



# Comparative Analysis of Seed Transcriptome and Coexpression Analysis Reveal Candidate Genes for Enhancing Seed Size/Weight in *Brassica juncea*

Shikha Mathur<sup>1</sup>, Kumar Paritosh<sup>2</sup>, Rajesh Tandon<sup>3</sup>, Deepak Pental<sup>2</sup> and Akshay K. Pradhan<sup>2\*</sup>

<sup>1</sup>Department of Genetics, University of Delhi South Campus, New Delhi, India, <sup>2</sup>Centre of Genetic Manipulation of Crop Plants, University of Delhi South Campus, New Delhi, India, <sup>3</sup>Department of Botany, University of Delhi, New Delhi, India

## OPEN ACCESS

### Edited by:

Deepmala Sehgal,  
International Maize and Wheat  
Improvement Center (Mexico), Mexico

### Reviewed by:

Raju Datla,  
Global Institute for Food Security  
(GIFS), Canada  
Chuang Ma,  
Northwest A&F University, China

### \*Correspondence:

Akshay K. Pradhan  
pradhancgmcp@gmail.com

### Specialty section:

This article was submitted to  
Plant Genomics,  
a section of the journal  
Frontiers in Genetics

Received: 13 November 2021

Accepted: 12 January 2022

Published: 24 February 2022

### Citation:

Mathur S, Paritosh K, Tandon R,  
Pental D and Pradhan AK (2022)  
Comparative Analysis of Seed  
Transcriptome and Coexpression  
Analysis Reveal Candidate Genes for  
Enhancing Seed Size/Weight in  
*Brassica juncea*.  
Front. Genet. 13:814486.  
doi: 10.3389/fgene.2022.814486

Seed size/weight is a multigenic trait that is governed by complex transcriptional regulatory pathways. An understanding of the genetic basis of seed size is of great interest in the improvement of seed yield and quality in oilseed crops. A global transcriptome analysis was performed at the initial stages of seed development in two lines of *Brassica juncea*, small-seeded EH-2 and large-seeded PJ. The anatomical analyses revealed significant differences in cell number and cell size in the outer layer of the seed coat between EH-2 and PJ. Pairwise comparisons at each developmental stage identified 5,974 differentially expressed genes (DEGs) between the two lines, of which 954 genes belong to different families of transcription factors. Two modules were found to be significantly correlated with an increased seed size using weighted gene coexpression network analysis. The DEG and coexpression datasets were integrated with the thousand seed weight (Tsw) quantitative trait loci (QTL) mapped earlier in the EPJ (EH-2 × PJ) doubled haploid (DH) population, which identified forty potential key components controlling seed size. The candidate genes included genes regulating the cell cycle, cell wall biogenesis/modification, solute/sugar transport, and hormone signaling. The results provide a valuable resource to widen the current understanding of regulatory mechanisms underlying seed size in *B. juncea*.

**Keywords:** *Brassica juncea*, seed size, thousand seed weight, transcriptomics, RNA-seq, coexpression

## INTRODUCTION

Seed size/weight is one of the most important traits from the evolutionary and breeding perspectives, as it is directly related to the overall yield of a crop. Identification of factors for regulation of seed size is a major challenge for crop improvement programs, particularly in oilseed crops where the seed is the major source of the product. Seed size is determined by coordinated cell proliferation and/or cell expansion in the triploid endosperm, the diploid maternal tissue of the ovule, and the diploid embryo and is controlled by both maternal and zygotic genetic factors (Chaudhury and Berger, 2001; Sun et al., 2010). Seed development in *Arabidopsis thaliana* includes stages of embryogenesis, seed maturation, accumulation of storage compounds, and consumption of free nuclear endosperm by expanding cotyledons followed by rapid desiccation (Sun et al., 2010).

The controlling mechanisms of seed size in plants have been extensively studied in *A. thaliana* and *Oryza sativa* (rice) and include maternally imprinted genes, IKU pathway, ubiquitin–proteasome pathway, mitogen-activated protein kinase (MAPK signaling), G-protein signaling, phytohormones, and transcriptional regulatory factors, which have been well described in some reviews (Li and Li, 2016; Savadi, 2018; Li et al., 2019). It has been shown that the time of cellularization of syncytial endosperm, mitotic activity at initial stages of seed development, and partitioning of carbohydrates and photoassimilates play a major role in the determination of seed size (Borrás et al., 2004; Zhou et al., 2009; Wang et al., 2010; Ruan et al., 2012). It has been suggested that transcriptional trajectories at the early stages of seed development are mostly responsible for the determination of final seed size via restricting cell numbers (Ruan et al., 2012; Du et al., 2017).

A comparative transcriptome analysis of varieties contrasting in seed size at different stages of seed development in *Brassica napus* (Geng et al., 2018), *Brassica rapa* (Niu et al., 2020), soybean (Lu et al., 2016; Du et al., 2017), and maize (Sekhon et al., 2014) has identified promising candidate genes for improvement of the trait. However, no such analysis has been performed to dissect the molecular mechanisms controlling seed size/weight in *Brassica juncea* despite the fact that *B. juncea* germplasm harbors greater variation for seed size than *B. napus* and *B. rapa*. The genomic regions and candidate genes regulating seed size in *B. juncea* have been identified using marker-based bi-parental quantitative trait loci (QTL) mapping or genome-wide association studies (Ramchiary et al., 2007; Dhaka et al., 2017; Sra et al., 2019; Kumar and Bisht, 2020; Kang et al., 2021).

*B. juncea* (AABB) is an important oilseed crop in the Indian subcontinent, which is grown in more than six million hectares during the winter seasons in the northwestern part of the region. Heterosis breeding in *B. juncea* exploits the high genetic diversity between the east European and Indian gene pools to develop hybrid varieties with enhanced yields (Pradhan et al., 1993). Seed size is an important yield influencing parameter in *B. juncea*, and large seed size is desirable to fetch a high market price. The first commercial hybrid of *B. juncea* released in India, DMH -1 (Pusa bold × EH-2), is reported to give a 30% higher yield than the check varieties (Sodhi et al., 2006), but it has a small seed size (thousand seed weight (Tsw) ~ 3.9 g) as compared to Pusa Bold (Tsw ~ 6.4 g). The smaller seed size of the hybrid is attributed to the characteristic small seed size of the east European parent-EH-2, which is however a good combiner resulting in high heterosis when crossed with Indian varieties. Therefore, one of the major objectives of genetic improvement of *B. juncea* in India lies in the improvement of the east European gene pool through a precise transfer of some beneficial alleles of seed size from the Indian gene pool without compromising heterosis.

The east European gene pool line, EH-2, and the Indian gene pool line, Pusa Jaikisan (PJ), represent the highest phenotypic variability in seed size in *B. juncea* germplasm reported so far. Dhaka et al. (2017) carried out a QTL analysis of Tsw in *B. juncea* using a bi-parental doubled haploid (DH) population EPJ, derived from F1 of a cross between PJ and EH-2. The study identified twenty-five QTL with the logarithm of odds (LOD)

scores greater than 2.5 in the EPJ DH population from three environments. The beneficial alleles in all Tsw QTL were contributed by PJ.

The present study attempts to dissect the transcriptional regulatory pathways controlling seed size/weight in *B. juncea* lines EH-2 and PJ by conducting a comparative transcriptome analysis of the initial developmental stages of seed. The inference of coexpression networks identified coreregulated gene sets and hub genes associated with an increase in seed size, which could help deduce the gene regulatory networks underlying the regulation of seed size. Integration of our transcriptome dataset with the known genomic regions associated with Tsw in the EPJ DH population identified putative transcriptional regulators of seed size in *B. juncea*.

## MATERIALS AND METHODS

### Plant Material and Sample Collection

Two *B. juncea* lines, PJ and EH-2, with contrasting seed weights were grown in the field during the winter season of 2017–2018. Unopened flower buds on the main shoot were artificially self-pollinated and tagged for both the cultivars. Two pods per plant from 10 plants were harvested from each line at 5, 10, 15, 20, 25, and 30 days after pollination (DAP), referred to as stages S0, S1, S2, S3, S4, and S5, respectively. Self-pollinated seeds from 10 biological replicates were harvested at the mature dry stage, referred to as stage S6 from each line for measurement of seed size. At least 120 seeds from each stage were photographed, and seed size was determined using ImageJ (Schneider et al., 2012) software. Tsw and seed oil content of mature dry open-pollinated seeds were measured following Ramchiary et al. (2007). Levene's test and two-sample *t*-test were used to compare variances and means, respectively.

### Anatomical Analysis

For the anatomical analysis, harvested seeds were fixed in Karnovsky's fixative (Karnovsky, 1965) for 24 h at 4°C. Fixed seeds were washed and stored in 0.2 M of sodium cacodylate buffer, pH 7.4 at 4°C. The seeds were dehydrated in series with 2-methoxy ethanol, ethanol, and *n*-propanol for 24 h each followed by incubation in *n*-butanol for 12 h. The seeds were subsequently treated with freshly prepared glycol methacrylate monomer twice for 24 h each. The seeds were then embedded in the desired orientation in gelatin capsules containing the fresh monomer, which was allowed to polymerize at 40°C for 24 h followed by subsequent incubations at 50°C, 55°C, and 60°C for 24 h each (Feder and O'Brien, 1968). Sections (2 and 3 μm) of the resin-embedded seeds were cut using a rotary microtome (American Optical Co., United States, model 820) using glass knives. Sections were stained using Toluidine blue O, prepared in sodium benzoate buffer, pH 4.4 (O'Brien and McCully, 1981), mounted in DPX, and then observed under a bright-field microscope. At least 10 seeds sampled at stages S0–S5 from 10 different individuals were analyzed for each cultivar. For seeds sampled at the mature green stage (S5), the

cell number and cell size of the outermost epidermal layer of the seed coat were measured. Seeds were dissected to isolate seed coat, which was fixed overnight in FAA at 4°C. The tissue was then cleared by incubation in choral hydrate solution for 2 h. Differential interference contrast (DIC) microscopy was employed to take observations. The desired region of sections was photographed, a number of cells in the outer layer of the seed coat were counted, and the mean area of 500 cells was determined based on ten individuals using the ImageJ software.

## Sample Collection and RNA Isolation

Unopened floral buds on the main shoot were tagged 12–14 h before anthesis and were artificially self-pollinated. Siliques were opened using fine forceps/needles, and intact seeds were immediately frozen in liquid nitrogen. Samples were harvested from both PJ and EH-2 on the same day at approximately mid time of photoperiod, within a 2-h interval (1,100–1,300 h) for each stage. As the seed size is known to vary with the branch position in *B. napus* (Ren et al., 2017), seeds were collected from different positions (whorls) on the main shoot of five individuals for each stage. An equal number of seeds were pooled from each of the five individuals for one biological replicate. Seeds were collected in three biological replicates at stages S1–S3 and were stored at –80°C until further processing.

