Characterization of differentially expressed sequence tag (EST) from sugary phenotype of storage root of cassava (*Manihot esculenta* Crantz)[1]

Luiz Joaquim Castelo Branco Carvalho[2*], Claudia Regina Batista de Souza[2,3], Julio Cezar de Mattos Cascardo[4] and Marco Antonio Valle Agostini[2].

[1]Research financed by The Rockefeller Foundation (RF96010#25/RF9707#26). Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq (45 2714-98.2). Programa. Nacional de Pesquisa em Biotecnologia, CENARGEN (Project Nº 060302058). [2]EMBRAPA-Genetic Resources and Biotechnology, Brasilia-DF. Brazil. [3]Current address: Departamento de Genética Universidade Federal do Pará, Belém-PA, Brasil. [4]Departamento de Biologia CelularUniversidade Estadual de Santa Cruz, Ilhéus-BA, Brasil. *Author for correspondence: carvalho@cenargen.embrapa.br.

ABSTRACT

The sugary storage root phenotype was, recently identified from traditional clones cultivated in Amazon. At the level of above ground morphology there is no particular distinction of sugary cassava from the farina and table cassava (see accompany poster). Major distinction among these types of cassava is observed only in the storage root, which shows profound alteration in storage root size and density in water, secondary xylem cells development, parenchyma cells size and turgidity, diverse starch structure presence, and osmotic parenchyma cells environment. Therefore it is expected to observe a profound effect on genes being expressed and regulated differentially in the sugary cassava in comparison with the farina type. We are using cDNA subtractive library to construct an EST database with silenced and expressed genes for this kind of storage root. Macro array and conventional mRNA blots are being used for differential gene expression analyzes. The BLAST and clustering analyses of 1200 high quality EST sequences gave 267 cDNA clones that were tested with macro array and conventional RNA blot analyses. Preliminary gene sequence annotation for 953 cDNA clones indicated that 25% are non hits, 22% codes for unknown protein, 17% codes for carbohydrate related metabolism, and other minor functions. cDNA clones are being tested for gene expression analyzes to gain information on tissue specificity, tissue age effect as well as phenotype specificity.

INTRODUCTION

A distinct type of cassava (*Manihot esculenta* Crantz) storage root with high water content has been known and used since pre-Columbian times by Brazilian Aborigines in the Amazon, under the name of Manipueira (Travaço, 1596; Lisboa, 1631; da Cunha, 1978). This type of cassava plant was named Mandiocaba by the early researchers in the Amazon and its cultivation was abandoned because of its low dry matter content, extremely low starch (2%) and high water content (de Albuquerque, 1969) in relation to cultivars used to produce farina (cassava flour) fresh consumption (table cassava). In spite of this lengthy association with man, no specific biochemical characterization of this type of cassava has been made for centuries. Recently, L.J.C.B. Carvalho and his group (Carvalho *et al.,* 2000) announced several visits to the proposed center of origin and domestication of cassava in Brazil (Allem, 1994; Olsen and Schaal, 1999) to search for additional diversity on this trait. They have reported the organization of a GENEBANK with a large number of rare clones, with a storage root showing novel features previously unknown to conventional cassava cultivars, including high free sugar content (Carvalho *et al*., 2004). Major distinction among these types of cassava is observed only in the storage root, which shows profound alteration in storage root size and density in water, secondary xylem cells development, parenchyma cells size and turgidity, diverse starch structure presence, and osmotic parenchyma cells environment. Therefore it is expected to observe a profound effect on genes being expressed and regulated differentially in the sugary cassava in comparison with the farina type.

We are using cDNA subtractive library to construct an EST database with silenced and expressed genes for storage root with sugary phenotype. Macro array and conventional mRNA blots are used to identify differentially expressed gene between sugary and normal storage root phenotypes. In this document we report our preliminary results to validate these two technologies toward the identification of differentially expressed genes in sugary phenotype of storage root of traditional cassava clones from Amazon.

