Naturally occurring variation in the promoter of the fruit-specific *CYC-B* gene in tomato can be used to modulate levels of β-carotene using marker-assisted backcross breeding

Caleb Orchard[1], Jessica Cooperstone[2], Elisabet Gas[1], Steve Schwartz[2], David Francis[1].
[1]Horticulture and Crop Sciences, The Ohio State University, OARDC, Wooster, Ohio
[2]Food Science and Technology, The Ohio State University, Columbus, Ohio

ABSTRACT

β-carotene is an important carotenoid for human health due to its pro-vitamin A activity. We examined the carotenoid profiles of vintage and contemporary tomato (*S. lycopersicum*) varieties to identify sources of high β-carotene. Red tomatoes had a range from 0.2 – 0.97 mg/100 g fresh weight of β-carotene, while several orange fruited varieties had 1.67 – 4.0 mg/100 g. The *B* gene (*CYC-B*) encodes a fruit-specific lycopene-β-cyclase which converts *trans*-lycopene to β-carotene. We used high-throughput genotyping to detect known genetic variation and *de novo* sequencing to discover new variation in *B*. The non-transcribed region 5′ to the *B* gene (promoter) contains significant nucleotide variation, with nine unique haplotypes across 1850 bp of sequence. Seven unique alleles occurred in high β-carotene varieties. Association mapping and non-parametric statistical approaches suggest two single nucleotide changes (SNPs) as the most likely cause(s) of high β-carotene, presumably through their influence on transcription of the gene. Analysis of the sequence data using clustering techniques suggested that the *B* promoter found in vintage varieties, contemporary breeding lines, and hybrids was originally derived from wild tomato species. A marker-assisted backcross breeding scheme leveraging genome-wide SNPs was used to rapidly develop a series of genetic resources containing different alleles of *CYC-B* in a uniform genetic background. Replicated field trials demonstrated that distinct alleles can be used to modulate the levels of β-carotene in tomato. These genetic resources are available to develop β-carotene enriched food products or to study dietary adsorption and utilization of carotenoids in the food matrix.

OBJECTIVES

Our goal was to quantify variation in β-carotene levels in tomato including wild, vintage (heirloom), contemporary inbred lines, and hybrids; develop new genetic resources with altered carotenoid content for future research and crop improvement; and describe the genetic basis for variation in β-carotene.

MATERIALS AND METHODS

Fig. 1 – Fig. 3 illustrate the steps taken in order to 1) assess variation in β-carotene and lycopene levels in tomato germplasm; 2) develop new genetic resources with altered carotenoid content; 3) describe the genetic basis of high β-carotene. Briefly, hexane extractions of ripe fruit were prepared and quantified by HPLC using a C30 column for separation. High β-carotene accessions were used as a source of DNA for sequencing of *CYC-B* and the 5' untranscribed region. Distinct sources of high β-carotene were chosen for population development, with a backcross (BC) scheme using OH8245 as the recurrent parent. BC progeny were screened for the *CYC-B* allele, and ~380 selections were then genotyped with 96 SNP markers distributed across the genome. Selections with a high percentage of the OH8245 genome were then evaluated in multi-location field trials.

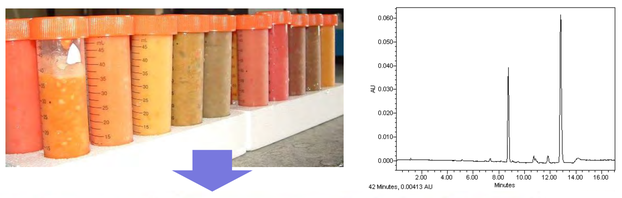


Fig. 1. β-carotene evaluation.
- Hexane extraction
- HPLC separation (C30)
- Choose parents for population development

Fig. 2. Population development.
- Jaune Flammee x OH8245
- (M82 x LA 716) x OH8245
- Backcross each hybrid to OH8245

