Nucleotide diversity of a major carotenoid biosynthetic pathway gene in wild and cultivated *Solanum* (Section *Lycopersicon*) species

Ana H. Araújo[1,2,3], Maria Esther de N. Fonseca[1]* and Leonardo S. Boiteux[1,4]

*[1]Embrapa Vegetable Crops, CP 0218, 70359-970 Brasília-DF, Brazil. [2]University of Brasília, Cell and Molecular Biology Graduate Program, Brazil, [3]Fellowship CAPES, [4]Fellowship CNPq. *Corresponding author: mesther@cnph.embrapa.br*

Received: 19 September 2007; Returned for revision: 08 October 2007; Accepted: 07 December 2007

A polymerase chain reaction (PCR)-based approach was employed to isolate putative alleles of the chromoplast-specific lycopene beta-cyclase (CYCB) gene from wild and cultivated tomatoes [*Solanum* (Section *Lycopersicon*)]. The objective of this work was to establish an effective PCR protocol by testing DNA samples from distinct germplasm accessions with a primer pair designed to selectively target conserved regions present in the available CYCB sequences. This PCR optimization allowed the amplification of 1219 out 1666 bp of the gene in six taxa: *S. cheesmaniae*, *S. peruvianum*, *S. neorickii*, *S. pennellii*, *S. pimpinellifolium* and *S. lycopersicum*. Sixty-three mutation sites (31 transitions, 18 transversions and 14 single base deletions/insertions) were detected in these accessions when compared to *S. lycopersicum* AF 254793 (used as reference sequence). The polymorphisms were found predominantly in green-fruited species (20 in *S. neorickii*, 20 in *S. peruvianum*, and 32 in *S. pennellii*). Lower levels of polymorphisms were found in yellow-fruited (three in *S. cheesmaniae*) and red-fruited species (eight in *S. pimpinellifolium* and none in the *S. lycopersicum*). The higher levels of nucleotide diversity in the CYCB-like gene sequences in accessions of green-fruited species as well as the phylogenetic tree agreed with the previous taxonomic studies based upon the granule-bound starch synthase gene phylogeny. Sequence analyses of the amplicons obtained via heterologous PCR indicated the CYCB gene-specificity of the primers. Therefore, this PCR-based strategy might be useful to isolate CYCB-like amplicons from other species within the genus *Solanum* and to develop molecular markers for assisted breeding.
Key words: CYCB, lycopene, lycopene-beta-cyclase, nucleotide diversity, tomatoes, molecular markers

Diversidade de seqüência de um gene componente da via biossintética de carotenóides em espécies selvagens e cultivadas de *Solanum* (Seção *Lycopersicon*): Uma estratégia baseada em PCR ("polymerase chain reaction") foi empregada para isolarem-se alelos do gene da ciclase do licopeno específica de cromoplasto (CYCB) em espécies selvagens e domesticadas de tomate [*Solanum* (Secção *Lycopersicon*)]. A otimização das condições de reação de PCR permitiu a amplificação de uma seqüência de 1219 dos 1666 pares de base do gene CYCB em seis taxa: *S. cheesmaniae*, *S. peruvianum*, *S. neorickii*, *S. pennellii*, *S. pimpinellifolium* e *S. lycopersicum*. Sessenta e três mutações (31 transições, 18 transversões e 14 inserções/deleções) foram detectadas nestes acessos, em estudos comparativos com a seqüência referência de *S. lycopersicum* (AF 254793). Estes polimorfismos foram detectados predominantemente em espécies de fruto maduro de cor verde (20 em *S. neorickii*, 20 em *S. peruvianum* e 32 em *S. pennellii*). Um número menor de polimorfismos foi detectado em acessos de fruto maduro amarelo (três em *S. cheesmaniae*) e acessos de espécies de fruto maduro vermelho (oito em *S. pimpinellifolium* e nenhum polimorfismo no acesso *S. lycopersicum*). A árvore filogenética gerada a partir da comparação das seqüências similares ao gene CYCB está de acordo com estudos taxonômicos prévios baseados na filogenia do gene que codifica a enzima de sintase de amido ("granule-bound starch synthase"). As análises das seqüências obtidas indicaram também a seletividade dos "primers", permitindo o

desenvolvimento de marcadores moleculares bem como a potencial utilização dessa estratégia para isolamento de amplicons contendo seqüências similares ao CYCB em outras espécies dentro do gênero *Solanum*.
Palavras-chave: CYCB, diversidade de nucleotídeos, licopeno, licopeno-beta-ciclase, marcadores moleculares, tomate