MATERIALS AND METHODS

Plant material: Storage root of 3cm diameter were harvest freshly in the field plots of the GENEBAK at EMBRAPA-Genetic Resources and Biotechnology in Brasilia. The roots were conditioned on ice box, brought to the laboratory, washed, peeled off and stored in -80°C until used. Two sugary cassava clones and one farina cassava type were used in all the experiments. Clone Cas36.1, which accumulate glycogen like starch, clone Cas36.0 which accumulate *waxy* starch and commercial variety IAC 12-829 were used to construct the subtracted libraries and expression analysis.

RNA extraction: Total RNA was extracted from cassava plant storage roots as described by de Souza *et al.*, 2002.

Construction of subtracted cDNA libraries: Polyadenylated RNA was purified from 500 µg of total RNA using the mRNA Purification kit (Pharmacia). Double-stranded cDNAs were synthesized and subtractive PCR was conducted using the PCR-Select cDNA Subtraction Kit (Clontech). All procedure steps to obtain the subtracted cDNA were performed according to the manufacturer's protocol. Three subtracted cDNA library were constructed. Library MAGL used Cas36.1 as tester and IAC 12-829 as driver, library MAAC used Cas36.0 as tester and IAC 12-829 as driver, and library MAGLR used IAC 12-829 as tester and Cas36.1 as driver. cDNAs were synthesized separately with 2 µg of polyA+ RNA, and the testers and driver cDNAs were separately digested with *RsaI*. The tester sample then was divided into two populations and each one ligated with adaptor 1 or 2R. Each tester population was hybridized separately with excess driver cDNA (33-fold). The two hybridization reactions were combined (with additional driver cDNA) for a second subtractive hybridization. The hybridizations led to equalization and enrichment of differentially expressed sequences. Finally, the differentially expressed cDNAs were amplified by PCR and cloned into the pGEM-Teasy vector (Promega).

cDNA clone sequence: Single bacterial colonies of the result of libraries MAGL, MAAC, and MAGLR were inoculated in 96-well microtiter plates containing LB and ampicillin (100 mg/L) and allowed to grow overnight with shaking at 37°C (18 hr at 320 RPM). The DNA was purified by standard alkaline lysis method with one modification at the end of the procedure, where the supernatant was passed through a multi-screen filter (Millipore) prior to DNA precipitation. The purified DNA was ressuspended in water. The sequencing reactions were performed with 200ng of DNA using standard protocols of the ThermoSequenase II dye terminator cycle sequencing kit (Amersham-Pharmacia Biotech) using universal M13 forward and reverse primers and the capillary sequencer MegaBACE 1000. The samples were electro injected by 2 KV for 100 seconds and the fragments were separated at 9 KV for 100 minutes.

Sequence analysis of ESTs: The base-calling software PHRED (Ewing *et al.,* 1998) read DNA sequencer trace and sequences portion of poor quality were trimmed and low quality sequence (PHREED >20) removed from further analysis. Vector, primers and adaptors sequences were eye trimmed from the high quality sequences. Only ESTs whose sequences exceed 250bp after trimming were considered for further analyses. Homology searches of high quality ESTs were performed using BLASTN and BLASTX against NCBI and TAIR data base to obtain the putative function for each EST. ESTs of identical putative function were assembled into groups in EXCEL software and the multiple sequence alignment algorithm of the software MALIGN used to identify ESTs of identical sequences or ESTs whose sequence hits twice the same sequence accession number in the data base. Finally, identical sequences were reassembled into non redundant sequence groups in EXCEL to construct a dedicated data base for the sugary cassava. The database will be public released and available for consultation at web site of Laboratory of Biochemistry and Biophysics at EMBRAPA-Genetic Resources and Biotechnology (look in the near future at www.cenargen.embrapa.br).

Macroarray of cDNAs: Selected sequenced clones listed in Table 1 and 2 were amplified through PCR directly from colony using the universal primers in 96 well plates. The PCR products (15 l) was mixed with 200 l of 0.4M NaOH and 10mM EDTA and boiled for 10 min, for denaturing and immediately cooled on ice for 3 min and blotted (100 l) on positively charged membrane using dot-blotting apparatus (Milipore, Belford, MA). The membrane was rinsed in 2x SSC buffer at room temperature for 5 min and air-dried and the DNA permanently fixed in an UV cross-linker. Arrays design included all the selected clones from the cDNA library, positive controls (Actin, Tubulin, empty vector, and Mec1 gene), negative control (blank) and concentration gradient of Mec1. mRNA samples from IAC and Mirasol storage roots were used to produce radiolabeled cDNA probes with reverse transcription SuperScript Kit (Invitrogen) as described by the manufacture. Conventional hybridization reaction techniques and procedures were used as described by de Souza *et al.*, 2002.

