Data in Brief 54 (2024) 110349

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Transcriptome dataset for comparative analysis of differentially expressed genes between wild type and transgenic potato plants overexpressing Nuclear Factor Y subunit A7

Bimpe Suliyat Azeez[a], Dool-Yi Kim[b], Jong-Kuk Na[a,c,*]

[a] *Department of Agriculture and Industries, Graduate School, Kangwon National University, Chuncheon 24341, South Korea*
[b] *National Institute of Crop Science, RDA, Wanju 55365, South Korea*
[c] *Department of SmartFarm and Agricultural Industries, Kangwon National University, Chuncheon, Kangwon 24341, South Korea*

ARTICLE INFO

Article history:
Received 16 January 2024
Accepted 14 March 2024
Available online 21 March 2024

Dataset link: StNF-YA7 RNA-sequencing (Original data)

Keywords:
Solanum tuberosum
Illumina NovaSeq 6000
RNA-seq
Transgene
Drought stress

ABSTRACT

Nuclear Factor Y (NF-Y) is divided into three different types of subunits, A, B, and C. NF-Ys play crucial roles in plants for controlling gene expression associated with various developmental processes and abiotic stresses, but it is mostly unknown the downstream genes regulated by NF-Ys in plant. One of the potato NF-Y genes, *StNF-YA7*, increased potato's drought tolerance when overexpressed under the control of constitutive *CaMV 35S* promoter. Therefore, it was of interest what genes are regulated by the increased expression level of *StNF-YA7*. To investigate the downstream genes of *StNF-YA7*, the transcriptome sequencing was carried out for four potato lines, including *Solanum tuberosum* L 'Superior' as wild type (WT), empty vector control (VC), and two *StNF-YA7* overexpressor lines (designated to *StNF-YA7* #19 & #26). The RNA sequencing data was produced by the Illumina NovaSeq 6000 sequencing system. The number of total raw reads obtained from the RNA sequencing was 36.7 million for WT, 36.2 for VC, 29.3 for StNF-YA7 #19, and 29.5 million for *StNF-YA7* #26,

* Corresponding author at: Department of Agriculture and Industries, Graduate School, Kangwon National University, Chuncheon 24341, South Korea.
E-mail address: jongkook@kangwon.ac.kr (J.-K. Na).

https://doi.org/10.1016/j.dib.2024.110349

respectively. The length of total raw reads for each sample was between 5.92 Gb (*StNF-YA7* #19) and 7.42 Gb (WT), and after filtering raw quality reads, the total length was between 5.81 Gb (*StNF-YA7* #19) and 7.29 Gb (WT). Each filtered clear read set of four transcriptomes was mapped on the potato reference genome, SolTub_3.0, and the percentage of mapped reads ranged from 89.8 % (VC) to 90.3 % (WT). GC contents range between 43.01 % (*StNF-YA7* #19) and 42.44 % (*StNF-YA7* #26). Q20 quality score ranges between 98.63 % (*StNF-YA7* #26) and 98.74 % (VC).

Specifications Table

Subject	Agronomy and Crop Science
Specific subject area	Potato Transcriptome analysis; RNA sequencing; potato NF-YA7
Data format	Raw data: fastq sequencing files, Filtered data
Type of data	Table, Figure
Data collection	RNA sequencing data was generated by paired-end sequencing using Illumina NovaSeq 6000 system (Illumina, USA), and FastQC (ver. 0.11.7) was used for quality check. Trimmomatic v0.38 was used to remove adapter sequences and trim poor quality bases. HISAT2 program (ver. 2.1.0) was used for mapping the clear reads on potato reference genome, SolTub_3.0.
Data source location	Kangwon National University, Chuncheon, Gangwon Province, South Korea
Data accessibility	Repository name: NCBI BioProject BioProject data accession number: PRJNA1044573
Related research article	Na JK, Kim KH, Seong ES, Kim BG, Cho KS, Cho JH, Park SK, Kim DY (2017) Overexpression of Nuclear Factor Y subunit StNF-YA7 enhances Drought Tolerance in Potato. Hortic. Environ. Biotechnol. 58(2): 170-177. DOI: 10.1007/s13580-017-0200-7

1. Value of the Data

- The dataset includes the transcriptomes of two transgenic potato lines overexpressing *StNF-YA7*, which can be used as the transcriptome reference to investigate the roles and the downstream genes of plant Nuclear Factor Y (NF-Y).
- The dataset provides a key information for a comparative analysis of differentially expressed genes between wild type potato and the *StNF-YA7* overexpressor lines.
- In addition, the dataset can be used to examine the functional analysis of genes regulated by *StNF-YA7* in potato, and also it could provide very useful information to narrow down specific genes associated with the drought tolerance in potato and relevant plant species.

