

StCoExpNet: a global co-expression network analysis facilitates identifying genes underlying agronomic traits in potatoes

Venkata Suresh Bonthala & Benjamin Stich

Article - Version of Record



Suggested Citation:

Bonthala, V. S., & Stich, B. (2024). StCoExpNet: a global co-expression network analysis facilitates identifying genes underlying agronomic traits in potatoes. *Plant Cell Reports*, 43(5), Article 117.
<https://doi.org/10.1007/s00299-024-03201-2>

Wissen, wo das Wissen ist.



UNIVERSITÄTS- UND
LANDESBIBLIOTHEK
DÜSSELDORF

This version is available at:

URN: <https://nbn-resolving.org/urn:nbn:de:hbz:061-20241129-121545-1>

Terms of Use:

This work is licensed under the Creative Commons Attribution 4.0 International License.

For more information see: <https://creativecommons.org/licenses/by/4.0>



StCoExpNet: a global co-expression network analysis facilitates identifying genes underlying agronomic traits in potatoes

Venkata Suresh Bonthala¹ · Benjamin Stich^{1,2,3,4}

Received: 29 November 2023 / Accepted: 18 March 2024 / Published online: 15 April 2024
© The Author(s) 2024

Abstract

Key message We constructed a gene expression atlas and co-expression network for potatoes and identified several novel genes associated with various agronomic traits. This resource will accelerate potato genetics and genomics research.

Abstract Potato (*Solanum tuberosum* L.) is the world's most crucial non-cereal food crop and ranks third in food production after wheat and rice. Despite the availability of several potato transcriptome datasets at public databases like NCBI SRA, an effort has yet to be put into developing a global transcriptome atlas and a co-expression network for potatoes. The objectives of our study were to construct a global expression atlas for potatoes using publicly available transcriptome datasets, identify housekeeping and tissue-specific genes, construct a global co-expression network and identify co-expression clusters, investigate the transcriptional complexity of genes involved in various essential biological processes related to agronomic traits, and provide a web server (StCoExpNet) to easily access the newly constructed expression atlas and co-expression network to investigate the expression and co-expression of genes of interest. In this study, we used data from 2299 publicly available potato transcriptome samples obtained from 15 different tissues to construct a global transcriptome atlas. We found that roughly 87% of the annotated genes exhibited detectable expression in at least one sample. Among these, we identified 281 genes with consistent and stable expression levels, indicating their role as housekeeping genes. Conversely, 308 genes exhibited marked tissue-specific expression patterns. We exemplarily linked some co-expression clusters to important agronomic traits of potatoes, such as self-incompatibility, anthocyanin biosynthesis, tuberization, and defense responses against multiple pathogens. The dataset compiled here constitutes a new resource (StCoExpNet), which can be accessed at <https://stcoexpnet.julius-kuehn.de>. This transcriptome atlas and the co-expression network will accelerate potato genetics and genomics research.

Keywords Transcriptome atlas · Co-expression network · *Solanum tuberosum* · Agronomic traits · Tuberization

Communicated by Maike Petersen.

✉ Venkata Suresh Bonthala
bonthala@hhu.de

- ¹ Institute of Quantitative Genetics and Genomics of Plants, Heinrich Heine University of Düsseldorf, Düsseldorf, Germany
- ² Present Address: Julius Kühn-Institut (JKI), Institute for Breeding Research On Agricultural Crops, Rudolf-Schick-Platz 3a, OT Groß Lüsewitz, 18190 Sanitz, Germany
- ³ Max Planck Institute for Plant Breeding Research, Köln, Germany
- ⁴ Cluster of Excellence On Plant Sciences, From Complex Traits Towards Synthetic Modules, Düsseldorf, Germany

Introduction

Potato (*Solanum tuberosum* L.) is a highly heterozygous autotetraploid species and is the world's most crucial non-cereal food crop (Bao et al. 2022). It ranks third in food production after wheat and rice, with an annual global production exceeding 376 million tons (FAO 2021). Biotechnological techniques have gained traction due to the escalating food demand and global climate change, fueled by the expanding human population, to generate better cultivars (Iizumi et al. 2014). To develop improved cultivars, researchers have employed diverse omics approaches, which have been instrumental in augmenting crop productivity (Yang et al. 2021). A milestone in potato omics-based research was the availability of several reference-quality,

chromosome-scale and haplotype-resolved genome assemblies, which helped in understanding the complexity and evolution of the potato genome (Potato Genome Sequencing Consortium 2011; Tang et al. 2022; Sun et al. 2022; Hoopes et al. 2022; Bao et al. 2022; Freire et al. 2021; Leisner et al. 2018; Zhou et al. 2020). These potato whole-genome sequencing projects have also contributed to the significant rise in potato transcriptome studies and reported spatiotemporal changes occurring in various potato tissues using RNA-seq (e.g., Massa et al. 2011; Chandrasekar et al. 2022; Tiwari et al. 2020; Pieczynski et al. 2018; Chen et al. 2019; Cao et al. 2020; Tai et al. 2020).

The Potato Genome Sequencing Consortium (2011) reported the sequencing of many tissues of two potato genotypes, such as DM1-3 516 R44 (DM) and RH89-039-16 (RH), under diverse stress conditions. Numerous studies ensued to investigate transcriptional dynamics, such as those covering various biotic and abiotic conditions and cultivars. For example, Massa et al. (2011) used 32 DM RNA-Seq libraries and quantified expression levels of 60% of DM genes under biotic and abiotic stress conditions. Tiwari et al. (2020) investigated the transcriptome of potato tissues generated under varying nitrogen supplies. Their results suggested that the genes from the glutaredoxin gene family, among others, played an important role in conferring nitrogen stress tolerance to potatoes. Chen et al. (2019) analyzed the transcriptional responses upon drought, rehydration and re-dehydration in the drought-tolerant potato landrace Jancko Sisu Yari. They observed that the drought- and rehydration-responsive genes are mainly involved in flavonoid, lipid and sugar metabolism, among others. Chandrasekar et al. (2022) investigated the transcriptional dynamics between resistant and susceptible cultivars against potato cyst nematode (PCN) to identify resistant mechanisms induced by PCN. They identified several disease-resistance genes and transcription factors (TFs) up-regulated in a resistant cultivar (Kufri Swarna).

The availability of plant transcriptomic data in public databases like the Sequence Read Archive (SRA) at the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/sra>) has led to the creation of consolidated collections or atlases. These have been developed for several crop species, including *Oryza sativa* (Xia et al. 2017), *Solanum lycopersicum* (Fernandez-Pozo et al. 2017), and *Glycine max* (Machado et al. 2020). These atlases are contributing to understanding the global transcriptional dynamics across cultivars/genotypes/landraces under various stress conditions or between tissues and deciphering the molecular mechanisms that govern biological processes. However, despite the availability of the data at public databases like NCBI SRA from several potato transcriptome studies, an effort has yet to be put into developing a global transcriptome atlas for potatoes.

Housekeeping (HK) genes are those genes expressed relatively stable across all tissue types under various conditions (Czechowski et al. 2005; Bustin et al. 2009). Several of these genes have also been used as internal reference genes in potato real-time quantitative polymerase chain reaction (qPCR) assays. However, many genes considered as HK genes do not exhibit uniform expression across various experimental conditions (Nicot et al. 2005; Hu et al. 2009; Tang et al. 2017). Hence, choosing appropriate reference genes is critical in potato qPCR assays. With the emergence of next-generation sequencing technology, RNA-Seq data can be used to evaluate commonly used reference genes and propose new ones (Yim et al. 2015; Machado et al. 2020). Although numerous transcriptome datasets are available at public repositories such as NCBI SRA from several potato transcriptome studies, no attempt has yet been made to assess the commonly used reference genes and identify new ones to improve the precision of potato qPCR assays across various experimental conditions.

Tissue-specific (TS) genes are those expressed and function in a specific tissue preferentially over the other tissues. Identifying these genes helps better understand tissue-gene relationships (Xiao et al. 2010). For example, the combinatory action of MADS and AP/ERF family transcription factors regulates the development of distinct floral parts in *Arabidopsis thaliana* (Chi et al. 2017). Machado et al. (2020) identified several TS genes specific to nodules, endosperm and flowers in soybean using many RNA-Seq datasets. Despite the availability of numerous RNA-Seq datasets, a systematic identification of TS genes in potato is lacking.

Gene co-expression networks (GCN) provide a robust method to explore transcriptomic data. These networks are undirected graphs of nodes that correspond to genes and are interconnected by edges based on significant co-expression between them, representing transcriptionally coordinated genes often involved in the same biological process (Stuart et al. 2003). GCNs are effective tools in functional genomics as they enable the inference of putative gene functions and regulatory mechanisms through gene co-expression (Ballouz et al. 2015). Additionally, GCNs permit the simultaneous identification and classification of numerous genes with similar expression patterns (Serin et al. 2016). For example, GCNs have been employed in specific areas of plant research, such as investigating the genetic basis of plant natural products (Wisecaver et al. 2017), nitrogen metabolism for plant growth (Gaudinier et al. 2018), cell wall development (Rao et al. 2019), and resistance responses to powdery mildew (Zhang et al. 2016a). GCNs have been constructed and explored gene co-expressions to understand the transcriptional regulation of various biological processes in several plants, such as *Arabidopsis thaliana* (Burks et al. 2022), *Oryza sativa* (Sircar et al. 2022), *Zea mays* (Yu et al.

2017), *Hordeum vulgare* (Lee et al. 2020), and *Glycine max* (Almeida-Silva et al. 2020).

A few GCNs have also been constructed for potatoes using publicly available transcriptomic datasets. Massa et al. (2011) constructed a GCN using RNA-Seq data from 32 DM libraries. They identified 18 co-expression clusters, representing genes with highly correlated expression profiles in a biological process. Ramšak et al. (2018) constructed a GCN using two microarray datasets to understand immune signaling in potatoes better. They discovered a link between ethylene (ET) and salicylic acid (SA) signaling pathways. Specifically, they found that activating the ET signaling module via the Ethylene Insensitive3 gene triggers the expression of the Nonexpressor of PR Genes1, a critical regulator of the SA pathway. Yan et al. (2018) constructed a GCN using 16 RNA-Seq datasets covering 11 cultivars to investigate the resistance of potatoes. This GCN analysis revealed that 134 genes were significantly enriched and exhibited high levels of co-expression in Andigena, particularly concerning potato disease and stress resistance. This finding highlighted the significant impact of evolutionary pressures during artificial potato domestication. In addition, several studies used GCNs to investigate transcriptional regulation under various stress conditions using datasets generated in respective studies. Qin et al. (2022) unraveled cultivar-specific rooting depth responses to drought stress in potatoes using GCN. Despite the growing availability of gene expression datasets (RNA-Seq) that provide unbiased representations of gene expression patterns across various potato cultivars worldwide, the GCN analyses conducted so far have focused on case-control experiments to address specific objectives or have used small datasets. This limited approach has hindered the ability to uncover the global transcriptional landscape of potatoes in different tissues and conditions.

The objectives of our study were to (i) construct a global expression atlas for potatoes using publicly available transcriptome datasets, (ii) identify housekeeping and tissue-specific genes, (iii) construct a global co-expression network and identify co-expression clusters, (iv) investigate the transcriptional complexity of genes involved in various essential biological processes related to agronomic traits, and (v) provide a web server to easily access the newly constructed expression atlas and co-expression network to investigate the expression and co-expression of genes of interest.

Materials and methods

Potato genome and annotation data

We used the genomic sequence and annotation data for the potato reference genome, dAg, from our recent study (Bonthala and Stich 2022). The gene annotation

contained 39,088 and 53,352 genes and transcripts, respectively. We used the gene annotation's exon-intron boundaries (gff3 format) as a reference guide in read mapping. From the annotation data, we used functional annotation such as gene description, gene ontology (GO) terms, Pfam domains, InterProScan descriptions, and Arabidopsis ortholog descriptions.

Potato RNA-Seq data, processing and quality control

We searched the NCBI SRA database (<https://www.ncbi.nlm.nih.gov/sra>) for potato transcriptome datasets. We exported the metadata using Run Selector (as of June 2022) with the following parameters: AssayType: RNA-Seq, LibrarySource: TRANSCRIPTOMIC, Organism: *Solanum tuberosum*, Common name: potato, and Platform: Illumina and BGISEQ. In addition, we also searched extensively for additional potato transcriptome datasets in the literature (as of June 2022), and we added the metadata of new datasets to exported metadata of NCBI. Using this metadata, we downloaded experiment details using NCBI e-fetch (Leinonen et al. 2010). Using these experiment details, we excluded samples showing technical issues, such as empty FASTQ files, pair-end samples with single-end reads, and pair-end samples with an unequal length of reads. Finally, we downloaded 3,227 SRA files and converted them into FASTQ files using SRA-TOOLKIT v3.0.0 (Leinonen et al. 2010). We performed the quality assessment of FASTQ using FastQC v0.12.1 (Andrews 2010). We removed the low-quality reads, i.e., those with an average base quality of less than 20 or containing adapter sequences, using Trimmomatic v0.39 (Bolger et al. 2014). We inferred the library strandedness for each sample by applying the approach presented by Zheng et al. (2019). This approach involves mapping 100,000 reads for each sample using Kallisto v0.46.1 (Bray et al. 2016) onto dAg genome under all three library types (–fr-stranded, –fr-stranded and none) separately, followed by comparing the obtained results across all three library types.

