A genomic assisted breeding program for cassava to improve nutritional quality and industrial traits of storage root

Luiz Joaquim Castelo Branco Carvalho1*, Eduardo Alano Vieira2, Josefino de Freitas Fialho2 and Claudia Regina Batista de Souza3

Received 4 July 2011
Accepted 11 September 2011

ABSTRACT – Cassava is cultivated for two ends proposals: “sweet cassava” as fresh consumes and “industry cassava” as source of starch and farina. Landraces were used to discover “spontaneous mutations” and to develop evolutionary and breeding perspective of gene function. Genomic and Proteomic resources were obtained. Gene expression by RNA blot and Microarray analysis were performed to identify differentially expressed genes. A new sugary cassava was identified to be related to missing expression of BEI and a nonsense mutation in GBSSI gene leading to amylose free starch. A pink phenotype showed no expression of CasLYB gene, and a yellow phenotype a down regulation of CasHYb. Proteomic analysis of carotenoid-protein complex together with gene expression analysis of CAP4 revealed a heteroduplex double strand cDNA associated with high carotenoid content. GBSSI gene sequencing identified 22 haplotypes and large nucleotide diversity. Segregating populations by crossing differential biochemical phenotypes and parents adapted to Cerrado’s Region were obtained.

Key words: cassava, landraces, genome, microarray and proteomics.

INTRODUCTION

Cassava landraces are the earliest form of the modern cultivars and represents the first step in cassava domestication. We have been use this resource to discover spontaneous mutations in the sucrose/starch and carotenoid synthesis/accumulation and to develop both evolutionary and breeding perspective of gene function related to those traits because relatively few major genes are involved. Currently, such functional analysis use either forward or backward genetic approaches. Studies on genetics of cassava are rare, incomplete and most of the time difficult, because of long life cycle of the plant. Mutants have not been found in cassava. Some attempts have been reported the identification of phenotypes variants in landraces for starch type without genetic analysis (Carvalho et al. 2004) and laboratory-induced mutant in starch (Raemkers et al. 2005) and linamarin cleavage (Sirirungta et al. 2004) in a restrict number of genetic backgrounds.

An alternative to laboratory-induced mutants, in cassava, is identifying biochemical phenotypes in landraces in its center of origin and domestication (Schaal et al. 2005). Since cassava is not a selfing species most of the collected

1 Embrapa Cenargen, C.P. 02372, 70.770-970, Brasília, DF, Brazil. *E-mail: carvalho@cenargen.embrapa.br
2 Embrapa Cerrados, C.P. 08223, 73.310-970, Planaltina, DF, Brazil
3 Universidade Federal do Pará, Instituto de Ciências Biológicas, Laboratório de Biologia Molecular, Belem, PA, Brazil
plants are highly heterozygous being practically impossible to find an inbred line in nature that could carry a recessive character like a mutation in phenotype using conventional genetic analysis. Therefore exploitation of that genetic variation for either quantitative or qualitative trait needs molecular species specific tools for its genetic analysis. In considering the constrains to produce offspring and the cassava’s genome complexity our two complementary strategies is based on association genetics analysis using natural populations and single segregating population, the last involving more time and field experiments. Biochemical phenotypes variant was found in the Genebank that may serve for the definition of gene functions; understand gene regulation and generate marker for breeding program as well. Because resources are often limited, our study is performed in a two-stage approach by using a subset of samples to identify biochemical phenotypes and SNPs. Instead of genotype hundred of controls for the characterization of haplotype tag SNPs (htSNPs), we are genotyping a sample cases and carry out preliminary tests of association to aid the selection of htSNPs. Once the subset has been genotyped the whole set of loci will be tested for equilibrium to proceed. In addition cross populations based on a modified backcross breeding design to obtain single segregating populations are being prepared. Here, we summarize advances on this systematic exploitation of the naturally occurring variation as a complementary resource for the functional analysis of the cassava genome, to generate tools usefull for a breeding program to improve nutritional value of storage root and new utilization for industrial process as well.

