Commercial Tomato PlexSeq™ SNP Panel
By AgriPlex Genomics

Introduction

The tomato, Solanum lycopersicum (Solanaceae, Nightshades family) is one of the most widely consumed fruits. The global tomato seed market was valued at $1.45 billion in 2020 and is expected to reach $2.05 billion by 2026, growing at a Compound Annual Growth Rate of 6.2% (1). The tomato market is segmented into hybrid seeds and open-pollinated seeds and by market niches. The United States is one of the largest markets for tomato seeds globally and is divided into three main segments: field grown Fresh Market and greenhouse grown Fresh Market which, as the name implies are primarily used for fresh consumption.
Commercial Processing tomatoes are used for producing products such as salsa, sauces, juice, paste, and ketchup. According to a 2020 USDA report, the value of Fresh Market Tomato production in the United States was USD 2.21 billion, while the value of processing tomatoes was USD 1.07 billion (2). Fresh Market Tomatoes are generally larger, more consistent in shape, and have a higher sugar content. Commercial Processing Tomatoes have a higher acid content, have thicker skin, and are more viscous making them easier to mechanically harvest, peel, and process (3).

The early domestication of tomato from its wild ancestor, Solanum pimpinellifolium is thought to have occurred in South America followed by a secondary event in Mesoamerica around 2,500 years ago (4, 5). Archaeological evidence suggests that tomatoes were cultivated by the Maya and Aztec civilizations 500 BC, and likely were primarily used for medicinal purposes rather than food (3). Over time, human selection led to the development of sweeter, more palatable tomatoes (6, 7). As tomatoes were introduced to different parts of the world, they were subjected to different selection regimes. For instance, when tomatoes were introduced to Europe in the 16th century, there was a preference for larger fruit. In Asia, there was a preference for smaller fruit.

The domesticated tomato, S. lycopersicum, is a diploid plant species with 2n=24 chromosomes and a haploid complement genome size of approximately 950 Mb (8). There is considerable genetic variation within the tomato clade of the genus Solanum. The wild relatives include nine species, which are diploid (2n=24) (9). Tomato has a high degree of synteny with other solanaceous crops such as potato, pepper, and eggplant, however, tomato has undergone a whole-genome triplication event that is not present in potato or pepper (10). The tomato genome also possesses some unique characteristics, such as the presence of a high number of genes encoding for cell wall-modifying enzymes, flavor and aroma compounds, and carotenoid biosynthesis enzymes. These characteristics make the tomato a valuable model for studying fruit development, ripening, and nutritional traits (11).

The first tomato genome sequence was published in 2012 (10). Since then, multiple high quality genome assemblies have been developed, including the latest release of the reference genome, SL4.0, in 2020 (12). Several other important genomic resources have been developed for tomato research, including comprehensive transcriptome databases (13, 14), high-density genetic maps (15, 16, 17), and various types of molecular markers (18).

Molecular markers as a technology for interrogating genetic variation have progressed in recent decades and many DNA molecular marker systems were developed. Consequently, so did the resolution of the genomic picture the markers can depict. Molecular markers have been used in tomato breeding starting with Restriction Fragment Length Polymorphisms (RFLPs) in the early 1990s (15), and Simple Sequence Repeats (SSRs or microsatellites) starting from the late 1990s on (18, 19). In the early 2000s, Single Nucleotide Polymorphisms (SNPs) emerged as the ultimate molecular marker, gaining popularity in tomato breeding (20).

SNPs are single nucleotide changes that are heritable, codominant, and distributed at high frequency throughout eukaryotic genomes. A SNP can be the causative mutation that directly affects a phenotype or can be associated by linkage to a causative mutation. SNPs also present the operational advantages of high-throughput genotyping and ease of use. These advantages have become more pronounced with the advent of Next Generation Sequencing (NGS), which has enabled the identification of large number of SNPs, such as the Solanaceae Coordinated Agricultural Project (SolCap) SNP array that is made of more than 7000 SNPs. NGS methods have also facilitated the development of high throughput, rapid, and economical genotyping platforms such as PlexSeq (see following).

