Protein and Peptide Letters, Vol. 11, No. 1, pp. 1-6, 2004
Bentham Science Publishers Ltd.
0929-8665/04 $ 45.00 + .00

SEED PROTEIN VARIATION AMONG PEPPER (*CAPSICUM* SP.) GENOTYPES REVEALED BY MALDI-TOF ANALYSIS

Marco Aurélio Caldas de Pinho Pessoa Filho[1], Carlos Bloch Junior[2], Danilo Fernandes da Silva Filho[3], Alexsandro Sobreira Galdino[1], Rodrigo Maranguape Silva da Cunha[1], Maria Aparecida Oliveira Alves[1], and Thalles Barbosa Grangeiro[1,*]

[1]Laboratório de Citogenética e Genética Molecular, Depto. de Biologia, bloco 906, Campus do Pici, Universidade Federal do Ceará (UFC), Av. Humberto Monte s/n, 60.451-970; Fortaleza-CE, Brasil; *corresponding author's e-mail: thalles@ufc.br; [2]EMBRAPA/CENARGEN, Brasília-DF, Brasil; [3]INPA/CPCA, Manaus-AM, Brasil

Abstract: A method for seed proteome analysis using MALDI-TOF mass spectrometry is described. The data were used to estimate the genetic diversity degree among twelve genotypes of pepper (*Capsicum*). The resulting spectra were converted into a binary matrix consisting of 23 protein data sets, and genetic similarity values were calculated with the FreeTree software and Jaccard's coefficient of similarity. We have also been able to identify the presence of certain proteins in the extracts, by checking their masses on on-line databases.

Key Words: Mass spectrometry – Time of Flight – Germplasm – *Capsicum* – Genetic variability.

INTRODUCTION

The genus *Capsicum*, a member of the Solanaceae plant family, is a major crop of great economic importance. The group has been originated in the New World (Central and South America) and five species are cultivated: *C. annuum, C. frutescens, C. baccatum, C. pubescens* and *C. chinense* [1]. Brazil represents an important secondary center of domesticated *Capsicum* species, where a considerable degree of diversity of *C. annuum* var. *annuum*¸ *C. baccatum* var. *pendulum*, *C. frutescens* and *C. chinense* is found. *C. pubescens* is practically not represented in Brazil, while *C. chinense* is probably the most important cultivated species in the east of the Andes. The Amazon constitutes the area with the highest degree of diversity [2]. Around 12,000 ha of land are used for cultivating *Capsicum* species in Brazil, involving US$ 1.5 million in resources only for commercializing seeds [3].

Genetic diversity studies have been widely used for the determination of strategies in breeding programs. These programs using hybrid varieties of vegetables have been applied in order to develop

superior hybrids with respect to homogeneity of the crop. They involve the selection of inbred lines, which produce F_1 hybrids exhibiting heterosis (a better performance in terms of vigor and yield than the mean of its parents' performance) [4].

Assessments of genetic similarity of various vegetable species and crops worldwide have been performed using morphological data, quantitative genetics [5], allozyme analyses, DNA studies [6-9], and proteome analyses [4]. In the era of genomics in which we are living, researchers are turning to methods of analyses to determine protein functional information on a large scale with a high quantity of results. Functional proteomics evolved from the need to understand and investigate expressed proteins of an organism [10].

Mass spectrometry has catapulted protein analyses and characterization. With several important and recent innovations extending its capability, this technique is now being applied in diverse research approaches: from protein structure elucidation studies to functional proteomics [11,12], from bacterial taxonomy [13] to DNA sequencing [14].

Here we describe a method for seed proteome analysis using MALDI-TOF mass spectrometry. The analyses aimed the assessment of genetic diversity among twelve *Capsicum* sp. genotypes from the Amazon region, which belong to the germplasm collection of the Instituto Nacional de Pesquisas da Amazônia (INPA), as a part of an ongoing breeding and characterization program.

MATERIALS AND METHODS

Seeds from twelve different pepper genotypes were used for the protein polymorphism survey. The seed samples were obtained from the germplasm collection of the Instituto Nacional de Pesquisas da Amazônia (INPA). The genotypes comprised three species: *C. annuum* (PIH-96, PIH-98 and PIH-99), *C. chinense* (PIH-94, PIH-95 and PIH-105) and *C. frutescens* (PIH-82 and PIH-97). Four genotypes have only been identified at the genus level as *Capsicum* sp. (PIH-100, PIH-101, PIH-102, and PIH-104).

