UNIVERSIDADE ESTADUAL DO CENTRO-OESTE

METABOLOMIC STUDY OF EXTRACTS OF BRAZILIAN GEOPROPOLIS BY UHPLC-HRMS (ORBITRAP), ^{1}H AND ^{13}C NMR, AND GC-MS

JOÃO FÁBIO TURCO

GUARAPUAVA

2023

UNIVERSIDADE ESTADUAL DO CENTRO-OESTE DO PARANÁ – UNICENTRO

PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA – PPGQ/DOUTORADO

Estudo Metabolômico de Extratos de Geoprópolis Brasileira por CLUP-EMAR (Orbitrap), ^{1}H e ^{13}C RMN e CG-EM

CANDIDATO: Mr. JOÃO FÁBIO TURCO

ORIENTADORA: Prof.ª Dr.ª YOHANDRA R. TORRES

CO-ORITENTADOR: Dr. JOÃO B. MOKOCHINSKI

Trabalho apresentado ao Colegiado de Pós-Graduação em Química, categoria Doutorado, da Universidade Estadual do Centro-Oeste, a fim de lograr a qualificação necessária para obtenção do título de Doutor em Química.

GUARAPUAVA

2023

Catalogação na Publicação
Rede de Bibliotecas da Unicentro

Turco, João Fábio

T933m Metabolomic study of extracts of Brazilian geopropolis by UHPLC-HRMS (ORBITRAP), ^{1}H and ^{13}C NMR, and GC-MS / João Fábio Turco. – – Guarapuava, 2023.
xv, 162 f. : il. ; 28 cm

Tese (doutorado) - Universidade Estadual do Centro-Oeste, Programa de Pós-Graduação em Química. Área de Concentração: Química, 2023.

Orientadora: Yohandra R. Torres
Co-orientador: João B. Mokochinski
Banca Examinadora: Jaine Hortolan Luiz, Jan Schripsema, Daiane Finger Ferreira, Giancarlo Di Vaccari Botteselle

Bibliografia

1. Geopropolis. 2. Fingerprints. 3. Metabolomics. 4. Lipidomics. 5. Multivariate analysis. 6. Table of similarities. 7. UHPLC-HRMS. 8. Orbitrap. 9. GC-MS. I. Title. II. Postgraduate Program in Chemistry.

CDD 540

JOÃO FÁBIO TURCO

METABOLOMIC STUDY OF GEOPROPOLIS OF BRAZILIAN STINGLESS BEES BY UHPLC-HRMS (ORBITRAP), ^{1}H AND ^{13}C NMR AND GC-MS

Tese apresentada à Universidade Estadual do Centro-Oeste, como parte das exigências do Programa de Pós-Graduação em Química, para a obtenção do título de Doutor.

Aprovado em 30 de novembro de 2023.

Documento assinado digitalmente
gov.br
JAINE HONORATA HORTOLAN LUIZ
Data: 30/11/2023 17:53:06-0300
Verifique em https://validar.iti.gov.br

Profª. Drª. Jaine Hortolan Luiz (UNIFAL-MG)

Documento assinado digitalmente
gov.br
JAN SCHRIPSEMA
Data: 30/11/2023 23:49:31-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Jan Schripsema (UENF – RJ)

Documento assinado digitalmente
gov.br
DAIANE FINGER FERREIRA
Data: 05/12/2023 14:55:25-0300
Verifique em https://validar.iti.gov.br

Profª. Drª. Daiane Finger Ferreira (UNICENTRO)

Documento assinado digitalmente
gov.br
GIANCARLO DI VACCARI BOTTESELLE
Data: 06/12/2023 09:27:41-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Giancarlo Di Vaccari Botteselle (UNICENTRO)

Documento assinado digitalmente
gov.br
YOHANDRA REYES TORRES
Data: 30/11/2023 13:49:51-0300
Verifique em https://validar.iti.gov.br

Profª. Drª. Yohandra Reyes Torres (Orientadora)

GUARAPUAVA-PR

2023

Abstract

The stingless bees (*Meliponini*) make propolis by mixing resins, waxes, and oils from vegetal sources, with salivary secretions rich in enzymes. Small portions of soil particles are added to the mixture to give geopropolis the needed consistence to be used as part of many internal and external hive structures. In order to protect the colony from external pathogens, geopropolis present diverse pharmacological effects, especially antibiotic, antiviral, antifungal. Commonly in poor countries, communities of less assisted rural workers, and indigenous people benefit from these characteristics of geopropolis. Its chemical composition is complex and related to its geographic origin, which botanic sources are available, and the species of the stingless bees. In general, geopropolis is rich in flavonoids, phenolic compounds, phenylpropanoids, sugars, and lipids. Despite the importance of this natural product, studies on geopropolis are still scarce and focused on punctual characteristics, on few species, rather than on a broader approach on larger numbers of species. The objective of this study was to employ metabolomics and lipidomics approaches, through chromatographic hyphenated techniques, and NMR in order to investigate the composition and possible biomarkers of geopropolis. Firstly, ethanolic extracts of geopropolis (EEGs) from 14 species and two sub-species of Brazilian stingless bees, from five Brazilian States (Paraná, Pernambuco, Maranhão, São Paulo, and Sergipe) were made. Samples were given by local beekeepers. The EEGs were kept in freezer until transported, at room temperature, to the Proteomics and Metabolomics Facility at London Institute of Medical Sciences (Imperial College London, campus Hammersmith), United Kingdom. The mass fingerprints of the EEGs were acquired using a high-resolution mass spectrometer (HRMS, Orbitrap), firstly by direct flow injection technique (FIA) and analysed using multivariate analysis such as PCA and HA. Also, possible correlations between the mass fingerprint data and total

flavonoid content (TFC) and antioxidant capacity in terms of quercetin equivalent (DPPH) and ascorbic acid equivalents (VCEAC) were assessed by PLS regression. After FIA, EEGs were injected into an UHPLC-HRMS and subsequent partition with water and chloroform allowed the investigation of the $CHCl_3$ lipidic rich fraction. The water-soluble fraction was derivatised and analysed by GC-MS. In this way, the potentialities of each analytical technique were explored to make it possible a comprehensive analysis of geopropolis. Using GC-MS identification was based on spectral data comparison with the instrument internal library, also the calculated retention index. In the case of UHPLC-HRMS compound annotation was based on sequential mass spectrometry experiments and carried out by the Compound Discoverer V. 3.3 (Thermo) software, using specialised local and online libraries. The composition analysis by UHPLC-HRMS and GC-MS revealed the presence of flavonoids, sugars, esters, terpenes, phenolics, organic acids, and phenylpropanoids. The lipidomic analysis revealed the presence of fatty acids, fatty acyls, phenolic lipids, steroids, and resorcinols. Exploratory multivariate analysis indicated that bee species and genus strongly affect the chemical composition of geopropolis as well as the geographical origin. The extension of the later factor depends on the bee specie. On one hand, it was observed that geopropolis from *Tetragonisca angustula* have similar chemical composition regardless of geographical origin indicating that this bee gathers similar vegetal resins throughout Brazil to make its propolis. On the other hand, *Melipona quadrifasciata* and *Melipona marginata* seen to have generalist collection patterns as their geopropolis have a more variable composition depending on the geographical origin.

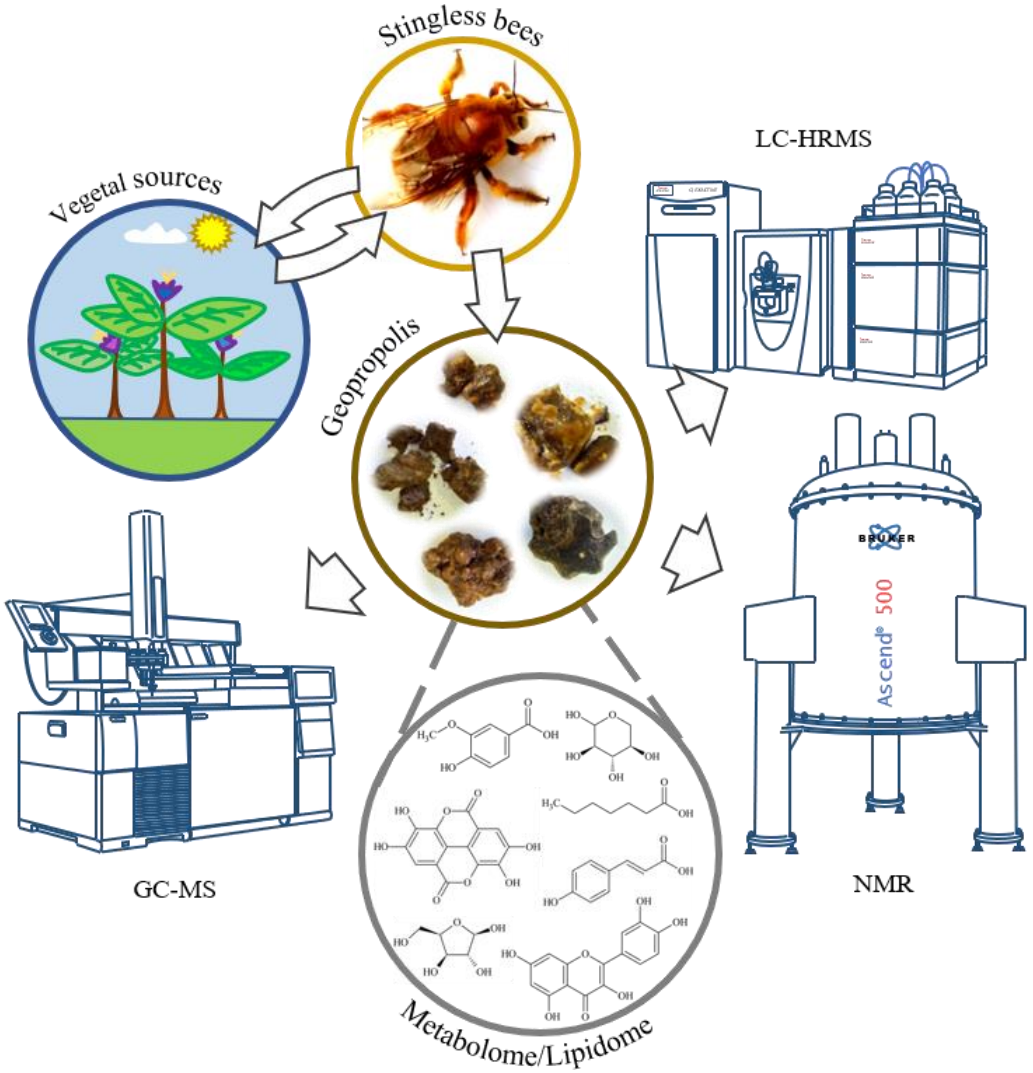
NMR analysis of the $CDCl_3$ extract (CEG) of geopropolis allowed confirmation of chemical classes already identified by chromatography hyphenated techniques and also

corroborated the importance of bee species and geographical origin on the chemical composition of geopropolis.

Finally, the application of PLS models to FIA mass fingerprints, using less than 3 latent variables, produced accurate models with low values for the errors of calibration and prediction (RMSEC < 0.79 $mg.g^{-1}$; RMSECV < 2.662 $mg.g^{-1}$; RMSEP < 1,0448 $mg.g^{-1}$). Additionally, acceptable determination coefficients ($0.6613 < R^2 < 0.8815$) indicated that mass fingerprints accurately estimate TFC and antioxidant capacity of geopropolis.

Key-words: Geopropolis, Fingerprints, Metabolomics, Lipidomics, Multivariate Analysis, Table of Similarities, UHPLC-HRMS, Orbitrap, GC-MS, ^{1}H NMR, ^{13}C NMR

Graphical Abstract



List of abbreviations

ANOVA	Analysis of variance
AOC	Antioxidant capacity
CEG	Chloroformic extract of geopropolis
CUPRAC	Cupric reducing antioxidant activity
EEG	Ethanolic extract of geopropolis
FID	Free induction decay
FIA-HRMS	Flow injection analysis-high resolution mass spectrometry
FT	Fourier transform
GC-MS	Gas chromatography coupled to mass spectrometry
HCA	Hierarchical clustering analysis
HMBC	Heteronuclear multiple bond correlation
HRMS	High resolution mass spectrometry
HSQC	Heteronuclear single quantum correlation
J	Spin-spin coupling
MS	Mass spectrometry
MSC	Multiplicative scatter correction
MW	Molecular weight
m/z	Mass-to-charge ratio
Neg	Negative
NMR	Nuclear magnetic resonance
PCA	Principal component analysis
PLS	Partial least square regression
Pos	Positive
QC	Quality control

RMSEC	Root mean square error of calibration
RMSECV	Root mean square error of cross validation
RMSEP	Root mean square error of prediction
RPD	Residual prediction deviation
RSD	Relative standard deviation
RI	Retention index
RT	Retention time
SEP	Standard error of prediction
SNV	**S**tandard normal variate
TFC	Total flavonoids content
TIC	Total ion chromatogram
TMS	Tetramethylsilane
UHPLC-HRMS	Ultra-high performance liquid chromatography coupled to high resolution mass spectrometry
V.	Software version
VCEAC	Vitamin C equivalent antioxidant capacity

List of figures

List of tables

Summary

Chapter 1 – Scope of the current study

1.1 Introduction

Geopropolis is a natural product elaborated by stingless bees, a tribe of bees' species classified as *Meliponini*, a branch of the *Apidae* family closely related to the honeybees (*Apini* tribe) and the bumble bees (*Bombini* tribe). Geopropolis presents pharmacological and biological effects, such as antiviral [1], antibacterial [2], antibiotic, antifungal, anti-inflammatory [3], and antioxidant [2]. Due to such relevant biological activities, geopropolis is used by local communities of rural workers, and indigenous communities in Brazil [4].

Despite its pharmacological importance, few studies have been conducted with the intention of elucidating and correlating the chemical composition of geopropolis with its therapeutic activities. Little is known about how geopropolis extracts act in the human body, as well as which substances are directly responsible for these biological activities of interest. In addition, to the best of our knowledge, a lipid profile of geopropolis has never been reported in the literature. Also noteworthy is the small number of toxicological and antimicrobial studies, considering that the use of geopropolis is widespread in the folk medicine of multiple countries [5].

To produce geopropolis, the stingless bees collect resins and saps from diverse parts of plants and trees and bring them to the nest. There, the resins are mechanically mixed with enzymes, waxes, and salivary secretions in order to produce propolis, in a process still little known by the researchers; portions of soils are then added to propolis [6]. Even though the soil is not commonly the main component of geopropolis, this feature is the reason for the use of the prefix *geo* (from the Greek, meaning earth in the sense of "ground or soil"), which distinguishes this material from ordinary propolis [7,

8]. Inevitably, geopropolis composition is directly affected by environmental conditions and biological factors, varying in many aspects such as colour and odour, depending on its geographical origin, stingless bees' species, available plants, among others factors [6, 8, 9].

Brazil is a megadiverse country and has continental dimensions, and such factors will directly interfere in the geopropolis composition and physical properties, thus the characteristics of this natural product might vary both qualitatively and quantitatively [3]. As an example, propolis from the common honey bee (*Apis mellifera*) produced in Brazil, can be currently classified in 12 types, whose composition is directly linked to its geographic origin [10]. In the same way, studies about geopropolis chemical characterisation are important in order to provide key-information about the composition of each type of geopropolis, from different geographic origins, species, and biomes. In addition, those information allow improvements in the process of standardisation and quality control of geopropolis; At the moment in Brazil there is still no specific legislation regulating technical aspects for the production and/or composition of geopropolis from stingless bees [11].

Metabolomics and lipidomics are essential approaches that generate comprehensive information about the composition of geopropolis; also allow the comparison, and grouping, even of a large number of geopropolis extracts, when associated with multivariate analysis/chemometrics. As a relevant example, these analyses may give insights about chemical markers related to the bees' species, geographical origins, etc. Metabolomics and lipidomics can also reveal relationships between geopropolis chemical composition and its biological features. In this approach, ultra-high performance liquid chromatography-high resolution mass spectrometry (UHPLC-HRMS), and gas chromatography-mass spectrometry (GC-MS), were

employed to simultaneously detect and identify a large number compounds during an exploratory analysis of the extracts of geopropolis. The joint approach of metabolomics and lipidomics allows the coverage of different classes of compounds, making possible the geopropolis chemical profile to be traced in a more complete way. It should be noted that the analysis of lipids in geopropolis samples has not yet been reported in the literature.

To complement the exploratory analysis, nuclear magnetic resonance (NMR) was employed to help in the elucidation of chemical structures, confirming previous analysis, as well as pointing out other compounds that were missed.

1.2 Geopropolis composition

Studies about the chemical composition of the Brazilian stingless bees geopropolis are relatively scarce in the literature. However, this natural product is considered a promising source of compounds of interest [12].

The chemical profile of geopropolis is complex and variable, with several compounds being identified every year. In a rapid look in the literature, geopropolis composition is linked to phenols, flavonoids, amino acids, fatty acids, and vitamins. In addition, minerals such as Si, Mn, Cu, Ca, Al, V, Ni, Zn and Cr have also been reported. Those components are mixed in a waxy phase of geopropolis, (30% on average), rich in balms, essential oils, terpenoids, and phenolic derivatives [8].

Bankova et al. [13] analysed the chemical composition of hydroalcoholic extracts of geopropolis from three species of Brazilian stingless bees: *Melipona compressipes*, *Melipona quadrifasciata anthidioides*, and *Tetragona clavipes*. The authors used GC-MS to identify more than 50 different compounds, most of them being terpenoids, flavonoids, sugars, and phenolic acids. Kujumgiev et al. [14] analysed ethanolic extracts of *M.*

compressipes, and *M. quadrifasciata anthidioides*, produced in the State of Paraná, also employing GC-MS. Similarly to Bankova et al. [13], the authors reported the presence of diterpenic acids (3%), aromatic acids (5%), glycoflavonoids (10%) and triterpenes (10%).

Dutra et al. [15] analysed the composition of the hydroalcoholic extract of geopropolis from *M. fasciculata* from the State of Maranhão (Brazil). The objective was to characterise the material from a pharmacognostic point of view, using benchtop chromatographic techniques. Dutra et al. [15] used thin layer chromatography (TLC) to carry out a comparative analysis of geopropolis extracts. The authors observed a predominance of phenolic compounds in the chemical profile of the geopropolis extracts from *M. fasciculata*, especially flavonoids, and phenylpropanoids. Triterpenes, diterpenes and saponins were also reported, while the absence of alkaloids was highlighted by the authors. In the same study Dutra et al. [15] quantified the total flavonoid content (TFC) of the ethanolic extracts of geopropolis of *M. fasciculata* using the spectrophotometric method; the results ranged from 0.17 to 2.6%, being according to the authors comparatively similar to the quantities of TFC normally observed for propolis of *A. mellifera* (common honeybee).

Souza et al. [16] carried out an exploratory study on geopropolis produced by the species *M. subnitida*, from the semi-arid region of the State of Paraíba (Brazil). The authors used liquid column chromatography (CLC) techniques to isolate two phenylpropanoids from the methanolic extract of *M. subnitida* geopropolis: 6-O-cinnamoyl-1-O-p-coumaroyl-β-D-glucopyranose and 6-O-*p*-coumaroyl-D-galactopyranose.

Cardozo et al. [17] studied the chemical variability of geopropolis from three species of native bees: *M. quadrifasciata quadrifasciata*, *M. marginata*, and *T. angustula* donated by meliponiculturists from the Prudentópolis region (State of Paraná, Brazil).

Samples were collected at different times of the year. The researchers used ultra-high performance liquid chromatography (UPLC) coupled to mass spectrometry (MS) to identify chemical components in the extracts. The identification of compounds was based on comparison with authentic standards, analysis of ESI(-)-MS/MS data in comparison with data already reported in the literature. Cardozo et al. [17] reported the presence of vanillin, vanillic acid, caffeic acid, coumaric acid, and terpenic acids.

Ruiz et al. [18] used spectrophotometric techniques to assess the phenolic content, and bioactivity of (geo)propolis of the stingless bees *M. beecheii*, from diverse parts of Southeastern Mexico region. The authors employed multivariate analysis (HCA, and PCA) to distinguish, and classify, (geo)propolis produced by *M. beecheii*, using the similarity in terms of total phenolic content, and *in vitro* bioactivity potential, as a marker for geographical origin. According to Ruiz et al. [18] this strategy is suitable to be used to establish regulations, marketing, and industrial applications of geopropolis of *M. beecheii* .

The chemical composition of propolis is varied and significantly dependent on the plants available nearby where the beehive is housed. Samples of propolis produced in different countries and regions have different compositions, as a result of the botanical diversity of each location [19]. In North America, Europe and West Asia, the predominant source of resins for propolis production is the exudate of poplar bud (Populus sp.). In South America, plant species of the genus Populus do not exist natively, however there is an enormous botanical diversity, which is used by bees to collect resins [20]. However, in countries like Brazil, with a predominantly tropical climate and high biodiversity, there are a large number of plant species available for the collection of resins, which are used for the production of propolis. As a result, Brazilian propolis and geopropolis have a highly complex chemical composition, depending on their region of origin [21].

Therefore, the standardization and quality control of these bee products are difficult tasks, however, extremely necessary for the effective use of these materials for therapeutic purposes [22].

1.3 Objectives

1.3.1 General Objectives

The main goals of this work were to investigate the metabolomic, and lipidomic, profiles of extracts of Brazilian geopropolis (EEGs), from different regions, species, and genera of stingless bees. To achieve these goals UHPLC-HRMS, GC-MS, FIA-HRMS, and NMR spectroscopy were employed to gather relevant chemical information about extracts of geopropolis, on which chemometric tools, and bioinformatics, were applied to identify compounds, classify samples, and discover potential biomarkers. Figure 1 shows a flowchart detailing the path followed during the studies.

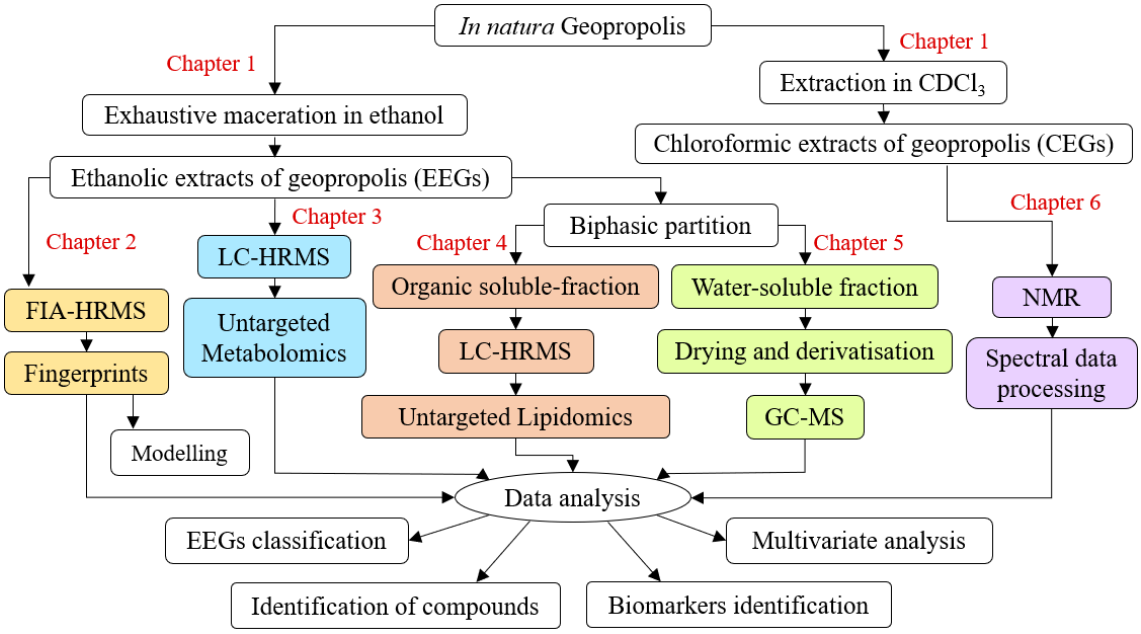


Figure 1 – Work path flowchart.

1.3.2 Specific Objectives

This work was built on an array of analytical goals thought to gather the maximum information possible about the chemical profile of the EEGs. Those goals where to:

I. Carry out the EEGs' HRMS fingerprints and search key information about the geopropolis composition and classification of geopropolis samples, through FIA-HRMS (Orbitrap) and chemometric tools;

II. Develop PLS models able to quickly predict relevant information about the EEGs with accuracy and robustness, based on the HRMS fingerprints, and previous results from TFC and AOC assays;

III. Perform a qualitative and untargeted analysis of EEGs, with samples from distinct geographical origins and different species of stingless bees, employing UHPLC-HRMS (Orbitrap) in a broad metabolomics approach;

IV. Investigate the composition of EEGs using GC-MS;

V. Perform the untargeted lipidomics analysis of the less polar fraction of EEGs, through the semi-targeted method, using LC-HMRS (Orbitrap);

VI. Identify chemical structures detected in the EEGs using LC-HMRS data, employing the Compound Discoverer V. 3.3 software (Thermo, USA), using in house and online spectral libraries;

VII. Use NMR to elucidate and/or confirm chemical structures in the CEGs, and to carry out the study of similarities based on spectral data;

VIII. Compare, classify, and differentiate geopropolis, employing multivariate analysis.

1.4 Sampling

In total 50 samples of *in natura* geopropolis, of 16 species of stingless bees, from diverse parts of Brazil were used during the study (Figure 2): *Friesomelitta doederleini*, *M. asilvai*, *M. bicolor*, *M. fasciculata*, *M. flavolineata*, *M. marginata*, *M. quadrifasciata*, *M. quadrifasciata quadrifasciata*, *M. scutellaris*, *M. seminigra*, *M. subnitida*, *Plebeia meridionalis*, *Scaptotrigona xanthotricha*, *Tetragona clavipes*, *T. angustula*, and *Trigona truculenta*.

Figure 2 – The origins and bee species of the geopropolis samples. The number of samples are between parenthesis, regions of origin are highlighted in bold letters.

All samples of geopropolis were obtained through donations from meliponiculturists and researchers, from diverse parts of Brazil. The samples were sent

to Laboratory of Chromatography and Natural Products (CRONAT), in Universidade Estadual do Centro-Oeste do Paraná (Unicentro), Brazil. The geopropolis samples were individually stored in labelled plastic bags, and then kept in freezer (-15°C) for future analysis. Table 1 presents information about the sample origins and the codes adopted for each geopropolis sample.

Table 1 – Origins of the geopropolis and codes – Continues on next page.

Species	Origin*	Code	Analysed in Chapter:					
			1	2	3	4	5	6
F. doederleini	Petrolina-PE	Fdoed	x	x	x	x	x	
M. asilvai	Prudentópolis-PR	Masil	x	x	x	x	x	
M. bicolor	Prudentópolis-PR	Mbic_01	x	x	x	x	x	
M. bicolor	Prudentópolis-PR	Mbic_02	x	x	x	x	x	
M. fasciculata	Pilar do Sul-SP	Mfasc_01	x	x	x	x	x	
M. fasciculata	B. do Corda-MA	Mfasc_02	x	x	x	x	x	
M. fasciculata	Prudentópolis-PR	Mfasc_03	x	x	x	x	x	
M. flavolineata	B. do Corda-MA	Mflav	x	x	x	x	x	
M. marginata	Prudentópolis-PR	Mmarg_01	x	x	x	x	x	
M. marginata	Curitiba-PR	Mmarg_02	x	x	x	x	x	
M. marginata	Prudentópolis-PR	Mmarg_03	x	x	x	x	x	x
M. marginata	Curitiba-PR	Mmarg_04	x	x	x	x	x	x
M. marginata	Prudentópolis-PR	Mmarg_05	x	x	x	x	x	x
M. marginata	Prudentópolis-PR	Mmarg_06	x	x	x	x	x	
M. marginata	Prudentópolis-PR	Mmarg_07	x	x	x	x	x	x
M. marginata	Prudentópolis-PR	Mmarg_08	x	x	x	x	x	x
M. marginata	Carirá-SE	Mmarg_09	x	x	x	x	x	x
M. marginata	Prudentópolis	Mmarg_10	x	x	x	x	x	
M. quadrifasciata	Pilar do Sul-SP	Mquad_01	x	x	x	x	x	
M. quadrifasciata	Prudentópolis-PR	Mquad_02	x	x	x	x	x	x
M. quadrifasciata	Prudentópolis-PR	Mquad_03	x	x	x	x	x	x

*The origin refers to the city and Brazilian State where geopropolis were collected.

Table 1 – Origins of the geopropolis and codes.

Species	Origin*	Code	Analysed in Chapter:					
			1	2	3	4	5	6
M. quadrifasciata	Curitiba-PR	Mquad_04	x	x	x	x	x	x
M. quadrifasciata	Prudentópolis-PR	Mquad_05	x	x	x	x	x	x
M. quadrifasciata	Prudentópolis-PR	Mquad_06	x	x	x	x	x	x
M. quadrifasciata	Prudentópolis-PR	Mquad_07	x	x	x	x	x	x
M. quadrifasciata	Prudentópolis-PR	Mquad_08	x	x	x	x	x	x
M. quadrifasciata	Prudentópolis-PR	Mquad_09	x	x	x	x	x	x
M. quadrifasciata	Prudentópolis-PR	Mquad_10	x	x	x	x	x	x
M. quadrifasciata	Prudentópolis-PR	Mquad_11	x	x	x	x	x	x
M. q. quadrifasciata	Paranaguá-PR	Mquad_12b	x	x	x	x	x	x
M. scutellaris	Pilar do Sul-SP	Mscut_01	x	x	x	x	x	
M. scutellaris	Mogi Mirim-SP	Mscut_02	x	x	x	x	x	
M. scutellaris	Prudentópolis-PR	Mscut_03	x	x	x	x	x	
M. scutellaris	Curitiba-PR	Mscut_04	x	x	x	x	x	
M. seminigra	Pilar do Sul-SP	Msemi	x	x	x	x	x	
M. subnitida	Prudentópolis-PR	Msubn	x	x	x	x	x	
P. meridionalis	Catanduvas-PR	Pmerid	x	x	x	x	x	
S. xanthotricha	Pilar do Sul-SP	Sxant	x	x	x	x	x	
T. clavipes	Pilar do Sul-SP	Tclav	x	x	x	x	x	
T. angustula	Prudentópolis-PR	Tangu_01	x	x	x	x	x	
T. angustula	Prudentópolis-PR	Tangu_02	x	x	x	x	x	x
T. angustula	Prudentópolis-PR	Tangu_03	x	x	x	x	x	x
T. angustula	Prudentópolis-PR	Tangu_04	x	x	x	x	x	
T. angustula	Q. Centenário-PR	Tangu_05	x	x	x	x	x	x
T. angustula	Prudentópolis-PR	Tangu_06	x	x	x	x	x	x
T. angustula	Curitiba-PR	Tangu_07	x	x	x	x	x	x
T. angustula	Curitiba-PR	Tangu_08	x	x	x	x	x	x
T. angustula	Prudentópolis	Tangu_09	x	x	x	x	x	x
T. angustula	Pilar do Sul-PR	Tangu_10	x	x	x	x	x	x
T. truculenta	Prudentópolis-PR	Ttruc	x	x	x	x	x	

*The origin refers to the city and Brazilian State where geopropolis were collected.

Overall, five samples were from the north-eastern region of Brazil, from the States of Maranhão (MA), Pernambuco (PE), and Sergipe (SE); Nine samples were from the south-eastern region, specifically from the State of São Paulo (SP); and 38 samples were from the southern region, specifically from the State of Paraná. Figure 3 andFigure **4** shows the samples of *in natura* geopropolis as they were when received from the beekeepers:

Figure 3 – Samples of *in natura* geopropolis, grouped according to the stingless bee species (in italic), and the unique sample code is at the bottom of its respective image.

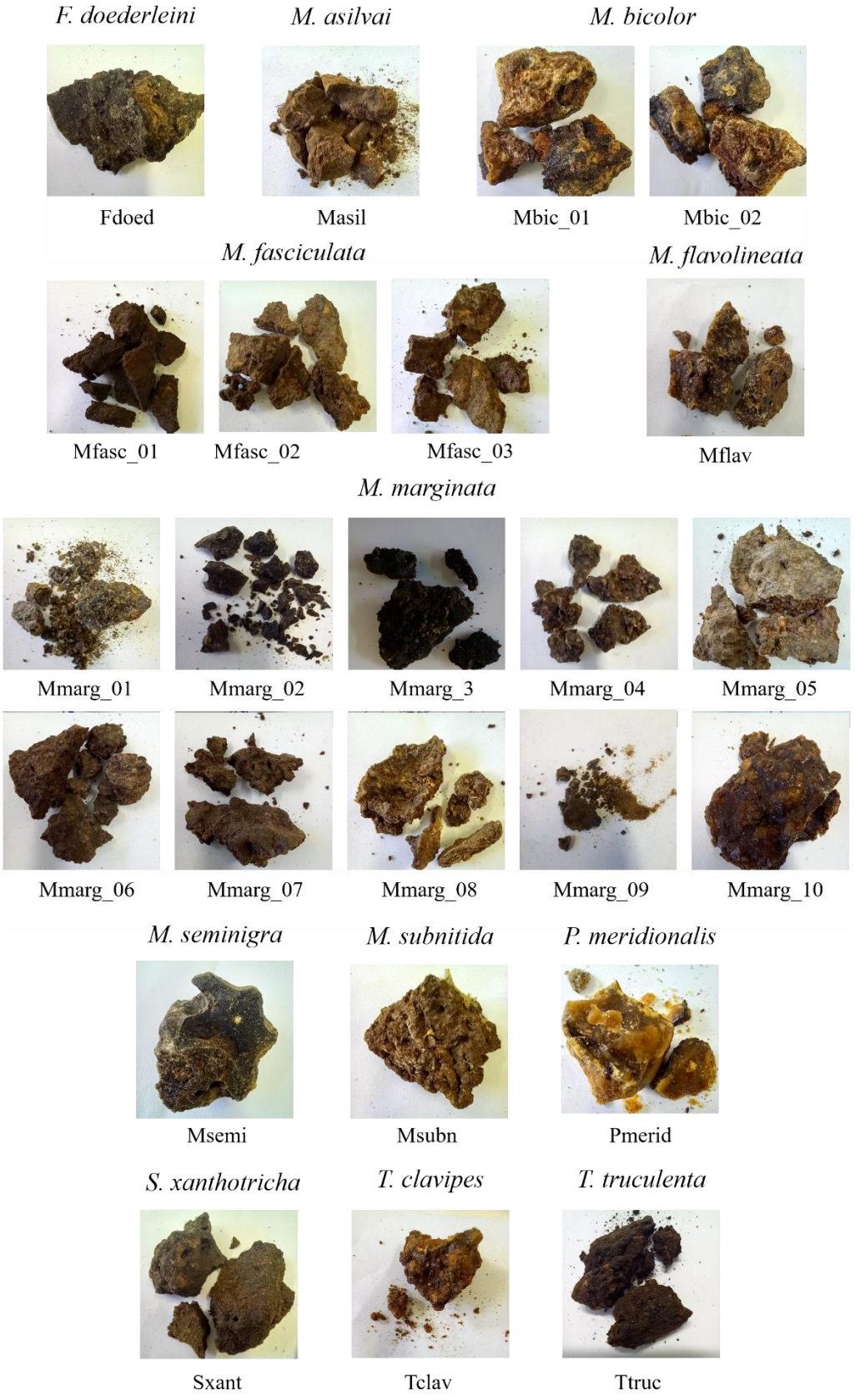


Figure 4 – Samples of *in natura* geopropolis, grouped according to the stingless bee species (in italic), and the unique sample code is at the bottom of its respective image.

1.5 Sample extraction and partition

The extraction of *in natura* geopropolis was carried out in two separate ways: firstly, ethanolic extracts of geopropolis (EEGs) were obtained to be used in analysis employing hyphenated techniques (FIA-HRMS, LC-HRMS, and GC-MS); lastly, chloroformic extracts of geopropolis (CEGs) were obtained using deuterated chloroform ($CDCl_3$) to be analysed by NMR.

All solvents, reagents, and standards used in this work were of analytic and/or chromatographic grades.

1.5.1 Obtention of the EEGs

For analysis using hyphenated techniques, the first geopropolis extracts were obtained by maceration using ethanol as extractive solvent. Figure 5 illustrates the steps followed to obtain the ethanolic extracts of geopropolis (EEGs): Portions of *in natura* geopropolis were crushed using mortar and pestle. Subsequently, 2 g of crushed geopropolis were added into flasks containing 20 mL of ethanol (HPLC degree) alone, and left stirring in an incubator at 180 rpm, and 25°C. After 24 h the EEGs were filtered under vacuum, and their volume was topped up to 25 mL using ethanol. Thereafter, the EEGs were stocked in labelled amber flasks, and stored in freezer at -15°C until the analysis took place.

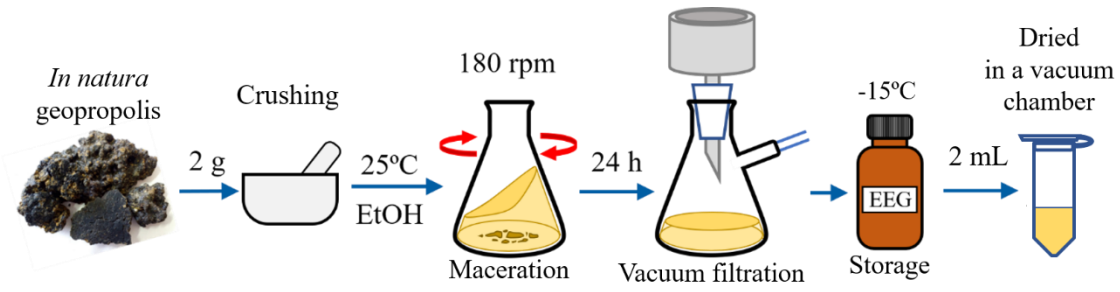


Figure 5 – Obtention of the EEGs.

Aliquots of 2 mL from each EEGs were dried out in a vacuum chamber, weighted, and then transported under room temperature to the Proteomics and Metabolomics Facility, Imperial College London, Faculty of Medicine, Hammersmith *campus*, London, United Kingdom. Once there, the samples were stored in freezer (-80°C) until the intended analysis were performed.

In the Proteomics and Metabolomics Facility, the EEGs were completely dried in a vacuum concentrator chamber and weighted thereafter. The EEGs were all resuspended to the concentration of 10 $\mu g.mL^{-1}$ with a 50/50% methanol/isopropanol solution. A quality control (QC) sample was created by pooling together 20 µL of each sample. These extracts were directly analysed by FIA-HRMS (Chapter 2) and UHPLC-HRMS (Chapter 3).

1.5.2 Biphasic partition of the EEGs

The biphasic partition was used to obtain a fraction rich in lipids. The methodology employed was described by Hall et al. [23], with small adaptations in volumes. This methodology is commonly employed method at the Metabolomics and Proteomics Facility (MRC London Institute of Medical Sciences, Imperial College London).

Exactly 20 µL were taken from each EEG, then diluted in 180 µL of a solution of MeOH/IPA (50:50). A QC sample was created by pooling 20 µL of each EEG together. The samples were left overnight to dry in a vacuum chamber.

The biphasic partition (or liquid-liquid extraction) started by adding 300 µL of methanol, followed by 100 µL of chloroform and 300 µL of ultrapure water. Each sample was shacked using vortex for approximately 15 s, until clear phases were formed. To complete the process the samples were centrifugated during 10 min at 14600 rpm. Finally,

two phases were visible, as shown in Figure 6. The remaining part of this fraction was then removed. In sequence, 200 µL of the organic lipid-rich (lipophilic) fraction was collected using a pipette, dried in a vacuum chamber, and stored in freezer until the lipidomics analysis was performed.

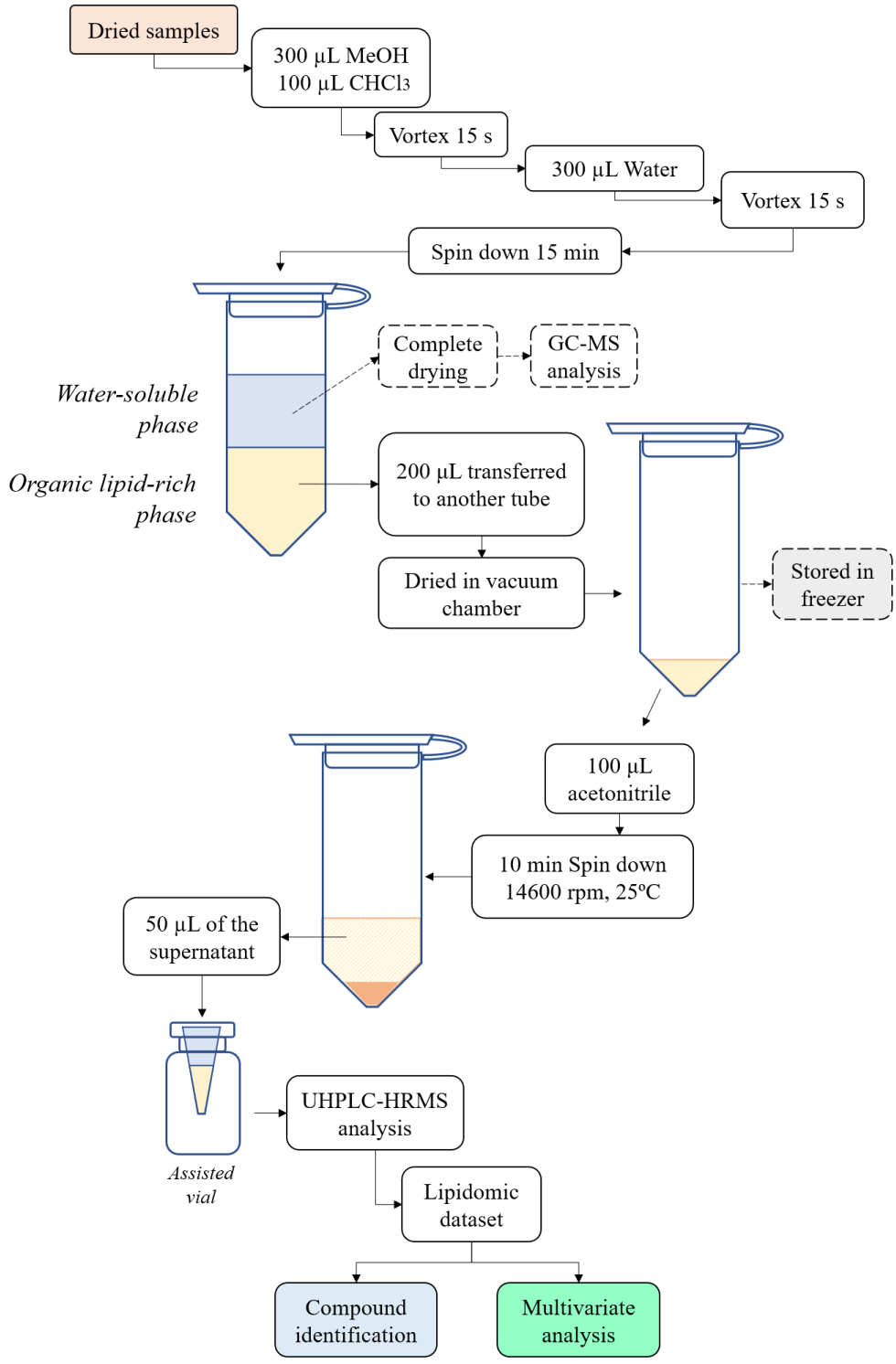


Figure 6 – Liquid-liquid extraction of the EEGs, and analysis of its lipid-rich phase.