Both molecular markers in general and SNPs specifically were used by geneticists to construct high-density genetic maps (2021), which in turn, were used to identify quantitative trait loci (QTLs) associated with important traits such as fruit quality, disease resistance, stress tolerance (21) and Enhanced Nutritional Content by increasing lycopene content (22, 23).

For breeders, the use of molecular markers permits accurate and early selection of individuals of interest, reduction of the number of selection cycles required (24), implementation of new breeding schemes such as marker-assisted selection, background genome selection (24, 25, 26), and genomic selection (27, 28). Overall, the utilization of molecular markers in breeding reduces the time to market of new lines at a lower cost of breeding.

We describe here a new, mid-density, multiplexed SNP panel designed for various research and breeding applications on the PlexSeq NGS genotyping platform.

Panel Characterization

The Initial Fresh Market Tomato Panel

The Commercial Tomato SNP panel was developed as an extension and update of a Fresh Market Tomato (FMT) panel that was created as part of an initiative of the Solanaceae Coordinated Agricultural Project (SolCap). This FMT panel contained 384 SNPs. SolCap was established with the help of the United States Department of Agriculture (USDA) to bridge the gap between breeding and genomics. The project has identified and validated 7,720 SNPs (Illumina's Tomato Infinium array) through whole transcriptome sequencing of cultivated tomato varieties and two wild species, S. pimpinellifolium and the weedy S. lycopersicum var. cerasiforme (https://emea.illumina.com/library-prep-array-kit-selector/kits-and-arrays/thesolcap-tomato-consortium.html). High density SNP arrays are an excessive and expensive genotyping tool for many applications; therefore, two 384 subsets of SNPs were further selected from the 7K SNP array. Molecular markers for these panels were chosen to:

- Provide coverage of all 12 chromosomes,
- Maintain a 0.2 cM or 0.2 Mbp distance between adjacent SNPS (corresponding to the expected recombination rate),
- The DNA sequences that are flanking the SNP lend them self to primer design.
- Include common SNPs between populations but also population specific SNPs.

One of these 384 SNPs subsets was created for Commercial Processing Tomatoes while the other was created for Fresh Market Tomatoes. The 384 SNP Fresh Market Tomato subset was introduced by AgriPlex Genomics in early 2018. The SNPs were multiplexed and validated as a PlexSeq panel and offered to tomato breeders and researchers since June of that year.

The Mid-Density, Commercial Tomato Panel

The new Commercial Tomato panel was designed by a consortium of leading tomato breeders and geneticists. Marker selection was conducted and lead by Dr. Tong Geon Lee of the University of Florida (Currently at Bayer) with contributions from Professor David Francis of The Ohio State University.

The need to expand upon the Fresh Market Tomato panel rose for several reasons:

- To obtain higher genome coverage by closing gaps of unqueried genomic intervals between adjacent markers in the current FMT panel.
- To augment the existing panel with more trait related markers
- To provide one panel that can be utilized by both Commercial Processing and Fresh Market Tomato studies and breeding.

The Commercial Tomato panel contains 1039 SNPs that span across all 12 tomato chromosomes. The additional markers were derived from genotypic datasets of diverse cultivated tomato inbreds, wild accessions, and their segregating populations. The various sources include (Table 1) transcriptome sequencing of Solanum lycopersicum, S. pimpinellifolium, S. pennellii (29, 30), whole genome sequencing of S. lycopersicum (31, 32, 33) and individual contributions of various tomato researchers.

The improvements in genome coverage of the Commercial Tomato Panel compared to the preceding FMT panel are depicted in Figure 1. The illustration represents the relative markers density in both panels. These improvements are further captured in Table 2:

- The marker density per chromosome has increased on average 2.7-fold, going from an average number of 32 markers per chromosome in the FMT panel to an average 87 SNPs per chromosome for the Commercial Tomato Panel.

- Consequently, the average inter-marker gap decreased to 0.773 Mbp which is a 64.9% improvement compared to the previous distance between adjacent markers in the FMT panel.

To provide continuity with the widely used 384 SNP FMT panel and to retain its quality in capturing the genetic variations within the Fresh Market Tomato class, 360 of the best performing SNPs (94%) of the Fresh Market Tomato Panel have been preserved in the Commercial Tomato Panel (Figure 1).

