Proton quantitative nuclear magnetic resonance and principal component analysis as tool for discrimination of guarana seeds from different geographic regions of Brazil

Guarana (Paullinia cupana) is an Amazonian native fruit, whose seeds are rich in caffeine and phenolic compounds. Due to the economic importance of guarana, it is necessary to improve the knowledge of its chemical characteristics, which may vary depending on several factors, such as soil characteristics, crop variety, and environmental conditions. In this context, this study describes the application of quantitative nuclear magnetic resonance associated with principal components analysis to evaluate the composition of guarana seeds from the two largest Brazilian producer regions: Amazonas and Bahia states. The principal component analysis discriminated the samples according to the source, revealing that the seeds from Bahia contain higher caffeine and phenolic contents (catechin, epi-catechin), but the seeds from Amazonas present higher content of fatty acids. Additionally, caffeine was quantified using quantitative nuclear magnetic resonance and liquid chromatography coupled to photodiode array detector, which corroborate the chemometrics. The results show a slightly larger content of caffeine for seeds from Bahia. Therefore, $q$NMR combined with chemometrics was important to detect the metabolic variability of guarana seeds from different regions.

Introduction

Paullinia cupana (Sapindaceae), also known as guarana, is a native plant from Amazon (Brazil), where was traditionally used by ancient indigenous people from Sateré-Mawé tribe in the preparation of a drink with medicinal properties (e.g. diuretic and for treating headache, fever, and cramps) as well as a stimulant drug. The Paullinia genus accounts for one hundred and ninety five species, besides many varieties. Until the 1970’s, guarana was produced on a commercial scale only in Maués county (Amazonas state). However, in 1980’s the Brazilian government encouraged the cultivation of guarana in the states of Bahia, Mato Grosso, Pará, Acre, and Rondônia. The most successful case of adaptation was in Bahia state that, currently, is the largest producer of guarana.

The guarana is commercialized in four different forms: as raw guarana (roasted grain), guarana stick (raw crushed guarana, dried, smoked and shaped as stick), guarana powder (roasted and ground seeds), and syrup (extract from guarana seed), which is the usual commercial form. About 70 % of guarana is utilized by the soft drink industry and the 30 % remaining are exported or used for domestic consumption, mainly as dietary supplement. Guarana seed is rich in caffeine (up to 7 %), presenting 4 to 6 times higher content than found in coffee beans, tea leaves, and cola nuts. The seeds also contain high concentration of polyphenols, especially proanthocyanidins as well as other methylxanthines (theophylline and theobromine), terpenes, epi-catechin, and starch. As these metabolites arise from the plant secondary metabolism, their levels can change in response to disease, genetic disorders or environmental conditions.

The “digital printing” or “fingerprint” techniques have become one of the most powerful approaches to food quality control, providing chemical information to characterize, for instance, the so-called geographical origin indication. Currently, several scientific studies have been published on the development of fingerprinting and geographical indications for food using various techniques. Traditionally, NMR spectroscopy is used for identification and elucidation of molecules, since it allows the unequivocal characterization of substances in all physical states. Nowadays, it has been used as a useful tool for analysis of complex mixture, such as plant metabolites and food for quality control. However, NMR generates highly complex spectroscopy matrices with elevated similarity, which makes infeasible the visual analysis of the data. Therefore, chemometrics can be used to solve chemical problems based on a large number of samples and measured variables. The aim of this work was to characterize the main compounds found in guarana seeds, and to differentiate guarana seeds produced in different geographical areas of Brazil. In addition, caffeine (major bioactive compound found in guarana) was quantified using quantitative nuclear magnetic resonance ($q$NMR) and high performance liquid chromatography coupled to photodiode array detector (HPLC-PDA) in order to attest its variation indicated in chemometric models.

