UV-visible scanning spectrophotometry and chemometric analysis as tools for carotenoids analysis in cassava genotypes (*Manihot esculenta* Crantz)

Rodolfo Moresco[1] [*], Virgílio G. Uarrota[1], Aline Pereira [1], Maíra Tomazzoli[1], Eduardo da C. Nunes[2], Luiz Augusto Martins Peruch[2], Jussara Gazzola[3], Christopher Costa[4], Miguel Rocha[4] and Marcelo Maraschin[1]

[1]Plant Morphogenesis and Biochemistry Laboratory, Federal University of Santa Catarina, Florianópolis, Brazil http://ufsc.br/

[2]Santa Catarina State Agricultural Research and Rural Extension Agency (EPAGRI), Experimental Station of Urussanga, Urussanga, Brazil http://www.epagri.sc.gov.br/

[3]Experimental Nutrition Laboratory, Federal University of Santa Catarina, Florianópolis, Brazil http://ufsc.br/

[4]Centre of Biological Engineering, University of Minho, Braga, Portugal, http://www.uminho.pt

Summary

In this study, the metabolomics characterization focusing on the carotenoid composition of ten cassava (*Manihot esculenta*) genotypes cultivated in southern Brazil by UV-visible scanning spectrophotometry and reverse phase-high performance liquid chromatography was performed. Cassava roots rich in β-carotene are an important staple food for populations with risk of vitamin A deficiency. Cassava genotypes with high pro-vitamin A activity have been identified as a strategy to reduce the prevalence of deficiency of this vitamin. The data set was used for the construction of a descriptive model by chemometric analysis. The genotypes of yellow-fleshed roots were clustered by the higher concentrations of *cis*-β-carotene and lutein. Inversely, cream-fleshed roots genotypes were grouped precisely due to their lower concentrations of these pigments, as samples rich in lycopene (red-fleshed) differed among the studied genotypes. The analytical approach (UV-Vis, HPLC, and chemometrics) used showed to be efficient for understanding the chemodiversity of cassava genotypes, allowing to classify them according to important features for human health and nutrition.

Keywords: Chemometrics, descriptive models, partial metabolome, cassava genotypes, carotenoids, RP-HPLC, UV-vis.

[*]To whom correspondence should be addressed. Email: rodolfo.moresco@posgrad.ufsc.br

1 Introduction

Cassava (*Manihot esculenta* Crantz, 1766) currently ranks as the third most important species as a source of calories in the world among the group of staple food crops, including rice and maize. It is primarily consumed in places where there are prevailing drought, poverty, and malnutrition. Diseases related to vitamin A deficiency are among the major nutritional problems in developing countries. It is estimated that 190 million children in preschool age have low retinol activity in plasma ($\leq$0.70 μmol.L-1), subclinical symptom of this deficiency [1]. Cassava is considered a staple food for many populations with risk of vitamin A deficiency and is predominantly produced by small-scale farmers with limited resources. Carotenoids come from plants secondary metabolism and belong to the class of terpenes, which are classified by the number of C5 units and also known as tetraterpenes (with 40 carbons or 8 units of C5). The carotenoids are more stable in *trans* isomeric form and could be divided in carotenes and xanthophylls. The former are hydrocarbons carotenoids, such as α- and β-carotene and lycopene, while xanthophylls are oxygenated derivatives of carotenes and include the compounds lutein and β-cryptoxanthin, for instance [2].

Some carotenoids are precursors to vitamin A [3]. The enzymatically conversion of pro-vitamin A carotenoids into vitamin A occurs mainly in the intestinal mucosa during the absorption of those precursors, protecting against vitamin A deficiency. Several health-promoting effects of carotenoids such as immune-enhancement and reduction of the risk of developing degenerative diseases, cancer, cardiovascular diseases, cataract, and macular degeneration have been claimed [4] [5].

Pro-vitamin A carotenoids contrast in their vitamin A activities because of their distinct chemical structures, being important to identify which carotenoid compounds form the food matrix and to check their actual activity in conversion to vitamin A. In cassava, β-carotene is the majoritarian pro-vitamin A carotenoid [6], but the amounts found in white and cream cassava roots (the most commonly consumed by populations) are usually low in comparison to yellow ones [3].

It is known that roots with yellow flesh are highly correlated with the concentration of carotenoids and the search for materials with higher pro-vitamin A activity is recognized as a strategy to reduce the prevalence of this deficiency [7]. Besides β-carotene, *M. esculenta* also contains small amounts of other carotenoids, e.g., lycopene and the xanthophylls, lutein and β-cryptoxanthin.

Because of the high importance of cassava crops in Brazil, genebanks of cassava collections have been established and maintained for the purpose of preserving wild genotypes, traditional landraces and commercial varieties. Thus, the identification and preservation of genotypes rich in carotenoids is thought to be relevant for the Brazilian and global food security and nutrition.

