

# Development of a Controlled Vocabulary and Software Application to Analyze Fruit Shape Variation in Tomato and Other Plant Species<sup>1[W]</sup>

Marin Talbot Brewer, Lixin Lang, Kikuo Fujimura, Nancy Dujmovic, Simon Gray, and Esther van der Knaap\*

Department of Horticulture and Crop Science, The Ohio State University/Ohio Agricultural Research and Development Center, Wooster, Ohio 44691 (M.T.B., E.v.d.K.); Department of Statistics (L.L.) and Department of Computer Science and Engineering (K.F.), The Ohio State University, Columbus, Ohio 43210; and Department of Mathematics and Computer Science, The College of Wooster, Wooster, Ohio 44691 (N.D., S.G.)

The domestication and improvement of fruit-bearing crops resulted in a large diversity of fruit form. To facilitate consistent terminology pertaining to shape, a controlled vocabulary focusing specifically on fruit shape traits was developed. Mathematical equations were established for the attributes so that objective, quantitative measurements of fruit shape could be conducted. The controlled vocabulary and equations were integrated into a newly developed software application, Tomato Analyzer, which conducts semiautomatic phenotypic measurements. To demonstrate the utility of Tomato Analyzer in the detection of shape variation, fruit from two  $F_2$  populations of tomato (*Solanum* spp.) were analyzed. Principal components analysis was used to identify the traits that best described shape variation within as well as between the two populations. The three principal components were analyzed as traits, and several significant quantitative trait loci (QTL) were identified in both populations. The usefulness and flexibility of the software was further demonstrated by analyzing the distal fruit end angle of fruit at various user-defined settings. Results of the QTL analyses indicated that significance levels of detected QTL were greatly improved by selecting the setting that maximized phenotypic variation in a given population. Tomato Analyzer was also applied to conduct phenotypic analyses of fruit from several other species, demonstrating that many of the algorithms developed for tomato could be readily applied to other plants. The controlled vocabulary, algorithms, and software application presented herein will provide plant scientists with novel tools to consistently, accurately, and efficiently describe two-dimensional fruit shapes.

Domestication of plant species was accompanied by profound changes in overall plant and organ morphology (Smith, 1997; Frary and Doganlar, 2003; Paris et al., 2003; Doebley, 2004). Domesticated lines of crops that were selected for improved fruit characters typically carry fruit that are more variable in shape, size, and color than their wild relatives (Grandillo et al., 1999; Paris et al., 2003). We are interested in understanding the genetic and molecular mechanisms that contribute to this variation in morphology with a specific focus on tomato (*Solanum lycopersicum*) fruit shape. Large-scale gene expression studies conducted on developing tomato fruit (Alba et al., 2005; Lemaire-Chamley et al., 2005) in addition to genetic analyses will permit the identification of genes controlling fruit

ontogeny and provide insight into developmental pathways regulating fruit formation. Furthermore, the identification of genes underlying traits characteristic of domesticated varieties may reveal patterns of selection. Tomato is an excellent model for fruit development and domestication studies owing to the tremendous genetic and genomic resources available for this species (Mueller et al., 2005). International efforts are under way to determine the sequence of the euchromatic portion of its genome, further improving the genomic toolbox for tomato.

Tomato fruit is classified according to 10 shape categories such as rounded, high rounded, ellipsoid, or pyriform (International Plant Genetic Resources Institute, 1996; International Union for the Protection of New Varieties and Plants, 2001). Additionally, the distal end of the fruit is categorized as indented, flat, or pointed, whereas the proximal end of the fruit is categorized as flat or indented (International Union for the Protection of New Varieties and Plants, 2001; International Plant Genetic Resources Institute, 1996). While these classifications are useful to group tomato varieties and describe cultivars, the classification scheme cannot be utilized to conduct precise quantitative measurements in a reliable and systematic manner. In addition, the terminology of fruit shape attributes is

<sup>1</sup> This work was supported by the National Science Foundation (grant no. DBI 0227541).

\* Corresponding author; e-mail vanderknaap.1@osu.edu; fax 330-263-3887.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors ([www.plantphysiol.org](http://www.plantphysiol.org)) is: Esther van der Knaap ([vanderknaap.1@osu.edu](mailto:vanderknaap.1@osu.edu)).

<sup>[W]</sup> The online version of this article contains Web-only data.

[www.plantphysiol.org/cgi/doi/10.1104/pp.106.077867](http://www.plantphysiol.org/cgi/doi/10.1104/pp.106.077867).

not described in sufficient detail and tends to be taxon specific. While the taxon-specific terminology may not pose a problem for intraspecies comparisons, cross-species comparisons may be hampered by the lack of agreed-upon terms describing common attributes. Thus, querying information across databases and facilitating retrieval would require development of a structured controlled vocabulary that is arranged in ontologies. Trait ontology provides a structured framework to describe and quantify plant phenotypes (Bruskiewich et al., 2002; Yamazaki and Jaiswal, 2005). Objective phenotypic evaluation can then be conducted for comparative purposes and genetic analysis of quantitative trait loci (QTL). Therefore, a first step toward a comprehensive fruit trait ontology database requires development of agreed-upon terms and associated descriptions for each attribute or trait.

To date, most phenotypic analyses consist of time-consuming manual measurements and subjective scoring of traits that limit the detection of underlying genes. Software-aided measurements of attributes such as height, width, area, and perimeter are most commonly conducted with ImageJ, a public domain program developed at the U.S. National Institutes of Health (program created by W.S. Rasband; <http://rsb.info.nih.gov/ij/>). ImageJ is a versatile program that allows the user to make minor adjustments to the image. However, objective measurements such as height and width are neither automated nor exported efficiently, and extensive and detailed phenotypic analyses that describe subtle differences in shape such as degree of circular shape and the slope along the boundary of the fruit require development of novel algorithms for the attributes. In addition, many of the descriptors necessary to characterize shape cannot be scored objectively without proper software tools. For example, the degree of distal end indentation would be difficult to rate on a scale, and the scoring of this trait would be inconsistent between years, plots, and persons. Moreover, precise phenotypic measurements are necessary to sufficiently characterize loci and the underlying genes that contribute to shape variation. Thus, an accurate and objective method for conducting phenotypic analyses combined with a concise and detailed set of descriptors and terms for fruit shape attributes is necessary.

In previous work, algorithms for some of the shape attributes were defined and shown to have a genetic basis (Van der Knaap and Tanksley, 2003). Here, additional trait terms and mathematical descriptors of shape attributes were developed to further improve phenotypic analyses. The terms were proposed for broader acceptance in the community and future incorporation into a trait ontology database. The mathematical descriptors were implemented in a software application, Tomato Analyzer. This program performs semiautomatic, objective, and quantitative measurements of attributes that will accelerate phenotypic analyses and eliminate subjective scoring of many fruit shape traits. Consequently, Tomato Analyzer can

