We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Metabolomics and Chemometrics as Tools for Chemo(bio)diversity Analysis - Maize Landraces and Propolis

Marcelo Maraschin et al.*
Plant Morphogenesis and Biochemistry Laboratory,
Federal University of Santa Catarina, Florianopolis, SC,
Brazil

1. Introduction

Developments in analytical techniques (GC-MS, LC-MS, ^{1}H-, ^{13}C-NMR, FT-MS, e.g.) are progressing rapidly and have been driven mostly by the requirements in the healthcare and food sectors. Simultaneous high-throughput measurements of several analytes at the level of the transcript (transcriptomics), proteins, (proteomics), and metabolites (metabolomics) are currently performed, producing a prodigious amount of data. Thus, the advent of *omic* studies has created an information explosion, resulting in a paradigm shift in the emphasis of analytical research of biological systems. The traditional approaches of biochemistry and molecular cell biology, where the cellular processes have been investigated individually and often independent of each other, are giving way to a wider approach of analyzing the cellular composition in its entirety, allowing achieving a *quasi*-complete metabolic picture.

The exponential growth of data, largely from genomics and genomic technologies, has changed the way biologists think about and handle data. In order to derive meaning from these large data sets, tools are required to analyze and identify patterns in the data, and allow data to be placed into a biological context. In this scenario, biologists have a continuous need for tools to manage and analyze the ever-increasing data supply. Optimal use of the data set, primarily of chemical nature, requires effective methods to analyze and manage them. It is obvious that all *omic* approaches will rely heavily upon bioinformatics for the storage, retrieval, and analysis of large data sets. Thus, and taking into account the multivariate nature of analysis in *omic* technologies, there is an increase emphasis in research on the application of chemometric techniques for extracting relevant information.

* Shirley Kuhnen[1], Priscilla M. M. Lemos[1], Simone Kobe de Oliveira[1], Diego A. da Silva[1],
Maíra M. Tomazzoli[1], Ana Carolina V. Souza[1], Rúbia Mara Pinto[2], Virgílio G. Uarrota[1], Ivanir Cella[2],
Antônio G. Ferreira[3], Amélia R. S. Zeggio[1], Maria B.R. Veleirinho[4], Ivone Delgadillo[4] and Flavia A. Vieira[4]
[1] *Plant Morphogenesis and Biochemistry Laboratory, Federal University of Santa Catarina, Florianopolis, SC, Brazil;*
[2] *EPAGRI – Florianópolis, SC, Brazil;*
[3] *NMR Laboratory, Federal University of São Carlos, São Carlos-SP;*
[4] *Chemistry Department, University of Aveiro – Campus Santiago, Aveiro - Portugal*

Metabolomics[*] and chemometrics[†] have been used in a number of areas to provide biological information beyond the simple identification of cell constituents. These areas include:

a. Fingerprinting of species, genotypes or ecotypes for taxonomic or biochemical (gene discovery) purposes;
b. Monitoring the behavior of specific classes of metabolites in relation to applied exogenous chemical and/or physical stimuli;
c. Studying developmental processes such as establishment of symbiotic associations or fruit ripening;
d. Comparing and contrasting the metabolite content of mutant or transgenic plants with that of their wild-type counterparts.

In general sense, strategies to obtain biological information in the above mentioned areas have focused on the analysis of metabolic differences that evidence responses to a range of extrinsic (ambient) and intrinsic (genetic) stimuli. Since no single analytical method has been found to obtain a complete picture of the metabolome of an organism, an association of advanced analytical techniques (GC-MS, LC-MS, FTIR, ^{1}H-, ^{13}C-NMR, FT-MS, e.g.) coupled to chemometrics, e.g., univariate (ANOVA, correlation analysis, regression analysis) or multivariate (PCA, HA, PLS) statistical techniques, has been performed in order to rapidly identify up- or down-regulated endogenous metabolites in complex matrices such as plant extracts, flours, starches, and biofluids, for instance. Plant extracts are recognized to be a complex matrix containing a wide range of primary and secondary metabolites that vary according to the environmental condition, genotype, developmental stage, and agronomic traits, for example. Such a complex matrix has long been used to characterize plant genotypes growing in a given geographic region and/or subjected to external stimuli, giving rise to additional information of interest, e.g., plant genetic breeding programs, local biodiversity conservation, food industry, and quality control in drug development/production processes.

In the former case, programs for genetic breeding of plants have often focused on the analysis of landraces[‡] genotypes (i.e., creole and local varieties), aiming at to identify individuals well adapted to specific local environmental conditions (soil and climate) and with superior agronomic performance and biomass yield. Indeed, the analysis and exploitation of the local genotypes' diversity has long been used as a strategy to improve agronomic traits by conventional breeding methods in plant crops of economical interest, as well as for stimulating the preservation of plant genetic resources. Taking into consideration that a series of primary (e.g., proteins and starch) and secondary metabolites (alkaloids, phenolic acids, and carotenoids, for instance) are well recognized compounds associated to the plants' adaptation mechanisms to their surroundings ecological factors, metabolomics and chemometrics have emerged as an interesting approach for helping the selection of

---

[*] *Metabolomics*: constitutes a quantitative and qualitative survey of the whole metabolites of an organism as well as a tissue, thus it reflects the genome and proteome of a sample.

[†] *Chemometrics*: according to the definition of the Chemometrics Society, it is the chemical discipline that uses mathematical and statistical methods to design or select optimal procedures and experiments, and to provide maximum chemical information by analyzing chemical data.

[‡] Landraces are genotypes with a high capacity to tolerate biotic and abiotic stress, resulting in high yield stability and an intermediate yield level under a low input agricultural system.

superior genotypes, as further described in the first part of this chapter for maize landraces developed and cultured in southern regions of Brazil.

In a second part of this chapter is described the adoption of a typical metabolomic platform, i.e., FTIR and UV-visible spectroscopies coupled to chemometrics, for discriminating propolis samples produced in southern Brazil, a region of huge plant biodiversity. Propolis is typically a complex matrix and has been recognized for its broad pharmacological activities (anti-inflammatory, antibacterial, antifungal, anticancer, and antioxidant, e.g.) since ancient times. Propolis (registration number chemical abstracts service - CAS 9009-62-5) is a beekeeping resinous and complex product, with a variable physical appearance, collected and transformed by honey bees, *Apis mellifera*, from the vegetation they visit. It may be ochre, red, brown, light brown or green; some are friable and steady, while the others are gummy and elastic.

Phenolics such as flavonoids and phenol-carboxylic acids are strategic components in propolis to render it bioactive against several pathogenic microorganisms, for instance as bacteriostatic and/or bactericidal agents. The flora (buds, twigs, bark, and less importantly, flowers) surrounding the hive is the basic source for the phenolics stuff and thus exerts an outstanding importance on the propolis final composition and on its physical, chemical, and biological properties. Although the wax component is an unquestionable supplement provided by the bee secretory apparatus by far less is known about the degree of intensity that these laborious insects play changing all the other chemical constituents collected in the Nature including minor ingredients like essential oils (10%), most of them responsible for the delicate and pleasant odor. All this flora contribution to propolis and the exact wax content may then explain physical properties such as color, taste, texture, melting point, and more importantly, from the health standpoint, a lot of pharmaceutical applications. However, for purpose of industrial applications, the propolis hydroalcoholic extract needs to meet specific composition in order to guarantee any claimed pharmacological activity. One common method used by the industry for quality control is analyzing the propolis sample for the presence of chemical markers known to be present in the specific propolis product they market. Even though this has been the acceptable method for quality control, the presence of the chemical markers do not always guarantee an individual is getting the actual propolis stated by the product label, especially if the product has been spiked with the chemical markers. The quantitation method for the chemical markers will confirm the compounds presence, but it may not confirm the presence of the propolis known to contain the chemical markers. Authentication of the propolis material may be possible by a chemical fingerprint of it and, if possible, of its botanical sources. Thus, chemical fingerprinting, i.e., metabolomics and chemometrics, is an additional method that has been claimed to be included in the quality control process in order to confirm or deny the propolis sample quality being used for manufacturing of a derived product of that resinous and complex matrix. The second part of this chapter aims to demonstrate the possibility of a FTIR and UV-vis metabolomic-chemometrics approach to identify and classify propolis samples originating from nineteen geographic regions (Santa Catarina State, southern Brazil) in different classes, on the basis of the concerted variation in metabolite levels detected by those spectroscopic techniques. Exploratory data analysis and patterns of chemical composition based on, for instance, principal component analysis, as well as discriminating models will be described in order to unravel propolis chemotypes produced in southern Brazil.

