the plant journal

The Plant Journal (2023) 116, 1136–1151 doi: 10.1111/tpj.16268

RESOURCE

Multi-year field trials provide a massive repository of trait data on a highly diverse population of tomato and uncover novel determinants of tomato productivity

Itay Zemach[1,†], Saleh Alseekh[2,3,†], Roni Tadmor-Levi[1], Josef Fisher[1], Shai Torgeman[1], Shay Trigerman[1], Julia Nauen[2], Shdema Filler Hayut[4], Varda Mann[4], Edan Rochsar[1], Richard Finkers[5], Regina Wendenburg[2], Sonia Osorio[2,6], Susan Bergmann[2], John E. Lunn[2], Yaniv Semel[7], Joseph Hirschberg[4], Alisdair R. Fernie[2,3,*] and Dani Zamir[1,*]

[1]The Robert H Smith Faculty of Agriculture, Food and Environment, Hebrew University of Jerusalem, Rehovot, Israel,
[2]Max-Planck-Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany,
[3]Center of Plant Systems Biology and Biotechnology, 4000 Plovdiv, Bulgaria,
[4]Department of Genetics, Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem 9190401, Israel,
[5]Plant Breeding, Wageningen Plant Research, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands,
[6]Department of Molecular Biology and Biochemistry, Instituto de Hortofruticultura Subtropical y Mediterranea "La Mayora", University of Malaga-Consejo Superior de Investigaciones Cientıficas, Malaga, Spain, and
[7]Phenome Networks, 10 Plaut Street, Science Park, 76706 Rehovot, Israel

Received 1 December 2022; revised 21 April 2023; accepted 29 April 2023; published online 7 May 2023.
*For correspondence (e-mail fernie@mpimp-golm.mpg.de; dani.zamir@mail.huji.ac.il).
†These authors contributed equally to this work.

SUMMARY

Tomato (*Solanum lycopersicum*) is a prominent fruit with rich genetic resources for crop improvement. By using a phenotype-guided screen of over 7900 tomato accessions from around the world, we identified new associations for complex traits such as fruit weight and total soluble solids (Brix). Here, we present the phenotypic data from several years of trials. To illustrate the power of this dataset we use two case studies. First, evaluation of color revealed allelic variation in *phytoene synthase 1* that resulted in differently colored or even bicolored fruit. Secondly, in view of the negative relationship between fruit weight and Brix, we pre-selected a subset of the collection that includes high and low Brix values in each category of fruit size. Genome-wide association analysis allowed us to detect novel loci associated with total soluble solid content and fruit weight. In addition, we developed eight F2 biparental intraspecific populations. Furthermore, by taking a phenotype-guided approach we were able to isolate individuals with high Brix values that were not compromised in terms of yield. In addition, the demonstration of novel results despite the high number of previous genome-wide association studies of these traits in tomato suggests that adoption of a phenotype-guided pre-selection of germplasm may represent a useful strategy for finding target genes for breeding.

Keywords: genome-wide association, soluble solids, tomato yield.

INTRODUCTION

Tomato (*Solanum lycopersicum*) is an important fruit crop that is consumed in both fresh and processed forms. For both types of product, the total soluble solid content of the fruits (mainly sugars and acids; Brix) is an important yield component. This is particularly true in processing tomatoes, where the field output can be presented as the product of total fruit yield multiplied by the Brix value (i.e., how many tomato sugars are harvested per unit area). In recent years considerable advances have been made in our understanding of how fruit yield can be increased by manipulating developmental aspects of the source–sink transition (Krieger et al., 2010; Park et al., 2014; Rodríguez-Leal et al., 2017; Soyk et al., 2019) and consequently upregulating metabolism. For example, a quantitative trait nucleotide within the gene encoding the cell wall invertase LIN5, underlying Brix, was identified in a *Solanum pennellii* introgression line population as well as in introgressions around the same genomic interval from other tomato wild relatives (Barrantes et al., 2016; Wang et al., 2020). This gene was similarly linked either to the soluble solid trait (Sauvage et al., 2014) or to sugar content in subsequent genome-wide association studies (GWAS) (Tieman et al., 2017). Similarly, the genetic architecture of other metabolites contributing to Brix, including aspartate, glutamate, malate, and alanine (Do et al., 2010; Sauvage et al., 2014; Schauer et al., 2006; Ye et al., 2017), as well as further primary metabolites (Do et al., 2010; Sauvage et al., 2014; Schauer et al., 2006; Ye et al., 2019), has been determined either by quantitative trait locus (QTL) mapping or via GWAS. In addition, a number of studies have focused on fruit weight (Chakrabarti et al., 2013; Frary et al., 2000; Mu et al., 2017; Prudent et al., 2009; Tanksley, 2004; Zhang et al., 2012), with three genes encoding regulators of this trait, namely *fw2.2*, regulating the carpel cell number, the cytochrome P450 gene *KLUH*, and *CELL SIZE REGULATOR*, which is particularly well characterized in this respect. In addition, *EXCESSIVE NUMBER OF FLORAL ORGANS* (*ENO*), which encodes an AP2/ERF transcription factor that regulates floral meristem activity, was additionally recently identified (Yuste-Lisbona et al., 2020). Intriguingly, the enhancement of both soluble solid contents and fruit weight has largely proven elusive although recent studies using the genetic approach of recurrent selection (Yamamoto et al., 2016), as well as computational modeling (Chen et al., 2021), suggest that it should be possible.

While multiplying the seed of the tomato diversity collection that is being established in Rehovot, we measured for each entry the fruit weight and the Brix value. The traits are negatively correlated and display considerable variance across the collection. In order to gain deeper insight into the genetic and metabolic factors underlying these two traits, we planted 400 lines that were characterized by a range of Brix values irrespective of the size range. Based on these results we selected 109 lines that were designated as the *Brix panel*, in which the Brix value, fresh weight, and primary metabolite content were subsequently measured. In parallel we characterized a broader set of 500 randomly selected genotyped lines; comparison of the GWAS results in the two panels revealed that several associations were apparent only in the Brix panel. Indeed, these analyses identified novel loci for both soluble solid content and fruit weight and confirmed that the negative correlation between these traits can be broken. Moreover, data from various independent intraspecific biparental F2 populations validated these and a handful of other associations. Importantly, this study indicates the utility of reducing the size of the population by pre-selecting genotypes based on their phenotypic variance as a means to find associations that would otherwise remain hidden.

RESULTS

Assembly of a tomato collection

The collection of tomato varieties is currently comprised of more than 7900 accessions. Their passport data are available in Table S1 and at http://unity.phenome-networks.com. Passport data include donor information, germplasm status, accession names and numbers, preliminary phenotypic data or descriptions, date of collection, and geographic details such as site, city, and country of collection. Germplasm status is sub-divided as shown in Figure 1. Irrespective of this, most accessions come from countries that are well known for their production and consumption of tomatoes such as Peru, USA, Mexico, Spain, and Italy; however, a surprisingly large number of 423 accessions were collected in the Russian Federation and donated by the IPK Gatersleben, likely reflecting the strong academic collaboration between the IPK genebank founders and the famous Russian researcher Nikolai Vavilov (Börner, 2012).

Phenotypic variation in the tomato collection

To evaluate the phenotypic diversity inherent in the collection, many of its accessions were characterized for morphological and agricultural traits. Specific categories were developed for each trait descriptor (Table S2) following the Solanaceae trait ontology (Bombarely et al., 2011). Phenotypes have been collected and documented since 2007, summing to a total of at least 20 phenotypic and 60 metabolic traits across different sub-panels (Figure 1). The distribution of qualitative traits revealed that the common characteristic is the classic tomato – red fruited, indeterminate, and 100–150 g in weight (Figure 2a–e). That said, for each trait descriptor there is an adequate representation for each category.

Tomato fruit color is a well-studied trait that represents an important component of fruit quality. Among the non-red fruited accessions, we find the well-known color variants *tangerine* (*t*) and yellow-flesh (*r*), whose functional structural genes involved in the carotenoid biosynthesis pathway have been cloned (Isaacson et al., 2002; Ronen et al., 2000). Other colors such as *green flesh* (*gf*) and a black external color are associated with anthocyanin and chlorophyll degradation mutants, respectively (Barry et al., 2008). Fruit shape also varied greatly across the collection, with 16 accessions being characterized by unique shapes (Figure 2d). Similarly, there was considerable variation in inflorescence structure, where the *compound inflorescence* (*s*) gene is the major functional gene (Lippman et al., 2008), but two further interacting MADS-box genes have been cloned from the collection that are also

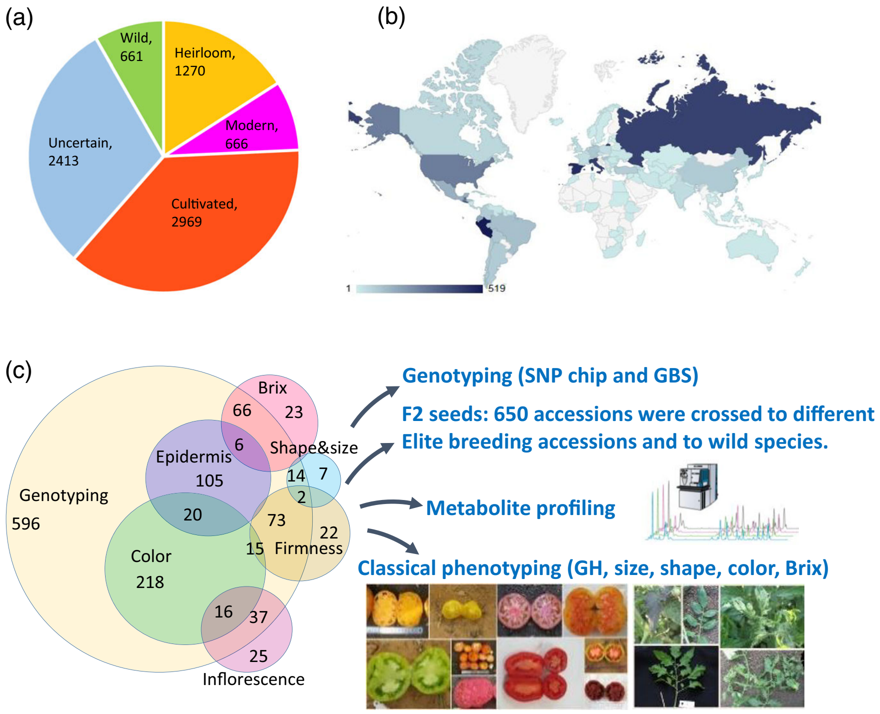

Figure 1. A description of the tomato collection.
(a) Pie chart of about 8000 tomato accessions, including 2969 cultivated varieties, 666 modern inbreds, 1270 heirloom, 661 wild species, and 2413 not clearly classified. (b) Geographic distribution of collection origin.
(c) Schematic representation of seven subset panels of the tomato collection. The Venn diagrams show the number of accessions in each panel and overlapping accessions between other panels. The accessions in the subset panels were subjected to different levels of genotyping using SNP-Chip and genotyping by sequencing (GBS), classical phenotyping for multiple growth seasons, and fruit metabolite profiling. About 650 accessions were crossed with different elite breeding accessions and with wild species to produce F2 seeds for crossing validation of QTLs.
For more information about accession numbers, passport data, accession names, geographical distributions, and preliminary phenotypic data, see Table S1 and http://unity.phenome-networks.com.
SNP, single nucleotide polymorphism; GBS, genotyping by sequencing; GH, growth habits.

important in determination of the highly branched inflorescence phenotype (Soyk et al., 2017).