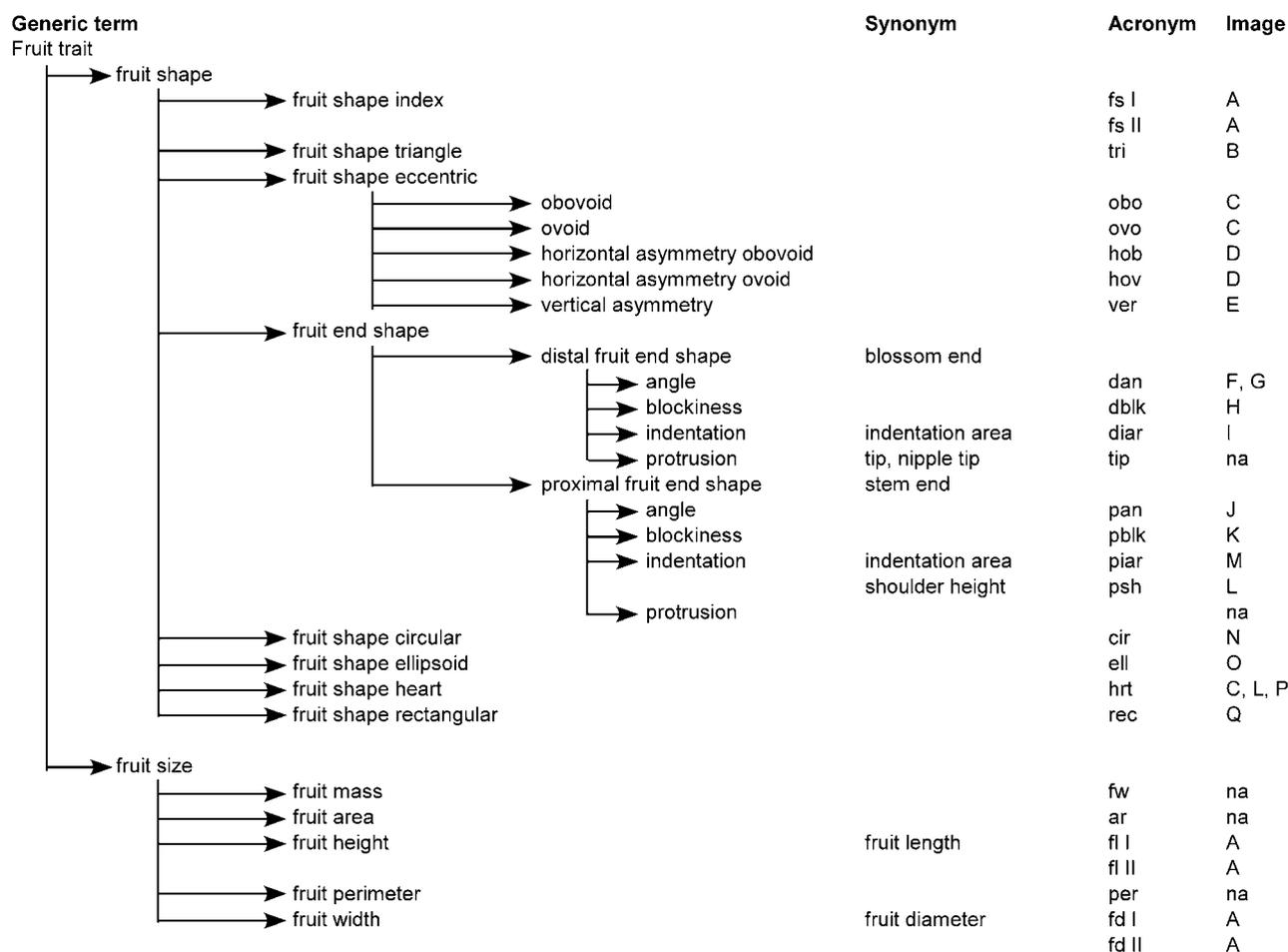
provide more accurate, consistent, and objective results and measure shape attributes that are impossible or impractical to determine manually. To validate that Tomato Analyzer functions properly and to show its use for fruit other than tomato, we used the software to collect phenotypic data on tomato fruit from segregating  $F_2$  populations as well as on a sample of fruit from other species.

## RESULTS

### Trait Ontology Terms and Mathematical Descriptors

An accurate description of fruit shape requires the development of a common vocabulary that encompasses a range of plant species. The structure of the terms needs to follow the True Path Rule that states that the pathway from a child term all the way up to the parent term must be accurate (Bruskiewich et al., 2002). As an example, Figure 1 displays the child term distal end fruit shape angle as an instance of distal end fruit shape. In turn, distal end fruit shape is an instance of fruit end shape and follows the path fruit trait, fruit shape, and fruit end shape. The trait terms and path relationships that were developed are listed in Figure 1. Instances of fruit trait are fruit shape and fruit size. Fruit shape is a parent term of fruit shape index, fruit shape triangle, fruit shape eccentric, and fruit end shape. Fruit shape eccentric and fruit end shape are parent terms of additional child terms (Fig. 1). The remaining fruit shape terms, fruit shape circular, ellipsoid, heart, and rectangular, describe how well the fruit surface depicts a circle, ellipse, etc. and determine the uniformity and homogeneity of objects. Instances of fruit size are fruit mass, fruit area, fruit height, fruit perimeter, and fruit width (Fig. 1). Tomato-specific synonyms for some of the terms are listed as well as the acronyms for each term. The acronyms will be used for detailed QTL analyses pertaining to fruit shape.

The development of a definition of each term and an associated mathematical descriptor would permit objective measurements of fruit shape attributes. Several terms and equations were adopted from previous published work on tomato fruit shape. These included fruit shape index, blockiness, and fruit shape triangle (Van der Knaap and Tanksley, 2003), fruit height, and fruit width (Lippman and Tanksley, 2001). Most acronyms applied in previous studies also remained the same. These included fs for fruit shape index, fl for fruit height, and fd for fruit width (Grandillo et al., 1999; Lippman and Tanksley, 2001; Van der Knaap and Tanksley, 2003). Terms and acronyms for distal and proximal end blockiness, as well as triangle (dblk, pblk, and tri), were renamed based on previously used blossom and stem end blockiness and heart shape terms, respectively (Van der Knaap and Tanksley, 2003). The remaining acronyms were newly developed for attributes first described and measured here. The definitions and mathematical descriptors for each trait



**Figure 1.** Trait ontology terms. Generic terms and their relationship as instance of other terms are indicated by the arrows. The synonyms are tomato-specific terms of the generic terms. The acronyms are used in QTL analyses. The Image column lists the Figure 2 section that displays how the mathematical descriptor for each term or acronym is measured. na, Not applicable.

term are presented below. A single equation was developed for each term or acronym.

#### Fruit Shape Index

Shape index is defined as the ratio of height to width (Fig. 2A). The software contains two acronyms for fruit shape index. The first acronym, fs I, is the ratio of maximum height to maximum width ( $H/W$ ). The second acronym, fs II, is the ratio of the height at mid width to the width at mid height ( $H_m/W_m$ ; Fig. 2A). For both descriptors, a value greater than 1 indicates an elongated fruit, equal to 1 indicates a round fruit, and less than 1 indicates a squat fruit. Typically, the results of these two measurements are very similar and the user can select which one describes the shape of a particular object most appropriately.