2. Maize: metabolomic and chemometric analyses for the study of landraces

Maize (*Zea mays* L.) was chosen as a model for metabolomic analysis because although most of this cereal produced worldwide is used for animal feeding, an important amount is also used in human diet and for industrial purposes, providing raw material for food, pharmaceuticals, and cosmetics production. The maize grain is composed of several chemicals of commercial value and the diversity of its applications depends on the differences in relative chemical composition, e.g. protein, oil, and starch contents, traits that show prominent genetic components (Baye et al., 2006; White, 2001). Over the last centuries, farmers have created thousands of maize varieties suitable for cultivation in numerous environments. Accordingly, it seems consensual that the maize landraces' phenotypes, e.g., morphological and agronomic traits and special chemical characteristics of grains are resultant of the domestication process. Thus, high throughput metabolomic analysis of maize genotypes could improve metabolic singularities knowledge about landraces, helping their characterization and evaluation, and indicating new alternatives for their use. In this context, to distinguish metabolic profiles it is necessary to consider the use of diverse analytical tools, such as spectroscopic and chromatographic techniques for instance. Techniques that are reproducible, stable with time, and do not require complex sample preparation such as infrared vibrational spectroscopy and nuclear magnetic resonance spectroscopy are desirable for metabolic profiling.

2.1 Metabolic profiling of maize landraces through FTIR-PCA – integral and degermed flours

Vibrational spectroscopy, and particularly Fourier transform infrared spectroscopy (FTIR) is thought to be interesting as one aims at discriminating and classifying maize landraces according to their chemical traits. FTIR is a physicochemical method that measures the vibrations of bonds within functional groups and generates a spectrum that can be regarded as a metabolic fingerprint. It is a flexible method that can quickly provide qualitative and quantitative information with minimal or no sample preparation of complex biological matrices (Ferreira et al., 2001). By other hand, a FTIR spectrum is complex, containing many variables per sample and making visual analysis very difficult. Hence, to extract useful information from the whole spectra, multivariate data analysis is needed, particularly through the determination of the principal components (PCA - Fukusaki & Kobayashi, 2005). Such a multivariate analysis technique could allow the characterization of the sample relationships (scores plans or axis) and the recovery of their subspectral profiles (loadings). This approach was applied to classify flour samples from whole (integral) and degermed maize grains of twenty-six landraces developed and cultivated by small farmers in the far-west region of Santa Catarina State, southern Brazil (Anchieta County – 26º31′11′′S, 53º20′26′′W).

Previously to multivariate analysis, FTIR spectra were normalized, baseline-corrected in the region of interest by drawing a straight line before resolution enhancement (k factor of 1.7) was applied using Fourier self deconvolution (Opus v. 5.0, Bruker Biospin, GmbH, Rheinstetten, Germany). Chemometric analysis used normalized, baseline-corrected (3000–600 cm^{-1}. 1700 data points) and deconvoluted spectra, which were transferred via a JCAMP format (OPUS v. 5.0, Bruker Biospin GmbH, Rheinstetten, Germany) into the data analysis software for PCA (The Unscramble v. 9.1, CAMO Software Inc., Woodbridge, USA).

Metabolomics and Chemometrics as Tools
for Chemo(bio)diversity Analysis - Maize Landraces and Propolis
257

Previously to PCA analysis each spectrum within the (3000–600 cm⁻¹) region was standard normal deviates corrected.

Figure 1 shows a PCA scores scatter plot for flour samples from whole and degermed grains using the whole FTIR spectral window data set (3000–600 cm⁻¹). The scores scatter plot (PC1 vs. PC2) that contains 93% of the data set variability shows a clear discrimination among flour samples of whole and degermed grains.
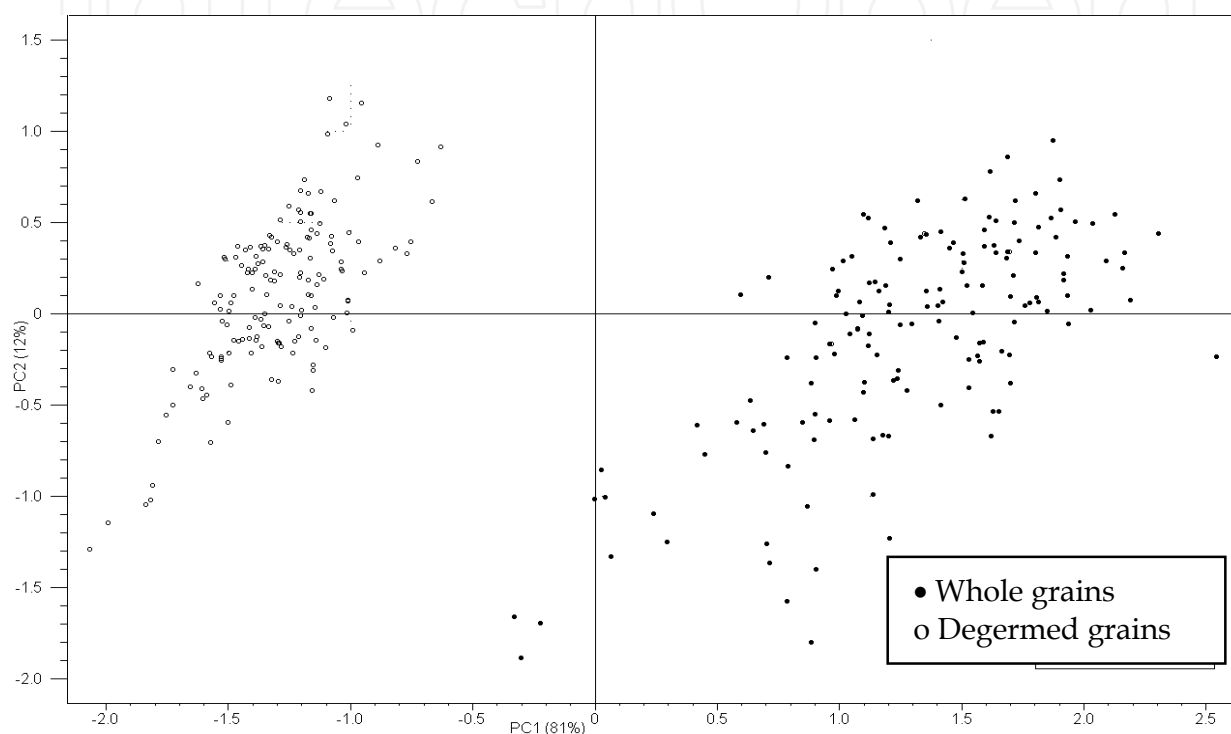
Fig. 1. Principal component analysis scores scatter plot of the FTIR data set in the spectral window of 3000–600 cm⁻¹ wavenumber of landrace maize flours of whole and degermed grain cultivated in the southern Brazil.

The samples of whole grains grouped in PC1+ axis seemed to be more discrepant in their chemical composition, appearing more scattered through the quadrants of the PCA representation. Figure 2 shows the loadings plot of PC1, revealing the most important wavenumbers which explain the distinction of the samples previously found (scores scatter plot). The loadings indicated a prominent effect of the lipid components (2924, 2850, and 1743 cm⁻¹) for the segregation observed. The two major structures of the grains are the endosperm and the germ (embryo) that constitute approximately 80 and 10% of the mature kernel dry weight, respectively. The endosperm is largely starch (approaching 90%) and the germ contains high levels of oil (30%) and protein (18% - Boyer & Hannah, 2001).

