

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Transcriptome data for tissue-specific genes in four reproductive organs at three developmental stages of micro-tom tomato



Seon-Hwa Bae^a, Jihee Park^b, Soon Ju Park^c, Jungheon Han^d, Jae-Hyeon Oh^{d,*}

^a Department of Horticulture, Institute of Agricultural Science & Technology, Jeonbuk National University, Jeonju 54896, Korea

^b Upland Crop Breeding Research Division, National Institute of Crop Science, Rural Development Administration, 20, Miryang-si, Gyeongnam 50424, Korea

^c Division of Biological Sciences and Research Institute for Basic Science, Wonkwang University, Iksan, Republic of Korea

^d Gene Engineering Division, National Institute of Agricultural Sciences, Rural Development Administration, 370, Jeonju-si, Jeollabuk-do, 54874, Republic of Korea

ARTICLE INFO

Article history: Received 8 September 2020 Revised 15 December 2020 Accepted 4 January 2021 Available online 8 January 2021

Keywords: Micro-tom tomato Solanum lycopersicum Transcriptome Developmental stages Tissue-specific Gene expression Gene candidate

ABSTRACT

Tomato belongs to the Solanaceae family of plants. It is a diploid plant with 12 chromosomes. Previous studies have reported that its genome size is 950 MB with 35,000 proteincoding genes. Micro-Tom Tomato is a miniature dwarf determinate tomato cultivar. It has a small-sized genome, a short lifecycle, and a short seed-setting under fluorescent light. These features are similar to those of Arabidopsis. Consequently, Micro-Tom Tomato is considered as a model cultivar of tomato (*Solanum lycopersicum*) suitable for research. We sequenced its transcriptomes to identify tissue-specific gene candidate profiles in different plant tissues (petals, sepals, pistils, and stamens) at developmental stages.

© 2021 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

E-mail addresses: jhoh8288@korea.kr, jhoh8288@gmail.com (J.-H. Oh).

https://doi.org/10.1016/j.dib.2021.106715

^{*} Corresponding author at: Jae-Hyeon Oh, Gene engineering Division, National Institute of Agricultural Science, RDA, 370, Nongsaengmyeong-ro, Deokjingu, Jeonju-si, Jeollabuk-do, Republic of Korea.

^{2352-3409/© 2021} Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Specifications Table

Subject	Biology
Specific subject area	Transcriptomics
Type of data	Table, Figure
How data were acquired	Illumina Hiseq X
Data format	Raw sequences (FASTQ)
Parameters for data collection	 Reproductive organs of Micro-Tom Tomato, i.e., petal, sepal, pistil, and stamen.
	2. Three developmental stages, i.e., R1, the bud state,
	flowering start stage; R2, the full-bloom flowering period;
	and R3, the flower closure stage of falling flowers before
	the fruit is produced.
Description of data collection	The samples collected for transcriptome analysis were immediately
	frozen in liquid nitrogen and stored at -70 °C. For each sample, the
	experiments were repeated in triplicate under the same conditions.
Data source location	National Institute of Agricultural Science, Republic of Korea
Data accessibility	Raw data of the RNA-Seq are available in the Sequence Read Archive
	(SRA) and were deposited in the NCBI under bio-project accession
	number PRJNA659891
	(https://www.ncbi.nlm.nih.gov/bioproject/PRJNA659891).

Value of the Data

- These data are important for investigating tissue-specific expression of Micro-Tom Tomato.
- These data provide trait-related candidate genes in tomato.
- These transcriptome data can be used to identify genetic differences in reproductive organs between *Solanum lycopersicum* and subspecies as well as related phenotypes in TWAS (transcriptome-wide association studies) for research on breeding features (fruit number, size, and shape) at the molecular level.
- Phenotypic data (flower color, petal arrangement, and peat count) can be used to identify related marker traits.
- Information provided in this study can be used to determine molecular factors relevant to breeding such as those affecting tomato fruit formation and male sterility and identify associated SNP markers.