The dried organic lipid-rich fractions of the EEGs were diluted in 100 µL of acetonitrile and then spined down during 10 min, at 14600 rpm, and room temperature. Subsequently, 50 µL of the supernatant was added into assisted vials and then sealed. The samples were kept in the instrument autosampler at 20 ºC during all the analysis.

1.5.3 Obtention of the CEGs

Chloroformic extracts of geopropolis (CEGs) were obtained using a simple and concise protocol commonly in the Metabolomics and NMR Laboratory (Universidade Estadual do Norte Fluminense Darcy Ribeiro):

Firstly, 100 mg of *in natura* geopropolis were completely dissolved in 500 µL of $CDCl_3$ with tetramethylsilane (TMS) as internal standard, in sequence the CEGs were stirred by hand, and filtered on qualitative paper to remove soil particles. Finally, the CEGs were added in thin-walled tubes suitable for NMR spectroscopy, and stored under refrigeration until analysis took place. Figure **7** illustrates the process for CEGs obtention.

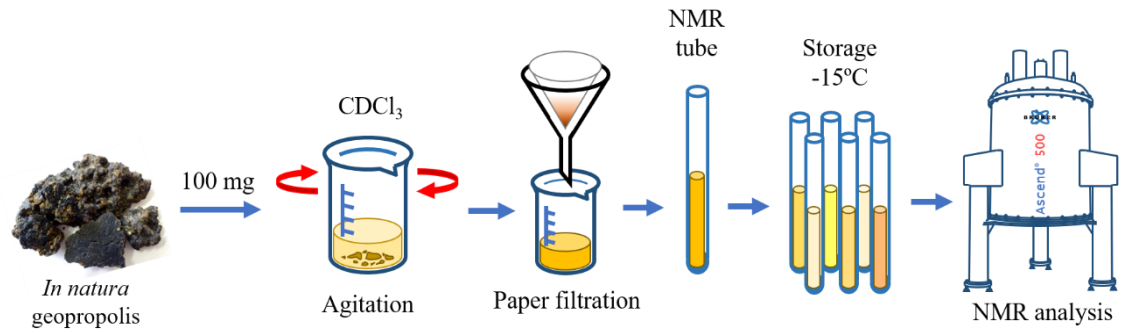


Figure 7 – Obtention of the CEGs.

Chapter 2 – Multivariate analysis of FIA-HRMS fingerprints of EEGs

2.1 Objectives

FIA–HRMS is a simpler technique and easier to implement if compared to LC-MS and provides mass fingerprints which may allow sample discrimination albeit huge ionization suppression. Then, in this part of the study EEGs were directly introduced into de mass spectrometer with the aim to carry out exploratory analysis of the whole fingerprints. The specific goals were to:

I. Investigate EEGs mass fingerprints looking for relationships or differences among geopropolis samples using only flow injection into the mass spectrometer.

II. Classify EEGs using the information given by their FIA-HR mass fingerprints, while determining which ions are responsible for these features.

II. Further explore FIA-HR fingerprints data by investigating their correlation with TFC and AOC, using mathematical tools.

IV. Build PLS models capable to estimate/predict TFC and AOC of EEGs with reasonable accuracy.

2.2 Methodology

2.2.1 Samples preparation for fingerprint acquisition

The fingerprints were acquired in both negative and positive modes of ionisation; therefore, two batches of samples were prepared, using two different microplates. In the plate for positive mode analysis, the wells received 180 µL of a 50/50% MeOH/IPA solution, also containing 0.1% (v/v) of formic acid and labelled arginine ($^{13}C_6H_{14}{}^{15}N_4O_2$)

250 ng.mL^{-1}. Finally, an aliquot of 20 µL of the samples were added. The negative mode was carried out in a similar way, with 20 µL of the samples being diluted in a 180 µL of a 50/50% MeOH/IPA solution, however, in this case, containing 0.1 ammonium hydroxide (v/v) and citric acid ($^{13}C_6H_8O_7$) 250 ng.mL^{-1}. The labelled internal standards, arginine and citric acid, were used in order to assess the mass accuracy, thus ensuring the data integrity. Formic acid and ammonium hydroxide were used in order to increase the ionisation process in positive and negative ionisation modes respectively.

Also, two blanks were prepared following the above-mentioned protocol, however instead of a sample, 20 µL of 50/50% MeOH/IPA solution were utilised. Figure 8 shows the workflow to obtain FIA-HR mass fingerprints:

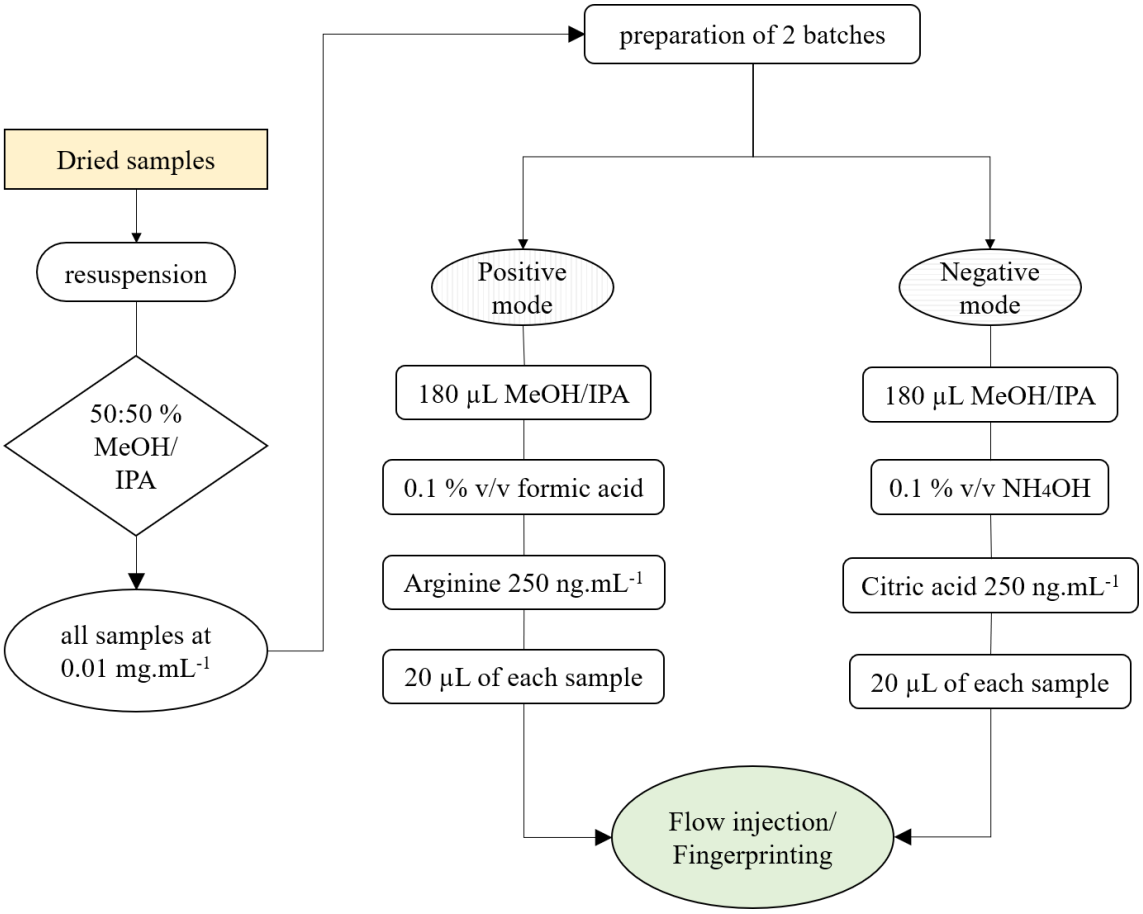


Figure 8 – Fingerprinting methodology – Sample preparation.

2.2.2 Instrumentation and fingerprinting

The fingerprints were acquired in a Q Exactive Quadrupole-Orbitrap mass spectrometer (Thermo). The instrument is equipped with an H-ESI (heated electron-spray ion source), S-lens ion optics technology, quadrupole mass filter, and Orbitrap mass analyser. The instrument is also connected to an ultra-high-performance liquid chromatograph, Vanquish UHPLC (Thermo).

The fingerprints were obtained using the technique of direct flow injection analysis (FIA), where the samples are directly injected in the mass spectrometer. Due the large number of samples, the injection was assisted by the LC autosampler (which was configured to bypass the chromatographic column, not using it). The spectra were acquired in full scan, recorded between 50 and 1000 m/z, and the mass resolution was 140 000. The infusion method consisted in two steps: First, 5 µL of a sample were infused during 30 s directly in the mass spectrometer. Second, a 50/50% MeOH/IPA solution was infused during 2.5 min in order to clean up the system and re-equilibrate the instrument to be ready for the next sample infusion. The QCs were injected in regular intervals, between each 10 injections.

2.2.3 Data mining/data pre-processing

Fingerprints data were pre-analysed with R V. 3.4.2, using R Studio V. 1.1.456. R is an opensource software, with a wide variety of packages available for data analysis. Therefore, R could be used by others in the field also interested in data visualisation, and machine learning. In this study proFIA R-package was employed to carry out the data pre-processing, and data framing, of the fingerprints obtained by FIA-HRMS.

Initially, the raw data is acquired from the instrument's computer, and converted from the original ".raw" format to the readable ".mzXML" format, using the MSConvert software V. 3.0.19287. Then, the files were processed using proFIA, which is able to extract the ion intensities in each sample, framing then into a single usable table, where columns represent the samples, and the rows stands for the ions (features). The resultant two tables, one for negative and another for positive ionisation modes, were merged to create a single data frame containing the maximum number of features obtained from the fingerprints, no mattering the ionisation mode. Figure 9 shows the path followed.

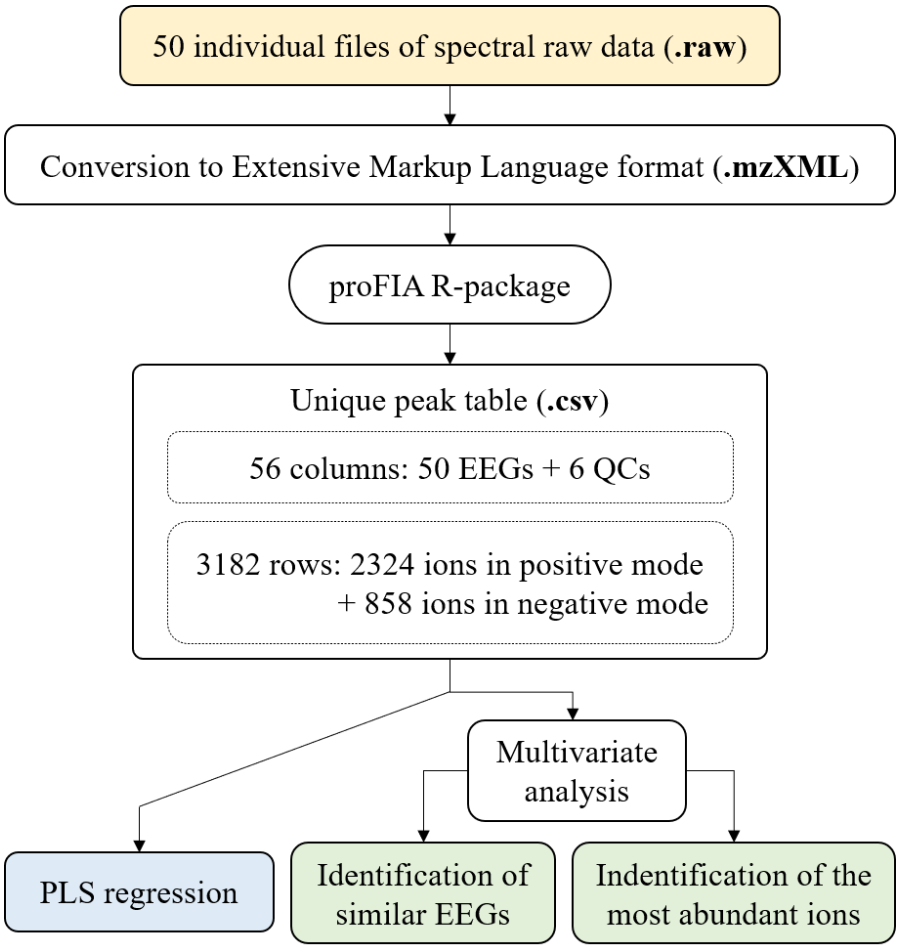


Figure 9 – Fingerprint data conversion and data processing workflow.

The resultant data-frame (peak table) acquired from the fingerprints was submitted to multivariate analysis, using the web-based platform MetaboAnalyst (V. 5.0) [24]. In

addition, the fingerprint data was used to build partial least squares (PLS) regression models, based on results previously obtained for geopropolis' total flavonoids content and antioxidant capacity.

2.2.4 Multivariate analysis of FIA–HRMS fingerprints of EEGs

The multivariate data analysis of the fingerprints data was performed using the MetaboAnalyst v. 5.0, a free online platform for bioinformatics [25]. Statistical tests such as ANOVA (analysis of variance), PCA (principal component analysis), HCA (hierarchical cluster analysis) were performed.

Due the large number of species represented by a small number of samples of each species, the EEGs were divided in two datasets and then analysed separately: the first one is composed by fingerprints of EEGs of *M. marginata, M. quadrifasciata*, *M. scutellaris*, and *T. angustula*, because more than three samples per species were available. The second dataset is composed of the reminiscent EEGs' fingerprints (*F. doederleini*, *M. fasciculata*, *M. seminigra, M. subnitida, S. xanthotricha, T. truculenta*), where samples from species of *Melipona* genera were gathered together in the "*Melipona*" group, and the others samples were reunited in the "non-*Melipona*" group. This way of analysis was adopted aiming to avoid an unbalanced and unnecessary complex analysis, with too many groups.

2.2.5 TFC and AOC

The EEGs' total flavonoid content (TFC), and antioxidant capacity (AOC), used in this study were assessed by stablished UV-Visible (UV-Vis) spectrometry-based techniques. EEGs' TFC was assessed using the technique employed by Woisky & Salatino [26]. The antioxidant capacity of the EEGs was measured by two different antioxidant assays: Vitamin C (ascorbic acid) equivalent antioxidant capacity (VCEAC),

described by Brand-Williams et al. [27]; and cupric ion reducing antioxidant capacity (CUPRAC), described by Apak et al. [28].

The results were previously published by Turco et al. [29] in the Microchemical Journal (volume 157, September 2020), under the title "*Could antioxidant capacity and flavonoid content of ethanolic extracts of geopropolis from Brazilian native bees be estimated from digital photos and NIR Spectra?*".

2.2.6 Partial least squares regression

In this Chapter, partial Least Squares (PLS) regression was employed to investigate the correlation between spectral data (mass fingerprints) and the values obtained by the reference tests: TFC, VCEAC and CUPRAC. The licence-free software ChemoStat V2 was used to carry out the PLS modelling [29].

Before the modelling process, three different mathematical pre-processing methods were separately applied and tested, in order to investigate different ways to minimise the noise effects, nonetheless enhancing the quality of the models. The functions were: standard normal variate (SNV), multiplicative scatter correction (MSC), and the Savitzyk-Golay (S-G) smoothing filter. Those pre-processing methods are among the most commonly used data transformation functions [31, 32]. The MSC and SNV are usually employed as scatter-corrective methods, while S-G is normally used to reduce the additive and multiplicative effects, including finite differences, in addition to a smoothing step [31].

Also, one model was tested without using any pre-processing to appraise the necessity of a data pre-processing step. At the end, four datasets were created. The datasets were submitted to a Hierarchical Cluster Analysis (HCA) in an effort to identify clusters among the samples, and as a way to select the training sample set (calibration)

and the validation samples set. From each identified cluster, one sample was chosen to make validation dataset; the remaining samples were used as calibration dataset; Figure 10 shows the PLS modelling pipeline:

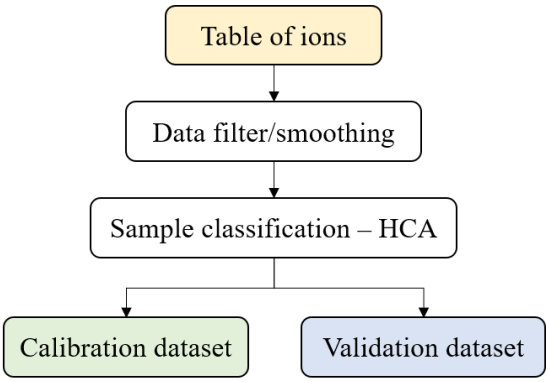


Figure 10 – PLS Regression workflow.

Although there is not a consensus to choose the number of LVs, five parameters were taken in account: the root mean square errors of: calibration (RMSEC), prediction (RMSEP), and cross-validation (RMSECV); as well the coefficient of determination (R^2) of the calibration and validation models. These parameters were analysed as a whole, aiming to avoid biasing or overfitting the models. In addition, these parameters were also observed to evaluate the quality and accuracy of the PLS models.

The PLS models were submitted to a validation process, in order to assess their prediction capabilities. The coefficient of determination (R^2) shows the proportion of variance in the set of the reference data that can be explained by the variation in the predicted data. An ideal PLS model would present a $R^2 \approx 1$, however, in reality models presenting $R^2 \geq 0.7$ are already considered models with high level of correlation. When cross validation is used, RMSECV expresses the square root of the cross-validation mean square error. The RMSECV is a quantitative measurement of the accuracy achieved in the process of predicting values for the samples, during cross-validation. Therefore, it is defined as the standard deviation of the differences between: spectral data and reference data, in the cross-validation sample set [33].

2.3 Results and discussion

2.3.1 EEGs FIA-HR mass fingerprints multivariate analysis

The mass fingerprints were obtained by injecting the samples directly into the mass spectrometer, without previous use of a chromatographic column to separate the compounds. Therefore, instead of the common chromatograms, this method generates the so-called "flowgrams", containing a single large "peak" with all the others convoluted within (Figure 11 A and C). The flowgrams of the QC samples are shown in Figure 11, where A and C are the flowgrams obtained in the positive and negative modes, respectively; B and D are the full range (70 – 1000 m/z) mass spectra (the fingerprint) extracted from the flowgrams.

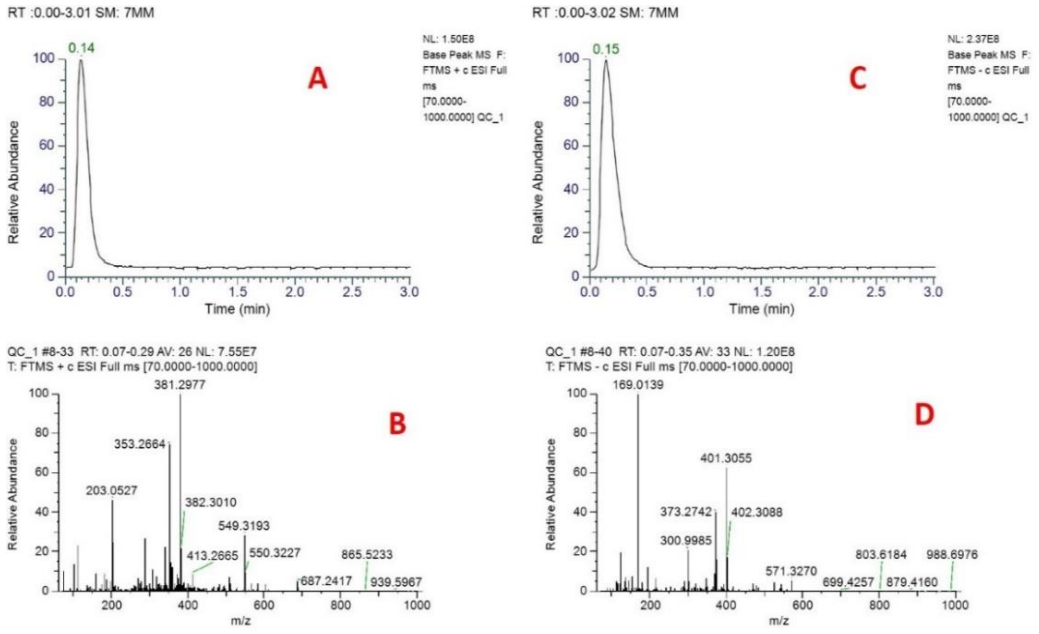


Figure 11 – Fingerprints of the QC samples. A and C: flowgrams. B and D fingerprints acquired in the positive and negative modes respectively.

An average mass spectra of 2326 ions were listed in positive mode of ionisation, and 858 ions in negative mode, totalising 3184 ions in the dataset, considering the resolution of four decimal places. Multivariate analysis was carried out with both positive

and negative datasets merged into a single table, which better represent the geopropolis features. Prior to multivariate analysis, exactly 1270 features (ions) were filtered out because they presented an RSD > 30%, based on QC samples (Part 1, section 1.5.1). This is in order to avoid interferences from baseline noises, that may cause distortions in the results [34].

Visually, it is relatively easy to perceive differences among the EEGs fingerprints; however, it is hard to point out what ions are the most characteristic of each species and/or geographical origin without applying statistical tests. However, FIA-HRMS produces a huge amount of information per sample, with long lists of ion weights and their abundances, which are different in size according to the complexity of the samples. Hence, it is extremely difficult to extract relevant information about the EEGs' characteristics without applying some effort to ordinate the dataset into a single data frame [35]. In this way, the ions/abundance lists of each sample should be ordered into a table (data frame), which contains all the EEGs' possible information, as a whole. Now, this data frame (or ions table) enables multiple types of multivariate analysis, such as Principal Component Analysis (PCA), Analysis of Variance (ANOVA), and Analysis of correlations. This table of ions was also posteriorly used to build the PLS models, by combining the ion intensities with the information of TFC, and antioxidant capacities (VCEAC and CUPRAC).

2.3.2 Principal component analysis (PCA) of the first dataset

Data pre-processing was carried out using square root transformation as data and data scaling was performed using the Pareto scaling function. The samples were grouped according to their species or, in some cases, their genus. In order to reduce the complexity of the analysis, the whole fingerprints dataset was split in two groups of geopropolis and

then analysed separately. The first dataset was composed by fingerprints of EEGs from the groups with the highest number of samples, as a consequence the *M. marginata*, *M. quadrifasciata*, *M. scutellaris*, and *T. angustula* groups were analysed separately from the others. The PC1 and PC2 were chosen to show the results of the analysis, shown by the Figure 12.

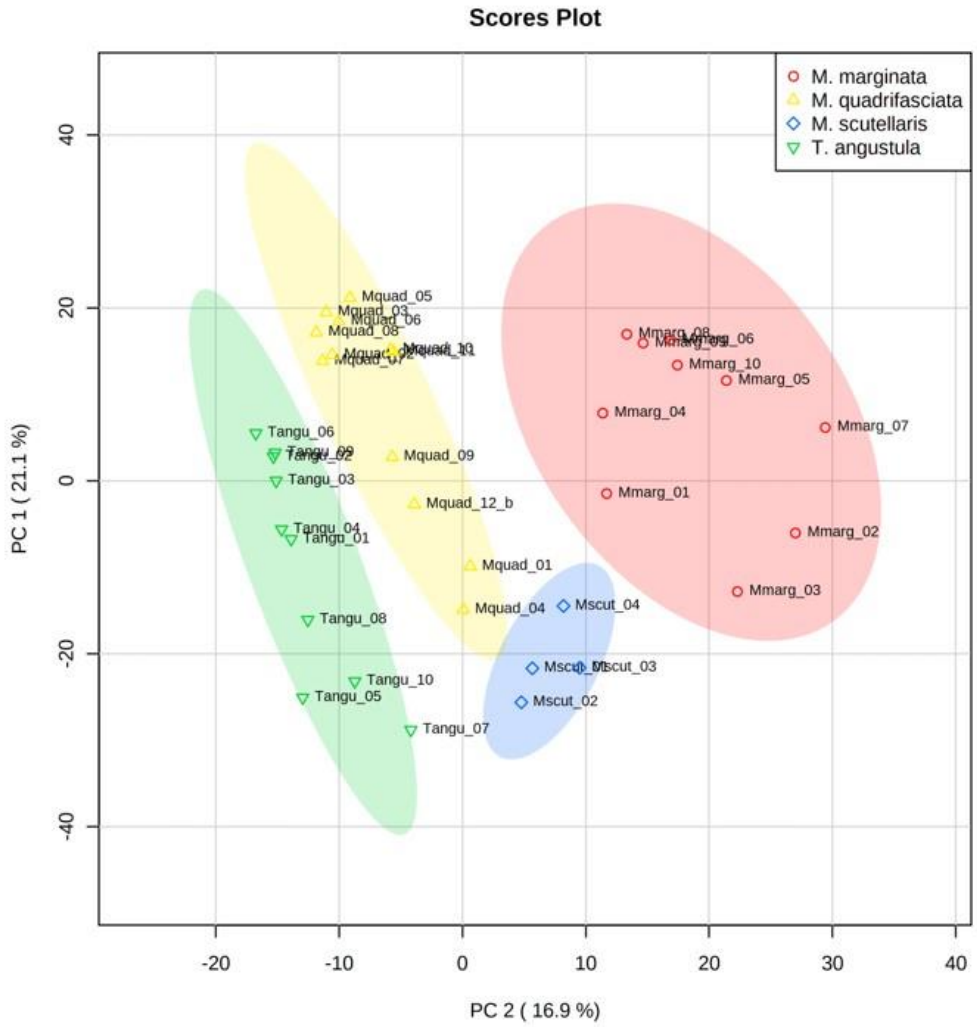


Figure 12 – Exploratory PCA of *M. marginata*, *M. quadrifasciata*, *M. scutellaris*, and *T. angustula*. Confidence interval of 95%.

The exploratory PCA of the fingerprints from the first dataset (Figure 12) makes visible that the features in the FIA-HRMS fingerprints could be associated with the stingless bee specie. In other words, EEGs from different species of stingless bees may

be distinguished using the whole mass fingerprint as they contain different protonated or deprotonated molecular ions of varying intensities. Also, the *M. marginata* group is detached from the others, which denotes that the EEGs of this species are probably chemically more distinct from the others.

It is known that the geographical origin also plays an important role in the geopropolis composition, consequently affecting the ions found in the fingerprints of the EEGs [5, 36–38]. To investigate this fact, Figure 13 shows PC1 vs PC2 replotted highlighting the sample origin.

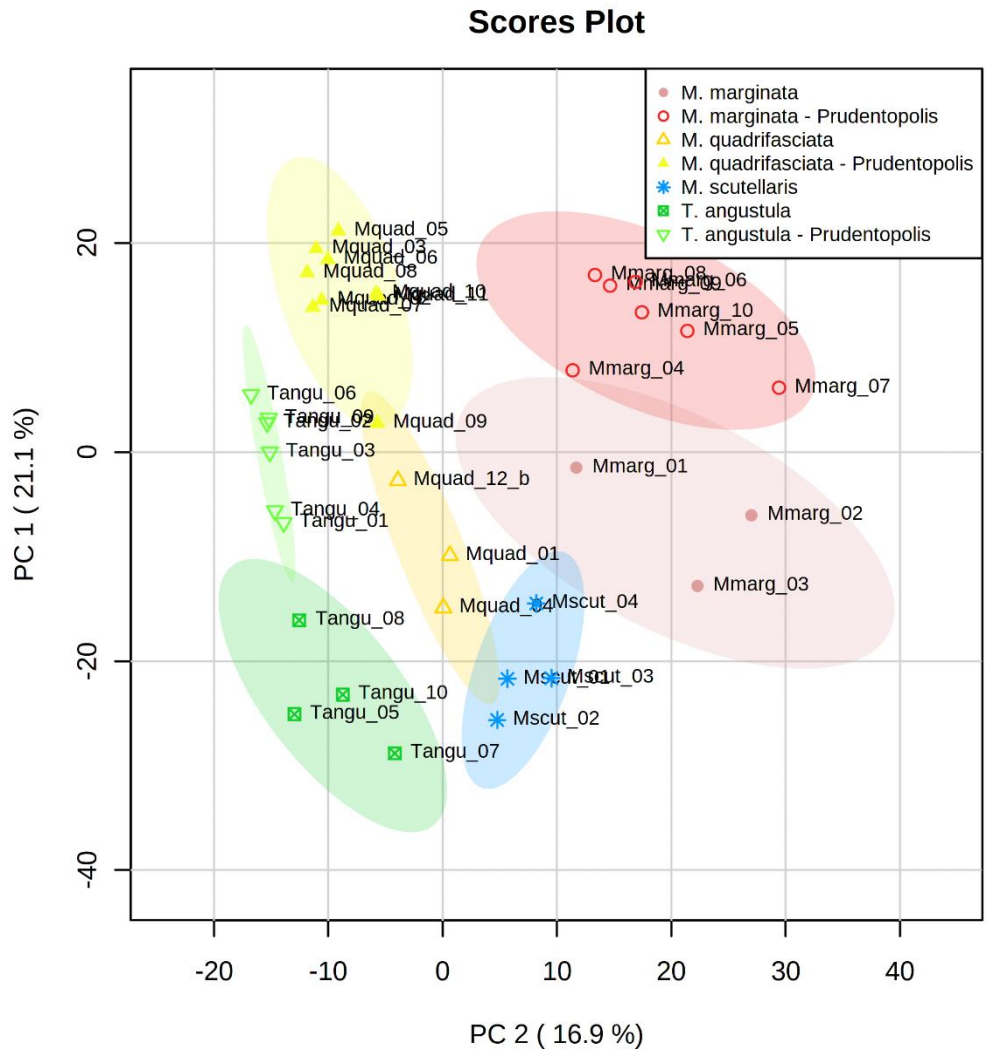


Figure 13 – PCA of the fingerprints of EEGs from *M. quadrifasciata*, *M. scutellaris*, *M. marginata*, and *T. angustula*, from different geographical origins.

Thereby, the EEGs' fingerprints could be characterised not only by the bees' species, but by a combination of both factors: the geographical origins and stingless bees' species.

The influence of the two factor, geographical origin and stingless bees' species, seems to vary from one species to another. EEGs of *M. marginata* from Prudentópolis presented m/z 367.1885 as their most characteristic ion, while the EEGs from the same species, but from other locations (Carirá – SE, Curitiba – PR, and Pilar do Sul – SP) presented the ion m/z 646.2349 as their common characteristic ion (Table 2). *M. quadrifasciata* from Prudentópolis presented m/z 428.2493, EEGs of samples from other regions presented m/z 695.4502. Although each sample of *M. scutellaris* geopropolis came from a distinct Brazilian region, as presented by Figure 2, their EEGs presented a more concise group when compared to the others, characterised by the ion m/z 286.0066. EEGs of *T. angustula* presented in their fingerprints the ion m/z 381.2977, however the most characteristic ions were m/z 709.4294 for geopropolis from Prudentópolis, and m/z 507.3816 for those produced outside this region (Curitiba and Quarto Centenario – Figure 2).

Even though the fingerprints were obtained using high resolution mass spectrometry (HRMS), it is still difficult to identify and exactly confirm which molecules (or fragments of molecules) are represented by each ion in a high level of confidence, once there is no possibility for trustful comparison with spectral libraries data [39]. Nonetheless, these ions are related to real compounds, but MS^2 spectral information is required to confirm their identity.

The second dataset, which was composed by groups represented by small numbers of samples, were difficult to group by their species only. Therefore, the genus was chosen to be used as a grouping factor, and the fingerprints of those EEGs were selected into two

groups: one group was formed only by species of the *Melipona* genus; the other "group" is a gathering of species of others genus (*Friesomelitta*, *Plebeia*, *Scaptotrigona*, *Tetragona*, and *Trigona*). The PC1 and PC2 were chosen to show the results of the analysis, shown by Figure 14.

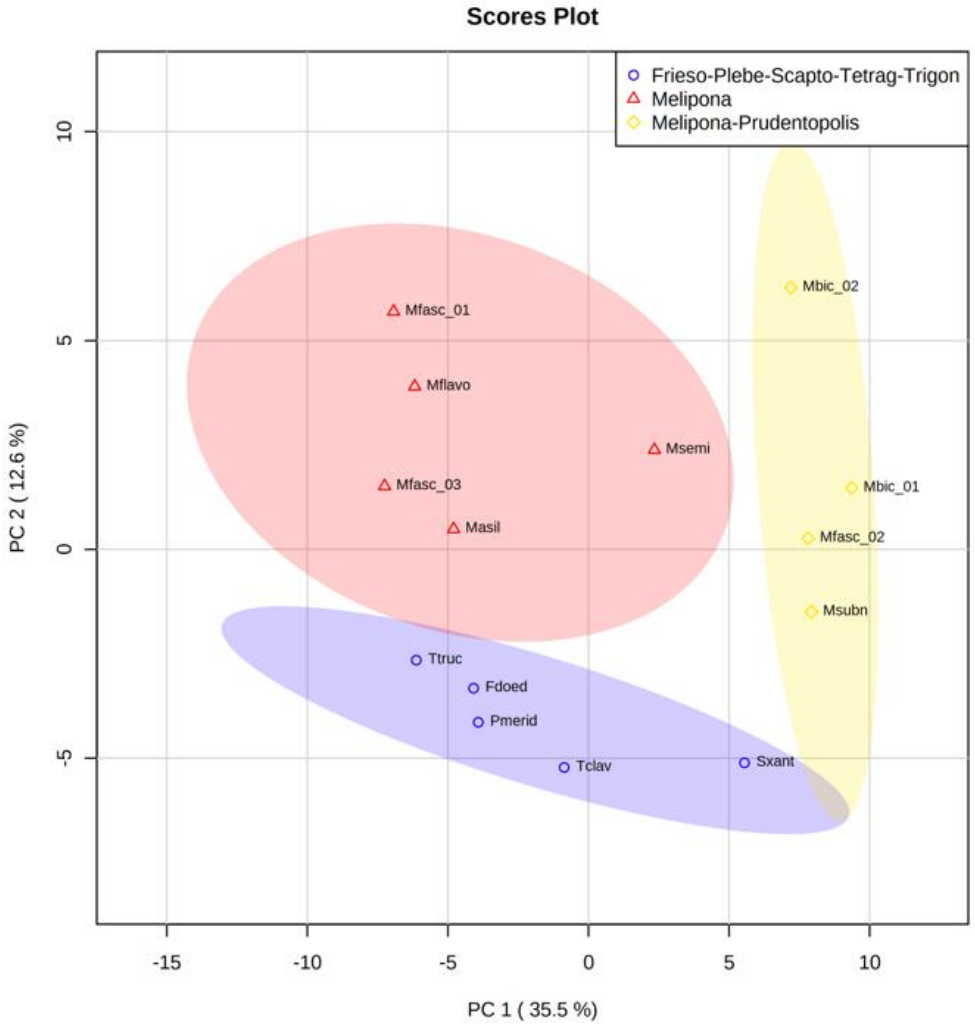


Figure 14 – PCA of the fingerprints of the EEGs of *Melipona* genus (*M. asilvai*, *M. bicolor*, *M. fasciculata*, *M. flavolineata*, *M. seminigra*, *M. subnitida*), and non-*Melipona* genus (*F. doederleini*, *P. meridionalis*, *S. xanthotricha*, *T. clavipes*, and *T. truculenta*), from different geographical locations.

The influence of the geographical origin on the chemical composition was further evident when a second PCA was carried out considering only the species with three or less samples (Figure 14). These geopropolis were tested separately, in order to avoid

unbalanced analysis, and misleading figures [33]. Henceforth, the EEGs were grouped according to the genus of the stingless bees. Furthermore, the EEGs from the *Melipona* group could be divided into two subgroups, based on their geographical origin. Figure 14 shows the PCA of fingerprints of EEGs from *Melipona* genus from Prudentópolis, *Melipona* genus from other regions, and the non-*Melipona* group.

2.3.3 Hierarchical clustering analysis (HCA)

The hierarchical clustering analysis (HCA), shown in Figure 15, indicates a discrimination of geopropolis by a combination of the producing bees, and geographical origins, to classify the groups. The Ward's linkage, using Euclidean distance measurement, was able to group almost all the samples according to their species and geographical origins.

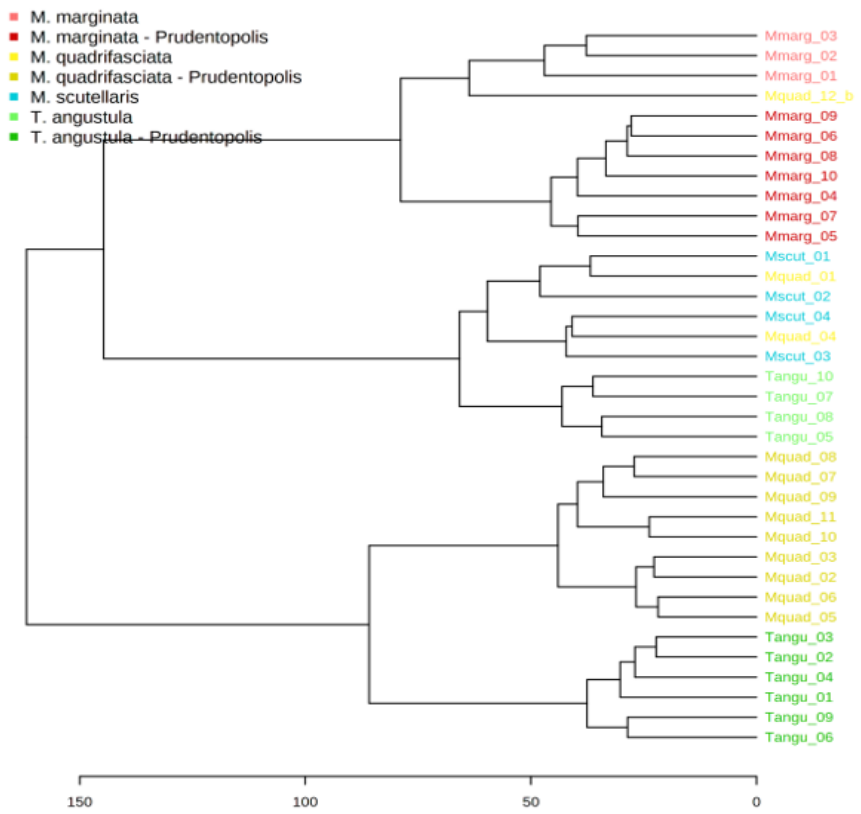


Figure 15 – HCA of the fingerprints of the EEGs of *M. marginata*, *M. quadrifasciata*, *M. scutellaris*, and *T. angustula*.

In the Figure 15, *T. angustula* EEGs are divided in two subgroups, being the large one composed by geopropolis from the Prudentópolis region, and the other composed by three samples coming from the regions of Curitiba (Tangu_05 and Tangu_09) and Quarto Centenário (Tangu_08), State of Paraná (refer to Figure 2). Similarly, *M. marginata* EEGs could be divided into two subgroups. Again, the largest group is formed by samples from Prudentópolis region. The large group is composed by EEGs which the geopropolis samples were produced in the state of Paraná; the other two samples, detached from this group are Mmarg_01 and Mmarg_02, from the States of São Paulo and Sergipe respectively. However, the other group is composed by two samples from Curitiba (Mmarg_02 and Mmarg_04) and one from Carira (margi_02), from the north-eastern region of Brazil (Figure 2). EEGs of *M. quadrifasciata* presented a subgroup of 9 samples from Prudentópolis and more four samples scattered (Figure 15).

2.3.4 Pearson correlation heat map

A Pearson correlations heat map, shown in Figure 16, helps us to observe the relationship between samples. The colour blue represents negative correlations, while red represents positive correlations. Fingerprints of the EEGs of *T. angustula* from the region of Prudentópolis presents a low, but yet noticeable, correlation with the other samples of the same species, with *p*-value around 0.3. This may denotate some similar chemical features among geopropolis of *T. angustula* regardless their distinct origin.

A nearly constant chemical composition regardless the geographical origin of the sample of the geopropolis produced by *T. angustula* was reported by Sawaya et.al. (2006). The authors compared the ESI(-)-MS fingerprints of *T. angustula* and *Apis mellifera* of samples collected throughout Brazil. While ESI(-)-MS fingerprints of *A. mellifera* propolis varied from region to region, those from *T. angustula* varied only in abundances

of the minor ions. The authors suggested that one plant is the main source of the resins for propolis of *T. angustula* [40].

On the other hand, the *M. quadrifasciata* group from Prudentópolis does not present any considerable correlation with its counterpart from other states (Mquad_01, Mquad_04, and Mquad_12_b), suggesting different vegetal sources of resins for each kind of geopropolis.

