

**Non-targeted metabolomic profiling of maize landraces (*Zea mays* L.)
combined with chemometric tools**

ABSTRACT

Grain samples of maize landraces were collected and subjected to Fourier Transform Infrared spectroscopy (FTIR) analysis combined with chemometric tools in order to discriminate them regarding the chemical composition. Principal component analysis (PCA) and hierarchical clustering (HCA) were applied on selected peaks of the spectral data. The most important chemical groups found in all maize landrace samples were monoterpenes, sesquiterpenes, tetraterpenes, aminoacids, polysaccharides, lipids and proteins. Trace signals of secondary metabolites were also found in cultivars according to year of harvest.

Keywords: Chemical composition; FTIR ;Maize landraces, Non-supervised techniques, R software.

1. INTRODUCTION

Commonly referred to as maize or corn, *Zea mays* ssp. *mays* is one of the world's most important crop plants, achieving a multibillion dollar annual revenue. In addition to its agronomic importance, maize has been a keystone model organism for basic research for nearly a century [1-2]. According to Dwivedi [1], the current industrial agriculture system may be the single most important threat to maize biodiversity. A serious consequence of biodiversity loss is the displacement of locally adapted maize

26 landraces with adaptation traits to future climates by mono-cropping with genetically
27 uniform hybrids and improved cultivars.

28

29 Maize landraces represent heterogeneous, local adaptations of domesticated species,
30 and thereby provide genetic resources that meet current and new challenges for
31 farming in stressful environments. These local ecotypes can show variable phenology
32 and low-to-moderate edible yield, but are often highly nutritious. The main
33 contributions of maize landraces to plant breeding have been traits for more efficient
34 nutrient uptake and utilization, as well as useful genes for adaptation to stressful
35 environments such as water stress, salinity, and high temperatures [1, 3-5].

36

37 Systematic maize landraces evaluation may define patterns of diversity, which will
38 facilitate identifying alleles for enhancing yield and abiotic stress adaptation, thus
39 raising the productivity and stability of staple crops in vulnerable environments [1, 3-
40 5]. In other hand, regarding the evaluation tools, metabolomics has recently been
41 claimed as a promising concept and valuable tool in biotechnology, given its extensive
42 range of applications in functional genomics and, more globally, in the characterization
43 of biological systems [6]. Indeed, the possible tasks include studying metabolic
44 systems, measuring biochemical phenotypes, understanding and reconstructing
45 networks and discriminating between samples [6]. In the targeted metabolomics
46 approach, specific metabolites of known identity are profiled; good quantitative
47 precision is typically obtained. On the other hand, untargeted metabolomics aims to
48 simultaneously measure as many metabolites as possible in a biological specimen [7] .

49 Taking into account that maize landraces are currently being replaced by new
50 improved cultivars with a narrow genetic basis in Brazilian agriculture and with a
51 purpose to add value of those important genetic materials, the present study was

52 designed with the main goals of investigating the chemical composition of maize
53 landraces using non-targeted metabolomic approach (Fourier Transformed Infrared
54 Spectroscopy-FTIR) combined with chemometric tools (PCA and HCA) as a rapid
55 diagnostic tool to ascertain changes in chemical composition.

56 **2. MATERIALS AND METHODS**

57 *2.1 Maize landraces*

58 Maize landraces (*Zea mays* L.) were kindly provided by small farmers of the Anchieta
59 County (the western part of Santa Catarina state, southern Brazil, 26°31'11" S,
60 53°20'26" W) and produced under agro-ecological management practices during three
61 different harvests (2007, 2008, 2009) Table 1 presents the details of the cultivars used
62 each year). Maize grains were collected, dried and crushed to obtain fine powder for
63 later analysis by Fourier transform infrared spectroscopy (FT-IR).

