DNA microarray-based discovery of molecular markers for the improvement of tomato color and nutritional quality

Audrey Darrigues[1], Wencai Yang[2], David M. Francis[1]

INTRODUCTION

Color is among the most important attributes of tomato for processing into whole and diced products. Both color and color uniformity are affected by yellow shoulder disorder (YSD), a ripening disorder that results in discoloration of the proximal end tissues of the fruit. We show that lycopene and β-carotene concentrations are reduced by 18% and 22%, respectively, in fruits affected by YSD. Variance partitioning suggests that YSD incidence and severity is affected by both genetics and environment. **The objectives of this project were to: (1) develop single nucleotide polymorphisms (SNPs) as molecular markers and (2) elucidate the genetic basis of YSD. We hypothesized that a QTL mapping approach would identify loci that affect both color and color uniformity.** SNP discovery in breeding populations was based on both analyses of large public EST databases and on hybridization to a custom oligonucleotide array. The array was hybridized with target cDNA from *S. lycopersicum* (Ohio 7814) and *S. pimpinellifolium* (LA1589). We developed algorithms to detect outliers and identified 1,296 potential SNPs. These putative SNPs are being verified by sequencing, screened for utility as markers on a collection of 99 *S. lycopersicum* lines and wild relatives, and applied to the genetic dissection of YSD. Implementing SNP-based marker technology has the potential to dramatically alter our approach to genetic characterization. Results and interpretation from this study will help bridge the gap between the goals of genetic and crop improvement research by facilitating the use of population structures that favor simultaneous genetic analysis and crop improvement.

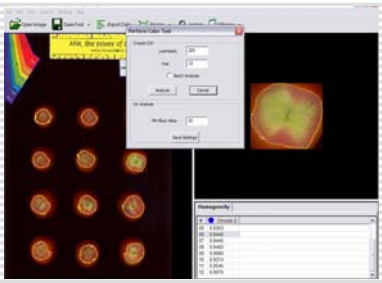

Figure 1. The color function of the Tomato Analyzer software (Brewer Talbot et al, in press) was used to obtain objective color measurements for YSD, RGB and L, a, b values. The color parameters were set at hue=15, luminosity=200 and minimum blue=25.

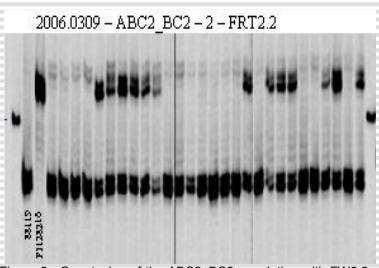

Figure 2. Genotyping of the ABC2_BC2 population with FW2.2 on acrylamide gel with the LiCor IR² sequencer.

Table 1. Effect of yellow shoulder disorder (YSD) on the carotenoid content of processing tomato varieties.

Genotype	Tissue Phenotype	Lycopene LSMean (mg/100g f.w.)	Beta-carotene LSMean (mg/100g f.w.)
FG00-118	Non-YSD	4.4814	1.3575
FG00-118	YSD	3.5150	1.0153
FG00-124	Non-YSD	4.6605	1.6373
FG00-124	YSD	3.4865	1.0472
FG99-36	Non-YSD	4.8590	1.6411
FG99-36	YSD	3.8212	1.3105
OH8245	Non-YSD	4.4919	1.7342
OH8245	YSD	3.7866	1.2206
PS696	Non-YSD	5.2618	1.6217
PS696	YSD	4.4027	1.4519
Overall means	Non-YSD	5.2146	1.6365
	YSD	4.2437	1.2803
	P-value	0.0069	<0.0001

METHODS

Quantification of lycopene and β-carotene in tomato varieties

Five processing tomato varieties were evaluated for their carotenoid content (Table 1). The proximal end of each fruit was cut and fruits were categorized as either non-affected by YSD or affected by YSD. Carotenoid extraction was carried out under red light following a hexane/acetone-based protocol (Ferruzzi et al., 1998). Separations were achieved using a Waters 2690 reverse-phase HPLC system equipped with a photodiode array detector and a C18 column (Vydac 201TP54; 4.6mm x 250 mm). The peak identification and the subsequent quantification of β-carotene and lycopene were achieved using the standard curve for each compound and their molar absorptivity coefficients (Nguyen et al., 2001).