Total RNA was isolated from EH-2 and PJ seeds, using Spectrum Plant Total RNA kit (Sigma-Aldrich, St. Louis, MO, United States) following the manufacturer's protocol. Each sample was given on-column DNase I treatment using RNase-free DNase (Qiagen, Valencia, CA, United States). RNA samples were checked for genomic DNA contamination by PCR amplification of a housekeeping gene (UBQ9) and were quantified using a NanoDrop spectrophotometer. The integrity of RNA was checked using Bioanalyser (Agilent Technologies, Santa Clara, CA, United States), and RNA samples with RIN  $\geq$  6 were used for sequencing experiments.

## Transcriptome Sequencing and Quality Assessment

The 18 mRNA libraries (two lines, three stages (S1–S3), and three biological replicates each) were developed from high-quality RNA samples using NEBNext® UltraTM II RNA Library Prep Kit (New England Biolabs, Ipswich, MA, United States) following the manufacturer's protocol and were sequenced on Illumina platform (HiSeq 10×) (Illumina, San Diego, CA, United States) to generate 150-nucleotide-long paired-end (150 bp  $\times$  2) sequence reads. At least 4 GB of data per sample was generated, and the quality of the raw sequence data was assessed using FastQC.<sup>1</sup> The reads were then filtered for quality, and adapters were removed using Trimmomatic (Bolger et al., 2014) with parameters TruSeq3-PE-2.fa:2:30:10 LEADING:3 SLIDINGWINDOW:4:15 TRAILING:3 MINLEN:60.

## Differential Expression Analysis

The high-quality reads were mapped on the *B. juncea* var. *Varuna* genome sequence (Paritosh et al., 2021) using STAR aligner (Dobin et al., 2013) with default settings. The raw feature counts obtained thereof were normalized using the *variance stabilizing transformation* (vst) as implemented in DESeq2 v1.30.1 (Love et al., 2014) using Bioconductor (Gentleman et al., 2004) in R (R Core Team, 2019) for biological quality assessment and data visualizations. The differential expression analysis was performed on raw feature counts using a generalized linear model with Wald's statistical test as implemented in the DESeq2 package using the model  $\sim$  line + stage, where line represents the two *B. juncea* lines and stage represents the seed developmental stages. The genes with “lfc threshold” of 1 and an adjusted *p*-value  $\leq$  0.05 after correcting *p*-value using Benjamini and Hochberg (BH) correction were considered to be significantly differentially expressed.

## Identification of Transcription Factors, Gene Ontology, and Pathway Enrichment

Transcription factors (TFs) were identified using a homology search (blastp, *evalue*  $1e-25$ ) against Plant TF database<sup>2</sup> (Plant TFDB v5.0). The genes were analyzed for Gene Ontology (GO) enrichment using Blast2GO<sup>3</sup>, and significantly overrepresented GO terms were identified with a false discovery rate (FDR) threshold of  $<0.1$ . The enrichment results were summarized using REVIGO<sup>4</sup> (Supek et al., 2011). KAAS annotation server (Moriya et al., 2007) was used to assign KO terms to *B. juncea* var. *Varuna* genes using blastp; and hypergeometric test in Blast2GO program was used for pathway enrichment with an FDR threshold of  $<0.1$ . Enrichment results were visualized using R package ggplot2 (Wickham, 2011). Additionally, Mercator4 V2.0 (Schwacke et al., 2019) was used to annotate *B. juncea* var. *Varuna* genes using default parameters, and MapMan was used for pathway enrichment (significance value  $\leq$  0.05) of the differentially expressed genes (DEGs).

## Coexpression Network Analysis

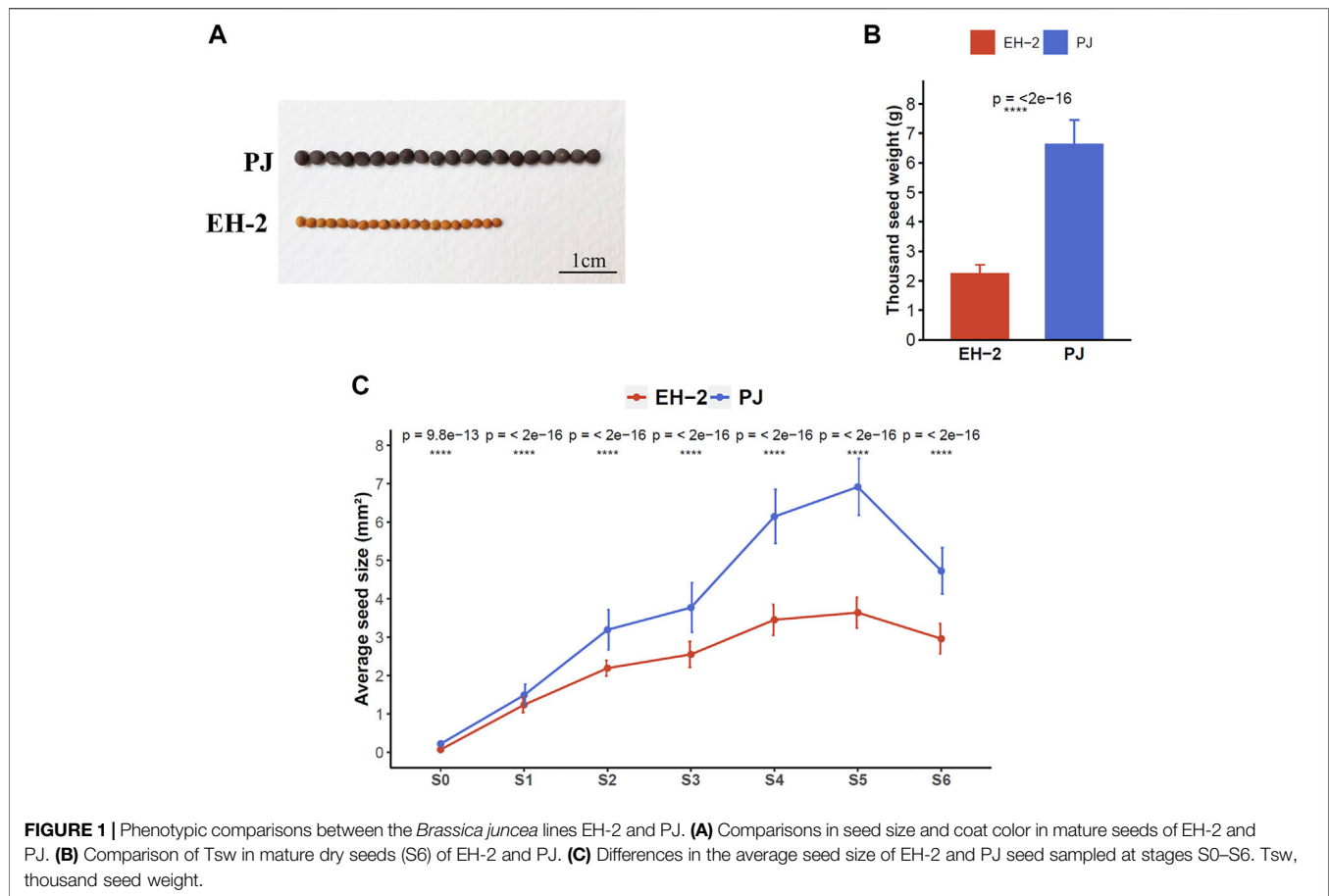
Coexpression networks were generated using the weighted gene coexpression network analysis (WGCNA) (Langfelder and Horvath, 2008) package in R to detect modules and key regulatory genes associated with seed size. The vst transformed expression data for expressed genes were used for coexpression analysis, and the genes with very low expression values were excluded from the analysis to avoid spurious nodes. A soft thresholding power of 18 was determined using pickSoftThreshold function based on the scale-free topology model fit ( $r^2 > 0.9$ ).

<sup>2</sup><http://plantfdb.cbi.pku.edu.cn>

<sup>3</sup><https://www.biobam.com/omicsbox>

<sup>4</sup><http://revigo.irb.hr>

<sup>1</sup><http://www.bioinformatics.babraham.ac.uk/projects/fastqc>



The `blockwiseModules` function was used to obtain coexpression modules with the following parameters: power, 18; network-type, signed; TOM-type, signed; minModuleSize, 50; maxBlockSize, 60,000; corType, bicor; maxPOutliers, 0.1; and mergeCutHeight, 0.25. `mergeCloseModules` function at `cutHeight = 0.2` was used to merge closely related modules after calculating the dissimilarity of module eigengenes. Pearson's correlation was used to correlate module eigengenes with trait values. Blast2GO was used for GO enrichment of modules. Cytoscape v3.8.2 (Smoot et al., 2011) was used to visualize the coexpression networks.

### Quantitative PCR Analysis

First-strand cDNA was synthesized from 2  $\mu$ g of total RNA using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA, United States) according to the manufacturer's instructions. Real-time quantitative reverse transcription-PCR (qRT-PCR) analyses were carried out using a PowerUp SYBR Green Master Mix (Applied Biosystems, United States) on QuantStudio 6 Flex 210 Real-Time PCR System (Applied Biosystems, United States). Expression profiling of the genes was carried out at stages S1–S3 for EH-2 and PJ to validate the transcriptome data. TIPS-41 gene (Chandna et al., 2012) was used as an internal control. The relative gene

expression levels were evaluated following the  $2^{-\Delta\Delta CT}$  method (Livak and Schmittgen, 2001).