The greatest chemical diversity observed in whole grains can be explained by genetic variation of embryos resulting from sexual reproduction. Some authors suggest that the high level of genetic and epigenetic diversity observed in maize could be responsible for its great adaptation capacity to a wide range of ecological factors. Lemos (2010) analyzing the

metabolic profile of maize landraces' leaf tissues from Anchieta County (southern Brazil) found a prominent chemical variability among individuals of same variety, although inter-variety variability has also been observed.
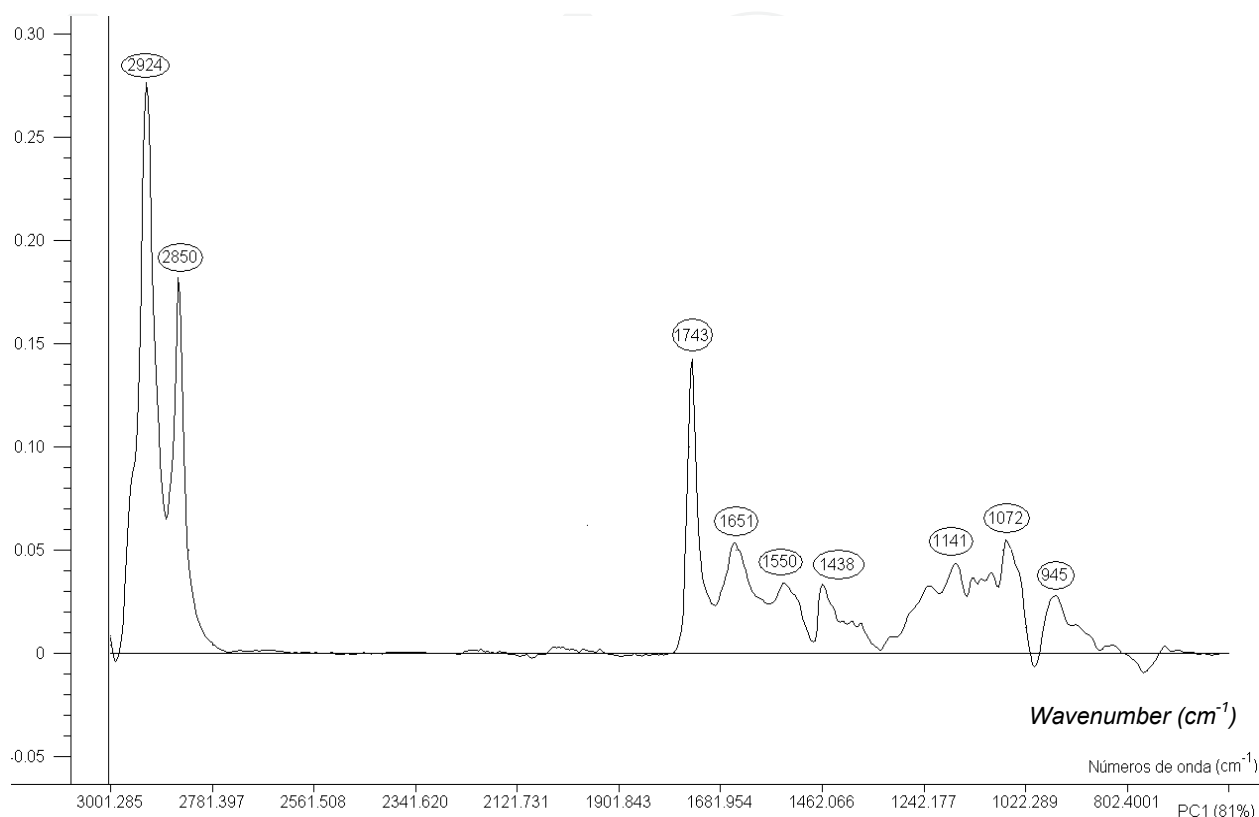

Fig. 2. PC1 loadings plot of the FTIR spectra of maize flours of whole and degermed grains in the 3000–600 cm^{-1} wavenumber region.

2.2 Starch recognition pattern of maize landraces by NMR spectroscopy and PCA

The composition of maize grains can be heterogeneous for both the quantity and quality of compounds from endosperm as starch, protein, and oil. In this context, metabolomics coupled to chemometrics approach was successfully applied to the discrimination of starches from the studied twenty-six maize landraces. The starches were extracted from flours with distilled water (1: 70, w/v) under reflux (80°C, 1 h), precipitated with ethyl alcohol (12 h, 4°C), and oven-dried (55°C until constant weight). Samples (50 mg) were dissolved in DMSO-δ_6 (0.6 mL) and ^{1}H-NMR spectra obtained under standard conditions. Sodium-3-trimethylsilylpropionate (TMSP-2, 2, 3, 3-d$_4$) was used as internal reference (δ_{ppm} 0.0). Spectra were processed using 32768 data points by applying an exponential line broadening of 0.3 Hz for sensitivity enhancement before Fourier transformation and were accurately phased, baseline adjusted, and converted into JCAMP format to build the data matrix. All calculations were carried out using the Pirouette software (v. 3.11, InfoMetrix, Woodinville, Washington, USA). The PCA analysis of the whole ^{1}H-NMR data set (32.000

Metabolomics and Chemometrics as Tools
for Chemo(bio)diversity Analysis - Maize Landraces and Propolis
259

data points) was performed including the spectra of amylose and amylopectin standards. The chemical structures and the purity of the standards of amylose and amylopectin were confirmed by ^{13}C-NMR spectroscopy. The PC1 (32%) vs. PC2 (28%) scores scatter plot allowed a clear segregation of the amylopectin standard and a discrimination for the maize flours samples in two groups (Fig. 3) by PC1 axis.
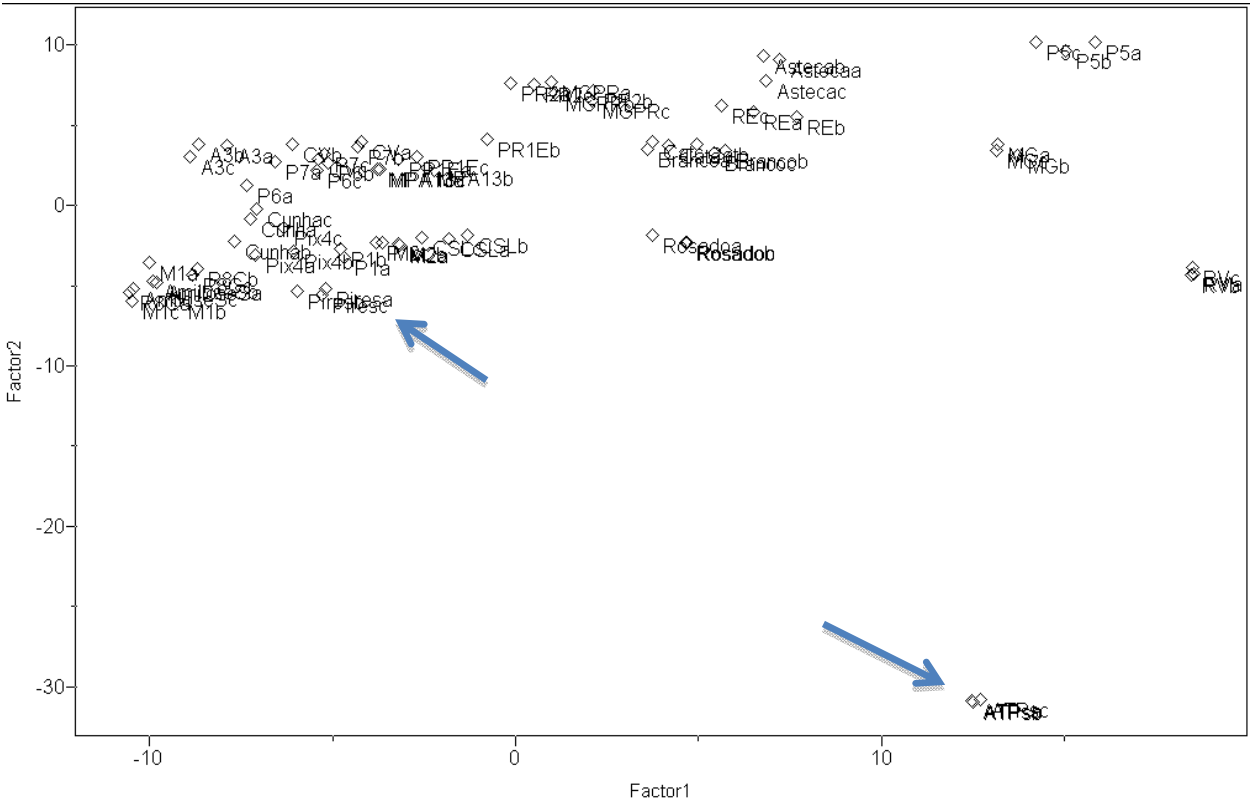

Fig. 3. Classification plot of starch fractions of maize landraces for ^{1}H-NMR after PCA analysis. The arrows shows the amylose and amylopectin stardands.