Transcript assembly and gene expression quantification

We aligned the high-quality reads of each library to the potato reference genome (dAg) using HISAT2 v2.2.1 (Kim et al. 2019) based on the default parameters. The log files were processed to obtain read mapping statistics. We performed transcript assembly and quantification of gene expression using StringTie v2.2.1 (Pertea et al. 2015) as follows: (1) The mapped reads in bam format were assembled into transcripts using StringTie for each sample with the following parameters: at least five reads supporting exon-junction boundary (–j 5), average read depth for a transcript of at least 10 (–c 10), and the inferred library strandedness

was considered. (2) Merging of assembled transcripts into tissue-wise separately for each of 15 tissues using stringtie-merge with the following parameters: minimum transcript length of 200 bp ($-m$ 200) and minimum isoform fraction of 0.5 ($-f$ 0.5). (3) Finally, transcriptome assemblies from each of the 15 tissues were merged into a single non-redundant transcriptome assembly using stringtie-merge with the earlier parameters. (4) Normalized expression was estimated in TPM using stringtie with the $-e$ option for each sample. In addition, raw read counts for each gene were calculated using the prepDE.py3 script (Pertea et al. 2015). Finally, Gffcompare v0.12.6 (Pertea and Pertea 2020) was used to compare the above-generated non-redundant transcriptome assembly with the reference transcripts (dAg).

Sample clustering

We assessed the sample clustering patterns by submitting genes with $\text{mean log}_2(\text{read count} + 1) \geq 1$ to hierarchical clustering based on Pearson's correlation matrices using R. We inspected the resulting tree for mislabeled samples.

Identification of novel genes and splicing isoforms

We relied on the Gffcompare v0.12.6 (Pertea and Pertea 2020) output files to identify novel genes and isoforms. Transcripts not overlapping with known reference transcripts were assigned to class U. The nucleotide sequences of the class U transcripts were extracted and translated using TRANSDCODER v5.7.0 (Haas et al. 2013). We predicted protein domains using HMMER v3.3.2 (Finn et al. 2015) with default parameters and the PFAM database v35 (Finn et al. 2016). We performed functional annotation using AHRD for class U transcripts (<https://github.com/groupschoof/AHRD>). We classified class J transcripts as putative novel isoforms.

Identification of housekeeping and tissue-specific genes

We used the data of 15 tissues to identify housekeeping (HK) genes in potato and assessed the variability in gene expression of HK genes using the approach of Hoang et al. 2017. The approach involves the following criteria: each gene is classified as expressed if $\text{TPM} \geq 1$ in at least one sample or otherwise not expressed. We calculated the mean TPM of genes expressed in all samples by taking the average gene expression across all samples, followed by computing the Coefficient of Variation (CoV). We calculated the ratio of the maximum to minimum (MFC) by dividing the largest by the smallest TPM values, followed by computing a product score (MFC-CoV) based on the product of CoV and

MFC for each gene. Finally, we classified genes with MFC-CoV scores within the first quartile as HK genes.

We used the log_2 transformed TPM values to identify tissue-specific (TS) genes. All 15 tissues were compared against each other to find significantly overexpressed genes using LIMMA v3.58.1 (Ritchie et al. 2015). We considered genes with $\text{log}_2(\text{fold-change}) \geq 2$ with adjusted $p < 0.05$ as significantly overexpressed. If gene G was overexpressed in tissue T compared with all other tissues, then gene G was considered specifically expressed in tissue T. Further, we assessed the tissue-specific expression of HK and TS genes using the Tau index as previously described. The Tau values scale from 0 to 1, where low and high values indicate widely expressed and more tissue-specific genes, respectively (Kryuch-kova-Mostacci and Robinson-Rechavi 2017).

Identification of genes encoding transcription factors and nucleotide-binding and leucine-rich repeats

We identified transcription factors (TFs) by feeding the protein sequences of the longest isoform of 39,088 genes to iTAK v1.7a (Zheng et al. 2016) in dAg. Using the NLR-Annotator v2, we identified the nucleotide-binding and leucine-rich repeat (NLR) encoding genes (Steuernagel et al. 2020).

Potato orthology map

We used the protein sequences of the longest isoform of the eight potato clones, for which chromosome-scale genome assemblies are available (Table S1), and fed them to OrthoFinder (Emms and Kelly 2019) to compute orthogroups across eight potato clones.

Network reconstruction, module detection and gene ontology enrichment

We constructed a Pearson correlation coefficient (PCC) based co-expression network for all genes expressed in at least one transcriptome library with a TPM of 1 using the pcc.py script of LSTrAP v1.3 (Goh and Mutwil 2021). We converted the PCC-based co-expression network into a Highest Reciprocal Rank (HRR) co-expression network using parameters of a maximum HRR of 50 and a PCC cut-off of 0.5 with a second-level neighborhood. We clustered the HRR co-expression network to detect co-expressed modules using the heuristic cluster chiseling algorithm (Mutwil et al. 2010) with default parameters. We performed gene ontology enrichment for each of the detected co-expression modules. We used the CoNekT v1.1.1 framework (Proost and Mutwil 2018) for network reconstruction, module detection and gene ontology enrichments. Finally, we developed

a web server to easily access the expression atlas and the co-expression network by adopting the CoNekT framework due to its rich features (Proost and Mutwil 2018).

Identification of homologs in the reference genome

Using BLAST (Altschul et al. 1990), we identified homologs in the reference genome (dAg) for a selected set of potato genes. We considered the first best hit as the homolog for respective genes. We used the CDS sequences of 14 Rpi genes (Armstrong et al. 2019) and the protein sequence of PhAN2 (UniProt ID: A4GRV2) (Laimbeer et al. 2020) to identify their homologs in the reference genome. We used the protein sequences of the tuber identity gene (IT1; ID: Soltu.DM.06G025210) and the SELF-PRUNING 6A (SP6A; ID: Soltu.DM.05G026370) genes to identify respective homologs in the reference genome (Tang et al. 2022). We used the CDS sequences of eight S-RNases involved in self-incompatibility, as mentioned in Dzidzienyo et al. (2016), to identify their homologs in the reference genome.

Results

Data collection, processing, mapping of reads and expression quantification

We performed extensive literature mining to gather as many potato RNA-Seq datasets as possible. We downloaded 3227 raw read sequencing files (.sra) from the National Center for Biotechnology Information (NCBI) Sequence Reads Archive (SRA) database and converted them into FASTQ format. We combined reads obtained from the same library in a single FASTQ file for single-end (SE) data or two files for paired-end (PE) data, resulting in 2,636 libraries (85.24% are PE and 14.75% are SE data) from 155 NCBI BioProjects comprising 20 broad tissue categories (Table S2).

We excluded reads containing adapter sequences or reads with an average quality of less than 20. We excluded 32 samples that contained less than 100,000 reads or for which less than 50% of reads remained after trimming. The reads from each sample were mapped onto the reference genome, followed by assembling transcripts and then performing quantification of transcript abundance. We used 2,604 samples containing an average of 23,060,781 read pairs per sample with PE data and 36,832,773 reads per sample with SE data for read mapping. Mapped and uniquely mapped reads corresponded to an average of 80.73% and 67.80%, respectively. We excluded 157 samples in which $\geq 50\%$ of reads failed to map or $\geq 40\%$ could not map uniquely. Finally, we excluded 106 samples which were made of combinations of multiple tissues, such as callus, plantlet, seedling, whole

plant and mixed tissues. In total, we kept 2341 samples from 147 NCBI BioProjects for downstream analyses (Table S3).

Leaf was the most abundant tissue representing 45.4% of the samples, while petiole tissue represented 0.21% (Table S4). We have also found that about 58% (1361 of 2341) of the libraries were unstranded. Finally, we assembled transcripts and estimated transcript abundances in raw read counts and transcripts per million (TPM) at the gene level (Figure S1).

Systematic analysis of thousands of potato RNA-Seq samples

In transcriptomics studies, the clustering of samples is instrumental in identifying broad transcriptional similarities between samples and identifying potential technical artefacts and mislabeled samples. Here we employed hierarchical clustering to identify mislabeled samples. The clustering analysis revealed two major clades comprising samples from aerial and underground tissues. However, interestingly, we found an additional cluster consisting of samples from pollen only (Fig. 1). In addition, we observed that seven samples from underground tissues were clustered with aerial tissues, while 35 aerial tissues clustered with underground tissues. In order to avoid the influence of these potentially mislabeled samples, we excluded these 42 samples from the downstream analyses.

In this study, we classified a gene as expressed if the gene had a minimum TPM threshold of 1 in at least one sample and found that across all samples about 87% of known potato reference genes (33,981 of 38,977) were expressed. An average of 18,589 genes were expressed per sample. The tissue with the highest number of expressed genes was leaf (31,427 genes), whereas pollen had the lowest number of expressed genes (12,801 genes) (Table S5). We found that 12,600 genes were expressed in at least 90% of samples, including 1121 genes in all 2299 samples. About 83% of all genes not expressed in any sample had coding sequences comprising < 300 codons (Figure S2).

Housekeeping and tissue-specific genes

Due to the availability of an extensive collection of RNA-Seq samples covering a wide range of tissues and environmental conditions, we also pursued identifying housekeeping (HK) genes for potatoes. In this study, we identified 281 HK genes (Table S6) using a previously described method (Hoang et al. 2017). We evaluated the expression levels of HK genes in all tissues and found that the genes had very low expression variation (Fig. 2A). Furthermore, we used the tissue-specific index Tau to estimate tissue-specificity and confirm whether the identified HK genes broadly

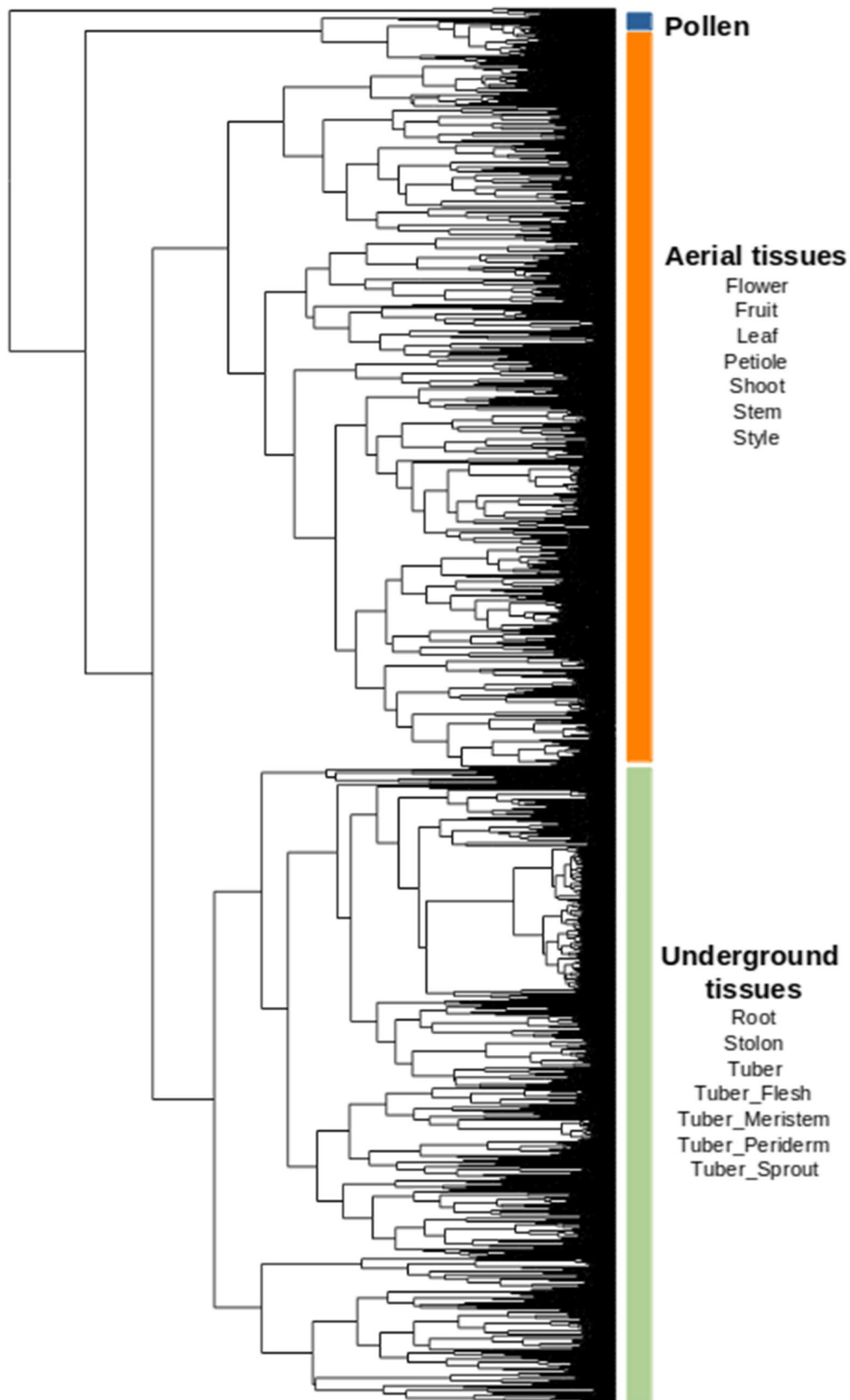


Fig. 1 Hierarchical clustering of samples using their transcriptional profiles. Raw read counts were used to compute Pearson's correlations, followed by hierarchical clustering in R. Samples were grouped into three major clades, i.e., aerial, underground tissues, and pollen clusters. A total of 42 samples were mislabeled. For e.g., 18 samples of underground tissues were grouped with samples of aerial tissues

expressed across all tissues. The Tau scores of the HK genes ranged from 0.058 to 0.282 (Fig. 2B).

