11/2011	24	Atlantic forest	Stigmaphylloid (J)
157	*Stigmaphyllon*	*angustilobum*	Almeida 503 (SP)	Divino de São Lourenço/ES; 17/11/2011	24	Atlantic forest	Stigmaphylloid (J)
158	*Stigmaphyllon*	*auriculatum*	Almeida 584 (HUEFS)	Miguel Calmon/BA; 13/08/2013	10	Caatinga	Stigmaphylloid (J)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
159	*Stigmaphyllon*	*blanchetii*	Almeida 532 (SP)	Conceição da Barra/ES; 23/01/2012	22	Atlantic forest	Stigmaphylloid (J)
160	*Stigmaphyllon*	*bonariense*	Queiroz 13530 (HUEFS)	Corrientes/Argentina; Dec/2007	30	Atlantic forest	Stigmaphylloid (J)
161	*Stigmaphyllon*	*caatingicola*	Almeida 577 (HUEFS)	Rio de Contas/BA; 23/07/13	12	Cerrado/Caatinga	Stigmaphylloid (J)
162	*Stigmaphyllon*	*cavernulosum*	Cardoso 2083 (HUEFS)	Rio de Contas/BA; Data	12	Cerrado/Caatinga	Stigmaphylloid (J)
163	*Stigmaphyllon*	*cavernulosum*	Gomes 1149 (HUEFS)	Salvador/BA; 04/06/2012	11	Atlantic forest	Stigmaphylloid (J)
164	*Stigmaphyllon*	*ciliatum*	Almeida 541 (SP)	Guarapari/ES; 29/01/2012	22	Atlantic forest	Stigmaphylloid (J)
165	*Stigmaphyllon*	*ciliatum*	Almeida 795 (HUEFS)	Itapema/SC; 25/11/2015	31	Atlantic forest	Stigmaphylloid (J)
166	*Stigmaphyllon*	*convolvulifolium*	Francener 1330 (SP)	Carlinda/MT; 14/03/2013	15	Amazon rainforest	Stigmaphylloid (J)
167	*Stigmaphyllon*	*gayanum*	Almeida 500 (SP)	Divino de São Lourenço/ES; 17/11/2011	24	Atlantic forest	Stigmaphylloid (J)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
168	*Stigmaphyllon*	*harleyi*	Santos 378 (HUEFS)	Morro do Chapéu/BA; 22/07/2005	10	Caatinga	Stigmaphylloid (J)
169	*Stigmaphyllon*	*hatschbachii*	Almeida 903 (RB)	Wenceslau Guimarães/BA; 2015	11	Atlantic forest	Stigmaphylloid (J)
170	*Stigmaphyllon*	*hatschbachii*	Sobrinho 10 (HUEFS)	Santa Teresinha/BA; 27/07/2000	11	Atlantic forest/Caatinga	Stigmaphylloid (J)
171	*Stigmaphyllon*	*hispidum*	Marinho 854 (HUEFS)	Wenceslau Guimarães/BA; 2015	11	Atlantic forest	Stigmaphylloid (J)
172	*Stigmaphyllon*	*lalandianum*	Almeida 816 (HUEFS)	Mogi Guaçu/SP; 12/08/2012	26	Atlantic forest/Cerrado	Stigmaphylloid (J)
173	*Stigmaphyllon*	*lalandianum*	Almeida 840 (HUEFS)	Serra do cipó/MG; 09/05/2018	23	Cerrado	Stigmaphylloid (J)
174	*Stigmaphyllon*	*macropodum*	Almeida 539 (SP)	Una/BA; 25/01/2012	21	Atlantic forest	Stigmaphylloid (J)
175	*Stigmaphyllon*	*occidentale*	Melo 7226 (HUEFS)	Mateiros/TO; 11/11/2009	13	Cerrado	Stigmaphylloid (J)
176	*Stigmaphyllon*	*palmatum*	Almeida 904 (RB)	Belém do Pará/PA; 15/04/2015	4	Amazon rainforest	Stigmaphylloid (J)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
177	*Stigmaphyllon*	*paralias*	R1928 (IAC 54701)	Chapada do Araripe/CE; Jul/2013	7	Caatinga	Stigmaphylloid (J)
178	*Stigmaphyllon*	*paralias*	Almeida 820 (HUEFS)	Santa Teresa/ES; 15/10/2015	22	Atlantic forest	Stigmaphylloid (J)
179	*Stigmaphyllon*	*puberulum*	Perdiz 732 (HUEFS)	Ilhéus/BA; 22/01/2010	21	Atlantic forest	Stigmaphylloid (J)
180	*Stigmaphyllon*	*puberulum*	Almeida 905 (RB)	Wenceslau Guimarães/BA; 30/11/2014	11	Atlantic forest	Stigmaphylloid (J)
181	*Stigmaphyllon*	*salzmannii*	Almeida 526 (SP)	São Mateus/ES; 21/01/2012	22	Atlantic forest	Stigmaphylloid (J)
182	*Stigmaphyllon*	*saxicola*	Bardini S/N (HUEFS 198029)	Ponte Nova/MG; 21/05/1980	23	Atlantic forest	Stigmaphylloid (J)
183	*Stigmaphyllon*	*saxicola*	Almeida 551 (HUEFS)	Catas Altas/MG; 22/07/2013	23	Atlantic forest	Stigmaphylloid (J)
184	*Stigmaphyllon*	*saxicola*	Almeida 580 (HUEFS)	Boa Nova/BA; 01/07/2013	12	Atlantic forest	Stigmaphylloid (J)
185	*Stigmaphyllon*	*sinuatum*	Amorim 8643 (CEPEC)	Barcelos/AM; 25/04/2014	1	Amazon rainforest	Stigmaphylloid (J)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(continuation)

Sample	Genus†	Species	Collection code (Herbarium)‡	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
186	*Stigmaphyllon*	*tomentosum*	Almeida 679 (HUEFS)	Jequitinhonha/MG; 10/12/2014	12	Atlantic forest	Stigmaphylloid (J)
187	*Stigmaphyllon*	*urenifolium*	Guedes 13932 (HUEFS)	Bom Jesus da Lapa/BA; 22/09/2007	10	Caatinga/Cerrado	Stigmaphylloid (J)
188	*Stigmaphyllon*	*urenifolium*	Almeida 728 (HUEFS)	Correntina/BA; 15/11/2014	12	Cerrado	Stigmaphylloid (J)
189	*Stigmaphyllon*	*vitifolium*	RBv 3820 (RB)	Rio de Janeiro Botanical Garden; 21/09/2018	24	Atlantic forest	Stigmaphylloid (J)
190	*Tetrapterys*	*mucronata*	Almeida 517 (SP)	Santa Teresa/ES; 17/01/2012	22	Atlantic forest	Tetrapteroid (H)
191	*Tetrapterys*	*phlomoides*	Moraes 40 (RB)	Petrópolis/RJ; 20/09/2018	24	Atlantic forest	Tetrapteroid (H)
192	*Tetrapterys*	*phlomoides*§	Moraes 40 (RB)	Petrópolis/RJ; 20/09/2018	24	Atlantic forest	Tetrapteroid (H)
193	*Tetrapterys*	*schiedeana*	Pace 402 (SPF)	Veracruz/Mexico; 11/08/2013	31	Seasonally dry forest	Tetrapteroid (H)
194	*Tetrapterys*	*xylosteifolia*	Almeida 800 (HUEFS)	Itapema/SC; 25/11/2015	31	Atlantic forest	Tetrapteroid (H)

Supplementary Box 1 – Detailed information of the Malpighiaceae sampling: species list, collection site, location and date, biome, and phylogenetic group classification. Samples consisted of plant leaves, unless specified otherwise. The collection site column is relative to the collection numbers in the map depicted in Figure 14A.

(conclusion)

Sample	Genus[†]	Species	Collection code (Herbarium)[‡]	Location and date (dd/mm/yyyy)	Collection site	Biome	Phylogenetic group
195	*Thryallis*	*latifolia*	Lombardi 685 (BHCB)	Belo Horizonte/MG; 09/05/2018	23	Atlantic forest	Bunchosioid (F)
196	*Verrucularina*	*glaucophylla*	Roque 4265 (HUEFS)	Palmeiras/BA; 24/05/2014	10	Cerrado/Caatinga	Byrsonimoid (A)
197	*Verrucularina*	*piresii*	Amorim 8619 (CEPEC)	Barcelos/AM; 20/04/2014	1	Amazon rainforest	Byrsonimoid (A)

[†]It is important to emphasize that very recently, some species have been synonymized: *Aenigmatanthera lasiandra* to *Niedenzuella lasiandra*; *Tetrapterys ambigua* to *Glicophyllum ambiguum*; *Tetrapterys cardiophylla* to *Glicophyllum cardiophyllum*; *Tetrapterys microphylla* to *Glicophyllum microphyllum*; *Tetrapterys ramiflora* to *Glicophyllum ramiflorum*.

[‡]**ALCB:** Herbarium Alexandre Leal Costa – Federal University of Bahia; **BHCB:** Herbarium of the Federal University of Minas Gerais; **CEPEC:** Herbarium André Maurício V. de Carvalho; **CGMS:** Herbarium of Federal University of Mato Grosso do Sul; **HUEFS:** Herbarium of State University of Feira de Santana; **IAC:** Agronomic Institute of Campinas; **RB:** Rio de Janeiro Botanical Garden Herbarium; **RBv:** Arboretum of Rio de Janeiro Botanical Garden (living collection); **SP:** Institute of Botany of São Paulo; **SPF:** Herbarium of the University of São Paulo; **US:** Smithsonian Institution United States National Herbarium.

[§]Bark.

Source: elaborated by the Author (2022).

Supplementary Box 2 – Characters retrieved from the ancestral characters reconstruction (all clades and genera) based on the classifications obtained *in silico* for Malpighiaceae samples. The ionization mode in which each classification was obtained is described (pos = positive ionization mode; neg = negative ionization mode; both = both ionization modes).

(to be continued)

	Classes present[†]	Classes absent
Clade A (Byrsonimoid clade)	Macrolactams (both); Sulfenyl compounds (pos)	–
Tribe Galphimieae (*Galphimia*+*Verrucularina*)	Naphthopyrans (pos); Oxazinanes (pos)	–
Galphimia	(3'→5')-dinucleotides and analogues (neg); Imidolactams (both); Keto acids and derivatives (neg); Metalloheterocyclic compounds (pos); Organofluorides (pos); Oxacyclic compounds (both)	Benzopyrans (both); Lactams (both); Lignan glycosides (both); Organochlorides (both); Peptidomimetics (both)[§], Pyrans (both); Pyrimidine nucleosides (both); Tetrahydrofurans (both); Thiocarbonyl compounds (pos)
Verrucularina	Benzofurans (neg); Biotin and derivatives (both); Diazanaphthalenes (pos); Dithioles (neg); Furanoid lignans (both); Hydroxy acids and derivatives (neg); Linear 1,3-diarylpropanoids (neg)	Naphthalenes (both)
Tribe Acmanthereae (*Acmanthera*)	Diazanaphthalenes (pos); Propargyl-type 1,3-dipolar organic compounds (pos)	Benzopyrans (both); Lactams (both); Lignan glycosides (both); Pyrimidine nucleosides (both); Pyrimidine nucleotides (both)[§], Saccharolipids (both)[§], Sulfenyl compounds (pos)
Acmanthereae+Byrsonimeae clade	–	(5'→5')-dinucleotides (both); Organic phosphonic acids and derivatives (pos); Pteridines and derivatives (both)

Supplementary Box 2 – Characters retrieved from the ancestral characters reconstruction (all clades and genera) based on the classifications obtained *in silico* for Malpighiaceae samples. The ionization mode in which each classification was obtained is described (pos = positive ionization mode; neg = negative ionization mode; both = both ionization modes).

(continuation)

	Classes present[†]	**Classes absent**
Tribe Byrsonimeae	Hydroxy acids and derivatives (neg); Imidolactams (both); Keto acids and derivatives (neg); Organic phosphoric acids and derivatives (both); Organofluorides (pos)	Oxanes (both)
Byrsonima	Furanoid lignans (both)	Macrolactams (both); Organochlorides (both)
Blepharandra+*Diacidia* clade	Anthracenes (pos); Pteridines and derivatives (both)	Thiocarbonyl compounds (pos)
Blepharandra	(5'→5')-dinucleotides (both); Benzothiadiazoles (pos); Diazanaphthalenes (pos); Dithioles (neg); Oxacyclic compounds (both); Oxanes (both); Oxazinanes (pos); Phenanthrenes and derivatives (both); Sulfoxides (pos); Tetralins (neg); Triazolopyrimidines (pos)[§]	–
Diacidia	Depsides and depsidones (neg); Isoflavonoids (both); Piperidines (pos); Quinolizines (pos); Tetrahydroisoquinolines (both)	Benzimidazoles (both); Coumarins and derivatives (both); Imidolactams (both); Keto acids and derivatives (neg); Lignan glycosides (both); Organofluorides (pos); Organothiophosphorus compounds (pos); Pyrans (both); Pyrimidine nucleosides (both); Sulfenyl compounds (pos); Tetrapyrroles and derivatives (both)
Clade B (Acridocarpoid clade)	Diazanaphthalenes (pos); Isoflavonoids (both); Oxacyclic compounds (both)	Tetrahydrofurans (both)
Clade B+remaining clades	Dithioles (neg); Furanoid lignans (both); Organic phosphoric acids and derivatives (both); Propargyl-type 1,3-dipolar organic compounds (pos)	–

Supplementary Box 2 – Characters retrieved from the ancestral characters reconstruction (all clades and genera) based on the classifications obtained *in silico* for Malpighiaceae samples. The ionization mode in which each classification was obtained is described (pos = positive ionization mode; neg = negative ionization mode; both = both ionization modes).

(continuation)

	Classes present†	**Classes absent**
Clade C (Mcvaughioid clade)	Linear 1,3-diarylpropanoids (neg)	Dithioles (neg); Indoles and derivatives (both)
Burdachia	Borinic acid derivatives (pos); Isocoumarins and derivatives (both); Keto acids and derivatives (neg); Organochlorides (both); Oxacyclic compounds (both); Oxazinanes (pos); Thiocarbonyl compounds (pos)	(5'→5')-dinucleotides (both); Anthracenes (pos); Benzopyrans (both); Propargyl-type 1,3-dipolar organic compounds (pos); Thiophenes (both)
Mcvaughia	Biotin and derivatives (both); Isoflavonoids (both); Sulfinyl compounds (pos)	Benzimidazoles (both); Furanoid lignans (both); Organic carbonic acids and derivatives (both); Organic phosphonic acids and derivatives (pos); Pyrimidine nucleosides (both)
Clade C+remaining clades	Anthracenes (pos); Hydroxy acids and derivatives (neg); Naphthopyrans (pos); Organofluorides (pos); Thiophenes (both)§	Organochlorides (both); Pyrans (both); Thiocarbonyl compounds (pos)
Clade D (Barnebyoid clade)	Diarylheptanoids (neg); Keto acids and derivatives (neg); Oxazinanes (pos)	Benzopyrans (both); Furanoid lignans (both); Glycerophospholipids (both)§, Lignan glycosides (both); Naphthalenes (both); Naphthopyrans (pos); Propargyl-type 1,3-dipolar organic compounds (pos); Pteridines and derivatives (both); Tetrahydrofurans (both)
Clade D+remaining clades	Metalloheterocyclic compounds (pos)	Oxanes (both)
Clade E (Ptilochaetoid clade)	2-arylbenzofuran flavonoids (neg); Dibenzylbutane lignans (neg); Isoflavonoids (both); Oxacyclic compounds (both); Oxanes (both); Pyrrolidines (neg); Thiocarbonyl compounds (pos)	Organothiophosphorus compounds (pos); Thiophenes (both)

Supplementary Box 2 – Characters retrieved from the ancestral characters reconstruction (all clades and genera) based on the classifications obtained *in silico* for Malpighiaceae samples. The ionization mode in which each classification was obtained is described (pos = positive ionization mode; neg = negative ionization mode; both = both ionization modes).

(continuation)

	Classes present†	**Classes absent**
Clade E+remaining clades	Benzofurans (neg); Biotin and derivatives (both); Macrolactams (both); Pyrans (both); Sulfenyl compounds (pos)	Organic thiophosphoric acids and derivatives (both); Organofluorides (pos)
Clade F (Bunchosioid clade)	Azolidines (pos); Benzodioxoles (both); Organochlorides (both); Quinolizines (pos)	Organic carbonic acids and derivatives (both); Organic phosphoric acids and derivatives (both)
Heladena	(3'→5')-dinucleotides and analogues (neg); 5'-deoxyribonucleosides (both); Halohydrins (neg)	Benzofurans (neg); Benzopyrans (both); Dithioles (neg); Furanoid lignans (both); Organic sulfuric acids and derivatives (both); Propargyl-type 1,3-dipolar organic compounds (pos); Pteridines and derivatives (both); Tetrahydrofurans (both)
Bunchosia	Lactams (both); Linear 1,3-diarylpropanoids (neg); Organic thiophosphoric acids and derivatives (both); Organofluorides (pos); Phenylpropanoic acids (neg); Sulfoxides (pos)	Benzodioxoles (both); Biotin and derivatives (both); Organochlorides (both); Pyrans (both); Quinolizines (pos); Tetrapyrroles and derivatives (both)
Bunchosia+*Thryallis* clade	Thiocarbonyl compounds (pos)	Sulfenyl compounds (pos); Thiophenes (both)
Thryallis	Diarylheptanoids (neg); Dibenzylbutane lignans (neg); Furans (pos)§, Lactones (neg)§, Metalloheterocyclic compounds (pos); Organic carbonic acids and derivatives (both); Organic phosphoric acids and derivatives (both); Oxanes (both); Sulfinyl compounds (pos)	Anthracenes (pos); Azolidines (pos); Naphthalenes (both); Pyrimidine nucleosides (both)
Clade F+remaining clades	Diazanaphthalenes (pos)	(5'→5')-dinucleotides (both); Lactams (both);

Supplementary Box 2 – Characters retrieved from the ancestral characters reconstruction (all clades and genera) based on the classifications obtained *in silico* for Malpighiaceae samples. The ionization mode in which each classification was obtained is described (pos = positive ionization mode; neg = negative ionization mode; both = both ionization modes).

(continuation)

	Classes present†	**Classes absent**
		Metalloheterocyclic compounds (pos)
Clade G (Hiraeoid clade)	Piperidines (pos)	Benzofurans (neg); Benzopyrans (both); Dithioles (neg); Furanoid lignans (both); Hydroxy acids and derivatives (neg); Naphthopyrans (pos); Pteridines and derivatives (both); Pyrimidine nucleosides (both)
Clade G+remaining clades	Borinic acid derivatives (pos); Organic thiophosphoric acids and derivatives (both)	Anthracenes (pos); Macrolactams (both)
Clade H (Tetrapteroid clade)	–	Organic phosphonic acids and derivatives (pos)
Glicophyllum	Arylnaphthalene lignans (neg); Lactams (both); Linear 1,3-diarylpropanoids (neg); Macrolactams (both); Organofluorides (pos)	Naphthopyrans (pos); Tetrahydrofurans (both); Thiophenes (both)
Glicophyllum+Niedenzuella clade	Organochlorides (both)	Sulfenyl compounds (pos)
Niedenzuella	Anthracenes (pos); Thiocarbonyl compounds (pos)	Borinic acid derivatives (pos); Organic thiophosphoric acids and derivatives (both); Oxanes (both); Pyrimidine nucleosides (both)
Tetrapterys	(5'→5')-dinucleotides (both); Benzothiadiazoles (pos); Depsides and depsidones (neg); Linear 1,3-diarylpropanoids (neg); Macrolactams (both); Phenylpropanoic acids (neg); Tetralins (neg)	Metalloheterocyclic compounds (pos); Naphthopyrans (pos); Organic thiophosphoric acids and derivatives (both); Propargyl-type 1,3-dipolar organic compounds (pos); Thiophenes (both)

Supplementary Box 2 – Characters retrieved from the ancestral characters reconstruction (all clades and genera) based on the classifications obtained *in silico* for Malpighiaceae samples. The ionization mode in which each classification was obtained is described (pos = positive ionization mode; neg = negative ionization mode; both = both ionization modes).

