J. Amer. Soc. Hort. Sci. 146(5):339–345. 2021. https://doi.org/10.21273/JASHS05066-21

DNA Variation in a Diversity Panel of Tomato Genetic Resources

Joanne A. Labate
Plant Genetic Resources Unit, Agricultural Research Service, U.S. Department of Agriculture, 630 W. North St., Geneva, NY 14456

Additional index words. genotyping by sequencing, germplasm, Plant Variety Protection, single nucleotide polymorphism, *Solanum lycopersicum*, *Solanum pimpinellifolium*

Abstract. **A diversity panel of 190 National Plant Germplasm System (NPGS) tomato (*Solanum lycopersicum*) accessions was genotyped using genotyping by sequencing. These originated from 31 countries and included fresh market, ornamental, processing, breeders' lines, landraces, and home gardening types, as well as six different accessions of the economically valuable cultivar San Marzano. Most of the 34,531 discovered single nucleotide polymorphisms were rare and therefore excluded from downstream analyses. A total of 3713 high-quality, mapped single nucleotide polymorphisms that were present in at least two accessions were used to estimate genetic distances and population structure. Results showed that these phenotypically and geographically diverse NPGS tomato accessions were closely related to each other. However, a subset of divergent genotypes was identified that included landraces from primary centers of diversity (South America), secondary centers of diversity (Italy, Taiwan, and France), and genotypes that originated from wild species through 20th century breeding for disease resistance (e.g., 'VFNT Cherry'). Extreme variant accessions produce cultivated fruit traits in a background that contains many wild or primitive genes. These accessions are promising sources of novel genes for continued crop improvement.**

Vegetable producers require an abundance of genetic diversity to remain competitive and meet increasing consumer demands for the beneficial vitamins, minerals, dietary fiber, and other nutrients provided by these crops. Access to germplasm for development of improved cultivars is critical given the narrow genetic base of some crops and risk of susceptibility to both recurrent and newly emerging biotic and abiotic stresses. Tomato (*Solanum lycopersicum*) constituted a $1.7 billion industry in 2020 in the United States [U.S. Department of Agriculture (USDA), National Agricultural Statistics Service, 2021] and a $48 billion industry in 2018 worldwide (Food and Agriculture Organization of the United Nations, 2020). Production systems include open-field and protected environments, including high-tunnel, greenhouse, and hydroponics. Cost of tomato production is relatively high per unit area and demands intensive agronomic management and considerable investment (Kelley et al., 2017).

Tomato is classified into one cultivated and 12 wild species (Peralta et al., 2008). The USDA, Agricultural Research Service (ARS), Plant Genetic Resources Unit (PGRU) conserves 6610 tomato accessions in the form of publicly available seed stocks (USDA, ARS, 2021a). These are distributed throughout the world by the USDA National Plant Germplasm System (NPGS) for purposes of breeding, research, and higher education. According to Genesys (2021) this is the second largest tomato collection in the world, after The World Vegetable Center in Taiwan.

Primary centers of diversity for cultivated tomato are Chile, Ecuador, Peru, and Mexico, with secondary centers throughout the world (Villand et al., 1998). Breeding of cultivated tomato has emphasized crosses with wild relatives due to its narrow genetic base, ease of interspecific crossing, production demands based on growing conditions and market niche, susceptibility to pests and diseases, and vulnerability to abiotic factors. Twentieth century breeding using interspecific crosses has increased the genetic diversity of elite germplasm (Bauchet and Causse, 2012). Recent consumer demand calls for more flavorful, colorful, and heirloom types available year-round. But new commercial cultivars will still require the resistances, shelf life, and durability of standard supermarket choices.

Given the global importance of the crop, large germplasm collections, and advanced genetic tools, there exists a rich literature on the subject of tomato genetic diversity (Bauchet and Causse, 2012; Zhao et al., 2019). Many of these reports have focused on taxonomy, domestication, and selection hypotheses (100 Tomato Genome Sequencing Consortium et al., 2014; Blanca et al., 2012, 2015; Sim et al., 2011, 2015) and trait-marker associations (Bauchet et al., 2017; Mata-Nicolás et al., 2020; Mazzucato et al., 2008; Sauvage et al., 2014). Understanding the patterns of genetic variation among accessions or stocks, the focus of many studies, is critical for breeders and other scientists to make efficient use of diverse germplasm (Cebolla-Cornejo et al., 2013; de los Ángeles Martínez-Vázquez et al., 2017; Hanson and Yang, 2016; Jayakodi et al., 2021; Jin et al., 2019; Kulus, 2018; Mata-Nicolás et al., 2020).

To describe genetic diversity patterns of cultivated tomato germplasm, a phenotypically broad panel of 190 PGRU accessions was assembled for genotyping by sequencing (GBS).

Received for publication 10 Mar. 2021. Accepted for publication 28 Apr. 2021. Published online 18 June 2021.
I acknowledge Susan Srmack and Paul Kisly for providing excellent technical support. This paper is dedicated to the memory of Dr. Larry D. Robertson, my collaborator in the initial phase of this study. U.S. Department of Agriculture is an equal opportunity provider and employer.
The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the U.S. Department of Agriculture or the Agricultural Research Service of any product or service to the exclusion of others that may be suitable.
J.A.L. is the corresponding author. E-mail: joanne.labate@usda.gov.

Types included categories such as home garden, fresh market, vintage, geodiversity, landraces from primary and secondary centers of diversity, fruit shape diversity, ornamental, processing, expired Plant Variety Protection (PVP) lines, lines from private companies and public breeding programs, lines with introgressed disease resistance, eight 'San Marzano' accessions (a widely popular commercial cultivar), and one *Solanum pimpinellifolium* sample. Many of the accessions fit into more than one of these categories.

The major objectives of this study were to 1) discover and score a set of high-quality, mapped single nucleotide polymorphisms (SNPs) in a diversity panel of tomato; 2) estimate heterozygosity, relationships, and population structure of the accessions; and 3) draw conclusions as to which accessions might harbor sources of novel alleles that could be valuable for further crop improvement.

