Mapeo genético y análisis de QTLs para carotenos en una población s_1 de yuca

Genetic mapping and QTL analysis for carotenes in an S_1 population of cassava

Jaime Alberto Marín Colorado[1], Hernando Ramírez[2], Martin Fregene[1].

[1]International Center for Tropical Agriculture –CIAT, AA 6713, Cali, Colombia.
[2]Facultad de Ciencias Agropecuarias, Universidad Nacional de Colombia, AA 237, Palmira, Valle. Colombia. Author for correspondence: m.fregene@cgiar.org

REC.: 21-08-08 ACEPT.: 06-02-09

RESUMEN

La población S_1 de la variedad tailandesa MTAI8 (AM320), la cual presenta patrones de segregación definidos para el contenido de carotenos totales (Beta-caroteno), se sometió a un análisis de agrupamiento segregante (B.S.A = Bulk Segregant Analysis) empleando 700 marcadores moleculares tipo microsatélites o SSR. Se generaron 25 grupos de ligamiento identificando 3 QTLs mayores asociados con una región del genoma de yuca con el contenido de carotenos totales. Tres marcadores SSR; explicaron el 37.2% (NS109), 32% (rSSRY251), 27.7% (rSSRY313) de la varianza fenotipica total, situados en el grupo de ligamiento D fuertemente asociados con el contenido de carotenos totales en la familia AM320 S_1, se estableció una fuerte correlación positiva entre color de pulpa de raíz y contenido de carotenos totales (r=0.81), y negativa entre carotenos totales y materia seca (r= -0.31). Los tres posibles QTLs presentaron efecto positivo y de carácter aditivo para el contenido de carotenos totales, puede ser la oportunidad para implementar selección asistida por marcadores para carotenos totales en yuca.

Palabras clave: *Manihot esculenta* Crantz; caroteno; QTLs; SSR.

ABSTRACT

The S_1 population of the Thai cassava variety MTAI8 (AM320), which shows patterns of segregation defined for the content of total carotenes (beta-carotene), underwent an analysis of segregating bulk (BSA = Bulk Segregating Analysis) using 700 molecular markers of the type microsatellites or SSR. 25 linkage groups were generated, identifying 3 major QTLs associated with the same region of the cassava genome as total carotene content. Three SSR markers explained 37.2% (NS109), 32% (rSSRY251), and 27.7% (rSSRY313) of the total phenotypic variance. They were placed in Linkage Group D and strongly associated with the total carotene content in the family AM320 S_1. A strong positive correlation was established between root pulp color and total carotene content (r = 0.81), and a negative correlation between total carotene content and dry matter (r = -0.31). The three possible QTLs showed a positive effect and were additive for the total

carotene content, and may be an opportunity to implement marker assisted selection for total carotenes in cassava.

key-words: *Manihot esculenta* Crantz; carotene, QTLs, SSR-microsatellites.

INTRODUCTION

Cassava (*Manihot esculenta* Crantz), a staple food in Latin America since the pre-Columbian period, and introduced to Africa in the 16th Century (FAO, 2000), is a poor source of macronutrients, and a few micronutrients: in particular vitamin A (Kawano, 2003); but also vitamins C, B2, B6, and magnesium and potassium.

The content of beta-carotene is a key precursor in the synthesis of vitamin A. It is characterized by high heritability and its presence is detected through root color. Genetic mapping and QTL analysis are basic steps required to understand the genetic behavior of this trait. With this information, and exploiting the high genetic variability that exists in the crop, attempts can be made to improve the carotene content in cassava through conventional methods, and also using genetic transformation approaches (Fregene *et al.,* 1997).

The objectives of this study were to: construct a genetic map of the total carotene content; locate markers strongly associated with total carotene content; and establish the regions in the genome (QTL) that are associated with carotene content.

MATERIALS AND METHODS

The study was carried out in CIAT (Palmira, Cauca Valley Department, Colombia; 3°16N, 76°32'E, & 965 MASL) with an S_1 population produced by a self-fertilization of a Thai variety MTAI-8.

The experiment was planted in 2005 and 2007, in a completely randomized block design with two replicates, in simple plots with six plants per furrow, with a distance of 0.8 m and 1 m between rows. The variables dry material (DM), total carotene content (TCC) and root pulp color (RPC) was evaluated after eleven months.