(continuation)

	Classes present†	**Classes absent**
Remaining clades	–	Benzofurans (neg); Benzopyrans (both); Diazanaphthalenes (pos); Furanoid lignans (both); Lignan glycosides (both); Pyrimidine nucleosides (both)
Christianella	Isoindoles and derivatives (both); Oxacyclic compounds (both); Tetrahydroisoquinolines (both)	Coumarins and derivatives (both); Dithioles (neg); Naphthopyrans (pos); Organothiophosphorus compounds (pos); Pteridines and derivatives (both); Pyrans (both); Tannins (neg)§, Tetrahydrofurans (both)
Christianella+remaining clades	Lactams (both); Organochlorides (both)	Sulfenyl compounds (pos)
Alicia	Anthracenes (pos); Benzothiadiazoles (pos); Benzothiazoles (both); Diazanaphthalenes (pos); Imidolactams (both); Organofluorides (pos); Thiocarbonyl compounds (pos)	Biotin and derivatives (both); Borinic acid derivatives (pos); Lactams (both); Organic thiophosphoric acids and derivatives (both); Organochlorides (both); Oxanes (both); Thiophenes (both)
Alicia+*Callaeum* clade	Arylnaphthalene lignans (neg); Macrolactams (both); Organic phosphonic acids and derivatives (pos)	Metalloheterocyclic compounds (pos)
Callaeum	5'-deoxyribonucleosides (both); Benzodioxoles (both); Benzofurans (neg); Benzopyrans (both); Depsides and depsidones (neg); Dibenzylbutane lignans (neg); Isocoumarins and derivatives (both); Lignan glycosides (both); Lignan lactones (pos); Phenylpropanoic acids (neg); Piperidines (pos); Sulfenyl compounds (pos)	–

Supplementary Box 2 – Characters retrieved from the ancestral characters reconstruction (all clades and genera) based on the classifications obtained *in silico* for Malpighiaceae samples. The ionization mode in which each classification was obtained is described (pos = positive ionization mode; neg = negative ionization mode; both = both ionization modes).

(continuation)

	Classes present†	**Classes absent**
Heteropterys	Lignan glycosides (both); Pyrimidine nucleosides (both)	Borinic acid derivatives (pos); Propargyl-type 1,3-dipolar organic compounds (pos)
Heteropterys+remaining clades	–	–
Dicella	Depsides and depsidones (neg); Furanoid lignans (both); Isoflavonoids (both); Oxacyclic compounds (both); Sulfoxides (pos)	–
Dicella+remaining clades	(5'→5')-dinucleotides (both); Benzopyrans (both); Quinolizines (pos); Tetrahydroisoquinolines (both)	Dithioles (neg); Oxanes (both)
Carolus	Isocoumarins and derivatives (both); Organochlorides (both); Oxanes (both); Phenylpropanoic acids (neg)	(5'→5')-dinucleotides (both); Coumarins and derivatives (both); Hydroxy acids and derivatives (neg); Indoles and derivatives (both); Metalloheterocyclic compounds (pos); Organic carbonic acids and derivatives (both); Organothiophosphorus compounds (pos); Pyrans (both); Sulfenyl compounds (pos); Tetrahydrofurans (both); Tetrapyrroles and derivatives (both); Thiophenes (both); Triazines (pos)§
Carolus+*Hiptage* clade	5'-deoxyribonucleosides (both)	–
Hiptage	Diazanaphthalenes (pos); Dithioles (neg)	Benzopyrans (both); Borinic acid derivatives (pos); Naphthopyrans (pos); Organic thiophosphoric acids and derivatives (both); Propargyl-type 1,3-dipolar organic compounds (pos); Pteridines and derivatives (both);

Supplementary Box 2 – Characters retrieved from the ancestral characters reconstruction (all clades and genera) based on the classifications obtained *in silico* for Malpighiaceae samples. The ionization mode in which each classification was obtained is described (pos = positive ionization mode; neg = negative ionization mode; both = both ionization modes).

(continuation)

	Classes present†	**Classes absent**
		Quinolines and derivatives (both)§, Quinolizines (pos)
Clade H+remaining clades	Metalloheterocyclic compounds (pos); Oxanes (both)	–
Clade I (Malpighioid clade)	(3'→5')-dinucleotides and analogues (neg); Imidolactams (both); Piperidines (pos)	Benzofurans (neg); Furanoid lignans (both); Lignan glycosides (both)
Ectopopterys	Organofluorides (pos); Sulfoxides (pos)	Benzimidazoles (both); Coumarins and derivatives (both); Dithioles (neg); Naphthopyrans (pos); Organic phosphonic acids and derivatives (pos); Organic sulfuric acids and derivatives (both); Organic thiophosphoric acids and derivatives (both); Pyrans (both); Thiocarbonyl compounds (pos)
Amorimia	–	Biotin and derivatives (both); Hydroxy acids and derivatives (neg); Metalloheterocyclic compounds (pos); Piperidines (pos); Pyrimidine nucleosides (both); Thiophenes (both)
Amorimia+remaining clades	–	Diazanaphthalenes (pos)
Malpighia	Anthracenes (pos); Diarylheptanoids (neg); Keto acids and derivatives (neg); Lignan glycosides (both); Linear 1,3-diarylpropanoids (neg); Organofluorides (pos); Quinolizines (pos); Sulfinyl compounds (pos)	Propargyl-type 1,3-dipolar organic compounds (pos); Sulfenyl compounds (pos)
Malpighia+*Mascagnia* clade	Borinic acid derivatives (pos); Isoflavonoids (both); Oxocins	(3'→5')-dinucleotides and analogues (neg);

Supplementary Box 2 – Characters retrieved from the ancestral characters reconstruction (all clades and genera) based on the classifications obtained *in silico* for Malpighiaceae samples. The ionization mode in which each classification was obtained is described (pos = positive ionization mode; neg = negative ionization mode; both = both ionization modes).

(continuation)

	Classes present†	**Classes absent**
	(neg)§	Benzopyrans (both)
Mascagnia	(5'→5')-dinucleotides (both); Benzodioxoles (both); Depsides and depsidones (neg); Isoindoles and derivatives (both); Lactams (both); Lignan lactones (pos); Macrolactams (both); Oxacyclic compounds (both); Phenanthrenes and derivatives (both)	Imidolactams (both)
Clades I+J	Thiocarbonyl compounds (pos)	Borinic acid derivatives (pos)
Clade J (Stigmaphylloid clade)	Macrolactams (both)	Biotin and derivatives (both); Sulfenyl compounds (pos)
Bronwenia	Anthracenes (pos); Azolidines (pos); Lactams (both); Linear 1,3-diarylpropanoids (neg); Organofluorides (pos); Phenylpropanoic acids (neg); Sulfoxides (pos)	Metalloheterocyclic compounds (pos); Organic carbonic acids and derivatives (both); Organic phosphoric acids and derivatives (both); Oxanes (both); Pyrans (both); Tetrapyrroles and derivatives (both); Thiophenes (both)
Diplopterys	(5'→5')-dinucleotides (both); 5'-deoxyribonucleosides (both); Azolidines (pos); Depsides and depsidones (neg); Imidolactams (both); Organofluorides (pos); Sulfenyl compounds (pos)	Benzofurans (neg); Dithioles (neg); Furanoid lignans (both); Indoles and derivatives (both); Naphthalenes (both); Naphthopyrans (pos); Organic phosphonic acids and derivatives (pos); Organochlorides (both); Oxanes (both); Propargyl-type 1,3-dipolar organic compounds (pos)
Stigmaphyllon	Biotin and derivatives (both); Pyrimidine nucleosides (both)	Macrolactams (both)
Diplopterys+*Stigmaphyllon*+re	Organochlorides (both)	Pyrimidine nucleosides (both)

Supplementary Box 2 – Characters retrieved from the ancestral characters reconstruction (all clades and genera) based on the classifications obtained *in silico* for Malpighiaceae samples. The ionization mode in which each classification was obtained is described (pos = positive ionization mode; neg = negative ionization mode; both = both ionization modes).

(continuation)

	Classes present†	**Classes absent**
maining genera		
Banisteriopsis	Diazanaphthalenes (pos)	Benzofurans (neg); Benzopyrans (both); Dithioles (neg); Furanoid lignans (both); Macrolactams (both); Metalloheterocyclic compounds (pos); Tetrahydrofurans (both)
Sphedamnocarpus	2-arylbenzofuran flavonoids (neg); 5'-deoxyribonucleosides (both); Anthracenes (pos); Arylnaphthalene lignans (neg); Biotin and derivatives (both); Borinic acid derivatives (pos); Depsides and depsidones (neg); Isoflavonoids (both); Lactams (both); Linear 1,3-diarylpropanoids (neg); Oxacyclic compounds (both); Pyrimidine nucleosides (both); Pyrrolidines (neg); Sulfenyl compounds (pos)	Thiophenes (both)
Banisteriopsis+*Sphedamnocarpus*+remaining clades	–	Diazanaphthalenes (pos)
Peixotoa	Dibenzylbutane lignans (neg); Linear 1,3-diarylpropanoids (neg)	Organic carbonic acids and derivatives (both); Organic phosphonic acids and derivatives (pos); Pyrans (both); Thiophenes (both)
Peixotoa+remaining clades	–	Thiocarbonyl compounds (pos)

Supplementary Box 2 – Characters retrieved from the ancestral characters reconstruction (all clades and genera) based on the classifications obtained *in silico* for Malpighiaceae samples. The ionization mode in which each classification was obtained is described (pos = positive ionization mode; neg = negative ionization mode; both = both ionization modes).

(conclusion)

	Classes present†	**Classes absent**
Aspicarpa	Depsides and depsidones (neg); Diazanaphthalenes (pos); Dithioles (neg); Imidolactams (both); Phenanthrenes and derivatives (both); Thiocarbonyl compounds (pos)	Biotin and derivatives (both); Lignan glycosides (both); Organic phosphoric acids and derivatives (both); Piperidines (pos); Sulfenyl compounds (pos); Tetrapyrroles and derivatives (both)
Aspicarpa+*Janusia* clade	Lactams (both)	Macrolactams (both)
Janusia	Metalloheterocyclic compounds (pos)	Naphthopyrans (pos); Organic carbonic acids and derivatives (both); Thiophenes (both)
Aspicarpa+*Janusia*+remaining clades	Biotin and derivatives (both); Piperidines (pos); Sulfenyl compounds (pos)	Benzofurans (neg); Dithioles (neg); Furanoid lignans (both); Metalloheterocyclic compounds (pos); Propargyl-type 1,3-dipolar organic compounds (pos)
Camarea	Benzothiazoles (both); Depsides and depsidones (neg); Diazanaphthalenes (pos); Halohydrins (neg); Keto acids and derivatives (neg); Propargyl-type 1,3-dipolar organic compounds (pos)	Benzopyrans (both); Organic phosphonic acids and derivatives (pos)
Camarea+*Gaudichaudia* clade	Phenylpropanoic acids (neg)	Organochlorides (both); Oxanes (both)
Gaudichaudia	(3'→5')-dinucleotides and analogues (neg); Benzodioxoles (both); Isoindoles and derivatives (both); Organofluorides (pos); Pyrimidine nucleosides (both); Quinolizines (pos); Thiocarbonyl compounds (pos)	–

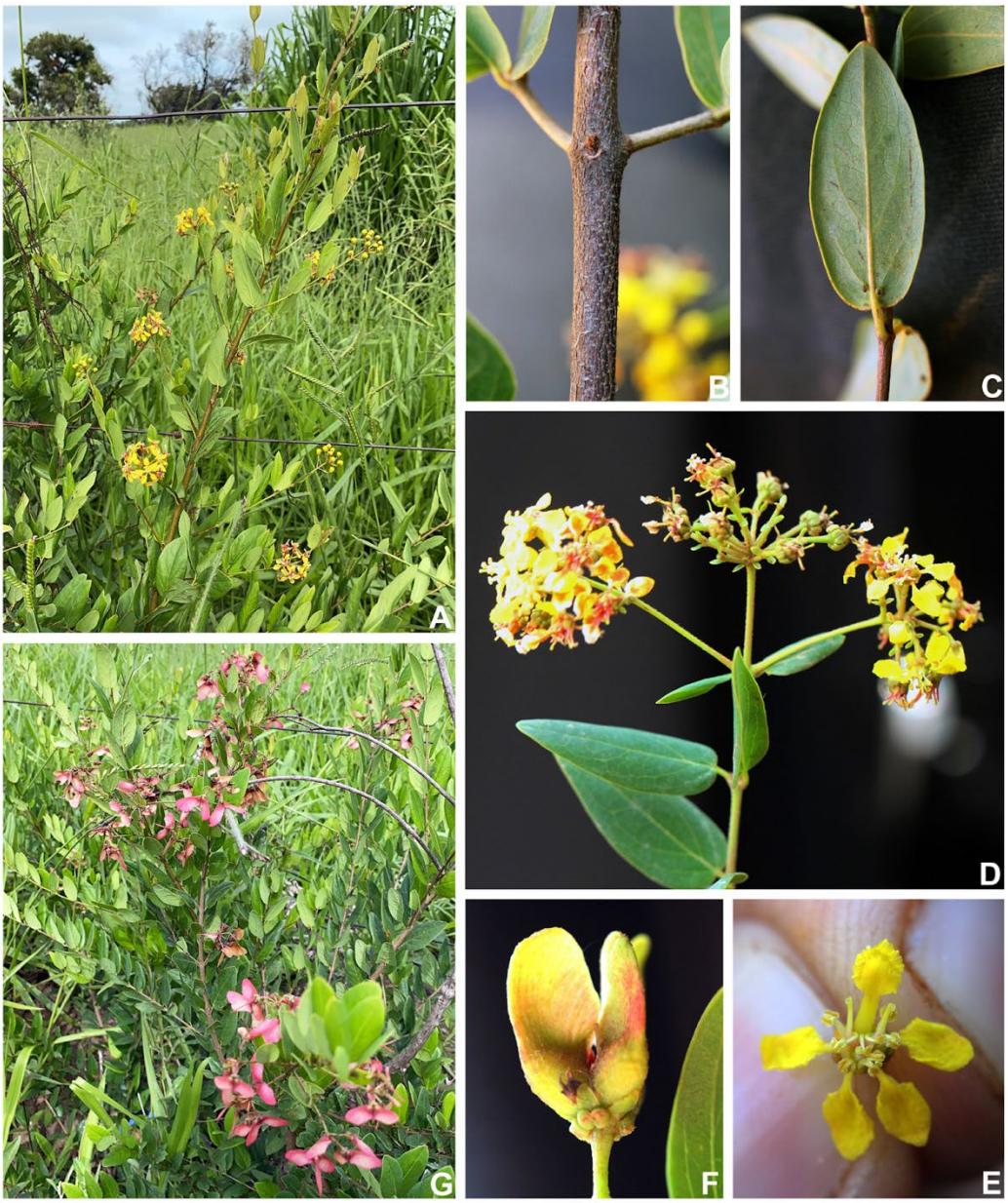
†**Classes retrieved as present in all Malpighiaceae clades:** Alkyl halides (both); Allyl-type 1,3-dipolar organic compounds (both); Aryl halides (both); Azacyclic compounds (both); Azoles (both); Benzene and substituted derivatives (both); Boronic acid derivatives (both); Carboxylic acids and derivatives (both); Cinnamic acids and derivatives (both); Diazinanes (pos); Diazines (both); Fatty Acyls (both); Flavonoids (both); Glycerolipids (both); Heteroaromatic compounds (both);

Imidazopyrimidines (both); Macrolides and analogues (pos); Organic metal salts (pos); Organic sulfonic acids and derivatives (both); Organonitrogen compounds (both); Organooxygen compounds (both); Phenol ethers (both); Phenols (both); Prenol lipids (both); Purine nucleosides (both); Purine nucleotides (both); Pyridines and derivatives (both); Sphingolipids (both); Steroids and steroid derivatives (both); Sulfonyls (pos); Thioethers (both).

[§]Synapomorphy.

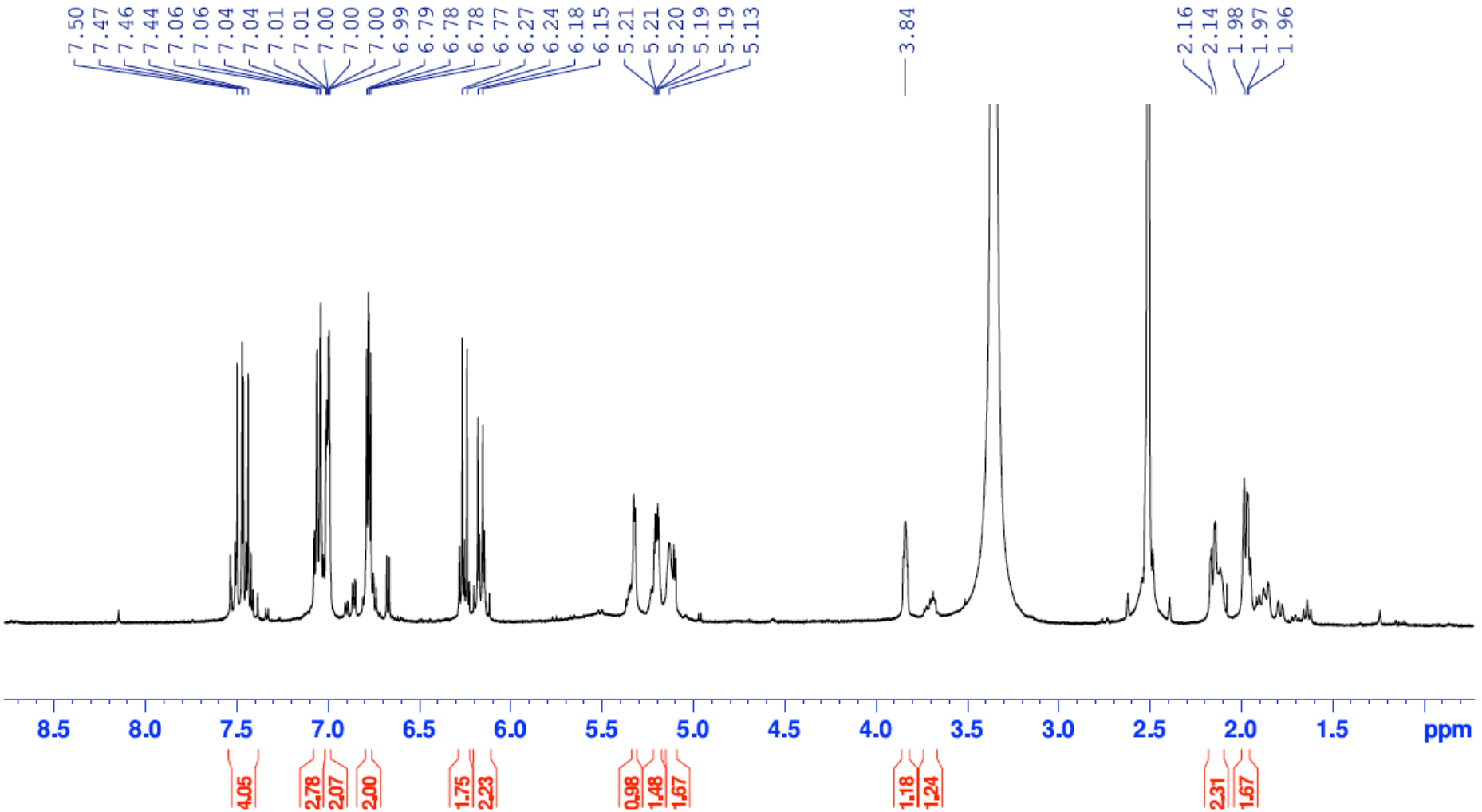
Source: elaborated by the Author (2022).