Macroarray data analyses: Macroarray technology is been used to assess the subtraction efficacy of our procedure and expression analyzes of differentially expressed gene in different tissue sources addressing questions such as tissue specificity, development and diversity of sugary cassava. Digital images of the autoradiogram of the arrays were generated by using a high resolution (1024x1024 pixel) Dual Scan (AGFA T2000) and transferred to a high capacity computer to further analyses with the ZERO_Scan DNA analyses package (Scanalytics). The generated data is assembled in EXCEL for calculations, comparisons and construct a dedicated data base for differentially expressed gene between sugary and farina type of cassava.

RESULTS AND DISCUSSIONS

CDNA subtraction library construction and assessment of efficacy of subtraction: Figure 1 displays an assessment efficacy of the subtraction procedure with the PCR enrichment step of differentially expressed cDNA population for the three library (MAGL, MAGLR, and MAAC) as well as its validation by the northern blot reverse. One constitutive gene (Mec1) of storage root of cassava, a low copy number gene (MyB), and two sucrose-starch conversion pathway (Sucrose synthase and Fructokinase) genes were visualized by reverse northern blot analyses. Subtracted and non subtracted populations of cDNA for each library (Panel A) showed that the overall fragment pattern of subtracted is quite distinct from the non-subtracted, indicating that selective fragment amplification had occurred. Panel B shows that similar intensity of blots signal for the low copy number gene (MyB) and enzyme coding genes were subtracted the same way indicating that the selective fragment amplification was efficient. In addition the Mec1 gene was completely subtracted in all libraries. These results indicate the efficacy of the PCR amplification step in improving the overall selective amplification step preceding the clone step.

Sequence, clustering and functional categorization of storage root ESTs: Sequences of libraries MAGL and MAAC were analyzed together to increase our efficiency in identifying differentially expressed genes in the sugary phenotype in relation to the farina type of cassava. The library MAGLR was analyzed separately to identify genes that were silenced with the appearance of the sugary phenotype. About 1033 high quality ESTs were obtained out of 1440 sequences. Two hundred sixty seven clones remained after putative function assignment and clone sequences redundancies identification and used for macroarray and northern blot analyzes. The major classes of putative function for selected sequences showed in Figure 2 indicates that about 25% and 22% of the clones are non hits and unknown proteins. The sequences from the MAGLR library will be reported in our database soon.

Macroarray analyses: Macroarray blot analyses showed in Figure 3 indicates that the hybridization signal of a blot sample of 96 selected clones for the library MAGL probed with a pool of cDNA from clone Mirasol varied in signal intensity. The three level of signal intensity indicate a partial subtraction of the overall population of mRNA and enhance the value of the PCR selective amplification step in cloning differentially expressed genes. It also suggests that additional screening may be required, in order to isolate solely differentially expressed associated with the sugary cassava. A larger number of clones are being tested to identify ESTs exclusively related to the sugary phenotype and confirmed with conventional northern blot analyzes.