The Roxo 41 variety located closer to the amylopectin standard suggesting the predominance of that polysaccharide in relation to amylose in its starchy fraction. This result is in accordance to PCA analysis of the IR data set of the fingerprint region of carbohydrates that diagnosed the starch granules from Roxo 41 with superior amylopectin amount in respect to their amylose content (data not shown). By other hand, the amylose standard was grouped with MPA1 and Rajado 8 Carreiras varieties, suggesting that the starch granules contain superior amount of that polysaccharide in its starchy fraction.

3. Propolis: ATR-FTIR and UV-visible spectrophotometry coupled to chemometrics as an analytical approach for pattern recognition

3.1 Attenuated total reflectance-Fourier transform infrared spectroscopy

Propolis (registration number chemical abstracts service - CAS 9009-62-5) is a sticky colored material, which honeybees collect from different plants exudates and modify in its hypopharyngeal glands, being used in the hive to fill gaps and to protect against invaders as insects and microorganisms. Raw propolis usually contains 50% resin and balsam, 30% wax,

10% aromatic oils, 5% pollen, and 5% other substances as inorganic salts and amino acids. This resin has been used by humanity since ancient civilizations like Egyptian, Assyrian, Greek, Roman, and Inca. In these days, a number of studies have confirmed important biological activities such as antibacterial, antifungal, antiviral, antioxidant, anti-inflammatory, hepatoprotective, and antitumoral (for review see Bankova, 2009; Banksota et al., 2001; Castaldo & Capasso, 2002).

The aspect, texture and the chemical composition of propolis is highly variable and depends on the climate, season, bee species and mainly the local flora which is visited by bees to collect resin (Markham et al., 1996). For this reason, comparing propolis samples from distinct regions might be the same as to compare extracts of two plants that belong to different taxonomical families (Bankova, 2005).

Propolis from Europe is the best known type of propolis. In European regions with temperate climate bees obtain resin mainly from the buds of *Poplus* species and the main bioactive components are flavonoids (Greenaway et al., 1990). In tropical countries, the botanical resources are much more variable in respect to temperate zones, so bees find much more possibilities of collecting resins and hence the chemical composition of tropical propolis are more variable and distinct from European ones (Sawaya et al., 2011). Different compounds have been reported in tropical propolis such as terpenoids and prenylated derivatives of *p*-coumaric acids in Brazilian propolis (Marcucci, 1995), lignans in Chilean samples (Valcic et al., 1998), and polyisoprenylated benzophenones in Venezuelan, Brazilian, and Cuban propolis (Cuesta-Rubio et al., 1999; Marcucci, 1995).

In order to be accepted officially into the main stream of the healthcare system and for industrial applications, propolis needs chemical standardization that guarantees its quality, safety, efficacy, and provenance. The chemical diversity mainly caused by the botanical origin makes the standardization difficult. Since the chemistry and biological activity of propolis depends on its geographical origin, a proper method to discriminate its origin is needed (Bankova, 2005).

Chromatographic methods (HPLC, TLC, GC, e.g.) are largely used to identification and quantification of propolis compounds, but it its becoming clear that to separate and evaluate all constituents of propolis is an almost impossible task (Sarbu & Mot, 2011). Even thought the presence of the chemical markers are considered an acceptable method for quality control, not always is guarantee about what is stated by the product label, especially if the product has been spiked with the chemical markers. Besides, literature has demonstrated that is not possible to ascribe the pharmacological activity solely to a unique compound and until now no single propolis component has shown to possess anti-bacterial activity higher than total extract (Kujumgiev et al., 1999; Popova et al., 2004). Thus, a possibility is offered by the fingerprinting methods that can analyze in a non-selective way the propolis samples as a whole.

Poplar propolis, for example, can be distinguished by UV-visible spectrophotometric determination of all three important components (flavones and flavonols, flavonones and dihydroflavonols, and total phenolics - Popova et al., 2004), but some constraints regarding such an analytical approach has been claimed for propolis from tropical regions (Bankova & Marcucci, 2000).

Metabolomics and Chemometrics as Tools
for Chemo(bio)diversity Analysis - Maize Landraces and Propolis
261

The search for faster screening methods capable of characterizing propolis samples of different geographic origins and composition has lead to the use of direct insertion mass spectrometric fingerprinting techniques (ESI-MS and EASI-MS), which has proven to be a fast and robust method for propolis characterization (Sawaya et al., 2011), although this analytical approach can only detect compounds that ionize under the experimental conditions. Similarly, Fourier transform infrared vibrational spectroscopy (FTIR) has also demonstrated to be valuable to chemically characterize complex matrices such as propolis (Wu et al., 2008).

In order to achieve the goal of treat propolis sample as a whole than just be focused only in marker compounds, chemometric methods are being considered an important tool to analyze the huge data sets generated by non-selective analytical techniques such as UV-vis, MS, NMR, and FT-IR, generating information not only about chemical composition of propolis but also discriminating its geographical origin.

Authentication of propolis material may be possible by a chemical fingerprint of it and, if possible, of its botanical sources. Thus, chemical fingerprinting, i.e., metabolomics and chemometrics, is an additional method that has been claimed to be included as a quality control method in order to confirm or deny the propolis sample being used for the manufacturing of a derived product of that resinous and complex matrix.

Over the last decades, infrared (IR) vibrational spectroscopy has been well established as a useful tool for structure elucidation and quality control in several industrial applications. Indeed, the development of Fourier transform (FT) IR and attenuated total reflectance (ATR) techniques have also evolved allowing rapid IR measurements of organosolvent extracts of plant tissues, edible oils, and essential oils, for example (Damm et al., 2005; Lai et al., 1994; Schulz & Baranska, 2007). In consequence of the strong dipole moment of water, IR spectroscopy applications have mostly focused on the analysis of dried or non-aqueous plant matrices and currently IR methods are widely used as a fast analytical technique for the authentication and detection of adulteration of vegetable oils.

ATR-FTIR spectroscopy was applied to propolis samples collected in the autumn-2010 and originated from nineteen geographic regions of Santa Catarina State (southern Brazil) in order to gain insights as to the chemical profile of those complex matrices. FTIR spectroscopy measures the vibrations of bonds within functional groups and generates a spectrum that can be regarded as a metabolic fingerprint. Similar IR spectral profiles (3000 – 600 cm^{-1}, figure 4) were found by a preliminary visual analysis for purpose of an exploratory overview of data, revealing typical signals of e.g., lipids (2910 – 2845 cm^{-1}), monoterpenes (1732, 1592, 1114, 1022, 972 cm^{-1}), sesquiterpenes (1472 cm^{-1}), and sucrose (1122 cm^{-1} - Schulz & Baranska, 2007) for all the studied samples. However, we were not able to identify by visual inspection of the spectra a clear picture regarding a discriminating effect of any primary or secondary metabolites among the propolis samples.