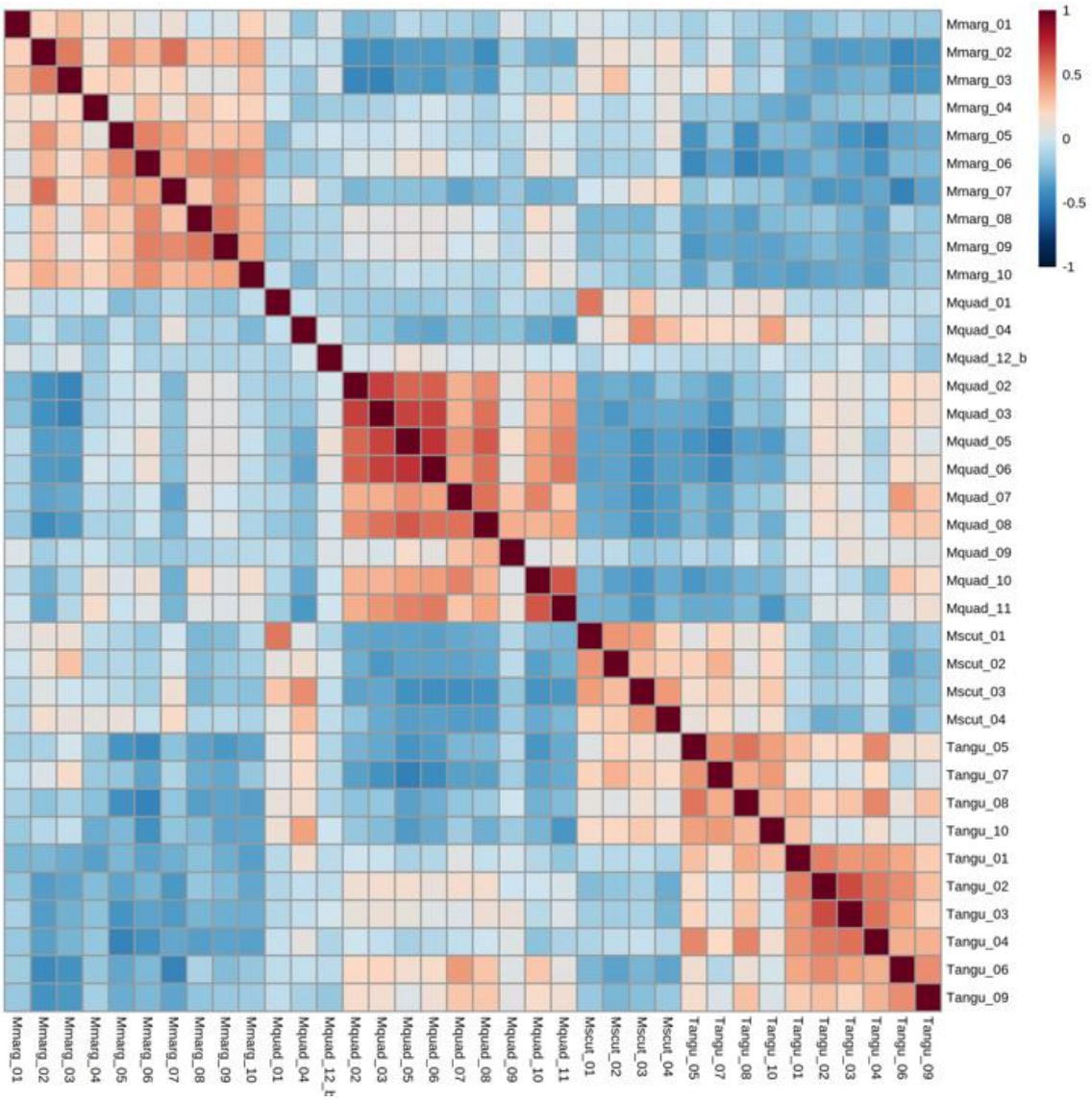


Figure 16 – Correlations map of the EEGs fingerprints of *M. marginata*, *M. quadrifasciata*, *M. scutellaris*, and *T. angustula*.

2.3.5 Most characteristic ions in the FIA-HRMS fingerprints of EEGs

In total 548 *m/z*'s were found relevant, but only the most characteristic were selected as markers: *M. marginata*; m/z 407.1467; *M. quadrifasciata*, m/z 707.4500; *M. scutellaris* m/z 283.0066; and *T. angustula*, m/z 415.3188. Figure 17 shows the boxplot graphics for ions abundance, presenting the original and the normalised total ion abundance. In theory, high resolution mass spectrometry (HRMS) would allow compound identification with high confidence degree, however without any other information (retention time, fragmentation data, comparison with standards, etc), only assumptive conclusions could be made.

Table 2 presents a summarised list of the most characteristic ions (*m/z*'s) in each group of EEGs, selected by picking the ions with highest values for $-\log_{10}(p)$, using the loadings of ANOVA and PCA.

Table 2 – The relatively most characteristic *m/z*'s of the EEGs – Continues on next page.

Specie/Genus /Group	Geographical origin	Exact mass [m/z]	Ionisation mode	$-\log_{10}(p)$	*f*–value
M. marginata	Carira Curitiba Pilar do Sul Prudentópolis	407.1467 623.3197	Pos Pos	4.2515 9.4203	5.5849 6.7862
M. quadrifasciata	Curitiba Paranaguá Pilar do Sul Prudentópolis	428.2493 707.4500	Pos Pos	5.284 12.6190	3.0580 23.6630
M. scutellaris	Curitiba Mogi Mirim Prudentópolis Pilar do Sul	283.0066	Neg	16.511	7.4529
T. angustula	Curitiba Quarto Centenário	507.3816	Pos	9.4990	25.0050

Table 2 – The relatively most characteristic *m/z*'s of the EEGs.

Specie/Genus /Group	**Geographical origin**	**Exact mass [m/z]**	**Ionisation mode**	**-log$_{10}$(*p*)**	***f*–value**
T. angustula	Prudentópolis	544.3336	Pos	5.9593	8.0536
Melipona*	Barra do Corda Pilar do Sul Prudentópolis	572.6178	Pos	9.5449	6.8580
Non-*Melipona* **	Petrolina Catanduvas Pilar do Sul Prudentópolis	517.3721	Pos	1.8844	2.7604

*Species: *M. asilvai*, *M. bicolor*, *M. fasciculata*, *M. flavolineata*, *M. seminigra*, *M. subnitida*.

Genus: *Friesomelitta*, *Plebeia*, *Scaptotrigona*, *Tetragona*,** and *Tetragonisca*.

Each group presents a variety of ions that could be considered characteristic to then, however only the most significative was chosen as their marker ion. As such, an analysis of variance (ANOVA) was performed. By observing the *p*-values, it was possible to determine which ions were the most abundant and unique in each group, and as consequence, the most characteristic. The graphic of loadings was used to find out, based on the *p*-value, what ions were most significatively characteristic for each species. The relative abundances of these ions are represented in Figure 17:

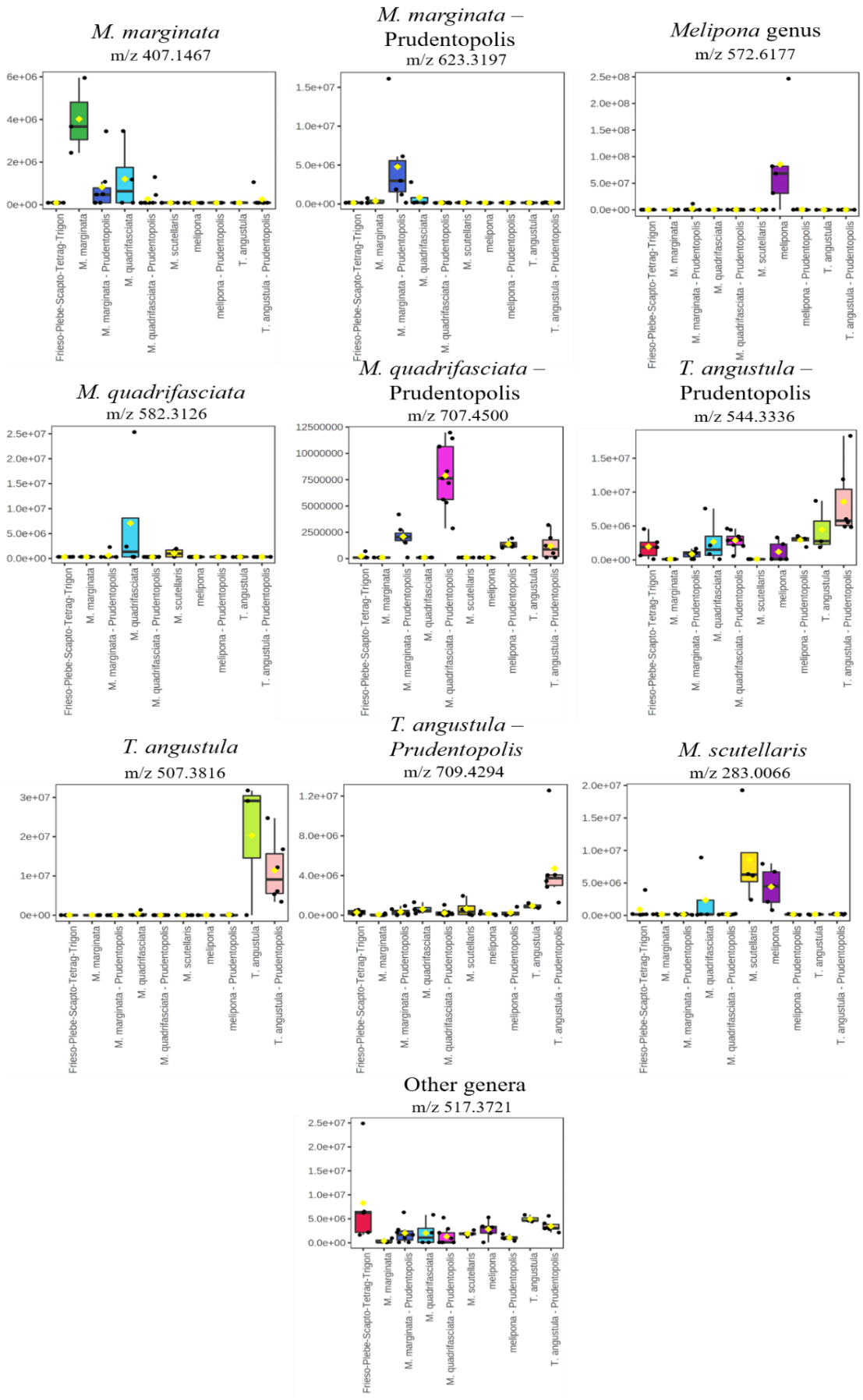


Figure 17 – Relative abundance of the most characteristic ions of each EEGs.

2.3.6 PLS regression

PLS regression was employed to find out if it were possible to associate FIA-HRMS data to the total flavonoid content and antioxidant capacity of EEGs [41]. Positive mode ions table displays approximately 2327 features per EEGs, while the negative mode given approximately 858. The two tables were merged to make a more representative dataset, which was used during modelling phase. The final table presented about 3185 ions for each EEG and then was used to build up PLS models.

The validation of the PLS models was carried out by internal and external validation methods [42]. Firstly, the dataset was submitted to a hierarchical clustering analysis (HCA), in order to observe how many groups were present (Annex I). Approximately 10 groups were observed. Subsequently, one sample was removed from each group to be used in the validation dataset. The remaining samples were used to compose the calibration model (also called "training" model). The Cross-validation method was used to assess the predictive ability of the models. This is a method of data resampling to test the ability of predictive models to estimate values to each sample in the tested group, being used to avoid overfitting [43]. In addition, three pre-processing methods (MSC, SNV, and S-G) were also tested and compared to the models without any data pre-treatment. Figure 18 presents the results.

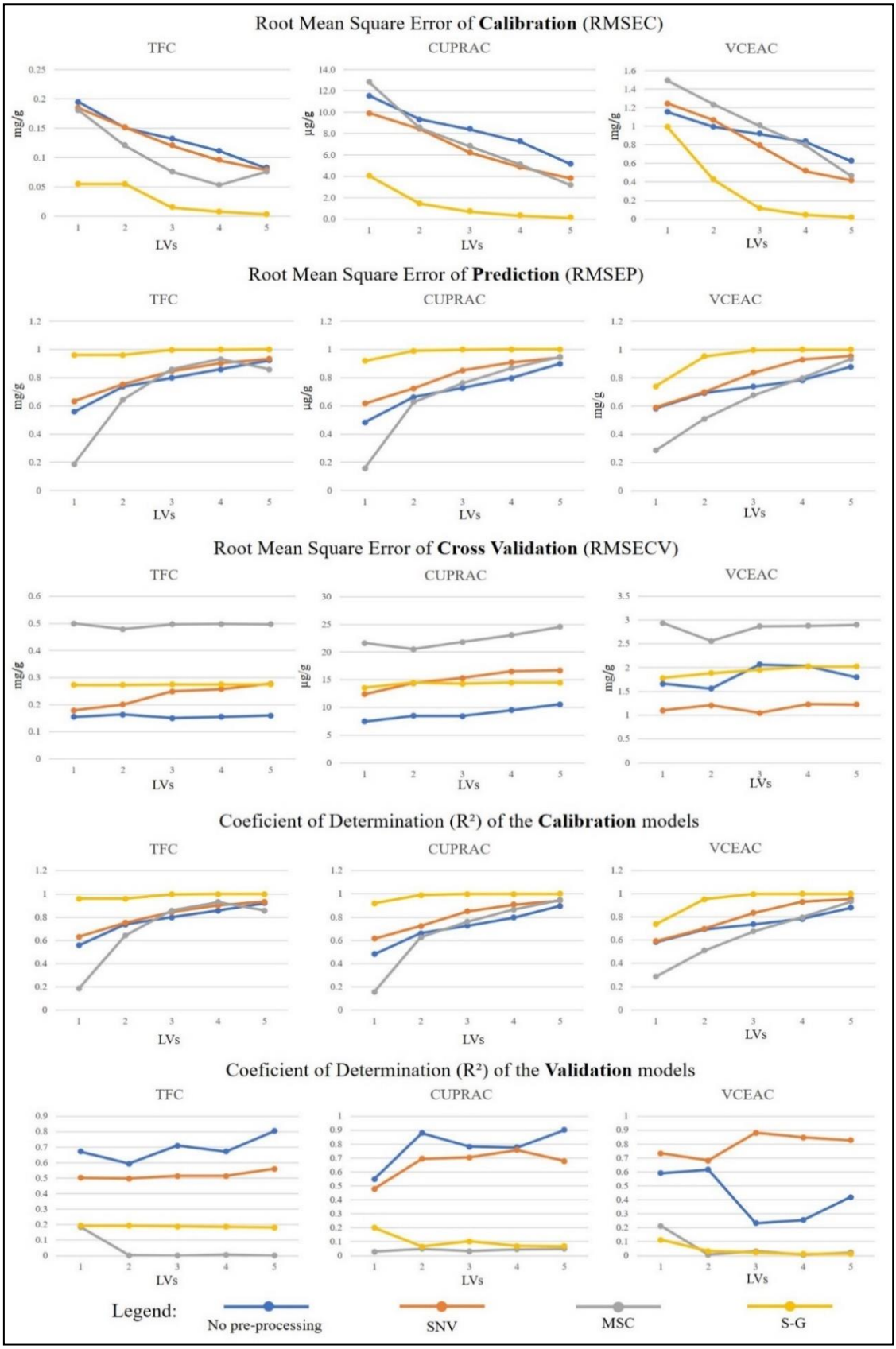


Figure 18 – Comparison of PLS models using the following pre-processing methods: SNV, MSC, and S-G; *vs* the number of LVs used to build the models.

As shown in Figure 18, the models using Savitzyk-Golay (S-G) smoothing algorithm always presented the best results for the calibration models. Those models presented the lowest RMSEC for TFC, CUPRAC and DPPH models; in addition to the highest values for R^2 of calibration (close to 1.0 in all cases), using a low number of latent variables (LVs). However, the S-G algorithm also presented the highest values for RMSEP and the lowest values for R^2 of validation ($0.2 > R^2 > 0.001$). The results of the internal cross validation (RMSECV) showed intermediate values when compared to the results achieved by those models in which pre-processing methods were applied. In addition, the results using the S-G method also presented the tendency of getting worse when more LVs are used. These evidences are indicating that this pre-processing method would give to the models a poor predictive capacity [33], and unsatisfactory accuracy.

MSC pre-processing had a similar pattern of results of the S-G method, but slightly worse in the case of TFC models. The MSC pre-processing also presented the highest values for the internal cross validation error (RMSECV), when compared to all other tested methods. Also, MSC pre-processing presented the lowest coefficients of determination, especially for the validation models ($0.01 < R^2 < 0.20$).

On the one hand, the S-G and MSC pre-processing methods provided better results for the calibration models. However, on the other hand they showed completely unsatisfactory results for the validation models, demonstrating a poor predictive capability. Apparently, not applying any pre-processing method made the models to present intermediate values for RMSEC, with slightly better R^2 of validation, and the lowest RMSECV for TFC, and CUPRAC, models. Thus, not applying any pre-processing method was chosen to be the best option for the TFC and CUPRAC models. The VCEAC, differently from the other models, presented more satisfactory results during the validation step when SNV was applied as a pre-processing method. The SNV method

provided the lowest values for RMSECV and the highest for R^2 of validation for VCEAC, as well as highly satisfactory values for RMSEP. The final PLS models features are summarised in the Table 3.

Table 3 – PLS modelling: analytical results and figures of merit.

Assays	***Pre-processing***	***LVs***	***RMSEC (mg/g)***	***RMSECV (mg/g)***	R^2_{Cal}	***RMSEP (mg/g)***	R^2_{Val}
TFC	None	**3**	0,1315	0,1500	**0,7988**	0,2646	**0,7096**
CUPRAC	None	**2**	$9{,}3429.10^{-3}$	$15{,}0902.10^{-3}$	**0,6613**	$8{,}4806.10^{-3}$	**0,8802**
VCEAC	SNV	**3**	0,7900	2,0662	**0,8355**	1,0448	**0,8815**

The models were built using a small number of LVs, which is another evidence of the quality of the models [33]. The RMSECV is defined as the standard deviation of the differences between the reference data (TFC, and AOC measurements), and the values predicted through the spectral dataset by PLS modelling. Consequently, RMSECV is interpreted as being a quantitative measurement of the accuracy achieved by the process of prediction, during the cross-validation. Therefore, the parameters of the calibration models were fundamentally focused on the minimisation of the RMSECV [33].

The final models are the ones that presented the lowest values for RMSEC, RMSEP, and RMSECV. In addition, the models also present highly satisfactory values for R^2, for both calibration and validation datasets. The models are represented in Figure ***19***, the red circles represent the calibration models, and the blue circles are the validation models.

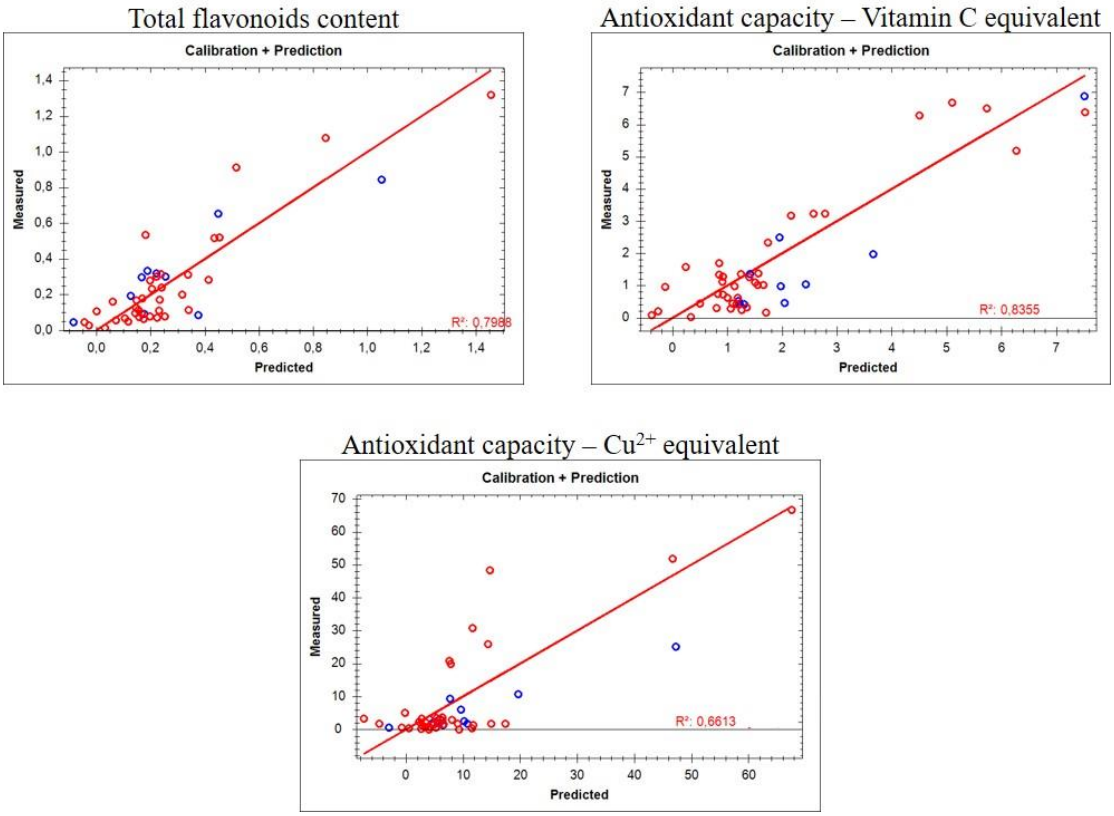


Figure 19 – PLS models based on EEGs fingerprints dataset for TFC, and AOC, reference tests.

According to the results observed for PLS modelling (Table 3), it is possible to estimate the TFC and AOC (DPPH and CUPRAC) of the EEGs using their mass fingerprints.

2.4 Partial conclusions of FIA-HRMS (fingerprints analysis)

FIA-HRMS fingerprints associated to PCA, HCA and Pearson correlations evidence that:

I. EEGs from different stingless bee species may be distinguished due to different metabolomic patterns.

II. Within geopropolis produced by the same species there are features associated to the geographic origin but the influence of this factor seems to have different extension according to the actual bee specie.

III. Biomarkers were assigned to geopropolis from each bee species and within the species to geopropolis from Prudentópolis region, as the larger number of samples came from this region.

Additionally, based on the spectral features of the EEGs, it was possible to build PLS models with considerable accuracy and good predictive abilities. By using PLS regression and mass fingerprints associated with chemometrics it is possible to characterise the EEGs. The regression models demonstrated capability to be used in preliminary analysis, of unknown samples, for predicting the properties: total flavonoid contents and antioxidant capacity in terms of vitamin C equivalent and cupric ion reduction power. Thus, it is possible to have a preliminary idea about the composition (in terms of flavonoid content) and antioxidant activity of an EEG by acquiring a mass fingerprint spectrum, being an accurate and reliable estimate.

Although HRMS is an expensive analytical technique, once a workflow is established, the cost per sample will decrease. Additionally, the advantage of this method is that there is no need to carry out classical tests on new samples, consequently decreasing the use of organic solvents and chemical reagents, reducing the exposure of the analyst to potentially dangerous substances. This methodology also shows potential to be used for the quality control of geopropolis extracts, once it allows the analysis of massive quantities of samples at once, in a relatively short period of time, and not requires complicated sample preparation steps.

Chapter 3 – Untargeted Metabolomic Analysis of the EEGs by UHPLC–HRMS (Orbitrap)

3.1 Objectives

Part 3 of this study aims to explore the general chemical profile of the EEGs employing UHPLC-HRMS, and bioinformatic tools, to identify compounds of interest.

I. Explore the chemical profile of the EEGs, employing UHPLC-HRMS, and bioinformatic tools, to obtain qualitative information about the EEGs main composition, through an untargeted metabolomic approach;

II. Classify the EEGs based on their metabolomic profile, regarding the species of the stingless bees, and geographical origins, using multivariate analysis;

III. Highlight the specificities in the composition of each group of EEGs, using multivariate analysis to encounter characteristic compounds, and possible biomarkers.

3.2 Methodology

3.2.1 EEGs preparations for UHPLC-HRMS analysis

The untargeted metabolomic analysis were performed in both negative and positive modes of ionisation, and two batches of samples were prepared, using 96-wells microplates. Firstly, 100 µL of the EEGs solutions were dried in vacuum chamber, then the dried portion was resuspended by adding 100 µL of methanol. To ensure a complete dissolution, ultrasonic bath was used for 5 min and then spin down during 10 min at 14600 rpm. Subsequently, 20 µL of the supernatant were added in a 96-well plate, followed by

the addition of 180 µL of methanol to a final concentration of 10 µg.mL^{-1}. Five microlitres were taken automatically to chromatographic analysis. Figure 20 shows the workflow followed to the LC-HMRS analysis.

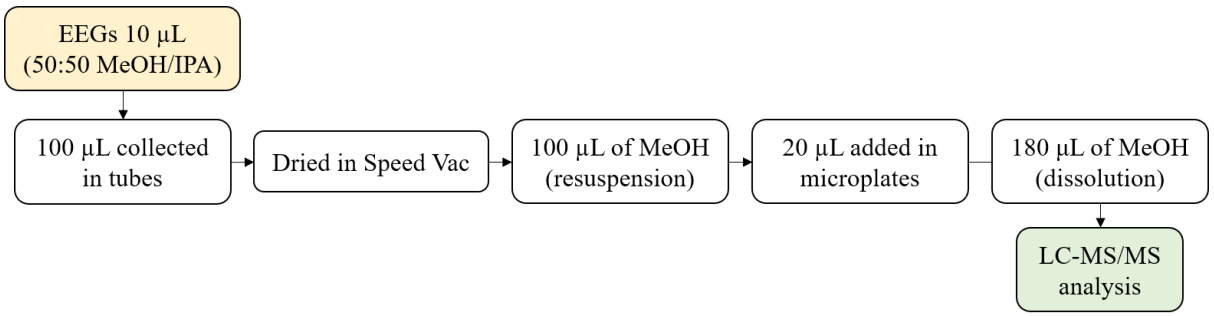


Figure 20 – EEGs preparation for UHPLC-HRMS analysis workflow.

3.3.2 Instrument setup and spectra acquisition

Ultra-high-performance liquid chromatography was performed using the instrument Q Exactive Quadrupole-Orbitrap mass spectrometer (Thermo), equipped with a binary pump, an autosampler, an online vacuum degasser, and a temperature-controlled column compartment. The high-resolution mass spectrometry was performed on the Orbitrap Q-Exactive mass spectrometer (Thermo), already cited in the Part 3 of this work. A Restek UHPLC C18 (2.1 × 46 mm, 1.9 µm, USA) reversed-phase column was used for the analysis. The time of analysis was 11.5 min and the flow rate was 0.80 mL.min^{-1}. The mobile phase was composed by acid solutions of water (A) and acetonitrile (B), both with 0.1% v/v formic acid. The gradient of elution was the following: 45% of B from 0 to 1 min; from 1 to 8 min constantly increasing of B until reach 100%; staying in 100% of B from 8 to 11.5 min and analysis finishes; from 11.5 to 12 min the amount of B decreases to 45% of B; from 12 to 15 min the system returns to its initial conditions, in order to be prepared for another injection. The spectra were acquired in full scan, recorded between 50 and 1000 m/z, and the mass resolution was 140 000.

The source conditions for every experiment were as follows: ionisation mode, positive or negative; sheath gas, 60 AU; auxiliary gas, 10 A.U.; sweep gas, 3 A.U.; spray voltage, 3.50 kV in positive mode and −4.0 kV in negative mode; heater temperature, 320 °C; ion transfer capillary temperature, 320 °C; and S-lens RF level, 50.0, scan range of 75–1500 m/z; polarity, negative or positive; spectrum data type centroid. Thermo Xcalibur V. 3.0.63 was used for data handling.

3.3.3 Compound identification/annotation

Compound identification was carried out by the Compound Discoverer V. 3.3 software, installed on a Dell computer equipped with an Intel Xeon, CPU E5-1650 V. 4 @ 3.60GHz, and 64 GB of memory RAM, Windows 10 Professional V. 20H2, 64-bit operating system.

Compound identification/annotation took in consideration the degrees of confidence described by Rochat [44], shown in Figure 21. According to Rochat [44], compounds identified using real standards, and/or identification by NMR spectroscopy, are tier one of confidence, being attested as confirmed compounds; tier two is for the "high confidence" identifications/annotations, when the spectra acquired from the matrix of study is compared to a spectral library, and enough similarities are found; tier three considers the annotations made by exact mass, without comparing the any spectral data; tier four are made by assumption, when only the formula of the compound is identified through calculations, without any other type of information matching the unknown compound.

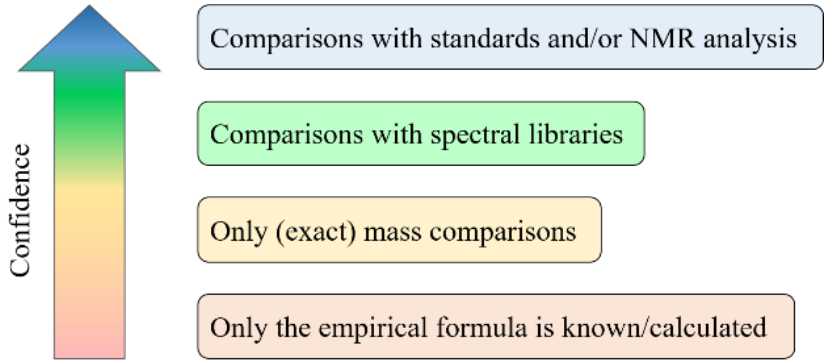


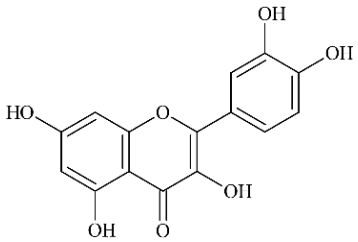
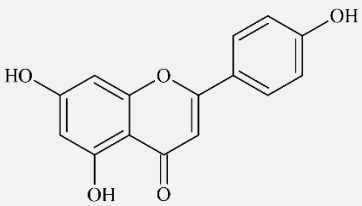
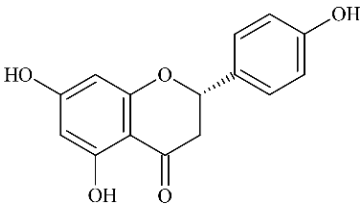
Figure 21 – Scale proposed by Rochat [44] for the confidence degree for compound identification/annotation.

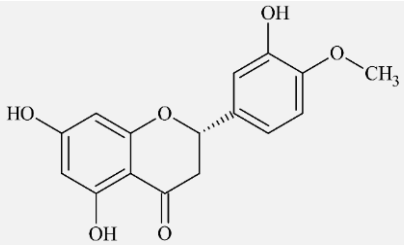
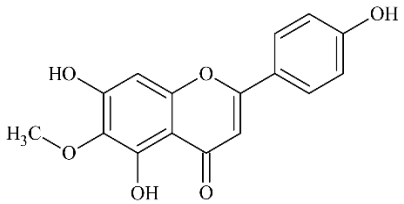
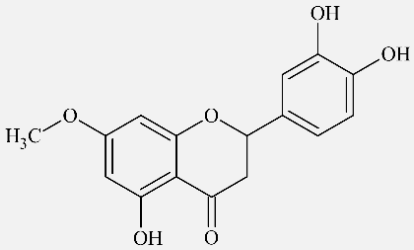
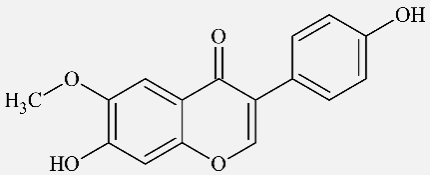
3.3 Results and discussion

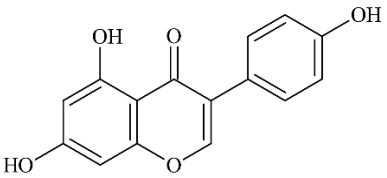
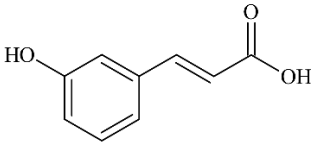
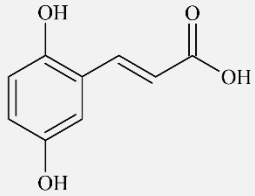
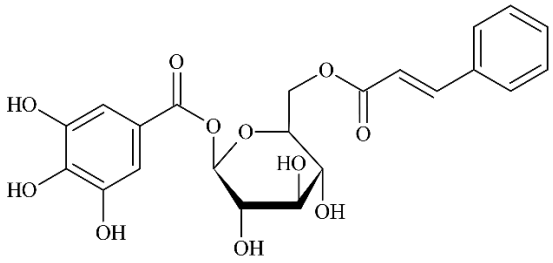
3.3.1 Composition of the ethanolic extract of geopropolis

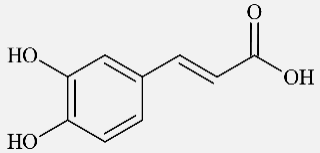
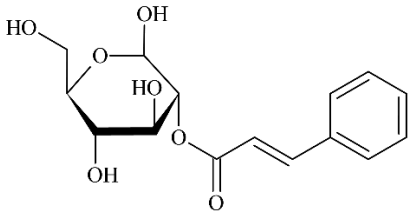
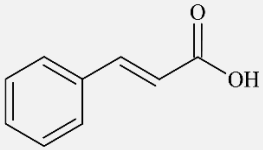
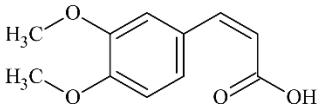
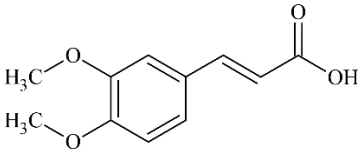
From the analysis of the EEGs by UHPLC-HRMS, 68 compounds were identified, based on MS data, with elevated degree of confidence [44], and presented in Table 4.The composition of geopropolis is complex, and in this study compounds belonging to several classes such as flavonoids, isoflavones, phenylpropanoids, phenolic compounds, coumarins, retinoids, sugars and acid sugars, aurones, aromatic hydrocarbons, and tannins were found.

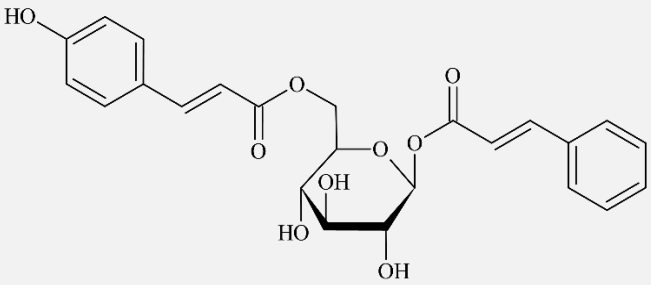
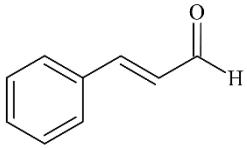
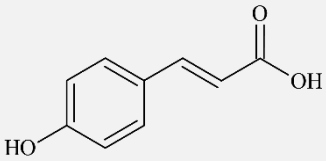
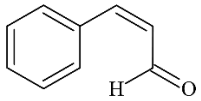
Table 4 – Untargeted metabolomic analysis: compounds identified with the most elevated level of confidence possible – continues on next page.

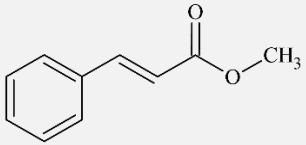
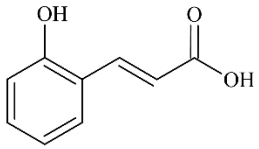
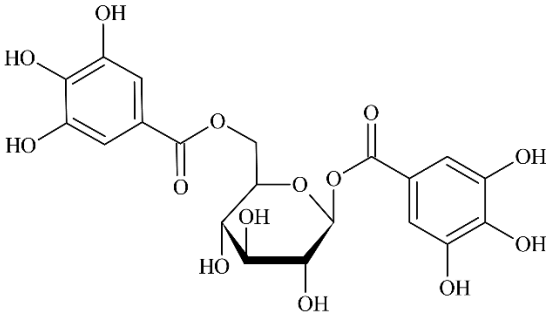
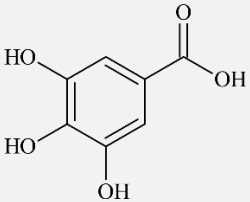
RT [min]	Identified compound	Molecular formula	MW	Δmass [ppm]	Reference
	Flavonoids				
2.969	Quercetin 	$C_{15}H_{10}O_7$	302.0427	0.95	[45]
3.632	Apigenin 	$C_{15}H_{10}O_5$	270.0528	0.87	[45]
3.818	Naringenin 	$C_{15}H_{12}O_5$	272.0685	0.61	[46]
3.958	Hesperitin	$C_{16}H_{14}O_6$	302.079	0.74	

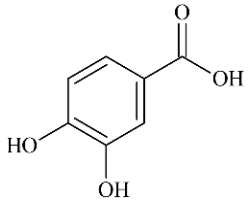
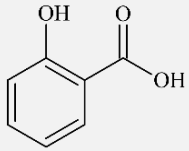
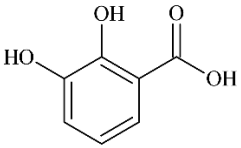
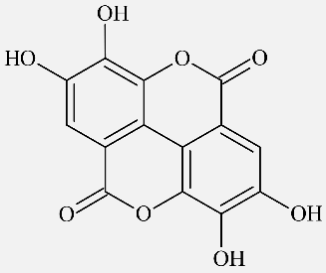
					
	Hispidulin				
4.088		$C_{16}H_{12}O_6$	300.0634	1.02	[45]
	Sternbin				
4.344		$C_{16}H_{14}O_6$	302.079	0.24	
5.743	*Isoflavones*				
	Glycitein				
7.271		$C_{16}H_{12}O_5$	284.0685	0.83	
	Genistein	$C_{15}H_{10}O_5$	270.0528	0.52	

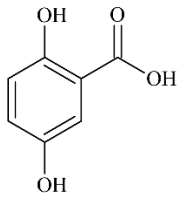
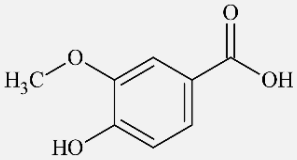
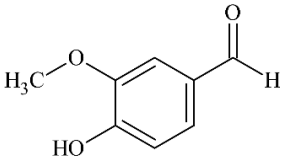
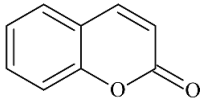
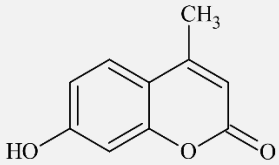
					
	Phenylpropanoids				
	m-Coumaric acid				
1.619		$C_9H_8O_3$	164.0473	0.31	
	2,5-Dihydroxycinnamic acid				
1.670		$C_9H_8O_4$	180.0423	0.74	
	6-O-[(2E)-3-Phenyl-2-propenoyl]-1-O-(3,4,5-trihydroxybenzoyl)-β-D-glucopyranose				
2.097		$C_{22}H_{22}O_{11}$	462.1162	0.56	
2.133	Caffeic acid	$C_9H_8O_4$	180.0423	1.00	[47]

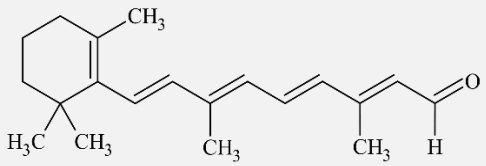
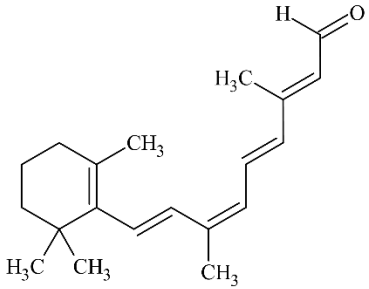
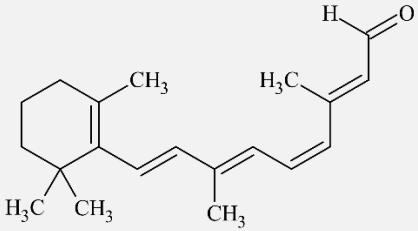
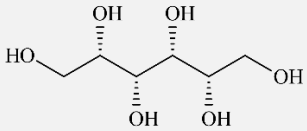
					
	[(2R,3R,4S,5S,6R)-2,4,5-trihydroxy-6-(hydroxymethyl)oxan-3-yl] (E)-3-phenylprop-2-enoate				
2.947		$C_{15}H_{18}O_7$	310.1053	0.14	
	Cinnamic acid				
2.970		$C_9H_8O_2$	148.0524	0.55	[48]
	(Z)-3,4-dimethoxycinnamic acid				
2.969		$C_{11}H_{12}O_4$	208.0736	0.74	
	Phenylpropanoids				
	o-Methylferulic acid				
3.134		$C_{11}H_{12}O_4$	208.0736	0.58	

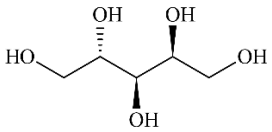
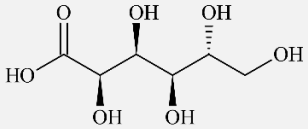
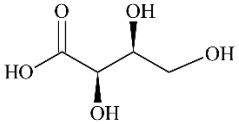
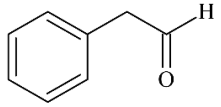
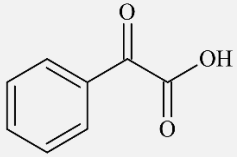
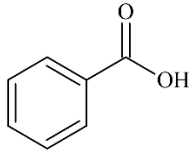
3.244	6-O-[(2E)-3-(4-Hydroxyphenyl)-2-propenoyl]-1-O-[(2E)-3-phenyl-2-propenoyl]-β-D-glucopyranose 	$C_{24}H_{24}O_9$	456.142	0.64
3.947	*trans*-Cinnamaldehyde 	C_9H_8O	132.0575	1.05
4.932	*p*-Coumaric acid 	$C_9H_8O_3$	164.0473	1.12
4.940	*cis*-Cinnamaldehyde 	C_9H_8O	132.0575	1.13
5.065	Methyl-cinnamate	$C_{10}H_{10}O_2$	162.0681	1.20