Allelic variation of *phytoene synthase 1* (*Psy1*)

One of the traits examined during the phenotypic characterization of the collection is fruit internal color (Figure 2e). It was measured by looking at the external color of the ripe fruit and then at its cross-section. Most lines in the collection, whose color is red or yellow, showed a uniform color, except 27 lines that displayed a yellow/red *bicolor* characteristic (Figure 3a; Table S3). The whole fruit of these lines looked yellow but often red color could be seen in the center of the fruit placenta (Figure 3). The *bicolor* phenotype originated from heirlooms, and therefore the genetic background (the original line where the mutation occurred) is unknown. In addition, it is important to note that these mutant lines are known for many years and yet, this mutation has not yet been studied in depth. A literature review suggests that this mutation may have been identified in the 1950s (Young, 1956). The study was conducted on accession Rumi Banjan, which exists in the collection (CC4418), and the author hypothesized that it was an additional gene (modifier) that affected the yellow flesh allele r, which is a known mutant of the *PSY1* gene. During the phenotyping of the collection, CC4418 was characterized as displaying a *bicolor* phenotype and bearing the missing mutation (Figure 3; Table S3). Carotenoid profiling was performed separately for red and yellow sectors of the CC0383 fruits (accessions displaying similar phenotypes and the same mutation as CC4418); the results revealed that red sectors accumulated lycopene, while yellow sectors showed low levels of carotenoids similarly to known *yellow flesh* (*r*) mutants (Figure 3b; Figure S1). We further performed an allelism test by crossing the *bicolor* accessions CC0282, CC0299, and CC0383 to known color mutations and demonstrated the *bicolor* phenotype to be a *Psy1* allele. Single nucleotide polymorphism (SNP) genotyping of the F2 population derived from the cross of *bicolor* accession CC0383 with the *yellow flesh* mutant e3756m2, together with F3 progeny tests for each F2

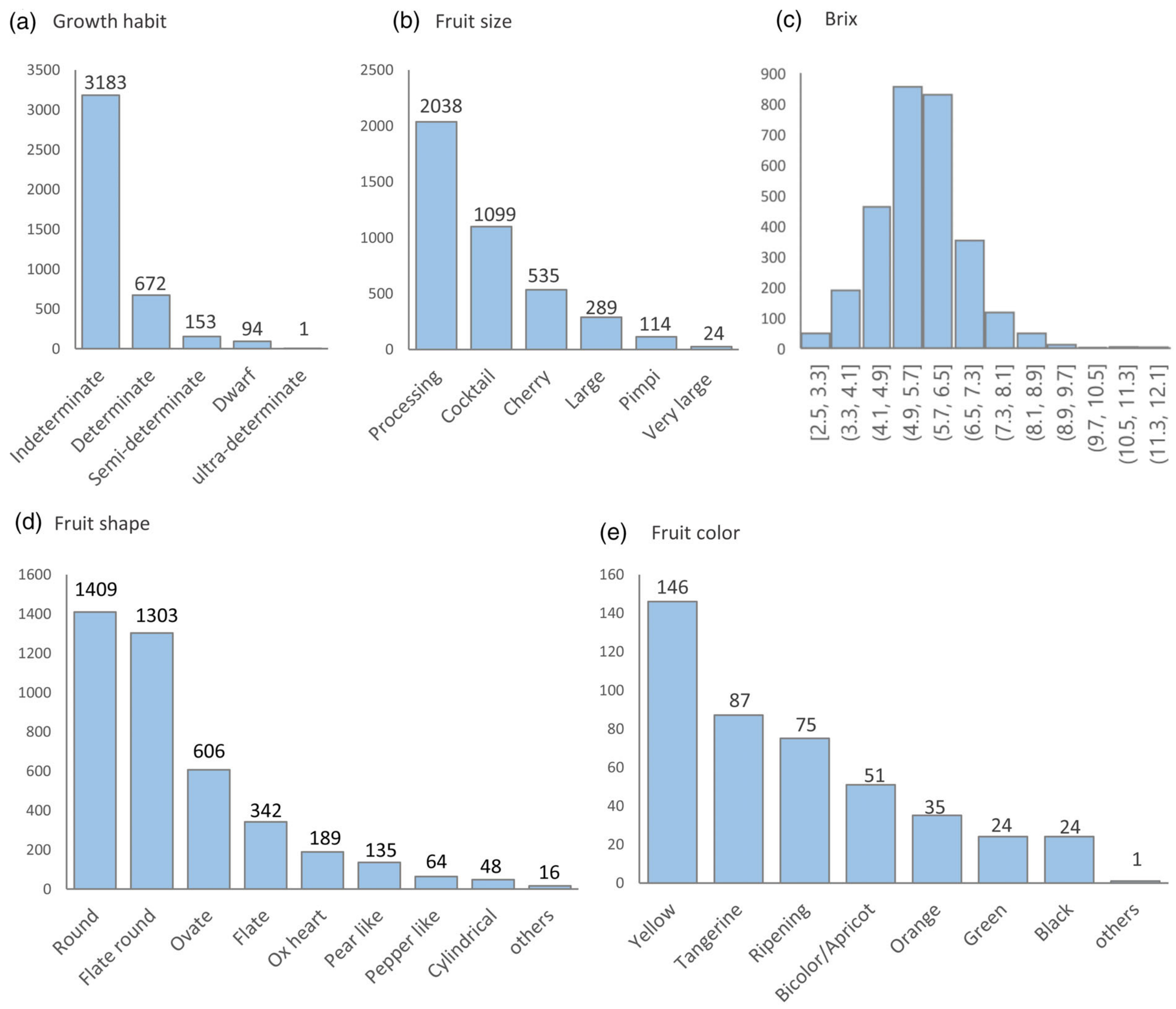

Figure 2. Germplasm status and qualitative trait distribution of the tomato collection.
(a) Growth habits. (b) Fruit size. (c) Frequency distribution of Brix across 3135 varieties. (d) Fruit shape. (e) Fruit internal color. Numbers above columns represent the number of accessions scored per descriptor. Traits and categories were based on Solanaceae trait ontology (Bombarely et al., 2011). The *y*-axis represents the number of accessions in each descriptor.

individual, led us to suspect that the *bicolor* allele causative mutation is a deletion in the *Psy1* region, which was later confirmed by sequencing of the gene region. The results show that a large deletion of 3786 bp, most of it in the promoter region of *Psy1* and part in the first two exons of the gene, occurred in the *bicolor* lines (Figure 3c). Primers were designed to capture the deletion region by PCR, and using this marker we genotyped 31 lines which were suspected to have the *bicolor* phenotype. We found that all *bicolor* accessions harbor the same deletion, and phylogenetic analyses led us to conclude that the *bicolor* accession background is similar (Figure 3d). Furthermore, as the collection was multiplied and phenotyped over more than 10 years, the analysis of the *bicolor* trait indicated that the collection was maintained in an orderly manner.

Genetic diversity and population structure of the collection

To estimate the genetic variation in the collection we analyzed 1200 accessions that were genotyped for 384 SNPs spanning the genome (Figure 4a–c) (http://unity.phenome-networks.com). According to the Evanno method, the ideal number of sub-populations K is where the ΔK value is the highest (Earl & von Holdt, 2012). These analyses revealed four sub-populations, corresponding to wild, heirloom, cultivated, and modern cultivars (Figure 4c). Such a large population of lines is cumbersome to handle for most researchers and plant breeders. For this reason, 10 sub-panels of the collection were established by selecting accessions that provided maximum variation for specific traits (color, Brix, shape, size, firmness, inflorescence).

