NMR metabolomics and metabolic pathways analysis of cassava genotypes at different harvesting times and cooking characteristics

Elenilson G. Alves Filho[a], Lorena Mara A. Silva[b]; Robson M. Martins[c], Willyane J.D.J. Oliveira[c], Cristine V. Soares[c], Luciana A. de Oliveira[d], Edy S. de Brito[b]*

[a] Universidade Federal do Ceará, Departamento de Engenharia de Alimentos, Bloco 858, Campus do Pici, 60440-900, Fortaleza - CE, Brazil.

[b] Embrapa Agroindústria Tropical, Rua Dra. Sara Mesquita, 2270, Pici, 60511-110, Fortaleza - CE, Brazil.

[c] Universidade Federal do Ceará, Departamento de Química, Bloco 940, Campus do Pici, 60440-900, Fortaleza - CE, Brazil

[d] Embrapa Mandioca e Fruticultura, 44380-000, Cruz das Almas, BA, Brazil.

* Corresponding author

E-mail: edy.brito@embrapa.br

Abstract

Cassava is an important staple food for low income countries. However, its cooking characteristics are especially affected by genotype. In this study, two groups of genotypes, namely hard to cook (HTC) and easy to cook (ETC), harvested at different times (9 and 15 months) were evaluated by ^{1}H NMR and chemometrics. Additionally, the lignin of these materials was studied by ^{1}H-^{13}C HSQC. The carbohydrates were the most important class of compounds to differentiate the cassava genotypes. The correlation of NMR with cooking time and starch content showed that the higher content of primary metabolites, mostly glucose, can be associated to longer cooking times and reduction of starch corroborating the metabolic pathways analysis. Furthermore, it was observed that the lignin from cell wall did not differentiate the cooking performance of the genotypes.

Keywords: cassava; lignin; *Manihot esculenta* Crantz; multivariate analysis; NMR, pathway analysis

1. Introduction

Cassava (*Manihot esculenta* Crantz) is a starchy root that provides calories and nutrition for more than half a billion people (Siebers, Catarino, & Agusti, 2017). It is the world's sixth vital crop (Zainuddin, Fathoni, Sudarmonowati, Beeching, Gruissem, & Vanderschuren, 2018) because it can be cultivated on marginal soils in several conditions as severe drought and high temperatures (Bredeson, Lyons, Prochnik, Wu, Ha, Edsinger-Gonzales, et al., 2016; Siebers, Catarino, & Agusti, 2017). The roots are used as raw, after boiled or processed for human and animal nutrition and they are also used as source for energy production (El-Sharkawy, 2003). The genetic diversity of cassava (Bredeson, et al., 2016) represents a broad genomic base for crop breeding programs, which focus on increasing resistance to pests and diseases, adaptation to edaphoclimatic conditions, as well as on the reduction of constrains that limit the yields and quality of cassava.

Cassava roots can be harvested anytime from 8 to 24 months after planting (Bellotti & Arias, 2001), and they are an important source of starch, which accounts from 20 to 40% of its weight (Lima, Rocha Viana, Sousa Sabino, Silva, Silva, & Sousa, 2017). In Africa, long-term harvested and bitter cassava are often used for processing, whereas short-term harvested and sweet types are direct boil-and-eat (Ngeve, 2003). During the cooking process, starch undergoes modifications as gelatinization, water absorption, and volume increase of the granulates, resulting in a product with important characteristics for consumer acceptability (Butarelo, Beleia, Fonseca, & Ito, 2004). Efforts have been focused on breeding programs with the aims of conferring disease tolerance and increasing yield of the root. However, the development of cultivars that present important characteristics for consumers as cooking time is also relevant.

Emerging metabolomics evaluation by means of nuclear magnetic resonance (NMR) spectroscopy allows obtaining comprehensive profiles of the primary metabolism during plant physiological activities (Filho, Braga, Silva, Miranda, Silva, Canuto, et al., 2018). NMR also enables partial characterization of several matrices as lignin that is a hydrophobic, three-dimensional and highly branched natural biopolymer of phenolic monomers (Inkrod, Raita, Champreda, & Laosiripojana, 2018). Therefore, the aim of this work was to evaluate hard to cook (HTC) and easy to cook (ETC) cassava genotypes by metabolomics-based discrimination through ^{1}H NMR (nuclear magnetic resonance) and chemometrics in order to point metabolites associated to the cooking time. In addition, we also evaluate the lignin portion of the root through ^{1}H-^{13}C HSQC (heteronuclear single quantum correlation) NMR to evaluate the influence of the cell wall material in the cooking time.

2. Materials and methods

2.1. Sampling

The cassava (*Manihot esculenta*) genotypes HTC (2009 02-13, 2009 02-16, 2009 09-05 and 2009 12-20) and ETC (BRS Brasil, BRS Dourada, Eucalipto and Saracura) were grown at Embrapa Mandioca e Fruticultura (Cruz das Almas-BA, Brazil). The experiment was carried out with plants arranged in a completely randomized 8×4 design, with four replications. It was used a spacing of 1.0 m × 0.7 m fertilized with P_2O_5. The experiment was conducted from May-2016 (planting) to February-2017 and August-2017 (harvesting). After the harvesting, the material was cleaned, peeled, chopped, grinded and freeze-dried. The dried samples were used for the subsequent steps.

92

93 *2.2. NMR spectroscopy*

94 An amount of 30 mg of dried cassava was soaked in a mixture of 600 μL of D_2O
95 and 1 mM of TMSP-d_4 (sodium-3-trimethylsilylpropionate-2,2,3,3-d_4). The solutions
96 were mixed for 2 min at room temperature and centrifuged at 804.6 *g* for 2 min. The
97 supernatants were transferred to 5 mm NMR tubes. The ^{1}H NMR spectra were acquired
98 in triplicate under quantitative parameters (Freitas, Alves Filho, Silva, Zocolo, de Brito,
99 & Gramosa, 2018). The PRESAT pulse sequence was applied for non-deuterated water
100 suppression, and the spectra were acquired under controlled temperature (299.1 K) with
101 24 scans using 48,000 of time domain points for a spectral window of 16.0 ppm. The
102 RF pulse was calibrated to 90° (7.75 μs pulse length at 58 dB of power), acquired with
103 acquisition time of 5.0 s and recycling delay of 25.0 s (determined by the inversion-
104 recovery pulse sequence). The ^{1}H NMR spectra processing was performed by applying
105 exponential Lorentzian broadening of 0.3 Hz and zero filling to 16k points before the
106 Fourier transformation. The phase correction was performed manually, and the
107 automatic baseline correction was applied over the entire spectral range. The spectra
108 were referenced to chemical shift at δ 0.0 from TMSP-d_4 ((trimethylsilyl)propionic-
109 2,2,3,3-d_4 acid sodium salt) simplet signal.