DNA extraction was performed using the mini-extraction protocol (Dellaporta, 1983). Bulk segregant analysis (BSA) was carried out (Michelmore *et al*., 1991), which generated two groupings: orange roots (10 genotypes), and white roots (10 genotypes). The DNAs from the groupings and from the parents were genotyped with 700 variable SSR markers for cassava (Mba *et al.*, 2001). In the evaluations where polymorphism was maintained in the individual as well as the groupings, the markers were evaluated in all the family, S_1 AM320 (TAI8), consisting of 229 plants. 140 polymorphic SSR were submitted to genetic mapping and QTL analysis.

Linkage analysis and genetic mapping was performed using a LOD score of 4.0, and a recombination fraction of 0.3 (Fregene *et al*., 1997). The genetic map was constructed with 140 SSR loci, using the computer program MAPMAKER 2.0, run on a Macintosh Centris 650, and a version, Mapmaker 3.0 (Lander *et al*., 1987). For the QTL mapping in the segregant S_1 population single marker analysis was applied with the software packet QGENE (Nelson, 1997). A QTL was established with a significance of $P \leq 0.005$. The results of R^2 represent the total phenotypic variance explained. For the QTL analysis, the T-test statistical model was used (Liu y Liu, 1995), in which three or more linked markers were submitted to multiple regression analysis.

For the QTL analysis a study was established with three models: S.M.A. = Single Marker Analysis; S.I.M. = Simple Interval Mapping; C.I.M. = Compound Interval Mapping. This type of analysis establishes three categories to confirm a QTL in a genomic region. For the QTL mapping, Interval Mapping was used employing a LOD algorithm ≥ 3.0 with the software packets MAPMAKER/QTL y QGENE.

RESULTS Y DISCUSSION

17 possible QTLs were found in the linkage groups (1, 2, 3, 4, 6, 11, 15, 16, 19, 22) (Tables 1, 2, 3), but only three QTLs for CCT-2008 (Total Carotene Content); four QTLs for CPR-2006-2008 (Root Pulp Color), located in the linkage group 1, and presenting significance of $P \leq 0.001$. Linkage group 1 had a LOD score of 33 for CCT-2008, LOD 68 for CPR-2006, and LOD 78 for CPR-2008 (Figure 1a). Three QTLs explained 32.3% of the phenotypic variance for the total carotene content. Four QTLs that affected the root pulp color – 2006 explained 41.3%, and four QTLs for root pulp color -2008 explained 37.7% (Table 4).

TablE 1. Marker loci associated with total carotenes (2008) for simple regression analysis with QGENE.

MARKER	L.G.	P	V.E.
rSSRY313	3	0.0000	27.7%
rSSRY251	3	0.0000	32.0%
NS109	3	0.0000	37.2%
SSRY66	3	0.0000	16.4%
NS717	3	0.0000	16.9%
SSRY9	3	0.0000	12.7%
NS980	3	0.001	7.3%
SSRY92	3	0.0031	6.5%
rSSRY60	3	0.054	3.3%
rSSRY95	6	0.0359	3.4%
rSSRY223	11	0.0424	3.3%
SSRY177	11	0.0275	4.4%
SSRY272	18	0.0347	3.5%
SSRY242	24	0.0109	4.7%

L.G. = Linkage Group; P = probability value; V.E. = Variance explained
*Statistical significance at the levels ≤0.01.

Table 2. Marker loci associated with root pulp color (2006-2008) for simple regression analysis with QGENE.

MARKER	L.G.	P	V.E.
R.P.C.-2006			
rSSRY313	3	0.0000	41.4%
rSSRY251	3	0.0000	50.2%
NS109	3	0.0000	45.0%
SSRY66	3	0.0000	26.7%
NS717	3	0.0000	28.5%
SSRY9	3	0.0000	23.5%
NS980	3	0.0001	10.0%
SSRY92	3	0.0042	6.6%
rSSRY60	3	0.0085	5.7%
NS189	5	0.0249	4.1%
rSSRY241	6	0.0462	4.4%
rSSRY223	11	0.0129	5.0%
NS272	11	0.0334	3.9%
NS306	14	0.0524	3.4%
SSRY94	20	0.0424	4.1%
SSRY242	24	0.0424	3.7%
R.P.C.-2008			
rSSRY313	3	0.0000	37.2%
rSSRY251	3	0.0000	44.7%
NS109	3	0.0000	42.2%
SSRY66	3	0.0000	26.1%
NS717	3	0.0000	26.7%
SSRY9	3	0.0000	21.5%
NS980	3	0.0001	9.5%
SSRY92	3	0.0051	5.9%
rSSRY60	3	0.0260	4.1%
rSSRY241	6	0.0424	4.1%
rSSRY62	9	0.0266	4.2%
SSRY205	9	0.0366	3.3%
rSSRY223	11	0.0312	3.6%
SSRY181	15	0.0229	4.2%
SSRY272	18	0.0272	3.8%