Commercial Tomato Panel			
Number of SNPs (%)	**Species**	**Source**	**Reference**
360 (35%)	Solanum lycopersicum, S. pimpinellifolium, S. pennellii	TS	29, 30
509 (49%)	S. lycopersicum	WGS	31, 32, 33
170 (16%)	NA	NA	Individual contributions
Total: **1039**			

Table 1. Summary of the SNP numbers by species and source of the Commercial Tomato Panel. TS, transcriptome sequencing, WGS, whole-genome resequencing, NA, not available.

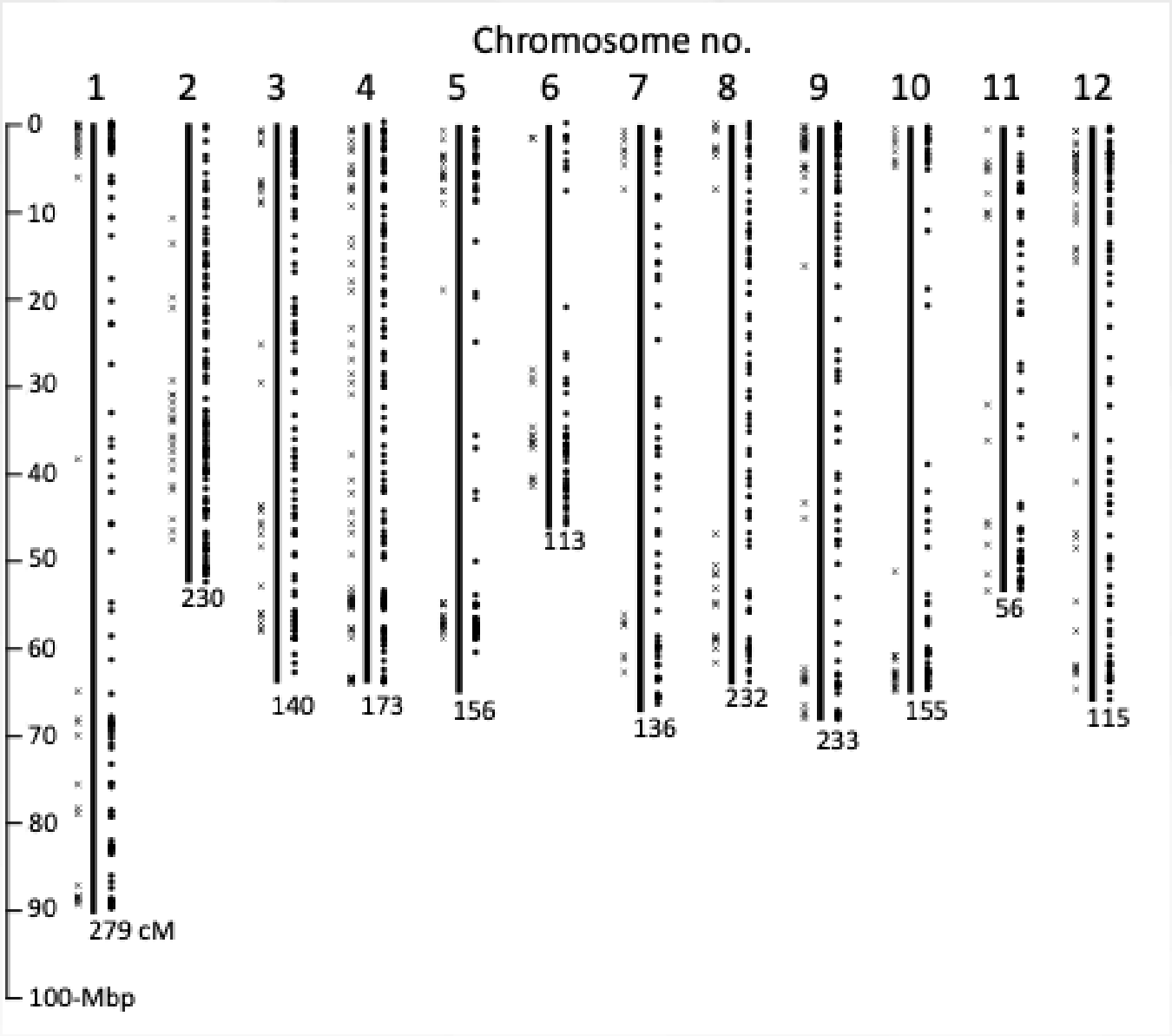

Figure 1. Chromosomal distribution of 1,039 SNPs in the Commercial Tomato SNP panel. The physical positions of SNPs are indicated by circles to the right of the chromosome. x symbol, left to the chromosome mark the positions of markers that were already included in the initial FMT panel. Estimated genetic map lengths (cM) based on contemporary U.S. Fresh Market Tomato populations (31,32,33) are indicated immediately below to the corresponding chromosome. Courtesy of Tong Geon Lee (University of Florida; Bayer).

	Fresh Market Tomato Panel		Commercial Tomato Panel	
Chrom.	**N**	**Average inter-marker distance (Mbp)**	**N**	**Average inter-marker distance (Mbp)**
1	34	2.691719	99	0.924163
2	25	1.902987	110	0.485786
3	30	1.960087	99	0.646472
4	66	0.969940	119	0.544230
5	37	1.737220	74	0.881625
6	18	2.378442	60	0.797141
7	18	3.649099	66	1.021638
8	22	2.931501	85	0.760103
9	40	1.733449	95	0.727663
10	28	2.380937	72	0.905713
11	21	2.634511	60	0.909432
12	45	1.466472	100	0.665945
Total	384		1039	
Average	32	2.2030	87	0.7725
%CV	43.18	292	23.02	21.30

Table 2. Number of markers per chromosome, and average distance (Mbp) between adjacent markers in the initial Fresh Market Tomato panel and the updated Commercial Tomato Panel.

The Commercial Tomato panel contains 384 trait-associated markers, 274 of which are associated with gene coding for major biochemical functions such as transcription and replication regulation, cell division and cellular structure proteins, photosynthesis, disease resistance to fusarium, late blight resistance, and Verticillium wilt disease resistance protein (please see Appendix A for a complete list). The remaining 110 SNPs are associated with genes of unknown factors.

A study was conducted by AgriPlex Genomics and a consortium of tomato researchers and breeders to characterize and validate the Commercial Tomato panel; 2726 samples were submitted for this purpose from four academic tomato genetics and breeding programs.