Materials and Methods

Plant material. A total of 190 diverse tomato cultivars (Supplemental Table 1) were sampled from the USDA, ARS, PGRU germplasm collection. The accessions originated from 31 countries in Africa, Asia, Europe, North America, Central America, and South America, and were originally received by NPGS between 1932 (PI 97538, 'Cherry') and 2002 (PI 647566, 'Flora-dade'). Seventeen of the accessions were considered highly unique based on PVP certificates (USDA, Agricultural Marketing Service, 2021) (Supplemental Table 1) as processing, fresh market, or breeding types. 'San Marzano' accessions originated from Argentina, Hungary, Italy (two accessions), South Africa, Spain, and the United States (two accessions). Disease resistance data from NPGS GRIN-Global (USDA, ARS, 2021b) and passport data from NPGS GRIN-Global (USDA, ARS, 2021a), the Genebank Information System of the IPK Gatersleben (Genebank Information System of the IPK Gatersleben, 2021), and web resources, such as Vegetable Cultivar Descriptions for North America (Wehner, 2016), were used to annotate the accessions including their reported disease resistances (Supplemental Table 1).

Genotyping by sequencing. Plants were grown to the seedling stage in a greenhouse in Geneva, NY. One true leaf was sampled per accession and DNA was extracted using the DNeasy 96 plant kit (Qiagen, Hilden, Germany). GBS libraries were prepared at Cornell University's Genomic Diversity Facility using the restriction enzyme PstI and the standard barcode and common adaptor sets (Elshire et al., 2011). GBS was performed using an Illumina HiSEq. 2000 (San Diego, CA). Output files were analyzed using TASSEL 3.0 and TASSEL 4.0 analysis pipelines (Glaubitz et al., 2014) (Supplemental Table 2). Briefly, tags were aligned to *S. lycopersicum* whole genome assembly build 2.50 (The Tomato Genome Consortium, 2012) using Burrows-Wheeler Aligner version 0.7.13 software with default settings (Li and Durbin, 2009). For the DiscoverySNPCallerPlugin, parameters were set to retain rare alleles (i.e., minimum minor allele frequency equaled zero and minimum minor allele count equaled 1). In addition, at least 10% of DNA samples were required to be scored in order for a site to be retained. The software VCFtools version 0.1.12a (Danecek et al., 2011) was used to exclude sites with more than two alleles or an insertion/deletion and to exclude all genotypes with a quality below a threshold of 98. TASSEL 5 GUI (Bradbury et al., 2007) was used to filter out poorly performing sites by requiring at least 10% of genotypes for that SNP to be present. Finally, the minor allele was required to be present in at least two or more independent DNA samples. No imputation (prediction) of missing data were performed. All raw data are available through the National Center for Biotechnology Information sequence read archive as study SRP308407, BioProject PRJN705205, and BioSample numbers SAMN18077515–SAMN18077706.

Statistical analyses. TASSEL 5 GUI (Bradbury et al., 2007) was used to estimate allele frequencies, proportion missing data, heterozygosities, and pairwise genetic distances. The latter was estimated as the proportion of SNPs different between a pair of samples, ignoring missing data. Population groups were inferred using STRUCTURE version 2.3.4 (Pritchard et al., 2000) with 200,000 burn-in iterations followed by 500,000 iterations. Output from three runs for K = 1 to K = 10 clusters were used as input for Structure Harvester core version A.2 (Earl and von Holdt, 2012) to find the maximum rate of change in the log probability of data between successive K values, Delta K (Evanno et al., 2005). DISTRUCT version 1.1 (Rosenberg, 2004) was used to graph STRUCTURE results. To illustrate overall genetic similarity among genotypes, Phylip version 3.69 software package was used to generate a genetic distance-based neighbor-joining (NJ) tree (Saitou and Nei, 1987) based on 100 bootstrap replicates using the procedures outlined in Labate and Robertson (2015). The resultant majority rule consensus tree was drawn using FigTree version 1.4 (Rambaut, 2012).

Results

The accessions in this study were not a conventional core collection (Brown, 1995), but rather, a diversity panel intended to capture genetic variation with an emphasis on germplasm that may have value for breeding (Hardigan et al., 2015). This global sample of germplasm was heavily weighted toward accessions from the United States (n = 80); the second largest number of accessions originated from Italy (n = 37) (Supplemental Table 1). All other countries were represented by fewer than 10 accessions, and three accessions were of unknown geographical origin. Primary centers of diversity included a total of 15 accessions originating from Chile, Ecuador, and Peru, and six accessions from Mexico. The remaining countries were represented by one to three accessions each, except for Canada (n = 8) and Spain (n = 5). To ensure a broad range of morphological diversity, 45 accessions that were used in fruit shape research were part of this panel. These included heirloom and modern lines; landraces from Italy, Latin America, and Spain; and mutant stock 'LA0330' (Rodriguez et al., 2011). A set of 43 of the accessions were categorized as heirlooms by the SolCAP project (SolCap, 2013). Primary, secondary, and countries contiguous with primary centers of diversity were represented by 50 accessions from a geodiversity and landrace set (Baldo et al., 2011; Villand et al., 1998). One ornamental line, several breeding lines, and many commercially important fresh market and processing lines were also sampled. These were developed by private companies, universities, agricultural experiment stations, and the USDA. The popular 'San Marzano' processing cultivar was represented by eight accessions originating from Argentina, Hungary, Italy (n = 2), South Africa, Spain, and the United States (n = 2). The U.S. versions included 'Pink San Marzano' (PI 303775) and 'P.A. Young SV 616C' (PI 279566); the latter

being a line bred for resistance to heat sterility. The single *S. pimpinellifolium* sample LA2102 was collected from El Lucero, Loja, Ecuador, in 1981.

GBS in two 96-plex runs yielded more than 225 × 10^6 and 192 × 10^6 good, barcoded reads per run (Supplemental Table 2). The requirement of tags to appear at least five times across 190 barcoded samples yielded 981,444 tags, of which 793,018 aligned to unique positions in the genome. Initially, 34,531 mapped, high-quality sites were identified. However, many of the minor alleles in this data set were very rare (only one or two observations). Therefore, a minimum requirement of at least three counts for an allele was applied to ensure that at least two DNA samples were scored for all minor alleles. This gave a total of 3713 high-quality mapped SNP sites in 190 DNA samples. Proportion of missing data were 0.33 in this set of 1.41 × 10^6 diploid genotype data points. Average heterozygosity equaled 0.08 and ranged from 0.03 in 'Mataverde' (PI 505317), a cultivar collected in Colombia in 1986 to 0.28 in '422' (PI 128586), a landrace collected in Chile in 1938. Heterozygosity of the single *S. pimpinellifolium* sample equaled 0.19. The maximum genetic distance between the 17,766 pairs of *S. lycopersicum* samples was 0.37 for 'T932' (G 33075) vs. 'LYC449' (G 33061), the minimum was 0.018 for 'San Marzano' (PI 262910) and 'Pomodoro San Marzano-Lampadina' (PI 647487), and the mean genetic distance between all pairs was 0.11 (Supplemental Table 3). Approximately 50% of the pairwise genetic distances were ≤0.10.