110 Two-dimensional (2D) NMR experiments were acquired using the standard
111 spectrometer library pulse sequences. The ^{1}H-^{1}H gradient correlation spectroscopy
112 (*g*COSY) experiments were obtained with spectral width of 9,615.4 Hz in both
113 dimensions; 1442 × 200 data matrix; 16 scans per t1 increment and relaxation delay of
114 1.0 s. The ^{1}H-^{13}C *g*HSQC experiments were acquired with an evolution delay of 3.425
115 ms (transfer delay) for a coupling constant one-bond proton-carbon [^1J(C,H)] of 146
116 Hz; 1442 × 200 data matrix; 48 scans per t1 increment; spectral widths of 9,615.4 Hz

in f2 and 30,154.5 Hz in f1, and relaxation delay of 1.0 s. The ^{1}H-^{13}C *g*HMBC (gradient heteronuclear multiple bond correlation) experiments were recorded with an evolution delay of 62.5 ms for coupling constant ^{LR}J(C,H) of 8 Hz; 1442 × 200 data matrix; 96 scans per t1 increment; spectral widths of 9,615.4 Hz in f2 and 36,182.7 Hz in f1, and relaxation delay of 1.0 s.

2.3. Multivariate analysis of the ^{1}H NMR dataset

A total data of 135 ^{1}H NMR spectra (chemical shifts between δ 0.7 and 9.2) from biological triplicate of 8 different cassava genotypes under two different harvest times (9 and 15 months) were converted to American Standard Code for Information Interchange (ASCII) files and imported by Origin™ 9.4 program for numerical matrix construction. The resultant matrix was imported by the PLS-Toolbox package (version 8.6.2 – Eigenvector Research Inc., Manson, WA USA) under the Matlab™ programming language (R2019a; The MathWorks Inc., Natick, MA) to perform the unsupervised chemometric method by Principal Component Analysis (PCA), as well as supervised analysis by Partial Least Squares (PLS) and Partial Least Squares Discriminant Analysis (PLS-DA) under confidence level of 95%. Initially, a general evaluation was developed using the total number of spectra, which resulted in a numerical matrix with dimensionality of 1,127,598 data points: 138 cassava samples × 8,171 variables from each spectrum. In order to detail the cassava discrimination according to the species, additional PCA were developed considering each harvest time separately, which resulted in two numerical matrices with dimensionality of 563,799 data points: 69 cassava samples × 8,171 variables.

Algorithms for baseline correction and normalization were applied over the variables, as well as the variables alignment using COW (Correlation Optimized

142 Warping) with segment of 50 data points and a slack of 5 data points (Sousa,
143 Magalhães, & Ferreira, 2013). The samples data were mean-centered and the Singular
144 Value Decomposition (SVD) algorithm was applied to decompose the matrices. For
145 PLS and PLS-DA, the Simplified PLS (SIMPLS) algorithm was applied to modeling
146 classification analysis. The number of Latent Variables (LV) was chosen based on the
147 following statistical parameters: RMSEC (Root Mean Squared Error of Calibration);
148 RMSECV (Root Mean Squared Error of Cross Validation); and similarity index
149 (RMSEC / RMSECV) higher than 0.75 (Ballabio & Consonni, 2013; Freitas, Alves
150 Filho, Silva, Zocolo, de Brito, & Gramosa, 2018).
151
152 *2.4. Metabolic pathway analysis*
153 In order to evaluate the metabolic pathways associated to different harvesting
154 times (9 and 15 months) and the cooking characteristics (hard and easy to cook) of the
155 aforementioned cassava species, the pair wise comparison was performed to both
156 characteristics by classification model using orthogonal partial least squares
157 discriminant analysis (OPLS-DA) algorithm in the PLS Toolbox™ program. the
158 loadings and coefficient plots were analyzed and the variables important for projection
159 (VIP) with value higher than 1 were quantified and used as input for metabolic pathway
160 analysis using MetaboAnalyst 4.0 (*http://www. metaboanalyst.ca*) (Chen, Wu, Li, Liu,
161 Zhao, & Yang, 2019; Liu, Wu, Lim, Aggarwal, Yang, & Wang, 2017; Liu, Wu, Lim,
162 Lai, Lee, & Yang, 2018; Xia, Psychogios, Young, & Wishart, 2009).
163
164 *2.5. Determination of starch (fresh and dry weight), cooking time, and °Brix value*
165 The starch analysis was performed according to the methodology described by
166 Holm et al (1986) in the dry samples. The starches were hydrolyzed by the action of

the enzymes α-amylose and amyloglucosidase; and the glucose content quantified spectrophotometrically.

The cooking time was determined using the modified Mattson apparatus. The apparatus consisted of a support formed by two parallel plates with 12 holes, and each supporting 12 cylindrical aluminum connecting rods of 90 g contained in the needle tip. Ten plants per plot were harvested and 10 roots were selected, washed, cut into pieces of 6 cm (cylinders) and peeled. The cylinders were washed, drained and 12 cylinders were weighed and thermally treated in boiling, distilled water (1 kg L^{-1}). The cooking time was recorded after 2 cm penetration of needle tips in the 10 cylinders. Softening time was evaluated in two replicates.

For °Brix value 100 mg of the dried sample were mixed with 0.9 mL of water, stirred in ultrasound bath for 2 min and centrifuged for 2 min. The sample supernatant was collected, and a drop was placed onto a digital refractometer (Atago, model Pocket PAL-3, Tokyo, Japan) for measurement at 25 °C.