64

65

66

67

68

69

70

71

72

73

74

75

76

Year	Maize cultivar analyzed	Acronym used
2007	Cultivar7	CULT7
	Cultivar10	CULT10
	Cultivar11	CULT11
	Cultivar12	CULT12
	Cultivar13	CULT13
	Cultivar15	CULT15
	Cultivar17	CULT17
2008	Lingua De Papagaio	LP0
	Mato Grosso-Palha Roxa	MG0
	Movimento De Pequenos Agricultores	MPA10
	Palha Roxa	PR0
	Rajado 8 Carreiras	R8C0
	Rajado	RJ0
	Roxo	RX0
	Roxo De Emílio	RXE0
2009	Lingua De Papagaio	LP1
	Mato Grosso-Palha Roxa	MG1
	Movimento De Pequenos Agricultores	MPA11
	Palha Roxa	PR1
	Rajado 8 Carreiras	R8C1
	Roxo	RX1
	Roxo De Emílio	RXE1

78 2.2. *Flour sample preparation*

79 Maize grains (250g – dry weight) were selected from each maize genotype and ground
80 to pass a 0.5mm sieve using a laboratory cyclone mill (MB Braeski C.Q).

81 2.3 *Fourier Transform infrared spectroscopy*

82 FTIR spectra of maize flours were recorded in a Bruker IFS-55 (Model Opus v. 5.0,
83 Bruker Biospin, Germany) spectrometer with a DTGS detector equipped with a golden
84 gate single reflection diamond attenuated total reflectance (ATR) accessory (45°
85 incidence-angle). A background spectrum of the clean crystal was acquired and
86 samples (100 mg) were spread and measured directly after pressing them on the
87 crystal. The spectra were recorded at the absorbance mode from 4000 to 500 cm⁻¹ at
88 the resolution of 4cm⁻¹. Five replicate spectra were collected for each sample. For pre-
89 processing, the spectra were normalized, baseline-corrected in the region of interest
90 by drawing a straight line before resolution enhancement (*k* factor of 1.7). The
91 assumed line shape was Lorentzian with a half width of 19 cm⁻¹ [8-11].

92

93 2.4 *Chemometric analysis*

94 FT-IR spectra were acquired and subjected to chemometric analysis using the R
95 software [12]. Non-targeted metabolomics was used to find variations in the chemical
96 composition of different cultivars, cultivated under different harvests and
97 environmental conditions. FT-IR spectra were acquired in transmittance mode and
98 then transformed to absorbance mode, as spectra objects in the ChemoSpec package
99 [13] using the equation 1.

100

$$\left[A = -\log_{10}(T) \right] \quad (\text{Eq. 1})$$

where A is the absorbance and T the transmittance

The spectra object was then converted to a hyperSpec [14] object using the bridging package hyperChemoBridge [15]. Before applying chemometric tools, the spectra were baseline corrected, intensity vector normalized and smoothed for subsequent analysis of PCA and HCA [6] for sample classification according to their biochemical status.

3. Results and discussion

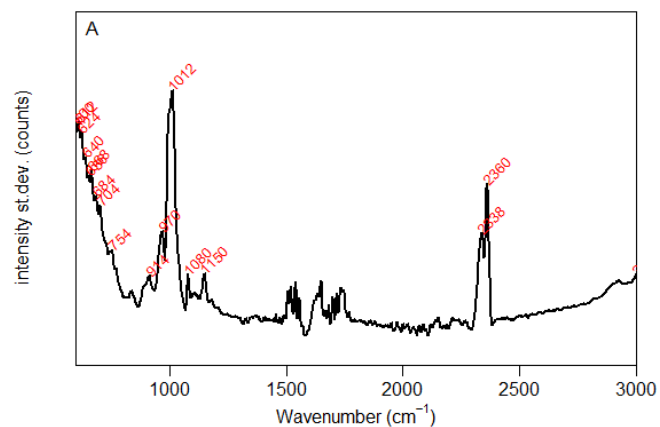
3.1 Peak selection and multivariate analysis

Results of the assignments for the most characteristic FTIR bands found in maize landraces and related chemical group are summarized in Table 2. The most important chemical groups found in all maize landrace samples were monoterpenes, sesquiterpenes, tetraterpenes, aminoacids, polysaccharides, lipids and proteins (Table 2).