A FTIR spectrum is complex, containing many variables per sample and making visual analysis very difficult. Hence, to extract extra useful information, i.e., latent variables, from the whole spectra chemometric analysis was performed considering the whole FTIR data set using principal components analysis (PCA) for an exploratory overview of data. This method could reveal similarity/dissimilarity patterns among propolis samples, simplifying
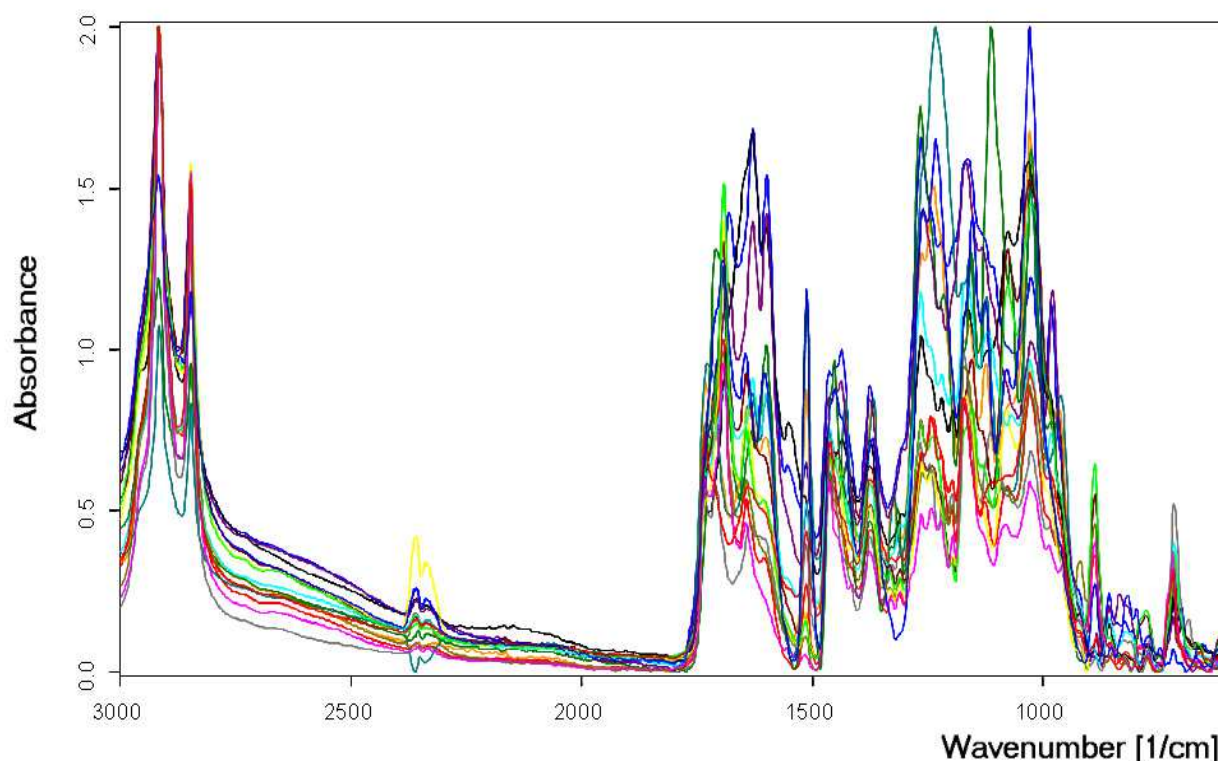
Fig. 4. IR spectral profile of propolis samples (autumn, 2010) produced in southern Brazil, according to the geographic regions of origin in Santa Catarina State. IR spectra are shown from top to bottom following the geographic precedence, i.e. 19 counties, of the propolis samples: Angelina *(ANG)*, Balneário Gaivotas *(BG)*, Bom Retiro *(BR₁ and BR₂)*, Caçador *(Cç)*, Campo-Erê *(CE)*, Canoinhas *(CA)*, Campos Novos *(CN)*, Descanso *(DS)*, José Boiteux *(JB)*, Porto União *(PU)*, Serra Alta *(SA)*, São Joaquim *(SJ₁ and SJ₂)*, São José do Cerrito *(SJC)*, Urupema *(URU)*, Vidal Ramos *(VR)*, Florianópolis *(FLN)*, and Xaxim *(XX)*.

data dimensions and results interpretation, without missing the more relevant information associated to them (Fukusaki & Kobayashi, 2005; Leardi, 2003). The covariance was choose for matrix construction in PCA calculation, since all variables considered were expressed in the same unit. By doing so, the magnitude differences were maintained, i.e., data were not standardized and variables contribution to samples distribution along axes was direct proportional to their magnitude. For the purpose of the propolis chemical profile analysis this kind of information is thought to be very useful, because wavenumber with higher absorbances (higher metabolites concentration) contribute more significantly with objects distribution into PCA, introducing quantitative information beside the compositional information of the sample data.

The principal component analysis (PCA) of the whole spectral data (3000 – 600 cm^{-1}, 1700 data points) revealed that PC1 and PC2 defined 88% of the variability from the original IR data and a peculiar pattern of lipids (2914 cm^{-1} and 2848 cm^{-1} - C-H stretching vibrations) for the samples from the northern region (*NR*) of Santa Catarina State. The climate in that region is typically mesothermic, humid subtropical with a mild summer and an annual temperature average of 17.2ºC – 19.4ºC. On the other hand, the propolis produced in the

highlands (1360m altitude, annual maximum and minimum temperatures average of 18.9°C and 9.3°C, respectively) were discrepant regarding their monoterpene (1114 cm^{-1} and 972 cm^{-1} - ϖ-CH2) and sesquiterpene (1472 cm^{-1} - δ CH$_2$) compounds (Schulz & Baranska, 2007) – Figure 5. In despite of *NR$_1$ and NR$_2$* propolis samples have grouped in PC1- and PC2+, they differ somewhat in respect to their chemical composition, an effect attributed to the flora composition found in those regions, e.g., mostly Atlantic Rainforest in *NR$_1$* as *NR$_2$* shows extensive areas covered by artificial reforestations i.e., *Eucalyptus* spp and *Pinus* spp, furnishing distinct raw materials for propolis production.
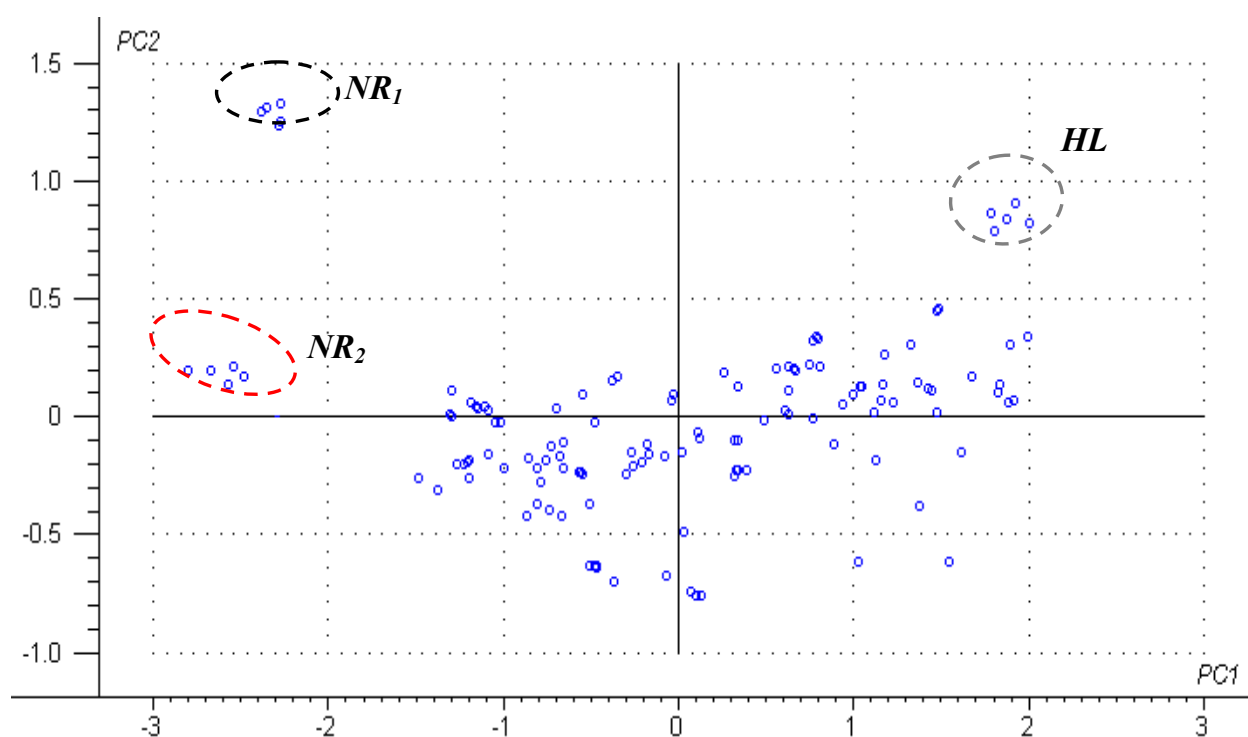

Fig. 5. Principal component analysis scores scatter plot of the FTIR data set in the spectral window of 3000–600 cm^{-1} wavenumber (1700 data points) of propolis samples produced in the southern Brazil (Santa Catarina State). *NR$_1$*, *NR$_2$* and *HL* refer to propolis samples originated from northern and highland regions of Santa Catarina State. The calculations were carried out using The Unscrambler software (v. 9.1, Oslo - Norway). PC1 and PC2 accounts for 88% of the variance preserved.