Carotenoid pigments are essential structural and physiological components of the plant cell, playing a major role during photosynthesis when they are able to dissipate excess light energy absorbed by the photosynthetic system (Demmig-Adams and Adams, 2002). These pigments are also precursors of important compounds including vitamins and plant hormones and they also represent important ecological mechanisms allowing attraction of pollinators and seed-dispersing animals (Hirschberg, 2001). In human nutrition, the carotenoid pigments beta-carotene, alpha-carotene and cryptoxanthin are converted to vitamin A in the digestive tract. The carotenoids lycopene (the typical red tomato pigment), lutein, and zeaxanthin function as powerful antioxidants being associated with fewer incidences of chronic diseases as well as certain types of cancer in humans (Rao and Agawal, 2000; Smidt and Burke, 2004).

The carotenoid biosynthetic pathway is one of the most extensively known in plants with an array of structural genes and gene products already characterized (Cunningham and Gantt, 1998; Isaacson et al., 2004). Molecular markers based upon those genes have been extensively employed (Thorup et al., 2000; Fanciulino et al., 2007; Just et al., 2007). In the cultivated tomato (*Solanum lycopersicum* L.), two lycopene cyclases [lycopene beta-cyclase (LCYB) and lycopene epsilon-cyclase (LCYE)] (EC 1.14.–.–) are involved in converting lycopene to either beta-carotene or alpha-carotene. Two lycopene beta-cyclases have also been described in tomato; one is LCYB and the other is the chromoplast-specific lycopene beta-cyclase (CYCB) (EC 1.14.–.–). The characterization of two yellow/orange fruit mutants (*delta* and *Beta*) revealed that these traits are controlled by distinct lycopene cyclase genes (Ronen et al., 1999, 2000). The *Beta* mutation controls the up regulation of the CYCB gene, so that the fruit accumulates the orange pigment beta-carotene. The *delta* corresponds to an up-regulated LCYE, which results in the accumulation of the yellow pigment delta-carotene. A third mutation (*old-gold*) controls lycopene accumulation in fruits and petals due to a frame-shift mutation in the CYCB gene (Ronen et al., 1999, 2000).

Even though the genetics of the carotenoid synthesis is well studied in the cultivated tomato, there is a scarce amount of information available about the diversity of CYCB genes in closely related wild tomato species. Due to the vital function of these pigments, the genes coding for the carotenoid pathway enzymes are evolutionarily conserved among flowering plants (Hirschberg, 2001). The main goal of our research was to develop a specific primer and employ it in PCR assays aiming to amplify CYCB allelic variants in a germplasm collection comprising species of *Solanum* (section *Lycopersicon*). This germplasm collection has a wide array of mature fruit colors varying from green (*S. habrochaites*, *S. chilense*, *S. peruvianum*, *S. neorickii*, and *S. pennellii*), to yellow/orange (*S. galapagense* and *S. cheesmaniae*) and red/deep red (*S. lycopersicum* var. *cerasiforme*, *S. lycopersicum*, and *S. pimpinellifolium*). A second goal was to evaluate the usefulness of the sequence data obtained from the CYCB amplicon to examine the phylogenetic relationships of the wild and cultivated tomato species.

The full-length CYCB sequence (AF 254793) was retrieved from the GenBank database to serve as reference for primer design. The 24-mer forward primer is located in the position 170-193 (5´-TAGCAC CCACATCAAAGCCAGAGT-3´) and the 26-mer reverse primer (5´-CCGAAAAGACACACAAGCTGAGTAAA-3´) is located in the position 1380-1403. The optimal PCR conditions that yielded a sharp and distinct amplicon with the expected size (1219 bp-data not shown) were set up as follows: the 25 µL reaction mixture consisted of 0.2 µM each primer, 0.25 mM each dNTP, 1XPCR buffer, 2.0 mM $MgCl_2$, 1 U *Thermus aquaticus* DNA polymerase (Invitrogen, USA) and 80 ng DNA template. The reaction was performed on a 9700 thermocycler (Applied Biosystems, USA) with the following program: denaturation for 5 min at 94°C; 30 cycles of denaturation for 0.5 min s at 96°C, annealing for 1 min at 56°C and extension for 2 min at 68°C; and a final extension of 10 min at 72°C. The amplicons were resolved in agarose gel, gel-purified and ligated into pGEM-T Easy-T vector

(Promega, USA). After transformation into competent *E. coli* XL1Blue (Clontech, USA), three clones from seven germplasm accessions were randomly selected and sequenced. The quality of the sequences obtained in the trace files was first checked visually and also screened using the SeqMan program (Lasergene, Madison, WI, USA). There was no variability within species. The sequence obtained for each accession was aligned in the region defined by the two primers using the Clustal method available in the Megalign program (Lasergene, Madison, WI, USA). Phylogenetic analysis was done on matrices of aligned sequences using the bootstrap (5,000 replications) options of PAUP 4.0 software (Swofford, 1998). A consensus tree obtained by parsimony is shown in Figure 1.