## RESULTS

### Phenotypic and Anatomic Variations Between the Two *B. juncea* Lines

The two *B. juncea* lines, EH-2 and PJ, showed significant phenotypic contrasts in seed traits including seed size, Tsw, oil content, fatty acid composition, coat color, and glucosinolate content at maturity (Figure 1, Supplementary Table S1). The average seed size and Tsw of mature seed (S6) in PJ (seed size,  $4.73 \pm 0.6$  mm<sup>2</sup>; Tsw,  $6.62 \pm 0.84$  g) was significantly higher than those in EH-2 (seed size,  $2.95 \pm 0.4$  mm<sup>2</sup>; Tsw,  $2.24 \pm 0.3$  g) (Figures 1A,B). Phenotypic measurements for seed size and anatomical analysis were carried out at six stages (designated as S0–S5) encompassing major events in seed development, to correlate seed size with seed developmental stages in EH-2 and PJ (Figure 1C, Supplementary Figure S1). Anatomical analysis revealed stages S1–S3 as the initial stages of embryogenesis with 2- to 4-celled, globular, and transition/heart stage embryos observed at stages S1, S2, and S3, respectively. We observed cellularized micropylar

endosperm at the S3 stage (**Supplementary Figure S1**). Cellularization of syncytial endosperm begins between stages S2 and S3 in both EH-2 and PJ. The fresh seeds of PJ were significantly ( $p$ -value < 0.05) larger in size ( $\text{mm}^2$ ) than those of EH-2 at all the six developmental stages identified (**Figure 1C**). The seed size increased from S0 ( $0.22 \pm 0.07 \text{ mm}^2$  in PJ and  $0.07 \pm 0.02 \text{ mm}^2$  in EH-2) to S5 ( $6.92 \pm 0.74 \text{ mm}^2$  in PJ and  $3.64 \pm 0.4 \text{ mm}^2$  in EH-2) and decreased thereafter due to seed desiccation to a final seed size of  $4.73 \pm 0.6 \text{ mm}^2$  in PJ and  $2.95 \pm 0.4 \text{ mm}^2$  in EH-2 at S6. The maximum increase in seed size in EH-2 was observed at stages S0–S1 ( $1.17 \pm 0.03 \text{ mm}^2$ ) followed by stages S1–S2 ( $0.96 \pm 0.03 \text{ mm}^2$ ), and that in PJ was at stages S3–S4 ( $2.37 \pm 0.09 \text{ mm}^2$ ) followed by stages S1–S2 ( $1.7 \pm 0.05 \text{ mm}^2$ ) (**Supplementary Table S1**), indicating that the final seed size in *B. juncea* is mostly determined at the early stages of seed development. The average cell size and cell number of the outermost epidermal layer of seed coat in PJ seeds sampled at S5 were greater than in EH-2 (**Supplementary Figure S2**), showing that differential cell division and cell expansion in seed coat might partly contribute to the observed phenotypic differences in the seed size.

## Analysis of Transcriptome of Early Seed Development in the Two *B. juncea* Lines

As we observed a substantial and differential increase in seed size in EH-2 and PJ through stages S1–S3, a comparative transcriptome profiling of the seeds was performed in triplicates at stages S1, S2, and S3 to identify the factors controlling seed size. More than  $252.63 \times 10^6$  high-quality reads were generated from a transcriptome sequencing of EH-2 and PJ and mapped to the *B. juncea* var. Varuna genome with an average unique mapping rate of 88.74 (**Supplementary Table S2**). The raw feature counts (uniquely mapped reads) for each sample were normalized using variance stabilizing transformation (vst) as implemented in the DESeq2 package (Love et al., 2014) for biological quality assessment. Sample PS3c (PJ, S3) clustered with EH-2 stage S3 samples and was excluded from downstream analysis. Pearson's correlation between the biological replicates at different stages varied from 0.97 to 0.99, indicating high reproducibility among the biological replicates (**Supplementary Figure S3A**). Principal component analysis (PCA) assigned 17 samples to 6 distinct groups and identified *B. juncea* line from which tissue was sampled and the stage of seed development as the first two principal components explaining 44.1 and 29.23% variance (**Supplementary Figure S3B**). A total of ~77.5% of genes were expressed in at least one of the samples.

To validate the expression pattern of DEGs in the transcriptome analysis, a reverse transcription–quantitative PCR (RT-qPCR) analysis was performed for 9 DEGs in EH-2 and PJ at the S1–S3 stages of seed development. The relative expression levels (log<sub>2</sub>fold change) of the tested genes were found to follow the same expression trend as observed in the RNA-seq data (**Supplementary Figure S4**), indicating that the

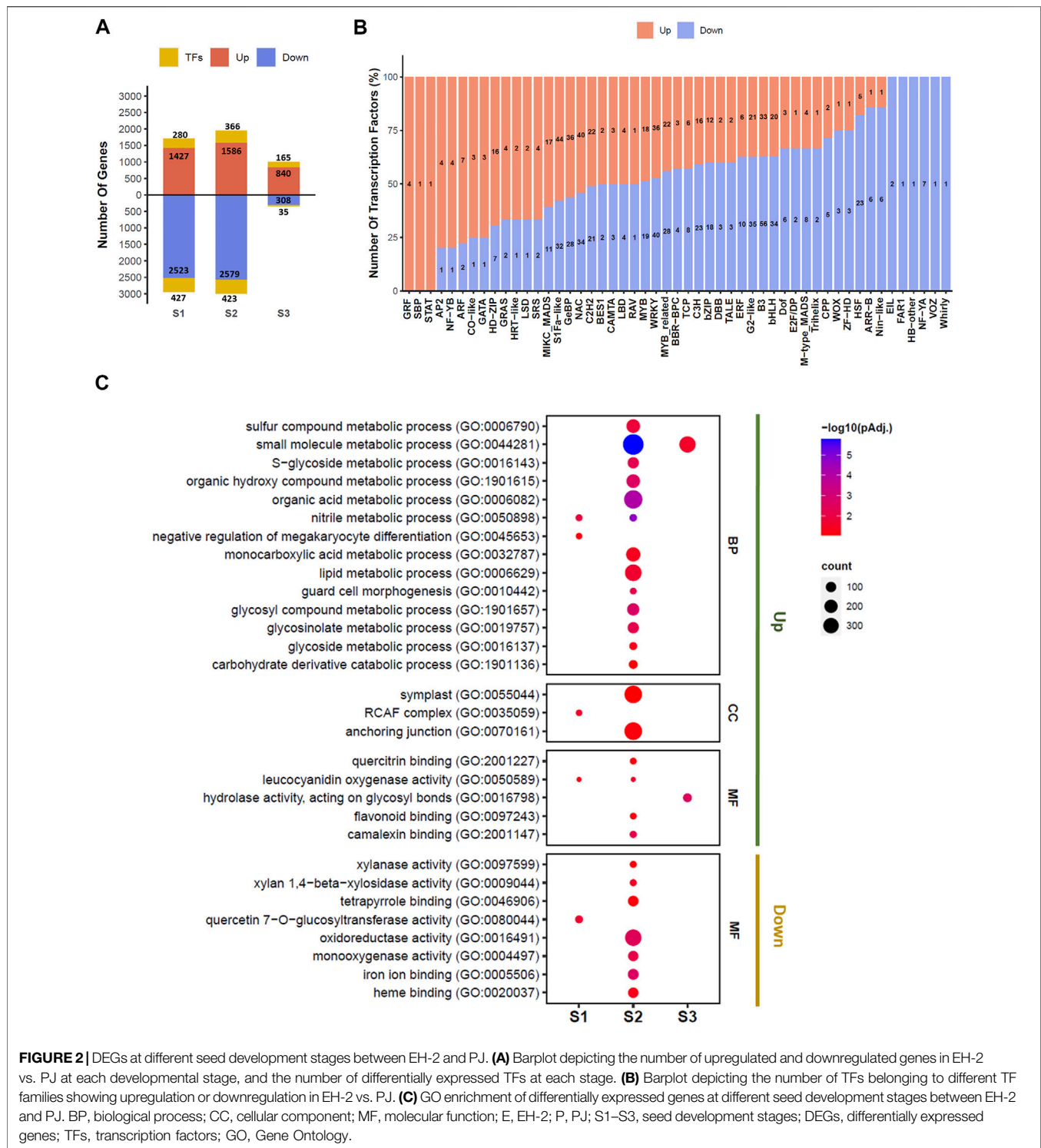
transcriptome data accurately reflected the transcript abundance levels.

## Differentially Expressed Genes During Initial Stages of Seed Development

A total of 5,974 genes were significantly differentially expressed (lfcthreshold = 1, padj < 0.05) in pairwise comparisons at the S1, S2, and S3 stages of seed development between EH-2 and PJ (**Supplementary Table S3**). A total of 2,447 genes including 441 TFs were significantly upregulated, whereas 3,527 genes including 513 TFs were significantly downregulated in EH-2 as compared with PJ at the S1–S3 stages of seed development. The highest number of genes (4,954) was found to be differentially expressed at the S2 stage, and the least number of genes (1,348) was differentially expressed at the S3 stage between EH-2 and PJ (**Figure 2A**).

A total of 18,596 TFs were identified in *B. juncea* similar to TFs documented in Plant TFDB v5.0 (**Supplementary Table S4**). Out of these, 954 TFs belonging to 51 TF families were differentially expressed between EH-2 and PJ at the S1–S3 stages of seed development (**Figure 2B**; **Supplementary Table S5**). B3, S1Fa-like, WRKY, and NAC TF families seem to play relatively important roles in the initial stages of seed development in *B. juncea*, as 89, 76, 76, and 74 members belonging to these families were differentially expressed in EH-2 and PJ. Members of several TF families, WRKY, ARR-B, GRF, MYB, ARF, B3, AP2, bHLH, bZIP, E2F/DP, MADS, MIKC-MADS, and NAC, were found to be differentially expressed in EH-2 and PJ. These TF families have been implicated in the regulation of seed size in different plant species (Mathew et al., 2016; Li et al., 2019). A total of 7 and 14 TFs belonging to CPP and TCP TF families, respectively, were differentially expressed in EH-2 and PJ (**Figure 2B**). CPP and TCP TF families regulate organ growth via the regulation of cell proliferation (Yang et al., 2008; Martín-Trillo and Cubas, 2010). Members of the ARF TF family, which regulate auxin-responsive gene expression (Guilfoyle and Hagen, 2007), were upregulated in EH-2, whereas members of the ARR-B TF family, the regulators of cytokinin signal transduction (Sakai et al., 2001), were upregulated in PJ. A total of 64 genes belonging to the GePB TF family of cytokinin-responsive genes (Chevalier et al., 2008) were differentially expressed in EH-2 and PJ. These results suggest that the levels of auxin and cytokinins are differentially regulated during the S1–S3 stages of seed development in EH-2 and PJ. The expression profiles of differentially expressed TFs are given in **Supplementary Table S5**.