25 mg of ground seeds from each genotype were mixed with 500 µL of HPLC buffer (50% Acetonitrile, 0.1% TFA) and extraction was carried out for 12 hours at 4ºC on a shaker. The samples were spun at 18,000 rpm for 20 minutes and the supernatants were transferred to clean tubes and dried. The dried down pellets for each genotype were then resuspended on a 0.1% TFA solution (1 mL). They were kept at 4ºC until they were used in the mass spectrometry analysis.

The analyses were performed by MALDI-TOF mass spectrometry using an Applied Biosystems Voyager DE-STR instrument operating at a 25-kV accelerating voltage in the reflector mode. The sample matrix was trans-4-hydroxy-3-methoxy cinnamic acid. One µL of the extract was diluted in 3 µL of 0.1% TFA, and then mixed with an equal volume (4 µL) of the sample matrix. One µL of this mixture was spotted and air-dried before analysis.

The data revealed by MALDI-TOF/MS spectrum analyses of the seed protein extracts were converted to a binary coded matrix in which the presence of a specific protein with a certain molecular mass was coded as 1, and the absence of this same protein was coded as 0. Since no mass standards were used during the analyses, only relatively high differences between two masses were regarded as distinct proteins

(for instance, two proteins with masses of 5,106.54 Da and 5,111.86 Da were interpreted as being the same protein during the construction of the binary matrix). The data in the binary matrix were used to calculate all pair wise genetic similarities among the different genotypes. Genetic similarities were expressed as Jaccard's coefficient of similarity using the following formula and the software FreeTree [15]:

$$S_J = \frac{a}{a+b+c}$$

In this formula, the value S is the genetic similarity between any two genotypes, i and j, *a* is the number of proteins present in both i and j (1 1 counts in the binary matrix), *b* is the number of proteins present in i and absent in j (1 0), and *c* is the number of proteins absent in i and present in j (0 1). The genetic similarity matrix was then used to construct a dendrogram employing the software FreeTree [15]. The dendrogram shows the relationships among the genotypes and is graphically represented as a tree. The tree was visualized with the program TreeView [16].

RESULTS AND DISCUSSION

The twelve different mass spectra presented a variation in the pattern of expressed proteins within the investigated range of mass values (1,500 to 15,000 Da). Nevertheless, some proteins have shown a constant presence in all analyzed samples: the proteins with masses around 4,500 Da and 5,100 Da are two examples of these (Figure 1). The mass spectrometry analyses of the twelve pepper genotypes resulted in a binary matrix, which was used to perform the genetic similarity analysis. This matrix contained 23 protein data sets (with each reproducible mass regarded as one particular protein). The degree of similarities between all pairs of genotypes was calculated using the variability of these data sets (Table 1). Using the genetic similarity matrix, a dendrogram relating the pepper genotypes was constructed (Figure 2).

Table 1 – Pairwise Jaccard's genetic similarity coefficients of the twelve pepper (*Capsicum*) genotypes based on seed protein mass profiles as measured by MALDI-TOF spectrometry.

	82	**94**	**95**	**96**	**97**	**98**	**99**	**100**	**101**	**102**	**104**	**105**
82	1											
94	0.4666	1										
95	0.4615	0.4705	1									
96	0.3333	0.6190	0.4761	1								
97	0.3500	0.7368	0.5000	0.8571	1							
98	0.4000	0.5000	0.5000	0.5714	0.6000	1						
99	0.2352	0.5882	0.6000	0.5714	0.6842	0.3684	1					
100	0.3571	0.5625	0.2941	0.4761	0.5000	0.6000	0.3333	1				
101	0.4210	0.7368	0.5789	0.8571	0.9000	0.6842	0.6842	0.5789	1			
102	0.3157	0.6315	0.4000	0.6087	0.7142	0.4285	0.6666	0.4000	0.7142	1		
104	0.3529	0.6111	0.4444	0.7500	0.7894	0.6470	0.5555	0.4444	0.7894	0.6000	1	
105	0.2941	0.6470	0.4705	0.7000	0.7368	0.4210	0.6875	0.3888	0.7368	0.7222	0.7058	1

The twelve genotypes possess different levels of similarity (Table 1), ranging from 0.2352 (between genotypes PIH-99 and PIH82) to 0.9000 (between genotypes PIH-101 and PIH-97). The average similarity value was 0.5368 (±0.1537), a value smaller than those obtained from RAPD (Random Amplified Polymorphic DNA) and AFLP (Amplified Fragment Length Polymorphism) analyses by Paran and co-workers [17] for *Capsicum annuum* genotypes, and also smaller than the value obtained for the same accessions present in this study by RAPD analyses performed by our group (unpublished data). A breeding experiment with the association of this polymorphism with the agronomic performance of the genotypes would validate our results. This characterization is part of an ongoing project at the Instituto Nacional de Pesquisas da Amazônia (INPA). The grouping of genotypes belonging to different species (Figure 2) may be due to the high level of cross-pollination between different *Capsicum* species and also by the transport of foreign plant material among local human populations.

