LncRAnalyzer: Uncovering Long Non-Coding RNAs and Exploring Gene Co-expression Patterns in Sorghum Genomics

Shinde Nikhil[1], Habeeb Shaikh Mohideen[2], Raja Natesan Sella[1*]

1. Membrane Protein Interaction Lab, Department of Genetic Engineering, SRM Institute of Science and Technology, Chengalpattu District, Tamil Nadu 603203 (India). Ph: +91 9940194847, E-Mail: raja.clri@gmail.com
2. Entomoinformatics Lab, Department of Genetic Engineering, SRM Institute of Science and Technology, Chengalpattu District, Tamil Nadu 603203 (India). Ph: +91 9940051100, E-Mail: habeebm@srmist.edu.in

*corresponding author

Abstract:

Sorghum (*Sorghum bicolor L. Moench*) is a versatile crop with significant phenotypic and genetic diversity. Despite the availability of multiple re-sequenced sorghum genomes, non-coding regions remain underexplored. Long non-coding RNAs (lncRNAs), a major class of non-coding RNAs, with low expression levels, complex expression patterns, and can be identified using RNA-seq. However, distinguishing lncRNAs from protein-coding genes is challenging due to their low abundance and tissue-specific expressions. This study developed an automated pipeline for identifying lncRNAs and Novel Protein Coding Transcripts (NPCTs) using RNA-seq datasets. Using a dual sorghum reference genome scheme, we identified 8770 lncRNAs in BTX642 (33.35% cultivar-specific) and 8869 lncRNAs in RTX430 (30.72% cultivar-specific). Some lncRNAs were linked to pre- and post-flowering drought tolerance in RTX430 and BTX642, respectively. The NPCTs encoded elements such as RE-1, TNT, Zinc finger (zf) protein, FAR1 related, and Xylan glucosyltransferase suggesting their involvement in drought-specific phenotype development. Target predictions for lncRNAs revealed both cis- and trans-regulated genes were associated with peroxidase, glutathione S-transferase, and MYB TFs during drought, emphasizing their role in drought tolerance. Differentially expressed lncRNAs in leaf and root tissues (p-value < 0.05, $|\log2fold| > 2$), revealed that the most abundant lncRNAs during drought regulate TFs such as C2H2-GATA, DUF domain, and WD-40. Downstream gene co-expression analysis showed time point-specific lncRNA interactions with TFs, highlighting their roles in drought response in sorghum. Our developed pipeline LncRAnalyzer identified lncRNAs and NPCTs, and integrated them into gene co-expression networks provided deeper insights into drought tolerance in sorghum.

Keywords:

LncRAnalyzer, RNA-seq, long non-coding RNA, Sorghum genomics, gene co-expression network

Introduction:

Sorghum *(Sorghum bicolor L. Moench*) is an important crop for food security, the staple food for more than 500 million people across the globe (FAO, 2011). Sorghum comprises 34,129 genes, exhibiting many Structural Variants (SVs) within its core gene pool [1,2]. A pan-genome analysis reported 44079 gene families in sorghum with 64% of dispensable gene families, suggesting that sorghum is more genetically diverse than maize [3]. Previous analysis reports SVs especially deletions were major players in local adaptations and plant's response to stresses such as drought [4]. This includes deletions in MYB/SANT-like DNA-binding domain protein which was linked to drought tolerance in sugarcane [5]. Past studies reported overexpression of several genes such as molybdenum cofactor biosynthesis protein [6], expansin [7], bZIP [8], anthocyanidin 5,3-O-glucosyltransferase [9], penta-tricopeptide repeat family [10], etc. enhance drought tolerance in plants. However, the regulation of these genes is poorly understood. The loss of function or overexpression of a few genes is insufficient to address plant responses to environmental stimuli and the genetics underlying stress tolerance. The previous analysis reported long non-coding RNAs (lncRNAs) regulate the expression of several functional genes and confer stress tolerance to plants [11,12]. Thus, it emphasizes its importance in the growth and development of plants.

Briefly, the lncRNA is the major class of non-coding RNAs that impart plants with a wide range of characteristics [13]. They are >200 nucleotides long, do not encode proteins, and affect gene expression at several levels, including epigenetic, transcriptional, post-transcriptional, translational, and post-translational mechanisms [14]. Most lncRNAs are tissue and condition-specific and play an important role in physiology and development [15]. Although, they occupy the major portion of non-coding RNAs, their low expression levels are crucial for ensuring specificity to their regulated targets [16]. In addition to exhibiting low expression levels, lncRNAs have low sequence conservation and were initially regarded as transcriptional noise [16,17]. LncRNAs were categorized into five categories based on their position in the genome: sense, antisense, bidirectional, intronic, and intergenic lncRNAs [18]. They could be polyadenylated or non-polyadenylated transcripts with low coding potentials [19]. Past studies reported a novel group of novel non-polyadenylated transcripts with 50-300 nt length in Arabidopsis and rice [20,21]. These transcripts were named as intermediate-sized ncRNAs (im-ncRNAs) [20]. Most of the intergenic regions in Arabidopsis,

rice, and maize transcribe lncRNAs that become polyadenylated and are thought to be transcribed by RNA polymerase II and some by Pol IV or Pol V [22–24]. This also includes several unstable lncRNAs. In widely accepted view most stable lncRNAs were transcribed by RNA polymerase II [25]. As per the understanding, lncRNAs are polyadenylated, capped [26], transcribed by RNA polymerase II, and stable which can be detected using RNA-seq datasets [27].

The present investigation aimed to develop an automated pipeline named LncRAnalyzer to report lncRNAs associated with drought response in sorghum using RNA-seq datasets. We implemented this workflow on 389 RNA-seq accessions of pre- and post-flowering drought-tolerant sorghum genotypes RTX430 and BTX642 to report known and cultivar-specific lncRNAs. The downstream gene co-expression analysis reported co-expressed TFs and lncRNAs that help this crop to sustain under pre- and post-flowering drought.

Methods:

LncRAnalyzer Overview

The LncRAnalyzer is an automated pipeline for identifying long non-coding RNAs (lncRNAs) and Novel Protein coding Transcripts (NPCTs) using RNA-seq datasets. This protocol takes inputs as raw *.fastq (*.gz), rRNA sequences (*.fasta), reference genome (*.fasta), and annotations (*.gtf). Initially, quality trimming and removal of rRNA reads will be performed followed by alignment, reference guided assembly, size selection (>200 nt), annotation comparison, and selection of transcripts belonging to category intronic (i), exonic but opposite orientation (x), multi exonic junctions (j), other gene overlap (o) and intergenic (u). Once the class-codes selected transcripts are obtained, the lncRNAs and NPCTs will be identified based on their coding potentials or features using CPC2 [28], RNAsamba [29], CPAT [30], PLEK [31], and LGC [32]. In addition to this, the use of FEELnc [33] is employed to report intergenic and mRNA-spliced lncRNAs. The putative transcripts that show similarity with protein families will be removed using PfamScan [34]. Finally, true lncRNAs and true NPCTs are obtained by taking the intersection of the techniques mentioned above in the Venn diagram representation. If additional information about non-coding genomes will provided in the form of small nucleolar (snoRNAs), small nuclear (snRNAs), micro RNAs (miRNAs), and LiftOver files [https://genome.ucsc.edu/cgi-bin/hgLiftOver], conservation analysis between species will be also performed using Slncky [35]. The entire workflow of lncRNAs and NPCTs identification is automated with Bipipe [36] and easily implementable by updating/creating conda environments **(Figure 1(a)).**

Data Sources

The raw RNA-seq accessions of sorghum pre-flowering drought tolerant genotype RTX430 and post-flowering drought tolerant genotype BTX642 were downloaded from NCBI Sequence Read Archive (SRA) from past studies [37] **(Supplemental Data S1)**. The sorghum reference genome (v3.1), RTX430 cultivar reference (v1.1), and BTX642 cultivar reference (v1.1) were downloaded from Phytozome [https://phytozome-next.jgi.doe.gov] to report known and cultivar-specific lncRNAs associated with drought tolerance.

Workflow Implementation

1. Data pre-processing

The analysis workflow includes quality trimming of raw *.fastq reads using Fastp [38]. Aligning quality trimmed *.fastq reads to plant's rRNA sequences retrieved from RNAcental [https://rnacentral.org] using HISAT2 [39] followed by Sequence Alignment Map (SAM) to Binary Alignment Map (BAM) conversions and BAM sorting with SAMtools [40]. Finally, un-mapped reads to rRNA sequences were obtained via the *bamToFastq* method using BEDTools [41] and compressed using *gzip*.

2. Alignment, reference guided assembly, and class-codes selection

The un-mapped reads to rRNA sequences were aligned first to sorghum reference genomes using HISAT2 allowing a maximum of 2 mismatches. Reference-guided transcripts from individual *.bam files were assembled in conservative mode with transcript length >200, default splice junctions, and minimum locus gap separation using StringTie [42]. The transcripts *.gtfs from StringTie were merged into a single *.gtf file using the *stringtie --merge* option with minimum mapped reads >3 and TPM values >1. Further, uni-exonic transcripts were discarded from the merged *.gtf file using GFFRead [43]. Annotation comparison was performed using GFFCompare [43], and transcripts belonging to class codes "i", "o", "u", "x" and "j" were obtained with GFFRead.

3. Identification of lncRNAs and NPCTs

The coding potential, features, and sequence similarity-based categorization of lncRNAs/NPCTs transcripts was performed using CPC2, CPAT, RNAsamba, LGC, PLEK, and PfamScan. The CPC2 used sequence information alone for coding potential-based characterization of putative lncRNAs/NPCTs transcripts. Subsequently, CPAT was employed to extract coding potentials by constructing a hexamer table and logit model utilizing known mRNA sequences and putative lncRNAs/NPCTs transcripts. RNAsamba is a neural network-based approach that uses known mRNAs and putative lncRNAs/NPCTs transcripts to train datasets

into HDF5 files. Later, *rnasamba classify* method was used to estimate their coding potentials. LGC is a novel approach that uses sequence information alone to differentiate lncRNAs from protein-coding genes based on ORF length and GC content feature relationships. Next, the PLEK uses an improved k-mer scheme to train Support Vector Machine (SVM) models using putative lncRNAs/NPCTs and known mRNA sequences and classify them into coding or non-coding categories. In addition, PfamScan was also performed to discard transcripts that showed sequence similarities with known protein families. This will omit transcripts that could be translated into protein domains and increase the probability of lncRNA identification.

We utilized FEELnc, a random forest-based approach to report mRNA spliced and intergenic lncRNAs. First, known mRNA transcripts were filtered from lncRNAs using *FEELnc_filter.pl*. Next, coding potential estimation was performed with the reference genome (*.fasta), annotations (*.gtf), and merged *.gtf using *FEELnc_codpot.pl*. Here, *FEELnc_codpot.pl* uses two approaches shuffle: which identifies mRNA spliced lncRNAs, and intergenic: which identifies intergenic lncRNAs. Finally, *FEELnc_classifier.pl* was used to assign the classes i.e. genic/intergenic, divergent/convergent, sense/antisense, etc. to the reported lncRNA transcripts. The present workflow integrated the Slncky tool to identify high-confidence lncRNAs and perform conservation analysis between species. The conservation analysis requires LiftOver files and non-coding gene information for query and target species. We built LiftOver files by aligning sorghum (BTX623) and maize (B73) genomes using LAST [44] followed by building chain files using ucsc-axtChain tool [45]. In summary, we employed eight approaches namely CPC2, CPAT, LGC, PfamScan, RNAsamba, PLEK, Slncky, and FEELnc in the present workflow of which seven have been used lncRNAs identification and five for NPCTs identification. To report cultivar-specific lncRNAs, the unmapped reads to the sorghum reference genome were aligned to the individual genomes of BTX642 and RTX430 cultivars, and the LncRAnalyzer procedure was repeated **(Figure 1(b)).**

4. Summary

The summary report enumerates the lncRNA and NPCTs identified by each method in Tab-Separated Values (*.TSV) format. The intersection of all core methods was taken to obtain true lncRNAs and NPCTs which were written as *.gtf (lncRNAs) and *.fa (NPCTs) respectively. The Venn diagram describing the intersection of all methods was prepared using the Venn [46] R package.

5. Performance

The lncRNAs and NPCTs identified based on their coding potentials with the above approaches, namely FEELnc, RNAsamba, CPAT, LGC, and CPC2 were subjected to supervised machine learning (independent of the pipeline) using Python-based Sci-Kit-Learn library [47]. To test the superiority of the approach, datasets with predicted labels and coding potentials for each method were used to train Logistic Regression (LR), Support Vector Machines (SVM), Decision Tree (DT), Random Forest (RF), and, K-Nearest Neighbors (KNN) models. Subsequently, the performance of these models (testing) was evaluated by comparing their predictions to the actual labels i.e. coding potentials. This allows the determination of differences between the true and predicted labels. The ROC curve, a visual representation of the performance of all the approaches used in LncRAnalyzer was prepared using a custom R script. The present pipeline also tests the performance of various approaches used for lncRNA and NPCTs identification using R, as previously described for superTranscript [48], which outputs an ROC curve.

Downstream Analysis

1. Lnc RNAs Conservation analysis

The lncRNAs conservation analysis between sorghum (BTX623) and maize (B73) with LiftOver and non-coding gene information was performed using Slncky. Additionally, the conserved lncRNAs between cultivars were obtained by comparing the genomic coordinates of lncRNAs using GFFCompare. The class code "=" transcripts were conserved because they have exact exonic overlaps or intron chain matches with reference. Further, known lncRNAs for sorghum were obtained from PLncDB [49] and significant hits (e-value $< 10^{-5}$) were considered as conserved lncRNAs.

2. Lnc RNAs Target Predictions

LncRNAs target prediction was done using the methodology previously described for the maize [50]. Briefly, cis-targets for lncRNAs were identified based on the genomic location of Protein Coding Genes (PCG) 100 kb upstream and 100 kb downstream of lncRNAs. Further, Protein Coding Transcripts (PCTs) that show $|PCC| > 0.6$ and p-value < 0.05 with lncRNAs based on their expressions were considered cis-targets for the same.

A triplex Forming Oligos (TFO) i. e. lncRNAs or mi-RNAs forms RNA: DNA triple helix and regulates the expression of nearby or distant genes via chromatin modifications in eukaryotes [51,52]. To report lncRNAs triple-helices with reference genome we used triplextor with parameters (minimum guanine rate 50%, score >15, and maximum error rate 5%) [53]; RNA:DNA triplex located 1000 bp upstream and 200 bp downstream of Transcription Start Site (TSS) were considered potential target regions for lncRNAs. The

PCTs that shared |PCC| > 0.9 and p-value < 0.01 with lncRNAs based on their expressions and reported with DNA:RNA triplex at promotor region were considered trans-targets for lncRNAs.

The cis and trans target genes were combined and KEGG orthology terms were searched for these genes using KAAS [54]. This step was essential to identify biochemical pathways for regulated target genes.

3. Functional annotation of NPCTs

NPCTs identified were directly looked for homologies with plant protein sequences from SwissProt using BLASTX. Next, NPCTs were predicted Open Reading Frames (ORFs) using TransDecoder [55] followed by homology searches with SwissProt and Pfam using BLASTP and HMMER respectively. Finally, a functional annotations table from homology searches was generated using Trinotate as described previously [56,57].

Time-ordered transcriptomics and gene co-expression studies

1. Transcription factors (TFs) prediction

The known mRNA sequences from reference genomes were obtained using GFFRead. TFs prediction was done using PlantTFCat [58], PlantTFDB [59], and iTAK [60] web servers. Result tables from above all three methods were combined in a single TSV file using R.

2. Differential gene expression and count normalization

The transcript level abundance estimates for mRNAs, known lncRNAs, and cultivar-specific lncRNAs were performed using featureCounts [61] from *.bam, mRNA, and lncRNA *.gtf files as input. Timepoint count normalization for mRNAs, lncRNAs, and cultivar-specific lncRNAs was performed using DESeq2 [62] with Likelihood Ratio Test (LRT) for individual tissue types and conditions. The Differentially Expressed (Pvalue < 0.05 and |log2fold| > 2|) LncRNAs (DELs) were retrieved and plotted on a volcano plot using an enhanced volcano plot [63] R package. Timepoint replication means from DESeq2 normalized counts were taken, TF/lncs, and all gene matrices in TSV format were prepared for gene co-expression studies using R.

3. Time-ordered gene co-expression and network visualization

The TO-GCN [64] package was used to report co-expressed genes in leaf and root under control and pre-flowering in RTX430 as well as control and post-flowering drought in BTX642. First, TF/lncs and all gene matrices obtained in the previous step were used to estimate Pearson's Correlation Coefficient (PCC) cutoffs. Next, a gene co-expression network with obtained cutoff values was constructed; this leads to eight co-expression types (C1+C2+, C1-C2+, C10C2+, …, C1-C2-) for conditions i.e. control vs pre-flowering drought and control vs. post-flowering drought, etc. Finally, time-ordered levels for each condition were

assigned by taking the top 10 highly expressed TF/lncs in the first time point as a seed for Breadth First Search (BFS). To report changes in the expression of TF/lncs with time, the TO-GCN levels for condition C1+C2+ were further mapped against actual time point expressions using z-score statistics, as described previously [64]. A heatmap showing changes in the expression of TF/lncs with time was prepared with the *geom_tile* function of the *ggplot2* R package. Finally, time-ordered gene co-expression networks were visualized Cytoscape [65].

Results:

Workflow Implementation

With LncRANalyzer, we have preprocessed 389 RNA-seq single-end libraries using Fastp to remove adapter contamination. Of 8565 million reads, 0.34% from both genotypes were reported with adapter contamination and removed. Next, unmapped reads to the plant rRNA sequences were also extracted, which reported an average of 0.80% of reads in both genotypes **(Supplemental Data S2).** Unmapped reads to rRNA sequences were aligned to the sorghum reference genome (v3.1) and showed an average alignment rate of 93.40% and 91.81% for BTX642 and RTX430 respectively. The rest of the reads were unmapped and were further aligned to the cultivar-specific reference genome. This reported average alignment rates of 47.06% and 44.98% for BTX642 and RTX430 respectively **(Supplemental Data S3)**. Reference-guided transcripts were assembled and merged using StringTie. These reported several new transcripts, of which 15.70% were uni-exonic and discarded **(Table 1 (a))**. An annotation comparison was done for the class code selection of transcripts. BTX642 reported a total of 42724 and 14060 transcripts belonging to class codes “i”, “o”, “u”, “x” and “j” on the sorghum reference genome and own cultivar-specific reference respectively. RTX430 has 437481 and 12574 transcripts that belong to class codes “i”, “o”, “u”, “x” and “j” on the sorghum reference genome and own cultivar-specific reference respectively **(Table 1(b)).** This reported 8351 and 9349 intergenic transcripts in BTX642 and RTX430 respectively **(Supplemental Figure S1(a)).** Further, class codes selected transcripts were evaluated for coding potentials, features, and sequence similarities with protein domains and classified into lncRNAs and NPCTs. CPAT reported the highest number of lncRNAs in both genotypes. A total of 14497 mRNA spliced and 7138 intergenic in BTX642. The number of mRNAs spliced and intergenic lncRNAs in RTX430 were 14896 and 7730 respectively **(Table 2).** This suggests that lncRNAs identified in RTX430 were slightly higher than in BTX642 and could be associated with pre-flowering drought tolerance. Most of the lncRNAs reported by FEELnc were spliced from mRNAs only. This highlights the role of mRNA splicing in gene regulation of sorghum. The

intersection of all methods is summarized as a Venn diagram which shows true lncRNAs and true NPCTs across datasets **(Supplemental Figures S1(b)-S1(j)).** The pipeline's output included common lncRNAs (*.gtf) and NPCTs (*.fa) predicted by all methods, along with Venn diagrams illustrating the number and distribution of lncRNAs and NPCTs (*.tiff). Among all the above coding potentials and feature-based lncRNA/NPCTs prediction methods, we found LGC and CPC2 give better estimates of lncRNAs in three trained machine learning models out of five **(Figure 1(c) and 1(d); Supplemental Data S4).**

Downstream Analysis

1. Lnc RNAs conservation analysis

The conserved lncRNAs were identified across the cultivars while looking for the exact match for the intron chain. Among the total lncRNAs identified on the sorghum reference genome in both genotypes, 1316 had matching intron chains and the rest were novel **(Figure 2(a))**. Results suggest that fewer lncRNAs existed between these two genotypes with exact intron chain match, and the rest could be spliced from the same gene or transcribed intergenic regions **(Supplemental Data S5, Sheet 1)**. This highlights the diversity in lncRNA splicing patterns and potential regulatory roles of intergenic lncRNAs between the genotypes. BTX642 reported a total of 5845 and 2925 lncRNAs on the sorghum reference genome and the cultivar reference genome, respectively. Among them, 2606 (44.58%) and 1074 (36.71%) lncRNAs corresponded to entries in PLncDB, while the rest were novel. Similarly, RTX430 reported 6144 and 2725 lncRNAs on the sorghum and cultivar reference genome, respectively. Of these, 3067 (49.91%) and 957 (35.11%) lncRNAs were found in PLncDB, with the rest being novel. The results suggest that most of the lncRNAs identified using a dual reference scheme were novel and could play a role in drought tolerance in both sorghum genotypes **(Figure 2(b); (Supplemental Data S5, Sheet 2)**.

2. Lnc RNAs Target Predictions

Cis-targets of lncRNAs were reported across tissue types in both genotypes. In BTX642, 1364 expressed lncRNAs in leaf tissues were found to be linked with 7162 target genes located 100 kb upstream/downstream. Whereas, 1502 expressed lncRNAs in roots were associated with 7973 target genes within the same distance. Regarding regulated transcripts, 10942 were from leaves and 11870 from roots. Similarly, RTX430 has 1342 expressed lncRNAs in the leaf tissues and was found to be linked with 8620 target genes located 100 kb upstream/downstream. Root tissues of RTX430 reported 1420 expressed lncRNAs with 8941 100 kb upstream/downstream target genes. This includes 12793 regulated transcripts in leaf and 12768 regulated transcripts in root tissues **(Table 3 (a)).** We found that these lncRNAs were

regulators of major TFs such as C2H2, MYB, AP2/ERF, B3, WD40, bHLH, NAC, and GRAS in leaf and roots of BTX642; suggesting that they may have a crucial role in stay-green phenotype. Similarly, some lncRNAs regulated CCHC(Zn), WD40, C2C2-GATA, bHLH, C2H2, WRKY, FAR1, MYB, and NAC TFs in the leaf and roots of RTX430; inferring that they may have been involved in pre-flowering drought tolerance (**Supplemental Data S6)**. We used a triplexator, a tool based on Hoogsteen base pairing with DNA (minimum score=15 and guanine rate >50%) to eliminate the ambiguity in trans-target prediction. The promoter region of a gene is defined as the sequence ranging from 1000bp upstream to 200bp downstream of the TSS [66]. Hence, we investigated the formation of DNA: RNA triple helices within this region. Triplexator reported 1198 lncRNAs with 14316 triplexes in BTX642 and 1092 lncRNAs with 11241 triplexes in RTX430 **(Table 3 (b); Supplemental Data S7).** The given gene was considered trans-target for lncRNAs when it met the given parameters (|PCC| > 0.9 and p-value < 0.01) and formed triplex 1000bp upstream and 200bp downstream of TSS. We reported 74 lncRNAs with 102 and 58 lncRNAs with 81 trans-targets in leaf and root tissues of BTX642 respectively. While RTX430 reported 105 lncRNAs with 151 and 82 lncRNAs with 104 trans-targets in leaf and root tissues respectively. These also include several TFs that were regulated by lncRNAs in the leaf and roots of both genotypes (**Table 3 (c); Supplemental Data S8)**. The results demonstrate that a greater number of lncRNAs form triplex with the promoter region of genes in BTX642 and regulation of these genes could be the reason for a stay-green phenotype. Whereas, RTX430 has a comparatively smaller number of genes with triplex-forming regions with lncRNAs. This suggests that triplex forming sites for lncRNAs were either missing or carry mutations due to which lncRNAs were not forming triplex. Additionally, we searched for KO terms related to lncRNA-regulated target genes using KAAS. The analysis revealed that peroxidase, glutathione S-transferase, and MYB transcription factor genes were more abundant in the leaf tissues of BTX642 under post-flowering drought conditions **(Supplemental Figure S2(a))**. In the root tissues of BTX642, there was a greater number of lncRNA-regulated genes linked to peroxidase, serine carboxyl peptidase, homeobox leucine zipper protein, histone H2A, and H2B during post-flowering drought **(Supplemental Figure S2(b))**. Likewise, in RTX430 leaf tissues, genes regulated by lncRNAs involved in ferredoxin, auxin-responsive factor (ARF), and centromeric protein E were noted under pre-flowering drought **(Supplemental Figure S2(c))**. In contrast, the root tissues exhibited an increase in regulated genes such as peroxidase, MYB transcription factor activity, speckle type POZ protein, nicotinamide transporter, glutathione S-transferase, and glyoxylate/hydroxypyruvate reductase under pre-flowering drought conditions **(Supplemental Figure**

S2(d)). The results validate that the increased activity of MYB TF, peroxidase, and glutathione S-transferase could be associated with drought tolerance in BTX642 and RTX430.

3. Functional annotation of NPCTs

Out of 54 NPCTs, 34 were annotated using Trinotate across leaf and root tissues in both cultivars and the rest were unknown protein-coding transcripts. These transcripts were not previously annotated on sorghum reference as well as cultivar reference genomes. NPCTs' shared location on the sorghum reference genome was considered as known or conserved however NPCTs exclusively located on the cultivar reference genome were cultivar-specific ones. Their functional annotation identified that they code for elements such as RE-1, FAR1, TNT, Zinc finger (zf) protein, receptor-like protein (RLP-6), Xylan glucosyltransferase, etc. in both genotypes with varied numbers. Compared to BTX642, RTX430 reported more NPCTs most of which were located on the sorghum reference genome and very few were on the cultivar reference genome. These include some TNT, Zinc finger proteins, putative disease resistant protein, putative ribonuclease H, lactate dehydrogenase, Cyclic nucleotide-gated ion channels, etc., and may be involved in pre-flowering drought tolerance. On the other hand, BTX642 has RLP-6, Xylan glucosyltransferase, LRR serine/threonine protein kinase, and Histone H2B encoding transcripts, inferring that they could be hidden players in drought response and stay-green phenotype **(Figure 3; Supplemental Data S9)**.

Time-ordered transcriptomics and gene co-expression studies

1. Transcription factors (TFs) prediction

Several transcription factors (TFs) were predicted across sorghum reference genome and cultivar reference genomes using iTAK, PlantTFCat, and PlantTFDB online servers. These numbers were 4410, 4217, and 4094 on the sorghum reference genome, BTX642, and RTX430 cultivar reference genomes respectively **(Supplemental Data S10).** PlantTFCat reported the highest number of predicted TF genes across all three reference genomes **(Table 4).** Compared to cultivar reference genomes, the sorghum reference genome has more predicted TFs, and reduced numbers of TF genes might be associated with drought tolerance in BTX642 and RTX430.

2. Differential gene expression and count normalization

Differential gene expression analysis for leaf and root tissues of BTX623 and RTX430 was performed using DESeq2. This reports 16958 and 2251 core genes that were differentially expressed (P-value < 0.05) in the leaf and roots of BTX642 **(Supplemental Figures S4(a) and S4(b)).** Since, they were expressed under all given conditions i.e. control, pre- and post-flowering drought, they are required for normal growth and

development of the BTX642. Further, we reported 1455 genes in leaf and 620 genes in roots that were exclusive to post-flowering drought in BTX642. These genes confer a stay-green phenotype to BTX642 under post-flowering drought. This includes 172 TFs along with 264 lncRNAs in leaf tissues and 39 TFs with 145 lncRNAs in root tissues **(Table 5)**. The top 50 highly expressed TFs and lncRNAs unique to the post-flowering drought were listed in the leaf and root tissues (**Supplemental Figures S4(c)-S4(f))**. We found that lncRNA BTX642.nov01G228.1 one of the highly expressed in leaf tissues of BTX642 and a potential hub for several genes. This lncRNA positively regulates genes such as serine/threonine protein phosphatase (PP2A), extended synaptotagmin (eSYT), glycosyltransferase (GTs), calcium-dependent protein kinases (CDPKs), Fructose-1,6-bisphosphatase (F1-6BP), etc. while negatively regulating genes such as NudC domain encoding (NudC), NAD(P)H-hydrate epimerase (NAD(P)H-HE), histone deacetylase (HDAC), T-complex protein 1 subunit theta (CCT8), DUF domain protein (DUF), OXA1/albino3-like protein (OXA1), etc. in the leaf under post-flowering drought in BTX642. The results suggest that BTX642.nov01G228.1 one of the players expressed under post-flowering drought in the leaf confers a stay-green phenotype to BTX642. Similarly, BTX642.nov07G21079.3 one of the lncRNA expressed exclusively under post-flowering drought with higher expressions in root tissues and positively regulating genes such as C2C2-GATA-12, SnoaL-like domain (SnoLD), Sentrin-specific protease (SENP), and Histone H1/H5. This infers that up-regulation of C2C2-GATA and other genes was involved in root growth and highlights its significance in the stay-green phenotype in BTX642 **(Supplemental Data S11, Sheets 2 and 3)**.