Structure Harvester showed the optimal number of clusters was k = 3 for the 190 sampled accessions based on maximum Delta K (Evanno et al., 2005) (Supplemental Fig. 1). Accessions were sorted and plotted based on membership coefficients (Fig. 1, Supplemental Table 4) with *S. pimpinellifolium* placed at the origin of the x-axis with membership coefficients of 0.6663, 0.3333, and 0.0004 in clusters one, two, and three, respectively. An Italian landrace with obovoid fruit shape fell at the opposite extreme of the x-axis with membership coefficients of 0.1298, 0.2593, and 0.6109 in clusters one, two, and three, respectively. The graph showed a steep slope close to the x-origin that indicated a large (>40%) but decreasing proportion of cluster one in 14 samples (sample numbers 650 through 678 in Fig. 1, Supplemental Table 4). The graph was then relatively flat with the predominance of cluster two (mean = 65%). Moving along the x-axis, cluster three gradually increased to its maximum proportion of 0.6109. The NJ tree (Fig. 2) showed similar results with *S. pimpinellifolium* being the most divergent and clustering with the identical group of 12 samples with the omission of number 780 PI 647523 'VFNT Cherry'. The fruits of these 12 samples were either small cherry or grape types, or highly fasciated in their appearance (USDA, ARS, 2021a). Although 'VFNT Cherry' was a divergent genotype in the NJ tree, it did not cluster with *S. pimpinellifolium*. Similarly, the divergent cluster described as traditional Italian landraces in the NJ tree (sample numbers 630, 632, and 640 in Fig. 2, Supplemental Table 4) and close samples 614, 617, 620, 623, and 639 (Italian origin plus the Polish heirloom G 33045 'Opalka') were grouped together at the right end in Fig. 1.

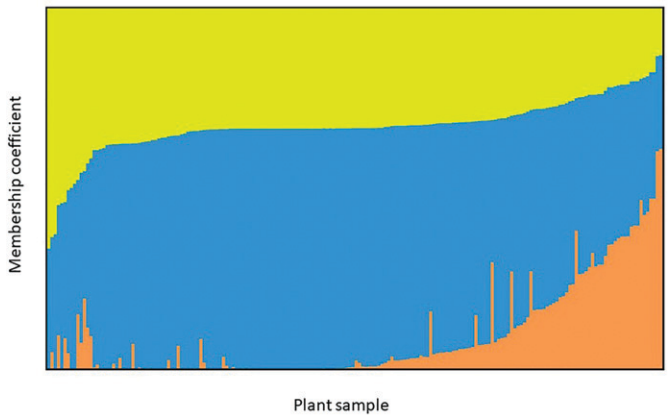

Fig. 1. Population structure of 190 tomato samples for K = 3 groups based on 3713 single nucleotide polymorphisms. Results were sorted with *Solanum pimpinellifolium* at the x-origin. Identities of the 190 accessions along the x-axis are reported in Supplemental Tables 1 and 4.

The NJ tree drawn as a bifurcating cladogram with bootstrap values showed very little support for genetic structure in the sample of 190 accessions with a few exceptions (Supplemental Fig. 2). First, there was 85% bootstrap support for the split between the root of the tree and the majority of accessions. The group at the root consisted of *S. pimpinellifolium* and landraces 646 ('T932' from Italy), 683 ('LA0394' from Peru), and 720 ('W-C 1050' from Ecuador). A group of 15 accessions, which included 12 processing lines and 11 lines categorized as PVP, were supported with 92% bootstrap value (Supplemental Fig. 2). Four of the eight 'San Marzano' accessions clustered together with 95% bootstrap support (785 'San Marzano' from Italy, 790 'Pomodoro San Marzano-Lampadina' from Italy, 786 'San Marzano' from Spain, and 667 'San Marzano' from Argentina) with the other four 'San Marzano' accessions dispersed throughout the NJ tree (Supplemental Fig. 2). The only additional clusters with >90% bootstrap support were four pairs of accessions, namely, {669, 682} both from Ecuador; {651, 762} 'LA0330', 'Burbank'; {606, 778} 'Costoluto Genovese', 'Pomodoro Superselezione di Marmande'; and {694, 767} 'Manalucie', 'Venus' from Florida and North Carolina agricultural experiment stations, respectively.

Discussion

In this study, there were several sources of potential bias in the SNP data. The GBS method will under-call heterozygotes due to low sequencing coverage. Accordingly, the proportion of missing data across the samples in this study was negatively correlated with the proportion of heterozygous sites (r^2 = 0.2119). Additional inherent error associated with GBS is irregular PCR amplification with a bias toward certain alleles or loci. Restriction enzyme biases include mutation in a restriction site preventing digestion, and methylation sensitivity; the latter can introduce bias against intergenic regions (Scheben et al., 2017). However, the results supported several previous observations such as the prevalence of rare alleles, genetic similarity between accessions based on shared pedigrees, and the admixture of certain accessions with *S. pimpinellifolium* (Blanca et al., 2015, see below). Therefore, the first objective to discover and score a set of high-quality, mapped tomato SNPs was met.