We compared the global expression patterns between tissues to identify tissue-specific genes (Figure S3). All 15 tissues were compared pairwise, resulting in 308 genes with a significantly higher expression in a single tissue compared with all the others (Fig. 2C and Table S7). Interestingly, more than 90% (278 of 308) of these genes had Tau indexes > 0.8 and a median Tau of 0.97005 (Fig. 2B). Given their solid preferential expression in particular tissues, we called these genes tissue-specific (Tau > 0.8). The tissue-specific genes ranged from 11 in roots to 137 in pollen. Interestingly, 18 tissue-specific genes belonged to ten transcription factor (TF) gene families (Table S8). The number of tissue-specific TF genes ranged from one in fruit, root and style to nine in pollen.

Identification of novel transcripts

We compared the genomic coordinates of the transcripts assembled in our study with the reference transcripts (dAg) using Gffcompare (Pertea and Pertea 2020) and categorized them into 15 classes (Table S9). We found that 99.22% (58,274 of 58,734) of the transcripts precisely matched the exon–intron splice junctions of known transcripts (class “=”). We also investigated class-J and class-U categories, which account for 17,312 and 30,832 transcripts, respectively. Class-J comprises multi-exon transcripts with at least one known exon junction, while class-U encompasses transcripts located in intergenic regions. While class-J transcripts include new isoforms of known genes, those from class-U identify potentially new genes. We found that approximately 84% (14,476 of 17,312) of the class-J transcripts and about 11% (3489 of 30,832) of the class-U transcripts contain a complete open reading frame (ORFs) (Table S9). We found that 14,476 class-J transcripts belong to about 30% of reference genes (11,736 out of 39,088). In addition, we found 608 transcription factors belonging to 59 TF families (Table S10) and 94 NLR genes (Table S11) in class-J transcripts. The gene ontology enrichment analysis revealed that the class-J transcripts were enriched with several biological processes (FDR < 0.05). The top five enriched biological processes were “response to abscisic acid” (GO:0009737), “salt stress” (GO:0009651), “water deprivation” (GO:0009414), “cold response” (GO:0009409), and “positive regulation of transcription, DNA-templated” (GO:0045893) (Table S12).

On the other hand, we found 1150 non-transposon genes within 3489 class-U transcripts. Interestingly, we found 108 transcription factors (TF) belonging to 26 families (Table S13) and five NLR genes (Table S14) in the class-U transcripts. However, we did not find significantly enriched gene ontology terms in these transcripts.

Co-expression network construction and detection of co-expression clusters

To determine if our co-expression network has a scale-free architecture (Barabási and Bonabeau 2003), we calculated the Pearson correlation coefficient (PCC) for each pair of genes with a threshold of 0.5 and determined the number of times a particular gene is co-expressed with other genes at this threshold (node degree). We plotted the resulting power law distribution, which showed a negative correlation between node frequency (the number of genes with a certain number of connections) and node degree (the number of connections per gene). This distribution confirms the scale-free topology of our network (Figure S4). We constructed an HRR-based co-expression network using the above computed gene–gene PCC values using the CoNekT framework (Proost and Mutwil 2018). The constructed network contained 28,388 nodes representing genes and 4,57,580 edges representing associations between two nodes, such as HRR and PCC (Table S15). Using the heuristic cluster chiseling algorithm (Mutwil et al. 2010), we identified 853 clusters of co-expressed genes with the size of modules ranging from 2 to 285. We found that about 51% of co-expression clusters contained just two genes, while about 32% contained more than ten genes (Table S15). We visually assessed the quality of the identified clusters by inspecting the deviation of expression patterns of individual genes against the average expression pattern of the respective cluster. In this study, we considered genes with a Z score smaller than ± 1 as a tight co-expression in respective clusters. Based on these criteria, we found that an average of 85.39% of the genes across all clusters showed a tight co-expression (Figures S5 & S6). To understand the relationships between the identified clusters and tissues, we plotted a heatmap for the Z scores of the average expression level (TPM) per module at each tissue (Fig. 3), and we found about 54% (461) clusters that showed distinct expression patterns across tissue, i.e., Z score larger than ± 1 in at least three tissues. To understand the function of these clusters, we conducted an enrichment analysis, which revealed that more than 65% of the clusters contained at least one significantly enriched (corrected p-value < 0.05) biological process (Table S15). The identified clusters, thus, effectively grouped the genes that may participate in the same biological pathways and constitute the basis for identifying gene co-expression clusters underlying various agronomic traits.

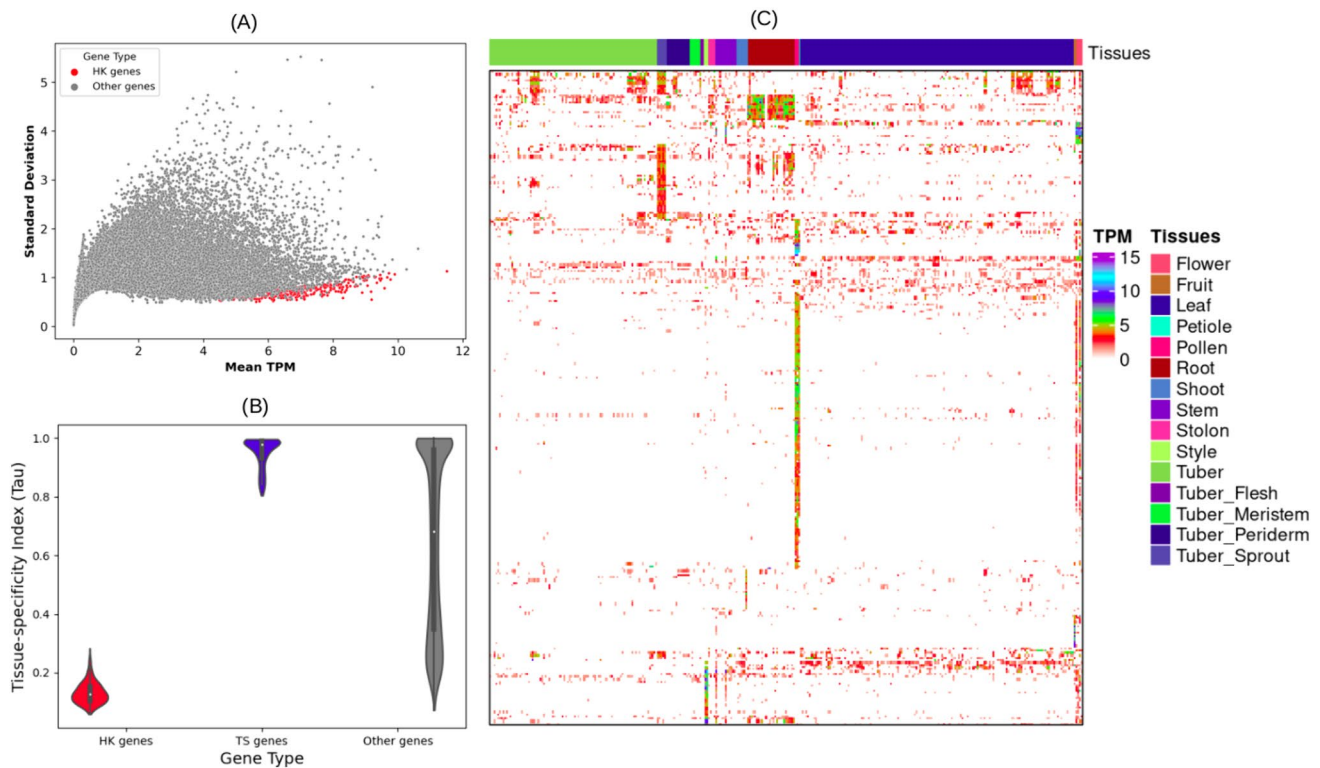


Fig. 2 Housekeeping (HK) and tissue-specific (TS) genes. **A** Scatter plot of mean versus standard deviation of transcripts per million (TPM) showing uniform and stable expression of 281 HK genes. Grey dots represent all the non-HK-expressed genes (TPM ≥ 1 in at least one sample). **B** Violin plot of the distribution of Tau indexes for housekeeping (HK), tissue-specific (TS), and the other genes. Low

Tau values indicate a stable and constitutive expression, while higher values support tissue-specificity (Tau > 0.8). **C** Global transcriptional patterns of tissue-specific (TS) genes. Expression values are represented as log2 transformed transcripts per million (TPM) values in all tissues (2299 samples)

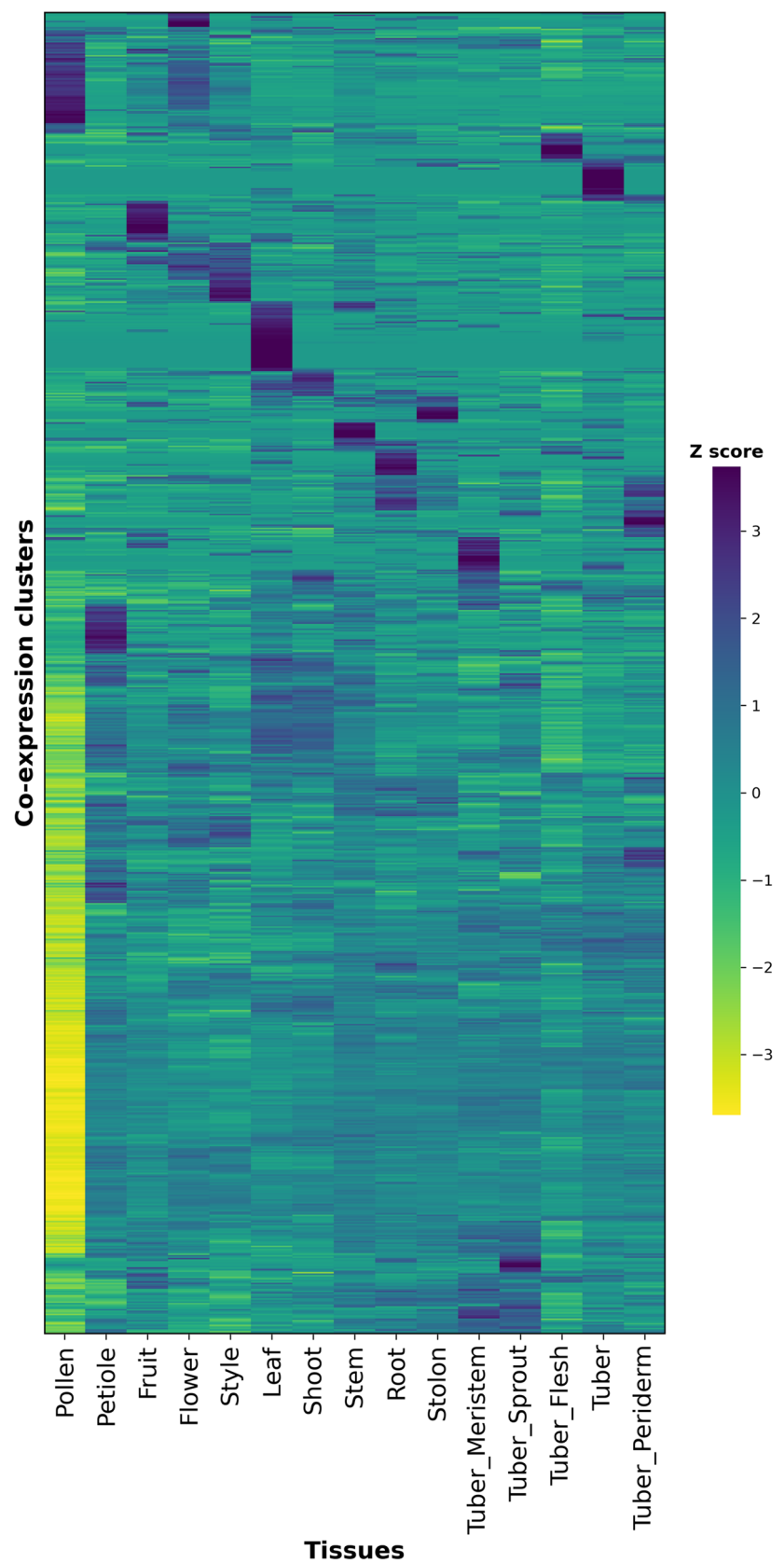
Co-expression clusters related to anthocyanin biosynthesis

We searched for co-expression modules associated with anthocyanin biosynthesis using the gene ontology (GO) term “anthocyanin-containing compound biosynthetic process (GO:0009718)” in our co-expression network. We found a single co-expression cluster, Cluster_90 (corrected p -value $< 8.7e-05$), containing genes including the ones that encode the structural enzymes involved in the anthocyanin biosynthesis (Table S16; Table S17), except the primary regulator gene, R2R3 MYB TF, the homolog of PhAN2 (R2R3 MYB TF), present on chromosome 10 (Jung et al. 2009). Nonetheless, in this cluster, we found three new MYB TFs (SOLTUB.AGRIA.G00000008919, SOLTUB.AGRIA.G00000017419, and SOLTUB.AGRIA.G00000019730) mapping to chromosomes 2 and 5. In addition, we also found 19 TFs belonging to bHLH, MADS, B3, C2H2 and AP2 TF families and two WD40 repeat-containing proteins in Cluster_90 (Fig. 4A; Table S18). Moreover, various GO terms such as “flavonoid biosynthetic process” (GO:0009813), “organic substance biosynthetic process” (GO:1,901,576),

“pigment metabolic process” (GO:0042440), “DNA-binding transcription factor activity” (GO:0000981), and “anthocyanin-containing compound metabolic process” (GO:0046283) were significantly enriched (corrected p -value < 0.05) in this cluster (Table S19).

