Food Research International 61 (2014) 75–82

Contents lists available at ScienceDirect

Food Research International

journal homepage: www.elsevier.com/locate/foodres

Potential markers of coffee genotypes grown in different Brazilian regions: A metabolomics approach

José Henrique da Silva Taveira [a,*], Flávio Meira Borém [a], Luísa Pereira Figueiredo [a], Nádia Reis [b], Adriana S. Franca [b], Scott A. Harding [c], Chung-Jui Tsai [c]

[a] PPGEA, Universidade Federal de Lavras, Campus Universitário UFLA, S/N, 37200-000 Lavras, MG, Brasil
[b] DEMEC, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, 31270-901 Belo Horizonte, MG, Brasil
[c] Warnell School of Forestry and Natural Resources, Department of Genetics, University of Georgia, Athens, GA, USA

ARTICLE INFO

Article history:
Received 30 November 2013
Accepted 23 February 2014
Available online 12 March 2014

Keywords:
Metabolite profiling
Coffea arabica L.
GC-Q/MS
PLS-DA

ABSTRACT

Seeds from different coffee species and coffee from different continents or countries have very distinct chemical composition. However, the differences between genotypes grown at micro-regional levels with similar geographical characteristics are still unclear. In this study, we highlighted the need of using metabolite profiling instead of the usual targeted analysis as a more powerful tool to describe the slight differences between coffees of the same species grown in close origins. Thus, our study focused on finding potential metabolite markers to describe differences of *Coffea arabica* L. genotypes (Mundo Novo and Bourbons) grown in Brazilian coffee producing municipalities (Lavras, Santo Antônio do Amparo—SAA, and São Sebastião da Grama—SSG). Using the metabolomics approach, 44 metabolites were identified, and some showed great potential for origin and genotype differentiation. The Partial Least Square Discriminant Analysis — PLS-DA model showed that the SAA coffee samples had the most differentiated metabolite profile (approximately 95% accuracy) compared to the other municipalities. The samples from Lavras and SGG had similar profiles (model accuracy of approximately 50%). Potential metabolite markers for the SAA samples included galactinol, fructose, malic acid, oxalic acid, D-glucose, D-sorbitol, galactinol, and myo-inositol. The model used to differentiate the Bourbon and the MN genotype showed 100% accuracy indicating very different metabolite profiles. The features that were most influential in differentiating genotype were: 5-CQA, oxalic acid, galactinol, nicotinic acid, caffeine, and caffeic acid (Bourbon) and myo-inositol, quinic acid, malic acid, fructose, and D-glucose (MN). Enhancing subtle differences in the data by combining information from GC-Q/MS and multivariate analysis resulted in the identification of coffee origins and genotypes as well as the identification of potential markers.

1. Introduction

Coffee is one of the most common commercially traded commodities in the world. The demand for specialty coffees in the market has been growing much greater compared to regular coffee and are characterized by high quality, great flavor potential after roasting, absence of any defects, and their relationship to origin, crop or specific genotypes. Primarily, coffee quality has been determined by genetic (Villarreal et al., 2009), environmental (Alonso-Salces, Serra, Reniero, & Héberger, 2009; Avelino et al., 2005; Joët et al., 2010), and post-harvest (Bytof et al., 2007; Duarte, Pereira, & Farah, 2010; Joët, Laffargue, et al., 2010; Joët, Salmona, Laffargue, Descroix, & Dussert, 2010; Knopp, Bytof, & Selmar, 2005) methods. Although Brazil is the largest worldwide coffee producer, its coffee is known as regular and flat. Conversely, coffees originating from different areas, such as Central America, Africa, and Asia, are very well known for their high quality and their sensory attributes. Coffee beans originating from these areas have very different metabolite profiles (Choi, Choi, Park, Lim, & Kwon, 2010). Because these regions are located in different parts of the world, it is expected that they would have quite different metabolite profiles due to significant environmental differences. *Coffea arabica* L. and *Coffea canephora* species also have large differences in their chemical composition (Alonso-Salces et al., 2009; Wei et al., 2012). Therefore, we highlighted the need for analyses that are capable of discerning samples from origins that are geographically very close to each other and between genotypes from the same species. Studies have already been conducted comparing coffees from two different locations, with different altitude levels and slope exposures in Costa Rica (Avelino et al., 2005). Although the results were obtained from coffee samples harvested during one