**NATURAL VARIATION**

Agriculture, in the center of origin and domestication of cassava (CODC) in Brazil, is very different from the way cassava is grown in many other parts of the globe. Farms that plant “modern” elite varieties cultivate the crop in monoculture as a source of starch. Moreover, many anthropologists have noted that cassava is consumed in a variety of different ways in CODC region, from granular meal added to sauces (farinha), to flatbreads, fermented drinks, smoked cassava cake, starch egg balls, yellow juice (TUCUPI) and as a leafy vegetable. The multiple uses of cassava and the diverse modes of being cultivated in the Amazon suggest that the crop might have high levels of genetic diversity in this region as reported by Carvalho et al. (2001), and further suggest that the Amazon region may be a source of other traits for breeding program that would be particularly useful for addressing some of the nutritional needs of the developing world. We conducted several seasons of fieldwork in the lower Amazon basin to observe the uses of cassava and to collect the diverse cultivars associated with these different uses.

If various landraces are prepared in different ways and have different uses in CODC, there may be underlying biochemical variation that could benefit efforts to improve the modern crop. An astounding level of diversity was found for cassava landraces that occur in the southern Amazon region (Carvalho et al. 2000). Cassava roots consumed in this region vary tremendously, not only in shape, size, color, but also in anatomical structure and carbohydrate diversity and content. The biochemical bases for these differences were investigated and have already yielded several novel polysacharides that are not observed in the modern varieties. Examples include landraces with high free sugar (especially glucose), amylose-free starch, phytoglycogen, and other carbohydrates not typically found in cultivated cassava (Carvalho et al. 2004). Often these varieties are associated with specific food uses. Examples include sugary cassava, which is used to make glucose syrup for desserts, and the phytoglycogen-rich cassava, which is prepared as a baby food (phytoglycogen is much easier to digest than normal starch from modern cassava varieties). In addition of finding variants in carbohydrates composition and polymer structure, landraces in the Amazon vary widely in root pigments as well, some of which have important nutritional value. Other landraces sequester the antioxidants lycopene and lutein in high amounts as well. Variation in root pigments has been observed in germplasm collections previously; however, the Amazon varieties are unusual in the high amounts of pigment that are sequestered as well as in the type of pigments. As with the carbohydrate variants, the specific landraces that are high in pigments are associated with specific uses. For example, the high β-carotene landrace is used to make tucupi, a yellow juice that is used in soup; the roots are also boiled and served as a vegetable. This yellow variant is of special interest since it not only accumulate β-carotene, but also has 40 % higher protein content than the white cassava storage root (Carvalho et al. 2004). Thus, this variety has the potential to address some of the nutritional challenges in modern cassava crop to provide high quality food such as protein and pro-vitamin A content.
BIOCHEMICAL PHENOTYPES IN THE STORAGE ROOT

Either candidate gene or whole genome scanning approach can be used to develop molecular marker-trait association in a breeding program. However, a strategy considering candidate gene is more practical and appropriated to generate marker-trait associated to specific biochemical phenotype and its application in a breeding program for cassava. This approach requires a considerable knowledge of the physiology and biochemistry of the phenotype. This knowledge is available for starch and carotenoid accumulation in model plants as well as grain crops and has been applied successfully in the case of carotenoid candidate gene analysis in Solanaceae (Thourp et al. 2000). However, the biology of biochemical phenotypes are usually species-specific and vary with the organ and storage tissue studied as well as plant development. Consequently, different mechanisms of regulating starch and carotenoid accumulation are involved, including genetic background of the cultivar as well as general environmental. In our current study we focus on three biochemical phenotype (sucrose/starch pathway, carotenoid synthesis/accumulation and protein content) because those traits are determined by a relative small number of genes.

Sugary cassava

High free sugar content cassava storage root together with low starch and dry matter content define the variant sugary cassava (Carvalho et al. 2004) found in the Amazon. In addition this biochemical phenotype also presents variants in the amylose/amylopectin proportion as well as a new identified variant in the structure of α-polyglucan with short branch length and high density branching pattern. So far these structural variants have observed to be due to missing expression of the gene coding for SBEI and a nonsense mutation in the sequence of the gene coding for the enzyme GBSSI (Carvalho et al. 2004). Low yield of the sugary landrace, in Cerrado region of Brazil, has been associated with its low resistance to cassava bacterial blight (Xanthomonas axonopodis pv. manihotis) and local adaptation due to the seasonality of the regional weather. Crosses with local elite adapted cultivar have allowed us to improve the adaptation as well as the transfer of the sugary phenotype to local adapted variety.