2. Background

An increasing number of NF-Ys have been identified in plant species, and many of them have been investigated for their roles, revealing that they are involved in various developmental processes and abiotic stress responses. However, it is still largely unknown the downstream target genes of NF-Ys and their roles in plant. To identify the downstream genes of NF-Ys, the transcriptome analysis can be carried out on the transgenic plants with target NF-Y gene under suppression or overexpression. Potato *StNF-YA7* has been identified as a gene that enhances potato drought tolerance when overexpressed under the control of duplicated *CaMV 35S* promoter in the associated article, DOI: 10.1007/s13580-017-0200-7 [1]. Transgenic potato plants

Table 1
Summary of RNA sequencing data.

Description of sequence data	Sample id			
	WT	Empty vector	*StNF-YA7* #19	*StNF-YA7* #26
Total number of raw reads	36,715,767	36,198,186	29,306,007	29,518,791
Total length of raw read (bp)	7,416,584,934	7,312,033,572	5,919,813,414	5,962,795,782
GC content of raw reads (%)	43.02	43.1	43	43.42
Total number of filtered reads	36,230,369	35,640,379	28,855,243	28,981,749
Total length of filtered read (bp)	7,292,573,295	7,171,187,537	5,805,870,093	5,829,274,911
GC content of filtered reads (%)	43.04	43.11	43.01	43.44
Percentage of filtered quality reads	98.33	98.07	98.08	97.76
Number of mapped reads*	65,431,007	64,014,638	51,881,334	52,282,871
Overall mapping ratio (%)	90.3	89.81	89.9	90.2
Q20 (%)	98.84	98.74	98.79	98.63
Q30 (%)	95.82	95.52	95.67	95.24

* Mapping of the clear reads was carried out against the potato reference genome, SolTub_3.0.

overexpressing *StNF-YA7* was used for RNA sequencing to obtain differentially expressed genes by altered expression of *StNF-YA7*. The resulting transcriptome data can be useful to identify potential downstream genes of *StNF-YA7* that involved in drought tolerance of potato.

3. Data Description

The transcriptome data were produced from four different potato lines, including one-month-old wild type potato (WT), empty vector control (VC), and two transgenic lines overexpressing *StNF-YA7* (*StNF-YA7* #19 & *StNF-YA7* #26). The raw sequence read data were deposited at Sequence Read Archive (SRA) database of NCBI BioProject under the accession of PRJNA1044573. SRA accession numbers for four transcriptome data are SRR26945113 (WT), SRR26945114 (*StNF-YA7* #26), SRR26945115 (*StNF-YA7* #19), and SRR26945116 (VC), respectively. The raw and filtered read data were summarized in Table 1. Based on the mapping of the filtered reads on potato reference genome SolTub_3.0, the number of differentially expressed genes (DEGs) are obtained and are represented in the Venn diagrams (Fig. 1). In total 1239 and 2276 DEGs were identified

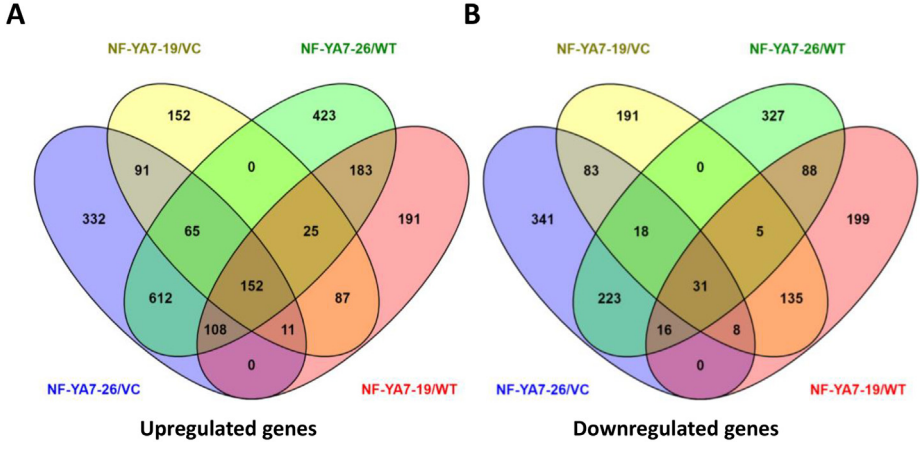

Fig. 1. Venn diagrams showing the comparison of differentially expressed genes in two *StNF-YA7* transgenic plants compared to empty vector control (VC) or wild type *Solanum tuberosum* L 'Superior' (WT). **A**. Upregulated genes **B**. Downregulated genes.

from *StNF-YA7* #19 and *StNF-YA7* #26 compared to wild type, respectively. On the other hand 1054 and 2091 DEGs were found in *StNF-YA7* #19 and *StNF-YA7* #26 compared to vector control.