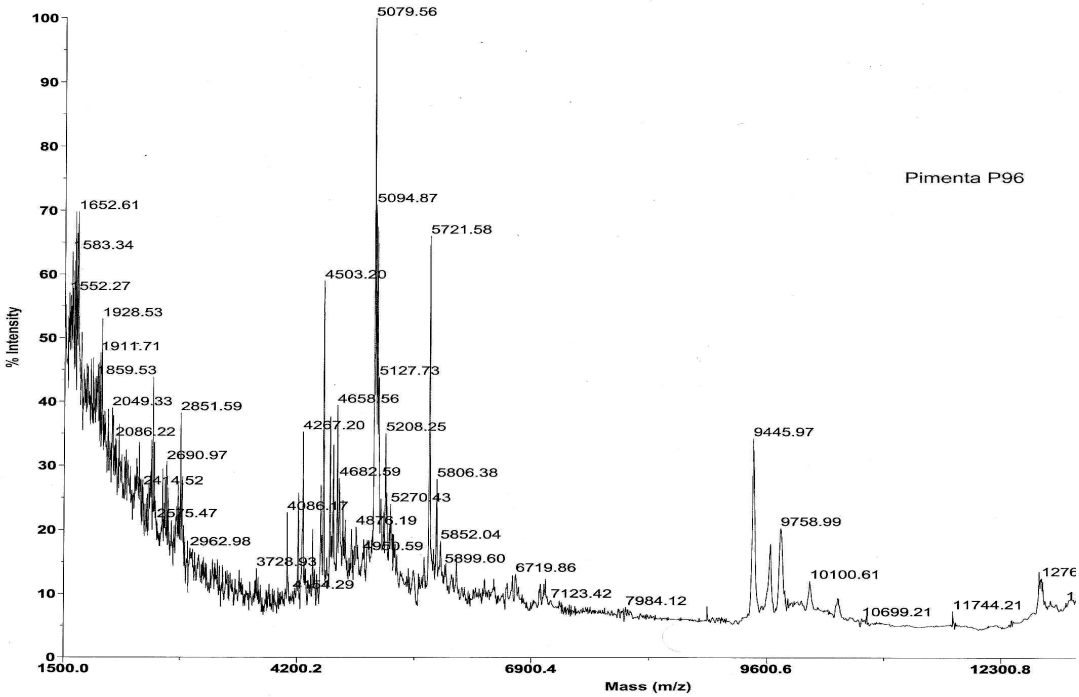

Figure 1 - Matrix assisted laser-desorption ionization time-of-flight (MALDI-TOF) mass spectra of seed protein extract from pepper (*Capsicum annuum*) genotype PIH-96. The mass values observed in all genotypes analyzed are indicated.

Posch and co-workers [4] performed a similar study with *Capsicum annuum* genotypes using 2-D electrophoresis for the detection of seed protein polymorphism. In their study 102 proteins were analyzed (including water-soluble and urea/detergent-soluble proteins). In comparison with 2-D electrophoresis analysis, MALDI-TOF represents a much faster and cheaper methodology, with the generation of results in only a few minutes. Even though the number of protein data sets in our study is likely to be

underestimated, because the mass spectrometry analysis only allows a specified range of masses to be analyzed each time for the conditions used, further studies involving the analyses of a wider range of masses would assure a better general view of the functional proteome of the samples, and consequently a more reliable genetic diversity assessment.