In RTX430, we reported 15970 and 22993 core genes that showed differential (P-value < 0.05) expression in leaf and root tissues under control, pre- and post-flowering drought conditions **(Supplemental Figures S4(g) and S4(h)).** This report genes involved in the normal growth and development of RTX430. Further, 2121 DEGs in leaf and 848 DEGs in roots were exclusive to pre-flowering drought. These genes were associated with pre-flowering drought tolerance and recovery. This includes 219 TFs in the leaf and 73 TFs in the roots. A total of 364 and 188 lncRNAs were exclusive to pre-flowering in the leaf and root tissues of RTX430 **(Table 5)**. The top 50 highly expressed TFs and lncRNAs exclusive to the pre-flowering drought condition in the leaf and root tissues of RTX430 were identified **(Supplemental Figures S(i)-S(l)).** These reported RTX430.nov09G22826.1 under the top 50 categories in leaf and positively regulates genes such as WD40-like-120, DUF domain protein (DUF), E3-ubiquitin ligase (E3 ligase), RNA binding protein (RBM25), etc. under pre-flowering drought. Whereas, RTX430.nov08G21197.1 was one of the members under the top 50 categories in roots and positively regulates genes such as C2H2-556, retinoblastoma-

related (RB-4), autophagy-related protein (ATG), proteasome component (ECM29), LYST-interacting protein (LIP5), etc. under pre-flowering drought. The results demonstrate that lncRNAs with abundant expression regulate genes such as WD40-like, DUF, C2H2, and RB-4 which along with other genes confers pre-flowering drought tolerance to RTX430 (**Supplemental Data S11, Sheets 4 and 5)**.

Additionally, we found 770 lncRNAs in the leaf and 688 lncRNAs in the roots of BTX642 that were expressed differentially (P-value <0.05) with |log2fold| > 2 under post-flowering drought. Interestingly, these numbers were ~4 fold lower when the crop was under control and pre-flowering drought conditions **(Figures 4(a) and 4(b))**. Similarly, RTX430 reported 291 lncRNAs in leaves and 210 lncRNAs in roots that showed differential expression with |log2fold| >2 under pre-flowering drought. These numbers were slightly lower than those of the leaf and root tissues under control conditions and approximately 2.5-3 times lower compared to post-flowering drought **(Figures 4(c) and 4(d))**. The results demonstrate that BTX642 displays a greater number of lncRNAs in the leaf and a reduced number in root tissues compared to RTX430 under post-flowering drought circumstances, with |log2fold| > 2. This observation is associated with a stay-green phenotype in BTX642 during post-flowering drought. While RTX430 reported comparatively more lncRNAs with |log2fold| > 2 in leaf and root tissues than BTX642 under pre-flowering drought. This infers that additional lncRNAs in leaf and root tissues were conferring pre-flowering drought tolerance to RTX430.

3. Time-ordered gene co-expression and network visualization

Time-ordered gene co-expression analysis compared control conditions with post-flowering drought for BTX642 and pre-flowering drought for RTX430. This gave co-expressed TFs/lncRNAs under different levels. These levels correspond to time points under which genes were co-expressed. In BTX642, a total of 373 TFs and 172 lncRNAs were co-expressed in 9 levels under control conditions in the leaf tissues. These numbers were 1095 and 374 for TFs and lncRNAs with 15 co-expression levels under post-flowering drought **(Figure 5(a))**. The results suggest that there were more co-expressed TFs and lncRNAs in the leaf of BTX642 under post-flowering drought than in control conditions. These include 54 lncRNAs and 155 TFs such as C2H2 (24), WRKY (18), AP2/ERF-ERF (16), MYB (10), bHLH (11), WD40-like (5), PHD (7), NAC (3), GRAS (3), etc. that were co-expressed in 9 and subsequent levels under post-flowering drought. These TFs and lncRNAs were associated with stay-green phenotype under post-flowering drought in BTX642 **(Figure 5(b))**. In root tissues of BTX642, we reported 1565 TFs and 464 lncRNAs that were co-expressed in 13 levels under control conditions. These numbers were 1037 and 340 for TFs and lncRNAs in

13 co-expression levels under post-flowering drought **(Figure 5(c))**. The results demonstrate that the number of co-expressed TFs and lncRNAs was reduced under post-flowering drought conditions. This inferred that drought tolerance may be caused by decreased co-expressed TFs and lncRNAs in roots during post-flowering drought **(Figure 5(d))**.

RTX430 reported a total of 507 TFs and 142 lncRNAs that were co-expressed in 9 levels in the leaf under control conditions. These numbers were 1106 and 408 for TFs and lncRNAs with 20 co-expression levels for pre-flowering drought plants **(Figure 5(e))**. These include 92 lncRNAs and 307 TFs such as C2H2 (43), WD40-like (28), PHD (19), bHLH (15), MYB (12), bZIP (11), B3 (9), C3H (8), CCHC(Zn) (8), SET (8), AP2/ERF-ERF (7), NAC (7), etc. that were co-expressed 9 and subsequent levels under pre-flowering drought **(Figure 5(f))**. In root tissues, 1614 TFs and 547 lncRNAs were co-expressed in 13 levels under control conditions. While in the pre-flowering drought, there were 1861 TFs and 584 lncRNAs in 11 co-expression levels **(Figure 5(g))**. This suggests that, in comparison to control roots, there were more co-expressed TFs and lncRNAs during the pre-flowering drought (**Figure 5(h))**. This includes TFs (control/pre-flowering drought) such as C2H2 (250/315), WD40-like (112/113), bHLH (76/79), PHD (69/90), AP2/ERF-ERF (60/67), bZIP (59/63), MYB (54/68), WRKY (50/72), NAC (48/61), CCHC(Zn) (37/53), etc. The results validate that more numbers of co-expressed TFs and lncRNAs in roots were associated with pre-flowering drought tolerance in RTX430 **(Supplemental Data S12)**.

Discussion:

The present analysis developed an automated pipeline for lncRNAs and NPCTs identification using RNA-seq datasets through the use of eight different methods namely CPC2, CPAT, RNAsamba, LGC, PLEK, PfamScan, Slncky, and FEELnc. A similar approach was used before in maize [50]. Of eight, seven approaches namely CPC2, CPAT, RNAsamba, PfamScan, Slncky, LGC, and FEELnc were tested on sorghum RNA-seq datasets because of their availability, superiority, and easy installation via updating or creating a conda environment. Our four strategies namely CPC2, CPAT, RNAsamba, and FEELnc rely on machine-learning methods, essential for identifying and classifying lncRNAs from RNA-seq datasets. The CPC2 used sequence information alone to classify sequences using the SVM algorithm [28]. CPAT was developed using a logistic regression method that trains datasets into hexamer table and logit model using mRNAs and putative lncRNAs/NPCTs transcripts [30]. The RNAsamba is a neural network-based approach to classify lncRNAs from protein-coding transcripts [29]. FEELnc is a random forest-based approach that uses known mRNAs and putative transcripts in *.gtf to report mRNA spliced and intergenic lncRNAs [33].

Also, PLEK and LGC were integrated into the present workflow. PLEK utilizes an improved k-mer scheme combined with SVM [31], while LGC employs a novel algorithm based on ORF length and GC content [32]. However, we suppressed the use of PLEK due to its time-consuming nature with bpipe and one can activate it by modifying the primary script named *Main.groovy* in the pipeline. We also incorporated PfamScan to exclude potential transcripts that resemble protein sequences during the lncRNA identification. The current workflow includes conducting conservation analysis of lncRNAs across species using Slncky, contingent upon the availability of additional data files such as LiftOver [https://genome.ucsc.edu/cgi-bin/hgLiftOver], snoRNAs, snRNAs, and miRNAs. Furthermore, we have supplied two additional scripts: *find_species_in_ensembl.py* and *download_datasets_ensembl.py*. The first investigate species in Ensembl Plants [http://plants.ensembl.org], while the latter will acquire necessary datasets including reference genomes (*.fa), annotations (*.gtf), and rRNA sequences (*.fa) using pybiomart [https://pypi.org/project/pybiomart/]. We also included scripts named *Liftover.py* and *ensembl_gtf2bed.py*, the first will generate LiftOver files and later will produce requisite annotations (*.bed) for Slncky conservation analysis. We tested the superiority of five coding potential-based approaches on sorghum RNA-seq datasets using five different supervised machine-learning models which reported LGC and CPC2 were performing better than others. Similar results for lncRNA identification were obtained using human RNA-seq datasets [67].

Plant-LncPipe, a recent approach was developed to report lncRNAs using various tools in horticultural crops [68]. However, this approach doesn't report NPCTs that have a significant role in plant responses such as drought. LncPipe, a Nextflow-based approach used for the lncRNAs identification from RNA-seq [69]. This uses CPAT and PLEK to classify lncRNAs from mRNA sequences but doesn't employ quality control with rRNA, reporting of NPCTs, and conservation studies using Slncky. A lncEvo is an automated workflow that utilizes Nextflow to incorporate Slncky for conservation analysis [70]. However, its implementation solely depends on either Docker or AWS cloud. In addition to conda profiles, the LncRAnalyzer can be implemented using Docker. We further conducted testing of this workflow using Docker and did not report any issues. UClncR, another workflow for identifying lncRNAs from RNA-seq data, can be applied to both stranded and un-stranded RNA-seq protocols [71]. The present workflow does not depend on the strandedness of RNA-seq data, but further exploration in this direction is possible. ICAnnoLncRNA, an automated workflow for RNA-seq-based lncRNA identification [72]. However, relying solely on LncFinder to identify lncRNAs may limit the results, and incorporating additional tools could potentially enhance the

accuracy and comprehensiveness of the findings. We incorporated a total of eight approaches for identifying lncRNAs. This could be increased depending on availability, easy installation, and execution through Bpipe [36].

The lncRNAs conservation analysis between BTX642 and RTX430 was performed which reported 1316 lncRNAs with exact matching intron chains and the rest were novel. We found that 47.24% of lncRNAs on the sorghum reference genome and 35.91% of lncRNAs on the cultivar reference genome exhibited conservation with PLncDB sequences in both genotypes. This infers that cultivar-specific lncRNAs in both genotypes were less conserved and might associated with drought tolerance. A past study on lncRNA conservation using BLASTN reported that lncRNAs exhibit lower sequence conservation than protein-coding genes [73]. This supports our findings. Previous reports indicate reduced lncRNA conservation among many plant species [74–76]. The lower conservation of lncRNAs might be due to extensive mRNA splicing and transcription of intergenic regions. Past research showed that lncRNAs were the product and precursors of alternative splicing in cancer, influencing disease progression through cancer-related splice variants [77].

We reported several lncRNAs that were cis and trans-acting targets for several TFs such as AP2/ERF, NAC, WD40-like, MYB, WRKY, FAR1, B3, bHLH, etc. They may play a crucial role in pre- and post-flowering drought tolerance. Similar results for TF-induced drought tolerance in plants were reported previously [78]. Additionally, we reported several lncRNAs that form triplex 1000 bp upstream and 200 bp downstream of TSS of important TFs such as WD-40, MADS-MIKC, C2H2, NAC, TRAF, bZIP, MYB, GRAS, CCHC(Zn), B3, WRKY, etc. This highlights the important role of lncRNAs in TF regulation. However, the number of TFs with triplex-forming sites was reduced in RTX430, suggesting that the mutations and genome organization of RTX430 might influence triplex formation. The research showed that lncRNA: DNA triplexes form specific chromosomal regions and are essential for predicting Topologically Associating Domains (TADs) [79]. Therefore, lncRNA: DNA triplexes affect both gene regulation as well as the organization of three-dimensional chromatin architecture. The KEGG orthology analysis of target genes indicates that many of the lncRNAs function as regulators for genes related to the activity of MYB transcription factors, peroxidase, and glutathione S-transferase in response to drought conditions. This highlights the role of these genes in drought tolerance in sorghum. Previous work on Arabidopsis [80], tobacco [81], and wheat [82] suggests that increased activity of glutathione S-transferase confers drought tolerance. While overexpression of MYB TF in Rice [83–85] and Arabidopsis [86–88] increased drought

tolerance. In addition, regulation of peroxidase genes enhances drought tolerance by reducing the accumulation of ROS in Rice [89].

A total of 54 NPCTs were reported in two genotypes, these sequences were transcribed and not previously annotated on genomes. Mostly, they encode for elements such as RE-1, FAR1, TNT, Zinc finger (zf) protein, receptor-like protein (RLP-6), Xylan glucosyltransferase, etc. This highlighted their important role in stress response. Previous analysis showed TNT-1 expressions were associated with abiotic factors such as wounding, freezing, and others and were known to induce defense plant mechanisms [90]. Moreover, the activity of Transposable Elements (TEs) was influenced by stress and led to an increased tolerance in plants [91]. This supports our findings, suggesting that the relationship between transposase activity and drought tolerance in sorghum can be further explored.

We found 172 TFs along with 264 lncRNAs in leaf tissues and 39 TFs with 145 lncRNAs in root tissues of BTX642 which were exclusive to post-flowering drought. This includes TFs such as WD-40, MYB, NAC, SNF, PHD, FAR1, bromodomain, A20-like, bZIP, AP2/ERF, BED-type (Zn), bHLH, etc., and highlights their role in the development of stay-green phenotype. Past analysis reported NAC, WD-40, MYB, bZIP, and FAR1 were associated with drought tolerance in sorghum [92–98]. In particular, the elevated expressions of NAC genes in leaf tissues were linked with post-flowering drought tolerance in sorghum [99]. The lncRNA BTX642.nov01G228.1 was one of the top 50 highly expressed in leaf tissues and positive regulator for genes such as serine/threonine protein phosphatase (PP2A), extended synaptotagmin (eSYT), calcium-dependent protein kinases (CDPKs), Fructose-1,6-bisphosphatase (F1-6BP), glycosyltransferase (GTs), etc. The PP2A served as an internal control for qPCR assays in sorghum [100,101]. However, reduced expression of synaptotagmin-5 (OsSYT-5) and F1-6BP induced drought tolerance in rice and sorghum [102,103]. The CDPK activity was increased in polyethylene glycol (PEG) induced sorghum and maize [104]. Increased activity of GTs in the leaf and roots leads to susceptibility to drought stress and less efficient recovery in sorghum genotypes BTX642 and RTX430 [105]. We examined this relationship and correlated expression patterns of GT, CDPK, eSYT, and F1-6BP with BTX642.nov01G228.1 in leaf tissues of BTX642, which showed decreased expression of these genes with BTX642.nov01G228.1 during post-flowering drought **(Supplemental Figure S6(a))**. This suggests BTX642.nov01G228.1 key player controlling post-flowering drought tolerance in leaf tissues. Likewise, BTX642.nov07G21079.3 was expressed exclusively under post-flowering drought in root tissues of BTX642. This gene positively regulates the expression of C2C2-GATA-12, SnoaL-like domain (SnoLD), Sentrin-specific protease

(SENP), and Histone H1/H5 **(Supplemental Figure S6(b))**. Past studies reported that the SnoaL-like polyketide cyclase region showed 50-fold expression in roots under drought in *Citrullus colocynthis* [106]. The activity of C2C2-GATA TFs was linked to drought tolerance in barley plants exposed to long-term drought [107]. In RTX430, RTX430.nov09G22826.1 is one of the top 50 highly expressed lncRNAs in leaf tissues during pre-flowering drought and positively regulates genes WD40-like-120, DUF domain protein (DUF), E3-ubiquitin ligase (E3 ligase), and RNA binding protein (RBM25) **(Supplemental Figure S6(c))**. Expression of WD-40, DUF domains, and E3-ubiquitin ligase in leaf tissues was linked to drought tolerance in sorghum [108–111]. Similarly, RTX430.nov08G21197.1 one of the key players in the root tissues of RTX430 under pre-flowering drought positively regulates C2H2-556, autophagy-related protein (ATG), proteasome component-related (ECM29), and LYST-interacting protein (LIP5), etc **(Supplemental Figure S6(d))**. The previous report showed that C2H2 zinc finger, LIP5, ATG, and proteasome component-related were linked to drought stress responses in plants [112–118].

Our time-ordered transcriptomics reported a total of 917 lncRNAs and 2804 TFs that were exclusively expressed during drought stress in leaf and root tissues of BTX642 and RTX430 **(Supplemental Data S12).** These include several TF families such as C2H2, AP2/ERF-ERF, WD40-like, Hap3/NF-YB, bZIP, MYB, PHD, bHLH, WRKY, GRAS, WRKY, NAC, CCHC(Zn), etc. with varied numbers between tissue types which highlight their roles in drought tolerance. Similar stress-responsive genes/TFs were identified previously in various plants, supporting our findings [119–123].

Conclusion:

The present analysis developed a workflow named LncRAnalyzer for lncRNAs and NPCTs identification using RNA-seq datasets. We integrated eight different approaches into the present workflow, of which seven approaches were tested on 389 RNA-seq accessions of BTX642 and RTX430 from previous studies [37]. In our analysis, LGC demonstrated superior performance over CPC2, FEELnc, CPAT, and RNAsamba when trained datasets using different machine learning models. We reported several known and cultivar-specific lncRNAs and NPCTs that have a role in pre- and post-flowering drought tolerance. We selected four condition-specific candidate lncRNAs namely BTX642.nov01G228.1, BTX642.nov07G21079.3, RTX430.nov09G22826.1, and RTX430.nov08G21197.1 were under the top 50 highly expressed categories in leaf and root tissues of BTX642 and RTX430 and could be associated with pre- and post-flowering drought tolerance. Time-ordered gene co-expression analysis revealed timepoint co-expressed lncRNAs and

TFs and identified several drought-responsive TF families in sorghum. The study will enrich the existing set of lncRNAs identified and provide insight into the regulation of drought-responsive genes in sorghum.

Availability of data and materials:

The LncRAnalyzer [https://github.com/nikhilshinde0909/LncRAnalyzer] and LncRAnalyzer-Docker version [https://github.com/nikhilshinde0909/LncRAnalyzer-Docker] both are available on GitHub. The docker image of LncRAnalyzer is available and can be pulled from DockerHub [https://hub.docker.com/repository/docker/nikhilshinde0909/lncranalyzer/general]. The Python and R codes used for training datasets using machine learning and preparing ROC curves will be provided upon request. The supplemental data, tables, pipeline, and user manual have been submitted along with the manuscript.

Acknowledgments:

We thank the SRM Institute of Science and Technology, Kattankulathur, Chennai, India, for offering fellowships and the facilities required to conduct the present research. We also acknowledge SRM HPCC for providing the computational analytic infrastructure.

Author Contributions:

1. Conceptualization: SN, RNS
2. Methodology: SN, HSM
3. Investigation: SN, RNS
4. Visualization: SN, RNS, HSM
5. Supervision: RNS, HSM
6. Writing-original draft: SN
7. Writing-review & editing: SN, RNS

Conflict of Interest:

The authors declare that they have no competing interests.

References:

[1] R.F. McCormick, S.K. Truong, A. Sreedasyam, J. Jenkins, S. Shu, D. Sims, M. Kennedy, M. Amirebrahimi, B.D. Weers, B. McKinley, A. Mattison, D.T. Morishige, J. Grimwood, J. Schmutz, J.E. Mullet, The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome

atlas, and signatures of genome organization, Plant J. 93 (2018). https://doi.org/10.1111/tpj.13781.

[2] E.A. Cooper, Z.W. Brenton, B.S. Flinn, J. Jenkins, S. Shu, D. Flowers, F. Luo, Y. Wang, P. Xia, K. Barry, C. Daum, A. Lipzen, Y. Yoshinaga, J. Schmutz, C. Saski, W. Vermerris, S. Kresovich, A new reference genome for Sorghum bicolor reveals high levels of sequence similarity between sweet and grain genotypes: implications for the genetics of sugar metabolism, BMC Genomics 2019 201 20 (2019) 1–13. https://doi.org/10.1186/S12864-019-5734-X.

[3] Y. Tao, H. Luo, J. Xu, A. Cruickshank, X. Zhao, F. Teng, A. Hathorn, X. Wu, Y. Liu, T. Shatte, D. Jordan, H. Jing, E. Mace, Extensive variation within the pan-genome of cultivated and wild sorghum, Nat. Plants 7 (2021). https://doi.org/10.1038/s41477-021-00925-x.

[4] K. Songsomboon, Z. Brenton, J. Heuser, S. Kresovich, N. Shakoor, T. Mockler, E.A. Cooper, Genomic patterns of structural variation among diverse genotypes of Sorghum bicolor and a potential role for deletions in local adaptation, G3 Genes, Genomes, Genet. 11 (2021). https://doi.org/10.1093/g3journal/jkab154.

[5] F. Salvato, P. Loziuk, E. Kiyota, G.S. Daneluzzi, P. Araújo, D.C. Muddiman, P. Mazzafera, Label-Free Quantitative Proteomics of Enriched Nuclei from Sugarcane (Saccharum ssp) Stems in Response to Drought Stress, Proteomics 19 (2019). https://doi.org/10.1002/pmic.201900004.

[6] Y. Lu, Y. Li, J. Zhang, Y. Xiao, Y. Yue, L. Duan, M. Zhang, Z. Li, Overexpression of Arabidopsis Molybdenum Cofactor Sulfurase Gene Confers Drought Tolerance in Maize (Zea mays L.), PLoS One 8 (2013). https://doi.org/10.1371/journal.pone.0052126.

[7] P. Lü, M. Kang, X. Jiang, F. Dai, J. Gao, C. Zhang, RhEXPA4, a rose expansin gene, modulates leaf growth and confers drought and salt tolerance to Arabidopsis, Planta 237 (2013). https://doi.org/10.1007/s00425-013-1867-3.

[8] S. Ying, D.F. Zhang, J. Fu, Y.S. Shi, Y.C. Song, T.Y. Wang, Y. Li, Cloning and characterization of a maize bZIP transcription factor, ZmbZIP72, confers drought and salt tolerance in transgenic Arabidopsis, Planta 235 (2012). https://doi.org/10.1007/s00425-011-1496-7.

[9] Z.H. Cui, W.L. Bi, X.Y. Hao, P.M. Li, Y. Duan, M.A. Walker, Y. Xu, Q.C. Wang, Drought stress enhances up-regulation of anthocyanin biosynthesis in Grapevine leafroll-associated virus 3-infected in vitro grapevine (Vitis Vinifera) leaves, Plant Dis. 101 (2017). https://doi.org/10.1094/PDIS-01-17-0104-RE.

[10] S.C. Jiang, C. Mei, S. Liang, Y.T. Yu, K. Lu, Z. Wu, X.F. Wang, D.P. Zhang, Crucial roles of the

pentatricopeptide repeat protein SOAR1 in Arabidopsis response to drought, salt and cold stresses, Plant Mol. Biol. 88 (2015). https://doi.org/10.1007/s11103-015-0327-9.

[11] T.R. Mercer, M.E. Dinger, J.S. Mattick, Long non-coding RNAs: Insights into functions, Nat. Rev. Genet. 10 (2009). https://doi.org/10.1038/nrg2521.

[12] M. Xin, Y. Wang, Y. Yao, N. Song, Z. Hu, D. Qin, C. Xie, H. Peng, Z. Ni, Q. Sun, Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing, BMC Plant Biol. 11 (2011). https://doi.org/10.1186/1471-2229-11-61.

[13] U.C. Jha, H. Nayyar, R. Jha, M. Khurshid, M. Zhou, N. Mantri, K.H.M. Siddique, Long non-coding RNAs: Emerging players regulating plant abiotic stress response and adaptation, BMC Plant Biol. 20 (2020). https://doi.org/10.1186/s12870-020-02595-x.

[14] D. Panda, L. Ananda Lekshmi, R. Lissy Vargheese, N. Premalatha, M. Kumar, L. Mahalingam, Non-Coding RNA and Its Prospective Utilization in Plant Breeding, in: Case Stud. Breed. Strateg. Major Plant Species, 2023. https://doi.org/10.5772/intechopen.106429.

[15] L. Statello, C.J. Guo, L.L. Chen, M. Huarte, Gene regulation by long non-coding RNAs and its biological functions, Nat. Rev. Mol. Cell Biol. 22 (2021). https://doi.org/10.1038/s41580-020-00315-9.

[16] J.S. Mattick, P.P. Amaral, P. Carninci, S. Carpenter, H.Y. Chang, L.L. Chen, R. Chen, C. Dean, M.E. Dinger, K.A. Fitzgerald, T.R. Gingeras, M. Guttman, T. Hirose, M. Huarte, R. Johnson, C. Kanduri, P. Kapranov, J.B. Lawrence, J.T. Lee, J.T. Mendell, T.R. Mercer, K.J. Moore, S. Nakagawa, J.L. Rinn, D.L. Spector, I. Ulitsky, Y. Wan, J.E. Wilusz, M. Wu, Long non-coding RNAs: definitions, functions, challenges and recommendations, Nat. Rev. Mol. Cell Biol. 24 (2023). https://doi.org/10.1038/s41580-022-00566-8.

[17] X. Wang, H. Fan, B. Wang, F. Yuan, Research progress on the roles of lncRNAs in plant development and stress responses, Front. Plant Sci. 14 (2023). https://doi.org/10.3389/fpls.2023.1138901.

[18] H.L. V. Wang, J.A. Chekanova, Long noncoding RNAs in plants, in: Adv. Exp. Med. Biol., 2017. https://doi.org/10.1007/978-981-10-5203-3_5.

[19] J. Liu, H. Wang, N.H. Chua, Long noncoding RNA transcriptome of plants, Plant Biotechnol. J. 13 (2015). https://doi.org/10.1111/pbi.12336.

[20] Y. Wang, X. Wang, W. Deng, X. Fan, T.T. Liu, G. He, R. Chen, W. Terzaghi, D. Zhu, X. Wang Deng, Genomic features and regulatory roles of intermediate-sized non-coding rnas in arabidopsis, Mol. Plant

7 (2014). https://doi.org/10.1093/mp/sst177.

[21] T.T. Liu, D. Zhu, W. Chen, W. Deng, H. He, G. He, B. Bai, Y. Qi, R. Chen, X.W. Deng, A global identification and analysis of small nucleolar RNAs and possible intermediate-sized non-coding RNAs in or, Mol. Plant 6 (2013) 830–846. https://doi.org/10.1093/mp/sss087.

[22] T. Lu, C. Zhu, G. Lu, Y. Guo, Y. Zhou, Z. Zhang, Y. Zhao, W. Li, Y. Lu, W. Tang, Q. Feng, B. Han, Strand-specific RNA-seq reveals widespread occurrence of novel cis-natural antisense transcripts in rice, BMC Genomics 13 (2012). https://doi.org/10.1186/1471-2164-13-721.

[23] L. Li, S.R. Eichten, R. Shimizu, K. Petsch, C.T. Yeh, W. Wu, A.M. Chettoor, S.A. Givan, R.A. Cole, J.E. Fowler, M.M.S. Evans, M.J. Scanlon, J. Yu, P.S. Schnable, M.C.P. Timmermans, N.M. Springer, G.J. Muehlbauer, Genome-wide discovery and characterization of maize long non-coding RNAs, Genome Biol. 15 (2014). https://doi.org/10.1186/gb-2014-15-2-r40.

[24] F. Ariel, T. Jegu, D. Latrasse, N. Romero-Barrios, A. Christ, M. Benhamed, M. Crespi, Noncoding transcription by alternative rna polymerases dynamically regulates an auxin-driven chromatin loop, Mol. Cell 55 (2014). https://doi.org/10.1016/j.molcel.2014.06.011.

[25] I. Ulitsky, D.P. Bartel, Ulitsky, Igor, and David P Bartel. 2013. “lincRNAs: Genomics, Evolution, and Mechanisms.” Cell 154 (1): 26–46. doi:10.1016/j.cell.2013.06.020.lincRNAs: genomics, evolution, and mechanisms., Cell 154 (2013).

[26] A.C. Tuck, D. Tollervey, XA transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs, Cell 154 (2013). https://doi.org/10.1016/j.cell.2013.07.047.