2.6. Lignin extraction

The lignin extraction was performed by pressurized liquid extraction (Dionex ASE 350, Thermo Fisher Scientific, Massachusetts, EUA). The biological replicates were unified to give rise to one representative genotype sample. Therefore, the lignin was extracted twice for ETC (BRS Brasil, Saracura, BRS Dourada, Eucalipto) and HTC (2009 02-13, 2009 02-16, 2009 09-05 and 2009 12-20) genotypes. Dried cassava (5.7 g) was mixed with 1.7 g diatomaceous earth and placed in stainless steel extraction cells of 34 mL. The extractions were performed with 60% ethanol/water (v/v), acidified with 30 mM H_2SO_4, at 190 °C for 75 min. All the extractions were performed in triplicate. The resultant dissolved lignin extract (organosolv liquor) was evaporated and

192 precipitated upon dilution with water (10:1 w/w dilution ratio H_2O: organosolv liquor,
193 at 27 ºC for 48 h). The lignin was recovered by vacuum filtration using qualitative filter
194 paper of 80G. The paper with the lignin were dried at 60 °C and weighted to obtain the
195 lignin yield. Lignin (50 mg) was mixed with 600 µL of DMSO-d_6, inserted in 5 mm
196 NMR tubes, and dipped into an ultrasonic bath for 24 h. The ^{1}H-^{13}C HSQC were
197 acquired with spectral widths of 30165.9 Hz and 9615.4 for ^{13}C and ^{1}H dimensions,
198 respectively. In addition, it was recorded as 962 complex points with a recycle delay of
199 0.5 s for ^{1}H dimension. It was obtained 64 transients, with 256 complex points for ^{13}C
200 dimension with one-bond ^{1}J X-H coupling constant of 146 Hz. The data were processed
201 by means of VNMRJ™, using standard parameters. The characteristic signals of the
202 lignin components were assigned in accordance with previously reported data (Bai,
203 Xiao, Shi, & Sun, 2013; Mansfield, Kim, Lu, & Ralph, 2012; Pinheiro, Soares,
204 Santaella, Silva, Canuto, Cáceres, et al., 2017). A relative quantitative method along
205 with the integrated areas of the HSQC cross peaks from syringyl and guaiacyl was used
206 to determine the proportion of the aforementioned compounds.
207
208 3. Results and discussion
209
210 *3.1. NMR characterization*
211 The identification of the main organic compounds in different genotypes of
212 cassava was performed. Figures 1Sa and 1Sb present the ^{1}H NMR spectra of the ETC
213 cassava genotype (Dourada harvested after 9 months) and HTC genotype (2009_1220
214 harvested after 9 months), respectively. In addition, the Table 1S describes the
215 structures, ^{1}H and ^{13}C chemical shifts (δ), multiplicity, correlations and constant
216 coupling (J in HZ) of the correspondent compounds identified. Therefore, cassava

comprises high levels of sugars, amino acids, and short chain organic acids, regardless the genotypes.

3.2. Metabolomic and metabolic pathway analysis of cassava harvesting time

In general, slights differences among the compounds from cassava with different cooking characteristics were observed, which depended on the genotype. Therefore, due to the number of identified compounds within the cassava genotypes, as well as the inherent similarity among the samples composition, unsupervised multivariate evaluation by PCA was applied to investigate the composition variability according to the cassava genotypes under different harvesting time. Figure 1 presents the PCA results, with the cassava genotypes harvested after 9 months of planting in red and 15 months in blue. In addition, the HTC genotypes are illustrated with stars, and the ETC in circles.

Figure 1

The PC1 was the main responsible for the cassava separation mainly according to the harvest time, with cassava harvested after 9 months of planting at positive scores, and those harvested after 15 months of planting at negative scores. It is also observed that the genotypes harvested after 15 months presented higher chemical variation, since those samples were more disperse at the PC2 axis. The loadings graph showed the root harvested after 15 months with higher content of glucose, arginine, succinic acid, while roots harvested after 9 months of planting manly presented higher content of sucrose.

3.3. Metabolomic and metabolic pathway analysis of cassava with different cooking characteristics

In order to observe the chemical variation among the HTC and ETC samples, it was performed distinct PCA for genotypes harvested after 9 and 15 months. Figures 2a and 2b presents the scores and loadings for genotypes harvested after 9 months respectively; and Figures 2c and 2d those harvested after 15 months. The HTC genotypes are illustrated in red and the ETC in blue.

Figure 2

For genotypes harvested after 9 months (Fig. 2a and 2b), the PC2 was the main axis for cassava genotypes chemical distinction according to the cooking performance. For genotypes harvested after 15 months (Fig. 2c and 2d) the PC1 was the relevant axis. In general, the loadings graph (Fig. 2b and 2d) for both harvesting periods shows that ETC cassava genotypes presented higher amounts of sucrose than the HTC genotypes. On the other hand, the HTC genotypes presented higher amounts of glucose. In addition, for ETC roots harvested after 9 months, it was observed a higher content of amino acids arginine, glutamine, threonine, tryptophan and tyrosine. For roots harvested after 15 months, the discrepancy in amino acids content was reduced; and tryptophan and tyrosine also increased in ETC samples.

For a comprehensive analysis of cassava harvested with different cooking characteristics, the data from supervised OPLS-DA (Izquierdo-García, Villa, Kyriazis, del Puerto-Nevado, Pérez-Rial, Rodriguez, et al., 2011) was used as input for metabolic pathway analysis using MetaboAnalyst (Chen, Wu, Li, Liu, Zhao, & Yang, 2019; Liu, Wu, Lim, Aggarwal, Yang, & Wang, 2017; Liu, Wu, Lim, Lai, Lee, & Yang, 2018;

Xia, Psychogios, Young, & Wishart, 2009). Figure 3a shows the pathways associated with the cooking characteristics for cassava harvested at 9 months, and Figure 3b shows the pathway at 15 months. The metabolites colored from deep red to yellow indicate an increased concentration of metabolites and the size indicates the impact on pathway (–log(p)). The most significant metabolic pathways with a false discovery rate (FDR) lower than 6.9×10^{-8} and 2.33×10^{-4}, for 9 and 15 months respectively, and with metabolites with impact at the route higher then 0.03 were considered for both harvesting periods.

