Subtractive cDNA library and macro-array analysis to isolate specific genes related to the color diversity in the storage root of cassava (*Manihot esculenta* Crantz)[1].

Claudia Regina Batista de Souza[2,3], Luiz Joaquim Castelo Branco Carvalho[2*], 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: Dept. de Genética-Universidade Federal do Pará, Belém-PA, Brazil. [4]Deptº. de Biologia CelularUniversidade Estadual de Santa Cruz, Ilhéus-BA, Brazil. *Author for correspondence: carvalho@cenargen.embrapa.br.

ABSTRACT

This study was performed to evaluate cDNA subtractive library in combination with macro-array technology to identify genes using their transcriptionally defective mutant in the diversity of pigmented storage root of cassava. Diversity of color in cassava storage root is due to the presence of a variety of carotenoids, including -carotene, lycopene, phytoene and lutein that are synthesized and accumulated in chromoplast of this type of root. Three putative mutants in our GENEBANK were identified to accumulate solely lycopene, -carotene, or lutein. If this differential accumulation is in the synthesis pathway or in the process of accumulation is not yet known. Three cDNA subtractive libraries were prepared and the generated ESTs are being tested for its tissue specificity, traditional clone specificity and for the tissue age effect in the storage root. Preliminary results indicate that the number of genes is differentially distributed among cellular and molecular processes including 20% coding for complex carbohydrate metabolism related genes, 16% coding for senescent cell process, 15% coding for protease as the major classes of genes. Secondary metabolism derived compounds genes code for 8% of the genes identified. Color phenotype and tissue age in the storage root are being tested for the differentially expressed EST using macro array and conventional mRNA blot as well as a set of eight genes coding for enzymes of the carotenoid syntheses pathway.

INTRODUCTION

The storage root (SR) of cassava (*Manihot esculenta* Crantz) has been considered in the past, solely as a source of starch. However, to improve cassava for carotenoid pigmentation is an open and interesting area of research considering the beneficial effects of this compound class acting as antioxidants and provitamin A and given the prevalence of vitamin A deficiency diseases in the tropics and Brasil (FAO, 1998; WHO/UNICEF, 1995; www.saude.gov.br.). Genetic diversity for cassava SR color is available in several germplasm collections maintained in India (Moorthy *et al.,* 1990), Brazil (Ortega-Flores, 1991; Guimaraes & Barros, 1971; Marinho et al. 1996) and Colombia (Iglesias *et al.,* 1997). Studies with an orange-pigmented SR clone indicated that two cycles of selection and recombination could improve the amount of -carotene as the major provitamin A carotene (Jos et al. 1990). The inheritance of -carotene accumulation was found to be controlled by the action of two genes with complete dominance of orange over white SR trait (Iglesias *et al.,* 1997). These reports underscore the high potential of cassava for development towards as an improved staple food combining macronutrients (starch) with provitamin A as a micronutrient. However, the few SR pigmented clones obtained so far may not represent promising genetic materials to develop provitamin A rich cassava because of their relatively low levels of -carotene (ranging from 0.75 to 4.69mg/100g FW). More recently, expeditions to the center of origin and domestication of cassava in the Brazilian Amazon has been able to collecte additional diversity for the SR carotenoid content trait, including clones with the color varying from white, intense yellow, yellow, cream and pink (Carvalho *et al.,* 2000a; Carvalho *et al.*, 2000b).

Carotenoid biosynthesis represents a molecularly well-characterized biochemical pathway in plants. To date, practically all structural genes involved in plant carotenogenesis are known; only -ring hydroxylation remains to be elucidated. While there is a growing wealth of data available showing how differential expression of these structural genes affects carotenoid content or pattern, no regulatory genes have been identified so far.

