

Distribution of *SUN*, *OVATE*, *LC*, and *FAS* in the Tomato Germplasm and the Relationship to Fruit Shape Diversity^{1[C][W][OA]}

Gustavo R. Rodríguez, Stéphane Muñoz, Claire Anderson, Sung-Chur Sim, Andrew Michel, Mathilde Causse, Brian B. McSpadden Gardener, David Francis, and Esther van der Knaap*

Department of Horticulture and Crop Science (G.R.R., C.A., S.-C.S., D.F., E.v.d.K.), Department of Entomology (A.M.), and Department of Plant Pathology (B.B.M.G.), Ohio State University/Ohio Agricultural Research and Development Center, Wooster, Ohio 44691; and Institut National de la Recherche Agronomique, UR1052 Unité de Génétique et d'Amélioration des Fruits et Légumes, Montfavet 84 143, France (S.M., M.C.)

Phenotypic diversity within cultivated tomato (*Solanum lycopersicum*) is particularly evident for fruit shape and size. Four genes that control tomato fruit shape have been cloned. *SUN* and *OVATE* control elongated shape whereas *FASCIATED* (*FAS*) and *LOCULE NUMBER* (*LC*) control fruit locule number and flat shape. We investigated the distribution of the fruit shape alleles in the tomato germplasm and evaluated their contribution to morphology in a diverse collection of 368 predominantly tomato and tomato var. *cerasiforme* accessions. Fruits were visually classified into eight shape categories that were supported by objective measurements obtained from image analysis using the Tomato Analyzer software. The allele distribution of *SUN*, *OVATE*, *LC*, and *FAS* in all accessions was strongly associated with fruit shape classification. We also genotyped 116 representative accessions with additional 25 markers distributed evenly across the genome. Through a model-based clustering we demonstrated that shape categories, germplasm classes, and the shape genes were nonrandomly distributed among five genetic clusters ($P < 0.001$), implying that selection for fruit shape genes was critical to subpopulation differentiation within cultivated tomato. Our data suggested that the *LC*, *FAS*, and *SUN* mutations arose in the same ancestral population while the *OVATE* mutation arose in a separate lineage. Furthermore, *LC*, *OVATE*, and *FAS* mutations may have arisen prior to domestication or early during the selection of cultivated tomato whereas the *SUN* mutation appeared to be a postdomestication event arising in Europe.

Tomato (*Solanum* section *Lycopersicon*) is native to western South America, from Ecuador and Peru to Bolivia and northern Chile. Cultivated tomato (*Solanum lycopersicum*) is postulated to have been domesticated in Mexico (Jenkins, 1948) with Peru suggested as an alternative location (De Candolle, 1886). The fruit of tomato var. *cerasiforme*, also known as cherry tomato, are typically larger than fruit of the wild species but smaller than those of cultivated tomato. Wild cherry tomato is hypothesized to be a direct progenitor of cultivated tomato (Rick and Holle, 1990). However,

others consider var. *cerasiforme* a revertant from cultivation (i.e. feral plants) or a possible hybrid between wild and weedy taxa (Peralta et al., 2008). Indeed, previous studies have shown that most accessions of tomato var. *cerasiforme* are more closely related to cultivated tomato than to wild relatives and others that are an admixture between tomato and the wild relative *Solanum pimpinellifolium* possibly resulting from the frequent hybridizations between them (Nesbitt and Tanksley, 2002; Ranc et al., 2008).

Some of the most important changes that occurred during the domestication and improvement of tomato were increased fruit weight and the emergence of variable fruit shapes and colors (Paran and van der Knaap, 2007). Tomato varieties have been classified based on fruit morphology into shape categories described by the International Union for the Protection of New Varieties of Plants (UPOV) and the International Plant Genetic Resources Institute (IPGRI; IPGRI, 1996; UPOV, 2001). In addition to the fruit shape categories, tomatoes have also been categorized into germplasm classes based on geographic origin and/or age. However, definitions for germplasm classes are neither accepted by all nor clearly delineated, with categories partially overlapping. Tomatoes classified as regional or landraces are farmer or gardener selected and are

¹ This work was supported by the National Science Foundation (grant nos. DBI 0227541 [to E.v.d.K.] and IOS 0922661 [to E.v.d.K. and B.B.M.G.]) and by the U.S. Department of Agriculture (grant no. 2004–35300–14651 to E.v.d.K. and D.F.).

* Corresponding author; e-mail vanderknaap.1@osu.edu.

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) is: Esther van der Knaap (vanderknaap.1@osu.edu).

^[C] Some figures in this article are displayed in color online but in black and white in the print edition.

^[W] The online version of this article contains Web-only data.

^[OA] Open Access articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.110.167577

adapted to the local environment typically in areas of local subsistence (Male, 1999). The term vintage and contemporary (modern) tomato refers to tomato accessions released before (vintage) or after (contemporary) a certain year (Williams and Clair, 1993; Park et al., 2004; Sim et al., 2009). Breeding and elite breeding lines are used in current breeding programs that seek to develop commercially competitive varieties. U. S. heirloom tomatoes comprise a diverse and loosely defined group. Heirlooms have been referred to as accessions handed down from generation to generation, old commercial varieties, contemporary varieties created to fill niche markets, those of mysterious origin, or treasured (Male, 1999). Many heirloom varieties were brought to North America by European settlers and therefore could be considered as regional accessions or landraces in the country from where they originated.

Fruit shape and locule number are quantitatively inherited characters with estimates of quantitative trait loci (QTL) number ranging from four to 17 with the major loci explaining 19% to 79% of the genetic variation (Barrero and Tanksley, 2004; Brewer et al., 2007; Gonzalo and van der Knaap, 2008). The *sun* and *ovate* loci control fruit elongation and the underlying genes are known. *SUN* encodes a protein that is a positive regulator of growth resulting in elongated fruit and is hypothesized to alter hormone or secondary metabolite levels (Xiao et al., 2008). The mutation is the result of a gene duplication event that was mediated by the retrotransposon *Rider* (Xiao et al., 2008; Jiang et al., 2009). *OVATE* encodes a negative regulator of growth, presumably by acting as a repressor of transcription and thereby reducing fruit length (Liu et al., 2002; Hackbusch et al., 2005; Wang et al., 2007). The *OVATE* allele that conditions an elongated fruit carries a premature stop codon and is presumed to be a null allele (Liu et al., 2002). Locule number, which has a pleiotropic effect on fruit shape and size, is controlled by the *fasciated* (*fas*) and *locule number* (*lc*) loci. *FAS* encodes a YABBY transcription factor and down-regulation of the gene is caused by a large insertion in the first intron (estimated to be 6–8 kb), resulting in fruits with high locule number (Cong et al., 2008). The molecular nature of *LC* was recently identified (S. Muñoz, N. Ranc, E. Botton, A. Berard, S. Rolland, P. Duffe, Y. Carretero, M. Le Paslier, C. Delalande, M. Bouzayen, D. Brunel, and M. Causse, unpublished data). Two single-nucleotide polymorphisms (SNPs) were found to be critical in controlling the locule number phenotype and were located approximately 1,200-bp downstream of the stop codon of a gene encoding a WUSCHEL homeodomain protein, members of which regulate stem cell fate in plants (Mayer et al., 1998).

The goals of this study were to determine whether allelic distribution of *SUN*, *OVATE*, *LC*, and *FAS* was associated with fruit shape, genetic background, as well as geographical and historical origin in a diverse collection of cultivated accessions. This information

would offer important insights into the number of genes involved and their effect on fruit shape in the tomato germplasm. Moreover, the knowledge of the distribution of the fruit shape gene alleles would allow us to examine the molecular events that accompanied domestication and selection of this important crop. We showed that the diversity in tomato fruit morphology was explained to a large extent by mutations in the *SUN*, *OVATE*, *LC*, and *FAS* genes. We analyzed the genetic clustering of this dataset relative to fruit shape category and germplasm class, and demonstrated nonrandom distribution of the major fruit shape alleles. Moreover, our data suggested that *FAS* and *SUN* arose in the *LC* mutant background. *OVATE* on the other hand arose in a different ancestral population. Finally, our studies offered valuable insights into the evolution of tomato from a round berry to a fruit with diverse shapes.

