![data logo] **data**

Transcriptome Dataset of Soybean (*Glycine max*) Grown under Phosphorus-Deficient and -Sufficient Conditions

Hengyou Zhang [1,2], Shanshan Chu [1] and Dan Zhang [1,]*

[1] Collaborative Innovation Center of Henan Grain Crops, College of Agronomy, Henan Agricultural University, Zhengzhou 450002, China

[2] Department of Biological Sciences, University of North Carolina at Charlotte, Charlotte, NC 28223, USA

* Correspondence: zhangdan8006@163.com

Abstract: This data descriptor introduces the dataset of the transcriptome of low-phosphorus tolerant soybean (*Glycine max*) variety NN94-156 under phosphorus-deficient and -sufficient conditions. This data is comprised of the transcriptome datasets (four libraries) acquired from roots and leaves of the soybean plants challenged with low-phosphorus, which allows further analysis whether systemic tolerance response to low phosphorus stress occurred. We describe the detailed procedure of how plants were prepared and treated and how the data were generated and pre-processed. Further analyses of this data would be helpful to improve our understanding of molecular mechanisms of low-phosphorus stress in soybean.

Data Set: https://www.ncbi.nlm.nih.gov/sra/?term=SRP097715

Data Set License: There is no specific license.

Keywords: soybean; *Glycine max*; low-phosphorus stress; RNA-seq

---

1. Introduction

Phosphorus (P) is an essential component of fertilizers in a soil that is critical for plant production. Because there is no known chemical or technological substitute for P in either natural ecosystems or in the agroecosystems and because the mobility of P in soil is very limited [1], P has been a global constraint on crop productivity in some crop-growing areas, especially regions without sufficient P content in soil and financial support to obtain P-containing fertilizers. The use of fertilizers can provide season-long P for crop growth, but it is not an economically effective method for farmers. The development and use of crops with higher P-use efficiency is an economical and environmentally friendly method of providing sustainable crop production [2].

Soybean (*Glycine max* (L.) Merri) is one of the most important crops, providing about half of the global demand for vegetable oils and proteins. It has been demonstrated that the growth and development of soybean require more P compared with other crops such as rice, corn, and wheat [3]. In soybean, low-P stress may decrease soybean nodule development, increase flower/pod abscission, and impair overall plant growth, consequently limiting yield and seed quality [4–7]. Thus, low-P stress is more problematic than other nutrient deficiencies or toxicities in soybean [8].

Soybean tolerance to low-P stress is a complex trait involved in a number of genes, some of these genes may have small effects. Knowledge of a genetic and molecular basis of soybean resistant to low-P stress, thus far, has been obtained from the identification of a number of quantitative trait loci (QTLs) associated with P-efficiency [4,5,9] and a major gene, *GmACP1*, encoding an acid phosphatase involved in regulating P efficiency [5]. Despite great efforts, it remains challenging to pinpoint P-efficiency

genes underlying previously identified QTLs that have a relatively large confidence interval. Here, we expanded our previous studies [4,5,9] by conducting a transcriptome analysis of a low-P tolerant Nannong 94-156, under phosphorus-deficient and -sufficient conditions. Our objective was to better understand the genetic and molecular basis of low-P tolerance in soybean. Further studies might functionally identify the candidate genes and introduce the low-P tolerance genes into soybean.

2. Result

A total of four RNA libraries (Roots+P, Leaves+P, Roots−P, and Leaves−P) for four conditions were prepared and sequenced. Transcriptome sequencing generated approximate 21.1 million (M), 23.5 M, 21.3 M, and 22.3 M reads for Roots+P, Leaves+P, Roots−P, and Leaves−P conditions, respectively. In this dataset, 94% of the samples had quality scores of greater than 30. Preprocess analyses using TopHat [10] showed that approximately 77%–86% of high-quality reads were mapped to unique locations on the reference soybean genome. The sequencing and preprocessing results are summarized in Table 1. An initial comparative analysis of the transcriptome dataset has revealed that acid phosphatases might be involved in enhancing P efficiency in low-P tolerance in soybean, and gene expression analysis and functional analysis has indicated the robustness of the dataset [11]. However, further analyses using the data is needed—for example, identifying the underlying alternative splicing genes and regulatory network as well as identifying the differentially expressed genes systemically involved in low-P tolerance in roots and leaves—which will further improve our comprehensive understanding of the molecular mechanism underlying plant adaptations to phosphate deficiency.

Table 1. Summary of sequencing reads and transcriptome statistics.