1. Data Description

Data presented in this article show differentially expressed genes between developmental stages of four tomato tissue samples (petals, sepals, pistils, and stamens) (Fig. 1). Transcriptomic data for each sample of Solanum lycopersicum were obtained by sequencing using an Illumina HiSeq X platform. Sequencing generated a total of 54,714,695 bp of paired-end data in FASTQ format. Sequencing data were deposited in NCBI Sequence Read Archive (Accession SUB8058188) under a bio-project accession number PRJNA659891 as shown in Table 1. Complete reference transcriptome has been employed for transcriptome assemblies previously [1]. Pre-processed reads were mapped to Solanum lycopersicum genome transcript using references obtained from Sol Genomics Network (https://solgenomics.net/organism/Solanum_lycopersicum/genome) [2,3]. Annotation was performed using the Sol Genomics Network database (DB). Of 35,768 standard genes used in the analysis, 27,392 were expressed, of which 27,392 (100%) genes had functional descriptions and percent (%) mean annotated gene values (Table 2 and Table S1). Trimmed reads were mapped to the transcript reference genome to obtain gene expression values. The average mapping rate was 87.25% (Table 3) based on HISAT analysis [4]. From sequences that passed the pre-processing process and mapped to clean reads, gene expression value for each tissue was calculated using HISeq [5]. Tissue-specific gene expression profiles of those transcripts in tissues

Table 1			
Sequencing data	of Micro-Tom	Tomato	transcriptome.

Sample name	Total length (bp) ^a	Accession link
R1 stage of petal1	2,189,364,847	https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=run_browser&run=SRR12560335
	2,063,691,153	
R2 stage of petal2	2,396,872,503	https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=run_browser&run=SRR12560334
	2,277,360,906	
R3 stage of petal3	2,115,025,526	https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=run_browser&run=SRR12560331
	2,009,649,471	
R1 stage of sepal1	2,489,405,347	https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=run_browser&run=SRR12560330
	2,357,158,492	
R2 stage of sepal2	2,327,853,296	https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=run_browser&run=SRR12560329
	2,205,392,002	
R3 stage of sepal3	2,180,206,125	https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=run_browser&run=SRR12560328
	2,069,454,324	
R1 stage of pistil1	1,880,758,438	https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=run_browser&run=SRR12560327
	1,789,393,852	
R2 stage of pistil2	2,297,041,939	https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=run_browser&run=SRR12560326
	2,176,359,982	
R3 stage of pistil3	2,293,205,403	https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=run_browser&run=SRR12560325
	2,176,601,195	
R1 stage of stamen1	2,755,513,327	https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=run_browser&run=SRR12560324
	2,671,128,527	
R2 stage of stamen2	2,600,248,153	https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=run_browser&run=SRR12560333
	2,514,497,620	
R3 stage of stamen3	2,483,916,013	https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=run_browser&run=SRR12560332
	2.394.596.748	

^a Base pair.



Fig. 1. Micro-Tom Tomato materials used for experiments. (A) Micro-Tom Tomato. (B) R1, as a bud state, (C) R2, the full bloom flowering period, (D) R3, the flower closure stage of falling flowers before the fruit is produced.

Table 2

Statistical results of tissue-specific gene clusters.

# of reference genes	# of expressed genes	Total annotation (Sol Genomics Network DB ^a)
35,768	27,392	27,392 (100%)

The number of reference genes, the number of expressed genes, and the total annotation. ^a Data base.

Fig. 2. Results of DEGs. Up-regulated and down-regulated genes were determined by comparing gene expression levels in different tissues. (A) petal vs. stamen. (B) sepal vs. stamen, and (C) pistil vs. stamen.

from tomato at different developmental stages were assessed with two selection criteria: 1) selection criteria for expressed genes, average normalized read counts from three repetitions \geq 500); 2) non-expressed gene selection criteria, average normalized read counts from three repetitions \leq 50. The analysis of genes with tissue-specific expression is shown in Table 4. Gene expression profiles for four tissues showed 616 significant genes (Table 4). In addition, we identified descriptions of these tissue-specific genes in the Sol Genomics Network Database (Table S2). Sequence data were used to identify differentially expressed genes (DEGs) in petals, sepals, pistils, and stamen. We identified DEGs in different tissues independently and classified them as up-regulated or down-regulated genes. Numbers of up-regulated genes in petal vs. pistil, sepal vs. pistil, and stamen vs. pistil were 7364, 9908, and 10,659, respectively. Numbers of downregulated genes petal vs. pistil, sepal vs. pistil, and stamen vs. pistil were 2994, 3051, and 2907, respectively (Fig. 2). Finally, a Venn diagram was created using a sample that commonly contains genes for down-regulation for the combination of DEGs of petal vs. pistil, sepal vs. pistil, and stamen vs. pistil (Fig. 3). Functional classification of these DEGs was performed using a gene

^{12,000} 10,659 9,908 10,000 Number of DEGs 8,000 7,364 6,000 4,000 2.994 3,051 2,907 2.000 0 Down Up Down Up Down Up A в С

Table 3						
Statistics	of r	eads	mapped	to	the	reference.