4. Experimental Design, Materials and Methods

4.1. Sample collection

Plant samples included *Solanum tuberosum* 'Superior' as wild type (WT), empty vector control plant (VC) with *2 × 35S:smGFP*, and two transgenic lines (*StNF-YA7* #19 & #26) with *2 × 35S:StNF-YA7* gene. Potato plants were grown *in vitro* for four weeks in the magenta box at 22 °C under a 13 h light/11 h dark photoperiod of growth chamber. Whole plant samples were collected, immediately frozen in liquid nitrogen, and stored at −80 °C. Total RNA was extracted from each sample using GeneAll Ribospin plant reagent (GeneAll Biotechnology Co., Ltd, Seoul, South Korea) according to the manufacturer's instructions. RNA purity was checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA).

4.2. Library construction and production of sequencing data

RNA concentration was measured again using Quan-it^{TM} RiboGreen RNA Assay kit (Invitrogen, cat. #R11490). The integrity of the total RNA was assessed to obtain RNA integrity number (RIN) of each sample using the TapeStation RNA screentape (Agilent, #5067-5576), and only high-quality RNA samples with RIN greater than 7.0 were used for RNA library construction. For library construction, 1 ug of total RNA was used to purify mRNA using Illumina TruSeq Stranded mRNA Sample Prep Kit (Illumina, Inc., San Diego, CA, USA, #RS-122-2101). Briefly, the purified mRNA was sheared, and the sheared mRNA fragments were used to prepare the first strand cDNA using SuperScript II reverse transcriptase with random primers (Invitrogen, #18064014). The second strand cDNA synthesis was followed using DNA Polymerase I, RNase H, and dUTP. Then, an end repair and 'A' tailing processes were carried out. The processed cDNAs were purified and enriched with PCR to create the final cDNA library. The libraries were quantified using KAPA Library Quantification kits (Kapa Biosystems, #KK4854) and checked for quality using the TapeStation D1000 ScreenTape (Agilent Technologies, # 5067-5582). Finally, the libraries were used for the paired-end (2 × 100 bp) sequencing by an Illumina NovaSeq 6000 system (Illumina, Inc., San Diego, CA, USA), from which paired-end sequencing reads were generated and trimmed by removing adapter sequences and bases with poor base quality using Trimmomatic v0.38 [2].

4.3. Differentially expressed genes (DEGs)

The clear reads were aligned to the Solanum tuberosum reference genome (SolTub_3.0) using HISAT (ver. 2.1.0) implemented with Bowtie2 [3]. The reference genome sequence and gene annotation data were downloaded from NCBI Genome assembly and NCBI RefSeq database, respectively. Aligned data (SAM file format) were sorted and indexed using SAMtools (ver. 1.9). After alignment, the transcripts were assembled and quantified using StringTie v2.1.3b [4]. Differential gene expression was analysed by edgeR software (ver. 3.40.2) [5]. The statistical significance of differential expressed gene (DEG) was determined using edgeR exactTest. Fold change and *p*-value were extracted from the result of exactTest. DEGs with significance were selected on the basis of the criteria: |fold change|≥2 and raw *p*-value < 0.05.

Limitations

None.

Ethics Statement

The current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

Data Availability

StNF-YA7 RNA-sequencing (Original data) (NCBI Bioproject).

CRediT Author Statement

Bimpe Suliyat Azeez: Data curation, Writing – original draft; **Dool-Yi Kim:** Resources, Data curation; **Jong-Kuk Na:** Conceptualization, Supervision.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (Grant number: 2021R1I1A3047975132218210213010 3).

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.

References

[1] J.-K. Na, K.H Kim, E.-S. Seong, B.-G. Kim, K.S. Cho, J.H. Cho, S.-K Park, D.-Y. Kim, Overexpression of Nuclear Factor Y Subunit StNF-YA7 enhances drought tolerance in potato, Hortic. Environ. Biotechnol. 58 (2) (2017) 170–177, doi:10.1007/s13580-017-0200-7.
[2] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for illumine sequence data, Bioinformatics 30 (15) (2014) 2114–2120, doi:10.1093/bioinformatics/btu170.
[3] D. Kim, B. Langmead, S.L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, Nat. Methods 12 (4) (2015) 357–360, doi:10.1038/nmeth.3317.
[4] M. Pertea, G.M. Pertea, C.M. Antonescu, T.-C. Chang, J.T. Mendell, S.L. Salzberg, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads, Nat. Biotechnol. 33 (3) (2015) 290–295, doi:10.1038/nbt.3122.
[5] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (1) (2010) 139–140, doi:10.1093/bioinformatics/btp616.