

# Trait Diversity and Potential for Selection Indices Based on Variation Among Regionally Adapted Processing Tomato Germplasm

Heather L. Merk<sup>1</sup>

*Department of Horticulture and Crop Science, The Ohio State University, OARDC, 1680 Madison Avenue, Wooster, OH 44691*

Shawn C. Yarnes and Allen Van Deynze

*Seed Biotechnology Center, University of California, Extension Center Drive, Davis, CA 95616*

Nankui Tong<sup>2</sup>

*Campbell Soup Company, Agricultural Research Center, 28605 County Road 104 Davis, CA 95618*

Naama Menda and Lukas A. Mueller

*Boyce Thompson Institute for Plant Research, Tower Road, Ithaca, NY 14853*

Martha A. Mutschler

*Department of Plant Breeding and Genetics, Cornell University, 14 Laura Lane, Ithaca, NY 14853*

Steven A. Loewen

*University of Guelph Ridgetown Campus, 120 Main Street East, Ridgetown, Ontario N0P 2C0, Canada*

James R. Myers

*Department of Horticulture, Oregon State University, 4017 Ag and Life Sciences Building, Corvallis, OR 97331*

David M. Francis<sup>3</sup>

*Ohio State University, Department of Horticulture and Crop Science, 1680 Madison Avenue, Wooster, OH 44691*

**ADDITIONAL INDEX WORDS.** *Solanum lycopersicum*, standardized phenotyping, digital phenotyping, plant genetic resources, variance components, best linear unbiased predictor, BLUP, multitrait selection indices, MTI

**ABSTRACT.** For many horticultural crops, selection is based on quality as well as yield. To investigate the distribution of trait variation and identify those attributes appropriate for developing selection indices, we collected and organized information related to fruit size, shape, color, soluble solids, acid, and yield traits for 143 processing tomato (*Solanum lycopersicum* L.) lines from North America. Evaluation of the germplasm panel was conducted in a multiyear, multilocation trial. Data were stored in a flat-file format and in a trait ontology database, providing a public archive. We estimated variance components and proportion of variance resulting from genetics for each trait. Genetic variance was low to moderate (range, 0.03–0.51) for most traits, indicating high environmental influence on trait expression and/or complex genetic architecture. Phenotypic values for each line were estimated across environments as best linear unbiased predictors (BLUPs). Principal components (PC) analysis using the trait BLUPs provided a means to assess which traits explained variation in the germplasm. The first two PCs explained 28.0% and 16.2% of the variance and were heavily weighted by measures of fruit shape and size. The third PC explained 12.9% of the phenotypic variance and was determined by fruit color and yield components. Trait BLUPs and the first three PCs were also used to explore the relationship between phenotypes and the origin of the accessions. We were able to differentiate germplasm for fruit size, fruit shape, yield, soluble solids, and color based on origin, indicating regional breeding programs provide a source of trait variation. These analyses suggest that multitrait selection indices could be established that encompass quality traits in addition to yield. However, such indices will need to balance trait correlations and be consistent with market valuation.

Received for publication 3 May 2012. Accepted for publication 25 July 2012.

This work was supported by an award from the U.S. Department of Agricultural National Institute of Food and Agriculture (2009-85606-05673) and through gifts from the Mid-American Food Processors Association.

We thank the field staff at the Campbell Soup Company in Davis, CA, and at The Ohio State University North Central Agricultural Research Station in Fremont, OH, for assistance in the coordination of field experiments. We gratefully acknowledge the assistance of Troy Aldrich in organizing, distributing, and seeding the trials. We thank Dr. Stephanie Wedryk for providing critical feedback of an early manuscript draft. We also thank Steve Smith, Red Gold, Inc.; Charles Rivara, California Tomato Research Institute; Mike Montna, California Tomato Growers Association; and Al Krueger, Ontario Processing Vegetable Growers, for discussion and or provision of resources relevant to contract structures.

<sup>1</sup>Present address: Syngenta, 11055 Wayzata Boulevard, Minnetonka, MN 55305.

<sup>2</sup>Present address: BHN Research of Gargiulo Inc., 25675 Immokalee Road, Immokalee, FL 34142.

<sup>3</sup>Corresponding author. E-mail: francis.77@osu.edu.

Breeders of horticultural crops and agronomic crops have often adopted different strategies and systems of selection. Breeders of grain crops have a long tradition of quantitative approaches and of collecting objective data from large populations. This practice is facilitated in grain breeding by high-density planting, stability of the grain in the field, and the mechanization of harvest. In contrast, many horticultural crops require a labor-intensive harvest of a perishable commodity. Evaluation is often based on attributes beyond yield with appearance and quality receiving significant attention during selection. Both cost and time constrain the collection of objective data in horticultural crops, including tomato, and breeding often defaults to a qualitative decision. Thus, differences between commodities have affected approaches to selection, and the challenge remains for breeding programs targeting horticultural crops to develop the capacity to collect, store, and analyze objective trait data across multiple environments and generations in a high-throughput manner.

Plant breeders are beginning to consider estimated breeding value, the merit of an individual as determined by the performance of its progeny rather than actual cultivar performance as criteria for selection (Heffner et al., 2009). Estimates of breeding value can be derived from phenotypic data and pedigree information or genome-wide selection models (GWSs) that combine phenotype and genotype (Crossa et al., 2010; de los Campos et al., 2009). Estimating a breeding value or building robust GWS models requires integrating pedigree, genotypic, and phenotypic data for large populations. Phenotypes are recorded over multiple generations, locations, and years, often with unbalanced experimental designs. To account for spatial variation between environments, unbalanced data, and pedigree relationships, best linear unbiased predictors of phenotypes are used in place of arithmetic means. Phenotypic data on the scale and scope required to estimate breeding values have recently been summarized for several agronomic crop species, including barley [*Hordeum vulgare* L. (Lorenz et al., 2010; Wang et al., 2012)], maize [*Zea mays* L. (Kump et al., 2011; Riedelsheimer et al., 2012; Tian et al., 2011)], and soft wheat [*Triticum aestivum* L. (Souza et al., 2012)].

Trends in breeding are also affecting how trait data are managed. Historically, traits were organized based on categorical descriptors. For example, tomato fruit shape is often classified based on categories described by the International Union for the Protection of New Varieties of Plants (2001) and the International Plant Genetic Resources Institute (1996). These systems retain some use and overlay well with objective measures but are not entirely consistent with each other nor amenable to quantitative analysis (Rodriguez et al., 2011). The use of ontology terms has been suggested as a way to organize phenotypes in a standardized and quantitative format that is also amenable to storage in databases (Brewer et al., 2006; Jung et al., 2011; Milc et al., 2011). In addition, organizing traits into a standardized format with a quantitative scale allows comparative queries across experiments. Archives of phenotypic data are the biological complement to open access genomic data.

The process of measuring traits should be reliable, consistent, and objective if genotypic differences are to be detected and selection optimized. Digital phenotyping has emerged as one method to accomplish these goals (Hartmann et al., 2011). Such methods are helping to drive the transition from categorical to quantitative phenotyping that links ontology terms and trait descriptors. Tomato Analyzer software has emerged as

a tool to quantify fruit size, shape, and color in a semiautomated fashion (Brewer et al., 2007; Darrigues et al., 2008; Gonzalo et al., 2009; Gonzalo and van der Knaap, 2008). When applied to a structured breeding population, the precise phenotypic quantification of color increased the proportion of variance that could be ascribed to genetic factors (Darrigues et al., 2008). Plant breeders may therefore realize increased gain under selection from efforts to collect and store quantitative phenotypic data.

All crop improvement, whether marker-assisted, genome-wide, or phenotype-based, is grounded on our ability to accurately partition trait variance into environmental and genetic components. To address a lack of baseline data for important tomato traits, we collected extensive data for a diverse collection of processing tomato breeding lines from North America in a multiyear, multilocation trial. Our specific objectives were to examine the range of variation for important traits, to estimate the genetic contribution to these traits, examine correlations between traits, determine how variation is distributed within and between subpopulations within the germplasm, and integrate this information to begin developing multitrait selection indices. Phenotypes were collected using standardized, quantitative methods: analysis of digital images using Tomato Analyzer (Brewer et al., 2006), chemical tests of fruit quality, and components of yield. Traits were classified into Solanaceae Phenotype Ontology terms (Jung et al., 2011; Menda et al., 2008) and stored in the Sol Genomics Network (2012) database. We determined that significant variation exists for economically important traits and that regionally adapted germplasm may serve as a reservoir for trait variation.