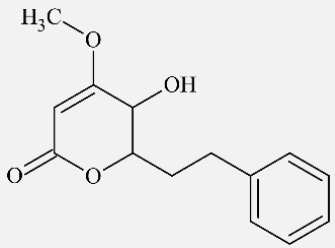
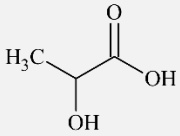
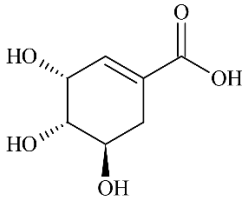
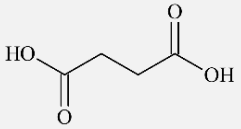
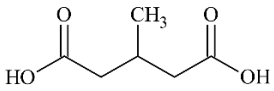
					
	o-Hydroxycinnamic acid				
5.133		$C_9H_8O_3$	164.0473	1.55	
	Phenolic Compounds				
	1-O,6-O-digalloyl-β-D-glucose				
1.456		$C_{20}H_{20}O_{14}$	484.0853	0.42	
	Gallic acid				
1.582		$C_7H_6O_5$	170.0215	0.88	[49]
1.668	Protocatechuic acid	$C_7H_6O_4$	154.0266	0.57	[49]

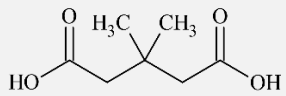
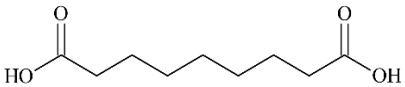
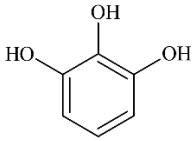
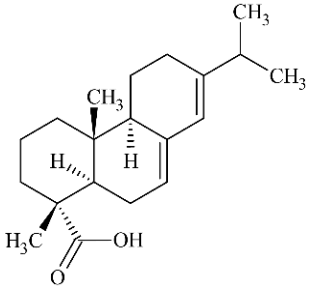
					
	Salicylic acid				
1.773		$C_7H_6O_3$	138.0317	1.07	
	2,3-Dihydroxybenzoic acid (*o*-Pyrocatechuic acid)				
2.044		$C_7H_6O_4$	154.0266	0.58	
	Ellagic acid				
2.066		$C_{14}H_6O_8$	302.0063	0.63	[50]
2.202	2,5-Dihydroxybenzaldehyde (Gentisaldehyde)	$C_7H_6O_3$	138.0317	0.97	

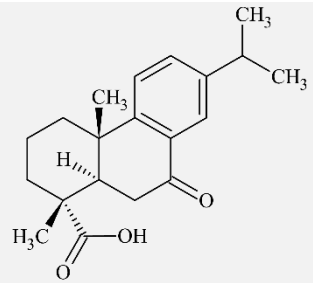
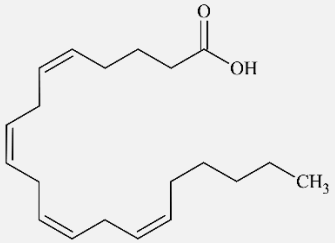
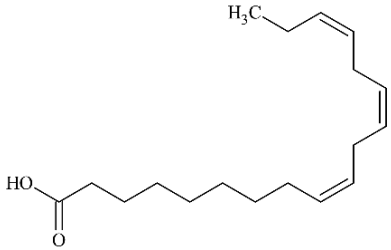
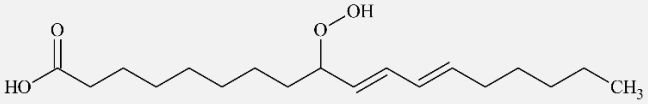
					
	Vanillic acid				
2.391		$C_8H_8O_4$	168.0423	1.69	[48]
	Vanillin				
2.787		$C_8H_8O_3$	152.0473	0.50	[51]
	Coumarins				
	Coumarin				
1.586		$C_9H_6O_2$	146.0368	1.11	
	4-Methylumbelliferone hydrate (Hymecromone hydrate)				
6.437		$C_{10}H_8O_3$	176.0473	0.78	
	Retinoids				

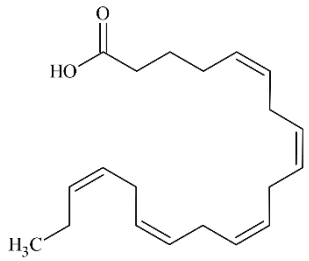
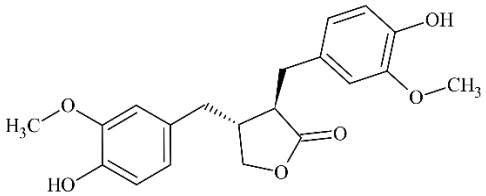
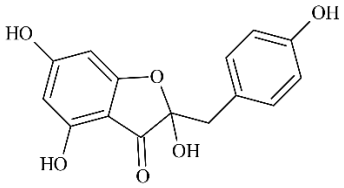
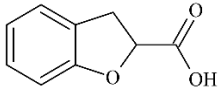
	All-trans-Retinal			
6.736		$C_{20}H_{28}O$	284.214	0.02
	9-*cis*-Retinal			
7.371		$C_{20}H_{28}O$	284.214	0.29
	11-*cis*-Retinal			
8.039		$C_{20}H_{28}O$	284.214	0.2
	Sugars			
	L-Iditol			
1.406		$C_6H_{14}O_6$	182.079	0.73
1.426	L-(-)-Arabitol	$C_5H_{12}O_5$	152.0685	0.91

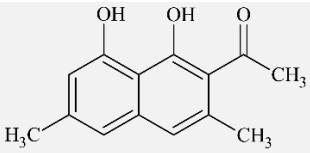
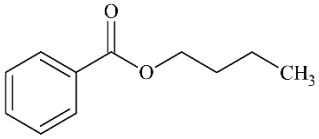
					
	Gluconic acid				
1.439		$C_6H_{12}O_7$	196.0583	0.98	
	L-Threonic acid				
1.470		$C_4H_8O_5$	136.0372	1.16	
	Aromatichydrocarbons				
	Phenylacetaldehyde (Hyacinthin)				
2.288		C_8H_8O	120.0575	1.9	
	Phenylglyoxylic acid				
2.567		$C_8H_6O_3$	150.0317	0.89	
	Benzoic acid				
2.651		$C_7H_6O_2$	122.0368	0.89	[48]

5.482	Dihydrokawain-5-ol 	$C_{14}H_{16}O_4$	248.1049	0.25	
	Organic acids				
1.412	DL-Lactic acid 	$C_3H_6O_3$	90.03169	0.87	[13]
1.476	Shikimic acid (3,4,5-trihydroxycyclohex-1-ene-1-carboxylic acid) 	$C_7H_{10}O_5$	174.0528	1.36	
2.066	Succinic acid 	$C_4H_6O_4$	118.0266	1.05	
2.082	3-Methylglutaric acid 	$C_6H_{10}O_4$	146.0579	0.61	

2.305	3,3-Dimethylglutaric acid 	$C_7H_{12}O_4$	160.0736	0.58	
2.456	Azelaic acid 	$C_9H_{16}O_4$	188.1049	0.77	
	Tannins				
1.415	Pyrogallol 	$C_6H_6O_3$	126.0317	0.74	[51]
	Diterpenoids				
6.894	Abietic acid 	$C_{20}H_{30}O_2$	302.2246	0.10	[52]
8.027	7-Oxodehydroabietic acid	$C_{20}H_{26}O_3$	314.1882	0.20	[52]

					
	Fatty acids				
7.581	Arachidonic acid 	$C_{20}H_{32}O_2$	304.2402	0.06	[13]
8.291	α-Linolenic acid 	$C_{18}H_{30}O_2$	278.2246	0.16	[53]
8.32	(10E,12E)-9-Hydroperoxy-10,12-octadecadienoic acid 	$C_{18}H_{32}O_4$	312.2301	0.16	
9.041	Timnodonic acid (eicosapentaenoic acid)	$C_{20}H_{30}O_2$	302.2246	0.10	

				
	Lignans			
	Matairesinol			
3.939		$C_{20}H_{22}O_6$	358.1416	0.00
	Aurones			
	Maesopsin			
1.711		$C_{15}H_{12}O_6$	288.0634	0.55
	Othercompounds			
	2,3-Dihydro-1-benzofuran-2-carboxylic acid			
2.287		$C_9H_8O_3$	164.0473	0.09
5.978	1-(1,8-dihydroxy-3,6-dimethyl-2-naphthyl)-ethan-1-one	$C_{14}H_{14}O_3$	230.0943	0.46

				
	Butylbenzoate			
7.358			$C_{11}H_{14}O_2$	178.0994

3.3.2 Exploratory multivariate analysis of the UHPLC-HRMS data

Firstly, PCA was tried to visualise relationships among EEGs. Figure 22 presents the PCA plotted graphic, for the four species that are represented by more than four samples of geopropolis (*M. marginata*, *M. quadrifasciata*, *M. scutellaris*, and *T. angustula*). The PC1 and the PC2 were chosen to represent the analysis and together accomplished for more than 40 % of the total data variance. The 95% confidence interval ellipses were not showed just to enhance the visualisation of the PCA results.

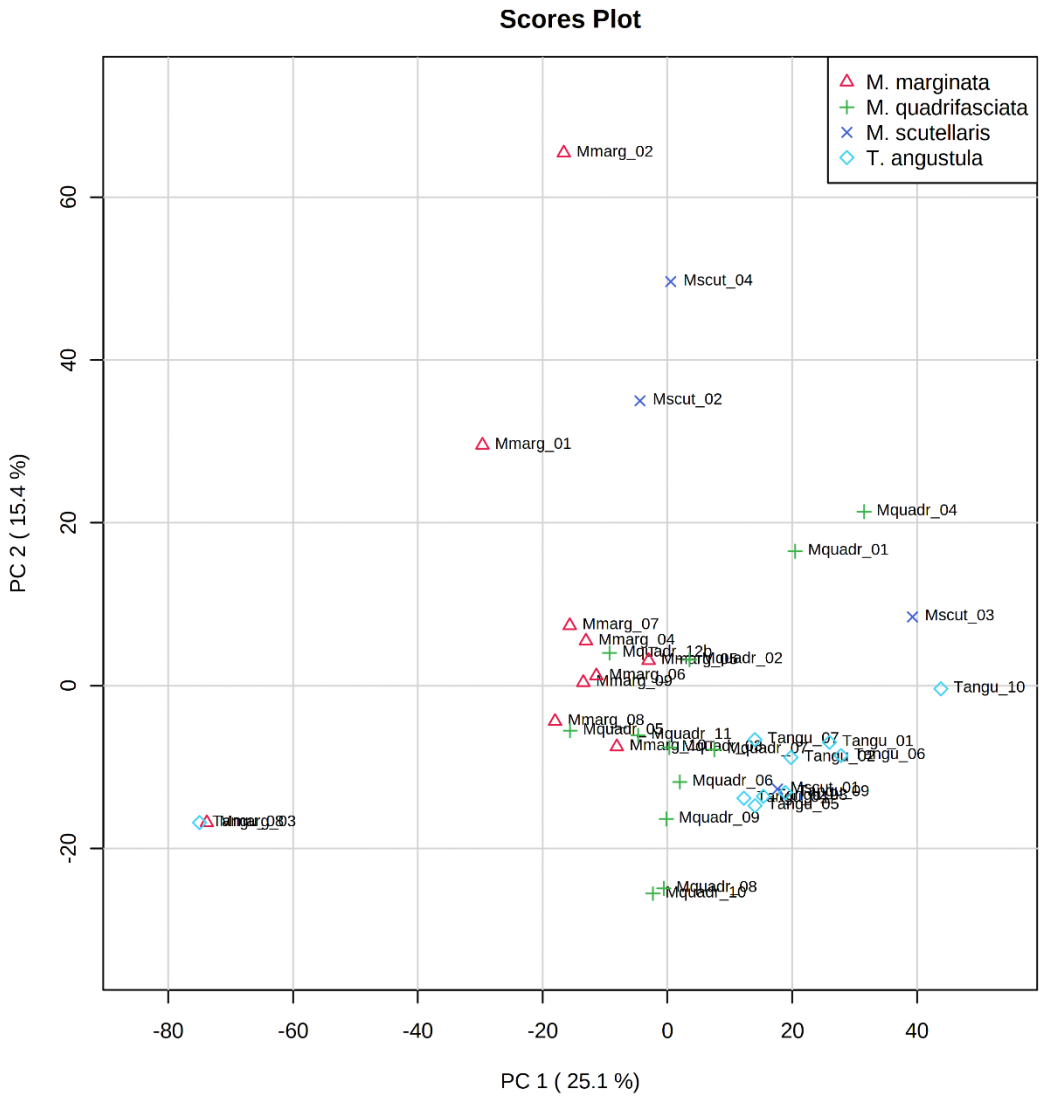
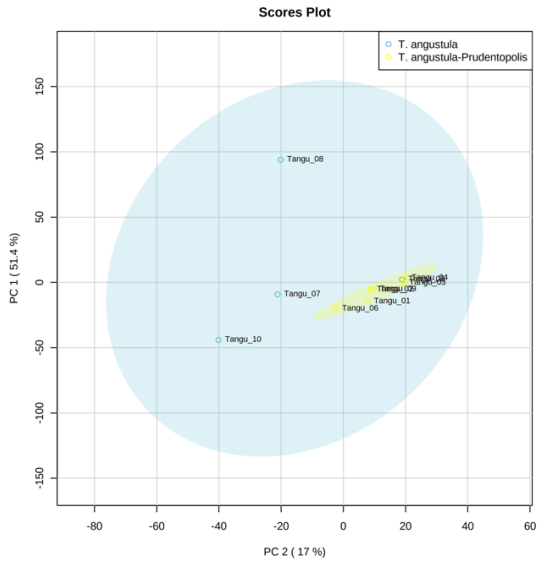


Figure 22 – PCA based on the UHPLC-HRMS dataset of the four groups with more than three samples.

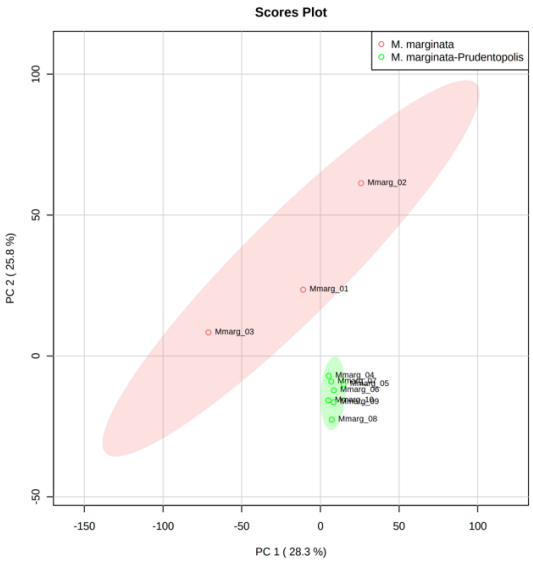
The PCA (Figure 22) shows the EEGs tendency to form clusters together, with groups following the species of the producer stingless bee; the only exception might be the *M. scutellaris*. However, in the same graphic, it is noticeable that a portion of the EEGs are spread around the main clusters, and this behaviour is not likely to be aleatory, as well as demonstrated in the Part 2 of this study: the geographic and the biologic factor play important roles in the geopropolis composition. The three larger groups of EEGs (*M. marginata*, *M. quadrifasciata*, and *T. angustula*) presented clustered groups relatively near to each other, and those are representatives of EEGs of geopropolis produced in the Prudentópolis region in Brazil. The other EEG were made out of samples from diverse regions of Brazil, as depicted by the Figure 2.

In the case of *M. marginata* EEGs, the difference between the "subgroups" is rather noticeable, with three samples from other regions spread along the PC2, albeit in the bottom of the graphic there is a cluster of EEGs made of geopropolis produced in Prudentópolis region. To compare with the *M. marginata* grouping pattern, the EEGs from *T. angustula* have only a small trend of division and have not statistical differences between samples from Prudentópolis and other regions, in a 95% interval of confidence, thus reinforcing our assumption about a selective gathering pattern of this bee when looking for vegetal resins to make its geopropolis. To further illustrate this fact, shown in Figure **23**, a PCA were plotted considering EEGs from *T. angustula*, *M. quadrifasciata* and *M. marginata* separately.

T. angustula
X
T. angustula - Prudentópolis



M. marginata
X
M. marginata - Prudentópolis



M. quadrifasciata
X
M. quadrifasciata - Prudentópolis

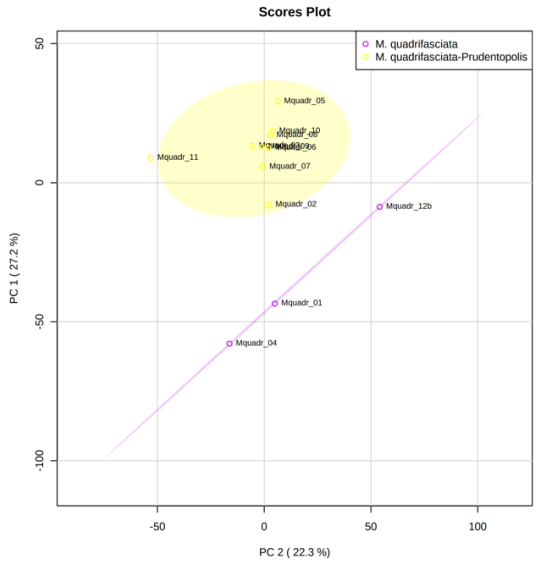


Figure 23 – PCA score plots contrasting the geopropolis origins.

The geographic region where the samples of geopropolis were produced is probably the main reason for the observed differences between the subgroups of species; mainly for *M. marginata* and *M. quadrifasciata*. However, this geographic effect on the composition was not statistically significant to composition of EEG of *T. angustula*.

The second group of EEGs analysed were the ones formed by smaller number of samples from each bee specie, and they were also analysed in an exploratory PCA. The graphic of PCA is depicted by the Figure 24, the two groups, *Melipona* genus and the other species.

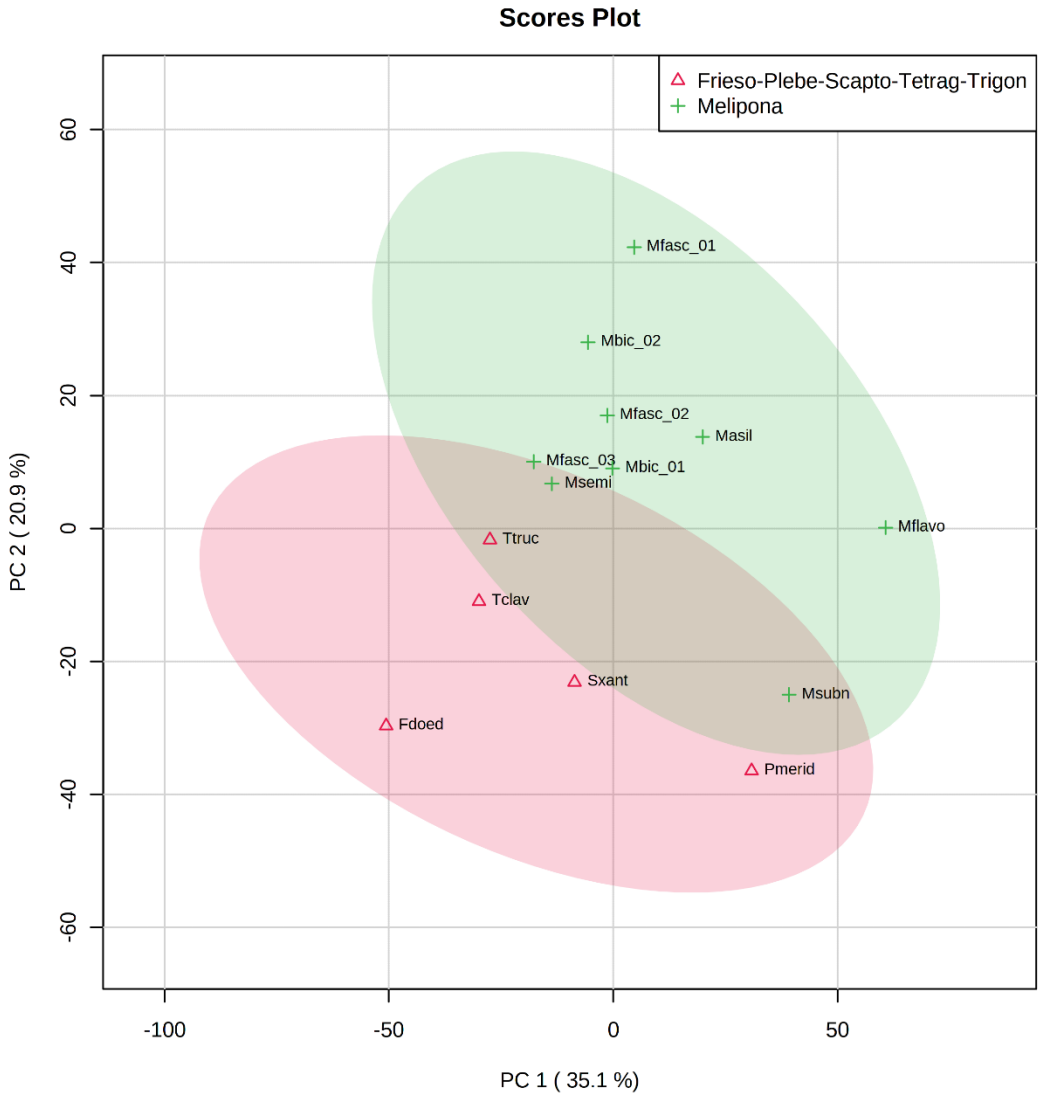


Figure 24 – PCA based on the UHPLC-HRMS dataset form the *Melipona* group and samples from others genus.

Although the two groups present themselves slightly overlapped, they present significative differences when a t-test was applied, a similar result to those of the three previous Parts of this study, for the same two groups. The t-tests showed that 87 features were found significant, with p-value > 0.05, to distinguish the two groups from each other.

3.3.3 EEGs most characteristic compounds assessed by UHPLC-HRMS

The most characteristic compounds found the EEGs were described in Table 5. The compounds were chosen by picking the lowest p-value, using the ANOVA's table of loadings. The compounds were annotated according to the Compound Discoverer V. 3.3 software, by comparison between the detected exact mass, isotopic pattern, and MS^2 spectral matching with available libraries.

Table 5 – Most characteristic compounds in the EEGs identified by LC-HRMS.

Specie	Compound name	Formula	RT (min)	Δmass (ppm)	$-Log_{10}(p)$	*f*–value
M. marginata	4'-O-methyldavidigenin	$C_{16}H_{16}O_4$	7.104	-0.45	2.204	3.455
M. marginata* - *Prudentópolis	Ikarisoside D	$C_{28}H_{30}O_{11}$	3.852	-3.32	6.650	10.181
M. quadrifasciata	-	$C_{31}H_{22}O_{10}$	2.095	0.26	3.729	5.385
M. quadrifasciata* - *Prudentópolis	Salicylic acid	$C_7H_6O_3$	1.773	-1,07	2.434	3.646
M. scutellaris	-	$C_8H_6O_6$	2.807	-0.04	2.547	7.791
T. angustula	Melilotoside A2	$C_{36}H_{58}O_9$	8.405	-0.72	1,981	3.075
Melipona*	-	$C_{23}H_{28}O_5$	11.103	0.01	3.961	4.273
Non-*Melipona* **	-	$C_{30}H_{54}O_5$	9.497	0.76	2.284	3.456

Figure 25 shows the boxplots of the loadings from ANOVA. The results were shown as total ion abundance, in each EEG group. In Table 5, 4'-O-methyldavidigenin,

Ikarisoside D, (E)-4,6-dioxo-2-Octenedioate, and Melilotoside A2 were annotated by Compound Discoverer V 3.3 based only on exact mass comparison, thus an identification classified as intermediate by the confidence degree levels for compound identification described in Part 4, section 4.4.3. Using the same parameters, salicylic acid was identified with strong confidence degree, by comparisons with mass spectra from inhouse and online libraries, by Compound Discoverer V 3.3 software.

The compounds annotated as main characteristic for each EEG have not been yet reported in the geopropolis literature. This is the first time those compounds were identified in any kind of geopropolis, which represents an advance in the comprehension and knowledge of geopropolis composition.

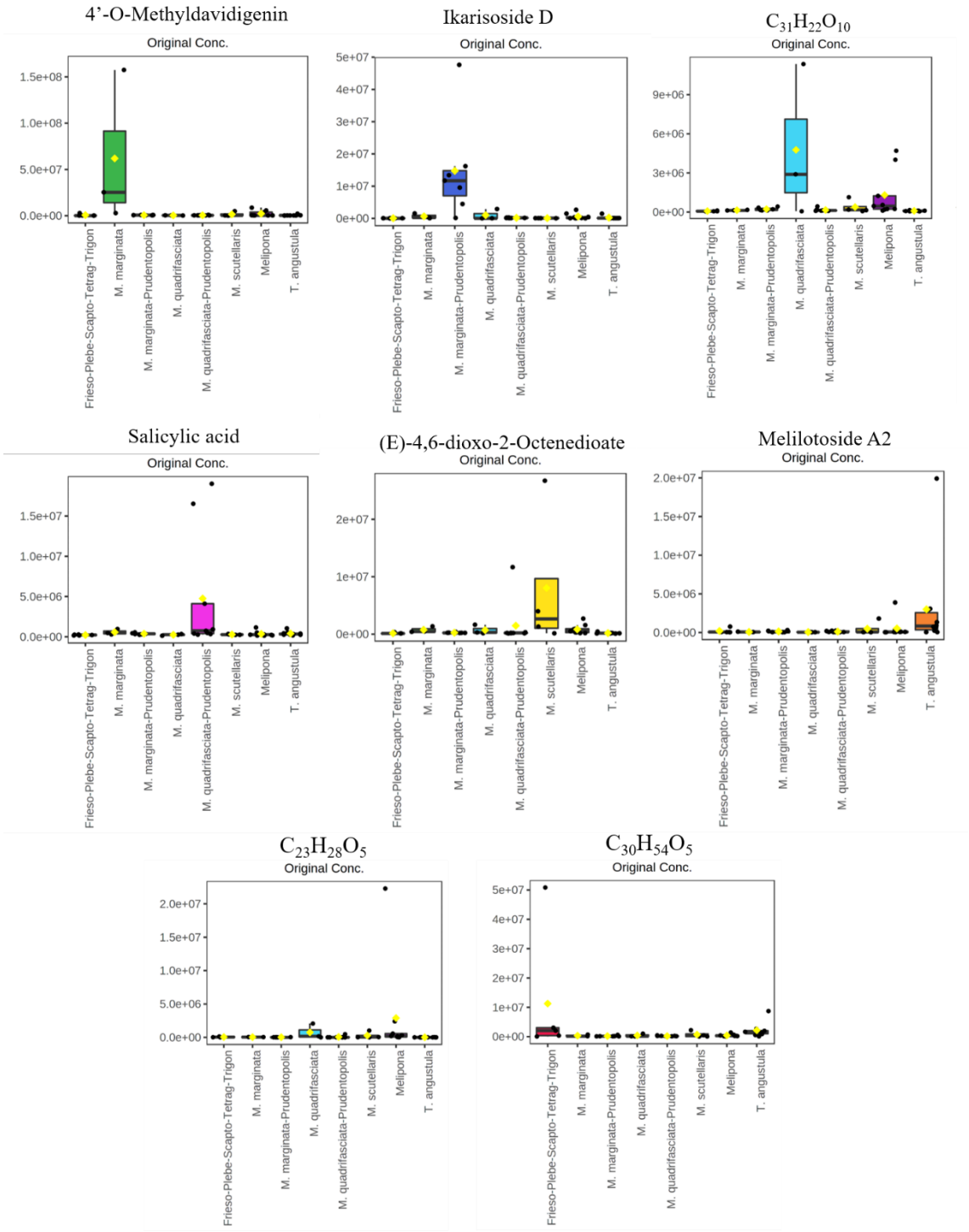


Figure 25 – Characteristic compounds of each EEG base in the ANOVA.

3.4 Partial conclusions from UHPLC-HRMS analysis of EEGs

The untargeted metabolomic approach of geopropolis revealed several compounds. Many of them, for the first time detected in geopropolis, such as retinoids,

kahweol, sorbicillin, prostaglandins, longistylin, among others. Also, several kinds of flavonoids, phenolics, and sugars were identified, which are classes of compounds expected to exist in geopropolis composition [45, 54, 55].

Many studies about propolis and geopropolis pointed out that the geographical origin have a huge impact on geopropolis composition. However, during this study we observed that the bee's species and genus also affect the stingless bee preferences for sources of resins, to make geopropolis. The factor species strongly influences the geopropolis composition.

One of the main advantages of using the metabolomic approach altogether to UHPLC-HRMS techniques, is the possibility of comparison of the datasets to spectral libraries, which is impossible to FIA-HRMS technique. Thus, the metabolomic analysis of the EEGs allowed the identification, with high degree of confidence [44], of 68 compounds.

Chapter 4 – Untargeted Lipidomic analysis of the less polar fraction of EEGs by UHPLC–HRMS (Orbitrap)

4.1 Objectives

Part 4 of this study aims to explore the lipidomic profile of the EEGs, while identifying the main, most abundant, and/or the characteristic compounds in each group of EEGs. The main objectives were to:

I. Identify compounds, from diverse lipidic classes, in the less polar fraction of the EEGs using UHPLC-HRMS (Orbitrap) to assess the lipidomic profile of the EEGs;

II. Classify the EEGs based on their lipid profile, employing chemometric tools to identify which compounds characterise and/or differentiate each group.

4.2 Methodology

4.2.1 Lipidomic analysis by UHPLC-HRMS (orbitrap)

The lipidomic analysis was carried out using the methodology developed by Hall et al. [23] with small adaptations. This methodology is commonly employed as standard method in the Metabolomics and Proteomics facility (MRC London Institute of Medical Sciences). The lipidomics analysis of the lipophilic fraction of the EEGs was carried out using a Thermo Q-Exactive (Orbitrap) mas spectrometer coupled to Vanquish LC (Thermo, USA). Five microliters of sample were injected onto a Restek C18 column (Restek Ltd., USA; 503 2.1 mm, 1.7 mm) maintained at 55 ºC. The mobile phase A was 10 mM ammonium formate in acetonitrile/water (60:40) and mobile phase B was 10 mM ammonium formate in isopropanol/acetonitrile (90:10). The flow rate was 0.5 $mL.min^{-1}$.

A heated electrospray ionisation source was maintained at 40 ºC, the desolvatation temperature was 380 ºC, and the desolvatation gas flow was set at 40 arbitrary units. Spectra were acquired in positive and negative ion mode in the range of 100-2000 m/z.

4.2.2 Compound identification

Compound identification/annotation was carried out by Compound Discoverer V. 3.3, using online and in house specialised spectral libraries.

4.3 Results and discussion

4.3.1 Composition of the organic lipid-rich fraction of the EEGs

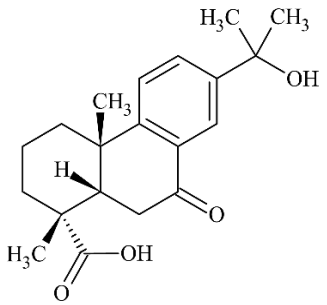
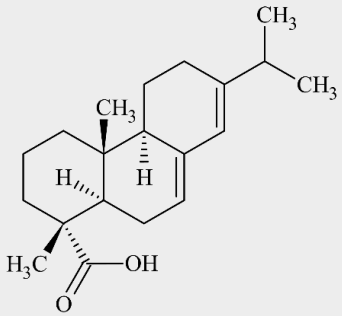
During the lipidomic analysis 23254 features were detected throughout the composition of the EEGs. However, only 61 compounds were identified with strong degree of confidence, following the criteria proposed by Rochat [44], briefly described in the item 4.4.3 of the previous section. The compounds are presented in Table 6.

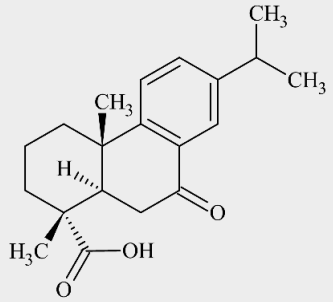
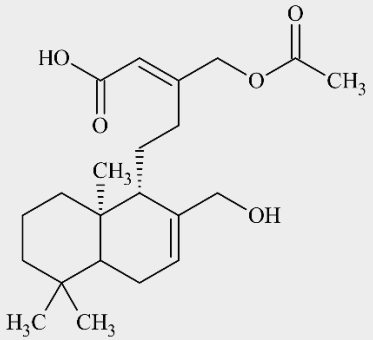
According to Lavinas et al. [5], terpenoids, mainly the mono and sesquiterpenoids, obtained directly from plant resins are highly relevant for the stingless bees. In plants, diterpenic acids present an important role as defensive compounds against potential pathogens and predatory herbivores [56]. Natural abietanoids (cyclic diterpenoids) have demonstrated a large spectrum of biologically interesting activities, in different studies and reviews in the literature. Perhaps, antimicrobial, antiulcer and cardiovascular activities are the most representative pharmacological effects associated to this class of diterpenoids [57].

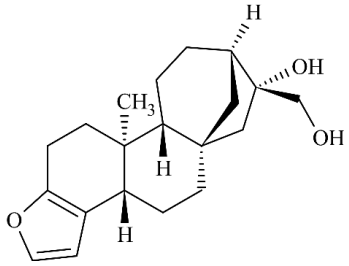
Two of the cyclic diterpenoids found in the EEGs organic lipid-rich fraction were kahweol and cafestol (Figure 26). The molecular structure of both compounds is

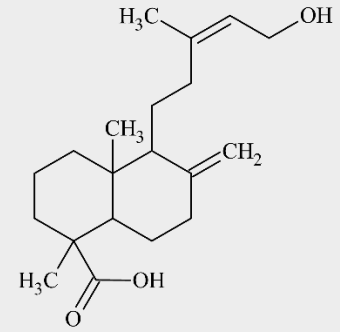
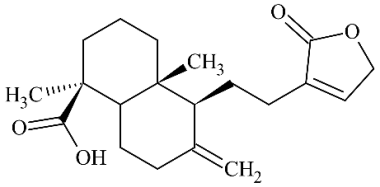
extremely similar, with kahweol having an extra double bond in the second ring. These diterpenes are commonly found in coffee [58], and presents notable benefits to human health: antioxidant activity, anti-inflammatory effect, and protection against cancer and toxic substances. In addition to the effect of raising serum lipid, *in vitro* and *in vivo* experimental results have revealed that the two diterpenes demonstrate multiple potential pharmacological actions such as anti-inflammation, hepatoprotective, anti-cancer, anti-diabetic, and anti-osteoclastogenesis activities. The most relevant mechanisms involved are down-regulating inflammation mediators, increasing glutathione inducing apoptosis of tumorous cells and anti-angiogenesis. Cafestol and kahweol show similar biological activities, however those compounds are not exactly the same, which might be due to the presence of one conjugated double bond on the furan ring in kahweol [58].

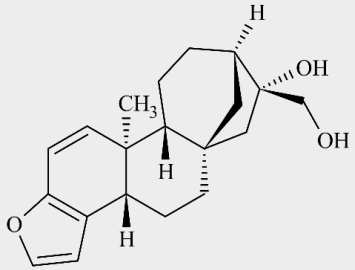
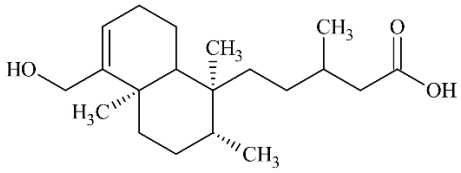
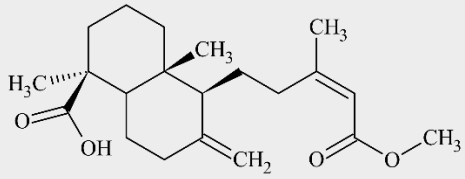
Table 6 – Compounds identified in the organic lipid-rich fraction of the EEGs using LC–HRMS (Orbitrap). Compound classes are in italics. – Part I.

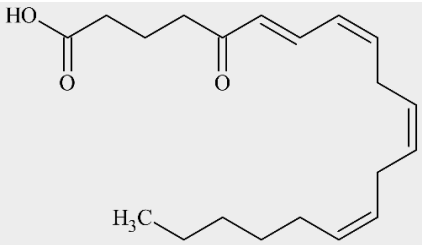
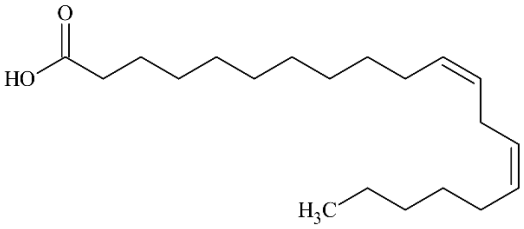
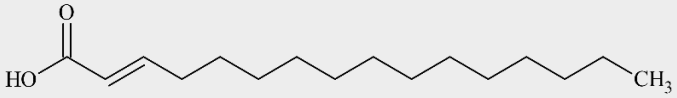
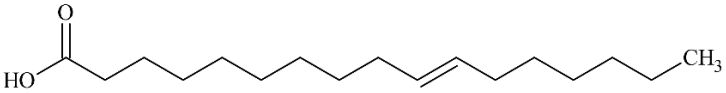
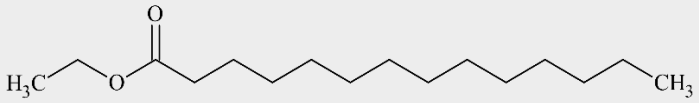
RT [min]	Compound	Molecular formula	Ionization mode	Detected mass	ΔMass [ppm]	Compound Discoverer matching (%)
	Abietanoids					
0.446	(1R,4aS)-7-(2-Hydroxypropan-2-yl)-1,4a-dimethyl-9-oxo-3,4,10,10a-tetrahydro-2H-phenanthrene-1-carboxylic acid 	$C_{20}H_{26}O_4$	Pos	330.18338	0.82	83.1
0.792	Abietic acid 	$C_{20}H_{30}O_2$	Neg	302.22431	0.91	75.9

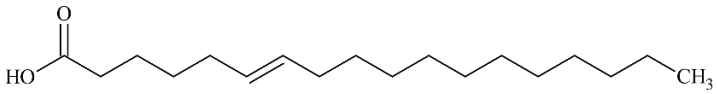
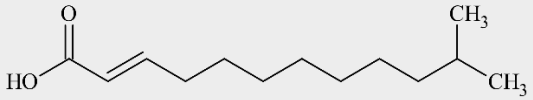
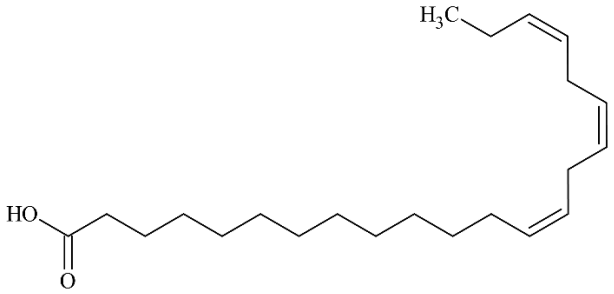
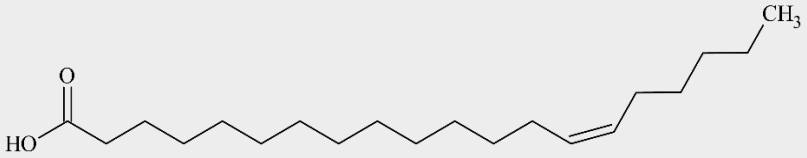
1.513	7-Oxodehydroabietic acid 	$C_{20}H_{26}O_3$	Pos	314.18863	1.39	98.2
	Diterpenoids					
0.353	(2E)-3-(Acetoxymethyl)-5-[(1S,4aR,8aR)-2-(hydroxymethyl)-5,5,8a-trimethyl-1,4,4a,5,6,7,8,8a-octahydro-1-naphthalenyl]-2-pentenoic acid 	$C_{22}H_{34}O_5$	Neg	378.24029	0.89	76.1
0.358	Cafestol	$C_{20}H_{28}O_3$	Pos	316.20425	1.27	86.5

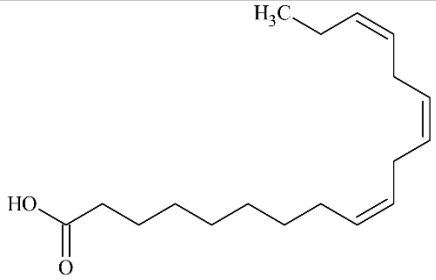
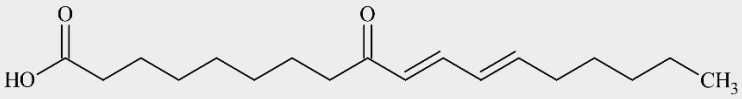
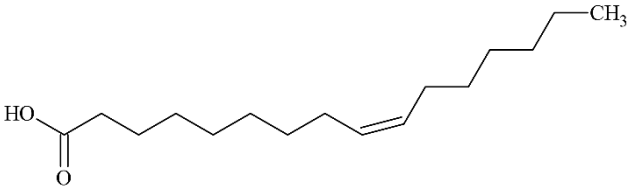


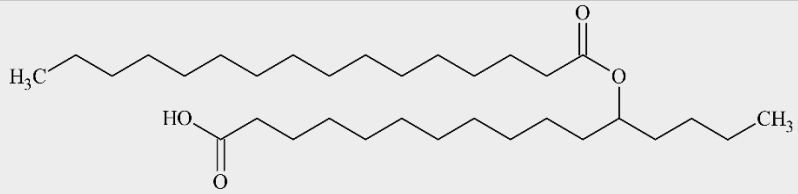
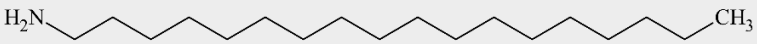
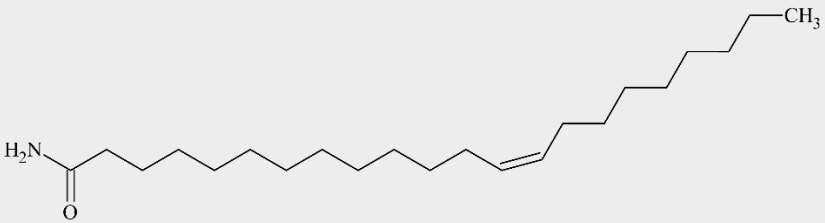
	5-[(3Z)-5-Hydroxy-3-methyl-3-penten-1-yl]-1,4a-dimethyl-6--methylenedecahydro-1-naphthalenecarboxylic acid					
0.387		$C_{20}H_{32}O_3$	Neg	320.23475	1.25	94.5
	(1S,4aR,5S)1,4a-Dimethyl-6-methylene-5-[2-(2-oxo-2,5-dihydro-3-furanyl)ethyl]-decahydro-1-naphthalenecarboxylic acid					
0.400		$C_{20}H_{28}O_4$	Pos	332.19903	0.82	84.8
0.438	Kahweol	$C_{20}H_{26}O_3$	Pos	314.18859	1.24	78.8