Phenotypic evaluation of ABC2_BC2

A BC2 breeding population, *S. lycopersicum* (OH88119) x *S. pimpinellifolium* (PI128216), was evaluated for color in replicated trials. At harvest, the cut surface of the proximal end of 12 fruits from each plot were scanned at 200dpi. The images were analyzed with the color function of the Tomato Analyzer software (Brewer Talbot et al, *in press*) and objective measurements for color were collected (Figure 1). The color parameters of the Tomato Analyzer are based on the RGB color system. Algorithms were developed to convert the RGB values to L, a, b values, which were then used to calculate chroma, $\sqrt{(a^2+b^2)}$, and hue, $180/\pi*\arccos(a/\sqrt{(a^2+b^2)})$. The L coordinate represents lightness to darkness. Chroma is defined as saturation or intensity of color and hue represents the colors of the color wheel. The extent of YSD was obtained from the number of pixels that fall outside of the range hue=15, which is considered non-red tissue. To encompass both color and color uniformity, an index was computed from L, hue and YSD values.

Marker development and verification

A custom DNA microarray consisting of 15,925 'genes' was hybridized with target cDNA from *S. lycopersicum* (OH7814) and *S. pimpinellifolium* (LA1589). Each gene is represented by 12 24-base oligonucleotide probes. We developed algorithms to detect outliers (i.e. probe outliers within an individual gene) and identified 1,296 potential SNPs. Subsequent sequence verification confirmed 52% of the features identified as outliers.

Genotypic evaluation

A total of 42 markers were tested in the population, of which 15 were developed with the oligo array and the other 27 were from our database of molecular markers. Genotyping was performed using either a LiCor IR² sequencer with acrylamide gels (Figure 2) or agarose-based electrophoresis (Yang et al, 2004).

Statistical analyses

Marker-trait associations were determined using the general linear model procedure. The model was $Y_{ijk} = \mu + R_i + M_j + G_k(M_j) + \varepsilon_{ijk}$, where Y_{ijk} is the trait value, μ is the population mean, R_i is the effect of the i^{th} replication, M_j is the effect of the j^{th} marker, $G_k(M_j)$ is the effect of the k^{th} genotype within the j^{th} marker class, and ε_{ijk} is the experimental error. The appropriate F-test for marker-trait associations was $M_j/G_k(M_j)$. The total phenotypic variation explained by marker loci (Vm/Vp_total) was determined from variance component estimates using the MIXED model of analysis of variance with the restricted maximum likelihood (REML) method. All statistical procedures were implemented using SAS software.

Table 2. Molecular markers significantly associated with tomato color and color uniformity in a BC2 introgressed population, *Solanum lycopersicum* (OH88119) x *S. pimpinellifolium* (PI128216)

Marker	Marker Type (RE)	Chromosome (location)[x]	Trait	*P* value	Genotypic means of marker classes[y] LL	LP	Vm/Vp[z]
LEOH17	SNP (BseNI)	1 or 3	HUE	0.0278	46.0355	46.8963	0.0401
			AVGL	0.0087	41.8047	42.6714	0.0820
SSR32	SSR	2 (67.9)	AVGL	0.0481	42.3095	41.7678	0.0338
LEOH348	SNP (Tail)	2 (72.3)	CHROMA	0.0080	42.0643	40.8968	0.0951
SSR26	SSR	2 (76.8)	CHROMA	0.0055	42.0908	40.7603	0.1217
			AVGL	0.0288	42.3808	41.6904	0.0522
SSR349A	SSR	2 (.)	INDEX	0.0168	3.7223	4.3256	0.0940
			HUE	0.0232	45.8168	47.1679	0.0929
			CHROMA	0.0127	42.2021	40.766	0.1513
			YSD	0.0074	10.9676	14.6463	0.1197
CosOH7	SNP (HinfI)	2 (.)	CHROMA	0.0378	41.9181	40.9631	0.0605
fw2.2	SSR	2 (.)	CHROMA	0.0001	42.5426	40.6727	0.2473
			AVGL	0.0304	42.4515	41.8193	0.0446
LEOH127	SNP (HincII)	3 (113)	INDEX	0.0196	4.0094	3.6884	0.0327
			HUE	0.0171	46.5489	45.8318	0.0337
			YSD	0.0145	12.6976	10.7617	0.0386
LEOH200	SNP (EcoRV)	6 (69.4)	INDEX	0.0016	4.0913	3.5268	0.0845
			HUE	0.0107	46.662	45.5745	0.0625
			YSD	0.0016	13.1032	9.9108	0.0862
			AVGL	0.0012	42.4871	41.3832	0.1144
LEOH112	SNP (Tail)	6 (72.2)	INDEX	0.0039	4.1059	3.6299	0.0629
			HUE	0.0320	46.6439	45.7688	0.0382
			YSD	0.0028	13.2382	10.426	0.0703
			AVGL	0.0021	42.523	41.5527	0.0983
LEOH147	SNP (Tsp45I)	8 (19)	CHROMA	0.0120	42.0014	39.2814	0.0877
			AVGL	0.0024	42.3946	43.1014	0.0910
LEOH197	INDEL	12 (81.3)	CHROMA	0.0131	41.035	42.1154	0.0846