## Materials and Methods

**PLANT MATERIALS.** A panel of 143 processing tomato lines (genotypes) representing breeding germplasm in North America was assembled by the Solanaceae Coordinated Agricultural Project (SolCAP) (Table 1). Ninety-five lines originated from breeding programs in the Great Lakes region of North America (midwestern United States and Ontario, Canada) and were considered “humid”-adapted. Twenty-six lines were derived from germplasm adapted to Oregon or California and were included as “arid”-adapted. The arid-adapted germplasm included 14 lines from the Cornell University breeding program that were developed from California-adapted germplasm with selection in alternate generations in California or Sinaloa, Mexico. Germplasm adapted to the production environments in the Great Lakes region or west coast of North America represent distinct genetic subpopulations (Sim et al., 2011). Pedigree records were not available for 22 lines; adaptation for these lines was reported as “undetermined.”

Seedlings were grown inside a greenhouse and transplanted to the field 6 to 8 weeks after sowing. Transplants were spaced 0.3 m apart on raised beds with 1.54 m between beds. Ohio trials were conducted at The Ohio State University North Central Agricultural Research Station in Fremont, OH, which is located in an area of commercial tomato production. Production practices were as recommended for commercial growers. California trials were conducted at the Campbell's Soup Company research station in Davis, CA, also using standard procedures for commercial growers.

**EXPERIMENT DESIGN.** Field trials were conducted with an unbalanced design across three years. Control cultivars were

Table 1. Processing tomato germplasm panel, including accession identification, donor institution, and regional adaptation.

| Donor no./cultivar name | Donor <sup>2</sup> | Adaptation | Donor no./cultivar name | Donor      | Adaptation        |
|-------------------------|--------------------|------------|-------------------------|------------|-------------------|
| 2K1-1439                | Ohio State         | Humid      | OH05-8179               | Ohio State | Humid             |
| 2K1-2019                | Ohio State         | Humid      | OH05-8181               | Ohio State | Humid             |
| 2K1-2029                | Ohio State         | Humid      | OH05-8184               | Ohio State | Humid             |
| 2K1-2054                | Ohio State         | Humid      | OH05-8185               | Ohio State | Humid             |
| CULBPT04-1              | Cornell            | Arid       | OH05-8186               | Ohio State | Humid             |
| CULBPT04-2              | Cornell            | Arid       | OH05-8187               | Ohio State | Humid             |
| CULBPT04-3              | Cornell            | Arid       | OH05-8188               | Ohio State | Humid             |
| CULBPT04-4              | Cornell            | Arid       | OH05-8193               | Ohio State | Humid             |
| CULBPT04-5              | Cornell            | Arid       | OH05-8197               | Ohio State | Humid             |
| CULBPT-05-10            | Cornell            | Arid       | OH05-8206               | Ohio State | Humid             |
| CULBPT-05-11            | Cornell            | Arid       | OH05-8210               | Ohio State | Humid             |
| CULBPT-05-15            | Cornell            | Arid       | OH05-8214               | Ohio State | Humid             |
| CULBPT-05-18            | Cornell            | Arid       | OH08-5201               | Ohio State | Und. <sup>y</sup> |
| CULBPT-05-20            | Cornell            | Arid       | OH08-5202               | Ohio State | Und.              |
| CULBPT-05-21            | Cornell            | Arid       | OH08-5203               | Ohio State | Und.              |
| CULBPT-05-22            | Cornell            | Arid       | OH08-5204               | Ohio State | Und.              |
| CULBPT-05-9             | Cornell            | Arid       | OH08-5205               | Ohio State | Und.              |
| CULBPT-A46-2            | Cornell            | Arid       | OH08-5206               | Ohio State | Und.              |
| E3259                   | Ohio State         | Humid      | OH08-5207               | Ohio State | Und.              |
| E6203/LA4024            | Ohio State         | Arid       | OH08-5210               | Ohio State | Und.              |
| F02-7530                | Ohio State         | Humid      | OH08-5211               | Ohio State | Und.              |
| F03-6331                | Ohio State         | Humid      | OH08-5213               | Ohio State | Und.              |
| F03-7463                | Ohio State         | Humid      | OH08-5215               | Ohio State | Und.              |
| F06-1013-1              | Ohio State         | Humid      | OH08-5216               | Ohio State | Und.              |
| F06-1014-1              | Ohio State         | Humid      | OH08-7438               | Ohio State | Und.              |
| F06-2041                | Ohio State         | Humid      | OH08-7439               | Ohio State | Und.              |
| F06-2054                | Ohio State         | Humid      | OH08-7454               | Ohio State | Und.              |
| F06-2058                | Ohio State         | Humid      | OH08-7457               | Ohio State | Und.              |
| FG02-188                | Ohio State         | Humid      | OH08-7458               | Ohio State | Und.              |
| 'Heinz 1706'/LA4345     | TGRC               | Humid      | OH08-7459               | Ohio State | Und.              |
| 'Hunt 100'/LA3144       | TGRC               | Arid       | OH08-7460               | Ohio State | Und.              |
| 'M82'/LA3475            | TGRC               | Arid       | OH08-7466               | Ohio State | Und.              |
| OH03-6439               | Ohio State         | Humid      | OH08-7469               | Ohio State | Und.              |
| OH05-8018               | Ohio State         | Humid      | OH08-7470               | Ohio State | Und.              |
| OH05-8022               | Ohio State         | Humid      | OH2641                  | Ohio State | Humid             |
| OH05-8025               | Ohio State         | Humid      | OH3614                  | Ohio State | Humid             |
| OH05-8027               | Ohio State         | Humid      | OH5-8127                | Ohio State | Humid             |
| OH05-8028               | Ohio State         | Humid      | OH5-8157                | Ohio State | Humid             |
| OH05-8030               | Ohio State         | Humid      | OH5-8164                | Ohio State | Humid             |
| OH05-8192               | Ohio State         | Humid      | OH7814                  | Ohio State | Humid             |
| OH05-8036               | Ohio State         | Humid      | OH7870                  | Ohio State | Humid             |
| OH05-8040               | Ohio State         | Humid      | OH7983                  | Ohio State | Humid             |
| OH05-8044               | Ohio State         | Humid      | OH8243/PI 601423        | Ohio State | Humid             |
| OH05-8046               | Ohio State         | Humid      | OH8245                  | Ohio State | Humid             |
| OH05-8048               | Ohio State         | Humid      | OH832                   | Ohio State | Humid             |
| OH05-8053               | Ohio State         | Humid      | OH8446                  | Ohio State | Humid             |
| OH05-8059               | Ohio State         | Humid      | OH8556                  | Ohio State | Humid             |
| OH05-8062               | Ohio State         | Humid      | OH86120                 | Ohio State | Humid             |
| OH05-8064               | Ohio State         | Humid      | OH8614-1                | Ohio State | Humid             |
| OH05-8068               | Ohio State         | Humid      | OH87160                 | Ohio State | Humid             |
| OH05-8069               | Ohio State         | Humid      | OH88119                 | Ohio State | Humid             |
| OH05-8070               | Ohio State         | Humid      | OH9241                  | Ohio State | Humid             |
| OH05-8072               | Ohio State         | Humid      | OH9242                  | Ohio State | Humid             |
| OH05-8074               | Ohio State         | Humid      | OH981049                | Ohio State | Humid             |
| OH05-8078               | Ohio State         | Humid      | OH981067                | Ohio State | Humid             |
| OH05-8079               | Ohio State         | Humid      | OH981136                | Ohio State | Arid              |
| OH05-8087               | Ohio State         | Humid      | OH981205                | Ohio State | Humid             |
| OH05-8090               | Ohio State         | Humid      | OH987034                | Ohio State | Humid             |

*Continued next page*

Table 1. Continued.

| Donor no./cultivar name | Donor <sup>2</sup> | Adaptation | Donor no./cultivar name | Donor        | Adaptation |
|-------------------------|--------------------|------------|-------------------------|--------------|------------|
| OH05-8095               | Ohio State         | Humid      | 'Oroma'                 | Oregon State | Arid       |
| OH05-8104               | Ohio State         | Humid      | 'Pearson'/LA0012        | TGRC         | Arid       |
| OH05-8109               | Ohio State         | Humid      | Peto95-43/LA3528        | TGRC         | Arid       |
| OH05-8117               | Ohio State         | Humid      | R001                    | Ridgetown    | Humid      |
| OH05-8124               | Ohio State         | Humid      | R9812                   | Ridgetown    | Humid      |
| OH05-8139               | Ohio State         | Humid      | R996                    | Ridgetown    | Humid      |
| OH05-8142               | Ohio State         | Humid      | R997                    | Ridgetown    | Humid      |
| OH05-8144               | Ohio State         | Humid      | 'Saucy'                 | Oregon State | Arid       |
| OH05-8151               | Ohio State         | Humid      | 'Sun1642'/TA491         | Public       | Arid       |
| OH05-8162               | Ohio State         | Humid      | UC-204B/LA4437          | TGRC         | Arid       |
| OH05-8163               | Ohio State         | Humid      | UC-82/LA1706            | TGRC         | Arid       |
| OH05-8165               | Ohio State         | Humid      | 'Unilever265'           | Public       | Arid       |
| OH05-8171               | Ohio State         | Humid      | 'PS696'/'Perfect Peel'  | Seminis      | Humid      |
| OH05-8175               | Ohio State         | Humid      |                         |              |            |

<sup>2</sup>Ohio State = The Ohio State University, Wooster; Cornell = Cornell University, Ithaca, NY; TGRC = Tomato Genetics Resource Center, Davis, CA; Oregon State = Oregon State University, Corvallis; Ridgetown = University of Guelph Ridgetown Campus, Ridgetown, Ontario, Canada; Public = germplasm in the public domain; Seminis = Seminis Vegetable Seeds, Inc., St. Louis.