Accumulation of -carotene in the cassava storage root might be the result of a variety of regulatory mechanisms (for general review see Cunningham and Grantt, 1998). In lipophilic product being delivered into the lipid-bilayer. The massive carotenoid deposition will ultimately lead to deleterious alteration of the physico-chemical properties of membrane, which needs to be avoided. In fact, chromoplasts develop sites of carotenoid deposition in the stroma. These are lipid globules (plastoglobules) and/or proteins organizing fibrillar supramolecular structures containing carotenoid and lipids. The formation of carotenoid crystals is another principle of carotenoid sequestration as well as the proliferation of plastid membranes (for review, see Camara et al., 1995). To identify and to provide such cellular sinks for carotenoid may add to the arsenal of technology currently available to modify this biosynthetic pathway in transgene approaches such as in rice grains (Ye et al., 2000).

In this document we are reporting preliminary results toward the exploitation of the natural genetic diversity of SR color in cassava to identify and isolate genes particularly associated with three color phenotypes isolated in traditional clones from the Amazon. Our approaches includes the description of the occurrence of carotenoid types in cassava storage root, their differential accumulation during root development, the identification of chromoplast-associated proteins, and finally the isolation of candidate genes controlling the differential biosynthesis of carotenoid using the diversity of storage root.

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 -80ºC until used. Two sugary cassava clones and one farina cassava type were used in all the experiments. Clone Klinazik, which accumulate -carotene and lutein, clone BMG19 which accumulates -carotene, clone Mirasol which accumulate several forms of carotenoid, and white 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 MALC used clone Mirasol as tester and IAC 12-829 as driver, library MABT used cloned BGM19 as tester and IAC 12-829 as driver, and library MALT used clone Klinazik as tester and IAC 12-829 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 MALC, MABC, and MALU were inoculated in 96-well microtiter plates containing LB and ampicillin (100 mg/L) and allowed to grow overnight with shaking at 37ºC (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 (PHRED >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 pigmented 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 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 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 pigmented cassava. Digital images of the auto radiogram 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 pigmented and white cassava.

RESULTS AND DISCUSSIONS

cDNA subtraction library construction and subtraction validation: Figure 1 displays an assessment efficacy of the subtraction procedure with the PCR enrichment step of differentially expressed cDNA population for the three library (MACL, MABT, and MALT) as well as its validation by the northern blot reverse. One constitutive gene (Mec1) of storage root of cassava, and a low copy number gene (ARP) 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. Exception is observed for the case of library MALC that showed similar band pattern. Panel B shows that similar intensity of hybridization signal for the low copy number gene (ARP) was subtracted in the library MALC in contrast to the overall fragment observed in the gel pattern and the expression of Mec1 gene. Further investigation on the efficacy of these libraries is under way.

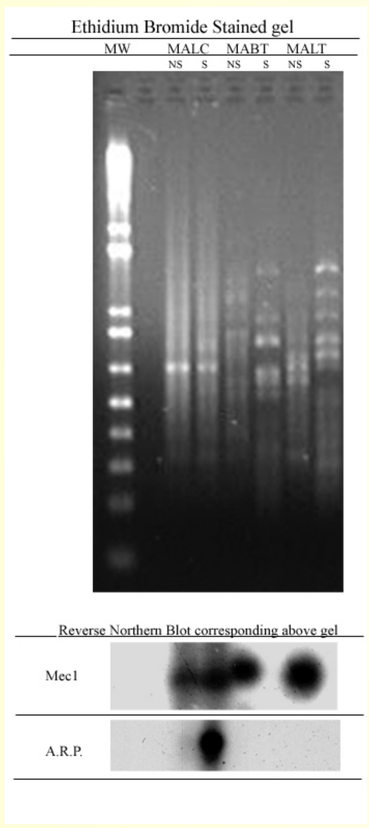

Figure 1 - Assessment of subtraction efficiency. Subtracted (S) and non subtracted (NS) cDNA pools as PCR amplified products. Mec1 and ARP clones used as probe. The Mec1 is the constitutive positive control gene. The gene ARP is genes fully subtracted.