The PCR optimization allowed the amplification of 1219 out of 1666 bp of the CYCB gene from six taxa: *S. peruvianum*, *S. neorickii*, *S. pennellii*, *S. cheesmaniae*, *S. pimpinellifolium*, and *S. lycopersicum*. Sixty-three mutation sites (31 transitions, 18 transversions and 14 single base deletions/insertions) were observed in the six species when compared with *S. lycopersicum* AF 254793 sequence. Thirty-one of the polymorphisms represented transitions (6 T/C, 5 C/T, 8 G/A and 12 A/G), eighteen transversions (2 A/C, 3 A/T, 4 G/T, 4 T/A and 5 T/G) and fourteen single base deletions/insertions (InDels). These results are shown in Table 1. The polymorphisms were found predominantly in green-fruited species (20 in *S. neorickii*, 20 in *S. peruvianum*, and 32 in *S. pennellii*) and only a few in yellow (three in *S. cheesmaniae*) and red-fruited species (eight in *S. pimpinellifolium* and none in the *S. lycopersicum* 'Santa Clara'). There were no long deletions/insertions in the entire sequence of all the putative CYCB alleles. Bootstrap analysis (Figure 1) gave a good support (63%) for the branch formed by the green and yellow-fruited species and for the branch formed only by the green-fruited species (64%). The sequence similarity between the putative wild species CYCB genomic amplicons with the mRNA sequence available on GenBank indicates this gene is intronless. Translation of this sequence and analysis of the predicted polypeptide showed identity to the AF 254793 sequence (Ronen et al., 2000) varying from 95.5% (*S. neorickii*) to 100% (*S. lycopersicum* 'Santa Clara'). The results support the specificity of the pair of primers for the CYCB gene sequences, since no other PCR product was obtained in the described assay conditions. Interestingly, the frequency of polymorphisms in green-fruited species sequences as well as the phylogenetic tree agreed with the previous taxonomic studies based upon the granule-