#### Fruit Shape Triangle

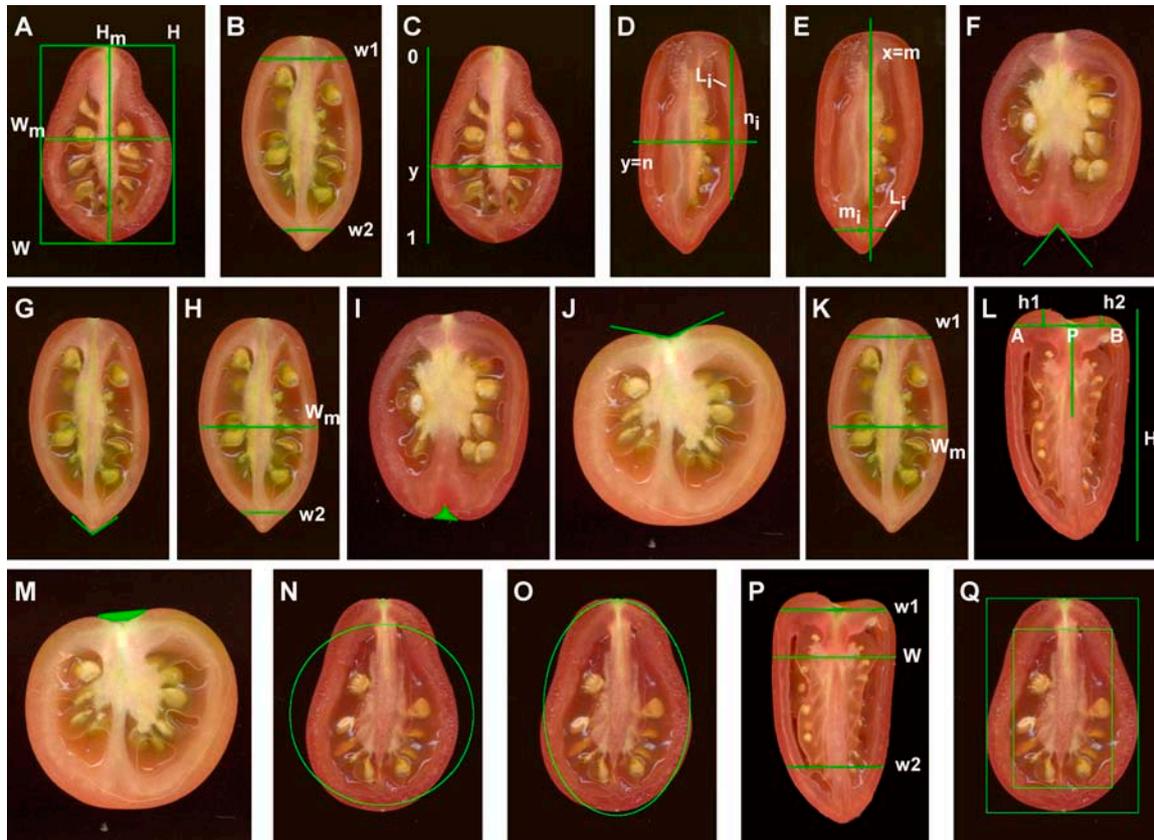
The software defines triangle as the ratio of the proximal end width to distal end width,  $w_1/w_2$  (Fig. 2B). The width is measured at user-defined distances

from the proximal end. A fruit shape triangle value greater than 1 indicates that the proximal end of the fruit is wider than the distal end of the fruit, while a value less than 1 indicates that the distal end of the fruit is wider.

#### Fruit Shape Eccentric

The ovoid and obovoid functions describe how top or bottom heavy a fruit is, respectively. Thus, the fruit displayed in Figure 2C is considered obovoid shaped. The degree of ovoid or obovoid is calculated by identifying the position of the widest section,  $y$ , of the fruit (Fig. 2C). The following formula is then used to describe obovoid:  $4 \times (y - 0.5)$  if  $y > 0.5$ ; 0 otherwise, indicating the fruit is not obovoid. The following formula describes ovoid:  $-4 \times (y - 0.5)$  if  $y < 0.5$ ; 0 otherwise, indicating the fruit is not ovoid.

Horizontal asymmetry and vertical asymmetry describe how asymmetric a fruit is when divided along a horizontal or vertical axis, respectively. The horizontal or vertical axes that divide the fruit are termed  $n$  and  $m$ , respectively (Fig. 2, D and E). The position of the



**Figure 2.** Descriptors of fruit morphology traits. A, Fruit shape index I: ratio of maximum height to width,  $H/W$ . Fruit shape index II: ratio of mid height to mid width,  $H_m/W_m$ . B, Fruit shape triangle: the ratio of proximal width to distal width,  $w_1/w_2$ . C, Fruit shape eccentric obovoid: the position of the widest width of the fruit,  $4 \times (y - 0.5)$  if  $y > 0.5$ ; 0 otherwise. D, Fruit shape eccentric horizontal asymmetry,  $(\sum |n - n_i|)/\text{number of columns } L$ . E, Fruit shape eccentric vertical asymmetry,  $(\sum |m - m_i|)/\text{number of rows } L$ . F, Distal fruit end shape angle at position 5% above the tip from the fruit. G, Distal fruit end shape angle at position 5% above the tip from the fruit. H, Distal fruit end blockiness: ratio of fruit width at the distal end to mid width,  $w_2/W_m$ . I, Distal end indentation area relative to total fruit area. J, Proximal fruit end shape angle. K, Proximal fruit end blockiness: ratio of fruit width at the proximal end to mid width,  $w_1/W_m$ . L, Proximal fruit end indentation: shoulder height,  $(h_1 + h_2)/2H$ . M, Proximal fruit end indentation: area, indentation area relative to total fruit area. N, Fruit shape circular, fitting precision  $R^2$ . O, Fruit shape ellipsoid, fitting precision  $R^2$ . P, Fruit shape heart: taperness function,  $1 - w_2/W + w_1/W$ .  $w_2$  = average width below widest width  $W$ ;  $w_1$  = average width above widest width  $W$ . Q, Fruit shape rectangular: the ratio of maximum area inscribing the rectangle to the minimum area of the enclosing rectangle,  $S_{in}/S_{out}$ .

horizontal axis,  $n$ , is determined by finding the topmost and bottommost points of the fruit and dividing by two to find the center. Likewise, the position of the vertical axis,  $m$ , is determined by finding the leftmost and rightmost points of the fruit and dividing in half to find the center. To compute horizontal asymmetry, each column of pixels, termed  $L_{i_v}$  is determined, and the midpoint of the column,  $n_{i_v}$  is found (Fig. 2D). Next, the difference between  $n_{i_v}$  and  $n$  is calculated and recorded. Once every column is examined, the sum of the differences is determined and divided by the number columns. Thus, the general formula for horizontal asymmetry is  $(\sum |n - n_i|)/\text{number of columns}$ . For vertical asymmetry, each row of pixels, termed  $L_{i_v}$  is determined, and the midpoint of the column,  $m_{i_v}$  is found (Fig. 2E). Next, the difference between  $m_{i_v}$  and  $m$  is calculated and recorded. Once every row is examined, the sum of the differences is determined and

divided by the number rows. Thus, the formula for vertical asymmetry is  $(\sum |m - m_i|)/\text{number of rows}$ . Vertical and horizontal asymmetry values of 0 signify a perfectly symmetric shape. Horizontal asymmetry ovoid is defined by the general formula for horizontal asymmetry if there is more area above the horizontal axis  $n$  than below it; otherwise, horizontal asymmetry ovoid equals 0. Similarly, horizontal asymmetry obovoid is defined by the general formula for horizontal asymmetry if there is more area below the horizontal axis  $n$  than above it; otherwise, horizontal asymmetry obovoid equals 0.

