



OPEN

DATA DESCRIPTOR

RNAseq of diverse spring wheat cultivars released during last 110 years

Saman Maqbool^{1,2}, Samar Naseer², Nageen Zahra³, Fatima Rasool³, Humaira Qayyum², Khawar Majeed², Muhammad Jahanzaib^{2,4}, Muhammad Sajjad⁵, Muhammad Fayyaz⁶, Muhammad Kashif Naeem³, Muhammad Ramzan Khan³, Hao Zhang^{1,7}, Awais Rasheed^{1,2}✉ & Huihui Li^{1,7}✉

Here, we performed RNA-seq based expression analysis of root and leaf tissues of a set of 24 historical spring wheat cultivars representing 110 years of temporal genetic variations. This huge 130 tissues RNAseq dataset was initially used to study expression pattern of 97 genes regulating root growth and development in wheat. Root system architecture (RSA) is an important target for breeding stress-resilient and high-yielding wheat cultivars under climatic fluctuations. However, root transcriptome analysis is usually obscured due to challenges in root research due to their below ground presence. We also validated the dataset by performing correlation analysis between expression of RSA related genes in roots and leaves with 25 root traits analyzed under varying moisture conditions and 10 yield-related traits. The Pearson's correlation coefficients between root phenotypes and expression of root-specific genes varied from -0.72 to 0.78 , and strong correlations with genes such as *DRO1*, *TaMOR*, *ARF4*, *PIN1* was observed. The presented datasets have multiple uses such as a) studying the change in expression pattern of genes during time, b) differential expression of genes in two very important tissues of wheat i.e., leaf and roots, and c) studying customized expression of genes associated with important phenotypes in diverse wheat cultivars. The initial findings presented here provided key insights into understanding the transcriptomic basis of phenotypic variability of RSA in wheat cultivars.

Background & Summary

Bread wheat (*Triticum aestivum*) is one of the most important staple food crops providing 55% of carbohydrates to the world population. The grain yield of wheat has to increase at an average annual rate of $\sim 2\%$ in a limited area of cultivated land to meet the world food demand¹. A deeper understanding of wheat genetics is required to address the primary challenge of sustaining food security in the context of climate change to feed the growing population. It is critical to deepen the knowledge of the wheat genomics and its genetic composition as well as the broad range of sequencing and transcriptomics data to understand genetic basis of wheat adaptability to target environments². Identification and functional characterization of genes that regulate developmental stages critical for withstanding climatic fluctuations is an important aspect of this area of research. Similarly, it is central to functional genetic studies to analyze dynamic expression patterns of each gene contributing to plant development in various tissues and response to various environmental stimuli³.

Roots are significant for the production of food grains such as wheat and rice⁴. A variety of morphological and physiological traits expressed by root systems facilitate the uptake of water and nutrients. Similar to above-ground traits, there must be an understanding of unique root system architecture (RSA) for optimum

¹State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences (CAAS), CIMMYT-China office, Beijing, China. ²Department of Plant Sciences, Quaid-i-Azam University, Islamabad, 45320, Pakistan. ³National Institute of Genomics and Advanced Biotechnology (NIGAB), National Agriculture Research Center (NARC), Islamabad, Pakistan. ⁴Oilseeds Research Program, Crop Sciences Institute (CSI), National Agriculture Research Center (NARC), Islamabad, Pakistan. ⁵Department of BioSciences, COMSATS University, Islamabad, Pakistan. ⁶Crop Disease Research Institute, National Agriculture Research Center (NARC), Islamabad, Pakistan. ⁷Nanfan Research Institute, CAAS, Sanya, Hainan, China. ✉e-mail: arasheed@qau.edu.pk; lihuihui@caas.cn