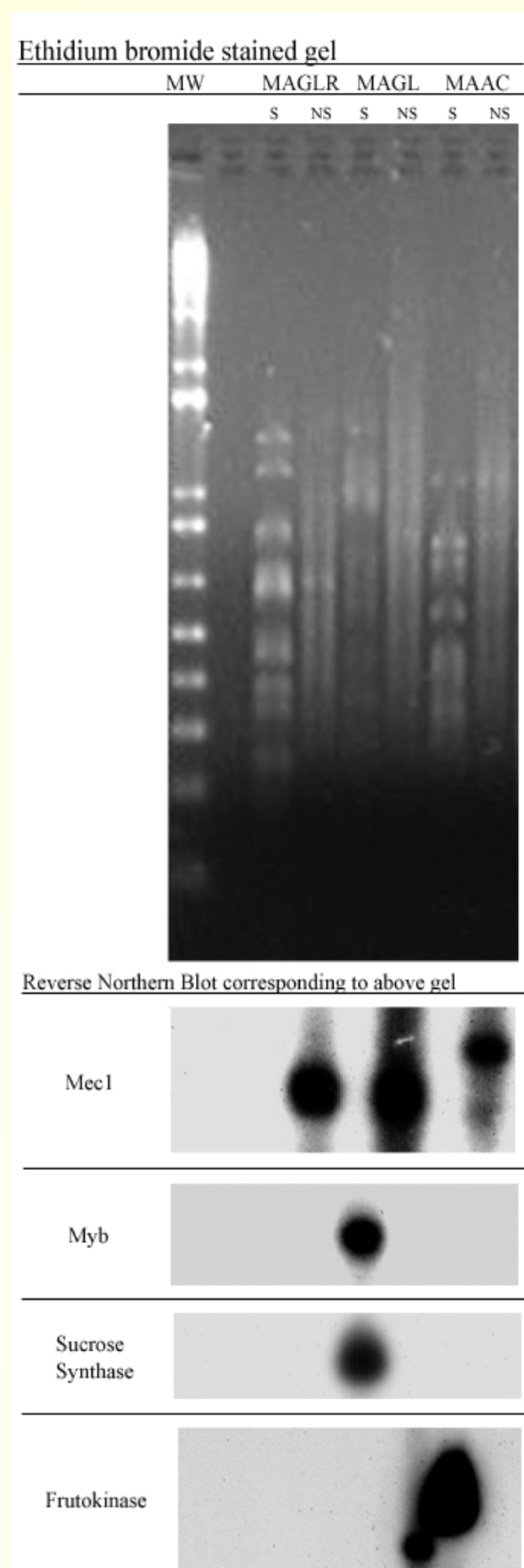

Figure 1 - Assessment of subtraction efficacy. **Panel A** ethidium bromide stained gel for total and subtracted cDNA pools PCR amplified. **Panel B** Reverse northern blot of Panel A probed with three classes of genes. MAGL, MAGLR and MAAC are the libraries code. S Subtracted cDNA pool, NS Non-subtracted total cDNA pool, MW is the molecular marker. The Mec1 is the non-subtracted control gene. The gene, MyB, Sucrose synthase and fructokinase were genes fully subtracted.