Figure 3

The different pathways associated to cooking characteristics of the roots were similar regardless harvesting time. The lower content of succinic acid in ETC roots might induce the down regulation of citrate cycle (TCA cycle) (1) and sulfur metabolism (2). The TCA cycle provides carbon skeleton for biosynthesis of several compounds (Vega-Mas, Cukier, Coleto, González-Murua, Limami, González-Moro, et al., 2019) and the sulfur metabolism is essential for plant growth, development, and response to environmental changes. Therefore, the suppression of both pathways might affect aerial plant growth (van der Merwe, Osorio, Araújo, Balbo, Nunes-Nesi, Maximova, et al., 2010). The lower content of α and β-glucose in ETC roots also might induce the down regulation of galactose metabolism (3). This metabolism is linked to the synthesis of raffinose family oligosaccharides (RFOs) (Zhang, Song, & Bartels, 2016). The RFOs protect plant cells from oxidative damage caused by various types of stress conditions (Nishizawa, Yabuta, & Shigeoka, 2008; Peshev, Vergauwen, Moglia, Hideg, & Van den Ende, 2013) and as carbon transport and storage (Turgeon & Wolf,

291 2009). which also might affect plant development. Finally, it is observed the upper
292 regulation of sucrose in ETC roots. Sucrose metabolism is linked to starch and sucrose
293 metabolism (4) changing properties of starch in grains (Chang, Liu, Lin, Li, Wang,
294 Chien, et al., 2017) and the upper regulation might induce the accumulation of starch
295 in the root.

296 The free sugars were the more important components to differentiate the
297 genotypes and their cooking performance. Therefore, in order to observe the correlation
298 of the carbohydrates (observed at NMR spectrum as main components, Fig. 1SI) with
299 the cooking performance and starch content, it was employed a PLS regression
300 modeling using cooking time and percentage of starch (fresh and dry weight). The
301 regression modeling was developed for genotypes from the different harvesting periods
302 (9 and 15 months) separately, in order to maximize the covariance between the
303 independent variables (^{1}H NMR dataset i.e. X matrix) and the dependent variables –
304 cooking time and percentage of starch.

305 Table 1 describes the respective statistical parameters of the regressions. In
306 general, the models possess high correlation coefficients of calibration (above 0.9) and
307 validation (above 0.8); very low bias values from the modeling; relative low calibration
308 and validation errors; and relative high similarity criterion (proximity) between the
309 calibration and validation modeling (Ballabio & Consonni, 2013; Freitas, Alves Filho,
310 Silva, Zocolo, de Brito, & Gramosa, 2018). Despite the satisfactory figures of merit,
311 the cooking parameter could be better since the cooking procedure was interrupted at
312 50 min (standard time for cooking essay) and consequently the genotypes with longer
313 cooking times might have this parameter under estimated.

314 Table 1

315 In general, statistical parameters presented elevate the correlation between the
316 cooking time variations for cassava genotypes at different harvesting time (9 and 15
317 months). The compositional variability of the entire ^{1}H NMR spectra showed a close
318 correlation among the composition and cooking characteristics of the root. Since starch
319 and sucrose metabolism pathways was trigged, the study of the correlation of NMR and
320 the starch accumulation at the roots is important. The model shows that there is a
321 correlation between NMR spectrum and starch content at the root. For both variables
322 (percentage of starch at fresh and dry weight) the model for 9 months was better
323 adjusted. The PCA shows that cassava harvested after 15 months presents higher
324 content α and β-glucose while roots harvested at 9 months mainly presents higher
325 content of sucrose. The free sugars accumulation in cassava were previously correlated
326 with disruption in starch synthesis pathway by enzyme activity (Carvalho, Souza,
327 Cascardo, Junior, & Campos, 2004). On this behalf, glucose was found to be the major
328 free sugar in cassava with low starch content and reduced levels of amylose as well
329 (Carvalho, Souza, Cascardo, Junior, & Campos, 2004).

330 In starch biosynthesis, the enzyme granule-bound starch synthase I (GBSSI)
331 (Denyer, Johnson, Zeeman, & Smith, 2001) polymerizes amylose from the donor
332 substrate ADP-glucose. The inhibition of this enzyme produce amylose-free starches
333 (Raemakers, Schreuder, Suurs, Furrer-Verhorst, Vincken, de Vetten, et al., 2005) which
334 melts at higher temperature and leads to weaker gels (Raemakers, et al., 2005).
335 Therefore, the lower content of α and β-glucose along with higher content of sucrose in
336 ETC roots might be correlated to higher content of starch in this roots that possess better
337 cooking characteristics. In addition, these data shows that the roots harvested at 9
338 months might possess higher content of starch and better cooking characteristics.
339

340 *3.4. Lignin characterization of different genotypes of cassava*

341 In order to evaluate the influence of the cell wall material on the cooking time,
342 lignin of genotypes with different cooking characteristics was analyzed. The Figure 4a
343 shows the representative spectrum from BRS Dourada (ETC) lignin and the Figure 4b
344 shows the hybrid 2009 12-20 (HTC) lignin.

345 Figure 4

346 It was observed the presence of the main compounds of lignin as syringyl and
347 guaiacyl. It was also noticed the oxidized syringyl, p-coumarate, cinnamic aldehyde,
348 and the polysaccharides residues as aryl ether linkage as α–*O*–4, β–*O*–4, and the
349 anomeric signal, highlighted with the square in the Figure 4a and 4b. The ratio of
350 syringyl/guaiacyl was also evaluated for the genotypes with different cooking
351 characteristics. The information about the ratio of syringyl and guaiacyl is an important
352 parameter for matrix understanding. Lignin with higher amount of syringyl is more
353 easily removed during the delignification process (Gutiérrez, Rodríguez, & del Río,
354 2006) since the less-reactivity of the C5 aromatic carbon from syringyl implies in a less
355 condensed structure which increase the lignin solubility (Sette, Wechselberger, &
356 Crestini, 2011). Therefore, our hypothesis is that the high content of guaiacyl might
357 impose a greater constrain in the matrix inducing a stiffening of the system. The Table
358 2 displays the S/G ratio of the lignin obtained from the different genotypes.

359 Table 2

360 The Table 2 shows that the S/G ratio did not changed with respect to the cooking
361 characteristics or genotype of cassava. Therefore, the cooking performance of the root
362 might be more associated to the disruption in the starch synthesis pathway than to the
363 cell wall adhesive characteristics (lignin).

364

4. Conclusion

The choice of harvesting time and genotypes plays important role on cooking performance and, consequently, final destination of cassava roots. The root harvested at 15 months presents higher content of glucose, arginine, succinic acid, while roots harvested at 9 months of planting mainly presents higher content of sucrose. In general, regardless the harvesting time, the ETC cassava genotypes presented higher amounts of sucrose and lower amount of glucose than HTC. Those metabolites were associated to important pathways as galactose, sucrose and starch metabolism. Roots with higher content of glucose presents reduced content of starch and higher cooking time. Starch is a gelatinization agent and its low content can be associated to longer cooking times as corroborated with the trigged pathways and the multivariate regression of NMR data, and percentage of starch. Therefore, clones harvested after 9 months and the ones pointed as ETC (Saracura, Dourada, Eucalipto and Brasil) might possess better cooking characteristics. In addition, the syringyl/guaiacyl ratio in lignin did not correlate to the cooking performance of the genotypes. This works contributes to the understanding of the biosynthetic mechanism that lead to the different cooking characteristics of cassava.