Name	Year of Release	Pedigree
T9	1911	Landrace
C518	1933	T9/8 A
C217	1944	C 516 X C 591
C271	1957	C230/IP165
Drik	1958	C-271/WILLET-DWARF//SONORA-64
Maxipak-65	1965	PJ/GB55 or PJ62/GB55
Photowar-70	1970	BURT/KENYA//QUETA(L)/3/NAD63
Pari-73	1973	CNO67//SN64/KLRE/3/8156
Pak-81	1981	KVZ/BUHO//KAL/BB
Barani-83	1983	BB/GLL/3/GTO/7 C//BB/CNO67
Chakwal 86	1986	FORLANI/ACC//ANA
Rawal 87	1987	MAYA/MON//KVZ/TRM
Pasban-90	1990	INIA F66/TH.DISTICHUM//INIAF66/3/GENARO T81
Inquilab-91	1991	LR64A/NAI60
Parwaz-94	1994	C271/WT(E)/SN64
Punjab-96	1996	URES/BOW'S
Chakwal-97	1998	BUC'S/FCT'S'
G.A.2002	2002	DWL5023/SNB//SNB
Seher-2006	2006	CNO67//SN64/KLRE/3/8156
Chakwal-50	2008	ATTILA/3/HUI/CARC//CHEN/CHTO/4/ATTILA
NARC-2009	2009	INQALAB 91*2/TUKURU
Dharabi-2011	2011	HXL-7573/2*BAGULA//PASTOR
Pakistan-13	2013	MEX94.27.1.20/3/Sokoll//Attila/3*BCN
Ujala-2016	2016	KIRITATI/4/2*WEAVER/TSC//WEAVER/3/WEAVER

Table 1. List of historical spring wheat cultivars with release year and pedigree.

resource acquisition⁵. Since roots are important components of breeding programs, it is crucial to understand the molecular mechanisms involved in root formation especially under challenging conditions.

In order to pinpoint the genetic components influencing the root growth in maize, rice and other crops, a variety of forward and reverse genetics techniques including transcriptomics and functional genomics have been applied⁶. The transcriptome studies using next-generation sequencing (NGS) technologies have paved the way in linking genotype to phenotype and can detect the molecular mechanisms underlying plant responses to abiotic stress⁷. Presently, several population-wide transcriptome analyses have been conducted in cereal crops including rice⁸, wheat⁹, and barley¹⁰. These studies unravelled the associations between gene expression and traits; however, field studies have generally been restricted to transcriptomics of above-ground shoots due to the challenge of sampling root tissues in field conditions. RNA-seq previously known as whole transcriptome shotgun sequencing has excitingly shaped whole transcriptome profiling⁷. It can identify transcript levels, expressed polymorphisms, and splicing isoforms. The development of high-throughput next-generation RNA-seq technologies provides new insights into transcriptome analysis such as a detailed expression profile, higher sensitivity to genes expressing at both high and low extremes, and no limitation by the lack of prior genome knowledge¹¹. RNA-seq studies in wheat are increasing rapidly owing to the reconstruction of the entire transcriptome using the short paired-end (PE) assembly of de novo reads¹² and provide a precise measurement of transcript levels. In wheat, some large-scale RNAseq studies available where transcriptome of multiple tissues from a single cultivar are reported like in Chinese Spring and Azhurnaya¹³. In this study, we conducted transcriptome profiling using RNA-seq on a set of 24 bread wheat varieties with diverse phenotypes supported by their large-scale phenotypic variation in agronomic and RSA traits^{14,15}. We initially analyzed the dataset to identify expression variation of potential transcripts or genes involved in RSA and validated by correlation analysis with RSA phenotypes.

Methods

Plant material. A panel of 24 historical spring wheat cultivars released in Pakistan was selected for this study. The cultivar name, year of release, and pedigree are given in Table 1. These cultivars selected based on the year of release to represent the cultivated diversity over the course of 110 years.

Growth and RNA isolation. The seeds of 24 wheat cultivars were surface sterilized using 3% NaOCl and were sown in triplicates in plastic trays containing peat moss. Two weeks after germination (at Zadoks stage 2), seedling leaf and root tissues were collected and subjected to total RNA extraction. RNA extraction was performed using EasyPure Plant RNA Kit (ER301-01) following the instructions provided by manufacturer and quantified using Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, USA).