Pigmented cassava

Naturally occurring color variation associated with carotenoid accumulation was observed in the storage root of landraces of cassava from the center of origin and domestication in Brazil. Carotenoid separation, identification and quantification by HPLC analyses showed that total β-carotene is the major carotenoid form present and accounts for 54 % to 77 % of the total carotenoid in cassava roots. The carotenoid biosynthetic pathway is fully activated in cassava storage root, including the white phenotype. No detection of α-carotene in 24 landraces studied was observed, but variable amounts of lutein (an α-ring xanthophyll) were present. Yellow color intensity variation was associated with the accumulation of different carotenoid. Landraces with white storage root showed a profile with eight types, whereas intense yellow showed 17 types. Variation in total β-carotene content ranged from none in landrace Mirasol (pink color) that accumulates only lycopene (99.81 μg g⁻¹ DWt), to 49.91 (μg g⁻¹ DWt) in landrace MC008 (deep yellow color) that accumulates β-carotene (66 %) its colorless precursor phytoene (31 %) and it’s derived intermediate β-criptoxanthin (3 %). Variations, in μ and β-ring xanthophyll content, preferentially accumulate high amounts of lutein or violaxanthin, together with β-carotene, such as in landraces MC002 and MC016, respectively. The possible mechanisms responsible for this genetic variation include genetic regulation of genes coding for enzymes in the syntheses pathway as observed with the pink landraces, variants in the yellow that differentially synthesizing β-carotene and β-ring xanthophylls (violaxanthin, neoxanthin, luteoxanthin and auroxanthin) but not α-ring xanthophylls (lutein) and carotenoid cleavage enzymes. In addition the differential accumulation of β-carotene, ranging from 0.07 to 49.91 μg g⁻¹ DWt, thought possible associated with the sink capacity of chromoplast related to carotenoid sequestering mechanism involving carotenoid-proteins association.

Protein content enhancement and its association with carotenoid accumulation

Protein content in cassava storage root is usually low (1-2 %) in the commercial varieties. Recently (Carvalho et al. 2008) we reported a six fold higher value for proteins content in a plastid fraction of intense yellow cassava root when compared to white cassava. This observation correspond to an increase in more the 40 % in protein content in the bulk of the storage root of cassava. Correlation studies between total carotenoid and protein content indicates a highly significant positive correlation.
between root protein content in the plastid fraction and total carotenoid content.

Biochemical tools so far developed or used in this breeding program includes free sugar detection kit (specifically for glucose) for fast field test, starch iodine field test to differentiate amylase free starch, glycogen like starch, etc. A diagnostic test kit to differentiate high carotenoid accumulation and protein content are being designed and developed.

In addition, the diversity observed in those biochemical phenotypes are the subject of our studies using genomic and proteomic technologies to establish resources, develop tools and technology to determine gene function to assist a breeding program to improve cassava storage root quality and new industrial utilization.

**GENOMIC RESOURCES AND TOOLS TO STUDY DIVERSITY**

Cytogenetic studies using conventional staining techniques (looking at chromosome number or morphology) and cytological markers (C-band pattern, CMA/DAPI fluorochrome bands, prophase chromosome condensation pattern, rDNA sites and the maximum number of nucleoli) to identify cytogenetic differences between Manihot species shows that cassava exhibit a chromosome number of n = 18 or 2n = 36 and a genome size of 770Mbp. Mitotic karyotypes show a very similar chromosome sizes, varying from 1.23 to 2.41 mm per karyotype and a variable number of satellites (0-4) chromosomes (Carvalho and Guerra 2002). C-banding pattern display an identical band pattern, fluorochrome staining revealed eight chromosomes with a CMA+band, four of them coinciding with the satellites. In situ hybridization revealed one 45S rDNA site in the terminal regions of six chromosomes. Four sites showed a similar size, whereas the other two sites were smaller. All satellites observed with DAPI were labeled with the probes SK 18S+25S. A 5S rDNA site was observed in a sub terminal position of two medium sized chromosomes. Together, this information will facilitate to adopt a chromosome sorting technology to sequence cassava genome. The chloroplast genome of *M. esculenta* is 161,453 bp in length which includes a pair of inverted repeats (IR) of 26,954 bp with 128 protein coding genes being 96 single copy and 16 duplicated (Daneil et al. 2008). The genome consists of 49.82 % protein coding, 1.7 % tRNA, and 5.6 % rRNA genes, and 42.87 % non-coding sequence. The G+C and A+T contents in the cassava chloroplast genome are 35.87 and 64.13 %, respectively. The overall A+T content is similar to poplar (63.26 %), tobacco (62.2 %), citrus (61.52 %), maize (61.5 %), and rice (61.1 %). The cassava chloroplast genome has the ancestral angiosperm genome organization and in particular is co-linear with Populus, its closest fully sequenced relative.