Samples	Raw Reads		Filtered Reads		NCBI SRA ID	Bioproject ID
	Total Reads	Q30	Total Reads	Q30		
Roots+P	21,096,949	86.42%	18,064,338	94.95%	SRR5281856	PRJNA356948
Roots−P	21,305,403	85.85%	18,258,283	95.33%	SRR5281855	PRJNA356948
Leaves+P	23,577,620	87.37%	20,519,365	95.29%	SRR5281858	PRJNA356948
Leaves−P	22,291,144	86.07%	18,967,411	94.93%	SRR5281857	PRJNA356948

Archived Data Accessible to Users

This is the first publicly available dataset of soybean transcriptome change to low-P stress. Filtered sequence reads for the four libraries were submitted to the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/) sequence read archive (SRA) database in compressed fastq format. The sequencing data is available with accession numbers SRR5281855-58. This dataset could be optionally retrieved using an SRA Toolkit (http://www.ncbi.nlm.nih.gov/sra) and could be transformed into tool-specified data formats. Further analysis of the data set could be performed using software tools and databases summarized at OmicTools (https://omictools.com).

3. Materials and Methods

3.1. Plant Material and Treatment

Soybean cultivar, NN94-156 (B20), is a low-P tolerance soybean that has been used for linkage mapping in our previous studies, and several QTLs with a high explanation of phenotypic variation have been identified [4,5]. Soybean treatment for phosphorus-deficient and -sufficient conditions were followed as previously described [5,11]. Briefly, seeds were surface-sterilized with 0.5% sodium hypochlorite for no more than 4 minutes, rinsed twice with sterile water, and then germinated in sterile vermiculite. The seedlings with fully expanded cotyledons were transferred into a 60-hole hydroponic tank (70 × 50 × 30 cm) [12] filled with 50% Hoagland's nutrient solution [13] (pH 5.8) supplemented with 500 μM (+P, KH_2PO_4), with two plants per hole. Three days after transplanting,

60 plants were transferred to a separate tank containing 50% Hoagland's nutrient solution lacking P supply (−P, 5 μM P), and the remaining 60 plants remained in +P solution as controls. Planting more plants than actually needed for the RNA-seq assays allows us to select phenotypically identical plants for sample collection, which is helpful in reducing variation between individuals. All plants were placed in the hydroponics box using a completely randomized block design. The used solution was replaced with the corresponding fresh solution every three days. Whole roots and the uppermost mature trifoliate leaf were separately collected from controls and treated plants, respectively, at 1, 3, 7, and 14 days after transferring for treatment and flash frozen in liquid nitrogen and stored at −70 °C in a refrigerator until they were used. We chose these four time points because we found that *GmACP1* [5], an acid phosphatase-encoding gene that contributes to soybean tolerance to low-phosphorus stress, was significantly upregulated after treatment at all of the four time points, with the most dramatic difference occurring at 7 days after treatment (data not shown). To capture a maximum of transcriptome change associated with low-P tolerance, we collected the tissues from treatments and controls, respectively, at the four time points and pooled them for RNA-seq assay. Tissues collected from three phenotypically identical plants at each time were pooled and powdered, and an equal amount of tissues (by weight) from four-time points per condition were pooled for RNA extraction. Therefore, a total of four samples including roots (+P), roots (−P), leaves (+P), and leaves (−P) were prepared. All plants were grown under controlled conditions (10 h light/14 h dark, day/night temperature of 28/20 °C) in a growth chamber.

3.2. RNA Isolation, Library Preparation, and RNA Sequencing

Total RNA was extracted using Trizol Kit according to manufacturer's protocol (Promega, Fitchburg, WI, USA). Library construction, RNA integrity, purity, and concentration were assessed using an Agilent 2100 Bioanalyzer, (Agilent Technologies, Santa Clara, CA, USA). Sequencing libraries were prepared from 1 μg RNA per sample using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (New England Biolabs, Ipswich, MA, USA) and NEBNext® Multiplex Oligos for Illumina Kit (New England Biolabs, Ipswich, MA, USA) as previously described [14]. The final quantified libraries were sequenced using Illumina HiSeq™ 2500, and 100-bp paired-end raw reads were generated.

3.3. Data Preprocessing

The quality of all reads was accessed by running the FastQC program (version 0.11.5, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). For quality control of raw reads, a user preference-based Perl program was written to select clean reads by removing low-quality sequences (there were more than 50% bases with quality lower than 20 in one sequence), reads with more than 5% N bases (bases unknown) and reads containing adaptor sequences. Alternatively, similar results in quality control could also be obtained by using several traditional bioinformatics tools, such as Trimmomatic [15], FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), and cutadapt [16]. The clean reads from each library were aligned to the Williams 82 reference soybean genome (Wm82.a2.v1) using TopHat [10] with minimum intron size i = 30 and the maximum intron size I = 15000 as previously described [17], and the rest of parameters were set as default. The identification of differentially expressed genes could be achieved using one of the following toolkits, such as cufflinks [18], DESeq [19], and EdgeR [20] following the users' manual. For those who are not familiar with command-line operations, these bioinformatics tools could also be run in user-friendly web interfaces, such as Galaxy (https://usegalaxy.org) and Discovery Environment at CyVerse (http://www.cyverse.org).