Spectroscopic methods using UV-Visible (UV-Vis) wavelengths are rapid, cheap, and provide metabolic fingerprints that can be processed enabling pattern recognition between samples. UV-Vis scanning spectrophotometry requires little sample amounts and preparation, and rapidly provides valuable and robust information about the presence of particular classes of metabolites, such as carotenoids. Chromatographic methods (e.g., reverse phase-high performance liquid chromatography – RP-HPLC) are also largely used for the identification and quantification of

plant extracts' compounds. These compounds can be used as biochemical markers assisting to discriminate samples with peculiar characteristics.

This study, in connection with the metabolomics characterization of the genebank's cassava accesses, emphasizes the carotenoid profile in root samples using a typical analytical platform, i.e., UV-Vis spectrophotometry and reverse phase-high performance liquid chromatography (RP-HPLC). The data set afforded (i.e., the spectrophotometric profiles and the chromatographic quantification of each carotenoid compound) was used to build descriptive and classification models by calculation of the principal components and cluster analysis. Such an analytical approach allows the rapid and effective extraction of relevant and non-redundant information from a set of complex data, enabling a more detailed and robust understanding of possible differences and/or similarities in the studied samples, as well as their improved discrimination. In practical terms, this study develops and applies biotechnological approaches, by coupling the use of biochemical markers with bioinformatics tools to gain insights to more rationally support genetic breeding programs of cassava.

2 Material and Methods

2.1 Selection of cassava genotypes

Roots of ten genotypes of *M. esculenta* (2010/2011 season) from the EPAGRI's germplasm bank (Urussanga Experimental Station, 28º31'18"S, 49º19'03"W, Santa Catarina, southern Brazil) were used in this study. Traditionally, they have been called *Apronta mesa*, *Pioneira*, *Oriental*, *Amarela*, *Catarina*, *IAC576-70 - (Instituto Agronômico de Campinas)*, hereafter designed as *IAC5760-70*, *Salézio*, *Estação*, *Videira*, and *Rosada* and were selected based on their economic and social importance. The samples were classified considering a visual analysis of colors of root parenchymal tissues. Cream-fleshed and yellow-fleshed, with lower and higher intensity of the yellow color, respectively, and red-fleshed, samples with reddish color.

2.2 Extraction and analysis of carotenoid compounds

Carotenoids were extracted from fresh roots as described in 2004 by Rodriguez-Amaya & Kimura [2] using an Ultra-Turrax (Janke & Kunkel IKA - T25 basic) and organic solvents: acetone and petroleum ether (v/v, 1:1). The absorbances of the organosolvent extracts were collected on a UV-visible spectrophotometer (Gold Spectrum lab 53 UV-Vis spectrophotometer, BEL photonics, Brazil) using a spectral window from 200 to 700nm. Aliquots (10μl) of the extracts were also injected into a liquid chromatograph (LC-10A Shimadzu) system equipped with a C18 reversed-phase column (Vydac 201TP54, 250mm x 4.6mm, 5μm Ø, 35ºC) coupled to a pre-column (C18 Vydac 201TP54, 30mm x 4.6mm, 5μm Ø) and a spectrophotometric detector (450nm). Methanol: acetonitrile (90: 10, v/v) was used for elution at a rate of 1 ml/min. The identification and quantification of compounds of interest was carried out via co-chromatography and comparison of retention times of samples with those of standard compounds (Sigma–Aldrich, USA) under the same experimental conditions.

2.3 Statistical and chemometric analysis

All procedures were performed in triplicate (n = 3). Data were collected, summarized, and submitted to analysis of variance (ANOVA) followed by post-hoc Tukey's test ($p \leq 0.05$) for mean comparison. The processing of the spectrophotometric profile considered the definition of the spectral window of interest (200-700nm), baseline correction, normalization, and optimization of the signal/noise ratio (smoothing). The processed data set was initially subjected to multivariate statistical analysis, by applying Principal Component Analysis (PCA) and clustering methods. Further, these data were also used for fold-change and Volcano plot with t-test univariate analyses, considering samples with pigmented roots (red and yellow fleshed) and non-pigmented roots (cream-fleshed) as the two groups to compare. Furthermore, the spectral data set and the amounts of the target carotenoids determined by RP-HPLC were used for calculation of the Principal Components, supported by scripts written in R language (v. 3.1.1) [8] using tools defined by our research group and some functions from the packages Chemospec [9], HyperSpec [10], and ggplot2 [11]. Finally, in order to extract latent information from the UV-Vis dataset, classification models were built by applying supervised classification and feature selection methods, e.g., PLS-DA (Partial Least Squares Discriminant Analysis) and RF (Random Forest). All R scripts, raw data, and additional chemometric analysis are available in supplementary material, in http://darwin.di.uminho.pt/metabolomicspackage/ as well as the data analysis report automatically generated from the R scripts using the features provided by R Markdown http://darwin.di.uminho.pt/metabolomicspackage/cassava-carotenoids.html. This allows anyone to fully reproduce and document the experiments.