Experimental Sample

Guarana seeds (Figure 1a) were collected from the two main production states of Brazil: Amazonas and Bahia (highlighted in Figure 1b). The samples were processed according to the “Serviço Brasileiro de Respostas Técnicas – SBRT (2008)” guidelines with modifications. The fruits of guarana were fermented for two days into plastic bags at room temperature. Thereafter, the seeds were...
wash nitrogen and was transferred to 5 mm NMR tubes. The NMR experiments were acquired on an Agilent DD2 600-MHz spectrometer (14.1 T), equipped with 5 mm (H-F/OD-d4) inverse detection One ProbeTM with actively shielded z-gradient. The 1H NMR experiments were acquired under quantitative conditions: 90° pulse calibrated to 7.45 μs; acquisition time of 3.3 s, and relaxation delay of 23.0 s. It was acquired 32 transients in a spectral width of 10.0 ppm, and the supernatant was transferred to 5 mm NMR tubes. The NMR experiments were acquired with the compounds identified. Legend: GABA – gamma-aminobutyric acid; Caffeine – caffeine; GABA – gamma-aminobutyric acid; and Malic acid – malic acid.

Results and discussions

Characterization

The 1H NMR spectrum of the guarana seeds is displayed in Figure 2. The extraction method (aqueous methanol) was used in order to favor the extraction of primary and secondary metabolites from the seeds. Therefore, signals were noticed from corresponding low molecular weight primary metabolites such as sugars, organic acids, fatty acids, and cyanolipids, as well as the secondary metabolites caffeine, cathechin, and epi-catechin. In the high field region it was observed the characteristic signals of metabolites as fatty acids (1.31 ppm for \(-\text{CH}_2\); 0.88 ppm for \(-\text{CH}_3\); and 5.42 ppm for \(-\text{CH}ceil\rceil\rceil_2\) hydrogen); cyanolipids (1.31 ppm for \(-\text{CH}_2\); 0.88 ppm for \(-\text{CH}_3\); 4.85 ppm for \(-\text{CH}_3\) of the acylglycerol moiety; and 5.42 ppm for methylene hydrogen); valin (1.02 ppm for \(-\text{CH}_2\); and 1.07 ppm for the other \(-\text{CH}_3\) ethanol (1.24 ppm for \(-\text{CH}_2\); threonine (1.33 ppm for \(-\text{CH}_2\); alanine (1.49 ppm for \(-\text{CH}_2\); gamma-aminobutyric acid named as GABA (alpha- \(-\text{CH}_2\); at 2.3 ppm, and gamma- \(-\text{CH}_2\); at 3.02 ppm); acetic acid (1.93 ppm for \(-\text{CH}_3\); cathechin and epi-catechin with \(-\text{CH}_2\) groups at region from 2.50 to 2.90 ppm; malic acid with the \(-\text{CH}_2\) group overlapped with catechin and epi-catechin signals. At the sugar region was observed the characteristic signal of malic acid (4.42 ppm for \(-\text{CH}_2\); \(-\alpha\)-glucose (5.23 ppm for \(-\text{CH}_2\); and was reduced back to 0 % of B and held there for 5 min. Quantification was performed with an external calibration curve of caffeine ranging from 10.0 mg.L\(^{-1}\) to 100.0 mg.L\(^{-1}\). The results were expressed as percentage of caffeine per grams of sample.

Figure 1. a) Paulinia cupana (guarana) fruit. Source: Embrapa archive; b) geographic regions of guarana sampling.

Figure 2. 1H NMR spectrum of guarana seeds showing signals labelled with the compounds identified. Legend: GABA – gamma aminobutyric acid.
Chemometrics

In order to identify potential relationships between the geographical origin and chemical composition of guarana samples, the \(^1\)H NMR spectra were used as input data for PCA analysis. Figure 3a presents the biplot PC1 vs. PC2 showing the separation of the seeds according to the geographical origin. The samples were symbolized regarding the region: Amazonas as gray square; south of the state of Bahia as circle; Baixo sul region as black asterisks, and commercial from Bahia as triangle. At positive scores of PC1 (48.3 % of the total variance) are located the majority of the samples collected in Amazonas. On the other hand, the samples collected at Bahia (from Baixo Sul region and the south of Bahia) are located in negative scores of PC1. It was also observed that the commercial samples (dark triangles) are distributed all over the PCA graphic. This might indicate the poor quality control of the powder, with a mixture of powders from Bahia and Amazonas.

