Comparative Proteomical Analysis of Zygotic Embryo and Endosperm from Coffea arabica Seeds
Li#via L. Koshino, Clarissa P. Gomes, Luciano P. Silva, Mirian T. S. Eira, Carlos Bloch Jr., Octa#vio L. Franco, and A#ngela Mehta


Downloaded from http://pubs.acs.org on November 28, 2008

More About This Article

Additional resources and features associated with this article are available within the HTML version:

• Supporting Information
• Access to high resolution figures
• Links to articles and content related to this article
• Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML
Comparative Proteomical Analysis of Zygotic Embryo and Endosperm from Coffea arabica Seeds

L´IVIA L. KOSHINO,† CLARISSA P. GOMES,§ LUCIANO P. SILVA,# MIRIAN T. S. EIRA,⊥ CARLOS BLOCH JR.,# OCT ´AVIO L. FRANCO,§,‡ AND ANGELA MEHTA*

Universidade de Brası´lia; Embrapa Recursos Genetı´cos e Biotecnologia, Embrapa Café; Centro de Ana´lises Proteoˆmicas e Bioquı´micas, Programa de Pós-Graduac¸a˜o em Ciências Genoˆmicas e Biotecnologia, Universidade Católica de Brası´lia, Brası´lia, Brazil and Universidade Federal de Juiz de Fora, Juiz de Fora, Minas Gerais, Brazil

During coffee seed development, proteins are predominantly deposited in cotyledons and in the endosperm. Reserve proteins of the 11S family are the most abundant globulins in coffee seeds, acting as a nitrogen source during roasting and guaranteeing flavor and aroma. The aim of the present study was to compare the protein profiles of endosperm and zygotic embryos of coffee seeds. Proteins were extracted from whole seed as well as from embryo and endosperm, separately. Total proteins were analyzed by two-dimensional electrophoresis (2-DE) followed by identification by mass spectrometry (MS). The most abundant spots observed in the gels of coffee seeds were excised, digested with trypsin, and identified by MS as subunits of the 11S globulin. Spots with identical pI and molecular masses were also observed in the protein profiles of coffee endosperm and embryo, indicating that 11S protein is also highly expressed in those tissues. Peptide sequence coverage of about 20% of the entire 11S globulin was obtained. Three other proteins were identified in the embryo and endosperm 2D profiles as a Cupin superfamily protein, an allergenic protein (Pru ar 1), exclusive to the endosperm 2D map, and a hypothetical protein, observed only in the zygotic embryo profile.

KEYWORDS: 2-DE; mass spectrometry; zygotic embryo; endosperm; Coffea arabica

INTRODUCTION

Coffee is one of the most important agricultural products worldwide and is cultivated in more than 60 countries. Brazil is responsible for more than one-third of the global coffee production and exportation and, together with Vietnam and Colombia, accounts for about 50% of the world production. The main cultivated coffee species are Coffea arabica and Coffea canephora. Several studies of protein/gene expression have been performed mostly related to flavor and aroma in these species (1–6).

Coffee seeds contain specialized storage tissues (cotyledons and endosperm) in which proteins are formed and deposited during maturation. After development, these proteins constitute a major seed protein fraction. Seed storage proteins comprise groups of multiple isoforms encoded by gene families, including globulins, albumins, glutelins, and prolamins (7). These proteins have shown multiple functions such as storage (8) as well as a key role in the plant defense system (9, 10). Globulins constitute the most widely distributed group of seed proteins, which are stored as hexameric structures (11). In coffee, the 11S globulin has been extensively studied by using genomic and proteomic approaches (5, 12, 13). This protein is formed by two subunits linked by a disulfide bond: a large acid α subunit and a small basic β subunit derived from a single 11S precursor, which exists as multiple isoforms (7, 14). These abundant proteins are probably important for the quality of the beverage, because they act as the main source of nitrogen during roasting (15, 16).

Proteomic analysis has been increasingly employed to better understand biological processes. By using this approach, a high number of proteins expressed in a specific tissue or under certain biological conditions can be visualized, allowing the determination of quantitative data regarding protein abundance, as well as qualitative differences related to protein mass and isoelectric point (pI). Currently, proteomic analyses are being performed by coupling two-dimensional electrophoresis (2-DE) and high-throughput mass spectrometry (MS). In 2-DE, proteins are separated by protein mass and pI, and therefore a high resolution can be obtained. Protein spots can be further analyzed by MS and identified by peptide mass fingerprinting (PMF) or de novo
sequencing. Peptide sequencing gives more detailed information and allows the identification of proteins from organisms with noncharacterized or nonpublic genomes, which is the case of the Coffee Genome database. Proteomic analyses have been applied for the study of coffee grains (5, 6, 17); however, still little is known regarding coffee seed proteins accumulated in different tissues, especially in the embryo. In summary, this paper aims to compare the global protein profiles of coffee zygotic embryo and endosperm by two-dimensional gel electrophoresis (2-DE) and mass spectrometry (MS).