		Aligned	times	Aligned exactly 1 time		Aligned ≥ 1 time		Aligned (discordantly or single)		Mapping rate	
Sample ID	Total reads	Reads (each)	Percent (%)	Reads (each)	Percent (%)	Reads (each)	Percent (%)	Reads (each)	Percent (%)	Reads (each)	Percent (%)
R1 stage of petal1	16,228,567	1,977,546	12.19	11,941,434	73.58	157,754	0.97	2,151,832	13.26	14,251,021	87.81
R2 stage of petal2	17,687,223	2,332,128	13.19	12,820,346	72.48	176,324	1.00	2,358,424	13.33	15,355,095	86.81
R3 stage of petal3	15,687,678	2,015,073	12.84	11,340,974	72.29	154,337	0.98	2,177,293	13.88	13,672,605	87.16
R1 stage of sepal1	18,439,779	2,392,650	12.98	13,255,331	71.88	195,342	1.06	2,596,456	14.08	16,047,129	87.02
R2 stage of sepal2	17,213,216	1,924,299	11.18	12,928,657	75.11	185,119	1.08	2,175,140	12.64	15,288,917	88.82
R3 stage of sepal3	16,091,023	1,838,825	11.43	12,040,917	74.83	175,141	1.09	2,036,139	12.65	14,252,198	88.57
R1 stage of pistil1	13,888,577	1,857,155	13.37	9,953,884	71.67	146,040	1.05	1,931,497	13.91	12,031,422	86.63
R2 stage of pistil2	16,988,926	2,239,810	13.18	12,292,865	72.36	181,993	1.07	2,274,258	13.39	14,749,116	86.82
R3 stage of pistil3	16,950,140	2,082,582	12.29	12,444,246	73.42	181,805	1.07	2,241,506	13.22	14,867,558	87.71
R1 stage of stamen1	19,867,525	2,667,189	13.42	14,727,089	74.13	283,306	1.43	2,189,940	11.02	17,200,336	86.58
R2 stage of stamen2	18,705,506	2,548,267	13.62	13,851,656	74.05	265,738	1.42	2,039,845	10.91	16,157,239	86.38
R3 stage of stamen3	17,898,152	2,378,900	13.29	13,386,974	74.80	261,948	1.46	1,870,329	10.45	15,519,252	86.71
Total	205,646,312	26,254,424	12.75	150,984,373	73.38	2,364,847	1.14	26,042,659	12.73	179,391,888	87.25

Tissue	No. of genes
Peral	22
Sepal	97
Pistil	60
Stamen	437



The number of genes.



Fig. 3. Venn diagrams showing the number of differentially expressed genes (DEGs) for up-regulated and down-regulated genes in Micro-Tom Tomato tissues. (A) Up-regulated genes. (B) Down-regulated genes.

ontology (GO) analysis tool [6]. These DEGs were classified into three GO categories: biological processes, molecular functions, and cellular compartments (Tables S3 – S5). Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was also performed (Tables S3 – S5). In conclusion, this data set and reported DEGs can provide insight into the expression by tissue-specific of tomato species to analyze tissue-specific. Furthermore, these data can be used for functional genomic studies and genetic/genomic studies of tomato species.

2. Experimental Design, Materials and Methods

2.1. Plant samples

For research on molecular breeding factors such as male sterility, petal, sepal, pistil, and stamen tissue samples were obtained by dividing plant reproductive organs of Micro-Tom Tomato into three stages (R1, as the bud state, flowering start stage; R2, the full bloom flowering period, and R3, the flower closure stage of falling flowers before the fruit is produced.). Plant reproductive organs for these three stages are shown in Fig. 1. Samples collected for transcriptomic analysis were immediately frozen in liquid nitrogen and stored at -70 °C.

