

DATA NOTE

Open Access



Genome sequence data of the contemporary fresh-market tomatoes

Juhee Kim^{1,2} and Tong Geon Lee^{3*}

Abstract

Objective The fresh-market tomato (*Solanum lycopersicum*) is bred for direct human consumption. It is selected for specific traits to meet market demands and production systems, and unique genetic variations underlying fresh-market tomato yields have been recently identified. However, DNA sequence variant-trait associations are not yet fully examined even for major traits. To provide a rich genome sequence resource for various genetics and breeding goals for fresh-market tomato traits, we report whole genome sequence data of a pool of contemporary U.S. fresh-market tomatoes.

Data description Eighty-one tomatoes were nominated by academic tomato breeding programs in the U.S. Of the 81 tomatoes, 68 were contemporary fresh-market tomatoes, whereas the remaining 13 were relevant fresh-market tomato breeding and germplasm accessions. Whole genome sequencing (WGS) of the 81 tomatoes was conducted using the Illumina next-generation sequencing technology. The polymerase chain reaction (PCR)-free, paired-end sequencing libraries were sequenced on an average depth per sequenced base of 24× for each tomato. This data note enhances visibility and potential for use of the more diverse, freely accessible whole genome sequence data of contemporary fresh-market tomatoes.

Keywords *Solanum lycopersicum*, Genetic variants, Nucleotide variability, Whole genome sequencing

Objective

Tomato (*Solanum lycopersicum*) is widely consumed worldwide [1], providing micronutrients in the human diet. The fresh-market tomato is one of the most consumed types of contemporary tomatoes [2], destined for fresh food ingredients such as tomato slice. There has been a significant improvement in the traits of the cultivated tomato, notably disease resistance and fruit quality [3]. Importantly, the fresh-market tomato is selected for

distinguishable fruit traits (such as fruit size/shape/firmness), which are desirable in the industry (consumer market and production systems) [3]. However, the genetic architecture (characteristics of DNA sequence variations responsible for traits) in the contemporary fresh-market tomato needs to be further explored [4]. This is evidenced by the recent findings, 1) the positive effects from both previously known loci, most likely associated with tomato domestication and/or historical improvement, and new associations for fresh-market tomato yield and 2) the phenotypic variations for important, but not yet fully investigated, traits such as flavor in the contemporary fresh-market tomato germplasm [4]. Given this, exploiting DNA sequence variations responsible for traits has been of interest in the (applied) tomato research community. Furthermore, contemporary germplasm accessions are in high demand as such resources can be beneficial for rapidly incorporating favorable genotype

*Correspondence:

Tong Geon Lee
tonggeonlee@gmail.com

¹ Agriculture and Life Sciences Research Institute, Kangwon National University, Chuncheon, Republic of Korea

² Gulf Coast Research and Education Center, University of Florida, Wimauma, FL, USA

³ Horticultural Sciences Department, University of Florida, Gainesville, FL, USA



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	Tomato sequence note	Excel file (.xlsx)	figshare (https://doi.org/10.6084/m9.figshare.21799967) [7]
Data file 2	Tomato sequence FastQC	HTML file (.html)	figshare (https://doi.org/10.6084/m9.figshare.25499491) [12]
Data set 1	Genome sequence data of tomato	FASTQ file (.fastq)	NCBI Sequence Read Archive (https://identifiers.org/ncbi/insdc.sra:SRP484668) [11]

combinations to achieve industry-driven traits (discussion in Bhandari et al. [4]). Whole genome sequence data had previously been used to identify the genetic recombination and DNA sequence variant-trait association in the contemporary fresh-market tomato [4–6], but sequence datasets (i.e., FASTQ files created using the Illumina platform) of previous studies were not published. To provide a rich genome sequence resource for future tomato research, we report the sequence datasets for 64 previously examined U.S. contemporary fresh-market tomatoes and 17 newly sequenced tomatoes.

Data description

We selected 81 tomatoes that were nominated by academic tomato breeding programs located in Florida and North Carolina, major fresh-market tomato-producing areas of the U.S. (Table 1, Data file 1) [7]. Of the 81 tomatoes, 68 were contemporary fresh-market tomatoes, whereas the remaining 13 were relevant fresh-market tomato breeding and germplasm accessions. Of the 13 accessions, 12 accessions showed various plant architectures, such as a *brachytic* plant-like short architecture, and one accession (PI 128654) was known to carry an original Tomato spotted wilt virus resistance gene. 64 tomato sequence datasets (FASTQ files) were previously generated [4–6]. To further diversify the genetic resources of this tomato class, we sequenced the genomes of additional 17 tomatoes (14 contemporary and three relevant fresh-market tomato breeding and germplasm accessions; [8, 9]) using the same technical conditions and quality control as described in our previous study [4]. Plants were grown in the greenhouse as previously described [10], and the leaf tissue was collected approximately six weeks after sowing. Total genomic DNA was extracted from a single plant per each tomato using a DNeasy Plant Mini Kit (Qiagen). The PCR-free, paired-end libraries (DNA fragment length was 350 bp) were prepared from the extracted DNA, and sequenced using the Illumina next-generation sequencing technology. Illumina raw reads with adapter contamination and/or uncertain nucleotides constitute (Ns; >10% of either read) were removed. Using the quality-controlled reads, we estimated genome coverage to be an average depth

per sequenced base of $24\times$ [with Fla. 7060 and Micro-Tom showing the highest ($31\times$) and lowest ($5\times$) depths, respectively] (Table 1, Data file 1, Data set 1) [7, 11]. All sequenced tomatoes passed FastQC's quality control (www.bioinformatics.babraham.ac.uk/projects/fastqc), both the mean quality score and per sequence quality score (Table 1, Data file 2) [12]. In addition, BWA (version 0.7.17; [13]) was used to map reads to the tomato reference genome sequence SL4.0 [14] to assess the mapping quality. The high mapping rate (>98%) and mapping quality score (>35) were calculated (Table 1, Data file 1) [7].