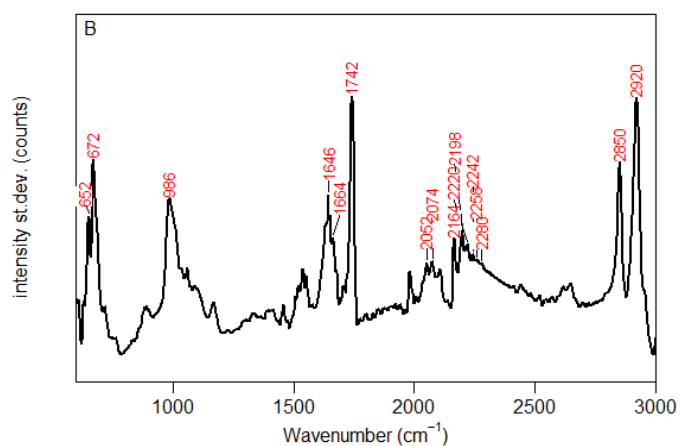
Table 2: Assignment for the most characteristic ATR-FTIR bands found in maize landraces. Peaks and chemical group related is represented.

Year of harvest	Chemical groups	Representative	ATR-IR wavenumber (1/cm)
2007	Monoterpenes Sesquiterpenes Tetraterpenes Aminoacids	Lutein, beta-carotene Methionine	600
			610
			624
			640
			668
			684
			704
	Polysacharides Lipids/fatty acids	Lutein Amylose, amylopectin Amylose, amylopectin	744
			968
			998
			1012
			2338
			2360
			3000
2008	Monoterpenes Tetraterpenes Polysacharides Proteins Lipids	Amylose, amylopectin	652
			672
			986
			1646
			1664
			1742
			2052
	Lipids		2164
			2198
			2220
			2242
			2258
			2850
2009	Polysacharides	Amylose, amylopectin	2920
	Proteins		998
	Lipids		1650
			1710
			1742
			2852
			2922

129 The intensity of peaks is also shown in the Figure 1 (A-C) for maize samples from 2007,
130 2008 and 2009 respectively. The selected peaks found in figure 1 were subjected for
131 further multivariate analysis aiming to find similarities or differences between maize
132 samples.



133



134

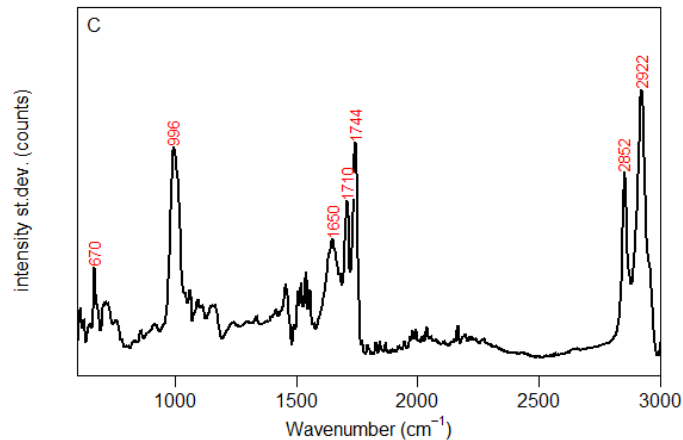
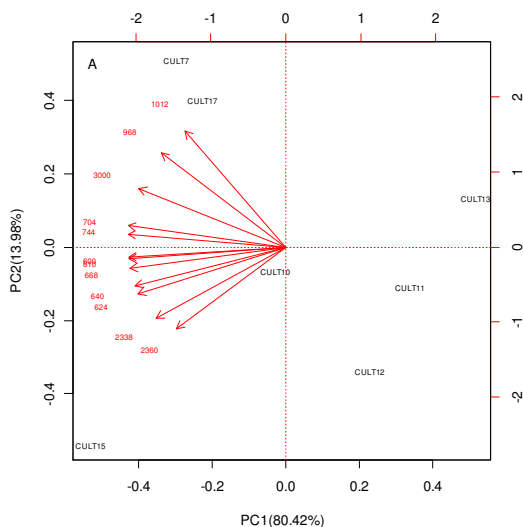