<sup>3</sup>Undetermined adaptation.

replicated in each block to provide the ability to analyze the data as an augmented design (Federer and Raghavarao, 1975). In 2009, two locations (Ohio and California) were arranged as randomized complete block designs with two replications. In 2010, plots were also organized as randomized complete blocks with two replications intended for the Ohio and California locations. However, the second Ohio replicate was not harvested as a result of field conditions. With the exception of yield, data for all traits were obtained from seven plots. For yield, one replication was harvested from California in 2009 and one replication was harvested from Ohio in 2010. To obtain yield data from a third environment, an augmented design was grown and harvested in Ohio in 2011.

**EVALUATION OF PHENOTYPE.** Plots were harvested when 80% of fruit in a plot was red ripe. Plots were hand-harvested with the exception of one replicate in California in 2009, which was machine-harvested. Yield from this replicate was measured as total harvested weight. In the Ohio environments, yield was measured as total harvested weight and also as marketable yield based on the total weight minus cull tomatoes and green tomatoes.

A sample of 50 ripe fruit from each plot was used for measuring shape, size, color, and soluble solids and acid quality-associated traits. Five to 10 fruit from each plot were sliced longitudinally and the same number were cut along a latitudinal axis. Half of each fruit was placed on a flat-bed scanner and scanned to create a single digital image with multiple fruit for each plot. Images were saved in the .jpg compressed image format at a resolution of 100 dpi. Digital images of each plot were analyzed using Tomato Analyzer software [Version 2.2 (Brewer et al., 2006)]. The external fruit border, pericarp border (latitudinal slice only), and rotation of the fruit were adjusted manually as necessary. Six size traits and 28 shape traits were obtained from the longitudinal slices. Three shape traits and 11 color traits, including color uniformity as defined by the percentage of surface area that was "red" (hue 0–50) or "yellow" (hue 70–100), were obtained from the latitudinal slices. Four color traits (R, G, B, and luminosity) were obtained using the RGB color scale used by computers and seven ( $L^*$ ,  $a^*$ ,  $b^*$ , hue, chroma, and the two color

uniformity traits) were obtained using the  $L^*a^*b^*$  color space, which is a universal color space with defined standards (Commission Internationale de l'Eclairage, 1978).

Although shape traits have been previously defined (Brewer et al., 2006; Gonzalo et al., 2009; Gonzalo and van der Knaap, 2008), the descriptions are not intuitive; therefore, we briefly describe them here for important traits. The measure "external index I" describes fruit shape elongation as the ratio of length to width. "Proximal angle macro" measures pointedness at the stem end of the fruit as the angle at which best-fit lines drawn on either side of the end pass through 20% of the perimeter. "Circular" is a measure of how well the fruit fit a circle of diameter that is equivalent to the width and provides an estimate of how round the fruit are. "Shoulder height" is the ratio of average height of the fruit above the stem scar to the maximum height. "Proximal area indentation" is expressed as a ratio of the area bounded by the fruit shoulders on either side, the shoulder height at the top, and the stem scar at the bottom to the area of the fruit. "Ovoid" is a measure of asymmetry; it is the ratio of the average width above and below the maximum width.

Ten fruit per plot were quartered, and a section from each fruit was combined and homogenized in a blender to evaluate soluble solids, pH, and titratable acidity. Soluble solids were measured using a refractometer (Mark II Plus; ABBE, Depew, NY). Puree was filtered through a laboratory wipe, and refractive index was recorded. A sample of homogenate was diluted by adding 40 mL of distilled water to 10 mL of homogenate to measure pH and titratable acids using a standard laboratory pH meter ( $\phi$  45; Beckman, Brea, CA). Titratable acidity, which is predominantly the result of citric acid in tomato, was measured by titrating with 0.1 N NaOH to pH 8.2. Percent citric acid was calculated using the following equation:

$$\text{percent citric acid} = \left\{ \left[ (\text{mL NaOH} \times \text{N NaOH} \times 64 \text{ mg/meq}) / \text{mL sample titrated} \right] \times \text{dilution} \right\} \times 100$$

where 64 mg/meq is the equivalent weight of citric acid (Friedrich, 2001).



**STATISTICAL ANALYSIS.** All statistical analyses were performed using R software Version 2.13 for Mac (R Development Core Team, 2011). For each trait, components of phenotypic variance were estimated from analysis of variance using restricted maximum likelihood methods. The linear mixed effects “lmer” command in the lme4 package (Bates et al., 2011) was used to estimate variance components. For each trait except yield, the following linear model was used:

$$Y_{ijk} = \mu + G_i + L_j + Y_k + R(Y_k * L_j) + GL_{ij} + GY_{ik} + \varepsilon_{ijk}$$

where  $Y_{ijk}$  is the trait measured,  $\mu$  is the overall mean,  $G_i$  is the effect resulting from the  $i$ th genotype,  $L_j$  is the effect resulting from the  $j$ th location,  $Y_k$  is the effect resulting from the  $k$ th year,  $R(Y_k * L_j)$  is the effect resulting from replicate within year  $\times$  location,  $GL_{ij}$  is the effect resulting from genotype  $\times$  location interaction,  $GY_{ik}$  is the effect resulting from genotype  $\times$  year interaction, and  $\varepsilon_{ijk}$  is the residual error (effect resulting from experimental error). All effects were treated as random.

For yield measured across environments (location  $\times$  year), the following model was used:

$$Y_{ij} = \mu + G_i + E_j + GE_{ij} + \varepsilon_{ij}$$

where  $Y_{ij}$  is yield,  $\mu$  and  $G_i$  are as described previously,  $E_j$  is the effect resulting from the  $j$ th environment,  $GE_{ij}$  is the effect resulting from genotype  $\times$  environment interaction, and  $\varepsilon_{ij}$  is the residual error.

Best linear unbiased predictors were estimated for each line for each trait using the same models used to estimate variance components. The random effect “ranef” command in the lme4 package was used to estimate BLUPs for all terms in the model (Bates et al., 2011). The BLUPs for genotypes were extracted and used in principal components analysis (PCA) and correlation analysis. PCA was performed using the “prcomp” command. PCA identified which traits explained the most phenotypic variation among the germplasm evaluated. Pearson correlation coefficients and corresponding probability values were calculated between all pairwise combinations of traits. The “correlation” function in the agricolae package (de Mendiburu, 2010) was used because it provides an improved structure of the output relative to the core package.

To decide which traits were most relevant in the germplasm with respect to variability and correlation with other traits, eigenvector loadings from PCA, Pearson correlation coefficients, and proportion of variance resulting from genetics (i.e., genotype) were considered. Color traits measured using the  $L^*a^*b^*$  color space were preferentially retained because this is a standard for universal color measurement, whereas RGB values are subject to software interpretation (Darrigues et al., 2008). Eigenvector loadings from PCA were inspected and traits that contributed high positive or negative loadings to the first three PCs were retained. Pearson correlation coefficients were used to determine traits that were highly correlated. When two traits were highly correlated ( $r > 0.80$ ) and one of the traits had a higher proportion of variance resulting from genetics, the trait with the higher value was retained. Finally, any traits that had a proportion of variance resulting from genetics greater than or equal to 0.05 and that were not already selected based on the preceding criteria were retained for further analysis.

Two approaches were used to evaluate differences between regional processing germplasm. First, analysis of variance (ANOVA) was performed for each of the first three PCs with

the score for each line as the dependent variable and germplasm groups treated as levels. Levels were assigned to each line according to whether the line was adapted to California/Oregon (arid), the Great Lakes (humid), or was of undetermined adaptation. ANOVA was also performed on the BLUPs for the retained traits with levels as described previously. In both cases, a simple linear model was evaluated. When the effect of adaptation was significant ( $P < 0.05$ ), a Tukey’s test was performed using the multcomp package (Hothorn et al., 2008).