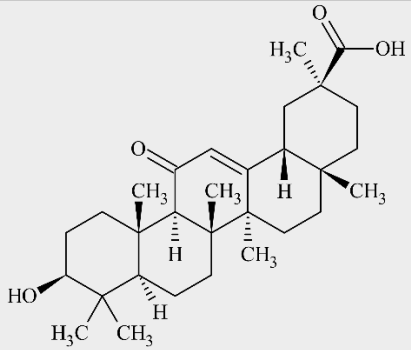
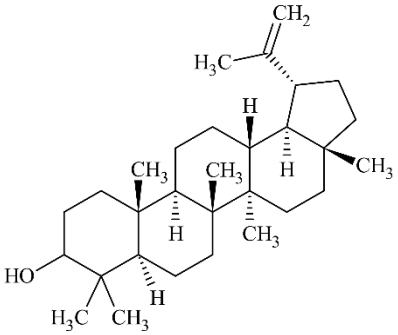
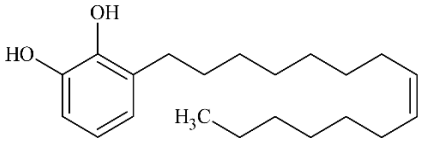
						
	5-[(1S,2R,4aR)-5-(Hydroxymethyl)-1,2,4a-trimethyl-1,2,3,4,4a,7,8,8a-octahydro-1-naphthalenyl]-3-methylpentanoic acid					
0.450		$C_{20}H_{34}O_3$	Pos	322.25107	0.84	77.9
	(1S,4aR,5S)-5-[(3E)-5-Methoxy-3-methyl-5-oxo-3-penten-1-yl]-1,4a-dimethyl-6-methylenedecahydro-1-naphthalenecarboxylic acid					
0.583		$C_{21}H_{32}O_4$	Neg	348.2298	0.75	76.1
	Fatty acids					
0.389	5-Keto-Eicosatetraenoic acid	$C_{20}H_{30}O_3$	Neg	318.21902	1.50	79.9

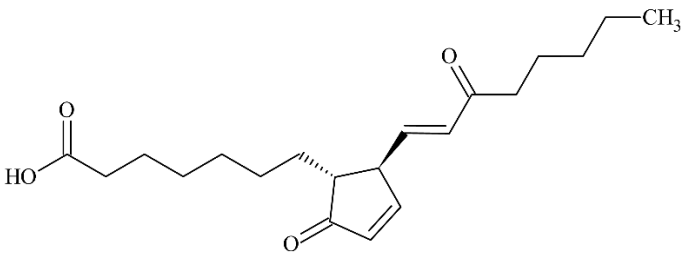
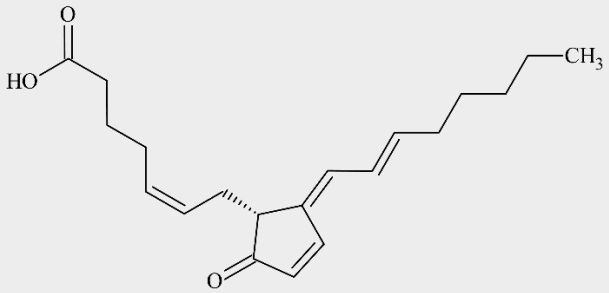
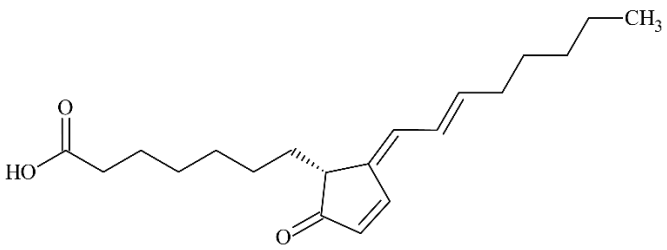
						
	11(Z),14(Z)-Eicosadienoic acid					
0.481		$C_{20}H_{36}O_2$	Neg	308.27137	0.53	77.8
	trans-δ^2-Hexadecenoic acid					
0.729		$C_{16}H_{30}O_2$	Neg	254.22428	1.19	75.3
	trans-10-Heptadecenoic acid					
0.882		$C_{17}H_{32}O_2$	Neg	268.23995	1.05	80.7
	Ethyl myristate					
0.998		$C_{16}H_{32}O_2$	Neg	256.24005	0.69	83.3

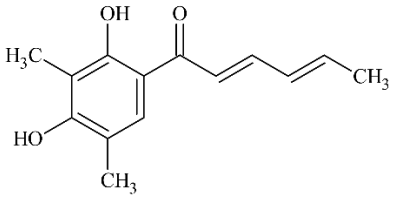
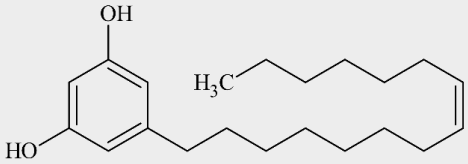
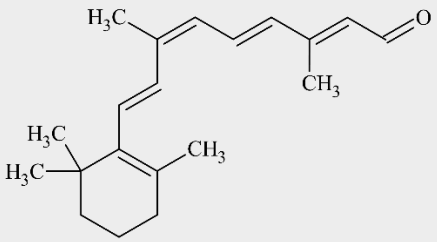
	trans-Petroselinic acid					
1.055		$C_{18}H_{34}O_2$	Neg	282.25566	0.78	78.6
	trans-δ^2-11-Methyldodecenoic acid					
1.177		$C_{13}H_{24}O_2$	Pos	212.17789	1.21	91.6
	Docosatrienoic acid					
1.280		$C_{22}H_{38}O_2$	Neg	334.28694	0.70	82.4
	14(Z)-Eicosenoic acid					
1.572		$C_{20}H_{38}O_2$	Neg	310.28722	0.13	75.0
1.879	α-Linolenic acid	$C_{18}H_{30}O_2$	Pos	278.22504	1.64	75.4

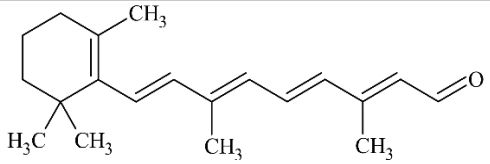
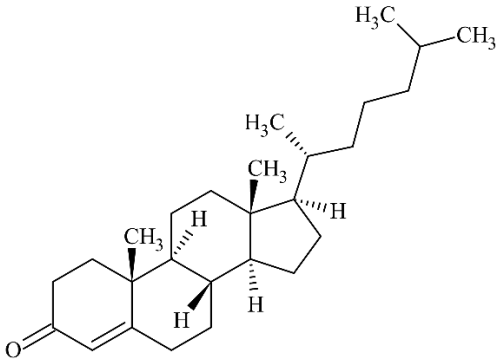
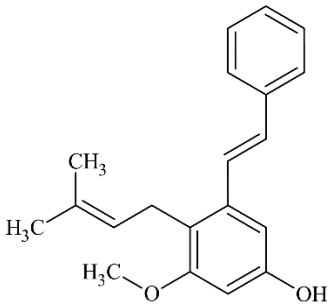
						
	9-Oxo-10(E),12(E)-Octadecadienoic acid					
6.601		$C_{18}H_{30}O_3$	Pos	294.22001	1.74	90.4
	Palmitoleic acid					
6.592		$C_{16}H_{30}O_2$	Pos	254.22497	1.53	88.3
	Fatty Acid Esters of Hydroxy Fatty Acids					
0.698	FAHFA 40:9	$C_{40}H_{60}O_4$	Neg	604.44906	0.17	86.2
3.919	FAHFA 32:0	$C_{32}H_{62}O_4$	Neg	510.46492	0.22	92.6

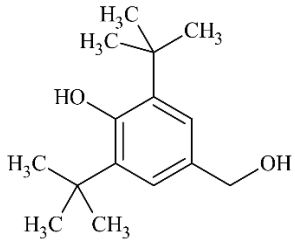
						
3.958	FAHFA 34:1	$C_{34}H_{64}O_4$	Neg	536.48085	0.73	81.2
5.183	FAHFA 36:2	$C_{36}H_{66}O_4$	Neg	562.49631	0.35	76.9
	Fatty amines					
0.757	Octadecanamine 	$C_{18}H_{39}N$	Pos	269.30822	0.10	75.9
	Fatty amides					
1.914	Erucamide 	$C_{22}H_{43}NO$	Pos	337.33473	0.80	92.5
	Pentacyclic triterpenoids					
0.833	18-β-Glycyrrhetinic acid	$C_{30}H_{46}O_4$	Pos	470.34041	1.69	76.5

						
	Lupeol					
2.058		$C_{30}H_{50}O$	Pos	426.38708	2.14	75.3
	Phenolic lipids					
	Urushenol					
1.399		$C_{21}H_{34}O_2$	Neg	318.25585	0.11	82.4
	Prostaglandins					

15-Keto-Prostaglandin A_1						
0.345		$C_{20}H_{30}O_4$	Neg	334.21421	0.58	84.7
15-Deoxy-$\delta^{12,14}$-Prostaglandin A_2						
0.394		$C_{20}H_{28}O_3$	Neg	316.20383	0.05	85.6
15-Deoxy-$\delta^{12,14}$-Prostaglandin A_1						
0.557		$C_{20}H_{30}O_3$	Pos	318.21984	1.10	78.0
Resorcinols						

0.341	Sorbicillin 	$C_{14}H_{16}O_3$	Pos	232.11027	1.38	78.0
0.941	Bilobol 	$C_{21}H_{34}O_2$	Neg	318.25532	1.76	78.9
	Retinoids					
0.518	9-*cis*-Retinal 	$C_{20}H_{28}O$	Pos	284.2143	1.00	77.4
0.924	*All-trans*-Retinal	$C_{20}H_{28}O$	Pos	284.21416	0.50	74.8

						
	Steroids					
	Cholest-4-en-3-one					
2.670		$C_{27}H_{44}O$	Pos	384.33964	1.10	89.1
	Stilbenes					
	Longistyline C					
0.442		$C_{20}H_{22}O_2$	Pos	294.16234	1.23	77.4

Other compounds							
0.548	3,5-Di-*tert*-Butyl-4-hydroxybenzyl alcohol 	$C_{15}H_{24}O_2$	Neg	236.17741	0.95	75.0

The EEGs fraction presented a large number of lipidic classes: fatty acids, phenolic lipids, fatty acyls, abietanoids and other diterpenoids, pentacyclic triterpenoids, prostaglandins, retinoids, resorcinols, steroids, fatty amines and amides, and other least polar compounds.

Diverse lipidic compounds were found in the organic lipid-rich fraction of the EEGs. Figure 26 shows the structures of some of these compounds:

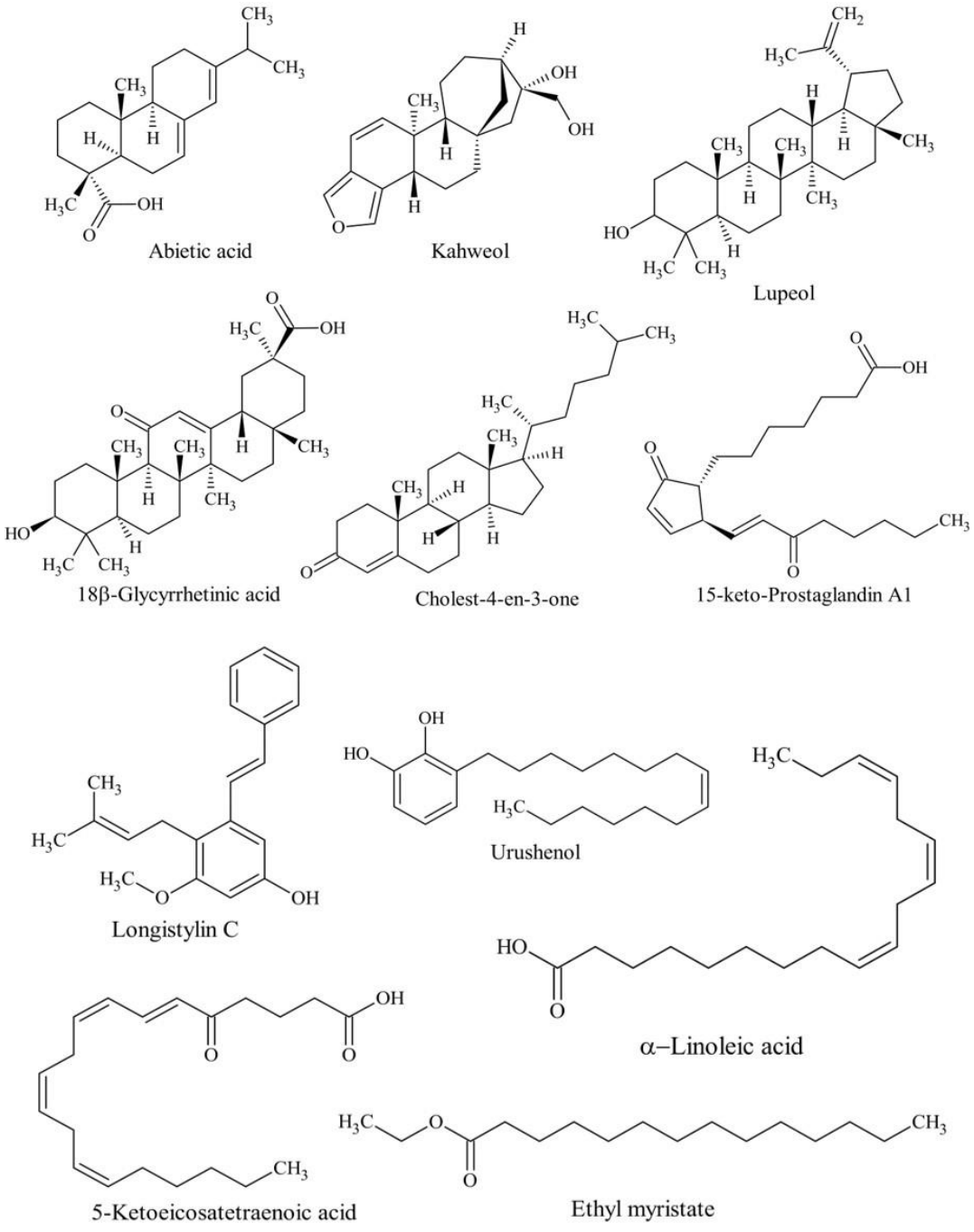


Figure 26 – Examples of lipids and lipid-like compounds found in the EEG.

Abietic acid is a tricyclic diterpenoid, associated to antiallergic, anti-inflammatory, and anticonvulsant activities [59, 60]. This compound is also classified as a resin diterpene [61], which the main source is, as well as others abietanoids, the oleoresins extracted from pine trees saps. Abietic and dehydroabietic acids are found in resins of diverse families of pine trees, such as *Araucariaceae*, *Cupressaceae*, *Phyllocladaceae*, *Pinaceae*, and *Podocarpaceae* [62].

Lupeol (Figure 26) in another example of pentacyclic triterpene, secondary metabolite of many plants, reported to possess beneficial pharmacological effects, especially its promising anti-inflammatory, and cancer potential [63]. In the last decades, several reports showed that triterpenes are directly able to inhibit tumoral cells growth, also decreasing cell cycle progression, and to induce the apoptosis of those cells, in both *in vitro* and *in vivo* situations [64]. In addition, it is noteworthy that Lupeol was reported to exhibit strong anti-mutagenic activity under in vitro and in vivo systems [65].

Longistylin C (Figure 26) is a plant originated prenylated phenolic compound, classified as a stilbene, that presented interesting antidepressant effects in *in vivo* studies with rats [66]. Liu et al. [66] confirmed longistylin C antidepressant effect through behavioural tests in mice. In addition, the authors also observed a neuroprotective activity against glutamate-induced injury in PC12 (pheochromocytoma) cells. The author concluded that those results imply that longistylin C will represent a reference for the development of new antidepressant drugs [66]. The 18-β-Glycyrrhetinic acid is a pentacyclic terpenoid that exhibits potent antitumor effects against the colorectal cancer. Wang et al. [67], demonstrated that 18-β-Glycyrrhetinic acid is capable to inhibit the proliferation and migration (metastasis) of the carcinogenic cells *in vivo* and *in vitro* tests.

α-Linoleic acid is a polyunsaturated fatty acid, essential to plants and animals, being required during the biological processes involved in the growing and developing.

This fatty acid is responsible for around 2% of the daily energy consumed by an adult person [68]. The presence of α-linoleic acid in geopropolis has been reported in diverse studies [69], being an alternative source of this compound. Another fatty acid found in the EEGs were the eicosapentaenoic acid, an omega-3 polyunsaturated fatty acid with anti-inflammatory property.

Retinoids are a class of compounds which include 9-*cis*-Retinal and *all-trans*-Retinal, and are precursors and/or derivatives of the A vitamins group. Retinoids have diverse pharmacological uses, for example being used in the treatment of vitamin A deficiency, photosensitivity, acne, and have been tested in for treatment of malignant neoplasms [70].

4.3.2 Multivariate analysis of the UHPLC–HRMS lipidomic dataset

A PCA analysis was performed using the whole lipidomic dataset, containing all the 23254 features. The PC1 and PC3 were chosen as the best components to show the analysis results, and the exploratory PCA is presented in Figure 27.

In the PCA analysis is possible to see diverse clusters of samples, some are closer to each other, while other presented a rather different pattern. The *M. marginata*'s EEGs split into two clusters: one cluster for samples from Prudentópolis, and another sparser cluster, that groups the EEGs made of geopropolis from other regions. The EEGs from *M. quadrifasciata* also presented the same feature, with samples from Prudentópolis grouping together while the other samples present their own different cluster. This geographic pattern was also observed during the fingerprint analysis (Chapter 2). According to Silva Cruz et al. [71], the geopropolis' characteristics are influenced by its geographical origins, once this represents a different source of vegetation available to be visited by the stingless bees. However, other aspects are also relevant, namely the bees'

species, geographic position of the hive, soil type, etc [5, 71]. Differently from the previous discussed groups, the EEGs of *T. angustula* geopropolis have not presented a clustering pattern based on simple geographical origin, once the samples from Prudentópolis and other regions are clustered together (Figure 27), giving further evidences that this species has "a favourite" vegetal sources to make its geopropolis.

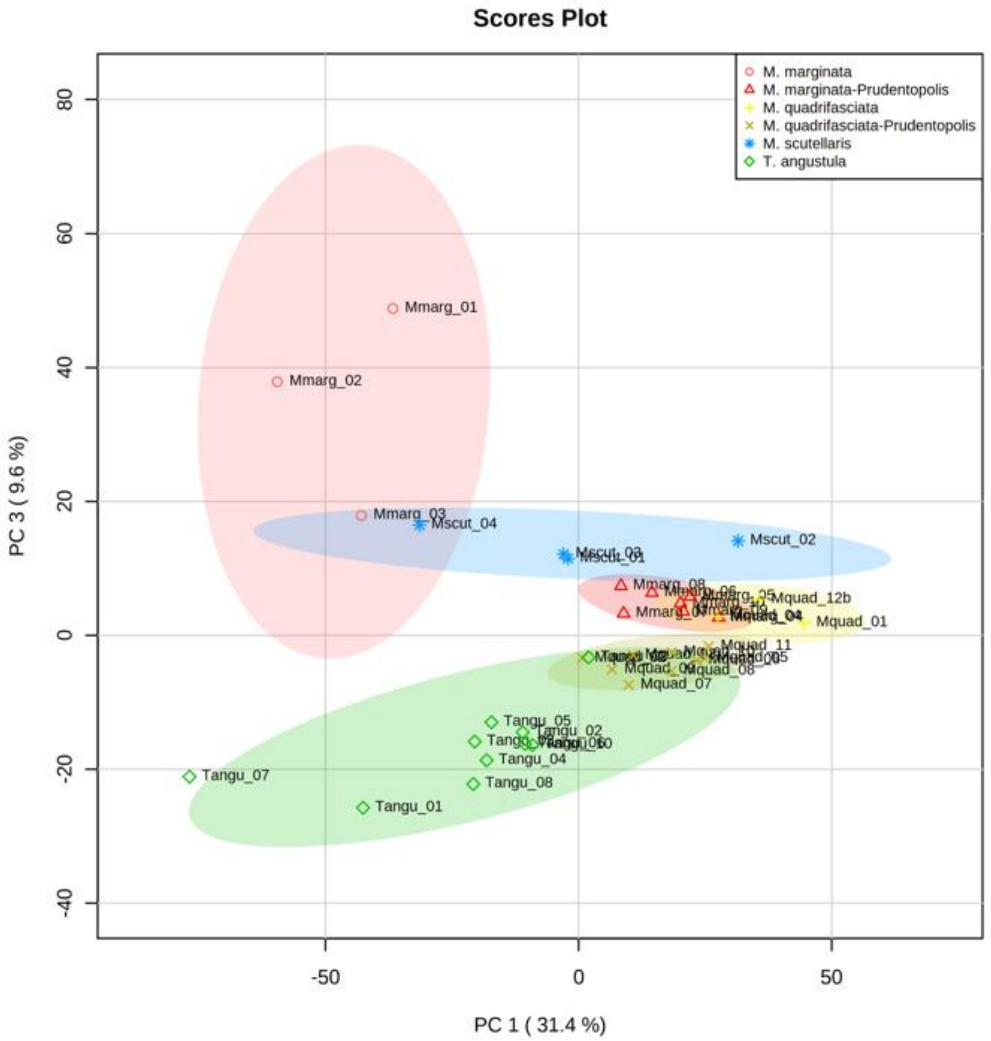


Figure 27 – Exploratory PCA of LC–HRMS lipidomic dataset of *M. marginata*, *M. quadrifasciata*, *M. scutellaris*, and *T. angustula*. Ellipses represent a confidence interval of 95%.

M. scutellaris group is composed only by four samples, thus it is difficult to observe patterns within this group due the small number of EEGs that represented this species during this study.

A second PCA, shown in Figure 28, was plotted to investigate the EEGs of the species that are represented by a smaller number of samples. The PC1 and PC3 were once again chosen to represent the PCA. The EEGs are clustered partially according to its genus; The first group is composed by EEGs which the geopropolis produced by species of the *Melipona* genus. The second group is a gathering of the rest of the EEGs, which are unitarian samples, produced by species of others genus. Both, *Melipona* and the non-*Melipona* groups have diverse geographic origin, as it is noticeable in Figure 2.

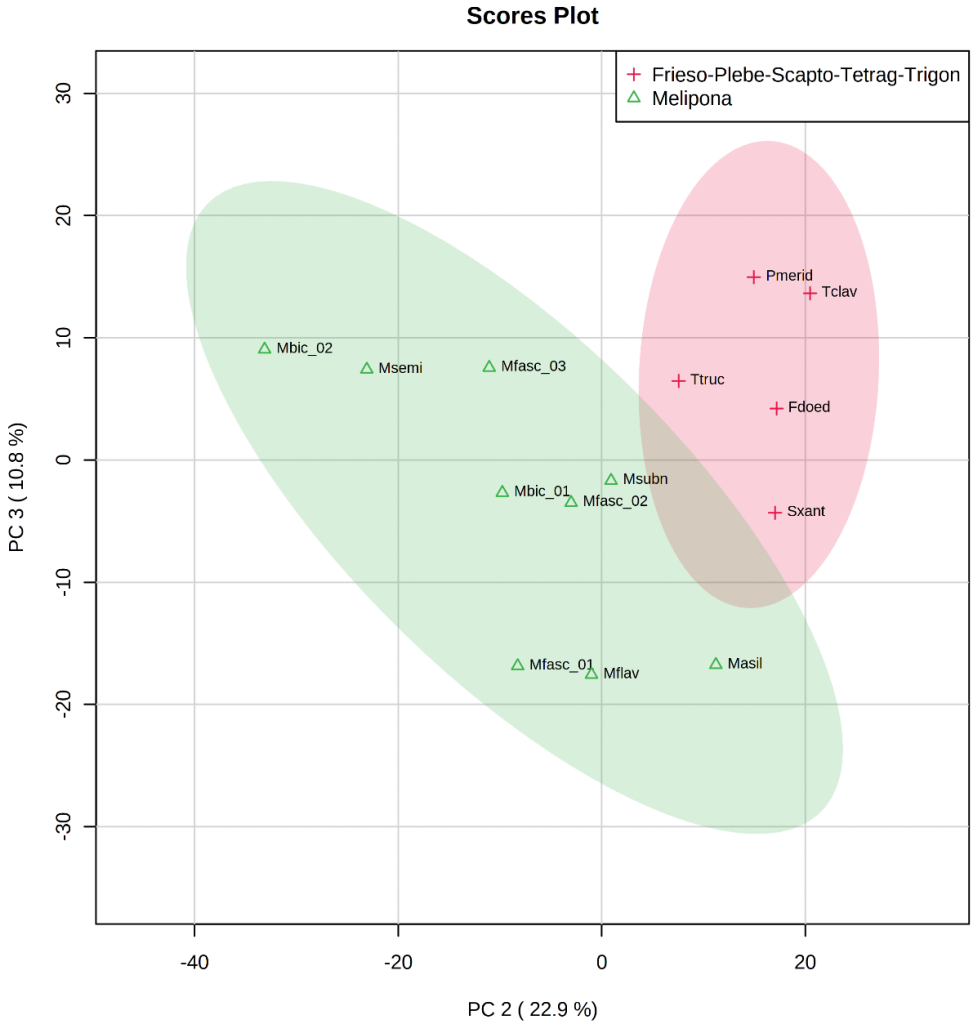


Figure 28 – Exploratory PCA of LC–HRMS lipidomic dataset of *Melipona* genus stingless bees (*M. asilvai*, *M. bicolor*, *M. fasciculata*, *M. flavolineata*, *M. seminigra*, *M. subnitida)*, and non-*Melipona* genus (*F. doederleini*, *P. meridionalis*, *S. xanthotricha*, *T. clavipes*, and *T. truculenta)*. Ellipses represent a confidence interval of 95%.

A t-test was applied to compare *Melipona* and the non-*Melipona* EEGs, and showed significative differences between the two groups. Using a confidence interval of 95%, the t-test revealed that at least 128 features are significant to distinguish the two groups.

4.3.3 Most characteristic compounds in the chloroform fraction of EEGs

One-way ANOVA was employed aiming to encounter more information about what features are responsible to cause statistically significant differences between the means of the relative concentrations of the compounds in the EEGs. Consequently, the most common compounds are the most abundant compounds found in the organic lipid-rich fraction of the EEGs and are summarised in Table 7.

Table 7 – Most characteristic compounds found in the lipophilic fraction of the EEGs. Continues on next page.

Specie	Origin	Formula	RT (min)	-Log$_{10}$(*p*)	*f*–value
M. marginata	Carira Curitiba Pilar do Sul	$C_{46}H_{89}O_9P$	3.712	12.987	27.383
	Prudentópolis	$C_{38}H_{63}N_5O_4$	2.436	8.029	12.995
M. quadrifasciata	Curitiba Paranaguá Pilar do Sul	$C_{55}H_{114}N_9P$	7.767	5.647	3.912
	Prudentópolis	$C_{39}H_{60}N_6O_4$	1.937	7.513	4.210
M. scutellaris	Curitiba Mogi Mirim Prudentópolis Pilar do Sul	$C_{51}H_{72}N_4$	0.899	2.660	2.678
T. angustula	Curitiba Quarto Centenário Prudentópolis	$C_{40}H_{66}N_6O_2$	3.523	13.791	30.539

Table 7 – Most characteristic compounds found in the organic lipid-rich fraction of the EEGs.

Specie	Origin	Formula	RT (min)	-$Log_{10}(p)$	*f*–value
*Melipona genus species**	Barra do Corda Pilar do Sul Prudentópolis	$C_{22}H_{24}O_6$	1.809	2.206	3.359
*Other genus species***	Petrolina Catanduvas Pilar do Sul Prudentópolis	$C_{47}H_{81}N_2OP$	6.191	9.835	17.369

*Species: *M. asilvai, M. bicolor, M. fasciculata, M. flavolineata, M. seminigra, M. subnitida.*

**Genus*: Friesomelitta, Plebeia, Scaptotrigona, Tetragona*, and *Tetragonisca*.

Compound Discover V. 3.3 identified only the empirical formula of the compounds presented in Table 7, and consequently they remained unnominated.

The analysis of variance (ANOVA) was performed in order to observe which compounds were more characteristic to each specie. The graphic of loadings was used to find out, based on the *p*-value, which ions were significatively more characteristic for each species. In other words, which compounds would make the EEGs unique. Figure 29 shows graphically the relative abundances of these compounds and the relative abundances of them.

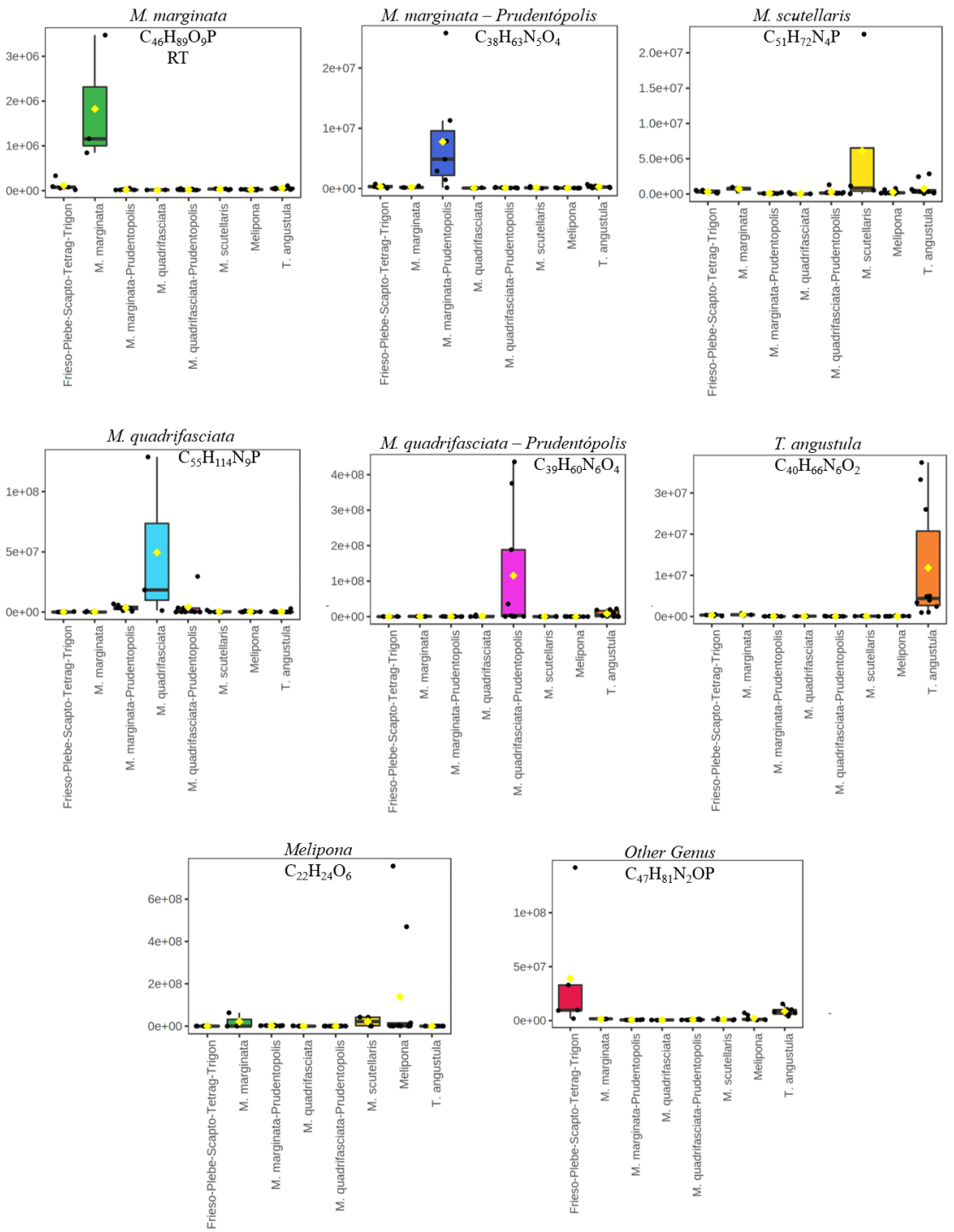


Figure 29 – Relative abundance presented by the most characteristic compound in the organic lipid-rich fraction of the EEGs.

Each EEGs presents a variety of compounds that might be considered characteristic to their kind, but only the most characteristic, was chosen as their marker compound.

4.4 Partial conclusions from the UHPLC–HRMS lipidomic analysis of the EEGs

Several classes of lipids were observed during the lipidomic analysis of the organic lipid-rich fraction of the EEGs. The presence of lipids in the geopropolis composition is not only restricted to waxes and fatty acids; as demonstrated in this study, geopropolis might present fatty amides and amines, phenolic lipids, resorcinols, retinoids, abietanoids, diterpenoids, among other compounds considered as lipids in its composition.

The untargeted lipidomic approach allowed the identification and classification of the extracts of geopropolis, revealing a new way to classify geopropolis, based on their least polar fraction. The lipidomic profile of EEGs is also affected by the bee species and the geographic origin.

Chapter 5 –GC–MS analysis of EEGs

5.1 Objectives

Them main goal of this part of the study is to assess the qualitative chemical profile of EEGs, using gas chromatography coupled to mass spectrometry (GC-MS) as a complementary analytical tool to LC-MS:

I. Classify the EEGs based on its GC-MS profile, using statistical tools (HCA, PCA, heatmaps);

II. Identify possible marker compounds, employing multivariate analysis to find the most characteristic compounds in each group of geopropolis.

5.2 Methodology

5.2.1 Sample preparation and liquid-liquid extraction for GC-MS analysis

From 20 µL were taken from each EEG and diluted using 180 µL of a solution of MeOH/IPA (50:50). A QC was created by pooling 20 µL of each EEG together. The samples were left overnight to dry in vacuum chamber.

The liquid-liquid extraction (biphasic partition) started by adding 300 µL of methanol, followed by 100 µL of Chloroform and 300 µL of ultrapure water. Each sample was shacked using vortex for approximately 15 s. To complete the process the samples were centrifugated during 15 min at 14600 rpm until clear phases were formed. Finally, two phases were visible: the aqueous phase (containing mainly polar compounds) on top and the organic phase (containing mainly nonpolar/hydrophobic compounds) below, as illustrated by Figure 30. From the aqueous phase 50 µL were deposited into assisted vials and subsequently left to dry completely using vacuum chamber. With the samples dried

out, the vials were sealed and organised in plates to be derivatised online and then injected. Details about the derivatisation step were presented in the next topic (5.2.2).

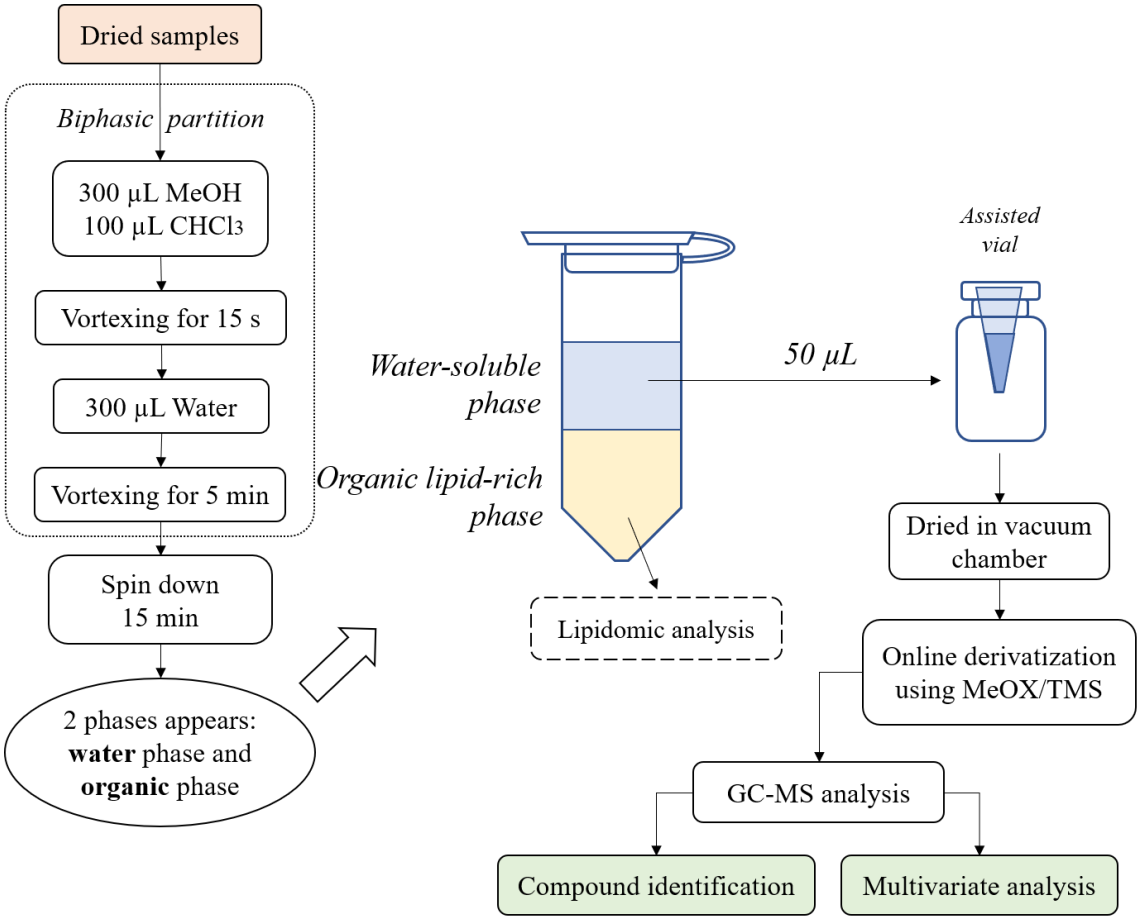


Figure 30 –EEGs pretreatment for GC-MS analysis.

5.2.2 Derivatisation of the water-soluble fraction of EEGs for GC-MS analysis

The derivatisation reagents, methoxyamine hydrochloride in pyridine (20.0 $mg.mL^{-1}$), and MSTFA, were placed into 2 mL vials in the reagents tray. The vials with the dried organic lipid-rich fraction of the EEGs were capped with magnetic caps and placed on the samples tray. The derivatisation process was carried out online, and injections automatised, using a Gerstel MultiPurpose Sampler (Germany).

To start the derivatisation process, a vial with the sample was moved to the agitator and 40 µL of methoxyamine hydrochloride in pyridine was added to dissolve the dried

sample. The sample was mixed for 90 min at 750 rpm and 37 ºC in the agitator. 40 µL of the second reagent, MSTFA (N-methyl-N-(trimethylsilyl trifluoroacetamide)) was then added to the sample and mixed for 30 min again at 750 rpm and 37 ºC. Two hours after finishing the derivatisation.

5.2.3 GC-MS Analysis

GC-MS analysis was carried out in an Agilent 8890 gas chromatograph coupled to a 5977B Mass Selective Detector (MSD) (Agilent, USA), using an Agilent 122-5532G DB5ms fused silica capillary column (40 m x 250 µm x 0.25 µm of film thickness). Helium was used as carrier gas and was adjusted to column velocity flow of 1.1 $mL.min^{-1}$. Other GC-MS conditions are: ion source temperature, 250 ºC; quadrupole temperature, 150 ºC, pressure, 13.071 psi; column temperature started at 60 ºC for 5.9 min then changed at the rate of 10 ºC/min to the maximum temperature of 325 ºC. Injection volume, 1.0 µL, spitless mode. The total time of elution was 37.5 min

Spectrometric data were acquired and processed using the Agilent MassHunter V. 10.0 software.

5.2.4 Data analysis

GC-MS data was processed using R Studio V. 1.1.456, and in house developed scripts.

Principal component analysis (PCA), heatmaps, and analysis of variance (ANOVA), were performed using the web interface MetaboAnalyst V. 5.0 (http://www.metaboanalyst.ca), created by the University of Alberta, AB, Canada [72].

5.2.5 Compound identification

The compound identification was carried out using Agilent MassHunter V. 10.0. The main way to identify a compound was by comparison of the retention time (RT), as well as the mass spectral information, of a query compound to the spectral libraries Fiehn and NIST (National Institute of Standards and Technology, USA). The mass spectra of unknown compounds were interpreted based on their fragmentation patterns.

Fiehn GC-MS library is based on a fatty-acid methyl ester retention index (RI) system, allowing the calculation of the Kovats' Index. Also, the Fiehn library spectra have been acquired using Agilent retention time locking (RTL) feature based on the RT of the myristic acid (trimethylsilylated), using an Agilent ZORBAX DB- 5MS column. The RTL software is able to generate all the retention times which enables universal measurements, as long as the same GC-MS method and column are used. Therefore, any user laboratory can theoretically reproduce the same results by locking the retention times to the mass spectral library. Those features provide a high credibility compound identification with low error incidence.

Significant differences among the 6 groups of geopropolis varieties for each of the volatile constituents were determined by one-way ANOVA using the free online platform MetaboAnalyst 5.0 [72].