#### Distal Fruit End Shape

The angle of the distal fruit tip refers to the intersection of two lines where the slope is measured via regression along the boundary of the fruit on both

sides at a user-defined distance from the distal end of the fruit (Fig. 2, F and G). The slope is determined by the regression measured at  $\pm 5\%$  when the user-defined positions are between 5% and 50% from the distal end (macro setting), whereas the slope is determined by the regression measured at  $\pm 2\%$  when the user-defined positions are between 2% and 10% from the distal end (micro setting). The angle is measured at the point where the lines intersect and is expressed in degrees, where  $180^\circ$  is flat, greater than  $180^\circ$  is indented (Fig. 2F), and less than  $180^\circ$  is pointed (Fig. 2G). The term blockiness is referred to as squared or box-like shapes. Blockiness is calculated as the ratio of the width at a user-selected proportion of the height closest to the distal end of the fruit to the mid width,  $w_2/W_m$  (Fig. 2H). Indentation refers to the area of indentation at the distal end of the fruit (Fig. 2I). The distal indentation area is determined by finding the most indented distal end position, the lowest distal end on both sides of the indented position, and by the boundary of the fruit along the indented area. The computational method is described in more detail below (shoulder height; Fig. 2L). The distal end indentation area is the ratio of indentation area to total fruit area.

#### *Proximal Fruit End Shape*

The angle of the proximal fruit end refers to the angle from the shoulder points to the site of pedicel attachment or the proximal end, where  $180^\circ$  is flat and greater than  $180^\circ$  is concave (Fig. 2J). Blockiness is calculated as the ratio of the width at a user-selected proportion of the height from the top of the fruit to the mid width,  $w_1/W_m$  (Fig. 2K). Indentation is measured by two methods, either as the "shoulder height, psh" (Fig. 2L) or as the "indentation area, piar" (Fig. 2M). Shoulder height is calculated by first locating the most indented point ( $P$ ) at the proximal end. A straight vertical line from point  $P$  to the center of gravity is drawn. Next, a line perpendicular to the vertical line is drawn through the boundary, selecting the intersection points  $A$  and  $B$ . Finally, shoulder height points  $h_1$  and  $h_2$  are defined to be the maximal distance from the arc  $A-P$  and  $B-P$  to the line  $AB$ , respectively. Shoulder height is defined as  $(h_1 + h_2)/2H$  (Fig. 2L). The larger the shoulder height value is, the more indented the fruit at the proximal end. The indentation area is calculated by finding the area between the two shoulder points, the lowest position  $P$ , and the boundary. The "proximal end indentation area, piar" is the ratio of the indentation area to the total fruit area.

#### *Fruit Shapes Circular, Ellipsoid, Heart, and Rectangular*

These functions are related to homogeneity and uniformity, that is, similarity of the object to the common shapes circle, ellipse, heart, and rectangle. The method developed for circular and ellipsoid determines the fitting precision  $R^2$ . This value represents the coefficient

of determination and reflects how well the actual shape fits a circle or ellipse based on regression (Fig. 2, N and O). The closer the value to 1, the more similar the fruit is to a circle or ellipse. Heart shape is a function of three object characteristics: the location of the maximum width  $y$  (Fig. 2C), the shoulder height (Fig. 2L), and the taperness (Fig. 2P). To calculate heart, the location of the maximum width is described as  $1 - y$ , where  $y$  is the widest point. The shoulder height is described as  $(h_1 + h_2)/2H$  (see above). The taperness is described as  $1 - w_2/W + w_1/W$ .  $w_1$  and  $w_2$  are the average width above and below the widest width, respectively. The weight of these individual components is expressed as  $0.25 \times [(1 - y) \times (1 - w_2/W + w_1/W)] + 20 \times [(h_1 + h_2)/2H]$  and is returned as the heart shape value. Rectangular is calculated as the ratio of  $S_{in}/S_{out}$  where  $S_{out}$  is the minimum area of an enclosing rectangle and  $S_{in}$  is the maximum area of the inscribing rectangle (Fig. 2Q). Thus, the closer the value is to 1, the more rectangular the shape of the object.

#### *Fruit Size*

Fruit size measurements calculated by Tomato Analyzer include area, width, height, and perimeter. For width, two descriptors are available, the widest width (fd I, largest horizontal cross section) and the width measured at the midpoint of the height (fd II; Fig. 2A). There are also two descriptors for height available: the highest height (fl I, largest vertical cross section) and the height measured at the midpoint of the width (fl II; Fig. 2A).

#### *Tomato Analyzer Application*

The Tomato Analyzer application requires digital images of cut fruit saved in jpeg format. When loaded into the application, the entire image appears in the left viewing window (Fig. 3). The software is designed to recognize objects of a certain size and image resolution, measured in dots (pixels) per inch (dpi). Generally, the smaller the object, the higher the resolution required to provide accurate analysis. The implementation of the equations with the software relies on obtaining the  $x$  and  $y$  coordinates of a pixel in a jpeg image of the fruit objects. The software automatically determines the boundaries of fruit in a scanned image. The object boundary is extracted through contour tracing, which results in a list of adjacent points describing the border of an object in an image. All phenotypic measurements are calculated based on the boundaries.

Prior to phenotypic analysis, if fruit are positioned at an angle or if an attached object distorts the boundary, the position of individual fruit can be manually adjusted using the software. Occasionally, the distal and proximal ends of the fruit are not correctly identified, resulting in aberrant angle and indentation values. Therefore, Tomato Analyzer also contains a