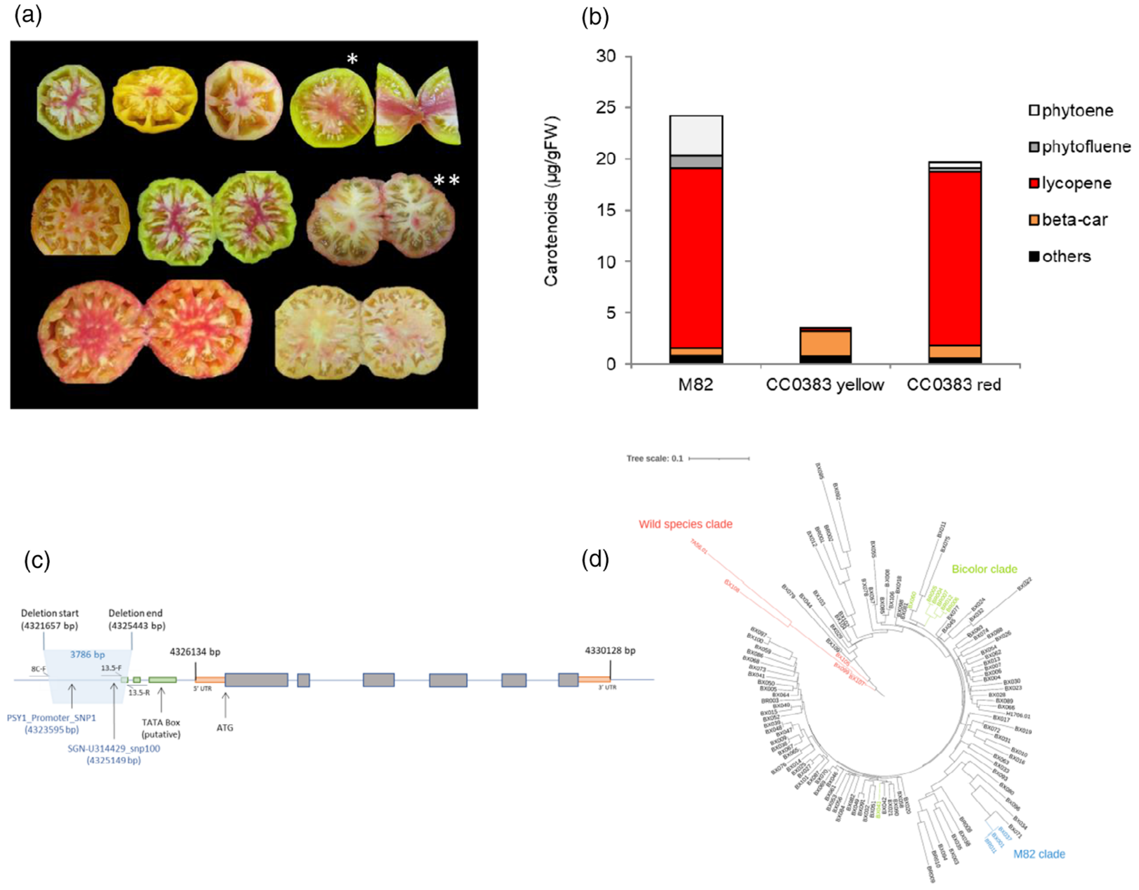

Figure 3. Allelic variation of *phytoene synthase 1* (*Psy1*) and the yellow/red *bicolor* characteristic.
(a) Pictures of selected fruit *bicolor* mutants characterized in the tomato collection. (b) Carotenoid profiling of red and yellow sectors of accession CC0383. (c) Schematic illustration of *Psy1* highlighting a large deletion of 3786 bp, most of it in the promoter region. (d) Phylogenetic analysis based on PCR and sequence analysis of the deletion region using 140 accessions including the *bicolor* mutants.
*A cross-section showing red color in the center of the fruit placenta and yellow margins in the pericarp area. **Fruit cross-sections of accession Rumi Banjan (CC4418) identified in the 1950s and displaying a *bicolor* phenotype and bearing a mutation similar to that of CC0383.

Alternatively, accessions were grouped according to other criteria (Cerasiforme, Wild, Modern Inbreds, and the Resequenced collection). The number of accessions in each panel and other genotyping and phenotyping information are shown in Figure 1c. The Brix panel was constructed to investigate the total soluble solids trait in tomato fruit in a manner that will neutralize the effect of fruit size as much as possible. As a first approach, phenotypic and genotypic data were collected from a primary panel of 400 accessions that was constructed over time while the collection was increased. The 400 accessions were selected to include high and low Brix levels in each fruit size category and the collection was further reduced to 109 lines in the hope of overcoming the well-known negative correlation between Brix value and fruit weight (Figure 5a). The 109-member Brix panel was planted in two different environments: in autumn 2009 in the greenhouse for growth during the Israeli winter and secondly in the following summer under open field conditions. The Brix value, fruit weight, locule number, color, size, and shape were recorded and fruit primary metabolite levels were assessed (Table S4). To better understand the relations between the primary metabolite contents and the Brix value, we created a correlation network between the average of the phenotypic and metabolic traits (Figure 5b). In keeping with previous studies (Magwaza & Opara, 2015), this network revealed that Brix is positively correlated with sucrose, glucose, and fructose as well as the sugar alcohols inositol, maltitol, and dehydroascorbate. Moreover, proline is at the juncture of amino acid and sugar clusters, perhaps explainable by the fact that both proline and the sugar alcohols are highly responsive to stress conditions (Yoshiba et al., 1997). Intriguingly, no significant correlation was found between Brix and fruit weight, indicating that this correlation was partially broken by our selection of the accessions in the Brix panel. However, fruit weight remains negatively correlated with sucrose and glutamate contents (Figure 5b,c).

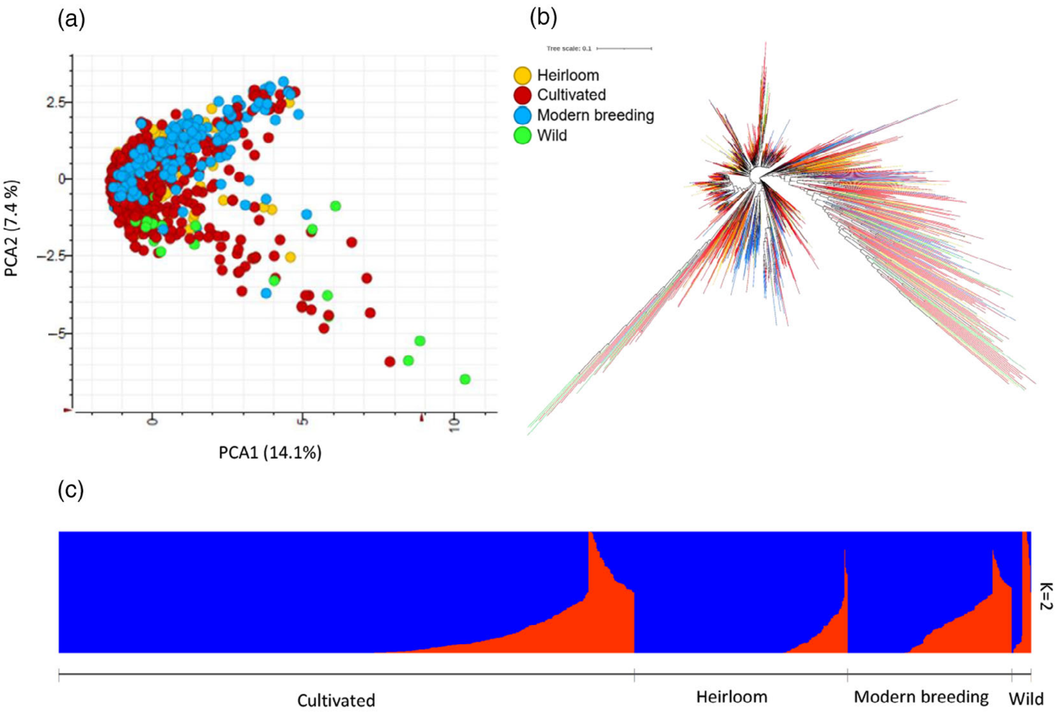

Figure 4. Genomic diversity of the collection of 1200 tomato accessions.
(a) Principal component analysis (PCA) based on 384 SNPs across 1200 tomato accessions classified as cultivated, modern inbreds, heirloom, and wild species. (b) Neighbor-joining phylogenetic tree of the tomato collection. (c) Population structure based on 384 SNP markers analyzed in 1200 accessions. STRUCTURE analysis was conducted with $K = 2$, estimated according to (Earl & von Holdt, 2012). Each accession is represented by a single column, with the color indicating cluster membership. PCA was computed by Genedata Expressionist® Analyst using a covariance matrix and positive correlation ($1 - r$). The neighbor-joining tree was generated with Clustal Omega and viewed in ITOL (Letunic & Bork, 2007). In PCA, tree accessions are color-coded as follows: red (cultivated), green (wild species), orange (Heirloom), and blue (modern inbreds).

Population structure and association mapping in the Brix panel

In order to reduce the number of false positive GWAS associations, a population structure correction was performed. This indicated an optimal cluster number of seven, with the clusters including the two main populations of cultivated and wild species, the admixed cerasiforme accessions, and four further sub-populations – determinant, indeterminant, modern inbreds, and vintage varieties (Figure S2a,b). This number of populations was used to estimate the kinship relationships among the panel. Within the 109 Brix panel accessions, the average LD was $R^2 = 0.38$ (Figure S2c), consistent with previous studies (Sauvage et al., 2014). There are two different common approaches in GWAS. The first requires a preliminary haplotype division of the genotyping panel of markers by estimating the LD between markers and subsequently treats groups of tightly linked markers as a single representative marker. The second, which was adopted in this study, ignores the LD between markers and performs GWAS on the basis of single marker analysis (Mackay & Powell, 2007). However, the LD data between markers were used in the design of primers for SNP markers for in-house genotyping during the F2 validation process.

A total of 600 common associations (LOD > 7) were found with both GenABEL and GAPIT algorithms (complete list available at https://unity.phenome-networks.com). Given the large number of trait–marker associations we used the following filtering criteria to identify the most promising candidate genes: (i) we selected significant QTLs with minor allele frequency (MAF) > 0.05; (ii) we selected QTLs harboring only a few genes within the associated genomic interval; and (iii) candidate genes were screened according to their annotation and defined function. This process led us to focus on 85 significant associations that were identified for a total of 11 metabolites as well as Brix and fruit weight (Table S5). Additionally, we found a handful of other associations that did not fulfill all the above criteria but were nevertheless of high interest. For example, although the association between glutamate and solcap_snp_sl_18959 has an LOD score of only 3.27, it was also taken through the validation process as the marker solcap_snp_sl_18959 harbors the gene *glutamate synthase*. Returning to the chosen 85 associations, their genetic effects varied between 38 and 500%, whilst their