[27] T. Doughty, E. Kerkhoven, Extracting novel hypotheses and findings from RNA-seq data, FEMS Yeast Res. 20 (2020). https://doi.org/10.1093/femsyr/foaa007.

[28] Y.J. Kang, D.C. Yang, L. Kong, M. Hou, Y.Q. Meng, L. Wei, G. Gao, CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features, Nucleic Acids Res. 45 (2017). https://doi.org/10.1093/nar/gkx428.

[29] A.P. Camargo, V. Sourkov, G.A.G. Pereira, M.F. Carazzolle, RNAsamba: Neural network-based assessment of the protein-coding potential of RNA sequences, NAR Genomics Bioinforma. 2 (2020). https://doi.org/10.1093/nargab/lqz024.

[30] L. Wang, H.J. Park, S. Dasari, S. Wang, J.P. Kocher, W. Li, CPAT: Coding-potential assessment tool using an alignment-free logistic regression model, Nucleic Acids Res. 41 (2013). https://doi.org/10.1093/nar/gkt006.

[31] A. Li, J. Zhang, Z. Zhou, PLEK: A tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme, BMC Bioinformatics 15 (2014). https://doi.org/10.1186/1471-2105-15-311.

[32] G. Wang, H. Yin, B. Li, C. Yu, F. Wang, X. Xu, J. Cao, Y. Bao, L. Wang, A.A. Abbasi, V.B. Bajic, L. Ma, Z. Zhang, Characterization and identification of long non-coding RNAs based on feature relationship, Bioinformatics 35 (2019). https://doi.org/10.1093/bioinformatics/btz008.

[33] V. Wucher, F. Legeai, B. Hédan, G. Rizk, L. Lagoutte, T. Leeb, V. Jagannathan, E. Cadieu, A. David, H. Lohi, S. Cirera, M. Fredholm, N. Botherel, P.A.J. Leegwater, C. Le Béguec, H. Fieten, J. Johnson, J. Alföldi, C. André, K. Lindblad-Toh, C. Hitte, T. Derrien, FEELnc: A tool for long non-coding RNA annotation and its application to the dog transcriptome, Nucleic Acids Res. 45 (2017). https://doi.org/10.1093/nar/gkw1306.

[34] W. Li, A. Cowley, M. Uludag, T. Gur, H. McWilliam, S. Squizzato, Y.M. Park, N. Buso, R. Lopez, The EMBL-EBI bioinformatics web and programmatic tools framework, Nucleic Acids Res. 43 (2015). https://doi.org/10.1093/nar/gkv279.

[35] J. Chen, A.A. Shishkin, X. Zhu, S. Kadri, I. Maza, M. Guttman, J.H. Hanna, A. Regev, M. Garber, Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs, Genome Biol. 17 (2016). https://doi.org/10.1186/s13059-016-0880-9.

[36] S.P. Sadedin, B. Pope, A. Oshlack, Bpipe: A tool for running and managing bioinformatics pipelines, Bioinformatics 28 (2012). https://doi.org/10.1093/bioinformatics/bts167.

[37] N. Varoquaux, B. Cole, C. Gao, G. Pierroz, C.R. Baker, D. Patel, M. Madera, T. Jeffers, J. Hollingsworth, J. Sievert, Y. Yoshinaga, J.A. Owiti, V.R. Singan, S. DeGraaf, L. Xu, M.J. Blow, M.J. Harrison, A. Visel, C. Jansson, K.K. Niyogi, R. Hutmacher, D. Coleman-Derr, R.C. O'Malley, J.W. Taylor, J. Dahlberg, J.P. Vogel, P.G. Lemaux, E. Purdom, Transcriptomic analysis of field-droughted sorghum from seedling to maturity reveals biotic and metabolic responses, Proc. Natl. Acad. Sci. U. S. A. 116 (2019). https://doi.org/10.1073/pnas.1907500116.

[38] S. Chen, Y. Zhou, Y. Chen, J. Gu, Fastp: An ultra-fast all-in-one FASTQ preprocessor, in: Bioinformatics, 2018. https://doi.org/10.1093/bioinformatics/bty560.

[39] D. Kim, J.M. Paggi, C. Park, C. Bennett, S.L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype, Nat. Biotechnol. 37 (2019). https://doi.org/10.1038/s41587-019-0201-4.

[40] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The Sequence Alignment/Map format and SAMtools, Bioinformatics 25 (2009). https://doi.org/10.1093/bioinformatics/btp352.

[41] A.R. Quinlan, I.M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features, Bioinformatics 26 (2010). https://doi.org/10.1093/bioinformatics/btq033.

[42] M. Pertea, G.M. Pertea, C.M. Antonescu, T.C. Chang, J.T. Mendell, S.L. Salzberg, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads, Nat. Biotechnol. 33 (2015). https://doi.org/10.1038/nbt.3122.

[43] G. Pertea, M. Pertea, GFF Utilities: GffRead and GffCompare, F1000Research 9 (2020). https://doi.org/10.12688/f1000research.23297.2.

[44] M.C. Frith, M. Hamada, P. Horton, Parameters for accurate genome alignment, BMC Bioinformatics 11 (2010). https://doi.org/10.1186/1471-2105-11-80.

[45] D. Karolchik, A.S. Hinricks, T.S. Furey, K.M. Roskin, C.W. Sugnet, D. Haussler, W.J. Kent, The UCSC table browser data retrieval tool, Nucleic Acids Res. 32 (2004). https://doi.org/10.1093/nar/gkh103.

[46] A. Dusa, Venn: Draw Venn Diagrams, Available at: Http://CRAN.R-Project.Org/Package=venn (2017).

[47] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011).

[48] N.M. Davidson, A.D.K. Hawkins, A. Oshlack, SuperTranscripts: A data driven reference for analysis and visualisation of transcriptomes, Genome Biol. 18 (2017). https://doi.org/10.1186/s13059-017-1284-1.

[49] J. Jin, P. Lu, Y. Xu, Z. Li, S. Yu, J. Liu, H. Wang, N.H. Chua, P. Cao, PLncDB V2.0: A comprehensive encyclopedia of plant long noncoding RNAs, Nucleic Acids Res. 49 (2021). https://doi.org/10.1093/nar/gkaa910.

[50] P. Liu, Y. Zhang, C. Zou, C. Yang, G. Pan, L. Ma, Y. Shen, Integrated analysis of long non-coding RNAs and mRNAs reveals the regulatory network of maize seedling root responding to salt stress, BMC Genomics 23 (2022). https://doi.org/10.1186/s12864-021-08286-7.

[51] Y. Li, J. Syed, H. Sugiyama, RNA-DNA Triplex Formation by Long Noncoding RNAs, Cell Chem.

Biol. 23 (2016). https://doi.org/10.1016/j.chembiol.2016.09.011.

[52] E. Matveishina, I. Antonov, Y.A. Medvedeva, Practical guidance in genome-wide RNA: DNA triple helix prediction, Int. J. Mol. Sci. 21 (2020). https://doi.org/10.3390/ijms21030830.

[53] F.A. Buske, D.C. Bauer, J.S. Mattick, T.L. Bailey, Triplexator: Detecting nucleic acid triple helices in genomic and transcriptomic data, Genome Res. 22 (2012). https://doi.org/10.1101/gr.130237.111.

[54] Y. Moriya, M. Itoh, S. Okuda, A.C. Yoshizawa, M. Kanehisa, KAAS: An automatic genome annotation and pathway reconstruction server, Nucleic Acids Res. 35 (2007). https://doi.org/10.1093/nar/gkm321.

[55] B. Haas, A. Papanicolaou, TransDecoder, Https://Github.Com/TransDecoder/TransDecoder (2017).

[56] A. Sayadi, E. Immonen, H. Bayram, G. Arnqvist, The de novo transcriptome and its functional annotation in the seed beetle callosobruchus maculatus, PLoS One 11 (2016). https://doi.org/10.1371/journal.pone.0158565.

[57] N. Ghaffari, A. Sanchez-Flores, R. Doan, K.D. Garcia-Orozco, P.L. Chen, A. Ochoa-Leyva, A.A. Lopez-Zavala, J.S. Carrasco, C. Hong, L.G. Brieba, E. Rudiño-Piñera, P.D. Blood, J.E. Sawyer, C.D. Johnson, S. V. Dindot, R.R. Sotelo-Mundo, M.F. Criscitiello, Novel transcriptome assembly and improved annotation of the whiteleg shrimp (Litopenaeus vannamei), a dominant crustacean in global seafood mariculture, Sci. Rep. 4 (2014). https://doi.org/10.1038/srep07081.

[58] X. Dai, S. Sinharoy, M. Udvardi, P.X. Zhao, PlantTFcat: An online plant transcription factor and transcriptional regulator categorization and analysis tool, BMC Bioinformatics 14 (2013). https://doi.org/10.1186/1471-2105-14-321.

[59] A.Y. Guo, X. Chen, G. Gao, H. Zhang, Q.H. Zhu, X.C. Liu, Y.F. Zhong, X. Gu, K. He, J. Luo, PlantTFDB: A comprehensive plant transcription factor database, Nucleic Acids Res. 36 (2008). https://doi.org/10.1093/nar/gkm841.

[60] Y. Zheng, C. Jiao, H. Sun, H.G. Rosli, M.A. Pombo, P. Zhang, M. Banf, X. Dai, G.B. Martin, J.J. Giovannoni, P.X. Zhao, S.Y. Rhee, Z. Fei, iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases, Mol. Plant 9 (2016). https://doi.org/10.1016/j.molp.2016.09.014.

[61] Y. Liao, G.K. Smyth, W. Shi, FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features, Bioinformatics 30 (2014). https://doi.org/10.1093/bioinformatics/btt656.

[62] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data

with DESeq2, Genome Biol. 15 (2014). https://doi.org/10.1186/s13059-014-0550-8.

[63] Blighe K, Rana S, Lewis M, Publication-ready volcano plots with enhanced colouring and labeling, 2022.

[64] Y.M. Chang, H.H. Lin, W.Y. Liu, C.P. Yu, H.J. Chen, P.P. Wartini, Y.Y. Kao, Y.H. Wu, J.J. Lin, M.Y.J. Lu, S.L. Tu, S.H. Wu, S.H. Shiu, M.S.B. Ku, W.H. Li, Comparative transcriptomics method to infer gene coexpression networks and its applications to maize and rice leaf transcriptomes, Proc. Natl. Acad. Sci. U. S. A. 116 (2019). https://doi.org/10.1073/pnas.1817621116.

[65] D. Otasek, J.H. Morris, J. Bouças, A.R. Pico, B. Demchak, Cytoscape Automation: Empowering workflow-based network analysis, Genome Biol. 20 (2019). https://doi.org/10.1186/s13059-019-1758-4.

[66] C.P. Yu, S.C.C. Chen, Y.M. Chang, W.Y. Liu, H.H. Lin, J.J. Lin, H.J. Chen, Y.J. Lu, Y.H. Wu, M.Y.J. Lu, C.H. Lu, A.C.C. Shih, M.S. Ben Ku, S.H. Shiu, S.H. Wu, W.H. Li, Transcriptome dynamics of developing maize leaves and genomewide prediction of cis elements and their cognate transcription factors, Proc. Natl. Acad. Sci. U. S. A. 112 (2015). https://doi.org/10.1073/pnas.1500605112.

[67] H. Zheng, A. Talukder, X. Li, H. Hu, A systematic evaluation of the computational tools for lncRNA identification, Brief. Bioinform. 22 (2021). https://doi.org/10.1093/bib/bbab285.

[68] X.C. Tian, Z.Y. Chen, S. Nie, T. Le Shi, X.M. Yan, Y.T. Bao, Z.C. Li, H.Y. Ma, K.H. Jia, W. Zhao, J.F. Mao, Plant-LncPipe: a computational pipeline providing significant improvement in plant lncRNA identification, Hortic. Res. 11 (2024). https://doi.org/10.1093/hr/uhae041.

[69] Q. Zhao, Y. Sun, D. Wang, H. Zhang, K. Yu, J. Zheng, Z. Zuo, LncPipe: A Nextflow-based pipeline for identification and analysis of long non-coding RNAs from RNA-Seq data, J. Genet. Genomics 45 (2018). https://doi.org/10.1016/j.jgg.2018.06.005.

[70] O. Bryzghalov, I. Makałowska, M.W. Szcześniak, lncEvo: automated identification and conservation study of long noncoding RNAs, BMC Bioinformatics 22 (2021). https://doi.org/10.1186/s12859-021-03991-2.

[71] Z. Sun, A. Nair, X. Chen, N. Prodduturi, J. Wang, J.P. Kocher, UClncR: Ultrafast and comprehensive long non-coding RNA detection from RNA-seq, Sci. Rep. 7 (2017). https://doi.org/10.1038/s41598-017-14595-3.

[72] A.Y. Pronozin, D.A. Afonnikov, ICAnnoLncRNA: A Snakemake Pipeline for a Long Non-Coding-RNA Search and Annotation in Transcriptomic Sequences, Genes (Basel). 14 (2023).

https://doi.org/10.3390/genes14071331.

[73] S. Sang, W. Chen, D. Zhang, X. Zhang, W. Yang, C. Liu, Data integration and evolutionary analysis of long non-coding RNAs in 25 flowering plants, BMC Genomics 22 (2021). https://doi.org/10.1186/s12864-021-08047-6.

[74] P. Deng, S. Liu, X. Nie, S. Weining, L. Wu, Conservation analysis of long non-coding RNAs in plants, Sci. China Life Sci. 61 (2018). https://doi.org/10.1007/s11427-017-9174-9.

[75] I. Ulitsky, Evolution to the rescue: Using comparative genomics to understand long non-coding RNAs, Nat. Rev. Genet. 17 (2016). https://doi.org/10.1038/nrg.2016.85.

[76] H. Hezroni, D. Koppstein, M.G. Schwartz, A. Avrutin, D.P. Bartel, I. Ulitsky, Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species, Cell Rep. 11 (2015). https://doi.org/10.1016/j.celrep.2015.04.023.

[77] J. Ouyang, Y. Zhong, Y. Zhang, L. Yang, P. Wu, X. Hou, F. Xiong, X. Li, S. Zhang, Z. Gong, Y. He, Y. Tang, W. Zhang, B. Xiang, M. Zhou, J. Ma, Y. Li, G. Li, Z. Zeng, C. Guo, W. Xiong, Long non-coding RNAs are involved in alternative splicing and promote cancer progression, Br. J. Cancer 126 (2022). https://doi.org/10.1038/s41416-021-01600-w.

[78] Y. Hu, X. Chen, X. Shen, Regulatory network established by transcription factors transmits drought stress signals in plant, Stress Biol. 2 (2022). https://doi.org/10.1007/s44154-022-00048-z.

[79] B. Soibam, A. Zhamangaraeva, LncRNA:DNA triplex-forming sites are positioned at specific areas of genome organization and are predictors for Topologically Associated Domains, BMC Genomics 22 (2021). https://doi.org/10.1186/s12864-021-07727-7.

[80] J. Xu, X.J. Xing, Y.S. Tian, R.H. Peng, Y. Xue, W. Zhao, Q.H. Yao, H. Zhang, Transgenic Arabidopsis plants expressing tomato glutathione S-transferase showed enhanced resistance to salt and drought stress, PLoS One 10 (2015). https://doi.org/10.1371/journal.pone.0136960.

[81] W. Ji, Y. Zhu, Y. Li, L. Yang, X. Zhao, H. Cai, X. Bai, Over-expression of a glutathione S-transferase gene, GsGST, from wild soybean (Glycine soja) enhances drought and salt tolerance in transgenic tobacco, Biotechnol. Lett. 32 (2010). https://doi.org/10.1007/s10529-010-0269-x.

[82] Á. Gallé, J. Csiszár, M. Secenji, A. Guóth, L. Cseuz, I. Tari, J. Györgyey, L. Erdei, Glutathione transferase activity and expression patterns during grain filling in flag leaves of wheat genotypes differing in drought tolerance: Response to water deficit, J. Plant Physiol. 166 (2009). https://doi.org/10.1016/j.jplph.2009.05.016.

[83] Y. Tang, X. Bao, Y. Zhi, Q. Wu, Y. Guo, X. Yin, L. Zeng, J. Li, J. Zhang, W. He, W. Liu, Q. Wang, C. Jia, Z. Li, K. Liu, Overexpression of a myb family gene, Osmyb6, increases drought and salinity stress tolerance in transgenic rice, Front. Plant Sci. 10 (2019). https://doi.org/10.3389/fpls.2019.00168.

[84] X. Yin, Y. Cui, M. Wang, X. Xia, Overexpression of a novel MYB-related transcription factor, OsMYBR1, confers improved drought tolerance and decreased ABA sensitivity in rice, Biochem. Biophys. Res. Commun. 490 (2017). https://doi.org/10.1016/j.bbrc.2017.07.029.

[85] H. Xiong, J. Li, P. Liu, J. Duan, Y. Zhao, X. Guo, Y. Li, H. Zhang, J. Ali, Z. Li, Overexpression of OsMYB48-1, a novel MYB-related transcription factor, enhances drought and salinity tolerance in rice, PLoS One 9 (2014). https://doi.org/10.1371/journal.pone.0092913.

[86] X. Dai, Y. Xu, Q. Ma, W. Xu, T. Wang, Y. Xue, K. Chong, Overexpression of an R1R2R3 MYB gene, OsMYB3R-2, increases tolerance to freezing, drought, and salt stress in transgenic Arabidopsis, Plant Physiol. 143 (2007). https://doi.org/10.1104/pp.106.094532.

[87] Y. Wang, Q. Wang, M.L. Liu, C. Bo, X. Wang, Q. Ma, B. Cheng, R. Cai, Overexpression of a maize MYB48 gene confers drought tolerance in transgenic arabidopsis plants, J. Plant Biol. 60 (2017). https://doi.org/10.1007/s12374-017-0273-y.

[88] Y. Qin, M. Wang, Y. Tian, W. He, L. Han, G. Xia, Over-expression of TaMYB33 encoding a novel wheat MYB transcription factor increases salt and drought tolerance in Arabidopsis, Mol. Biol. Rep. 39 (2012). https://doi.org/10.1007/s11033-012-1550-y.

[89] C. Zheng, X. Wang, Y. Xu, S. Wang, X. Jiang, X. Liu, W. Cui, Y. Wu, C. Yan, H. Liu, Y. Lu, J. Chen, J. Zhou, The peroxidase gene OsPrx114 activated by OsWRKY50 enhances drought tolerance through ROS scavenging in rice, Plant Physiol. Biochem. 204 (2023). https://doi.org/10.1016/j.plaphy.2023.108138.

[90] C. Mhiri, J.B. Morel, S. Vernhettes, J.M. Casacuberta, H. Lucas, M.A. Grandbastien, The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress, Plant Mol. Biol. 33 (1997). https://doi.org/10.1023/A:1005727132202.

[91] M. Thieme, A. Brêchet, Y. Bourgeois, B. Keller, E. Bucher, A.C. Roulin, Experimentally heat-induced transposition increases drought tolerance in Arabidopsis thaliana, New Phytol. 236 (2022). https://doi.org/10.1111/nph.18322.

[92] A. Kebede, G. Dugassa, B. Mulugeta, K. Taye, Multi - locus genome - wide association analysis for root and shoot traits at seedling stage in Ethiopian sorghum (Sorghum bicolor (L .) Moench)

accessions, Genet. Resour. Crop Evol. (2024). https://doi.org/10.1007/s10722-024-02066-4.

[93] J. Wang, J. Zhou, B. Zhang, J. Vanitha, S. Ramachandran, S.Y. Jiang, Genome-wide Expansion and Expression Divergence of the Basic Leucine Zipper Transcription Factors in Higher Plants with an Emphasis on Sorghum, J. Integr. Plant Biol. 53 (2011). https://doi.org/10.1111/j.1744-7909.2010.01017.x.

[94] R. Joshi, S.H. Wani, B. Singh, A. Bohra, Z.A. Dar, A.A. Lone, A. Pareek, S.L. Singla-Pareek, Transcription factors and plants response to drought stress: Current understanding and future directions, Front. Plant Sci. 7 (2016). https://doi.org/10.3389/fpls.2016.01029.

[95] A. Katiyar, S. Smita, S.K. Muthusamy, V. Chinnusamy, D.M. Pandey, K.C. Bansal, Identification of novel drought-responsive microRNAs and trans-acting siRNAs from Sorghum bicolor (L.) Moench by high-throughput sequencing analysis, Front. Plant Sci. 6 (2015). https://doi.org/10.3389/fpls.2015.00506.

[96] M. Lu, D.F. Zhang, Y.S. Shi, Y.C. Song, T.Y. Wang, Y. Li, Expression of SbSNAC1, a NAC transcription factor from sorghum, confers drought tolerance to transgenic Arabidopsis, Plant Cell. Tissue Organ Cult. 115 (2013). https://doi.org/10.1007/s11240-013-0375-2.

[97] X. Jin, Y. Long, S. Xiong, Z. Yang, W. Chen, A. Hawar, X. Chi, Y. Chen, H. Luo, J. Qi, G. Lu, L. Dai, Y. Yang, B. Sun, SbNAC2 enhances abiotic stress tolerance by upregulating ROS scavenging activities and inducing stress-response genes in sorghum, Environ. Exp. Bot. 192 (2021). https://doi.org/10.1016/j.envexpbot.2021.104664.

[98] D. feng ZHANG, T. ru ZENG, X. yang LIU, C. xi GAO, Y. xiang LI, C. hui LI, Y. chun SONG, Y. su SHI, T. yu WANG, Y. LI, Transcriptomic profiling of sorghum leaves and roots responsive to drought stress at the seedling stage, J. Integr. Agric. 18 (2019). https://doi.org/10.1016/S2095-3119(18)62119-7.

[99] S. Sanjari, R. Shirzadian-Khorramabad, Z.S. Shobbar, M. Shahbazi, Systematic analysis of NAC transcription factors' gene family and identification of post-flowering drought stress responsive members in sorghum, Plant Cell Rep. 38 (2019). https://doi.org/10.1007/s00299-019-02371-8.

[100] R. Kumar Singh, A. Prasad, M. Prasad, Protein Phosphatases of Cereals and Millets: Identification, Structural Organization, and Their Involvement in the Regulation of Abiotic Stresses, in: Protein Phosphatases Stress Manag. Plants Funct. Genomic Perspect., 2020. https://doi.org/10.1007/978-3-030-48733-1_13.

[101] P. Sudhakar Reddy, D. Srinivas Reddy, K. Sivasakthi, P. Bhatnagar-Mathur, V. Vadez, K.K. Sharma,

Evaluation of Sorghum [Sorghum bicolor (L.)] reference genes in various tissues and under abiotic stress conditions for quantitative real-time PCR data normalization, Front. Plant Sci. 7 (2016). https://doi.org/10.3389/fpls.2016.00529.

[102] S. Shanmugam, V.A. Boyett, M. Khodakovskaya, Enhancement of drought tolerance in rice by silencing of the OsSYT-5 gene, PLoS One 16 (2021). https://doi.org/10.1371/journal.pone.0258171.

[103] A. Ramachandra Reddy, Fructose 2,6-bisphosphate-modulated photosynthesis in Sorghum leaves grown under low water regimes, Phytochemistry 43 (1996). https://doi.org/10.1016/0031-9422(96)00052-0.

[104] A. Pestenácz, L. Erdei, Calcium-dependent protein kinase in maize and sorghum induced by polyethylene glycol, Physiol. Plant. 97 (1996). https://doi.org/10.1034/j.1399-3054.1996.970221.x.

[105] T. Scavuzzo-Duggan, N. Varoquaux, M. Madera, J.P. Vogel, J. Dahlberg, R. Hutmacher, M. Belcher, J. Ortega, D. Coleman-Derr, P. Lemaux, E. Purdom, H. V. Scheller, Cell Wall Compositions of Sorghum bicolor Leaves and Roots Remain Relatively Constant Under Drought Conditions, Front. Plant Sci. 12 (2021). https://doi.org/10.3389/fpls.2021.747225.

[106] Y. Si, C. Zhang, S. Meng, F. Dane, Gene expression changes in response to drought stress in Citrullus colocynthis, Plant Cell Rep. 28 (2009). https://doi.org/10.1007/s00299-009-0703-5.

[107] Y. Wang, H. Li, C. Zhao, C. Yang, Q. Xu, H. Yuan, H. Yang, X. Zeng, Identification of a novel transcription factor under long-term drought resistance in highland barley: a DNA affinity purification sequencing-based transcriptomic analysis, Chem. Biol. Technol. Agric. 10 (2023). https://doi.org/10.1186/s40538-022-00376-2.

[108] J. Chen, R. Chopra, C. Hayes, G. Morris, S. Marla, J. Burke, Z. Xin, G. Burow, Genome-Wide Association Study of Developing Leaves' Heat Tolerance during Vegetative Growth Stages in a Sorghum Association Panel, Plant Genome 10 (2017). https://doi.org/10.3835/plantgenome2016.09.0091.

[109] M. Nagaraju, P.S. Reddy, S.A. Kumar, A. Kumar, G. Rajasheker, D.M. Rao, P.B. Kavi Kishor, Genome-wide identification and transcriptional profiling of small heat shock protein gene family under diverse abiotic stress conditions in Sorghum bicolor (L.), Int. J. Biol. Macromol. 142 (2020). https://doi.org/10.1016/j.ijbiomac.2019.10.023.

[110] Y. Fang, Q. Du, Q. Yang, J. Jiang, X. Hou, Z. Yang, D. Zhao, X. Li, X. Xie, Identification, characterization, and expression profiling of the putative U-box E3 ubiquitin ligase gene family in

Sorghum bicolor, Front. Microbiol. 13 (2022). https://doi.org/10.3389/fmicb.2022.942302.

[111] J. Cui, G. Ren, Y. Bai, Y. Gao, P. Yang, J. Chang, Genome-wide identification and expression analysis of the U-box E3 ubiquitin ligase gene family related to salt tolerance in sorghum (Sorghum bicolor L.), Front. Plant Sci. 14 (2023). https://doi.org/10.3389/fpls.2023.1141617.

[112] H. Cui, J. Chen, M. Liu, H. Zhang, S. Zhang, D. Liu, S. Chen, Genome-Wide Analysis of C2H2 Zinc Finger Gene Family and Its Response to Cold and Drought Stress in Sorghum [Sorghum bicolor (L.) Moench], Int. J. Mol. Sci. 23 (2022). https://doi.org/10.3390/ijms23105571.

[113] Z. Xia, Y. Huo, Y. Wei, Q. Chen, Z. Xu, W. Zhang, The arabidopsis LYST INTERACTING PROTEIN 5 acts in regulating abscisic acid signaling and drought response, Front. Plant Sci. 7 (2016). https://doi.org/10.3389/fpls.2016.00758.

[114] J. Tang, D.C. Bassham, Autophagy during drought: function, regulation, and potential application, Plant J. 109 (2022). https://doi.org/10.1111/tpj.15481.

[115] M. Yang, L. Wang, C. Chen, X. Guo, C. Lin, W. Huang, L. Chen, Genome-wide analysis of autophagy-related genes in Medicago truncatula highlights their roles in seed development and response to drought stress, Sci. Rep. 11 (2021). https://doi.org/10.1038/s41598-021-02239-6.

[116] D. Kuzuoglu-Ozturk, O.C. Yalcinkaya, B.A. Akpinar, G. Mitou, G. Korkmaz, D. Gozuacik, H. Budak, Autophagy-related gene, TdAtg8, in wild emmer wheat plays a role in drought and osmotic stress response, Planta 236 (2012). https://doi.org/10.1007/s00425-012-1657-3.

[117] H. Sharma, R. Batra, S. Kumar, M. Kumar, S. Kumar, H.S. Balyan, P.K. Gupta, Identification and characterization of 20S proteasome genes and their relevance to heat/drought tolerance in bread wheat, Gene Reports 27 (2022). https://doi.org/10.1016/j.genrep.2022.101552.

[118] S.L. Stone, The role of ubiquitin and the 26S proteasome in plant abiotic stress signaling, Front. Plant Sci. 5 (2014). https://doi.org/10.3389/fpls.2014.00135.

[119] L. Wang, L. Liu, Y. Ma, S. Li, S. Dong, W. Zu, Transcriptome profilling analysis characterized the gene expression patterns responded to combined drought and heat stresses in soybean, Comput. Biol. Chem. 77 (2018). https://doi.org/10.1016/j.compbiolchem.2018.09.012.

[120] S. Sazegari, Z. Zinati, A. Tahmasebi, Dynamic transcriptomic analysis uncovers key genes and mechanisms involved in seed priming-induced tolerance to drought in barley, Gene Reports 21 (2020). https://doi.org/10.1016/j.genrep.2020.100941.