MATERIALS AND METHODS

Plant Material. *C. arabica* (cultivar Catuaí Vermelho IAC-99) fruits were collected 210–220 days after anthesis, corresponding to the stage when the germinability is acquired (13, 18). Seeds were isolated from the fruits and further separated into embryo and endosperm. Whole seeds were also used for comparison, and all tissues were stored at −80 °C.

Protein Extraction. Whole seed, endosperm, and embryo samples collected from two consecutive years were separately homogenized in a liquid nitrogen precooled mortar by using a pestle, and total proteins were extracted according to the method described by de Mot and Vanderleyden (19). Approximately 0.1 g of the resulting tissue powder was put into a 1.5 mL tube with 0.75 mL of extraction buffer consisting of 0.7 M sucrose, 0.5 M Tris-HCl, 30 mM EDTA, 0.1 M KCl, and 40 mM DTT. An equal volume of buffer-saturated phenol (Invitrogen) was added, and after 15 min of vigorous shaking, the sample was centrifuged and the supernatant recovered. Two additional rounds of extraction were performed using an equal volume of extraction buffer each time. Proteins in the phenol phase were precipitated in 5 volumes of chilled 0.1 mM ammonium acetate in methanol at −20 °C for 2 h. Precipitates were washed in acetone and resuspended in 40 µL of lysis buffer [9.8 M urea, 0.2% (v/v) Nonidet P-40 (Sigma), 100 mM DTT, and 2% (v/v) of a mixture of ampholytes pH 5–8 and pH 3–10 (Bio-Rad) in the ratio of 5:1]. The total protein content was estimated according to the method of Bradford (20).

Two-Dimensional Electrophoresis. Isoelectric focusing of coffee proteins was performed according to the method of de Mot and Vanderleiden (19). Polyacrylamide gels containing 3.6% acrylamide, 0.21% bisacrylamide, 7.2% ampholyte pH 5–7 and 3–10 in the proportion of 5:1 (v/v), 2% Nonidet P-40, and 55% urea were used. Approximately 150 µg of proteins was loaded onto the 11 cm gel after a prerun. Electrophoresis was performed in a vertical system (Invitrogen) at 400 V for 18 h, using 20 mM NaOH in the upper compartment of the chamber and 10 mM H3PO4 in the lower. Molecular mass separation was performed according to the method of Laemmli (21) in 18 × 18 cm gels 12%. Electrophoresis was performed for 6 h at 120 V using glycine buffer in the upper and lower compartments of the chamber, and at least four repetitions for each tissue were obtained. After running, 2D gels were fixed overnight in a solution containing 50% ethanol and 12% acetic acid. Silver staining was carried out according to the method of Blum et al. (22). Gels were digitalized using the scanner HP Scanjet model 8290 and further analyzed with BioNumerics software v. 4.5 (Applied-Maths). First, calibration with a gray scale was necessary to transform gray levels into values for each pixel of the gel picture. A calibration curve from BioNumerics software was used, and all gel pictures were analyzed as tiff files. The six gel images were placed in one folder, and the wizard detection method proposed by the software was used for spot detection. Automatically detected spots were manually checked, and some of them were manually added or removed according to size (>0.2 cm), format (circular), and density (>2 pixels cm⁻²). Following the detection procedure, the normalization step was carried out to attribute a common protein identity for identical spots derived from different images. For this procedure, a reference gel was constructed, and automatic matching options of

![Figure 1. Two-dimensional electrophoresis (left) and three-dimensional gel representations (right) of proteins from whole seed, endosperm, and embryo, as indicated.](image-url)
Table 1. MS/MS Protein Identification of Spots from 2D Gels of Whole Seed, Embryo, and Endosperm

<table>
<thead>
<tr>
<th>Spot</th>
<th>Peptide Sequence</th>
<th>Protein Identification</th>
<th>Accession No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1S, Sp12</td>
<td>LNQEPSPFR</td>
<td>11S globulin</td>
<td>AAC61983</td>
</tr>
<tr>
<td>2S</td>
<td>LNAQPSFR</td>
<td>Cupin family protein</td>
<td>NP_195388</td>
</tr>
<tr>
<td>4S</td>
<td>TNDNAMIN</td>
<td>major allergen Pru ar 1</td>
<td>O50001</td>
</tr>
<tr>
<td>9S, Emb 1, Emb 2, Emb 3, Sp28, Sp29</td>
<td>LSENIGLPOEADVFNPR</td>
<td>hypothetical protein</td>
<td>EAZ14610</td>
</tr>
<tr>
<td>Sp20</td>
<td>IPILSSQLSLAER</td>
<td>Cupin family protein</td>
<td>NP_195388</td>
</tr>
<tr>
<td>Sp6</td>
<td>FGVEEGDFAVQR</td>
<td>major allergen Mal d 1</td>
<td>O50001</td>
</tr>
<tr>
<td>Sp31</td>
<td>APVTSYEVFTNAPPR</td>
<td>major allergen</td>
<td>O50001</td>
</tr>
<tr>
<td>Emb26</td>
<td>SSWNSPYDTSYGAG-SGGGGGGG</td>
<td>hypothetical protein</td>
<td>EAZ14610</td>
</tr>
</tbody>
</table>

Bionumerics software was used. For each sample, when a protein was detected in all gel images, this protein was automatically added to the reference gel.