L.G. = Linkage Group; P = probability value; V.E = variance explained.
*statistical significance at levels ≤0.01.

Table 3. Marker loci associated with dry material (2006 – 2008) for simple regression analysis with QGENE.

MARKERS	L.G.	P	V.E
M.S-2006			
SSRY12	2	0.0281	4.4%
NS717	3	0.0024	7.2%
SSRY9	3	0.0003	10.6%
rSSRY251	3	0.0168	4.4%
NS109	3	0.034	5.0%
rSSRY313	3	0.0048	5.8%
rSSRY70	4	0.0472	3.4%
SSRY141	6	0.0355	3.8%
SSRY183	6	0.0162	4.5%
M.S-2008			
NS980	3	0.0342	4.2%
NS717	3	0.0034	7.5%
SSRY9	3	0.0191	5.7%
rSSRY251	3	0.0157	4.8%
rSSRY313	3	0.0175	4.8%
NS306	14	0.0494	3.6%
NS194	25	0.0085	6.1%

L.G. = Linkage Group; P = probability value; V.E = variance explained.
*statistical significance at levels ≤0.01.

Table 4. Individual QTLs that affect the total carotene content (TCC) and root pulp color (RPC) in the S_1 population, through analysis of MIC-1000 permutations, with WinQTLCart.

Trait	QTL	SSR Interval	L.G.	LOD
TCC	**QCCT1**	**rSSRY66-rSSRY313**	**3**	**8**
	QCCT2	**rSSRY313-NS109**	**3**	**33**
	QCCT3	**NS109-NS717**	**3**	**17**
RPC-2006	**QCPR6-1**	**rSSRY66-rSSRY313**	**3**	**28**
	QCPR6-2	**rSSRY313-NS109 NS109-**	**3**	**68**
	QCPR6-3	**rSSRY251 rSSRY251-**	**3**	**34**
	QCPR6-4	**NS717**	**3**	**30**
RPC-2008	**QCPR8-1**	**rSSRY66-rSSRY313**	**3**	**33**
	QCPR8-2	**rSSRY313-NS109 NS109-**	**3**	**78**
	QCPR8-3	**rSSRY251 rSSRY251-**	**3**	**31**
	QCPR8-4	**NS717**	**3**	**28**