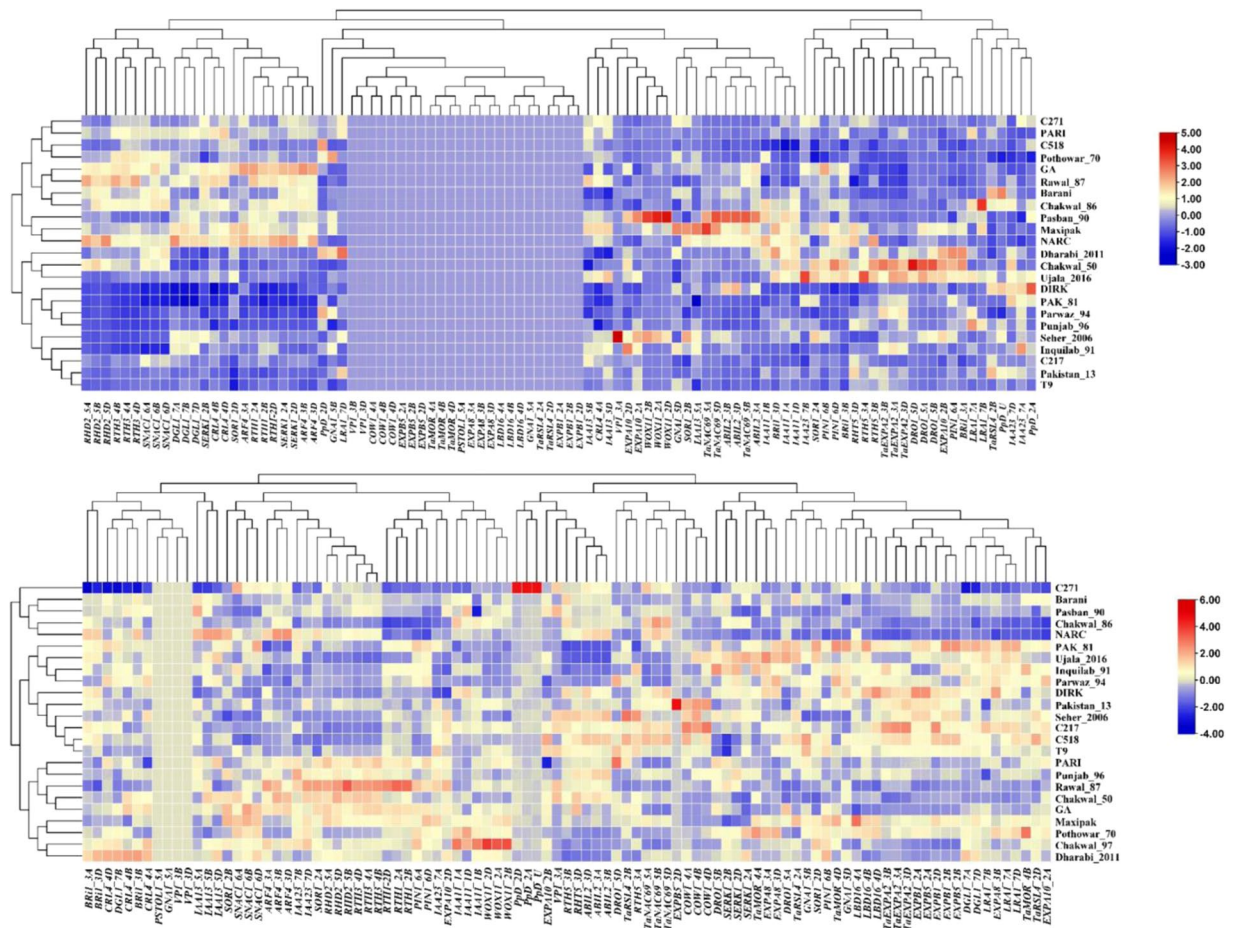


Fig. 1 Differential expression of RSA-related genes in (A) leaf and (B) root tissues of 24 bread wheat cultivars. The original gene names have been used while Traes IDs can be found in the associated excel file available at FigShare under <https://doi.org/10.6084/m9.figshare.23292389>.

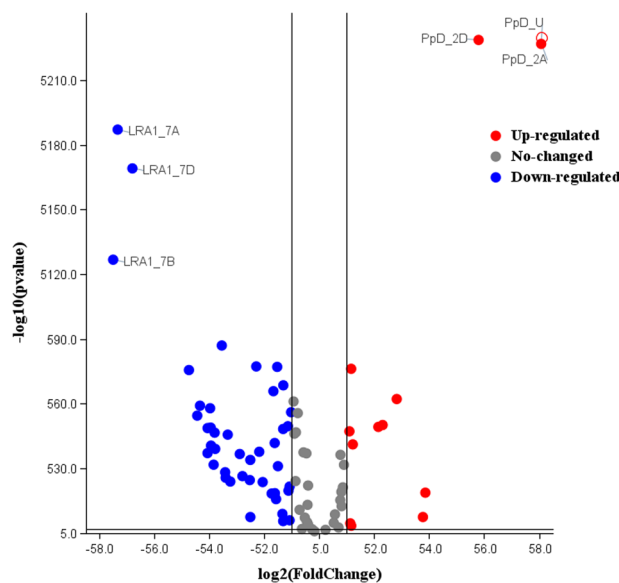


Fig. 2 Volcano plot showing up- and down-regulation (leaf vs root) of root-related genes. Only top six genes are labelled.