## Results

Raw data in flat file format are available through the Solanaceae Coordinated Agricultural Project (2012) and organized according to phenotype ontologies in the Sol Genomics Network (2012) database. We narrowed the number of traits for which we present data from 52 to 22 (Table 2). Thirty-one shape traits, 11 color traits, three chemical traits, six size traits, and total yield were reduced to seven shape traits, six color traits, three chemical traits related to fruit quality, five size traits, and total yield. Traits were selected based on the proportion of variance resulting from genetics, high positive or negative loadings to the first three PCs, and correlations between traits. In general, the five traits that explained the most variation for each of the first three PCs were retained with the exceptions of average green (fruit color), two redundant shape indices, and the fruit shape traits “proximal blockiness,” “triangle,” and “horizontal asymmetry.ov” (Table 2). Average green was eliminated because it is described in the RGB color space, which is subject to software interpretation, and was highly correlated with color measures in the  $L^*a^*b^*$  color space. The three shape indices were perfectly correlated; therefore, only “external index I” was retained. Three shape traits were not retained because the proportion of genotypic variance was less than 0.05. Yield was retained as a result of its economic importance despite a low proportion of variance resulting from genetics. Of the 22 traits retained, BLUPs were estimated on a per-genotype basis. As expected, there was a strong positive linear relationship between BLUPs and means with shrinkage of the BLUPs toward the population average (Fig. 1).

Principal components analysis was conducted to determine which traits were the major sources of variation within the germplasm panel. Cumulatively, the first three PCs explained 57.1% of the variation. PC1 accounted for 28.0% of the variation and was weighted toward traits describing shape: “proximal angle macro,” “external index I,” and “circular.” PC2 was also weighted toward traits describing shape (“horizontal asymmetry.ov,” “triangle,” “proximal blockiness,” “ovoid”) as well as color ( $L^*$ ). The second PC accounted for 16.2% of the variation. PC3 accounted for 12.9% of the variation and was weighted toward the traits describing color and size: percent yellow tissue,  $a^*$ , hue, perimeter, and area.

Pearson correlation coefficients were calculated between each pair of traits using the estimated BLUPs (Table 3). There were significant correlations among soluble solids and acid quality, yield, color, shape, and size traits. In Ohio, total harvested samples were sorted into red-ripe, cull (cracked or diseased fruit), or green fruit. Total yield and marketable (red-ripe) yield measured in Ohio were significantly correlated ( $r = 0.92$ ,  $P < 0.00001$ ; data not shown). Soluble solids were negatively correlated with yield ( $r = -0.31$ ,  $P = 0.0002$ ),

Table 2. Variance components for important yield and quality traits in tomato.

| Trait category |                                  | Variance component (proportion of total) <sup>z</sup> |                     |               |      |          | Replication within year × location | Residual |
|----------------|----------------------------------|---|---------------------|---------------|------|----------|------------------------------------|----------|
|                |                                  | Genotype  | Genotype × location | Genotype × yr | Year | Location |                                    |          |
| Yield          | Yield (kg·ha <sup>-1</sup> )     | 0.03  | 0.00                | NA            | NA   | 0.85     | NA                                 | 0.11     |
| Quality        | Soluble solids concentration (%) | 0.28  | 0.01                | 0.01          | 0.05 | 0.27     | 0.10                               | 0.37     |
|                | pH                               | 0.16  | 0.00                | 0.03          | 0.11 | 0.20     | 0.00                               | 0.51     |
|                | Titrateable acids (g/100 mL)     | 0.13  | 0.00                | 0.02          | 0.04 | 0.61     | 0.06                               | 0.21     |
| Fruit color    | Proportion red tissue            | 0.20  | 0.11                | 0.03          | 0.00 | 0.26     | 0.01                               | 0.40     |
|                | Proportion yellow tissue         | 0.41  | 0.25                | 0.03          | 0.01 | 0.01     | 0.00                               | 0.29     |
|                | L*                               | 0.29  | 0.07                | 0.03          | 0.10 | 0.00     | 0.01                               | 0.50     |
|                | a*                               | 0.12  | 0.02                | 0.01          | 0.18 | 0.53     | 0.00                               | 0.13     |
|                | b*                               | 0.17  | 0.10                | 0.04          | 0.02 | 0.27     | 0.00                               | 0.40     |
|                | Hue                              | 0.18  | 0.04                | 0.02          | 0.09 | 0.52     | 0.01                               | 0.16     |
|                |                                  |   |                     |               |      |          |                                    |          |
| Fruit size     | Perimeter (cm)                   | 0.08  | 0.07                | 0.06          | 0.00 | 0.31     | 0.00                               | 0.48     |
|                | Area (cm <sup>2</sup> )          | 0.10  | 0.07                | 0.06          | 0.00 | 0.33     | 0.00                               | 0.48     |
|                | Width midheight (cm)             | 0.06  | 0.06                | 0.05          | 0.01 | 0.38     | 0.00                               | 0.45     |
|                | Height midwidth (cm)             | 0.30  | 0.08                | 0.05          | 0.02 | 0.08     | 0.00                               | 0.48     |
|                | Maximum height (cm)              | 0.05  | 0.06                | 0.05          | 0.01 | 0.38     | 0.00                               | 0.45     |
| Fruit shape    | External index I                 | 0.51  | 0.03                | 0.01          | 0.05 | 0.10     | 0.07                               | 0.29     |
|                | Proximal angle macro (degrees)   | 0.07  | 0.03                | 0.00          | 0.33 | 0.16     | 0.02                               | 0.40     |
|                | Distal angle macro (degrees)     | 0.09  | 0.00                | 0.00          | 0.30 | 0.18     | 0.00                               | 0.43     |
|                | Circular                         | 0.46  | 0.03                | 0.02          | 0.16 | 0.02     | 0.00                               | 0.32     |
|                | Shoulder height                  | 0.08  | 0.05                | 0.01          | 0.02 | 0.19     | 0.03                               | 0.65     |
|                | Proximal area indentation        | 0.08  | 0.04                | 0.02          | 0.03 | 0.32     | 0.00                               | 0.52     |
|                | Ovoid                            | 0.05  | 0.06                | 0.02          | 0.00 | 0.20     | 0.00                               | 0.67     |
|                |                                  |   |                     |               |      |          |                                    |          |

<sup>z</sup>Yield was measured for 143 processing lines (genotypes) in three environments; one replication in California in 2009, one replication in Ohio in 2010, and one replication in Ohio in 2011. Fruit soluble solids, acid quality (pH, titrateable acids), color, shape, and size were measured in 2009 and 2010 in Ohio and California with two replications per location per year.

NA = not applicable.

although positively correlated with titrateable acids ( $r = 0.43$ ,  $P < 0.0001$ ). There were also significant correlations between color traits. L\* was negatively correlated with a\* and positively correlated with b\* ( $r = -0.64$  and  $0.77$ ,  $P < 0.001$  for both comparisons). The color uniformity traits were correlated with L\*, a\*, and b\*. The correlations between percent yellow tissue and a\*, percent red tissue and L\*, and percent red tissue and b\* were negative (range of  $r = -0.76$  to  $-0.93$ ,  $P < 0.0001$ ). Hue was also correlated with the color uniformity traits ( $r = -0.83$  and  $0.95$ ,  $P < 0.0001$ ). For processing tomatoes, low values of L\* and hue and high values of percent red tissue are generally desirable.

Size and shape traits were also significantly correlated. The size traits were generally positively correlated (range of  $r$  for significant correlations =  $0.47$  to  $1.00$ ,  $P < 0.0001$ ); width midheight and maximum height were perfectly correlated. Height midwidth and “external index I” were positively correlated ( $r = 0.85$ ,  $P < 0.0001$ ); all other significant correlations with shape were negative (range of  $r = -0.25$  to  $-0.97$ ,  $P \leq 0.02$ ).

Variance components resulting from genetic, environmental, and genetic-by-environment interaction factors were estimated for each retained trait using REML (Table 2). Yield had the lowest proportion of genetic variance (0.03), whereas fruit quality (variance range =  $0.13$ – $0.28$ ) and color traits (variance range =  $0.12$ – $0.41$ ) had relatively high levels of genetic variance. Fruit size and shape traits ranged from  $0.05$  to  $0.51$ . Notable shape traits with high genetic variance were “external index I,” which had the highest proportion of variance resulting from genetics ( $0.51$ ), height midwidth ( $0.30$ ), and “circular”

( $0.46$ ). Overall, fruit soluble solids and acid quality, color, and several size and shape traits had a relatively high proportion of variance resulting from genetic effects with low proportions of variance resulting from genetic interaction effects.