Sequence, clustering and functional categorization of storage root ESTs: Sequences of three libraries (MALC, MABT and MALT) were analyzed together to increase our efficiency in identifying differentially expressed genes in the pigmented phenotype in relation to the white type of cassava. Only preliminary results from the library MALC is showed in the present document. ESTs were screened against current GeneBank and TAIR database using BLASTX algorithm. Putative functions were assigned after apply stringency level of e-value of 10^{-8}. A total of 335 out of 576 ESTs from the MALC library showed significant similarity to any known sequence in the data bank. Figure 2 shows the categorization of high quality selected ESTs based on their putative function assigned from the homology analyze. The distribution of ESTs from the library MALC indicates three major cellular processes most activated in the pigmented cassava. Cell death (16%) followed by protease syntheses (15%) and complex carbohydrate metabolism (14%). Few non-hits (3%) were also observed. Soon, an EST database for pigmented cassava will be public released and available for consultation in the web site of the Laboratory of Biochemistry and Biophysics at EMBRAPA-Genetic Resources and Biotechnology (look in the near future at www.cenargen.embrapa.br).

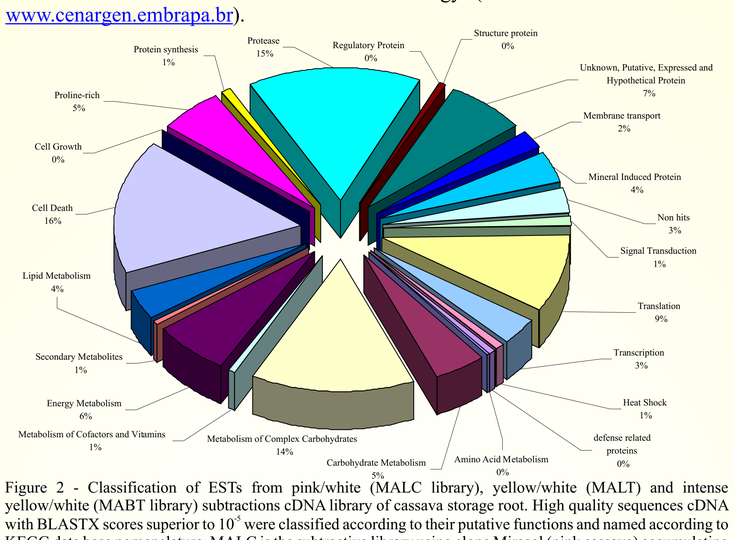

Figure 2 - Classification of ESTs from pink/white (MALC library), yellow/white (MALT) and intense yellow/white (MABT library) subtractions cDNA library of cassava storage root. High quality sequences cDNA with BLASTX scores superior to 10^{-5} were classified according to their putative functions and named according to KEGG data base nomenclature. MALC is the subtractive library using clone Mirasol (pink cassava) accumulating lycopene. MALT is the subtractive library using clone Klinazik (intense yellow cassava) accumulating lutein. MABT is the subtractive library using clone BGM19 accumulating b-carotene.

Macroarray analyses: Macroarray blot analyses in Figure 3 is the hybridization signal of a blot sample of 88 selected clones for the library MAGL probed with a pool of cDNA from white cassava IAC 12-829. Only few hybridization spots were observed (2 out of 88), indicating the high efficacy of the subtraction procedure used. Additional screening is required, in order to isolate solely differentially expressed genes associated with pigmented cassava. A larger number of clones are being tested to identify ESTs exclusively related to the pigmented phenotype and confirm with conventional northern blot analyzes.

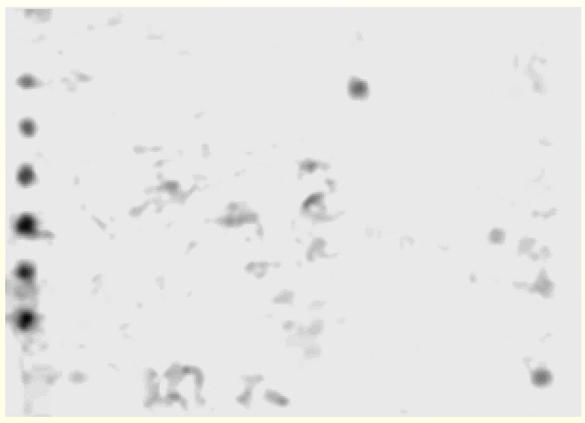

Figure 3 - Hybridization signals of 96 cDNA clones inserts from storage root of pigmented cassava (MALC library) probed with the driver cDNA pools of white cassava IAC 12-829.