The success rate varied between 76.8 % to 96.2% among the different submissions and averaged 86.48%. When addressing the 4 different breeding programs as populations, the average % polymorphism of the panel ranged from 62.6% to 75.3%, and averaged 66.96%, or on the average 696 markers were polymorphic. The Average Minor Allele Frequency was 0.19 ± 0.166 (corresponding to a coefficient of variance of 86.5%).

It is important to note that the germplasm used for these experiments consisted of a variety of cultivated lines as well as wild tomato species; thus, the observed variation in performance parameters is expected for a panel designed to include common SNPs between populations as well as population specific SNPs.

PlexSeq™: The mid-density, multiplexed, SNP genotyping platform

Several attributes of the PlexSeq™ process contribute to its unique value as a genotyping platform:

- The proprietary multiplexing algorithm, PlexForm™: The software designs all possible primers around all requested SNPs. Artificial intelligence algorithms identify the optimal sets of primers that can be mixed in one PCR amplification reaction.

- Once the amplifications are completed, the amplicon mixture is equivalent to barcoded libraries produced from other NGS methods. The process is unique because the samples produce amplicon libraries that are equivalent in concentration and do not require any additional equalization steps. A mixture of all the libraries is subjected to one bead cleanup and are loaded onto the sequencer. The process saves time, plasticware, and expenses.

- The method requires only minute quantities of crude DNA that can be isolated from a variety of tissues, enabling a quick and inexpensive DNA isolation process to start the genotyping workflow.

The PlexSeq™ workflow consists of:

- Crude DNA isolation
- Primary PCR: highly multiplexed, low volume (3ul) PCR amplifications
- Secondary, barcoding PCR amplifications
- Pooling: barcoded amplicons are combined into one tube, purified, and quantitated
- Sequencing on an NGS sequencer
- This relatively simple workflow is amenable to automation; all steps can be carried out on liquid handlers and high-capacity thermocyclers. This enables high throughput genotyping.