A long-recognized problem in developing improved cultivars has been how to efficiently exploit novel variation for traits of interest from gene banks (Corak et al., 2019; Reeves et al., 2012). Recommendations have been made to NPGS to broaden crop core collections to develop sets that are more representative of the whole, and to increase the number of SNPs to facilitate Genome Wide Association Studies (Kuzay et al., 2020). Reeves

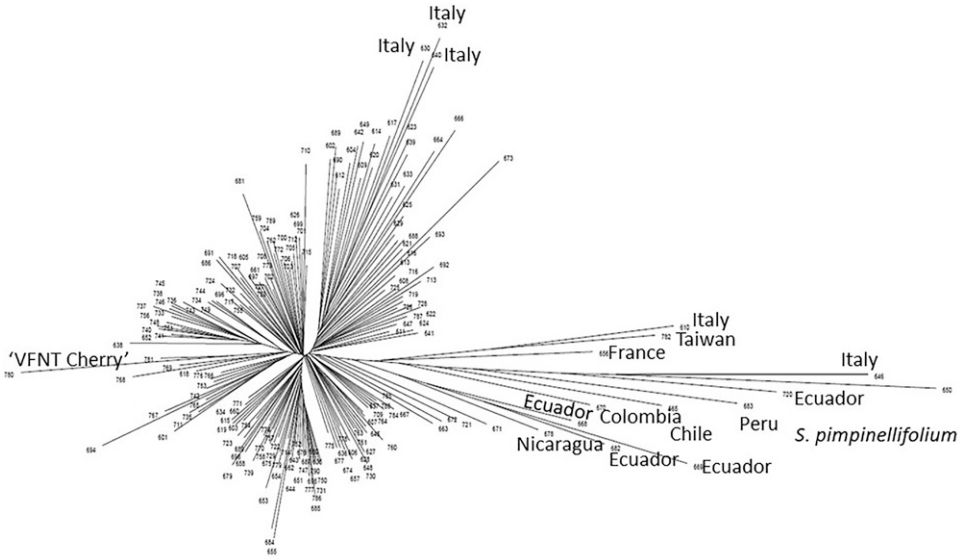

Fig. 2. Neighbor-joining tree of 190 tomato samples based on 3713 single nucleotide polymorphisms. Fifteen genetically divergent accessions are labeled with country of origin; 'VFNT Cherry' contains multiple interspecific introgressions for disease resistances.

et al. (2012) warned against applying neutral marker loci as predictive tools to estimate the diversity of agronomically important loci. This is because natural or artificial selection will shift allele frequencies of the latter away from neutral distributions. This raises the question of how to capture valuable functional diversity from germplasm collections. Geographic origins and genetic relatedness are not reliable predictors of shared traits (Corak et al., 2019, and references therein). By comparing various methods to construct core subsets to maximize phenotypes and minor alleles, an in-depth analysis showed that inclusion of accessions based on random sampling was largely as effective as selecting accessions based on molecular markers (Corak et al., 2019). In contrast, SNPs did support a priori groups in common bean [*Phaseolus vulgaris* (Kuzay et al., 2020)], sorghum [*Sorghum bicolor* (Cuevas and Prom, 2020)], walnut [*Juglans regia* (Bernard et al., 2020)], and 21 potato species [*Solanum* section *Petota* (Hardigan et al., 2015)]. An amplified fragment length polymorphisms (AFLP) core set captured most desirable traits in wild potato species *Solanum microdontum* (Bamberg and del Rio, 2014) and *Solanum demissum* (del Rio and Bamberg, 2020).

In the present study, SNP markers were not good predictors of phenotypes or geographic origins. Despite the range of fruit diversity and country of origin, pairwise genetic distances were low. This is understandable because differences across a small number of loci can impart large differences in fruit morphology (Snouffer et al., 2020). Geographic origin was applied successfully to develop core subsets of potato landraces (*Solanum tuberosum*) held at the International Potato Center (CIP) (Gopal et al., 2013; Huamán et al., 2000); however, tomato germplasm can disperse rapidly on a global scale due to its popularity and broad adaptation, and the concept of geographic origin becomes obfuscated. Surprisingly, the two most distantly related accessions in this study (excluding *S. pimpinellifolium*) were both described as landraces originating from Italy. The mean pairwise genetic distance of 0.11 in this tomato panel was similar to the value of ≤0.10, which was considered to be a threshold for redundancy in a sorghum core collection (Cuevas and Prom, 2020). Redundant samples of 'San Marzano' were more closely related than this by an order of magnitude. Clearly, documentation of phenotypic traits and their underlying genetic architecture should receive high priority in this and other tomato germplasm collections.

A challenge for plant genetic resource managers and breeders is to anticipate target phenotypes for crop improvement. Increased pest, pathogen, and weed pressures, temperature and rainfall fluctuations, drought, salinization of soils, and human encroachment of agricultural lands are all seen as impending threats to crop production (Ceccarelli and Grando, 2020). Whatever breeding strategies are used, the understanding of natural genetic variation and its partitioning remains crucial for success (Ebert, 2020).

For the second objective of this study, namely, to estimate relationships and population structure, overall results showed a lack of clear-cut genetic distinctions among types; fresh market, ornamental, processing, breeders' lines, landraces, and home gardening types were not genetically distinct as categories. Rather, the NJ tree showed many short branches connected at the internal node (star-like tree) which is the signal of a founder effect. This supported longstanding observations of the relatively low genetic diversity within cultivated tomato due to its mating system and natural history (Miller and Tanksley, 1990; Rick and Fobes, 1975). This lack of diversity has led to an emphasis in using wild species relatives for crop improvement, especially for early sources of disease resistance (Rick, 1982).

The third major objective of this study was to draw conclusions as to which accessions might harbor sources of novel alleles that could be valuable for further crop improvement. Three Italian landraces, 'VFNT Cherry', and a dozen cherry or highly fasciated types, showed genetic divergence from most accessions. These small islands of diversity were not predicted based on passport data, usage type, or phenotype. Several of these accessions were classified as "mixture" between *S. lycopersicum* and *S. pimpinellifolium* by Blanca et al. (2015). The authors discussed the possibility of natural gene flow between northern Ecuadorian *S. pimpinellifolium* and Ecuadorian *S. lycopersicum* cherry tomato due to dispersal by animals or humans. Large flattened, fasciated fruit resulting from the fusion of multiple ovaries (example shown in Supplemental Fig. 3) may have been a consequence of interspecific incompatibility, and this fortuitously generated a larger fruited phenotype for humans to select. Several of the accessions in the current study supported this scenario. For example, PI 129026, PI 129033, and PI 129084 were all collected in 1938 and have either retained some of the diversity of *S. pimpinellifolium* based on shared ancestry or originated from recent interspecific hybridization in nature. This implies that NPGS accessions that were collected during the same expedition, originally 287 *S. lycopersicum* accessions collected by H.L. Blood (USDA, ARS, 2021a) could also be sources of diverse alleles. The finding of extensive novel genomic

and morphological variation in cherry tomato accessions from Ecuador and Peru suggested their untapped potential as breeding material (Mata-Nicolás et al., 2020). In comparison with the Heinz 1706 reference genome, 1.2 to 1.9 million mutations were identified in four cherry tomato accessions (Gramazio et al., 2020). The complex genetic relationships among large-fruited cultivated tomatoes, *S. pimpinellifolium,* and cherry tomato also supported the assertion that unmined alleles for crop improvement are available from the latter two taxa, which will easily cross-fertilize with modern cultivars (Razifard et al., 2020).