[x](.)The relative map position of the marker is not available.
[y]LL=homozygous *Solanum lycopersicum*; LP=heterozygous *S. lycopersicum* x *S. pimpinellifolium*
[z]Proportion of total phenotypic variation explained by marker locus

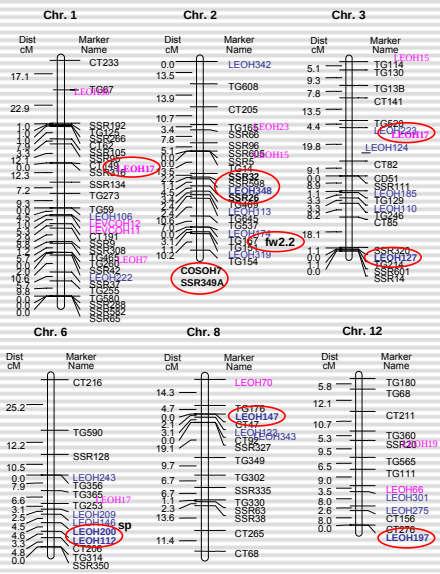

Figure 3. Distribution of markers that are significantly associated with traits of color and color uniformity. Only the chromosomes with significant markers are shown. In blue are the markers that were developed via the oligo array and in purple are the markers that were developed via EST analysis.

RESULTS & CONCLUSIONS

Effect of YSD on carotenoids

Lycopene was significantly reduced by 18.6% (*P*=0.007) and β-carotene by 21.8% (*P*<0.0001). These results show that by reducing the incidence of YSD, the potential for health benefits of tomato and tomato products can be optimized.

Association of marker and loci that affect color

Of the 12 markers with significant associations with color parameters, six are SNPs and INDELs developed from the oligonucleotide array. These 12 markers define a minimum of seven putative QTL on six chromosomal regions (Figure 3).

There are seven putative QTL associated with chroma (the higher the value, the more intense the color), five QTL for hue (the lower the angular value, the more red), four QTL for both YSD and INDEX, which account for color quality and uniformity (the lower the value, the better). The presence of QTL with positive alleles from both *S. lycopersicum* and *S. pimpinellifolium* suggests the possibility of obtaining transgressive segregants with improved color traits.

Five of the markers are located on chromosome 2, where a QTL associated with L coordinate had been reported previously (LEOH23 marker; Yang et al, 2004). Two markers, LEOH200 and LEOH112, are located on chromosome 6 and their positive associations with color traits is due to the *S. pimpinellifolium* allele.

Two loci introgressed from wild species, fruit weight (*fw2.2*) and self-pruning (*sp*), are mapped to chromosomes 2 and 6, respectively, in proximity of the markers described above (Figure 3). **Recovering recombinants with the color QTL without these specific loci will reduce the negative effects and increase the efficiency of marker-assisted selection.**

Genome sequencing of tomato is underway; accurate mapping of QTL for YSD and other traits of agricultural importance will allow us to apply candidate gene approaches as a more efficient breeding tool in the near future.

REFERENCES

Brewer Talbot, M., L. Lang, K. Fujimura, N. Dujmovic, S. Gray and E. van der Knaap. In press. Development of a controlled vocabulary and a software application to analyze fruit shape variation in tomato and other plant species. Plant Physiology.

Ferruzzi, M. G., L. C. Sander, C. L. Rock, and S. J. Schwartz. 1998. Carotenoid determination in biological microsamples using liquid chromatography with a coulometric electrochemical array detector. Analytical Biochemistry 256: 74-81

Nguyen, M., D. M. Francis and S. Schwartz. 2001. Thermal isomerisation susceptibility of carotenoids in different tomato varieties. J. Sci. Food Agric. 8:910-917

Yang, W., X. Bai, E. Kabelka, C. Eaton, S. Kamoun, E. van der Knaap and D. Francis. 2004. Discovery of single nucleotide polymorphisms in Lycopersicon lycopersicum by computer aided analysis of expressed sequence tags. Molecular Breeding 14: 21-34.

ACKNOWLEDGEMENTS

The authors thank the College of Wooster – AMRE team for the development of the Tomato Analyzer, especially Simon Gray and Nancy Dujmovic for their contribution to the color analysis component. The custom array was designed in collaboration with NimbleGen Systems, Inc, and is part of a joint project with Esther van der Knaap and Saskia Hogenhout (OSU/OARDC). The carotenoid analyses were carried out in the lab of Steven Schwartz, OSU. This project was partly funded by the OARDC Research Enhancement Competitive Grant awarded to AD.