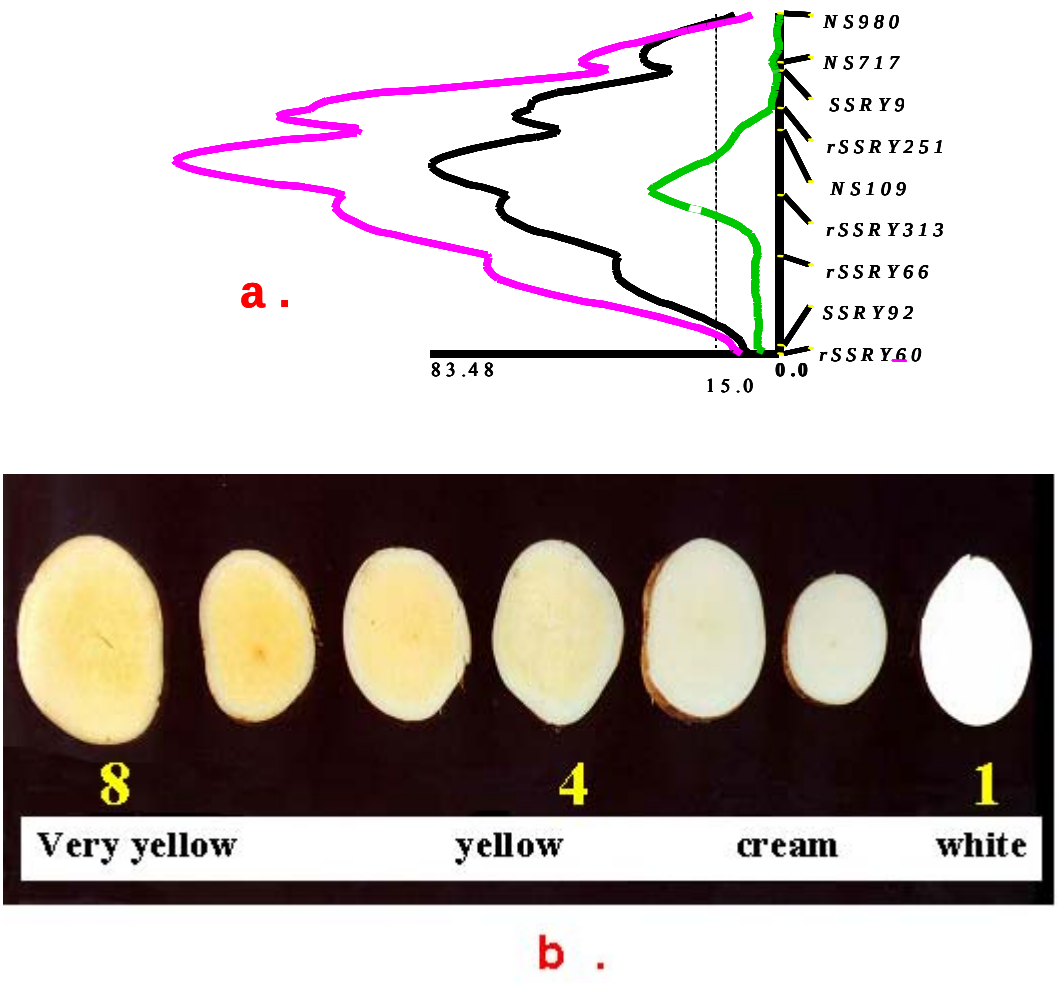


Figure 1. a. Three possible QTLs in linkage group 1, explaining between 20 - 40% of the phenotypic variation, and presenting an additive effect.
b. Range of colors of the cassava root with a numerical range for cassava carotene content.