The loadings graph of PC1 (Fig. 3b) presented seeds from Bahia with highest amount of carbohydrates, phenolic compounds as catechin and epi-catechin, besides the caffeine. In addition, the samples from Amazonas exhibited the highest amount of fatty acids and cyanolipids. These differences were clearly observed on the expansion of the \(^1\)H NMR spectra (from 0.6 to 3.2 ppm) of the samples (Figure 3 c, d, and e). This finding indicated that the edaphoclimatic conditions of Bahia (tropical eastern north climate) when compared with Amazon (equatorial climate) seem to have ideal conditions for guarana cultivation (mostly higher phenolic content). Additionally, the temperature (°C) and total rainfall (mm) of the sampling period (September of 2013 to February of 2015)\(^{11}\) was examined. The variance analysis showed significant differences for the temperature wherein Maués-area presented higher temperatures than Bahia-area while rainfalls in both regions were quite similar.

The temperature difference might influence the chemical composition of guarana seeds.

Additionally, after the evaluation of the Figure 3c and Figure 3d, it was noticed that the seeds from Baixo sul region (Fig. 3c) and from south of Bahia (Fig 3d) also presented variations on the chemical composition. Therefore, the PCA was also performed for detailed evaluation of these samples. The biplot PC1 vs. PC2 presented in Figure 4a shows a tendency of separation of the guarana seeds according to the geographic origin, with samples from Baixo sul region of Bahia (star symbol) at negative scores of PC1, and samples from south of Bahia (circle symbols) at positive scores of PC1. In the PC1 loadings plot (Figure 4b) it is possible to observe that the seeds from Baixo sul region of Bahia presented higher amount of the phenolic compounds, while samples from south of Bahia contained higher amount of carbohydrates. The variance analysis for climate data was also examined showing similarity between the climates for both regions, which shows that this variability can be associates to other parameter beyond the edaphoclimatic condition.

Quantification of caffeine

About 70 % of the guarana production is absorbed by the beverage industry, located in Amazon state\(^2\). This high demand is driven by the energetic properties of the seed that is mostly attributed to the elevated caffeine content. Therefore, in order to obtain the content of caffeine for the different regions, \(^1\)H qNMR and HPLC-UV were employed to quantify caffeine in guarana seeds. The Figure 5 shows the box-and-whisker graphs for the quantification of caffeine.

The caffeine content obtained from HPLC-UV method ranges from 1.5 to 5 % per gram of seed while through qNMR, the caffeine content ranges from 1 to 3.5 % per gram of seed. Both methods were in accordance with publish data that shows that the caffeine ranges from 2.5 to 6 %\(^2\). However, the quantification using the HPLC-UV method presented values higher than the ones obtained through the \(^1\)H qNMR. This occurs because the HPLC-UV experiments were performed with samples that undergo exhaustive extraction yielding up to 98 % of caffeine extraction (data not shown). In addition, according to statistical analysis by ANOVA for
Discrimination of Guarana Seeds from Different Geographic Regions of Brazil

Proceedings of the XIII International Conference on the Applications of Magnetic Resonance in Food Science 2016

www.impublications.com/mrfss2016

1H qNMR, the caffeine contents found in the samples from Amazonas and Bahia are significantly different with 95 % of significance level. However, the HPLC-UV results showed that the caffeine content from both regions is statistically equal, with the same significance level. The higher dispersion of the data obtained by HPLC-UV induces a tendency of similarity among the values. On the other hand, the results obtained by 1H qNMR has lower deviation given a small discrepancy among the samples highlighting the differences of caffeine content. Therefore, the caffeine content of samples from different geographic origin is statistically different corroborating the chemometrics.

Conclusions
Chemometric analysis discriminated the samples according to the geographic origin. Guarana seeds from Bahia state presented the highest phenolic content (catechin and epi-catechin), and guarana seeds from Amazon state presented the highest content of fatty acids and cyanolipids. Based on the qNMR results (less dispersed data), the content of caffeine from Bahia and Amazonas are significantly different. Therefore, as the methodologies were performed with biological replicates and qNMR corroborate the chemometrics, this study emphasizes the remarkable advantage of 1H qNMR for food analysis based on the identification and quantification of compounds in complex mixtures without the use of standards.

References

Figure 4. a) PC1 × PC2 of guarana seeds extract from the Bahia state; and b) loadings graph plotted in line from PC1 axis (top) and 1H NMR spectra (bottom) to illustrate the NMR regions.

Figure 5. Box and whisker graph for quantification of caffeine through HPLC and qNMR. Legend: AM – Amazon (Maués); BS – Baixo Sul region of Bahia; CO – Commercial from Bahia; SB – South of Bahia.