3 Results and Discussion

Carotenoids typically show maximum absorption at 450nm [6] and as depicted in Fig.1, all the spectral profiles (200-700nm) of the yellow and red cassava root extracts revealed prominent absorbance peaks between 400-500nm, indicating that the organosolvent system used was efficient to extract the target metabolites. Lower absorbance values were found in cream-fleshed roots, precisely because they have low concentrations of carotenoids as the *Rosada* genotype (red-fleshed) showed the highest absorbance values at 450nm.

Fold change analysis applied to samples with pigmented roots (red and yellow-fleshed) and non-pigmented roots (cream-fleshed) discriminated important points between 400 and 500 nm (carotenoids fingerprint region). Besides, some signals around 300 nm, a typical region of phenolic compounds absorption, also showed to be significant in their relative signal intensity defined as a $\geq$ 2-fold change (Fig.2). This approach is consistent with the univariate analysis expressed through statistical significance of the descriptive p-value model. Indeed, wavelengths with lower p-values and therefore representing more significant differences between the samples studied were detected in the 440-470 nm spectral window (data shown on the analysis report).

Because of this, a volcano plot analysis was applied, associating the statistical differences found by the t-test and fold change analysis (Fold change threshold = 2 and p value $\leq$ 0.001). This

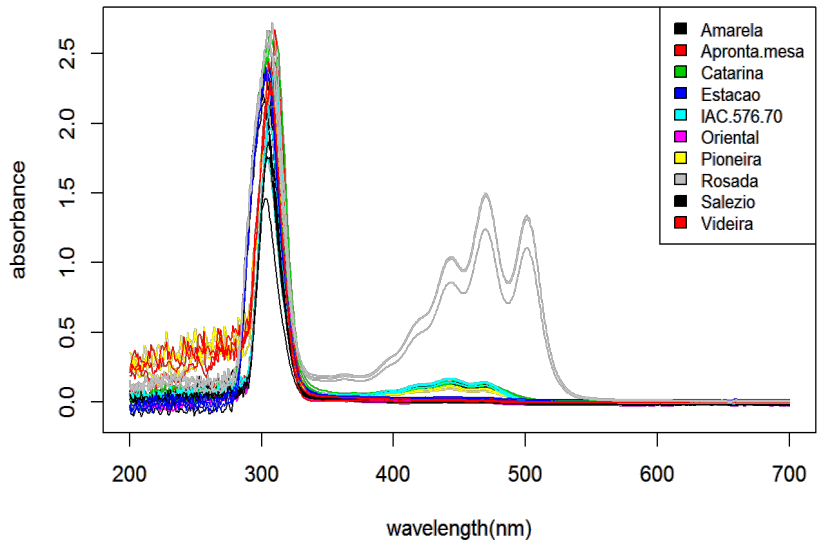

Figure 1: Typical UV-Vis spectrophotometric profiles (λ = 200-700 nm, acetone: petroleum ether - v/v) of root parenchymal tissues of ten cassava genotypes cultivated in southern Brazil.