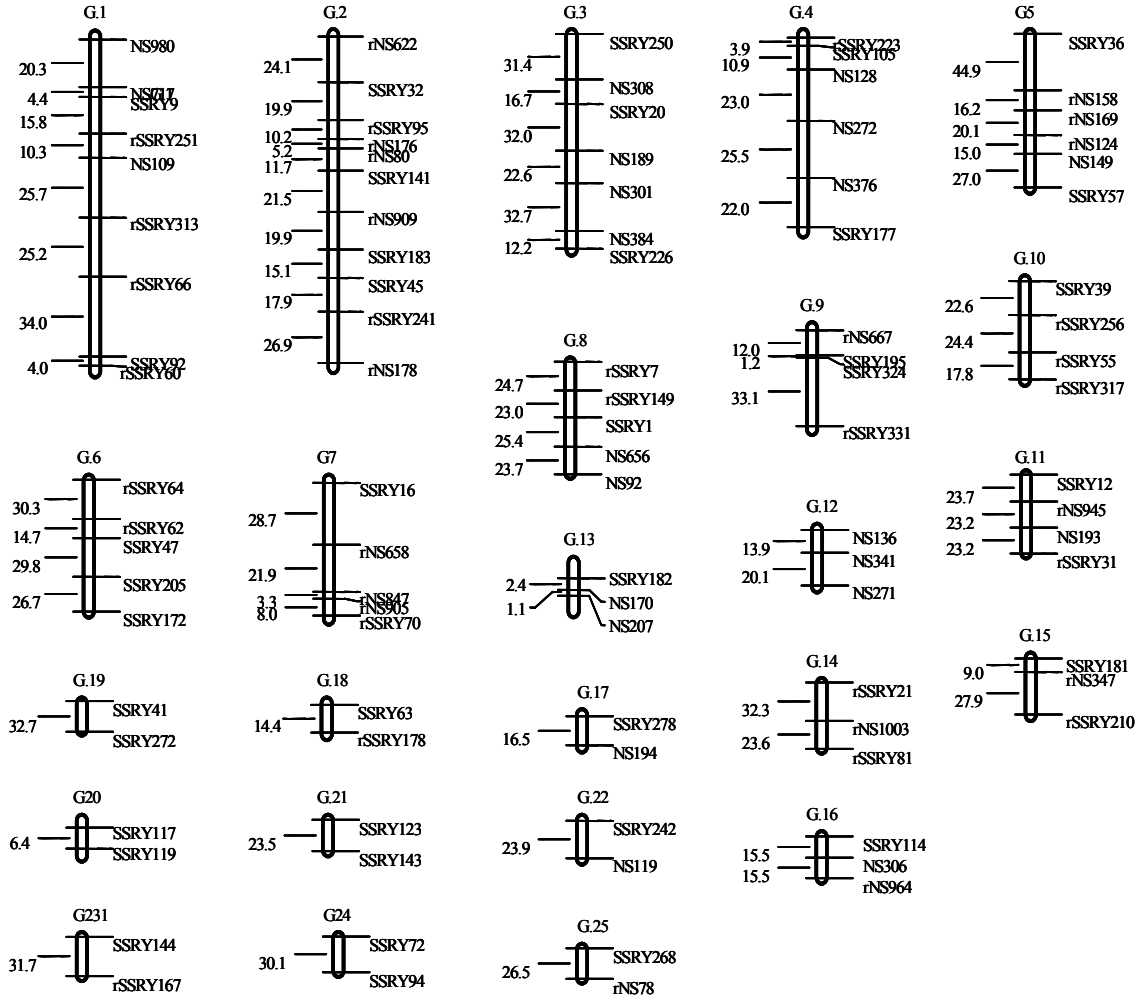


Figure 2. Molecular genetic map for cassava carotene content based on microsatellites for the S_1 population (AM320) with 229 individuals.

The genetic map, which consisted of 25 linkage groups, covered 1476.6 cM, with an average distance between markers of 21.5 cM (Figure 2); 37 SSR markers remained unlinked. 12 microsatellites (11.6%) segregated normally with a significance of 0.001, 43 microsatellites (41.7%) with a significance of 0.01, 22 microsatellites (21.3%) with a segregation of 0.05 significance and 26 microsatellites (25.2%) did not segregate with the expected ratio of 1:2:1. Only 12 groups segregated normally. The most extreme cases of

segregation distortion in the population S_1 were observed for the marker NS980 in the linkage group 1, where only 5 plants of 229 were homozygote for B.

In order to construct the map, the 26 microsatellites that did not segregate as expected were also used, given that biological factors exist, such as: systematic markers; genotype classification errors; chromosomal rearrangements (Gebhardt *et al.*, 1991; Kianian y Quiros, 1992); gene defects; and the effects of two or more recessive genes for lethality (Liu, 1998). Additionally, segregation is a consequence of natural selection during fertilization, plant growth, gametogenesis and seed development (Jung *et al*., 1996; Xu, *et al*., 1997; Nacagahra, 1972; Harushima, *et al.,* 1996; Garris, *et al.,* 2003), all factors that cause deviations from the proportions expected under Mendelian laws. In the genomic composition of the S_1 population, 27% of homozygote individuals presented the allele (**A**), 46.3% of heterozygote (**H**), and 26.6% were homozygote individuals with the allele (**B**).