CONCLUDING REMARKS

The ESTs data presented here is the first overview of genes that are differentially expressed the pigmented cassava in relation to white type of cassava. These genes can be exploited to unravel regulatory networks involved in differentially accumulation of carotenoid in pigmented storage root of cassava. This EST data will make it feasible for molecular breeders to develop new varieties of cassava with high carotenoid content as well as protein content.

AKNOWLEDGEMENTS

Special aknowledges 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 060302058 for LJCBC), the Small Grant program of Cassava Biotechnology Network (CBN) and The Rockefeller Foundation (RF96010#25/RF9707#26 for LJCBC). 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

Camara, B, Hugueney, P, Bouvier, F, Kuntz, M, Moneger, R. 1995. Biochemistry and Molecular Biology of chromoplast development. Int Rev Cytol. 163:175-235.

Carvalho, LJCB, Cabral, GB, and Campos, L. 2000a. Raiz de reserva de mandioca: Um sistema biológico de múltiplas utilidade. EMBRAPA-Recursos Genéticos e Biotecnologia. Serie Documentos # 44: 16p. Brasilia,DF. Brazil

Carvalho, LJCB, Boiteux, MENF, and Rocha, TL. 2000b. Carotenoids synthesis and accumulation in cassava storage root. Plant Biology (ASPB-USA), Abs. 397.

Cunningham, F.X. and Gantt, E. 1998 Genes and Enzymes of carotenoid biosynthesis in plants. Annual Review in Plant Physiology and Plant Molecular Biology 49: 557-583.

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.

FAO 1998 Food outlook. December 1998. FAO via delle terme de carcalla. 00100 Rome, Italy.

Gimaraes ML, Barros, MSC 1971 Sobre a ocorrência de caroteno em variedades de mandioca amarela. Bol. Tec. Div. Tecnol. Agri. Aliment. 4:1-4.

Iglesias, C., Mayer, J., Chavez, L. and Calle, F. 1997 Genetic potential and stability of carotene content in cassava roots. Euphytica 94: 367-373.

Jos, JS, Nair SG, Moorthy SN, and Nair RB. 1990 Carotene enhancement in cassava. J Root Crops. 16:5-11.

Marinho, HA, Xavier, JJBN, Miranda, RM, Castro, JS. 1996 Estudos sobre carotenóides com atividade de pro-vitamina "A" em cultivares de mandioca (*Manihot esculenta* Crantz) em ecossistema de terra firme de Manaus, Amazonas, Brasil. ACTA Amazônica 1:127-136.

Moorthy, S.N., Jos, J.S., Nair, R.B. and Sreekumari, M.T. 1990 Variability of -carotene content in cassava germplasm. Food Chemistry 36:233-236.

Ortega-Flores, CI. 1991 Carotenoides com atividade pro-vitamica A e teores de cianeto em diferentes cultivares de mandioca no Estado de São Paulo. MSc. Thesis. Universidade de São Paulo, Brazil. 77p.

WHO/UNICEF 1995 Global Prevalence of Vitamin A Deficiency. Geneva, World Health Organization, P. 1-11.

Ye X, Al-Babili S, Klöti A, Zhang J, Lucca P, Beyer P, Potrykus I (2000): Engineering the Provitamin A (ß-Carotene) Biosynthetic Pathway into (Carotenoid-Free) Rice Endosperm. Science 287: 303-305.

Zagalsky PF. 1976. Carotenoid-protein complexes. Pure Appl Chemistry. 47:103.

Zagalsky PF, Haxo F, Hertsberg S, Liaaen-Jensen S. 1989. Studies on the blue caroteneoprotein, Linchiacyanin, isolated from the starfish *Linckia laevigata* (Hexinodermatai aesteriodea).