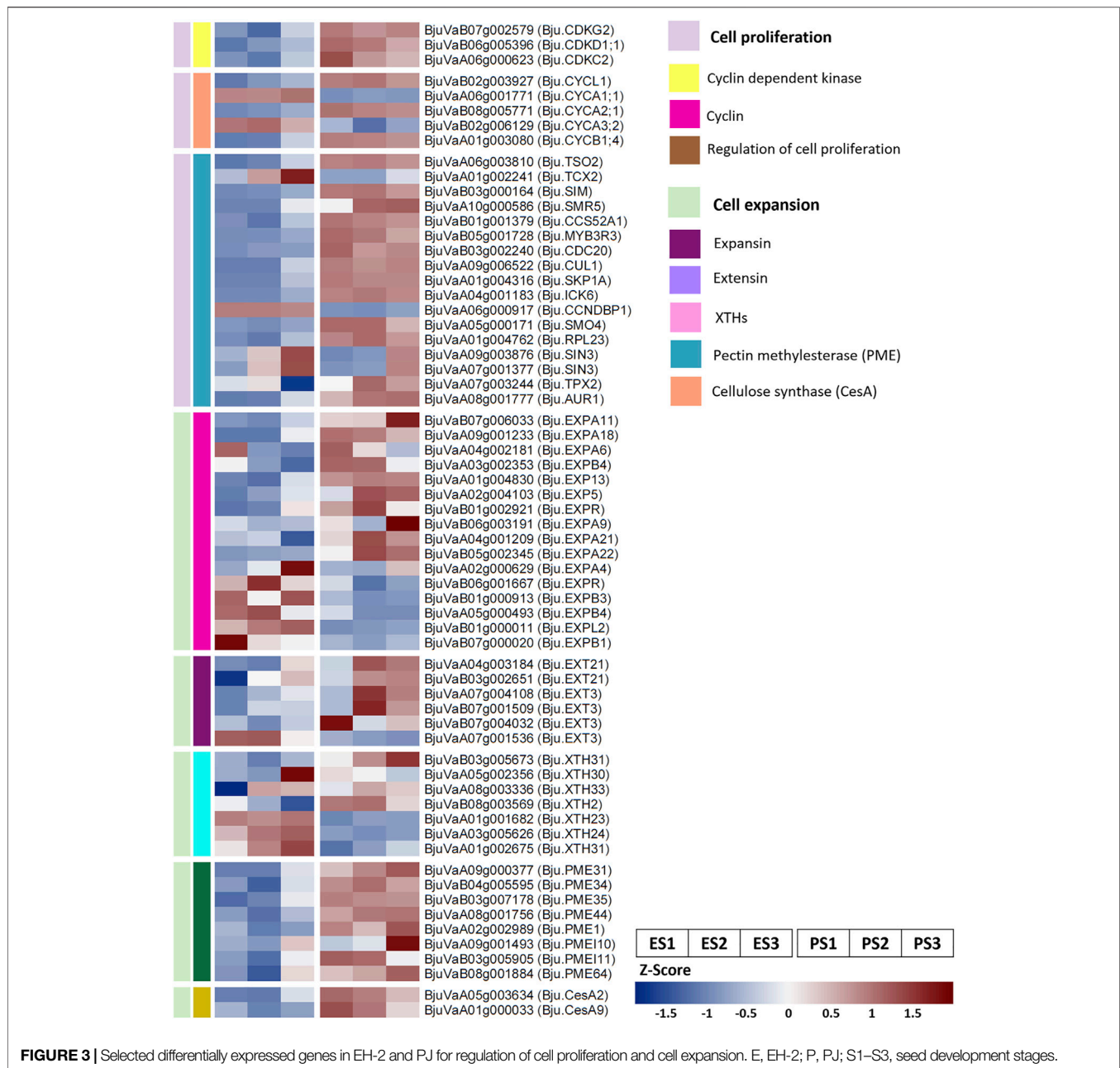
The GO enrichment analysis of DEG revealed several GO terms either commonly or uniquely overrepresented at the S1–S3 stages of seed development (**Figure 2C**; **Supplementary Table S6**). The majority of the GO terms were uniquely overrepresented at the S2 stage, suggesting the onset of a major developmental reprogramming of seed development. We observed a consistent overrepresentation of GO terms “lipid metabolic process (GO:0006629)” and “glycosinolate metabolic process (GO:0019757)” in genes upregulated in EH-2 at the S2 stage, suggesting the importance of this stage in differential regulation of fatty acid and glucosinolate content in EH-2 and PJ. The GO terms overrepresented in genes downregulated in EH-2 were “xylanase activity (GO:0097599)” and “xylan 1–4 beta



xylosidase activity (GO:0009962),” indicating increased activity of cell wall-modifying enzymes in PJ. The Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology (KO) enrichment analysis revealed pathways differentially regulated in EH-2 and PJ at the S1–S3 stages of seed development (**Supplementary Table S7**). The KO pathway “Flavonoid biosynthesis (ko00941)”

was overrepresented in genes significantly upregulated in PJ, which is consistent with the brown coat color in PJ.

The DEGs between EH-2 and PJ at stages S1–S3 were analyzed using the MapMan tool to further investigate the metabolic pathways and cellular functions differentially regulated in EH-2 and PJ (**Supplementary Figure S5**). The genes involved in cell wall



biosynthesis and modification showed higher expression in PJ at the S1–S2 stages and might be involved in regulating the differences in cell size of seed epidermis as observed in EH-2 and PJ.

## Differential Regulation of Genes Determining Seed Size

We identified the DEGs involved in key cellular processes regulating seed size including cell proliferation, cell expansion, endosperm development, and embryonic development (Figure 3; Supplementary Table S8). In addition, we identified the *B. juncea* homologs of genes reported to regulate seed size in heterologous systems

showing differential expression in EH-2 and PJ (Supplementary Table S9).

EH-2 and PJ exhibited differential regulation of cell proliferation at the S1–S3 stages of seed development. Bju.SMO4 (BjuVaA05g000171) was significantly upregulated in PJ at S2. SMO4 is a key regulator of organ growth, and mutation in SMO4 decreases organ size by restricting cell proliferation (Zhang et al., 2015). Interestingly, we found an upregulation of all five paralogs of Bju.AN3 (BjuVaA02g004301, BjuVaA06g003464, BjuVaA09g000483, BjuVaB02g000498, and BjuVaB06g001001) in small-seeded EH-2. AN3 has been reported to influence seed size in *A. thaliana* as a result of increased cell proliferation or cell size (Meng et al., 2016; Liu et al., 2020). Several genes involved in the regulation

of the cell cycle were also found to be differentially expressed in EH-2 and PJ. Cyclins Bju.CYCB1;4 (BjuVaA01g003080), Bju.CYCL1 (BjuVaB02g003927), and Bju.CYCA2;1 (BjuVaB08g005771) were upregulated in PJ at S1 and S2. On the other hand, A-type cyclins Bju.CYCA1;1 (BjuVaA06g001771) and Bju.CYCA3;2 (BjuVaB02g006129) were upregulated in EH-2 at S1 and S2. Bju.TSO2 (BjuVaA06g003810) was found to be upregulated in PJ at S1 and S2. It is predominantly expressed in the S phase of the cell cycle and plays a role in dNTP biosynthesis in actively dividing cells (Wang and Liu, 2006). Bju.TCX2 (BjuVaA01g002241) was found to be upregulated in EH-2 at S2. It has been reported as a negative regulator of the cell cycle (Clark et al., 2019). We observed a differential expression of DNA replication machinery in EH-2 and PJ. There was an upregulation of Bju.SIM (BjuVaB03g000164), Bju.SMR5 (BjuVaA10g000586), and Bju.CCS52A1 (BjuVaB01g001379) in PJ at the S1 and S2 stages. SIM and CCS52A1 act together to increase endoreplication in *A. thaliana* trichomes (Kasili et al., 2010).

Gene enrichment analysis revealed increased activity of cell wall modification enzymes in PJ. We found a consistent upregulation of the genes involved in plant cell wall biogenesis that included cellulose synthases (CeSA), xyloglucan endotransglucosylases/hydrolases (XTHs), xyloglucan xylosyltransferase (XXTs), and pectin methylesterases (PMEs) in PJ (Figure 3; Supplementary Table S8). XTHs are implicated in the regulation of cell wall loosening and cell expansion (Rose et al., 2002). We observed a significant upregulation of several genes encoding extensin and expansin proteins in PJ (Figure 3), which also suggests a higher rate of cell expansion in PJ. These data suggest that upregulation of genes is associated with an increase in cell size and cell proliferation in PJ as compared to EH-2 and might explain the higher cell size and cell number in the outermost epidermal layer of seed coat as observed in PJ at S5 (Supplementary Figure S2).

## Identification of Coexpression Modules Regulating Seed Size

In an independent analysis, coexpressed gene sets were identified using WGCNA to investigate the transcriptional networks controlling seed size. A total of 18 modules that grouped genes with similar expression patterns were identified after merging closely related modules (Supplementary Figure S6C; Supplementary Table S10). The module eigengene of each coexpression module was associated with stages of seed development, genotype, and an increase in seed size via calculation of Pearson's correlation coefficient (Supplementary Figure S6D). Several modules were correlated with more than one developmental stage; however, some modules were specific to just one stage, indicating stage-specific coexpression networks. For example, cyan, royalblue, and brown modules showed high positive correlations with stages S1, S2, and S3, respectively. The list of genes and GO terms associated with each WGCNA module is given in Supplementary Tables S10, S11, respectively. We identified the hub genes in our module of interest using gene connectivity ( $K_{\text{within}}$ ) measure. The genes with the top 10%  $K_{\text{within}}$  in a module were identified as the hub genes in that module.

Module eigengenes of pink ( $r^2 = -0.8$ ,  $p$ -value =  $1e-04$ ) and darkred ( $r^2 = 0.79$ ,  $p$ -value =  $1e-04$ ) modules were found to be