RESULTS

Tomato Germplasm and Fruit Shape Categorization

We visually classified 368 tomato accessions according to eight fruit shape categories: flat, round, rectangular, ellipsoid, heart, long, obovoid, and oxheart (Fig. 1; Supplemental Table S1). These accessions represented eight germplasm classes based on geographic origin and/or history (Table I). Some of the fruit shape categories were represented by accessions from all germplasm classes whereas other shape categories contained accessions from only a few classes (Table I). For example, fruit in the ellipsoid category was found in nearly all germplasm classes with the exception of the seven wild accessions, which produced only round fruits. In contrast, long tomatoes were commonly found among U.S. heirloom and regional Spanish accessions and rectangular shape was represented mostly by Italian accessions (Table I).

Phenotypic Diversity Analyzed by Tomato Analyzer

To analyze fruit shape objectively, and determine whether the visual categories were supported by quantitative shape measurements, we obtained the values for 36 fruit attributes using image analysis. A subset of 120 accessions (hereafter called the subcollection) that represented the diversity of the fruit shape and germplasm classes observed in the larger collection of 368 accessions was phenotypically evaluated using Tomato Analyzer (TA; Brewer et al., 2006; Gonzalo et al., 2009). Iterations of linear discriminant analysis (LDA) led to the identification of the seven most important attributes that define the fruit shape categories. The final set of attributes identified for their predictive value were: fruit shape index, distal end protrusion, widest width position, the proximal end blockiness value at 20% from the proximal end, rectangular, distal angle at 20% along the boundary from the tip of the fruit, and proximal eccentricity. The

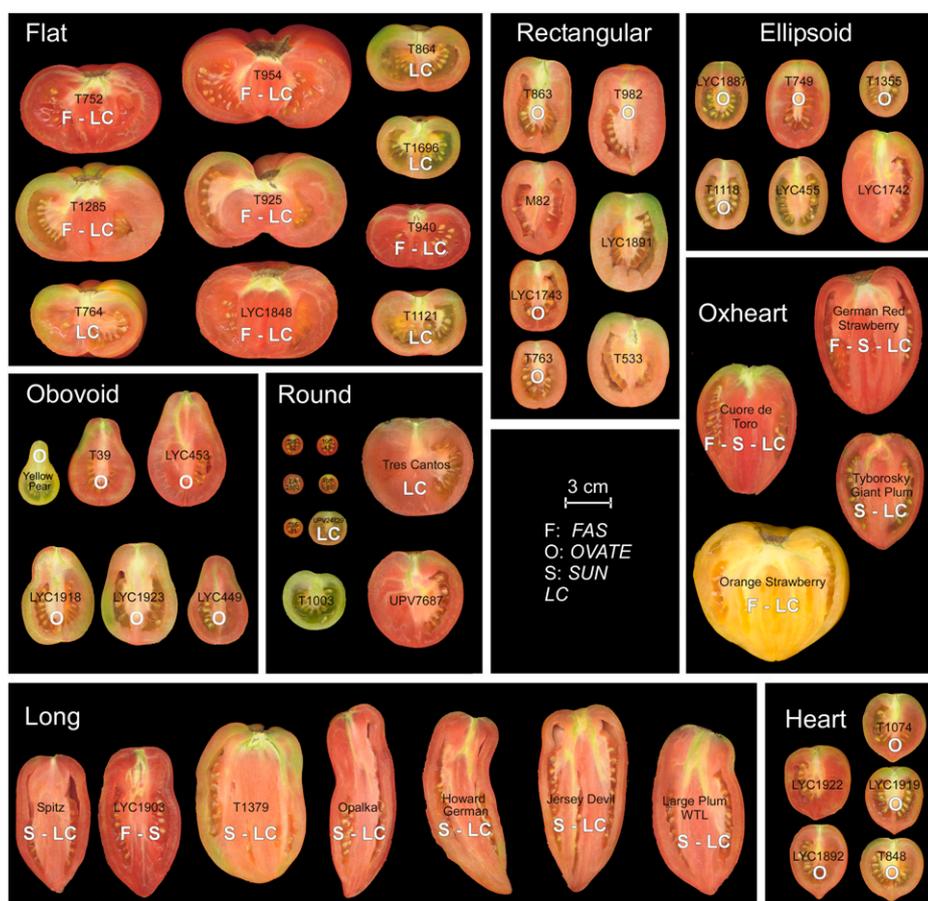


Figure 1. Tomato fruit shape categories adapted from UPOV (2001) and IPGRI (1996). Each fruit is identified by variety name (information available at <http://solgenomics.net/>) and presence of mutation in the *SUN*, *OVATE*, *LC*, and/or *FAS* genes (abbreviated as S, O, LC, and F, respectively).

range of these attributes for each shape category is listed in Supplemental Table S2. Using the measurements of these attributes, 83% of fruit could be accurately classified by a simple linear discriminant function according to the cross-validation test (Table II). The relatively few discrepancies between the visual classification and the objective classification based on TA attributes arose with fruits classified as ellipsoid, obovoid, or rectangular. Each of these three categories had some accessions classified as one or the other. Upon closer inspection, approximately one-third of these were misclassified during the initial visual scoring of the collection. The remaining misclassified accessions were the result of transitioning and overlapping values for the seven attributes in the ellipsoid, obovoid, and rectangular shape categories (Supplemental Table S2). Regardless, the objective assessment using the seven attributes in TA provided a robust classification of fruit shape categories consistent with subjective international descriptors.

Allelic Distribution of *SUN*, *OVATE*, *LC*, and *FAS* According to Fruit Shape Category and Germplasm Class

We determined the alleles for *SUN*, *OVATE*, *LC*, and *FAS* in the 368 accessions comprising the entire collection. The data showed that all obovoid, and many of the

ellipsoid (83%), rectangular (59%), and heart (48%) tomatoes carried the mutant allele of *OVATE* whereas most of the long (88%) and oxheart (83%) tomatoes carried the mutant allele of *SUN* (Table I). The most frequent mutation in flat tomatoes was *LC* (82%) followed by *FAS* (28%). Many of the long tomatoes also carried the mutation in the *LC* gene (63%), a finding that was supported by genetic evidence for the *lc* QTL in a population that segregated for elongated fruit shape (Gonzalo et al., 2009). All oxheart tomatoes carried the *LC* mutation in addition to *SUN* and/or *FAS*. Most round tomatoes carried the wild-type allele at the four shape loci, with the *LC* mutation most prevalent at 33%. The majority of the round tomatoes with the *LC* mutation were tomato var. *cerasiforme* lines (Supplemental Table S1). To evaluate whether fruit shape category and shape gene mutations were correlated to one another, we conducted a χ^2 test. The test corroborated the lack of independence between fruit shape categories and alleles of *SUN*, *OVATE*, *LC*, and *FAS* ($\chi^2 = 790$, degrees of freedom = 84, $P < 0.0001$), indicating that the tomato shape mutations have a major impact on fruit morphology. These results were further supported by a quantitative association of fruit shape alleles with specific shape attributes (Supplemental Table S3).

The mutant alleles of *LC* and *OVATE* were well represented in all germplasm classes with the excep-

Table I. Morphological diversity and allelic distribution of *SUN*, *OVATE*, *FAS*, and *LC* genes in tomato based on fruit shape category and germplasm class

Number of accessions in each category is given, including the percentage from the total (in parentheses).