Further chemometric analysis took into consideration the fact that propolis is a very well known source of phenolic compounds, e.g., phenolics acids and flavonoids. Indeed, phenolic compounds occur ubiquitously in most plant species and take part of the chemical constitution of propolis worldwide. IR spectroscopy allows to identify phenolic compounds since they demonstrate strong IR bands due to C-H wagging vibrations between 1260 – 1180 cm^{-1} and 900 – 820 cm^{-1} (Schulz & Baranska, 2007). The principal

component calculations were performed for both 1260 – 1180 cm^{-1} and 900 – 820 cm^{-1} spectral windows and PC1 and PC2 resolved about 96% of the spectral data variability. An interesting discrimination profile was detected where samples from the far-west (*FW*) region grouped distinctly in respect to northern (*NR₁*) ones, which also differed from the highlands (*HL*) propolis samples. Such findings can be explained in any extension based on the flora composition of the studied geographic regions. In the northern and far-west regions of Santa Catarina State the Atlantic Rainforest is typically found, but the floristic composition varies according to the altitude, e.g., 240 m altitude – *NR₁* and 830 m – *FW*. Besides, as a mesothermic humid subtropical climate is found in *NR₁*, the *FW* region is characterized by a temperate climate that determines a discrepant composition of plant species. Finally, the *HL* region (1360m altitude, temperate climate) is covered by the Araucaria Forest, where parana pine (*Araucaria angustifolia, Gymnospermae, Araucariaceae*) occurs as a dominant plant species. *A. angustifolia* produces a resinous exudate rich in guaiacyl type lignans, fatty acids, sterols (Anderegg & Rowe, 2009), phenolics, and terpenic acids that is thought to be used by honey bee (*Apis mellifera*) for the propolis production. Since the plant species populations influence the propolis chemical composition, the discrimination profile detected by ATR-FTIR coupled to chemometrics seems to be an interesting analytical approach to gain insights as to the effect of the climatic factors and floristic composition on the chemical traits of that complex matrix.

3.2 Ultraviolet-visible scanning spectrophotometry

Combination of UV-visible spectrophotometric wavelength scans and chemometric (PCA) analysis seems to be a simple and fast way to prospect plant extracts. This analytical strategy revealed to be fruitful for discrimination of *habanero* peppers according to their content of capsaicinoids, substances responsible for the pungency of their fruits (Davis et al., 2007).

Chemometric analysis was performed considering the absorbance values of the total UV-visible data set (200 ηm to 700 ηm, 450 data points) for the propolis samples in study, by using principal components analysis (PCA) for an exploratory overview of data.

In a first approach, principal components analysis (PCA) was tested by both correlation and covariance matrices of calculations. If correlation is used, the data set is standardized (mean-centered and columns scaled to the unit of variance), decreasing the effect of differences in magnitude between variables and leading to a distribution of objects (*eigenvalues*) with equal influence from all variables. On the other hand, if covariance is used, data is only mean-centered; retaining its original scale. The resulting distribution is then determined either by composition and magnitude of variables, leading to a PCA representation more influenced by larger observed values (Manetti et al., 2004). A similar distribution of objects was found for both correlation and covariance matrices in PC calculations, as PC1 and PC2 resolved 91.2% and 96.3%, respectively of the variability of the spectrophotometric data set. Thus, the covariance matrix was chosen for PCA calculations, since all variables considered were expressed in the same unit. By doing so, the magnitude differences were maintained, i.e., data were not standardized and variables contribution to samples distribution along axes was direct proportional to their magnitude. For the purpose of the chemical profile analysis of the propolis samples this kind of information is thought to be very useful, because

Metabolomics and Chemometrics as Tools
for Chemo(bio)diversity Analysis - Maize Landraces and Propolis
265

wavelengths with higher absorbances (higher metabolites concentration) contribute more significantly with objects distribution into PCA, introducing quantitative information beside the compositional information of the sample data.

Principal component analysis was performed using The Unscrambler software (v. 9.1, Oslo - Norway) and revealed mostly a distribution of the propolis samples along the PC1 axis (91% sample total variability), as PC2 (5% sample total variability) seemed to be lesser discriminator of the objects. A clear separation of the samples according to the east-west axis of Santa Catarina State could be found, where propolis produced near coastal regions (CR_1 and CR_2) grouped in PC1+/PC2-, as the sample from the far-west region *(FW)* was detected in the opposite side of PC1 axis, along with the samples from the northern region (*PU, Cç,* and *CA* - Figure 6). Interestingly, propolis samples from the counties *SJ, URU, BR* (highlands counties), and *ANG*, which shown geographic proximities and a certain common floral composition, seemed to be similar in their chemical profiles as determined by UV-visible scanning spectrophotometry, grouping in PC1+/PC2+.
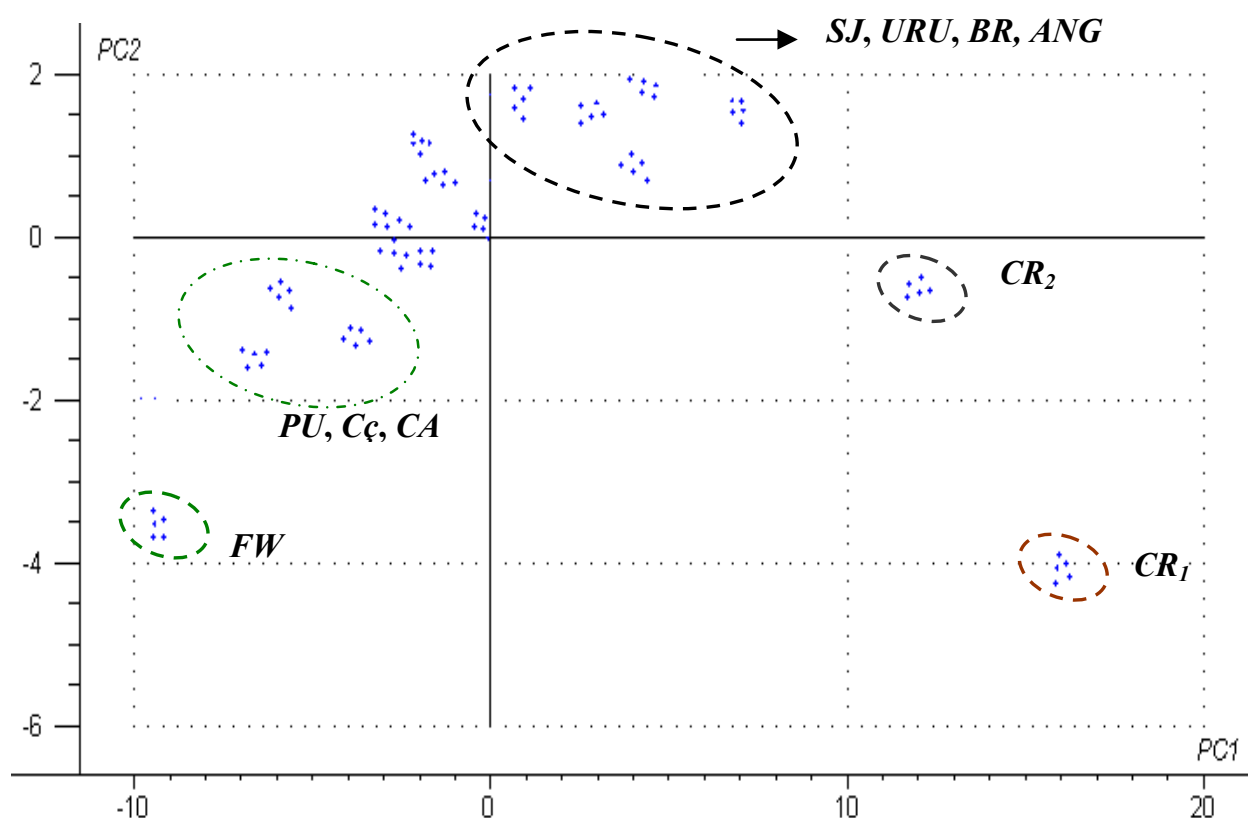

Fig. 6. Principal component analysis scores scatter plot of the UV-visible data set in the spectral window of 200 ηm to 700 ηm (450 data points) of propolis samples produced in the southern Brazil (Santa Catarina State). *CR₁, CR₂,* and *FW* refer to propolis samples originated from coastal (*BG* and *FLN* Counties) and far-west (*CE* County) regions, respectively, of Santa Catarina State. The sample grouping of propolis with similar UV-visivel scanning profiles regarding their (poly)phenolic composition is detached in the PC1+/PC2+ quadrant. PC1 and PC2 resolved 96% of the total variability of the spectral data set.