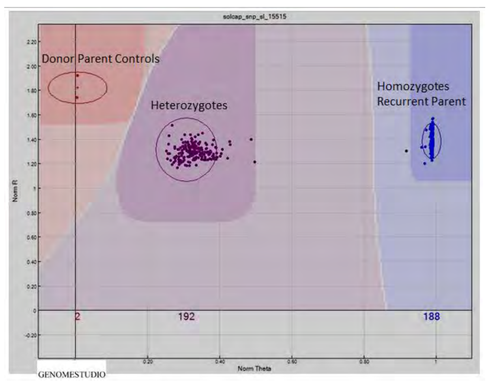


Fig 3. Genotyping platforms for selection. 96 Single Nucleotide Polymorphisms (SNPs) distributed across the genome were commercialized for use with the Illumina BeadXpress (for BC_1) or LGC genomics KASP™ (Kompetitive Alelle Specific PCR) platforms (for BC_2).

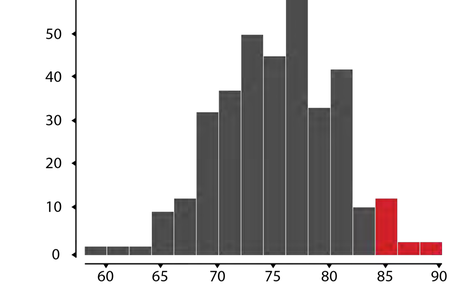


Fig 4. Selections in red have a high percentage of OH8245 SNPs (85-90%). On average, BC_1 progeny show ~75% of the recurrent parent. High β–carotene selections can be identified with 85%-90% in the first BC; repeat cross to OH8245 and repeat SNP selection in BC_2.

RESULTS

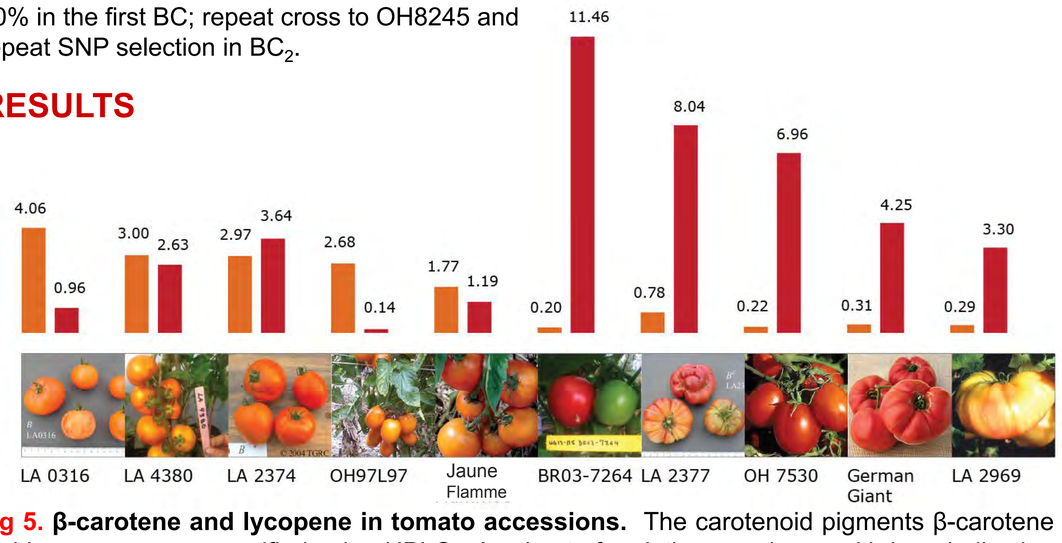


Fig 5. β-carotene and lycopene in tomato accessions. The carotenoid pigments β-carotene and lycopene were quantified using HPLC. A subset of varieties are shown with bars indicating the relative amounts in mg/100 g fresh weight of fruit. β-carotene is indicated in orange; lycopene, in red.

ACKNOWLEDGMENTS

This project was supported in part by the USDA NIFA AFRI Plant Breeding, Genetics, and Genome Solanaceae Coordinated Agricultural Project (SolCAP) grant 2009-85606-05673 and Hatch project OHO1287.