To find the primary regulatory gene of anthocyanin biosynthesis R2R3 MYB TF in our reference genome, we searched the reference protein sequences using the protein sequence of PhAN2 (UniProt ID: A4GRV2) as a query using BLASTP (Altschul et al. 1990). This search resulted in the identification of four homologs of PhAN2 in potatoes, mapping to chromosome 10. Three of them are present in three different clusters, namely Cluster_133, Cluster_78, and Cluster_85. In contrast, the fourth homolog does not have a cluster assignment. Among the three homologs, only one homolog (SOLTUB.AGRIA.G00000035098) is present in a cluster, Cluster_78, in which the GO terms “phenylpropanoid metabolic” (GO:0009698) and the “proanthocyanidin biosynthetic” (GO:0010023) processes were significantly enriched (corrected p -value < 0.05) (Table S20). Hence, we consider this gene as the homolog of PhAN2, which regulates the early biosynthetic genes in our reference genome

Fig. 3 The clusters-tissues relationship heatmap. The heatmap shows that the tissues are on the x-axis, and the co-expression clusters are on the y-axis. The heatmap was colour-coded with a Z score of averaged expression level (TPM) per module at each tissue. Python was used to cluster the rows and columns hierarchically



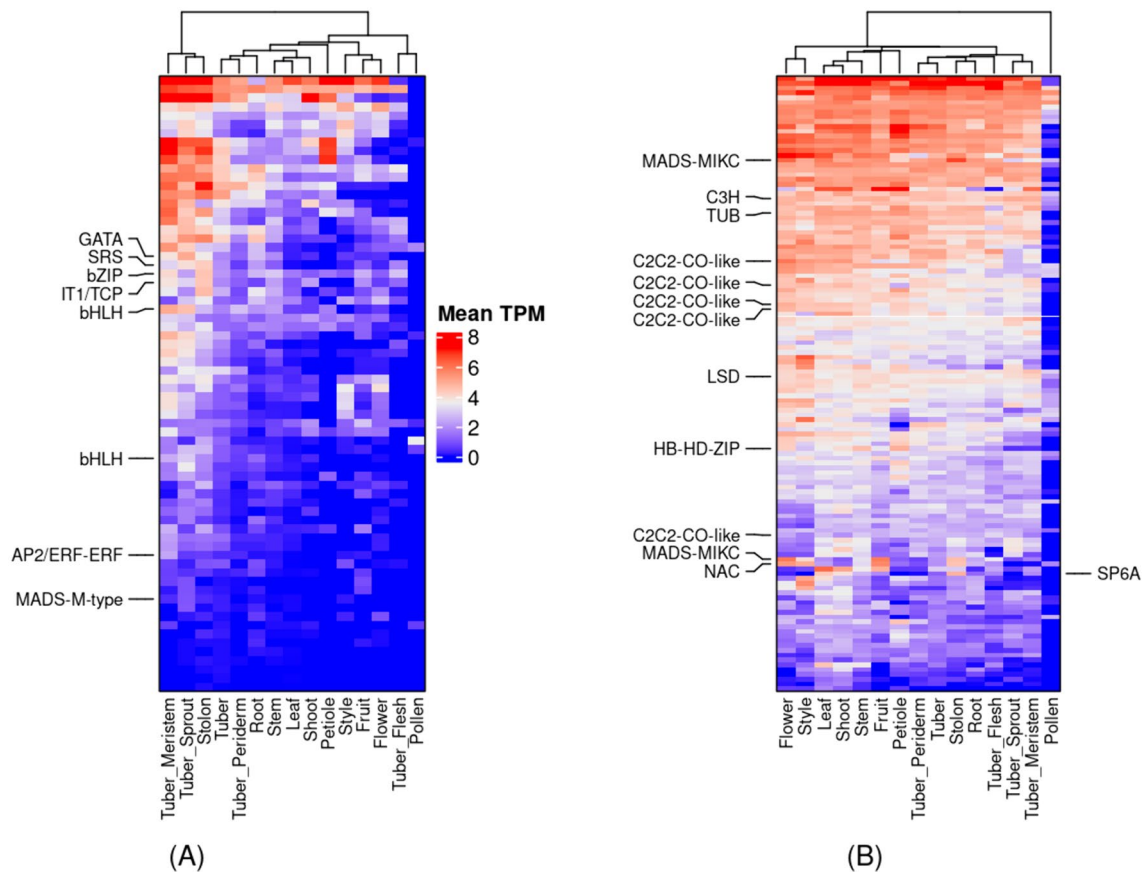


Fig. 5 The heatmaps of co-expression clusters harboring tuberization genes, the tuber identity gene (IT1) and SELF-PRUNING 6A (SP6A). **A** Cluster_23. **B** Cluster_97. The heatmaps show that the tissues are on the x-axis, and the genes are on the y-axis. The heatmaps were color-coded with log₂ transformed mean TPM values for each tissue. We clustered the rows and columns hierarchically in Python.

We marked the transcription factors, including IT1, on the left side of the heatmap and the SP6A on the right side. The interactive clusters for Cluster_23 and Cluster_97 can be accessed at <https://stcoexpnet.julius-kuehn.de/cluster/graph/473> and <https://stcoexpnet.julius-kuehn.de/cluster/graph/689>, respectively

Co-expression clusters related to defense responses

In this study, we identified 578 genes which belong to different classes of nucleotide-binding (NB) domain and leucine-rich repeat (LRR) (NLR) genes (Table S27) using NLR-Annotator (Steuernagel et al. 2020). A total of 432 out of 578 NLR genes were assigned to 119 co-expression clusters which contain 1–44 NLR genes per cluster. Among the 119 co-expression clusters, 43 were enriched for at least one biological process involved in defense mechanisms (Table S28), such as “response to biotic stimulus (GO:0009607)”, “defense response (GO:0006952)”, “response to fungus (GO:0009620)”, “defense response to fungus (GO:0050832)”, “response to bacterium (GO:0009617)”, “defense response to bacterium (GO:0042742)”, “response to virus (GO:0009615)”, and “defense response to virus (GO:0051607)”.

We found eight of 14 known NLRs effective against *Phytophthora infestans* (Rpi genes) (Armstrong et al. 2019) in

three co-expression clusters, Cluster_223, Cluster_210, and Cluster_103, while the remaining Rpi genes were either not assigned to any cluster or clusters did not enrich for any of the above-mentioned biological processes (Table S29). Cluster_223 contains 58 genes, of which 32 encode NLRs (Table S30; Fig. 6A). In this cluster, we found four Rpi genes, Rpi-R3b, Rpi-R9a, Rpi-vnt1.1, and Rpi-vnt1.1 A2056, mapping to two homologs in the reference genome, SOLTUB.AGR1A.G00000038927 and SOLTUB.AGR1A.G00000032822. In this cluster, the biological process “defense responses” to fungi, bacteria and viruses were enriched (Table S31). Cluster_103 contains 92 genes, of which 44 encode NLRs (Table S32; Fig. 6B). In this cluster, we found two Rpi genes, Rpi-R8 and Rpi-ber, mapping to two homologs in the reference genome, SOLTUB.AGR1A.G00000032965 and SOLTUB.AGR1A.G00000035214. In this cluster, the biological process “defense responses” to fungi and bacteria were enriched (Table S33). Cluster_210 contains 92 genes, of which 11 encode NLRs (Table S34;

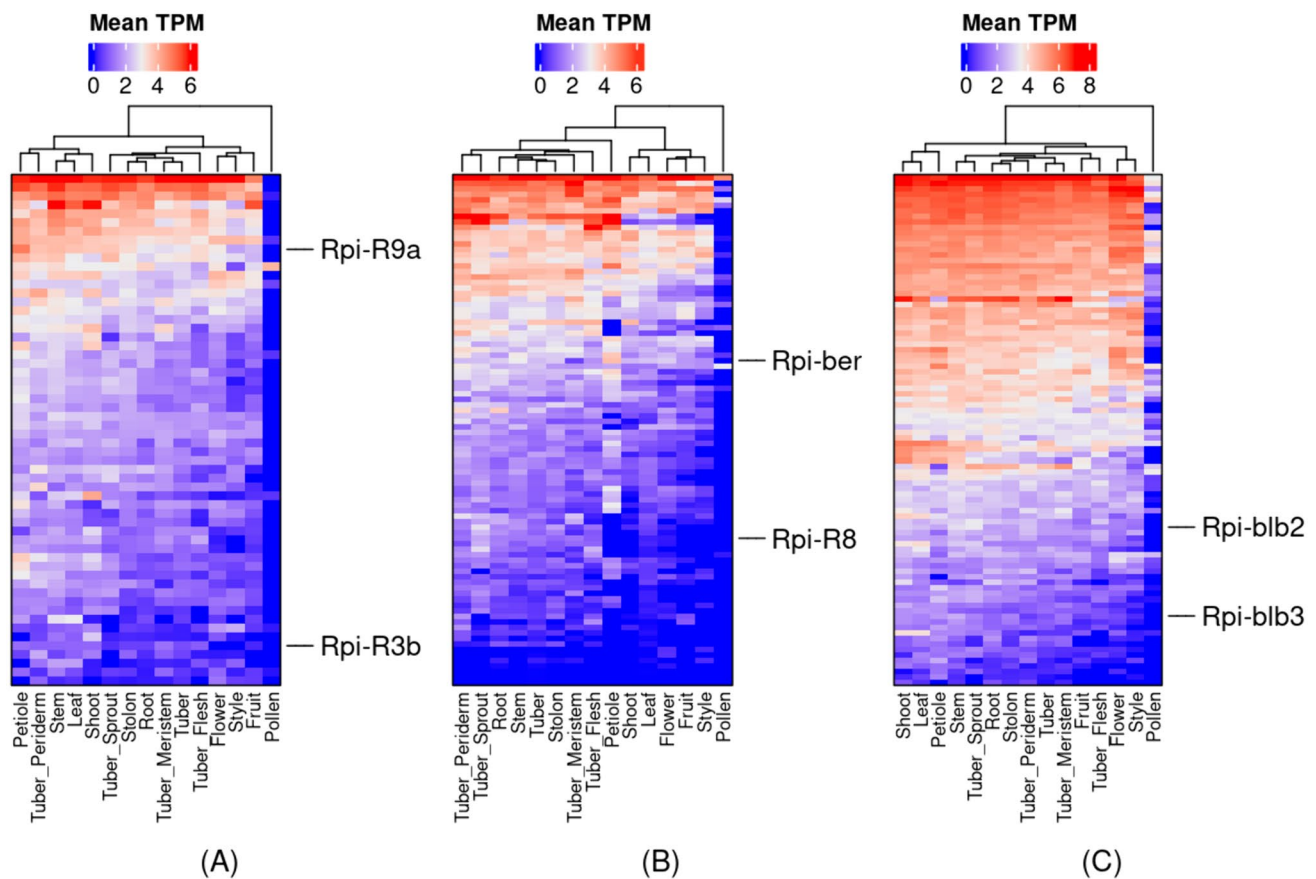


Fig. 6 The heatmaps of co-expression clusters harboring genes involved in resistance against late blight disease. **A** Cluster_223; **B** Cluster_103; **C** Cluster_210. The heatmaps show that the tissues are on the x-axis, and the genes are on the y-axis. The heatmaps were color-coded with log₂ transformed mean TPM values for each tissue. We clustered the rows and columns hierarchically in Python. We marked the identified NLRs effective against *Phytophthora*

infestans (Rpi genes) on the right side of the heatmap. Rpi-vnt1.1 and Rpi-vnt1.1-A2056 are homologs to Rpi-R9a. The interactive clusters for Cluster_223, Cluster_103, and Cluster_210 can be accessed at <https://stcoexpnet.julius-kuehn.de/cluster/graph/327>, <https://stcoexpnet.julius-kuehn.de/cluster/graph/479>, and <https://stcoexpnet.julius-kuehn.de/cluster/graph/322>, respectively

Fig. 6C). Similarly, this cluster contained two Rpi genes, Rpi-blb2 and Rpi-blb3, mapping to two homologs in the reference genome, SOLTUB.AGR1A.G00000044086 and SOLTUB.AGR1A.G00000013669. In this cluster, defense responses related to a specific pathogen were not enriched, but “response to biotic stimulus” and “defense response” were enriched (Table S35).

Co-expression cluster related to self-incompatibility

We searched in our co-expression network for clusters harboring genes involved in self-incompatibility (Dzidzienyo et al. 2016). We found one co-expression cluster, Cluster_30, containing the S-RNase gene SOLTUB.AGR1A.G00000001844 which showed an extreme expression in style samples (mean TPM of 5783.82). This cluster contained 99 more genes and the majority of these genes showed high mean expression in style samples (Table S36; Fig. 7). However, surprisingly, we

found no enriched GO terms in this cluster. Furthermore, this cluster contained two genes belonging to two TF gene families: GATA and bHLH.

Data availability through a web server

The data presented above are easily accessible by researchers to explore the expression atlas of 2299 transcriptome samples and the co-expression network interactively via a web server called StCoExpNet. This web server is freely available at <https://stcoexpnet.julius-kuehn.de>.