Once the sequencing is complete, a proprietary allele frequency-based genotype calling analysis software, PlexCall™, provides an automated sequencer to data workflow. This Java- based software is tuned for each assay and is fully automated based on only the sequencing output files and a sample sheet indicating sample location on the plate.

Two other features make PlexSeq™ a unique fit for molecular breeding and seed QA. These applications typically require the genotyping of a large number of individuals.

AgriPlex Genomics' extensive collection of unique barcode combinations allows for the simultaneous sequencing of thousands of individuals, limited only by the sequencer's capacity. Similarly, molecular breeding may require the addition or substitution of only some of the SNP markers as the individual breeding program advances or among programs as the parental-lines genetic makeup evolves. Those changes in SNP composition are also required for Quality Assurance (QA) applications as the diversity of the germplasm changes. The panel, being a collection of PCR primers not tethered to a surface (e.g.: chips) provides the flexibility to dynamically customize and alter the composition of the markers in the panel so it best fits the germplasm or the application. The Commercial Tomato Panel is available as a service from AgriPlex Genomics and is also available as a kit to be used by in-house genotyping laboratories.

Applications

Genomic selection:

The Commercial Tomato SNP panel will enable genomic selection; the level of polymorphism observed indicates that there is a high likelihood of obtaining sufficient polymorphic markers.

for imputation to a higher marker density level. The average Genomic Selection predictive ability will vary for different combinations of parental lines and in different years.

The combination of rapid, cost-effective genotyping of a prediction population during the last generation of line fixation saves expenses on the cost of field space for seed increase and allows rapid recycling of progeny as parents.

Marker-assisted backcrossing:

The Commercial Tomato SNP panel can be used for background recovery estimates in marker assisted backcrossing programs. The combination of informative background markers and a selection of trait markers allows for the accurate estimation of background recovery, ensures recovery of valuable genes from the recipient line, and can provide additional confirmation that a target gene is carried by the selected progeny. Background selection can reduce by 2 or more the number of backcross generations required to achieve >95% recipient parent recovery.

QTL mapping:

While not the primary target application, the Commercial Tomato SNP panel can be used for biparental mapping purposes. The density of polymorphic markers may, in some cases, be lower than what is desirable (largely dependent on the parents involved), which may lead to gaps in the linkage map. However, the panel does provide an option for genotyping much of the genome. Any remaining gaps could then be filled in with other marker systems or by further panel customization.

Trait profiling:

The Commercial Tomato SNP panel contains high-value trait markers (Appendix A), covering a range of traits related to disease resistance, enzymatic profile, abiotic stress tolerance, taste, and others. These markers are designed to characterize the targeted genes and QTLs across different Solanum species and accessions including some wild relatives and are highly informative within the larger tomato U.S. germplasm, enabling interrogation of traits from/into lines of interest.

Seed purity and hybridity:

The genome coverage of the Commercial Tomato SNP panel includes the number of markers that will allow for the identification of diagnostic marker subsets for seed quality applications such as genetic purity testing (uniformity) and varietal identification in commercial production operations.

Conclusions

AgriPlex Genomics' implementation of the Commercial Tomato SNP panel is providing an excellent, cost-effective alternative for applications requiring mid-density SNP numbers over any number of sample throughput. The panel fits with rapid line fixation protocols due to its low cost-per-sample and rapid turnaround time. This panel will allow for major-locus selection as part of genomic selection, or backcross introgressions amongst tomato accessions, or the

transfer of genes and QTLs of interest from wild tomato species into cultivated germplasm. The panel presents a valuable tool for saving critical time, expenses, and shortening the "time to market." The panel primarily enables molecular breeding applications. However, the suite of trait markers and genome coverage expands its usefulness in a range of other research applications.

The flexibility of the PlexSeq platform permits continual revision and upgrading of the panel, ensuring the process keeps pace with the current breeding and seed production needs of the tomato breeding and research communities.