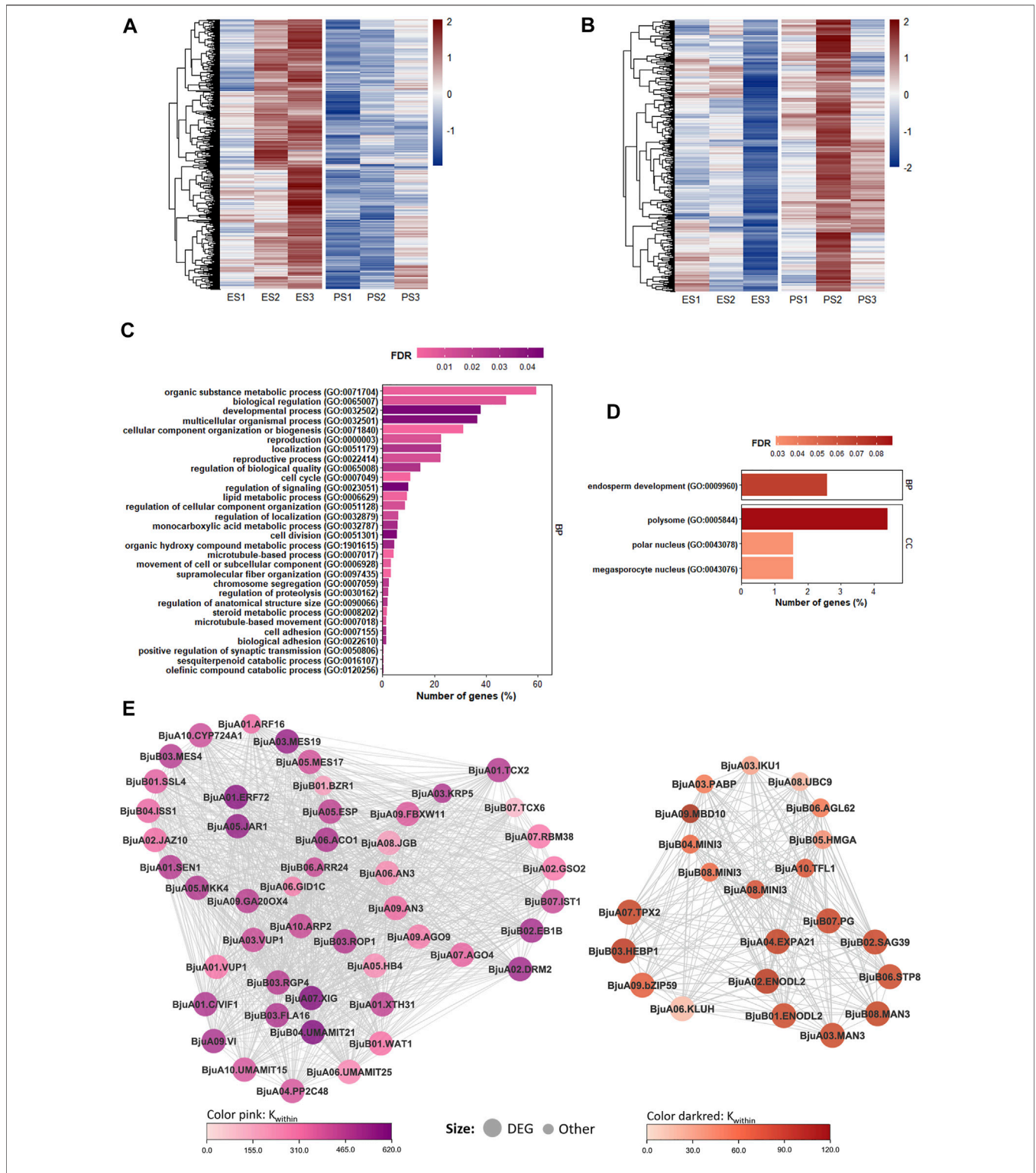
highly associated with an increase in seed size, and these modules harbored genes with high gene significance for an increase in seed size (Supplementary Figure S6E; Supplementary Table S10). The pink module harbored 2,383 genes with higher expression in EH-2 at the S2 and S3 stages and showed a negative correlation with the increase in seed size (Figure 4A). The  $K_{\text{within}}$  values in the pink module ranged from 3.9 to 616.6. A total of 250 genes in the pink module including 42 TFs were significantly ( $\text{padj} < 0.05$ ,  $\text{lfcthreshold} = 1$ ) upregulated in EH-2. GO enrichment analysis of pink module revealed a significant overrepresentation of GO terms "reproductive process (GO:0022414)," "cell cycle (GO:0022414)," "cell division (GO:0051301)," and "regulation of anatomic structure size (GO:0090066)" (Figure 4C). Several DEGs identified as regulators of cell proliferation and expansion (Figure 3) also showed up as hub genes in the pink module (Figure 4E). Bju.TCX2 (BjuVaA01g002241,  $K_{\text{within}} = 485.4$ ) was identified as a hub gene in the pink module and was significantly upregulated in EH-2 (Figure 4E). TCX2 restricts cell proliferation via the regulation of genes in the DREAM complex (Clark et al., 2019). Genes involved in small RNA mediated DNA methylation in CHH context; Bju.AGO4 (BjuVaA07g001798), Bju.AGO9 (BjuVaA09g003432), and Bju.DRM2 (BjuVaA02g000576) were upregulated in EH-2, indicating their relevance in the negative regulation of seed size in *B. juncea*. We identified several DEGs involved in hormone signaling in the pink module with Bju.ERF72 (BjuVaA01g004064,  $K_{\text{within}} = 585.1$ ) and Bju.JAR1 (BjuVaA05g000146,  $K_{\text{within}} = 555.02$ ) as the hub genes. Together, the coregulated gene sets in the pink module provide insights on transcriptional programs, restricting seed growth in *B. juncea*.

The darkred module harbored 386 genes with higher expression in PJ at the S2 and S3 stages and showed the highest positive correlation with an increase in the seed size (Figure 4B). The  $K_{\text{within}}$  values in the darkred module ranged from 4.6 to 114.3. A total of 27 genes in the darkred module including five TFs were significantly ( $\text{padj} < 0.05$ ,  $\text{lfcthreshold} = 1$ ) upregulated in PJ. The GO enrichment of the darkred module revealed an overrepresentation of biological process "endosperm development (GO:0009960)" (Figure 4D). Based on  $K_{\text{within}}$  scores, we identified DEGs, Bju.ENODL2 (BjuVaA02g001491,  $K_{\text{within}} = 104.2$ ), Bju.HEBP1 (BjuVaB03g001743,  $K_{\text{within}} = 103.59$ ), Bju.EXP21 (BjuVaA04g001209,  $K_{\text{within}} = 99.07$ ), and Bju.TPX2 (BjuVaA07g003244,  $K_{\text{within}} = 95.89$ ) as the hub genes in the darkred module (Figure 4E). Bju.KLUH (BjuVaB03g001743) was significantly upregulated in PJ at S3. KLUH has been reported to increase seed size and oil content in *A. thaliana* seeds (Adamski et al., 2009). The coexpression of these DEGs with key regulators of endosperm development and seed size including Bju.AGL62, Bju.MINI3, Bju.IKU1, and Bju.UBC9 warrants a further investigation of their roles in seed development with respect to regulation of seed size in *B. juncea*.

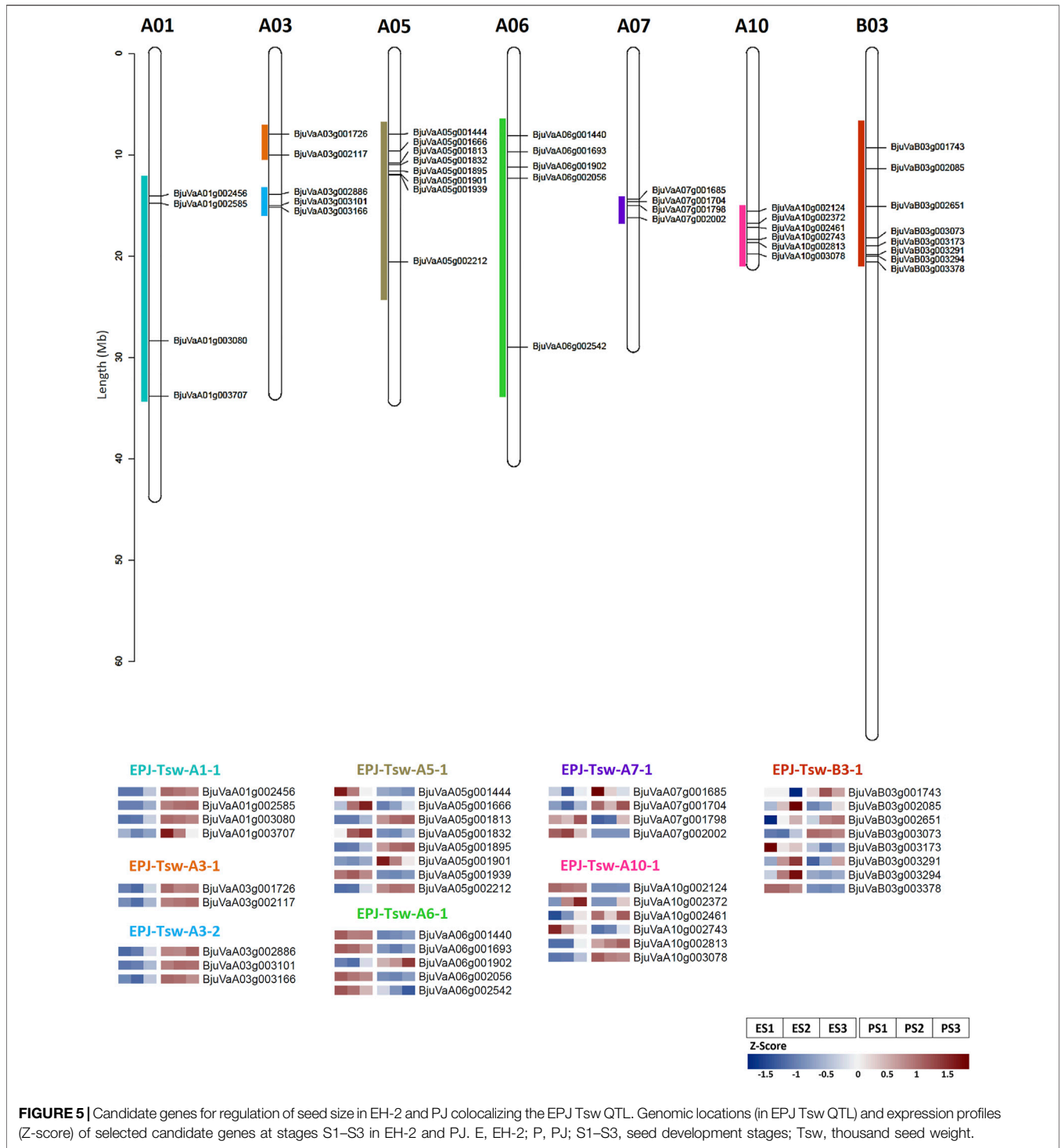
## Candidate Genes Underlying the Quantitative Trait Locus Governing Seed Size

We identified the DEGs colocalizing EPJ Tsw QTL (Dhaka et al., 2017) to further shortlist candidate genes involved in the regulation of the seed size in EH-2 and PJ. A total of 588





**FIGURE 4 |** Coexpression modules with the highest correlation with increase in seed size. **(A)** Heatmap showing expression profiles of the pink module. The color scale on the right represents Z-scores. **(B)** Heatmap showing expression profiles of the darkred module. The color scale on the right represents Z-scores. **(C)** GO enrichment (top 30 biological process enriched terms) of the pink module. **(D)** GO enrichment of the darkred module. **(E)** Sub-networks connecting genes related to cell cycle and cell division in the pink module and endosperm development in the darkred module to hub genes in EPJ Tsw QTL. E, EH-2; P, PJ; S1–S3, seed development stages; GO, Gene Ontology; Tsw, thousand seed weight.



genes encompassing eight Tsw QTL exhibited differential expression in the EH-2 and PJ seeds (**Supplementary Table S13**). We identified a total of 40 DEGs colocalizing the eight EPJ Tsw QTL as candidates for the regulation of seed size on the basis of their association with seed size/weight, cell proliferation and cell wall biogenesis/modification, and/or hub gene status in pink or darkred modules with gene significance for an increase in

seed size  $>\pm 0.7$  (**Figure 5; Supplementary Table S14**). These 40 candidate genes include 15 TFs belonging to the ARF, GeBP, WRKY, EIL, E2F/DP, GRF, CPP, C3H, bHLH, MYB-related, NAC, and DoF TF families. The candidate genes for seed size include genes regulating cell proliferation/division, cell wall biogenesis/modification, solute/hormone transporters, DNA methylation, auxin, cytokinin, ethylene, and JA responsive

genes (**Supplementary Table S14**) highlighting the importance of these pathways in differential regulation of seed size in *B. juncea*.