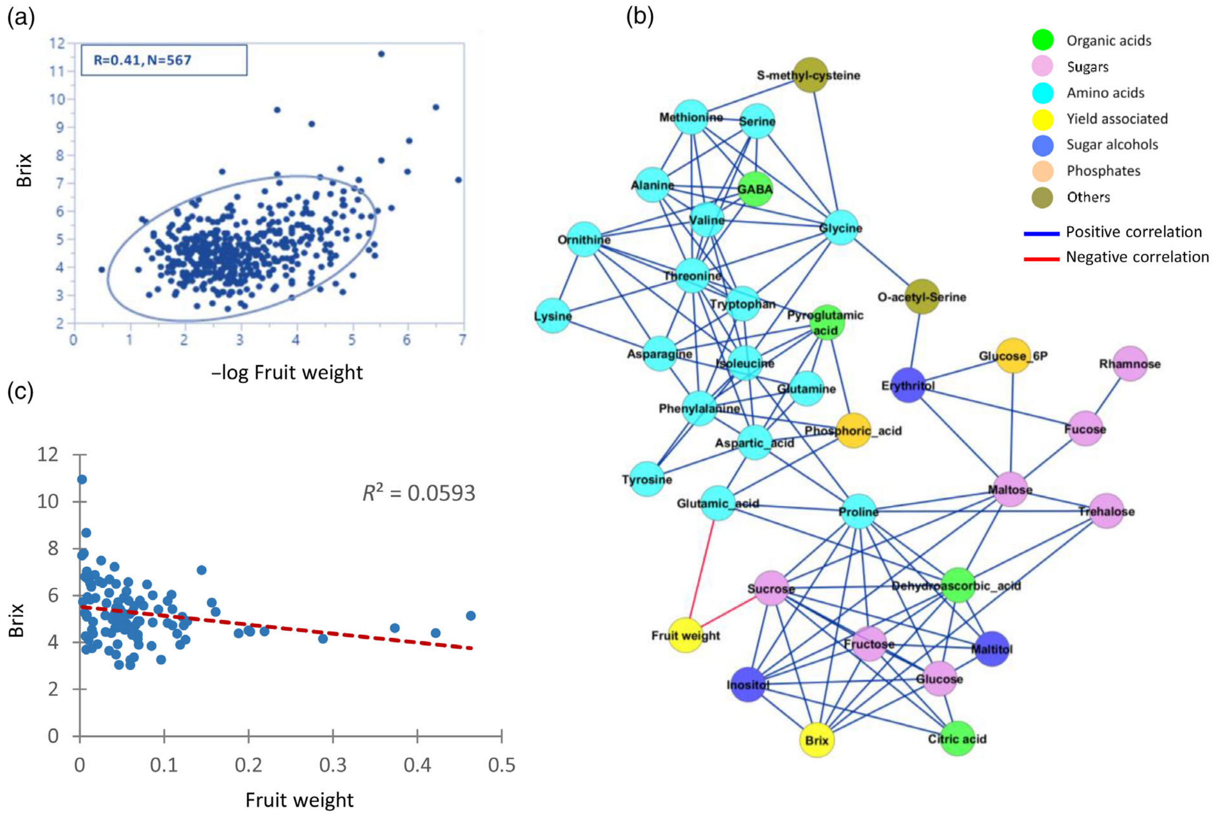

Figure 5. Correlation analysis between fruit weight, Brix, and primary metabolites.

(a) Scatter plot and linear regression of fruit weight (−log(fresh weight)) and Brix data measured for 567 accessions of the tomato collection, $R = 0.41$. (b) Correlation network of primary metabolites measured by GC-MS, fruit weight, and Brix across the 109 Brix panel lines for multiple seasons. (c) Scatter plot and linear regression of fruit weight and Brix measured for 109 Brix panel lines, $R = -0.06$.

Spearman's rank correlation test with Bonferroni correction for multiple comparisons with a significance level of 0.05. Cytoscape was used to generate graphs. Each trait (node) is represented by a circle, and different colors represent different metabolite classes. Interactions are indicated with lines: blue represents a positive correlation, and red represents a negative correlation.

heritability varied between 12 (for aspartate) and 68% (glucose). Within the 85 associations there were three main associations for Brix (Figure 6a); for example, Brix was associated with solcap_snp_39725, which is located on chromosome 9 within the well-characterized gene *LIN5* (Fridman et al., 2004). In an earlier study, Fridman et al. found three polymorphic SNPs in exon 2 of *LIN5* but focused on the one originating from the wild species *S. pennellii* (Fridman et al., 2004). Here we revealed that solcap_snp_39725 had an effect of 38% on Brix, and this association has an LOD score of 9.37 (Figure 6a). Moreover, the marker solcap_snp_sl_53191 had an effect of 40% with an LOD of 6.42, whilst the marker solcap_snp_sl_31961 had an LOD value of 5.01 and hence did not meet our filtering criteria. We decided to validate all three associations nonetheless, since solcap_snp_sl_31961 also displayed putative pleiotropic effects as it was strongly associated with citrate, proline, sucrose, and glutamine levels. One novel association for fruit weight was identified, namely with solcap_snp_sl_21280, which is located on chromosome 1, with an LOD score of 10 and an effect of 140%. This SNP was additionally not associated with Brix, providing further evidence for the reduction of the correlation between fresh weight and Brix within this panel (Figure 7a).

Of the measured metabolites, aspartate was found to be associated with a region of chromosome 8 (with solcap_snp_sl_56404 being representative of this region) as well as a region of chromosome 6 (Figure S3a,b). However, the previously reported association with chromosome 4 (Sauvage et al., 2014) was not apparent in this subpopulation. By contrast, the previously reported association with the aluminum-activated malate transporter gene (Tieman et al., 2017; Ye et al., 2017) was identified here with an LOD of 7.86 and an effect of 100%. Glutamate was found in novel associations with SNPs on chromosomes 3, 8, and 12 displaying LODs of 8.5, 8.2, and 9.5 and effects of 232, 219, and 262%, respectively (Figure 8a,b). We further validated the association of glutamate with solcap_snp_sl_18959 on chromosome 3 since this region

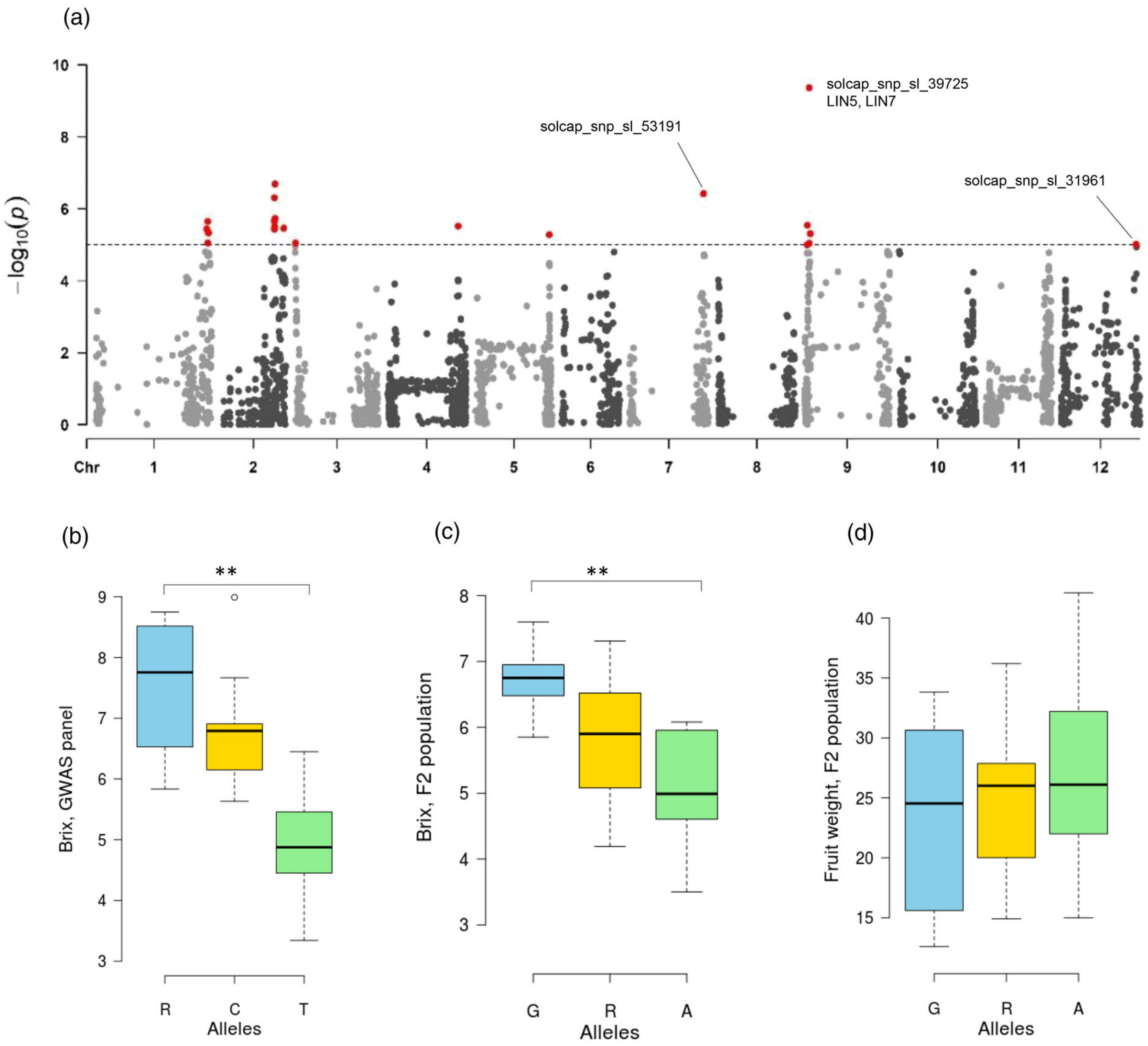

Figure 6. Genome-wide association study and F2 validation of Brix.
(a) Manhattan plot for Brix using the Brix panel. SNPs with significant associations ($P \le 0.05$) are colored in red. The significant associations are highlighted; SNPs were validated using F2 segregated populations. (b) Haplotype analysis of the SNP marker 'solcap_snp_sl_39725' at position 3 477 979 on chromosome 9 across the Brix panel. (c) Cross-validation of the association of Brix with the SNP marker 'solcap_snp_sl_39725' using the F2 segregated population. (d) Haplotype analysis of fruit weight in the F2 population showed no effect of the marker 'solcap_snp_sl_39725' on fresh weight.
Asterisks indicate significant differences as determined by Student's *t*-test. LIN, cell wall invertase. Letters (C, T, A, G, R) are SNPs representing different haplotypes.

corresponds to an interval harboring seven genes, including *Solyc03g083440*, which is annotated as a glutamate synthase gene. In addition to the above, a number of further associations were uncovered (Table S5) for aspartate, sucrose, proline, glucose, and glutamine, which were distributed across the tomato genome.