5.3 Results and discussion

5.3.1 Composition of the EEGs assessed by GC-MS analysis

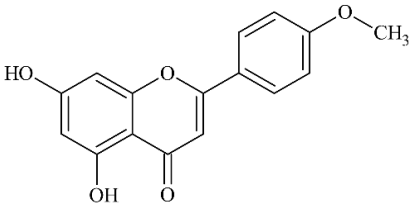
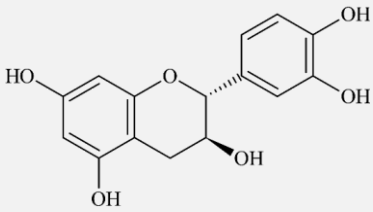
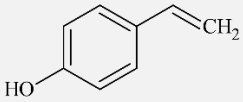
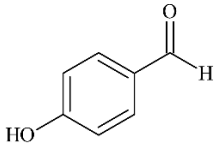
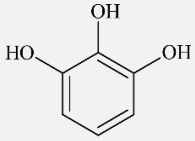
The analysis of the derivatised water-soluble fraction of EEGs via GC-MS allowed the identification of 90 compounds. The compounds were confirmed by the methodology described in the section 3.3.4, using the Agilent MassHunter V. 10.0 software to identify the compounds. Multiple compound classes such as (Table 8):

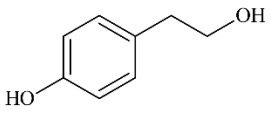
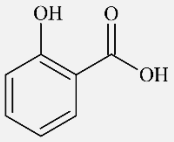
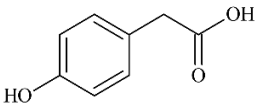
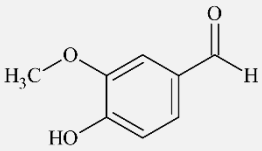
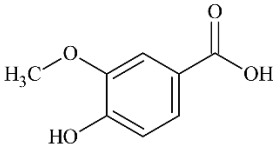
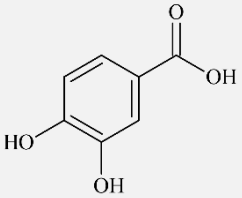
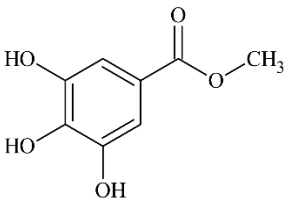
flavonoids, phenolic compounds, phenylpropanoids, phenolic aldehydes, sugars, tannins, quinones, furocoumarins, organic acids, lactones, polyols, and fatty acids were observed.

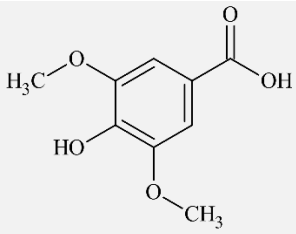
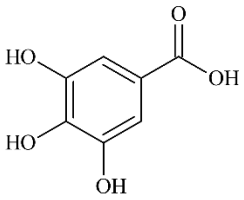
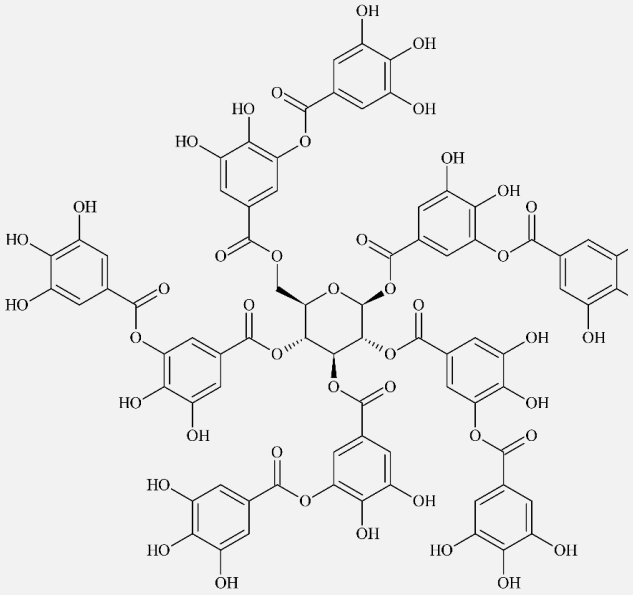
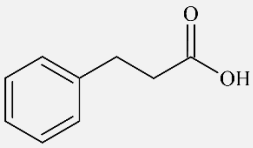
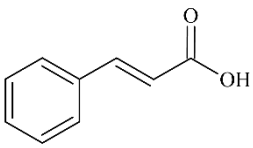
The stingless bees use geopropolis as a barrier to defend their colony, not only from physical, but also from biological threats [8]. Thus, the presence of compounds with strong antimicrobial and antifungal activities is expected in geopropolis, as well as other kinds of pharmacological and biological affects [45].

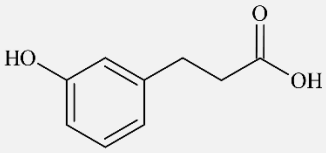
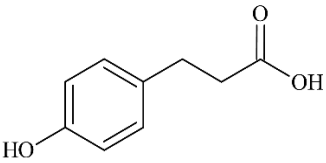
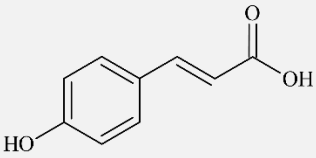
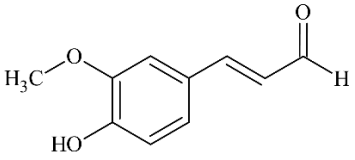
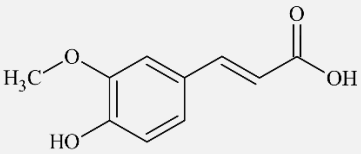
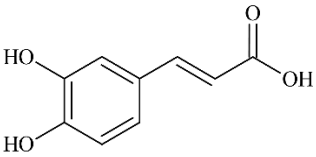
Diterpenoids, triterpenoids and phenolic compounds (mainly flavonoids) are commonly found in geopropolis from a wide variety of different stingless bees' species. These compounds are commonly found in different parts of the plants, influencing the stingless bees products composition [5]. The flavonoids catechin and acacetin were detected in the water-soluble fraction of EEGs and both present pharmacological effects of highly interest. Acacetin is an *O*-methylated flavone (Figure 31) present in low quantities in diverse plants. Diverse potentially therapeutic effect are addressed to acacetin, being the most common the prevention of infection, inflammation, and cancer, among other metabolic disorders [73].

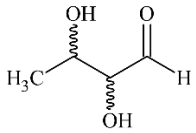
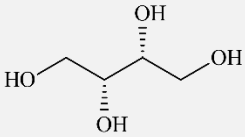
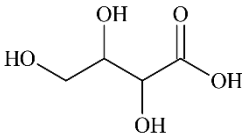
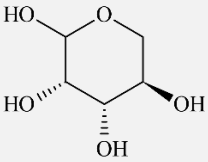
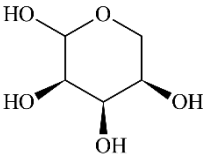
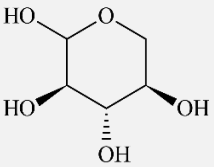
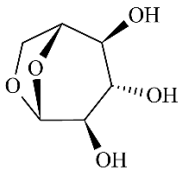
Table 8 – Compounds identified in the derivatised water-soluble fraction of the EEGs using GC-MS – Part I.

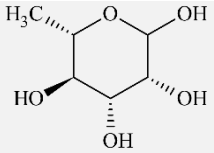
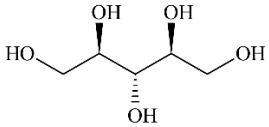
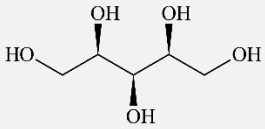
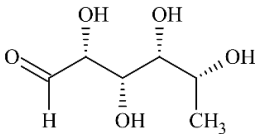
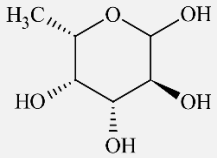
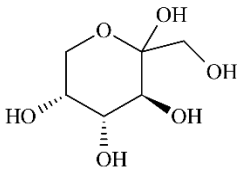
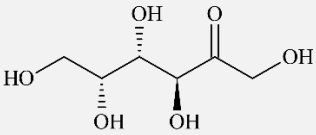
RT [min]	Identified compound name	Molecular formula	ΔRI	Ref.
	Flavonoids			
12.74	Acacetin (4-Methoxy-apigenin) 	$C_{16}H_{12}O_5$	9	
25.12	Catechin 	$C_{15}H_{14}O_6$	12	[45]
	Phenolic compounds			
9.65	*p*-Vinylphenol 	C_8H_8O	4	[13]
12.40	*p*-Hydroxybenzaldehyde 	$C_7H_6O_2$	4	
12.93	Pyrogallol 	$C_6H_6O_3$	15	
13.32	Tyrosol	$C_8H_{10}O_2$	10	

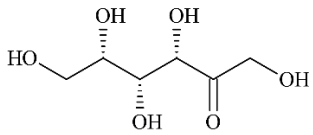
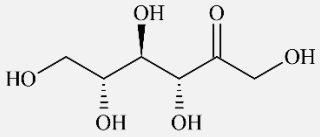
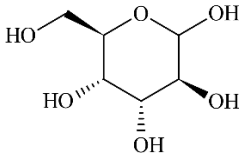
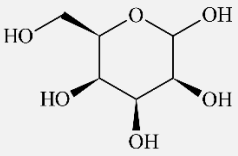
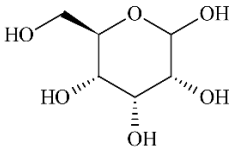
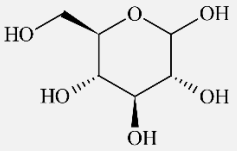
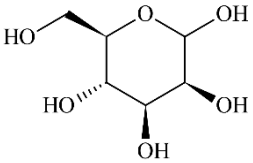
				
	Salicylic acid			
13.95		$C_7H_6O_3$	5	[45]
	p-Hydroxyphenylacetic acid			
14.11		$C_8H_8O_3$	7	[74]
	Vanillin			
14.20		$C_8H_8O_3$	6	[48]
	Vanillic acid			
15.50		$C_8H_8O_4$	16	[48]
	Protocatechuic acid (3,4-dihydroxybenzoic acid)			
16.09		$C_7H_6O_4$	8	[49]
	Methyl gallate			
16.79		$C_8H_8O_5$	6	
16.90	Syringic acid	$C_9H_{10}O_5$	6	[45]

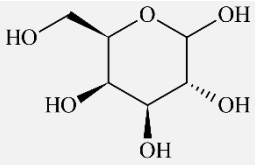
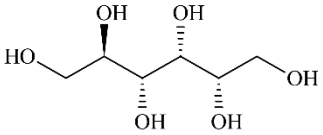
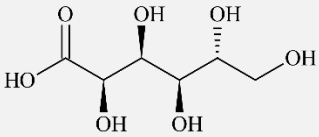
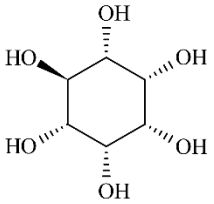
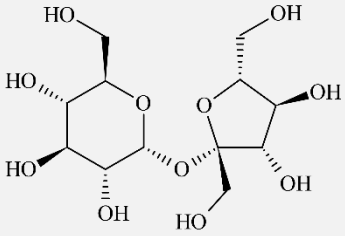
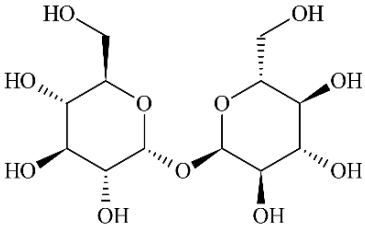
				
	Gallic acid			
17.48		$C_7H_6O_5$	13	[49]
	Tannic acid			
17.50		$C_{76}H_{52}O_{46}$	12	
	Phenylpropanoids			
	Hydrocinnamic acid			
11.44		$C_9H_{10}O_2$	11	[75]
	Cinnamic acid			
13.07		$C_9H_8O_2$	17	[48]

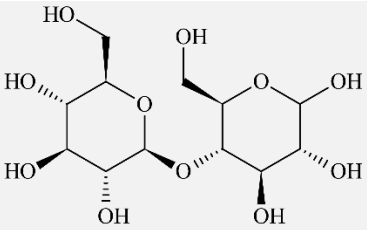
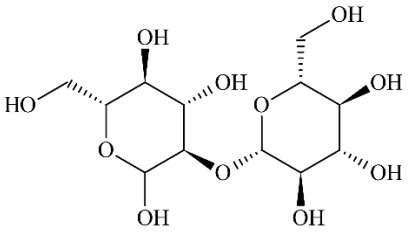
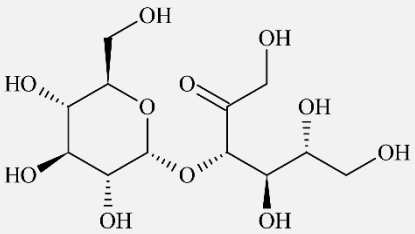
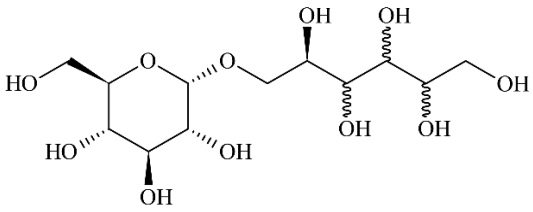
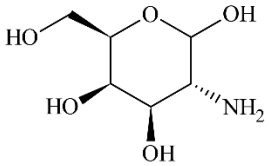
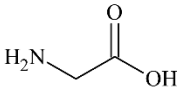
	Dihydro-3-coumaric acid			
15.17		$C_9H_{10}O_3$	6	
	Phloretic acid			
15.48		$C_9H_{10}O_3$	15	
	p-Coumaric acid (4-Hydroxycinnamic acid)			
17.36		$C_9H_8O_3$	17	[48]
	Coniferaldehyde (4-Hydroxy-3-methoxycinnamaldehyde)			
17.42		$C_{10}H_{10}O_4$	13	
	Ferulic acid			
18.82		$C_{10}H_{10}O_4$	4	[48]
	Caffeic acid			
19.23		$C_9H_8O_4$	13	[48]
	Sugars			
11.83	Threose	$C_4H_8O_5$	2	

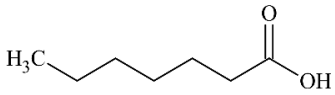
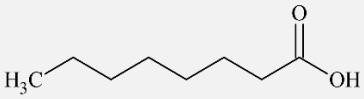
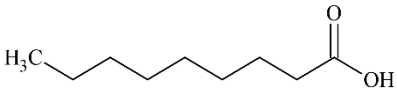
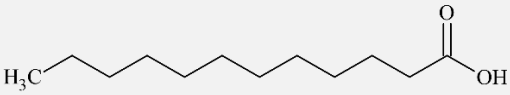
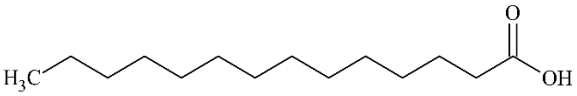
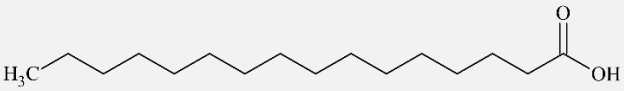
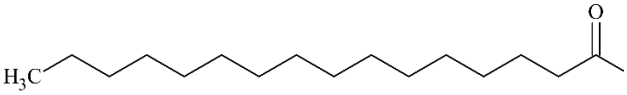
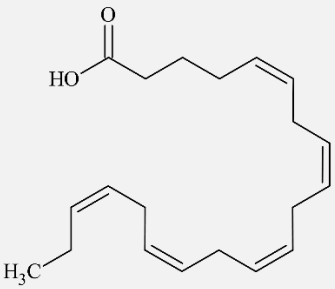
				
	D-Threitol			
12.64		$C_4H_{10}O_4$	16	
	Threonic acid			
12.96		$C_4H_8O_5$	1	
	D-Lyxose			
14.26		$C_5H_{10}O_5$	4	
	Ribose			
14.31		$C_5H_{10}O_5$	23	[49]
	Xylose			
14.35		$C_5H_{10}O_5$	20	[49]
	1,6-Anhydroglucose			
14.81		$C_6H_{10}O_5$	0	
14.96	Rhamnose	$C_6H_{12}O_5$	10	

				
	Ribitol			
14.98		$C_5H_{12}O_5$	3	[53]
	Xylitol			
15.00		$C_5H_{12}O_5$	13	[49]
	6-Deoxy-D-glucose			
15.05		$C_6H_{12}O_5$	2	
	L-(-)-Fucose			
15.08		$C_6H_{12}O_6$	1	[49]
	Fructose			
16.62		$C_6H_{12}O_6$	3	[49]
	Tagatose			
16.71		$C_6H_{12}O_6$	3	
16.71	L-Sorbose	$C_6H_{12}O_6$	0	[49]

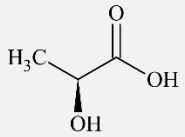
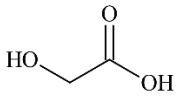
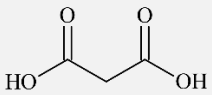
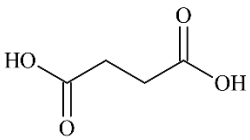
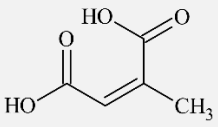
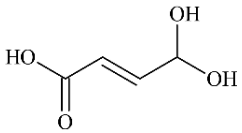
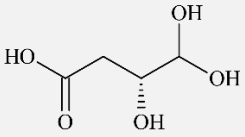
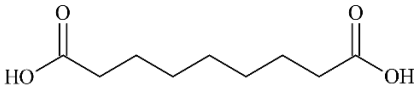
				
	Psicose			
16.72		$C_6H_{12}O_6$	6	
	D-(+)-Altrose			
16.81		$C_6H_{12}O_6$	6	
	Talose			
16.85		$C_6H_{12}O_6$	3	
	D-Allose			
16.88		$C_6H_{12}O_7$	13	
	D-Glucose			
16.88		$C_6H_{12}O_6$	3	[53]
	D-Mannose			
16.88		$C_6H_{12}O_6$	12	[49]
17.07	D-(+)-Galactose	$C_6H_{12}O_6$	9	[49]

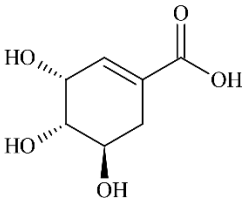
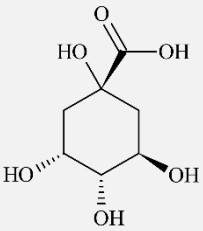
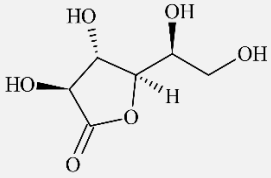
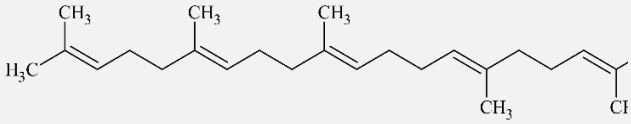
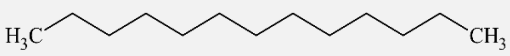
				
	D-sorbitol			
17.23		$C_6H_{14}O_6$	2	[49]
	Gluconic acid			
17.92		$C_6H_{12}O_7$	26	[49]
	myo-Inositol			
18.79		$C_6H_{12}O_6$	3	[49]
	Sucrose			
23.36		$C_{12}H_{22}O_{11}$	10	[76]
	D-(+)-Trehalose			
24.11		$C_{12}H_{22}O_{11}$	6	
24.16	Cellobiose	$C_{12}H_{22}O_{11}$	30	[76]

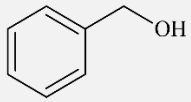
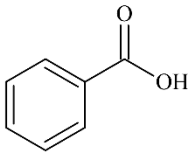
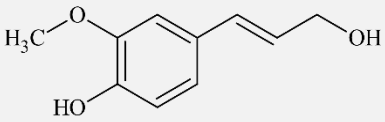
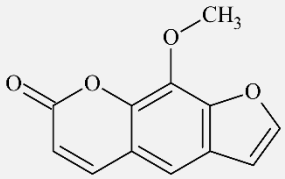
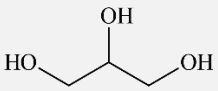
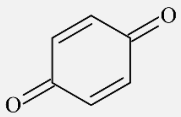
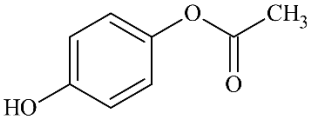
				
	Sophorose			
24.36		$C_{12}H_{22}O_{11}$	2	
	Turanose			
24.41		$C_{12}H_{22}O_{11}$	6	
	Palatinitol			
25.38		$C_{12}H_{24}O_{11}$	6	[77]
	Amino sugars			
	Galactosamine			
16.81		$C_6H_{13}NO_5$	5	
	Amino acids			
	Glycine			
9.96		$C_2H_5NO_2$	4	

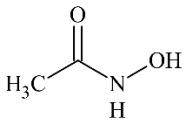
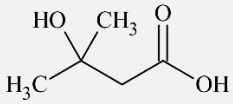
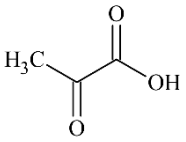
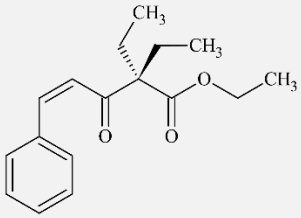
	Fatty acids			
	Enanthic (heptanoic acid)			
8.14		$C_7H_{14}O_2$	2	
	Caprylic (octanoic acid)			
9.42		$C_8H_{16}O_2$	5	
	Pelargonic (nonanoic acid)			
10.68		$C_9H_{18}O_2$	18	
	Lauric (dodecanoic acid)			
14.27		$C_{12}H_{24}O_2$	5	[13]
	Myristic (tetradecanoic acid)			
16.43		$C_{14}H_{28}O_2$	9	[13]
	Palmitic (hexadecanoic acid)			
18.40		$C_{16}H_{32}O_2$	14	[49]
	Stearic (octadecanoic acid)			
20.21		$C_{18}H_{36}O_2$	12	[49]
	Timnodonic (eicosapentaenoic acid)			
23.37		$C_{20}H_{30}O_2$	13	

Organic acids

	L-(+)-Lactic acid			
6.87		$C_3H_6O_3$	17	
	Glycolic acid			
6.95		$C_2H_4O_3$	10	[49]
	Malonic acid			
8.68		$C_3H_4O_4$	9	
	Succinic acid			
10.17		$C_4H_6O_4$	8	
	Citraconic acid			
10.50		$C_5H_6O_4$	13	
	Fumaric acid			
10.52		$C_4H_4O_4$	21	[13]
	D-Malic acid			
12.27		$C_4H_6O_5$	0	
	Azelaic acid			
15.85		$C_9H_{16}O_4$	10	

	Cyclohexanecarboxylic acids			
	Shikimic acid			
15.97		$C_7H_{10}O_5$	13	
	Quinic acid			
16.48		$C_7H_{12}O_6$	5	[49]
	Lactones			
	L-Galactono-γ-lactone			
17.15		$C_6H_{10}O_6$	8	
	Triterpene			
	Squalene			
24.79		$C_{30}H_{50}$	13	[78]
	Alkanes			
	Tridecane			
13.71		$C_{13}H_{28}$	19	
	Aromatic hydrocarbons			
7.94	Benzyl alcohol	C_7H_8O	0	[13]

				
	Benzoic acid			
9.23		$C_7H_6O_2$	8	[48]
	Coniferyl alcohol			
17.33		$C_{10}H_{12}O_3$	15	[13]
	Furocoumarins			
	Xanthotoxin			
20.22		$C_{12}H_8O_4$	8	
	Polyols			
	Glycerol			
9.68		$C_3H_8O_3$	15	[77]
	Quinones			
	1,4-Benzoquinone			
11.21		$C_6H_6O_2$	10	
	4-Acetoxyphenol (hydroquinone monoacetate)			
11.21		$C_8H_8O_3$	21	[13]
	Other organic compounds			
7.39	Acetohydroxamic acid	$C_2H_5NO_2$	5	

			
	β-Hydroxyisovalerate		
8.67		$C_5H_{10}O_3$	3
	Glyoxylic acid		
10.30		$C_2H_2O_3$	15
	4-Pentenoic acid, 2,2-diethyl-3-oxo-5-phenyl-, ethyl ester		
20.65		$C_{17}H_{22}O_3$	19

Acacetin is also known as 4-methoxy-apigenin, being a compound commonly associated to high quality *A. mellifera* (common bee) propolis, and its presence in geopropolis could be evidence of a common source of resins shared by the two kinds of bees. The other flavonoid found in the derivatised water-soluble fraction of EEGs was catechin. This flavonoid is most commonly found in wild fruits and shows similar pharmacological effects to those of acacetin; also presenting skin protective effects against damage caused by ultraviolet radiation [79]. Both compounds, catechin and acacetin also have considerable antioxidant capacities [73, 79]. Flavonoids originate from plants, being divided in several subgroups which includes flavones, isoflavones, flavanols, and chalcones. Flavonoids are versatile defenders of the plants, protecting the plants from biotic stresses, and acting as an UV filter [80].

A considerable variety of phenolic compounds were identified in the derivatised water-soluble fraction of EEGs, listed in Table 8 [75]. Phenolic compounds are an important class of secondary metabolites commonly found in plants, and in its derivatives, such as coffee, tea, and wine. Geopropolis, due to its vegetal origins, shows a composition rich in phenolic compounds [81]. Among the phenolic compounds identified in the EEGs during this study, there are some compounds with noticeable antioxidant capacities, such as the acids: protocatechuic, gallic, vanillic, *p*-salicylic, and syringic (Figure 31). Each of these compounds has a characteristic and pleasant aroma [82]. According to Campos et al. [75], the antioxidant activity of geopropolis is related to the presence of phenolic compounds in its composition. During antioxidant assays by DPPH scavenging technique, the phenolic compounds were able to stabilise the radicals via a mechanism of electronic donation also known as antirradicalar effect/capacity.

Vanillin could be considered a phenolic compound, more precisely a phenolic aldehyde. Vanillin as well as vanillic acid displays large number of biological and

pharmacological properties of interest, such as antioxidant, anticarcinogenic, antimutagenic and antimicrobial activities [83]. Due to its characteristics, it has been commonly used by the food and beverage industry as a flavouring, and a food preservative [84]. According to Srinivasan, Platel and Rao [85], individuals who daily consume small portions of vanillin may experiment loss of weight, through the decreasing of fat accumulation in the adipose tissue, followed by a reduction in the low-density lipoprotein (LDL) fraction of the cholesterol.

Gallic acid is a polyphenolic compound (Figure 31) with antiallergic, anti-inflammatory and anticarcinogenic effects. Being a biologically active compound, gallic acid acts as an strong antioxidant compound, preventing the destructive effects of free radicals in human cells [86]. Because of its low toxicity, gallic acid is often used by the pharmaceutical and the food industry, especially for the production of juices, being therefore safe and there is no established limit of use, in addition, the absorption of this compound in the human organism occurs more effectively if compared to other phenolic compounds, and yet its slightly sweetness pleases the palate [87].

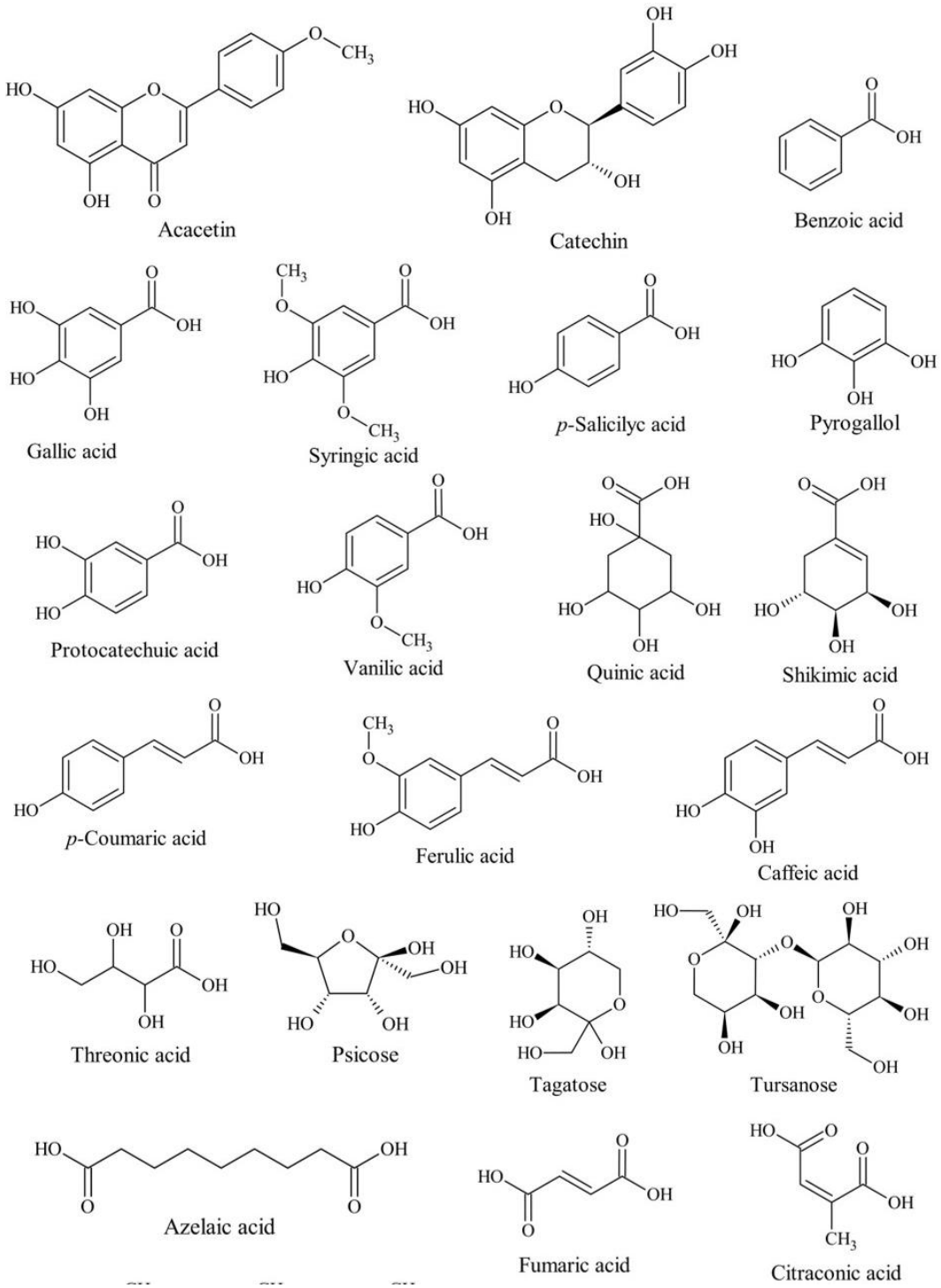


Figure 31 – Structures of some compounds detected in the derivatised water-soluble fraction of EEGs using GC-MSEEGs.

Phenylpropanoids are a class of phenolic compounds that presents a wide range of reported biological effects, such as antimicrobial, anticancer, anti-inflammatory, and others [88]. The *p*-coumaric acid is a phenylpropanoid that plays an important role in the

secondary metabolism of plants, being a precursor of other phenolic compounds, such as ferulic and caffeic acids (Figure 31), as well as flavonoids and polyphenols [89]. In addition, the class of the phenylpropanoids also presents compounds with relevant pharmacological activities, exhibiting neuroprotective, cardioprotective, renoprotective, and hepatoprotective effects [88]. Sharma et al. [90] observed, after treatment with *p*-coumaric acid, significant inhibition in the proliferation of human and mouse melanoma cells. Geopropolis presents skin protective effects [91] that could be linked to the presence of phenylpropanoids, such as ferulic acid, in its composition. According to Suzuki et al. [92], oral supplementation with ferulic acid in healthy man decreases sympathetic nervous activity; also, strengthens the skin barrier function, which helps to prevent the penetration of unwanted compounds from the exterior into the body through the epidermis.

In total 32 different sugars were observed in the EEGs. This large number of sugars was possible to be detected thanks the derivatisation step [93], described in the item 3.2.2. Some of the detected sugars are commonly observed in many bee's products, especially in honey. The most abundant sugars being sucrose, fructose, glucose, mannose, galactose, and xylose. The honey produced by *Meliponini* bees (stingless bees) is a condensed collection of many different sugars, in general, with the predominance of glucose and fructose [94]. However, less common kinds of sugars were also identified among the EEGs composition, such as sophorose, tagatose, D-threitol, turanose, psicose, and threonic acid, shown in the Figure 31. Some of those compounds were yet not reported in the literature for the composition of the Brazilian stingless bees geopropolis (Table 8). Tagatose (Figure 31) is considered a rare sugar (monosaccharides with limited availability in nature). This compound is considered as a low-calorie sweetener, also

presenting antibacterial activities, being capable of inhibiting the growth of crop pathogens such as *Phytophthora infestans* [95].

Traditionally, organic acids are related to food production and technology. However, recently claim the attention of the chemical industry, as a source of building blocks for polymers, and as cosubstrates for pharmaceuticals products. Lactic and glycolic acids are keratolytic compounds, that mean they are capable to break down the outer skin layers, decreasing the thickness of psoriatic plaques. Salicylic acid is also a keratolytic compound, alongside with other alpha-hydroxy acids such as citric, malic, glycolic, and tartaric acids. These compounds are also able to enhance the penetration of medications, during treatment by topical application on skin [96].

5.3.2 Multivariate analysis of the GC-MS data

5.3.2.1 Exploratory PCA

Four groups of EEGs are partially overlapped in PCA based on GC-MS profile (Figure 32). The PC 1 and PC 2 were chosen to represent the analysis. By using the 95% confidence interval ellipse as a reference, at least three groups are visually distinguishable from each other.

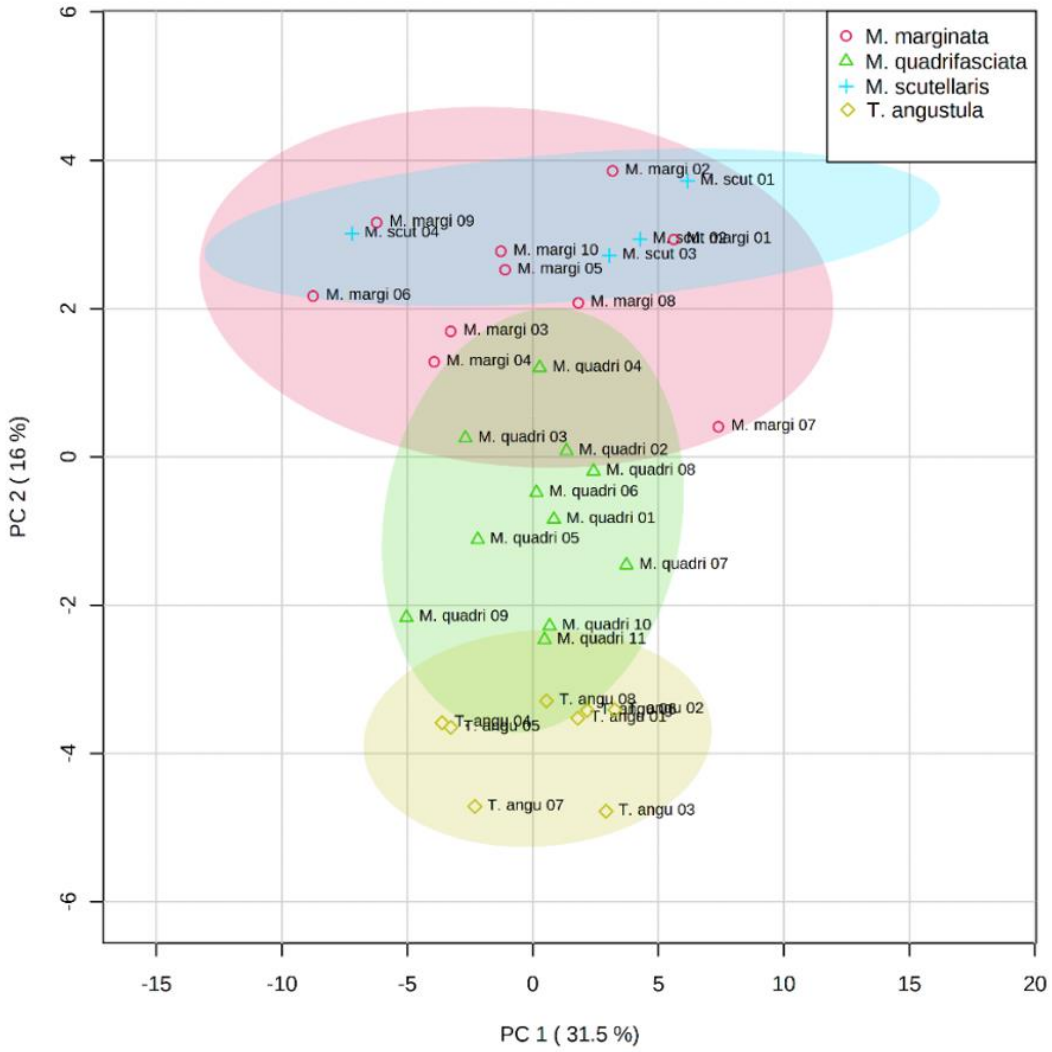


Figure 32 – PCA of the GC-MS profiles of EEGs from *M. marginata*, *M. quadrifasciata*, *M. scutellaris*, and *T. angustula.*

The geopropolis produced by *Tetragonisca angustula* presented a rather distinctive composition from that produced by *Melipona marginata* and *Melipona scutellaris*, only showing a slightly similarity with *Melipona quadrifasciata*. Both *M. scutellaris* and *M. marginata* have not presented a significant distinction, probably those species share a resemblance at least among their polar compounds. The species of the stingless bees plays a significant role in the geopropolis composition, being a relevant feature to characterise geopropolis.

The EEGs from *M. quadrifasciata* and *M. marginata* and *T. angustula*, form distinctive groups with only few samples sharing similar features in their composition with other groups (Figure 32). The QC group was located in the centre of the PCA, and each QC injection overlapped each other, indicating excellent reproducibility of the injections.

The rest of the EEGs, which do not present more than three samples per species, were divided into two classes: EEGs which the geopropolis were produced by bees belonging to the *Melipona* genus and those from geopropolis produced by another genera (*Friesomelitta, Plebeia, Scaptotrigona, Trigona*, and *Tetragonisca*). After a PCA analysis (Figure 33) an already expected result appears: the EEGs presented a distinction, based on a feature related with the bees' species.

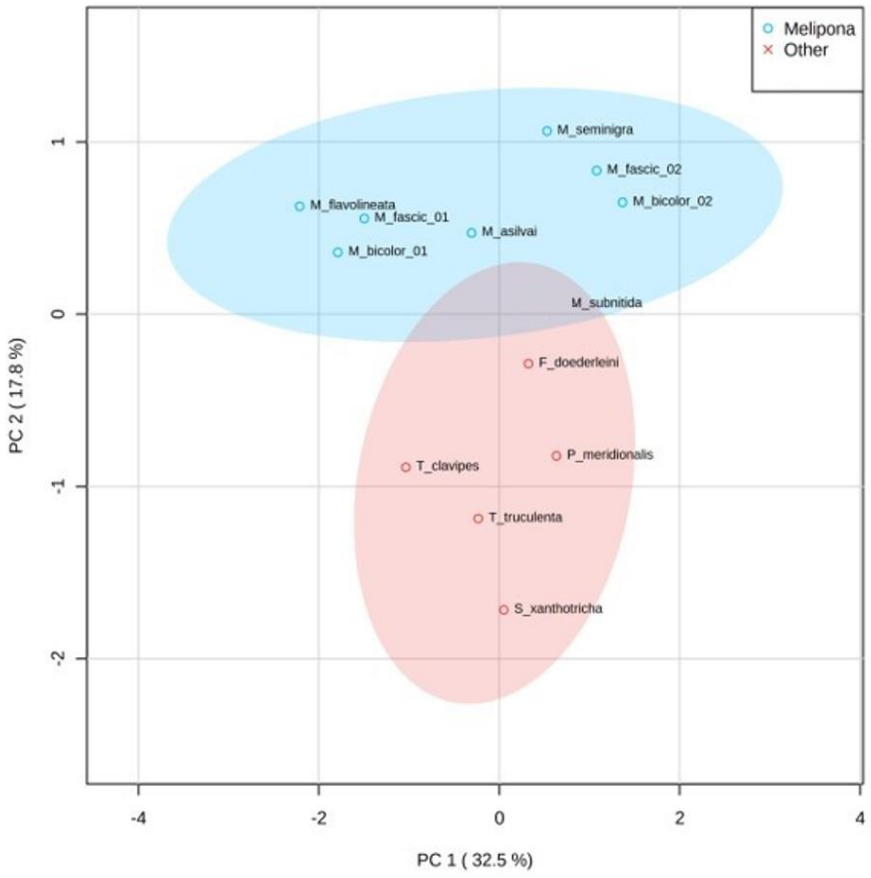


Figure 33 – PCA of the GC-MS profiles of EEGs from *Melipona* genus group and the other non-*Melipona* group of species.

However, there are not two groups, but the *Melipona* group and another gathering of species. This gathering of species/genera works like a contrast for the *Melipona* group. In fact, these non-*Melipona* species are not necessarily sharing similar compounds compositions, but they enhance the evidence of chemical differences between geopropolis produced by stingless bees from *Melipona* genus and others types. This is another evidence that the bee type heavily influences geopropolis' composition.

A t-test revealed that there are significant differences ($p < 0.05$) between EEGs from *Melipona* genus and the other genera (*Friesomelitta*, *Plebeia*, *Scaptotrigona*, *Tetragona*, and *Trigona*) (Figure 34).