Author Contributions: Dan Zhang and Hengyou Zhang performed the data analysis; Shanshan Chu performed the experiment; Hengyou Zhang and Dan Zhang wrote the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Metson, G.S.; Wayant, K.A.; Childers, D.L. Introduction to P Sustainability. In *Phosphorus, Food, and Our Future*; Wyant, K.A., Corman, J.E., Elser, J.J., Eds.; Oxford University Press: Oxford, UK, 2013.
2. Gaxiola, R.A.; Sanchez, C.A.; Paez-Valencia, J.; Ayre, B.G.; Elser, J.J. Genetic manipulation of a "Vacuolar" H^+-PPase: From salt tolerance to yield enhancement under phosphorus-deficient soils. *Plant Physiol.* **2012**, *159*, 3–11. [CrossRef] [PubMed]
3. Li, X.; Chang, W.; Zhang, C. Advances of soybean (*Glycine max L.*) phosphorus nutrition and high P-efficient germplasms screening in China. *Soybean Sci.* **2011**, *30*, 322–327.
4. Zhang, D.; Liu, C.; Cheng, H.; Kan, G.; Cui, S.; Meng, Q.; Gai, J.; Yu, D. Quantitative trait loci associated with soybean tolerance to low phosphorus stress based on flower and pod abscission. *Plant Breed.* **2010**, *129*, 243–249. [CrossRef]
5. Zhang, D.; Song, H.N.; Cheng, H.; Hao, D.R.; Wang, H.; Kan, G.Z.; Jin, H.X.; Yu, D.Y. The acid phosphatase-encoding gene GmACP1 contributes to soybean tolerance to low-phosphorus stress. *PLoS Genet.* **2014**, *10*, e1004061. [CrossRef] [PubMed]
6. Cassman, K.G.; Whitney, A.S.; Stockinger, K.R. Root-growth and dry-matter distribution of soybean as affected by phosphorus stress, nodulation, and nitrogen-source. *Crop Sci.* **1980**, *20*, 239–244. [CrossRef]
7. Olivera, M.; Tejera, N.; Iribarne, C.; Ocana, A.; Lluch, C. Growth, nitrogen fixation and ammonium assimilation in common bean (*Phaseolus vulgaris*): Effect of phosphorus. *Physiol. Plant.* **2004**, *121*, 498–505. [CrossRef]
8. Jones, G.D.; Lutz, J.A.; Smith, T.J. Effects of phosphorus and potassium on soybean nodules and seed yield. *Agro J.* **1977**, *69*, 1003–1006. [CrossRef]
9. Zhang, D.; Cheng, H.; Geng, L.Y.; Kan, G.Z.; Cui, S.Y.; Meng, Q.C.; Gai, J.Y.; Yu, D.Y. Detection of quantitative trait loci for phosphorus deficiency tolerance at soybean seedling stage. *Euphytica* **2009**, *167*, 313–322. [CrossRef]
10. Trapnell, C.; Pachter, L.; Salzberg, S.L. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **2009**, *25*, 1105–1111. [CrossRef] [PubMed]
11. Zhang, D.; Zhang, H.; Chu, S.; Li, H.; Chi, Y.; Triebwasser-Freese, D.; Lv, H.; Yu, D. Integrating QTL mapping and transcriptomics identifies candidate genes underlying QTLs associated with soybean tolerance to low-phosphorus stress. *Plant Mol. Biol.* **2017**, *93*, 137–150. [CrossRef] [PubMed]
12. Li, H.Y.; Yang, Y.M.; Zhang, H.Y.; Chu, S.S.; Zhang, X.G.; Yin, D.M.; Yu, D.Y.; Zhang, D. A genetic relationship between phosphorus efficiency and photosynthetic traits in soybean as revealed by QTL analysis using a high-density genetic map. *Front. Plant Sci.* **2016**, *7*, 924. [CrossRef] [PubMed]
13. Hoagland, D.R.; Arnon, D.I. *The Water-culture Method for Growing Plants without Soil. Circular*, 2nd ed.; California Agricultural Experiment Station: Berkeley, CA, USA, 1950.
14. Zou, H.D.; Tzarfati, R.; Hubner, S.; Krugman, T.; Fahima, T.; Abbo, S.; Saranga, Y.; Korol, A.B. Transcriptome profiling of wheat glumes in wild emmer, hulled landraces and modern cultivars. *BMC Genom.* **2015**, *16*, 777. [CrossRef] [PubMed]
15. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
16. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2011**, *17*, 10–12. [CrossRef]
17. Shen, Y.T.; Zhou, Z.K.; Wang, Z.; Li, W.Y.; Fang, C.; Wu, M.; Ma, Y.M.; Liu, T.F.; Kong, L.A.; Peng, D.L.; et al. Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell* **2014**, *26*, 996–1008. [CrossRef] [PubMed]
18. Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D.R.; Pimentel, H.; Salzberg, S.L.; Rinn, J.L.; Pachter, L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **2012**, *7*, 562–578. [CrossRef] [PubMed]
19. Anders, S.; Huber, W. Differential expression analysis for sequence count data. *Genome. Biol.* **2010**, *11*, R106. [CrossRef] [PubMed]

20. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. EdgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [CrossRef] [PubMed]