	Fruit Shape Category								TOTAL	<i>SUN</i> ^a	<i>OVATE</i> ^b	<i>FAS</i> ^c	<i>LC</i> ^d	Wild type ^e
	Flat	Round	Rectangular	Ellipsoid	Heart	Long	Obovoid	Oxheart						
Heirloom	7	3	1	10	1	9	8	8	47	19 (40)	17 (36)	10 (21)	22 (47)	2 (4)
Latin American														
Regional	6	1	0	8	1	0	4	0	20	0	12 (60)	2 (10)	8 (40)	0
<i>Cerasiforme</i>	19	20	0	6	0	0	1	0	46	0	6 (13)	4 (9)	26 (57)	12 (26)
Wild	0	7	0	0	0	0	0	0	7	0	0	0	0	7 (100)
Regional Spanish	2	2	1	4	0	8	4	1	22	9 (41)	8 (36)	3 (14)	12 (55)	1 (5)
Regional Italian	71	7	23	25	16	6	12	2	162	5 (3)	61 (38)	30 (19)	61 (38)	35 (22)
Contemporary														
Italian	9	1	7	9	3	1	7	1	38	1 (3)	18 (47)	2 (5)	12 (32)	8 (21)
American	8	0	1	3	1	0	0	0	13	1 (8)	2 (15)	1 (8)	7 (54)	3 (23)
Other	0	2	1	5	1	0	4	0	13	3 (23)	7 (54)	0	2 (15)	2 (15)
TOTAL	122	43	34	70	23	24	40	12	368					
<i>SUN</i> ^a	0	0	0	5 (7)	0	21 (88)	1 (3)	10 (83)	37 (10)					
<i>OVATE</i> ^b	0	2 (5)	20 (59)	58 (83)	11 (48)	5 (21)	40 (100)	0	136 (37)					
<i>FAS</i> ^c	34 (28)	1 (2)	0	2 (3)	0	2 (8)	2 (5)	10 (83)	51 (14)					
<i>LC</i> ^d	100 (82)	14 (33)	5 (14)	2 (3)	2 (9)	15 (63)	0	12 (100)	150 (41)					
Wild type ^e	16 (13)	25 (58)	11 (32)	8 (11)	10 (44)	0	0	0	70 (19)					

^aAccessions that carry the *SUN* gene duplication. ^bAccessions that carry the premature stop codon mutation in the *OVATE* gene. ^cAccessions that carry the large insertion mutation in the *FAS* gene. ^dAccessions that carry the mutations near the *LC* gene. ^eAccessions that carry the wild-type alleles at four loci.

tion of the wild species (Table I). The *FAS* mutation was often present in U.S. heirloom and regional Italian collection whereas the *SUN* mutation was often present in U.S. heirloom and regional Spanish accessions. The most common mutant alleles in the Italian collection were for *OVATE* and *LC*. The *FAS* mutation was more common in the regional than in the contemporary Italian collection. There were very few Italian accessions that carried the mutation in the *SUN* gene (Table I). We also investigated the distribution of alleles in the Latin American germplasm. As expected, only wild-type alleles for the fruit shape genes were found in the wild species (Table I). Moreover, none of the Latin American regional and *cerasiforme* accessions carried the mutation in *SUN*, suggesting that the gene duplication resulting in elongated fruit shape did not occur in South or Central America. In contrast, the mutations in *OVATE*, *LC*, and *FAS* were found in Latin American regional and *cerasiforme* accessions. The most common mutant allele was for *LC*, followed by *OVATE* and *FAS* (Table I). Two *cerasiforme* accessions (LA1655 and VIR739) carried both the *LC* and *FAS* mutations and exhibited flat fruits and higher locule number (6.8 and 7.5, respectively) than the *LC* mutation alone (3.4 locules) in this background (Supplemental Table S1).

Genetic Analysis of the Subcollection

To evaluate genetic structure in relation to germplasm and morphological categories, we genotyped 120 accessions with an additional 25 loci that were

randomly chosen based on their distribution across the genome. With the exception of two *cerasiforme* accessions, these 118 represented cultivated tomato. After the genotypic evaluations, we removed four accessions because they were identical to another accession in the subcollection. The number of alleles per locus ranged from two to nine for 10 single sequence repeat (SSR) markers with a mean of 5.5 (Supplemental Table S4). The remaining markers had two alleles per locus with the exception of the markers SP and LeOH16.2 that had three. The resulting dataset of 114 cultivated and two *cerasiforme* accessions and 29 markers (including the four fruit shape genes) was analyzed with the STRUCTURE 2.2 software (Pritchard et al., 2000). We tested population structure for $K = 1$ to 15 and determined that the best number of clusters is 5 (Evanno et al., 2005; Supplemental Figs. S1 and S2). We tested the consistency among five different runs at $K = 5$ after which we determined the ranking of inferred ancestry among accessions within each cluster as well as the stability of accessions in the same cluster (Fig. 2). This analysis indicated that none of the accessions changed from one to another cluster and that the grouping of the accessions was robust. The analysis was repeated with identical settings for burn in and iterations but only including the 25 randomly chosen markers and omitting the four fruit shape genes. This analysis did not define a best K for the number of clusters (Supplemental Fig. S1). Therefore, the alleles of *SUN*, *OVATE*, *LC*, and *FAS* appeared to be essential in determining genetic structure for this dataset, suggesting that selection for fruit shape was

Table II. Summary of classification with cross validation from LDA for fruit shape categories and seven fruit shape attributes measured by TA

No. (N) of Accessions	Fruit Shape Category ^a								Total
	Ellipsoid	Flat	Heart	Long	Obovoid	Oxheart	Rectangular	Round	
N	25	36	9	11	13	5	9	12	120
N correct	19	32	8	9	10	4	6	12	99
Proportion	0.76	0.89	0.89	0.82	0.77	0.800	0.67	1.00	0.83

^aFruit shape category for each accession was visually defined (see “Materials and Methods” and Supplemental Table S1). Attributes included in LDA: fruit shape index, distal end protrusion, widest width position, the proximal end blockiness at 20% of the height from the top of the fruit, rectangular, distal angle at 20% along the boundaries from the tip of the fruit, and proximal eccentricity. A description of the attributes is given in Brewer et al. (2006) and Gonzalo et al. (2009).

responsible for differentiating statistically distinct subpopulations within the collection of 116 cultivated accessions.

Pairwise F_{ST} and Nei's Genetic Distances of the STRUCTURE Clusters

To test the significance of the genetic clusters at $K = 5$, we conducted pairwise $F_{ST}(\theta)$; Weir and Cockerham, 1984) analyses. We found that the STRUCTURE clusters were significantly different from one another, strongly supporting the genetic grouping (Supplemental Table S5). Based on Nei's genetic distances (Nei, 1978), one of the most distinct STRUCTURE groups was represented by cluster 5, comprised of contemporary accessions, sharing the fewest common alleles with other clusters. This finding is not surprising because recent introgression of disease resistance genes from wild relatives have enhanced the diversity in this germplasm class (Ruiz et al., 2005; Sim et al., 2009). Clusters 1 and 2 as well as clusters 3 and 4 shared more alleles relative to other pairwise comparisons (Supplemental Table S5).

Nonrandom Distribution of Shape Categories, Germplasm Class, and Alleles of the *SUN*, *OVATE*, *LC*, and *FAS* Genes

Germplasm classes including regional Italian, regional Spanish, and U.S. heirloom lines and fruit shape categories including flat, ellipsoid, and obovoid were associated with genetic clusters at $P < 0.001$ (Supplemental Table S6, A and B). Specifically, heirloom accessions were significantly overrepresented in cluster 3, regional Spanish accessions were significantly overrepresented in cluster 4, whereas regional Italian accessions were significantly overrepresented in cluster 1 ($P < 0.05$; Supplemental Table S6A). In addition, three of four contemporary U.S. lines and four of nine contemporary Italian lines grouped together in cluster 5 (Fig. 2). We did not statistically analyze these two germplasm classes because there were fewer than 10 in each group. Nevertheless, these results suggested that the contemporary accessions, whether American or Italian, carried similar alleles, implying that the accessions were related to one another or that selection

through breeding favored the same allele at multiple loci. With respect to shape category, the flat category was significantly overrepresented in cluster 3 and 4, and the ellipsoid category in cluster 1 and 2. The obovoid category was randomly distributed among the different genetic clusters ($P > 0.05$; Supplemental Table S6B).

The accessions that carry the mutation in the *LC* gene were overrepresented in cluster 3 and 4 (Supplemental Table S6C), which was consistent with the overrepresentation of flat fruit in these clusters (Supplemental Table S6B). *OVATE* was significantly overrepresented in cluster 1 and 2 (Supplemental Table S6C), which was consistent with the overrepresentation of ellipsoid fruit in these clusters (Supplemental Table S6B). The mutations in *FAS* and *SUN* were significantly overrepresented in cluster 3, representing all the oxheart, and most of the flat and long fruit (Fig. 2).