Other divergent accessions in this study, such as 'AVRDC #6', which was collected in Taiwan from Tainan Agricultural Experiment Station in 2001, and 'VFNT Cherry' with a pedigree tracing to the University of California breeding program, are better explained by intentional interspecific breeding. Variation found in nature and residual linkage drag from cultivar development can both provide potential sources of new alleles (Labate and Robertson, 2012).

Fifteen accessions formed a cluster of which 12 were known processing types, whereas many additional accessions classified as processing fell outside of this cluster and were broadly scattered throughout the cladogram. The cluster of 15 closely related accessions included 10 PVP cultivars that shared UC-82 in their pedigrees. This demonstrated that the novel horticultural traits of PVP cultivars are not good indicators of their underlying genetic backgrounds, and the traits likely arise from a particular novel combination of a few common alleles. In the absence of detailed pedigree data, other passport information, such as the year collected or the year developed, could be useful to identify potential sources of diverse germplasm in processing types.

The authentic 'San Marzano' cultivar originated as a local Italian cultivar and is highly renowned for its commercial significance (Loiudice et al., 1995). Although our study did not attempt to verify a true 'San Marzano' based on known molecular markers (Caramante et al., 2009; Rao et al., 2006), four of the accessions were strong candidates to be representative of the original genotype. This set included two Italian accessions, one donated from an experiment station and the other from a seed company, as well as two accessions collected in 1938 in Argentina and 1960 from Spain, respectively. The provenances and close relatedness among these accessions suggest that this genotype is authentic. The other four versions of 'San Marzano' in this study (from the United States, Hungary, and South Africa) are likely to be somewhat similar in phenotype but not representative in their genetic background. Because tomato seed is efficient to store due to its small size, and is long-lived under repository conditions, there is no impetus to eliminate this small degree of redundancy in name or genotype of an accession. However, such genotypic details should be made readily apparent for users deciding on which accessions to request from the repository. These principles can be applied to similar occurrences in the NPGS tomato collection as well as other crops with small, long-lived seed.

Conclusions

GBS has proven to be a valuable technique to apply to crop germplasm collections, with or without a reference genome (Egea et al., 2017; Lee et al., 2020; Niu et al., 2019; Pavan et al., 2019; Shi et al., 2017). SNP genotypes provide a robust foundation with which to interpret and complement passport, pedigree, and phenotypic trait data. Rather than using molecular markers as a basis with which to assemble a core subset (van Hintum et al., 2000), the PGRU tomato diversity panel was developed based on passport data and subsequently analyzed with molecular markers. This showed that the genotypes were not largely associated with a priori groupings.

A recent study (Alfonso and Bamberg, 2020) demonstrated improved discriminating power among wild potato (*Solanum fendleri*) populations by use of adaptive loci identified using F_{ST} outlier tests (Foll and Gaggiotti, 2008). These adaptive loci revealed associations between the populations and natural habitats including climate variables, and will be valuable to construct core subsets for mining certain traits such as drought tolerance. A similar approach would be appropriate for wild tomato species [e.g., *Solanum chilense* (Böndel et al., 2015)] or feral and landrace populations of cultivated tomato. Metabolomics represents another method to characterize germplasm based on functional diversity (Reeves et al., 2020). Many hundreds of tomato quantitative metabolic loci for fruit and yield traits have been documented (Schauer et al., 2006; Tohge et al., 2020). Adaptive loci or metabolic loci may confer greater power to discern groups and assemble core subsets for domesticated tomato compared with random, anonymous SNPs.

Literature Cited

100 Tomato Genome Sequencing Consortium, S. Aflitos, E. Schijlen, H. de Jong, D. de Ridder, S. Smit, R. Finkers, J. Wang, G. Zhang, and N. Li. 2014. Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. Plant J. 80:136–148, doi: https://doi.org/10.1111/tpj.12616.

Alfonso, H. and J.B. Bamberg. 2020. Detection of adaptive genetic diversity in wild potato populations and its implications in conservation of potato germplasm. Amer. J. Plant Sci. 11:1562, doi: https://doi.org/10.4236/ajps.2020.1110113.

Baldo, A.M., D.M. Francis, M. Caramante, L.D. Robertson, and J.A. Labate. 2011. AlleleCoder: A PERL script for coding co-dominant polymorphism data for PCA analysis. Plant Genet. Resources 9:528–530, doi: https://doi.org/10.1017/S1479262111000839.

Bamberg, J. and A. del Rio. 2014. Selection and validation of an AFLP marker core collection for the wild potato *Solanum microdontum*. Amer. J. Potato Res. 91:368–375, doi: https://doi.org/10.1007/s12230-013-9357-5.

Bauchet, G. and M. Causse. 2012. Genetic diversity in tomato (*Solanum lycopersicum*) and its wild relatives, p. 133–162. In: M. Caliskan (ed.). Genetic diversity in plants. IntechOpen, London, UK.

Bauchet, G., S. Grenier, N. Samson, J. Bonnet, L. Grivet, and M. Causse. 2017. Use of modern tomato breeding germplasm for deciphering the genetic control of agronomical traits by genome wide association study. Theor. Appl. Genet. 130:875–889, doi: https://doi.org/10.1007/s00122-017-2857-9.

Bernard, A., T. Barreneche, A. Donkpegan, F. Lheureux, and E. Dirlewanger. 2020. Comparison of structure analyses and core collections for the management of walnut genetic resources. Tree Genet. Genomes 16:1–14, doi: https://doi.org/10.1007/s11295-020-01469-5.