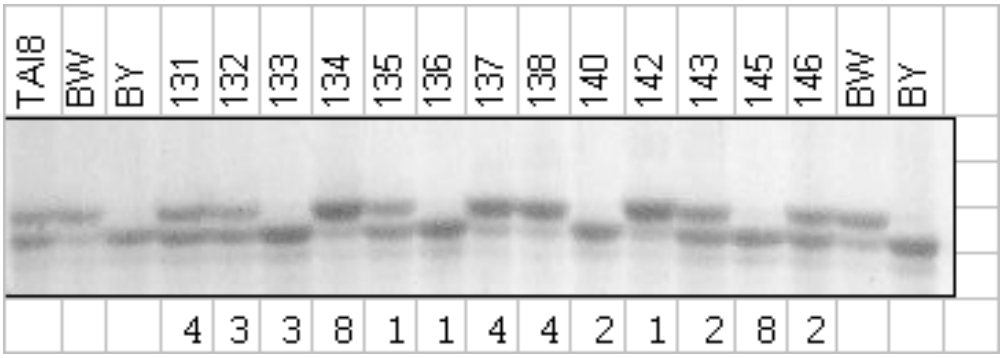
Three QTLs were detected in linkage group 1, consisting of 6 SSRs that explained: 27.7% (rSSRY313), 32% (rSSRY251), 37.2% (NS109), 16.4% (SSRY66), 16.9% (NS717), 12.7% (SSRY9) of the total phenotypic variance. One of the markers strongly associated with total carotene content was SSRY251, which explained 32% of the phenotypic variance. The segregation of this marker and the color evaluation (Figs. 1b, 3a y 3b) were highly correlation with total carotene content ($r \geq 0.8$). It can be seen that, depending on the color intensity, when the color range is 8, it is associated with homozygosity for the small allele of SSRY251 in this family. Three possible QTLs were located in the linkage group 1 between the intervals: QCCT1 (rSSRY66-rSSRY313); QCCT2 (rSSRY313-NS109); QCCT3 (NS109-rSSRY251), which pertain to group D of the cassava genetic map (Fregene *et al*. 1997). It could be said that a region of the cassava genome is associated with total carotene content.

Total phenotypic variation for the mapped S_1 population was 32.3% for the total carotene trait, and 39.5% for root pulp color. Three QTLs were detected for TCC (Total carotene content): QCCT1, QCCT2, QCCT3, and 4 QTLs with an additive effect for RPC (root pulp color): QCPR1, QCPR2, QCPR3, QCPR4. These genetic components represent the

most important QTLs for explaining phenotypic variation. As a consequence of the presence of the three possible gene states in this S_1 population, the effect of the different gene states on the phenotype could be determined.

Five probable secondary QTLs affected all the traits (TCC, RPC, DW), indicating the presence of regions in the linkage groups with multiple QTL loci for correlated traits. A number of the QTLs identified in more than one trait coincide in genomic region and direction of phenotypic effect, suggesting pleiotropy, a type of effect that could be responsible for both negative and positive correlations between the traits: the negative correlation resulting from a pleiotropic effect being TCC-DM; RPC-DM (-0.31; -0.31 ≤ 0.01) and a positive correlation being TCC-RPC (0.81 ≤ 0.01).

Primary and secondary interactions produced amongst QTLs (digenic interactions) can be explained by epistatic factors, as it is possible that some of the QTLs for the traits that constitute total carotenes could have a behavioral effect as modifying factors, contributing to the genotypic correlation. Epistasis will produce genetic erosion, and reduce vigor in related species. In many studies it has been observed that the production of progeny from crosses between parents of the same species or subspecies is lower than that of the parents (Li *et al*., 1996). In this study epistatic effects produced among many digenic interactions could be a good genetic basis for the carotene trait in the cassava crop. The results of the digenic interactions tend to increase the carotene content.



a.

b.