The occurrence of mutant alleles in certain genetic clusters might indicate separate origin of the mutations. The most widespread mutations in the tomato germplasm were for the *OVATE* and *LC* genes. These mutations were found in distinct genetic clusters, indicating separate origin. The data showed that accessions with both *OVATE* and *LC* mutations were indeed rare (Table III). It is possible that the lack of coinheritance was reinforced by repulsion phase linkage of *OVATE* and *LC* on chromosome 2. Accessions that carry both *SUN* and *OVATE* as well as *FAS* and *OVATE* were also found less often than would be expected by chance (Table III). This observation suggested an independent origin for the *OVATE* mutation and genetic isolation of lineages carrying the mutant allele. On the other hand, *SUN*, *LC*, and *FAS* are found more often together in the same accession than expected by chance, indicating that the mutations arose in the same ancestral population (Table III).

Presence of Fruit Shape Genes in Commercially Grown Fresh Market Tomato Varieties

Nearly all tomato accessions evaluated in this study were heirloom, regional, and contemporary accessions with a few exceptions (Supplemental Table S1). As was evident from our analysis, selection for mutant alleles of the four fruit shape genes played a key role in

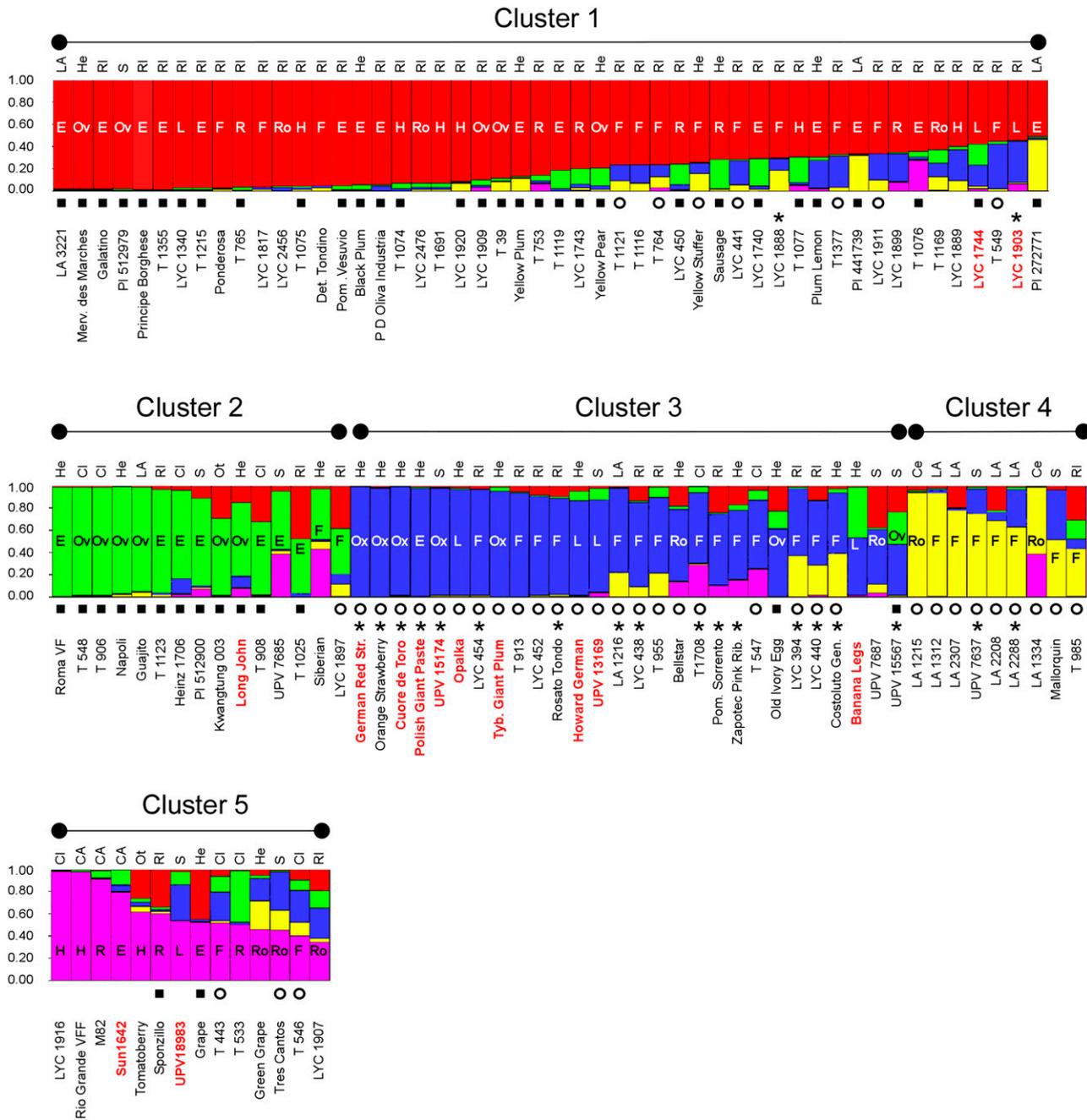


Figure 2. Population structure of 116 tomato accessions using STRUCTURE software and 29 markers (25 randomly distributed markers and *SUN*, *OVATE*, *LC*, and *FAS*). The coefficients of estimated ancestry per accession in each cluster were represented by an individual bar, where each color refers to a distinct cluster. The name of the accession is below the bar whereas the notation above the bar indicates germplasm class. CA, Contemporary U.S.; Ce, *cerasiforme*; RI, regional Italian; CI, contemporary Italian; LA, regional Latin American; S, regional Spanish; He, U.S. Heirloom; and Ot, other. The black squares indicate accessions carrying *OVATE*; the white circles indicate accessions carrying *LC*; the stars indicate accessions carrying *FAS*; and accessions carrying *SUN* are listed in red. The fruit shape category is written on each bar. Ov, Obovoid; R, rectangular; Ro, round; H, heart; Ox, oxheart; L, long; F, flat; and E, ellipsoid.

defining population structure in the subpopulation dataset. To determine the relevance of these mutant alleles in commercial germplasm, we obtained varieties from local retail stores and evaluated them for shape and the alleles of the shape genes. We found that

all the store-bought varieties carried one or more mutant alleles for the four shape genes. For example, ellipsoid-shaped grape tomatoes, which occupy a distinct niche market, carried the null allele of *OVATE*, indicating that the introduction of this group was

Table III. The expected and observed combinations of fruit shape alleles in the 368 tomato accessions

Fruit Shape Gene	N ^a	LC	FAS	SUN
OVATE	136	2 (55) 137.9****	5 (19) 18.7****	5 (14) 9.7**
LC	150		39 (21) 31.3****	27 (15) 17.3****
FAS	51			11 (6.7) 8.7**
SUN	37			

^aRefers to number of accessions that carried the mutant alleles of the gene listed in the first column. χ^2 analyses were conducted to determine whether the observed combination of alleles (number) were higher or lower than expected (number in parentheses). The χ^2 values obtained in each paired analysis of the mutant alleles appears in the second row at ** $P < 0.01$ and **** $P < 0.0001$. The χ^2 values obtained in the analysis of combinations of three or more fruit shape genes were all highly significant.

accompanied by selection for the mutation in this gene. Unexpectedly, while the heirloom Roma carried the *OVATE* mutation (Supplemental Table S1), all three store-bought ellipsoid-shaped Roma-type tomatoes carried the mutation in *SUN* instead. Two of the three store-bought Roma tomatoes also carried the *LC* mutation. Flat-shaped and high-loculed varieties carried the *LC* mutation and one of the three carried the *FAS* mutation as well. Similar to our findings with the older accessions, several round and low-loculed varieties carried *LC*, indicative of modifiers of this mutation. It appears that the *SUN*, *LC*, *FAS*, and *OVATE* mutant alleles are not unique to heirloom, regional, and contemporary accessions but are found in commercially grown varieties sold at retail stores and, thus, are highly relevant today.