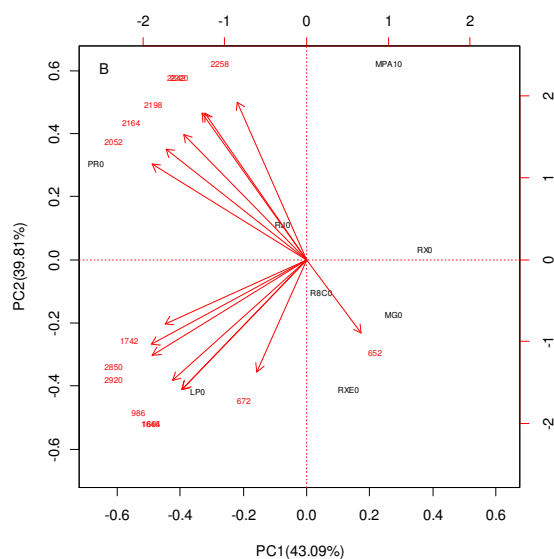
Figure 1: Main peaks found in FTIR spectra of (A) maize landraces produced in 2007, (B) 2008, (C) maize landraces produced during 2009 harvest , respectively

Principal component analysis of the 2007 maize landraces shown in the Figure 2A. Differences between cultivars were observed. Cultivars (11 and 12) and (7, 17) grouped together. The total variance explained by the first 2 PCs was 94.4%, being 80.42% for PC1 and 13.98% for PC2. Cultivars 7 and 17 grouped together due to similarities in amylose, amylopectin and lutein contents. Cultivar 10 grouped alone due to its content in fatty acids.

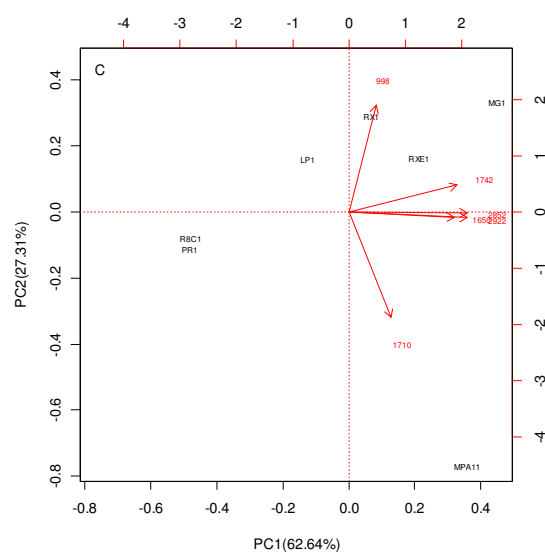


146
 147 Figure 2: Score and loading plots of Principal component analysis (PCA) of the whole
 148 spectral range ($600\text{--}3000\text{ cm}^{-1}$) showing the sample clustering according to their
 149 similarities. (A) Maze landraces produced during 2007 year.

150
 151 PCA of 2008 samples is represented in the Figure 2B. PC1 explained 43.09% and PC2
 152 39.81%, respectively. A good separation was found between PRO, MPA10 and LP0
 153 cultivars. PRO was most correlated to lipids or fatty acids and LP0 correlated to
 154 amylose, amylopectin, monoterpenes and proteins. PCA on 2009 maize samples is
 155 represented in Figure 2C. PC1 explained 62.64% and PC2 (27.31%) of the total
 156 variance. The loading plot indicated that cultivars MG1, LP1, RX1 and RXE1 are similar
 157 in chemical composition due to their values of amylose, amylopectin and fatty acids.
 158 Cultivars R8C1 and PR1 showed to be similar due to their lower levels of proteins and
 159 lipids.



