Naturally occurring variation in the promoter of the fruit-specific \textit{CYC-B} gene in tomato can be used to modulate levels of \beta-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

\section*{ABSTRACT}
\beta-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 (\textit{S. lycopersicum}) varieties to identify sources of high \beta-carotene. Red tomatoes had a range from 0.2 – 4.0 mg/100 g fresh weight of \beta-carotene, while several orange fruited varieties had 1.67 – 4.0 mg/100 g. The \textit{B} gene (\textit{CYC-B}) encodes a fruit-specific lycopene-\beta-cyclase which converts trans-lycopene to \beta-carotene. We used high-throughput genotyping to detect known genetic variation and de novo sequencing to discover new variation in \textit{B}. The non-transcribed region 5' to the \textit{B} gene (promoter) contains significant nucleotide variation, with nine unique haplotypes across 1850 bp of sequence. Seven unique alleles occurred in high \beta-carotene varieties. Association mapping and non-parametric statistical approaches suggest two single nucleotide changes (SNPs) as the most likely cause(s) of high \beta-carotene, presumably through their influence on transcription of the gene. Analysis of the sequence data using clustering techniques suggested that the \textit{B} promoter found in vintage varieties, contemporary breeding lines, and hybrids was originally derived from \textit{S. pennellii} in 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 \textit{CYC-B} in a uniform genetic background. Replicated field trials demonstrated that distinct alleles can be used to modulate the levels of \beta-carotene in tomatoes. These genetic resources are available to develop \beta-carotene enriched food products or to study dietary adsorption and utilization of carotenoids in the food matrix.

\section*{OBJECTIVES}
Our goal was to quantify variation in \beta-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 \beta-carotene.

\section*{MATERIALS AND METHODS}
Fig. 1 – Fig. 3 illustrates the steps taken in order to 1) assess variation in \beta-carotene and lycopene levels in tomato germplasm; 2) develop new genetic resources with altered carotenoid content; 3) describe the genetic basis of high \beta-carotene. Briefly, hexane extractions of ripe fruit were prepared and quantified by HPLC using a C30 column for separation. High \beta-carotene accessions were used as a source of DNA for sequencing of \textit{CYC-B} and the 5' untranscribed region. Distinct sources of high \beta-carotene were chosen for population development, with a backcross (BC) scheme using OH8245 as the recurrent parent. BC progeny were screened for the \textit{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.

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

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