DISCUSSION

In contrast to wild relatives that carry round and two-loculed fruit, cultivated tomato fruit is highly diverse in shape. In this study, we demonstrate that the diversity in fruit morphology in the cultivated germplasm is explained to a large extent by mutations in the *SUN*, *OVATE*, *LC*, and/or *FAS* genes. Individually, the alleles of these genes explain as much as 71% of the observed variation for specific fruit shape attributes (Supplemental Table S3). At the same time, it is evident that interactions between genes and uncharacterized modifiers also affect fruit shape. For example, store-bought Roma types and Sun1642 carry the duplication of the *SUN* gene and exhibit an ellipsoid instead of a long-shaped fruit. Sun1642 is a contemporary U.S. accession that clusters genetically with similar accessions such as M82 and Rio Grande, demonstrating that the mutant *SUN* allele was introgressed from another accession, most likely an heirloom. Moreover, differences in fruit shape of varieties carrying the *OVATE*,

FAS, and *LC* mutations also suggest that suppressors and enhancers of these genes are present within the cultivated germplasm. For example, accessions that carry the *OVATE* mutation display a range of fruit shapes from long (e.g. LYC1340) and obovoid (e.g. Yellow Pear) to round (e.g. Gold Ball Livingston) whereas accessions carrying *LC* mutation produce long (e.g. Howard German), oxheart (e.g. Cuore de Toro), round (e.g. LA1215), or flat (e.g. Druzba) fruit. Although Howard German fruit has on average five locules controlled by *LC* (Gonzalo et al., 2009), the effect of *SUN* is dominant over *LC* in controlling overall fruit shape (i.e. long), even though locule number is impacted. When adding the *FAS* mutation to the *SUN-LC* mutant background, the fruit is less elongated and, instead, oxheart in shape. Fruit shape in rectangular, ellipsoid, and heart-shaped varieties that do not carry the *OVATE* or *SUN* mutation might be controlled by genes such as those underlying the shape QTL *fs8.1* and/or *tri2.1/dblk2.1* controlling fruit elongation (Grandillo et al., 1996; Ku et al., 2000; Brewer et al., 2007; Gonzalo and van der Knaap, 2008). However, without the knowledge of the underlying genes at *fs8.1* and *tri2.1/dblk2.1* and the mutations that gave rise to the altered fruit shape phenotype, it is not possible to survey the alleles at these QTL throughout the germplasm. Alternative explanations to describe variation in fruit shape are also plausible. While it is likely there are more than two alleles for each fruit shape gene with the exception of *SUN*, it is not clear whether any of the other alleles would result in the fruit shape changes described for the known mutant alleles. For example, *LC* exhibits multiple alleles but only two SNPs are associated with changes in locule number (S. Muñoz, N. Ranc, E. Botton, A. Berard, S. Rolland, P. Duffe, Y. Carretero, M. Le Paslier, C. Delalande, M. Bouzayen, D. Brunel, and M. Causse, unpublished data). These two SNPs were genotyped in our collection. Limited sequencing of *OVATE* in several accessions also showed there are more than two alleles. However, only one SNP was associated with elongated fruit shape (G.R. Rodríguez and E. van der Knaap, unpublished data). The *SUN* mutation was likely to have occurred recently based on findings presented herein. In fact, there are no nucleotide differences between the ancestral and derived locus with the exception of the template switch that accompanied the transposition event (Xiao et al., 2008). Thus, the existence of other alleles of *SUN* that feature elongated fruit shape is extremely unlikely. A third allele has been reported for *FAS* (Cong et al., 2008), although we did not find this allele in any of our accessions. Because we did not search for additional alleles of *FAS*, the existence of more than two alleles that would result in highly loculed fruit is possible albeit unlikely.

The population structure analysis resulted in the identification of five genetic clusters, some exhibiting significant associations with fruit shape category and germplasm class. The fruit shape alleles of *SUN*,

OVATE, *FAS*, and *LC* are responsible for the observed clustering, indicating that selection for fruit shape is responsible for the underlying genetic structure in tomato. Genetic groupings according to tomato fruit morphology have been reported in other studies (Mazzucato et al., 2008), supporting the notion that selection of diverse fruit shapes played a critical role in tomato domestication. We demonstrate that the fruit shape controlled by *SUN* and *FAS* is significantly overrepresented in STRUCTURE cluster 3, *LC* in clusters 3 and 4, and *OVATE* in clusters 1 and 2 (Fig. 2; Supplemental Table S6C). In addition, the analysis of Nei's genetic distances shows that clusters 1 and 2 share a large number of alleles as do clusters 3 and 4. This observation and the distribution of fruit shape alleles suggest a separate origin of the *OVATE* mutation (Supplemental Table S5). The *OVATE* lineage remained separate from the lineages carrying mutations in the other fruit shape genes during the domestication and selection of tomato, otherwise random interbreeding would have resulted in more accessions carrying mutations in *OVATE* and one of the other genes. However, it is also conceivable that the combination of these two mutations results in seed sterility and/or reduced plant viability that would preclude the formation of lineages that carry both *OVATE* and *LC* mutations.

Evidence about the consumption of tomato before and immediately after the discovery of America by Christopher Columbus is extremely limited. As described in the Florentine Codex writings by de Sahagún who lived between 1499 and 1590, tomatoes were eaten with salt and chile pepper (*Capsicum* spp.; De Sahagún, 1959). Historical evidence demonstrated that tomato arrived from Mexico to Spain and Italy following Columbus' exploration of the Americas. The first written record of tomato in Europe was in 1544 where it was described as flat and segmented fruit (Matthioli, 1544). Other descriptions of fasciated fruit followed soon thereafter (Oellinger, 1553). The fasciated phenotype of those tomatoes suggests that the earliest tomatoes that arrived in Europe carried both *LC* and *FAS* mutations. It was not until 1813 that Dunal documented in *Solanum* the pear-shaped fruit and named it *Lycopersicum piriforme*, the earliest mention of a different shape of tomato (Dunal, 1813). Therefore, it is plausible that tomatoes carrying the *OVATE* mutation traveled to Europe centuries after the first wave of tomatoes that carried the *LC* and *FAS* mutations. The mutant allele subsequently spread in Italy where, at present time, 71 out of 109 elongated accessions carry the *OVATE* mutation (Table I). The latter finding demonstrates that the *OVATE* mutation underlying the classical Italian paste tomato is in fact very widespread in its germplasm. All four tomato fruit shape gene mutations are widespread in the cultivated tomato germplasm, including commercial fresh market varieties sold at present time in grocery stores. This finding clearly shows that the fruit shape gene mutations, whether maintained for curiosity's sake, for

cultural and culinary purposes, or to develop high yielding and uniquely shaped varieties, played key roles in the selection of tomato and that these mutations are still highly relevant today.

Although speculative, it has not escaped our notion that the distribution of the four fruit shape genes in the tomato germplasm enables us to develop a model for the domestication and selection of tomato. Based on the data presented in this work, we hypothesize how tomato evolved from a spherical to a variably shaped fruit, and where and when the fruit shape mutations arose (Fig. 3). Accessions of tomato var. *cerasiforme* represent an admixture of the remnant ancestral gene pool from which tomato may have been domesticated as well as feral derivatives of cultivated varieties (Nesbitt and Tanksley, 2002; Peralta et al., 2008; Ranc et al., 2008). If we were to assume that tomato var. *cerasiforme* is the direct progenitor of cultivated tomato, the *FAS*, *LC*, and *OVATE* mutations would have arisen prior to domestication since the mutant alleles were found in the progenitor's gene pool. On the other hand, if we were to assume that the tomato var.