SUPPLEMENTARY MATERIAL C – CHAPTER 3

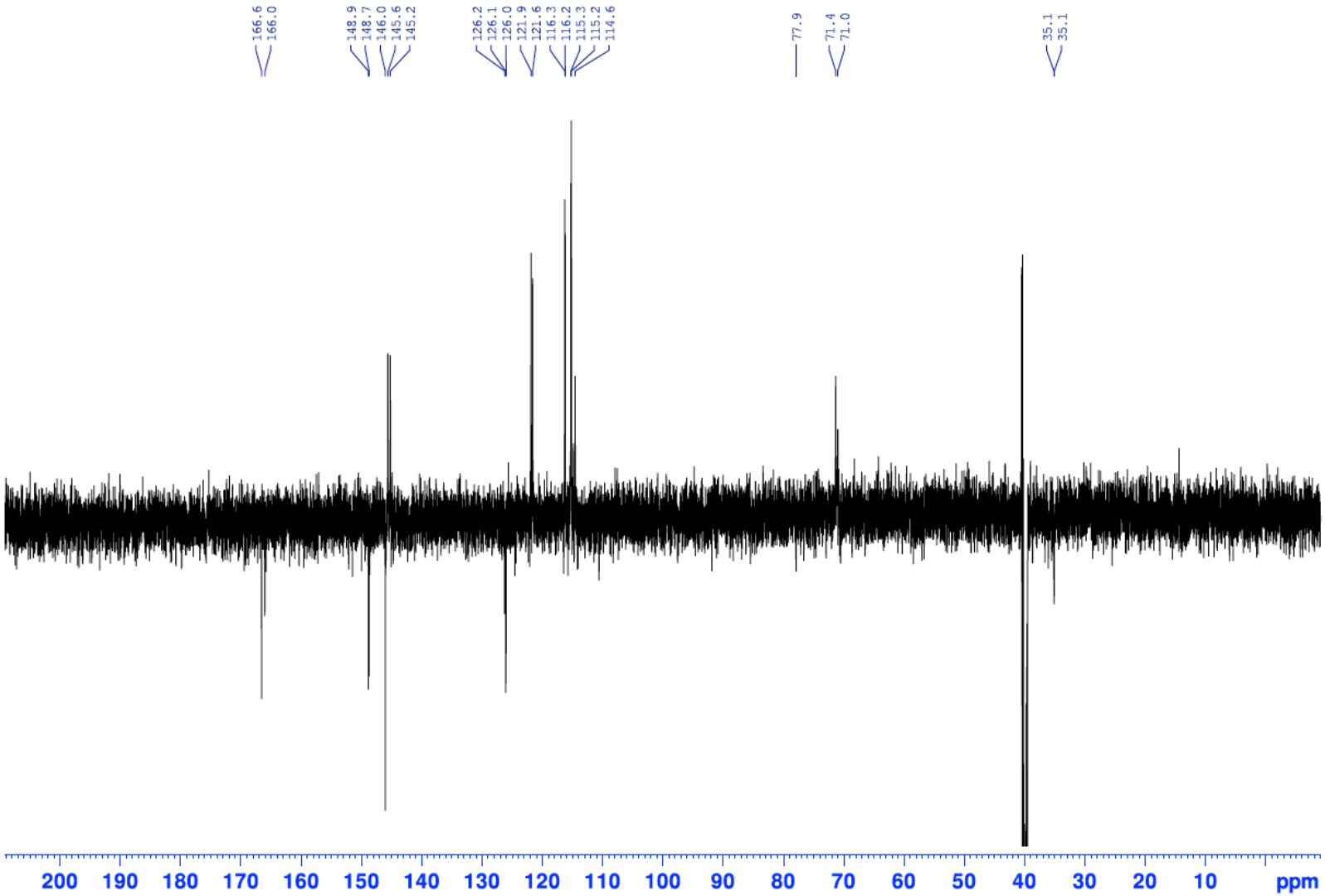


Supplementary Figure 7 – Field photographs of *Heteropterys umbellata* in the Cerrado biome. (**A**) Erect shrubs with flowers. (**B**) Detail of the stem. (**C**) Detail of the abaxial surface of a leaf. (**D**) Terminal umbels in lateral view. (**E**) Flower in frontal view. (**F**) Fruits in lateral view. (**G**) Fruiting branches. (**A**, **G** by H.Mannochio-Russo, and **B**–**F** by R.F.Almeida).

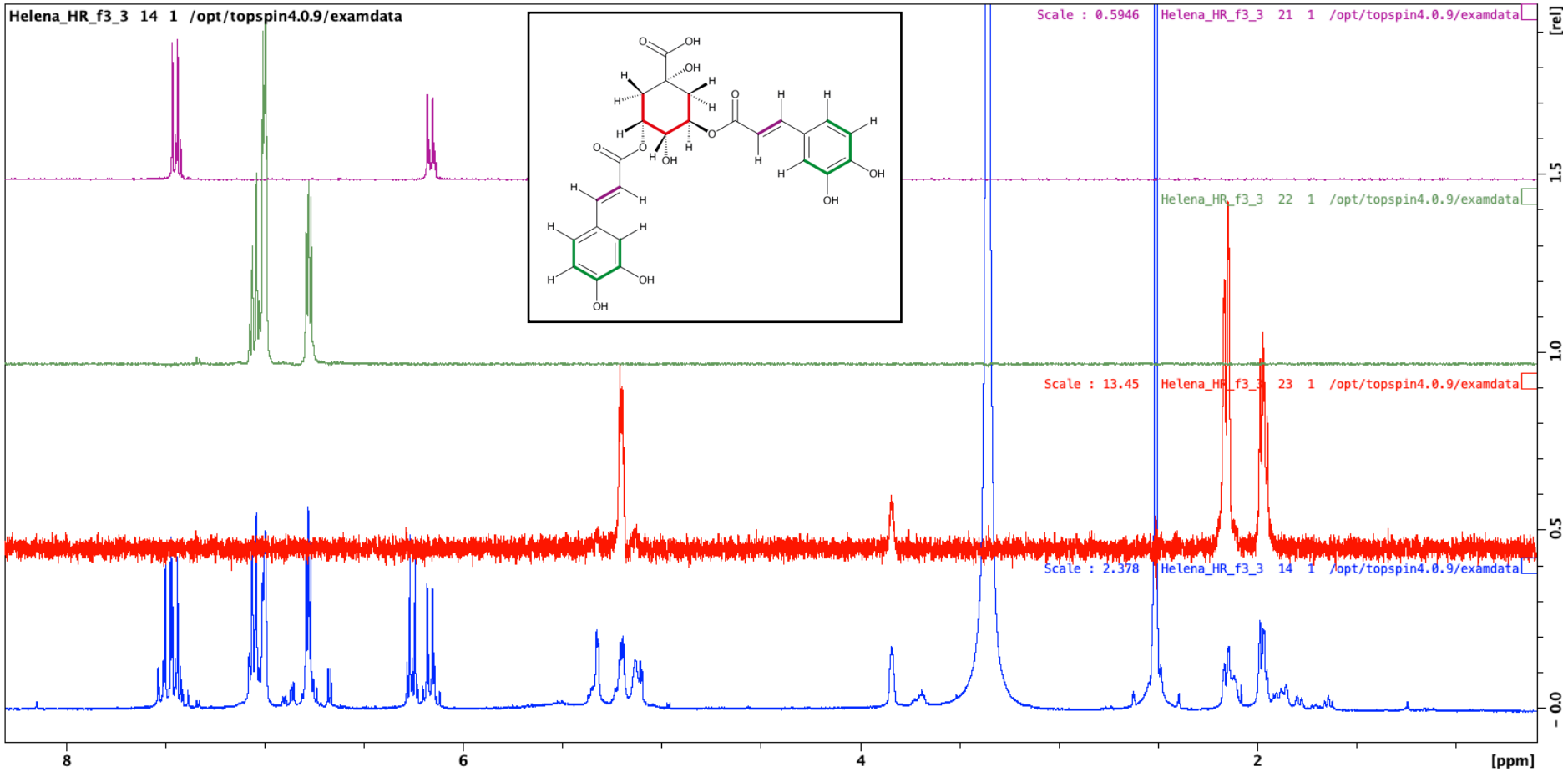
Source: elaborated by the Author (2022).



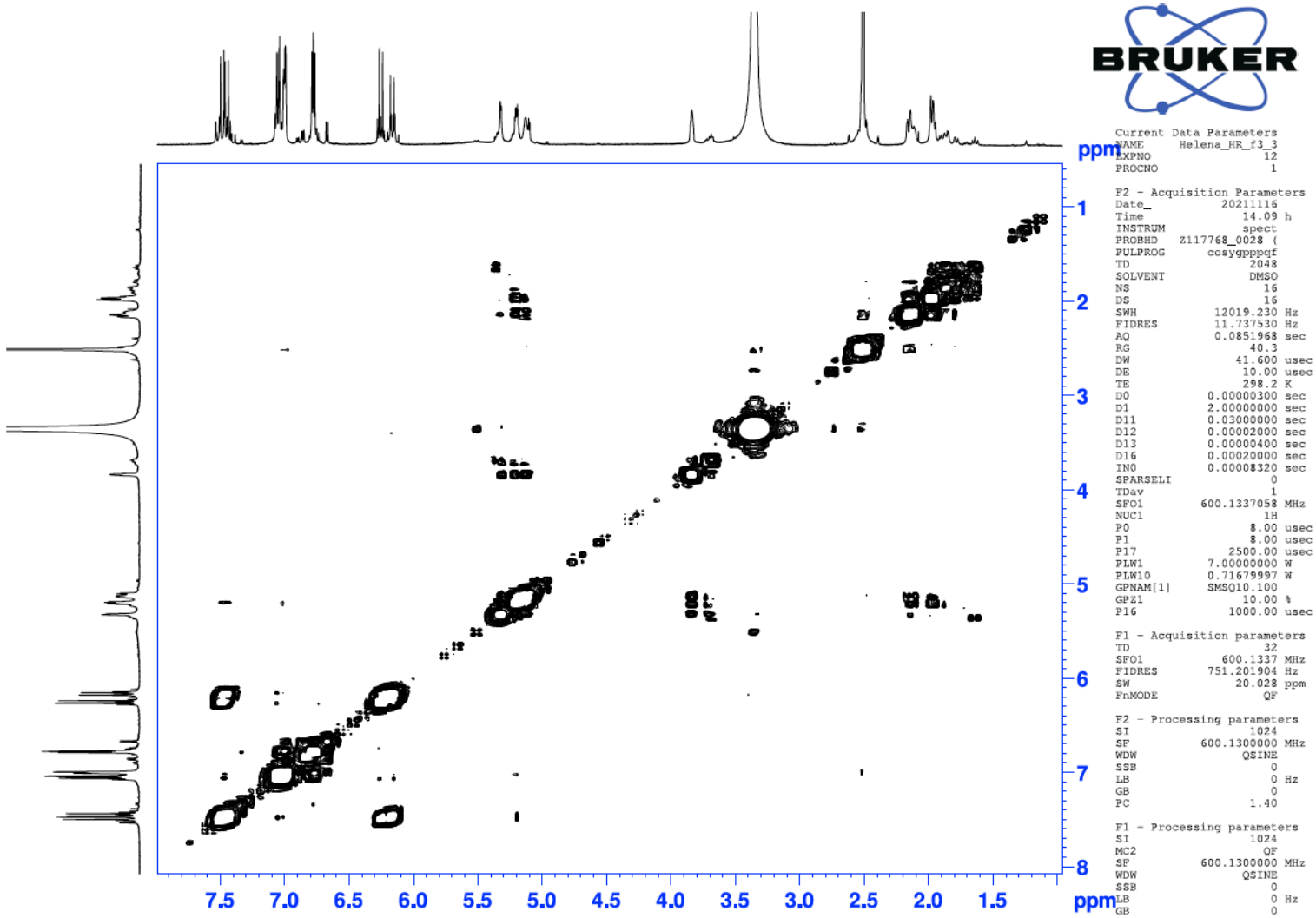
Supplementary Figure 8 – ^{1}H NMR (600 MHz, DMSO-d_6) spectrum of compound **1**.



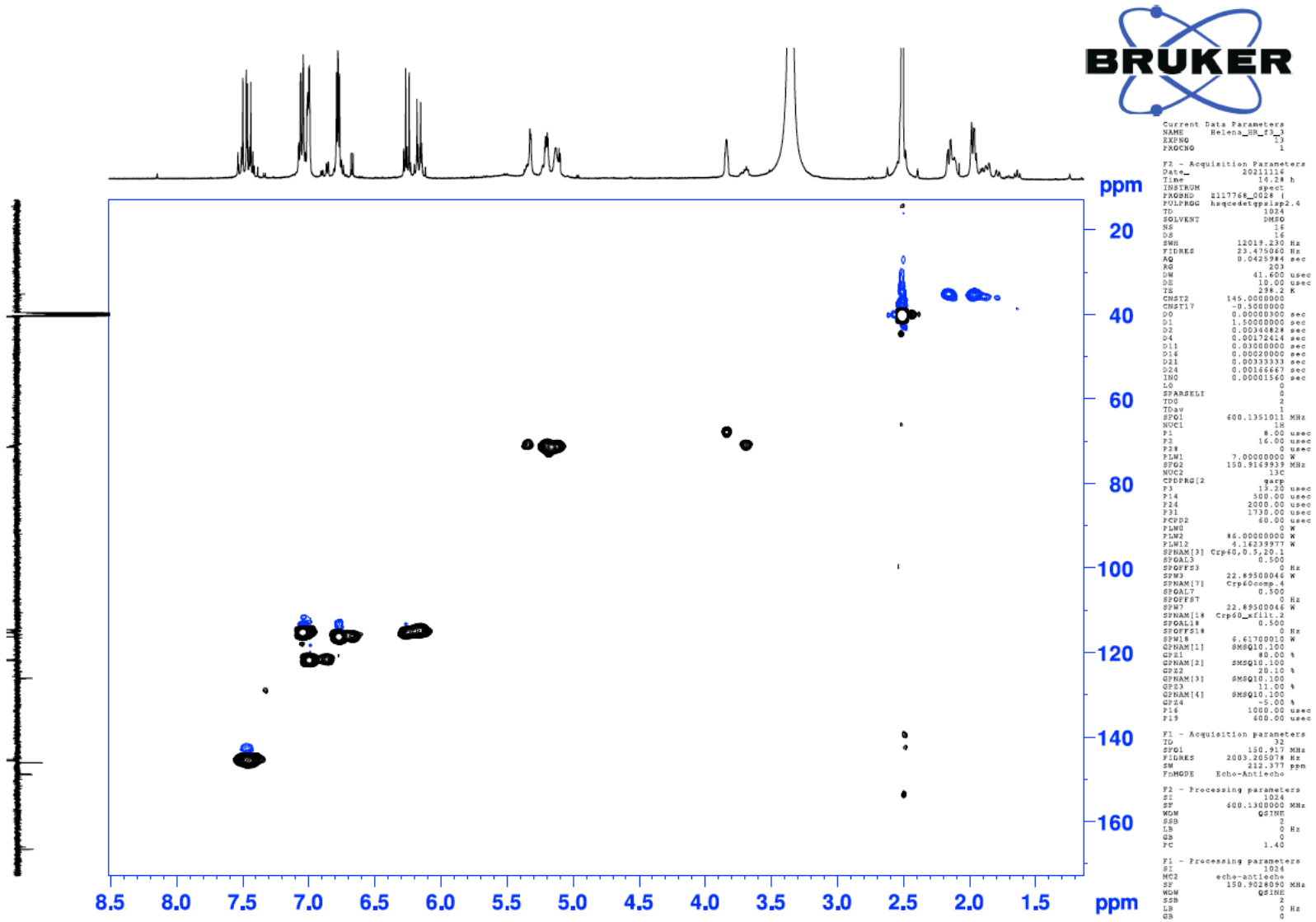
Supplementary Figure 9 – DEPTQ NMR (150 MHz, DMSO-d_6) spectrum of compound **1**.



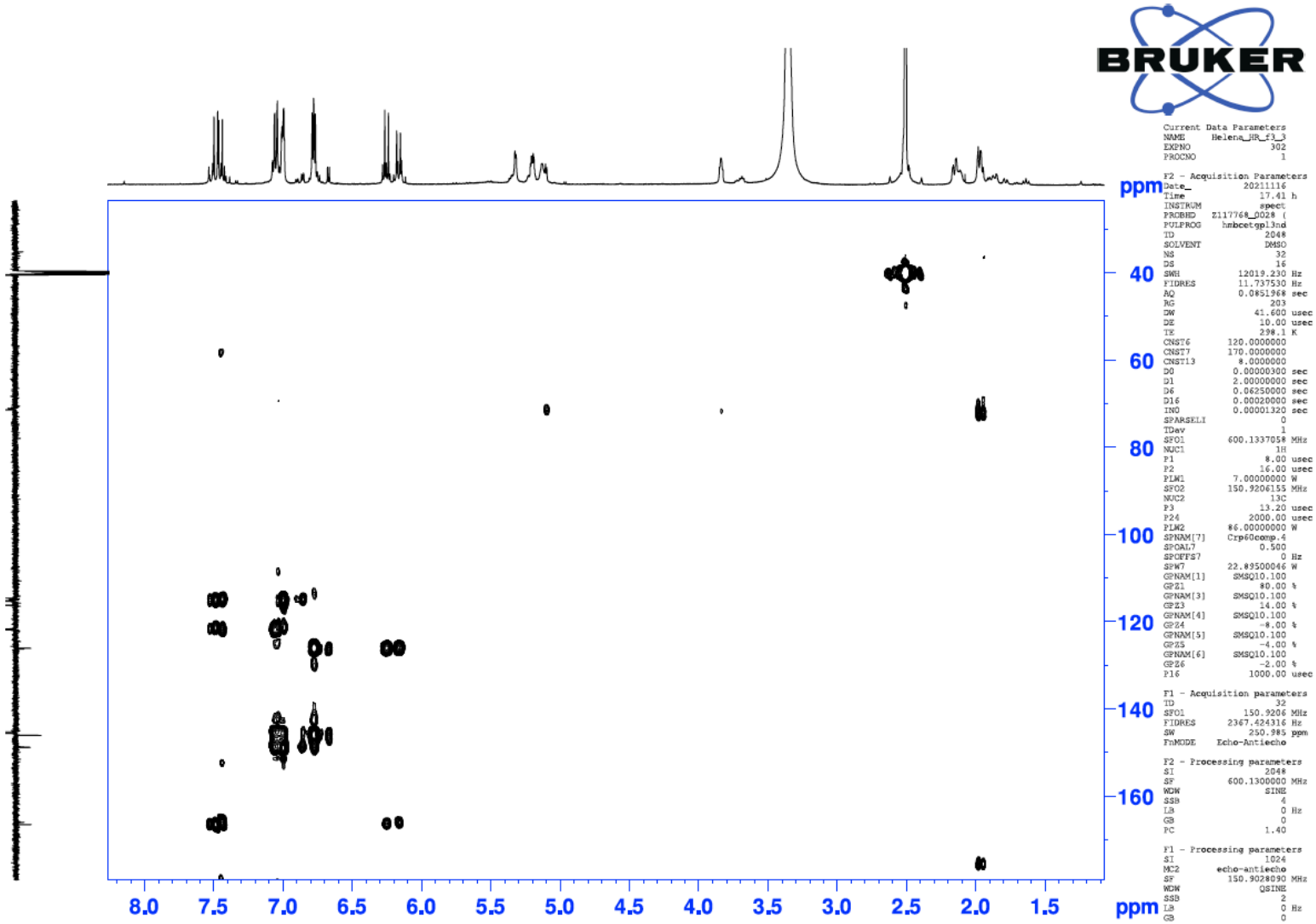
Supplementary Figure 10 – TOCSY 1D NMR (600 MHz, DMSO-d_6) spectrum of compound **1**. The structure of compond **1** showing all the spin systems is displayed.