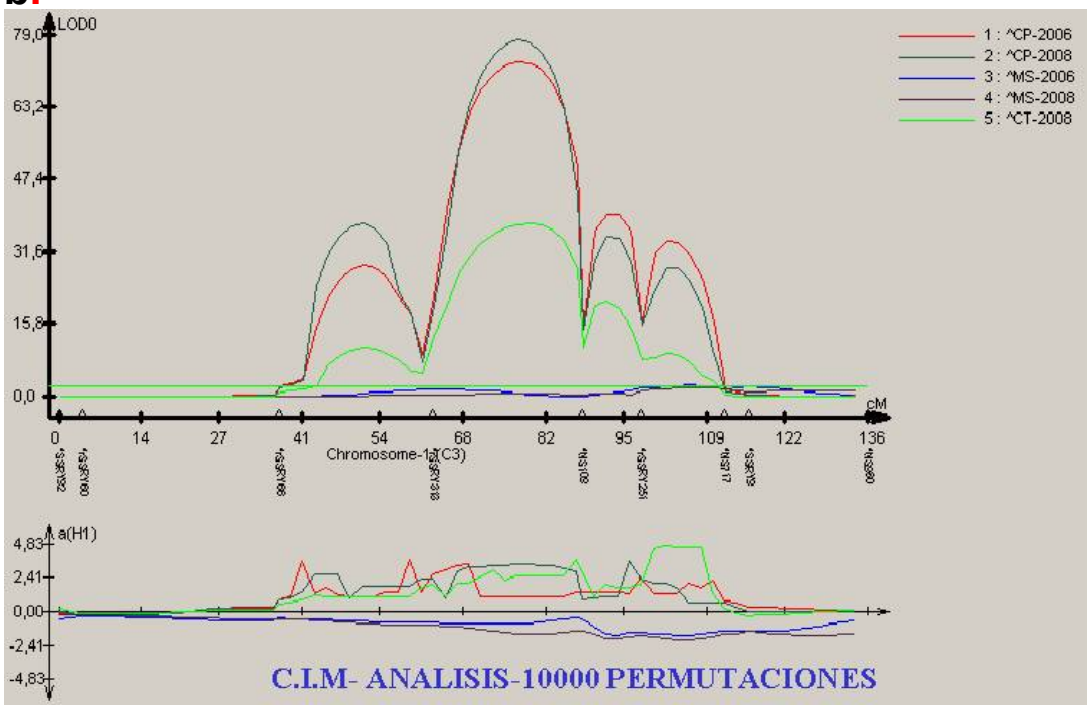


Figure 3. a. Silver-stained polyacrylamide gel showing the amplification of the SSR marker SSRY251 in some individuals of the family AM320. Root color codes are: 8 = orange/pink. 1-2 = white. 4-5 = cream.
b. C.I.M. (Compound interval mapping) analysis.

The S_1 population used in this study is very effective in detecting multiple QTLs in a single linkage group that affect the same trait. Additionally, with this type of population, additive factors can be distinguished from dominant ones. As it is probable that the S_1 population could be misleading as a consequence of the considerable variation in quantitative traits, especially when there is maximal linkage disequilibrium, it is necessary to continue studies of these effects in recombinant generations in order to determine with greater confidence the positive, negative and neutral effects produced by each allele, and to determine whether they are based on digenic interactions combined in the analysis of different regions.

Two genes are involved in the biosynthesis of beta-carotene: phytoene synthase and phytoene desaturase (Hirschberg, 1998). Correa *et al*. (2004), using the same S_1 population (AM320) with the yellow variety MTAI8, mapped phytoene synthase. This locus explained 30% of the phenotypic variance, and was situated within a different

linkage group (5 for S_1 and G for the cassava F_1) to the possible QTLs associated with the cDNA SSR markers established in this study. In contrast, the simple regression analysis between phytoene synthase, as an independent variable, and the beta-carotene content, as the dependent variable, did not show association with the gene that explained 30% of the phenotypic variance for beta-carotene content. The results suggest the existence of other genes that act on the inheritance of beta-carotene content in cassava.

QTLs identified and linked to the total carotene in cassava are a valuable addition to the Marker Assisted Selection scheme for the improvement of nutritional traits related to human health. Marker Assisted Selection for total carotenes could contribute to the identification and selection of superior genotypes, with the objective to reduce time and costs in a crop improvement program.

CONCLUSIONS

Using the cassava population S_1, a genetic map was constructed for total carotene content using microsatellite (SSR) markers. This map served as a base for QTL analysis.

Seven SSR markers (rSSRY313, rSSRY251, NS109, SSRY66, NS717, SSRY9, NS980) located in linkage group 1 and associated with total carotene content in the family S_1 AM320 explained between 30 and 40% of the phenotypic variance.

Three major QTLs for total carotene were detected and localized, with the respective action and effect on the trait.

ACKNOWLEDGEMENTS

To the CIAT project “Harvest Plus” which financed the Master’s thesis of the biologist Jaime Alberto Marín Colorado, from which was derived this article. To Dr. Martin Fregene for believing in this work. To the members of the Cassava Genetic Laboratory, especially to Janneth Gutiérrez, Adriana Alzate, Adriana Núñez y Diana Falla. To the

field workers in the Cassava Genetic Program, Jairo Valencia, Armando Bedoya, Antonio López for their collaboration and advice.

To the CIAT Biochemical Laboratory, especially to Alba Lucia Chávez, for their collaboration with the analysis of phenotypic and total carotene data.

BIBLIOGRAPHY

1. Correa, A.; Barrera, E.; Castelblanco, W.; Morante, N.; Ceballos, H.; Tohme, J.; Fregene, M. 2004. Genetic Mapping of Genes Involved in the Biosynthesis of Beta-carotene. Inform Annual. Cali, Colombia: CIAT. p30-38.