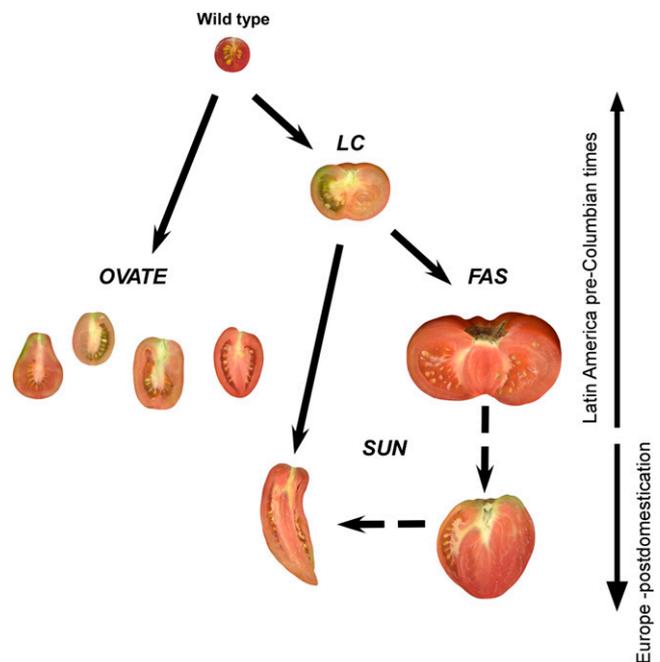


Figure 3. Model of the evolution of fruit shape variation in tomato. The tomato domestication most likely occurred in Latin America starting with round wild-type tomato var. *cerasiforme* accessions. *LC* arose first in the *cerasiforme* background, resulting in round or flat tomatoes. The *OVATE* mutation arose in a different ancestral population than *LC* but also in the *cerasiforme* background, resulting in ellipsoid fruit. *FAS* arose in the *cerasiforme* *LC* background around the same time as *OVATE*, resulting in flat and highly lobed fruit. The presence of *OVATE*, *LC*, and *FAS* mutant alleles in the Latin American germplasm suggest they arose early or prior to domestication of tomato. The *SUN* mutation arose postdomestication in Europe, most likely in the *LC* background of a cultivated tomato and resulting in long fruit. It is also possible that *SUN* arose in the *LC* and *FAS* mutant background, resulting in an oxheart fruit. [See online article for color version of this figure.]

cerasiforme genetic pool is a transition between the wild and the cultivated germplasm, selection of the *LC*, *FAS*, and *OVATE* mutations would be directly involved in the domestication of tomato. Based on the distribution of *LC* in the *cerasiforme* germplasm, where over half of them carry the mutant allele, the mutation probably arose prior to *FAS* and *OVATE* (Fig. 3). The notion that *LC* is an old mutation is supported by data presented by Muños and colleagues (S. Muños, N. Ranc, E. Botton, A. Berard, S. Rolland, P. Duffe, Y. Carretero, M. Le Paslier, C. Delalande, M. Bouzayen, D. Brunel, and M. Causse, unpublished data). However, increased selection for the *LC* mutation and genetic drift could also explain the broad presence of the mutant allele in the progenitor's germplasm. Approximately 12% of the *cerasiforme* accessions carry the mutant alleles of *OVATE* or *FAS*, suggesting that they arose at roughly the same time but in different lineages. It is likely that *FAS* arose in the *LC* background because many accessions that carry *FAS* (80%) also carry *LC* mutations. However, some accessions including two *cerasiforme* lines only carry *FAS*, thus the mutation could have also arisen independently in the wild-type *LC* background.

The *SUN* mutation arose much later than the *OVATE* and *FAS* mutations. None of the *cerasiforme* or regional Latin American accessions carry this mutation and therefore, *SUN* arose most likely postdomestication and in Europe (Fig. 3). The *SUN* mutation is found in approximately half of the U.S. heirloom and Spanish regional accessions with elongated fruit. Many of the U.S. heirloom accessions carrying *SUN* are of northern European origin such as Opalka, Spitz, German Red Strawberry, and Tyboroski Giant Plum (Supplemental Table S1). This result supports the notion that *SUN* arose in Europe and came to North America as an heirloom. It is highly unlikely however, that the *SUN* mutation arose in Italy since only six out of 109 elongated fruit accessions from the Italian collection carry *SUN*.

It is likely that *SUN* arose in the *LC* background in cultivated tomato. In general, the older accessions representing the heirloom and regional accessions carry both *SUN* and *LC* mutations. The exceptions are Sun1642, UPV18983, Long John, Banana Legs, Orange Banana, Tegucigalpa, LYC1903, LYC1744, and T923 (Supplemental Table S1). Sun1642 and UPV18983 are considered contemporary accessions and the result of recent breeding efforts (Fig. 2). Banana Legs is a created heirloom that results from a deliberate cross (Male, 1999), and represents an admixture genotype in our cluster analysis. We also assume the admixture genotype and deliberate crosses that generated Long John, Orange Banana, LYC1903, LYC1744, and T923 accessions (Fig. 2). The latter three are Italian accessions and since the *SUN* mutation is quite rare in this germplasm, it is likely that the mutation was bred from an accession that originated elsewhere. In all, these findings suggest that like *FAS*, *SUN* most likely arose in the *LC* mutant background.

MATERIALS AND METHODS

Plant Material, DNA Extraction, and Fruit Scanning

A total of 368 tomato (*Solanum lycopersicum*) accessions were grown in the field in Wooster, OH in the summers of 2005 to 2007 (Supplemental Table S1). The collection includes U.S. heirloom (47 accessions), contemporary U.S. (13), regional Spanish (22), regional Latin American (20), tomato var. *cerasiforme* (46), and wild (seven) accessions. The Italian germplasm was obtained from two sources (Supplemental Table S1) and was divided into regional (162) and contemporary (38) based on the presence of the uniform ripening locus *u* located on chromosome 10 (Kinzer et al., 1990; Philouze, 1991). Older tomato accessions such as those found in the heirloom and regional categories often carry fruit with green shoulders when unripe, whereas accessions in the contemporary category lack green shoulders and ripen evenly. The seeds were obtained from a variety of sources (Supplemental Table S1 and in the Tomato Cultivars and Heirlooms section found at the Sol Genomics Network [http://solgenomics.net/]). Approximately eight fruit from each plant were cut longitudinally through the center, placed cut-side down on a scanner and digitalized at 100 dots per inch as previously described (Brewer et al., 2006). Total genomic DNA was isolated from young leaves as described previously (Bernatzky and Tanksley, 1986; Fulton et al., 1995).

Fruit Shape Categories

The fruit shape terms and the number of categories in UPOV (UPOV, 2001) and IPGRI (IPGRI, 1996) classification systems are not completely consistent. Moreover, terms from an older version of UPOV (1992) are not the same as those in the most recent version (2001). The UPOV and IPGRI fruit shape terms are also inconsistent with the prevailing ontology terms (http://solgenomics.net/tools/onto/; SP:0001000, Solanaceae phenotype ontology). Therefore, to maintain consistency with terms present in the trait ontology database, we renamed categories and merged ones for which varieties were often classified in both (Supplemental Table S7). We merged the flattened and slightly flattened categories into just one category entitled flat. The term round is used for spherical shaped fruit. The category ellipsoid represented oval-shaped fruit. The heart-shaped category in UPOV and IPGRI was renamed oxheart. Fruit categorized as oxheart tended to be large and tapered with prominent shoulders. The term heart represented fruit that were larger toward the proximal end than the distal end, had less prominent shoulders than oxheart, and had a distinct tip at the distal end. Instead of pear shaped or pyriform, we adopt the term obovoid. The category long included varieties that produced very elongated, cylindrical, tapered, and often slightly curved fruit. The term rectangular remained the same. Thus the eight categories are flat, round, rectangular, ellipsoid, heart, long, obovoid, and oxheart. After this modification of the fruit shape categories, we classified our germplasm accordingly (Supplemental Table S1).

Phenotypic Analysis of the Tomato Subcollection of 120 Accessions

We selected 120 representative cultivated tomato accessions from the larger set of 368 examined. This selection was balanced to equally represent the eight shape categories as well as regional representation. Two tomato var. *cerasiforme* accessions were also included whereas more distant wild relatives were not included. Phenotypic data were collected using TA software program (Brewer et al., 2006; Gonzalo et al., 2009). The TA attribute values were subjected to multiple iterations of LDA to determine which attributes were most important for defining shape. First, all 36 attributes were subjected to LDA after which seven were removed because they showed high correlations with one or more attributes that were kept in the analysis. Then, several combinations of three attributes were subjected to LDA. The highest value of the proportion of correct assignments in the cross-validation test was obtained when fruit shape index, distal end protrusion, and width at the widest position were included. Additional attributes were added one by one and kept only if they led to an increase in the correct proportion of assignments in the cross-validation test. This process led to the selection of seven TA-defined attributes for an objective classification scheme (Supplemental Table S2). The predictive accuracy of an objective image-based classification scheme using a fixed linear discriminant function based on these seven attributes was then assessed. The analyses were carried out with MINITAB 15.1.0.0 software.