Blanca, J., J. Cañizares, L. Cordero, L. Pascual, M.J. Diez, and F. Nuez. 2012. Variation revealed by SNP genotyping and morphology provides insight into the origin of the tomato. PLoS One 7:e48198, doi: https://doi.org/10.1371/journal.pone.0048198.

Blanca, J., J. Montero-Pau, C. Sauvage, G. Bauchet, E. Illa, M.J. Díez, D. Francis, M. Causse, E. Van der Knaap, and J. Cañizares. 2015. Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. BMC Genomics 16:1–19, doi: https://doi.org/10.1186/s12864-015-1444-1.

Böndel, K.B., H. Lainer, T. Nosenko, M. Mboup, A. Tellier, and W. Stephan. 2015. North–south colonization associated with local

adaptation of the wild tomato species *Solanum chilense*. Mol. Biol. Evol. 32:2932–2943, doi: https://doi.org/10.1093/molbev/msv166.
Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635, doi: https://doi.org/10.1093/bioinformatics/btm308.
Brown, A.H.D. 1995. The core collection at the crossroads, p. 3–19. In: T. Hodgkin, A. Brown, T. Hintum, and E. Morales (eds.). Core collections of plant genetic resources. Wiley, Chichester, UK.
Caramante, M., R. Rao, L.M. Monti, and G. Corrado. 2009. Discrimination of 'San Marzano' accessions: A comparison of minisatellite, CAPS and SSR markers in relation to morphological traits. Sci. Hort. 120:560–564, doi: https://doi.org/10.1016/j.scienta.2008.12.004.
Cebolla-Cornejo, J., S. Roselló, and F. Nuez. 2013. Phenotypic and genetic diversity of Spanish tomato landraces. Scientia Hort. 162:150–164, doi: https://doi.org/10.1016/j.scienta.2013.07.044.
Ceccarelli, S. and S. Grando. 2020. Evolutionary plant breeding as a response to the complexity of climate change. iScience 23(12):101815: doi: https://doi.org/10.1016/j.isci.2020.101815.
Corak, K., S. Ellison, P. Simon, D. Spooner, and J. Dawson. 2019. Comparison of representative and custom methods of generating core subsets of a carrot germplasm collection. Crop Sci. 59:1107–1121, doi: https://doi.org/10.2135/cropsci2018.09.0602.
Cuevas, H.E. and L.K. Prom. 2020. Evaluation of genetic diversity, agronomic traits, and anthracnose resistance in the NPGS sudan sorghum core collection. BMC Genomics 21:88, doi: https://doi.org/10.1186/s12864-020-6489-0.
Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, and S.T. Sherry. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158, doi: https://doi.org/10.1093/bioinformatics/btr330.
de los Ángeles Martínez-Vázquez, E., A. Hernández-Bautista, R. Lobato-Ortiz, J.J. García-Zavala, and D. Reyes-López. 2017. Exploring the breeding potential of Mexican tomato landraces. Scientia Hort. 220:317–325, doi: https://doi.org/10.1016/j.scienta.2017.03.031.
del Rio, A. and J. Bamberg. 2020. A core subset of the *ex situ* collection of *S. demissum* at the US Potato Genebank. Amer. J. Potato Res. 97:505–512, doi: https://doi.org/10.1007/s12230-020-09799-9.
Earl, D.A. and B.M. von Holdt. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4:359–361, doi: https://doi.org/10.1007/s12686-011-9548-7.
Ebert, A.W. 2020. The role of vegetable genetic resources in nutrition security and vegetable breeding. Plants 9:736, doi: https://doi.org/10.3390/plants9060736.
Egea, L.A., R. Mérida-García, A. Kilian, P. Hernandez, and G. Dorado. 2017. Assessment of genetic diversity and structure of large garlic (*Allium sativum*) germplasm bank, by diversity arrays technology "genotyping-by-sequencing" platform (DArTseq). Front. Genet. 8:98, doi: https://doi.org/10.3389/fgene.2017.00098.
Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379, doi: https://doi.org/10.1371/journal.pone.0019379.
Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. 14:2611–2620, doi: https://doi.org/10.1111/j.1365-294X.2005.02553.x.
Foll, M. and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics 180:977–993.
Food and Agriculture Organization of the United Nations. 2020. FAOSTAT Food and Agriculture Organization of the United Nations statistical databases: Value of agricultural production. 26 Apr. 2021. <http://www.fao.org/faostat/en/#data/QC/>.
Genebank Information System of the IPK Gatersleben. 2021. GBIS/I. 26 Apr. 2021. <https://gbis.ipk-gatersleben.de/gbis2i/faces/index.jsf/>.
Genesys. 2021. Genesys global portal on plant genetic resources. 26 Apr. 2021. <www.genesys-pgr.org/>.
Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One 9:e90346, doi: https://doi.org/10.1371/journal.pone.0090346.
Gopal, J., V. Kumar, R. Kumar, and P. Mathur. 2013. Comparison of different approaches to establish a core collection of Andigena (*Solanum tuberosum* Group Andigena) potatoes. Potato Res. 56:85–98, doi: https://doi.org/10.1007/s11540-013-9232-2.
Gramazio, P., L. Pereira-Dias, S. Vilanova, J. Prohens, S. Soler, J. Esteras, A. Garmendia, and M.J. Díez. 2020. Morphoagronomic characterization and whole-genome resequencing of eight highly diverse wild and weedy *S. pimpinellifolium* and *S. lycopersicum* var. *cerasiforme* accessions used for the first interspecific tomato MAGIC population. Hort. Res. 7:1–16, doi: https://doi.org/10.1038/s41438-020-00395-w.
Hanson, P.M. and R.Y. Yang. 2016. Genetic improvement of tomato (*Solanum lycopersicum* L.) for phytonutrient content at AVRDC - The World Vegetable Center. Ekin J. 2(2):1–10.
Hardigan, M.A., J. Bamberg, C.R. Buell, and D.S. Douches. 2015. Taxonomy and genetic differentiation among wild and cultivated germplasm of *Solanum* sect. Petota. Plant Genome 8(1):plantgenome2014-06, doi: https://doi.org/10.3835/plantgenome2014.06.0025.
Huamán, Z., R. Ortiz, and R. Gómez. 2000. Selecting a *Solanum tuberosum* subsp. *andigena* core collection using morphological, geographical, disease and pest descriptors. Amer. J. Potato Res. 77:183–190, doi: https://doi.org/10.1007/BF02853943.
Jayakodi, M., M. Schreiber, N. Stein, and M. Mascher. 2021. Building pan-genome infrastructures for crop plants and their use in association genetics. DNA Res. 28:dsaa030, doi: https://doi.org/10.1093/dnares/dsaa030.
Jin, L., L. Zhao, Y. Wang, R. Zhou, L. Song, L. Xu, X. Cui, R. Li, W. Yu, and T. Zhao. 2019. Genetic diversity of 324 cultivated tomato germplasm resources using agronomic traits and InDel markers. Euphytica 215:1–16, doi: https://doi.org/10.1007/s10681-019-2391-8.
Kelley, W.T., G.E. Boyhan, K.A. Harrison, P.E. Sumner, D.B. Langston, A.N. Sparks, S. Culpepper, W.C. Hurst, and E.G. Fonsah. 2017. Commercial tomato production handbook. Univ. Georgia Ext. Bul. 1312.
Kulus, D. 2018. Genetic resources and selected conservation methods of tomato. J. Appl. Bot. Food Qual. 91:135–144, doi: https://doi.org/0.5073/JABFQ.2018.091.019.
Kuzay, S., P. Hamilton-Conaty, A. Palkovic, and P. Gepts. 2020. Is the USDA core collection of common bean representative of genetic diversity of the species, as assessed by SNP diversity? Crop Sci. 60:1398–1414, doi: https://doi.org/10.1002/csc2.20032.
Labate, J.A. and L.D. Robertson. 2012. Evidence of cryptic introgression in tomato (*Solanum lycopersicum* L.) based on wild tomato species alleles. BMC Plant Biol. 12:133, doi: https://doi.org/10.1186/1471-2229-12-133.
Labate, J.A. and L.D. Robertson. 2015. Nucleotide diversity estimates of tomatillo (*Physalis philadelphica*) accessions including nine new inbred lines. Mol. Breed. 35:106, doi: https://doi.org/10.1007/s11032-015-0302-9.
Lee, H.-Y., J.-G. Kim, B.-C. Kang, and K. Song. 2020. Assessment of the genetic diversity of the breeding lines and a genome wide association study of three horticultural traits using worldwide cucumber (*Cucumis* spp.) germplasm collection. Agronomy 10:1736, doi: https://doi.org/10.3390/agronomy10111736.
Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760, doi: https://doi.org/10.1093/bioinformatics/btp324.
Loiudice, R., M. Impembo, B. Laratta, G. Villari, A.L. Voi, P. Siviero, and D. Castaldo. 1995. Composition of San Marzano tomato varieties. Food Chem. 53:81–89, doi: https://doi.org/10.1016/0308-8146(95)95791-4.
Mata-Nicolás, E., J. Montero-Pau, E. Gimeno-Paez, V. Garcia-Carpintero, P. Ziarsolo, N. Menda, L.A. Mueller, J. Blanca, J. Cañizares, and E. Van der Knaap. 2020. Exploiting the diversity of tomato: The development