Abbreviations: 4-CQA, 4-caffeoylquinic acid; 5-CQA, 5-caffeoylquinic acid; BFP, Bourbon Fazenda Paixão; BFT, Bourbon Fazenda Toriba; BIACJ9, Bourbon IACJ9; CGAs, chlorogenic acids; GC-Q/MS, gas chromatography quadrupole mass spectroscopy; GLSW, generalized least squares weighting; HPLC, high performance liquid chromatography; MC, mean centering; MN, Mundo Novo; NMR, nuclear magnetic resonance; PCA, principal component analysis; PLS-DA, partial least-squares discriminant analysis; SAA, Santo Antônio do Amparo; SSG, São Sebastião da Grama.

* Corresponding author. Tel.: +55 35 38291596; fax: +55 35 38291362.
E-mail address: henriquetaveira@yahoo.com.br (J.H. da Silva Taveira).

http://dx.doi.org/10.1016/j.foodres.2014.02.048
0963-9969/

agricultural crop season (2003), consistent differences in their chemical and sensory profiles were reported. These results indicated that some of the most influential metabolites used to differentiate between the geographical locations included caffeine, trigonelline, chlorogenic acids, and sucrose as potential markers. In addition, Bertrand et al. (2008), used chemometric analysis to differentiate between coffee varieties grown in different *terroirs* and harvested in one agricultural crop; they found that fatty acids and chlorogenic acids are consistent markers that can be used to differentiate coffee samples.

Some studies used to describe metabolic pathways of coffee seeds have focused on environmental conditions (altitude, temperature, rainfall, and irradiance) during one agricultural crop season (2006) at the Reunion Islands (Joët, Laffargue, et al., 2010; Joët, Salmona, et al., 2010; Joët et al., 2009). In addition to describing environmental interference on coffee seed metabolism, these studies have determined consistent and important pathways for chlorogenic acid and polysaccharide formation.

A recent study has been developed by Figueiredo et al. (2013), which describes the potential for quality of coffee genotypes grown in different Brazilian regions. Their results involved sensory and chemical attributes that were consistent, even though the coffee samples were harvested in one agricultural crop season.

In addition, Alonso-Salces et al. (2009) used chemometric tools for the botanical and geographical characterization of green coffee (*C. arabica* L. and *C. canephora*) during three different crop seasons. These authors also concluded that there were no significant differences in the chemical composition of the coffee beans between years of harvest.

A better understanding of the chemical composition of coffee genotypes settled in different sites at Brazil could bring new challenges and provide motivation for changes in the production chain. Additionally, establishing marker metabolites will help with the classification of coffee origins, which has recently become a central component for agricultural promotion. Additionally, these markers will provide additional support in fighting against forgery and coffee adulteration at the regional, national and international levels.

Metabolomics is a comprehensive and non-target analysis that covers a broad range of metabolites; it plays a key role in describing precursor compounds responsible for quality and improving strategies for coffee breeding programs. These chemical precursor compounds release more than 800 new aromatic constituents (Leroy et al., 2006) that affect the flavor of the roasted coffee beans. Because of the large number of chemical compounds that can be found in coffee beans, quality cannot be attributed to a single compound or a specific environmental parameter. Differences in metabolite profiles are complicated and not easy to explain. Many more studies are needed to understand the relationship between quality and metabolite profiles.

Recently, it has become easier to distinguish materials from different origins in the world or different species using sophisticated metabolite profiling tools, such as HPLC (Alonso-Salces et al., 2009), GC-Q/MS (Du, Wang, Yu, Liu, & Huang, 2011; Frost, Nyamdari, Tsai, & Harding, 2012) and NMR (Wei, Furihata, Hu, Miyakawa, & Tanokura, 2010; Wei et al., 2012). Because rapid analytical methods are needed to determine coffee origins, this study focused on showing the potential of using metabolite profiling of coffee seeds coupled with multivariate analysis, PCA and PLS-DA, as techniques to identify coffee samples according to their botanical and geographical origins in Brazil. We also focused on finding key metabolites that are responsible for such differences.