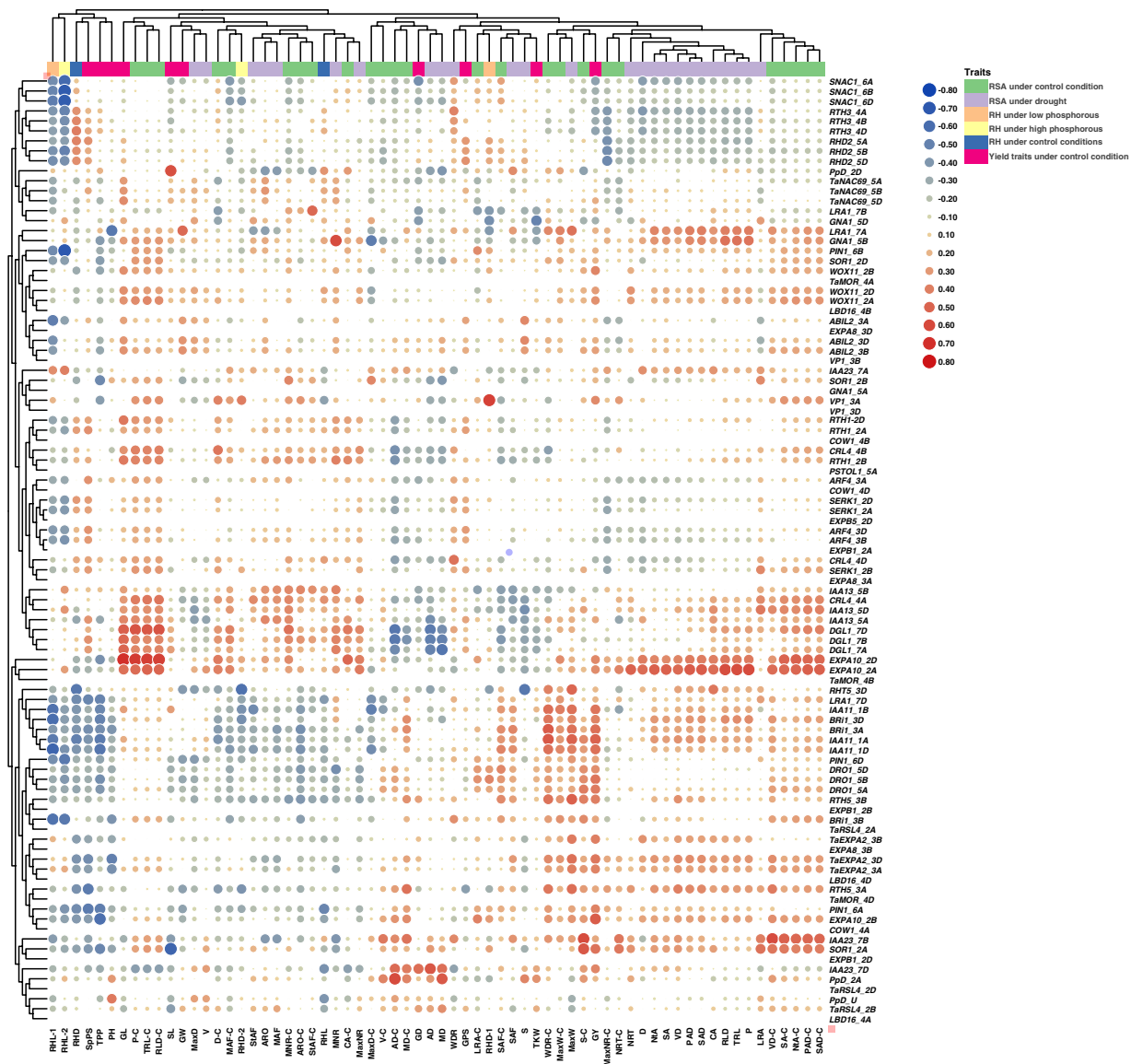


Fig. 3 Heatmap showing significant correlations between expression of RSA-related genes in leaf, yield-related traits, and root traits under control and drought stress conditions. The size of the circle explains the extent of correlation. Traits are abbreviated as; Maximum weight (MaxW), Maximum diameter (MaxD), Lower root area (LRA), median number of roots (MNR), steep angle frequency (StAF), solidity (S), volume diameter (VD), surface area (SA), network area (NtA), projected area diameter (PAD), surface area diameter (SAD), median angle frequency (MAF), average root orientation (ARO), shallow angle frequency (SAF), depth (D), width to depth ratio (WDR), maximum number of roots (MaxNR), number of root tips (NRT), volume (V), perimeter (P), total root length (TRL), root length diameter (RLD), convex area (CA), average diameter (AD), median diameter (MD), spikes per spike (SpPS), plant height (PH), tillers per plant (TPP), grain yield (GY), grain length (GL), thousand kernel weight (TKW), grain density (GD), grains per spike (GPS), spike length (SL), and grain weight (GW).

RNA Sequencing and identification of differentially expressed genes. The RNA samples were sequenced from Beijing Genomics Institute (BGI), China. For cDNA synthesis, the oligo (dT) method was used. The 50-bp single-end sequencing libraries were constructed, and BGISEQ-500 platform was used for sequencing using standard protocols. ‘Clean data’ was produced as FastQ data files using SOAPnuke version 2.1.6. Mapping with reference genome of bread wheat¹⁶ was done using HISAT2 software v 2.2.1¹⁷. Bowtie software was used for alignment of reference sequence with reads¹⁸. The reads were then quantified using featureCounts software and differentially expressed genes (DEGs) were identified using DeSEQ_2 in R v 4.1.1. The threshold value for filtering of DEGs was set at 0.1. All the DEG files were then culminated into a single file used for further analysis¹⁹. The R codes were used to generate heatmaps directly from the normalized count file²⁰, or phenotypic data from the diversity panel was used to calculate correlation values and plot correlation values as heatmaps²⁰.