Marker Development for the Fruit Shape Genes

Two alleles are known for the *SUN*, *OVATE*, and *LC* genes (Gonzalo and van der Knaap, 2008; S. Muñoz, N. Ranc, E. Botton, A. Berard, S. Rolland, P. Duffe, Y. Carretero, M. Le Paslier, C. Delalande, M. Bouzayen, D. Brunel, and M. Causse, unpublished data; LC sequence accession nos. JF284938 and JF284939). *FAS* is represented by more than two alleles. For this study however, we focused on the allelic variant that carried the proposed 6- to 8-kb insertion in the first intron that is underlying the *FAS* gene mutation in the cultivated germplasm. Another *FAS* allelic variant (Cong et al., 2008) appears to be unique since the allele has not been found in other accessions.

For *OVATE*, a derived cleaved amplified polymorphic sequence marker was developed using a fluorescently labeled M13 F primer (5'-CAC-GACGTTGTAACGAC-3') in combination with primers EP158 (5'-CAC-GACGTTGTAACGAC-3') and EP159 (5'-CGTCGGTTTCTACGTCATCA-3'). After amplification, the products were digested with *DdeI* and separated on a LI-COR IR2 4200 (LI-COR Biosciences). The null allele of the *OVATE* gene, resulting in an elongated fruit, yielded a fragment of 125 bp whereas the wild-type allele yielded a fragment of 135 bp.

For *FAS*, we first determined part of the sequence of the large insertion present in the first intron of the *YABBY* gene. This large insertion is indicative of the allele that causes the increased locule number phenotype (Cong et al., 2008). Genomic DNA was digested with *HindIII* and separated on a 0.8% agarose gel. After transfer of the DNA to Hybond N+, the blots were hybridized to a DNA fragment amplified with EP1016 (5'-CGAAGAGTGC-TAATTGATGCT-3') and EP1017 (5'-TTTCGATTTTATGGAACTTTTGA-3') corresponding to the 3' region of intron 1 in the *FAS* gene. The presence of a band of approximately 4.2 kb coincided with the high-locule phenotype whereas a band of approximately 4.0 kb coincided with the low-locule phenotype. Digested DNA samples from tomato accessions Zapotec Pinked Ribbed, Costoluto Genovese, and LYC281 carrying the *fas* insertion allele were separated on a 0.8% low-melt agarose gel overnight and the 4.2-kb fragments were extracted from the gel using β -agarose (NEB). The *HindIII* DNA fragments were self ligated overnight. Amplification across the *HindIII* site was conducted using primers EP1031 (5'-AGCATCAATTAGCACTCTTCG-3') and EP1032 (5'-GCTGCAAAGGCAACAGTACA-3'), resulting in a 4-kb band that was sequenced in its entirety. This analysis permitted the identification of the breakpoint of the 3' end of the insertion as well as part of the insertion sequence. Southern-blot analysis using the insertion fragment as probe revealed that the inserted region was unique in the tomato genome and that it had rearranged (moved, not duplicated) from a region that mapped very close to *fas* (data not shown). Based on the sequence analysis of the insert with that of the genomic DNA, we developed three primers: EP1069 (5'-CCAATGATA-ATTAAGATAITGTGACG-3'), EP1070 (5'-ATGGTGGGGTTTTCTGTTC-3'), and EP1071 (5'-CAGAAATCAGAGTCCAATTCCA-3'). When the insertion is present, EP1069 and EP1071 will amplify a band of 466 bp; when the insertion is absent (wild type), EP1070 and EP1071 will amplify a band of 335 bp. The amplification products were separated on 2% agarose gels.

For *SUN*, we were unable to develop a reliable PCR-based marker. Instead, we used Southern-blot analysis to detect the alleles at this locus. DNA digested with *EcoRV* was hybridized with a probe amplified with primers EP45 (5'-TTTACCCGATGTGAAAACGA-3') and EP46 (5'-CATCAATAGTC-CAAGGGGAAA-3'). An extra 4.3-kb fragment signifies the presence of the gene duplication that leads to an elongated fruit shape whereas the absence signifies the wild-type allele at *sun*.

For *LC*, we developed four primers: lcn-SNP695-F (5'-GTCTCTGGAT-GATGACTATTGCACTTT-3'), lcn-SNP695-R (5'-TCAGCGCCTCATTTTCTA-TAGTATTTGT-3'), lcn-SNP695-F-cer (5'-CTTTCTCTAAAAGATTTGGCAT-GAGGT-3'), and lcn-SNP695-R-lev (5'-AAAGTAGTACGAATTGTCCAAT-CAGTCAG-3') that are included in the same PCR master mix. When the cultivated allele is present, lcn-SNP695-F and lcn-SNP695-R-lev will amplify a band of 533 bp; when the wild-type allele is present lcn-SNP695-F-cer and lcn-SNP695-R will amplify a band of 395 bp.

Marker Selection

To genotype the 120 accessions, we assessed alleles for 10 SSR, 14 CAPS, and one InDel marker. These markers were selected from the Tomato Mapping Resource Database (<http://www.tomatomap.net/>) and chosen based on their polymorphisms within cultivated tomato as well as their random distribution across the genome. Consequently, two or three markers per chromosome were employed with the exception of chromosome 10 that was only genotyped with one marker. CAPS markers were scored on 2% to 4% agarose gels whereas the

InDel and SSR markers were scored on the LI-COR IR2 4200 (LI-COR Biosciences). Details on the markers used are given in Supplemental Table S4.

Genetic Cluster Analysis

Clusters of similar genotypes were delineated using STRUCTURE version 2.2 (Pritchard et al., 2000). To avoid redundancy and bias in our subcollection, we removed four accessions that had the same genotype based on the alleles for the 25 randomly distributed markers and four fruit shape genes. The accessions that were removed were Jersey Devil (identical to Howard German), PI513088 and PI513036 (identical to Opalka), and UPV11936 (identical to Yellow Plum). A model assuming admixture and independent allele frequencies was selected. We used a burn-in period of 500,000 Markov Chain Monte Carlo iterations and then 1,000,000 iterations after burn in to estimate the parameters. The selected run length was much longer than suggested by Pritchard and colleagues (Pritchard et al., 2007) to minimize the effect of the starting configuration as well as to obtain the most accurate parameter estimates. Twenty independent runs were done for K (= number of clusters) varying from 1 to 15. The K optimum was defined according to the method proposed by Evanno et al. (2005). The strong modal signal at the true $K = 5$ (Supplemental Fig. S1) was also supported by the plateau observed for parameter $P(X|K)$ at $K = 5$ (Supplemental Fig. S2), the rate of change of the likelihood distribution, and the absolute values of the second order.

The assignment of individuals to clusters was quite robust when compared to predefined classes. Some shape categories, genotype classes, and fruit shape genes are overrepresented in some particular cluster (Supplemental Table S6). For verification of the STRUCTURE groupings, we estimated pairwise F_{ST} (θ ; Weir and Cockerham, 1984) and Nei's standard genetic distance (Nei, 1978) using the Microsatellite analyzer V4.05 (Dieringer and Schlotterer, 2003). The P value for the F_{ST} was calculated based on 10,000 permutations and a Bonferroni correction was applied.

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers JF284938 and JF284939.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Estimation of optimum number of clusters (K ; Evanno et al., 2005).

Supplemental Figure S2. The graph for the parameter L (K) and number of clusters (K).

Supplemental Table S1. Description and data of Tomato Germplasm Collection.

Supplemental Table S2. Ranges of variation for attributes related to fruit shape categories.

Supplemental Table S3. Associations among fruit shape genes and fruit shape attributes measured by Tomato Analyzer.

Supplemental Table S4. Molecular markers used in this study.

Supplemental Table S5. Estimates of Nei's standard genetic distance and pair-wise θ between five clusters of tomato accessions.

Supplemental Table S6. Contingency tables associating STRUCTURE-based clusters and different categorical values of categories with more than 10 accessions.

Supplemental Table S7. Comparison among tomato fruit shape categories according to IPGRI (1996), UPOV (2001), and the classification used in this study.