Several DEGs related to the regulation of cell wall biosynthesis/modification including Dof zinc finger protein DOF3.1 (BjuVaA01g003707), galactosyl transferase family protein XXT3 (BjuVaA10g002813), a CPP family TF endoglucanase 25 (BjuVaA05g001901), EXT21 (BjuVaB03g002651), FLA16 (BjuVaB03g002085), and RGP4 (BjuVaB03g003291) colocalized EPJ Tsw QTL and are candidates for the regulation of seed size. We observed an upregulation of a C3H TF, Bju.CYAB1;4 (BjuVaA01g003080) colocalizing EPJ-Tsw-A1-1 in PJ at S1 and S2, suggesting higher rates of mitotic activity in PJ at stages S1 and S2. An increased expression of CYAB1;4 has been shown to increase seed size in *A. thaliana* (Ren et al., 2019). Several auxin-responsive DEGs including Auxin transporter-like protein 1 (BjuVaA03g002117), Bju.AGO10 (BjuVaA01g002456), Bju.Dof3.1 (BjuVaA03g001726), Bju.ARF2 (BjuVaA06g002542), Bju.IAA22 (BjuVaA06g001440), and Bju.PATL5 (BjuVaB03g003378) colocalized EPJ Tsw QTL and suggest a role of auxin signaling in differential regulation of seed size in *B. juncea*. EPJ-Tsw-A3-2 QTL harbored a cytokinin-responsive gene, a two-component response regulator ARR22-like protein (BjuVaA03g003166) with higher transcript abundance in PJ as a putative candidate for regulation of seed size. A component of the ethylene signaling cascade, Bju.EIL1 (BjuVaB03g003073), was significantly upregulated in PJ at S1 and S2 and colocalize with EPJ-Tsw-B3-1 QTL, suggesting a role of ethylene signaling in the regulation of the seed size. *O. sativa* homolog of Bju.EIL1 has been reported to increase grain size via mutant analysis (Yang et al., 2015). EPJ-Tsw-A10-1 QTL harbored Bju.SWEET15 (BjuVaA10g002461), upregulated in PJ, is one of the candidates for the regulation of seed size. SWEET genes play critical roles in sugar transport from seed coat to developing embryo during seed filling, and a *sweet11;12;15* triple mutant has been shown to reduce dry seed weight in *A. thaliana* (Chen et al., 2015). EPJ-Tsw-A5-1 QTL harbored Bju.FAD3 (BjuVaA05g001444) gene, which was significantly upregulated in EH-2. Silencing of gmFAD3 has been reported to increase seed size in soybean (Singh et al., 2011). Based on these results, it is possible to target these genes for the enhancement of seed size in *B. juncea*.

## DISCUSSION

Seed size is a multigenic trait that is predominantly controlled by cues from maternal and zygotic tissues with some environmental influence (Li and Li, 2016). *B. juncea* holds the maximum diversity in seed size/weight among the cultivated Brassicas, with PJ (Indian gene pool) and EH-2 (east European gene pool) showing high phenotypic contrasts for the trait; however, the mechanism of regulation of seed size in *B. juncea* is poorly understood. QTL analysis for Tsw in the EPJ DH population identified large genomic regions harboring several genes. Furthermore, a large number of small-effect Tsw

QTLs were identified in the study (Dhaka et al., 2017), indicating a role of multiple genetic factors in differential regulation of seed size in EH-2 and PJ, which limited the scope of Tsw QTL dissection in the EPJ population using the candidate gene approach. Therefore, a comparative transcriptomics approach was undertaken to identify putative candidate genes and to understand the regulatory mechanisms underlying the trait.

We performed a comprehensive transcriptome profiling of the initial stages of seed development (S1–S3) in EH-2 and PJ. The expression data showed high reproducibility and distinct clustering of seed development stages in the PCA plot, reaffirming the results from the anatomical analysis that embryogenesis progresses at a comparable pace in both large- and small-seeded *B. juncea* lines. We identified 5,974 DEGs including 954 TFs belonging to 51 TF families with an overrepresentation of B3, S1Fa-like, WRKY, and NAC TF families in regulating early seed development in *B. juncea*. The role of these TF families has been established in seed development (Le et al., 2010; Agarwal et al., 2011). Members of several TF families, WRKY, ARR-B, GRF, MYB, ARF, B3, AP2, bHLH, bZIP, E2F/DP, MADS, MIKC-MADS, NAC, CPP, and TCP showed differential expression in EH-2 and PJ in our dataset (**Figure 2B**). These TF families have been implicated in the control of the seed size and/or organ growth (Yang et al., 2008; Martín-Trillo and Cubas, 2010; Mathew et al., 2016; Li et al., 2019). However, the exact role of this TF in the regulation of seed size remains to be established.

Seed size is mostly regulated via differential cell proliferation and cell expansion in different seed compartments, which influence the final seed size (Savadi, 2018; Li et al., 2019). Differential increase in seed size at different stages in EH-2 and PJ (**Figure 1C**) and increased cell size and cell number in the epidermal layer of seed coat in PJ seeds sampled at S5 as compared to EH-2 (**Supplementary Figure S2**) suggested that differential regulation of cell proliferation and cell expansion coordinately explain the observed phenotypic variation in seed size in these lines. It has been hypothesized that common genetic components regulate the size of seeds and other organs via affecting cell size and cell number (Linkies et al., 2010). We observed an upregulation of genes related to cell wall biogenesis and/or expansion including expansins, extensins, XTHs, PMEs, and CeSA (**Figure 3; Supplementary Table S3**) in PJ, suggesting that higher rates of cell wall biosynthesis and/or expansion in PJ might be responsible for its larger seed size via an increase in cell size. We also observed differential regulation of cell cycle and DNA replication machinery in EH-2 and PJ (**Figure 3; Supplementary Table S8**). There was an upregulation of CYCB1;4 (BjuVaA01g003080) gene in PJ at stages S1 and S2, indicating that increased cell proliferation at these stages might result in increased seed size in PJ. Altogether, higher cell expansion accompanied by an increased mitotic activity in PJ might explain its larger seed size.

We identified coexpressed gene sets/modules correlated with an increase in the seed size. The GO analysis of

modules correlated with an increase in seed size, which suggested an influence of genes involved in the cell cycle, cell division, and endosperm development in the determination of the final seed size in EH-2 and PJ. We identified Bju.TCX2 (BjuVaA01g002241) as hub gene in pink coexpression module, negatively associated with seed size. TCX2 is a homolog of the DREAM complex gene LIN54 and restricts stem cell proliferation in *A. thaliana* root (Clark et al., 2019). The role of the DREAM complex in the regulation of seed size has been hypothesized via transcriptome and coexpression analysis of seed development in *B. rapa* (Niu et al., 2020). Bju.TPX2 was upregulated in PJ and was identified as a hub gene in the darkred module, which is positively correlated with an increase in the seed size. TPX2 is a positive regulator of cell proliferation (Tomašíková et al., 2020), and its coexpression with genes associated with endosperm development warrants detailed analysis of its role in seed development in *B. juncea* with respect to the determination of seed size.

An interplay of multiple plant hormones, including auxins, cytokinins, gibberellins, brassinosteroids, ethylene, and JA, have been shown to regulate different aspects of seed development including seed size (Savadi, 2018; Li et al., 2019; Hu et al., 2021). Pathway analysis suggested differential hormone regulation in EH-2 and PJ during the S1–S3 stages (**Supplementary Figure S5**). Additionally, *B. juncea* homologs of several hormonal regulators of seed size in *A. thaliana* and *O. sativa* were differentially expressed in EH-2 and PJ (**Supplementary Table S9**). We observed that 7 out of 9 ARF family TFs were upregulated in EH-2 (**Figure 2B**). An auxin-induced JA responsive gene Bju.JAR1 was significantly upregulated in EH-2 and was identified as a hub gene in the pink module (**Figure 4E**). There is limited evidence of the role of JA signaling in the regulation of seed size. However, a recent study by Hu et al. (2021) has shown that increased levels of JA repress seed size via negative regulation of the integument cell proliferation. We identified genes in the ethylene signaling pathway, Bju.ACO1 and Bju.ERF2, as hub genes in the pink coexpression module (**Figure 4E**). Ethylene has been shown to influence the seed size in the early stages of seed development in *B. napus* in a dose-dependent manner (Walton et al., 2012); however, the mechanism of control of seed size via ethylene signaling is not well characterized. An overexpression of ERF72 has been reported to inhibit hypocotyl elongation in *A. thaliana* by inhibition of cell elongation and links auxin, brassinosteroid, and ethylene signaling (Liu et al., 2018). Further functional studies on the hub genes in the coexpression networks might reveal novel genes for improving seed size and could be crucial to understanding the regulatory mechanisms underlying the trait.

An integration of the DEG and coexpression datasets with eight EPJ Tsw QTL identified forty candidate genes for engineering seed size in *B. juncea*, including genes regulating cell cycle, cell wall biogenesis/modification, solute/sugar transport, and hormone signaling (**Supplementary Table**

**S14**). A total of five EPJ Tsw QTL, including EPJ-Tsw-A3-1, EPJ-Tsw-A3-2, EPJ-Tsw-A7-1, EPJ-Tsw-A10-1, and EPJ-Tsw-B3-1, were part of the consensus Tsw QTL detected in four biparental DH mapping populations, TD, VH, DE, and EPJ (Dhaka et al., 2017), providing broad application prospects to the 23 candidate genes harbored in these QTL. Eight DEGs identified as hub genes in coexpression modules highly correlated with an increase in seed size and also colocalized EPJ Tsw QTL including Bju.ESP, Bju.MKK4, Bju.AGO4, Bju.CYP72A1, Bju.FLA16, Bju.RGP4, Bju.HEBP1, and Bju.MES4. Some of these genes have not been reported earlier to directly regulate seed size and could be novel candidates for the improvement of the trait.

Direct orthologues of a few candidate genes colocalizing EPJ Tsw QTL have been reported to regulate seed size in *A. thaliana* and other crops; however, their precise roles in the regulation of seed size in *B. juncea* are yet to be established. Auxin response factor Bju.ARF2 (BjuVaA06g002542) was upregulated in EH-2 at stages S1–S3 and colocalized with EPJ-Tsw-A06-1 QTL. Overexpression of the ARF2 gene has been reported to decrease the seed size in *A. thaliana* by restricting cell proliferation and expansion of the integuments (Schruff et al., 2006). *A. thaliana* homolog of Bju.AGO10 (upregulated in PJ, colocalizing EPJ-Tsw-A1-1) has been shown to downregulate ARF2 expression during embryogenesis (Roodbarkelari et al., 2015), but its direct role in the regulation of seed size has not been elucidated. Homologs of two candidate genes identified in the study, ARF2 and SWEET15, have also been reported as Tsw candidates using analysis of putative selective sweeps in *B. juncea* (Kang et al., 2021). Altogether, these genes are promising candidates for the enhancement of the seed size in *B. juncea*; however, they need to be functionally validated to establish their precise roles in the regulation of seed size.