160



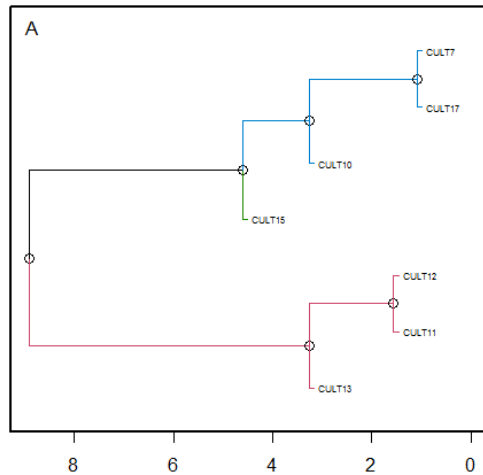
161

162 Figure 2: Score and loading plots of Principal component analysis (PCA) of the whole
 163 spectral range ($600\text{-}3000\text{ cm}^{-1}$) showing the sample clustering according to their
 164 similarities. (B) Maze landraces produced during 2008 year, (C) 2009 cultivars.

165

166 When FTIR data of 2007, 2008 and 2009 were subjected to hierarchical clustering
 167 (Figure 3A-C) aiming to find similarities or differences between samples and look for
 168 important wave bands related to sample clustering, an interesting result can be

169 observed and the main bands responsible for sample differences were found.
 170 Differences in the 2007 dataset were mainly due to the region of carbohydrates (998,
 171 1012 and 3000). Cultivar 15 grouped alone, and the most similar in chemical
 172 composition were cultivars 7 and 17, 11 and 12 respectively (Figure 3A).



173
 174 Figure 3: Hierarchical cluster dendrogram (HCA) of the main peaks found in the
 175 spectra. (A) maize landraces produced during 2007 year.

176
 177 For samples of 2008 harvest, the region of proteins, amylose and amylopectin was
 178 most important in sample dissimilarity. PR0 and MPA10 were the most dissimilar
 179 cultivars (Figure 3B). Two visible groups were found; being the first composed by LP0
 180 and RXE0 and second group by RX0, MG0, RJO and R8C0. For maize landraces of 2009
 181 (Figure 3C), proteins derived the sample clustering. MPA11 and LP1 were the most
 182 dissimilar cultivars confirming that observed previously in PCA analysis.

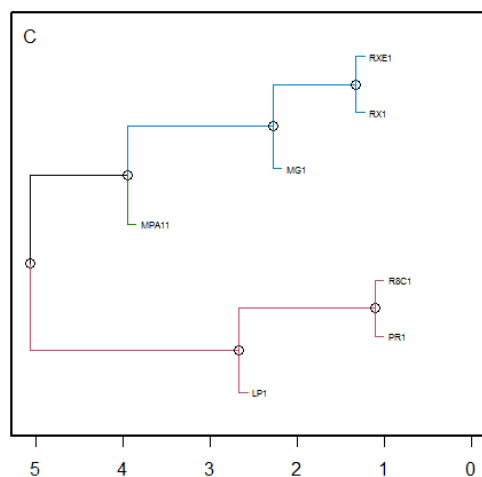
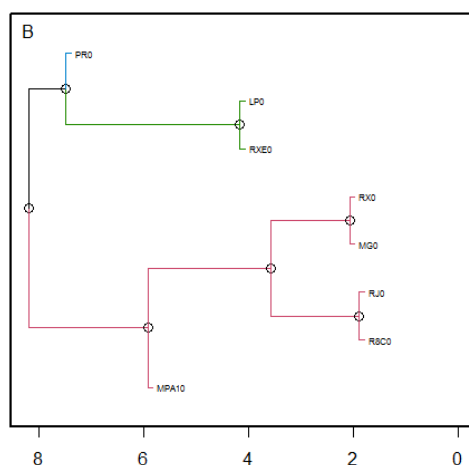


Figure 3: Hierarchical cluster dendrogram (HCA) of the main peaks found in the spectra. (B) 2008, (C) maize landraces produced during 2009 year.

CONCLUSIONS

FT-IR technique coupled with chemometric tools (PCA, HCA) were capable in finding changes in chemical composition of maize landraces produced over different years. Similarities in maize samples were also found due to carbohydrate composition. The major bands in FT-IR are related to carbohydrates, lipids and proteins. Trace signals of secondary metabolites were also found in cultivars according to year of harvest.