of a phenotypically and genetically detailed germplasm collection. Hort. Res. 7:1–14, doi: https://doi.org/10.1038/s41438-020-0291-7.
Mazzucato, A., R. Papa, E. Bitocchi, P. Mosconi, L. Nanni, V. Negri, M.E. Picarella, F. Siligato, G.P. Soressi, and B. Tiranti. 2008. Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (*Solanum lycopersicum* L.) landraces. Theor. Appl. Genet. 116:657–669, doi: https://doi.org/10.1007/s00122-007-0699-6.
Miller, J.C. and S.D. Tanksley. 1990. RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor. Appl. Genet. 80:437–448.
Niu, S., Q. Song, H. Koiwa, D. Qiao, D. Zhao, Z. Chen, X. Liu, and X. Wen. 2019. Genetic diversity, linkage disequilibrium, and population structure analysis of the tea plant (*Camellia sinensis*) from an origin center, Guizhou plateau, using genome-wide SNPs developed by genotyping-by-sequencing. BMC Plant Biol. 19:1–12, doi: https://doi.org/10.1186/s12870-019-1917-5.
Pavan, S., N. Bardaro, V. Fanelli, A.R. Marcotrigiano, G. Mangini, F. Taranto, D. Catalano, C. Montemurro, C. De Giovanni, and C. Lotti. 2019. Genotyping by sequencing of cultivated lentil (*Lens culinaris* Medik.) highlights population structure in the Mediterranean gene pool associated with geographic patterns and phenotypic variables. Front. Genet. 10:872, doi: https://doi.org/10.3389/fgene.2019.00872.
Peralta, I.E., S. Knapp, and D. Spooner. 2008. The taxonomy of tomatoes: a revision of wild tomatoes (*Solanum* section *Lycopersicon*) and their outgroup relatives (*Solanum* sections *Juglandifolium* and *Lycopersicoides*). Syst. Bot. Monogr. 84:1–186.
Pritchard, J.K., M. Stephens, N.A. Rosenberg, and P. Donnelly. 2000. Association mapping in structured populations. Amer. J. Hum. Genet. 67:170–181, doi: https://doi.org/10.1086/302959.
Rambaut, A. 2012. FigTree 1.4.0. 26 Apr. 2021. <http://tree.bio.ed.ac.uk/software/figtree/>.
Rao, R., G. Corrado, M. Bianchi, and A. Di Mauro. 2006. $(GATA)_4$ DNA fingerprinting identifies morphologically characterized 'San Marzano' tomato plants. Plant Breed. 125:173–176, doi: https://doi.org/10.1111/j.1439-0523.2006.01183.x.
Razifard, H., A. Ramos, A.L. Della Valle, C. Bodary, E. Goetz, E.J. Manser, X. Li, L. Zhang, S. Visa, and D. Tieman. 2020. Genomic evidence for complex domestication history of the cultivated tomato in Latin America. Mol. Biol. Evol. 37:1118–1132, doi: https://doi.org/10.1093/molbev/msz297.
Reeves, P.A., L.W. Panella, and C.M. Richards. 2012. Retention of agronomically important variation in germplasm core collections: Implications for allele mining. Theor. Appl. Genet. 124:1155–1171, doi: https://doi.org/10.1007/s00122-011-1776-4.
Reeves, P.A., H.M. Tetreault, and C.M. Richards. 2020. Bioinformatic extraction of functional genetic diversity from heterogeneous germplasm collections for crop improvement. Agronomy 10:593, doi: https://doi.org/10.3390/agronomy10040593.
Rick, C.M. 1982. A new self-compatible wild population of *L. peruvianum*. TGC Report 32:43–44.
Rick, C.M. and J.F. Fobes. 1975. Allozyme variation in the cultivated tomato and closely related species. Bull. Torrey Bot. Club 102:376–386, doi: https://doi.org/10.2307/2484764.
Rodriguez, G.R., S. Munos, C. Anderson, S.C. Sim, A. Michel, M. Causse, B.B. McSpadden Gardener, D. Francis, and E. Van der Knaap. 2011. Distribution of *SUN, OVATE, LC*, and *FAS* in the tomato germplasm and the relationship to fruit shape diversity. Plant Physiol. 156:275–285, doi: https://doi.org/10.1104/pp.110.167577.
Rosenberg, N.A. 2004. DISTRUCT: A program for the graphical display of population structure. Mol. Ecol. Notes 4:137–138, doi: https://doi.org/10.1046/j.1471-8286.2003.00566.x.
Saitou, N. and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425, doi: https://doi.org/10.1093/oxfordjournals.molbev.a040454.
Sauvage, C., V. Segura, G. Bauchet, R. Stevens, P.T. Do, Z. Nikoloski, A.R. Fernie, and M. Causse. 2014. Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. Plant Physiol. 165:1120–1132, doi: https://doi.org/10.1104/pp.114.241521.