Association validation

Whilst GWAS is an immensely powerful tool for discovering gene–trait associations, independent validation of its results is crucial due to a number of disadvantages inherent to the approach (Tam et al., 2019). For this purpose, an average of eight F2 biparental intraspecific populations were developed for each of the Brix panel accessions. These seeds are freely available as an important resource for validation. Each of the most significant GWAS associations was validated by selecting F2 populations segregating for the associated SNP using the linkage disequilibrium data between markers, in order to minimize the occurrence of false positives. For each F2 population, a total of 200 individuals were genotyped and 15 plants from each genotype (i.e., 15 homozygous common, 15 homozygous rare, and 15 heterozygous) were planted, with the heterozygous group helping to decipher the mode of

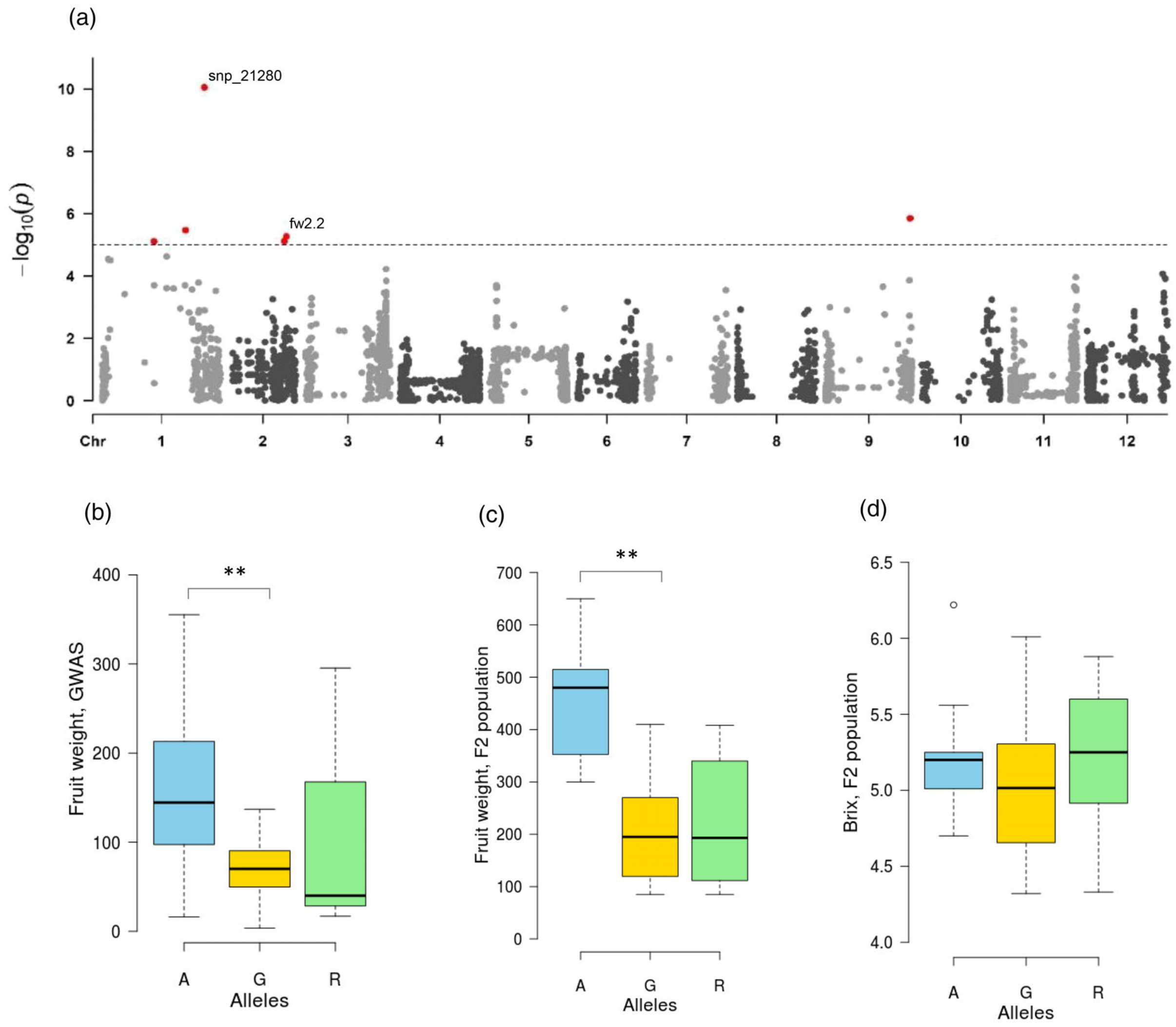

Figure 7. Genome-wide association study of fruit weight and validation using F2 segregated populations.
(a) Manhattan plot for fruit weight using the Brix panel. SNPs with significant associations ($P \leq 0.05$) are colored in red. For significant associations, SNP names are highlighted. (b) Haplotype analysis for novel fruit weight of the SNP marker 'snp_21280' at position 85 022 895 on chromosome 1 across the Brix panel. (c) Cross-validation of the association of fruit weight with the SNP marker 'snp_21280' using the F2 segregated population. (d) Haplotype analysis of fruit weight in the F2 population showed no effect of the marker 'snp_21280' on Brix.
Asterisks indicate significant differences as determined by Student's *t*-test. fw2.2, fruit weight. Letters (A, G, R) are SNPs representing different haplotypes.

inheritance of the desired marker–trait association. All validation results are summarized in Table 1. Brix and fruit weight were the first traits subjected to validation trials because of their importance to breeders. The fruit weight association was validated three times in three independent F2 populations (Figure 7c), with a significant increase in fruit weight resulting from the rare allele, in agreement with the GWAS results. Moreover, similar to the solcap_snp_sl_21280 marker, the F2 populations displayed unaltered Brix values (Figure 7c; Table S6). Finally, there was no significant difference in average fruit weight between the rare allele homozygotes and the heterozygotes, indicating a direct dominant effect of the rare allele on fruit weight (Figure 7b).

The Brix associations were validated by three independent F2 populations for solcap_snp_sl_39725 and one F2 population each for solcap_snp_sl_31961 and solcap_snp_sl_53191 (Figure 6b). The association with solcap_snp_sl_39725 was successfully validated in all populations, revealing a dominant effect of the rare allele. This is not surprising given its identification in previous GWAS studies, albeit being based on the glucose and fructose contents rather than the Brix value *per se* (Tieman et al., 2017), and the well-characterized role of *LIN5* as a determinant of Brix (Fridman et al., 2004; Zanor et al., 2009). The association of solcap_snp_sl_31961 with Brix was also validated despite the GWA being relatively low (Table 1). Interestingly, heterotic behavior was

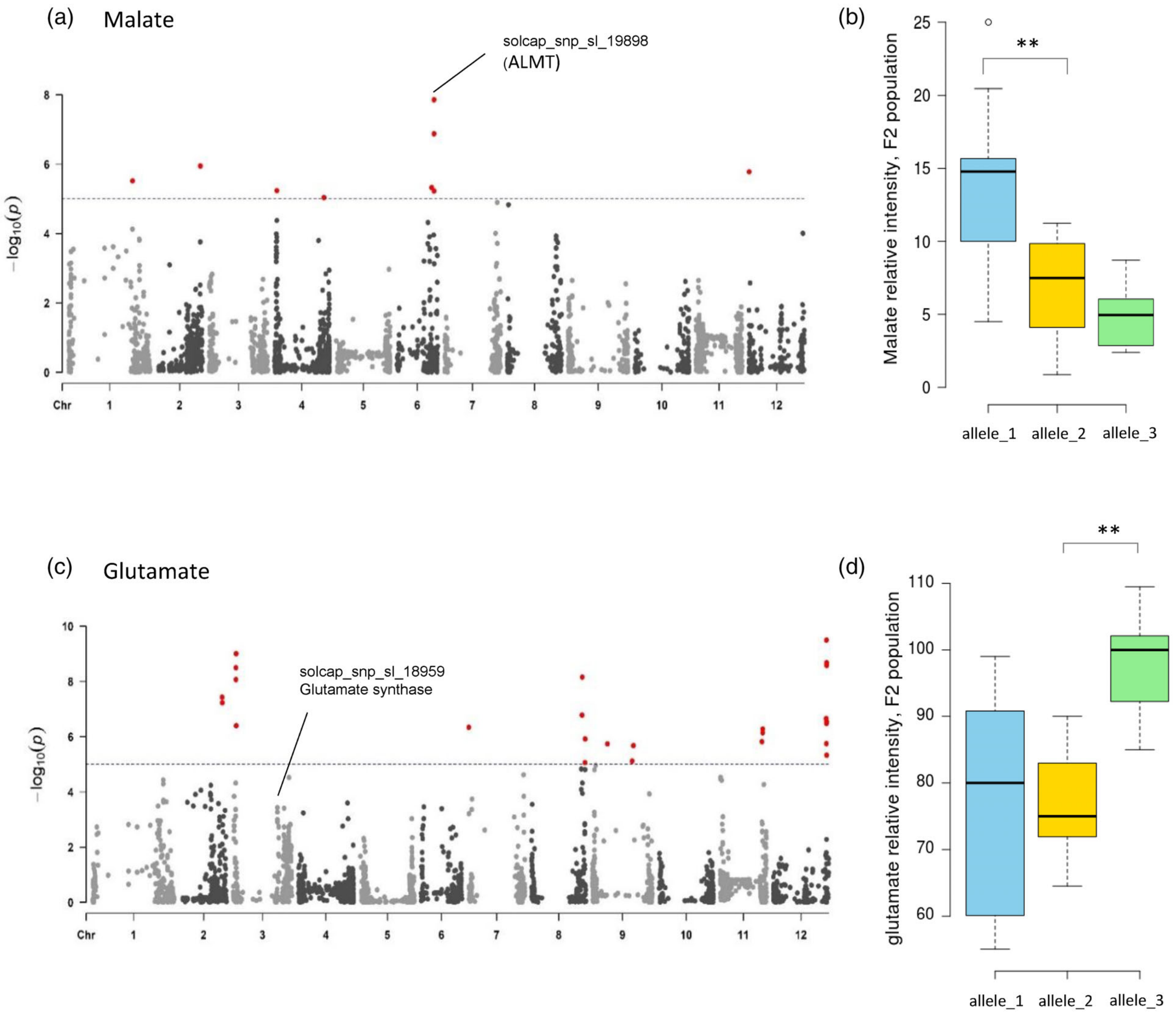

Figure 8. Genome-wide association study of malate and glutamate and cross-validations using F2 segregated populations.
(a) Manhattan plot for malate using the Brix panel. (b) Cross-validation of the malate association with the SNP marker 'solcap_snp_sl_19898' using the F2 segregated population. (c) Manhattan plot for glutamate using the Brix panel. (d) Cross-validation of the glutamate association with the SNP marker 'solcap_snp_sl_18959' using the F2 segregated population.
SNPs with significant associations ($P \leq 0.05$) are colored in red. Asterisks indicate significant differences as determined by Student's *t*-test. ALMT, aluminum-activated malate transporter.

observed in the F2 populations for both Brix and the contents of glucose and fructose. In the Brix panel, a novel association was found for Brix with a region on chromosome 7, where the best SNP (solcap_snp_sl_53191) had an LOD score of 6.42. This QTL was successfully validated using the F2 population (Table 1), whilst haplotype analysis indicated that this SNP was representative of the rare allele but was present at a higher frequency in the Brix panel than in the Resequencing panel. In this study we additionally detected a novel association for fresh weight with a region on the southern arm of chromosome 1 (Table 1 and Figure 7b) and a QTL for Brix associated with the solcap_snp_sl_53191 SNP on chromosome 7; this QTL was successfully validated using the F2 population (Table 1 and Figure 8b), whilst haplotype analysis indicated that this SNP was representative of the rare allele but was present at a higher relative frequency in the Brix panel than in the Resequencing panel.