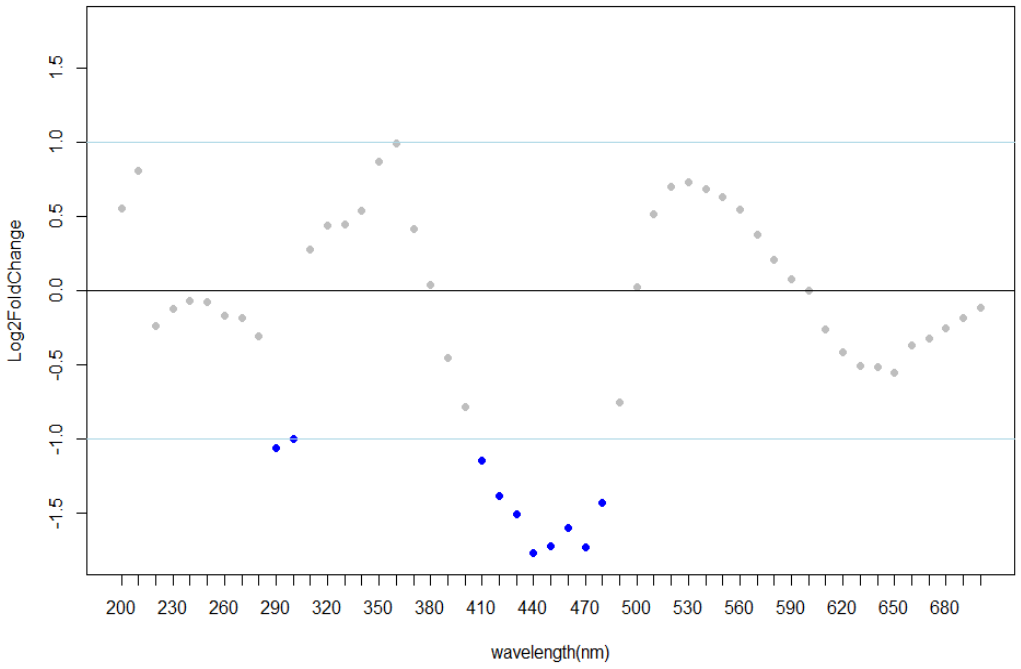

Figure 2: Fold change analysis of the UV-Vis data set (λ = 200-700 nm, acetone: petroleum ether - v/v). Significant differences (blue symbols) in the relative intensity (defined as a $\geq$ 2-fold change) of absorbance signals important for discriminating the samples where detected between 400 and 500 nm (carotenoids fingerprint region).

analysis is used to quickly identify differences in large metabolomics datasets. In this case, once again, the values that display large magnitude fold changes, as well as high statistical significance, occurred at 440-470 nm wavelengths, being related to typical maximum absorption bands of carotenoids (Fig.3).

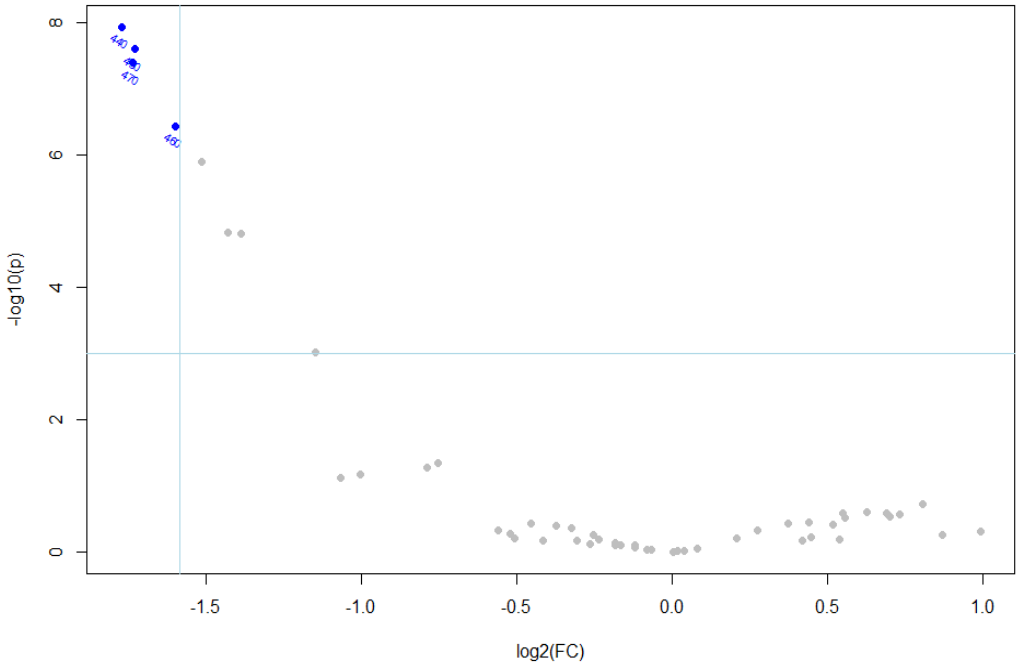

Figure 3: Volcano plot analysis (Fold change threshold = 2 and p $\leq$ 0.001). The larger magnitude fold changes and the higher statistical significance refer to wavelengths between 440-470 nm. (blue symbols).

In a second series of experiments, the principal components were calculated from the full spectrophotometric (λ = 200-700nm) data matrix. PC1 and PC2 contributed to explain 78.9% of the total variance of the data set, but a clear discrimination of the samples according to the carotenoid concentrations was not found. Only the *Rosada* genotype distinguished from the others by grouping in PC1 + / PC2 -. Genotypes with high (yellow) and low carotenoid contents (cream) were spread out over the factorial distribution plane, making difficult any discrimination (Fig.4).

Such findings prompted us to build a second analytical model by applying PCA to the carotenoid fingerprint region of the UV-Vis (400-500nm). In this case, PC1 and PC2 accounted for 99.97% of the variance, clearly revealing three groups according to their similarities (Fig.5). Interestingly, the samples were grouped according to their carotenoids contents determined by RP-HPLC and distributed according to the root flesh color. Cassava genotypes with yellow-fleshed roots (*Pioneira*, *Amarela*, *Catarina* and *IAC576-70*) were clustered along the PC2 + axis. Genotypes with cream-fleshed roots and lower carotenoid content (*Apronta mesa*, *Oriental*, *Salézio*, *Estação*, and *Videira*) were grouped in PC1/PC2 -. In its turn, the *Rosada* genotype (red) seems to have a metabolic profile occurring away from all the other samples.