The first three PCs could distinguish regionally adapted germplasm. Using ANOVA and a post hoc Tukey’s test with the eigenvector loading for each accession for PC1, PC2, or PC3 as the continuous variable, cultivars adapted to arid or humid production systems could be distinguished ( $P < 0.01$  for all comparisons) (Fig. 2). Regionally adapted germplasm could also be distinguished based on shape, size, color, and quality traits using the trait value for each accession as the continuous variable. “External index I” was greater in arid-adapted germplasm compared with humid-adapted germplasm ( $P = 0.0016$ ) (Fig. 1E). Conversely, “proximal angle macro” was greater in humid-adapted germplasm compared with arid-adapted germplasm ( $P = 0.024$ ). Regionally adapted germplasm could also be distinguished based on perimeter, area, and height midwidth. All three traits were greater in arid-adapted germplasm compared with humid-adapted germplasm ( $P < 0.0001$  for all comparisons) (Fig. 1). Humid-adapted germplasm had higher yield compared with arid-adapted germplasm ( $P = 0.0025$ ), although arid-adapted germplasm had higher soluble solids ( $P < 0.0001$ ) (Fig. 1A–B). The color traits percent red and L\* could be used to distinguish humid from arid-adapted germplasm ( $P = 0.0018$  and  $0.027$ , respectively). Humid-adapted germplasm had higher percent red (Fig. 1C) and lower L\*. Arid-adapted germplasm grown in Ohio had higher percent cull compared with humid-adapted germplasm ( $P < 0.05$ ; data not shown). Overall, shape, size, color, yield,

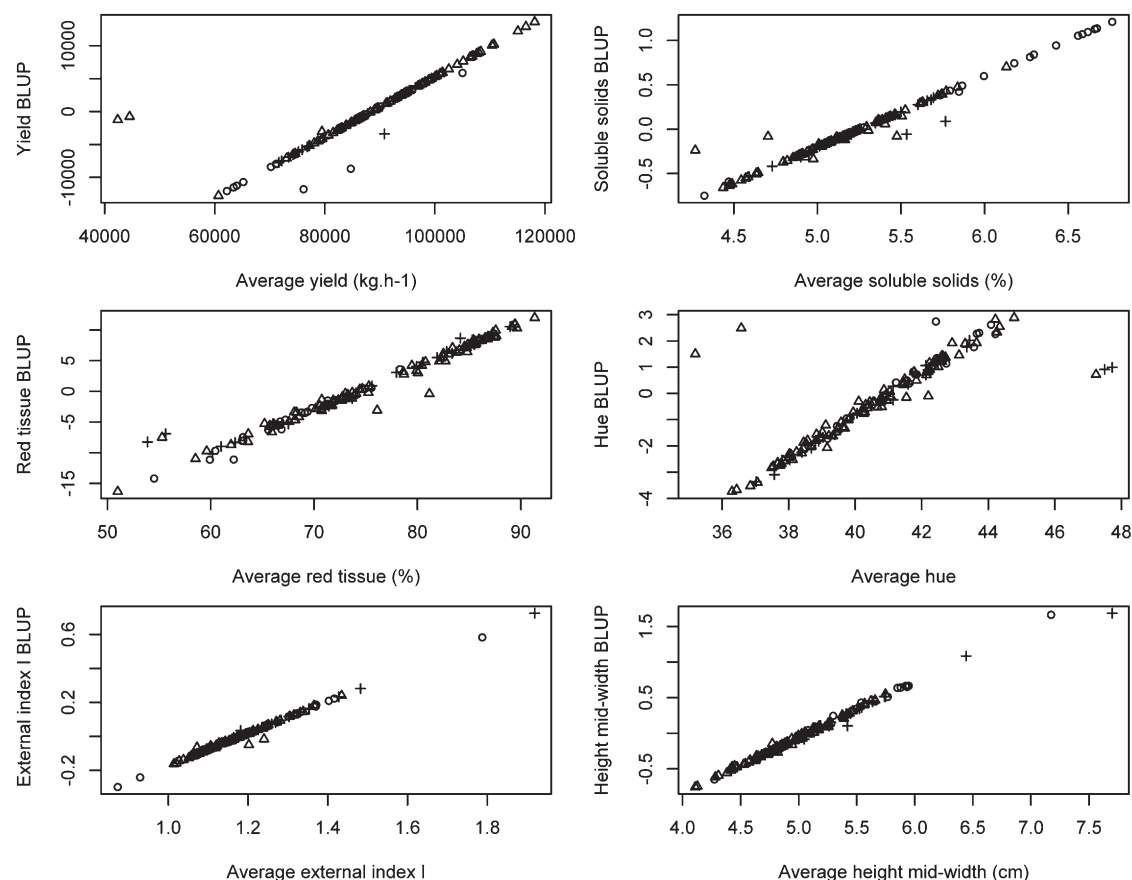


Fig. 1. Correspondence between best linear unbiased predictor (BLUP) and mean of each accession in the processing tomato germplasm for (A) yield, (B) soluble solids, (C) percent red tissue, (D) hue, (E) external index I, and (F) height midwidth. Humid-adapted accessions, arid-adapted accessions, and accessions with undetermined adaptation are labeled with triangles, circles, and crosses, respectively.

and soluble solids and acid quality traits could all be used to distinguish regional breeding programs.

## Discussion

We estimated variance components and proportion of variance resulting from genetics for 51 traits and have stored these data as flat files (Solanaceae Coordinated Agricultural Project, 2012) and in an ontology database (Sol Genomics Network, 2012). Many of these traits contain redundant information related to fruit shape, size, and color allowing us to reduce the number of highly relevant traits to 22 based on objective criteria. We estimated phenotypic values across environments as BLUPs, providing a baseline for trait values. An important outcome of our analysis is the discovery that regionally adapted germplasm provides a source of trait variation. The humid-adapted germplasm may serve as a source of color uniformity, deep red color, and shape variation. In contrast, the arid-adapted germplasm could serve as a source of increased soluble solids and fruit size. Differences in regional germplasm are likely the result of human selection for varying market needs and are reflected in genetic substructure (Sim et al., 2011).

The estimated genetic components of variance were low to moderate for most traits. The relatively low genetic variance may be a reflection of the fact that the germplasm panel consisted of cultivars and breeding lines that have experienced

strong selection. In addition, low genetic variance may reflect complex genetic inheritance and/or a high environmental influence on trait expression. Previous studies, conducted in the context of biparental mapping, suggest that the genetic basis of tomato fruit size, shape, and soluble solids and acid quality traits are the result of many quantitative trait loci [QTL (Azanza et al., 1994; Brewer et al., 2007; Chen et al., 1999; Gonzalo and van der Knaap, 2008; Grandillo et al., 1999; Osborn et al., 1987; Saliba-Colombani et al., 2001; Tanksley et al., 1996; Tanksley and Hewitt, 1988)].

Improvement of quantitative traits using conventional marker-assisted selection has had limited success relative to expectations (Bernardo, 2008; Heffner et al., 2009). An emphasis on trait and linkage discovery in biparental populations, small population sizes, marker density, and marker phase relative to QTL all contribute to the inability to translate discovery of linkage into marker-assisted breeding strategies (Bernardo, 2008; Xu and Crouch, 2008). Research implemented on commodity grains suggests several strategies that might improve breeding and marker-assisted breeding of horticultural crops, including tomato. Important features of these strategies include incorporating pedigree (Crossa et al., 2010) or kinship data (Heffner et al., 2011) to strengthen estimates of trait BLUPs, and experimental designs that shift the balance of genotypic and technical replication to increase population sizes (Federer and Raghavarao, 1975). Incorporation of pedigree information, increased population sizes, and

Table 3. Pearson correlation coefficients for each pair of traits for tomato yield, fruit soluble solids and acid quality, color, size, and shape measured for 143 processing lines.