[121] B.G. Mun, S.U. Lee, E.J. Park, H.H. Kim, A. Hussain, Q.M. Imran, I.J. Lee, B.W. Yun, Analysis of

transcription factors among differentially expressed genes induced by drought stress in Populus davidiana, 3 Biotech 7 (2017). https://doi.org/10.1007/s13205-017-0858-7.

[122] K. Satoh, K. Doi, T. Nagata, A. Hosaka, K. Suzuki, X. Ji, M. Raveendran, H. Leung, J. Bennett, S. Kikuchi, Gene expression analysis and data mining from microarray analysis applied to drought stress in rice, in: Drought Front. Rice Crop Improv. Increased Rainfed Prod., 2009. https://doi.org/10.1142/9789814280013_0015.

[123] M. Manna, T. Thakur, O. Chirom, R. Mandlik, R. Deshmukh, P. Salvi, Transcription factors as key molecular target to strengthen the drought stress tolerance in plants, Physiol. Plant. 172 (2021). https://doi.org/10.1111/ppl.13268.

Appendix S1- Figures

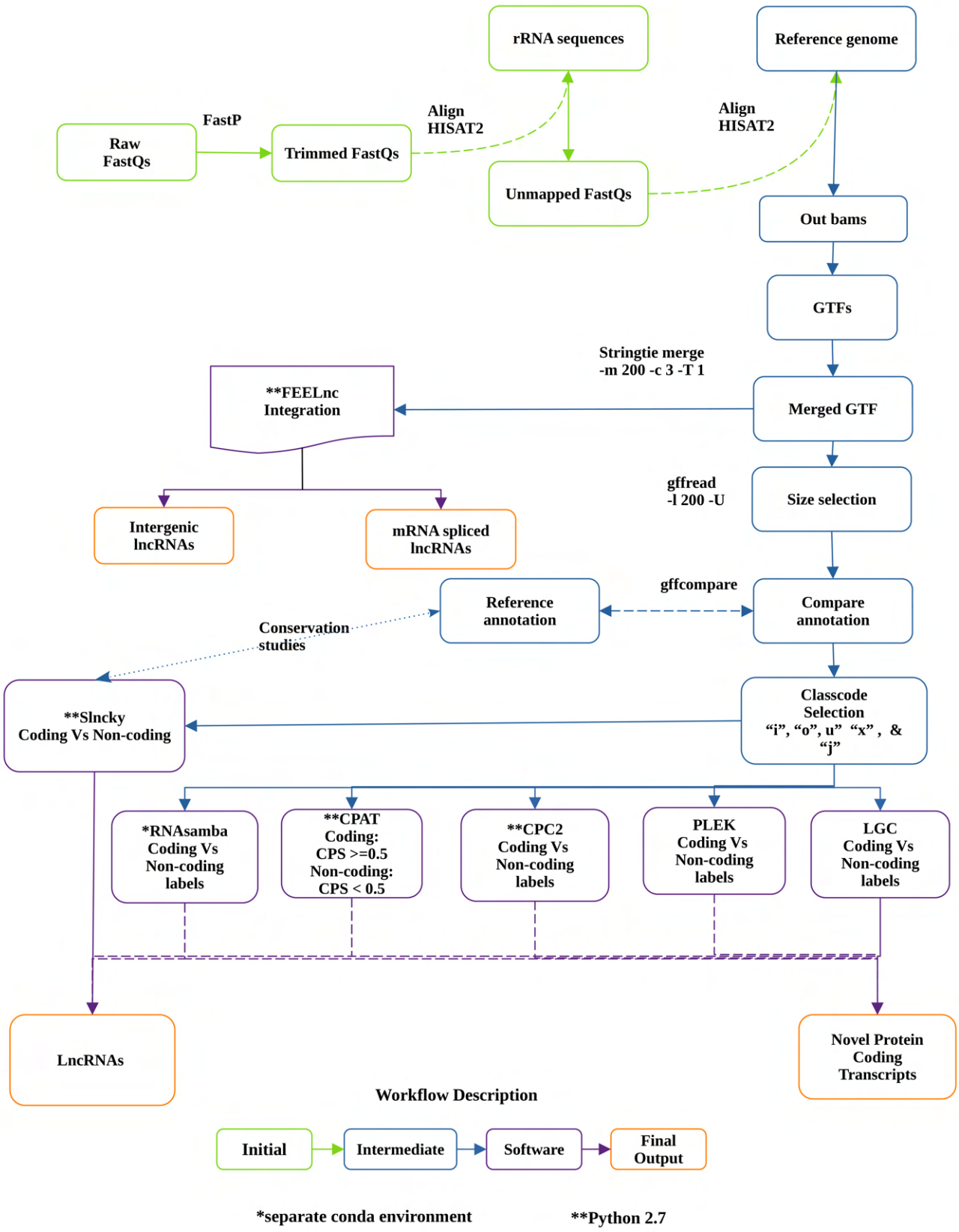


Figure 1(a): LncRAnalyzer workflow showing the integration of FEELnc, CPC2, CPAT, RNAsamba, PLEK, LGC, and Slncky tools in the automated pipeline for identification of lncRNAs and NPCTs using RNA-seq datasets.

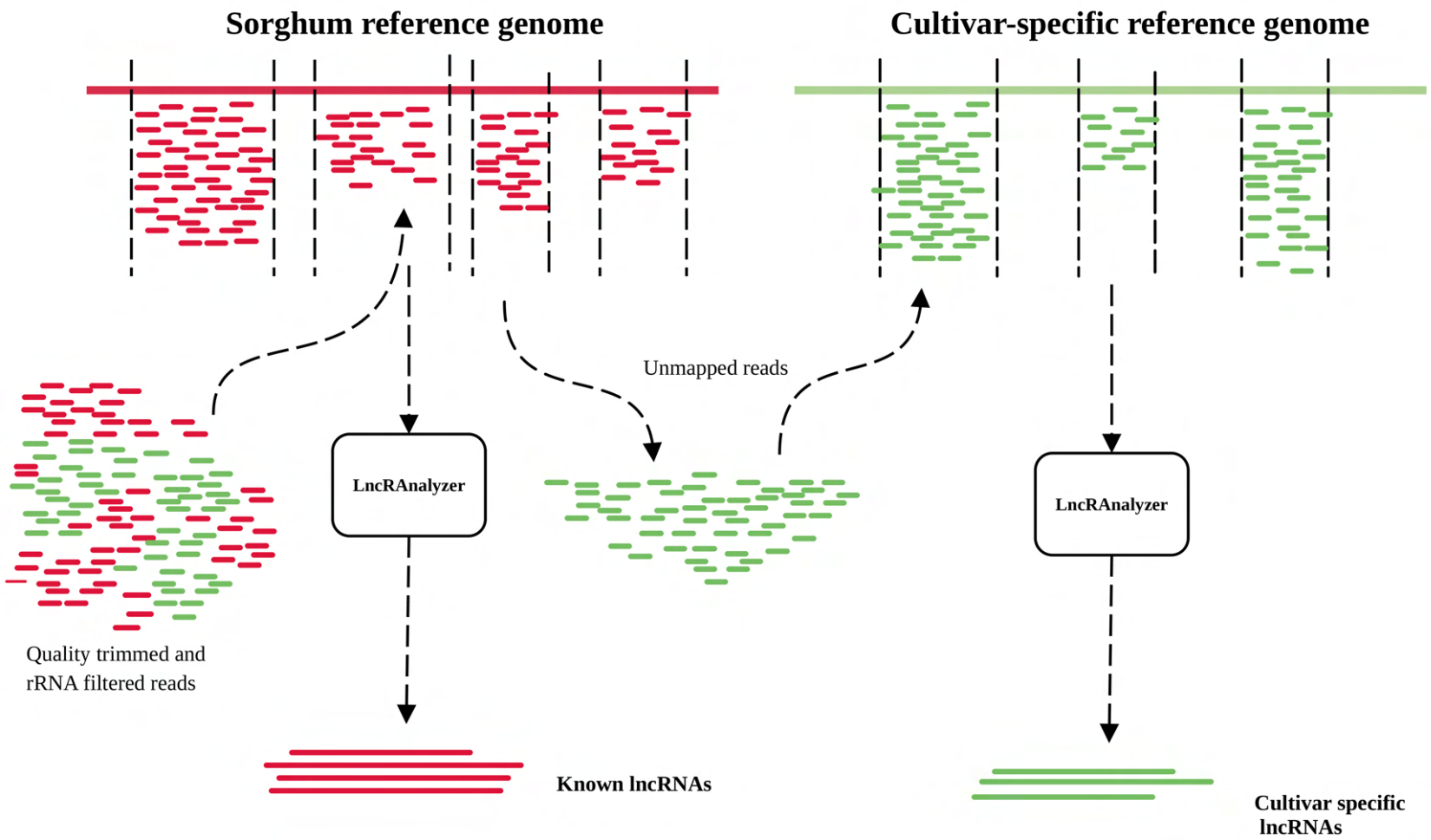


Figure 1(b): LncRAnalyzer implementation for identification of conserved and cultivar-specific lncRNAs using 389 RNA-seq accessions of BTX642 and RTX430.

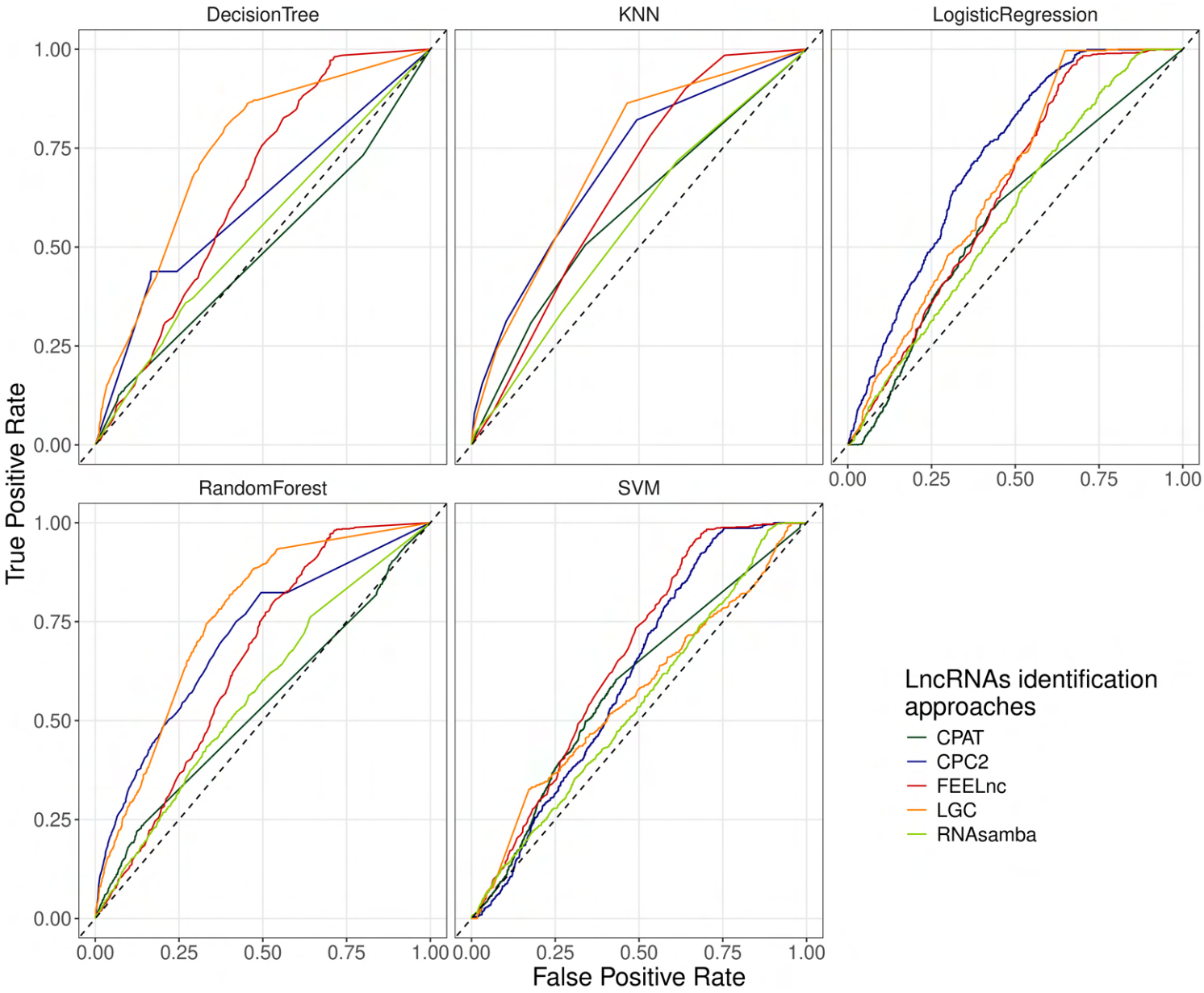


Figure 1(c): The performance of various tools used by LncRAnalyzer using the machine learning trained models for BTX642 RNA-seq datasets, reveals that CPC2 and FEELnc give more accurate predictions of lncRNAs than other methods.

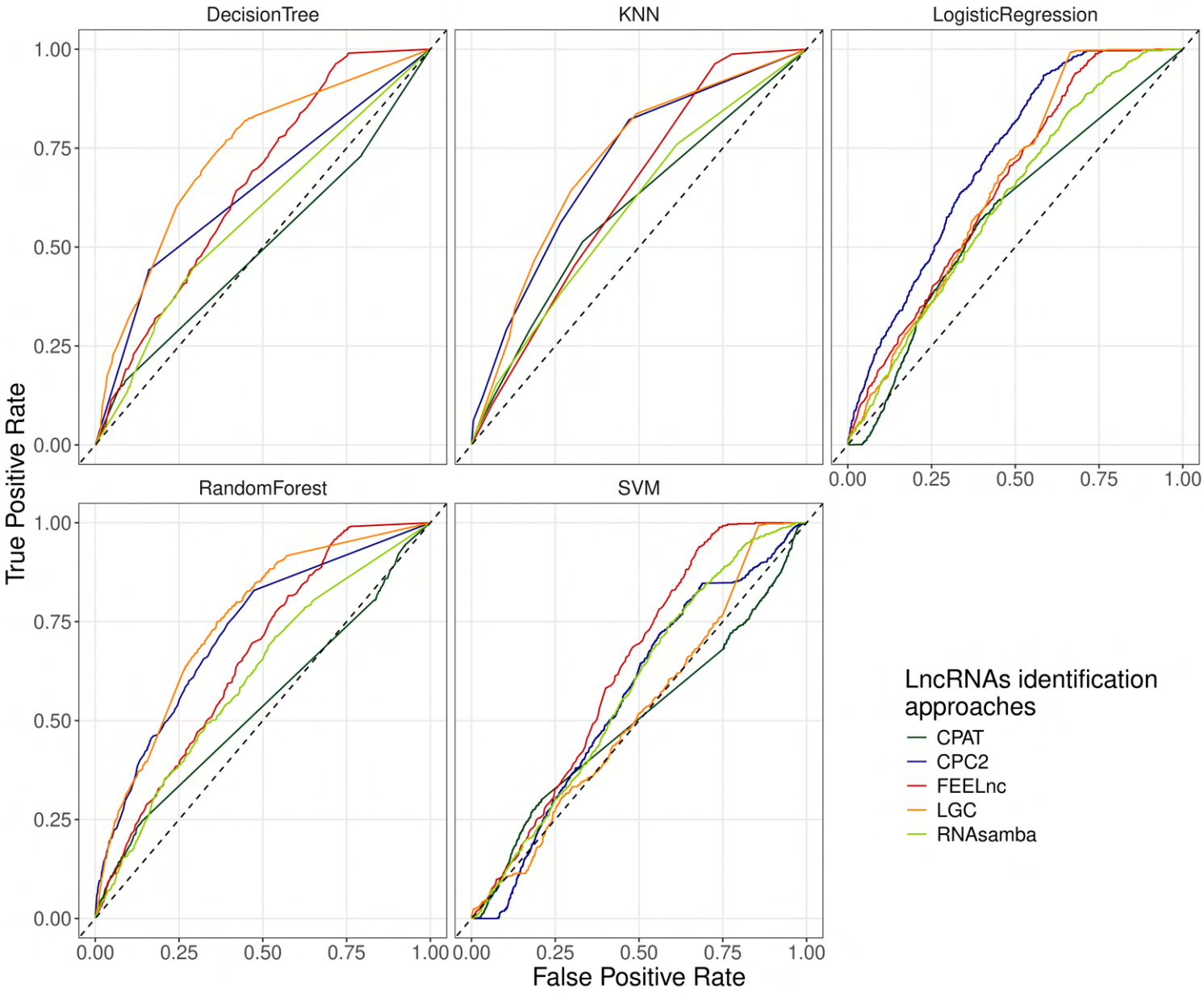


Figure 1(d): The performance of various tools used by LncRAnalyzer using the machine learning trained models for RTX430 RNA-seq datasets, reveals that CPC2 and FEELnc give more accurate predictions of lncRNAs than other methods.

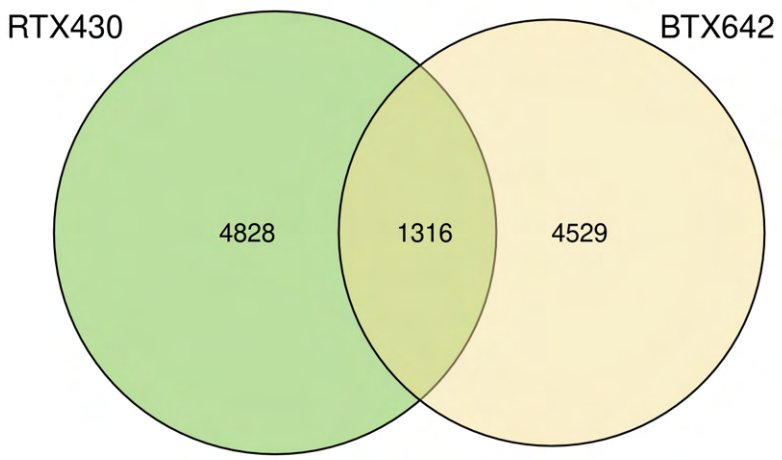


Figure 2(a): LncRNAs conservation analysis between BTX642 and RTX430, reports only 726 lncRNAs were conserved between these two genotypes.

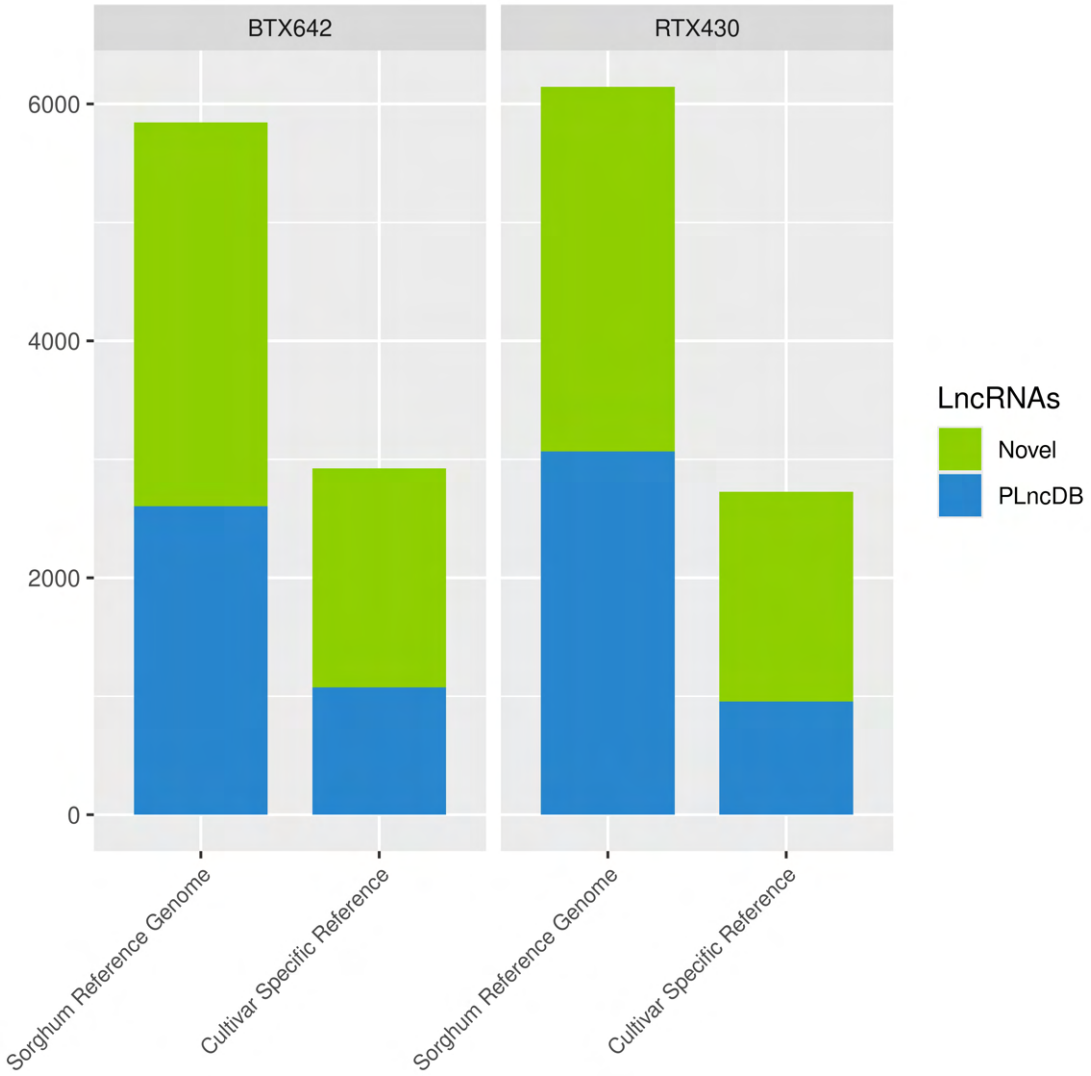


Figure 2(b): Conserved lncRNAs with PLncDB in BTX642 and RTX430 were identified using BLASTN (e-value < 10^{-5}), which reports several novel lncRNAs in these genotypes.

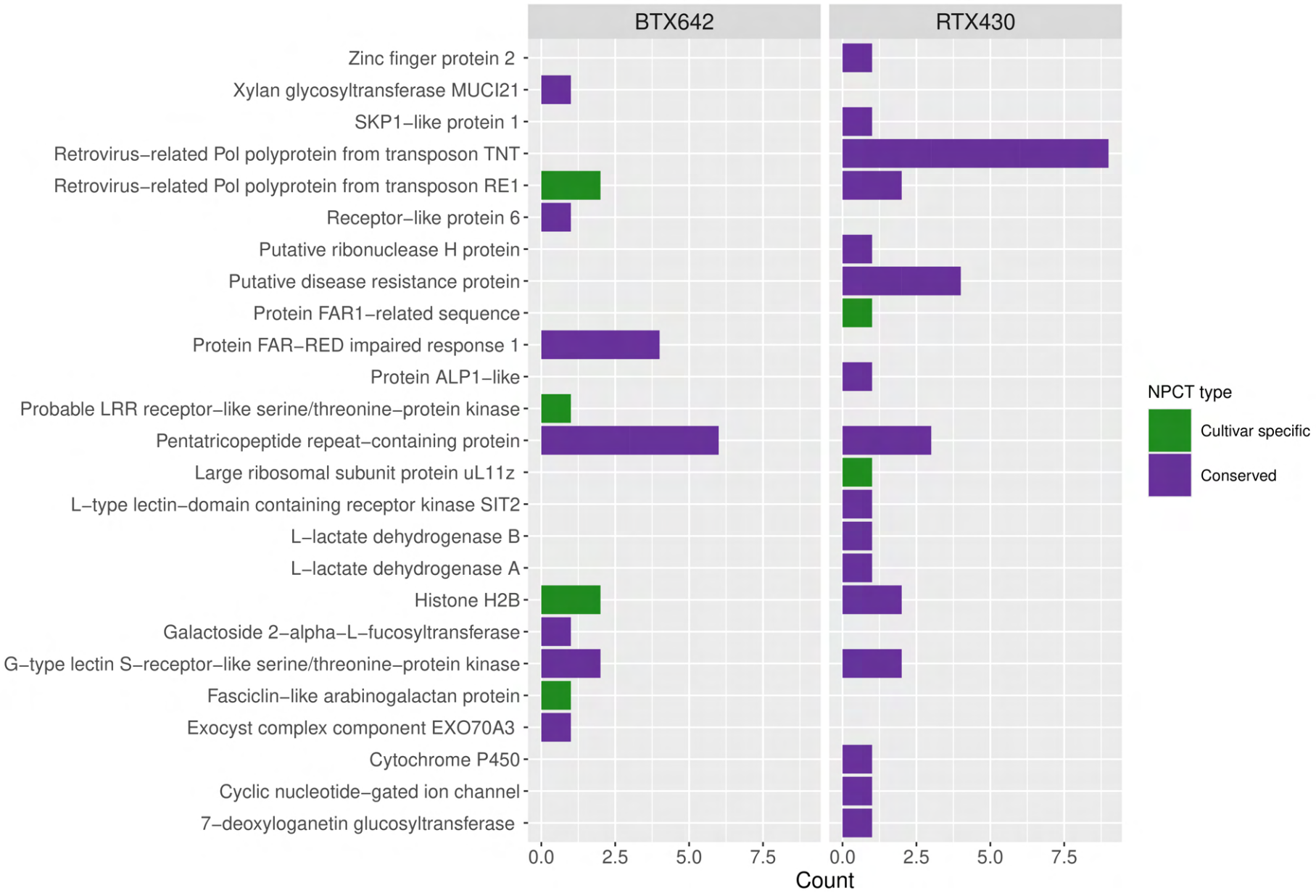


Figure 3: Functional annotations of NPCTs using Trinotate, reports conserved and cultivar-specific NPCTs in BTX642 and RTX430.

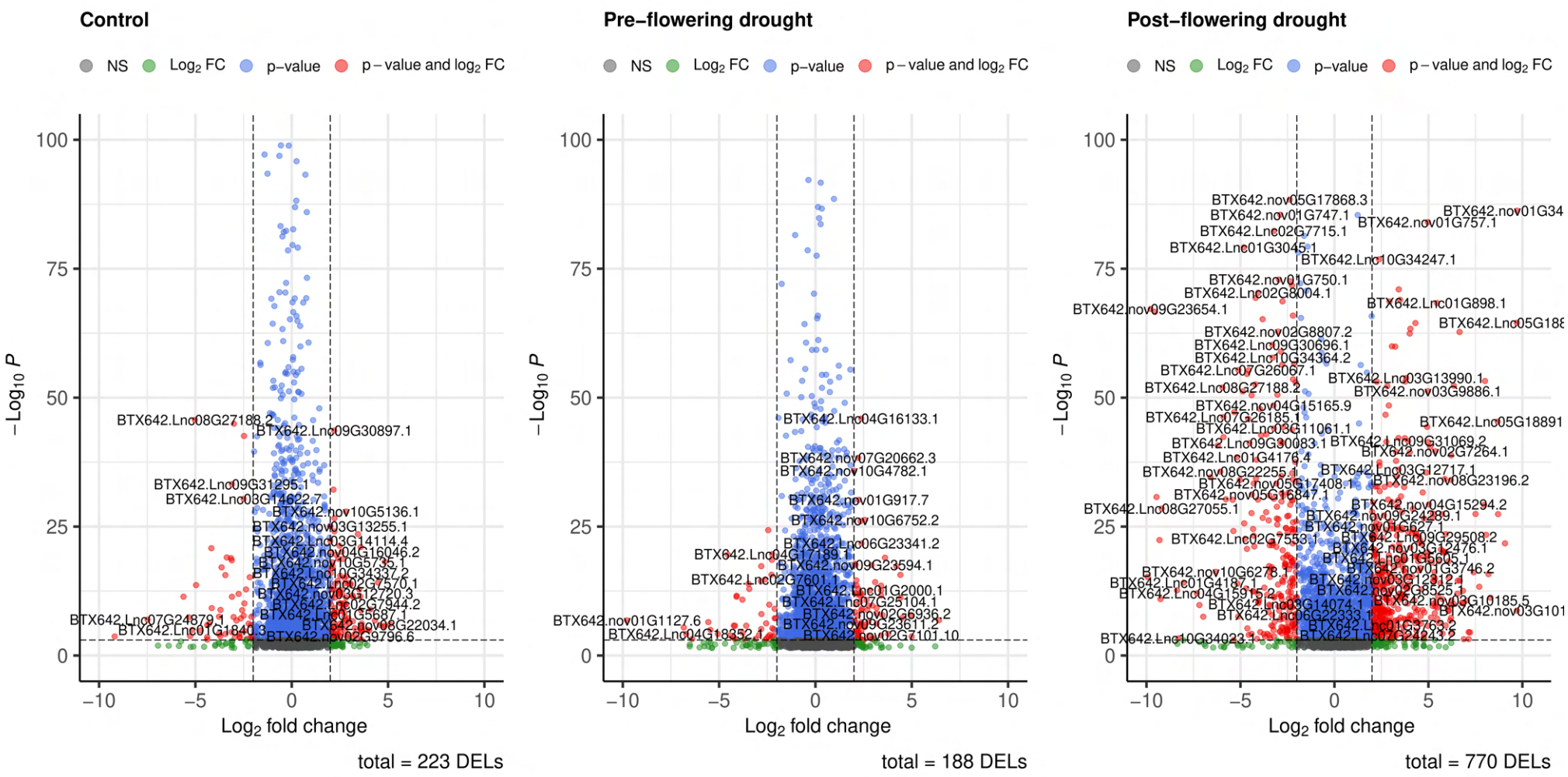


Figure 4(a): Number of Differentially Expressed LncRNAs (DELs) with P-value < 0.05 and |log2fold| > 2 under control, pre- and post-flowering drought conditions in the leaf tissues of BTX642.