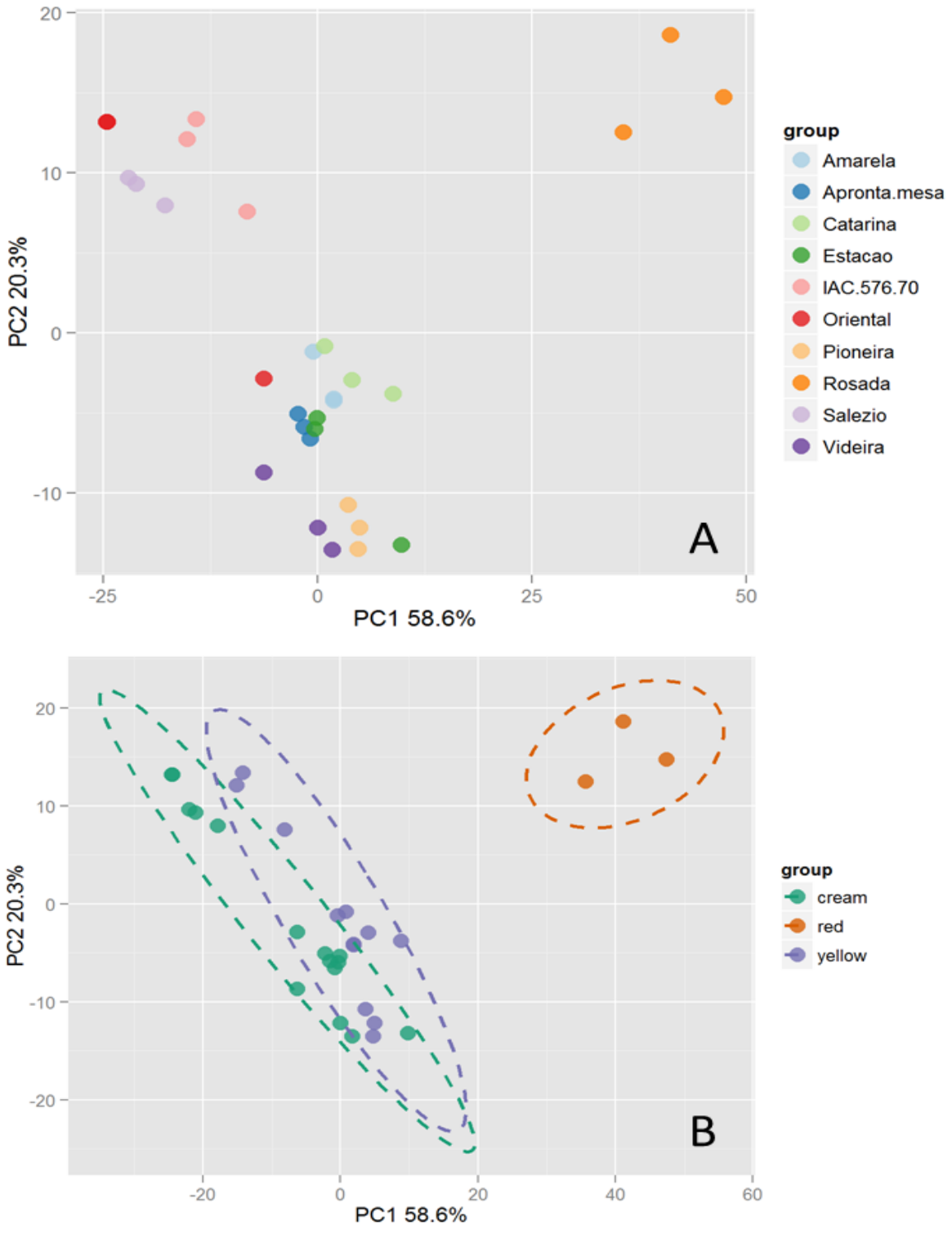

Figure 4: Factorial distribution (principal components 1 and 2) of the UV-Vis spectral data set (200-700 nm) of the organosolvent extract of roots of ten cassava genotypes (A). Graphical demonstration of factorial distribution (principal components 1 and 2) according to the root flesh color (B).