|                                  | Yield<br>(kg·h <sup>-1</sup> ) | Soluble<br>solids<br>concn (%) | pH       | Titratable<br>acids<br>(g/100 mL) | Proportion<br>of yellow<br>tissue | Proportion<br>of red<br>tissue | L*       | a*       | b*      | Hue     | Perimeter<br>(cm) | Area<br>(cm <sup>2</sup> ) | Width<br>midheight<br>(cm) | Ht<br>midwidth<br>(cm) | Maximum<br>ht<br>(cm) | External<br>index I | Proximal       |           | Distal         |           | Shoulder<br>ht | Proximal<br>area<br>indentation |
|----------------------------------|--------------------------------|--------------------------------|----------|-----------------------------------|-----------------------------------|--------------------------------|----------|----------|---------|---------|-------------------|----------------------------|----------------------------|------------------------|-----------------------|---------------------|----------------|-----------|----------------|-----------|----------------|---------------------------------|
|                                  |                                |                                |          |                                   |                                   |                                |          |          |         |         |                   |                            |                            |                        |                       |                     | angle<br>macro | (degrees) | angle<br>macro | (degrees) |                |                                 |
| Soluble solids concentration (%) | -0.31***                       |                                |          |                                   |                                   |                                |          |          |         |         |                   |                            |                            |                        |                       |                     |                |           |                |           |                |                                 |
| pH                               | -0.16                          | -0.06                          |          |                                   |                                   |                                |          |          |         |         |                   |                            |                            |                        |                       |                     |                |           |                |           |                |                                 |
| Titratable acids (g/100 mL)      | 0.08                           | 0.43***                        | -0.72*** |                                   |                                   |                                |          |          |         |         |                   |                            |                            |                        |                       |                     |                |           |                |           |                |                                 |
| Proportion of yellow tissue      | -0.04                          | 0.14                           | -0.16    | 0.21**                            |                                   |                                |          |          |         |         |                   |                            |                            |                        |                       |                     |                |           |                |           |                |                                 |
| Proportion of red tissue         | 0.05                           | -0.11                          | 0.27***  | -0.16                             | -0.69***                          |                                |          |          |         |         |                   |                            |                            |                        |                       |                     |                |           |                |           |                |                                 |
| L*                               | 0.06                           | 0.09                           | -0.23**  | 0.13                              | 0.70***                           | -0.76***                       |          |          |         |         |                   |                            |                            |                        |                       |                     |                |           |                |           |                |                                 |
| a*                               | 0.04                           | -0.13                          | 0.17*    | -0.19*                            | -0.93***                          | 0.81***                        | -0.64*** |          |         |         |                   |                            |                            |                        |                       |                     |                |           |                |           |                |                                 |
| b*                               | -0.04                          | 0.03                           | -0.14    | 0.12                              | 0.82***                           | -0.79***                       | 0.77***  | -0.79*** |         |         |                   |                            |                            |                        |                       |                     |                |           |                |           |                |                                 |
| Hue                              | -0.05                          | 0.11                           | -0.16    | 0.17*                             | 0.95***                           | -0.83***                       | 0.72***  | -0.97*** | 0.91*** |         |                   |                            |                            |                        |                       |                     |                |           |                |           |                |                                 |
| Perimeter (cm)                   | -0.01                          | 0.18*                          | 0.11     | -0.24**                           | -0.04                             | -0.01                          | 0.26**   | 0.12     | 0.01    | -0.07   | 0.98***           |                            |                            |                        |                       |                     |                |           |                |           |                |                                 |
| Area (cm <sup>2</sup> )          | 0.05                           | 0.14                           | 0.09     | -0.22**                           | -0.06                             | 0.03                           | 0.24**   | 0.15     | -0.02   | -0.1    | 0.47***           | 0.57***                    |                            |                        |                       |                     |                |           |                |           |                |                                 |
| Width midheight (cm)             | 0.26**                         | -0.02                          | -0.11    | 0.07                              | -0.11                             | 0.19*                          | -0.1     | 0.19*    | -0.17*  | -0.18*  | 0.87***           | 0.81***                    | 0.01                       |                        |                       |                     |                |           |                |           |                |                                 |
| Height midwidth (cm)             | -0.12                          | 0.17*                          | 0.19*    | -0.31***                          | 0.01                              | -0.1                           | 0.34***  | 0.06     | 0.11    | 0.01    | 0.52***           | 0.61***                    | 1.00***                    | 0.06                   |                       |                     |                |           |                |           |                |                                 |
| Maximum height (cm)              | 0.24**                         | -0.01                          | -0.11    | 0.05                              | -0.11                             | 0.18*                          | -0.09    | 0.19*    | -0.16*  | -0.18*  | 0.51***           | 0.39***                    | -0.51***                   | 0.85***                | -0.46***              |                     |                |           |                |           |                |                                 |
| External index I                 | -0.25**                        | 0.18*                          | 0.21**   | -0.28***                          | 0.08                              | -0.20*                         | 0.33***  | -0.07    | 0.18*   | 0.12    | 0.43***           | -0.32***                   | 0.51***                    | -0.79***               | -0.47***              | -0.94***            |                |           |                |           |                |                                 |
| Proximal angle macro (degrees)   | 0.22**                         | -0.11                          | -0.14    | 0.25**                            | -0.05                             | 0.20*                          | -0.32*** | 0.05     | -0.17   | -0.1    | -0.43***          | -0.32***                   | 0.51***                    | -0.79***               | -0.47***              | -0.94***            | 0.92***        |           |                |           |                |                                 |
| Distal angle macro (degrees)     | 0.22**                         | -0.32***                       | -0.19*   | 0.14                              | -0.12                             | 0.26**                         | -0.31*** | 0.25**   | -0.20*  | -0.24** | -0.54***          | -0.47***                   | 0.31***                    | -0.76***               | -0.76***              | -0.84***            | 0.90***        | 0.82***   |                |           |                |                                 |
| Circular                         | 0.30***                        | -0.25**                        | -0.18*   | 0.25**                            | -0.1                              | 0.22**                         | -0.33*** | 0.1      | -0.19*  | -0.13   | -0.55***          | -0.44***                   | 0.44***                    | -0.84***               | 0.40***               | -0.97***            | 0.90***        | 0.82***   | 0.26**         |           |                |                                 |
| Shoulder height                  | 0.05                           | 0.09                           | -0.14    | 0.17*                             | 0.03                              | -0.01                          | -0.09    | -0.04    | -0.06   | 0       | -0.04             | 0                          | 0.39***                    | -0.35***               | 0.37***               | -0.45***            | 0.57***        | 0.26**    | 0.35***        | 0.97***   |                |                                 |
| Proximal area indentation        | 0.02                           | 0.15                           | -0.13    | 0.17*                             | 0.05                              | -0.02                          | -0.05    | -0.05    | -0.05   | 0.01    | 0                 | 0.04                       | 0.40***                    | -0.32***               | 0.38***               | -0.43***            | 0.57***        | 0.23**    | 0.33***        | 0.33***   | 0.97***        |                                 |
| Ovoid                            | -0.01                          | 0.19*                          | 0.19*    | 0.01                              | 0.15                              | -0.01                          | -0.08    | -0.15    | -0.01   | 0.1     | -0.04             | 0                          | 0.21**                     | -0.18*                 | 0.20*                 | -0.25**             | 0.48***        | -0.03     | 0.20*          | 0.33***   | 0.34***        |                                 |

\*, \*\*, and \*\*\* indicate significance at the 0.05, 0.01, and 0.001 *P* levels, respectively, based on the *t* distribution.



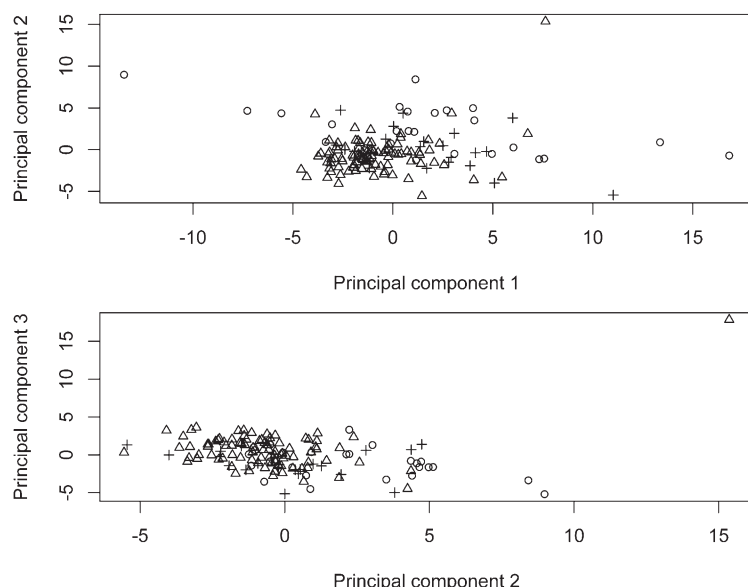


Fig. 2. Eigenvector loading for each accession in the processing tomato germplasm panel for (A) the first and second principal components and (B) the second and third principal components. Humid-adapted accessions, arid-adapted accessions, and accessions with undetermined adaptation are labeled with triangles, circles, and crosses, respectively.

use of BLUPs are driven by a desire to incorporate GWS into plant breeding (Meuwissen et al., 2001). Models for GWS in plants have been developed and evaluated for maize using simulated data and suggest that gain under selection can be improved, especially when selection is tied to increased population turnover (Bernardo, 2009; Bernardo and Yu, 2007). More recently, models for GWS have been evaluated using actual data or a combination of real and simulated data for barley (Iwata and Jannink, 2011), maize (Zhao et al., 2012), oat (*Avena sativa* L.) (Asoro et al., 2011), and wheat (Heffner et al., 2011). These analyses support the idea that gain under selection can be improved through GWS. Developing robust GWS approaches for horticultural crops will require not only a renewed commitment to collecting objective data, but also a commitment to increasing population sizes.