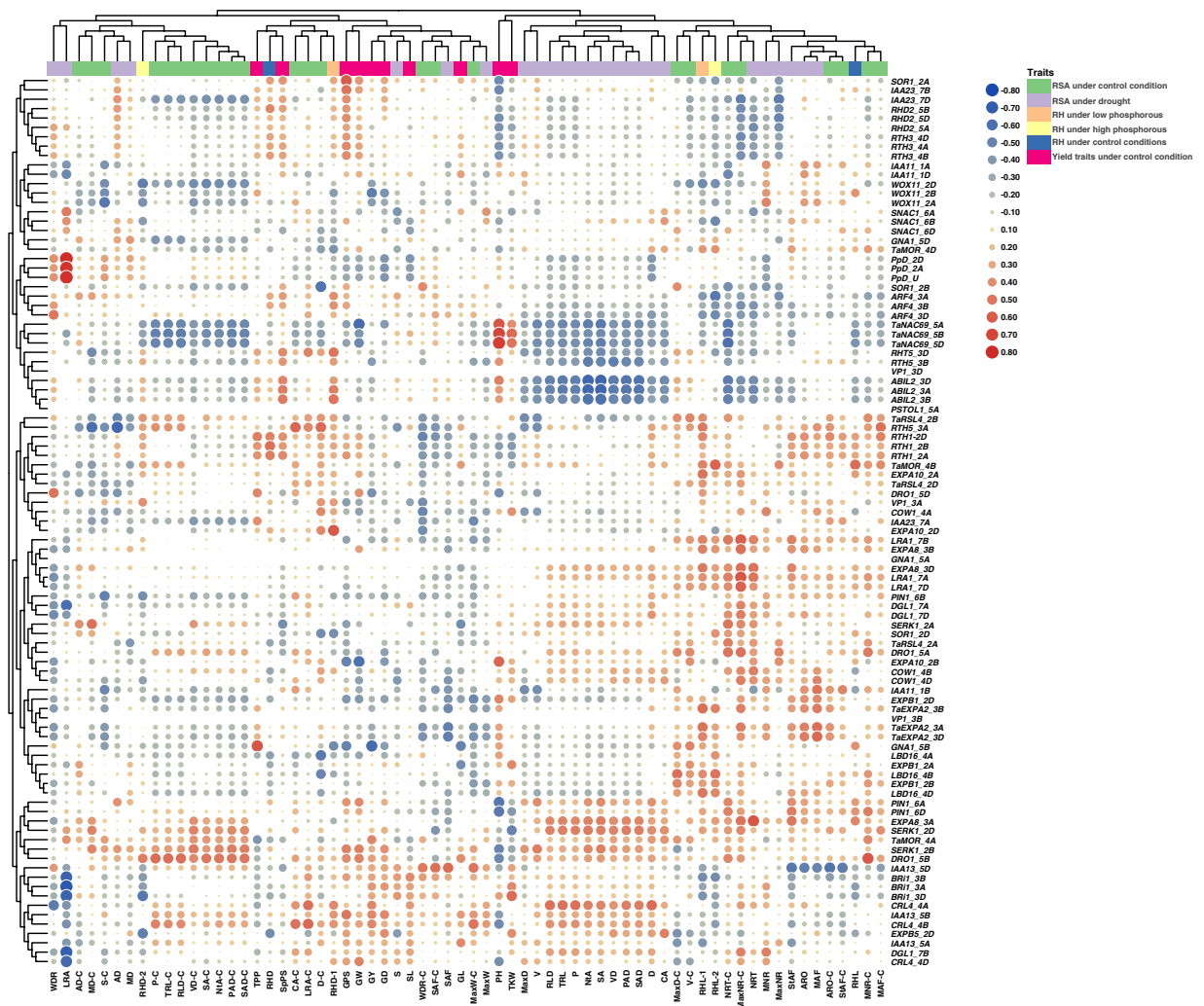


Fig. 4 Heatmap showing significant correlations between expression of RSA-related genes in roots, yield-related traits, and root traits under control and drought stress conditions. The size of the circle explains the extent of correlation. Traits are abbreviated as; Maximum weight (MaxW), Maximum diameter (MaxD), Lower root area (LRA), median number of roots (MNR), steep angle frequency (StAF), solidity (S), volume diameter (VD), surface area (SA), network area (NtA), projected area diameter (PAD), surface area diameter (SAD), median angle frequency (MAF), average root orientation (ARO), shallow angle frequency (SAF), depth (D), width to depth ratio (WDR), maximum number of roots (MaxNR), number of root tips (NRT), volume (V), perimeter (P), total root length (TRL), root length diameter (RLD), convex area (CA), average diameter (AD), median diameter (MD), spikes per spike (SpPS), plant height (PH), tillers per plant (TPP), grain yield (GY), grain length (GL), thousand kernel weight (TKW), grain density (GD), grains per spike (GPS), spike length (SL), and grain weight (GW).

Phenotyping for agronomic traits and root system architecture. The agronomic traits of the diversity panel were taken from our previous experiment²¹. Briefly, the diversity panel was planted at five locations and important agronomic traits were recorded. The phenotyping for RSA architecture traits has been described in detail¹⁴. The imaging platform consisting of RhizoVision crown hardware²² controlled by RhizoVision