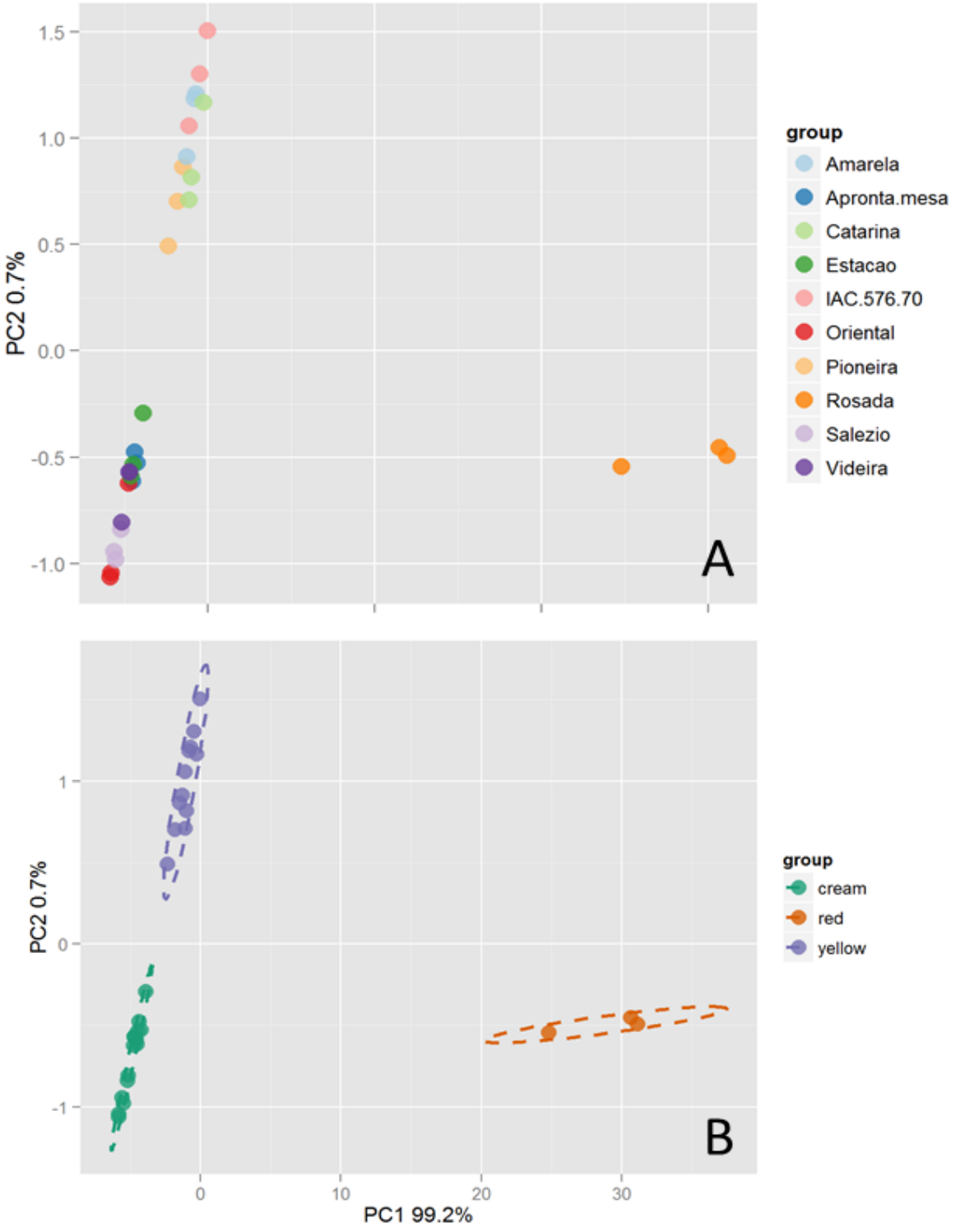

Figure 5: A - Factorial distribution (principal components 1 and 2) of the spectral data set of the fingerprint region of carotenoids (UV-Vis 400-500 nm, acetone: petroleum ether - v/v, 1:1). B - Graphical demonstration according to the root flesh color.

The chromatographic profiles of the organosolvent extracts identified *cis*- and *trans*-β-carotene, α-carotene, lutein, and β-cryptoxanthin in all the cassava genotypes analyzed. The presence of lycopene, a common precursor of the carotenoids above mentioned, was detected only in *Rosada* genotype, a fact that led us to speculate this is an important reason for its clear discrimination in respect to other genotypes. The isomer *trans* of β-carotene was the major compound regardless of the sample analyzed. In a second series of experiments, PCA was applied to the chromatographic data set revealing patterns of similarity of carotenoid composition among the studied genotypes.

These findings corroborate the results previously found by UV-Vis scanning spectrophotometry taking into account the fingerprint region of carotenoids (i.e. 400-500nm). Fig.6 depicts the grouping of genotypes after calculation of the principal components from the RP-HPLC quantification of carotenoids. PC1 and PC2 explain 97.8% of the total variance of the sample population under study.