2. Dellaporta, S.L.; Wood, J.; Hicks, J.R. 1983. A plant DNA minipreparation: version II. *Plant Mol Biol*. Rep 1: 19-21.

3. FAO, 2000. The global economy of Cassava, facts, tendencies and perspectives. Rome-Italy. p32-37.

4. Fregene, M.A.; Angel, F.; Gómez, R.; Rodríguez, F.; Roca, W.; Tohme, J.; Bonierbale, M. 1997. A molecular genetic map of cassava (*Manihot esculenta* Crantz). *Theor Appl Genet*. 95 (3): 431-441.

5. Garris, A. J.; Mccouch, S.R.; Kresovich, S. 2003. Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (*Oryza sativa* L). *Genetics* 165: 759-769.

6. Gebhardt, C.; Ritter, E.; Barone, A.; Debener, T.; Walkemeier, B.; Schachtschabel, U.; *et al*. 1991. RFLP maps of potato and their alignment with the homoelogus tomato genome. *Theor Appl Genet* 83(1): 49-57.

7. Harushima, Y.; Kurata, N.; Yano, M.; Nagamura, Y.; Sasaki, T.; Minobe, Y.; *et al*. 1996. Detection of segregation distortions in an indica-japonica rice cross using a high-resolution molecular map. *Theor Appl Genet* 92 (2): 145-150.

8. Hirschberg, J. 1998. Molecular biology of carotenoid biosynthesis. p177-180. *In*: Britton, G.; Liaaen-Jensen, S.; Pfander, H. (eds). Carotenoids. Vol 3. Berlin: Birkhäuser Verlag.

9. Jung, G.; Coyne, P. W.; Shroch, J.; Nienhuis, E.; Arnaud-Santana, J.; Bokosi, H. M.; *et al*. 1996. Molecular markers associated with plant architecture and resistance to common blight, web blight, and rust in common beans. *J Am Soc Hort Sci* 121 (5): 794-803.

10. Kawano, K. 2003. Thirty years of cassava breeding for productivity – biological and social factors for success. *Crop Sci* 43 (4): 1325-1335.

11. Kianian, S. F.; Quiros, C. F. 1992. Generation of a *Brassica Oleracea* composite RFLP map: linkage arrangements among various populations and evolutionary implications. *Theor App Genet* 84(5-6): 544-554.

12. Lander, E.S.; Green, P.; Abrahamson, J.; Barlow, A.; Dal, M. J.; Lincoln, S. E.; Newberg, L. 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1:174-181.

13. Li, Z. K.; Pinson, R. M.; Park, W. D.; Paterson, A. H.; Stansel, J. W. 1996. Epistasis for Three Grain Yield Components in Rice (*Oryza sativa L*). *Genetics* 145: 453-465.

14. Liu, B. H.; Liu, Y. Y. 1995. PGRI, a software for plant genome research. Plant Genome III. Conference abstract, p 105.

15. Liu, B. 1998. Statical genomics. Linkage, Mapping and QTL analysis. Boca Raton: CRC Press.

16. Mba, R.; Stephenson, P.; Edwards, K.; Melzer, S.; Mkumbira, J.; Gullberg, U.; *et al*. 2001. Simple Sequence Repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. *Theor Appl Genet* 102(1): 21-31.

17. Michelmore, R. W.; Paran, I.; Kesseli, R. V. 1991. Identification of markers linked to disease – resistant genes by bulked segregant analysis: a rapid method to detect markers in specific genome regions by using segregating populations. *Proc Natl Acad Sci* USA 88: 9828-9832.

18. Nacagahra, M. 1972. Genetic mechanism on the distorted segregation of markers gene belonging to the eleventh linkage group in cultivated rice. *J Breed*. 22: 232-238.

19. Nelson, J. C. 1997. Q-gene: software for marker based genome analysis and breeding. *Mol Breed* 3: 229-235.

20. Xu, Y.; Zhu, L.; Xiao, J.; Huang, N.; Mccouch, S. R. 1997. Chromosomal regions associated with segregation distortion of molecular markers in F2, backcross, doubled haploid and recombinant inbreed populations in rice (*Oyza sativa* L). *Mol Gener Genet* 253: 535-545.