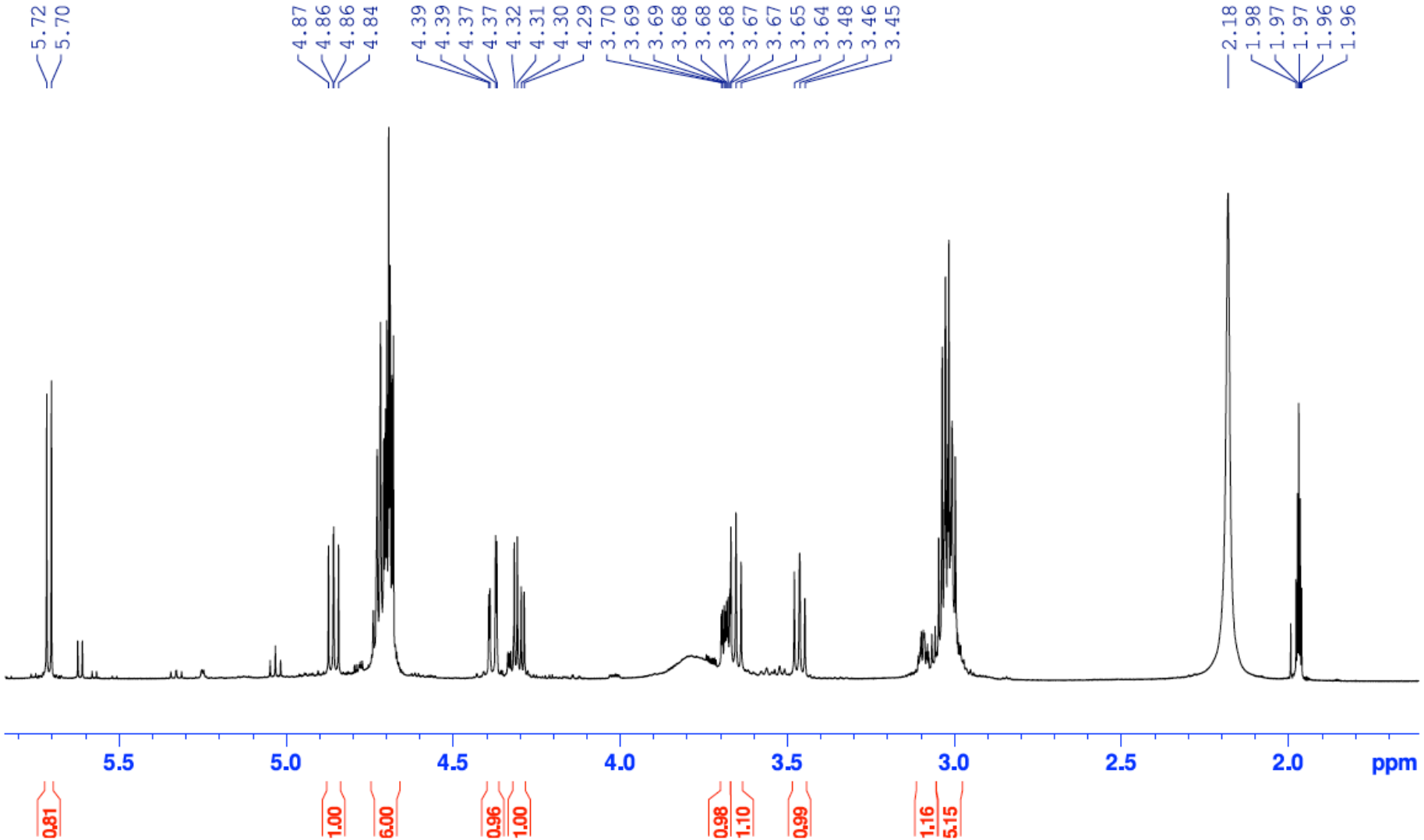
Supplementary Figure 11 – COSY NMR spectrum (600 MHz, DMSO-d_6) of compound **1**.



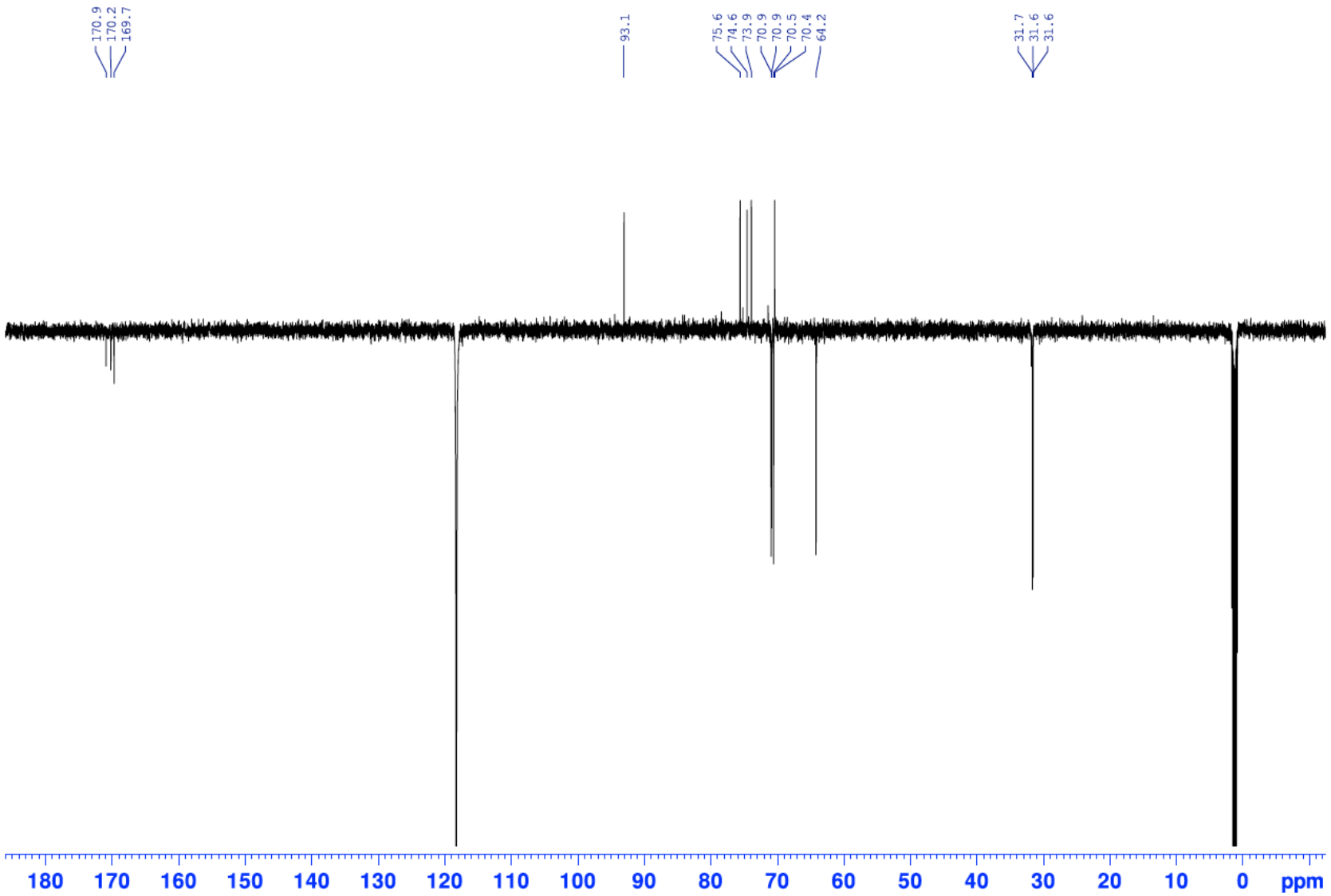
Supplementary Figure 12 – HSQC NMR (600 MHz, DMSO-d_6) spectrum of compound **1**.



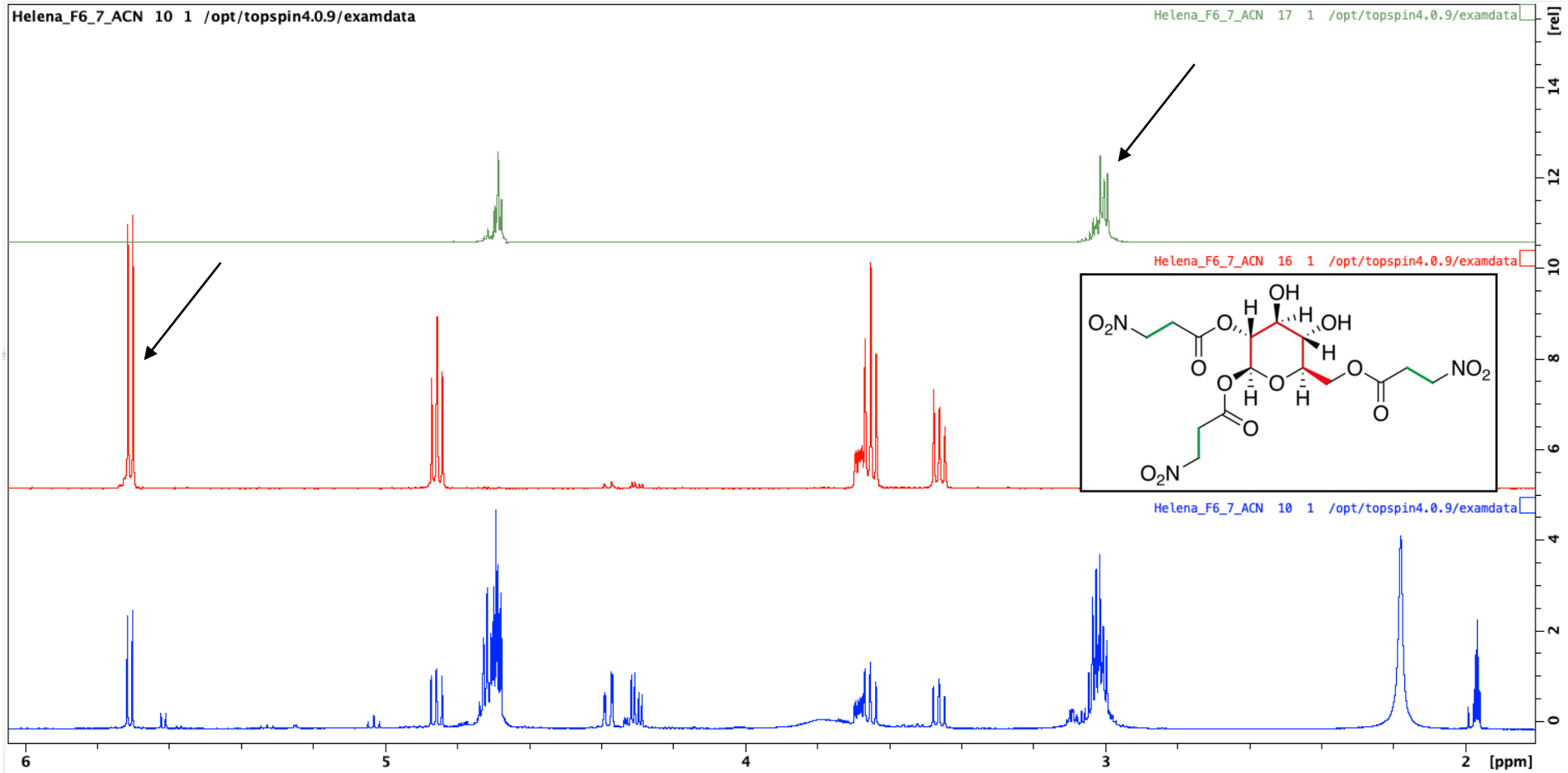
Supplementary Figure 13 – HMBC NMR (600 MHz, DMSO-d_6) spectrum of compound **1**.



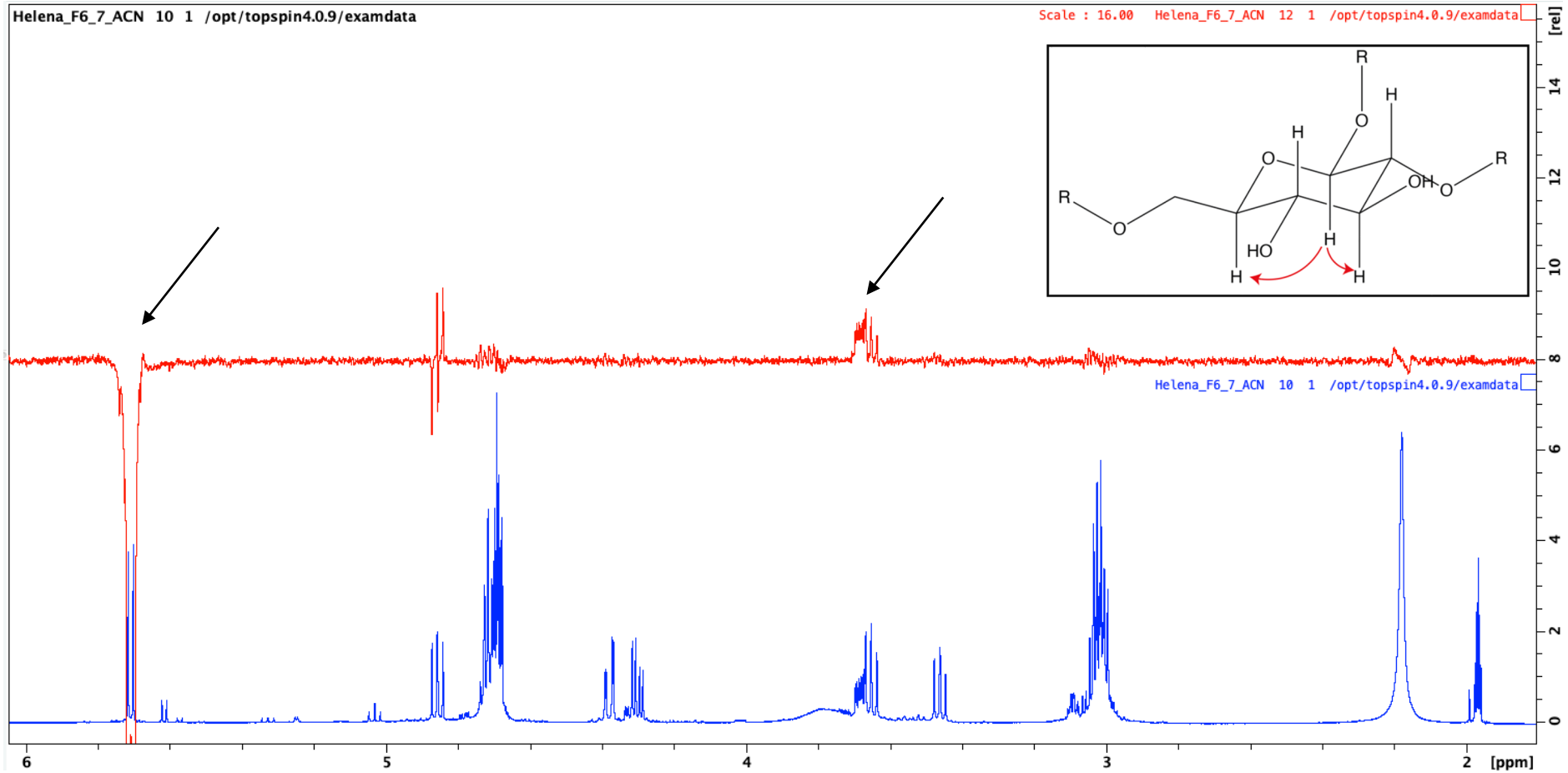
Supplementary Figure 14 – ^{1}H NMR (600 MHz, CD_3CN-d_3) spectrum of compound **2**.



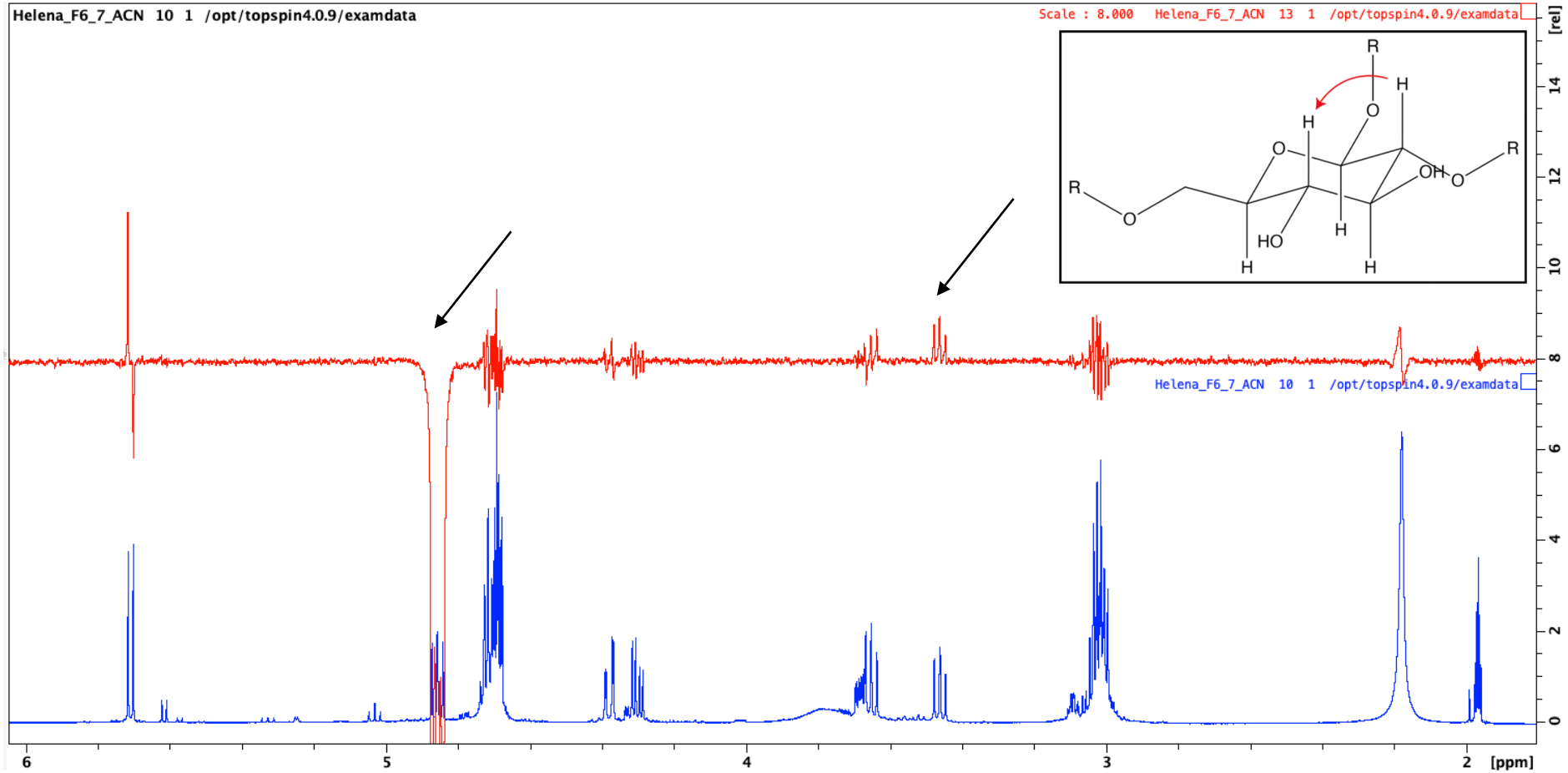
Supplementary Figure 15 – DEPTQ NMR (150 MHz, CD_3CN-d_3) spectrum of compound **2**.