CRediT authorship contribution statement

Elenilson G. Alves Filho: Methodology, Formal analysis, Visualization, Investigation, Writing - original draft. **Lorena M. A. Silva:** Conceptualization, Methodology, Formal analysis, Visualization, Investigation, Supervision, Writing - original draft, Writing – review & editing. **Robson M. Martins:** Formal analysis, Writing - original draft. **Willyane J.D.J. Oliveira:** Formal analysis, Writing - original draft. **Cristine V. Soares**: Formal analysis Writing - original draft. **Luciana A. de Oliveira**: Project administration, Conceptualization, Investigation, Funding acquisition Writing -

original draft. **Edy S. de Brito:** Project administration, Conceptualization, Investigation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is financially supported by Embrapa (03.15.01.002.00.00). The author EAF thank CNPq and FUNCAP for a scholarship (314737/2018-9).

References

Bai, Y.-Y., Xiao, L.-P., Shi, Z.-J., & Sun, R.-C. (2013). Structural variation of bamboo lignin before and after ethanol organosolv pretreatment. *International journal of molecular sciences, 14*(11), 21394-21413. https://doi.org/10.3390/ijms141121394.

Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: linear models. PLS-DA. *Analytical Methods, 5*(16), 3790-3798. https://doi.org/10.1039/C3AY40582F.

Bellotti, A. C., & Arias, B. (2001). Host plant resistance to whiteflies with emphasis on cassava as a case study. *Crop Protection, 20*(9), 813-823. https://doi.org/10.1016/S0261-2194(01)00113-2.

Bredeson, J. V., Lyons, J. B., Prochnik, S. E., Wu, G. A., Ha, C. M., Edsinger-Gonzales, E., Grimwood, J., Schmutz, J., Rabbi, I. Y., Egesi, C., Nauluvula, P., Lebot, V., Ndunguru, J., Mkamilo, G., Bart, R. S., Setter, T. L., Gleadow,

R. M., Kulakow, P., Ferguson, M. E., Rounsley, S., & Rokhsar, D. S. (2016). Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nature Biotechnology, 34*, 562. https://doi.org/10.1038/nbt.3535.

Butarelo, S. S., Beleia, A., Fonseca, I. C. d. B., & Ito, K. C. (2004). Hydration of cassava tissues and starch gelatinization during the cooking process. *Food Science and Technology, 24*, 311-315. https://doi.org/10.1590/S0101-20612004000300001.

Carvalho, L. J. C. B., Souza, C. R. B., Cascardo, J. C. M., Junior, C. B., & Campos, L. (2004). Identification and characterization of a novel cassava (Manihot esculenta Crantz) clone with high free sugar content and novel starch. *Plant Molecular Biology, 56*(4), 643-659. https://doi.org/10.1007/s11103-004-4873-9.

Chang, T. S., Liu, C. W., Lin, Y. L., Li, C. Y., Wang, A. Z., Chien, M. W., Wang, C. S., & Lai, C. C. (2017). Mapping and comparative proteomic analysis of the starch biosynthetic pathway in rice by 2D PAGE/MS. *Plant Molecular Biology, 95*(4-5), 333-343. 10.1007/s11103-017-0652-2.

Chen, L., Wu, J. E., Li, Z., Liu, Q., Zhao, X., & Yang, H. (2019). Metabolomic analysis of energy regulated germination and sprouting of organic mung bean (Vigna radiata) using NMR spectroscopy. *Food Chemistry, 286*, 87-97. https://doi.org/10.1016/j.foodchem.2019.01.183.

Denyer, K. A. Y., Johnson, P., Zeeman, S., & Smith, A. M. (2001). The control of amylose synthesis. *Journal of Plant Physiology, 158*(4), 479-487. https://doi.org/10.1078/0176-1617-00360.

El-Sharkawy, M. A. (2003). Cassava biology and physiology. *Plant Molecular Biology, 53*(5), 621-641. https://doi.org/10.1023/B:PLAN.0000019109.01740.c6.

Filho, E. G. A., Braga, L. N., Silva, L. M. A., Miranda, F. R., Silva, E. O., Canuto, K. M., Miranda, M. R., de Brito, E. S., & Zocolo, G. J. (2018). Physiological changes for drought resistance in different species of Phyllanthus. *Scientific Reports, 8*(1), 15141. 10.1038/s41598-018-33496-7.

Freitas, J. V. B., Alves Filho, E. G., Silva, L. M. A., Zocolo, G. J., de Brito, E. S., & Gramosa, N. V. (2018). Chemometric analysis of NMR and GC datasets for chemotype characterization of essential oils from different species of Ocimum. *Talanta, 180*, 329-336. https://doi.org/10.1016/j.talanta.2017.12.053.

Gutiérrez, A., Rodríguez, I. M., & del Río, J. C. (2006). Chemical characterization of lignin and lipid fractions in industrial hemp bast fibers used for manufacturing high-quality paper pulps. *Journal of Agricultural and Food Chemistry, 54*(6), 2138-2144. https://doi.org/10.1021/jf052935a.

Holm, J., Björck, I., Drews, A. and Asp, N.-G. (1986), A Rapid method for the analysis of starch. Starch/Stärke, 38: 224-226. doi:10.1002/star.19860380704

Inkrod, C., Raita, M., Champreda, V., & Laosiripojana, N. (2018). Characteristics of lignin extracted from different lignocellulosic materials via organosolv fractionation. *BioEnergy Research, 11*(2), 277-290. https://doi.org/10.1007/s12155-018-9895-2.

Izquierdo-García, J. L., Villa, P., Kyriazis, A., del Puerto-Nevado, L., Pérez-Rial, S., Rodriguez, I., Hernandez, N., & Ruiz-Cabello, J. (2011). Descriptive review of current NMR-based metabolomic data analysis packages. *Progress in*

Nuclear Magnetic Resonance Spectroscopy, 59(3), 263-270. https://doi.org/10.1016/j.pnmrs.2011.02.001.