DISCUSSION

Here we present a large collection of tomato varieties which was established to associate traits with SNPs. The collection database is freely accessible at https://unity.phenome-networks.com, where most accessions have phenotypic and genotypic records associated with them. Indeed, as little as 6 years ago an editorial article in *Nature*

Table 1 Summary of GWAS results and validation using the F2 segregated populations

Trait	SNP	chr	pos (SL2.50)	Mean (common)	Mean (rare)	LOD	effect	common#	rare#	F2	Mean (common)	Mean (heterozygous)	Mean (rare)	Significance	Mode
Fruit weight	solcap_snp_sl _21280	1	85 022 895	0.07	0.16	10.05	140	57	28	7002	0.056	0.067	0.069	0.33	–
										7004	0.042	0.0423	0.045	0.86	–
										7006	0.135	0.155	0.185	0.0026	Recessive
Brix	solcap_snp_sl _39725	9	3 477 979	4.96	6.83	9.37	38	78	9	7012	5.16	5.87	6.61	0.0061	Dominant
										7016	3.644	4.4	4.11	0.0256	Dominant
										7017	4.887	6.185	6	0.0132	Dominant
	solcap_snp_sl _53191	7	63 352 291	5.09	7.15	6.42	40	85	6	7019	8.83	8.08	7.78	0.1	–
Glutamate	solcap_snp_sl _18959	3	53 342 797	1.59	3.39	3.7	113	81	12	7025	55 042	63 047	53 480	0.14	–
										7027	50 543	46 855	41 306	0.32	
										7029	76 379	78 127	91 467	0.0068	–
Malate	solcap_snp_sl _19898	6	44 955 590	1.53	3.46	7	128	71	22	7031	13 712	7225	4838	0.0022	Common is recessive
Several metabolic traits (e.g., glucose, fructose)	solcap_snp_sl _31961	12	63 772 969	1.35	3.1	7.44	128	86	7	7035	7.1	8.68	6.6	0.0137	Heterotic increase
										7037	35 476	44 995	37 809	0.31	
Aspartate	solcap_snp_sl _56404	8	269 815	0.95	5.47	19.52	501	91	5	–	–	–	–	–	–
Proline	solcap_snp_sl _55681	12	64 182 643	2.99	16.95	9.46	476	89	10	7041	781.66	1441.7	1718	0.1264	–
										7043	5378.46	3338.38	3539	0.283	–

Genetics stated that '*not all crop data can be stored at germplasm centers, nor is there yet funding for detailed phenotypic data curation and structured databasing in one place*' (Anon, 2015). The tomato collection does just this with the deposition of large datasets based on both yield-associated traits collected over several years and visual descriptions, metabolomic data, and tools for analytical statistics in one platform. It is planned for the dataset to be continually updated as new accessions and other lines (e.g., introgression lines) are added. In addition to new genotypes, the phenotypic data will be expanded in new dimensions with the application of novel phenotyping technologies.

We believe that this resource will be of high value given that both QTL and GWAS mapping approaches have been enthusiastically embraced by tomato researchers (Leong et al., 2019; Liu et al., 2003; Szymański et al., 2020; Tanksley, 2004; Tieman et al., 2017; Ye et al., 2017; Zhu et al., 2018). The sub-panels, defined in Figure 1, were designed to deal with large variation of specific traits, and within this article we illustrated their utility by using the Brix panel as a case study. Intriguingly, the strategy of taking large GWAS panels, screening them on the basis of total soluble solids, and selecting a subset of lines (e.g., 109 in our Brix panel) covering the phenotypic variance to be regrown and phenotyped allowed us to identify novel loci for both Brix and fruit weight and confirmed previously known associations. For the Brix value, several QTLs have been identified, with that of the cell wall invertase gene *LIN5* being the best defined to date. Detailed genetic and biochemical studies revealed that a single quantitative trait nucleotide in the gene encoding this enzyme determined variance in total soluble solid content and that this was caused by modification in the sugar binding capacity of the enzyme (Fridman et al., 2004). In addition, the developmental regulator SELF PRUNING was also demonstrated to strongly influence Brix (Fridman et al., 2002), but of these, only LIN5 was identified to be associated with either total soluble solid or sugar content in previous GWAS (Sauvage et al., 2014; Tieman et al., 2017). In the current study we found three novel associations and validated one of these in F2 segregating populations (Figure 6). In addition to the detection of the two genes *fw2.2* (Frary et al., 2000) and *fasciated* (Yuste-Lisbona et al., 2020), which are well known to be associated with fruit weight, we also identified (though only in the reduced Brix panel) a previously unknown locus on chromosome 1 that is associated with this agronomically important trait. One of the candidates of interest at this locus encodes a pentatricopeptide repeat-containing protein that is of potential interest given that it has been demonstrated to underlie kernel yield in maize (*Zea mays*) (Huang et al., 2020).

The two traits discussed above, Brix (the total soluble solid content) and fruit weight, determine the agronomic yield of tomatoes that are processed, as well as constituting quality and yield traits of fresh salad varieties of tomato. The approach used here – that of pre-selecting a panel of accessions with a selected trait to break the strong negative correlation between yield and Brix (Schauer et al., 2006; Semel et al., 2006) – represents a powerful strategy to identify breeding material that allows to enhance the soluble solid content without conferring a yield penalty. Whether or not it is possible to separate these two traits has been a subject of debate for decades, but a recent mathematical modeling approach strongly advocates that it should be possible (Chen et al., 2021). This has been confirmed experimentally by an elegant study of Yamamoto et al., who demonstrated that breeding for both increased yield and increased Brix simultaneously is possible via successive rounds of recurrent selection (Yamamoto et al., 2016). The results presented here suggest that, provided a large enough population is used, these traits could be co-selected in a single generation, considerably reducing the time needed to produce elite lines compared to the recurrent selection approach. In addition to these traits, we also provide a large dataset of the genetic architecture of primary metabolite composition across the Brix panel, demonstrating that the pre-selection strategy also results in the identification of novel associations with the metabolites known to contribute considerably to the composite Brix trait. However, several of these are yet to be validated. This fact notwithstanding, the comparison of our results from the Brix panel with those obtained previously (Tieman et al., 2017; Ye et al., 2017) allows us to demonstrate that the differences observed are predominantly genetically encoded. Currently, we have not identified the mechanisms by which Brix can be increased without reducing yield or associated traits. However, careful analyses of the associated datasets should allow for the identification of haplotypes or combinations of haplotypes within a gene or genes which are responsible for this. Once such markers are identified it would be relatively trivial to either perform classical haplotype analyses of both traits or confirm these findings by the use of CRISPR-based prime editing approaches.

In summary, here we introduce a tomato seed collection of over 7900 lines and provide access to agronomic traits collected from several experiments. We additionally outline a strategy for phenotypic trait-driven selection of the genotypes to be included in GWAS study as a method that allows the identification of novel loci underlying variance in this trait. For this purpose we evaluated two of the sub-panels of the tomato core collection, namely the Resequencing panel, which was previously established with the goal of being representative of the genetic diversity of tomato (Tieman et al., 2017; Ye et al., 2017), and the Brix panel, which we established during the course of this work. In evaluating these two panels we were able to demonstrate that for both Brix and fruit weight – two of the most important agronomic traits in tomato – novel loci could be uncovered in the Brix panel that were not found when carrying out a GWAS on the Resequencing panel (either in this study or in previous work). Not only could we demonstrate the power of the approach in this manner, we also demonstrated that the negative correlation between these traits could be broken – effectively rendering it facile to breed for high Brix and increased fruit weight simultaneously.

EXPERIMENTAL PROCEDURES

Plant materials

Accession names are provided as CCXXXX, where XXXX represents a four-digit number. Accessions were contributed by several genebanks, heirloom collectors, breeding companies, and researchers. Maximum passport data are documented, including donor information, germplasm status, accession names and numbers, phenotypic data or descriptions, collection date, and geographic collection details such as site, city, and country.

Crosses were carried out in order to create segregating F2 populations and for the allelic tests. Due to the high levels of homozygosity, all crosses were performed on a plot rather than a single plant basis. Under the assumption that there was no effect of reciprocal crosses the selection of males/females was carried out technically. For female plots flowers were emasculated using tweezers 2 days before estimated anthesis – for this purpose all floral organs with the exception of the style were gently removed. For male plots tomato pollen were collected when the plant developed its second/third inflorescence. A minimum of 30 flowers were collected and anthers were isolated in 30 ml of *n*-heptane, hand-shaken for 10 sec, dried, and stored in the freezer. Finally, the desired pollen was placed on the emasculated female stigma when the average temperature was between 20°C and 26°C. F1 seeds were extracted from red ripe fruits.

Plant cultivation

Seeds were sown in Hishtil nurseries in Ashkelon, Israel. Seeds were sown in Hishtil trays and grown for 25–30 days in the nursery's greenhouse prior to transplantation either in the Western Galilee experimental farm near Akko under open field conditions or in Hatzav greenhouses of Menashe Baranes. All agricultural practices including irrigation, fertilization, lateral shoots pruning, crop protection, and pest management were carried out according to standard protocols.

DNA extraction and analysis

Five to eight young leaflets were collected in a 1.5-ml tube and stored at −80°C prior to extraction via the DNA miniprep protocol (Fulton et al., 1995). PCR was carried out with 34 amplification cycles each including 20 sec of denaturation at 94°C, 30 sec of annealing (at a primer-dependent temperature), and 75 sec of elongation at 72°C. Following all cycles the reaction was inactivated at 94°C. Each well contained 5 μl of DNA template, 8.4 μl of double distilled material, 2.5 μl reaction buffer (2 μl dNTPs [10 μM]), 3 μl $MgCl_2$, 2 μl of each primer [10 μM], and 0.125 μM Taq polymerase (Solis Biodyne).

DNA digestion was conducted using commercial restriction enzymes (New England Biolabs Inc.). DNA fragments were separated via agarose gel electrophoresis using standard protocols. DNA sequencing was performed by The Center for Genomic Technologies, The Hebrew University of Jerusalem, Givat Ram.