2. Materials and methods

2.1. Experimental site, biological material and processing

Experiments were performed on seeds of four *C. arabica* L. genotypes, i.e., MN and Bourbons BIACJ9, BFT, and BFP, grown in three different coffee producing municipalities of Brazil (Table 1). Three independent biological samples of coffee fruits were selectively harvested and chosen if completely ripe to ensure their uniformity, integrity and high quality. Then, the samples were processed using the wet method and dried at 11% (w.b.) according to methods established by Borém, 2008. After drying, the seeds were frozen, ground using an IKA A11 Basic Analytical mill, lyophilized, and then kept under −80 °C for further extraction.

Table 1
Location of coffee plots used as experimental sites.
Adapted from Figueiredo et al. (2013).

Growth location	Latitude	Longitude	Elevation
Lavras	21°14′43″S	44°59′59″W	950 m
SAA	20°56′47″S	44°55′08″W	1050 m
SSG	21°44′50″S	46°55′33″W	1300 m

2.2. Metabolite analysis

In total, 10 mg of the lyophilized tissue powder was extracted twice in 515 μl aqueous methanol 60% (v/v) in a 2.0 ml microcentrifuge tube containing internal standards (0.18 μl of 45 $mg \cdot ml^{-1}$ adonitol and 0.37 μl of 0.59 $mg \cdot ml^{-1}$ $^{13}C_6$ trans-cinnamic acid). The microcentrifuge tubes were incubated in a water bath set to 70 °C for 1 min and then transferred to a dry plate set to 70 °C while they were mixed. The mixtures were cooled down to room temperature and centrifuged at 2199 $rad \cdot s^{-1}$ for 5 min at room temperature. The supernatant from the first extraction phase was transferred into a new tube. The powder remaining in the tube was mixed with 515 μl aqueous methanol 60% (v/v), mixed for 15 min using sonication, and centrifuged at room temperature. The supernatant was added to the first extraction phase, and the total extract mixture was complete. A subsample (50 μl) of each extract was transferred to a glass micro-insert, taken to the derivatization, and then analyzed using GC-Q/MS according to the methodology described by Frost et al. (2012). Finally, data peaks obtained from the analysis of each sample were aligned using an in-house software program (MetaLab), which is available at http://aspendb.uga.edu.

2.3. Data analysis

The dataset was pretreated before the chemometric models were constructed to correct possible variations in the spectra that were not related to the samples' nature. These included variations in analysis conditions, such as environmental temperature and relative humidity, apparatus type, personnel, sample positioning in the equipment, and all possible perturbations. Thus, GLSW (alpha = 0.01) and MC were used to preprocess each class of dataset prior to the model calculation. The GLSW technique uses auto vectors and auto values of a covariance matrix to lower and smooth signals of inference or sample differences that should be equal (Wise et al., 2006). MC was used to prevent the most distant values from having a greater influence on the results compared to data values relatively close to each other. Pretreatment on the dataset was performed using MATLAB 7.13 (MathWorks™, MA, United States) and PLS Toolbox 6.5 (Eigenvector Research, Inc., WA, United States). The pretreated datasets were analyzed by PCA and PLS-DA, using the online software MetaboAnalyst (Xia, Mandal, Sinelnikov, Broadhurst, & Wishart, 2012; Xia, Psychogios, Young, & Wishart, 2009), which is available at http://www.metaboanalyst.ca/MetaboAnalyst/faces/Home.jsp. PLS-DA is a variation of PLS analysis. PLS-DA is considered a pair comparison analysis and is built to classify a group of samples as belonging or not belonging to a specific class (Berrueta, Alonso-Salces, & Héberger, 2007). The results from PLS-DA analysis are obtained by creating PLS regression model, from the original vector y, which contains dummy variables.