High loadings associated to the wavelengths 394 ηm, 360 ηm, 440 ηm, and 310 ηm seemed to influence the observed distribution of the propolis samples and could be associated to the presence of (poly)phenolic compounds. In fact, the λ_{max} for the cinnamic acid and its derivatives is near 310-320 ηm as for the flavonols is usually around 360 ηm (Tsao & Deng, 2004). Further chemical analysis of the total content of phenolics ad flavonoids in the propolis originated from the counties *SJ*, *URU*, *BR*, and *ANG* revealed similar contents, with average concentrations of 1411.52 µg/ml and 4.61 µg/ml of those secondary metabolites, respectively, in the hydroalcoholic (70: 30, v/v) extract. Such findings differed (P<0.05 – *Tukey* test) in respect to the concentrations detected for the propolis samples produced in the coastal (793.67 µg/ml – total phenolics and 2.82 µg/ml – flavonoids) and far-west (952.97 µg/mL – total phenolics and 0.59 µg/ml flavonoids) regions of Santa Catarina State, corroborating the PCA results herein shown.
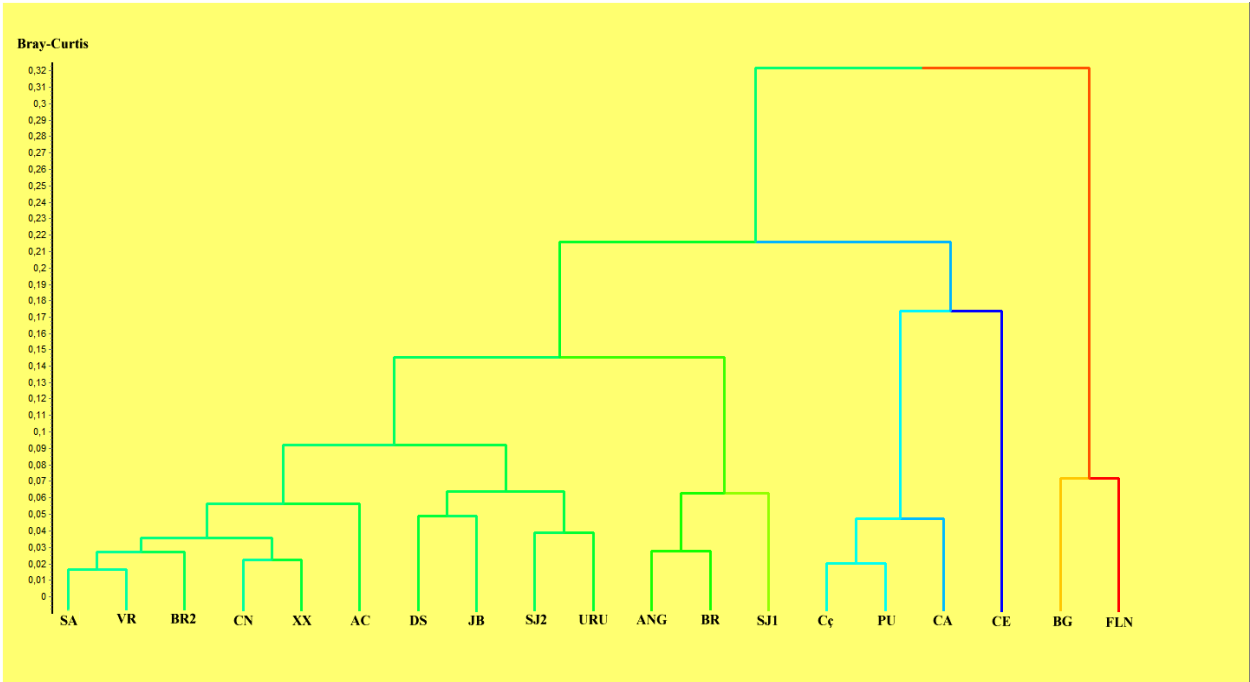

Fig. 7. Dendrogram of propolis samples using average linkage with Bray-Curtis dissimilarity measure. Data calculations were based on the absorbance values for the UV-visible spectral window of 200 ηm to 700 ηm of propolis samples produced in Santa Catarina State - southern Brazil, autumn-2010.

Metabolomics and Chemometrics as Tools
for Chemo(bio)diversity Analysis - Maize Landraces and Propolis
267

In order to check the chemical similarity pattern of propolis samples detected by PCA, further cluster analysis of the whole absorbance UV-vis data set, i.e., absorbance values of 200 ηm to 700 ηm (450 data points), was performed by using the Unweighted Pair Group Method with Arithmatic Mean (UPGMA) based on Bray-Curtis dissimilarity coefficient. UPGMA is a simple agglomerative or hierachical clustering method used in for the creation of phonetic trees, i.e., phenograms, hierarchical trees or dendrograms that indicate the similarity degree among samples/objects of interest, so that observations in the same cluster are similar in some sense. In UPGMA method after two objects with the least dissimilarity fuse together an arithmetic average of the dissimilarity of this new cluster and the rest of the objects is calculated. This leads to a reduction in the size of the original dissimilarity matrix. The procedure continues with the dissimilarity matrix being correspondingly reduced. When the average between an object and a cluster is calculated, the method gives equal weights to the members of the clusters when averaging, i.e., unweighted. Thus, in the progressive reduction of the dissimilarity matrix, only relationships between groups are considered, which are given equal weighting and this leads to loss of information about the relationships between pairs of objects (Legendre, 1998; Singh, 2008).

The hierarchical tree of the similarity of chemical profiles of the propolis samples is shown in figure 7. The findings suggest a resemblance of grouping as found by PCA calculations in respect to the *SJ*, *URU*, *BR*, and *ANG* samples, as well as for the propolis originated from the coastal (*BG* and *FLN*) and northern regions (*CA*, *PU*, and *Cç*). Additionally, UPGMA analysis also discriminate the propolis produced in the western (*AC, XX*, and *CN*) and far-west regions.

4. Conclusions

The chemo(bio)diversity analysis of maize landraces and propolis produced in southern regions of Brazil was successfully assessed by using a typical metabolomic platform involving spectroscopic techniques (FTIR, ^{1}H- and ^{13}C-NMR, and UV-visible) and chemometrics. The huge amount of data afforded by those spectroscopic techniques was analyzed using multivariate statistical methods such as principal component analysis and cluster analysis allowing obtaining extra information on the metabolic profile of the complex matrices in study.

The analytical approach described showed to be suitable when ones aim to discriminate maize flour samples from whole and degermed maize, an issue thought to be important for the food, cosmetic, and pharmaceutical industries regarding the usage and quality control process of that raw material. Similarly, the classification of maize landraces according to their starch traits is considered technologically relevant in order to optimize the usage of non-chemically modified starches in industrial process, for instance.

The classification of Brazilian propolis as to their chemical profiles and geographic regions seems to be relevant because that biomass is typically quite complex, making difficult and expensive to perform a complete characterization in that sense. By doing so, the propolis produced in southern Brazil might be better evaluated as to their potential usage in cosmetic and pharmaceutical industry, taking into consideration their secondary metabolite

constituents, e.g., mono/sesquiterpenes and phenolics. The coupling of chemometrics-spectroscopic techniques used is thought to be essential to allow detecting peculiar chemical traits of the propolis samples according to their geographic regions in a simple and fast way.

5. Acknowledgment

6. References

Anderegg, RJ & Rowe, JW. (2009). Lignans, the major component of resin from Araucaria angustifolia knots. *International Journal of the Biology, Chemistry, Physics and Technology of Wood,* Vol. 28, pp.171–175. ISSN 0018-3830

Bankova, V. (2005). Chemical diversity of propolis and the problem of standardization. *Journal of Ethnopharmacology,* Vol. 100, pp.114-117. ISSN: 0378-8741

Bankova, V & Marcucci, MC. (2000). Standardization of propolis: present status and perspectives. *Bee World,* Vol. 81, pp.182-188. ISSN: 0005-772X

Banksota, AH., Tezuka, Y & Kadota, S. (2001). Recent progress in pharmacological research of propolis. *Phytotherapy Research,* Vol. 15, pp.561-571. ISSN: 1099-1573

Baye, TM., Pearson, TC & Settles, AM. (2006). Development of a calibration to predict maize seed composition using single kernel near infrared spectroscopy. *Journal of Cereal Science*, Vol. 43, pp.236–243. ISSN: 0733-5210

Boyer, CD & Hannah, C. (2001). Kernel mutants of corn. In: Specialty corns. HALLAUER, AR. (Ed.). 2nd ed. pp. 153, CRC Press, London.