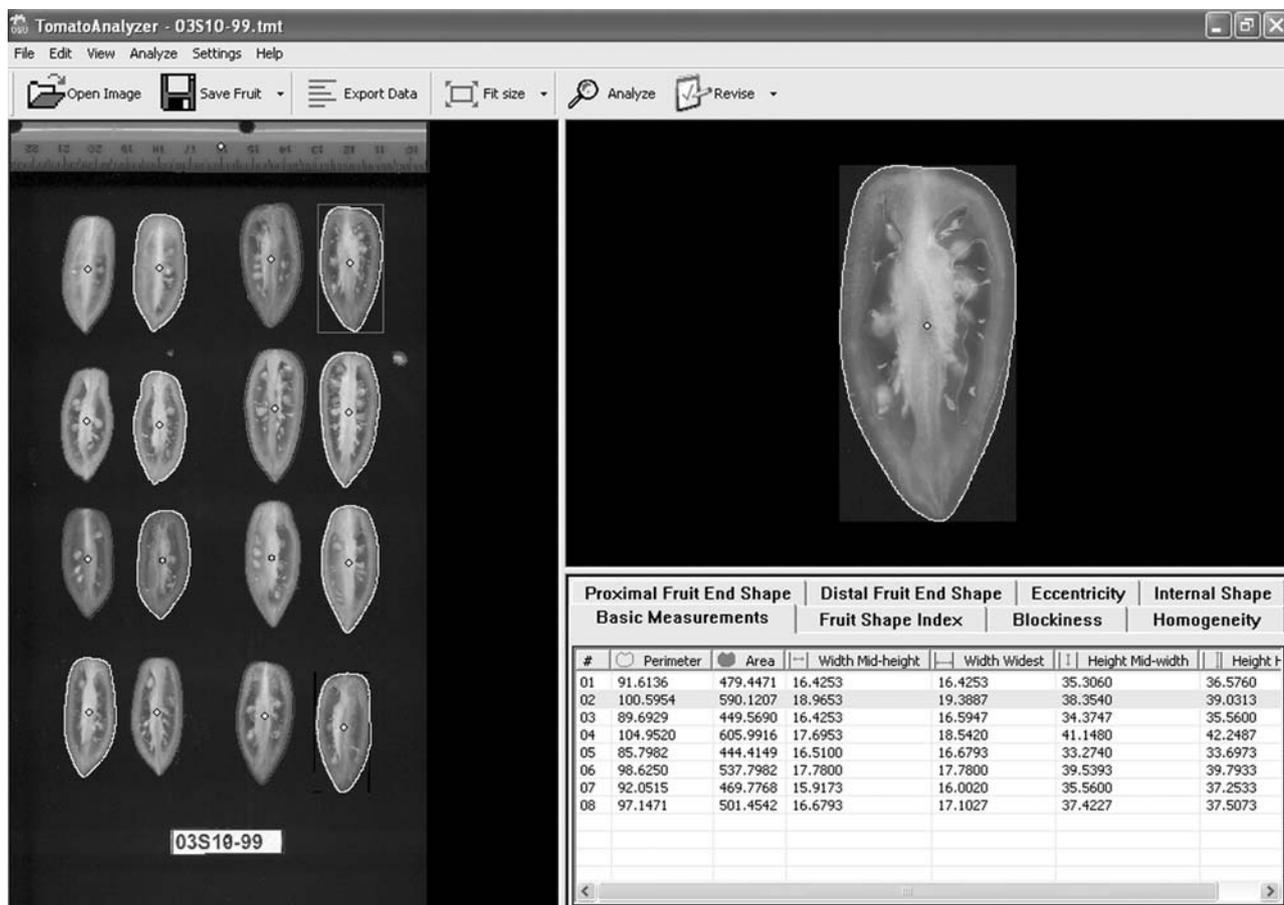


Figure 3. Tomato Analyzer application.

function to manually adjust the distal and proximal end points of the fruit. In addition, objects can be de-selected or selected.

The units used for the attribute values can be selected as pixels, centimeters, millimeters, or inches. As most of the measurements are ratios, selecting the appropriate dpi setting consistent with the resolution of original jpeg image is only required for the size measurements including perimeter, area, height, and width.

The measurements saved setting allows the user to select which attribute values to compute for display and export. Individual attributes or an entire measurement cluster can be selected or deselected. Shape attributes are divided into several clusters in Tomato Analyzer application and largely follow the grouping listed in Figure 1. Basic Measurements group comprises all the fruit size traits; Fruit Shape Index comprises fruit shape index I and II; Homogeneity comprises circular, ellipsoid, and rectangular; Distal Fruit End Shape comprises macro and micro angle, indentation, and protrusion; Proximal Fruit End Shape comprises angle, shoulder height, and indentation area; Eccentricity comprises all eccentricity attributes in addition to fruit shape heart; and Blockiness comprises distal and proximal fruit end shape blockiness and

fruit shape triangle. By selecting the corresponding group tab in the lower right corner of Tomato Analyzer, results for the selected group are displayed (Fig. 3).

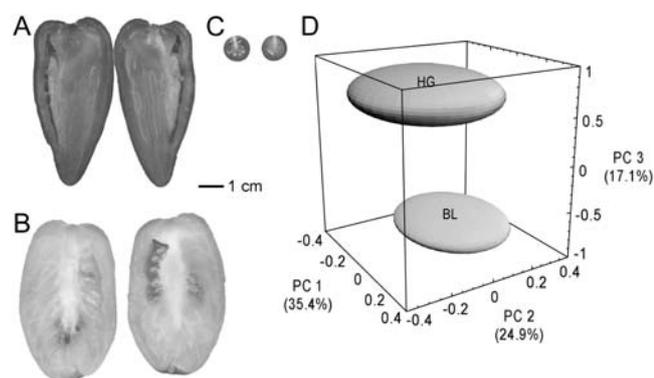
Some attributes allow the user to select settings that maximize phenotypic diversity. User-defined settings are offered for upper and lower blockiness positions. The upper position is used for calculating proximal end blockiness. The lower position is used to calculate distal end blockiness. In addition, these new values will affect fruit shape triangle. The values for the blockiness positions equal the percentage of the height from the top of the fruit. These values can be changed to any number as long as they are both between 0 and 1 and lower position is greater than upper position. There are two settings to calculate distal end angles, referred to as macro and micro. The setting for macro level determines the percentage of the perimeter from the bottom where the angle will be measured, ranging from 5% to 50%. The micro level setting determines where the proximal angle is measured, ranging from 2% to 10% from the tip of the fruit.

The save function allows the user to save the manual adjustments and analyzed fruit shape attributes. Subsequently, when a user opens the original image file, the saved file will be opened and will display all of the adjustments.

There are two methods to export data obtained by Tomato Analyzer. In the first, one image is analyzed and the data for the shape attributes of individual fruit is exported to a .csv file suitable for loading into a spreadsheet or statistical analysis package. The second method is called batch mode and allows more than one image to be loaded and analyzed. In this scenario, the data for each attribute is exported to a .csv file as an average of all fruit in a single image. This method of export is most useful when conducting QTL analyses.

#### Phenotypic Analyses Conducted by Tomato Analyzer

To validate the accuracy and demonstrate the utility of the Tomato Analyzer software, phenotypic and genetic analyses were conducted on two  $F_2$  populations derived from crosses between one of the extreme-shaped *S. lycopersicum* cultivars, Howard German or Banana Legs, and a small, round, wild relative, *Solanum pimpinellifolium* LA1589 (Fig. 4, A–C). Phenotypic data were collected for fruit from both populations using Tomato Analyzer. Principal components analysis (PCA) was conducted to determine the major sources of variation among the morphological traits within and between these populations (Fig. 4D). The variation within each of the two  $F_2$  populations could be explained by the first three principal components (PC), which combined represented 77.4% of the total variability (Fig. 4D). In addition, PC 3 demonstrated significant differences between the Howard German and Banana Legs  $F_2$  populations, whereas PC 1 and PC 2 were not significantly different (Fig. 4D; Supplemental Table I). PC 1, representing 35.4% of the variation, was predominantly affected by attributes that evaluate the tapered shape of fruit, including the traits fruit shape triangle, proximal end blockiness, and horizontal asym-



**Figure 4.** Phenotypic variation within the Howard German and Banana Legs populations. The fruit images in A, B, and C are depicted to scale. A, Image of *S. lycopersicum* cv Howard German. B, Image of *S. lycopersicum* cv Banana Legs. C, Image of *S. pimpinellifolium* accession LA1589. D, Graphical display of PC 1, 2, and 3 from analysis of fruit morphology traits in the Howard German (HG)  $\times$  LA1589, and Banana Legs (BL)  $\times$  LA1589  $F_2$  populations. Clouds represent the means and  $\pm$  SE of the first three PCs. The variation described by each of the PCs is listed in parentheses.

metry ovoid (loading values for contributing traits are listed in Supplemental Table II). PC 2, representing 24.9% of the variation, was primarily controlled by traits that contribute to fruit elongation such as fruit shape index, proximal end shape features, and distal end angle, as well as the fruit shape uniformity and homogeneity features circular, rectangular, and heart shape (Supplemental Table II). PC 3, representing 17.1% of the variation, was influenced by fruit size, homogeneity, and eccentricity characteristics, as well as distal end blockiness and indentation area (Supplemental Table II).