References

1. Mordor Intelligence. Tomato Seed Market - Growth, Trends, COVID-19 Impact, and Forecasts (2021 - 2026). June 2021.
2. USDA. Vegetables 2020 Summary. July 2021.3. Peralta, I. E. and D.M. Spooner. History, origin, and early cultivation of tomato (Solanaceae). In: J. Prohens and F. Nuez (eds.), Handbook of Tomato Breeding. Springer, New York, 2018.

3. Peralta, I. E. and D.M. Spooner. History, origin, and early cultivation of tomato (Solanaceae). In: J. Prohens and F. Nuez (eds.), Handbook of Tomato Breeding. Springer, New York, 2018.

4. Blanca J, Montero-Pau J, Sauvage C, Bauchet G, Illa E, Díez MJ, Francis D, Causse M, van der Knaap E, Cañizares J. Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. BMC Genomics. 2015 Apr 1;16(1):257. doi: 10.1186/s12864-015-1444-1.

5. Hamid Razifard, Alexis Ramos, Audrey L Della Valle, Cooper Bodary, Erika Goetz, Elizabeth J Manser, Xiang Li, Lei Zhang, Sofia Visa, Denise Tieman, Esther van der Knaap, Ana L Caicedo, Genomic Evidence for Complex Domestication History of the Cultivated Tomato in Latin America, Molecular Biology and Evolution, Volume 37, Issue 4, April 2020, Pages 1118-1132, https://doi.org/10.1093/molbev/msz297

6. Rick, C. Tomato. In: Evolution of Crop Plants, 2nd edition, edited by J. Smartt and N. W. Simmonds, pp. 415-418. Longman Scientific and Technical, 1995.

7. Sauvage C, Rau A, Aichholz C, Chadoeuf J, Sarah G, Ruiz M, Santoni S, Causse M, David J,
Glémin S. Domestication rewired gene expression and nucleotide diversity patterns in tomato. Plant J. 2017 Aug;91(4):631-645. doi: 10.1111/tpj.13592. Epub 2017 Jun 12. PMID: 28488328.

8. The Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature, 485(7400), 635-641.

9. Peralta, I. E., Spooner, D. M., and Knapp, S. (2008). Taxonomy of wild tomatoes and their relatives (Solanum sect. Lycopersicoides, sect. Juglandifolia, sect. Lycopersicon; Solanaceae). Systematic Botany Monographs, 84, 1-186.

10. Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature, 485(7400), 635-641.

11. Giovannoni, J. J. (2004). Genetic regulation of fruit development and ripening. The Plant Cell, 16(Suppl 1), S170-S180.

12. International Tomato Genome Sequencing Project; https://solgenomics.net/organism/Solanum_lycopersicum/genome

13. Chen, G., Hu, Z., Grierson, D., and Yang, Y. (2018). Transcriptome profiling reveals differential gene expression and a possible role for gibberellins in a shade-tolerant mutant of tomato. Horticulture research

14. Kudo T, Kobayashi M, Terashima S, Katayama M, Ozaki S, Kanno M, Saito M, Yokoyama K, Ohyanagi H, Aoki K, Kubo Y, Yano K. TOMATOMICS: A Web Database for Integrated Omics Information in Tomato. Plant Cell Physiol. 2017 Jan 1;58(1): e8. doi: 10.1093/pcp/pcw207. PMID: 28111364; PMCID: PMC5444566.

15. Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, et al. High-density molecular linkage maps of the tomato and potato genomes. Genetics. 1992 Dec;132(4):1141-60. doi: 10.1093/genetics/132.4.1141. PMID: 1360934; PMCID: PMC1205235.

16. Foolad, M. R. (2007). Genome mapping and molecular breeding of tomato. International journal of plant genomics, 2007.

17. Sim SC, Van Deynze A, Stoffel K, Douches DS, Zarka D, Ganal MW, Chetelat RT, Hutton SF, Scott JW, Gardner RG, Panthee DR, Mutschler M, Myers JR, Francis DM. High-density SNP genotyping of tomato (Solanum lycopersicum L.) reveals patterns of genetic variation due to breeding. PLoS One. 2012;7(9):e45520. doi: 10.1371/journal.pone.0045520.