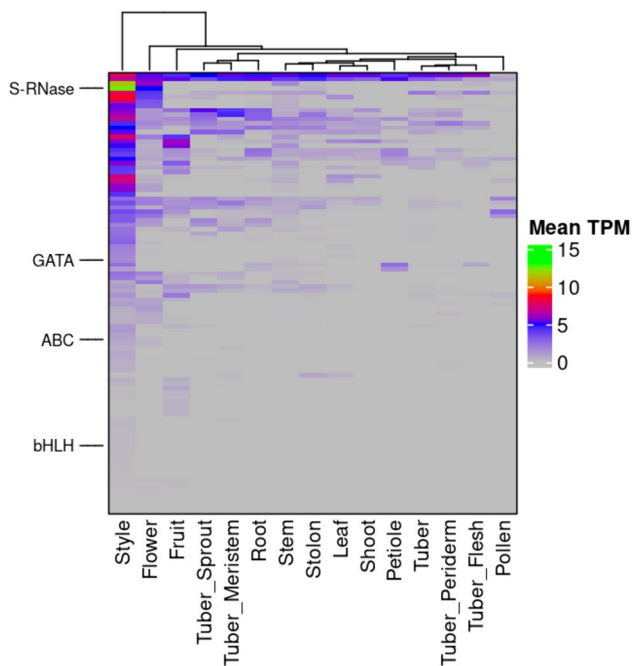


Fig. 7 The heatmap of Cluster_30 that harbors the S-RNase gene. The heatmap shows that the tissues are on the x-axis, and the genes are on the y-axis. The heatmap was color-coded with log₂ transformed mean TPM values for each tissue. We clustered the rows and columns hierarchically in Python. We marked this cluster's essential genes, S-RNase, ABC transporter (ABC), and two transcription factors (bHLH and GATA). The interactive cluster for Cluster_30 can be accessed at <https://stcoexpnet.julius-kuehn.de/cluster/graph/367>

Discussion

High-quality of publicly available potato RNA-Seq data

Publicly available RNA-Seq datasets, such as at NCBI SRA database, provide a wealth of information that can be used to investigate gene expression, alternative splicing, identify novel transcripts and identify functionally related genes in an organism. Researchers can use these datasets to test hypotheses, validate findings, and generate new insights into the mechanisms of various biological processes (Ferrari and Mutwil 2020; Wisecaver et al. 2017; Lin et al. 2017; Ramšak et al. 2018). In this study, we have performed extensive literature mining and constructed a global gene expression atlas for potatoes using thousands of publicly available RNA-Seq datasets (Figure S1). Our analyses revealed that these datasets clustered according to transcript abundance into three broad categories of tissues: pollen, aerial and underground tissues. Less than 2% of the analyzed RNA-Seq samples were excluded based on clustering analyses, indicating a high-quality level of the publicly available samples supported by the sample clustering in this study (Fig. 1). However, potato is highly heterozygous and in most cases tetraploid. When mapping reads from such samples onto

a single haploid reference genome (dAg), collapsing multiple alleles into one is expected but it will not negatively influence the result of our study, as we are aiming to make in our study conclusions regarding the expression of genes and not individual alleles.

New internal reference genes for qPCR experiments in potatoes

Housekeeping (HK) genes are those expressed constitutively across broad conditions and robustly (Czechowski et al. 2005; Bustin et al. 2009) and are used as internal reference genes in real-time quantitative polymerase chain reaction (qPCR) assays (Nicot et al. 2005; Hu et al. 2009; Tang et al. 2017). By utilizing the extensive compilation of RNA-Seq datasets presented in this study, one can assess the suitability of commonly utilized internal reference genes and put forward novel ones.

We identified 281 HK genes (Table S6) that showed a stable expression (Fig. 2A & B) across samples, supporting their suitability for use as internal reference genes in potato qPCR assays. The list of 281 HK genes includes three known reference genes, namely Elongation factor 1- α , 60 s ribosomal protein L8 (Nicot et al. 2005; Tang et al. 2017) and Ubiquitin-associated/translation elongation factor EF1B protein (Mariot et al. 2015), used as internal reference genes in qPCR potato experiments under a few different stress conditions. Because these three HK genes exhibit a consistently stable expression across thousands of samples generated under various experimental conditions, these are particularly recommended to be used as internal reference genes for qPCR assays of potatoes (Table 1). Further, we found homologs of known reference genes of other crops in our list of potato HK genes. For example, Heat shock protein 90 was validated as a reference gene in *Cajanus cajan* under heat and salt stress conditions (Sinha et al. 2015). Eukaryotic initiation factor 4A was validated as a reference gene in *Carica papaya* under different experimental conditions (Zhu et al. 2012). YT521-B-like protein family protein was validated as a reference gene in perennial ryegrass (Lee et al. 2010), and Ubiquitin-conjugating enzyme E2 was validated as a suitable reference gene for *Eucommia ulmoides* Oliv under different experimental conditions (Ye et al. 2018a, b). Therefore, given the high expression stability of these four HK genes across many samples generated under different experimental conditions (Table 1), these genes could be considered as novel reference genes for potato qPCR experiments.

The global co-expression network and co-expression clusters

Scale-free networks follow a power-law distribution, where a few genes are highly connected while most genes have

Table 1 List of known and novel internal reference genes used in potato qPCR assays identified in this study

Reference gene ID	Gene name	Mean TPM	SD	Tau index	References
SOLTUB.AGRIA.G00000040002*	60 s ribosomal protein l8 (L8)	8.926	0.996	0.0805	Tang et al. (2017)
SOLTUB.AGRIA.G00000020781*	Elongation factor 1-alpha (EF-1-alpha)	9.745	1.18	0.0838	Tang et al. (2017)
SOLTUB.AGRIA.G00000022993*	Ubiquitin-associated/translation elongation factor EF1B protein (C2)	5.303	0.552	0.1435	Mariot et al. (2015)
SOLTUB.AGRIA.G00000027916	Eukaryotic initiation factor 4A (EIF)	7.489	0.743	0.0816	Zhu et al. (2012)
SOLTUB.AGRIA.G00000026336	Heat shock protein 90 (HSP90)	10.909	1.319	0.0899	Sinha et al. (2015)
SOLTUB.AGRIA.G00000001694	ARATH YTH domain-containing protein ECT2	8.313	1.114	0.1246	Lee et al. (2010)
SOLTUB.AGRIA.G00000011879	Ubiquitin-conjugating enzyme E2	7.625	0.806	0.1065	Ye et al. (2018a, b)

Mean TPM, SD and Tau indexes indicate the mean TPM across all RNA-Seq samples, standard deviation, and tissue-specificity index, respectively

*Indicates known reference genes

only a few connections (Barabási and Bonabeau 2003). This structure was believed to be an evolved feature that ensures stability and robustness against genetic and environmental disturbances (Barabási and Oltvai 2004). We found that the potato's co-expression network follows this scale-free topology and supports the biological validity of our expression data (Figure S4). Most of the genes found in all clusters displayed a tight co-expression that suggests the genes in respective modules have similar expression patterns (Figure S5 and S6), indicating a high quality of the identified clusters. Further, we found several clusters positively related to specific tissues that may suggest that the genes within the co-expression cluster are actively involved in biological processes specific to that tissue. In contrast, many clusters are negatively related to pollens (Fig. 3).

Clusters of functionally related genes tend to have strong connections within the co-expression network. Identifying and examining these clusters can help to uncover the functional gene clusters of an organism (Mutwil et al. 2010; Rhee and Mutwil 2014; Aoki et al. 2016). In order to illustrate that the identified co-expression clusters are biologically interpretable, we discuss as proof of concept the ones related to important potato agronomic traits, including anthocyanin biosynthesis, tuberization, defense responses against multiple pathogens, and self-incompatibility.

Transcriptional complexity of anthocyanin production in potatoes

Anthocyanins are plant secondary metabolites that are responsible for the vibrant coloration of various plant tissues (Laimbeer et al. 2020). They have gained significant attention due to their numerous documented benefits for plants' physiological processes and human health (Stintzing and Carle 2004; De Pascual-Teresa and Sanchez-Ballesta 2008; Khoo et al. 2017; Schulz et al. 2016; Merzlyak and Chivkunova 2000). In solanaceous species, the early

biosynthetic genes (EBGs), which include chalcone synthase (CHS), chalcone isomerase (CHI), flavonoid 3-hydroxylase (F3H), and flavonoid 3' hydroxylase (F3'H) are regulated by the R2R3 MYB transcription factor (TF) (Jung et al. 2009). The late biosynthetic genes (LBGs), which include flavonoid 3'-5' hydroxylase (F3'5'H), dihydroflavonol 4-reductase (DFR), anthocyanidin synthase (ANS), glutathione S-transferase (GST), anthocyanin O-methyltransferase (AOMT), and glucosyl transferases (UGT), are regulated by a ternary protein complex called MBW in a spatiotemporal manner. The complex is formed from MYB, basic helix loop helix (bHLH) TFs, and WD40 repeat-containing proteins (Patra et al. 2013; Lin-Wang et al. 2010; Feller et al. 2011). Finally, the synthesized anthocyanins will be transported to the vacuole by the MATE transporter (Gomez et al. 2009). The genetic basis of the anthocyanin biosynthetic pathway has been studied in potatoes and identified several essential genes involved in this pathway (Jung et al. 2009; Zhang et al. 2009a; Zhang et al. 2009b; Śliwka et al. 2017; Laimbeer et al. 2020). However, so far, only a few studies have been conducted to investigate the transcriptional dynamics of the identified vital genes between different colored phenotypes (Laimbeer et al. 2020; Riveros-Loaiza et al. 2022). Moreover, these studies were conducted using single tissue of a small number of clones. Hence, limited information on the global transcriptional complexity of anthocyanin biosynthesis in potatoes is available. Therefore, we investigated the transcriptional complexity of anthocyanin production in potatoes using a global co-expression network in this study.

We found a single co-expression cluster (Cluster_90) that contains 24 TFs and 23 genes that encode various structural enzymes involved in the anthocyanin biosynthetic pathway (Fig. 4A; Tables S17 & S19). Hence, we associated this cluster with anthocyanin biosynthesis in potatoes. The sum of the TFs and genes belonging to Cluster_90 is more than double the number of genes identified in a recent study conducted to investigate the transcriptional dynamics

between genotypes with different colorations of flesh and skin (Riveros-Loaiza et al. 2022), illustrating that the global co-expression network approach is robust and efficient in identifying genes underlying agronomic traits. The newly identified TFs may play an essential role in anthocyanin biosynthesis in potatoes. For example, we identified three TFs belonging to the MYB TF family in this cluster, mapped to others than chromosome 10 (Jung et al. 2009). In addition, these genes showed above-average expression in multiple tissues (Figure S7). Therefore, we hypothesize that several homologs of PhAN2 (R2R3 MYB TF) may transcriptionally regulate the anthocyanin biosynthesis in different tissues spatiotemporally in potatoes. Further, we identified eight TFs belonging to the MADS-box TF family in this cluster. A SQUAMOSA-class MADS-box TF, VmTDR4, is associated with anthocyanin biosynthesis during normal ripening in bilberry (Jaakola et al. 2010). Hence, these MADS-box TFs may also play an essential role in potato anthocyanin biosynthesis. This cluster provided several genes that may help define future breeding strategies to develop new potato cultivars with high anthocyanin content.

Further, we identified the primary regulator of anthocyanin biosynthesis, R2R3 MYB TF, in a different cluster, Cluster_78. Cluster_78 contains several TFs (Fig. 4B; Tables S20 & S21) and many genes involved in the phenylpropanoid metabolic process, providing precursors for anthocyanin biosynthesis (Laimbeer et al. 2020). Thus, this cluster can be associated with the phenylpropanoid metabolic process and anthocyanin biosynthesis and illustrates the mechanistic interlink between both pathways which was not previously reported in potatoes.

Transcriptional complexity of tuberization in potatoes

In general, late-maturing cultivars (LMC) produce higher yields than early-maturing cultivars (EMC). However, abiotic stresses, such as heat waves and drought, negatively affect the tuber quality and yields of LMCs. In contrast, EMCs escape these stress conditions. The early induction of tuberization dictates the time to crop maturity and is an essential agronomic trait that lies in its ability to influence the overall yield over an extended period. On the molecular, it is known that leaves act as sensors for day length and generate a mobile signal known as tuberigen, which is then transported to the underground stems to trigger the process of tuberization (Zierer et al. 2021). The FLOWERING LOCUS T (FT) protein (StSP6A) controls potato tuberization (Navarro et al. 2011). In addition, a TCP TF, called *Identity of Tuber 1* (IT1), interacts with StSP6A and forms a protein complex which regulates the tuber initiation (Tang et al. 2022). Nevertheless, insights into the transcriptional complexity behind tuberization still need to be

discovered that may identify unknown genes playing an essential role in tuber development. Therefore, we investigated the transcriptional complexity of potato tuberization using a global co-expression network in this study.

In this study, we found two co-expression clusters based on the presence of the two essential genes involved in the regulation of potato tuberization, IT1 (Cluster_23; Fig. 5A) and StSP6A (Cluster_97; Fig. 5B). In addition, these two clusters enriched significantly for biological processes involved in the photoperiodic control of tuberization, day-length dependent tuberization, response to light stimulus, elongation of stolons, and transporting biomolecules (Table S24 & Table S26). Therefore, we associated these two clusters with the regulation of tuberization in potatoes. We found multiple TFs, such as bZIP, CO, and TCP, that are known to play an essential role in regulating tuberization by forming the tuberigen activation complex (TAC) (Teo et al. 2017) and other complexes similar to TAC (Tang et al. 2022). In addition, we found new TFs in these two clusters belonging to multiple TF families, such as C3H, TUB, LSD, NAC, SRS, bHLH, GATA, MADS-box, HB-HD-ZIP and AP2, and these TF may be directly or indirectly involved in the regulation of tuberization (Table S23; Fig. 5A & B). For example, researchers have discovered a MADS-box TF (IbSRD1) in sweet potatoes that responds to auxin and promotes the proliferation of metaxylem and cambium cells. The overexpression of IbSRD1 led to earlier thickening of storage roots, indicating that the gene is involved in regulating the initial growth of storage roots in an auxin-dependent manner (Noh et al. 2010). Therefore, the newly identified TFs provide us with new targets in breeding programs to improve the earliness of varieties and, thus, escape adverse abiotic stress conditions.