Limitations

The current genome sequence data was generated using the Illumina next-generation sequencing technology. Some sequence data might have substantial gaps in coverage likely across complex regions of genetic variation if aligning reads to a fully sequenced reference genome is applied. (Phased) long-read sequence data coupled with fully sequenced DNA molecules such as bacterial artificial chromosome and Fosmid can be required in order to discover sequence variants in such gaps (for example, Oxford Nanopore and Illumina NovaSeq technologies used to sequence-resolve the Fusarium wilt resistance gene introgression in this fresh-market tomato class [15]).

Abbreviations

BWA Burrows-Wheeler Aligner
 PCR Polymerase Chain Reaction
 WGS Whole Genome Sequencing

Acknowledgements

The authors thank members of the T.G.L. laboratory, especially Prashant Bhandari, Katherine Brown, and Claudia Jose, for laboratory assistance.

Authors' contributions

Conceptualization, TGL; investigation, JK & TGL; resources, JK & TGL; writing, JK & TGL; supervision, TGL. The authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The data described in this Data note can be freely and openly accessed on NCBI Sequence Read Archive SRP484668 (Data set 1) and figshare Datasets <https://doi.org/10.6084/m9.figshare.21799967> (Data file 1) and <https://doi.org/10.6084/m9.figshare.25499491> (Data file 2). Please see Table 1 and references [7, 11, 12] for details and links to the data.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 29 March 2024 Accepted: 24 June 2024

Published online: 02 July 2024

References

1. Food and Agriculture Organization of the United Nations (FAO). Agricultural production statistics. FAOSTAT, Rome, Analytical Briefs, (79) Rome. 2022. <https://doi.org/10.4060/cc9205en>.
2. U.S. Department of Agriculture. Tomatoes. 2016. www.ers.usda.gov/topics/crops/vegetables-pulses/tomatoes.
3. Scott JW, Myers JR, Boches PS, Nichols CG, Angell FF. Classical genetics and traditional breeding. In: Liedl BE, Labate JA, Stommel JR, Slade A, Kole C, editors. Genetics, genomics, and breeding of tomato. CRC Press; 2013. p. 37–74.
4. Bhandari P, Kim JH, Lee TG. Genetic architecture of fresh-market tomato yield. *BMC Plant Biol*. 2023;23:18. <https://doi.org/10.1186/s12870-022-04018-5>.
5. Bhandari P, Lee TG. A genetic map and linkage panel for the large-fruited fresh-market tomato. *J Am Soc Hortic Sci*. 2021;146(2):125–31. <https://doi.org/10.21273/JASHS04999-20>.
6. Bhandari P, Shekasteband R, Lee TG. A consensus genetic map and linkage panel for fresh-market tomato. *J Am Soc Hortic Sci*. 2022;147(1):53–61. <https://doi.org/10.21273/JASHS05110-21>.
7. Lee TG. Tomato sequence note figshare Dataset. 2024. <https://doi.org/10.6084/m9.figshare.21799967>.
8. Scott JW. University of Florida tomato breeding accomplishments and future directions. 1998;58:16–8.
9. Gardner RG. 'Mountain Spring' tomato; NC 8276 and NC 84173 tomato breeding lines. *J Amer Soc Hort Sci*. 1992;27:1233–4.
10. Lee TG, Hutton SF, Shekasteband R. Fine mapping of the brachytic locus on the tomato genome. *J Am Soc Hortic Sci*. 2018;143(4):239–47. <https://doi.org/10.21273/JASHS04423-18>.
11. NCBI Sequence Read Archive. Genome sequence data of tomato. 2024. <https://identifiers.org/ncbi/insdc.sra:SRP484668>.
12. Lee TG. Tomato sequence FastQC figshare Dataset. 2024. <https://doi.org/10.6084/m9.figshare.25499491>.
13. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinform*. 2010;26(5):589–95. <https://doi.org/10.1093/bioinformatics/btp698>.
14. Fernandez-Pozo N, Menda N, Edwards JD, Saha S, Teclé IY, Strickler SR, Bombarely A, Fisher-York T, Pujar A, Foerster H, Yan A, Mueller LA. The sol genomics network (SGN)—from genotype to phenotype to breeding. *Nucleic Acids Res*. 2015;43:D1036–41. <https://doi.org/10.1093/nar/gku1195>.
15. Lee TG. Long-read DNA sequencing leads to the more complete sequence characterization of the fruit size reducing region flanking a Fusarium wilt resistance gene. *Mol. Horticulture* 2022;2(16). <https://doi.org/10.1186/s43897-022-00037-w>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.