Imager software was used for root image acquisition and details have been described previously¹⁴. The RSA traits included in the study were maximum weight (MaxW), maximum diameter (MaxD), lower root area (LRA), median number of roots (MNR), steep angle frequency (StAF), solidity (S), volume diameter (VD), surface area (SA), network area (NtA), projected area diameter (PAD), surface area diameter (SAD), median angle frequency (MAF), average root orientation (ARO), shallow angle frequency (SAF), depth (D), width to depth ratio (WDR), maximum number of roots (MaxNR), number of root tips (NRT), volume (V), perimeter (P), total root length (TRL), root length diameter (RLD), convex area (CA), average diameter (AD), and median diameter (MD). The correlation between gene expression and various traits including RSA traits, root hair length and density under low and high phosphorous treatments, and yield-related traits was determined using 'psych' package in R version 4.2.1.

Data Records

In total, 130 RNA-seq datasets were generated which are deposited to the SRA repository of NCBI under BioProject PRJNA863398²³. The gene list of differentially expressed genes (DEGs) from each tissue is submitted to the Gene Expression Omnibus (GEO) repository under accession number GSE235844²⁴. The raw count data table from RNA-seq data of each tissue was generated and converted to normalized tpm values and deposited at FigShare²⁰ under <https://doi.org/10.6084/m9.figshare.23292389> and DryAd¹⁹ under <https://doi.org/10.5061/dryad.zs7h44jcs>.