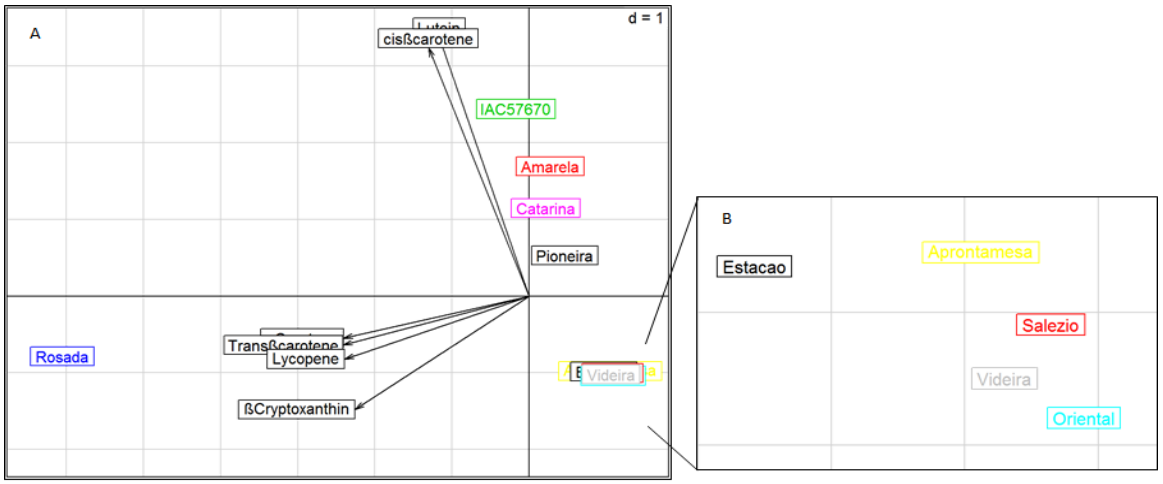

Figure 6: A - Scores scatter plot (PC1 and PC2) of the quantitative data of carotenoids determined by RP-HPLC in root samples of ten cassava genotypes (n = 3). B - Magnification to the overlapping samples at the PCA.

The genotypes with yellow-fleshed roots (*Pioneira*, *Amarela*, *Catarina*, and *IAC-576-70*) were grouped in PC2+, influenced by their higher concentration of *cis*-β-carotene and lutein. Inversely, the genotypes with cream-fleshed roots (*Apronta mesa*, *Oriental*, *Salézio*, *Estação*, and *Videira*) grouped in PC1 +/PC2 – due to their lower amounts of these pigments. Samples of red-fleshed roots (i.e., *Rosada*) showed higher dissimilarity among the studied genotypes, grouping into PC1/PC2 -. This result seems to be directly influenced by the presence of lycopene and the higher concentrations of *trans*-β-carotene, α-carotene, and β-cryptoxanthin. Finally, hierarchical cluster analysis was applied to the chromatographic data, affording similar results to the UV-Vis scanning spectrophotometry for the fingerprint region of carotenoids. Genotypes with the highest similarity in their carotenoid composition are represented by cluster hierarchical analysis in Fig.7. The similarities were defined based on the Euclidean distance between two samples, using the arithmetic average (UPGMA) to compute the resulting tree, the obtained cophenetic correlation was of 97.61%.

In order to extract latent information from the UV-visible dataset, classification models were built by applying supervised classification and feature selection methods. The methods used in

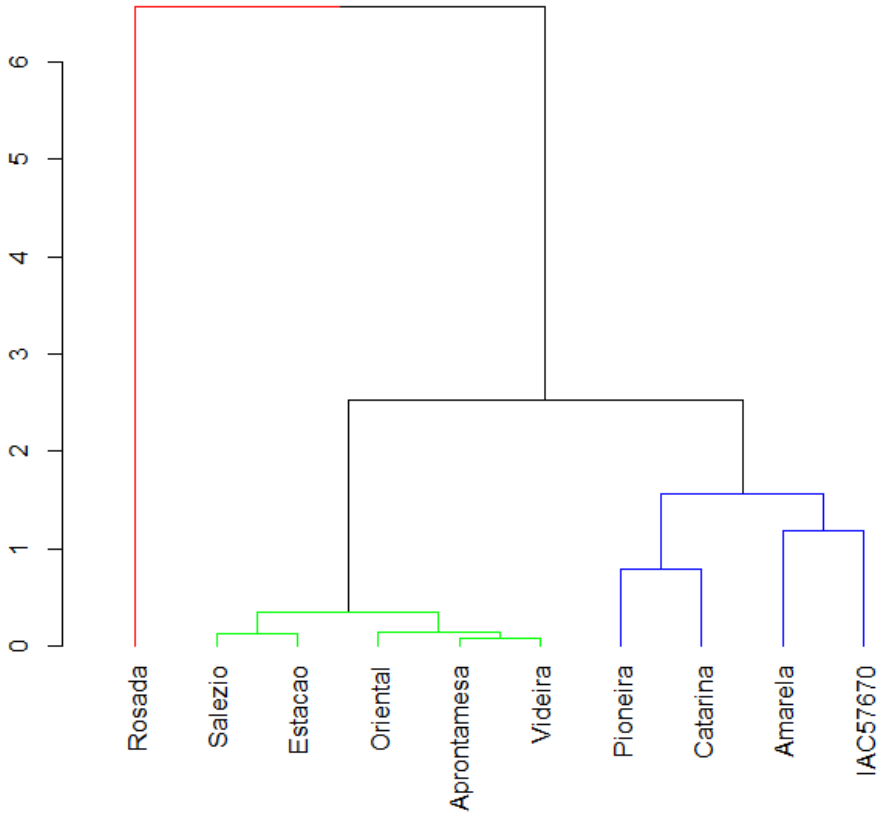

Figure 7: Dendrogram of cassava genotypes in respect to their carotenoid composition determined by RP-HPLC, followed by hierarchical clustering analysis (UPGMA method - 97.61% of cophenetic correlation). The similarity among genotypes of the same cluster is statistically significant ($p \leq 0.05$) when the branches in the dendrogram show the same color. Significance determined by Simprof analysis (Similarity Profile Analysis) from R Clustsig package in accordance with Clarke, Somerfield & Gorley (2008). [12]