196 **REFERENCES**

- 197 1. Dwivedi S, Ceccarelli S, Blair M, Upadhyaya H, Are A, Ortiz R. Landrace Germplasm
198 for Improving Yield and Abiotic Stress Adaptation. *Trends in Plant Science*, 2016; 21:31-
199 42.
- 200 2. Strable J, Scanlon M. Maize (*Zea mays*): A Model Organism for Basic and Applied
201 Research in Plant Biology. *Cold Spring Harbor Protocols*, 2009; 132.
- 202 3. Vasconcelos A, Bonatti M, Schlindwein S, D'Agostini L, Homem L, Nelson R.
203 Landraces as an adaptation strategy to climate change for smallholders in Santa
204 Catarina, Southern Brazil. *Land Use Policy*, 2013; 34:250-254.
- 205 4. Bellon M, Hellin J. Planting Hybrids, Keeping Landraces: Agricultural Modernization
206 and Tradition Among Small-Scale Maize Farmers in Chiapas, Mexico. *World*
207 *Development*, 2011; 39:1434-1443.
- 208 5. Mercer K, Martínez-Vásquez Á, Perales H. Asymmetrical local adaptation of maize
209 landraces along an altitudinal gradient. *Evolutionary Applications*, 2008;1:489-500.
- 210 6. Costa C, Maraschin M, Rocha M. An R package for the integrated analysis of
211 metabolomics and spectral data. *Computer Methods and Programs in Biomedicine*,
212 2016; 129:117-124.
- 213 7. Yi L, Dong N, Yun Y, Deng B, Ren D, Liu S, Liang Y. Chemometric methods in data
214 processing of mass spectrometry-based metabolomics: A review. *Analytica Chimica*
215 *Acta*, 2016; 914:17-34.

- 216 8. Uarrota VG, Amante E, Demiate I, Vieira F, Delgadillo I, Maraschin M.
217 Physicochemical, thermal, and pasting properties of flours and starches of eight
218 Brazilian maize landraces (*Zea mays* L.). *Food Hydrocolloids*, 2013; 30:614-624.
- 219 9. Kuhnen S, Ogliari J, Dias P, da Silva Santos M, Ferreira A, Bonham C, Wood K,
220 Maraschin M. Metabolic Fingerprint of Brazilian Maize Landraces Silk (Stigma/Styles)
221 Using NMR Spectroscopy and Chemometric Methods. *J. Agric. Food Chem.*, 2010;
222 58:2194-2200.
- 223 10. Schulz H, Baranska M. Identification and quantification of valuable plant
224 substances by IR and Raman spectroscopy. *Vibrational Spectroscopy*, 2007; 43:13-25.
- 225 11. Baranska M, Schulz H, Joubert E, Manley M. In Situ Flavonoid Analysis by FT-
226 Raman Spectroscopy: Identification, Distribution, and Quantification of Aspalathin in
227 Green Rooibos (*Aspalathus linearis*). *Analytical Chemistry*, 2006; 78:7716-7721.
- 228 12. R Core Team. R: A language and environment for statistical computing. R
229 Foundation for Statistical Computing, Vienna, Austria. 2016. URL [https://www.R-](https://www.R-project.org/)
230 [project.org/](https://www.R-project.org/).
- 231 13. Hanson BA. ChemoSpec: Exploratory Chemometrics for Spectroscopy. R package
232 version 4.3.28, 2016. github.com/bryanhanson/ChemoSpec
- 233 14. Beleites C, Sergo V. HyperSpec: a package to handle hyperspectral data sets in R', R
234 package version 0.98-20150304. 2016. <http://hyperspec.r-forge.r-project.org>
- 235 15. McManus C. HyperChemoBridge: Bridge between hyperSpec &ChemoSpec. R
236 package version 1.1.3. 2014. <https://github.com/Chathurga/HyperChemoBridge>