To identify regions of the genome responsible for the observed fruit morphology variation, QTL analysis was performed using the first three PCs as traits in the Banana Legs and Howard German  $F_2$  populations. A genetic map was constructed with molecular markers using MAPMAKER (Lander et al., 1987), and subsequent QTL analyses of each of the three PCs were conducted using QTL Cartographer (created in 2001 by S. Wang, C.J. Basten, and Z.B. Zeng [Department of Statistics, North Carolina State University]). In both populations, PC 1 was controlled by similarly located QTL on chromosomes 2, 3, and 7 (Table I). A QTL on chromosome 1 also controlled PC 1, but only in the Banana Legs population. PC 2 was controlled by a highly significant QTL on chromosome 7 in both populations. In addition, a QTL on chromosome 9 was present for PC 2 in the Banana Legs population. PC 3 differentiated the two populations, and showed one common QTL on chromosome 10. However, a QTL on the top of chromosome 7 controlled PC 3 only in the Howard German population, and a QTL on chromosome 11 affected PC 3 in only the Banana Legs population. These different QTL may underlie the subtle differences in fruit shape of the parental plants (Fig. 4, A and B). The PC 2 QTL located on chromosome 7 coincides with the *sun* locus that has been shown to control fruit shape index (Van der Knaap and Tanksley, 2001; Van der Knaap et al., 2004).

The following analysis was conducted with the attribute distal fruit end shape angle to demonstrate the flexibility of Tomato Analyzer. The user can select the location of the slope for the distal end angle measurement. Mapping of distal end angle at various distances from the tip of the fruit, from 2% to 20%, showed that the most significant QTL is associated with the angle measurement at 20% above the tip and that this QTL decreases in significance with angle measurements taken at positions closer to the tip of the fruit (Fig. 5). A similar trend was noted in the Howard German population at the same chromosomal location (data not shown). The results of the distal end angle measurements demonstrate that user-defined settings of Tomato Analyzer application permit phenotypic analyses that are optimized for each population or tailored to specific research questions. In addition, adjustment of settings is efficiently applied to all fruit in a population, while such a change would be labor intensive if manual measurements were taken.

**Table 1.** Regions of the genome responsible for fruit shape variation in two  $F_2$  populations as detected by PCA and subsequent QTL analysis

Population	QTL <sup>a</sup>	LOD	Most Significant Marker <sup>b</sup>	A <sup>c</sup>	D <sup>d</sup>	R <sup>2</sup>
Howard German	PC1.2	6.38	TG337	2.02*	-1.92*	0.19
	PC1.3	4.64	TG242	1.08	1.40	0.14
	PC1.7	4.31	CD57	2.01*	0.41	0.14
	PC2.7	22.01	COS103	-4.02*	0.33*	0.52
	PC3.7	7.09	TG342	3.55*	-3.31*	0.25
	PC3.10	4.10	CT234	1.37*	-0.25	0.12
Banana Legs	PC1.1	4.95	CT191	1.47*	-0.47	0.11
	PC1.2	8.89	TG337	2.16*	-0.95	0.22
	PC1.3	6.47	TG246	1.60*	-0.78	0.17
	PC1.7	7.86	COS103	2.61*	-0.34	0.18
	PC2.7	12.02	COS103	-2.47*	-0.11	0.30
	PC2.9	6.41	TG551	-1.54*	-0.05	0.16
	PC3.10	4.16	CT234	0.46	-1.07*	0.11
	PC3.11	4.74	TG546	1.10*	0.03	0.18

<sup>a</sup>QTL acronym reflects the PC for which it was detected (first number) and the chromosome where it was located (second number). <sup>b</sup>The map location of these markers can be found on the Solanaceae Genomics Network Web site (<http://www.sgn.cornell.edu>).

<sup>c</sup>An asterisk indicates a significant additive effect. A negative value indicates that an increase in the value of the attribute is due to *S. pimpinellifolium* allele, and a positive value indicates that an increase in the value of the trait is due to *S. lycopersicum* allele. <sup>d</sup>An asterisk indicates a significant dominant effect. A negative value indicates that the *S. pimpinellifolium* allele is dominant and a positive value indicates that the *S. lycopersicum* allele is dominant.

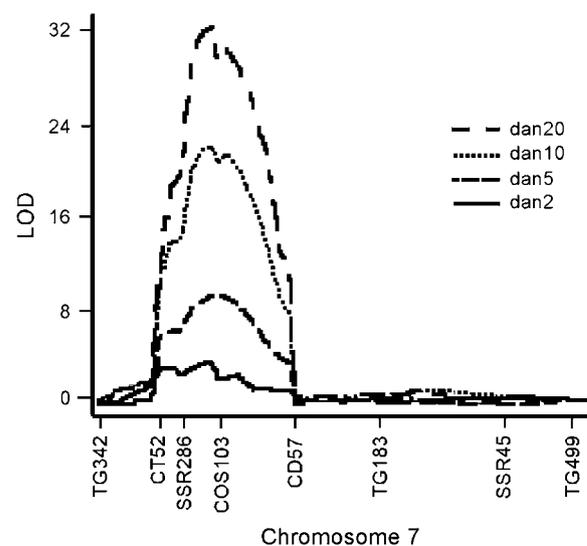
To demonstrate a wider utility of Tomato Analyzer in shape analysis, we evaluated its performance on fruit of different species (Fig. 6). The software could accurately determine the boundaries of fruit from as small as grape (*Vitis vinifera*) to as large as butternut squash (*Cucurbita moschata*) and of fruit of different colors. The output values provided by Tomato Analyzer were consistent with visual observations and manual measurements (Table II). For example, the software accurately measured distal end angle values from extremely pointed fruit, like the peppers (*Capsicum annuum*), to very rounded fruit, like the grape and Bartlett pear (*Pyrus communis*). In addition, the butternut and yellow squash (*Cucurbita pepo*) and pear had obovoid values, indicating that the largest width of the fruit was well below the midpoint in those fruit. Lastly, triangle shape at the 10% setting indicated that the peppers were the most triangular in shape, while butternut squash and Bartlett pear were the least triangular in shape. Thus, the Tomato Analyzer application is not limited to tomato fruit but can be applied to fruit morphology analyses of other plants.

## DISCUSSION

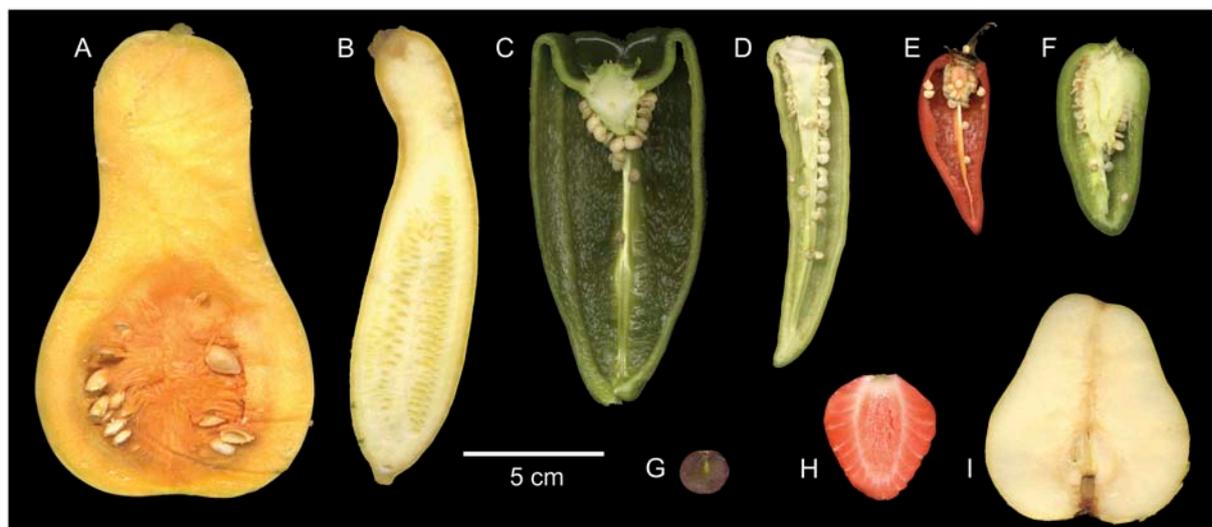
The development of structured, controlled vocabularies arranged in ontologies would provide great benefit to plant scientists (Bruskiewich et al., 2002).