2.2. Illumina library preparation and transcriptome sequencing

All experimental procedures performed in this study strictly followed the standard protocol provided in each product manual. Complete sequence library preparation and transcriptome sequencing for the Illumina HiSeq X protocols were conducted by Macrogen, Inc. (Seoul, Korea) (http://www.macrogen.com), an authorized sequence service provider for all individual samples (petal, sepal, pistil, and stamen). Sequencing control software was used to input and read the sequencing file to an output file that was created as a paired-end FASTQ file. Total transcriptome short reads from each sample underwent pre-processing using LengthSort and DynamicTrim in the SolexaQA package [1]. DynamicTrim is a short-read trimmer that individually crops each read to its longest contiguous segment for which quality scores are greater than the user-supplied quality cutoff value. LengthSort is a program to separate high-quality reads from low-quality reads (e.g., if the read is less than 25 bp, it is excluded from the analysis process). From reads that passed the pre-processing process and were mapped to clean reads, gene expression value was calculated. The transcript sequence was mapped to the reference sequence using HISAT2 software [4]. The total number of reads mapped to each gene to measure the expression was counted with HTSeq v.0.11.0 [5].

2.3. Identification of candidate genes by tissue-specific expression

To perform differential expression analysis, we compared one tissue relative to each of the other three tissues (e.g., petal vs. stamen, sepal vs. stamen, and pistil vs. stamen). DEG analysis was performed to identify log2-fold changes from RPKM values (Table S1). We identified gene function using the Sol Genomics Network database (https://solgenomics.net/organism/Solanum_lycopersicum/genome) [2, 3]. To identify tissue-specific expression of genes, the selection criterion for genes expressed in tissues was average of three iterations with normalized read counts \geq 500 while that for genes not expressed in tissues was average of three iterations with normalized read counts \leq 50. Through this, genes specifically expressed in each tissue were selected (Table 4). The complete procedure was conducted by SEEDERS, Inc. (seeders.co.kr) of Daejeon, South Korea.

2.4. Functional analysis of differentially expressed genes (DEGs) using tomato stamen tissue

Up-regulated and down-regulated genes were determined by comparing gene expression values of corresponding genes in the stamen. The number of DEGs with a log2-fold change based on RPKM value is presented in Tables S3 – S5.

2.5. Gene ontology (GO) analysis

To analyze functions of genes with conserved tissue-specific expression, we performed a GO enrichment analysis (Tables S3 – S5). GO analysis was used to classify functions of transcripts in clusters. The analysis was carried out using GO tool [6]. The significance level was designated at p < 0.05.

2.6. Dataset

Complete sequences generated in this study were submitted to GenBank sequence read archive (SRA) under bio-project number ID PRJNA659891 as shown in Table 1.

CRediT Author Statement

Seon-Hwa Bae: writing - original draft, visualization, writing - review and editing, and data curation; **Jihee Park:** formal analysis, investigation, project administration, and funding acquisition; **Soon Ju Park:** resources, writing - review and editing; **Jungheon Han:** idea conceptualization, writing - review & editing; **Jae-Hyeon Oh:** conceptualization, visualization, writing - review & editing, resources, data curation, supervision, and project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by a grant from the Next-Generation BioGreen 21 Program (Project No. PJ PJ01389402) funded by Rural Development Administration, Republic of Korea.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2021.106715.

References

- M.P. Cox, D.A. Peterson, P.J. Biggs, SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data, BMC Bioinf. (2010) 485, doi:10.1186/1471-2105-11-485.
- [2] N. Fernandez-Pozo, N. Menda, J.D. Edwards, et al., The sol genomics network (SGN)-from genotype to phenotype to breeding, Nucleic Acids Res. 43 (2015) D1036–D1041, doi:10.1093/nar/gku1195.
- [3] L.A. Mueller, T.H. Solow, N. Taylor, et al., The SOL genomics network: a comparative resource for Solanaceae biology and beyond, Plant Physiol. 138 (2005) 1310–1317, doi:10.1104/pp.105.060707.
- [4] D. Kim, B. Langmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, Nature Methods 12 (2015) 357-360, doi:10.1038/nmeth.3317.
- [5] S. Anders, P.T. Pyl, W. Huber, HTSeq-a Python framework to work with high-throughput sequencing data, Bioinforma. Oxf. Engl. 31 (2015) 166–169, doi:10.1093/bioinformatics/btu638.
- [6] M. Ashburner, et al., Gene ontology: tool for the unification of biology. The gene ontology consortium, Nat Genet 25 (1) (2000) 25–29, doi:10.1038/75556.