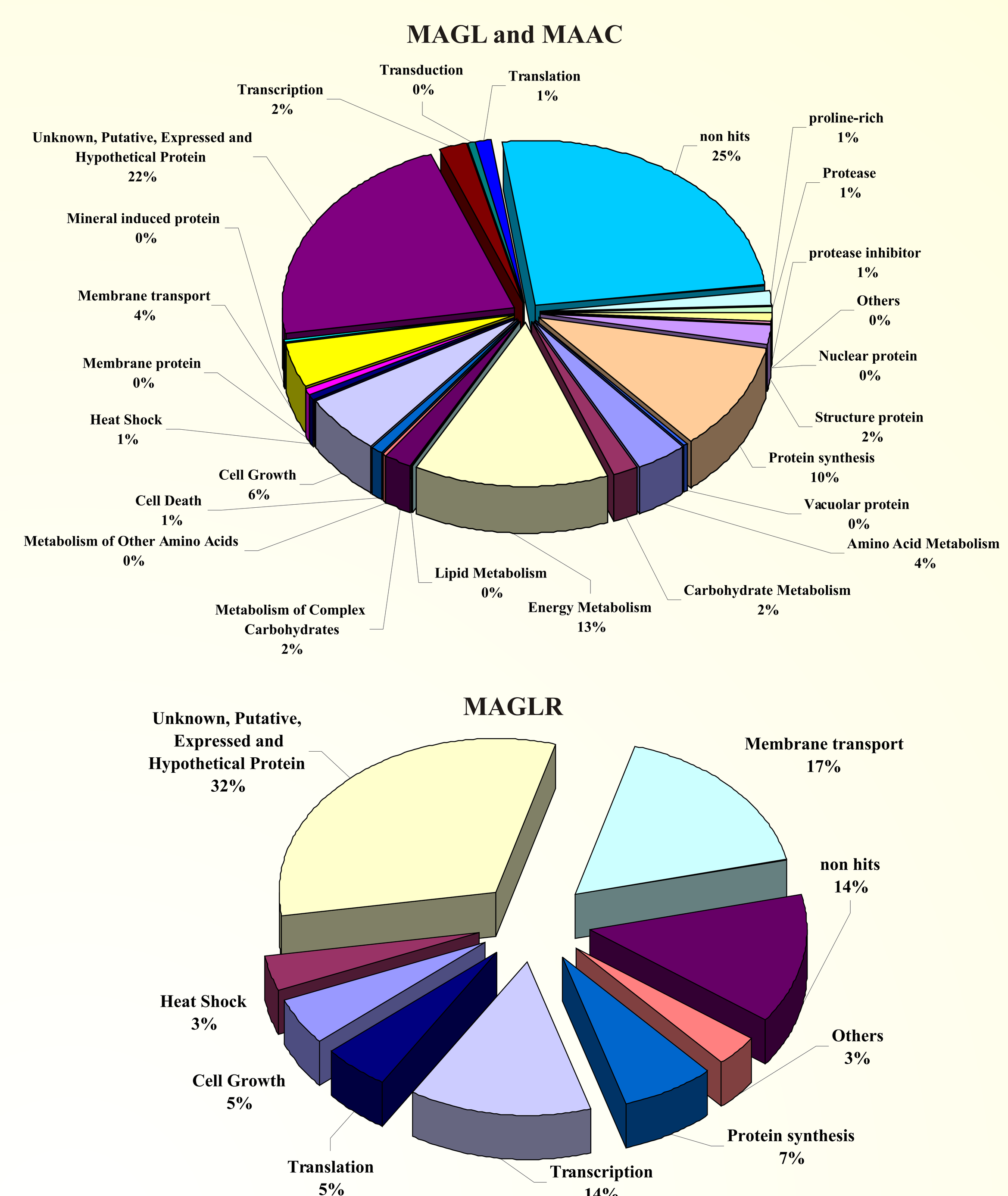

Figure 2 - Classification of ESTs from sugary/wt (MAGL and MAAC libraries) and wt/sugary (MAGLR library) subtractions cDNA library of cassava storage root. High quality sequences cDNA with BLASTX scores superior to 10^{-5} were classified according to their putative functions and named according to KEGG data base nomenclature. MAGL is the subtractive library using sugary cassava accumulating glycogen like starch as tester. MAGLR is the subtractive library using farina type of cassava as tester.