In addition to some structural changes in breeding strategies, attention will need to be directed toward multitrait selection indices (MTIs). For horticultural crops, selection is often based on quality traits in addition to yield. Multitrait selection models assign relative value to traits, yet determining relative value remains a challenge for many crops where perceived quality is important. In animal breeding, net merit is linked to economic value, and thus MTIs are directly modeled based on economic return (Hazel, 1943; Weller, 1994). When applying MTI, examining correlations between traits becomes particularly important. Like in numerous previous studies, soluble solids were strongly negatively correlated with yield (Azanza et al., 1994; Grandillo et al., 1999; Tanksley et al., 1996; Tanksley and Hewitt, 1988). For both traits, high values are desired. Determining how to balance the negative correlation can be done, like in animal breeding, based on economic value to the grower or processor. Alternatively, minimum standards can be set for one trait while selecting for improvement in another.

The challenge of developing MTI becomes clear when one examines what the market measures, perceptions of what traits are important, and what the market is actually willing to pay for

specific traits. Yield is directly valued in contracts between growers and processors. Depending on the region, yield may not be valued as a linear function. In Ontario, Canada, contracts for processing tomatoes include a productivity incentive designed to focus both the grower and processor on the goal of achieving predictable yields per hectare while providing adequate support for growers to produce a profitable crop under variable environmental conditions. When yield exceeds a threshold, growers receive a lower per-tonne payment. These contracts support a higher gross return per hectare and processors receive lower per-unit raw product costs when growers are successful. Contracts may also add incentives for quality. Fruit color is rewarded by contracts in Ontario and the midwestern United States with a tiered pricing structure paid per tonne depending on the grade option(s) chosen by the processor (Ontario Processing Vegetable Growers, 2011). Based on our analysis, selection for absolute color should focus on  $L^*$  (identified as a contributor to PC2) and  $hue^*$  (contributing to PC3). Because  $a^*$  and

$hue$  are strongly correlated (negatively), selection for both would be redundant. Selection for  $hue$  is favored as a result of a marginally higher genetic contribution and because, similar to  $L^*$ , a lower value is desirable. PC analysis also suggests that selection for color uniformity should focus on reducing the percentage of yellow tissue. Thus, the three most important color traits involve selection for lower numbers, a fact that will facilitate scaling. Ideally, the weight of the three color traits relative to yield and other quality traits would be determined by market forces as stipulated in contract structures.

Our analysis also detected significant variation for traits related to both shape and size of tomato fruit. Fruit size is a component of yield and will also affect market use. Fruit size and product recovery after peeling for value-added products may be related with larger fruit leading to higher recovery, thus affecting factory yield (Barringer et al., 1999; Garcia and Barrett, 2006). In contrast, whole peel tomato size is determined by can size, with a desired number of five fruit in a number 300 can (450 mL). Anecdotal information suggests that proximal or distal indentation may affect peel retention and thus processing costs for value-added products. Finally, shape is a key trait for highly specialized markets such as “Italian” whole peel tomatoes, characterized by extreme fruit-shape index (greater than 1). Incorporating shape and size information into a selection index will require that these traits be more directly linked to value either to growers or processors.

As we continue to incorporate genome-assisted selection methods into breeding programs, development of criteria for selection, including MTI based on market value, will be important to stimulate efforts to achieve gain under selection. The need for this link between trait and value is most clearly illustrated by the tradeoff between soluble solids and yield. The economic value of higher soluble solids remains inconsistently rewarded in contracts. Although growers and processors recognize the value of high soluble solids in ripe tomato fruit for the production of paste, contract rewards for high solids are

often revenue-neutral. In some production areas, a lack of incentive is by design. For example, under rain-fed agriculture in the Great Lakes region, there is no reward for high solids in the contract because of the complexity of management trade-offs, including weather and plant investment in foliar health. Growers could manage for higher solids but lose any gains resulting from precipitation at the end of the season. Thus, rewarding the value of solids in a contract above minimum market standards is not practical as a result of the growth environment. In the arid environment of California, contracts sometimes include incentives for soluble solids, the system rewards management, rather than choice of genetics. Cultivars that meet minimum industry standards are grouped. Solids are measured in each load delivered and the grower is either rewarded or penalized based on how the soluble solids level compares with the three- to five-year average for the cultivar group. Rates for soluble solids are negotiated based on the goals of the processor. The contract structure typically penalizes high tonnage growers with very low levels of soluble solids. For growers with very high soluble solids, the bonus for increased solids does not balance the lost revenue associated with lower yield. Although this approach maintains solids, it does not lead to gains over time (Grandillo et al., 1999). The contractual agreements between growers and processors are not always structured to reward what some members of the sector value. If the structure of contracts rewarded higher soluble solids based on cultivar choice, then breeders would place a much higher priority on this trait and we predict a corresponding shift in the soluble solids levels that has not yet occurred (Grandillo et al., 1999). Alternatively, soluble solid levels could be indexed to minimum standards in MTI, in which case we predict maintenance of the status quo. Over the past 10 years, soluble solids in California have not fluctuated much from an industry mean value of 5.2% soluble solids estimated across the top 50 cultivars [California Tomato Research Institute (CTRI), 2002, 2011]. This observation has led to a suggested industry minimum standard of 5.0% soluble solids. However, over 20% of the top 50 cultivars and three of the top 10 cultivars fall below this standard (CTRI, 2011).

The phenotypic data collected for this study provide a baseline data set describing trait variance, which will prove useful in developing selection models. The data collected are publicly available in spreadsheet format (Solanaceae Coordinated Agricultural Project, 2012) and in searchable database format on the Sol Genomics Network (2012). Significant genetic variation exists for fruit size, shape, color, soluble solids and acid quality, and yield in the germplasm panel. Lines from regional breeding programs may serve as a valuable source of genetic variation for germplasm improvement. A technological shift in focus to include genome-wide marker data and estimates of breeding value as selection criteria will likely also force the development of MTI for horticultural crops. These MTIs will need to balance correlated traits and must be consistent with market valuation. The fact that quality traits rewarded in contracts in the Great Lakes region are the traits for which lines adapted to the region tend to excel suggests that economic forces can shape selection. Despite this example, we believe a gap exists between what breeders select for and what the market is currently willing to pay for. Resolving this gap will require adjustments to how MTIs are incorporated into selection strategies or a shift in how the processing industry values traits.

## Literature Cited

- Asoro, F.G., M.A. Newell, W.D. Beavis, M.P. Scott, and J.-L. Jannink. 2011. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. *Plant Genome* 4:132–144.
- Azanza, F., T.E. Young, D. Kim, S.D. Tanksley, and J.A. Juvik. 1994. Characterization of the effect of introgressed segments of chromosome 7 and chromosome 10 from *Lycopersicon chmielewskii* on tomato soluble solids, pH, and yield. *Theor. Appl. Genet.* 87:965–972.
- Barringer, S.A., M.A. Bennett, and W.D. Bash. 1999. Effect of fruit maturity and nitrogen fertilizer on tomato peeling efficiency. *J. Veg. Crop Production* 5:3–11.
- Bates, D., M. Maechler, and B. Bolker. 2011. lme4: Linear mixed-effects models using eigen and jags. 1 Sept. 2011. <<http://cran.r-project.org/web/packages/lme4/index.html>>.
- Bernardo, R. 2008. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Sci.* 48:1649–1664.
- Bernardo, R. 2009. Genomewide selection for rapid introgression of exotic germplasm in maize. *Crop Sci.* 49:419–425.
- Bernardo, R. and J. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* 47:1082–1090.
- Brewer, M.T., L.X. Lang, K. Fujimura, N. Dujmovic, S. Gray, and E. van der Knaap. 2006. Development of a controlled vocabulary and software application to analyze fruit shape variation in tomato and other plant species. *Plant Physiol.* 141:15–25.
- Brewer, M.T., J.B. Moyseenko, A.J. Monforte, and E. van der Knaap. 2007. Morphological variation in tomato: A comprehensive study of quantitative trait loci controlling fruit shape and development. *J. Expt. Bot.* 58:1339–1349.
- California Tomato Research Institute. 2002. CTRI 2002 annual report. California Tomato Research Institute, Escalon, CA.
- California Tomato Research Institute. 2011. CTRI 2011 annual report. California Tomato Research Institute, Escalon, CA.
- Chen, F.Q., M.R. Foolad, J. Hyman, D.A. St Clair, and R.B. Beelaman. 1999. Mapping of QTLs for lycopene and other fruit traits in a *Lycopersicon esculentum* × *L. pimpinellifolium* cross and comparison of QTLs across tomato species. *Mol. Breed.* 5:283–299.
- Commission Internationale de l'Eclairage. 1978. Recommendations on uniform color spaces: Color-difference equations, psychometric color terms. Commission Internationale de l'Eclairage, Paris, France.
- Crossa, J., G. de los Campos, P. Perez, D. Gianola, J. Burgueno, J.L. Araus, D. Makumbi, R.P. Singh, S. Dreisigacker, J.B. Yan, V. Arief, M. Banziger, and H.J. Braun. 2010. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186:713–U406.
- Darrigues, A., J. Hall, E. van der Knaap, D.M. Francis, N. Dujmovic, and S. Gray. 2008. Tomato analyzer-color test: A new tool for efficient digital phenotyping. *J. Amer. Soc. Hort. Sci.* 133:579–586.
- de los Campos, G., H. Naya, D. Gianola, J. Crossa, A. Legarra, E. Manfredi, K. Weigel, and J.M. Cotes. 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182:375–385.
- de Mendiburu, F. 2010. Agricolae: Statistical procedures for agricultural research. 1 Sept. 2011. <<http://cran.r-project.org/web/packages/agricolae/index.html>>.
- Federer, W.T. and D. Raghuvarao. 1975. Augmented designs. *Biometrics* 31:29–35.
- Friedrich, J.E. 2001. Titratable activity of acid tastants. *Current Protocols Food Anal. Chem.* G2.1.1–G2.1.7.
- Garcia, E. and D.M. Barrett. 2006. Peelability and yield of processing tomatoes by steam or lye. *J. Food Processing Preservation* 30:3–14.
- Gonzalo, M.J., M.T. Brewer, C. Anderson, D. Sullivan, S. Gray, and E. van der Knaap. 2009. Tomato fruit shape analysis using morphometric and morphology attributes implemented in tomato analyzer software program. *J. Amer. Soc. Hort. Sci.* 134:77–87.