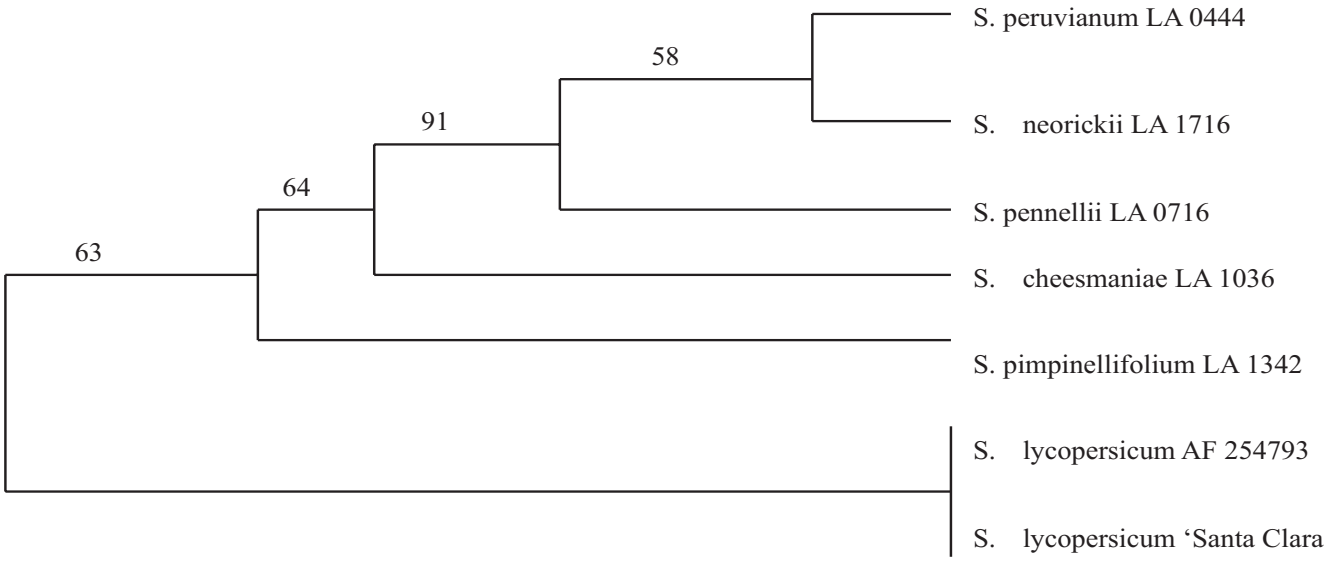

Figure 1. Phylogenetic tree obtained from the Clustal alignment of the chromoplast-specific lycopene cyclase gene fragment from *Solanum* species. *Solanum lycopersicum* 'Santa Clara' and *S. lycopersicum* (AF 254793) were found to be identical sequences. Percentage of 5,000 bootstrap replicates is given above branches.

Table 1. Sixty three mutations sites (indicated in italic boldface) detected in chromoplast-specific lycopene beta-cyclase gene-derived amplicons from seven *Solanum* (Section *Lycopersicon*) accessions. The numbers are the mutation sites found along the 1219 bp amplicon sequence in one or more accessions when compared to the reference sequence (*S. lycopersicum* AF 254793). The dots stand for nucleotide deletions. The types of mutations were identified as V (transversions) and T (transitions). Insertions/Deletions were identified as ID.

Accessions	63	64	65	80	92	117	149	176	221	237	290	293	356	372	428	517	540	547	548	549	552
S. lycopersicum "Santa Clara"	G	A	A	C	C	C	G	C	G	A	A	T	T	A	T	G	A	A	T	G	G
S. peruvianum LA 0444	G	***G***	A	C	***T***	C	***A***	***T***	***A***	***G***	A	T	***C***	***G***	***C***	G	A	A	T	G	G
S. cheesmaniae LA 1036	G	A	A	C	C	C	G	C	G	***G***	A	T	T	A	T	G	A	A	T	G	G
S. pennellii LA 0716	***T***	***G***	***G***	C	C	***T***	G	C	G	***G***	***G***	T	***C***	A	T	***A***	A	A	T	G	.
S. neorickii LA 1716	G	***G***	A	***T***	***T***	C	G	C	G	***G***	***G***	***C***	T	A	T	G	.	A	T	G	G
S. pimpinellifolium LA 1342	G	A	A	C	C	C	G	C	G	***G***	A	T	T	A	T	G	A	***C***	***A***	***T***	G
S. lycopersicum AF 254793	G	A	A	C	C	C	G	C	G	A	A	T	T	A	T	G	A	A	T	G	G
SNP type	V	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	ID	V	V	V	ID

Accessions	553	554	559	560	561	569	570	571	579	590	602	603	604	610	611	613	615	616	619	623	630
S. lycopersicum "Santa Clara"	.	T	G	A	A	A	T	A	A	.	.	C	T	A	T	G	A	G	.	C	.
S. peruvianum LA 0444	***G***	T	G	A	A	A	T	A	A	.	.	C	T	***G***	***A***	***T***	***G***	***A***	***G***	C	.
S. cheesmaniae LA 1036	.	T	G	A	A	A	T	A	A	.	.	C	T	A	T	G	A	G	***G***	C	.
S. pennellii LA 0716	.	.	G	***G***	A	A	T	A	***C***	.	.	.	***C***	A	T	G	A	G	.	***T***	.
S. neorickii LA 1716	***G***	T	***A***	A	***T***	***G***	***A***	***T***	A	.	.	C	T	A	T	G	A	G	***G***	C	***T***
S. pimpinellifolium LA 1342	.	T	G	A	A	A	T	A	A	***G***	***G***	C	T	A	T	G	A	G	***G***	C	.
S. lycopersicum AF 254793	.	T	G	A	A	A	T	A	A	.	.	C	T	A	T	G	A	G	.	C	.
SNP type	ID	ID	S	S	V	S	V	V	V	ID	ID	ID	S	S	V	V	S	S	ID	S	ID