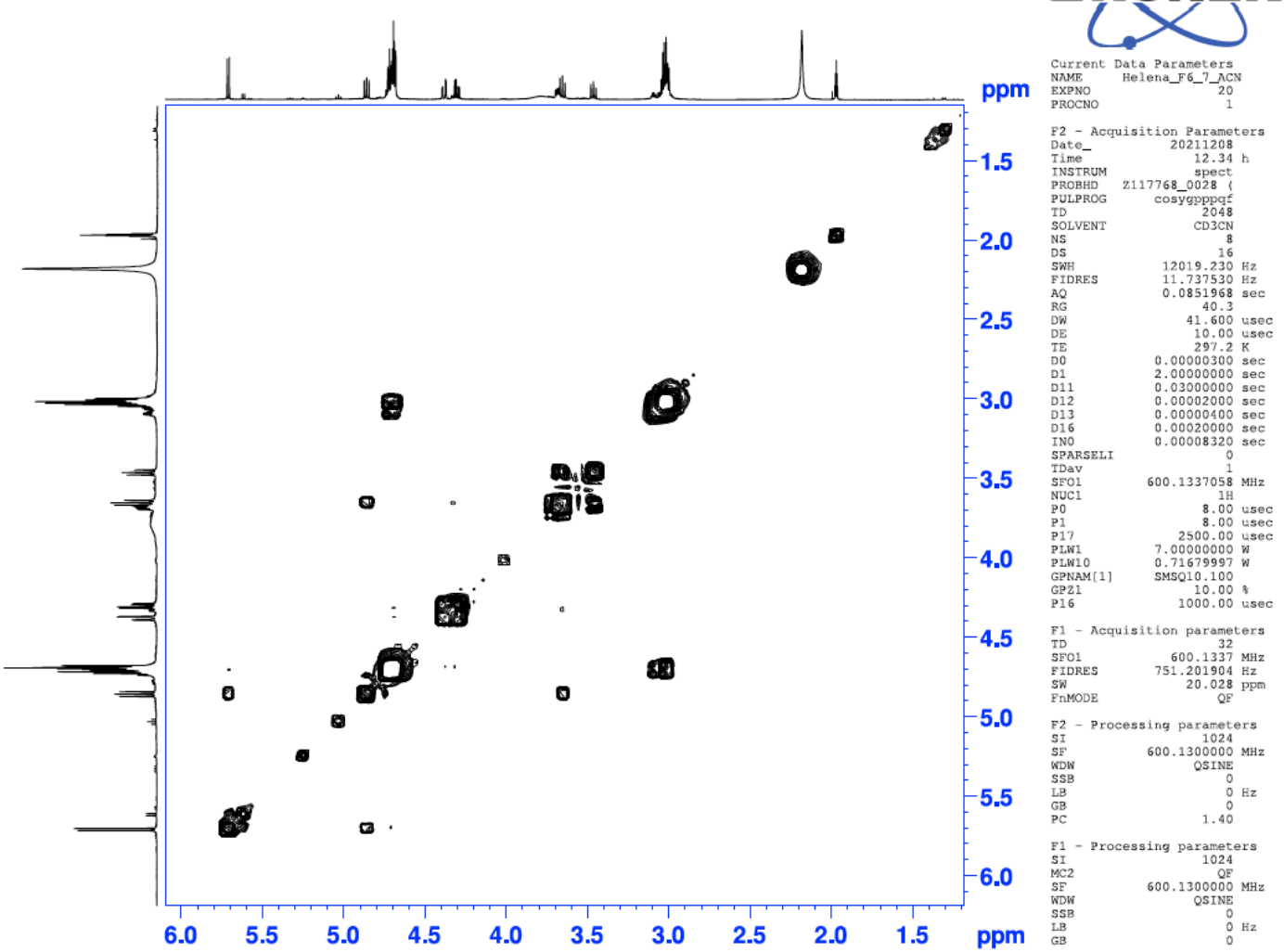
Supplementary Figure 16 – TOCSY 1D NMR (600 MHz, CD_3CN-d_3) spectrum of compound **2**. The structure of compond **2** showing all the spin systems is displayed. Arrows indicate the signals irradiated.



Supplementary Figure 17 – NOESY 1D NMR (600 MHz, CD_3CN-d_3) spectrum of compound **2**. The structure of compond **2** showing spatial interactions is displayed. Arrows indicate the signal irradiated and signals with spatial interaction.

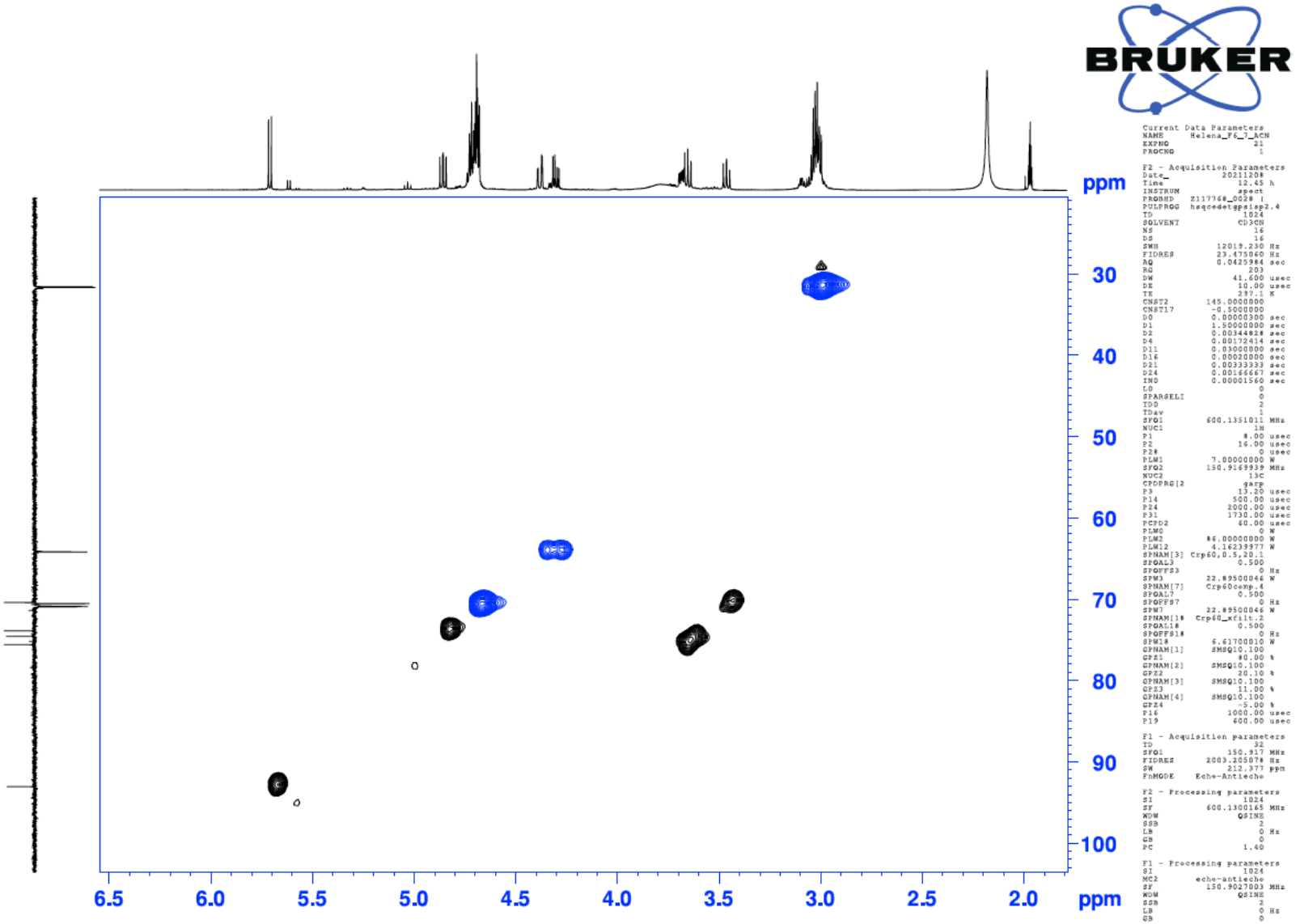


Supplementary Figure 18 – NOESY 1D NMR (600 MHz, CD_3CN-d_3) spectrum of compound **2**. The structure of compond **2** showing spatial interactions is displayed. Arrows indicate the signal irradiated and signals with spatial interaction.

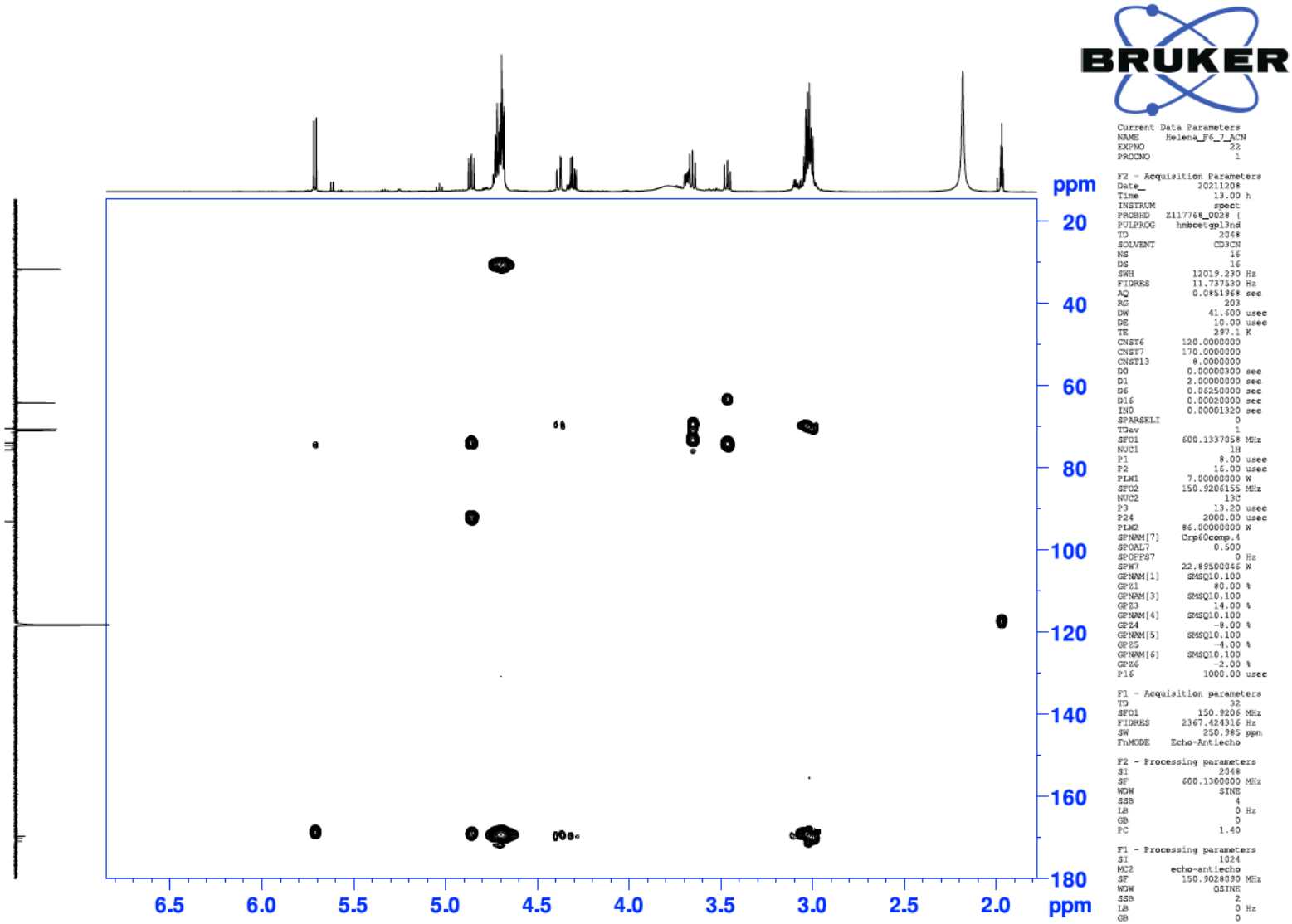



```
Current Data Parameters
NAME      Helena_F6_7_ACN
EXPNO                  20
PROCNO                  1

F2 - Acquisition Parameters
Date_            20211208
Time                12.34 h
INSTRUM             spect
PROBHD    Z117768_0028 (
PULPROG       cosygpppqf
TD                   2048
SOLVENT             CD3CN
NS                      8
DS                     16
SWH             12019.230 Hz
FIDRES          11.737530 Hz
AQ              0.0851968 sec
RG                   40.3
DW                 41.600 usec
DE                 10.00 usec
TE                 297.2 K
D0              0.00000300 sec
D1              2.00000000 sec
D11             0.03000000 sec
D12             0.00002000 sec
D13             0.00000400 sec
D16             0.00020000 sec
IN0             0.00008320 sec
SPARSELI                0
TDav                    1
SFO1          600.1337058 MHz
NUC1                   1H
P0                  8.00 usec
P1                  8.00 usec
P17              2500.00 usec
PLW1            7.00000000 W
PLW10           0.71679997 W
GPNAM[1]       SMSQ10.100
GPZ1               10.00 %
P16              1000.00 usec

F1 - Acquisition parameters
TD                     32
SFO1            600.1337 MHz
FIDRES          751.201904 Hz
SW                20.028 ppm
FnMODE                 QF

F2 - Processing parameters
SI                   1024
SF            600.1300000 MHz
WDW                 QSINE
SSB                     0
LB                      0 Hz
GB                      0
PC                   1.40

F1 - Processing parameters
SI                   1024
MC2                    QF
SF            600.1300000 MHz
WDW                 QSINE
SSB                     0
LB                      0 Hz
GB                      0
```

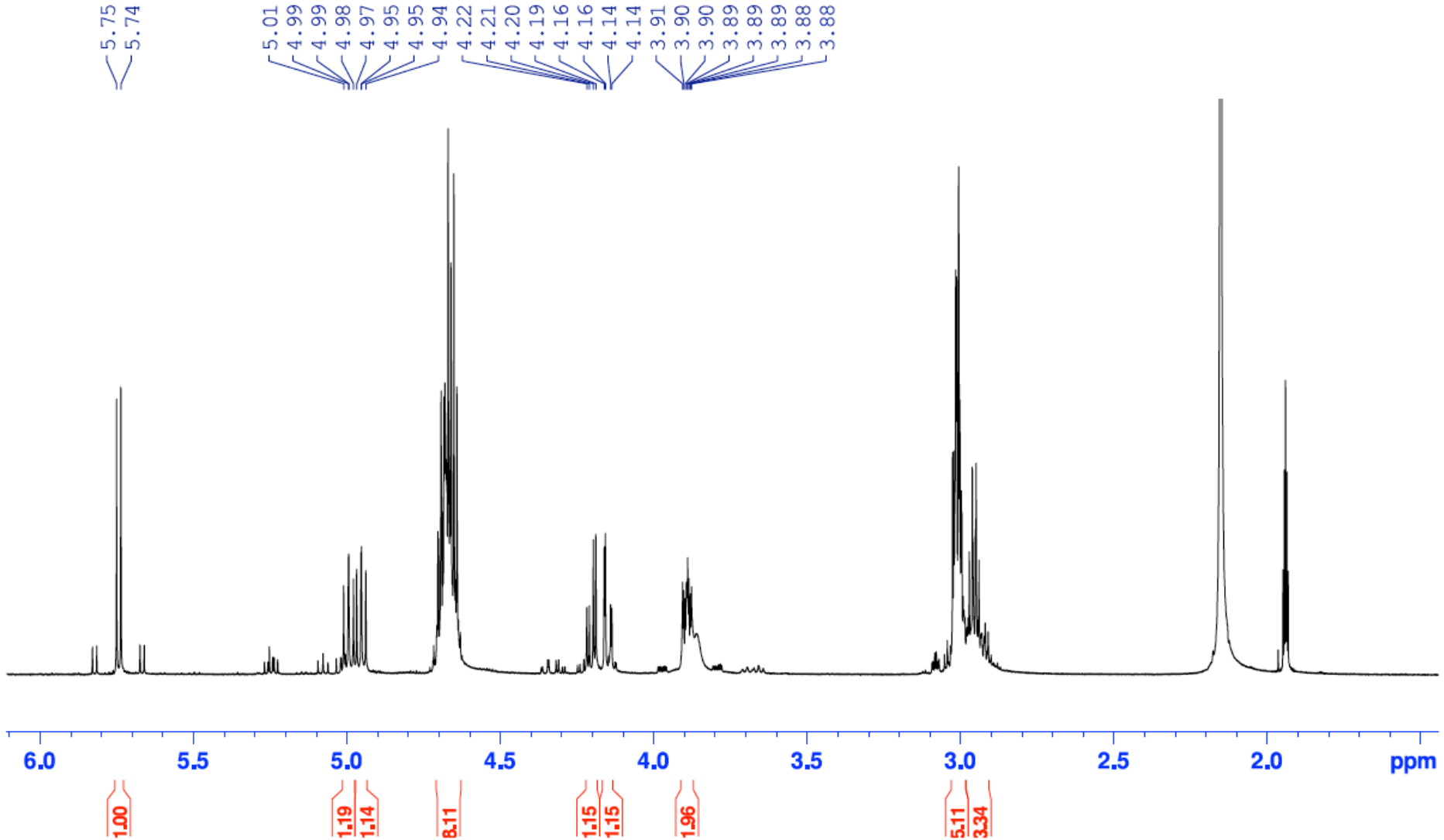
Supplementary Figure 19 – COSY NMR spectrum (600 MHz, CD_3CN-d_3) of compound **2**.



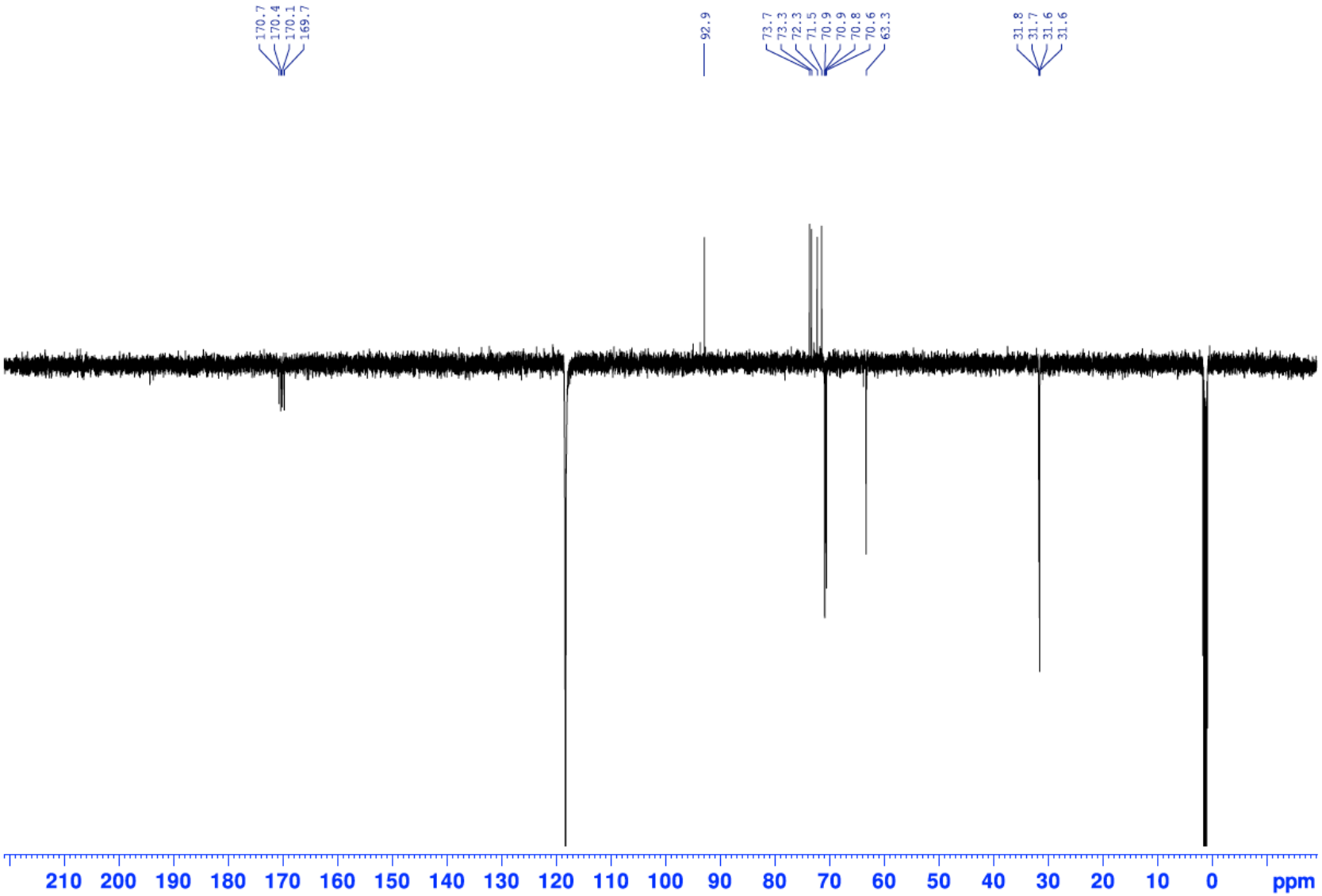
Supplementary Figure 20 – HSQC NMR (600 MHz, CD_3CN-d_3) spectrum of compound **2**.