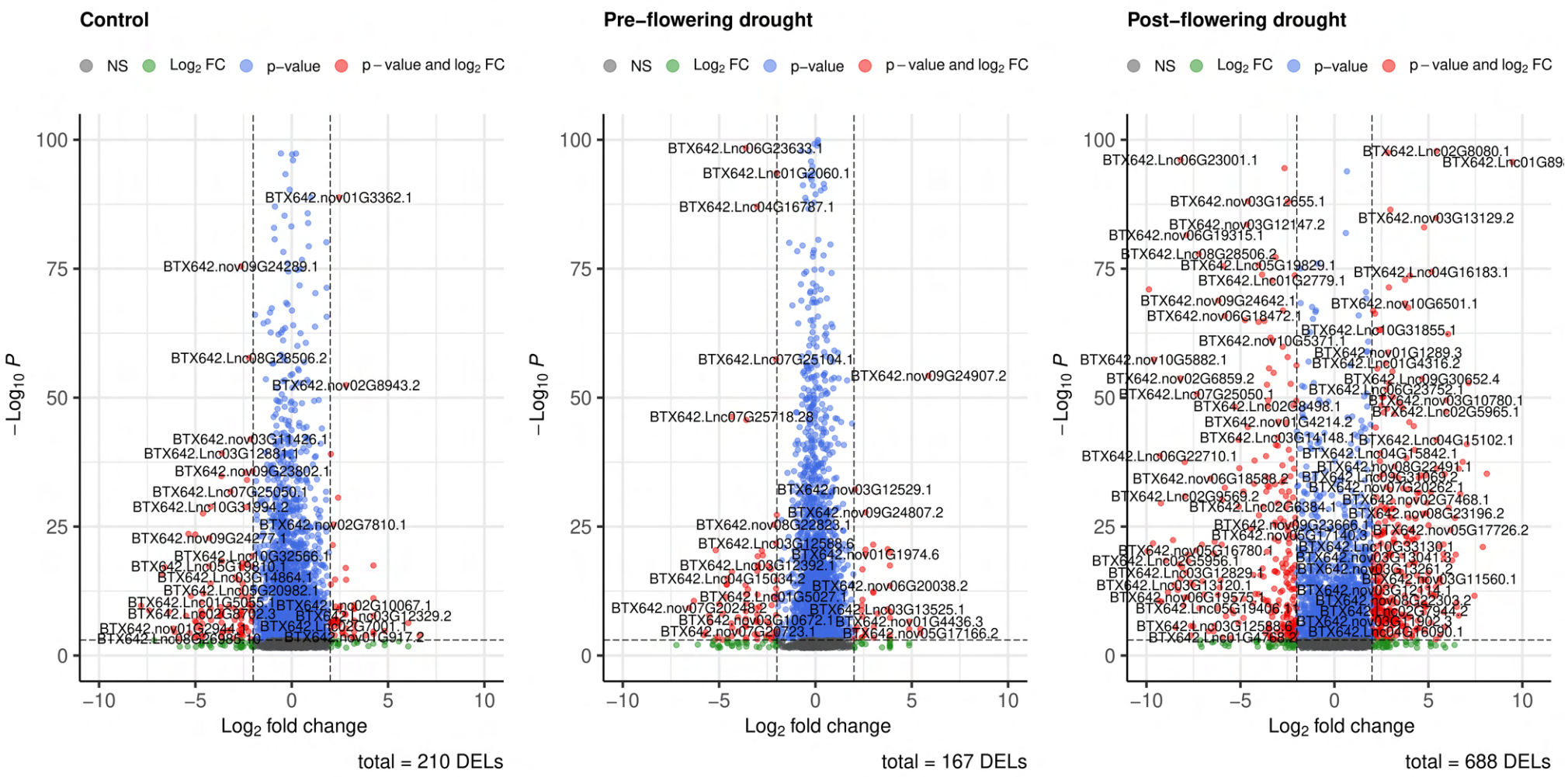


Figure 4(b): Number of Differentially Expressed LncRNAs (DELs) with P-value < 0.05 and |log2fold| > 2 under control, pre- and post-flowering drought conditions in the root tissues of BTX642.

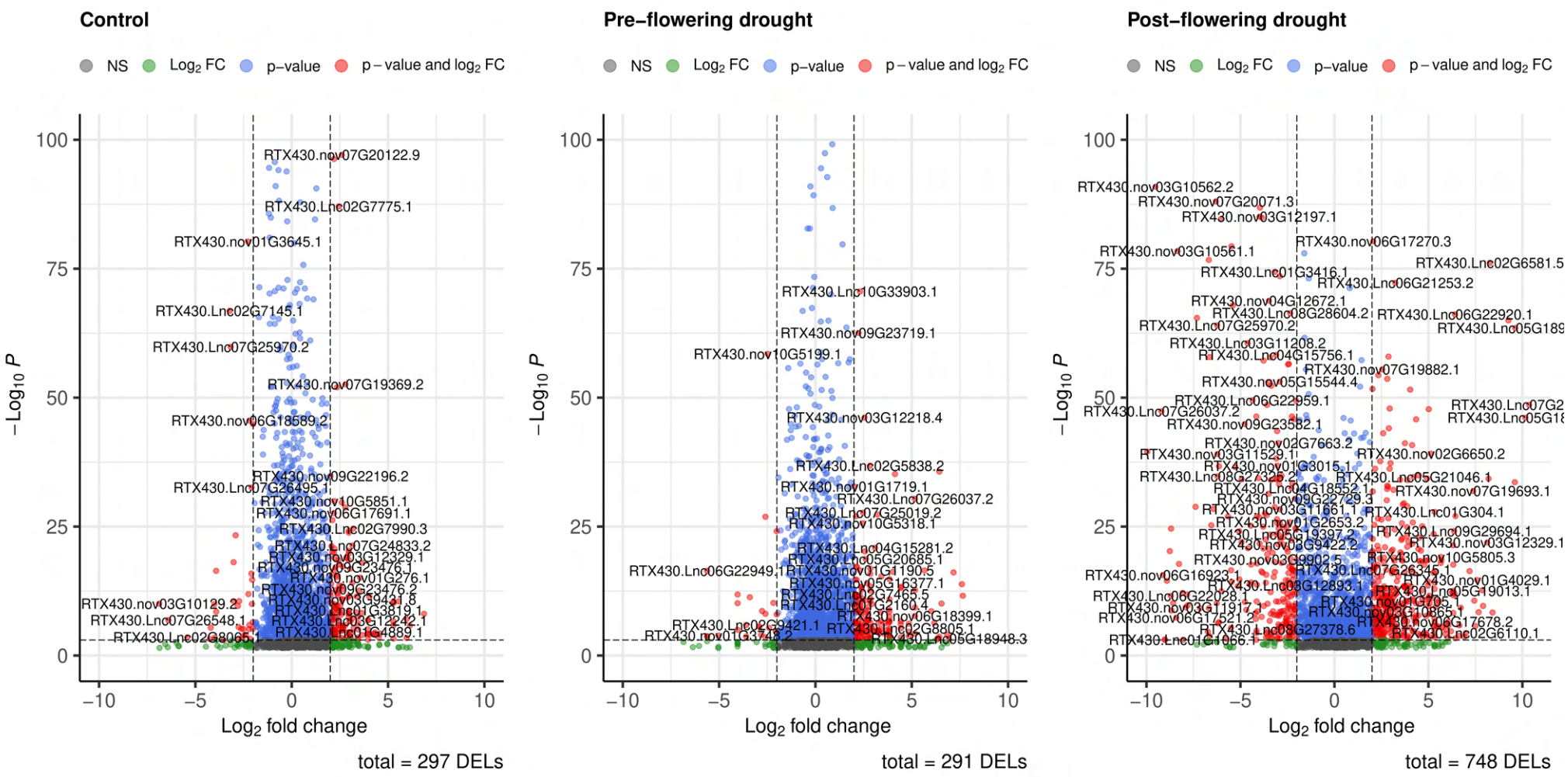


Figure 4(c): Number of Differentially Expressed LncRNAs (DELs) with P-value < 0.05 and |log2fold| > 2 under control, pre- and post-flowering drought conditions in the leaf tissues of RTX430

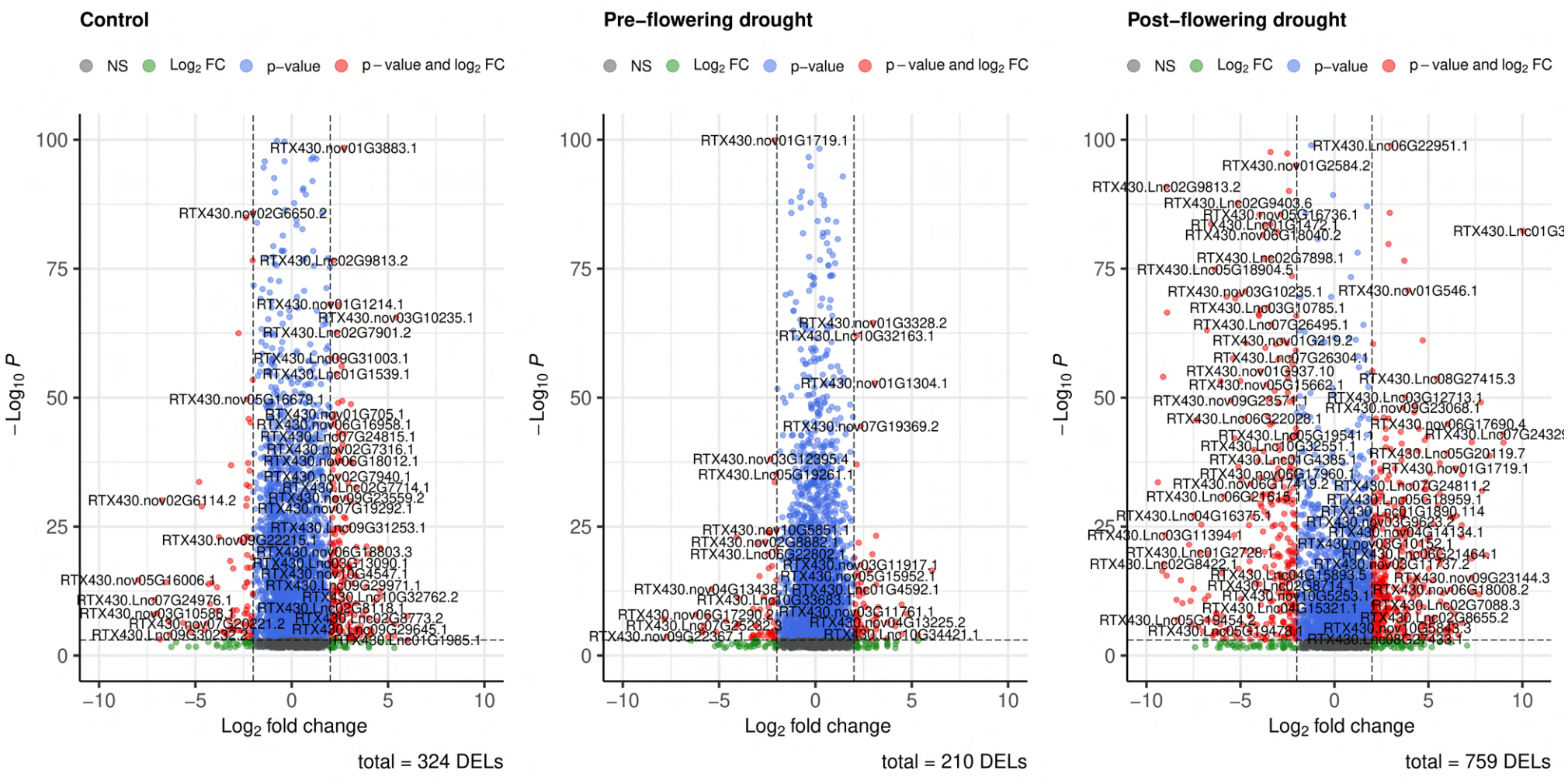


Figure 4(d): Number of Differentially Expressed LncRNAs (DELs) with P-value < 0.05 and |log2fold| > 2 under control, pre- and post-flowering drought conditions in the root tissues of RTX430

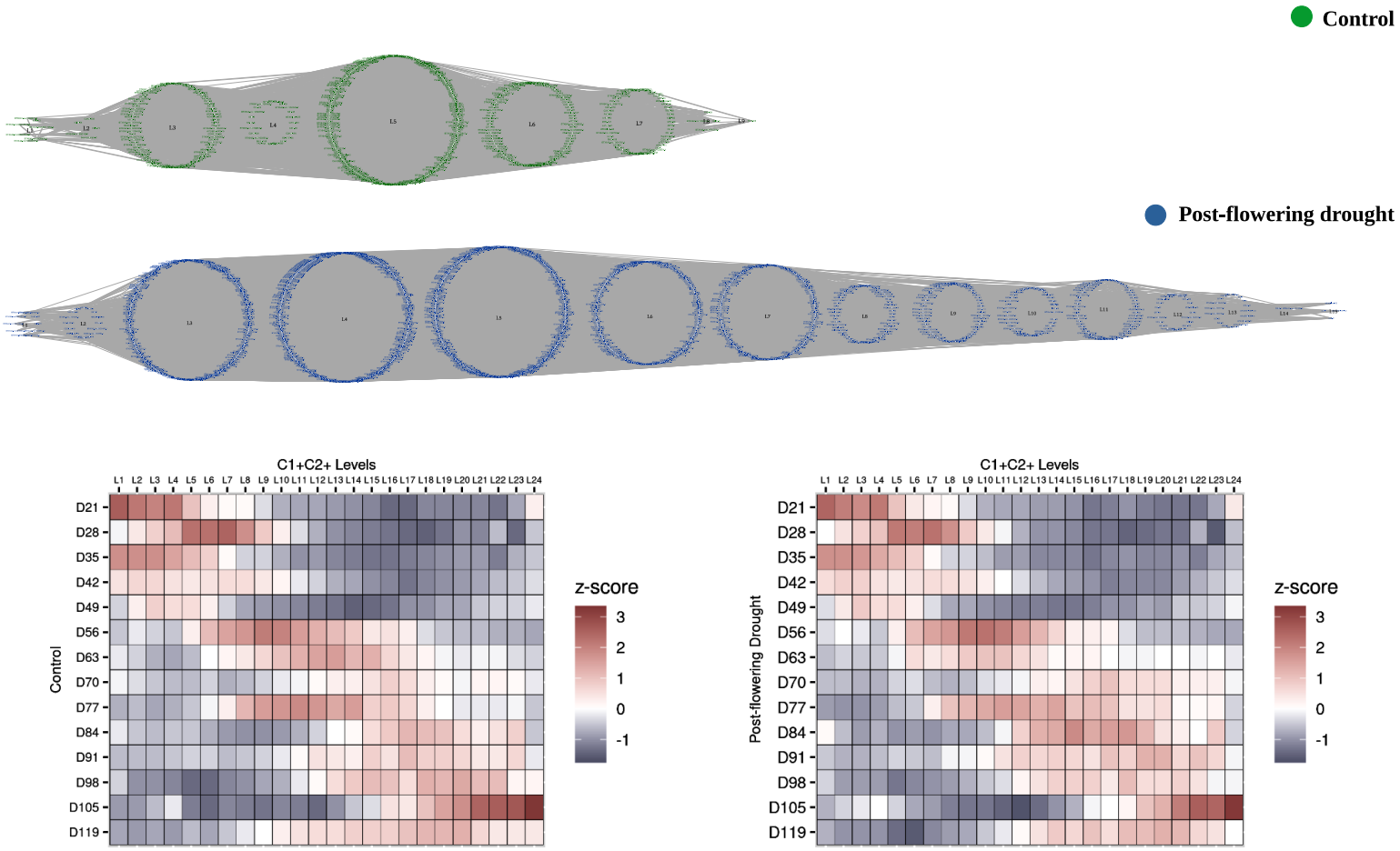


Figure 5(a): Time-ordered gene co-expression network showing co-expressed TFs and lncRNAs in different levels under control and post-flowering drought conditions in BTX642 leaf tissues.

Figure 5(b): Gene co-expression analysis of BTX642 leaf under control and post-flowering drought conditions.

i. Numbers of co-expressed TFs and lncRNAs in TO-GCN levels under control conditions.

ii. Numbers of various TFs co-expressed under control conditions.

iii. Numbers of co-expressed TFs and lncRNAs in TO-GCN levels under post-flowering drought conditions.

iv. Numbers of various TFs co-expressed under post-flowering drought conditions.

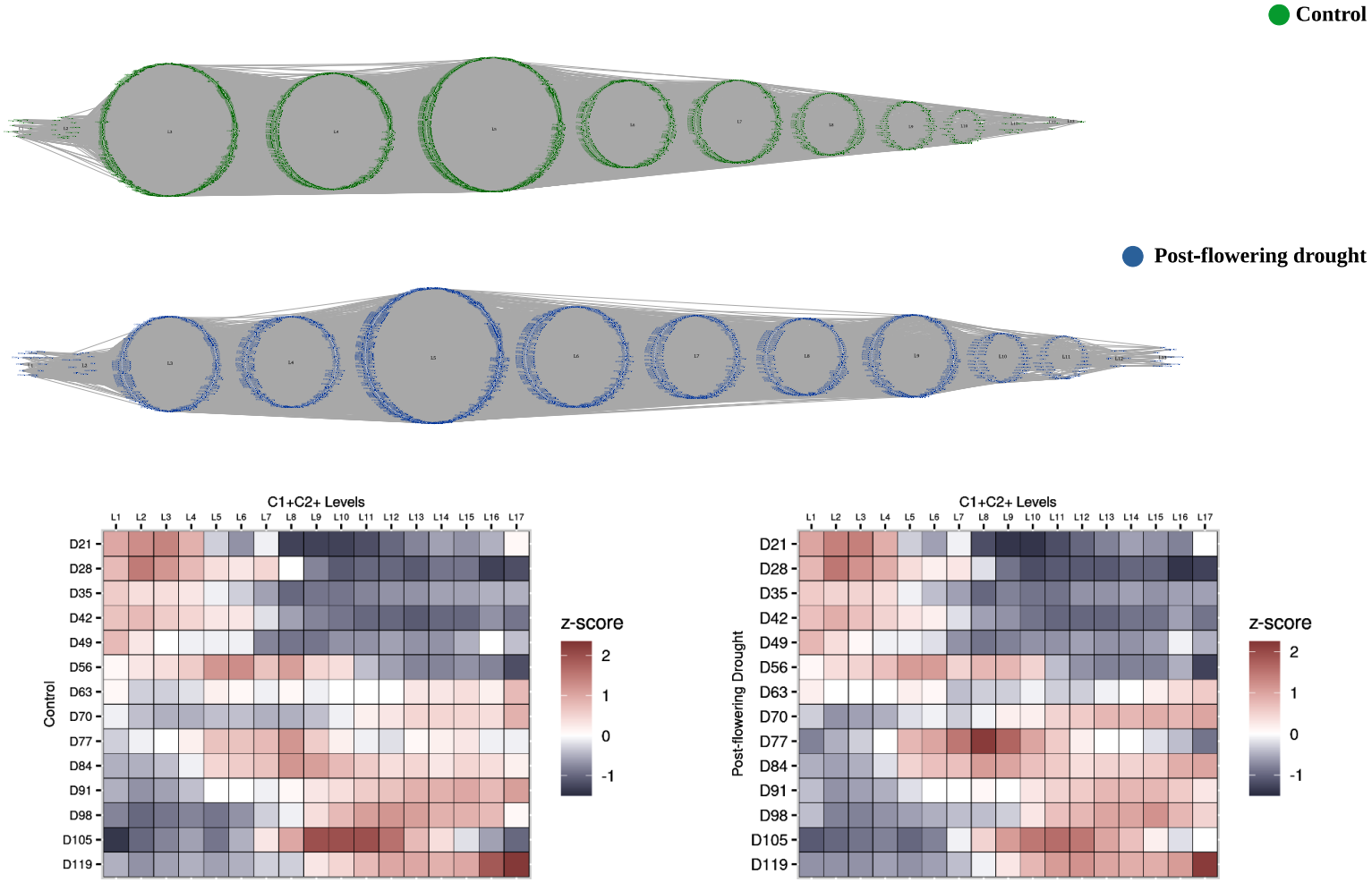


Figure 5(c): Time-ordered gene co-expression network showing co-expressed TFs and lncRNAs in different levels under control and post-flowering drought conditions in BTX642 root tissues.

Figure 5(d): Gene co-expression analysis of BTX642 roots under control and post-flowering drought conditions.

i. Numbers of co-expressed TFs and lncRNAs in TO-GCN levels under control conditions.

ii. Numbers of various TFs co-expressed under control conditions.

iii. Numbers of co-expressed TFs and lncRNAs in TO-GCN levels under post-flowering drought conditions.

iv. Numbers of various TFs co-expressed under post-flowering drought conditions.

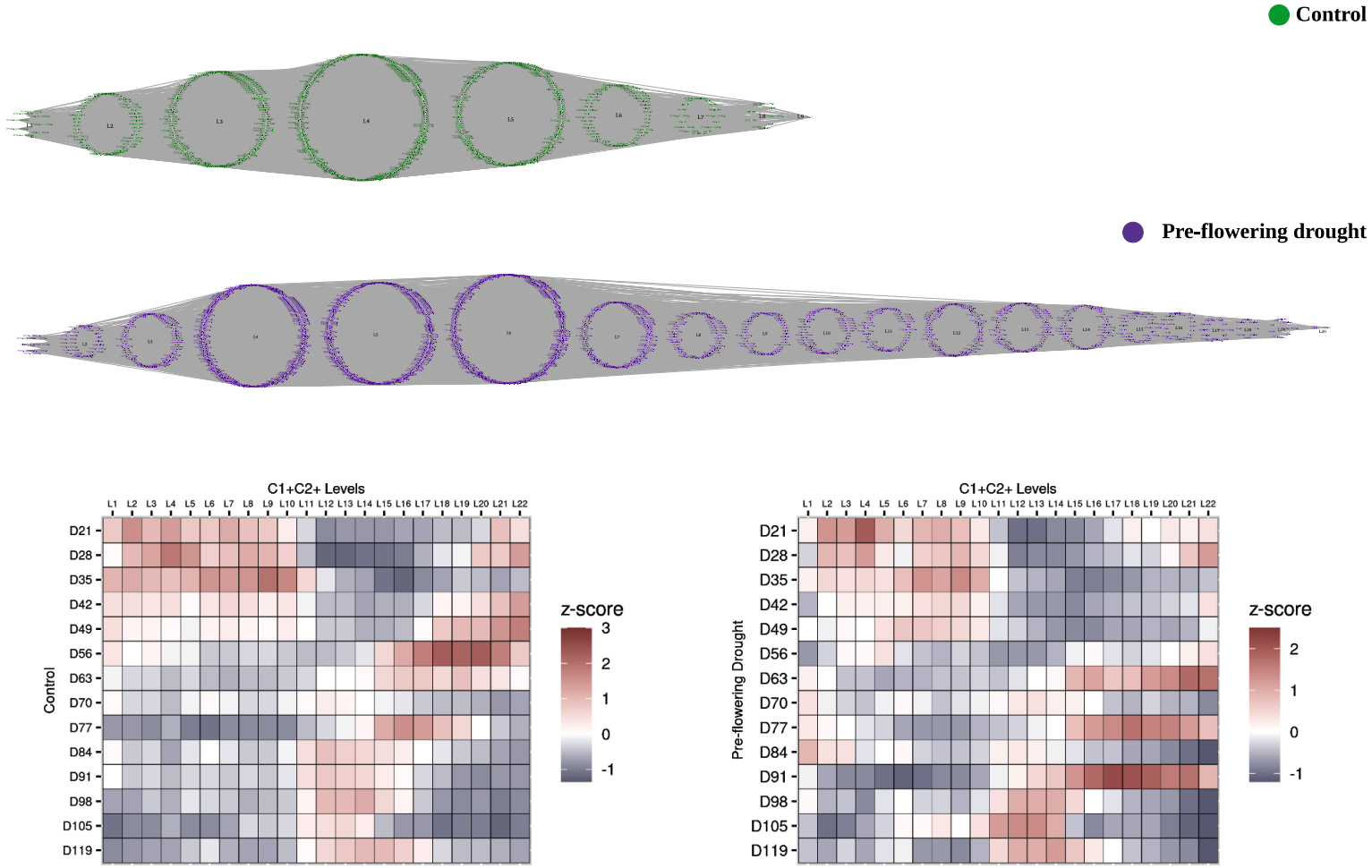


Figure 5(e): Time-ordered gene co-expression network showing co-expressed TFs and lncRNAs in different levels under control and pre-flowering drought conditions in RTX430 leaf tissues.

Figure 5(f): Gene co-expression analysis of RTX430 leaf under control and pre-flowering drought conditions.

i. Numbers of co-expressed TFs and lncRNAs in TO-GCN levels under control conditions.

ii. Numbers of various TFs co-expressed under control conditions.

iii. Numbers of co-expressed TFs and lncRNAs in TO-GCN levels under pre-flowering drought conditions.

iv. Numbers of various TFs co-expressed under pre-flowering drought conditions.

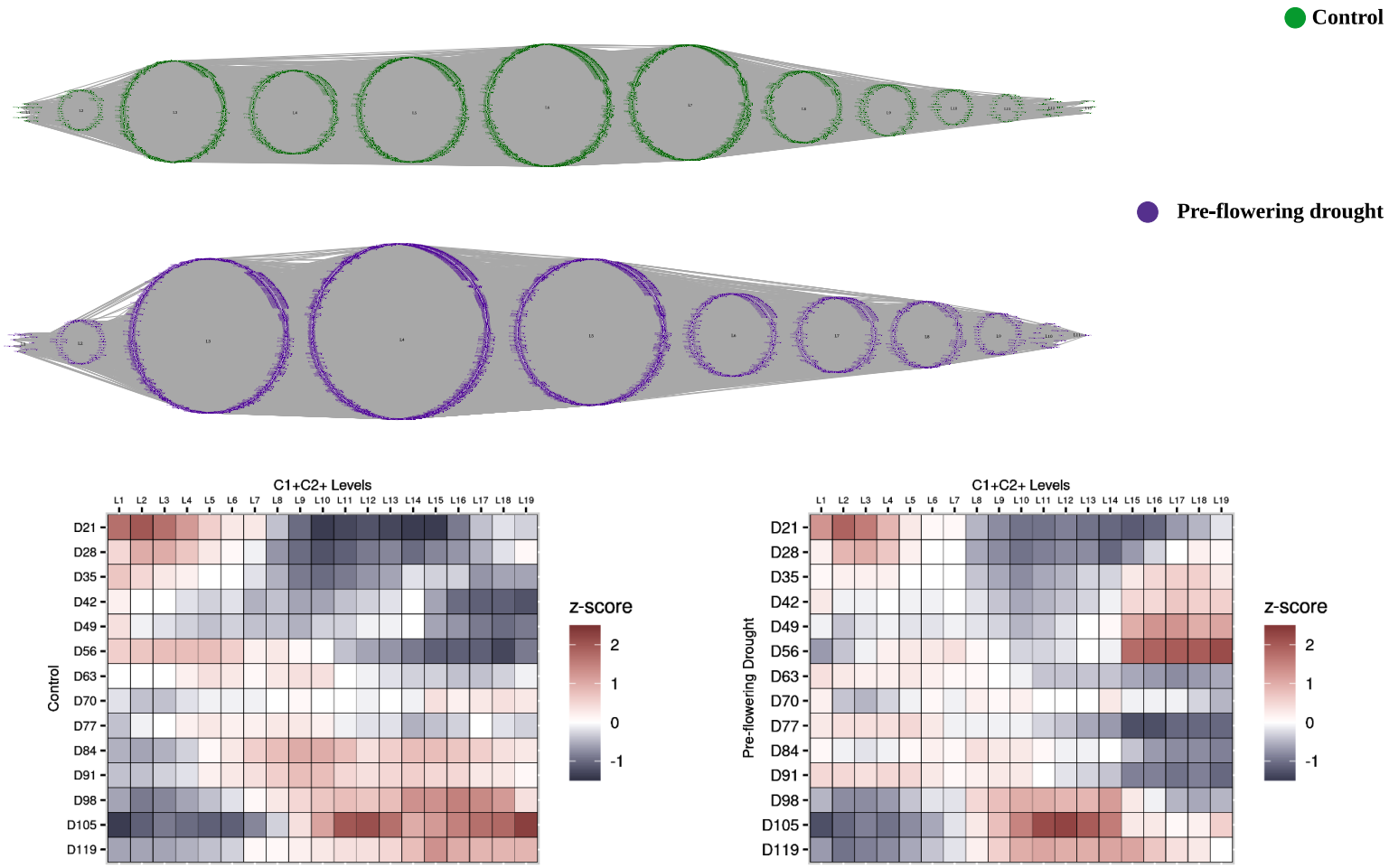


Figure 5(g): Time-ordered gene co-expression network showing co-expressed TFs and lncRNAs in different levels under control and pre-flowering drought conditions in RTX430 root tissues.