Transcriptional complexity of defense responses against multiple pathogens

Plants possess cell surface and intracellular receptors, which can detect molecules produced by pathogens and trigger defense responses. The nucleotide-binding (NB) domain and a leucine-rich repeat (NLR) genes are important but not the only defense responsive genes. All these genes accomplish the defense responses by detecting the molecules secreted by pathogens and activating a suite of dense response processes against the pathogens (Feehan et al. 2020). In this study, we identified 578 NLR genes (Table S27), which is significantly lower than the number of predicted NLR genes for most potato accessions (Tang et al. 2022) but slightly higher than for the wild relatives of sweet potato species, *Ipomoea trifida* (547 NLR genes) and *Ipomoea triloba* (569 NLR genes) (Wu et al. 2018). We found 226 NLR genes present in 43 co-expression clusters, which enriched for various biological processes involved in defense responses, of which

several co-expression clusters enriched for defense response processes against multiple pathogens (Fungi/Bacterium/Virus) (Table S28).

Phytophthora infestans is the major pathogen in potato and causes late blight disease. Several functional NLR genes effective against *Phytophthora infestans* (Rpi genes) have been successfully cloned (Armstrong et al. 2019; Paluchowska et al. 2022). Several transcriptomic studies have been conducted to identify differentially expressed genes between contrasting potato cultivars for late blight disease (Duan et al. 2020; Cao et al. 2020; Yang et al. 2018). However, the transcriptional regulation of these Rpi genes remains unknown. In this study, we identified eight of 14 NLRs that have been reported as effective against *Phytophthora infestans* (Rpi genes) in three co-expression clusters (Cluster_223, Cluster_103, and Cluster_210) along with 79 other NLR genes and five TFs belonging to EIL, C3H, C2H2, NAC and MYB TF families (Tables S30, S32 and S34; Fig. 6A, B and C).

The identified TFs may regulate the Rpi genes directly or indirectly to confer resistance against the pathogens. For example, an MYB TF increases resistance against the pathogen, *Botryosphaeria dothidea* in apples by regulating circular wax biosynthesis (Zhang et al. 2019). In addition, numerous studies have investigated the role of NAC transcription factors in plant immunity and identified dozens of NAC genes that function as positive or negative regulators of plant immunity, as well as modulators of hypersensitive response and stomatal immunity, or targets of pathogen effectors (Yuan et al. 2019). Furthermore, a novel protein elicitor (SsCut) from *Sclerotinia sclerotiorum* induces multiple defense responses in plants, *Arabidopsis*, soybean, rice, maize and wheat by causing hypersensitive response (HR). In addition, SsCut increases plant resistance to multiple pathogens, *S. sclerotiorum*, *Phytophthora nicotianae* and *Phytophthora sojae*. A Virus-induced gene silencing revealed that C2H2 TF acts as a regulator of SsCut-triggered immunity in *Nicotiana benthamiana* (Zhang et al. 2014, 2016b).

The newly identified TFs and the above-described NLR genes could be targeted in the breeding program to develop new potato cultivars with resistance to multiple pathogens, especially late blight disease.

Novel candidate genes to overcome self-incompatibility

Transforming the clonal crop potato into a diploid inbred/F1 hybrid variety presents an opportunity to employ efficient breeding techniques (Lindhout et al. 2011). Inbred potatoes could expedite the development of novel varieties with desired combinations of alleles for increased yield, tuber quality, and resistance traits (Jansky et al. 2016). However,

a major obstacle to this strategy is the prevalence of gametophytic self-incompatibility (SI) in most diploid potato germplasm, hindering the creation of diploid homozygous lines. SI is a reproductive isolation mechanism observed in plant species of about 60 plant families, including Solanaceae. In the Solanaceae, the style distinguishes between self and non-self-pollen to inhibit self-fertilization and promote outcrossing (Dzidzienyo et al. 2016). A single polymorphic locus, called the S-locus, governs the SI in potato (Fujii et al. 2016). This locus encompasses two distinct determinants: the female/pistil S-determinant, which is a cytotoxic S-ribonuclease known as S-RNase, and the male/pollen S-determinant, which consists of a group of pollen-specific S-locus F-box proteins called SLFs (McClure et al. 1989; Ushijima et al. 2003). The S-RNase functions by impeding the growth of self-pollen tubes through either ribosomal RNA (rRNA) degradation or disruption of the cytoskeleton's dynamic equilibrium (McClure et al. 1990; Roldán et al. 2012). During cross-pollination, based on the collaborative non-self-recognition system, the pollen-expressed SLFs recognize S-RNases and target them to the proteasomal degradation pathway, allowing pollen tube growth towards the ovaries where fertilization can take place (Kubo et al. 2010). On the other hand, the S-RNase is not degraded during self-pollination as it was not recognized by the self-SLFs that induce SI (Kubo et al. 2015).

Currently, two approaches are available to overcome the SI in potatoes and have been reported to confer self-compatibility (SC). (i) manipulation of S-RNase (Ye et al. 2018b; Enciso-Rodriguez et al. 2019). Although this method converted SI genotypes to SC, the SC mutant lines produced varying numbers of seeds per fruit (67–288) across mutant lines, raising concerns about the method's robustness. (ii) introgression of the S-locus inhibitor (Sli) gene from wild potatoes into commercial varieties through conventional breeding (Hosaka and Hanneman 1998; Birhman and Hosaka 2000). This method is both time-consuming and demanding in terms of labor. In addition, this method relies on introgressing an allele from a wild species characterized by extended stolons and elevated levels of toxic steroidal glycoalkaloids in tubers in cultivated potato (Leisner et al. 2018). Moreover, these approaches are S-RNase-centric and aim to inhibit the functions of S-RNase to solve SI in potatoes. Hence, we advocate redirecting attention away from S-RNase and towards other candidate genes implicated in potatoes' self-incompatibility (SI) mechanism. Consequently, it is imperative to devise effective methodologies centered on these alternative candidate genes and utilize genotypes possessing the desired traits.

In this study, we found a co-expression cluster (Cluster_30) comprising the S-RNase gene, and hence, we associated this cluster with SI (Table S36). The majority of the genes in this cluster showed a high average expression in

style tissue samples, suggesting the role of these genes in SI or biological processes related to SI (Fig. 7). Further, we analyzed this S-RNase gene's immediate neighborhood in the co-expression network (Fig. 7). We found a member of the ABC transporter family (SOLTUB.AGRIA.G00000013767), to be co-expressed with the S-RNase. This ABC transporter may be potentially involved in transporting the S-RNase from pistil to pollen to accomplish SI in potatoes similar to what was reported for apples (Meng et al. 2014). Therefore, we hypothesize that disrupting the ABC transporter gene's function by introducing mutations (Ye et al. 2018b; Enciso-Rodriguez et al. 2019) may block S-RNase transport to pollen from the pistil, leading to inducing SC.

Data availability through a web server

A web server StCoExpNet has been created for researchers to explore the constructed potato expression atlas and gene co-expression network by adopting the CoNekT framework. This platform was chosen as it allows rich visualization features along with a detailed graphical user manual (Proost and Mutwil 2018). Our web server has the potential to serve as a reference database for potato transcriptomic studies. Through this resource, one can prioritize genes based on their expression and co-expression for mutagenesis, QTL cloning and GWAS studies. In addition, this resource can be used to investigate the gene expression and co-expression of the whole gene family of interest at the genome scale. Further, the results obtained from this resource can be mapped to different potato reference genomes through the integrated ortholog relationships among eight potato genotypes. Moreover, the expression atlas and the co-expression network can be downloaded through this web interface for local use. We are confident that this website will enhance data reuse and assist research groups in their projects.

Conclusions

We have used an extensive collection of publicly available RNA-Seq datasets to construct a global transcriptome atlas for potatoes. We implemented a pipeline with state-of-the-art methods to map reads and quantify gene expression levels in 15 tissues. This atlas allowed us to identify housekeeping (HK) and tissue-specific (TS) genes. The HK genes might be used as internal reference genes in qPCR experiments, whereas TS genes might help researchers to test hypotheses in functional genomics studies. We also constructed a global gene co-expression network (GCN) for potatoes to explore the system-wide transcriptional landscape of potato tissues. We explored the functions of co-expression clusters using the gene ontology enrichment method. Several

of the identified co-expression clusters are strongly linked with various agronomic traits. Our analyses revealed several candidate genes for various agronomic traits, and these can be used in defining future potato breeding programs. Furthermore, the present GCN sheds light on the functions of multiple potato genes and co-expression clusters. These findings are likely significant not only for understanding the roles of these genes but also for identifying genes that contribute to relevant agronomic characteristics. To enhance the reusability of the collected data, we developed a user-friendly web interface that enables the community to access and navigate through the data quickly. This resource will serve as a valuable asset not just for fundamental research endeavors but also for advancing innovative approaches aimed at boosting potato yield to meet the ever-growing global food requirements.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00299-024-03201-2>.

Acknowledgements The authors acknowledge the computational infrastructure and support provided by the Center for Information and Media Technology at Heinrich Heine University Düsseldorf and the German Network for Bioinformatics Infrastructure (de.NBI, <https://www.denbi.de/>) that contributed to the research results reported within this study.

Author contributions VSB conceived, designed, performed the experiments and data analysis, and wrote the manuscript. BS contributed to data analysis and manuscript writing. All authors contributed to the article and approved the submitted version.

Funding Open Access funding enabled and organized by Projekt DEAL. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability The datasets analyzed during the current study are available in Supplementary Information. The interactive gene expression atlas and co-expression network are available at <https://stcoexpnet.julius-kuehn.de>.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

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/>.

References

- Almeida-Silva F, Moharana KC, Machado FB, Venancio TM (2020) Exploring the complexity of soybean (*Glycine max*) transcriptional regulation using global gene co-expression networks. *Planta* 252(6):104. <https://doi.org/10.1007/s00425-020-03499-8>
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Andrews S (2010) FastQC: a quality control tool for high throughput sequence data [Online]. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Aoki Y, Okamura Y, Tadaka S, Kinoshita K, Obayashi T (2016) ATTED-II in 2016: a plant coexpression database towards lineage-specific coexpression. *Plant Cell Physiol* 57(1):e5. <https://doi.org/10.1093/pcp/pcv165>
- Armstrong MR, Vossen J, Lim TY, Hutten RCB, Xu J, Strachan SM, Harrower B, Champouret N, Gilroy EM, Hein I (2019) Tracking disease resistance deployment in potato breeding by enrichment sequencing. *Plant Biotechnol J* 17(2):540–549. <https://doi.org/10.1111/pbi.12997>
- Ballouz S, Verleyen W, Gillis J (2015) Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. *Bioinformatics* 31(13):2123–2130. <https://doi.org/10.1093/bioinformatics/btv118>
- Bao Z, Li C, Li G, Wang P et al (2022) Genome architecture and tetrasomic inheritance of autotetraploid potato. *Mol Plant* 15(7):1211–1226. <https://doi.org/10.1016/j.molp.2022.06.009>
- Barabási AL, Bonabeau E (2003) Scale-free networks. *Sci Am* 288(5):60–69. <https://doi.org/10.1038/scientificamerican0503-60>
- Barabási AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5(2):101–113. <https://doi.org/10.1038/nrg1272>
- Birhman RK, Hosaka K (2000) Production of inbred progenies of diploid potatoes using an S-locus inhibitor (Sli) gene, and their characterisation. *Genome* 43(3):495–502. <https://doi.org/10.1139/g00-012>
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bonthala VS, Stich B (2022) Genetic divergence of lineage-specific tandemly duplicated gene clusters in four diploid potato genotypes. *Front Plant Sci* 13:875202. <https://doi.org/10.3389/fpls.2022.875202>
- Bray NL, Pimentel H, Melsted P, Pachter L (2016) Erratum: near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34(8):888. <https://doi.org/10.1038/nbt0816-888d>
- Burks DJ, Sengupta S, De R, Mittler R, Azad RK (2022) The *Arabidopsis* gene co-expression network. *Plant Direct* 6(4):e396. <https://doi.org/10.1002/pld3.396>
- Bustin SA, Benes V, Garson JA et al (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* 55(4):611–622. <https://doi.org/10.1373/clinchem.2008.112797>
- Cao W, Gan L, Shang K, Wang C, Song Y, Liu H, Zhou S, Zhu C (2020) Global transcriptome analyses reveal the molecular signatures in the early response of potato (*Solanum tuberosum* L.) to *Phytophthora infestans*, *Ralstonia solanacearum*, and Potato virus Y infection. *Planta* 252(4):57. <https://doi.org/10.1007/s00425-020-03471-6>
- Chandrasekar S, Natarajan P, Mhatre PH, Mahajan M, Nivitha S, Palanisamy VE, Reddy UK, Sundararaj P (2022) RNA-Seq of cyst nematode infestation of potato (*Solanum tuberosum* L.): a comparative transcriptome analysis of resistant and susceptible cultivars. *Plants* 11(8):1008. <https://doi.org/10.3390/plants11081008>
- Chen Y, Li C, Yi J, Yang Y, Lei C, Gong M (2019) Transcriptome response to drought, rehydration and re-dehydration in potato. *Int J Mol Sci* 21(1):159. <https://doi.org/10.3390/ijms21010159>
- Chi Y, Wang T, Xu G, Yang H, Zeng X, Shen Y, Yu D, Huang F (2017) *GmAGL1*, a MADS-box gene from soybean, is involved in floral organ identity and fruit dehiscence. *Front Plant Sci* 8:175. <https://doi.org/10.3389/fpls.2017.00175>
- Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol* 139(1):5–17. <https://doi.org/10.1104/pp.105.063743>
- De Pascual-Teresa S, Sanchez-Ballesta MT (2008) Anthocyanins: from plant to health. *Phytochem Rev* 7:281–299
- Duan Y, Duan S, Armstrong MR, Xu J, Zheng J, Hu J, Chen X, Hein I, Li G, Jin L (2020) Comparative transcriptome profiling reveals compatible and incompatible patterns of potato toward *Phytophthora infestans*. *G3 Genes Genom Genet* 10(2):623–634. <https://doi.org/10.1534/g3.119.400818>
- Dzidzienyo DK, Bryan GJ, Wilde G, Robbins TP (2016) Allelic diversity of S-RNase alleles in diploid potato species. *Theor Appl Genet* 129(10):1985–2001. <https://doi.org/10.1007/s00122-016-2754-7>
- Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20(1):238. <https://doi.org/10.1186/s13059-019-1832-y>
- Enciso-Rodriguez F, Manrique-Carpintero NC, Nadakuduti SS, Buell CR, Zarka D, Douches D (2019) Overcoming self-incompatibility in diploid potato using CRISPR-cas9. *Front Plant Sci* 10:376. <https://doi.org/10.3389/fpls.2019.00376>
- FAO (2021) Statistical data. Rome.
- Feehan JM, Castel B, Benthall AR, Jones JD (2020) Plant NLRs get by with a little help from their friends. *Curr Opin Plant Biol* 56:99–108. <https://doi.org/10.1016/j.pbi.2020.04.006>
- Feller A, Machemer K, Braun EL, Grotewold E (2011) Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J* 66:94–116. <https://doi.org/10.1111/j.1365-3113.2010.04459.x>
- Fernandez-Pozo N, Zheng Y, Snyder SI, Nicolas P, Shinozaki Y, Fei Z, Catala C, Giovannoni JJ, Rose JKC, Mueller LA (2017) The tomato expression atlas. *Bioinformatics* 33(15):2397–2398. <https://doi.org/10.1093/bioinformatics/btx190>
- Ferrari C, Mutwil M (2020) Gene expression analysis of *Cyanophora paradoxa* reveals conserved abiotic stress responses between basal algae and flowering plants. *New Phytol* 225(4):1562–1577. <https://doi.org/10.1111/nph.16257>
- Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR (2015) HMMER web server: 2015 update. *Nucleic Acids Res* 43(W1):W30–W38. <https://doi.org/10.1093/nar/gkv397>
- Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44(D1):D279–D285. <https://doi.org/10.1093/nar/gkv1344>
- Freire R, Weisweiler M, Guerreiro R, Baig N, Hüttel B, Obeng-Hinneh E, Renner J, Hartje S, Muders K, Truberg B, Rosen A, Prigge V, Bruckmüller J, Lübeck J, Stich B (2021) Chromosome-scale reference genome assembly of a diploid potato clone derived from an elite variety. *G3 Genes Genom Genet* 11(12):jkab330. <https://doi.org/10.1093/g3journal/jkab330>
- Fujii S, Kubo K, Takayama S (2016) Non-self- and self-recognition models in plant self-incompatibility. *Nat Plants* 2(9):16130. <https://doi.org/10.1038/nplants.2016.130>