In conclusion, a comprehensive transcriptome profiling of early seed development in *B. juncea* lines EH-2 and PJ revealed possible mechanisms of seed size regulation. A total of 5,746 genes including 954 TFs were differentially expressed. Our results suggest that differential regulation of cell proliferation and expansion as a result of differential expression of genes regulating cell cycle, cell wall biogenesis/modification, solute/sugar transport, and hormone signaling result in observed phenotypic contrasts in seed size in EH-2 and PJ. The study reports candidate genes for differential regulation of seed size in *B. juncea* lines of east European (EH-2) and Indian (PJ) gene pools. The 40 DEGs colocalizing with the eight EPJ Tsw QTL were identified as the most promising candidates for breeding efforts targeted at an improvement of the seed size of the east European gene pool lines of *B. juncea*. The candidate genes identified in the study will be functionally validated in the future to decipher their precise roles in the determination of final seed size/weight and to leverage their potential for improvement of the seed size in *B. juncea*. Overall, the study provides crucial insights to understand the biology of early seed development in *B. juncea*, which we anticipate will aid the improvement of seed-associated traits in oilseed mustard.

## DATA AVAILABILITY STATEMENT

The obtained RNA sequencing data were deposited to the NCBI BioProject public database and are accessible under the accession number PRJNA771260. The BioSample accession numbers for PJ and EH-2 are SAMN22260750 and SAMN22261840, respectively. The GenBank assembly accession of the reference genome used in the study is GCA\_015484525.1. The datasets supporting the conclusions of this article are included within the article/**Supplementary Material**.

## AUTHOR CONTRIBUTIONS

SM performed the sampling and experiments. RT supervised the anatomy experiments. SM and KP performed the transcriptome data analysis. AP and DP conceived and supervised the overall study. SM along with KP drafted the manuscript. All authors reviewed the manuscript and approved the final draft of the manuscript.

## REFERENCES

- Adamski, N. M., Anastasiou, E., Eriksson, S., O'Neill, C. M., and Lenhard, M. (2009). Local Maternal Control of Seed Size by KLUH/CYP78A5-dependent Growth Signaling. *Proc. Natl. Acad. Sci.* 106 (47), 20115–20120. doi:10.1073/pnas.0907024106
- Agarwal, P., Kapoor, S., and Tyagi, A. K. (2011). Transcription Factors Regulating the Progression of Monocot and Dicot Seed Development. *BioEssays* 33 (3), 189–202. doi:10.1002/bies.201000107
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30 (15), 2114–2120. doi:10.1093/bioinformatics/btu170
- Borrás, L., Slafer, G. A., and Otegui, M. E. (2004). Seed Dry Weight Response to Source-Sink Manipulations in Wheat, maize and Soybean: A Quantitative Reappraisal. *Field Crops Res.* 86 (2–3), 131–146. doi:10.1016/j.fcr.2003.08.002
- Chandna, R., Augustine, R., and Bisht, N. C. (2012). Evaluation of Candidate Reference Genes for Gene Expression Normalization in Brassica Juncea Using Real Time Quantitative RT-PCR. *PLoS ONE* 7 (5), e36918. doi:10.1371/journal.pone.0036918
- Chaudhury, A. M., and Berger, F. (2001). Maternal Control of Seed Development. *Semin. Cel Dev. Biol.* 12 (5), 381–386. doi:10.1006/scdb.2001.0267
- Chen, L.-Q., Lin, I. W., Qu, X.-Q., Sosso, D., McFarlane, H. E., Londoño, A., et al. (2015). A cascade of Sequentially Expressed Sucrose Transporters in the Seed Coat and Endosperm Provides Nutrition for the Arabidopsis Embryo. *Plant Cell* 27 (3), 607–619. doi:10.1105/tpc.114.134585
- Chevalier, F., Perazza, D., Laporte, F., Le He'nanff, G., Hornitschek, P., Bonneville, J.-M., et al. (2008). GeBP and GeBP-like Proteins Are Noncanonical Leucine-Zipper Transcription Factors that Regulate Cytokinin Response in Arabidopsis. *Plant Physiol.* 146 (3), 1142–1154. doi:10.1104/pp.107.110270
- Clark, N. M., Buckner, E., Fisher, A. P., Nelson, E. C., Nguyen, T. T., Simmons, A. R., et al. (2019). Stem-cell-ubiquitous Genes Spatiotemporally Coordinate Division through Regulation of Stem-cell-specific Gene Networks. *Nat. Commun.* 10 (1), 5574. doi:10.1038/s41467-019-13132-2
- Dhaka, N., Rout, K., Yadava, S. K., Sodhi, Y. S., Gupta, V., Pental, D., et al. (2017). Genetic Dissection of Seed Weight by QTL Analysis and Detection of Allelic Variation in Indian and East European Gene Pool Lines of Brassica Juncea. *Theor. Appl. Genet.* 130 (2), 293–307. doi:10.1007/s00122-016-2811-2
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* 29 (1), 15–21. doi:10.1093/bioinformatics/bts635
- Du, J., Wang, S., He, C., Zhou, B., Ruan, Y.-L., and Shou, H. (2017). Identification of Regulatory Networks and Hub Genes Controlling Soybean Seed Set and Size

## FUNDING

This research was supported by the Department of Biotechnology, Government of India, through the award of project DBTUDSC Partnership Center on Genetic Manipulation of Brassicas (BT/01/NDDDB/UDSC/2016).

## ACKNOWLEDGMENTS

SM acknowledges a research fellowship from the University Grants Commission (UGC) of India.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.814486/full#supplementary-material>

Using RNA Sequencing Analysis. *Exbotj* 68 (8), erw460–1972. doi:10.1093/jxb/erw460

- Dvořák Tomaštková, E., Rutten, T., Dvořák, P., Tugai, A., Ptošková, K., Petrovská, B., et al. (2020). Functional Divergence of Microtubule-Associated TPX2 Family Members in *Arabidopsis thaliana*. *Ijms* 21 (6), 2183. doi:10.3390/ijms21062183
- Feder, N., and O'Brien, T. P. (1968). PLANT MICROTUBULE: SOME PRINCIPLES AND NEW METHODS. *Am. J. Bot.* 55 (1), 123–142. doi:10.1002/j.1537-2197.1968.tb06952.x
- Geng, X., Dong, N., Wang, Y., Li, G., Wang, L., Guo, X., et al. (2018). RNA-seq Transcriptome Analysis of the Immature Seeds of Two Brassica Napus Lines with Extremely Different Thousand-Seed Weight to Identify the Candidate Genes Related to Seed Weight. *PLoS ONE* 13 (1), e0191297. doi:10.1371/JOURNAL.PONE.0191297
- Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., et al. (2004). Bioconductor: Open Software Development for Computational Biology and Bioinformatics. *Genome Biol.* 5 (10), R80. doi:10.1186/gb-2004-5-10-r80
- Guilfoyle, T. J., and Hagen, G. (2007). Auxin Response Factors. *Curr. Opin. Plant Biol.* 10 (5), 453–460. doi:10.1016/j.PBI.2007.08.014
- Hu, S., Yang, H., Gao, H., Yan, J., and Xie, D. (2021). Control of Seed Size by Jasmonate. *Sci. China Life Sci.* 64 (8), 1215–1226. doi:10.1007/s11427-020-1899-8
- Kang, L., Qian, L., Zheng, M., Chen, L., Chen, H., Yang, L., et al. (2021). Genomic Insights into the Origin, Domestication and Diversification of Brassica Juncea. *Nat. Genet.* 53 (9), 1392–1402. doi:10.1038/s41588-021-00922-y
- Karnovsky, M. J. (1965). A Formaldehyde-Glutaraldehyde Fixative of High Osmolality for Use in Electron Microscopy. *J. Cel Biol.* 27, 137A–138A.
- Kasili, R., Walker, J. D., Simmons, L. A., Zhou, J., De Veylder, L., and Larkin, J. C. (2010). SIAMESE Cooperates with the CDH1-like Protein CCS52A1 to Establish Endoreplication in *Arabidopsis thaliana* Trichomes. *Genetics* 185 (1), 257–268. doi:10.1534/genetics.109.113274
- Kumar, R., and Bisht, N. C. (2020). Heterotrimeric Ga Subunit Regulates Plant Architecture, Organ Size and Seed Weight in the Oilseed Brassica Juncea. *Plant Mol. Biol.* 104 (4–5), 549–560. doi:10.1007/s11103-020-01060-5
- Langfelder, P., and Horvath, S. (2008). WGCNA: An R Package for Weighted Correlation Network Analysis. *BMC Bioinformatics* 9. doi:10.1186/1471-2105-9-559
- Le, B. H., Cheng, C., Bui, A. Q., Wagmaster, J. A., Henry, K. F., Pelletier, J., et al. (2010). Global Analysis of Gene Activity during Arabidopsis Seed Development and Identification of Seed-specific Transcription Factors. *Proc. Natl. Acad. Sci.* 107 (18), 8063–8070. doi:10.1073/pnas.1003530107