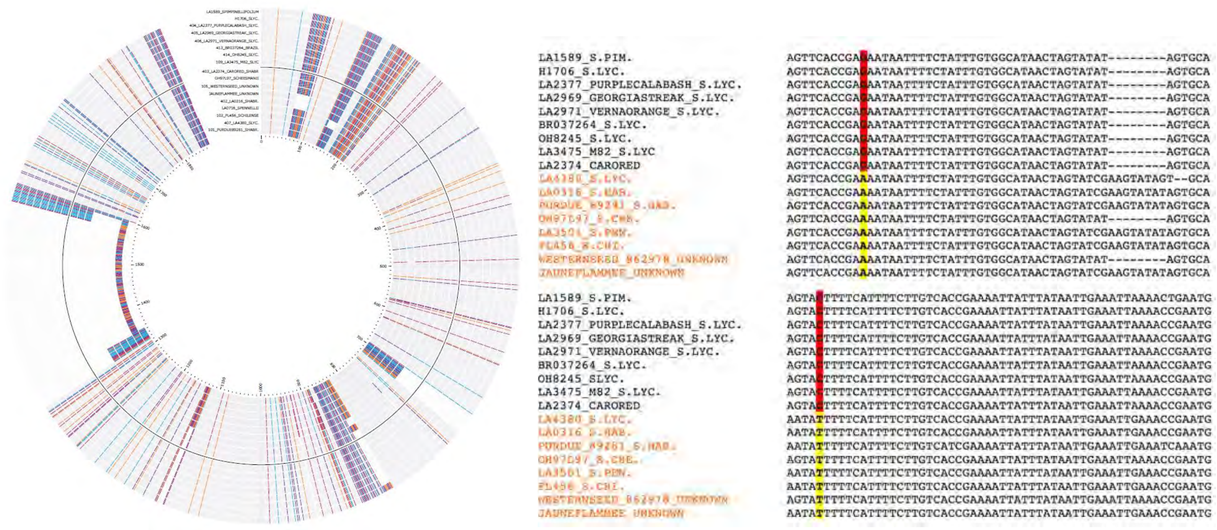


Fig 6. Variation in the DNA sequence 5' to the *B* gene. The figure to the left illustrates sequence variation among 17 accessions. Colors represent nucleotide differences, either insertion/deletion or single nucleotide polymorphisms. The figure to the right highlights two SNPs with complete correspondence to high β-carotene accessions. These are indicated by the highest probability in figure 6, below.

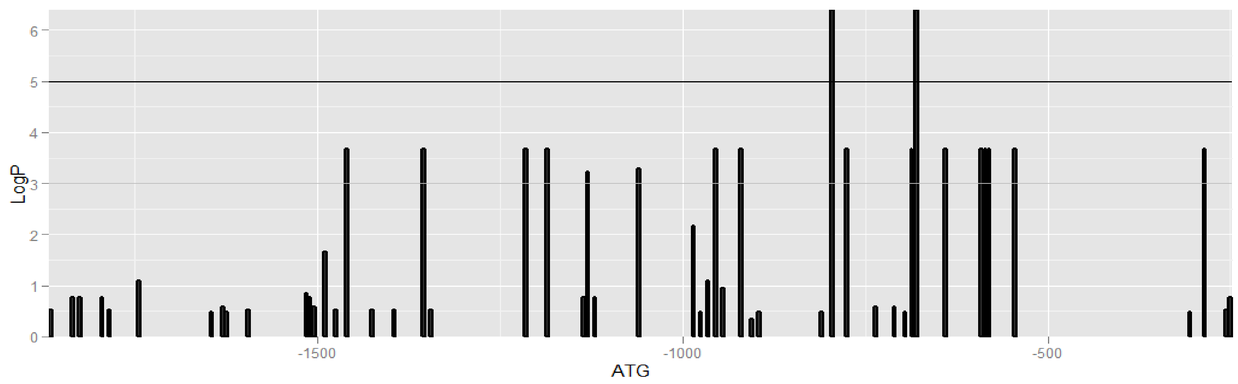


Fig 7. Association analysis to identify sequence variation most-likely causing high β-carotene. The graph is a "Manhattan plot" showing position of SNP or insertion/deletion variation relative to the *CYC-B* start codon (ATG = 0) vs –LogP based on statistical analysis for association with β-carotene levels in fruit. Grey bar indicates P = 0.05; black bar, P = 0.01 with a Bonferroni correction.