The summary statistics of raw reads from leaf and root tissues are given in Supplementary Table 1.

Technical Validation

RNA quality was initially determined using NanoDrop and samples <1.8 values of OD260/280 and OD260/230 were further processed for RNA integrity. RNA integrity was assessed with Agilent 2100 Bioanalyzer RNA Nano assay (Agilent Technologies, USA). The average RIN values were 7.3 ± 0.5 for leaf samples and 6.9 ± 0.4 for root tissues. The qRT-PCR validation of selected genes was also performed.

Quality assessment. The aligned FastQ data files were read using SAMTools software and basic statistics are presented in Supplementary Table 1. The average quality score which is ratio between the sum of base qualities and total length was >60 in all cases. The error rate which is mismatches per bases mapped was <0.001.

Initial analysis of differentially expressed genes. The gene list of differentially expressed genes (DEGs) from each tissue is submitted to the Gene Expression Omnibus (GEO) repository under accession number GSE235844. A total of 38 RSA related genes were identified from various cereal species including wheat, rice, maize, and barley from the published literature. The blastn analysis was carried out to identify their A-, B- and D-sub-genome homeologues of those genes. This process identified 95 homeologues in wheat which were then used in the subsequent analyses. The gene IDs, names, description and GO ontology are given in Supplementary Table 2. The GO enrichment analysis was performed using Triticeae-Gene Tribe, a homology database²⁵ (Supplementary Table 3). Initially, heatmaps were generated for 95 genes using gene IDs (TraesIDs) for tpm values in leaf (Fig. 1A) and root tissues (Fig. 1B). The expression profiles of the selected genes extracted from the main DEG file using the R code²⁰. The heatmaps were generated using TBTools. In leaf tissues, all genes were expressed except *TaMOR*, *PSTOL1*, *EXPA8*, *LBD16*, *EXPB1*, *COW1*, and *EXPB5*. No expression of *VPI* gene was observed in leaf tissues in all varieties except Seher-2006 which showed highest expression of *VPI_3A* (Fig. 1A). In contrast to leaf tissues, all genes exhibited differential expression patterns in roots except three homoeologues of *Ppd* gene and D homoeologue of *EXPB5* (Fig. 1B). Volcano plot showing differentially expressed gene across 24 cultivars is presented as Fig. 2. The expression of these genes was differential in across root tissues. Leaf gene expression was found significantly correlated with yield traits (Fig. 3). The differential expression of RSA concerning genes in roots was significantly correlated to all 25 root traits under optimum conditions with some variations (Fig. 4). The dataset and initial analysis proved very effective in culminating the differential gene expression in root tissues and due to the large number of samples, it was possible to associate gene expression data with the phenotypes. In conclusion, the differential expression of these genes in the roots provided a validation of the dataset.

Code availability

FigShare R code used to generate Figs. 3, 4 is available at <https://doi.org/10.6084/m9.figshare.23292389> along with the source file.

FigShare R code to generate heatmap of any given wheat gene IDs (using TraesIDs) using the normalized count data file is available at <https://doi.org/10.6084/m9.figshare.23292389>. The code is also available on GitHub repository as: <https://github.com/plantbiologyqau/R-code-for-gene-expression-heatmap/tree/main>

Received: 28 June 2023; Accepted: 20 November 2023;