Accessions which had been subjected to genotyping by sequencing were crossed with various inbred lines originating from breeding material generated in the Zamir lab and with wild-type accessions. Selfed progeny of each F1 hybrid (F2 seeds) was collected and stored with the rest of the collection. For each accession there are F2 seeds derived from the crosses to approximately eight different parental lines. Statistically, one of those populations will segregate for a specific SNP, and then it will be useful for validation.

Genotyping

SNP data were generated for the whole collection in three phases: (i) in 2009, 2957 accessions were genotyped for 17 SNPs via KeyGene SNPWave analysis (van Eijk et al., 2004), (ii) in 2010, a subset of 1205 representative accessions were genotyped for 384 SNP markers at Keygene, the Netherlands using Illumina BeadXpress technology, and finally (iii) 231 genotypes representing the greatest genotypic variance were hybridized to the SolCAP 10K SNP-Chip (Sim et al., 2012). In 2015, 650 accessions (Resequencing panel) were genotyped by sequencing (GBS) at Cornell University following an established protocol (Elshire et al., 2011) which yielded 16k SNP markers.

The Brix panel accessions were genotyped separately via use of the 10K SNP-Chip; however, GBS data are also available since the Brix panel accessions were all included within the Resequencing panel. All marker data are deposited in the Phenome Networks database (http://unity.phenome-networks.com). Primers used in this study are provided in Table S7.

Phenotyping

Phenotyping was performed across all growing seasons (2007–2010, 2015, and 2016) by lab members and assistants. Minimal qualitative descriptors were developed for 27 traits that were believed to be driven by a single gene and three further quantitative traits, namely fruit weight, Brix, and locule number, were also measured. The quantitative traits were generally scored by visual inspection while fruit weight and Brix were measured on at least 10 fruits by a digital scale and digital refractometry, respectively. In addition to phenotyping, each accession was photographed for its fruit appearance and other unique traits. More complete information is available at the Phenome Networks website (https://unity.phenome-networks.com).

Metabolic profiling

For the Brix panel, pericarp tissue was isolated from ripe fruits of the accessions in summer 2009 and winter 2009 harvests, snap frozen in liquid nitrogen, and stored at −80°C prior to extraction (Schauer et al., 2006) and GC-MS analysis using an established protocol (Alseekh et al., 2021; Lisec et al., 2006).

Correlation network

In each of the Brix panel experiments, 55 metabolites were measured. These values were normalized by dividing the mean value of each accession by the mean value of Heinz 1706, which served as a control in each set of GC-MS runs. Then, the results from the two harvests were averaged in order to develop a single correlation network. Pairwise correlation coefficients (ρ) were calculated using Spearman rank analysis and tested using two-way tests (ρ = 0.05).

Population stratification

The tomato collection was evaluated for its genetic variation and the presence of sub-populations using STRUCTURE software (version 2.3.4; Pritchard et al., 2000). In total, 1205 individuals from the collection, which were genotyped via the set of 384 SNP markers, were analyzed. In order to estimate the number of sub-populations (K), STRUCTURE was configured from $K = 1$ to $K = 10$ with a burn-in period of 50k and a Markov chain Monte Carlo (MCMC) method with a period of 100k was chosen with 10 iterations per K. These results were uploaded to STRUCTURE HARVESTER (Earl & von Holdt, 2012) and ΔK values were highest when $K = 2$. For this reason, an additional run for $K = 2$ was performed using the same burn-in length and MCMC period.

Phylogenetic analysis

Sequence analyses were performed using sequencher software 4.9 (Gene Codes Corporation, Ann Arbor, MI, USA; http://www.genecodes.com) and Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo). Phylogenetic trees were built with Interactive Tree Of Life (ITOL) (https://itol.embl.de/itol.cgi).

Linkage disequilibrium

To calculate linkage disequilibrium between SNPs tested in this study as well as the rest of the genome, the LD functions FullMatrix and SitebyAll of TASSEL software were used (Bradbury et al., 2007).

Genome-wide association mapping

The 109 Brix panel accessions were genotyped for the SolCAP tomato SNP-Chip and also for the 16 000 GBS markers. Association analysis was conducted through two statistical approaches each: the R package GAPIT (Lipka et al., 2012) using the compressed mixed linear model (CMLM) and the Phenome association tool based on the R package GenABEL (Aulchenko et al., 2007). Combining the results of both platforms, only common significant marker–trait associations ($-\log(P) > 7$) with MAF > 0.05 were selected for further study. For validation trials, associations related to Brix, fruit weight, and the metabolites important for breeding such as malate, glutamate, and proline were selected. Phenotypic effect represents the percentage additive phenotypic potential. It was calculated by dividing the mean phenotypic value of the common allele by the mean phenotypic value of minor the allele and subtracting 1: [(Minor allele mean/Common allele mean) × 100%] − 100%.

Statistics

Statistical analyses were performed using JMP12 (SAS institute, Cary, NC, USA), Microsoft Excel 2007, and RStudio (version 1.0.136) based on R statistical software version 3.2.2.

ACKNOWLEDGMENTS

SA and ARF acknowledge funding of the PlantaSYST project by the European Union's Horizon 2020 Research and Innovation Programme (SGA-CSA no. 664621 and no. 739582 under FPA no. 664620). Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

To access the data generated in this study, please go to https://unity.phenome-networks.com and log in.

Username: guest@huji.com.
Password: 123456

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Carotenoid profiling of red and yellow sectors of accession CC0383.

Red sectors accumulate lycopene, while yellow sectors show low levels of carotenoids. Accession CC0383 has a similar phenotype and the same mutation as accession Rumi Banjan (CC4418).

Figure S2. Population stratification and linkage disequilibrium in the Brix panel.

(a) Brix panel structure analysis. The peak of ΔK at $K = 7$ is indicated for three populations and four sub-populations. (b) Kinship plot of 109 Brix panel accessions. (c) Decay of LD over distance. The red line is the moving average of 10 nearby markers.

Figure S3. Genome-wide association study of aspartate.

(a) Manhattan plot describing the associations. (b) LD plot describing the high linkage between markers.

Figure S4. Genome-wide association study and F2 validation for metabolic traits mapped to SNP solcap_snp_sl_31961.

(a–c) Manhattan plots for fructose, glucose, and sucrose using the Brix panel. (d–f) Validation results showing the effects in the F2 segregated population on fructose, glucose, and Brix. Letters (C, T, Y) indicate SNPs representing different haplotypes.

Table S1. Current status and passport data of the collection of 7979 tomato accessions.

Table S2. Phenotypic trait descriptor categories used to characterized the tomato collection.

Table S3. The yellow/red *bicolor* characteristic found in the tomato collection.

Table S4. Phenotypic and metabolic traits of the Brix panel in different experiments.

Table S5. List of significant associations for yield-associated and metabolic traits based on the Brix panel.

Table S6. List of parents used to generate the F2 segregated populations for the GWAS validation.

Table S7. Primers sequences used in this study.

REFERENCES

Alseekh, S., Aharoni, A., Brotman, Y., Contrepois, K., D'Auria, J., Ewald, J.C. ***et al.*** (2021) Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nature Methods*, **18**, 747–756.

Anon. (2015) Growing access to phenotype data. *Nature Genetics*, **47**, 99.

Aulchenko, Y.S., Ripke, S., Isaacs, A. & van Duijn, C.M. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics (Oxford, England)*, **23**, 1294–1296.

Barrantes, W., López-Casado, G., García-Martínez, S., Alonso, A., Rubio, F., Ruiz, J.J. ***et al.*** (2016) Exploring new alleles involved in tomato fruit quality in an introgression line library of Solanum pimpinellifolium. *Frontiers in Plant Science*, **7**, 1172.

Barry, C.S., McQuinn, R.P., Chung, M.Y., Besuden, A. & Giovannoni, J.J. (2008) Amino acid substitutions in homologs of the STAY-GREEN protein are responsible for the green-flesh and chlorophyll retainer mutations of tomato and pepper. *Plant Physiology*, **147**, 179–187.

Bombarely, A., **Menda, N.**, **Tecle, I.Y.**, **Buels, R.M.**, **Strickler, S.**, **Fischer-York, T.** ***et al.*** (2011) The sol genomics network (solgenomics.Net): growing tomatoes using Perl. *Nucleic Acids Research*, **39**, D1149–D1155.

Börner, A. (2012) Nickolai Ivanovich Vavilov and his footprint on plant genetic resources conservation in Gemany. *SEL'SKOKHOZYAISTVENNAYA BIOLOGIA*, **5**, 20–30.

Bradbury, P.J., **Zhang, Z.**, **Kroon, D.E.**, **Casstevens, T.M.**, **Ramdoss, Y.** & **Buckler, E.S.** (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics (Oxford, England)*, **23**, 2633–2635.

Chakrabarti, M., **Zhang, N.**, **Sauvage, C.**, **Muños, S.**, **Blanca, J.**, **Cañizares, J.** ***et al.*** (2013) A cytochrome P450 regulates a domestication trait in cultivated tomato. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 17125–17130.

Chen, J., **Beauvoit, B.**, **Génard, M.**, **Colombié, S.**, **Moing, A.**, **Vercambre, G.** ***et al.*** (2021) Modelling predicts tomatoes can be bigger and sweeter if biophysical factors and transmembrane transports are fine-tuned during fruit development. *The New Phytologist*, **230**, 1489–1502.

Do, P.T., **Prudent, M.**, **Sulpice, R.**, **Causse, M.** & **Fernie, A.R.** (2010) The influence of fruit load on the tomato pericarp metabolome in a Solanum chmielewskii introgression line population. *Plant Physiology*, **154**, 1128–1142.

Earl, D.A. & **von Holdt, B.M.** (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Elshire, R.J., **Glaubitz, J.C.**, **Sun, Q.**, **Poland, J.A.**, **Kawamoto, K.**, **Buckler, E.S.** ***et al.*** (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, **6**, e19379.

Frary, A., **Nesbitt, T.C.**, **Grandillo, S.**, **Knaap, E.**, **Cong, B.**, **Liu, J.** ***et al.*** (2000) fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science (New York, N.Y.)*, **289**, 85–88.

Fridman, E., **Carrari, F.**, **Liu, Y.S.**, **Fernie, A.R.** & **Zamir, D.** (2004) Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science (New York, N.Y.)*, **305**, 1786–1789.