- Gaudinier A, Rodriguez-Medina J, Zhang L et al (2018) Transcriptional regulation of nitrogen-associated metabolism and growth. *Nature* 563(7730):259–264. <https://doi.org/10.1038/s41586-018-0656-3>
- Goh W, Mutwil M (2021) LSTrAP-Kingdom: an automated pipeline to generate annotated gene expression atlases for kingdoms of life. *Bioinformatics* 37(18):3053–3055. <https://doi.org/10.1093/bioinformatics/btab168>
- Gomez C, Terrier N, Torregrosa L, Viallet S, Fournier-Level A, Verries C et al (2009) Grapevine MATE-type proteins act as vacuolar h⁺-dependent acylated anthocyanin transporters. *Plant Physiol* 150:402–415. <https://doi.org/10.1104/pp.109.135624>
- Haas BJ, Papanicolaou A, Yassour M et al (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8(8):1494–1512. <https://doi.org/10.1038/nprot.2013.084>
- Hoang VLT, Tom LN, Quek XC et al (2017) RNA-seq reveals more consistent reference genes for gene expression studies in human non-melanoma skin cancers. *PeerJ* 5:e3631. <https://doi.org/10.7717/peerj.3631>
- Hoopes G, Meng X, Hamilton JP et al (2022) Phased, chromosome-scale genome assemblies of tetraploid potato reveal a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity. *Mol Plant* 15(3):520–536. <https://doi.org/10.1016/j.molp.2022.01.003>
- Hosaka K, Hanneman RE (1998) Genetics of self-compatibility in a self-incompatible wild diploid potato species *Solanum chacoense*. 1. Detection of an S locus inhibitor (Sli) gene. *Euphytica* 99:191–197. <https://doi.org/10.1023/A:1018353613431>
- Hu R, Fan C, Li H, Zhang Q, Fu YF (2009) Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. *BMC Mol Biol* 10:93. <https://doi.org/10.1186/1471-2199-10-93>
- Iizumi T, Luo JJ, Challinor AJ, Sakurai G, Yokozawa M, Sakuma H, Brown ME, Yamagata T (2014) Impacts of El Niño Southern oscillation on the global yields of major crops. *Nat Commun* 5:3712. <https://doi.org/10.1038/ncomms4712>
- Jaakola L, Poole M, Jones MO et al (2010) A SQUAMOSA MADS box gene involved in the regulation of anthocyanin accumulation in bilberry fruits. *Plant Physiol* 153(4):1619–1629. <https://doi.org/10.1104/pp.110.158279>
- Jansky SH, Charkowski AO, Douches DS, Gusmini G, Richael C, Bethke PC et al (2016) Reinventing potato as a diploid inbred line-based crop. *Crop Sci* 56:1412–1422. <https://doi.org/10.2135/cropsci2015.12.0740>
- Jung CS, Griffiths HM, De Jong DM et al (2009) The potato developer (D) locus encodes an R2R3 MYB transcription factor that regulates expression of multiple anthocyanin structural genes in tuber skin. *Theor Appl Genet* 120:45–57. <https://doi.org/10.1007/s00122-009-1158-3>
- Khoo HE, Azlan A, Tang ST, Lim SM (2017) Anthocyanidins and anthocyanins: colored pigments as food, pharmaceutical ingredients, and the potential health benefits. *Food Nutr Res* 61:1361779. <https://doi.org/10.1080/16546628.2017.1361779>
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37(8):907–915. <https://doi.org/10.1038/s41587-019-0201-4>
- Kryuchkova-Mostacci N, Robinson-Rechavi M (2017) A benchmark of gene expression tissue-specificity metrics. *Brief Bioinform* 18(2):205–214. <https://doi.org/10.1093/bib/bbw008>
- Kubo K, Entani T, Takara A, Wang N, Fields AM, Hua Z, Toyoda M, Kawashima S, Ando T, Isogai A, Kao TH, Takayama S (2010) Collaborative non-self recognition system in S-RNase-based self-incompatibility. *Science* 330(6005):796–799. <https://doi.org/10.1126/science.1195243>
- Kubo K, Paape T, Hatakeyama M, Entani T, Takara A, Kajihara K, Tsukahara M, Shimizu-Inatsugi R, Shimizu KK, Takayama S (2015) Gene duplication and genetic exchange drive the evolution of S-RNase-based self-incompatibility in *Petunia*. *Nat Plants* 1:14005. <https://doi.org/10.1038/nplants.2014.5>
- Laimbeer FPE, Bargmann BOR, Holt SH, Pratt T, Peterson B, Doullis AG, Buell CR, Veilleux RE (2020) Characterization of the F locus responsible for floral anthocyanin production in potato. *G3 Genes Genom Genet*. 10(10):3871–3879. <https://doi.org/10.1534/g3.120.401684>
- Lee JM, Roche JR, Donaghy DJ, Thrush A, Sathish P (2010) Validation of reference genes for quantitative RT-PCR studies of gene expression in perennial ryegrass (*Lolium perenne* L.). *BMC Mol Biol* 11:8. <https://doi.org/10.1186/1471-2199-11-8>
- Lee S, Lee T, Yang S, Lee I (2020) BarleyNet: a network-based functional omics analysis server for cultivated barley *Hordeum Vulgare* L. *Front Plant Sci* 11:98. <https://doi.org/10.3389/fpls.2020.00098>
- Leinonen R, Akhtar R, Birney E, Bonfield J et al (2010) Improvements to services at the European nucleotide archive. *Nucleic Acids Res* 38:D39–D45. <https://doi.org/10.1093/nar/gkp998>
- Leisner CP, Hamilton JP, Crisovan E, Manrique-Carpintero NC, Marand AP, Newton L, Pham GM, Jiang J, Douches DS, Jansky SH, Buell CR (2018) Genome sequence of M6, a diploid inbred clone of the high-glycoalkaloid-producing tuber-bearing potato species *Solanum chacoense*, reveals residual heterozygosity. *Plant J* 94(3):562–570. <https://doi.org/10.1111/tpj.13857>
- Lin H, Yu J, Pearce SP, Zhang D, Wilson ZA (2017) RiceAntherNet: a gene co-expression network for identifying anther and pollen development genes. *Plant J* 92(6):1076–1091. <https://doi.org/10.1111/tpj.13744>
- Lindhout P, Meijer D, Schotte T et al (2011) Towards F1 hybrid seed potato breeding. *Potato Res* 54:301–312. <https://doi.org/10.1007/s11540-011-9196-z>
- Lin-Wang K, Bolitho K, Grafton K, Kortstee A, Karunairetnam S, McGhie TK, Espley RV, Hellens RP, Allan AC (2010) An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. *BMC Plant Biol* 10:50. <https://doi.org/10.1186/1471-2229-10-50>
- Machado FB, Moharana KC, Almeida-Silva F, Gazara RK, Pedrosa-Silva F, Coelho FS, Grativol C, Venancio TM (2020) Systematic analysis of 1298 RNA-Seq samples and construction of a comprehensive soybean (*Glycine max*) expression atlas. *Plant J* 103(5):1894–1909. <https://doi.org/10.1111/tpj.14850>
- Mariot RF, de Oliveira LA, Voorhuijzen MM, Staats M, Hutten RC, Van Dijk JP, Kok E, Frazzon J (2015) Selection of reference genes for transcriptional analysis of edible tubers of potato (*Solanum tuberosum* L.). *PLoS ONE* 10(4):e0120854. <https://doi.org/10.1371/journal.pone.0120854>
- Massa AN, Childs KL, Lin H, Bryan GJ, Giuliano G, Buell CR (2011) The transcriptome of the reference potato genome *Solanum tuberosum* Group Phureja clone DM1-3 516R44. *PLoS ONE* 6(10):e26801. <https://doi.org/10.1371/journal.pone.0026801>
- McClure BA, Haring V, Ebert PR, Anderson MA, Simpson RJ, Sakiyama F, Clarke AE (1989) Style self-incompatibility gene products of *Nicotiana glauca* are ribonucleases. *Nature* 342(6252):955–957. <https://doi.org/10.1038/342955a0>
- McClure B, Gray J, Anderson M et al (1990) Self-incompatibility in *Nicotiana glauca* involves degradation of pollen rRNA. *Nature* 347:757–760. <https://doi.org/10.1038/347757a0>
- Meng D, Gu Z, Li W, Wang A, Yuan H, Yang Q, Li T (2014) Apple MdABCF assists in the transportation of S-RNase into pollen tubes. *Plant J* 78(6):990–1002. <https://doi.org/10.1111/tpj.12524>
- Merzlyak MN, Chivkunova OB (2000) Light-stress-induced pigment changes and evidence for anthocyanin photoprotection in apples.