ACKNOWLEDGMENTS

We thank the Tomato Genetics Resource Center (University of California, Davis); U.S. Department of Agriculture, Germplasm Resources Information Network (Geneva, NY); Institut für Pflanzengenetik und Kulturpflanzenforschung (Gatersleben, Germany); and Drs. Silvana Grandillo and Maria Jose Gonzalo for providing the seeds of the accessions used in this study. We

thank Dr. Maria Jose Gonzalo, Ms. Jenny Moysenko, and Ms. Na Zhang for fruit and image collections.

Received October 15, 2010; accepted March 24, 2011; published March 25, 2011.

LITERATURE CITED

- Barrero LS, Tanksley SD** (2004) Evaluating the genetic basis of multiple-locule fruit in a broad cross section of tomato cultivars. *Theor Appl Genet* **109**: 669–679
- Bernatzky R, Tanksley SD** (1986) Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. *Genetics* **112**: 887–898
- Brewer MT, Lang L, Fujimura K, Dujmovic N, Gray S, van der Knaap E** (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 MT, Moysenko JB, Monforte AJ, van der Knaap E** (2007) Morphological variation in tomato: a comprehensive study of quantitative trait loci controlling fruit shape and development. *J Exp Bot* **58**: 1339–1349
- Cong B, Barrero LS, Tanksley SD** (2008) Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. *Nat Genet* **40**: 800–804
- De Candolle A** (1886) *Origine des plantes cultivees*. F Alcan, Ed 3. Alglave, Paris, pp 290–292
- De Sahagún B** (1959) Florentine Codex: general history of the things of New Spain. Book 8: Kings and Lords. *In* Monographs of the School of American Research and the Museum of New Mexico, Vol 14, Pt IX. The School of American Research and The University of Utah, Santa Fe, NM, p 89
- Dieringer D, Schlotterer C** (2003) Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Mol Ecol Notes* **3**: 167–169
- Dunal MF** (1813) *Histoire Naturelle, Médicale et Economique des Solanum, et des Genres qui ont été Confondus Avec Eux*. Koenig, Paris
- Evanno G, Regnaut S, Goudet J** (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611–2620
- 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
- Gonzalo MJ, Brewer MT, Anderson C, Sullivan D, Gray S, van der Knaap E** (2009) Tomato fruit shape analysis using morphometric and morphology attributes implemented in Tomato Analyzer software program. *J Am Soc Hortic Sci* **134**: 77–87
- Gonzalo MJ, van der Knaap E** (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, Ku H-M, Tanksley SD** (1996) Characterization of fs8.1, a major QTL influencing fruit shape in tomato. *Mol Breed* **2**: 251–260
- Hackbusch J, Richter K, Müller J, Salamini F, Uhrig JF** (2005) A central role of Arabidopsis thaliana ovate family proteins in networking and subcellular localization of 3-aa loop extension homeodomain proteins. *Proc Natl Acad Sci USA* **102**: 4908–4912
- IPGRI** (1996) *Descriptors for Tomato (Lycopersicon spp.)*. International Plant Genetic Resources Institute, Rome, p 29
- Jenkins JA** (1948) The origin of cultivated tomato. *Econ Bot* **2**: 379–392
- Jiang N, Gao D, Xiao H, van der Knaap E** (2009) Genome organization of the tomato sun locus and characterization of the unusual retrotransposon Rider. *Plant J* **60**: 181–193
- Kinzer SM, Schwager SJ, Mutschler MA** (1990) Mapping of ripening-related or ripening-specific cDNA clones of tomato (*Lycopersicon esculentum*). *Theor Appl Genet* **79**: 489–496
- Ku HM, Grandillo S, Tanksley SD** (2000) fs8.1, a major QTL, sets the pattern of tomato carpel shape well before anthesis. *Theor Appl Genet* **101**: 873–878
- Liu J, Van Eck J, Cong B, Tanksley SD** (2002) A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. *Proc Natl Acad Sci USA* **99**: 13302–13306
- Male CJ** (1999) *100 Heirloom Tomatoes for the American Garden*. Workman Publishing, New York
- Matthioli PA** (1544) *Di pedacio dioscoride anazerbeo libri cinque della historia, et material medicinale trodotti in lingua vulgare Italiana*. *In* PDo Anazarbos, ed, *De materia medica*, Venetia de Bascaraini, Venetia, Italy, p 326
- Mayer KF, Schoof H, Haecker A, Lenhard M, Jürgens G, Laux T** (1998) Role of WUSCHEL in regulating stem cell fate in the Arabidopsis shoot meristem. *Cell* **95**: 805–815
- Mazzucato A, Papa R, Bitocchi E, Mosconi P, Nanni L, Negri V, Picarella ME, Siligato F, Soressi GP, Tiranti B, et al** (2008) Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (*Solanum lycopersicum* L.) landraces. *Theor Appl Genet* **116**: 657–669
- Nei M** (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590
- Nesbitt TC, Tanksley SD** (2002) Comparative sequencing in the genus *Lycopersicon*: implications for the evolution of fruit size in the domestication of cultivated tomatoes. *Genetics* **162**: 365–379
- Oellinger G** (1553) *Herbarium des Georg Oellinger, Anno 1553 zu Nürnberg*, Manuscript 2362, Plate 44 S541, S543, S545. Akademischer Gemeinschaftsverlag, Salzburg, Austria
- Paran I, van der Knaap E** (2007) Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. *J Exp Bot* **58**: 3841–3852
- Park YH, West MA, St Clair DA** (2004) Evaluation of AFLPs for germplasm fingerprinting and assessment of genetic diversity in cultivars of tomato (*Lycopersicon esculentum* L.). *Genome* **47**: 510–518
- Peralta IE, Knapp S, Spooner DM** (2008) Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicoides*, sect. *Juglandifolia*, sect. *Lycopersicon*; Solanaceae), Vol 84. American Society of Plant Taxonomists, Ann Arbor, MI
- Philouze J** (1991) Description of isogenic lines, except for one, or two, monogenically controlled morphological traits in tomato, *Lycopersicon esculentum* Mill. *Euphytica* **56**: 121–131
- Pritchard JK, Stephens M, Donnelly P** (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959
- Pritchard JK, Wen X, Falush D** (2007) Documentation for STRUCTURE software: version 2.2. <http://pritch.bsd.uchicago.edu/software/structure22/readme.pdf> (April 9, 2011)
- Ranc N, Muñoz S, Santoni S, Causse M** (2008) A clarified position for *Solanum lycopersicum* var. *cerasiforme* in the evolutionary history of tomatoes (solanaceae). *BMC Plant Biol* **8**: 130
- Rick CM, Holle M** (1990) Andean *Lycopersicon esculentum* var. *cerasiforme*: genetic variation and its evolutionary significance. *Econ Bot* **44**: 69–78
- Ruiz JJ, Garcia-Martinez S, Pico B, Gao MQ, Quiros CF** (2005) Genetic variability and relationship of closely related Spanish traditional cultivars of tomato as detected by SRAP and SSR markers. *J Am Soc Hortic Sci* **130**: 88–94
- Sim SC, Robbins MD, Chilcott C, Zhu T, Francis DM** (2009) Oligonucleotide array discovery of polymorphisms in cultivated tomato (*Solanum lycopersicum* L.) reveals patterns of SNP variation associated with breeding. *BMC Genomics* **10**: 466
- UPOV** (1992) *Guidelines for the conduct of tests for distinctness, homogeneity and stability*. Tomato (*Lycopersicon esculentum* L.). UPOV, Geneva, p 21
- UPOV** (2001) *Guidelines for the Conduct of Tests for Distinctness, Uniformity and Stability (Tomato)*. Geneva, Switzerland, p 24
- Wang S, Chang Y, Guo J, Chen JG** (2007) Arabidopsis Ovate Family Protein 1 is a transcriptional repressor that suppresses cell elongation. *Plant J* **50**: 858–872
- Weir BS, Cockerham CC** (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370
- Williams CE, Clair DA** (1993) Phenetic relationships and levels of variability detected by restriction fragment length polymorphism and random amplified polymorphic DNA analysis of cultivated and wild accessions of *Lycopersicon esculentum*. *Genome* **36**: 619–630
- Xiao H, Jiang N, Schaffner EK, Stockinger EJ, van der Knaap E** (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* **319**: 1527–1530