Lima, A. C. S., Rocha Viana, J. D., Sousa Sabino, L. B., Silva, L. M. R., Silva, N. K. V., & Sousa, P. H. M. (2017). Processing of three different cooking methods of cassava: Effects on in vitro bioaccessibility of phenolic compounds and antioxidant activity. *LWT - Food Science and Technology, 76*, 253-258. https://doi.org/10.1016/j.lwt.2016.07.023.

Liu, Q., Wu, J. E., Lim, Z. Y., Aggarwal, A., Yang, H., & Wang, S. (2017). Evaluation of the metabolic response of *Escherichia coli* to electrolysed water by 1H NMR spectroscopy. *LWT - Food Science and Technology, 79*, 428-436. https://doi.org/10.1016/j.lwt.2017.01.066.

Liu, Q., Wu, J. E., Lim, Z. Y., Lai, S., Lee, N., & Yang, H. (2018). Metabolite profiling of *Listeria innocua* for unravelling the inactivation mechanism of electrolysed water by nuclear magnetic resonance spectroscopy. *International Journal of Food Microbiology, 271*, 24-32. https://doi.org/10.1016/j.ijfoodmicro.2018.02.014.

Mansfield, S. D., Kim, H., Lu, F., & Ralph, J. (2012). Whole plant cell wall characterization using solution-state 2D NMR. *Nature Protocols, 7*, 1579. 10.1038/nprot.2012.064.

Ngeve, J. M. (2003). Cassava root yields and culinary qualities as affected by harvest age and test environment. *Journal of the Science of Food and Agriculture, 83*(4), 249-257. https://doi.org/10.1002/jsfa.1307.

Nishizawa, A., Yabuta, Y., & Shigeoka, S. (2008). Galactinol and raffinose constitute a novel function to protect plants from oxidative damage. *Plant Physiology, 147*(3), 1251-1263. 10.1104/pp.108.122465.

488 Peshev, D., Vergauwen, R., Moglia, A., Hideg, E., & Van den Ende, W. (2013).
489 Towards understanding vacuolar antioxidant mechanisms: a role for fructans?
490 *Journal of Experimental Botany, 64*(4), 1025-1038. 10.1093/jxb/ers377.
491 Pinheiro, F. G. C., Soares, A. K. L., Santaella, S. T., Silva, L. M. A. e., Canuto, K.
492 M., Cáceres, C. A., Rosa, M. d. F., Feitosa, J. P. d. A., & Leitão, R. C. (2017).
493 Optimization of the acetosolv extraction of lignin from sugarcane bagasse for
494 phenolic resin production. *Industrial Crops and Products, 96*, 80-90.
495 https://doi.org/10.1016/j.indcrop.2016.11.029.
496 Raemakers, K., Schreuder, M., Suurs, L., Furrer-Verhorst, H., Vincken, J.-P., de
497 Vetten, N., Jacobsen, E., & Visser, R. G. F. (2005). Improved Cassava Starch
498 by Antisense Inhibition of Granule-bound Starch Synthase I. *Molecular*
499 *Breeding, 16*(2), 163-172. https://doi.org/10.1007/s11032-005-7874-8.
500 Sette, M., Wechselberger, R., & Crestini, C. (2011). Elucidation of Lignin Structure
501 by Quantitative 2D NMR. *Chemistry – A European Journal, 17*(34), 9529-
502 9535. https://doi.org/10.1002/chem.201003045.
503 Siebers, T., Catarino, B., & Agusti, J. (2017). Identification and expression analyses
504 of new potential regulators of xylem development and cambium activity in
505 cassava (Manihot esculenta). *Planta, 245*(3), 539-548. 10.1007/s00425-016-
506 2623-2.
507 Sousa, S., Magalhães, A., & Ferreira, M. M. C. (2013). Optimized bucketing for
508 NMR spectra: Three case studies. *Chemometrics and Intelligent Laboratory*
509 *Systems, 122*, 93-102. https://doi.org/10.1016/j.chemolab.2013.01.006.
510 Turgeon, R., & Wolf, S. (2009). Phloem Transport: Cellular pathways and molecular
511 trafficking. *Annual Review of Plant Biology, 60*(1), 207-221.
512 10.1146/annurev.arplant.043008.092045.

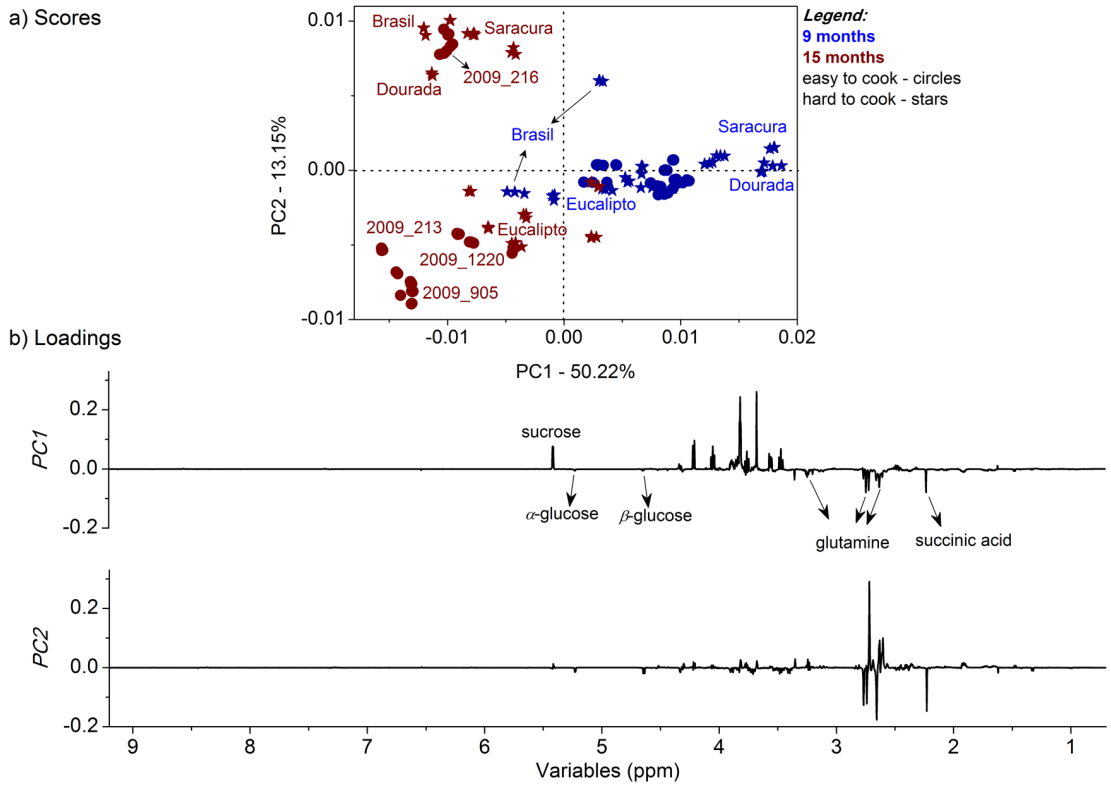
van der Merwe, M. J., Osorio, S., Araújo, W. L., Balbo, I., Nunes-Nesi, A., Maximova, E., Carrari, F., Bunik, V. I., Persson, S., & Fernie, A. R. (2010). Tricarboxylic acid cycle activity regulates tomato root growth via effects on secondary cell wall production. *Plant Physiology, 153*(2), 611-621. 10.1104/pp.109.149047.