**Mass Spectrometry.** Protein spots were excised from silver nitrate-stained gels, destained, and digested with trypsin according to the method of Shevchenko et al. (23). Peptides were analyzed in MALDI TOF-TOF Ultra Flex II (Bruker Daltonics) MS and de novo sequenced using CID and LIFT acquired MS/MS spectral data. The identification of protein spots was performed by sequence searches in public databases and in the Coffee Genome Project database using the Blastp and tBlastn programs, respectively.

**RESULTS AND DISCUSSION**

In this study, we have analyzed the proteomes of coffee seeds as well as zygotic embryo and endosperm. The main proteins observed in all three protein maps were identified by MS/MS experiments. Initially, the whole seed proteome was analyzed and revealed approximately 70 proteins ranging in mass from 4 to 8. Some of the main proteins observed (1S, 2S, 4S, and 9S) were excised from the gel and analyzed by mass spectrometry. The proteins identified corresponded to the 11S globulin, previously described (5, 6, 12). To verify the proteins expressed in the different compartments of the whole seed, we analyzed the endosperm and zygotic embryo separately by 2-DE. Comparison of all three 2D maps showed similar protein profiles (Figure 1). The 2-DE of embryos and endosperm revealed approximately 120 and 80 protein spots, respectively, showing higher protein diversity in the embryo. Several of the proteins observed in the seed could be visualized in the embryo and endosperm profiles (Table 1). Some of these proteins observed in the embryo (Emb1, Emb2, and Emb3) and endosperm (Sp12, Sp20, Sp28, and Sp29) were also excised from the gel and analyzed by MS to determine whether the proteins corresponded to the seed globulin. The results revealed that all protein spots corresponded to the same globulin 11S protein.

By using a proteomic approach, Rogers et al. (6) previously characterized the 11S globulin in coffee endosperm mainly by NH$_3$ terminal sequencing. In this work, we have analyzed several protein spots by MS/MS analysis and obtained 11S protein coverage of approximately 20% of the whole protein, considering the number of amino acid residues determined (Table 1). Yuffa et al. (17) previously analyzed zygotic and somatic embryos as well as whole seed of C. arabica and showed the expression of abundant polypeptides in all three protein profiles; however, no protein identification was performed. In this study, we identified protein spots from zygotic embryo, and our results show that 11S globulin is also highly expressed in this tissue.

It has been reported that the 11S globulin represents approximately 45% of total proteins in the endosperm tissue and 5–7% of coffee bean dry weight (24). Endosperm is a very important specialized storage organ and, therefore, study of the endosperm proteome can provide information on seed development. Nevertheless, this tissue is rich in storage proteins such as globulins, which hinders the identification of poorly abundant proteins. Indeed, a total of 11 different protein spots analyzed in this study were identified as the 11S globulin. A recent analysis of the rice endosperm proteome was performed, and a method to remove the highly abundant proteins was developed (25). This is an important strategy to detect and identify new proteins in the endosperm.

In this study, differentially expressed proteins were also excised from the gel and analyzed by MS/MS; however most proteins did not reveal reliable identification, probably due to the low amount of protein or low ionization capacity. One protein spot (Sp6), observed in the embryo and up-regulated in the endosperm (Figure 2), was identified as a Cupin superfamily protein. This family includes 11S and 7S plant seed storage proteins and germins (26, 27), which have been associated with allergy (28, 29). In coffee, allergy to caffeine (30–32) and the beans (33–35) has been reported. However, no protein has been directly associated with coffee allergies.

Another protein identified was spot Sp31, which was expressed only in the endosperm (Figure 2) and showed identity to Mal d 1, the major apple allergen, which belongs to a group of pathogenesis-related (PR) 10 proteins (36). It has been reported that the expression of PR10 genes occurs during conditions of stress and ripening (37, 38). Proteins from this family have been implicated in allergy in different consumed fruits such as peach, cherry, and oranges (39–41). Although there are no reports associating these proteins with coffee allergy, additional studies need to be performed to investigate their role in allergy.

A protein spot (Emb26) specifically expressed in the zygotic embryo was identified and showed 100% identity to a hypothetical protein from *Oryza sativa*. Although the masses
of both proteins were similar (30 kDa), the calculated pI of this protein was 10 as opposed to 4.5 for the coffee protein. Unexpectedly, this protein did not show a high identity to any protein present in the Coffee Genome database. It is possible that this protein is a new one, not present in the Coffee and public databases.

Overall, in this work we have identified several protein spots in the whole seed, embryo, and endosperm of coffee as the 11S globulin. We show that this protein is highly expressed in the endosperm and embryo tissues. Moreover, two allergenic proteins were identified, as well as a hypothetical protein. Further studies need to be performed to determine the functional role of these proteins in development and their possible relationship to quality, flavor, and aroma, as well as allergy.

LITERATURE CITED


Received for review June 5, 2008. Revised manuscript received September 2, 2008. Accepted September 17, 2008.