- J Photochem Photobiol B 55:155–163. [https://doi.org/10.1016/S1011-1344\(00\)00042-7](https://doi.org/10.1016/S1011-1344(00)00042-7)
- Mutwil M, Usadel B, Schütte M, Loraine A, Ebenhöf O, Persson S (2010) Assembly of an interactive correlation network for the Arabidopsis genome using a novel heuristic clustering algorithm. *Plant Physiol* 152(1):29–43. <https://doi.org/10.1104/pp.109.145318>
- Navarro C, Abelenda JA, Cruz-Oró E, Cuéllar CA, Tamaki S, Silva J, Shimamoto K, Prat S (2011) Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. *Nature* 478(7367):119–122. <https://doi.org/10.1038/nature10431>
- Nicot N, Hausman JF, Hoffmann L, Evers D (2005) Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *J Exp Bot* 56(421):2907–2914. <https://doi.org/10.1093/jxb/eri285>
- Noh SA, Lee HS, Huh EJ, Huh GH, Paek KH, Shin JS, Bae JM (2010) SRD1 is involved in the auxin-mediated initial thickening growth of storage root by enhancing proliferation of metaxylem and cambium cells in sweet potato (*Ipomoea batatas*). *J Exp Bot* 61(5):1337–1349. <https://doi.org/10.1093/jxb/erp399>
- Paluchowska P, Śliwka J, Yin Z (2022) Late blight resistance genes in potato breeding. *Planta* 255(6):127. <https://doi.org/10.1007/s00425-022-03910-6>
- Patra B, Schluttenhofer C, Wu Y, Pattanaik S, Yuan L (2013) Transcriptional regulation of secondary metabolite biosynthesis in plants. *Biochim Biophys Acta* 1829:1236–1247. <https://doi.org/10.1016/j.bbagr.2013.09.006>
- Pertea G, Pertea M (2020) GFF utilities: GffRead and GffCompare. *F1000Res*. <https://doi.org/10.12688/f1000research.23297.2>
- Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 33(3):290–295. <https://doi.org/10.1038/nbt.3122>
- Pham GM, Hamilton JP, Wood JC, Burke JT, Zhao H, Vaillancourt B, Ou S, Jiang J, Buell CR (2020) Construction of a chromosome-scale long-read reference genome assembly for potato. *Gigascience* 9(9):giaa100. <https://doi.org/10.1093/gigascience/giaa100>
- Pieczynski M, Wyrzykowska A, Milanowska K, Boguszewska-Mankowska D, Zagdanska B, Karłowski W, Jarmolowski A, Szwejkowska-Kulinska Z (2018) Genome wide identification of genes involved in the potato response to drought indicates functional evolutionary conservation with Arabidopsis plants. *Plant Biotechnol J* 16(2):603–614. <https://doi.org/10.1111/pbi.12800>
- Proost S, Mutwil M (2018) CoNekT: an open-source framework for comparative genomic and transcriptomic network analyses. *Nucleic Acids Res* 46(W1):W133–W140. <https://doi.org/10.1093/nar/gky336>
- Qin T, Ali K, Wang Y, Dormatey R, Yao P, Bi Z, Liu Y, Sun C, Bai J (2022) Global transcriptome and coexpression network analyses reveal cultivar-specific molecular signatures associated with different rooting depth responses to drought stress in potato. *Front Plant Sci* 13:1007866. <https://doi.org/10.3389/fpls.2022.1007866>
- Ramšak Ž, Coll A, Stare T, Tzfadia O, Baebler Š, Van de Peer Y, Gruen K (2018) Network modeling unravels mechanisms of cross-talk between ethylene and salicylate signaling in potato. *Plant Physiol* 178(1):488–499. <https://doi.org/10.1104/pp.18.00450>
- Rao X, Chen X, Shen H, Ma Q, Li G, Tang Y, Pena M, York W, Frazier TP, Lenaghan S, Xiao X, Chen F, Dixon RA (2019) Gene regulatory networks for lignin biosynthesis in switchgrass (*Panicum virgatum*). *Plant Biotechnol J* 17(3):580–593. <https://doi.org/10.1111/pbi.13000>
- Rhee SY, Mutwil M (2014) Towards revealing the functions of all genes in plants. *Trends Plant Sci* 19(4):212–221. <https://doi.org/10.1016/j.tplants.2013.10.006>
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43(7):e47. <https://doi.org/10.1093/nar/gkv007>
- Riveros-Loaiza LM, Benhur-Cardona N, Lopez-Kleine L, Soto-Sedano JC, Pinzón AM, Mosquera-Vásquez T et al (2022) Uncovering anthocyanin diversity in potato landraces (*Solanum tuberosum* L. Phureja) using RNA-seq. *PLoS ONE* 17(9):e0273982. <https://doi.org/10.1371/journal.pone.0273982>
- Roldán JA, Rojas HJ, Goldraij A (2012) Disorganisation of F-actin cytoskeleton precedes vacuolar disruption in pollen tubes during the in vivo self-incompatibility response in *Nicotiana glauca*. *Ann Bot* 110(4):787–795. <https://doi.org/10.1093/aob/mcs153>
- Schulz E, Tohge T, Zuther E, Fernie AR, Hinch DK (2016) Flavonoids are determinants of freezing tolerance and cold acclimation in *Arabidopsis thaliana*. *Sci Rep* 6:34027. <https://doi.org/10.1038/srep34027>
- Serin EA, Nijveen H, Hilhorst HW, Ligterink W (2016) Learning from co-expression networks: possibilities and challenges. *Front Plant Sci* 7:444. <https://doi.org/10.3389/fpls.2016.00444>
- Sinha P, Saxena RK, Singh VK, Krishnamurthy L, Varshney RK (2015) Selection and validation of housekeeping genes as reference for gene expression studies in *Pigeonpea (Cajanus cajan)* under heat and salt stress conditions. *Front Plant Sci* 6:1071. <https://doi.org/10.3389/fpls.2015.01071>
- Sircar S, Musaddi M, Parekh N (2022) NetREx: network-based rice expression analysis server for abiotic stress conditions. *Database (Oxford)*. <https://doi.org/10.1093/database/baac060>
- Śliwka J, Brylińska M, Stefańczyk E et al (2017) Quantitative trait loci affecting intensity of violet flower colour in potato. *Euphytica* 213:254. <https://doi.org/10.1007/s10681-017-2049-3>
- Steuernagel B, Witek K, Krattinger SG et al (2020) The NLR-annotator tool enables annotation of the intracellular immune receptor repertoire. *Plant Physiol* 183(2):468–482. <https://doi.org/10.1104/pp.19.01273>
- Stintzing FC, Carle R (2004) Functional properties of anthocyanins and betalains in plants, food, and in human nutrition. *Trends Food Sci Tech* 15:19–38. <https://doi.org/10.1016/j.tifs.2003.07.004>
- Stuart JM, Segal E, Koller D, Kim SK (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302(5643):249–255. <https://doi.org/10.1126/science.1087447>
- Sun H, Jiao WB, Krause K, Campoy JA, Goel M, Folz-Donahue K, Kukut C, Huettel B, Schneeberger K (2022) Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. *Nat Genet* 54(3):342–348. <https://doi.org/10.1038/s41588-022-01015-0>
- Tai HH, Lagüe M, Thomson S et al (2020) Tuber transcriptome profiling of eight potato cultivars with different cold-induced sweetening responses to cold storage. *Plant Physiol Biochem* 146:163–176. <https://doi.org/10.1016/j.plaphy.2019.11.001>
- Tang X, Zhang N, Si H, Calderón-Urrea A (2017) Selection and validation of reference genes for RT-qPCR analysis in potato under abiotic stress. *Plant Methods* 13:85. <https://doi.org/10.1186/s13007-017-0238-7>
- Tang D, Jia Y, Zhang J et al (2022) Genome evolution and diversity of wild and cultivated potatoes. *Nature* 609(7929):E14. <https://doi.org/10.1038/s41586-022-05298-5>
- Teo CJ, Takahashi K, Shimizu K, Shimamoto K, Taoka KI (2017) Potato tuber induction is regulated by interactions between components of a tuberigen complex. *Plant Cell Physiol* 58(2):365–374. <https://doi.org/10.1093/pcp/pcw197>
- Tiwari JK, Buckseth T, Zinta R, Saraswati A, Singh RK, Rawat S, Dua VK, Chakrabarti SK (2020) Transcriptome analysis of potato shoots, roots and stolons under nitrogen stress. *Sci Rep* 10(1):1152. <https://doi.org/10.1038/s41598-020-58167-4>
- Ushijima K, Sassa H, Dandekar AM, Gradziel TM, Tao R, Hirano H (2003) Structural and transcriptional analysis of the

- self-incompatibility locus of almond: identification of a pollen-expressed F-box gene with haplotype-specific polymorphism. *Plant Cell* 15(3):771–781. <https://doi.org/10.1105/tpc.009290>
- Wisecaver JH, Borowsky AT, Tzin V, Jander G, Kliebenstein DJ, Rokas A (2017) A global coexpression network approach for connecting genes to specialized metabolic pathways in plants. *Plant Cell* 29(5):944–959. <https://doi.org/10.1105/tpc.17.00009>
- Wu S, Lau KH, Cao Q et al (2018) Genome sequences of two diploid wild relatives of cultivated sweet potato reveal targets for genetic improvement. *Nat Commun* 9(1):4580. <https://doi.org/10.1038/s41467-018-06983-8>
- Xia L, Zou D, Sang J, Xu X, Yin H, Li M, Wu S, Hu S, Hao L, Zhang Z (2017) Rice Expression Database (RED): An integrated RNA-Seq-derived gene expression database for rice. *J Genet Genomics* 44(5):235–241. <https://doi.org/10.1016/j.jgg.2017.05.003>
- Xiao SJ, Zhang C, Zou Q, Ji ZL (2010) TiSGeD: a database for tissue-specific genes. *Bioinformatics* 26(9):1273–1275. <https://doi.org/10.1093/bioinformatics/btq109>
- Xu X, Pan S, Potato Genome Sequencing Consortium et al (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475(7355):189–195. <https://doi.org/10.1038/nature10158>
- Yan L, Lai X, Wu Y, Tan X, Wang H, Zhang Y (2018) Co-expression network-based analysis associated with potato initial resistance. *bioRxiv*. <https://doi.org/10.1101/496075>
- Yang X, Guo X, Yang Y, Ye P, Xiong X, Liu J, Dong D, Li G (2018) Gene profiling in late blight resistance in potato genotype SD20. *Int J Mol Sci* 19(6):1728. <https://doi.org/10.3390/ijms19061728>
- Yang Y, Saand MA, Huang L, Abdelaal WB, Zhang J, Wu Y, Li J, Sirohi MH, Wang F (2021) Applications of multi-omics technologies for crop improvement. *Front Plant Sci* 12:563953. <https://doi.org/10.3389/fpls.2021.563953>
- Ye J, Jin CF, Li N, Liu MH, Fei ZX, Dong LZ, Li L, Li ZQ (2018a) Selection of suitable reference genes for qRT-PCR normalisation under different experimental conditions in *Eucommia ulmoides* Oliv. *Sci Rep* 8(1):15043. <https://doi.org/10.1038/s41598-018-33342-w>
- Ye M, Peng Z, Tang D et al (2018b) Generation of self-compatible diploid potato by knockout of *S-RNase*. *Nat Plants* 4:651–654. <https://doi.org/10.1038/s41477-018-0218-6>
- Yim AK, Wong JW, Ku YS, Qin H, Chan TF, Lam HM (2015) Using RNA-Seq data to evaluate reference genes suitable for gene expression studies in soybean. *PLoS ONE* 10(9):e0136343. <https://doi.org/10.1371/journal.pone.0136343>
- Yu H, Jiao B, Liang C (2017) Systematic analysis of RNA-seq-based gene co-expression across multiple plants. *bioRxiv*. <https://doi.org/10.1101/139923>
- Yuan X, Wang H, Cai J et al (2019) NAC transcription factors in plant immunity. *Phytopathol Res* 1:3. <https://doi.org/10.1186/s42483-018-0008-0>
- Zhang Y, Cheng S, De Jong D, Griffiths H, Halitschke R et al (2009a) The potato R locus codes for dihydroflavonol 4-reductase. *Theor Appl Genet* 119:931–937. <https://doi.org/10.1007/s00122-009-1100-8>
- Zhang Y, Jung CS, De Jong WS (2009b) Genetic analysis of pigmented tuber flesh in potato. *Theor Appl Genet* 119:143–150. <https://doi.org/10.1007/2Fs00122-009-1024-3>
- Zhang H, Wu Q, Cao S, Zhao T, Chen L, Zhuang P, Zhou X, Gao Z (2014) A novel protein elicitor (SsCut) from *Sclerotinia sclerotiorum* induces multiple defense responses in plants. *Plant Mol Biol* 86(4–5):495–511. <https://doi.org/10.1007/s11103-014-0244-3>
- Zhang J, Zheng H, Li Y, Li H, Liu X, Qin H, Dong L, Wang D (2016a) Coexpression network analysis of the genes regulated by two types of resistance responses to powdery mildew in wheat. *Sci Rep* 6:23805. <https://doi.org/10.1038/srep23805>
- Zhang H, Zhao T, Zhuang P, Song Z, Du H, Tang Z, Gao Z (2016b) NbCZF1, a novel C2H2-type zinc finger protein, as a new regulator of SsCut-induced plant immunity in *Nicotiana benthamiana*. *Plant Cell Physiol* 57(12):2472–2484. <https://doi.org/10.1093/pcp/pcw160>
- Zhang YL, Zhang CL, Wang GL, Wang YX, Qi CH, Zhao Q, You CX, Li YY, Hao YJ (2019) The R2R3 MYB transcription factor MdMYB30 modulates plant resistance against pathogens by regulating cuticular wax biosynthesis. *BMC Plant Biol* 19(1):362. <https://doi.org/10.1186/s12870-019-1918-4>
- Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, Banf M, Dai X, Martin GB, Giovannoni JJ, Zhao PX, Rhee SY, Fei Z (2016) iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol Plant* 9(12):1667–1670. <https://doi.org/10.1016/j.molp.2016.09.014>
- Zheng H, Brennan K, Hernaez M, Gevaert O (2019) Benchmark of long non-coding RNA quantification for RNA sequencing of cancer samples. *Gigascience*. 8(12):giz145. <https://doi.org/10.1093/gigascience/giz145>
- Zhou Q, Tang D, Huang W et al (2020) Haplotype-resolved genome analyses of a heterozygous diploid potato. *Nat Genet* 52(10):1018–1023. <https://doi.org/10.1038/s41588-020-0699-x>
- Zhu X, Li X, Chen W, Chen J, Lu W, Chen L, Fu D (2012) Evaluation of new reference genes in papaya for accurate transcript normalization under different experimental conditions. *PLoS ONE* 7(8):e44405. <https://doi.org/10.1371/journal.pone.0044405>
- Zierer W, Rüscher D, Sonnewald U, Sonnewald S (2021) Tuber and tuberous root development. *Annu Rev Plant Biol* 72:551–580. <https://doi.org/10.1146/annurev-arplant-080720-084456>

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