Vega-Mas, I., Cukier, C., Coleto, I., González-Murua, C., Limami, A. M., González-Moro, M. B., & Marino, D. (2019). Isotopic labelling reveals the efficient adaptation of wheat root TCA cycle flux modes to match carbon demand under ammonium nutrition. *Scientific Reports, 9*(1), 8925. 10.1038/s41598-019-45393-8.

Xia, J., Psychogios, N., Young, N., & Wishart, D. S. (2009). MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Research, 37*(Web Server issue), W652-W660. https://doi.org/10.1093/nar/gkp356.

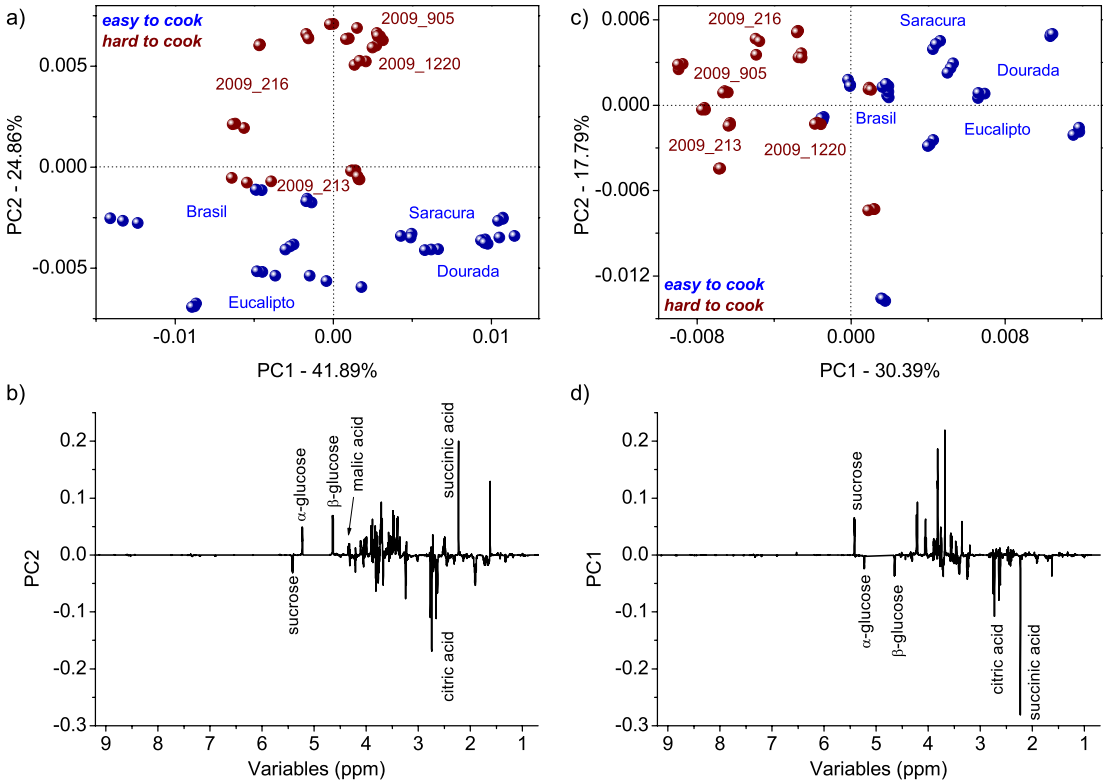
Zainuddin, I. M., Fathoni, A., Sudarmonowati, E., Beeching, J. R., Gruissem, W., & Vanderschuren, H. (2018). Cassava post-harvest physiological deterioration: From triggers to symptoms. *Postharvest Biology and Technology, 142*, 115-123. https://doi.org/10.1016/j.postharvbio.2017.09.004.

Zhang, Q., Song, X., Bartels, D. (2016). Enzymes and metabolites in carbohydrate metabolism of desiccation tolerant plants. *Proteomes*. *4*, 40. doi: 10.3390/proteomes4040040.



538

539 Figure 1 – PC1 × PC2 scores coordinate system (a) and respective loadings (b) of
540 different genotypes of cassava. Legend: cassava genotypes harvested after 9 months of
541 planting in red and 15 months of planting in blue; HTC genotypes are illustrated as stars
542 and the ETC as circles.



543

544 Figure 2 – Scores coordinate system (PC1 × PC2) from cassava genotypes harvested
545 after 9 months of planting (a) and after 15 months of planting (c). Relevant loadings
546 from cassava genotypes harvested after 9 months (b) and 15 months (d) plotted in lines
547 in the same intensity.