There are four rRNA genes and 30 distinct tRNAs, seven of which are duplicated in the IR. The infA gene, coding for translation initiation factor 1, is absent in the cassava chloroplast genome. The phylogenetic distribution of atpF intron is also missed in the cassava closely related Euphorbiaceae Hevea and Elateriospermum. Although this information will allow genome-based biotechnology using plastid genome we are still missing the fully nuclear genome sequence resources for cassava. Nuclear cassava genome sequence analysis, resources and tools development are in its infancy. A public institutional consortium with 14 members centralized a pilot project through the DOE-JGI Community Sequencing Program (CSP). This pilot has sequenced variety TMS30572 with the assemble of three Bacterial Artificial Chromosome (BAC) libraries using varieties TME3, AM560-2, MECW72. The deliverables so far are available at NCBI (http://www.phytozome.net/cassava). Complementary to these resources, EMBRAPA and CATAS/BIG have started to sequence the genome of the cassava ancestor (*M. falbellifolia*) and sugary cassava (Cas36.1).

Three molecular maps for cassava have been constructed using intraspecific F1 cross with 612 markers, a self-incompatible family with 100 markers, and a backcross family with 132 markers (http://www.ncbi.nlm.nih.gov/bioproject/15581). More than 300 RFLP, 600 SSR, 120 RAPD, and 9 isozyme markers have been employed to study genetic diversity and structure in a large collection of African and Latin American cassava varieties.

Several cDNA libraries have been constructed from varieties selected for high starch content (MPer183, CM523-7), bacterial blight resistance (MBra685, SG07-35, MCol11522), virus resistance (TME3), and drought tolerance (TME17). EST database (36261 entry) has been assembled in Japan (http://amber.gsc.riken.jp/cassava/), USA (http://titan.biotec.uiuc.edu/cassava/ and NCBI http://www.ncbi.nlm.nih.gov/bioproject/13644). Additional six cDNA libraries and EST database have been prepared using sugary cassava (Cas36.1 and Cas36.4) and pigmented cassava (Mirasol and MC008) in Brazil (http://genomaembrapa.br/genoma/).
Up to now the GENOMIC resources listed above have been used to manufacture a cDNA chip array tool with 25,392 elements and 23,937 unigenes. This cDNA array has been tested to understand the biochemical phenotype diversity observed above as well as to identify genes associated with traits related to domestication of cassava such as growth habit, flowering set, and storage root formation.

PROTEOMIC RESOURCES AND TOOLS TO STUDY DIVERSITY

Chromoplast enriched suspension, carotenoid-protein complex separation by conventional size exclusion chromatography (SEC), protein fractionation in SDS-PAGE, and shotgun proteomics technology were used to identify and characterize carotenoid-proteins complex from storage root of intense yellow cassava landrace. A non-denatured carotenoid-protein complex was separated in fractions 38-57 (peak 1) and a non-pigment-associated protein in fractions 109-134 (peak 2). Proteins separation in SDS-PAGE for each peak were fractionated, tryptic digested and peptide sequences established directly from the MSMS spectrum and peptide mass analysis using MASCOT search engine. Protein identification used public sequence databases available in local MASCOT server together with cDNA sequences from four cassava EST databases (ESTIMA, a domestic EST database, the RIKEN full-length cDNA database) that were translated into all six reading frames and protein name annotated in Arabidopsis data base. Peptide sequences gave 83 and 106 peptides from peak 1 and peak 2, respectively, and 26 and 39 identified proteins as well. Small Heat Shock Proteins (sHSP) of three classes were the most abundant protein present in the carotenoid-protein complex. Proteins sequences were performed with phylogenetic analysis to predict classes of protein function and family type using Pfam server. Proteins from SEC peaks (1 and 2) from white and intense yellow root were characterized by western blot using antibody against fibrillin and Or_protein. Results indicated that Fibrillin and Or_protein were detected in chromoplast suspension and peak 2 of intense yellow root but not in peak 1 and white root at all. A specific candidate carotenoid-associated protein, related to carotenoid accumulation, was identified and the expression of its corresponding coding gene analyzed across contrasting color phenotype white and intense yellow. Cumulatively the results support the correlation between carotenoid accumulation and protein content in the root tissue of pigmented cassava as previously observed (Carvalho et al. 2004) and may provide a new research avenue of two fold importance (i.e. increase proVitA carotenoid and protein content) to improve nutritional value for cassava.