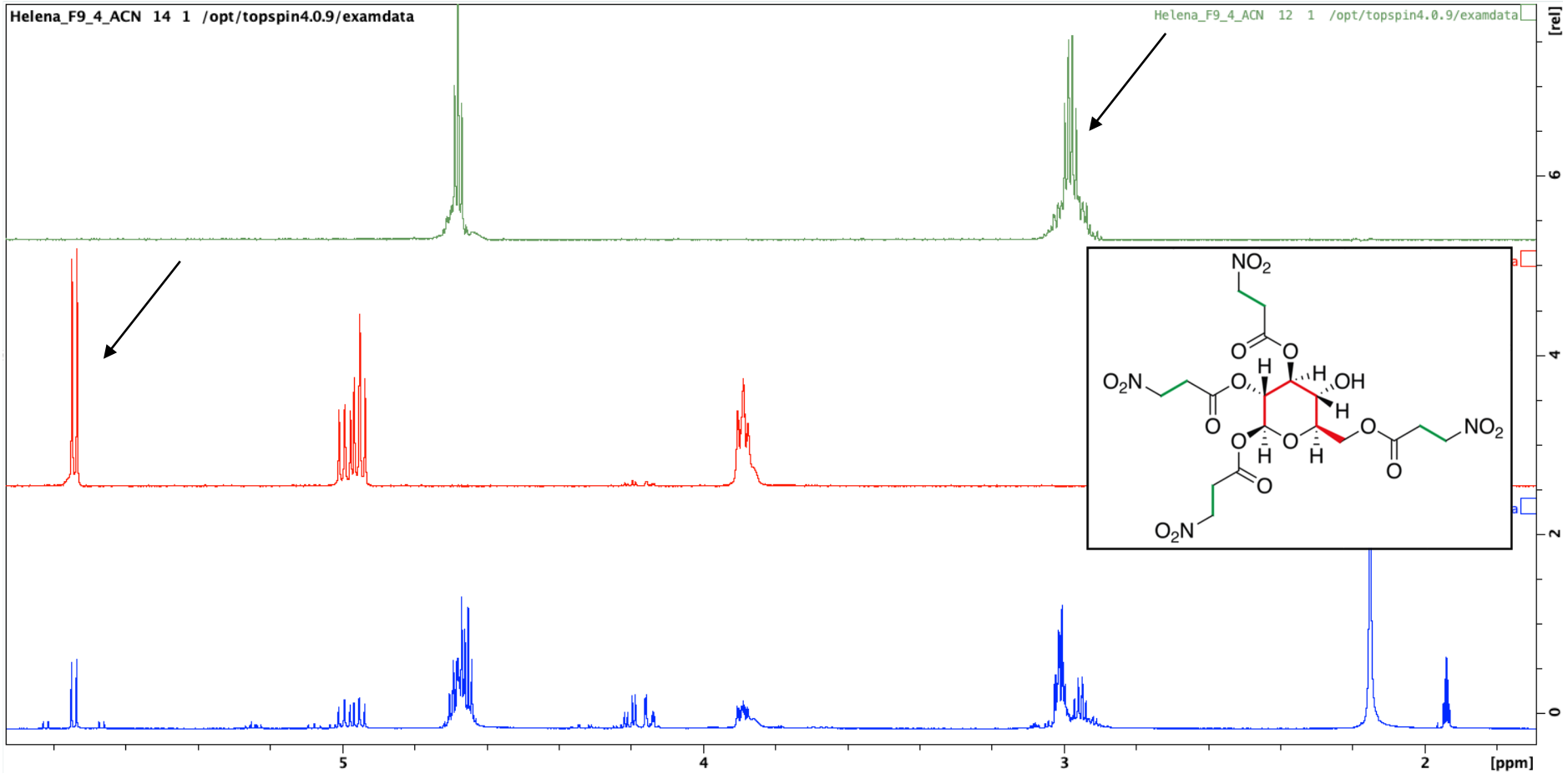
Supplementary Figure 21 – HMBC NMR (600 MHz, CD_3CN-d_3) spectrum of compound **2**.



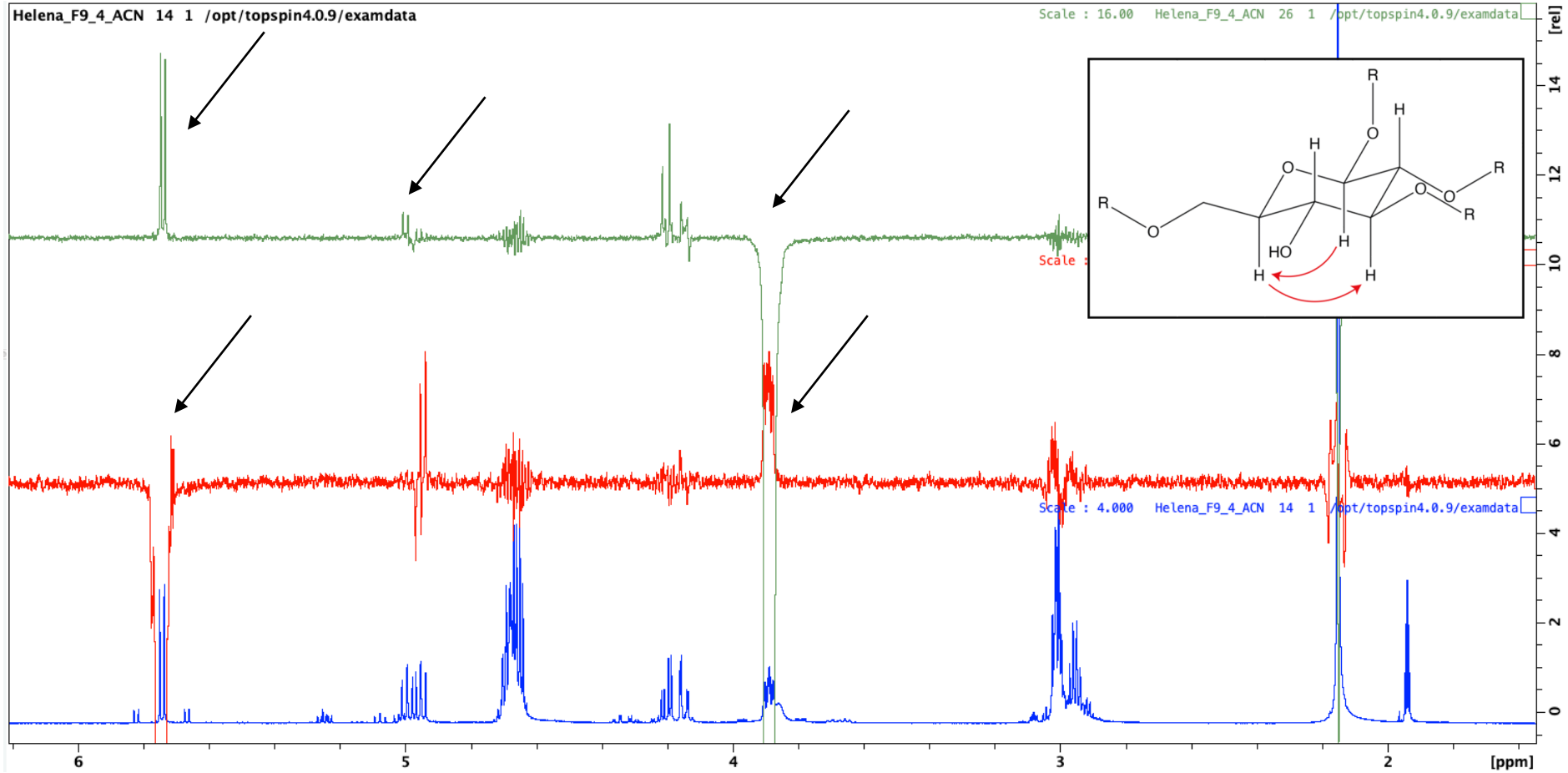
Supplementary Figure 22 – ^{1}H NMR (600 MHz, CD_3CN-d_3) spectrum of compound **3**.



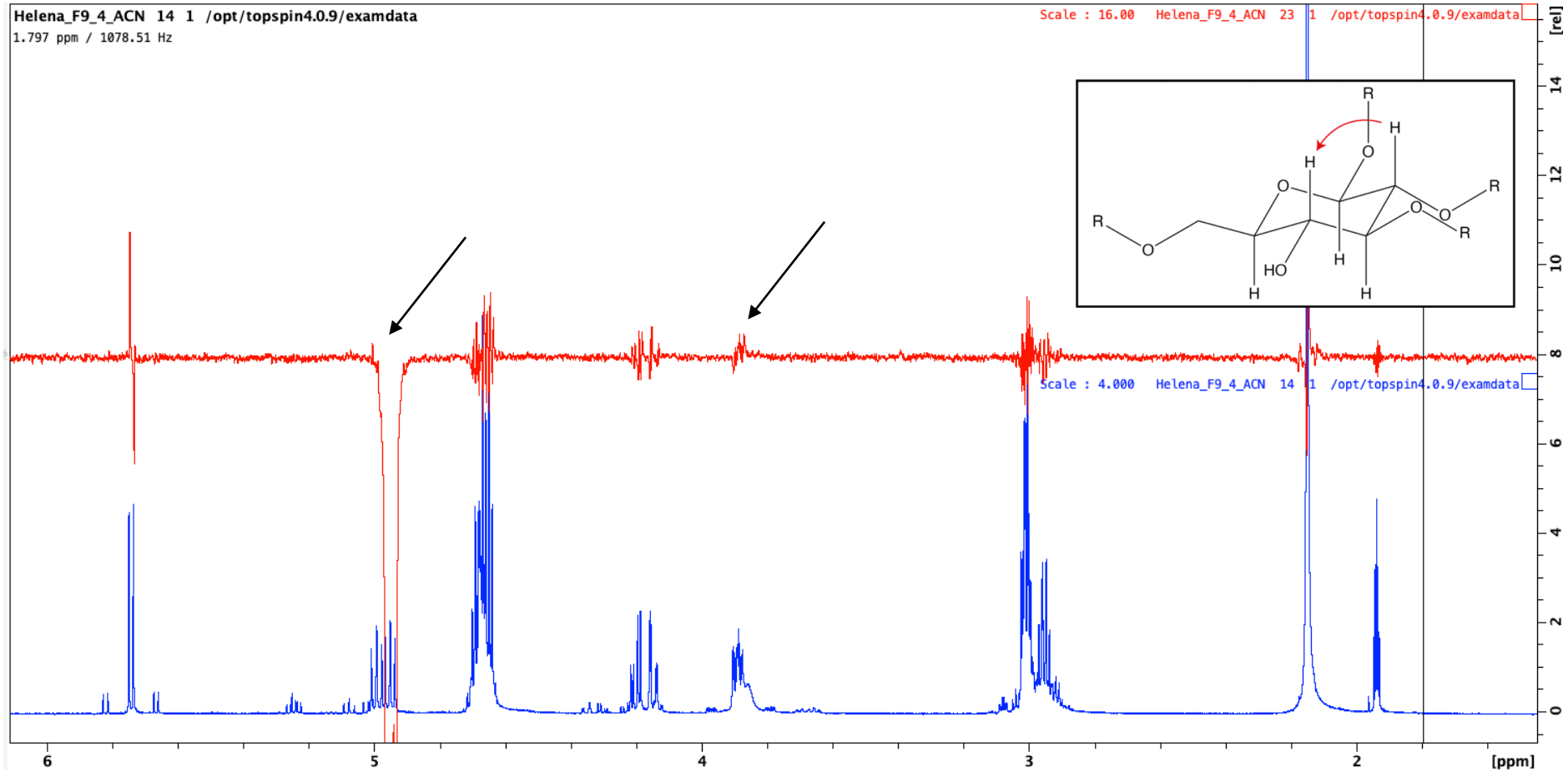
Supplementary Figure 23 – DEPTQ NMR (150 MHz, CD_3CN-d_3) spectrum of compound **3**.



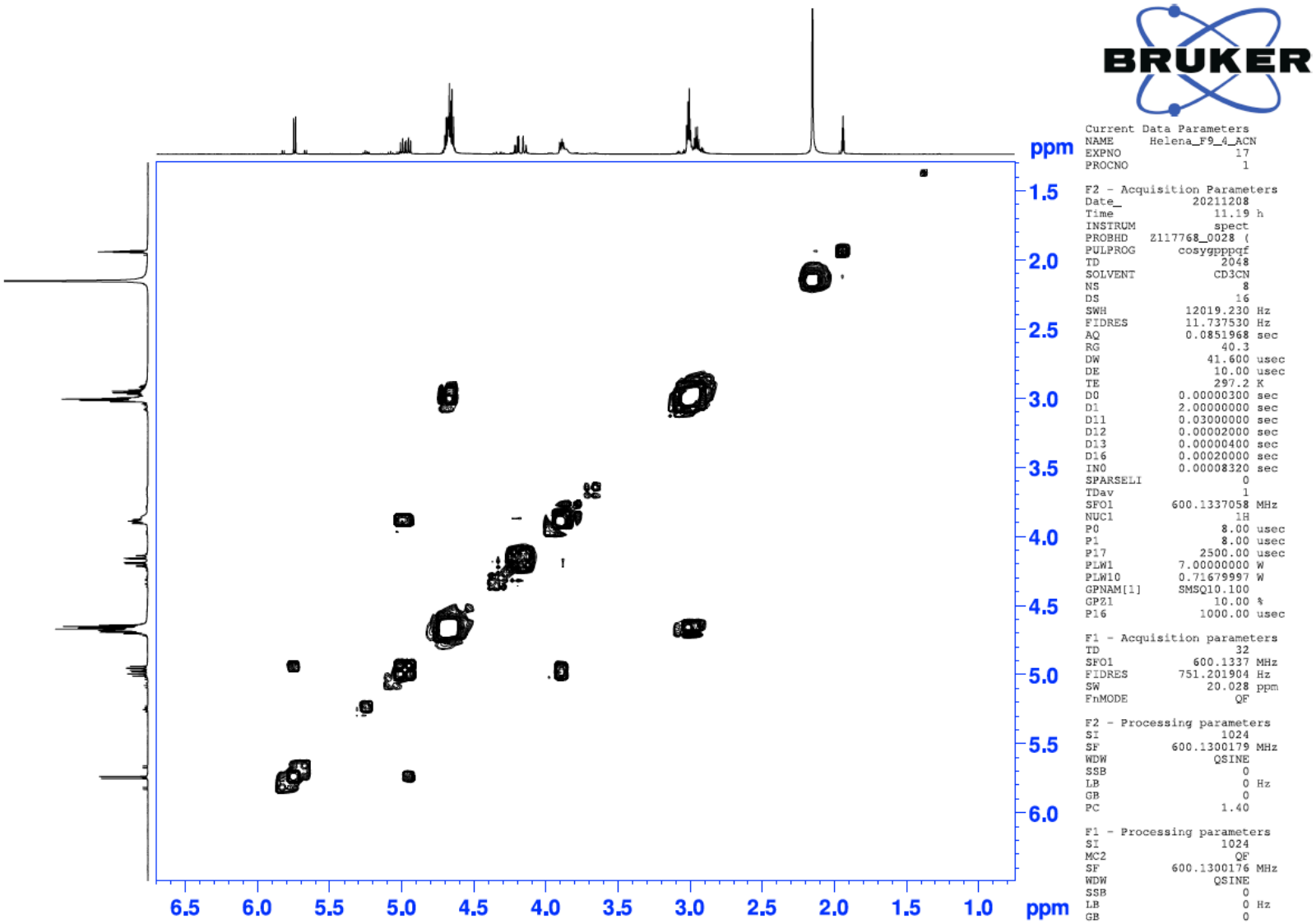
Supplementary Figure 24 – TOCSY 1D NMR (600 MHz, CD_3CN-d_3) spectrum of compound **3**. The structure of compond **3** showing all the spin systems is displayed. Arrows indicate the signals irradiated.



Supplementary Figure 25 – NOESY 1D NMR (600 MHz, CD_3CN-d_3) spectrum of compound **3**. The structure of compond **3** showing spatial interactions is displayed. Arrows indicate the signal irradiated and signals with spatial interaction.



Supplementary Figure 26 – NOESY 1D NMR (600 MHz, CD_3CN-d_3) spectrum of compound **3**. The structure of compond **3** showing spatial interactions is displayed. Arrows indicate the signal irradiated and signals with spatial interaction.