Castaldo, S & Capasso, F. (2002). Propolis, an old remedy used in modern medicine. *Fitoterapia,* Vol. 73, pp.S1-S6. ISSN: 0367-326X

Cuesta-Rubio, O., Cuellar, AC., Rojas, N., Velez, HC., Rastrelli, L & Aquino, R. (1999). A polyisoprenylated benzophenone from Cuban propolis. *Journal of Natural Products,* Vol. 62, pp.1013-1015. ISSN: 0974-5211

Damm, U., Lampen, P., Heise, HM., Davies, AN & Mcintyre, PS. (2005). Spectral variable selection for partial least squares calibration applied to authentication and quantification of extra virgin olive oils using Fourier transform Raman spectroscopy. *Applied Spectroscopy*, Vol. 59, pp.1286-1294. ISSN: 0003-7028

Davis, CB., Markey, CE., Busch, MA & Busch, KW. (2007). Determination of capsaicinoids in habanero peppers by chemometric analysis of UV spectral data. *Journal of Agricultural and Food Chemistry*, Vol. 55, pp. 5925-5933. ISSN: 0021-8561

Ferreira, D., Barros, A., Coimbra, MA & Delgadillo, I. (2001). Use of FTIR spectroscopy to follow the effect of processing in cell wall polysaccharide extracts of a sun-dried pear. *Carbohydrate Polymers*, Vol. 45, pp.175–182. ISSN: 0144-8617

Fukusaki, E & Kobayashi, A. (2005). Plant metabolomics: potential for practical operation. *Journal of Bioscience and Bioengineering*, Vol. 100, pp.347–354. ISSN: 1389-1723

Greenaway, W., Scaysbrook, T & Whately, FR. (1990). The composition and plant origin of propolis: a report of work at Oxford. *Bee World,* Vol. 71, pp. 107-118. ISSN: 0005-772X

Kujumgiev, A., Tsvetkova, I., Serkedjieva, YU., Bankova, V., Christov, R & Popov, S. (1999). Antibacterial, antifungal and antiviral activity of propolis of different geographical origins. *Journal of Ethnopharmacology*, Vol. 64, pp. 235-240. ISSN: 0378-8741

Lai, YW., Kemsley, EK & Wilson, J. (1994). Potential of Fourier transform infrared spectroscopy for the authentication of vegetable oils. *Journal of Agricultural and Food Chemistry*, Vol. 42, pp.1154-1159. ISSN: 0021-8561

Leardi, R. (2003). Chemometrics in data analysis. In: A user-friendly guide to multivariate calibration and classification. Naes, T., Isaksson, T., Fearn, T & Davies, T (eds). NIR Publications, West Sussex.

Legendre, P. (1998). *Numerical Ecology*. Elsevier Science, New York..

Lemos PMM (2010). Análise do metaboloma foliar parcial de variedades locais de milho (*Zea mays* L.) e dos efeitos anti-tumoral *in vitro* e na morfogênese embrionária de Gallus domesticus. PhD thesis, Federal University of Santa Catarina, Brazil.

Manetti, C., Bianchetti, C., Bizarri, M., Casciani, L., Castro, C., D´Ascenzo, G., Delfini, M., DI Cocco, ME., Laganà, A., Miccheli, A., Motto, M & Conti, F. (2004). NMR-based metabonomic study of transgenic maize. *Phytochemistry*, Vol. 65, pp.3187-3198. ISSN: 0031-9422

Marcucci, MC. (1995). Propolis: chemical composition, biological properties and therapeutic activity. *Apidologie*, Vol. 26, pp.83-99. ISSN: 1297-9678

Markham, KR., Mitchell, KA., Wilkins, AL., Daldy, JA & Lu, Y. (1996). HPLC and CG-MS identification of the major organic constituents in New Zealand propolis. *Phytochemistry* Vol. 42, pp.205-211. ISSN: 0031-9422

Popova, M., Bankova, V., Butovska, D., Petkov, V., Damynova, BN., Sabatini, AG., Marcazzan, GL & Bogdanov, S (2004). Validated methods for the quantifications of biologically active constituents of poplar-type propolis. *Phytochemical Analysis*, Vol. 15, pp.235-240. ISSN: 1099-1565

Sarbu, C & Mot, AC. (2011). Ecosystem discrimination and fingerprinting of Romain propolis by hierarchical fuzzy clustering and image analysis of TLC patterns. *Talanta*, Vol. 85, pp.1112-1117. ISSN: 0039-9140

Sawaya, ACHF., Silva, IB & Marcucci, MC. (2011). Analytical methods applied to diverse types of Brazilian propolis. *Chemistry Central Journal*, Vol. 5, pp.1-10. ISSN: 1752-153X

Schulz, H & Baranska, M. (2007). Identification and quantification of valuable plant substances by IR and Raman spectroscopy. *Vibrational Spectroscopy*, Vol. 43, pp.13-25. ISSN: 0924-2031

Singh, W. (2008). Robustness of three hierarchical agglomerative clustering techniques for ecological data. M.Sc. thesis, University of Iceland, Iceland.

Tsao, R & Deng, Z. (2004). Separation procedures for naturally occurring antioxidant phytochemicals. *Journal of Chromatography B*, Vol. 812, pp.85-99. ISSN: 1570-0232

Valcic, S., Montenegro, G & Timmermann, BN. (1998). Lignans from Chilean propolis. *Journal of Natural Products*, Vol. 61, pp.771-775. ISSN: 0974-5211

White, PJ. (2001). Properties of corn starch. In: Specialty corns. HALLAUER, AR. (Ed.). 2nd ed. pp. 189, CRC Press, London.

Wu, YW., Sun, SQ., Zhao, Y., Li, Q & Zhou, J. (2008). Rapid discrimination of extracts of
         Chinese propolis and poplar buds by FT-IR and 2D IR correlation spectroscopy.
         Journal of Molecular Structure, Vol. 884, pp.48-54. ISSN: 0022-2860
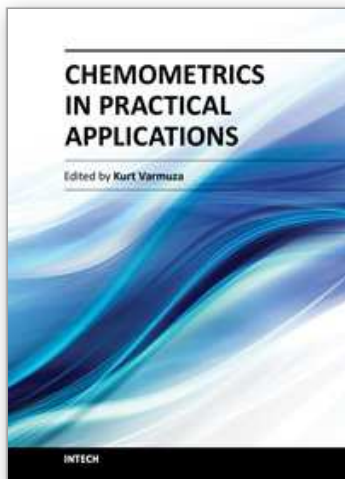
Chemometrics in Practical Applications

Edited by Dr. Kurt Varmuza

In the book "Chemometrics in practical applications", various practical applications of chemometric methods in chemistry, biochemistry and chemical technology are presented, and selected chemometric methods are described in tutorial style. The book contains 14 independent chapters and is devoted to filling the gap between textbooks on multivariate data analysis and research journals on chemometrics and chemoinformatics.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Marcelo Maraschin, Shirley Kuhnen, Priscilla M.M. Lemos, Simone Kobe de Oliveira, Diego A. da Silva, Maíra M. Tomazzoli, Ana Carolina V. Souza, Rúbia Mara Pinto, Virgílio G. Uarrota, Ivanir Cella, Antônio G. Ferreira, Amélia R.S. Zeggio, Maria B.R. Veleirinho, Ivone Delgadillo and Flavia A. Vieira (2012). Metabolomics and Chemometrics as Tools for Chemo(bio)diversity Analysis - Maize Landraces and Propolis, Chemometrics in Practical Applications, Dr. Kurt Varmuza (Ed.), ISBN: 978-953-51-0438-4, InTech, Available from: http://www.intechopen.com/books/chemometrics-in-practical-applications/metabolomics-and-chemometrics-as-tools-for-chemo-bio-diversity-analysis-maize-landraces-and-propolis

INTECH
open science | open minds