Schauer, N., Y. Semel, U. Roessner, A. Gur, I. Balbo, F. Carrari, T. Pleban, A. Perez-Melis, C. Bruedigam, and J. Kopka. 2006. Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat. Biotechnol. 24:447–454, doi: https://doi.org/10.1038/nbt1192.
Scheben, A., J. Batley, and D. Edwards. 2017. Genotyping-by-sequencing approaches to characterize crop genomes: Choosing the right tool for the right application. Plant Biotechnol. J. 15:149–161, doi: https://doi.org/10.1111/pbi.12645.
Shi, A., J. Qin, B. Mou, J. Correll, Y. Weng, D. Brenner, C. Feng, D. Motes, W. Yang, and L. Dong. 2017. Genetic diversity and population structure analysis of spinach by single-nucleotide polymorphisms identified through genotyping-by-sequencing. PLoS One 12:e0188745, doi: https://doi.org/10.1371/journal.pone.0188745.
Sim, S., M. Kim, S.-M. Chung, and Y. Park. 2015. Assessing the genetic variation in cultivated tomatoes (*Solanum lycopersicum* L.) using genome-wide single nucleotide polymorphisms. Hort. Environ. Biotechnol. 56:800–810, doi: https://doi.org/10.1007/s13580-015-0107-0.
Sim, S., M.D. Robbins, A. Van Deynze, A.P. Michel, and D.M. Francis. 2011. Population structure and genetic differentiation associated with breeding history and selection in tomato (*Solanum lycopersicum* L.). Heredity 106:927–935, doi: https://doi.org/10.1038/hdy.2010.139.
Snouffer, A., C. Kraus, and E. van der Knaap. 2020. The shape of things to come: Ovate family proteins regulate plant organ shape. Curr. Opin. Plant Biol. 53:98–105, doi: https://doi.org/10.1016/j.pbi.2019.10.005.
SolCap. 2013. Solanaceae Coordinated Agricultural Project. 26 Apr. 2021. <http://solcap.msu.edu/tomato_germplasm_data.shtml/>.
The Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635–641, doi: https://doi.org/10.1038/nature11119.
Tohge, T., F. Scossa, R. Wendenburg, P. Frasse, I. Balbo, M. Watanabe, S. Alseekh, S.S. Jadhav, J.C. Delfin, and M. Lohse. 2020. Exploiting natural variation in tomato to define pathway structure and metabolic regulation of fruit polyphenolics in the lycopersicum complex. Mol. Plant 13:1027–1046, doi: https://doi.org/10.1038/nature11119.
U.S. Department of Agriculture, Agricultural Marketing Service. 2021. Plant Variety Protection. 26 Apr. 2021. <https://www.ams.usda.gov/services/plant-variety-protection/issued-certificates>.
U.S. Department of Agriculture, Agricultural Research Service. 2021a. U.S. National Plant Germplasm System GRIN-Global. 26 Apr. 2021. <https://npgsweb.ars-grin.gov/gringlobal/search/>.
U.S. Department of Agriculture, Agricultural Research Service. 2021b. U.S. National Plant Germplasm System GRIN-Global. 26 Apr. 2021. <https://npgsweb.ars-grin.gov/gringlobal/uploads/images/npgs/ne9/tomato/TomatoResistances1975.pdf/>.
U.S. Department of Agriculture, National Agricultural Statistics Service. 2021. Vegetables 2019 summary (Feb. 2020). 26 Apr. 2021. <https://www.nass.usda.gov/>.
van Hintum, T.J.L., A.H.D. Brown, C. Spillane, and T. Hodgkin. 2000. Core collections of plant genetic resources. International Plant Genetic Resources Institute (IPGRI) Tech. Bul. No. 3.
Villand, J., P.W. Skroch, T. Lai, P. Hanson, C.G. Kuo, and J. Nienhuis. 1998. Genetic variation among tomato accessions from primary and secondary centers of diversity. Crop Sci. 38:1339–1347, doi: https://doi.org/10.2135/cropsci1998.0011183X003800050032x.
Wehner, T.C. 2016. Vegetable cultivar descriptions for North America. 26 Apr. 2021. <https://cucurbitbreeding.wordpress.ncsu.edu/2016/05/23/vegetable-cultivar-descriptions-for-north-america/>.
Zhao, J., C. Sauvage, J. Zhao, F. Bitton, G. Bauchet, D. Liu, S. Huang, D.M. Tieman, H.J. Klee, and M. Causse. 2019. Meta-analysis of genome-wide association studies provides insights into genetic control of tomato flavor. Nat. Commun. 10:1–12, doi: https://doi.org/10.1038/s41467-019-09462-w.