- Gonzalo, M.J. and E. van der Knaap. 2008. A comparative analysis into the genetic bases of morphology in tomato varieties exhibiting elongated fruit shape. *Theor. Appl. Genet.* 116:647–656.
- Grandillo, S., D. Zamir, and S.D. Tanksley. 1999. Genetic improvement of processing tomatoes: A 20 years perspective. *Euphytica* 110:85–97.
- Hartmann, A., T. Czauderna, R. Hoffmann, N. Stein, and F. Schreiber. 2011. Htphen: An image analysis pipeline for high-throughput plant phenotyping. *BMC Bioinformatics* 12:148.
- Hazel, L.N. 1943. The genetic basis for constructing selection indices. *Genetics* 28:476–490.
- Heffner, E.L., J.-L. Jannink, and M.E. Sorrells. 2011. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome* 4:65–75.
- Heffner, E.L., M.E. Sorrells, and J.L. Jannink. 2009. Genomic selection for crop improvement. *Crop Sci.* 49:1–12.
- Hothorn, T., F. Bretz, and P. Westfal. 2008. Simultaneous inference in general parametric models. *Biometrical J.* 50:346–363.
- International Plant Genetic Resources Institute. 1996. Descriptors for tomato (*Lycopersicon* spp.). Intl. Plant Genet. Resources Inst., Rome, Italy.
- International Union for the Protection of New Varieties and Plants. 2001. Guidelines for the conduct of tests for distinctness, homogeneity and stability (tomato). UPOV, Geneva, Switzerland.
- Iwata, H. and J.-L. Jannink. 2011. Accuracy of genomic selection prediction in barley breeding programs: A simulation study based on the real single nucleotide polymorphism data of barley breeding lines. *Crop Sci.* 51:1915–1927.
- Jung, S., N. Menda, S. Redmond, R.M. Buels, M. Friesen, Y. Bendana, L.A. Sanderson, H. Lapp, T. Lee, B. MacCallum, K.E. Bett, S. Cain, D. Clements, L.A. Mueller, and D. Main. 2011. The chado natural diversity module: A new generic database schema for large-scale phenotyping and genotyping data. *Database* 2011:bar051. 6 Aug. 2012. <<http://database.oxfordjournals.org/content/2011/bar051.full>>.
- Kump, K.L., P.J. Bradbury, R.J. Wissner, E.S. Buckler, A.R. Belcher, M.A. Oropeza-Rosas, J.C. Zwonitzer, S. Kresovich, M.D. McMullen, D. Ware, P.J. Balint-Kurti, and J.B. Holland. 2011. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nat. Genet.* 43:163–U120.
- Lorenz, A.J., M.T. Hamblin, and J.L. Jannink. 2010. Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. *PLoS One* 5:e14079.
- Menda, N., R.M. Buels, I. Tecle, and L.A. Mueller. 2008. A community-based annotation framework for linking Solanaceae genomes with phenomes. *Plant Physiol.* 147:1788–1799.
- Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Milc, J., A. Sala, S. Bergamaschi, and N. Pecchioni. 2011. A genotypic and phenotypic information source for marker-assisted selection of cereals: The cerealab database. *Database* 2011:baq038. 17 Aug. 2012. <<http://database.oxfordjournals.org/content/2011/baq038.full>>.
- Ontario Processing Vegetable Growers. 2011. Agreement for marketing the 2011 crop of tomatoes for processing under the farm products marketing act. Ontario Processing Vegetable Growers, London, Ontario, Canada.
- Osborn, T.C., D.C. Alexander, and J.F. Fobes. 1987. Identification of restriction-fragment-length-polymorphisms linked to genes-controlling soluble solids content in tomato fruit. *Theor. Appl. Genet.* 73:350–356.
- R Development Core Team. 2011. R: A language and environment for statistical computing. 1 Sept. 2011. <<http://www.r-project.org/>>.
- Riedelsheimer, C., A. Czedik-Eysenberg, C. Grieder, J. Lisec, F. Technow, R. Sulpice, T. Altmann, M. Stitt, L. Willmitzer, and A.E. Melchinger. 2012. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat. Genet.* 44:217–220.
- Rodriguez, G.R., S. Munos, C. Anderson, S.C. Sim, A. Michel, M. Causse, B.B.M. Gardener, D. Francis, and E. van der Knaap. 2011. Distribution of sun, ovate, lc, and fas in the tomato germplasm and the relationship to fruit shape diversity. *Plant Physiol.* 156:275–285.
- Saliba-Colombani, V., M. Causse, D. Langlois, J. Philouze, and M. Buret. 2001. Genetic analysis of organoleptic quality in fresh market tomato. 1. Mapping QTLs for physical and chemical traits. *Theor. Appl. Genet.* 102:259–272.
- Sim, S.C., M.D. Robbins, A. Van Deynze, A.P. Michel, and D.M. Francis. 2011. Population structure and genetic differentiation associated with breeding history and selection in tomato (*Solanum lycopersicum* L.). *Heredity* 106:927–935.
- Sol Genomics Network. 2012. Search SGN. 10 Apr. 2012. <<http://solgenomics.net/search/phenotypes>>.
- Solanaceae Coordinated Agricultural Project. 2012. Tomato phenotype data. 10 Apr. 2012. <[http://www.solcap.msu.edu/tomato\\_phenotype\\_data.shtml](http://www.solcap.msu.edu/tomato_phenotype_data.shtml)>.
- Souza, E.J., C. Sneller, M.J. Guttieri, A. Sturbaum, C. Griffey, M. Sorrells, H. Ohm, and D. Van Sanford. 2012. Basis for selecting soft wheat for end-use quality. *Crop Sci.* 52:21–31.
- Tanksley, S.D., S. Grandillo, T.M. Fulton, D. Zamir, Y. Eshed, V. Petiard, J. Lopez, and T. BeckBunn. 1996. Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L-pimpinellifolium*. *Theor. Appl. Genet.* 92:213–224.
- Tanksley, S.D. and J. Hewitt. 1988. Use of molecular markers in breeding for soluble solids content in tomato—A re-examination. *Theor. Appl. Genet.* 75:811–823.
- Tian, F., P.J. Bradbury, P.J. Brown, H. Hung, Q. Sun, S. Flint-Garcia, T.R. Rocheford, M.D. McMullen, J.B. Holland, and E.S. Buckler. 2011. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat. Genet.* 43:159–U113.
- Wang, M.H., N. Jiang, T.Y. Jia, L. Leach, J. Cockram, R. Waugh, L. Ramsay, B. Thomas, and Z.W. Luo. 2012. Genome-wide association mapping of agronomic and morphologic traits in highly structured populations of barley cultivars. *Theor. Appl. Genet.* 124:233–246.
- Weller, J.I. (ed.). 1994. Economic aspects of animal breeding. Kluwer, Chapman and Hall, London, UK.
- Xu, Y.B. and J.H. Crouch. 2008. Marker-assisted selection in plant breeding: From publications to practice. *Crop Sci.* 48:391–407.
- Zhao, Y.S., M. Gowda, W.X. Liu, T. Wurschum, H.P. Maurer, F.H. Longin, N. Ranc, and J. Reif. 2012. Accuracy of genomic selection in european maize elite breeding populations. *Theor. Appl. Genet.* 124:769–776.