Fridman, E., **Liu, Y.**, **Carmel-Goren, L.**, **Gur, A.**, **Shoresh, M.**, **Pleban, T.** ***et al.*** (2002) Two tightly linked QTLs modify tomato sugar content via different physiological pathways. *Molecular Genetics and Genomics*, **266**, 821–826.

Fulton, T.M., **Chuunwongse, J.** & **Tanksley, S.D.** (1995) Microprep protocol for extraction of DNA from tomato and other herbaceaous plants. *Plant Molecular Biology Reporter*, **13**, 207–209.

Huang, J., **Lu, G.**, **Liu, L.**, **Raihan, M.S.**, **Xu, J.**, **Jian, L.** ***et al.*** (2020) The kernel size-related quantitative trait locus qKW9 encodes a Pentatricopeptide repeat protein that Aaffects photosynthesis and grain filling. *Plant Physiology*, **183**, 1696–1709.

Isaacson, T., **Ronen, G.**, **Zamir, D.** & **Hirschberg, J.** (2002) Cloning of tangerine from tomato reveals a carotenoid isomerase essential for the production of beta-carotene and xanthophylls in plants. *The Plant Cell*, **14**, 333–342.

Krieger, U., **Lippman, Z.B.** & **Zamir, D.** (2010) The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nature Genetics*, **42**, 459–463.

Leong, B.J., **Lybrand, D.B.**, **Lou, Y.R.**, **Fan, P.**, **Schilmiller, A.L.** & **Last, R.L.** (2019) Evolution of metabolic novelty: a trichome-expressed invertase creates specialized metabolic diversity in wild tomato. *Science Advances*, **5**, eaaw3754.

Letunic, I. & **Bork, P.** (2007) Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics (Oxford, England)*, **23**, 127–128.

Lipka, A.E., **Tian, F.**, **Wang, Q.**, **Peiffer, J.**, **Li, M.**, **Bradbury, P.J.** ***et al.*** (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics (Oxford, England)*, **28**, 2397–2399.

Lippman, Z.B., **Cohen, O.**, **Alvarez, J.P.**, **Abu-Abied, M.**, **Pekker, I.**, **Paran, I.** ***et al.*** (2008) The making of a compound inflorescence in tomato and related nightshades. *PLoS Biology*, **6**, e288.

Lisec, J., **Schauer, N.**, **Kopka, J.**, **Willmitzer, L.** & **Fernie, A.R.** (2006) Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nature Protocols*, **1**, 387–396.

Liu, Y.S., **Gur, A.**, **Ronen, G.**, **Causse, M.**, **Damidaux, R.**, **Buret, M.** ***et al.*** (2003) There is more to tomato fruit colour than candidate carotenoid genes. *Plant Biotechnology Journal*, **1**, 195–207.

Mackay, I. & **Powell, W.** (2007) Methods for linkage disequilibrium mapping in crops. *Trends in Plant Science*, **12**, 57–63.

Magwaza, L.S. & **Opara, U.L.** (2015) Analytical methods for determination of sugars and sweetness of horticultural products - a review. *Scientia Horticulturae*, **184**, 179–192.

Mu, Q., **Huang, Z.**, **Chakrabarti, M.**, **Illa-Berenguer, E.**, **Liu, X.**, **Wang, Y.** ***et al.*** (2017) Fruit weight is controlled by cell size regulator encoding a novel protein that is expressed in maturing tomato fruits. *PLoS Genetics*, **13**, e1006930.

Park, S.J., **Jiang, K.**, **Tal, L.**, **Yichie, Y.**, **Gar, O.**, **Zamir, D.** ***et al.*** (2014) Optimization of crop productivity in tomato using induced mutations in the florigen pathway. *Nature Genetics*, **46**, 1337–1342.

Pritchard, J.K., **Stephens, M.** & **Donnelly, P.** (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Prudent, M., **Causse, M.**, **Génard, M.**, **Tripodi, P.**, **Grandillo, S.** & **Bertin, N.** (2009) Genetic and physiological analysis of tomato fruit weight and composition: influence of carbon availability on QTL detection. *Journal of Experimental Botany*, **60**, 923–937.

Rodríguez-Leal, D., **Lemmon, Z.H.**, **Man, J.**, **Bartlett, M.E.** & **Lippman, Z.B.** (2017) Engineering quantitative trait variation for crop improvement by genome editing. *Cell*, **171**, 470–480.e478.

Ronen, G., **Carmel-Goren, L.**, **Zamir, D.** & **Hirschberg, J.** (2000) An alternative pathway to beta -carotene formation in plant chromoplasts discovered by map-based cloning of beta and old-gold color mutations in tomato. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 11102–11107.

Sauvage, C., **Segura, V.**, **Bauchet, G.**, **Stevens, R.**, **Do, P.T.**, **Nikoloski, Z.** ***et al.*** (2014) Genome-wide Association in Tomato Reveals 44 candidate loci for fruit metabolic traits. *Plant Physiology*, **165**, 1120–1132.

Schauer, N., **Semel, Y.**, **Roessner, U.**, **Gur, A.**, **Balbo, I.**, **Carrari, F.** ***et al.*** (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nature Biotechnology*, **24**, 447–454.

Semel, Y., **Nissenbaum, J.**, **Menda, N.**, **Zinder, M.**, **Krieger, U.**, **Issman, N.** ***et al.*** (2006) Overdominant quantitative trait loci for yield and fitness in tomato. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 12981–12986.

Sim, S.C., **Durstewitz, G.**, **Plieske, J.**, **Wieseke, R.**, **Ganal, M.W.**, **Van Deynze, A.** ***et al.*** (2012) Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PLoS One*, **7**, e40563.

Soyk, S., **Lemmon, Z.H.**, **Oved, M.**, **Fisher, J.**, **Liberatore, K.L.**, **Park, S.J.** ***et al.*** (2017) Bypassing negative epistasis on yield in tomato imposed by a domestication gene. *Cell*, **169**, 1142–1155.e1112.

Soyk, S., **Lemmon, Z.H.**, **Sedlazeck, F.J.**, **Jiménez-Gómez, J.M.**, **Alonge, M.**, **Hutton, S.F.** ***et al.*** (2019) Duplication of a domestication locus neutralized a cryptic variant that caused a breeding barrier in tomato. *Nature Plants*, **5**, 471–479.

Szymański, J., **Bocobza, S.**, **Panda, S.**, **Sonawane, P.**, **Cárdenas, P.D.**, **Lashbrooke, J.** ***et al.*** (2020) Analysis of wild tomato introgression lines elucidates the genetic basis of transcriptome and metabolome variation underlying fruit traits and pathogen response. *Nature Genetics*, **52**, 1111–1121.

Tam, V., **Patel, N.**, **Turcotte, M.**, **Bossé, Y.**, **Paré, G.** & **Meyre, D.** (2019) Benefits and limitations of genome-wide association studies. *Nature Reviews. Genetics*, **20**, 467–484.

Tanksley, S.D. (2004) The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. *The Plant Cell*, **16**, S181–S189.

Tieman, D., **Zhu, G.**, **Resende, M.F., Jr.**, **Lin, T.**, **Nguyen, C.**, **Bies, D.** ***et al.*** (2017) A chemical genetic roadmap to improved tomato flavor. *Science (New York, N.Y.)*, **355**, 391–394.

van Eijk, M.J., **Broekhof, J.L.**, **van der Poel, H.J.**, **Hogers, R.C.**, **Schneiders, H.**, **Kamerbeek, J.** ***et al.*** (2004) SNPWave: a flexible multiplexed SNP genotyping technology. *Nucleic Acids Research*, **32**, e47.

Wang, X., **Gao, L.**, **Jiao, C.**, **Stravoravdis, S.**, **Hosmani, P.S.**, **Saha, S.** ***et al.*** (2020) Genome of Solanum pimpinellifolium provides insights into structural variants during tomato breeding. *Nature Communications*, **11**, 5817.

Yamamoto, E., **Matsunaga, H.**, **Onogi, A.**, **Kajiya-Kanegae, H.**, **Minamikawa, M.**, **Suzuki, A.** ***et al.*** (2016) A simulation-based breeding design that uses whole-genome prediction in tomato. *Scientific Reports*, **6**, 19454.

Ye, J., **Li, W.**, **Ai, G.**, **Li, C.**, **Liu, G.**, **Chen, W.** ***et al.*** (2019) Genome-wide association analysis identifies a natural variation in basic helix-loop-helix transcription factor regulating ascorbate biosynthesis via D-mannose/L-galactose pathway in tomato. *PLoS Genetics*, **15**, e1008149.

Ye, J., **Wang, X.**, **Hu, T.**, **Zhang, F.**, **Wang, B.**, **Li, C.** ***et al.*** (2017) An InDel in the promoter of Al-ACTIVATED MALATE TRANSPORTER9 selected during tomato domestication determines fruit malate contents and aluminum tolerance. *The Plant Cell*, **29**, 2249–2268.

Yoshiba, Y., **Kiyosue, T.**, **Nakashima, K.**, **Yamaguchi-Shinozaki, K.** & **Shinozaki, K.** (1997) Regulation of levels of proline as an osmolyte in plants under water stress. *Plant & Cell Physiology*, **38**, 1095–1102.

Young, P.A. (1956) Ry. A modifier gene for red color in yellow tomato fruit. *Report of the - Tomato Genetics Cooperative*, **33**, 6.

Yuste-Lisbona, F.J., **Fernández-Lozano, A.**, **Pineda, B.**, **Bretones, S.**, **Ortíz-Atienza, A.**, **García-Sogo, B.** ***et al.*** (2020) ENO regulates tomato fruit size through the floral meristem development network. *Proceedings of the National Academy of Sciences of the United States of America*, **117**, 8187–8195.

Zanor, M.I., **Osorio, S.**, **Nunes-Nesi, A.**, **Carrari, F.**, **Lohse, M.**, **Usadel, B.** ***et al.*** (2009) RNA interference of LIN5 in tomato confirms its role in controlling brix content, uncovers the influence of sugars on the levels of fruit hormones, and demonstrates the importance of sucrose cleavage for normal fruit development and fertility. *Plant Physiology*, **150**, 1204–1218.

Zhang, N., **Brewer, M.T.** & **van der Knaap, E.** (2012) Fine mapping of fw3.2 controlling fruit weight in tomato. *Theoretical and Applied Genetics*, **125**, 273–284.

Zhu, G., **Wang, S.**, **Huang, Z.**, **Zhang, S.**, **Liao, Q.**, **Zhang, C.** ***et al.*** (2018) Rewiring of the fruit metabolome in tomato breeding. *Cell*, **172**, 249–261.e212.