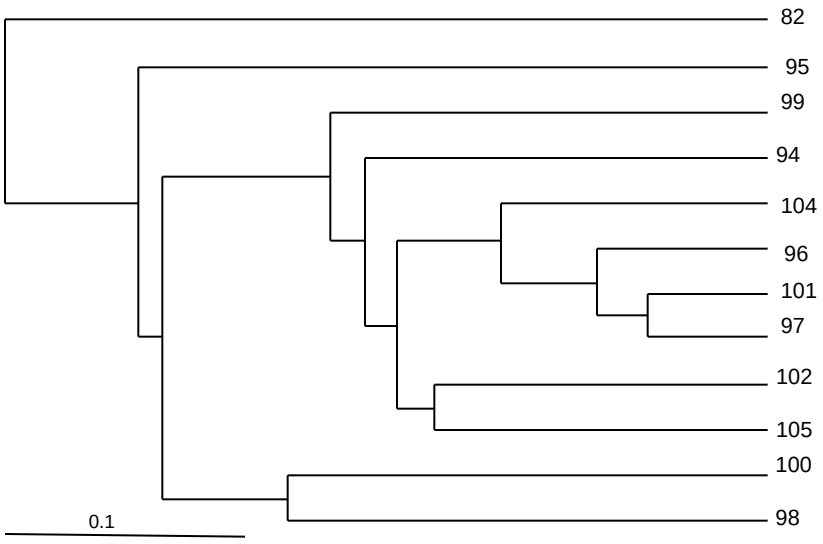

Figure 2. Dendrogram constructed from the genetic similarity (Jaccard's coefficient) data illustrating the relationships among the pepper (*Capsicum*) genotypes as revealed by their MALDI-TOF seed protein profiles.

An investigation on protein sequence databases indicated the existence of a few gamma-thionin sequences already determined for *Capsicum annuum* and a putative gamma-thionin precursor in *Capsicum chinense*. They possess a high degree of homology and their molecular masses seem to be present in the spectra produced by our analyses. A gamma-thionin from *C. annuum* (Entrez PID g1171503), with Mw 5,196.02 Da, seems to be present with high intensity on most of the genotypes (data not shown). The detection of predicted peptides from nucleotide sequences in seeds, and the prediction of post-translational modifications - by looking for their predicted masses in the spectra - might be an additional application of this methodology. However, other methods used in protein chemistry studies, such as the digestion with specific proteases (trypsin, for example) and further characterization of the resulting peptides would be necessary for the correct identification of particular proteins in the samples.

ACKNOWLEDGEMENTS

This work was supported by grants from Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq), EMBRAPA and Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP). The corresponding author (T. B. Grangeiro) has a research fellowship (PQ-2B) from CNPq.

REFERENCES

[1] Prince, J. P., Lackney, V. K., Angeles, C., Blauth, J. R., Kyle, M. M. **(1995)** *Genome 38,* 224-231.
[2] Ferreira, M. E. **(2000)** *Capsicum.* Pimentas e pimentões do Brasil Brasília: Embrapa Comunicação para Transferência de Tecnologia.
[3] Khatounian, C. A. **(1997)** *Horticultura Brasileira 15,* 199-205.
[4] Posch , A., van der Berg, B. M., Duranton, C, Görg, A. **(1994)** *Electrophoresis 15,* 297-304.
[5] Storfer, A. **(1996)** *Trends Ecol. Evol. 11,* 343-348.
[6] Hormaza, J. I., Pinney, K., Polito V. S. **(1998)** *Econ. Bot. 52, (1)* 78-87.
[7] Hogbin, P. M., Peakall, R. **(1999)** *Conserv. Biol. 13,* 514-522.
[8] Rodriguez, J.M., Berke, T., Engle, L., Nienhuis, J. **(1999)** *Theor. Appl. Genet. 99,* 147-156.
[9] Lefebvre, V., Goffinet, B., Chauvet, J. C., Caromel, B., Signoret, P., Brand, R., Palloix A. **(2001)** *Theor. Appl. Genet. 102,(5)* 741-750.
[10] Washburn, M. P., Yates III, J. R. **(2000)** *Curr. Opin. Microbiol. 3,* 292-297.
[11] Yates III, J. R. **(2000)** *Trends Genet. 16, (1)* 5-8.
[12] Gygi, S. P., Aebersold, R. (**2000**) *Curr. Opin. Chem. Biol. 4,* 489-494.
[13] Lay Jr, J. O. **(2000)** *Trends Anal. Chem. 19, (8)* 507-516.
[14] Köster, H., Tang, K., Fu, D. J., Braun, A., van der Boom, D., Smith, C. L., Cotter, R. J., Cantor, C. R. **(1996)** *Nat. Biotechnol. 14,* 1123-1128.
[15] Pavlicek, A., Hrda, S., Flegr, J. **(1999)** *Folia Biol. (Praha) 45,* 97-99.
[16] Page, R. D. M. **(1996)** *Comput. Appl. Biosci. 12,* 357-358.
[17] Paran, I., Aftergoot, E., Shifriss, C. **(1998)** *Euphytica 99,* 167-173.

Received on May 28, 2003, accepted on November 30, 2003.