- Li, N., and Li, Y. (2016). Signaling Pathways of Seed Size Control in Plants. *Curr. Opin. Plant Biol.* 33, 23–32. doi:10.1016/j.pbi.2016.05.008
- Li, N., Xu, R., and Li, Y. (2019). Molecular Networks of Seed Size Control in Plants. *Annu. Rev. Plant Biol.* 70 (1), 435–463. doi:10.1146/annurev-arplant-050718-095851
- Linkies, A., Graeber, K., Knight, C., and Leubner-Metzger, G. (2010). The Evolution of Seeds. *New Phytol.* 186 (4), 817–831. doi:10.1111/j.1469-8137.2010.03249.x
- Liu, K., Li, Y., Chen, X., Li, L., Liu, K., Zhao, H., et al. (2018). ERF72 Interacts with ARF6 and BZR1 to Regulate Hypocotyl Elongation in Arabidopsis. *J. Exp. Bot.* 69 (16), 3933–3947. doi:10.1093/jxb/ery220
- Liu, Z., Li, N., Zhang, Y., and Li, Y. (2020). Transcriptional Repression of GIF1 by the KIX-PPD-MYC Repressor Complex Controls Seed Size in Arabidopsis. *Nat. Commun.* 11 (1), 1846. doi:10.1038/s41467-020-15603-3
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2<sup>-ΔΔCT</sup> Method. *Methods* 25 (4), 402–408. doi:10.1006/meth.2001.1262
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8
- Lu, X., Li, Q.-T., Xiong, Q., Li, W., Bi, Y.-D., Lai, Y.-C., et al. (2016). The Transcriptomic Signature of Developing Soybean Seeds Reveals the Genetic Basis of Seed Trait Adaptation during Domestication. *Plant J.* 86 (6), 530–544. doi:10.1111/tpj.13181
- Martín-Trillo, M., and Cubas, P. (2010). TCP Genes: a Family Snapshot Ten Years later Trends in Plant Science. *Trends Plant Sci.* 15 (Issue 1), 31–39. doi:10.1016/j.tplants.2009.11.003
- Mathew, I. E., Das, S., Mahto, A., and Agarwal, P. (2016). Three rice NAC Transcription Factors Heteromerize and Are Associated with Seed Size. *Front. Plant Sci.* 7 (November), 1638. doi:10.3389/fpls.2016.01638
- Meng, L.-S., Wang, Y.-B., Loake, G. J., and Jiang, J.-H. (2016). Seed Embryo Development Is Regulated via an AN3-MINI3 Gene cascade. *Front. Plant Sci.* 7 (Nov 2016), 1645. doi:10.3389/fpls.2016.01645
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAA5: An Automatic Genome Annotation and Pathway Reconstruction Server. *Nucleic Acids Res.* 35 (Suppl. 2), W182–W185. doi:10.1093/nar/gkm321
- Niu, Y., Wu, L., Li, Y., Huang, H., Qian, M., Sun, W., et al. (2020). Deciphering the Transcriptional Regulatory Networks that Control Size, Color, and Oil Content in Brassica Rapa Seeds. *Biotechnol. Biofuels* 13 (1). doi:10.1186/s13068-020-01728-6
- O'Brien, T. P., and McCully, M. E. (1981). *The Study of Plant Structure Principles and Selected Methods*. Melbourne: Termarcaphi Pty. Ltd.
- Paritosh, K., Yadava, S. K., Singh, P., Bhayana, L., Mukhopadhyay, A., Gupta, V., et al. (2021). A Chromosome-scale Assembly of Allotetraploid Brassica Juncea (AABB) Elucidates Comparative Architecture of the A and B Genomes. *Plant Biotechnol. J.* 19 (3), 602–614. doi:10.1111/pbi.13492
- Pradhan, A. K., Sodhi, Y. S., Mukhopadhyay, A., and Pental, D. (1993). Heterosis Breeding in Indian Mustard (Brassica Juncea L. Czern & Coss): Analysis of Component Characters Contributing to Heterosis for Yield. *Euphytica* 69 (3), 219–229. doi:10.1007/BF00022368
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ramchiary, N., Padmaja, K. L., Sharma, S., Gupta, V., Sodhi, Y. S., Mukhopadhyay, A., et al. (2007). Mapping of Yield Influencing QTL in Brassica Juncea: Implications for Breeding of a Major Oilseed Crop of Dryland Areas. *Theor. Appl. Genet.* 115 (6), 807–817. doi:10.1007/s00122-007-0610-5
- Ren, D., Wang, X., Yang, M., Yang, L., He, G., and Deng, X. W. (2019). A New Regulator of Seed Size Control in Arabidopsis Identified by a Genome-wide Association Study. *New Phytol.* 222 (2), 895–906. doi:10.1111/nph.15642
- Ren, Y., Zhu, J., Wang, Y., Ma, S., Ye, G., and Hua, S. (2017). Seed Weight and Oil Content Response to branch Position Reveals the Importance of Carbohydrate and Nitrogen Allocation Among Branches in Canola (Brassica Napus L.) Genotypes. *J. Appl. Bot. Food Qual.* 90 (December), 359–367. doi:10.5073/JABFQ.2017.090.045
- Roodbarkelari, F., Du, F., Truernit, E., and Laux, T. (2015). ZLL/AGO10 Maintains Shoot Meristem Stem Cells during Arabidopsis Embryogenesis by Down-Regulating ARF2-Mediated Auxin Response. *BMC Biol.* 13 (1), 1–12. doi:10.1186/s12915-015-0180-y
- Rose, J. K. C., Braam, J., Fry, S. C., and Nishitani, K. (2002). The XTH Family of Enzymes Involved in Xyloglucan Endotransglucosylation and Endohydrolysis: Current Perspectives and a New Unifying Nomenclature. *Plant Cell Physiol.* 43 (12), 1421–1435. doi:10.1093/pcp/pcf171
- Ruan, Y.-L., Patrick, J. W., Bouzayen, M., Osorio, S., and Fernie, A. R. (2012). Molecular Regulation of Seed and Fruit set Trends in Plant Science. *Trends Plant Sci.* 17 (11), 656–665. doi:10.1016/j.tplants.2012.06.005
- Sakai, H., Honma, T., Aoyama, T., Sato, S., Kato, T., Tabata, S., et al. (2001). ARR1, a Transcription Factor for Genes Immediately Responsive to Cytokinins. *Science* 294 (5546), 1519–1521. doi:10.1126/science.1065201
- Savadi, S. (2018). Molecular Regulation of Seed Development and Strategies for Engineering Seed Size in Crop Plants. *Plant Growth Regul.* 84 (3), 401–422. doi:10.1007/s10725-017-0355-3
- Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 Years of Image Analysis. *Nat. Methods* 9 (7), 671–675. doi:10.1038/nmeth.2089
- Schruff, M. C., Spielman, M., Tiwari, S., Adams, S., Fenby, N., and Scott, R. J. (2006). The AUXIN RESPONSE FACTOR 2 Gene of Arabidopsis Links Auxin Signalling, Cell Division, and the Size of Seeds and Other Organs. *Development* 133 (2), 251–261. doi:10.1242/dev.02194
- Schwacke, R., Ponce-Soto, G. Y., Krause, K., Bolger, A. M., Arsova, B., Hallab, A., et al. (2019). MapMan4: A Refined Protein Classification and Annotation Framework Applicable to Multi-Omics Data Analysis. *Mol. Plant* 12 (6), 879–892. doi:10.1016/j.molp.2019.01.003
- Sekhon, R. S., Hirsch, C. N., Childs, K. L., Breitzman, M. W., Kell, P., Duvick, S., et al. (2014). Phenotypic and Transcriptional Analysis of Divergently Selected Maize Populations Reveals the Role of Developmental Timing in Seed Size Determination. *Plant Physiol.* 165 (2), 658–669. doi:10.1104/pp.114.235424
- Singh, A. K., Fu, D.-Q., El-Habbak, M., Navarre, D., Ghabrial, S., and Kachroo, A. (2011). Silencing Genes Encoding omega-3 Fatty Acid Desaturase Alters Seed Size and Accumulation of Bean Pod Mottle Virus in Soybean. *Mpmi* 24 (4), 506–515. doi:10.1094/MPMI-09-10-0201
- Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. (2011). Cytoscape 2.8: New Features for Data Integration and Network Visualization. *Bioinformatics* 27 (3), 431–432. doi:10.1093/bioinformatics/btq675
- Sodhi, Y. S., Chandra, A., Verma, J. K., Arumugam, N., Mukhopadhyay, A., Gupta, V., et al. (2006). A New Cytoplasmic Male Sterility System for Hybrid Seed Production in Indian Oilseed Mustard Brassica Juncea. *Theor. Appl. Genet.* 114 (1), 93–99. doi:10.1007/s00122-006-0413-0
- Sra, S. K., Sharma, M., Kaur, G., Sharma, S., Akhtar, J., Sharma, A., et al. (2019). Evolutionary Aspects of Direct or Indirect Selection for Seed Size and Seed Metabolites in Brassica Juncea and Diploid Progenitor Species. *Mol. Biol. Rep.* 46 (1), 1227–1238. doi:10.1007/s11033-019-04591-3
- Sun, X., Shanharaj, D., Kang, X., and Ni, M. (2010). Transcriptional and Hormonal Signaling Control of Arabidopsis Seed development Current Opinion in Plant Biology. *Curr. Opin. Plant Biol.* 13 (5), 611–620. doi:10.1016/j.pbi.2010.08.009
- Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). Revigo Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* 6 (7), e21800. doi:10.1371/journal.pone.0021800
- Walton, L. J., Kurepin, L. V., Yeung, E. C., Shah, S., Emery, R. J. N., Reid, D. M., et al. (2012). Ethylene Involvement in Silique and Seed Development of Canola, Brassica Napus L. *Plant Physiol. Biochem.* 58, 142–150. doi:10.1016/j.plaphy.2012.06.016
- Wang, A., Garcia, D., Zhang, H., Feng, K., Chaudhury, A., Berger, F., et al. (2010). The VQ Motif Protein IKU1 Regulates Endosperm Growth and Seed Size in Arabidopsis. *Plant J.* 63 (4), 670–679. doi:10.1111/j.1365-313X.2010.04271.x
- Wang, C., and Liu, Z. (2006). Arabidopsis Ribonucleotide Reductases Are Critical for Cell Cycle Progression, DNA Damage Repair, and Plant Development. *Plant Cell* 18 (2), 350–365. doi:10.1105/tpc.105.037044
- Wickham, H. (2011). ggplot2. *Wires Comp. Stat.* 3 (2), 180–185. doi:10.1002/wics.147

- Yang, C., Ma, B., He, S.-J., Xiong, Q., Duan, K.-X., Yin, C.-C., et al. (2015). Maohuzi6/Ethylene Insensitive3-Like1 and Ethylene Insensitive3-Like2 Regulate Ethylene Response of Roots and Coleoptiles and Negatively Affect Salt Tolerance in rice. *Plant Physiol.* 169 (1), 148–165. doi:10.1104/pp.15.00353
- Yang, Z., Gu, S., Wang, X., Li, W., Tang, Z., and Xu, C. (2008). Molecular Evolution of the CPP-like Gene Family in Plants: Insights from Comparative Genomics of Arabidopsis and rice. *J. Mol. Evol.* 67 (3), 266–277. doi:10.1007/s00239-008-9143-z
- Zhang, X.-R., Qin, Z., Zhang, X., and Hu, Y. (2015). Arabidopsis SMALL ORGAN 4, a Homolog of Yeast NOP53, Regulates Cell Proliferation Rate during Organ Growth. *J. Integr. Plant Biol.* 57 (10), 810–818. doi:10.1111/jipb.12424
- Zhou, Y., Zhang, X., Kang, X., Zhao, X., Zhang, X., and Ni, M. (2009). Short Hypocotyl Under Blue1 Associates with MINISEED3 and HAIKU2 Promoters *In Vivo* to Regulate Arabidopsis Seed Development. *Plant Cell* 21 (1), 106–117. doi:10.1105/tpc.108.064972

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Mathur, Paritosh, Tandon, Pental and Pradhan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.