Published online: 08 December 2023

References

- Wang, J. *et al.* Transcriptome analysis in roots and leaves of wheat seedlings in response to low-phosphorus stress. *Sci. Reports* 2019 9, 1–12 (2019).
- Aury, J. M. *et al.* Long-read and chromosome-scale assembly of the hexaploid wheat genome achieves high resolution for research and breeding. *Gigascience* 11, (2022).
- Pearce, S. *et al.* WheatExp: An RNA-seq expression database for polyploid wheat. *BMC Plant Biol.* 15, (2015).
- Gewin, V. Food: An underground revolution. *Nature* 466, 552–553 (2010).
- Maqbool, S. *et al.* Root system architecture in cereals: progress, challenges, and perspective. *Plant J.* <https://doi.org/10.1111/tj.15669> (2022).
- Deja-Muyile, A., Parizot, B., Motte, H. & Beekman, T. Exploiting natural variation in root system architecture via genome-wide association studies. *J. Exp. Bot.* 71, 2379–2389 (2020).
- Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63 (2009).
- Kawakatsu, T. *et al.* The transcriptomic landscapes of rice cultivars with diverse root system architectures grown in upland field conditions. *Plant J.* 106, 1177–1190 (2021).
- Rasool, F. *et al.* Transcriptome unveiled the gene expression patterns of root architecture in drought-tolerant and sensitive wheat genotypes. *Plant Physiol. Biochem. PPB* 178, 20–30 (2022).
- Osthoff, A., Donà Dalle Rose, P., Baldauf, J. A., Piepho, H. P. & Hochholdinger, F. Transcriptomic reprogramming of barley seminal roots by combined water deficit and salt stress. *BMC Genomics* 20, (2019).
- Oshlack, A., Robinson, M. D. & Young, M. D. From RNA-seq reads to differential expression results. *Genome Biology* vol. 11 (2010).
- Vicente, R. *et al.* De Novo Transcriptome Analysis of Durum Wheat Flag Leaves Provides New Insights Into the Regulatory Response to Elevated CO₂ and High Temperature. *Front. Plant Sci.* 10, 1 (2019).

13. Ramírez-González, R. H. *et al.* The transcriptional landscape of polyploid wheat. *Science* (80-). **361**, (2018).
14. Maqbool, S. *et al.* Root system architecture of historical spring wheat cultivars is associated with alleles and transcripts of major functional genes. *BMC Plant Biol.* **22**, (2022).
15. Shaukat, M. *et al.* Genetic gain for grain micronutrients and their association with phenology in historical wheat cultivars of Pakistan released between 1911 and 2016 in Pakistan. *Agronomy* **11**, (2021).
16. Appels, R. *et al.* Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* (80-). **361**, (2018).
17. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
18. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
19. Maqbool, S. & Rasheed, A. RNA-seq based expression of DEGs in roots and leaves of bread wheat, *Dryad*, <https://doi.org/10.5061/dryad.zs7h44jcs> (2022).
20. Rasheed, A. R codes and source files for wheat DEGs. *FigShare* <https://doi.org/10.6084/m9.figshare.23292389> (2022).
21. Hanif, U. *et al.* Genetic gain and G×E interaction in bread wheat cultivars representing 105 years of breeding in Pakistan. *Crop Sci.* **62**, 178–191 (2022).
22. Seethepalli, A. *et al.* RhizoVision Crown: An Integrated Hardware and Software Platform for Root Crown Phenotyping. *Plant Phenomics* **2020**, 1–15 (2020).
23. *NCBI Sequence Read Archive* <https://identifiers.org/ncbi/insdc.sra:SRP388693> (2023).
24. Rasheed, A. & Maqbool, S. *GEO*. <https://identifiers.org/geo/GSE235844> (2023).
25. Chen, Y. *et al.* A collinearity-incorporating homology inference strategy for connecting emerging assemblies in the triticeae tribe as a pilot practice in the plant pangenomic era. *Molecular Plant* **13**, 1694–1708 (2020).

Acknowledgements

This work was supported by the National Science Foundation of China (32022064), Nanfan special project, CAAS, Grant No: ZDXM2309 and YDLH2301, and Innovation Program of Chinese Academy of Agricultural Sciences.

Author contributions

S.M.: Experimentation, data analysis, manuscript writing. S.N.: Data analysis, revision of manuscript. N.Z.: Data analysis, DEGs estimation, heatmap preparation. F.R.: Data analysis, Sequence alignment, DEGs, Heatmap, manuscript writing. H.Q.: Data preparation, Phenotyping experiment, Manuscript editing, Data analysis. K.M.: Phenotyping experiment, Manuscript editing, Data analysis. M.J.: Conceived the idea, editing manuscript, funding, M.S.: Data analysis, DEGs estimation, heatmap preparation. M.F.: Data preparation, Phenotyping experiment, Manuscript editing, Data analysis. M. K. N.: Manuscript editing. M.R.K.: Revised the manuscript. H.Z. Data analysis, writing codes. A.R.: Conceived the idea and supervised, editing manuscript, funding. H.L.: Conceived the idea and supervised, editing manuscript, funding.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-023-02769-w>.

Correspondence and requests for materials should be addressed to A.R. or H.L.

Reprints and permissions information is available at www.nature.com/reprints.

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



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 Author(s) 2023