Figure 5(h): Gene co-expression analysis of RTX430 roots under control and pre-flowering drought conditions.

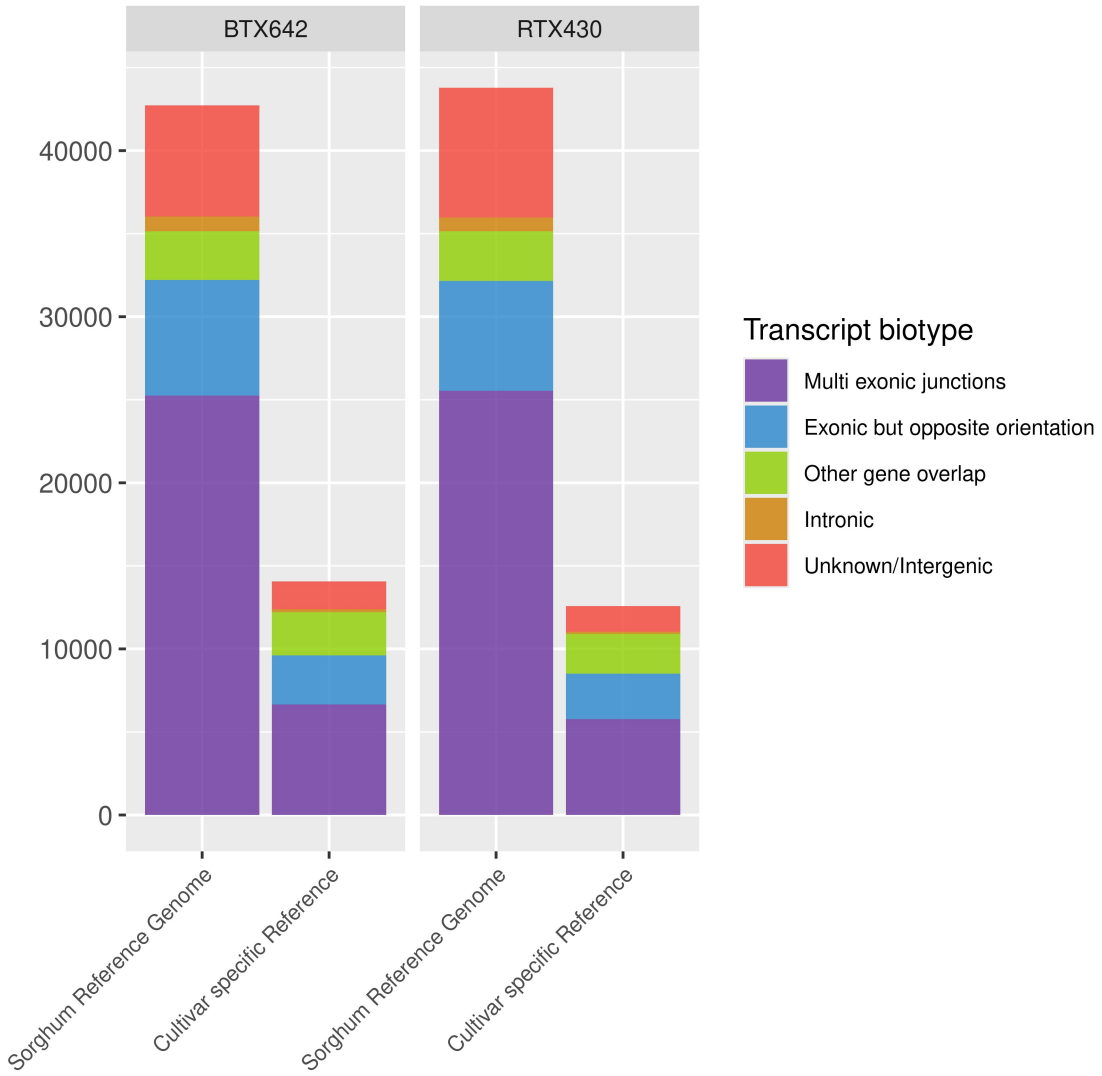
i. Numbers of co-expressed TFs and lncRNAs in TO-GCN levels under control conditions.

ii. Numbers of various TFs co-expressed under control conditions.

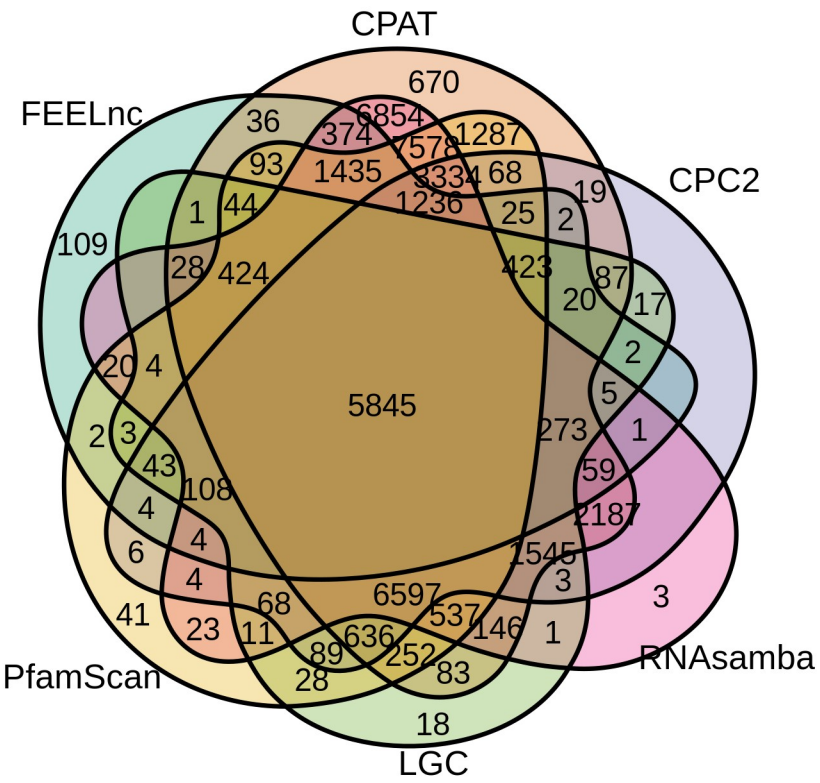
iii. Numbers of co-expressed TFs and lncRNAs in TO-GCN levels under pre-flowering drought conditions.

iv. Numbers of various TFs co-expressed under pre-flowering drought conditions

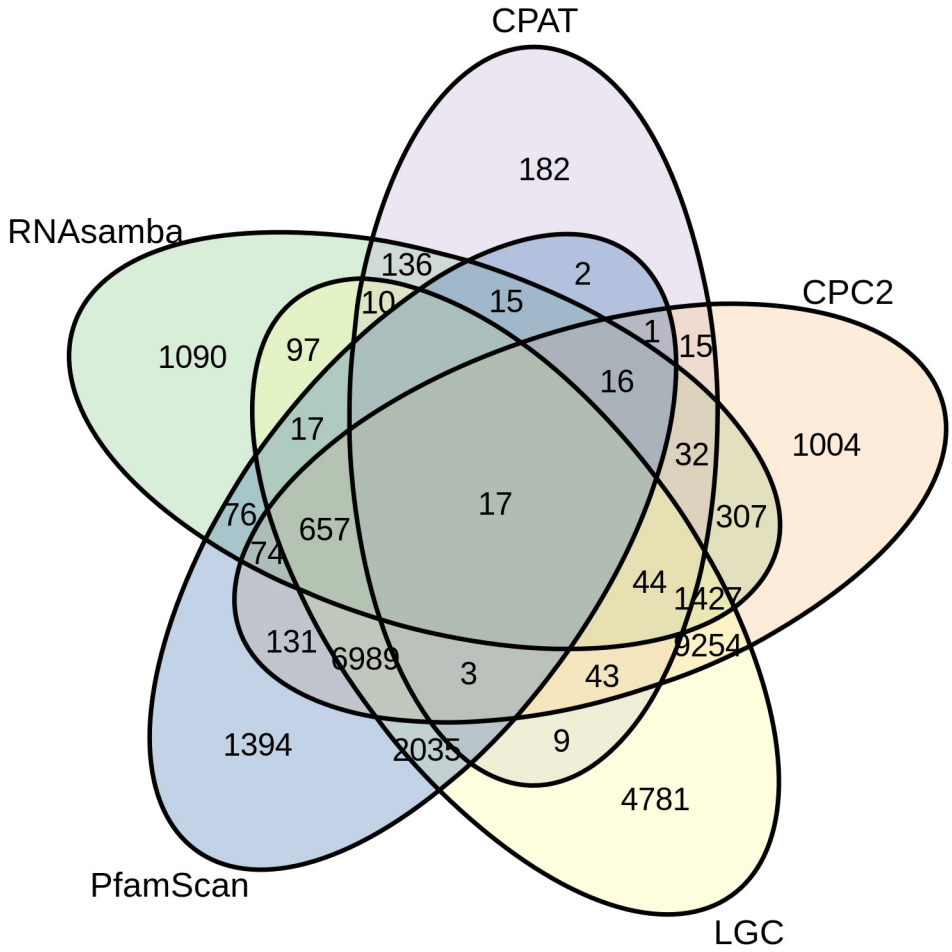
Appendix S1- Supplemental Figures



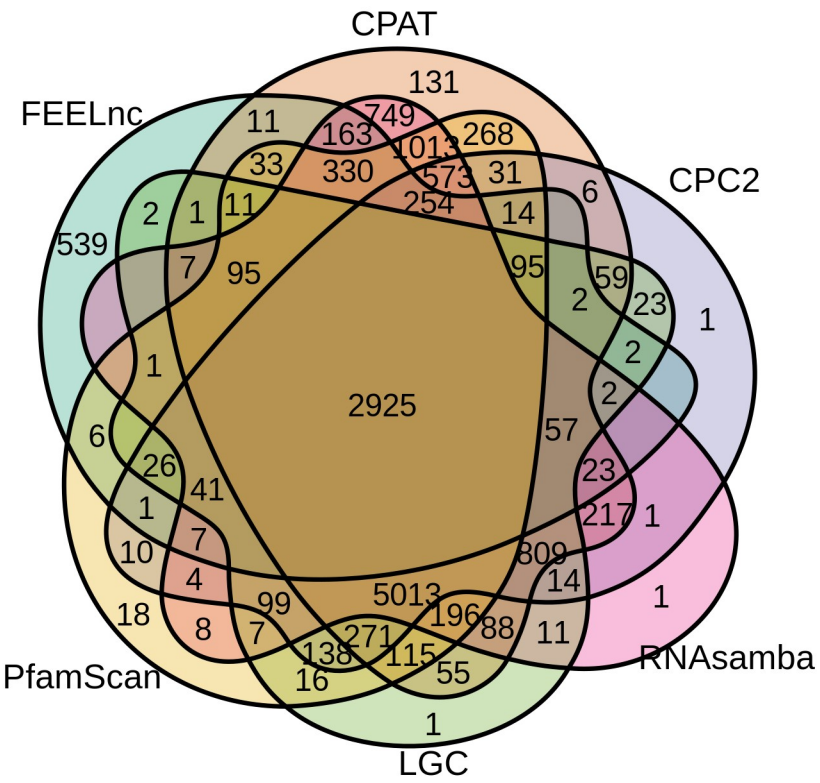
Supplemental Figure S1(a): The number of transcripts belonging to class codes ‘i’, ‘o’, ‘x’, ‘j’, and ‘u’ were selected for the identification of lncRNAs and NPCTs on the sorghum reference genome and the cultivar reference genome.



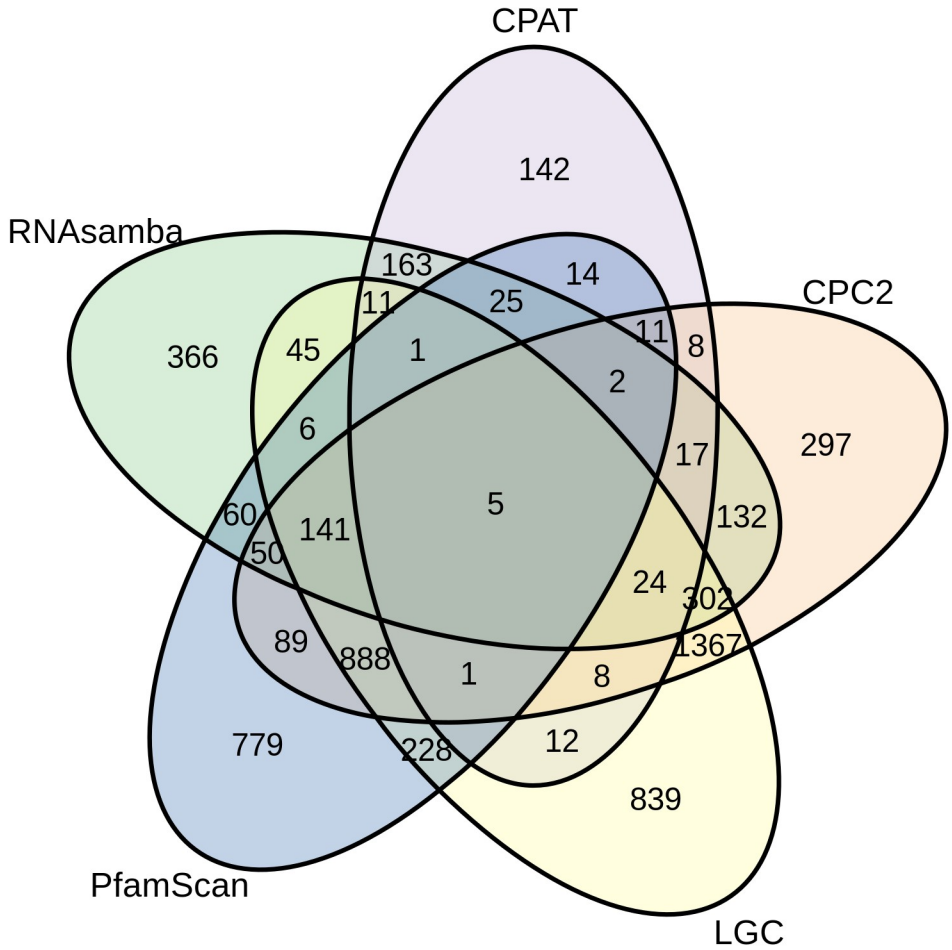
Supplemental Figure S1(b): The numbers of lncRNAs identified by LncRAnalyzer in BTX642 using four methods on the sorghum reference genome (BTX623).



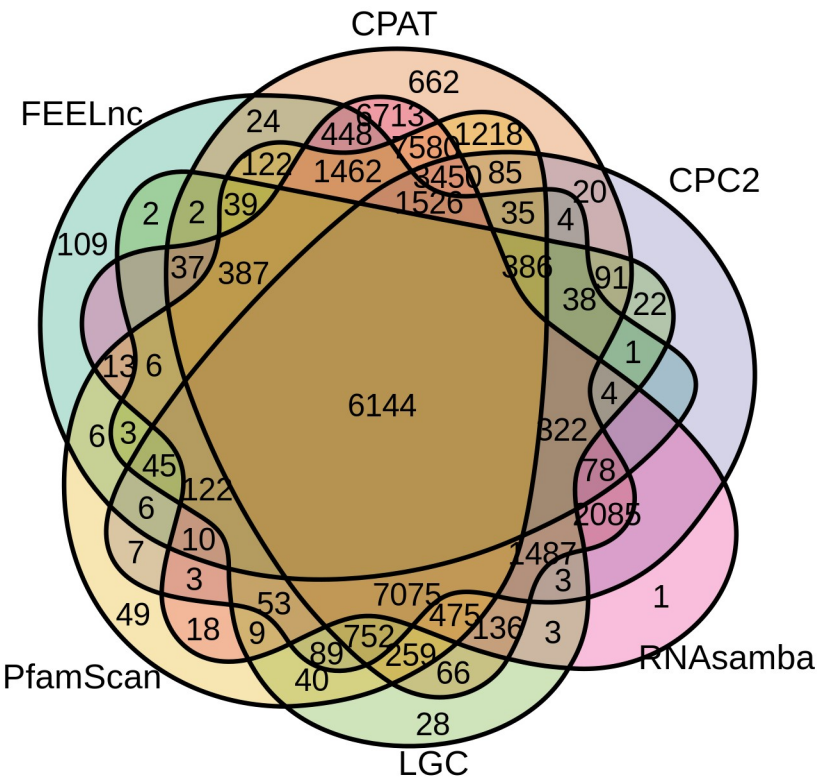
Supplemental Figure S1(c): The numbers of NPCTs identified by LncRAnalyzer in BTX642 using three methods on the sorghum reference genome (BTX623).



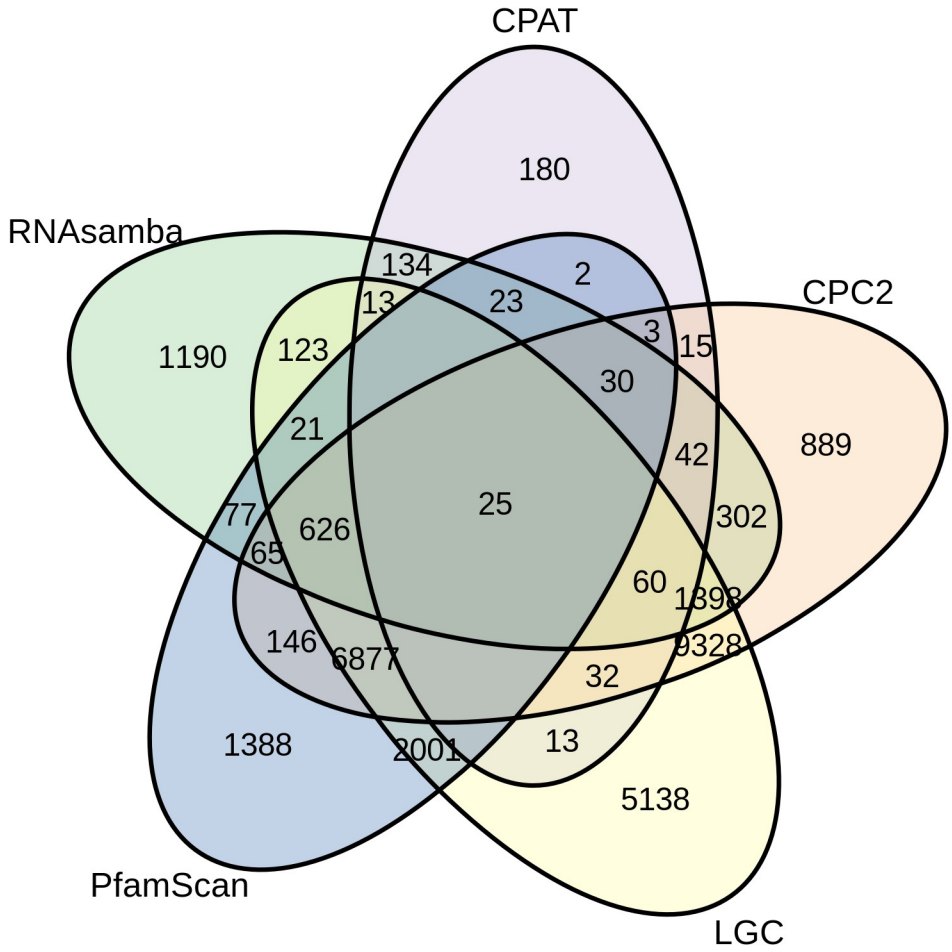
Supplemental Figure S1(d): The numbers of lncRNAs identified by LncRAnalyzer in BTX642 using four methods on the cultivar-specific reference genome.



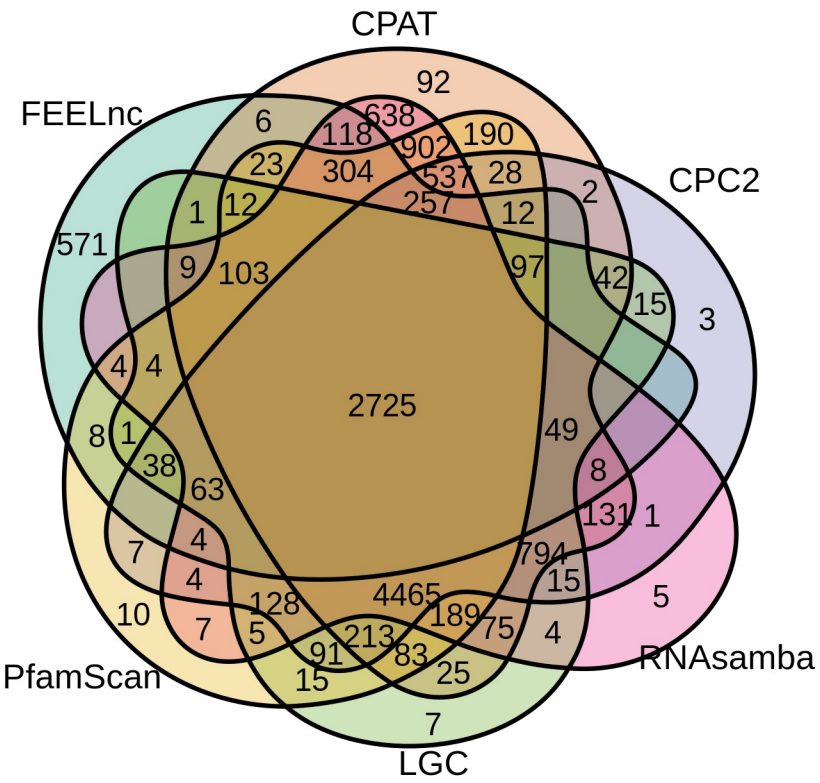
Supplemental Figure S1(e): The numbers of NPCTs identified by LncRAnalyzer in BTX642 using three methods on the cultivar-specific reference genome.



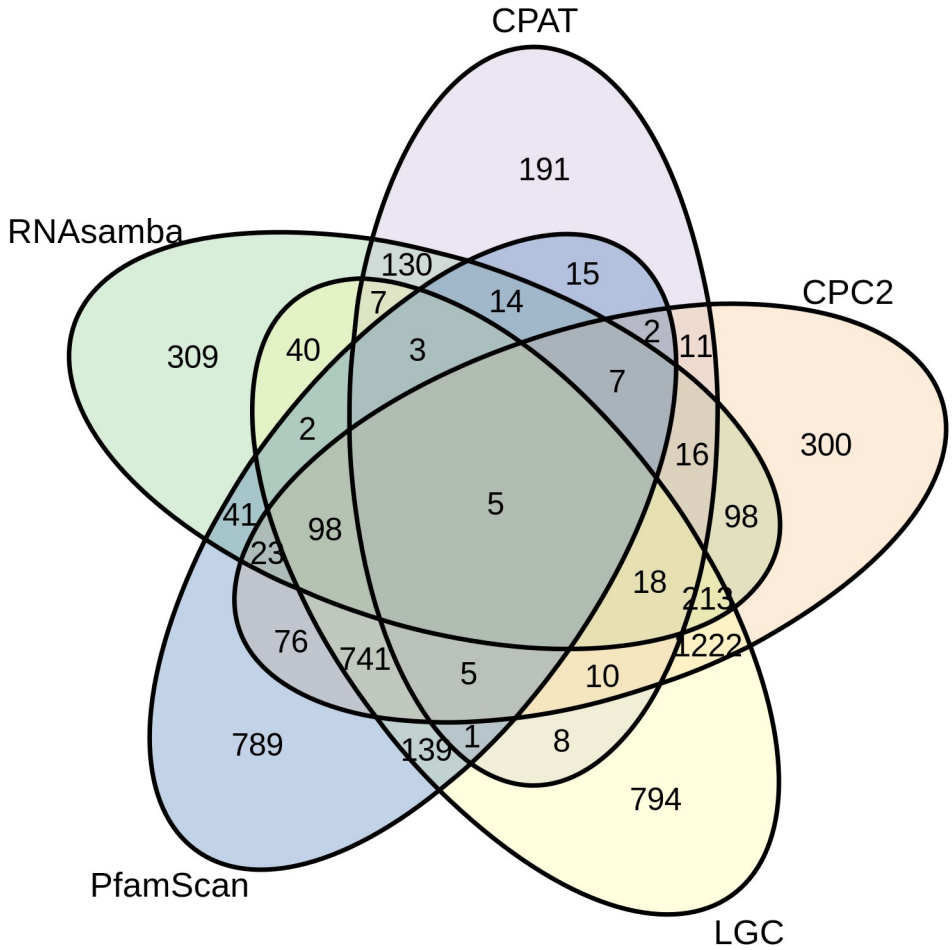
Supplemental Figure S1(f): The numbers of lncRNAs identified by LncRAnalyzer in RTX430 using four methods on the sorghum reference genome (BTX623).



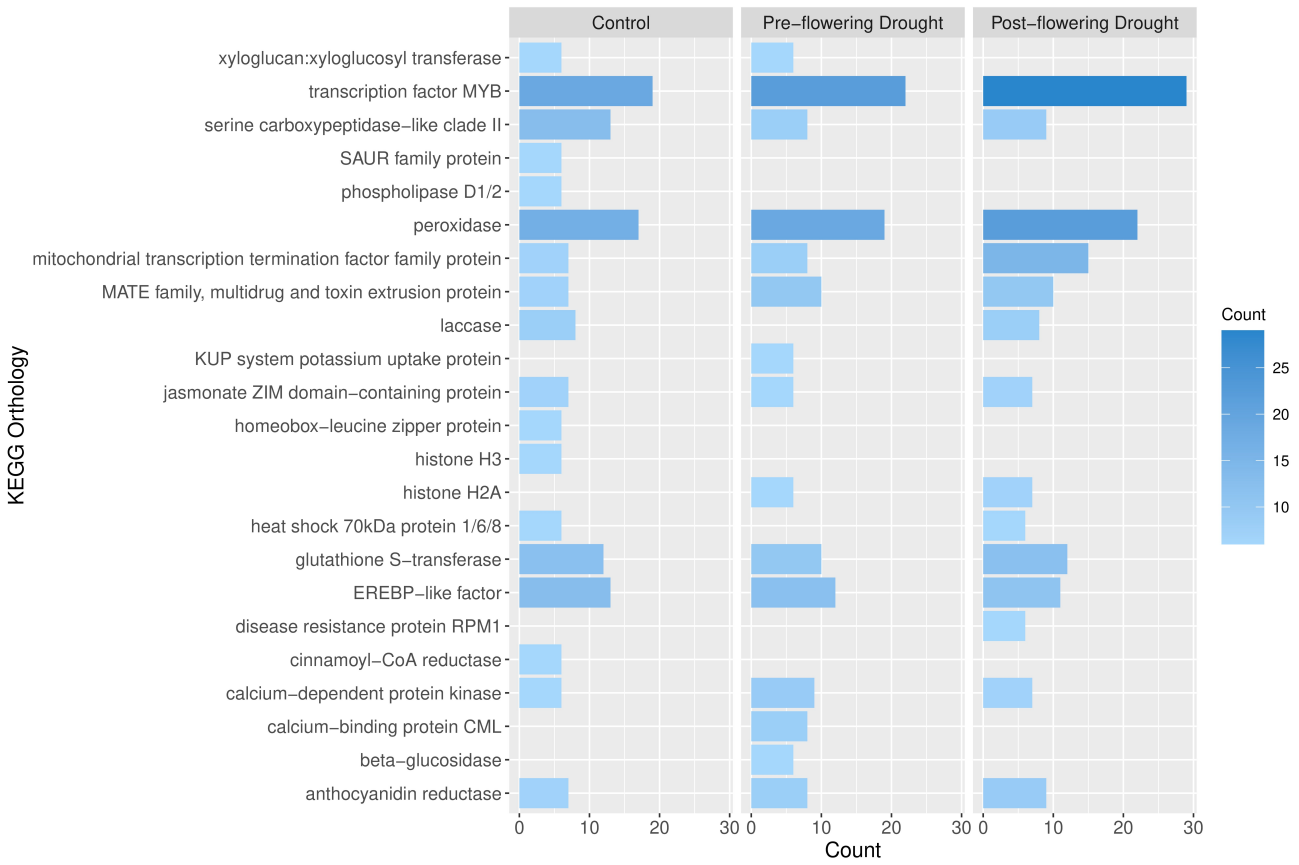
Supplemental Figure S1(g): The numbers of NPCTs identified by LncRAnalyzer in RTX430 using three methods on the sorghum reference genome (BTX623).