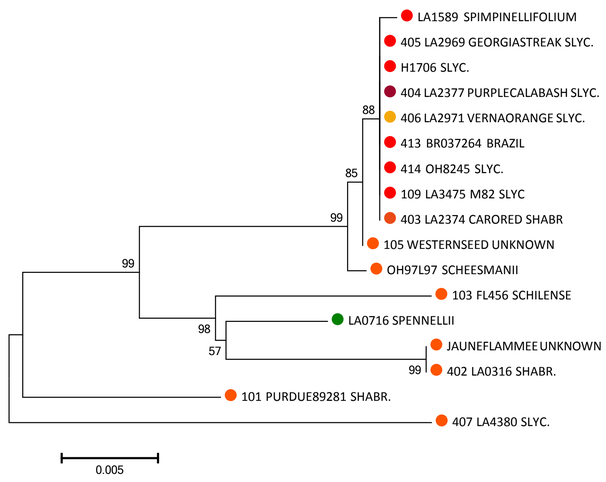
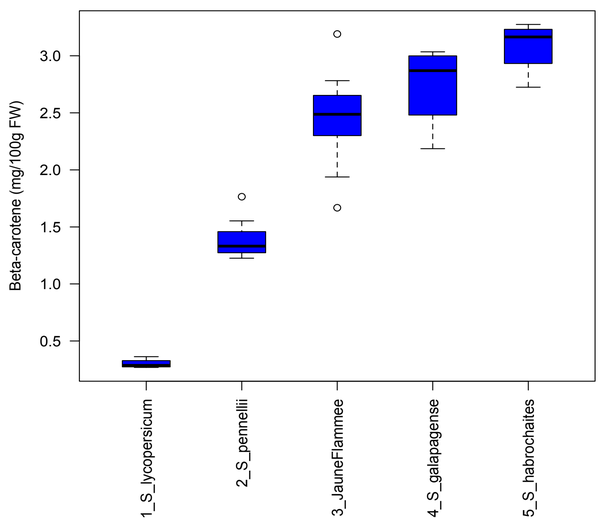


Fig 8. Alignment of DNA sequence 5' to the *CYC-B* gene. Sequence data were assembled using GAP4, and then aligned using MUSCLE (V 3.8). Alignments were clustered using MEGA (V 5.10). Bootstrapping was performed based on 10,000 replications of the phylogeny reconstruction. Bootstrap values are indicated at the tree nodes. Fruit color of the sequenced accessions are indicated by colored dots. Sequences of high β-carotene varieties cluster with sequences derived from wild species (either *S. cheesmanii* and *S. galapagense* on one branch or *S. chilense*, *S. pennellii* and *S. habrochaites* on the other).

Fig 9. High β-carotene selections compared to red-fruited control in multi-location field trials. Field evaluations were conducted in Fremont and Wooster, OH. Location differences were not significant, while the donor of the *B* allele was highly significant ($p < 0.001$) as a factor determining β-carotene levels in fruit.



CONCLUSIONS

- Red tomatoes had a range of 0.2 – 0.97 mg /100 g fresh weight of β-carotene, while several orange fruited varieties had 1.67 – 4.0 mg /100 g.
- We found nine unique haplotypes across 1850 bp of sequence; seven distinct alleles occurred in high β-carotene varieties.
- Association mapping and non-parametric statistical approaches suggest two SNPs as the most likely cause(s) of high β-carotene.
- The *B* promoter found in vintage varieties, contemporary breeding lines, and hybrids was originally derived from wild tomato species.
- Background genome selection identified BC_2 progeny with the equivalent "recurrent parent" (OH8245) genome expected in a BC_4.
- The source of the donor allele for *B* can be used to modulate β-carotene levels across a range from 2x to 12x of that found in typical red tomatoes.