this work were PLS-DA (Partial Least Squares Discriminant Analysis) and RF (Random Forest). The quantitative measure of the performance for PLS-DA and RF classification models showed accuracy values around 82% and 66%, respectively, revealing a good performance of the methods, mainly PLS-DA. The cross-validation for the PLS and RF models (10 fold, repeated 10 times) expressed by the confusion matrix showed that the method is able to differentiate the genotypes studied. Besides, taking into consideration the influence of the variables used by the models in the classification, the wavelength characteristic of carotenoids were again the most expressive to validate and distinguish the studied materials (data shown on report). Such findings demonstrate that the analytical approach used in this study is useful for distinguishing the genotypes of a population sample in a germplasm bank, for example.

4 Conclusions

The data set obtained by the analytical techniques employed in this work allowed a better understanding of the chemical variability associated with roots' carotenoid composition of the cassava genotypes. The large disparity in carotenoid contents reveals the chemical variability among the genotypes analyzed. Substantial amounts of carotenoids were detected in the cassava genotypes, indicating their potential as source of interesting compounds to human health and nutrition, given the presence of pro-vitamin A carotenoids (β-carotene, e.g.) and lycopene in roots of yellow and red color, respectively. The *Rosada* genotype was found to be discrepant because its richness in the carotenoids, in addition to the presence of lycopene in relevant amounts. The information obtained by coupling the analysis of biochemical markers for pro-vitamin A in cassava genotypes to bioinformatics tools revealed to be relevant for the rational design of biochemically-assisted cassava breeding programs. Indeed, the analytical approach adopted (i.e., UV-Vis/RP-HPLC/chemometrics) allowed to discriminate and classify the claimed genetic variability of the studied samples based on their biochemical traits, helping to identify/select cassava genotypes of interest to human health and nutrition.

Acknowledgements

To FAPESC (Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for financial support. The research fellowship from CNPq on behalf of the last author is acknowledged. The work is partially funded by Project PropMine, funded by the agreement between Portuguese FCT and Brazilian CNPq. The authors also thank the FCT Strategic Project of UID/BIO/04469/2013 unit.

References

[1] Who. Global prevalence of vitamin A deficiency in populations at risk 1995-2005. *World Health Organization Global Database on Vitamin A Deficiency*, 2009.

[2] D. Rodriguez-Amaya and M. Kimura. HarvestPlus Handbook for Carotenoid Analysis. *HarvestPlus Technical Monographs*, page 59, 2004.

[3] A. C. Kurilich and J. A. Juvik. Quantification of carotenoid and tocopherol antioxidants in Zea mays. *Journal of Agricultural and Food Chemistry*, 47(5):1948–1955, 1999.

[4] H. Tapiero, D. M. Townsend and K. D. Tew. The role of carotenoids in the prevention of human pathologies. *Biomedicine and Pharmacotherapy*, 58(2):100–110, 2004.

[5] S. Voutilainen, T. Nurmi, J. Mursu and T. H. Rissanen. Carotenoids and cardiovascular health 1 – 3. *The american journal of clinical nutrition*, 83:1265–1271, 2006.

[6] D. B. Rodriguez-Amaya. *A Guide to Carotenoid Analysis in Foods*. 2001.

[7] C. Iglesias, J. Mayer, L. Chavez and F. Calle. Genetic potential and stability of carotene content in cassava roots. *Euphytica*, 94(3):367–373, 1997.

[8] R Core Team. R : A Language and Environment for Statistical Computing. 1, 2014. URL `http://www.r-project.org/`.

[9] B. A. Hanson. ChemoSpec : An R Package for Chemometric Analysis of Spectroscopic Data and Chromatograms (Package Version 2 . 0-2). pages 1–41, 2014.

[10] C. Beleites. Import and Export of Spectra Files. pages 1–20, 2011.

[11] H. Wickham and W. Chang. Package 'ggplot2'. *The Comprehensive R Archive Network*, 2015.

[12] K. R. Clarke, P. J. Somerfield and R. N. Gorley. Testing of null hypotheses in exploratory community analyses: similarity profiles and biota-environment linkage. *Journal of Experimental Marine Biology and Ecology*, 366(1-2):56–69, 2008.