Current Data Parameters
NAME Helena_F9_4_ACN
EXPNO 17
PROCNO 1

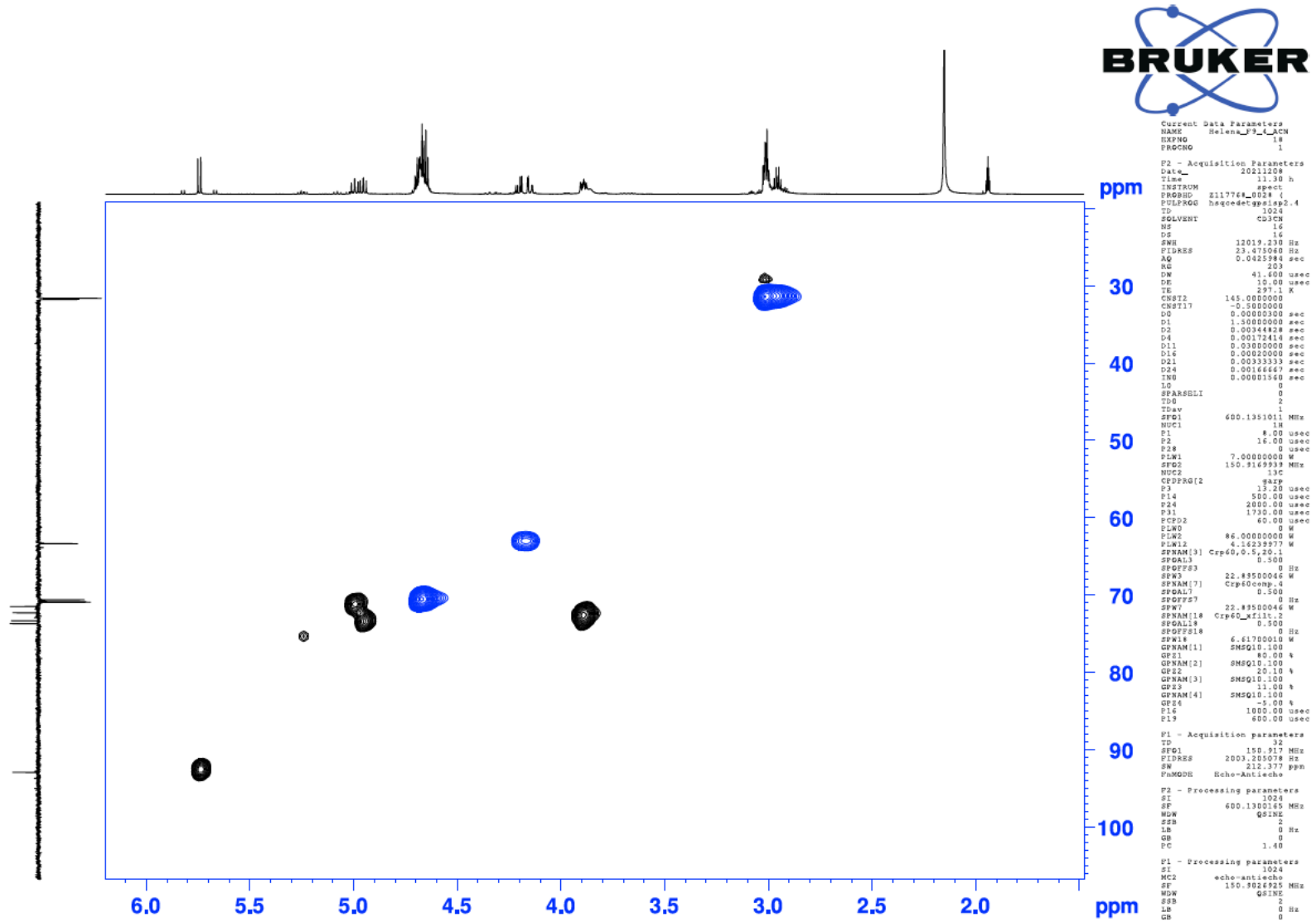
F2 - Acquisition Parameters
Date_ 20211208
Time 11.19 h
INSTRUM spect
PROBHD Z117768_0028 (
PULPROG cosygpppqf
TD 2048
SOLVENT CD3CN
NS 8
DS 16
SWH 12019.230 Hz
FIDRES 11.737530 Hz
AQ 0.0851968 sec
RG 40.3
DW 41.600 usec
DE 10.00 usec
TE 297.2 K
D0 0.00000300 sec
D1 2.00000000 sec
D11 0.03000000 sec
D12 0.00002000 sec
D13 0.00000400 sec
D16 0.00020000 sec
IN0 0.00008320 sec
SPARSELI 0
TDav 1
SFO1 600.1337058 MHz
NUC1 1H
P0 8.00 usec
P1 8.00 usec
P17 2500.00 usec
PLW1 7.00000000 W
PLW10 0.71679997 W
GPNAM[1] SMSQ10.100
GPZ1 10.00 %
P16 1000.00 usec

F1 - Acquisition parameters
TD 32
SFO1 600.1337 MHz
FIDRES 751.201904 Hz
SW 20.028 ppm
FnMODE QF

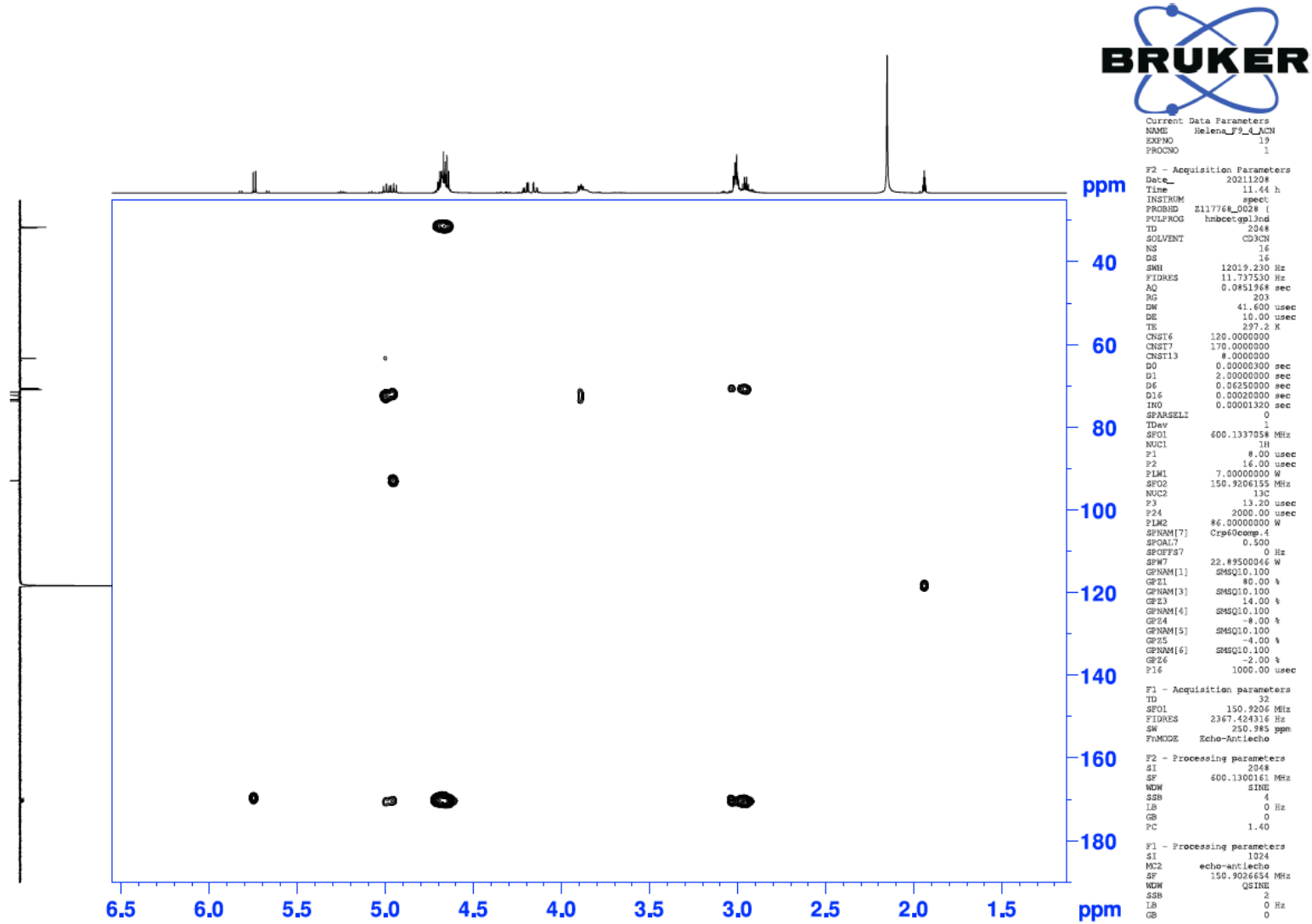
F2 - Processing parameters
SI 1024
SF 600.1300179 MHz
WDW QSINE
SSB 0
LB 0 Hz
GB 0
PC 1.40

F1 - Processing parameters
SI 1024
MC2 QF
SF 600.1300176 MHz
WDW QSINE
SSB 0
LB 0 Hz
GB 0

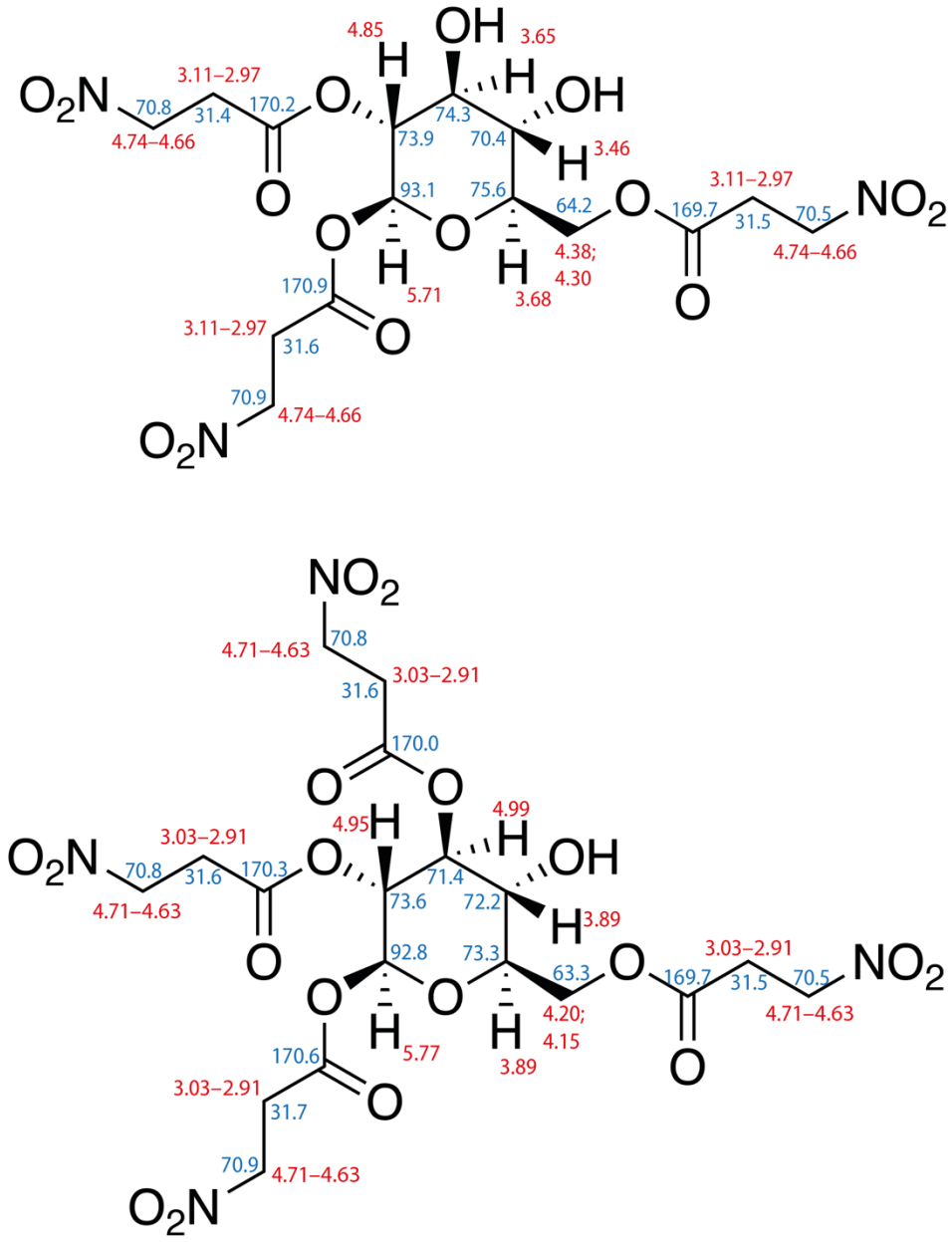
Supplementary Figure 27 – COSY NMR spectrum (600 MHz, CD_3CN-d_3) of compound **3**.



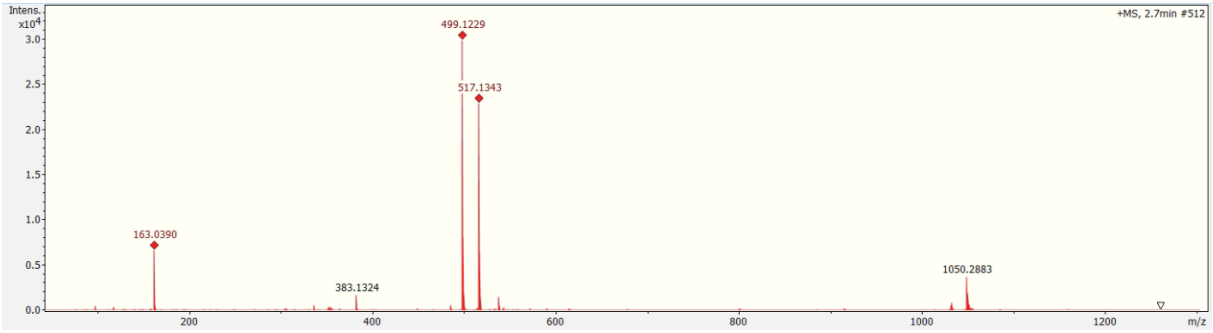
Supplementary Figure 28 – HSQC NMR spectrum (600 MHz, CD_3CN-d_3) of compound **3**.



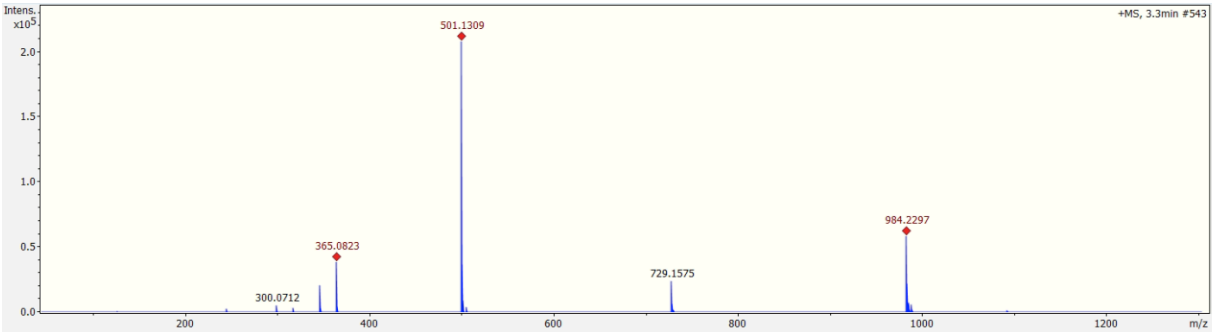
Supplementary Figure 29 – HMBC NMR spectrum (600 MHz, CD_3CN-d_3) of compound **3**.



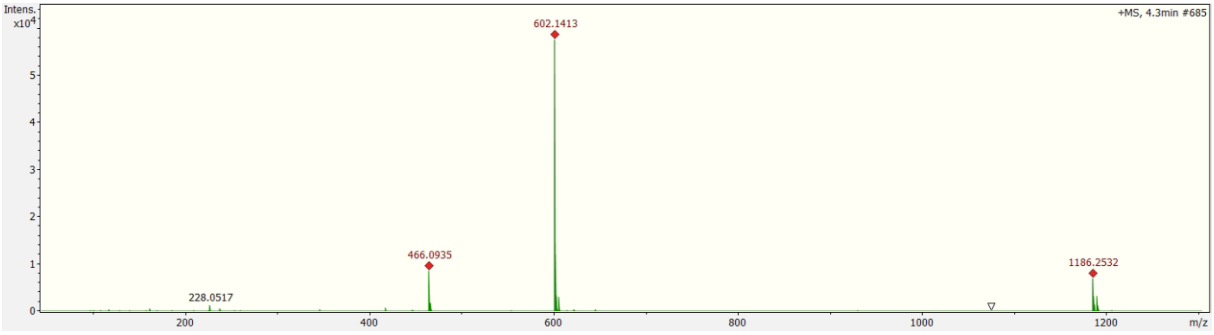
Supplementary Figure 30 – Comparison of the chemical shifts observed (by ^{1}H and DEPTQ) for compounds **2** (above) and **3** (below).



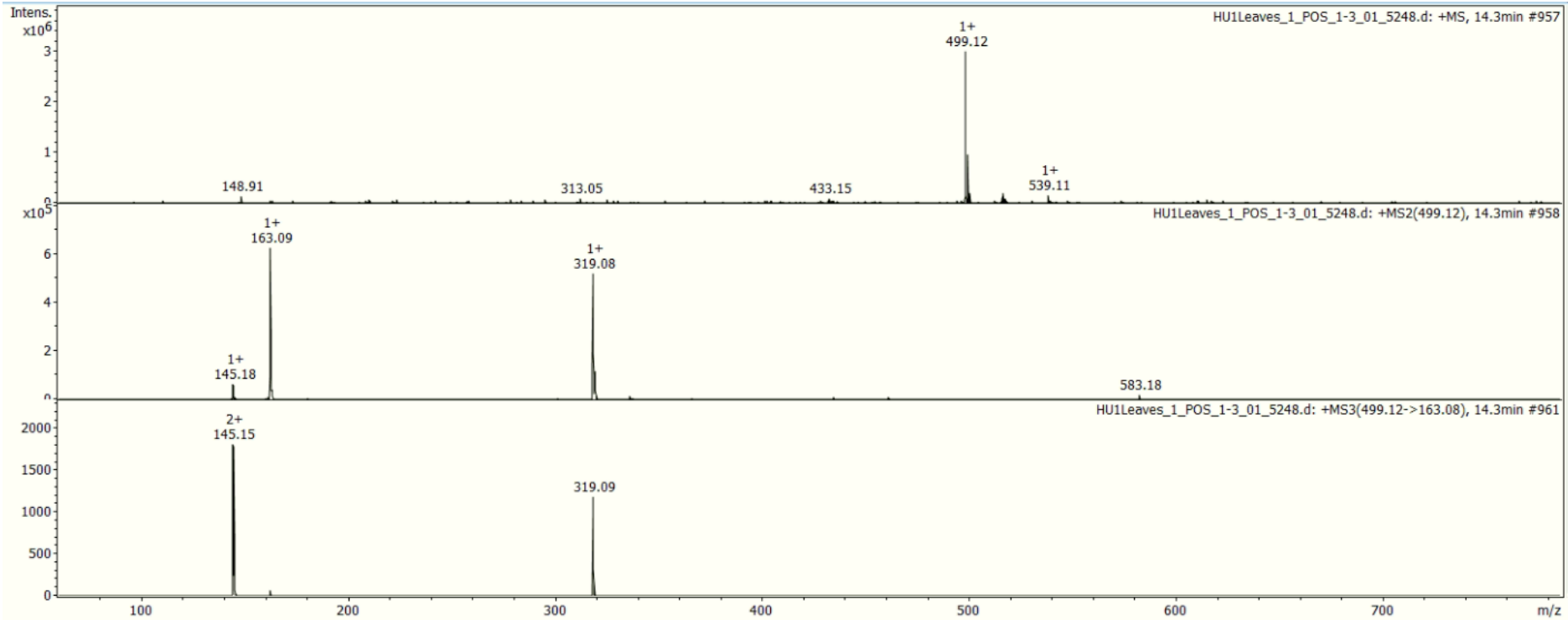
Supplementary Figure 31 – HRMS obtained for compound **1**.