Accessions	631	634	648	705	716	731	736	749	750	751	752	753	756	875	818	930	1200	1202	1204	1207	1208
S. lycopersicum "Santa Clara"	G	T	T	G	.	.	.	.	T	G	T	T	A	T	A	A	A	T	T	T	G
S. peruvianum LA 0444	***A***	T	.	***A***	.	.	.	.	T	G	T	***G***	A	T	A	A	***G***	T	T	T	G
S. cheesmaniae LA 1036	G	T	T	***A***	.	.	.	.	T	G	T	T	A	T	A	A	A	T	T	T	G
S. pennellii LA 0716	***A***	T	T	***A***	***T***	***T***	***T***	***T***	***G***	***T***	***G***	T	***T***	***C***	***G***	***G***	A	***G***	***G***	***A***	***A***
S. neorickii LA 1716	***A***	***C***	T	***A***	.	.	.	***C***	T	G	T	***G***	A	T	A	A	A	T	T	T	G
S. pimpinellifolium LA 1342	G	T	T	G	.	.	.	***C***	T	G	T	T	A	T	A	A	A	T	T	T	G
S. lycopersicum AF 254793	G	T	T	G	.	.	.	.	T	G	T	T	A	T	A	A	A	T	T	T	G
SNP type	S	S	ID	S	ID	ID	ID	ID	V	V	V	V	V	S	S	S	S	V	V	V	S

bound starch synthase gene phylogeny (Peralta and Spooner, 2001). Therefore, these primers might represent additional tools for use in taxonomic studies of natural populations within *Solanum* section *Lycopersicon*. The presence of mutations inside the coding region of putative CYCB genes of *Solanum* species unable to accumulate the red pigment lycopene provides a working hypothesis about the role of this gene in controlling this important tomato quality trait. Finally, the presence of natural genetic variation of the CYCB genes might greatly contribute to extend our ability to explore their diversity and function as either molecular markers in assisted-breeding or in transgenic strategies aiming to modify carotenoid content and/or profile in cultivated tomatoes.

REFERENCES

Cunningham FX, Gantt E (1998) Genes and enzymes of carotenoid biosynthesis in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 49:557-583.

Demmig-Adams B, Adams WW (2002) Antioxidants in photosynthesis and human nutrition. Science 298:2149-2153.

Fanciullino AL, Dhuique-Mayer C, Luro F, Morillon R, Ollitrault P (2007) Carotenoid biosynthetic pathway in the citrus genus: number of copies and phylogenetic diversity of seven genes. J. Agric. Food. Chem. 55:7405-7417.

Hirschberg J (2001) Carotenoid biosynthesis in flowering plants. Curr. Opin. Plant Biol. 4:210-218.

Isaacson T, Ohad I, Beyer P, Hirschberg J (2004) Analysis in vitro of the enzyme CRTISO establishes a poly-cis-carotenoid biosynthesis pathway in plants. Plant Physiol.136:4246-4255.

Just BJ, Santos CAF, Fonseca MEN, Boiteux LS, Oloizia BB, Simon PW (2007) Carotenoid biosynthesis structural genes in carrot (*Daucus carota*): isolation, sequence-characterization, single nucleotide polymorphism (SNP) markers and genome mapping. Theor. Appl. Genet. 114:693-704.

Peralta IE, Spooner DM (2001) Granule-bound starch synthase (GBSSI) gene phylogeny of wild tomatoes (*Solanum* L. Section *Lycopersicon* [Mill.] Wettst. Subsection *Lycopersicon*). Am. J. Bot. 88:1888-1902.

Rao AV, Agawal S (2000) Role of antioxidant lycopene in cancer and heart disease. J. Am. Coll. Nutr.19:563-569.

Ronen G, Cohen M, Zamir D, Hirschberg J (1999) Regulation of carotenoid biosynthesis during tomato fruit development: expression of the gene for lycopene epsilon cyclase is down regulated during ripening and is elevated in the mutant *delta*. Plant J. 17:341-351.

Ronen G, Carmel-Goren L, Zamir D, Hirschberg J (2000) An alternative pathway to b-carotene formation in plant chromoplasts discovered by map-based cloning of *beta* and *old-gold* color mutations in tomato. Proc. Natl. Acad. Sci. USA 97:1110-11107.

Smidt CR, Burke DS (2004) Nutritional significance and measurement of carotenoids. Curr. Top. Nutraceut. Res. 2:79-91.

Swofford DL (1998) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), version 4. Sinauer Associates, Sunderland.

Thorup TA, Tanyolac B, Livingstone KD, Popovsky S, Paran I, Jahn M (2000) Candidate gene analysis of organ pigmentation loci in the *Solanaceae*. Proc. Natl. Acad. Sci. USA 97:11192–11197.