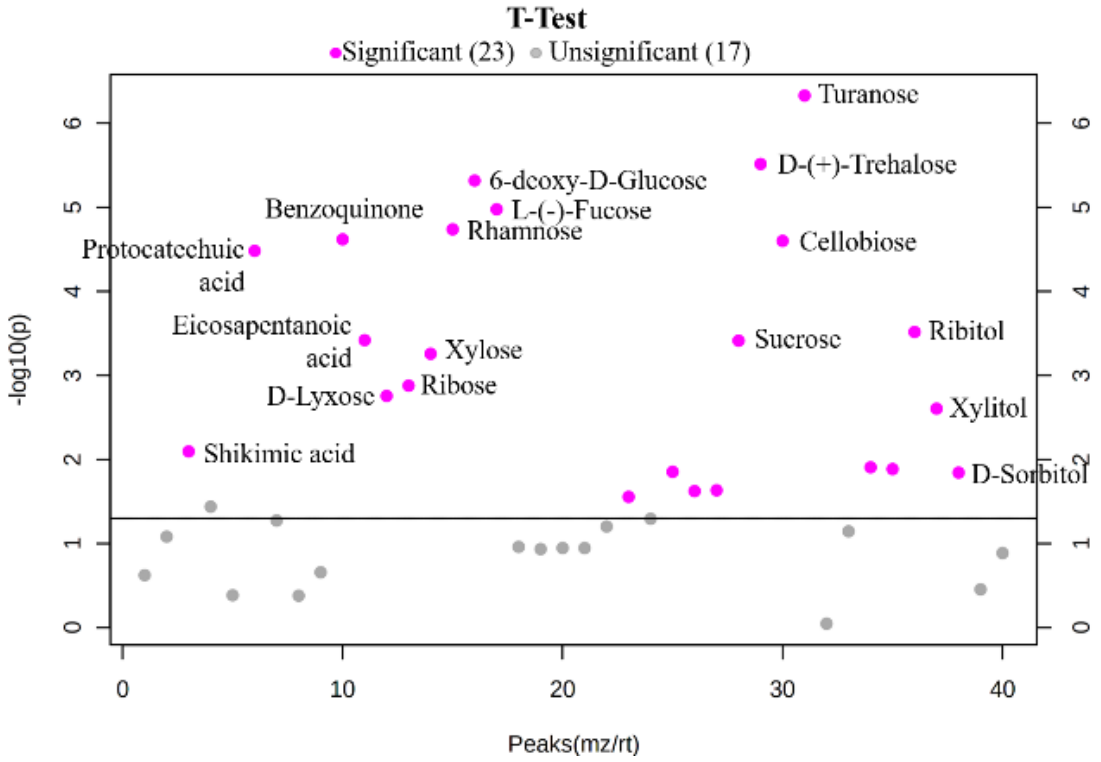


Figure 34 – t-Test, graphic of loadings. Comparison between *Melipona* genus and the other genus (*Friesomelitta*, *Plebeia*, *Scaptotrigona*, *Tetragona*, and *Tetragonisca*).

In total 23 compounds were found significant to distinguish the *Melipona* genus group samples from the *other genus* group.

5.3.2.2 Most characteristic compounds in the derivatised water-soluble fraction of EEGs

The One-Way ANOVA (Analysis of variance) was applied to determine which compounds are relatively the most abundant and statistically relevant to characterise the EEGs from each group of geopropolis. One-Way ANOVA was used aiming to compare the means of the groups, in order to determine whether there is statistical evidence of a significantly difference associated to the population's means. In all, 14 compounds were found relevant: threonic acid, rhamnose, 6-deoxy-D-glucose, L-(-)-fucose, D-(+)-trehalose, turanose, myo-inositol, cellobiose, sucrose, vanillin, vanillic acid, benzoquinone, 4-hydroxybenzaldehyde, and eicosapentaenoic acid. The most characteristic, and based on the *p*-value, the most characteristic compounds are shown Table 9.

Table 9 – Most characteristic compounds found in the derivatised water-soluble fraction of the EEGs.

Specie/Genus	Compound	*f*–value	*p*–value	$-\log_{10}(p)$
M. quadrifasciata	benzoquinone	20.957	8.73E-9	8.0589
M. marginata	vanillic acid	6.3802	6.08E-4	3.2157
M. scutellaris	D-(+)-trehalose	23.269	7.15E-8	7.1452
T. angustula	threonic acid	19.766	1.73E-8	7.7611
Melipona	6-deoxy-D-glucose	10.3770	1.33E-5	4.8761

5.3.2.4 Heatmap of the most significant compounds

A heatmap, presented by the Figure 35, shows the relatively most abundant compounds, based on the *p*-value, of each group of EEGs.

The EEGs from *M. scutellaris* have a composition particularly rich in sugars, presenting a variety of compounds, mainly sugars, such as: xylose, sucrose, ribose, D-allose, turanose, and threhalose. Also, the heatmap shows a strong significance of the glucuronic acid, the rare reductive sugar palatinitol, and fatty acid (lipid) eicosapentaenoic acid. The EEGs from the *Melipona* (*genus*) group presented similar compounds as the most significant, however the sugars D-mannose, D-sorbitol, D-glucose, and D-allose were not found significatively present in this group of samples. In addition, the sugar myo-inositol and *p*-hydroxybenzaldehyde were found as significant, differing the EEGs of *Melipona* and *M. scutellaris* species.

EEGs from *M. marginata* geopropolis presented as strongly significant vanillin, vanillic acid, shikimic acid, and 4-Hydroxy-3-methoxycinnamaldehyde. According to the heatmap, these compounds are statistically the most significant compounds to characterise the GC-MS profile of *M. marginata* EEGs. To complement, the sugars D-allose, D-mannose, D-sorbitol, and D-glucose are also significatively present in this group of EEGs.

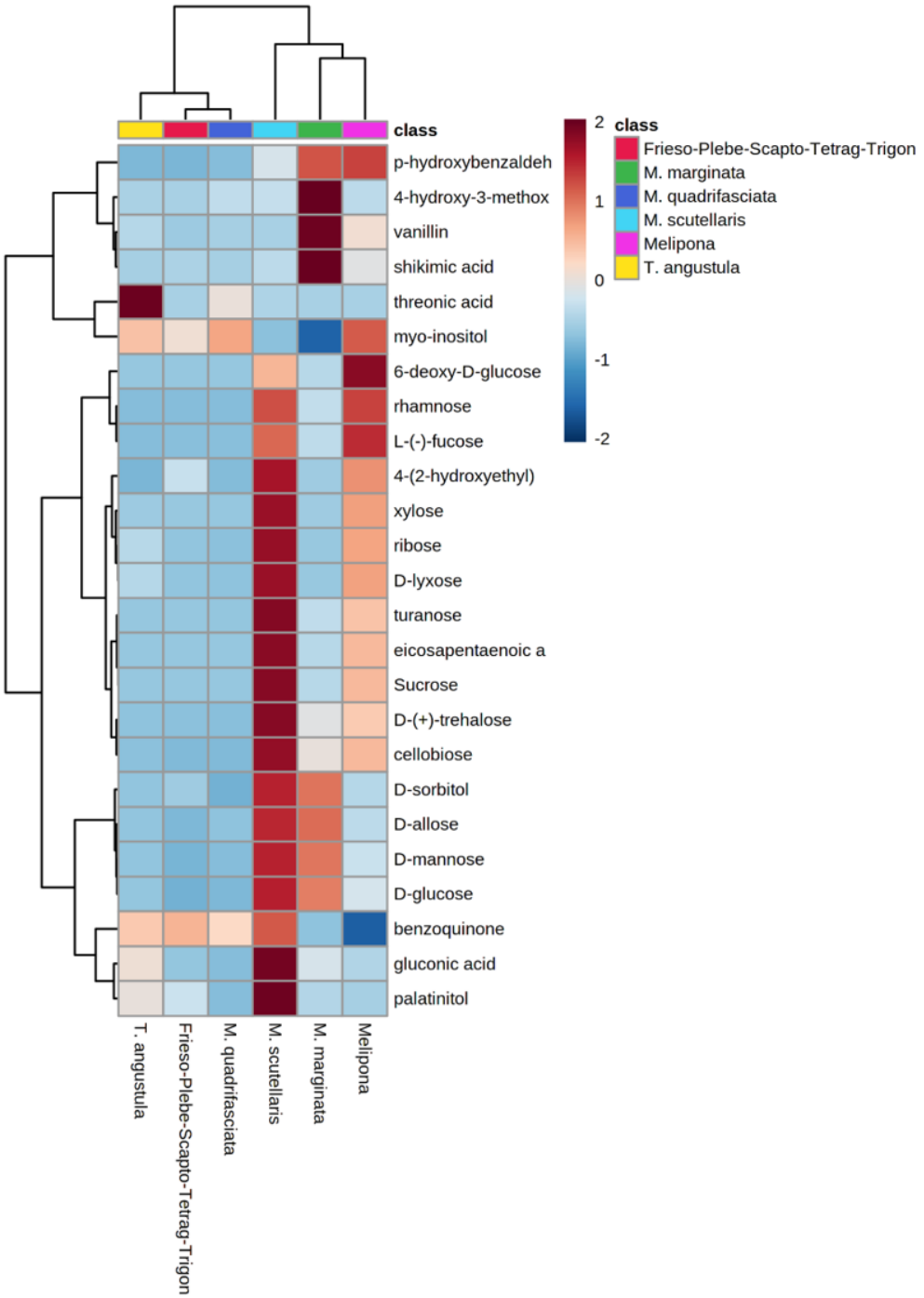


Figure 35 – Heatmap of GC-MS data from the EEGs. Red is for the most significative features; blue is for the less significant.

According to the Figure 35, the water-soluble fraction of the *T. angustula* EEGs only presented threonic acid as the most characteristic compound with high correlation. This sugar acid is a derivate of trehalose, and according to a report of Kwack et al [97], it may have potential in treatment of androgenic alopecia.

5.4 Partial conclusions from GC-MS analysis of the derivatised water-soluble fraction of the EEGs

Sugars were among the compounds most found in this water-soluble fraction of the EEGs although are a class of compounds rarely reported in propolis. Particularly, D-(+)-trehalose, threonic acid and 6-deoxy-D-glucose were relevant to distinguish the geopropolis from *M. scutellaris*, *T. angustula* and the Melipona, respectively. The sugars present in geopropolis play an important role in human feeding, especially for poor communities that rely on geopropolis as a supplementary source of energy.

Multivariate analysis of the GC-MC profile revealed the same tendencies already observed, pointing that factors such as species and genus are among the causes for the geopropolis to be chemically distinct. Also, the location plays an important role in the chemical composition of geopropolis.

Chapter 6 – NMR analysis of the chloroform extract of geopropolis

6.1 Introduction

In the last two decades, nuclear magnetic resonance (NMR) has emerged as one of the main analytical techniques for metabolomics, occupying a position of relevance alongside established techniques such as GC-MS and LC-MS [98]. Although chromatographic techniques, most commonly GC-MS and LC-MS, are still the most chosen for metabolomics, NMR presents several key advantages as it is a quantitative, unbiased, and non-destructive technique, that does not require previous separation or derivatisation steps, and provides objective information for compound identification [99].

NMR spectroscopy takes advantage of the interaction between nuclei, that are acting as subatomic-sized magnets, that when exposed to an external magnetic field, resonates emitting radiofrequencies. The signal generated from this interaction provides powerful means of probing the chemical bonding and environment of the nucleus. These phenomena are key to the applicability of ^{1}H and ^{13}C NMR to natural products.

When compared to chromatographic techniques, NMR is less sensitive, however, this technique provides concrete information about the chemical structures of the compounds presents in the analyte, especially when information is given by the combination of ^{13}C and ^{1}H NMR [99].

6.1.1 Usage of NMR in metabolomic analysis

NMR spectroscopy has a long history in successfully characterise, quantify, and determine the structure of small molecules from biological systems [100]. In addition, NMR also offers several advantages over other techniques on the metabolomic platforms [101], being a non-destructive, easily quantifiable, and unbiased technique, that requires

little sample preparation, with no need of derivatisation. NMR is also easily automated, which enhances reproducibility, agility, and efficiency, making more feasible, and reliable, high-throughput automated metabolomics studies. Nonetheless, NMR can help to identify compounds that are particularly challenging for LC-MS-based techniques, such as fatty acids, organic acid, polyols, sugars, and highly polar substances [100].

Metabolomic analysis usually generates large amounts of data, which are commonly handled using multivariate analysis. Generally, in NMR spectroscopy the spectral data pass through a binning process, being exported in tabular form, and then multivariate tools can be applied. However, information given by the minor signals are easily lost during the binning process [102].

Razali et al. [103] classified raw stingless bees honey using a metabolomic approach, associated to chemometric tools employing ^{1}H NMR, and UHPLC-HRMS techniques. According to the authors, the samples of raw honey were classified into three different groups, based on the bee species: d-Fructofuranose was found a as “marker” compound of *Heterotrigona itama*; β-d-Glucose, d-Xylose, α-d-Glucose were marker compounds of *Geniotrigona thoracica*; and l-Lactic acid, Acetic acid, l-Alanine were characteristic compounds in honey *Tetrigona apicalis*. Razali et al. [103] suggest that the quality, purity, and originality of the honey of those stingless bees can be quickly determined using ^{1}H NMR, and UHPLC-HRMS-based metabolomic approach, associated to statistical tools.

NMR-based metabolomic studies about geopropolis of stingless bees are rare in the literature; therefore, this is a promising and worthful filed of research, once this material is rich in several compounds of interest.

6.2 Objectives

The main goal of this part of the study was to explore the metabolomic profile of different kinds of geopropolis through their chloroformic extracts (CEGs), employing ^{1}H and ^{13}C NMR to:

I. Identify and confirm classes of compounds that were previously observed by chromatographic analysis made in Chapters 3, 4, and 5;

II. Identify compounds using HMBC, HSQC, and ^{1}H and ^{13} NMR spectral data;

III. Classify the CEGs based on their spectral information, employing multivariate analysis;

IV. Built a table of similarities based on spectral information in order to explore the similarities between CEGs.

6.3 Methodology

6.3.1 Sampling

NMR spectroscopy was performed using the same source material of i*n natura* geopropolis previously used to prepare the EEGs; remaining i*n natura* geopropolis were sent to the Metabolomics and NMR Laboratory, at the Universidade Estadual Norte Fluminense (UENF), Campos dos Goytacazes–RJ.

6.3.2 Obtention of the chloroformic extracts of geopropolis (CEGs)

The samples of *in natural* geopropolis analysed during this study were listed in Table 1. Chloroformic extracts of geopropolis (CEGs) were obtained using a simple and

concise protocol commonly in the Metabolomics and NMR Laboratory: 100 mg of *in natura* geopropolis were completely dissolved in 500 μL of $CDCl_3$ with TMS, after the CEGs were filtered using qualitative paper, as illustrated by Figure 36.

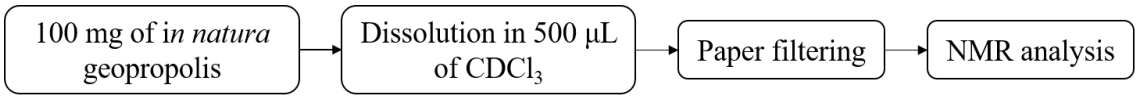


Figure 36 – Obtention of the chloroformic extracts of geopropolis (CEGs).

No internal standards other than TMS were added to the CEGs. After the measurements the chemical shifts were referenced to the residual solvent signals of $CDCl_3$ at 7.26 ppm for ^{1}H and 77.00 ppm for ^{13}C, and TMS signal was set as being 0.00 ppm.

6.3.3 NMR spectra data acquisition

NMR spectra were obtained following the methodology of Schripsema et al. [102]: with two different Bruker instruments, both Bruker Avance III HD systems, operating at 500 MHz for ^{1}H NMR, and 125.76 MHz for ^{13}C NMR. ^{1}H NMR spectra were acquired at 298 K using the standard Bruker pulseprogram zg30, with an acquisition time of 3.2 s and a relaxation delay of 1 s acquiring 64 K data points. ^{13}C NMR spectra were acquired at 298 K using the standard Bruker pulseprogram zgpg30, with an acquisition time of 1.1 s and a relaxation delay of 0.5 s acquiring 32 K data points. 8 scans were acquired for ^{1}H, and 256 scans for ^{13}C.

6.3.4 NMR spectra data processing

NMR spectra were processed using Spinworks V. 4.2.11 software, (Kirk Marat, University of Manitoba, Canada). The FIDs were multiplied with a Lorentz (Exponential)

window function with LB 0.300 Hz. The phase was adjusted manually and the fully automatic baseline correction was applied. The spectra were calibrated to the residual solvent signal. The data points were exported to Excel (Microsoft Office Excel 2016), and further processed.

6.3.5 Calculation of similarity and differential spectroscopy

The exported NMR spectra were processed in using MS Excel. ^{1}H NMR data were reduced to the spectral region from 11.00 to 0.00 ppm.

The spectra were aligned in relation to the average spectrum by calculation of the similarity. Only alignment of complete spectra was performed. After alignment the solvent signals were removed: In the 1 H NMR spectra the region from 7.30 to 7.22 ppm was set to zero. In the 13C NMR spectra the region from 77.4 to 76.6 ppm was set to zero. In the 13C NMR spectra the noise level was determined and all points with intensity below the noise level were set to zero (in this case points with an intensity below 0.0001). Subsequently the sum of all points was recalculated and adjusted to obtain a total sum of exactly one. In the final spreadsheet each 1 H NMR spectrum contained 35,073 points, 3278 points per ppm and each 13C NMR spectrum contained 57,324 points, 277 points per ppm. For the calculations of the similarity floating bins were applied with a width of 109 datapoints, corresponding to 0.033.

6.4 Results and discussion

6.4.1 Exploratory PCA and table of similarities based on ^{1}H-NMR profile of CEGs

Figure 37 shows the exploratory PCA of the ^{1}H NMR spectra of CEGs of *M. marginata*, *M. quadrifasciata*, and *T. angustula*. In Figure 37, the outline of three groups

appears based on the species of the stingless bees. In addition, the same PCA suggests that geopropolis of *M. marginata* is rather distinct from geopropolis of *M. quadrifasciata* and *T. angustula*, which is also highlighted by the table of similarities (Table 10).

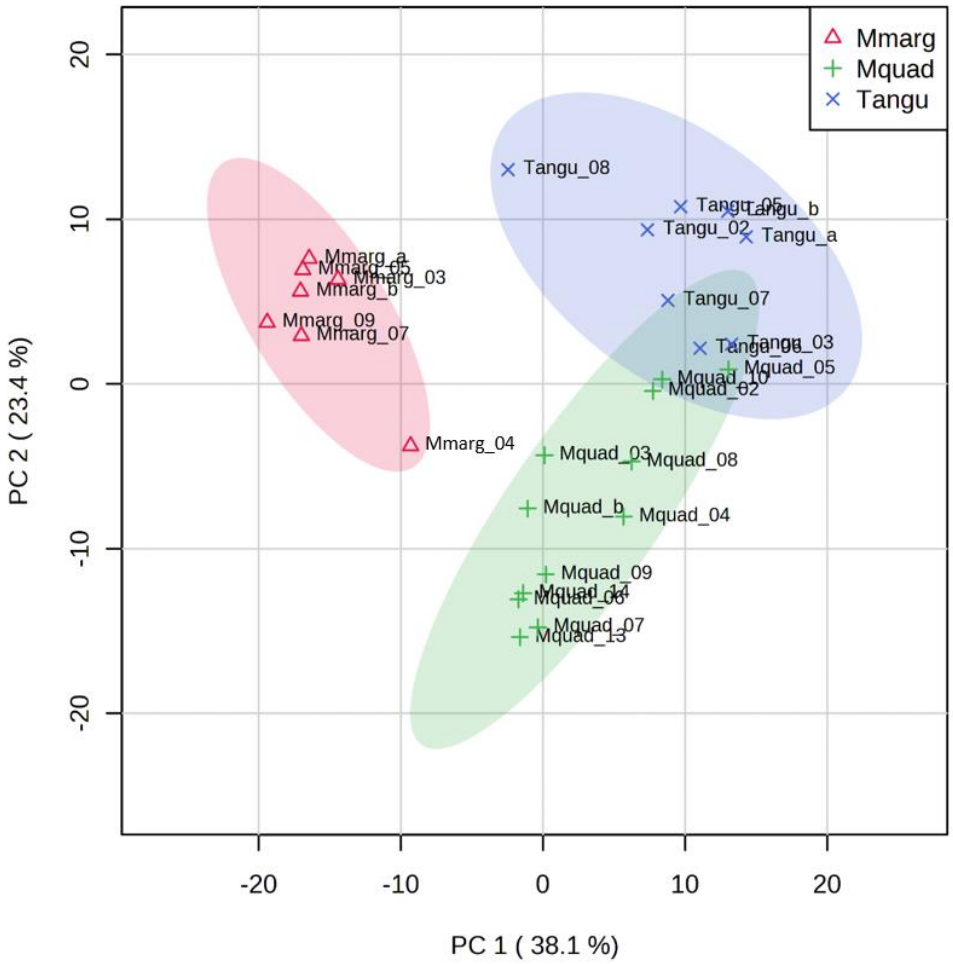


Figure 37 – PCA of the ^{1}H NMR spectra of geopropolis of *M. marginata* (Mmarg), *M. quadrifasciata* (Mquad), and *T. angustula* (Tangu). Ellipses represent a 95% confidence interval.

However, the table of similarities more clearly shows that also inside the *M. marginata* group there are differences between samples, even though being from the same species. Table 10 shows that sample Mm04 present the lowest level of similarities between the *M. marginata* group (around 0.24), followed by sample Mm07 (around 0.45%). These tendencies were expected since EEGs of geopropolis of *M. marginata* had presented the same tendencies in the previous parts of this work.

Table 10 – Table of similarities of ^{1}H NMR spectra of *M. quadrifasciata*, *M. marginata*, and *T. angustula*.

	Mm03	Mm04	Mm05	Mm06	Mm07	Mm09	Mq02	Mq03	Mq04	Mq05	Mq06	Mq07	Mq08	Mq09	Mq10	Mq13	Mq14	Mq12b	Ta02	Ta03	Ta05	Ta06	Ta07	Ta08	Ta09	Ta10
Mm03	1	0,28	0,70	0,61	0,43	0,54	0,57	0,69	0,60	0,59	0,61	0,56	0,64	0,63	0,65	0,54	0,60	0,78	0,69	0,62	0,61	0,69	0,72	0,72	0,58	0,56
Mm04	0,28	1	0,29	0,25	0,24	0,24	0,29	0,31	0,29	0,31	0,29	0,28	0,29	0,29	0,29	0,27	0,28	0,28	0,3	0,31	0,31	0,31	0,3	0,3	0,31	0,31
Mm05	0,70	0,29	1	0,60	0,48	0,59	0,53	0,68	0,53	0,53	0,55	0,51	0,58	0,58	0,59	0,48	0,55	0,71	0,62	0,57	0,56	0,61	0,61	0,54	0,52	0,51
Mm06	0,61	0,25	0,60	1	0,45	0,53	0,49	0,57	0,48	0,46	0,49	0,45	0,53	0,52	0,54	0,43	0,49	0,64	0,57	0,49	0,5	0,55	0,53	0,47	0,46	0,45
Mm07	0,43	0,24	0,48	0,45	1	0,68	0,42	0,43	0,38	0,36	0,39	0,37	0,42	0,39	0,43	0,34	0,41	0,44	0,36	0,37	0,33	0,37	0,35	0,3	0,37	0,35
Mm09	0,54	0,24	0,59	0,53	0,68	1	0,49	0,54	0,45	0,44	0,47	0,44	0,51	0,49	0,52	0,42	0,49	0,55	0,44	0,45	0,4	0,46	0,44	0,35	0,43	0,4
Mq02	0,57	0,29	0,53	0,49	0,42	0,49	1	0,71	0,74	0,75	0,7	0,67	0,75	0,69	0,77	0,64	0,68	0,62	0,66	0,73	0,64	0,72	0,67	0,53	0,73	0,7
Mq03	0,69	0,31	0,68	0,57	0,43	0,54	0,71	1	0,78	0,74	0,79	0,72	0,81	0,8	0,78	0,7	0,74	0,77	0,73	0,76	0,67	0,78	0,73	0,59	0,69	0,66
Mq04	0,6	0,29	0,53	0,48	0,38	0,45	0,74	0,78	1	0,81	0,85	0,87	0,86	0,87	0,78	0,82	0,83	0,67	0,66	0,76	0,62	0,76	0,7	0,53	0,71	0,66
Mq05	0,59	0,31	0,53	0,46	0,36	0,44	0,75	0,74	0,81	1	0,74	0,73	0,79	0,75	0,81	0,7	0,74	0,64	0,73	0,86	0,71	0,84	0,77	0,59	0,84	0,81
Mq06	0,61	0,29	0,55	0,49	0,39	0,47	0,7	0,79	0,85	0,74	1	0,87	0,83	0,85	0,77	0,81	0,81	0,69	0,65	0,72	0,6	0,73	0,67	0,52	0,67	0,64
Mq07	0,56	0,28	0,51	0,45	0,37	0,44	0,67	0,72	0,87	0,73	0,87	1	0,81	0,87	0,73	0,87	0,84	0,63	0,6	0,7	0,56	0,7	0,64	0,47	0,64	0,6
Mq08	0,64	0,29	0,58	0,53	0,42	0,51	0,75	0,81	0,86	0,79	0,83	0,81	1	0,86	0,87	0,76	0,81	0,71	0,69	0,79	0,64	0,79	0,72	0,55	0,72	0,68
Mq09	0,63	0,29	0,58	0,52	0,39	0,49	0,69	0,8	0,87	0,75	0,85	0,87	0,86	1	0,78	0,8	0,82	0,71	0,67	0,76	0,62	0,76	0,7	0,54	0,66	0,63
Mq10	0,65	0,29	0,59	0,54	0,43	0,52	0,77	0,78	0,78	0,81	0,77	0,73	0,87	0,78	1	0,7	0,76	0,7	0,71	0,79	0,66	0,79	0,73	0,57	0,75	0,73
Mq11	0,54	0,27	0,48	0,43	0,34	0,42	0,64	0,7	0,82	0,7	0,81	0,87	0,76	0,8	0,7	1	0,87	0,61	0,58	0,67	0,55	0,67	0,63	0,46	0,61	0,58
Mq12	0,6	0,28	0,55	0,49	0,41	0,49	0,68	0,74	0,83	0,74	0,81	0,84	0,81	0,82	0,76	0,87	1	0,67	0,62	0,7	0,57	0,71	0,65	0,49	0,65	0,62
Mq12b	0,78	0,28	0,71	0,64	0,44	0,55	0,62	0,77	0,67	0,64	0,69	0,63	0,71	0,71	0,7	0,61	0,67	1	0,75	0,67	0,63	0,74	0,73	0,6	0,61	0,58
Ta02	0,69	0,3	0,62	0,57	0,36	0,44	0,66	0,73	0,66	0,73	0,65	0,6	0,69	0,67	0,71	0,58	0,62	0,75	1	0,79	0,8	0,82	0,83	0,68	0,74	0,74
Ta03	0,62	0,31	0,57	0,49	0,37	0,45	0,73	0,76	0,76	0,86	0,72	0,7	0,79	0,76	0,79	0,67	0,7	0,67	0,79	1	0,8	0,9	0,83	0,63	0,81	0,81
Ta05	0,61	0,31	0,56	0,5	0,33	0,4	0,64	0,67	0,62	0,71	0,6	0,56	0,64	0,62	0,66	0,55	0,57	0,63	0,8	0,8	1	0,82	0,82	0,69	0,74	0,8
Ta06	0,69	0,31	0,61	0,55	0,37	0,46	0,72	0,78	0,76	0,84	0,73	0,7	0,79	0,76	0,79	0,67	0,71	0,74	0,82	0,9	0,82	1	0,9	0,69	0,77	0,78
Ta07	0,72	0,3	0,61	0,53	0,35	0,44	0,67	0,73	0,7	0,77	0,67	0,64	0,72	0,7	0,73	0,63	0,65	0,73	0,83	0,83	0,82	0,9	1	0,74	0,73	0,76
Ta08	0,72	0,3	0,54	0,47	0,3	0,35	0,53	0,59	0,53	0,59	0,52	0,47	0,55	0,54	0,57	0,46	0,49	0,6	0,68	0,63	0,69	0,69	0,74	1	0,59	0,62
Ta09	0,58	0,31	0,52	0,46	0,37	0,43	0,73	0,69	0,71	0,84	0,67	0,64	0,72	0,66	0,75	0,61	0,65	0,61	0,74	0,81	0,74	0,77	0,73	0,59	1	0,86
Ta10	0,56	0,31	0,51	0,45	0,35	0,4	0,7	0,66	0,66	0,81	0,64	0,6	0,68	0,63	0,73	0,58	0,62	0,58	0,74	0,81	0,8	0,78	0,76	0,62	0,86	1

*Blue colours indicate more similarity between samples, while shades of red points out less similarity.

The ^{1}H NMR spectra of geopropolis of *M. quadrifasciata*, and *T. angustula*, demonstrates a certain level of similarities, as shown by Table 10, being 0.46 the lowest level and 0.85 the highest.

6.4.2 Composition of geopropolis assessed by NMR

In the course of this work, the EEGs have presented a rich, varied, and complex composition with flavonoids, lipids, phenolic compounds, among others. The analysis of *in natura* geopropolis by NMR of geopropolis, using heavy chloroform ($CDCl_3$) as solvent, confirmed these features. In fact, NMR spectra acquired during this study implies that a considerable number of the structures "annotated" by MassHunter V. 10.0, and Compound Discoverer V. 3.3 in Parts 3, 4 and 5 respectively, were correct.

Geopropolis is a highly complex matrix, and as consequence many signals are mixed and overlapped, with several being considerably small and poorly defined, making it difficult to identify all the compounds, therefore this study was focused on the recognition of compound classes, as well as differences and similarities between the species.

6.4.3 Distinctive signals and compounds

Each species of stingless bees produces geopropolis with their own characteristic composition, as observed by the grouping of species in the PCA (Figure 37), and the clusters in the table of similarities (Table 10). An overall looking at the ^{1}H NMR spectra of geopropolis of *M. marginata, M. quadrifasciata*, and *T. angustula*, is sufficient to perceive differences among them. However, *M. quadrifasciata*, and *T. angustula* presents more similarities when compared to *M. marginata*. This condition was already observed in the previous chapters of this work.

Nonetheless, Figure 38, Figure **40**, and Figure **44**, briefly shows the ^{1}H NMR spectra *M. marginata*, *M. quadrifasciata*, and *T. angustula* "stacked" for a rapid comparative effect.

6.4.3.1 *Tetragonisca angustula*

According to the table of similarities (Table 10), geopropolis of *T. angustula* presents similar ^{1}H NMR spectra between samples regardless their geographical origin, being 0.90 the highest, and 0.56 the lowest similarities. Therefore, it is expected for *T. angustula* to present geopropolis with similar signals and spectra as Figure 38 shows. This is further evidence of the selective habits of *T. angustula* regarding resin collection to make propolis.

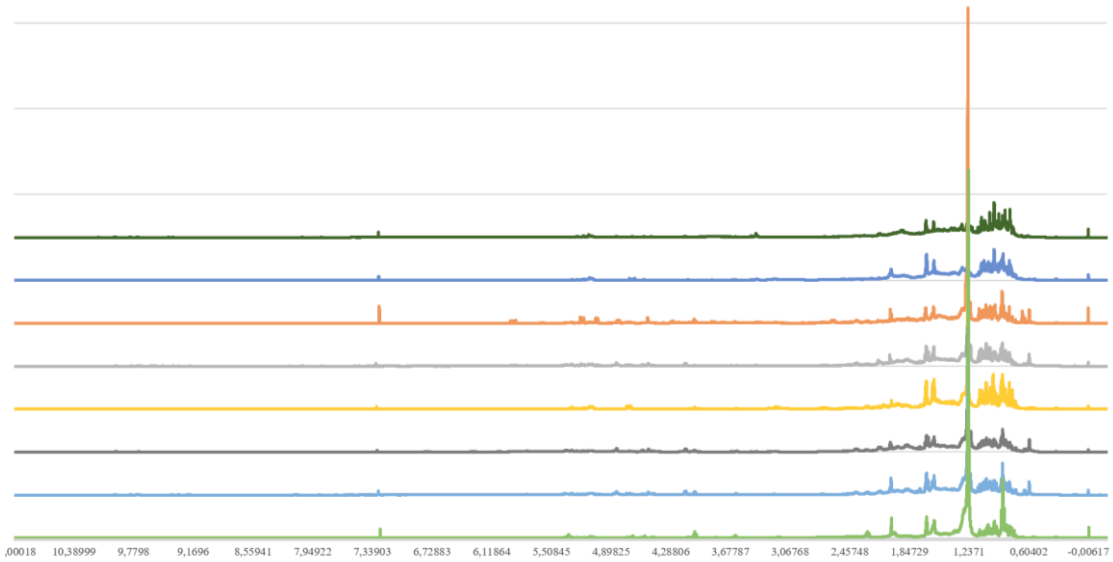


Figure 38 – ^{1}H NMR spectra of geopropolis of *T. angustula*, from -0.006 to 11.000 ppm.

In **Figure 39** – Section of ^{1H} NMR spectra of geopropolis from *T. angustula,* interval from 0.45 to 2.30 ppmFigure 39 there is a prominent peak with chemical shift of 1.26 ppm common to the spectra of all samples, preceded by a multiplet mostly related to the overlapping of CH_2 signals, characteristically for long carbonic chain fatty acids, such as

palmitic acid. Other signals are related to palmitic acid, shown in more detail in Figure 39, which are a multiplet in 1.64 ppm indicating H attached to α-carbon, and a triplet at 0.89 ppm related to the terminal methyl group [104].

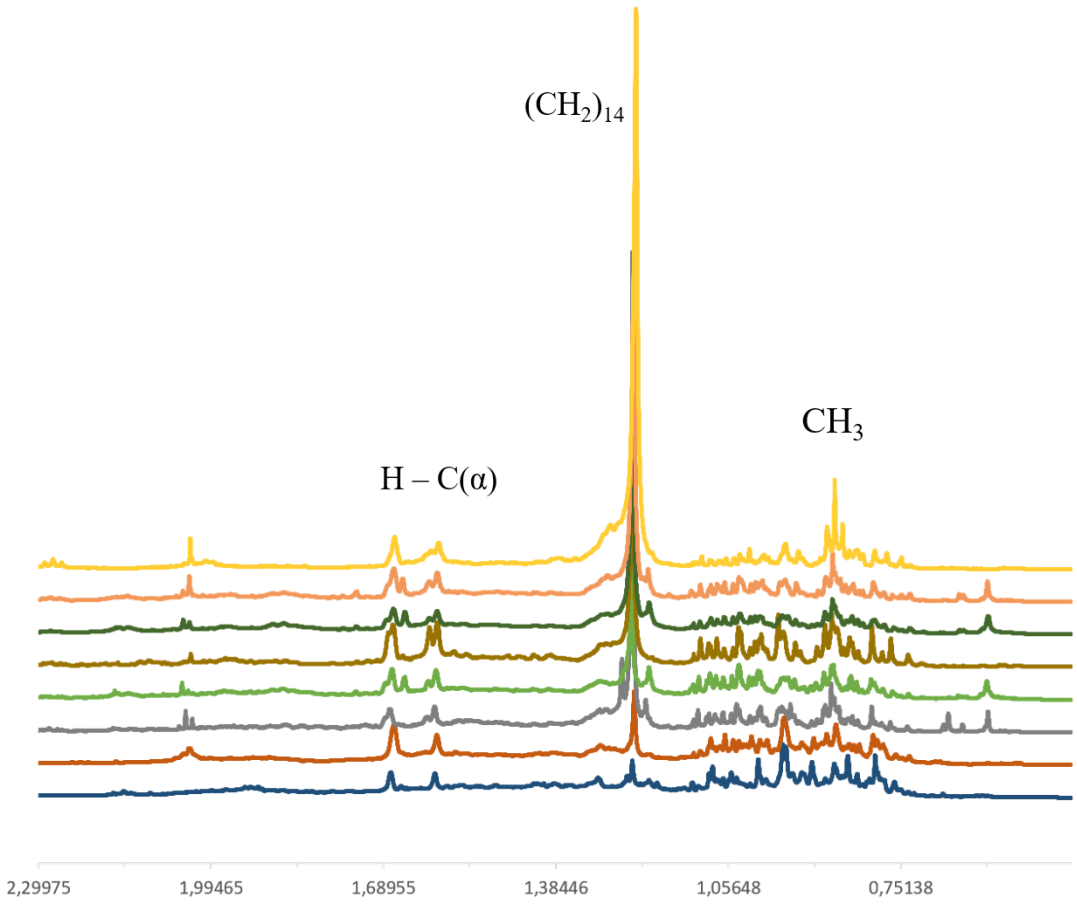


Figure 39 – Section of ^{1}H NMR spectra of geopropolis from *T. angustula*, interval from 0.45 to 2.30 ppm.

Also known as hexadecenoic acid, palmitic acid is a saturated fatty acid, and a lipid commonly found in plants, animals, and microorganisms. Naik et al. [105] identified fatty acids with carbon chains ranging from 4 to 24 carbons in propolis of *A. mellifera* from Jordan.

Turco et al. [106] reported fatty acids, with saturated and unsaturated carbonic chains, in extracts of geopropolis from several species, and also described in the Part 4 of

this work. The presence of characteristic signals of lipids is in concordance with previous results obtained during the lipidomic analysis (Part 4) of extracts of geopropolis.

6.4.3.2 *Melipona quadrifasciata*

The spectra of CEGs of *M. quadrifasciata* shown in Figure **40**, similarly to *T. angustula*, presented a variety of signals in the spectral region characteristically occupied by alkenyl hydrogens, which are found normally from 0.00 to 4.00 ppm (using TMS signal as reference). This is mostly from lipids which are naturally present in geopropolis, many already registered, during lipidomic analysis in Chapter 4. However, in this case the aromatic region presents more diversification of signals, therefore different compounds can possibly be found.

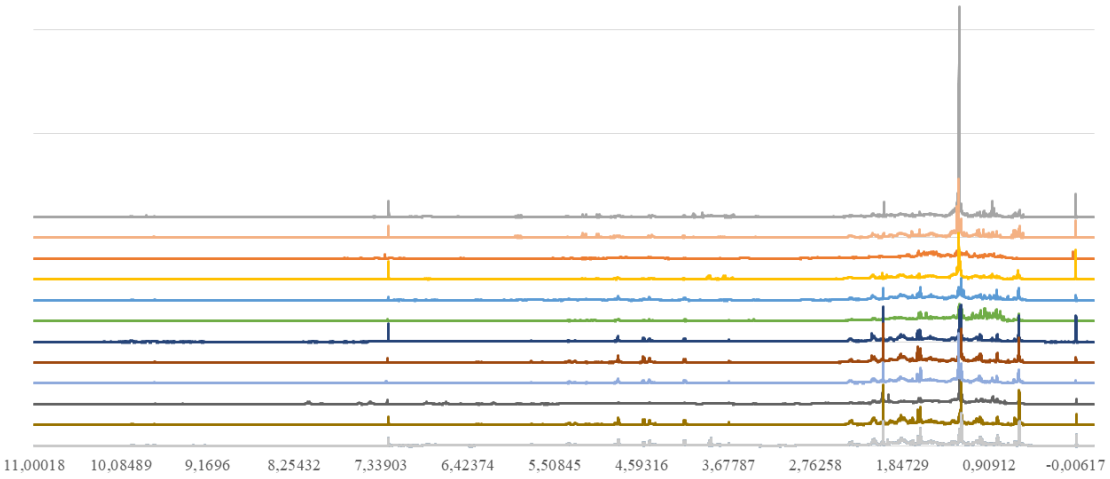


Figure 40 – Stacking of ^{1}H NMR spectra of geopropolis of *M. quadrifasciata*, from -0.006 to 11.000 ppm.

A doublet of doublets (dd) with shifting around 6.32 ppm (J = 17.4 Hz, 10.6 Hz) is an interesting group of signals observable in all the *M. quadrifasciata* ^{1}H NMR spectra, as shown by Figure 41.

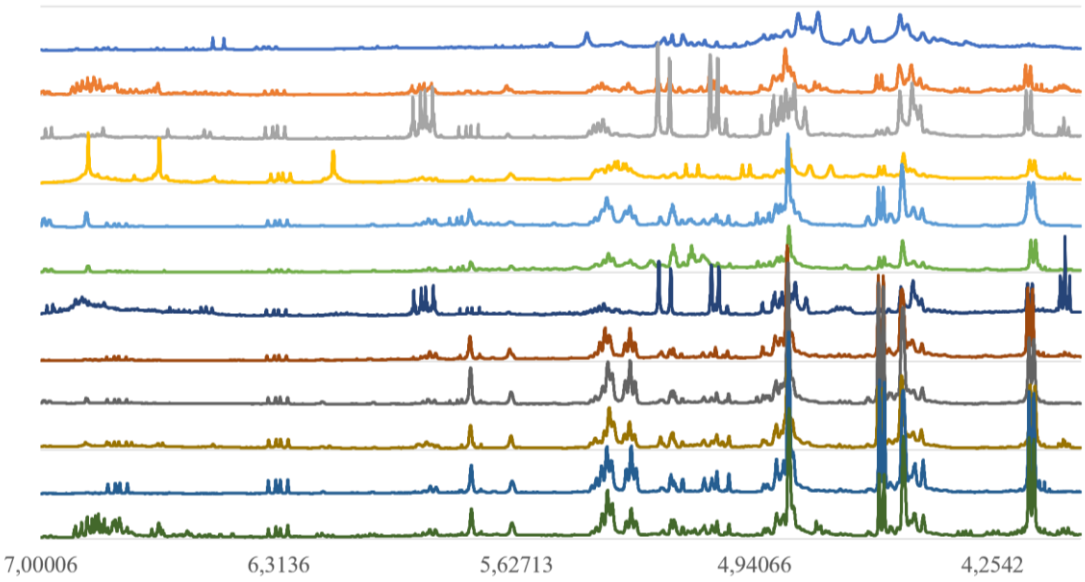


Figure 41 – Section between 4.00 and 7.00 ppm of the ^{1}H NMR spectra of geopropolis of *M. quadrifasciata*.

In NMR spectroscopy a doublet of doublets is a signal that is split into a doublet, and each line of this doublet split again into another doublet, occurring when a group of magnetically identical protons is coupled with two different protons, with different coupling constants [107].

α-Farnesene (Figure 42) is a sesquiterpene isoprenoid ($C_{15}H_{24}$) found in essential oils of several vegetal species, such as orange, and rose. α-Farnesene is versatile, being extensively applied in cosmetics, and medicine, due to its antibacterial and antifungal properties. In addition, farnesene can also be used in diesel fuel [108].

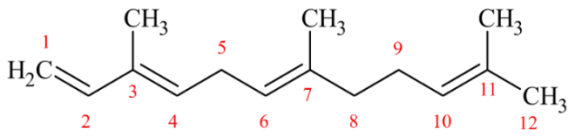


Figure 42 – Structure of α–Farnesene.

Table 11 shows the HSQC, and HMBC, signals associated to α–Farnesene, which were detected in all samples of geopropolis of *M. quadrifasciata*.

Table 11 – Signals assigned to α-Farnesene.