GENETIC ANALYSIS OF SELECTED TRAITS

Our forward genetic analysis is focused in two alternatives experimental designs. In the first one a population genetic analysis is carried out with candidate gene derived from the biochemical phenotype as describe above and tested under a evolutionary perspective in a subset sample of population including cassava ancestor (33 individuals) and landraces of cassava (121 individuals). In the second crossing populations were obtained for mapping, field evaluation and new cultivars are being prepared. Two candidate genes coding for starch synthesis (CasBEI and CasGBSSI) and three for carotenoid synthesis (CasPSY, CasLCYb and CasHYb) were selected as candidate gene to be sequenced across a population of 154 individuals. Our preliminary results analyzes in a subset sample for the N-terminal region of the CasGBSSI showed that large nucleotide diversity is observed in the cassava ancestor followed by the pigmented cassava. Tagima’s D value was highly significant for the combined subset sample indicating genetic neutrality. Haplotype number and diversity was also high for combined and ancestor subset. The likely genealogical history of the genetic diversity observed was inferred from RM networks constructed for haplotypes in the combined subset. An association of a specific set of haplotype was identified that confirmed the grouping pattern for normal cassava, pigmented cassava, sugary cassava and high HNC cassava. Provenance test varieties selection cycles in single segregating populations

Field test for agronomic performance and morphological descriptors has been performed in the Cerrado region of Brazil since 2002 in a provenance test (Figure 1). Results showed identical sugary and pigmented phenotype (Fialho et al. 2009, Vieira et al. 2009) as observed early in the center of origin and domestication. In 2006 a breeding program was initiated (Figure 2) to transfer these new identified traits to local varieties. In addition, RAPD and SSR markers were also applied and showed a large genetic diversity among the sugary accessions (Vieira et al. 2008). Agronomic performance (Vieira et al. 2009) indicated that the landrace Cas36.17 (sugary cassava) yielded about 9ton ha⁻¹ (12
month growth season base) while the best adapted local variety (cv. Japonesinha) yielded 28ton ha\(^{-1}\). All the other sugary access showed mainly lower yield due to a severe attack of *Xanthomonas axonopodis pv. manihotis*. This information was used to better orient the selection of landrace within individuals obtained in the crosses populations (F\(_1\) segregating, open pollinated crosses and self pollination population). Preliminary results for a half...
A genomic assisted breeding program for cassava to improve nutritional quality and industrial traits of storage root

sibling population (55 individuals) tested in of 2006/2007 growth season revealed segregation for sugary phenotype (93 % of the individuals), waxy phenotype (38 % of the individuals) and glycogen like starch (2 % of the individuals). Individuals (185) of the S1 population for the best adapted sugary landrace were planted in the 2007/2008 growth season in the Cerrado region and are being evaluated as above.

FINAL REMARKS

The combination of landrace diversity, biochemical phenotype characterization, gene expression analysis, gene sequence and protein sequence allowed us to
- Identify SNP and haplotypes in a subset sample leading their association with four distinct phenotypes.
- Identify phenotype mutation in the sucrose/starch conversion pathway associated with missing-expression of the gene coding for BEI and a nonsense mutation in the gene sequence coding for GBSSI.
- Identify phenotype mutation in the carotenoid synthesis associated with the missing-expression of genes coding for HyB that block the pathway and a synthesis pathway flux regulation.
- Identify a heteroduplex double strand cDNA associated with high carotenoid accumulation.
- Speed up the selection cycle in the conventional breeding program and obtaining of a new variety in five years.


Palavras-chave: mandioca, raças locais, genômica, proteômica, microarranjos.

REFERENCES


LJCB Carvalho et al.