548

549

550

551

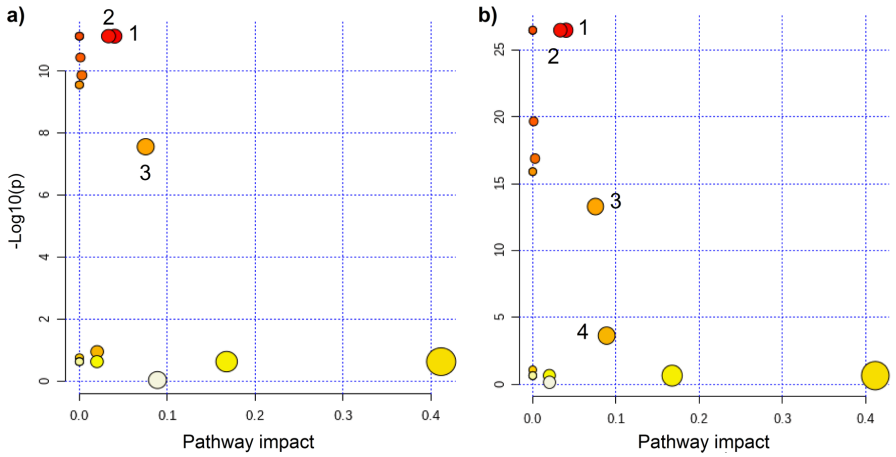


Figure 3 – Pathways associated with the metabolism response for HTC and ETC cassava over a) 9 months and b) 15 months of harvesting time. Legend: 1- Citrate cycle (TCA cycle); 2 - Sulfur metabolism; 3 - Galactose metabolism; 4 - Starch and sucrose metabolism.

571 Table 1 – Statistical parameters of the multivariate regression from PLS modeling of
572 the cooking time according to the cassava aging.

Cooking characteristics							
Model	*5 LV[a] (%)*	*Bias[b]*	*r^2 cal[c]*	*RMSEC[d]*	*r^2 val[e]*	*RMSECV[f]*	*RMSEC / RMSEV[g]*
9 months	88.35	-1.4×10^{-14}	0.91	3.31	0.86	4.00	0.83
15 months	74.17	-7.1×10^{-15}	0.91	3.42	0.88	4.10	0.83
Percentage of starch at fresh root							
Model	*8 LV[a] (%)*	*Bias[b]*	*r^2 cal[c]*	*RMSEC[d]*	*r^2 val[e]*	*RMSECV[f]*	*RMSEC / RMSEV[g]*
9 months	86.74	-3.5×10^{-15}	0.96	0.70	0.88	1.27	0.55
15 months	84.95	0	0.98	0.57	0.96	0.92	0.62
Starch at dried base							
Model	*8 LV[a] (%)*	*Bias[b]*	*r^2 cal[c]*	*RMSEC[d]*	*r^2 val[e]*	*RMSECV[f]*	*RMSEC / RMSEV[g]*
9 months	87.33	-2.8×10^{-14}	0.96	1.17	0.88	2.06	0.57
15 months	82.80	-2.8×10^{-14}	0.98	1.20	0.94	1.92	0.62

573 [a] The total variance percent in the X matrix refers to the first five Latent Variable (LV);
574 [b] Influenced modeling;
575 [c] Coefficient of correlation between the real times to cook and those predicted during
576 the calibration;
577 [d] Root Mean Square Error of Calibration;
578 [e] Coefficient of correlation between the real times to cook and those predicted during
579 the validation;
580 [f] Root Mean Square Error of the Cross Validation;
581 [g] Similarity criterion.
582

583

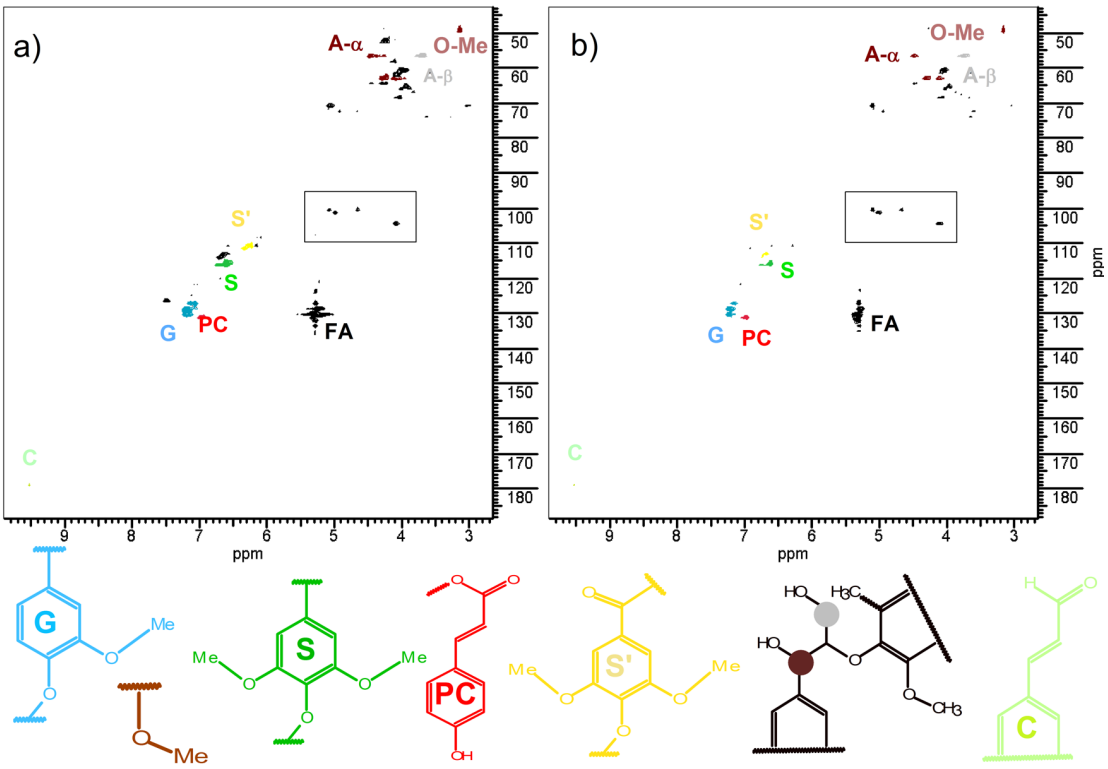


Figure 4 – ^{1}H-^{13}C HSQC of the lignin from cassava extracted in a) genotype BRS Dourada (ETC); b) hybrid 2009 12-20 (HTC). Legend: S: syringyl; S': oxidized syringyl; G: guaiacyl; OMe: methoxyl groups; PC: p-coumarate; C: cinnamic aldehyde; A: aryl ether with A–α: α–O–4 and A–β: β–O–4; FA: fatty acids bounded to lignin; square region: polysaccharides bounded to lignin.

598 Table 2 – Ratio of syringyl/guaiacil (S/G) from the different cassava genotypes.

Genotype	S/G
Saracura	0,6255
BRS Dourada	0,8385
Eucalipto	0,8655
BRS Brasil	0,79321
2009 02-13	0,8395
2009 02-16	0,8513
2009 09-05	0,7965
2009 12-20	0,8320

599

600

601