18. Fernandez-Moreno, P. T., Malpica, J. M., Lopez-Casado, G., and Cubero, J. I. (2013). Simple sequence repeats and diversity in cultivated and wild tomatoes. Genetic resources and crop evolution, 60(7), 2055-2070.

19. Bredemeijer, G., P. Arens and D. Wouters (1998): The use of semi-automated fluorescent microsatellite analysis for tomato cultivar identification. Theor. Appl. Genet., 97: 584-590

20. Hamilton, J.P., Sim, S.-C., Stoffel, K., Van Deynze, A., Buell, C.R. and Francis, D.M. (2012), Single Nucleotide Polymorphism Discovery in Cultivated Tomato via Sequencing by Synthesis. The Plant Genome, 5:. https://doi.org/10.3835/plantgenome2011.12.0033

21. Sim, S. C., Durstewitz, G., and Plieske, J. (2012). Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. PloS one, 7(7), e40563.

22. Kumar, R., Sharma, S., Kapoor, V., and Tyagi, A. K. (2016). Molecular breeding of tomato for nutritive value and processing. In Biotechnologies of crop improvement, volume 3 (pp. 57-87). Springer, Cham.

23. Eduardo Bernal, Debora Liabeuf, and David M. Francis 2020 Evaluating Quantitative Trait Locus Resistance in Tomato to Multiple Xanthomonas spp. Plant Disease 2020 104:2, 423-429

24. Majid R. Foolad and Dilip R. Panthee (2012) Marker-Assisted Selection in Tomato Breeding, Critical Reviews in Plant Sciences, 31:2, 93-123, DOI: 10.1080/07352689.2011.616057

25. Wencai Yang and David Francis, (2005) Marker-assisted selection for combining resistance to bacterial spot and bacterial speck in tomato, J. Amer. Soc. Hort Sci. 130(5): 716-721

26. Estrada-Ortiz, E., Trejo-Tapia, G., Guevara-Gonzalez, R. G., and Torres-Pacheco, I. (2019). Marker-assisted selection in tomato breeding for TYLCV resistance: A case study in Mexico. Molecular breeding, 39(2), 23.

27. Duangjit, J., Causse, M. and Sauvage, C. Efficiency of genomic selection for tomato fruit quality. Mol Breeding 36, 29 (2016). https://doi.org/10.1007/s11032-016-0453-3

28. Elisa Cappetta, Giuseppe Andolfo, Anna Guadagno, Antonio Di Matteo, Amalia Barone, Luigi Frusciante and Maria Raffaella Ercolano (2021) Tomato genomic prediction for good performance under high-temperature and identification of loci involved in thermotolerance response, Horticulture Research, Volume 8, 212, https://doi.org/10.1038/s41438-021-00647-3

29. Sim, S. C., Durstewitz, G., Plieske, J., Wieseke, R., Ganal, M. W., Van Deynze, A., Hamilton, J. P., Buell, C. R., Causse, M., Wijeratne, S., and Francis, D. M. (2012a). Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. PloS one, 7(7), e40563.

30. Sim, S. C., Van Deynze, A., Stoffel, K., Douches, D. S., Zarka, D., Ganal, M. W., Chetelat, R. T., Hutton S.F., Scott, J.W., Gardner, R.G., Panthee, D.R., Mutschler, M., Myers, J.R., and Francis, D.M. (2012b). High-density SNP genotyping of tomato (Solanum lycopersicum L.) reveals patterns of genetic variation due to breeding. PloS one, 7(9), e45520

31. Bhandari, P., and Lee, T. G. (2021). A genetic map and linkage panel for the large-fruited fresh-market tomato. Journal of the American Society for Horticultural Science J. Amer. Soc. Hort. Sci., 146(2), 125-131.

32. Bhandari, P., Shekasteband, R., and Lee, T. G. (2022). A consensus genetic map and linkage panel for fresh-market tomato. Journal of the American Society for Horticultural Science, 147(1), 53-61.

33. Bhandari, P., Kim, J., and Lee, T. G. (2023). Genetic architecture of fresh-market tomato yield. BMC Plant Biology, 23(1), 18.

Appendices

Appendix A