Currently, extensive trait ontologies are being developed for rice (*Oryza sativa*; Bruskiewich et al., 2002; Yamazaki and Jaiswal, 2005) and maize (*Zea mays*; Vincent et al., 2003). However, these species are monocotyledonous, and thus the present ontologies do not incorporate traits common to fruit typical of dicot species. Here, controlled vocabulary terms were developed to describe fruit shape traits. The terms must be defined, well structured, and biologically accurate to assist in information retrieval and provide meaningful comparative studies in synteny and homology (Bruskiewich et al., 2002; Vincent et al., 2003). The controlled vocabulary terms presented here were defined to facilitate consistent use within and across taxa. Additionally, mathematical descriptors were developed, which make the definitions even more robust. These terms are presented for acceptance by the plant scientific community and eventual incorporation into a trait ontology.

Software programs and computational methods have been developed to categorize and classify plant organ shapes such as fruit (Morimoto et al., 2000; Beyer et al., 2002) and seeds (Sako et al., 2001a, 2001b). Many of these applications were developed to describe shape uniformity and overall quality of fruit and seed lots. Despite the usefulness of categorizing objects for quality purposes, these methods could neither measure specific shape features, such as the degree of indentation of the proximal fruit end, nor be applied to quantitative genetic analyses. Recently, a method was employed to describe leaf shape by calculating the distance between coordinates along the leaf surface and subjecting the measurements to PCA to determine the greatest sources of variation (Langlade et al., 2005).



**Figure 5.** LOD curves for distal fruit end angle measurements on chromosome 7. Graphical depiction of the LOD curves for distal fruit end angle measurements at the 2%, 5%, 10%, and 20% above the distal tip of the fruit. The results were derived from the Banana Legs  $F_2$  population. The x axis displays the molecular markers along the chromosome; the y axis displays the LOD score.



**Figure 6.** Variety of fruit that were subjected to phenotypic evaluations by Tomato Analyzer. A, Butternut squash. B, Yellow squash. C, Large jalapeno. D, Banana pepper. E, Chili pepper. F, Jalapeno. G, Grape. H, Strawberry (*Fragaria* spp.). I, Bartlett pear.

In addition, the PCs were considered as traits in QTL analyses, and loci were identified contributing to the variation in leaf morphology (Langlade et al., 2005). The described method identifies overall changes in leaf shape but cannot identify individual features of shape, for instance the angle along the boundary or the position of the widest width of the leaf. Consequently, it is difficult to discern which locus controls a specific shape feature and vice versa which attribute is controlled by a specific locus. Tomato Analyzer, on the other hand, analyzes specific shape traits, which were combined and subjected to PCA and subsequent QTL analyses. Individual shape attributes that most significantly contribute to each PC and underlying QTL could be discerned (Supplemental Table II). One prime example of coincidence of individual trait QTL and PC QTL is offered by PC 2 and the attributes that most significantly contribute to this component. The major

PC 2 QTL is located on chromosome 7 near marker *cos103*. This result suggests that the attributes comprising PC 2 are controlled, at least in part, by the same locus. Indeed, when we map the individual traits such as fruit shape index, distal fruit end shape angles, and the proximal fruit end shape features, which comprise PC 2, we identify QTL for these traits at a similar location on chromosome 7 (Fig. 5; data not shown). In addition, these QTL coincide with the *sun* locus that is known from our previous studies to control fruit shape index (Van der Knaap and Tanksley, 2001; Van der Knaap et al., 2004). Another interesting finding is that PC 3 is controlled in part by traits affecting fruit size (Supplemental Table II). The major size QTL are typically located on chromosomes 1, 2, 3, 7, and 11 (Lippman and Tanksley, 2001; Van der Knaap and Tanksley, 2003), whereas the QTL controlling PC 3 in both populations is located on chromosome 10 and not

**Table II.** Output from Tomato Analyzer for select phenotypic traits on a variety of fruit

Images of these fruit appear in Figure 6.

Fruit	Fruit Shape Index I <sup>a</sup>	Area	Distal End Angle, 5% <sup>b</sup>	Distal End Angle, 20% <sup>b</sup>	Obovoid <sup>c</sup>	Triangle, 10% <sup>d</sup>
		cm <sup>2</sup>	Degrees	Degrees		
A, butternut squash	1.67	122.8	190	80	0.99	0.60
B, yellow squash	3.55	53.0	72	22	0.78	0.99
C, large jalapeno	2.04	70.5	83	48	0	2.21
D, banana pepper	4.00	25.1	28	12	0	2.04
E, chili pepper	2.51	12.0	47	17	0	2.75
F, jalapeno	2.04	18.5	98	35	0	1.71
G, grape	0.98	2.4	168	103	0	1.28
H, strawberry	1.13	13.9	158	87	0	1.43
I, Bartlett pear	1.12	46.6	228	125	0.73	0.59

<sup>a</sup>Fruit shape index I was measured as the ratio of the maximum height of the fruit to the maximum width. <sup>b</sup>Distal end angle was measured as the slope of two lines that were drawn at 5% or 20% distance, respectively, along boundary from the tip of the fruit. <sup>c</sup>Obovoid is a measure of pear shape. <sup>d</sup>Triangle, was measured as the ratio of the width 10% of the height from the proximal end of the fruit to the width 10% from the distal end of the fruit.

known to play a major role in controlling fruit size in tomato. This result suggests that employing PCA on several traits may lead to identification of novel QTL and/or improve significance of minor QTL.

In all, Tomato Analyzer is a flexible and comprehensive application that provides intuitive descriptors and output that facilitate the analysis of fruit morphology. Furthermore, our efforts to combine controlled vocabulary with mathematical descriptors into one software application make this a very useful tool for several applications. Tomato Analyzer allows for accurate and objective measurements of fruit shape attributes in a high-throughput manner and of traits that are nearly impossible to quantify manually. The application is specifically developed to analyze fruit shape QTL in tomato but could readily be applied to fruit of other species and other plant organs such as seed and leaves.

## MATERIALS AND METHODS

### Software Implementation

The software application has been implemented in C++ using Visual Studio 6.0 and runs on the Windows operating system. The image processing library Computer Vision and Image Processing (CVIP) 3.7c is used for image I/O. Modifications to the software were done using Visual Studio 2003 with source code control provided by SourceSafe. The source code cross indexer LXR was used to create an online indexable and searchable version of the software code. The program is free for academic purposes and can be downloaded from our laboratory Web site: <http://www.oardc.ohio-state.edu/vanderknaap/>.

### Plant Material

Two F<sub>2</sub> populations were constructed from crosses between one of two *Solanum lycopersicum* cultivars (Banana Legs or Howard German) and a wild species, *Solanum pimpinellifolium* accession LA1589. The Banana Legs population consisted of 99 plants, whereas the Howard German population contained 130 plants. Both populations were grown simultaneously in the same greenhouse in the summer of 2003. Up to eight fruits were harvested from each plant. Fruit were weighed, cut longitudinally, and scanned at 300 dpi resolution. The images were saved as jpeg image files for phenotypic analyses. Each image contained fruit from only one plant. If eight fruit per plant were harvested at once, these fruits were scanned and saved as one image. If fewer than eight fruit per plant were harvested at once and additional fruits were harvested later, both images were combined using Adobe Photoshop 7.0 (Adobe Systems) prior to analysis with Tomato Analyzer.