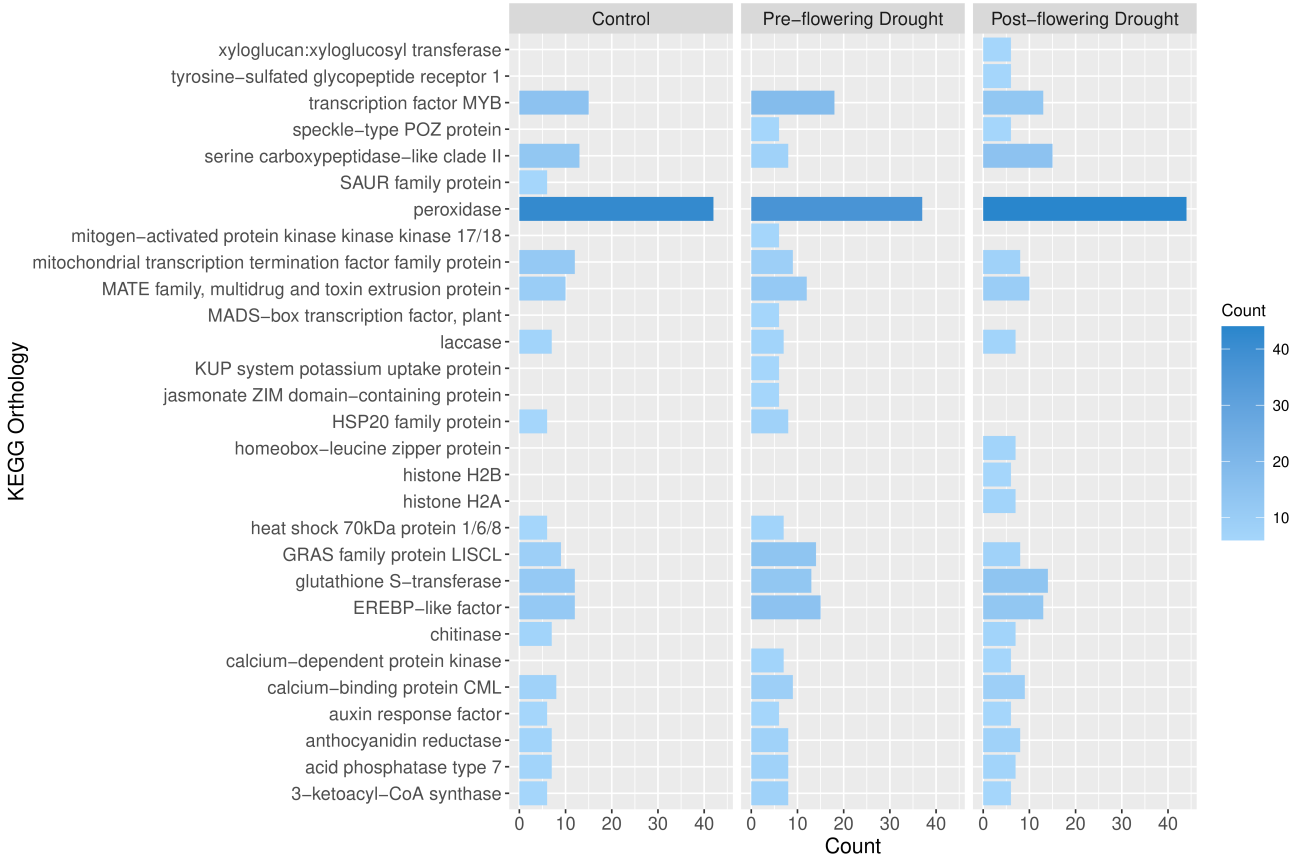
Supplemental Figure S1(h): The numbers of lncRNAs identified by LncRAnalyzer in RTX430 using four methods on the cultivar-specific reference genome.



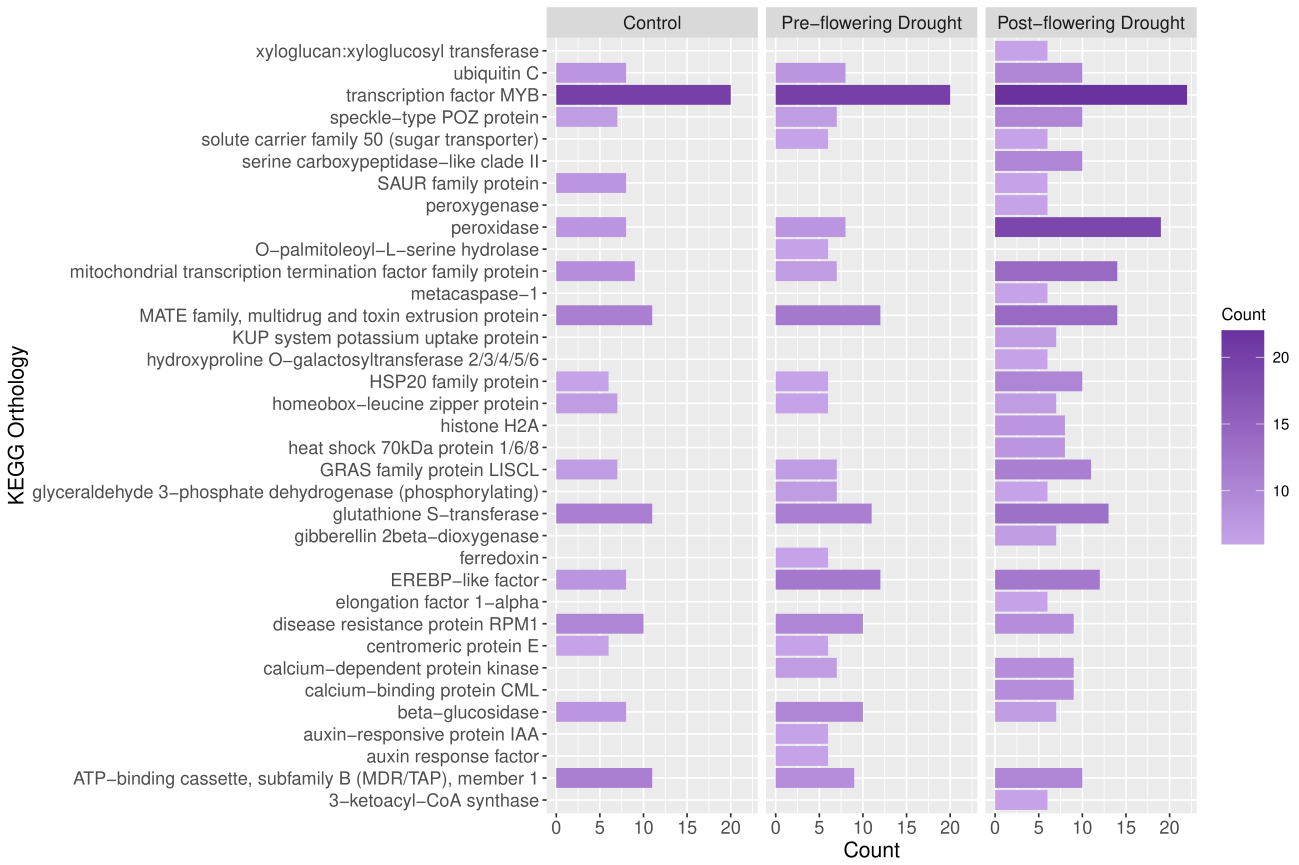
Supplemental Figure S1(i): The numbers of NPCTs identified by LncRAnalyzer in RTX430 using three methods on the cultivar-specific reference genome.



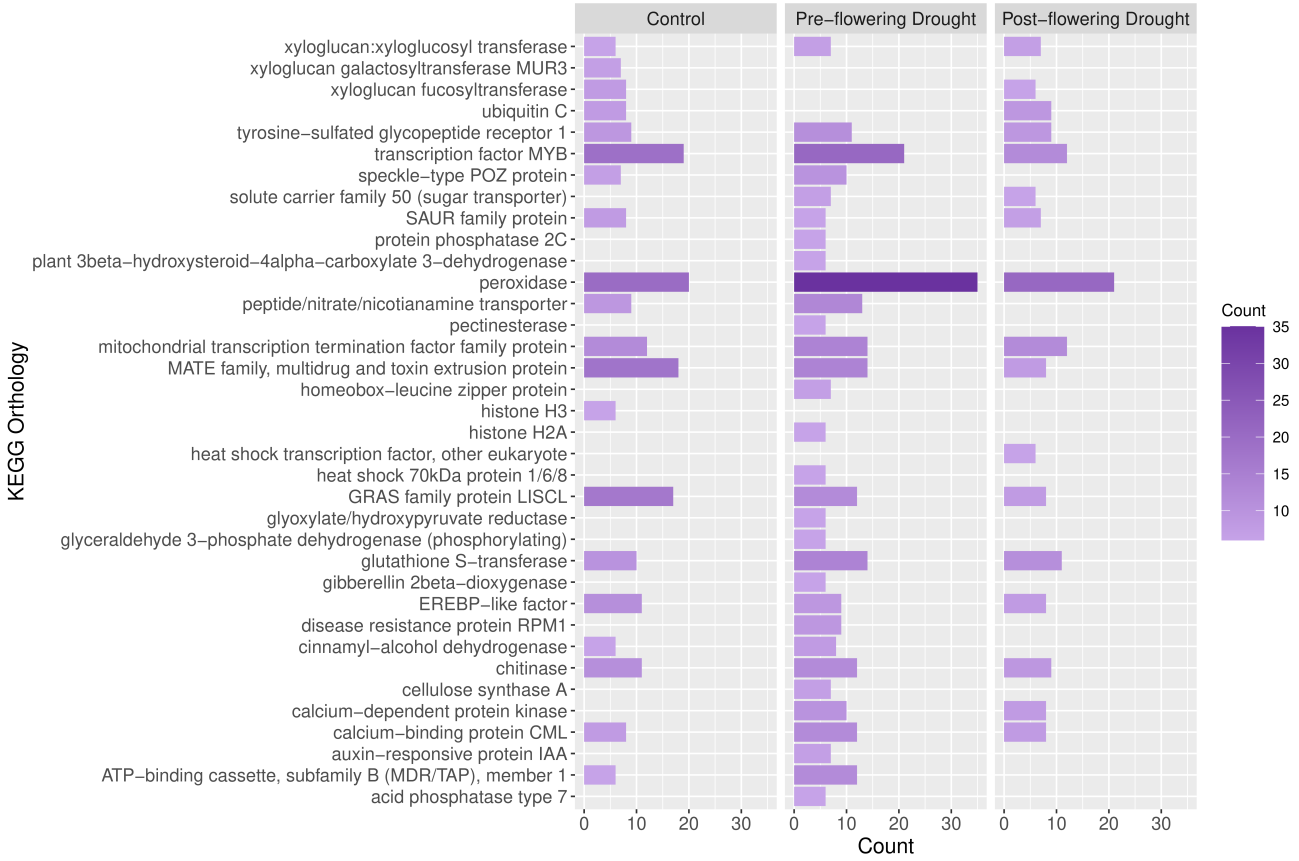
Supplemental Figure S2(a): The numbers of Cis and Trans regulated target genes by lncRNAs in BTX642 leaf tissues under control, pre-, and post-flowering drought.



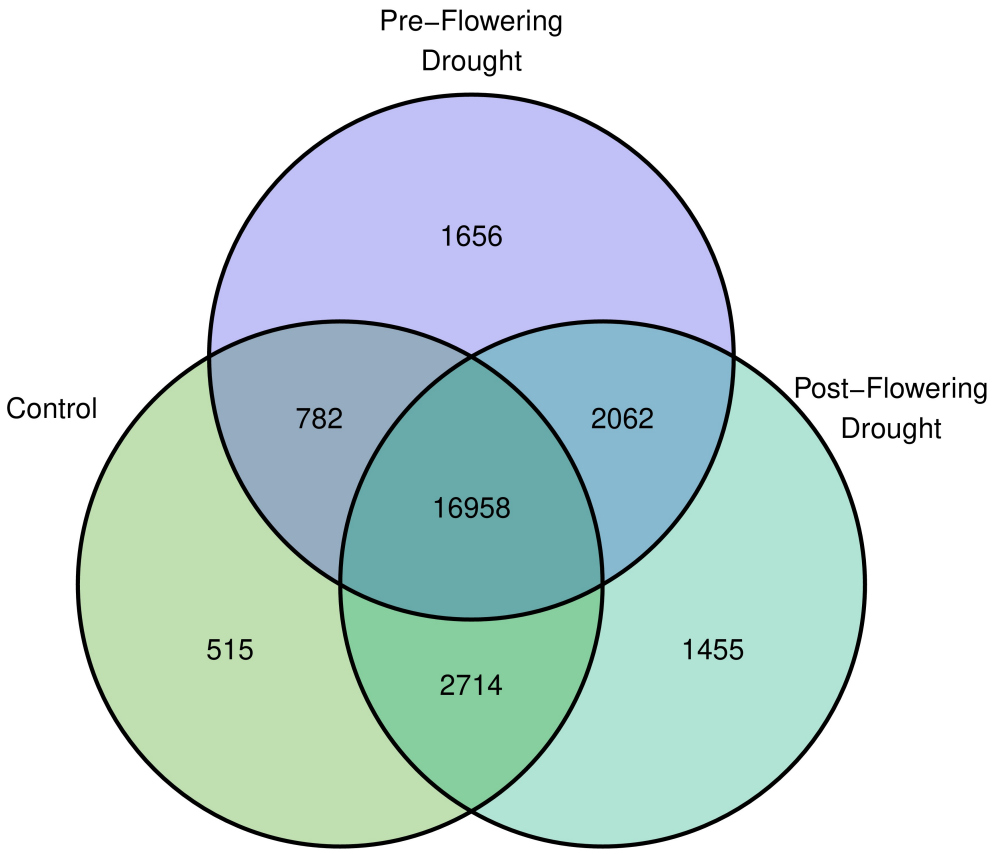
Supplemental Figure S2(b): The numbers of Cis and Trans regulated target genes by lncRNAs in BTX642 root tissues under control, pre-, and post-flowering drought.



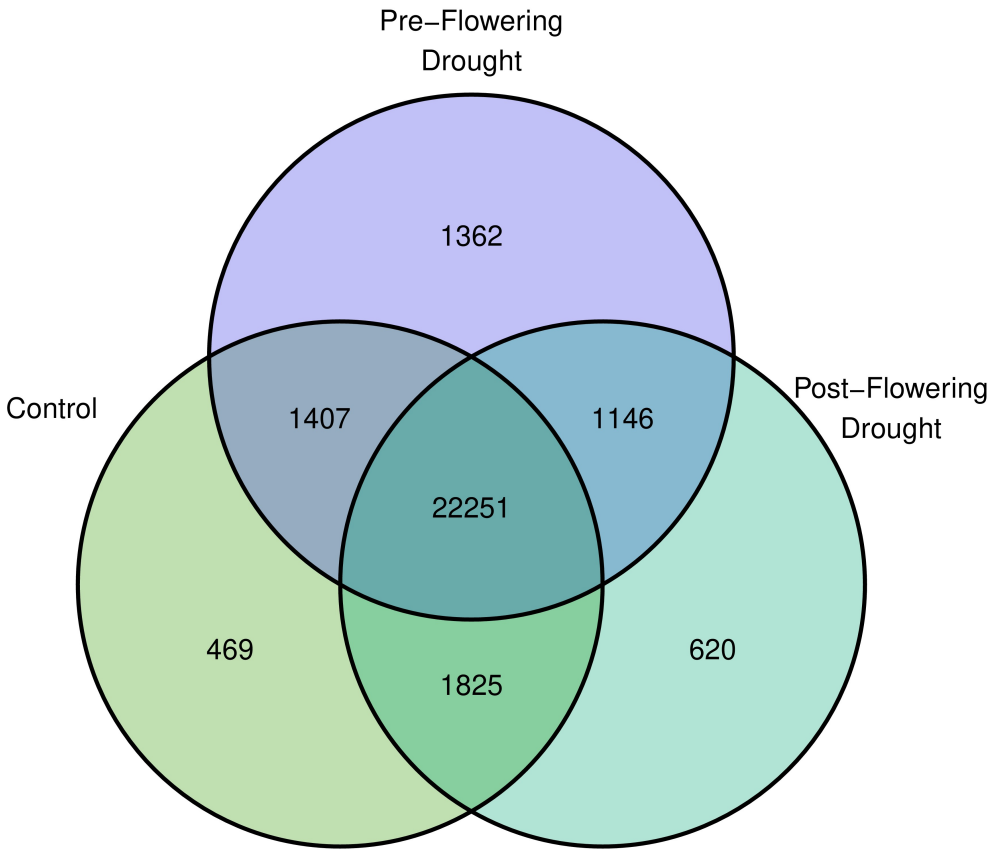
Supplemental Figure S2(b): The numbers of Cis and Trans regulated target genes by lncRNAs in RTX430 leaf tissues under control, pre-, and post-flowering drought.



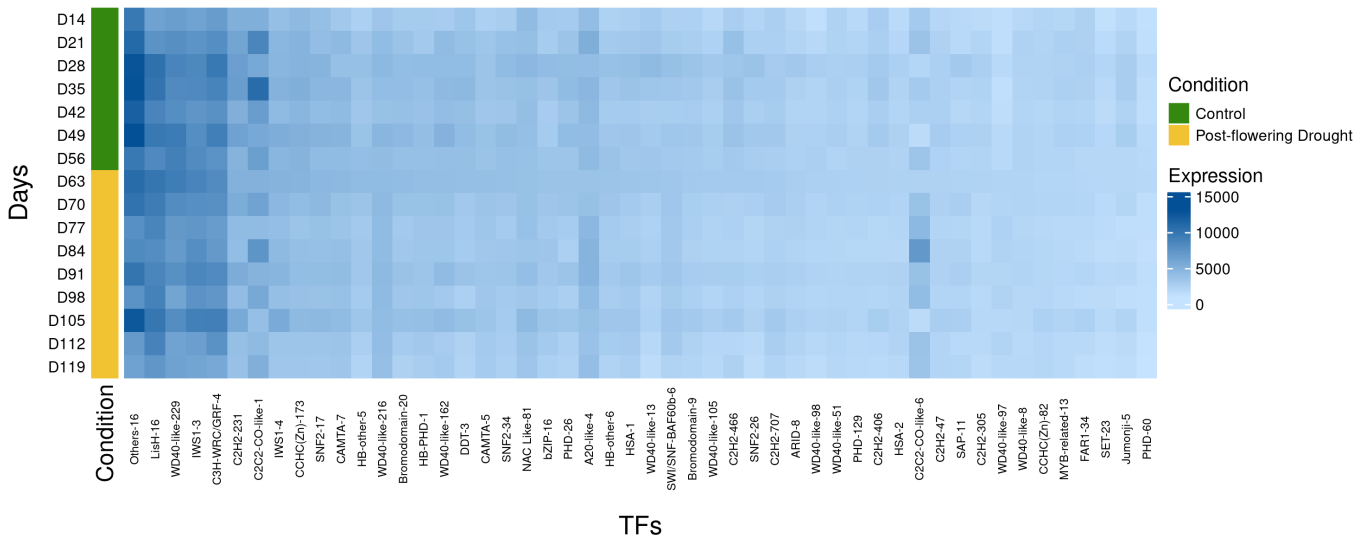
Supplemental Figure S2(b): The numbers of Cis and Trans regulated target genes by lncRNAs in RTX430 root tissues under control, pre-, and post-flowering drought.



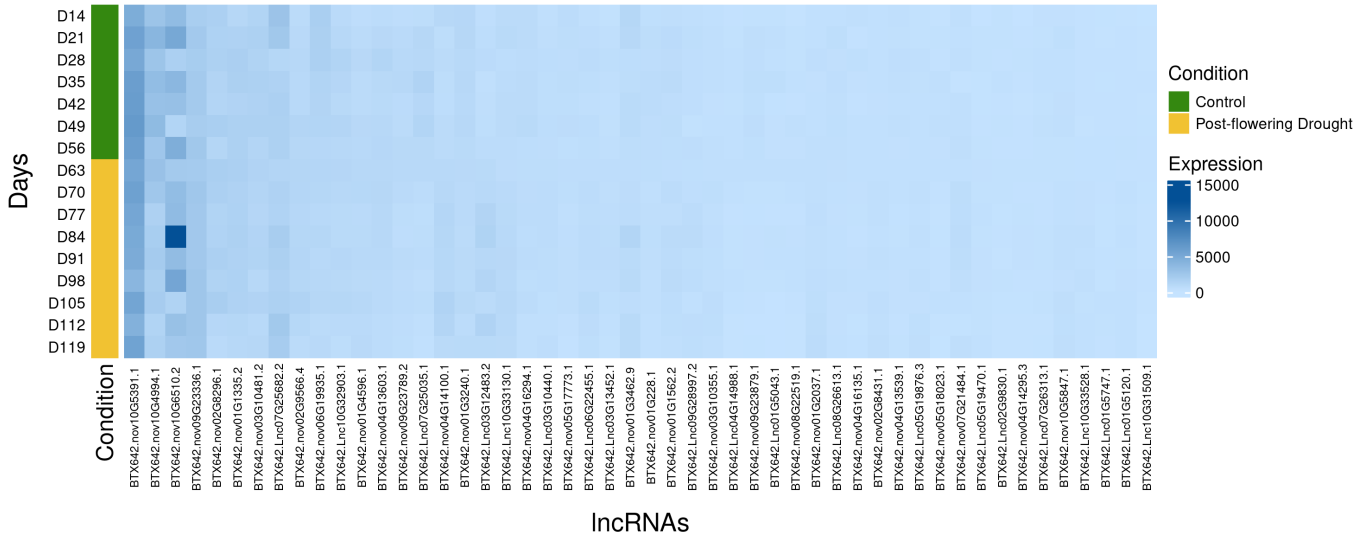
Supplemental Figure S4(a): Venn diagram showing numbers of DEGs in leaf tissues of BTX642 under control, pre-flowering, and post-flowering drought conditions.



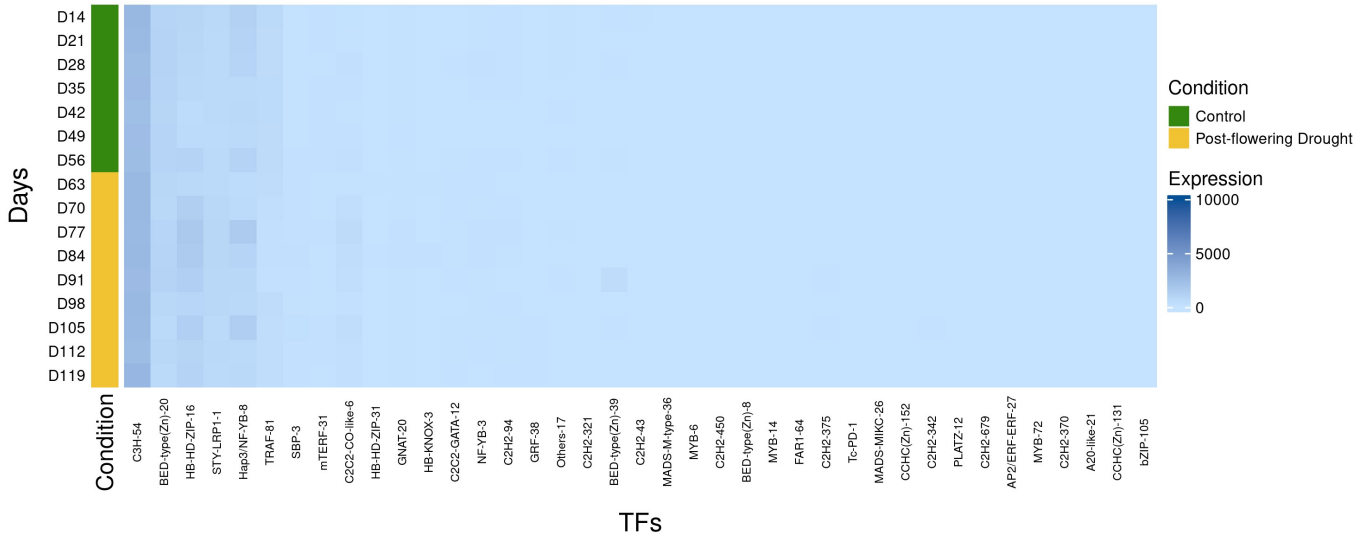
Supplemental Figure S4(b): Venn diagram showing numbers of DEGs in root tissues of BTX642 under control, pre-flowering, and post-flowering drought conditions.



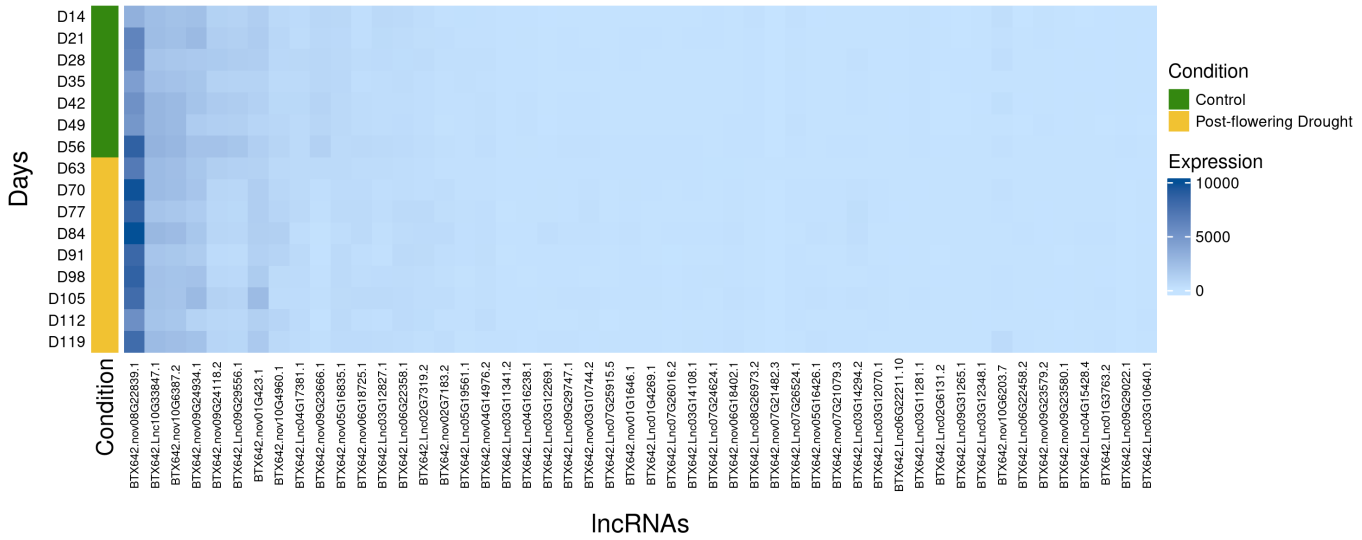
Supplemental Figure S4(c): A heatmap showing the top 50 highly expressed transcription factors exclusive to post-flowering drought conditions in BTX642 leaf tissues.