Position	^{1}H	^{13}C	HMBC	Multiplicity*, *J*
1	5.02, 5.06	110.01	2, 3, 4	dd (*J* = 10.6 Hz, 1.7 Hz)
2	6.32	141.60	3, 3’	dd (2H, *J* = 17.4 Hz, 10.6 Hz)
3		133.89		
3’	1.74	11.77	2, 3, 4	s (3H)
4	5.31	131.60	2, 3’	t (*J* = 6.4 Hz)
5	2.41	26.19	3, 4	t (2H, *J* = 6.4 Hz)
6	5.31	122.84	2, 5	
7		135.11		
7’	1.66	16.35	6, 8	s (3H)
8	2.17	38.22	9	t (1H)
9	2.18	26.10	8	
10	5.23	125.16	9	t (*J* = 7.2 Hz)
11		131.27		
11’	1.54	17.71	10, 11, 12	s (3H)
12	1.54	25.20	10, 11, 11’	s (3H)

*s – singlet, t – triplet, dd – doublet of doublets,

trans-Cinnamic acid was observed in ^{1}H spectrum of geopropolis of *M. quadrifasciata quadrifasciata* (Figure 43), which appear not so prominently in other spectra of *M. quadrifasciata*.

According to experimental data, ^{1}H NMR spectra of *trans*-cinnamic acid presents two doublets near to 6.46 and 7.69 (*J* = 15.6 Hz), and a multiplet around 7.40 ppm [109].

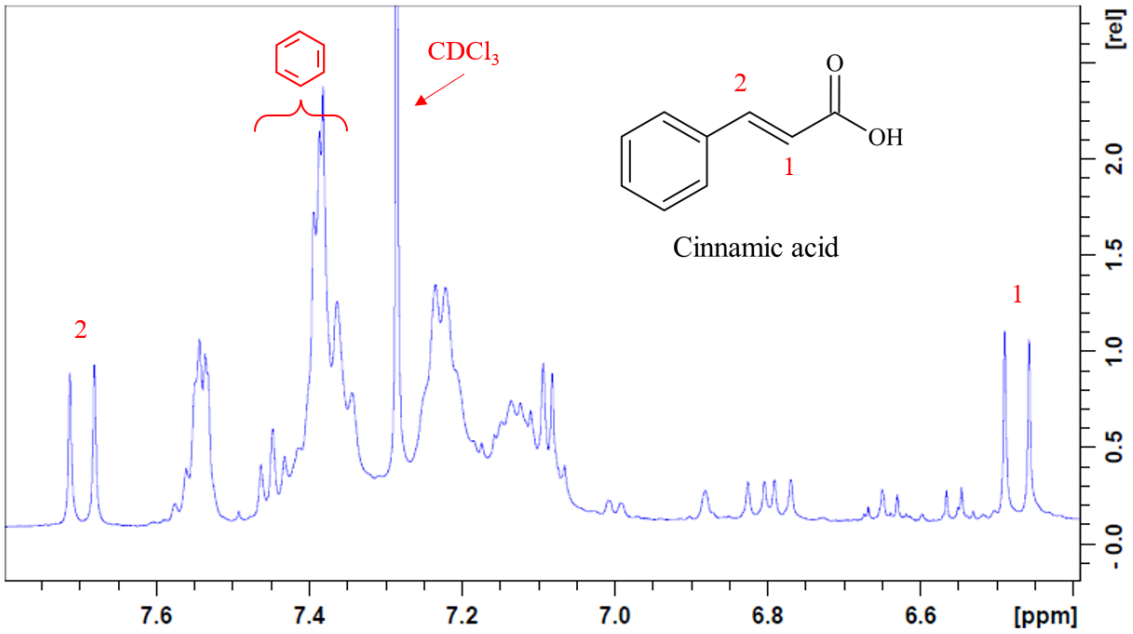


Figure 43 – Section of *M. quadrifasciata quadrifaciata* ^{1}H NMR spectrum. t*rans*-Cinnamic acid characteristic doublets are noticeable in 6.46 and 7.69 ppm.

6.5.4.3 *Melipona marginata*

Figure 44 shows that geopropolis of *M. marginata* presented more diverse spectra (and composition, consequently) between samples of the same group, when compared to *M. quadrifasciata* and *T. angustula*, which was already predicted by the table of similarities (Table 10).

The samples Mm04 and Mm09, from Curitiba-PR and Carirá-SE, respectively, presented less similarity with their counterparts from Prudentópolis-PR region. As different regions present different plants, variations in the geopropolis composition are expected for some stingless bees. In previous parts of this Work, geopropolis of *M. marginata* stingless bees also presented variances in their composition.

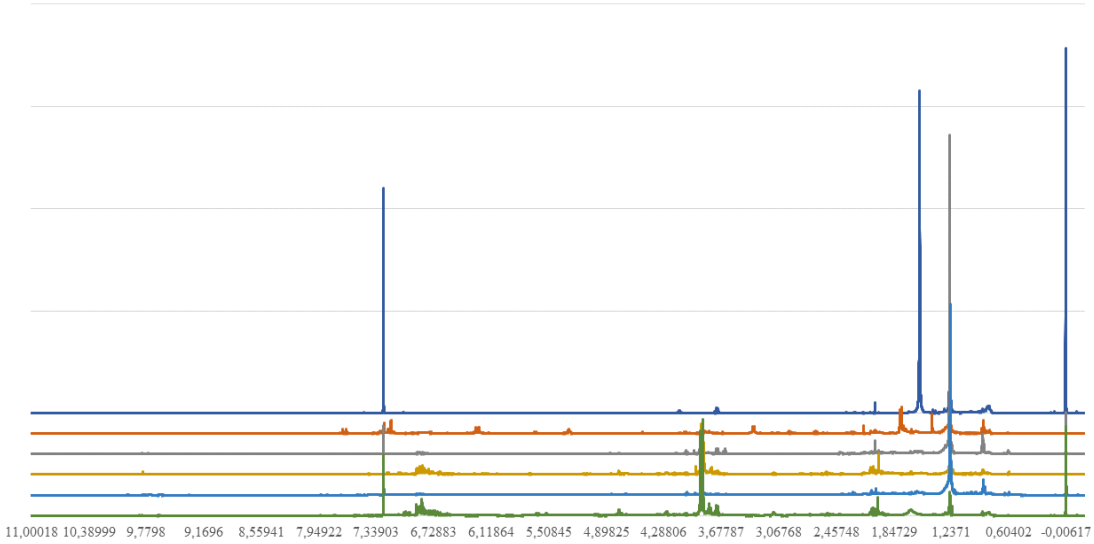


Figure 44 – Stacking of ^{1}H NMR spectra of *M. marginata*.

However, there are common signals in all *M. marginata* spectra, which are a prominent singlet at 1.27 ppm, a triplet in 0.90 ppm (J = 7.14 Hz, 3H), a short singlet in 2.067 ppm. According to experimental data [109], these signals characterise the capric acid (dodecanoic acid).

6.5 Partial conclusions from NMR analysis of CEGs

Geopropolis is rich in terpenoids, fatty acids, and other non-polar (or less polar) compounds. As each species of stingless bees presents different habits for collection of vegetal resins, deductively *M. quadrifasciata,* and *T. angustula*, had preferences for saps that are richer in less polar compounds, being *M. marginata* probably less selective.

NMR highlighted a considerable number of signals which are characteristics of terpenoids and other complex lipids, pointing to the relevant information about the metabolomic profile of the geopropolis, that is complex to observe using chromatographic techniques. Overall, geopropolis is a natural product rich in lipids, from several classes, and with different functions.

Through NMR analysis of CEGs it was possible to confirm the occurrence of chemical classes already identified through chromatography hyphenated techniques and multivariate analysis of ^{1}H- NMR spectra confirmed the strong effect of bee species on the chemical composition of geopropolis.

Final conclusions and future studies?

Although geopropolis composition is believed to be poorer or even inferior to *Apis mellifera* propolis, the results of this study showed a rather different reality. The composition of geopropolis ethanolic extracts (EEGs) is rich and complex, containing a large range of compounds and classes of molecules. Geopropolis presented in its composition: flavonoids, phenolic compounds, phenylpropanoids, coumarins, sugars fatty acids, lipids, terpenoids, steroids, tannins, organic acids, vitamins, among other classes of compounds. Other considerations should be added:

I. Geopropolis could be classified according to both bees' species and region of origin, once these factors affect their chemical composition (lipidomic and metabolomic).

II. The mass fingerprints of the ethanolic extracts of geopropolis could be used to predict its antioxidant capacity (using the VCEAC and CUPRAC methods), as well as the total flavonoids content.

IV. The water-soluble fraction of the EEGs presented a notable variety of phenolic compounds, sugars, and organic acids.

V. The organic lipid-rich fraction showed mostly fatty acids, terpenoids, and lipids.

VI. Depending on the specificity of each analytical technique herein used to study geopropolis, it was possible to point out the most characteristic features to each kind of EEGs. Table 12 summarizes these features.

The next steps of this study will include experiments and data analysis focused on the EEGs characterisation.

Table 12 – Unique features and characteristic/marker compounds of the EEGs.

Species, Genus or Group	**Origin/Region**	**FIA-HRMS**		**LC-HRMS (Metabolomics)**	**LC-HRMS (Lipidomics)**	**GC-MS**
		m/z	Mode	Compound	Formula	Compound
M. marginata	Prudentópolis	623.3197	Pos	Ikarisoside D	$C_{38}H_{63}N_5O_4$	Vanillic acid
	Other locations	407.1467	Pos	4'-O-methyldavidigenin	$C_{46}H_{89}O_9P$	
M. quadrifasciata	Prudentópolis	707.4500	Pos	Salicylic acid	$C_{39}H_{60}N_6O_4$	Benzoquinone
	Curitiba Mogi Mirim Pilar do Sul	428.2493	Pos	-	$C_{55}H_{114}N_9P$	
M. scutellaris	Curitiba Mogi Mirim Prudentópolis Pilar do Sul	283.0066	Neg	-	$C_{51}H_{72}N_4$	D-(+)-trehalose
T. angustula	Prudentópolis	544.3336	Pos	Melilotoside A2	$C_{40}H_{66}N_6O_2$	Threonic acid
	Other locations	507.3721	Pos	-		
Melipona	Petrolina	572.6178	Pos	-	$C_{22}H_{24}O_6$	6-deoxy-D-glucose
Non-*Melipona*	Catanduvas Pilar do Sul Prudentópolis	517.3721	Pos	-	$C_{47}H_{81}N_2OP$	-

Also, PLS regression will be applied to other datasets (GC-MS, metabolomics, and lipidomics), aiming to find out if there are significative correlations between them and the results from the reference tests (TFC, and AOC).

Undoubtedly, this study provided important insights about the geopropolis composition, and its characteristics, converging four distinct analytical ways into a single larger study, which can potentially be used in the future as reference compendium.

References

1. Coelho GR, Figueiredo CA, Negri G, Fernandes-Silva CC, Villar KDS, Badari JC, Oliveira MI De, Barbosa TF, Taniwaki NN, Namiyama GM, Mendonca RZ (2018) Antiviral Activity of Geopropolis Extract from Scaptotrigona Aff. Postica against Rubella Virus. J Food Res 7:91 . https://doi.org/10.5539/jfr.v7n6p91

2. dos Santos L, Hochheim S, Boeder AM, Kroger A, Tomazzoli MM, Dal Pai Neto R, Maraschin M, Guedes A, de Cordova CMM (2017) Chemical characterization, antioxidant, cytotoxic and antibacterial activity of propolis extracts and isolated compounds from the Brazilian stingless bees Melipona quadrifasciata and Tetragonisca angustula. J Apic Res 56:543–558 . https://doi.org/10.1080/00218839.2017.1371535

3. da Cunha MG, Franchin M, de Paula-Eduardo LF, Freires IA, Beutler JA, de Alencar SM, Ikegaki M, Tabchoury CPM, Cunha TM, Rosalen PL (2016) Anti-inflammatory and anti-biofilm properties of ent -nemorosone from Brazilian geopropolis. J Funct Foods 26:27–35 . https://doi.org/10.1016/j.jff.2016.07.009

4. Camargo RCR de, Oliveira KL de, Berto MI (2017) Mel de abelhas sem ferrão: proposta de regulamentação. Brazilian J Food Technol 20:1–6 . https://doi.org/10.1590/1981-6723.15716

5. Lavinas FC, Macedo EHBC, Sá GBL, Amaral ACF, Silva JRA, Azevedo MMB, Vieira BA, Domingos TFS, Vermelho AB, Carneiro CS, Rodrigues IA (2019) Brazilian stingless bee propolis and geopropolis: promising sources of biologically active compounds. Rev Bras Farmacogn 29:389–399 . https://doi.org/10.1016/j.bjp.2018.11.007

6. Dutra RP, Bezerra JL, Silva MCP da, Batista MCA, Patrício FJB, Nascimento FRF, Ribeiro MNS, Guerra RNM (2019) Antileishmanial activity and chemical composition from Brazilian geopropolis produced by stingless bee Melipona fasciculata. Brazilian J Pharmacogn 29:287–293 . https://doi.org/10.1016/j.bjp.2019.02.009

7. Popova M, Trusheva B, Bankova V (2021) Propolis of stingless bees: A phytochemist's guide through the jungle of tropical biodiversity. Phytomedicine 86:153098 . https://doi.org/10.1016/j.phymed.2019.153098

8. Sforcin JM (2017) Própolis E Geoprópolis Uma Herança Das Abelhas. 100

9. Salatino A, Salatino MLF, Negri G (2021) How diverse is the chemistry and plant origin of Brazilian propolis? Apidologie 52:1075–1097 . https://doi.org/10.1007/s13592-021-00889-z

10. Costa AS, Machado BAS, Umsza-Guez MA, Cirqueira MG, Nunes SB, Padilha FF (2013) Levantamento dos estudos realizados com a própolis produzida no estado da Bahia. SITIENTIBUS série Ciências Biológicas 13:1–7 . https://doi.org/10.13102/scb324

11. Villas-Bôas J (2018) Manual Tecnológico Mel de Abelhas Sem Ferrão, 2nd ed. Instituto Sociedade, População e Natureza (ISPN), Brasília

12. Cardozo D V., Mokochinski JB, Machado CS, Sawaya ACHF, Caetano IK, Felsner ML, Torres YR (2015) Chemical variability of geopropolis from Jataí, Mandaçaia and mandurí stingless bees. Rev Virtual Quim 7:2457–2474 . https://doi.org/10.5935/1984-6835.20150146

13. Bankova V, Christov R, Marcucci C, Popov S (1998) Constituents of Brazilian

geopropolis. Zeitschrift fur Naturforsch - Sect C J Biosci 53:402–406 . https://doi.org/10.1515/znc-1998-5-616

14. Kujumgiev A, Tsvetkova I, Serkedjieva Y, Bankova V, Christov R, Popov S (1999) Antibacterial, antifungal and antiviral activity of propolis of different geographic origin. J Ethnopharmacol 64:235–240 . https://doi.org/10.1016/S0378-8741(98)00131-7

15. Dutra RP, Nogueira AMC, Marques RRDO, Costa MCP, Ribeiro MNS (2008) Avaliação farmacognóstica de geoprópolis de Melipona fasciculata Smith da Baixada maranhense, Brasil. Brazilian J Pharmacogn 18:557–562 . https://doi.org/10.1590/S0102-695X2008000400010

16. De Souza SA, Camara CA, Da Silva EMS, Silva TMS (2013) Composition and Antioxidant Activity of Geopropolis Collected by Melipona subnitida (Jandaíra) Bees. Evidence-Based Complement Altern Med 2013:1–5 . https://doi.org/10.1155/2013/801383

17. Cardozo D V., Mokochinski JB, Machado CS, Sawaya ACHF, Caetano IK, Felsner ML, Torres YR (2015) Chemical Variability of Geopropolis from Jataí, Mandaçaia and Mandurí Stingless Bees. Rev Virtual Química 7:2457–2474 . https://doi.org/10.5935/1984-6835.20150146

18. Ruiz Ruiz JC, Pacheco López NA, Rejón Méndez EG, Samos López FA, Medina Medina L, Quezada-Euán JJG (2023) Phenolic Content and Bioactivity as Geographical Classifiers of Propolis from Stingless Bees in Southeastern Mexico. Foods 12:1434 . https://doi.org/10.3390/foods12071434

19. Bankova VS, de Castro SL, Marcucci MC (2000) Propolis: recent advances in chemistry and plant origin. Apidologie 31:3–15 .

https://doi.org/10.1051/apido:2000102

20. Maraschin M, Somensi-Zeggio A, Oliveira SK, Kuhnen S, Tomazzoli MM, Raguzzoni JC, Zeri ACM, Carreira R, Correia S, Costa C, Rocha M (2016) Metabolic Profiling and Classification of Propolis Samples from Southern Brazil: An NMR-Based Platform Coupled with Machine Learning. J Nat Prod 79:13–23 . https://doi.org/10.1021/acs.jnatprod.5b00315

21. Araújo KS da S, dos Santos Júnior JF, Sato MO, Finco FDBA, Soares IM, Barbosa R dos S, Alvim TDC, Ascêncio SD, Mariano SMB (2016) Physicochemical properties and antioxidant capacity of propolis of stingless bees (Meliponinae) and apis from two regions of Tocantins, Brazil. Acta Amaz 46:61–68 . https://doi.org/10.1590/1809-4392201501045

22. Bankova V (2005) Chemical diversity of propolis and the problem of standardization. J. Ethnopharmacol. 100:114–117

23. Hall Z, Bond NJ, Ashmore T, Sanders F, Ament Z, Wang X, Murray AJ, Bellafante E, Virtue S, Vidal-Puig A, Allison M, Davies SE, Koulman A, Vacca M, Griffin JL (2017) Lipid zonation and phospholipid remodeling in nonalcoholic fatty liver disease. Hepatology 65:1165–1180 . https://doi.org/10.1002/hep.28953

24. Pang Z, Chong J, Zhou G, Anderson De Lima Morais D, Chang L, Barrette M, Gauthier C, Jacques P-´ E, Li S, Xia J (2021) MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Res 49: . https://doi.org/10.1093/nar/gkab382

25. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, Wishart DS, Xia J (2018) MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Res 46:W486–W494 . https://doi.org/10.1093/nar/gky310

26. Woisky RG, Salatino A (1998) Analysis of propolis: some parameters and procedures for chemical quality control. J Apic Res 37:99–105 . https://doi.org/10.1080/00218839.1998.11100961

27. Brand-Williams W, Cuvelier ME, Berset C (1995) Use of a free radical method to evaluate antioxidant activity. LWT - Food Sci Technol 28:25–30 . https://doi.org/10.1016/S0023-6438(95)80008-5

28. Apak R, Güçlü K, Özyürek M, Karademir SE (2004) Novel total antioxidant capacity index for dietary polyphenols and vitamins C and E, using their cupric ion reducing capability in the presence of neocuproine: CUPRAC method. J Agric Food Chem 52:7970–7981 . https://doi.org/10.1021/jf048741x

29. Turco JF, do Nascimento CL, de Lima VA, Torres YR (2020) Could antioxidant capacity and flavonoid content of ethanolic extracts of geopropolis from Brazilian native bees be estimated from digital photos and NIR Spectra? Microchem J 157:105031 . https://doi.org/10.1016/j.microc.2020.105031

30. Helfer GA, Bock F, Marder L, Furtado JC, Da Costa AB, Ferrão MF (2015) Chemostat, um software gratuito para análise exploratória de dados multivariados. Quim Nova 38:575–579 . https://doi.org/10.5935/0100-4042.20150063

31. Peris-Díaz MD, Krężel A (2021) A guide to good practice in chemometric methods for vibrational spectroscopy, electrochemistry, and hyphenated mass spectrometry. TrAC Trends Anal Chem 135:116157 . https://doi.org/10.1016/j.trac.2020.116157

32. Rinnan Å, Berg F van den, Engelsen SB (2009) Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends Anal Chem 28:1201–1222 . https://doi.org/10.1016/j.trac.2009.07.007

33. Kumar S (2016) Analytical techniques for natural product research. CABI, Wallingford

34. Hackstadt AJ, Hess AM (2009) Filtering for increased power for microarray data analysis. BMC Bioinformatics 10:11 . https://doi.org/10.1186/1471-2105-10-11

35. Sawaya ACHF, da Silva Cunha IB, Marcucci MC, Aidar DS, Silva ECA, Carvalho CAL, Eberlin MN (2007) Electrospray ionization mass spectrometry fingerprinting of propolis of native Brazilian stingless bees. Apidologie 38:93–103 . https://doi.org/10.1051/apido:2006058

36. dos Santos CM, Campos JF, dos Santos HF, Balestieri JBP, Silva DB, de Picoli Souza K, Carollo CA, Estevinho LM, dos Santos EL (2017) Chemical Composition and Pharmacological Effects of Geopropolis Produced by Melipona quadrifasciata anthidioides. Oxid Med Cell Longev 2017:1–13 . https://doi.org/10.1155/2017/8320804

37. Jazmin M PR (2020) Advances in Chemical Composition and Biological Activity of Mexican Propolis. LOJ Pharmacol Clin Res 2:194–201 . https://doi.org/10.32474/lojpcr.2020.02.000137

38. Cunha MS, Dutras RP, Batista MCA, Abreu BV de B, Santos JR dos, Neiva V do A, Amaral FMM do, Ribeiro MN de S (2009) Padronização de extrativos de geoprópolis de Melipona fasciculata Smith (Tiúba). Cad Pesqui 16:31–38

39. Pan R, Guo F, Lu H, Feng W wei, liang Y zeng (2011) Development of the chromatographic fi{ligature}ngerprint of Scutellaria barbata D. Don by GC-MS combined with Chemometrics methods. J Pharm Biomed Anal 55:391–396 . https://doi.org/10.1016/j.jpba.2011.01.016

40. Sawaya ACHF, Cunha IBS, Marcucci MC, de Oliveira Rodrigues RF, Eberlin MN (2006) Brazilian Propolis of Tetragonisca angustula and Apis mellifera. Apidologie 37:398–407 . https://doi.org/10.1051/apido:2006011

41. Turco JF, do Nascimento CL, de Lima VA, Torres YR (2020) Could antioxidant capacity and flavonoid content of ethanolic extracts of geopropolis from Brazilian native bees be estimated from digital photos and NIR Spectra? Microchem J 157:105031 . https://doi.org/10.1016/j.microc.2020.105031

42. Tsanaktsidou E, Karavasili C, Zacharis CK, Fatouros DG, Markopoulou CK (2020) Partial Least Square Model (PLS) as a Tool to Predict the Diffusion of Steroids Across Artificial Membranes. Molecules 25:1387 . https://doi.org/10.3390/molecules25061387

43. Berrar D (2019) Cross-Validation. Encycl Bioinforma Comput Biol ABC Bioinforma 1–3:542–545 . https://doi.org/10.1016/B978-0-12-809633-8.20349-X

44. Rochat B (2017) Proposed Confidence Scale and ID Score in the Identification of Known-Unknown Compounds Using High Resolution MS Data. J Am Soc Mass Spectrom 28:709–723 . https://doi.org/10.1007/s13361-016-1556-0

45. Ferreira BL, Gonzaga LV, Vitali L, Micke GA, Baggio D, de Oliveira Costa AC, Fett R (2020) Dataset about Southern-Brazilian geopropolis: Physical and chemical perspectives. Data Br 29:105109 . https://doi.org/10.1016/j.dib.2020.105109

46. de Souza SA, da Silva TMG, da Silva EMS, Camara CA, Silva TMS (2018) Characterisation of phenolic compounds by UPLC-QTOF-MS/MS of geopropolis from the stingless bee *Melipona subnitida* (jandaíra). Phytochem Anal 29:549–558 . https://doi.org/10.1002/pca.2766

47. Zhao L, Yu M, Sun M, Xue X, Wang T, Cao W, Sun L (2017) Rapid Determination of Major Compounds in the Ethanol Extract of Geopropolis from Malaysian Stingless Bees, Heterotrigona itama, by UHPLC-Q-TOF/MS and NMR. Molecules 22:1935 . https://doi.org/10.3390/molecules22111935

48. Bonamigo T, Campos JF, Oliveira AS, Torquato HFV, Balestieri JBP, Cardoso CAL, Paredes-Gamero EJ, Souza K de P, Dos Santos EL (2017) Antioxidant and cytotoxic activity of propolis of Plebeia droryana and Apis mellifera (Hymenoptera, Apidae) from the Brazilian Cerrado biome. PLoS One 12: . https://doi.org/10.1371/journal.pone.0183983

49. Batista MCA, ABREU BV de B, DUTRA RP, CUNHA MS, AMARAL FMM do, TORRES LMB, RIBEIRO MN de S (2016) Chemical composition and antioxidant activity of geopropolis produced by Melipona fasciculata (Meliponinae) in flooded fields and cerrado areas of Maranhão State, northeastern Brazil. Acta Amaz 46:315–322 . https://doi.org/10.1590/1809-4392201600034

50. Dutra RP, Abreu BV de B, Cunha MS, Batista MCA, Torres LMB, Nascimento FRF, Ribeiro MNS, Guerra RNM (2014) Phenolic Acids, Hydrolyzable Tannins, and Antioxidant Activity of Geopropolis from the Stingless Bee Melipona fasciculata Smith. J Agric Food Chem 62:2549–2557 . https://doi.org/10.1021/jf404875v

51. Zhao L, Yu M, Sun M, Xue X, Wang T, Cao W, Sun L (2017) Rapid Determination of Major Compounds in the Ethanol Extract of Geopropolis from Malaysian Stingless Bees, Heterotrigona itama, by UHPLC-Q-TOF/MS and NMR. Molecules 22:1935 . https://doi.org/10.3390/molecules22111935

52. Cisilotto J, Sandjo LP, Faqueti LG, Fernandes H, Joppi D, Biavatti MW,

Creczynski-Pasa TB (2018) Cytotoxicity mechanisms in melanoma cells and UPLC-QTOF/MS2 chemical characterization of two Brazilian stingless bee propolis: Uncommon presence of piperidinic alkaloids. J Pharm Biomed Anal 149:502–511 . https://doi.org/10.1016/j.jpba.2017.11.038

53. Syed Salleh SNA, Mohd Hanapiah NA, Ahmad H, Wan Johari WL, Osman NH, Mamat MR (2021) Determination of Total Phenolics, Flavonoids, and Antioxidant Activity and GC-MS Analysis of Malaysian Stingless Bee Propolis Water Extracts. Scientifica (Cairo) 2021:1–11 . https://doi.org/10.1155/2021/3789351

54. Ferreira JM, Fernandes-Silva CC, Salatino A, Message D, Negri G (2017) Antioxidant Activity of a Geopropolis from Northeast Brazil: Chemical Characterization and Likely Botanical Origin. Evidence-Based Complement Altern Med 2017:1–6 . https://doi.org/10.1155/2017/4024721

55. Popova M, Trusheva B, Bankova V (2019) Propolis of stingless bees: A phytochemist's guide through the jungle of tropical biodiversity. Phytomedicine 153098 . https://doi.org/10.1016/j.phymed.2019.153098

56. Keeling CI, Bohlmann J (2006) Diterpene resin acids in conifers. Phytochemistry 67:2415–2423 . https://doi.org/10.1016/j.phytochem.2006.08.019

57. González MA (2015) Aromatic abietane diterpenoids: their biological activity and synthesis. Nat Prod Rep 32:684–704 . https://doi.org/10.1039/C4NP00110A

58. Wuerges KL, Dias RCE, Viegas MC, Benassi M de T (2020) Kahweol and cafestol in coffee brews: comparison of preparation methods. Rev CIÊNCIA AGRONÔMICA 51:1–6 . https://doi.org/10.5935/1806-6690.20200005

59. Helfenstein A, Vahermo M, Nawrot DA, Demirci F, İşcan G, Krogerus S, Yli-

Kauhaluoma J, Moreira VM, Tammela P (2017) Antibacterial profiling of abietane-type diterpenoids. Bioorg Med Chem 25:132–137 . https://doi.org/10.1016/j.bmc.2016.10.019

60. González MA, Correa-Royero J, Agudelo L, Mesa A, Betancur-Galvis L (2009) Synthesis and biological evaluation of abietic acid derivatives. Eur J Med Chem 44:2468–72 . https://doi.org/10.1016/j.ejmech.2009.01.014

61. Eksi G, Kurbanoglu S, Erdem SA (2020) Analysis of diterpenes and diterpenoids. In: Recent Advances in Natural Products Analysis. Elsevier, pp 313–345

62. González MA (2015) Aromatic abietane diterpenoids: total syntheses and synthetic studies. Tetrahedron 71:1883–1908 . https://doi.org/10.1016/j.tet.2015.01.058

63. Saleem M (2009) Lupeol, a novel anti-inflammatory and anti-cancer dietary triterpene. Cancer Lett 285:109–15 . https://doi.org/10.1016/j.canlet.2009.04.033

64. Bradford PG, Awad AB (2007) Phytosterols as anticancer compounds. Mol Nutr Food Res 51:161–70 . https://doi.org/10.1002/mnfr.200600164

65. Lira WDM, Dos Santos FV, Sannomiya M, Rodrigues CM, Vilegas W, Varanda EA (2008) Modulatory effect of Byrsonima basiloba extracts on the mutagenicity of certain direct and indirect-acting mutagens in Salmonella typhimurium assays. J Med Food 11:111–119 . https://doi.org/10.1089/JMF.2007.553

66. Liu Y, Zhao N, Li C, Chang Q, Liu X, Liao Y, Pan R (2017) Longistyline C acts antidepressant in vivo and neuroprotection in vitro against glutamate-induced cytotoxicity by regulating NMDAR/NR2B-ERK pathway in PC12 cells. PLoS One 12:e0183702 . https://doi.org/10.1371/journal.pone.0183702

67. Wang S, Shen Y, Qiu R, Chen Z, Chen Z, Chen W (2017) 18 β-glycyrrhetinic acid

exhibits potent antitumor effects against colorectal cancer via inhibition of cell proliferation and migration. Int J Oncol 51:615–624 . https://doi.org/10.3892/ijo.2017.4059

68. Taha AY (2020) Linoleic acid–good or bad for the brain? npj Sci Food 4:1 . https://doi.org/10.1038/s41538-019-0061-9

69. De Sousa DMN, Olinda RG, Martins CG, Abrantes MR, Coelho WAC, Da Silva JBA, De Morais SM, Batista JS (2015) Prospec????O fitoqu??mica, toxicidade in vitro e avalia????o das atividades anti-radicalar e antibacteriana da geopr??polis da abelha janda??ra. Acta Vet Bras 9:134–140

70. Clark S (2007) Retinoids. In: xPharm: The Comprehensive Pharmacology Reference. Elsevier, pp 1–2

71. Silva Cruz LF, Santos T de S, Souza CO, Santos LSM, Druzian JI, Tavares PPLG, Nascimento RQ, Bullos RB de A, Almeida LMR (2020) Determination of physicochemical characteristics and bioactive compounds in samples of pollen, geopropolis and honey from Melipona Scutellaris bee species. Brazilian J Dev 6:21484–21496 . https://doi.org/10.34117/bjdv6n4-353

72. Pang Z, Chong J, Zhou G, de Lima Morais DA, Chang L, Barrette M, Gauthier C, Jacques P-É, Li S, Xia J (2021) MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Res 49:W388–W396 . https://doi.org/10.1093/nar/gkab382

73. Singh S, Gupta P, Meena A, Luqman S (2020) Acacetin, a flavone with diverse therapeutic potential in cancer, inflammation, infections and other metabolic disorders. Food Chem Toxicol 145:111708 . https://doi.org/10.1016/j.fct.2020.111708

74. Abd Jalil MA, Kasmuri AR, Hadi H (2017) Stingless Bee Honey, the Natural Wound Healer: A Review. Skin Pharmacol Physiol 30:66–75 . https://doi.org/10.1159/000458416

75. Campos JF, dos Santos UP, Macorini LFB, de Melo AMMF, Balestieri JBP, Paredes-Gamero EJ, Cardoso CAL, de Picoli Souza K, Dos Santos EL (2014) Antimicrobial, antioxidant and cytotoxic activities of propolis from Melipona orbignyi (Hymenoptera, Apidae). Food Chem Toxicol 65:374–380 . https://doi.org/10.1016/j.fct.2014.01.008

76. Belina-Aldemita MD, Opper C, Schreiner M, D'Amico S (2019) Nutritional composition of pot-pollen produced by stingless bees (Tetragonula biroi Friese) from the Philippines. J Food Compos Anal 82:103215 . https://doi.org/10.1016/j.jfca.2019.04.003

77. Araújo MJAM, Bosco S de MG, Sforcin JM (2016) Pythium insidiosum: inhibitory effects of propolis and geopropolis on hyphal growth. Brazilian J Microbiol 47:863–869 . https://doi.org/10.1016/j.bjm.2016.06.008

78. Ragasa CY, Galian RAF, Ebajo VD, Aguda RM, Cervancia CR, Shen C-C (2015) Propolins and glyasperin A from stingless bee nests. Rev Bras Farmacogn 25:177–179 . https://doi.org/10.1016/j.bjp.2015.03.006

79. Bae J, Kim N, Shin Y, Kim S-Y, Kim Y-J (2020) Activity of catechins and their applications. Biomed Dermatology 4:8 . https://doi.org/10.1186/s41702-020-0057-8

80. Panche AN, Diwan AD, Chandra SR (2016) Flavonoids: an overview. J Nutr Sci 5:e47 . https://doi.org/10.1017/jns.2016.41

81. Pereira FAN, Barboza JR, Vasconcelos CC, Lopes AJO, Ribeiro MN de S (2021) Use of Stingless Bee Propolis and Geopropolis against Cancer—A Literature Review of Preclinical Studies. Pharmaceuticals 14:1161 . https://doi.org/10.3390/ph14111161

82. Muthukumaran J, Srinivasan S, Venkatesan RS, Ramachandran V, Muruganathan U (2013) Syringic acid, a novel natural phenolic acid, normalizes hyperglycemia with special reference to glycoprotein components in experimental diabetic rats. J Acute Dis 2:304–309 . https://doi.org/10.1016/S2221-6189(13)60149-3

83. Montes A, Merino R, De los Santos DM, Pereyra C, Martínez de la Ossa EJ (2017) Micronization of vanillin by rapid expansion of supercritical solutions process. J CO2 Util 21:169–176 . https://doi.org/10.1016/j.jcou.2017.07.009

84. Priefert H, Rabenhorst J, Steinbüchel A (2001) Biotechnological production of vanillin. Appl Microbiol Biotechnol 56:296–314 . https://doi.org/10.1007/s002530100687

85. Srinivasan K, Platel K, Rao MVL (2008) Hypotriglyceridemic effect of dietary vanillin in experimental rats. Eur Food Res Technol 228:103–108 . https://doi.org/10.1007/s00217-008-0911-1

86. Arabi M, Ghaedi M, Ostovan A (2017) Synthesis and application of in-situ molecularly imprinted silica monolithic in pipette-tip solid-phase microextraction for the separation and determination of gallic acid in orange juice samples. J Chromatogr B 1048:102–110 . https://doi.org/10.1016/j.jchromb.2017.02.016

87. Roidoung S, Dolan KD, Siddiq M (2016) Gallic acid as a protective antioxidant against anthocyanin degradation and color loss in vitamin-C fortified cranberry juice. Food Chem 210:422–427 . https://doi.org/10.1016/j.foodchem.2016.04.133

88. Neelam, Khatkar A, Sharma KK (2020) Phenylpropanoids and its derivatives: biological activities and its role in food, pharmaceutical and cosmetic industries. Crit Rev Food Sci Nutr 60:2655–2675 . https://doi.org/10.1080/10408398.2019.1653822

89. Ferreira PS, Victorelli FD, Fonseca-Santos B, Chorilli M (2019) A Review of Analytical Methods for p-Coumaric Acid in Plant-Based Products, Beverages, and Biological Matrices. Crit Rev Anal Chem 49:21–31 . https://doi.org/10.1080/10408347.2018.1459173

90. Sharma SH, Rajamanickam V, Nagarajan S (2018) Antiproliferative effect of p-Coumaric acid targets UPR activation by downregulating Grp78 in colon cancer. Chem Biol Interact 291:16–28 . https://doi.org/10.1016/J.CBI.2018.06.001

91. Batista JS, Salatino A, Negri G, Jara CEP, Paiva KAR de, Santos WLA dos, Teófilo T da S, Félix NS, Silva FHA, Rodrigues VH V. (2021) Photoprotective activity of geopropolis produced by Melipona subnitida (*Apidae*, *Meliponinae*) in the semiarid of the Brazilian Northeast. Res Soc Dev 10:e1121021305 . https://doi.org/10.33448/rsd-v10i2.12305

92. Suzuki A, Nomura T, Jokura H, Kitamura N, Fujii A, Hase T (2021) Beneficial effects of oral supplementation with ferulic acid, a plant phenolic compound, on the human skin barrier in healthy men. Int J Vitam Nutr Res. https://doi.org/10.1024/0300-9831/A000699

93. Ruiz-Matute AI, Hernández-Hernández O, Rodríguez-Sánchez S, Sanz ML, Martínez-Castro I (2011) Derivatization of carbohydrates for GC and GC-MS analyses. J Chromatogr B Analyt Technol Biomed Life Sci 879:1226–40 . https://doi.org/10.1016/j.jchromb.2010.11.013

94. da S. Sant'ana R, de Carvalho CAL, Oda-Souza M, de A. Souza B, de S. Dias F (2020) Characterization of honey of stingless bees from the Brazilian semi-arid region. Food Chem 327:127041 . https://doi.org/10.1016/j.foodchem.2020.127041

95. Chahed A, Nesler A, Navazio L, Baldan B, Busato I, Ait Barka E, Pertot I, Puopolo G, Perazzolli M (2020) The Rare Sugar Tagatose Differentially Inhibits the Growth of Phytophthora infestans and Phytophthora cinnamomi by Interfering With Mitochondrial Processes. Front Microbiol 11:1–12 . https://doi.org/10.3389/fmicb.2020.00128

96. Bodemer AA (2018) Psoriasis. In: Integrative Medicine, Fourth Edi. Elsevier, pp 726-738.e2

97. Kwack M-H, Ahn J-S, Kim M-K, Kim J-C, Sung Y-K (2010) Preventable effect of L-threonate, an ascorbate metabolite, on androgen-driven balding via repression of dihydrotestosteroneinduced dickkopf-1 expression in human hair dermal papilla cells. BMB Rep 43:688–692 . https://doi.org/10.5483/BMBRep.2010.43.10.688

98. Emwas A-H, Roy R, McKay RT, Tenori L, Saccenti E, Gowda GAN, Raftery D, Alahmari F, Jaremko L, Jaremko M, Wishart DS (2019) NMR Spectroscopy for Metabolomics Research. Metabolites 9:123 . https://doi.org/10.3390/metabo9070123

99. Dayrit FM, Dios AC de (2017) 1H and 13C NMR for the Profiling of Natural Product Extracts: Theory and Applications. In: Spectroscopic Analyses - Developments and Applications. InTech

100. Wishart DS, Cheng LL, Copié V, Edison AS, Eghbalnia HR, Hoch JC, Gouveia GJ, Pathmasiri W, Powers R, Schock TB, Sumner LW, Uchimiya M (2022) NMR and Metabolomics—A Roadmap for the Future. Metabolites 12:678 .

https://doi.org/10.3390/metabo12080678

101. Markley JL, Brüschweiler R, Edison AS, Eghbalnia HR, Powers R, Raftery D, Wishart DS (2017) The future of NMR-based metabolomics. Curr Opin Biotechnol 43:34–40 . https://doi.org/10.1016/j.copbio.2016.08.001

102. Schripsema J (2019) Similarity and differential NMR spectroscopy in metabolomics: application to the analysis of vegetable oils with 1H and 13C NMR. Metabolomics 15:39 . https://doi.org/10.1007/s11306-019-1502-9

103. Razali M, Zainal Z, Maulidiani M, Shaari K, Zamri Z, Mohd Idrus M, Khatib A, Abas F, Ling Y, Rui L, Ismail I (2018) Classification of Raw Stingless Bee Honeys by Bee Species Origins Using the NMR- and LC-MS-Based Metabolomics Approach. Molecules 23:2160 . https://doi.org/10.3390/molecules23092160

104. Wishart DS, Guo A, Oler E, Wang F, Anjum A, Peters H, Dizon R, Sayeeda Z, Tian S, Lee BL, Berjanskii M, Mah R, Yamamoto M, Jovel J, Torres-Calzada C, Hiebert-Giesbrecht M, Lui VW, Varshavi D, Varshavi D, Allen D, Arndt D, Khetarpal N, Sivakumaran A, Harford K, Sanford S, Yee K, Cao X, Budinski Z, Liigand J, Zhang L, Zheng J, Mandal R, Karu N, Dambrova M, Schiöth HB, Greiner R, Gautam V (2022) HMDB 5.0: the Human Metabolome Database for 2022. Nucleic Acids Res 50:D622–D631 . https://doi.org/10.1093/nar/gkab1062

105. Naik RR, Shakya AK, Oriquat GA, Katekhaye S, Paradkar A, Fearnley H, Fearnley J (2021) Fatty Acid Analysis, Chemical Constituents, Biological Activity and Pesticide Residues Screening in Jordanian Propolis. Molecules 26:5076 . https://doi.org/10.3390/molecules26165076

106. Turco JF, Mokochinski JB, Torres YR (2023) Lipidomic analysis of geopropolis of Brazilian stingless bees by LC-HRMS. Food Res Int 167:112640 .

https://doi.org/10.1016/j.foodres.2023.112640

107. BALCI M (2005) Spin-Spin Splitting in 1H-NMR Spectra. In: Basic 1H- and 13C-NMR Spectroscopy. Elsevier, pp 87–133

108. Sandoval CM, Ayson M, Moss N, Lieu B, Jackson P, Gaucher SP, Horning T, Dahl RH, Denery JR, Abbott DA, Meadows AL (2014) Use of pantothenate as a metabolic switch increases the genetic stability of farnesene producing Saccharomyces cerevisiae. Metab Eng 25:215–226 . https://doi.org/10.1016/j.ymben.2014.07.006

109. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, Fung C, Nikolai L, Lewis M, Coutouly M-A, Forsythe I, Tang P, Shrivastava S, Jeroncic K, Stothard P, Amegbey G, Block D, Hau DD, Wagner J, Miniaci J, Clements M, Gebremedhin M, Guo N, Zhang Y, Duggan GE, Macinnis GD, Weljie AM, Dowlatabadi R, Bamforth F, Clive D, Greiner R, Li L, Marrie T, Sykes BD, Vogel HJ, Querengesser L (2007) HMDB: the Human Metabolome Database. Nucleic Acids Res 35:D521-6 . https://doi.org/10.1093/nar/gkl923

Annex I – Hierarchical clustering analysis for each pre-processing during PLS regression.