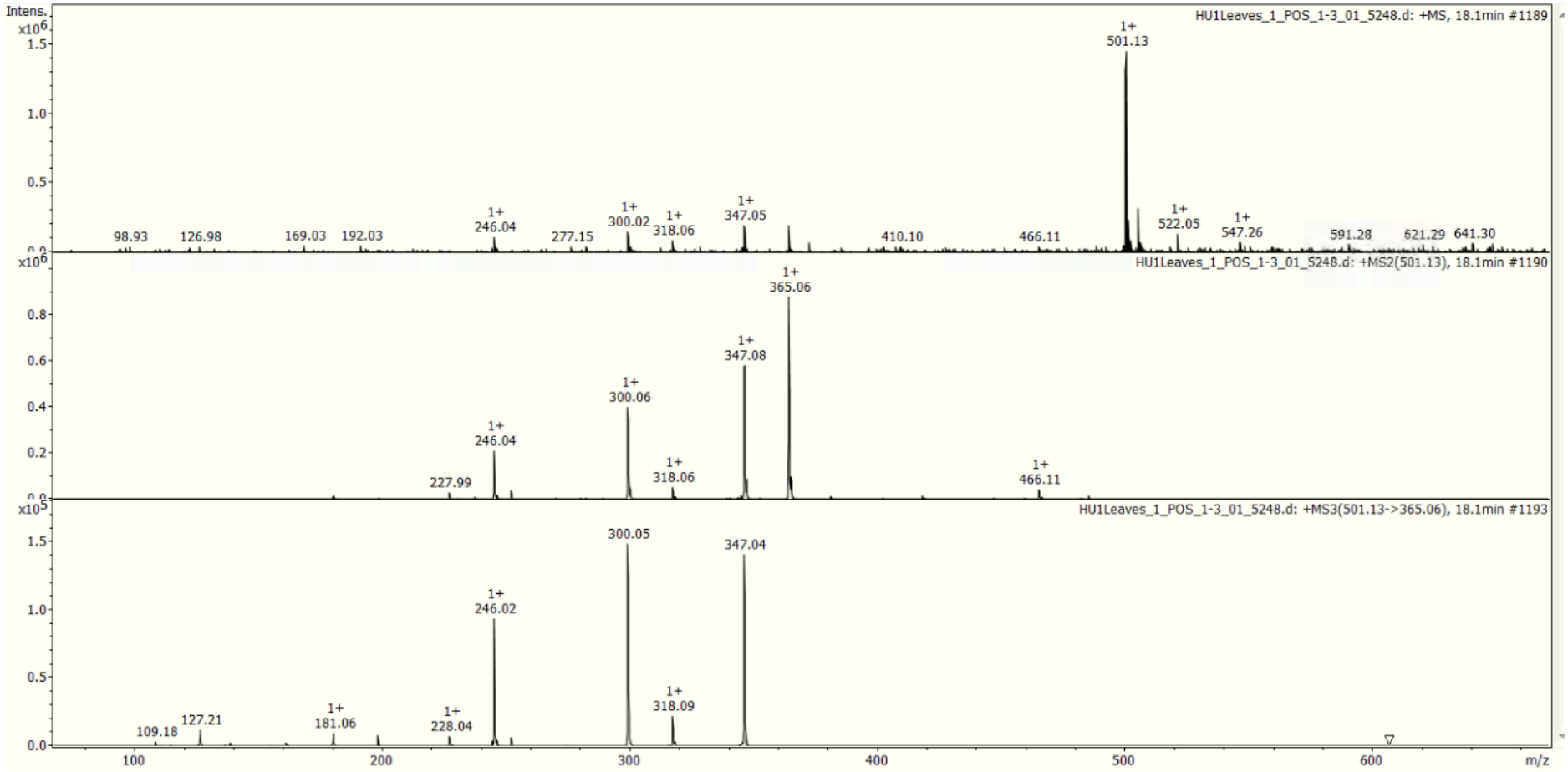
Supplementary Figure 32 – HRMS obtained for compound **2**.



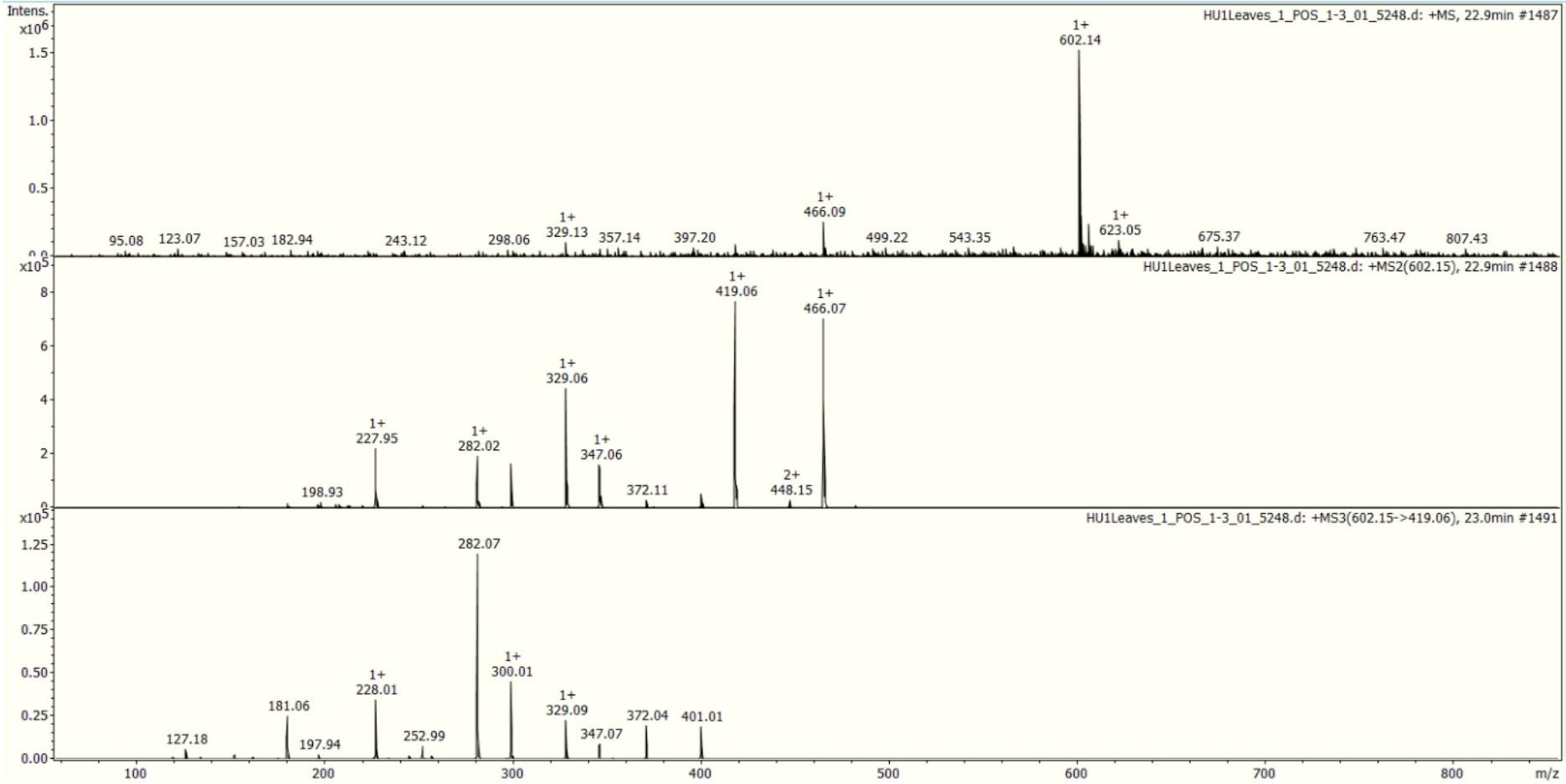
Supplementary Figure 33 – HRMS obtained for compound **3**.



Supplementary Figure 34 – Tandem mass spectra (MS^1, MS^2, and MS^3 – details are depicted in the figure) obtained for compound **1**.



Supplementary Figure 35 – Tandem mass spectra (MS^1, MS^2, and MS^3 – details are depicted in the figure) obtained for compound **2**.



Supplementary Figure 36 – Tandem mass spectra (MS^1, MS^2, and MS^3 – details are depicted in the figure) obtained for compound **3**.

Supplementary Box 3 – List of *Heteropterys umbellata* (HU) collections performed for this study. All vouchers are deposited in the herbarium of the Federal University of São Carlos (SPSC).

Collection code	Sample code	Collection sites	Collection dates	Description
H. Mannochio-Russo1	HU1Leaves HU1Stems	Jaboticatubas/MG (Serra do Cipó)	January 8th, 2021	Specimen with flowers
H. Mannochio-Russo2	HU2Leaves HU2 Stems	Gouveia/MG	January 9th, 2021	Specimen without flowers or fruits
H. Mannochio-Russo3	HU3Leaves HU3 Stems	Gouveia/MG	January 9th, 2021	Specimen without flowers or fruits
H. Mannochio-Russo4	HU4Leaves HU4 Stems	Pratânia/SP	January 11th, 2021	Specimen with fruits
H. Mannochio-Russo5	HU5Leaves HU5 Stems	Pratânia/SP	January 11th, 2021	Specimen with fruits
H. Mannochio-Russo6	HU6Leaves HU6 Stems	Pratânia/SP	January 11th, 2021	Specimen with fruits
H. Mannochio-Russo7	HU7Leaves HU7 Stems	Pratânia/SP	January 11th, 2021	Specimen with flowers

Source: elaborated by the Author (2022).

Supplementary Table 10 – Extraction yields obtained for *Heteropterys umbellata* leaves and stems in the seven collection points.

Sample code	Starting material (g)	Extraction yield	Sample code	Starting material (g)	Extraction yield
HU1Leaves	73.8	31%	**HU1Stems**	41.0	9%
HU2Leaves	21.0	29%	**HU2Stems**	24.0	7%
HU3Leaves	47.6	22%	**HU3Stems**	28.2	7%
HU4Leaves	54.9	21%	**HU4Stems**	31.9	7%
HU5Leaves	50.9	16%	**HU5Stems**	25.6	9%
HU6Leaves	26.4	28%	**HU6Stems**	16.1	8%
HU7Leaves	47.0	17%	**HU7Stems**	48.6	7%

Source: elaborated by the Author (2022).

Supplementary Table 11 – Metabolites annotated by HPLC-ESI-MS/MS and Molecular Network in the *Heteropterys umbellata* extracts (positive ionization mode).

#	Rt (min)	Precursor ion	MS/MS fragments (% abundance)	Metabolite
1	9.6	355 $[M+H]^+$	MS^2 (355): 163 (100) MS^3 (355→163): 145 (100), 135 (13), 117 (13)	Caffeoylquinic acid
2	11.3	449 $[M+H]^+$	MS^2 (449): 431 (73), 413 (22), 395 (24), 383 (100), 353 (53), 329 (41), 299 (6) MS^3 (449→383): 299 (100), 287 (6)	Luteolin-*C*-hexoside
3	11.8	449 $[M+H]^+$	MS^2 (449): 431 (80), 413 (43), 395 (12), 383 (51), 353 (14), 329 (23), 307 (38), 299 (7), 287 (100), 219 (29) MS^3 (449→287): 287 (100)	Luteolin-*C*-hexoside
4	12.2	595 $[M+H]^+$	MS^2 (595): 449 (100), 287 (33) MS^3 (595→449): 287 (100)	Luteolin-*O*-hexoside-deoxyhexoside
5	12.2	611 $[M+H]^+$	MS^2 (611): 465 (19), 303 (100) MS^3 (611→303): 303 (100), 257 (9), 229 (8)	Quercetin-*O*-hexoside-deoxyhexoside
6	12.5	433 $[M+H]^+$	MS2 (433): 415 (100), 397 (41), 379 (28), 367 (89), 337 (31), 313 (34), 283 (14) MS3 (433→415): 397 (62), 379 (26), 367 (82), 349 (8), 337 (100), 295 (24), 283 (15)	Apigenin-*C*-hexoside
7	12.8	465 $[M+H]^+$	MS^2 (465): 303 (100) MS^3 (465→303): 303 (100), 285 (7)	Quercetin-*O*-hexoside
8	12.9	449 $[M+H]^+$	MS^2 (449): 287 (100) MS^3 (449→287): 287 (100)	Luteolin-*O*-hexoside
9	14.3	517 $[M+H]^+$	MS^2 (517): 499 (100) MS^3 (517→499): 319 (94), 163 (100), 145 (6)	Dicaffeoylquinic acid
9	14.3	499 $[M-H_2O+H]^+$	MS^2 (499): 319 (100), 163 (90) MS^3 (499→319): 163 (100), 145 (5)	Dicaffeoylquinic acid

Source: elaborated by the Author (2022).

Supplementary Table 12 – ^{1}H (600 MHz) and ^{13}C (150 MHz) chemical shifts obtained for compound **1**, and comparison to the literature.

Position	Experimental, DMSO-*d*$_6$		Literature (ALKHATIB *et al.*, 2009), CD_3OD-*d*$_4$	
	δH[a] (600 MHz) (m[b]; J[c])	δC[a] (150 MHz)	δH[a] (300 MHz) (m[b]; J[c])	δC[a] (75 MHz)
1	-	77.9	-	81.0
2ax	2.15 dd (13.2, 3.4)	35.2	2.30 dd (14.0, 3.5)	36.5
2eq	1.97 dd (13.2, 8.5)	35.2	2.11 ddd (14.0, 7.0,1.5)	36.5
3	5.20 dt (8.5, 3.3)	71.4	5.54 dt (7.0, 3.5)	71.7
4	3.84 dd (11.8, 4.4)	68.0	3.91 dd (10.0, 3.5)	72.2
5	5.11 br d (3.3)	71.1	5.39 ddd (10.0, 7.5, 4.2)	73.6
6ax	1.97 dd (13.2, 8.5)	35.2	2.18 dd (14.0, 7.5)	39.9
6eq	2.15 dd (13.2, 3.3)	35.2	2.25 ddd (14.0, 4.2, 1.5)	39.9
COOH	-	175.8	-	178.1
α'	6.25 d (15.9)	115.2	6.42 d (16.0)	115.1
β'	7.49 d (15.9)	145.6	7.61 d (16.0)	146.4
1'	-	126.0	-	128.1
2'	7.06 d (1.9)	115.2	7.08 d (2.0)	114.2
3'	-	148.9	-	145.5
4'	-	146.1	-	148.3
5'	6.78 d (6.6)	116.3	6.79 d (9.0)	115.6
6'	7.00 dd (6.0, 1.9)	121.9	6.96 dd (9.0, 2.0)	122.1
CO	-	166.6	-	168.1
α''	6.16 d (15.9)	114.6	6.31 d (16.0)	115.1
β''	7.45 d (15.9)	145.2	7.59 d (16.0)	146.4
1''	-	126.1	-	128.1
2''	7.04 d (1.9)	115.3	7.07 d (2.0)	114.3
3''	-	148.7	-	148.3
4''	-	146.1	-	148.3
5''	6.77 d (6.5)	116.2	6.78 d (9.0)	115.6
6''	6.99 dd (6.6, 1.9)	121.6	6.95 d (9.0, 2.0)	122.1
CO	-	166.1	-	167.7

[a]Chemical shifts in ppm; [b]Multiplicity; [c]Coupling constant (J), in Hz.

Source: elaborated by the Author (2022).

Supplementary Table 13 – ^{1}H (600 MHz) and ^{13}C (150 MHz) chemical shifts obtained for compound **2**, and comparison to the literature.

	Experimental, CD_3CN-*d*$_3$		Literature (MAJAK et al., 1992), Acetone-*d*$_6$	
Position	**δH[a] (600 MHz) (m[b]; J[c])**	**δC[a] (150 MHz)**	**δH[a] (400 MHz) (m[b]; J[c])**	**δC[a] (100 MHz)**
1	5.71 d (8.4)	93.1	5.74 d (8.5)	93.3
2	4.83 dd (9.8, 8.4)	73.9	4.92 dd (8.5, 9.5)	74.3
3	3.65 dd (10.0, 9.2)	74.5	3.78 t (9.5)	75.1
4	3.46 dd (10.0, 9.3)	70.4	3.54 t (9.5)	70.7
5	3.68 ddd (9.9, 5.5, 2.2)	75.6	3.78 m	75.8
6a	4.38 dd (12.2, 2.2)	64.2	4.45 dd (12.0, 2.0)	64.5
6b	4.30 dd (12.2, 5.5)	64.2	4.32 dd (12.0, 5.8)	64.5
COOR (1)	-	170.9	-	170.7
COOR (2)	-	170.2	-	169.9
COOR (3)	-	169.7	-	169.4
2'	3.11 – 2.97 m	31.4 31.5 31.6	3.10 m	3 x 31.5 (overlapped)
3'	4.74 – 4.66 m	70.5 70.8 70.9	4.80	70.4 70.7 71.0

[a]Chemical shifts in ppm; [b]Multiplicity; [c]Coupling constant (J), in Hz.

Source: elaborated by the Author (2022).

Supplementary Table 14 – ^{1}H (600 MHz) and ^{13}C (150 MHz) chemical shifts obtained for compound **3**, and comparison to the literature.

	Experimental, CD_3CN-d_3		Literature (MAJAK *et al.*, 1994), Acetone-d_6	
Position	**δH[a] (600 MHz) (m[b]; J[c])**	**δC[a] (150 MHz)**	**δH[a] (400 MHz) (m[b]; J[c])**	**δC[a] (100 MHz)**
1	5.77 d (8.4)	92.8	5.94 d (8.3)	92.9
2	4.95 dd (10.1, 8.4)	73.6	5.05 dd (9.8, 8.3)	72.0
3	4.99 dd (10.1, 9.3)	71.4	5.35 t (9.8)	76.2
4	3.89 m	72.2	3.83 t (9.8)	68.8
5	3.89 ddd (10.2, 4.7, 2.2)	73.3	3.98 ddd (9.8, 5.2, 2.2)	75.5
6a	4.15 dd (12.5, 2.2)	63.3	4.45 dd (12.1, 2.2)	64.0
6b	4.20 dd (12.5, 4.7)	63.3	4.38 dd (12.1, 5.2)	64.0
COOR (1)	-	170.6	-	170.6
COOR (2)	-	170.3	-	169.8
COOR (3)	-	170.0	-	169.8
COOR (4)	-	169.7	-	169.2
2'	3.03 – 2.91	31.7 31.6 31.6 31.5	3.0 – 3.15 m	31.6 31.5 31.5 31.4
3'	4.71 – 4.63	70.9 70.8 70.8 70.5	4.80 m	70.8 70.7 70.6 70.4

[a]Chemical shifts in ppm; [b]Multiplicity; [c]Coupling constant (J), in Hz.

Source: elaborated by the Author (2022).

Supplementary Table 15 – Results obtained from the quantification of compounds **1**–**3** in *Heteropterys umbellata* leaves and stems extracts.

Sample code	1 (µg mL^{-1})	2 (µg mL^{-1})	3 (µg mL^{-1})	Sample code	1 (µg mL^{-1})	2 (µg mL^{-1})	3 (µg mL^{-1})
HU1Leaves	38.61	31.18	28.95	**HU1Stems**	58.77	15.75	16.69
	35.27	31.87	28.13		56.59	14.24	14.24
	35.83	32.14	30.38		63.92	14.34	15.18
HU2Leaves	61.88	11.07	13.08	**HU2Stems**	61.95	13.51	17.03
	65.64	12.38	13.25		58.52	12.66	16.81
	65.83	11.01	13.62		61.25	14.14	17.19
HU3Leaves	53.81	12.13	15.06	**HU3Stems**	42.33	9.92	15.37
	54.16	12.88	15.80		38.30	9.67	13.86
	46.24	12.27	15.35		38.60	9.42	14.52
HU4Leaves	13.53	0.00	0.00	**HU4Stems**	69.77	5.65	7.55
	16.37	0.00	0.00		63.06	5.44	7.59
	15.74	0.00	0.00		74.61	5.50	7.67
HU5Leaves	37.61	0.00	0.00	**HU5Stems**	93.04	4.70	7.39
	35.56	0.00	0.00		89.86	4.82	7.47
	33.74	0.00	0.00		85.62	4.88	7.43
HU6Leaves	28.04	7.68	9.20	**HU6Stems**	78.46	12.28	10.89
	31.90	7.91	9.20		77.93	12.25	10.67
	31.25	8.01	9.34		68.85	12.30	10.79
HU7Leaves	22.21	10.11	17.02	**HU7Stems**	59.44	12.49	14.54
	22.25	9.83	17.78		69.08	13.32	17.17
	22.67	10.03	18.34		63.63	14.06	15.90

Source: elaborated by the Author (2022).