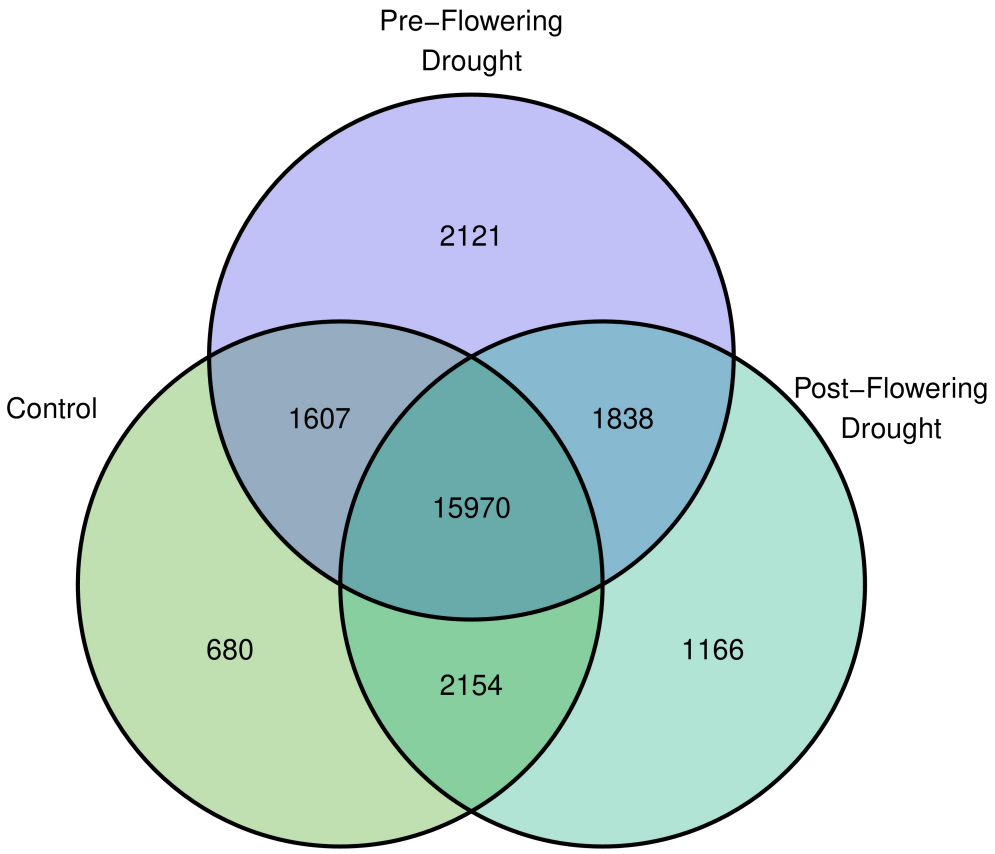
Supplemental Figure S4(d): A heatmap showing the top 50 highly expressed lncRNAs exclusive to post-flowering drought conditions in BTX642 leaf tissues.



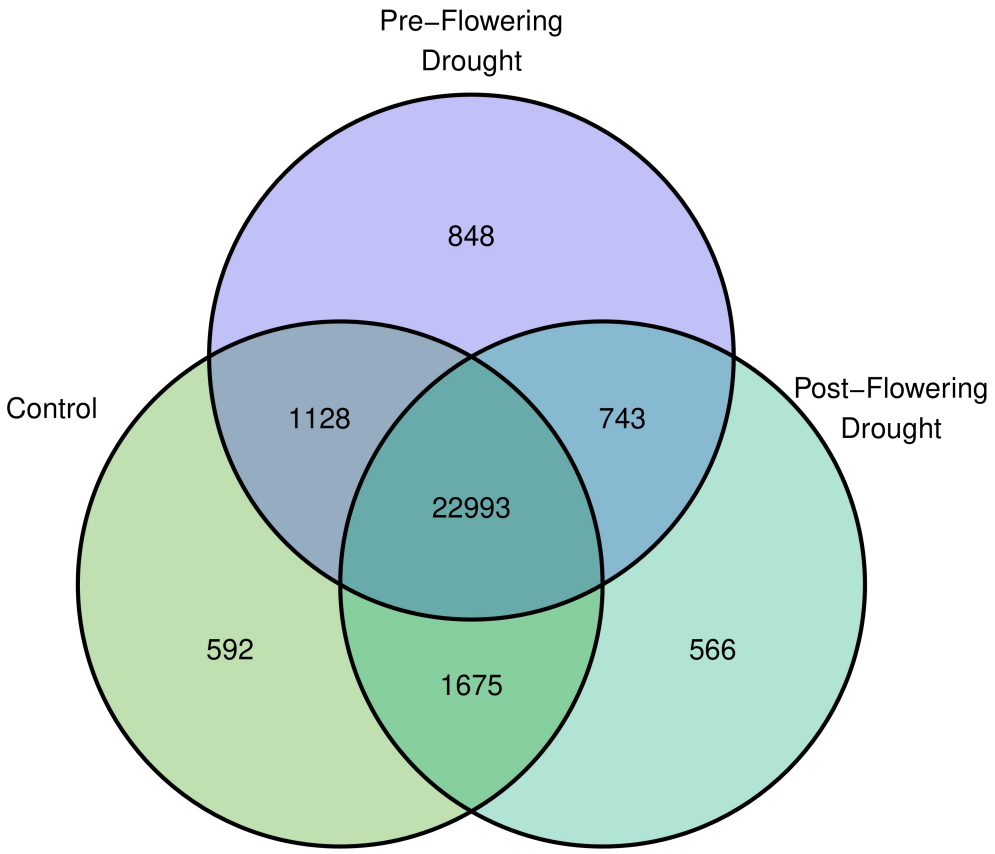
Supplemental Figure S4(e): A heatmap showing the top 50 highly expressed transcription factors exclusive to post-flowering drought conditions in BTX642 root tissues.



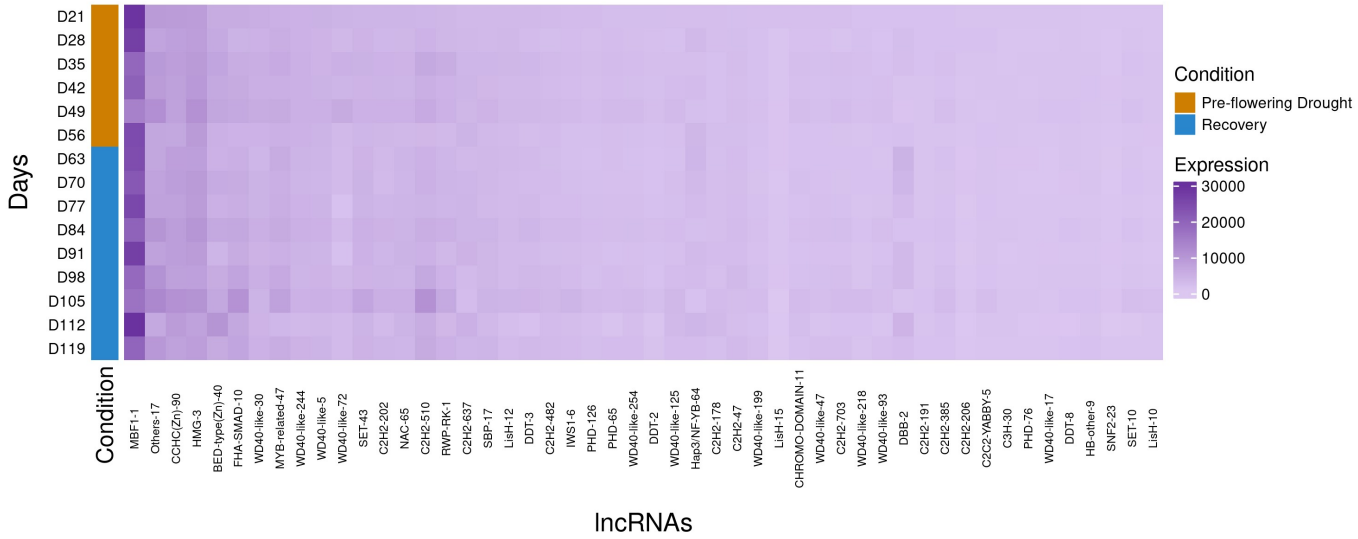
Supplemental Figure S4(f): A heatmap showing the top 50 highly expressed lncRNAs exclusive to post-flowering drought conditions in BTX642 root tissues.



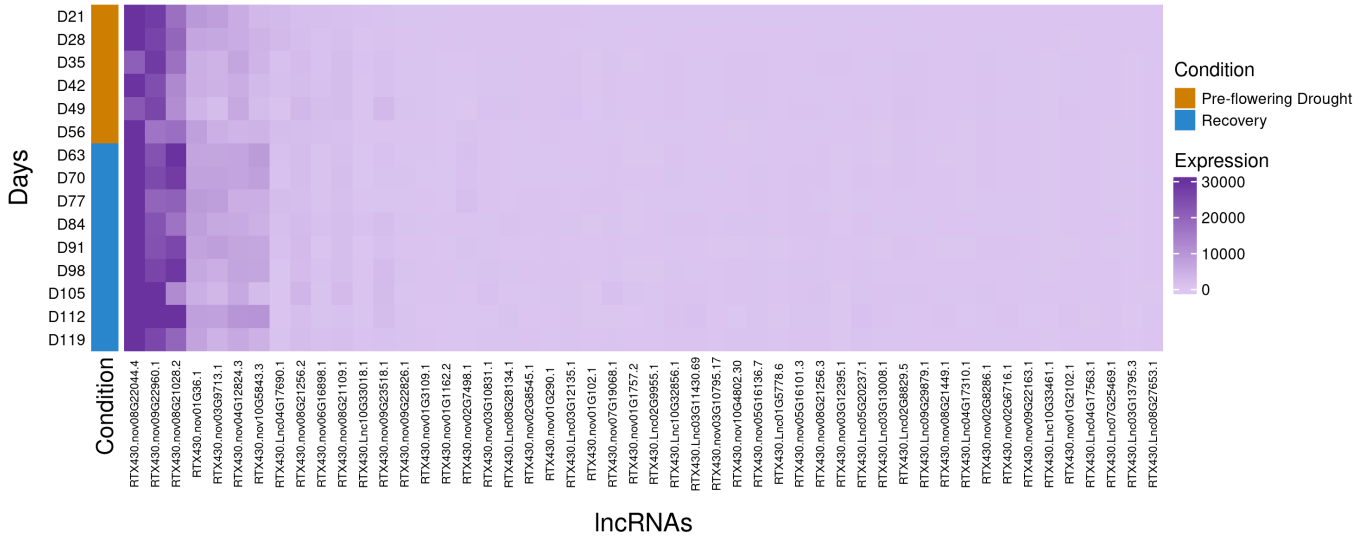
Supplemental Figure S4(g): Venn diagram showing numbers of DEGs in leaf tissues of RTX430 under control, pre-flowering, and post-flowering drought conditions.



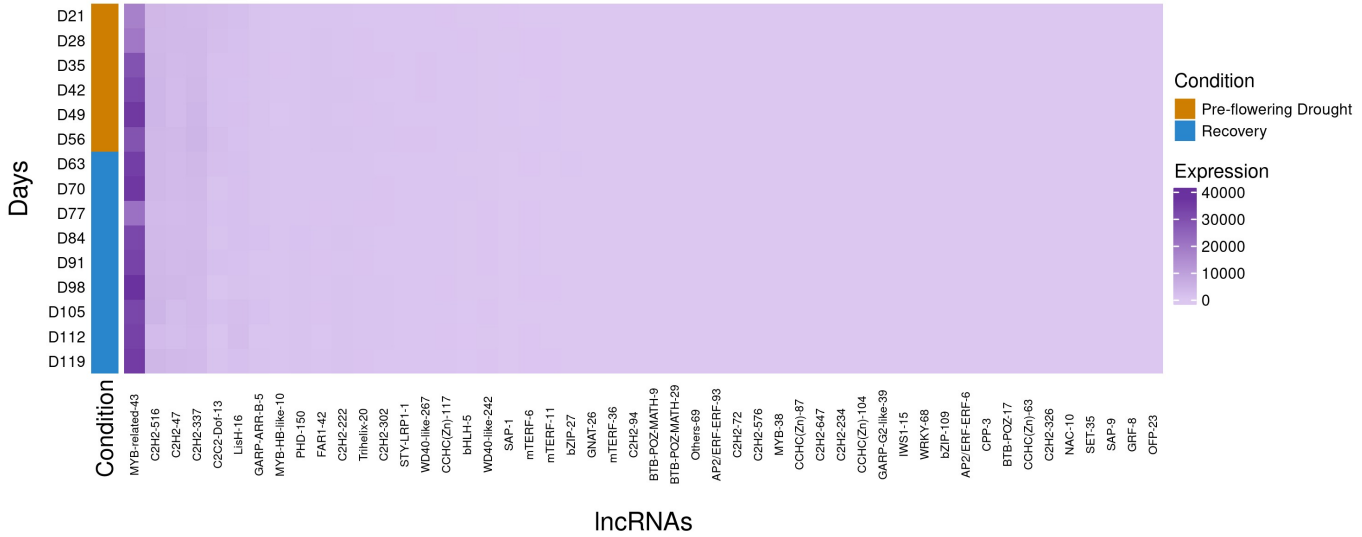
Supplemental Figure S4(h): Venn diagram showing numbers of DEGs in root tissues of RTX430 under control, pre-flowering, and post-flowering drought conditions.



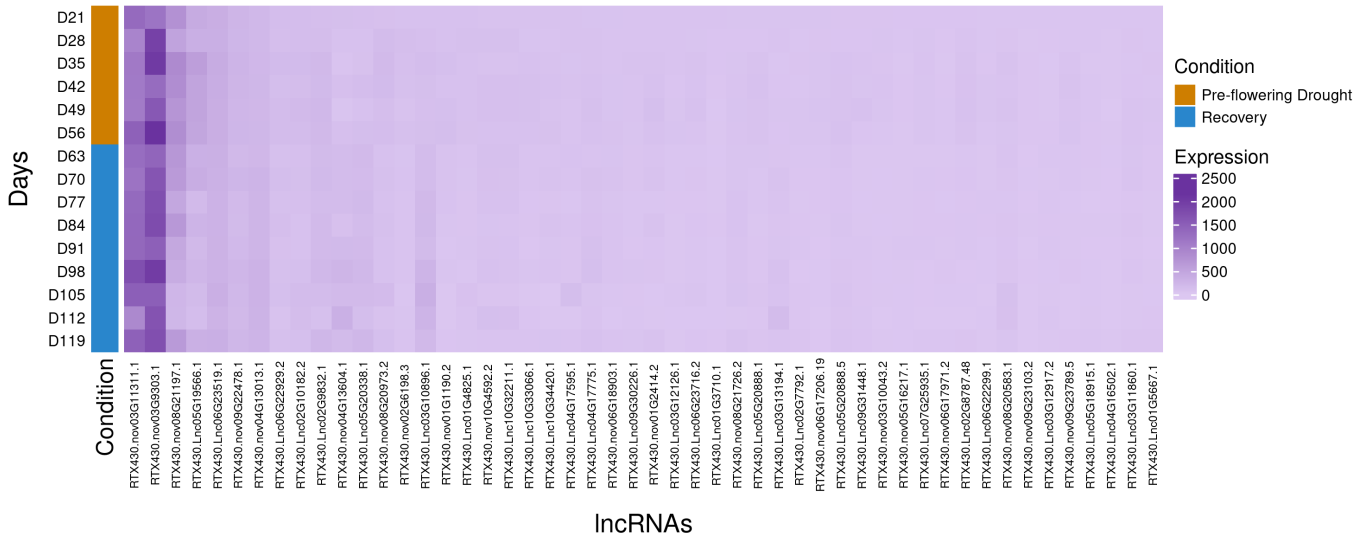
Supplemental Figure S4(i): A heatmap showing the top 50 highly expressed transcription factors exclusive to pre-flowering drought conditions in RTX430 leaf tissues.



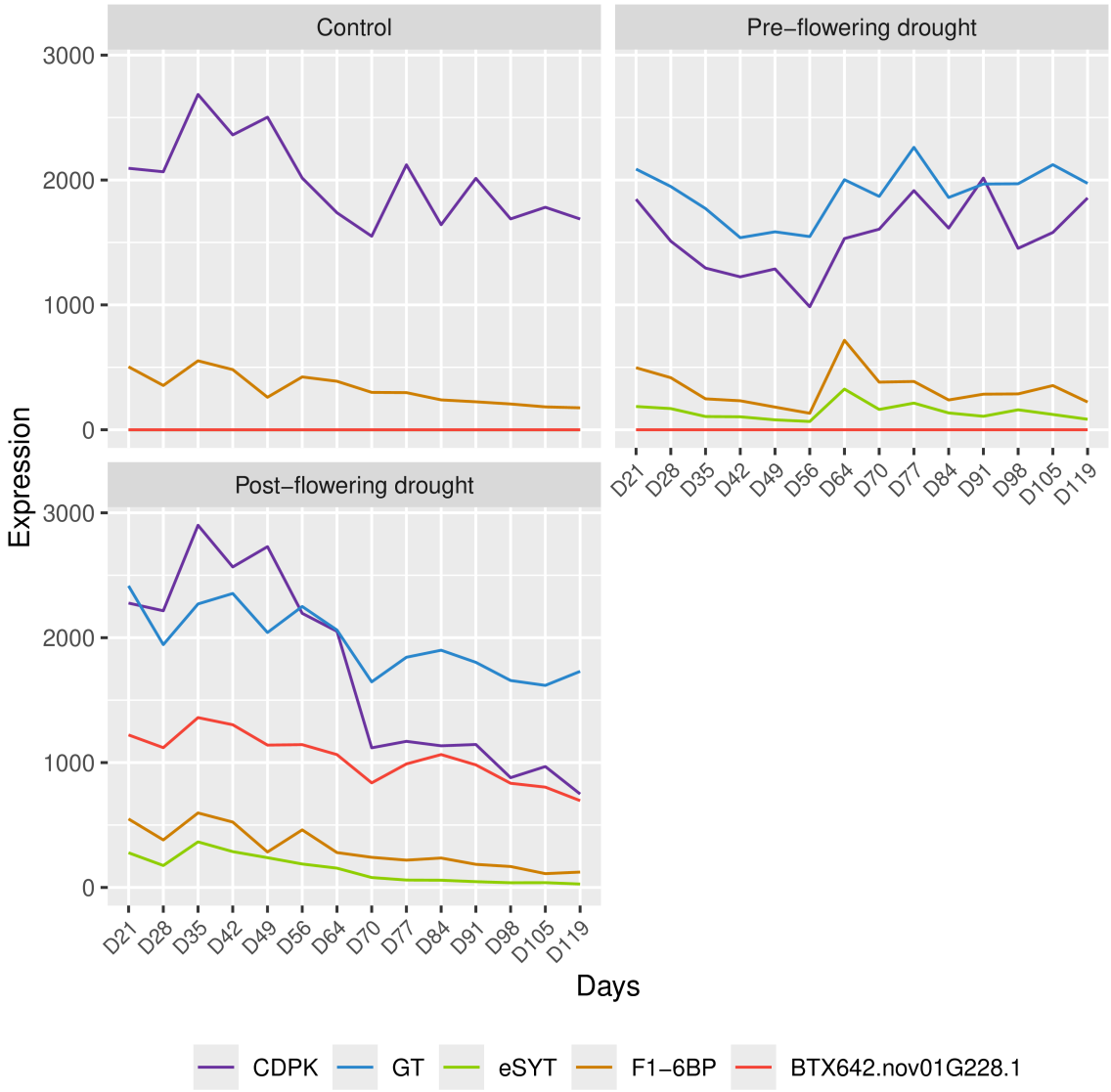
Supplemental Figure S4(j): A heatmap showing the top 50 highly expressed lncRNAs exclusive to pre-flowering drought conditions in RTX430 leaf tissues.



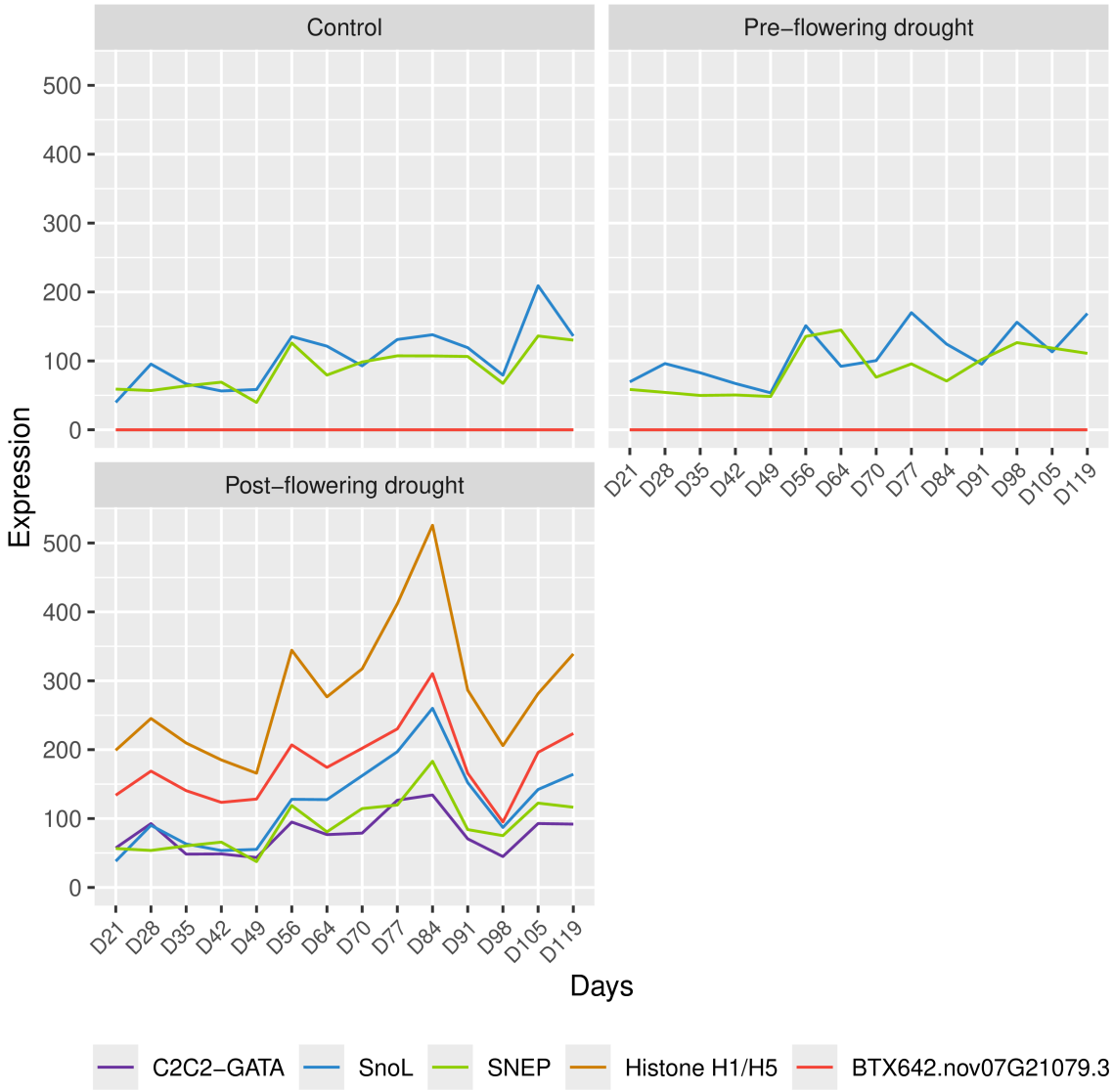
Supplemental Figure S4(k): A heatmap showing the top 50 highly expressed transcription factors exclusive to pre-flowering drought conditions in RTX430 root tissues.



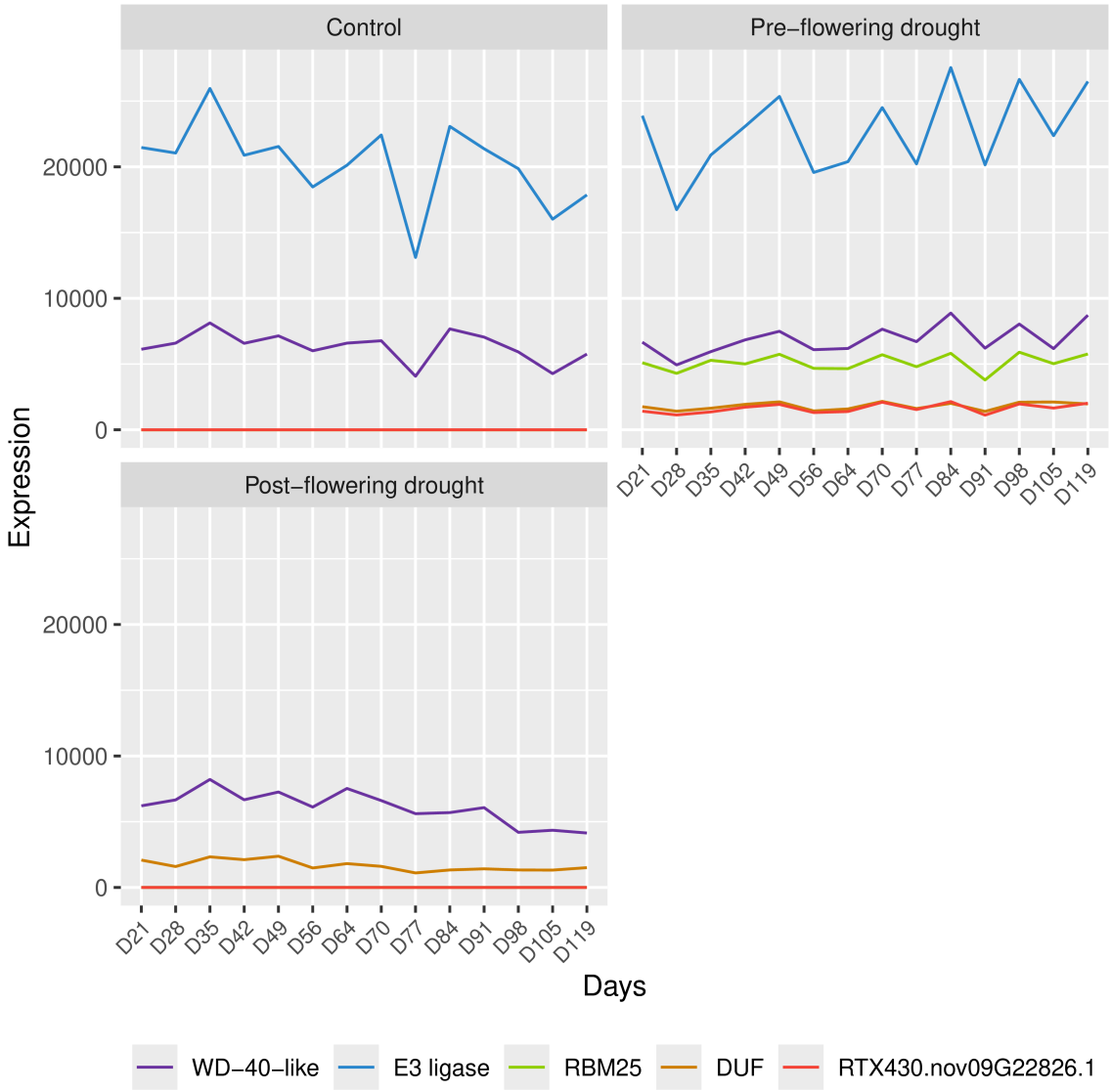
Supplemental Figure S4(l): A heatmap showing the top 50 highly expressed lncRNAs exclusive to pre-flowering drought conditions in RTX430 root tissues.



Supplemental Figure S6(a): BTX642.nov01G228.1 was one of top50 highly expressed lncRNAs under post-flowering drought in BTX642 leaf tissues and a key regulator of several genes such as CDPK, GT, eSYT, and F1-6BP.



Supplemental Figure S6(b): BTX642.nov07G21079.3 was one of top50 highly expressed under post-flowering drought in BTX642 roots tissues and a key regulator of several genes such as C2C2-GATA, SnoL, SNEP, and Histone H1/H5.



Supplemental Figure S6(c): RTX430.nov09G22826.1 was one of top50 highly expressed under pre-flowering drought in RTX430 leaf tissues and a key regulator of several genes such as WD-40, E3 ligase, RBM25, and DUF.

Supplemental Figure S6(d): RTX430.nov08G21197.1 was highly expressed under pre-flowering drought in RTX430 root tissues and key regulator of several genes such as C2H2, ATG, ECM29, and LIP5.