### Phenotypic Analyses

Tomato Analyzer was used to conduct the phenotypic analyses. Manual adjustments, including modification of fruit boundaries and rotation of individual fruits, were made to the images, if necessary. If boundaries were distorted, for example by an attached seed, they would be modified using the software. If fruit were scanned at an angle, manual rotation was required to properly align them. Adjusted images and associated results were saved as a separate file with the extension .tmt. Batch analyses allowed analysis and exportation of selected shape attribute data for at least 100 images at once. The values for each attribute were averaged for all fruit per image and exported into .csv file.

### Genotypic and Statistical Analysis

Total genomic DNA was isolated from young leaves as described by Bernatzky and Tanksley (1986) and Fulton et al. (1995). The genetic map was

constructed with a combination of RFLP and PCR-based markers. Additional information on RFLP and PCR-based markers, including map location and primer information, can be found on the Solanaceae Genomics Network Web site (<http://www.sgn.cornell.edu>). In addition, marker cos103 is the same marker as Lp103E7-R in Van der Knaap et al. (2004). A molecular linkage map was constructed with MAPMAKER v2.0 and the Kosambi mapping function (Kosambi, 1944; Lander et al., 1987).

PCA and analysis of variance were conducted with SAS V8 (SAS Institute). QTL analysis was performed by composite interval mapping (Zeng, 1993, 1994) using model six with five marker cofactors selected by forward regression and a 10-cm window size, as implemented in Windows QTL Cartographer v2.0. Log of the odds (LOD) scores greater than 4.0 were considered significant. Additive, dominance (d) effects and percent of phenotypic variance explained by the QTL ( $R^2$ ) were estimated with Windows QTL Cartographer at highest probability peaks.

## ACKNOWLEDGMENTS

Drs. R. Pratt and D.M. Francis are acknowledged for their critical reading of the manuscript, and Dr. P. Jaiswal for insightful comments and suggestions on the trait ontology terms. Also, we acknowledge J. Moysenko for excellent plant care and assistance in genotyping, and B. Stricker and R. Drushal for their contributions on Tomato Analyzer.

Received January 25, 2006; revised March 14, 2006; accepted March 15, 2006; published May 11, 2006.

## LITERATURE CITED

- Alba R, Payton P, Fei Z, McQuinn R, Debbie P, Martin GB, Tanksley SD, Giovannoni JJ (2005) Transcriptome and selected metabolite analyses reveal multiple points of ethylene control during tomato fruit development. *Plant Cell* **17**: 2954–2965
- Bernatzky R, Tanksley SD (1986) Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. *Genetics* **112**: 887–898
- Beyer M, Hahn R, Peschel S, Harz M, Knoche M (2002) Analyzing fruit shape in cherry (*Prunus avium* L.). *Sci Hortic (Amsterdam)* **96**: 139–150
- Bruskiewich R, Coe EH, Jaiswal P, McCouch S, Polacco M, Stein L, Vincent L, Ware D (2002) The Plant Ontology Consortium and plant ontologies. *Comp Funct Genomics* **3**: 137–142
- Doebley J (2004) The genetics of maize evolution. *Annu Rev Genet* **38**: 37–59
- Frary A, Doganlar S (2003) Comparative genetics of crop plant domestication and evolution. *Turk J Agric For* **27**: 59–69
- Fulton TM, Chunwongse J, Tanksley SD (1995) Microprep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Mol Biol Rep* **13**: 207–209
- Grandillo S, Ku HM, Tanksley SD (1999) Identifying the loci responsible for natural variation in fruit size and shape in tomato. *Theor Appl Genet* **99**: 978–987
- International Plant Genetic Resources Institute (1996) Descriptors for Tomato (*Lycopersicon spp.*). International Plant Genetic Resources Institute, Rome
- International Union for the Protection of New Varieties and Plants (2001) Guidelines for the Conduct of Tests for Distinctness, Homogeneity and Stability (Tomato). International Union for the Protection of New Varieties and Plants, Geneva
- Kosambi DD (1944) The estimation of map distances from recombination values. *Ann Eugen* **12**: 172–175
- Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**: 174–181
- Langlade NB, Feng X, Dransfield T, Copsey L, Hanna A, Thebaud C, Bangham A, Hudson A, Coen E (2005) Evolution through genetically controlled allometry space. *Proc Natl Acad Sci USA* **102**: 10221–10226
- Lemaire-Chamley M, Petit J, Garcia V, Just D, Baldet P, Germain V, Fagard M, Mouassite M, Cheniclet C, Rothan C (2005) Changes in

- transcriptional profiles are associated with early fruit tissue specialization in tomato. *Plant Physiol* **139**: 750–769
- Lippman Z, Tanksley SD** (2001) Dissecting the genetic pathway to extreme fruit size in tomato using a cross between the small-fruited wild species *Lycopersicon pimpinellifolium* and *L. esculentum* var. Giant Heirloom. *Genetics* **158**: 413–422
- Morimoto T, Takeuchi T, Miyata H, Hashimoto Y** (2000) Pattern recognition of fruit shape based on the concept of chaos and neural networks. *Comp Electron Agric* **26**: 171–186
- Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y, et al** (2005) The SOL genomics network: a comparative resource for Solanaceae biology and beyond. *Plant Physiol* **138**: 1310–1317
- Paris HS, Yonash N, Portnoy V, Mozes-Daube N, Tzuri G, Katzir N** (2003) Assessment of genetic relationships in *Cucurbita pepo* (Cucurbitaceae) using DNA markers. *Theor Appl Genet* **106**: 971–978
- Sako Y, McDonald MB, Fujimura K, Evans AF, Bennett MA** (2001a) A system for automated seed vigour assessment. *Seed Sci Technol* **29**: 625–636
- Sako Y, Regnier EE, Daoust T, Fujimura K, Harrison SK, McDonald MB** (2001b) Computer image analysis and classification of giant ragweed seeds. *Weed Sci* **49**: 738–745
- Smith BD** (1997) The initial domestication of *Cucurbita pepo* in the Americas 10,000 years ago. *Science* **276**: 932–934
- Van der Knaap E, Sanyal A, Jackson S, Tanksley SD** (2004) High-resolution fine mapping and fluorescence *in situ* hybridization analysis of *sun*, a locus controlling tomato fruit shape, reveals a region of the tomato genome prone to DNA rearrangements. *Genetics* **168**: 2127–2140
- Van der Knaap E, Tanksley SD** (2001) Identification and characterization of a novel locus controlling early fruit development in tomato. *Theor Appl Genet* **103**: 353–358
- Van der Knaap E, Tanksley SD** (2003) The making of a bell pepper-shaped tomato fruit: identification of loci controlling fruit morphology in Yellow Stuffer tomato. *Theor Appl Genet* **107**: 139–147
- Vincent PLD, Coe EH, Polacco ML** (2003) *Zea mays* ontology: a database of international terms. *Trends Plant Sci* **8**: 517–520
- Yamazaki Y, Jaiswal P** (2005) Biological ontologies in rice databases: an introduction to the activities in Gramene and Oryzabase. *Plant Cell Physiol* **46**: 63–68
- Zeng ZB** (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc Natl Acad Sci USA* **90**: 10972–10976
- Zeng ZB** (1994) Precision mapping of quantitative trait loci. *Genetics* **136**: 1457–1468