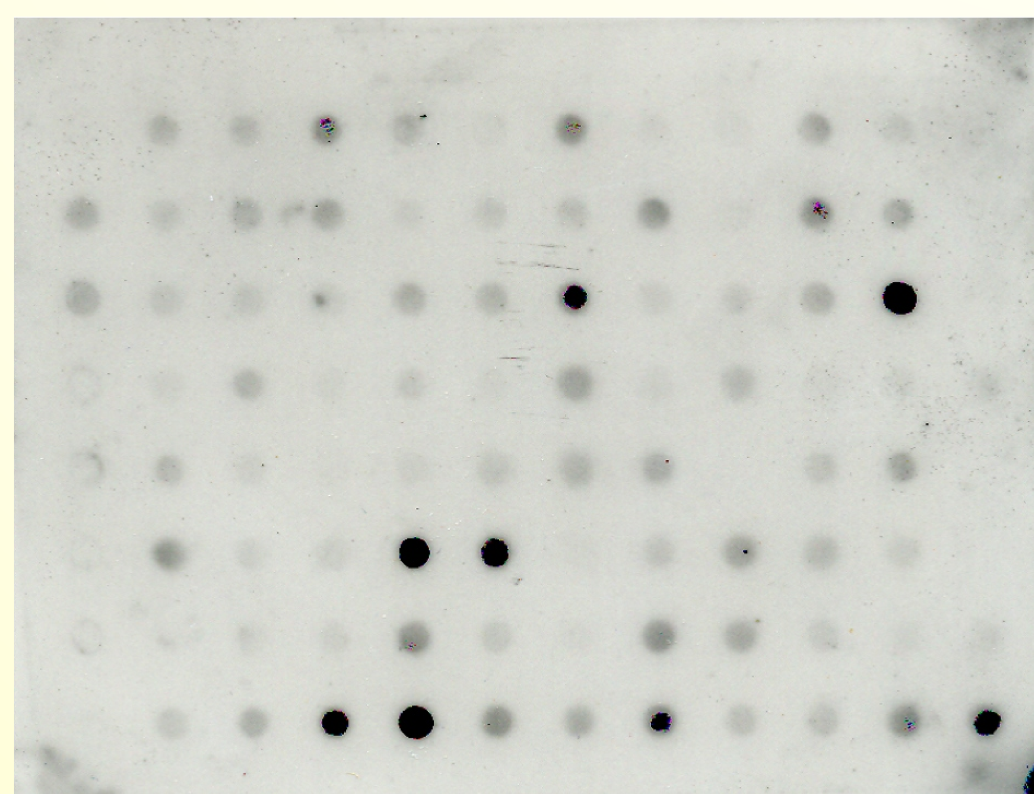

Figure 3 - Hybridization signals of 96 cDNA clones inserts from storage root of sugary cassava (MAGL library) probed with the driver cDNA pools for table cassava type (Mirasol).

CONCLUDING REMARKS

The ESTs data presented here is the first overview of genes that are differentially expressed and silenced in the sugary and farina type of cassava. These genes can be exploited to unravel regulatory networks involved in the formation of the storage root of the sugary cassava. This EST data will make it feasible for molecular breeders to develop new varieties of cassava with high free sugar content as well as novel starches.

AKNOWLEDGEMENTS

Special acknowledges are extended for the financial support provided by the Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (grant number 680.410/01-1 for LJCBC) of the Minister of Science and Technology of Brazil as well as to National Biotechnology Program of EMBRAPA (grant number 06.03.02.058 for LJCBC), the Small Grant program of Cassava Biotechnology Network (CBN) and The Rockefeller Foundation (RF96010#25/RF9707#26). Our special thanks to Tiago Oberda Carneiro Marques for the great help in the arts of the posters in the last moment is appreciated.

REFERENCES

Allem, A.C. 1994. The origin of *Manihot esculenta* Crantz (Euphorbiaceae). Genetic Resources and Crop Evolution. 41:133-150.

Carvalho LJCB, de Souza CRB, Cascardo JCM, Junior CB. 2004. Identification and characterization of a novel cassava (Manihot esculenta Crantz) clone with high free sugar content and novel starch. Plant Molecular Biology. (In press).

Carvalho, L.J.C.B., Cabral, G.B., and Campos, L. 2000. Raiz de reserva de mandioca: Um sistema biológico de múltiplas utilidade. EMBRAPA-Recursos Genéticos e Biotecnologia. Serie Documentos # 44. p16. Brasilia,DF. Brazil.

da Cunha, A.G. 1978. Dicionário Histórico das Palavras Portuguesas de Origem Tupi. Companhia Melhoramentos, Universidade de Brasília, Brasília, Brasil. 357p.

de Albuquerque, M. 1969. A mandioca na Amazônia. SUDAM, Belém-PA, Brasil. 277p.

de Souza CRB. Carvalho LJCB, de Almeida ERP, Gander ES. 2002. Towards the Identification of cassava storage root protein genes. Plant Foods for Human Nutrition 57: 353-363.

Ewing B, Hillier L, Wendl MC, and Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Research. 8:175-185.

Lisboa, C. 1631. História Animal e Árvores do Maranhão. Page 176.

Olsen, K., and B. A. Schaal. 1999. Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. Proc Natl. Acad. U.S.A. 96:55865591.

Travaços, S. 1596. Declaração do Brasil XLVII (Brazilian Declaration XLVII). Page 35.