

## Article

# Genome-Wide Identification of Long Non-Coding RNAs in Pearl Millet (*Pennisetum glaucum* (L.)) Genotype Subjected to Drought Stress

Baibhav Kumar <sup>1,†</sup>, Animesh Kumar <sup>1,†</sup>, Sarika Jaiswal <sup>1,\*</sup>, Mir Asif Iquebal <sup>1</sup>, Ulavappa B. Angadi <sup>1</sup>, Rukam S. Tomar <sup>2</sup>, Anil Rai <sup>1</sup> and Dinesh Kumar <sup>1,3</sup>

<sup>1</sup> Division of Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110012, India

<sup>2</sup> Department of Biochemistry and Biotechnology, Junagadh Agricultural University, Junagadh 362001, India

<sup>3</sup> Department of Biotechnology, Central University of Haryana, Mahendergarh 123031, India

\* Correspondence: sarika@icar.gov.in

† These authors contributed equally to this work.

**Abstract:** Pearl millet (*Pennisetum glaucum* L.) is affected by drought stress, affecting crop productivity and survival. Long non-coding RNAs (lncRNAs) are reported to play a vital role in the response to drought stress. lncRNAs represent a major part of non-protein coding RNAs and are present prevalently. These are involved in various biological processes, which may functionally act as RNA rather than getting transcribed as protein. We targeted genome-wide identification of lncRNAs in pearl millet from root and leaf tissues subjected to drought stress. A total of 879 lncRNAs were identified, out of which 209 (leaf control, root control), 198 (leaf treated, root treated), 115 (leaf control, leaf treated) and 194 (root control, root treated) were differentially expressed. Two lncRNAs were found as potential target mimics of three miRNAs from the miRBase database. Gene ontology study revealed that drought-responsive lncRNAs are involved in biological processes like ‘metabolic process’ and ‘cellular process’, molecular functions like ‘binding’ and ‘catalytic activities’ and cellular components like ‘cell’, ‘cell part’ and ‘membrane part’. lncRNA-miRNA-mRNA network shows that it plays a vital role in the stress-responsive mechanism through their activities in hormone signal transduction, response to stress, response to auxin and transcription factor activity. Only four lncRNAs were found to get a match with the lncRNAs present in the plant lncRNA database CANTATAdb, which shows its poorly conserved nature among species. This information has been cataloged in the pearl millet drought-responsive long non-coding RNA database (PMDlncRDB). The discovered lncRNAs can be used in the improvement of important traits, as well as CRISPR-Cas technology, in the editing of ncRNAs in plants for trait improvement. Such a study will increase our understanding of the expression behavior of lncRNAs, as well as its underlying mechanisms under drought stress in pearl millet.

**Keywords:** lncRNAs; pearl millet; drought; miRNAs; web resource



**Citation:** Kumar, B.; Kumar, A.; Jaiswal, S.; Iquebal, M.A.; Angadi, U.B.; Tomar, R.S.; Rai, A.; Kumar, D. Genome-Wide Identification of Long Non-Coding RNAs in Pearl Millet (*Pennisetum glaucum* (L.)) Genotype Subjected to Drought Stress.

*Agronomy* **2022**, *12*, 1976. <https://doi.org/10.3390/agronomy12081976>

Academic Editors: Margarita A. Vishnyakova, Eric J. Bishop von Wettberg and Maria G. Samsonova

Received: 5 February 2022

Accepted: 15 April 2022

Published: 22 August 2022

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The advent of advanced high-throughput genomic platforms has provided us with various insights into biological gene regulation. It opines that the expansion of the regulatory potential of the noncoding parts of the genome leads to the evolution of developmental processes that regulate the organism’s complexity [1,2]. Only a portion of the genomes (~1–2%) is found to be responsible for protein coding, while many of the regulatory elements get transcribed into non-coding RNA (ncRNA). Among these ncRNAs, long non-coding RNAs (lncRNAs) represent the most prevalent, involved in various biological processes, which may functionally act as RNA rather than getting translated into protein [3].

Noncoding RNAs which include microRNAs (miRNAs) and lncRNAs are the functional transcripts for gene regulation that are not translated into proteins. The lncRNAs are generally longer than 200 nucleotides and are involved in DNA methylation, chromatin remodeling or enhancing the expression of mRNA targeted by miRNA [4]. The ncRNAs can be divided into two categories: lncRNAs and ncRNAs (those being smaller than 200 nucleotides). lncRNAs are the RNA transcripts that act as either primary or as spliced transcripts, independent of their length threshold [5]. This definition supports the presence of some lncRNAs like BC1 and snaR in the lncRNAdb having a length of less than 200 nucleotides.

Different research shows lncRNAs to be involved in several important regulatory and functional roles [6,7] such as splicing, transcription, translation, cell cycle, protein localization, flowering, development of pollens, sex differentiation, cellular structure integrity, heat shock response, cancer progression, exhibit cell-specific expression and stress responses [8], chromatin regulation [9], development of floral organ and root [10], etc. In addition, lncRNAs are poorly conserved among different species as compared to mRNAs, snoRNAs (small nucleolar) and miRNAs (micro RNAs). Similar to mRNAs, lncRNAs are transcribed by RNA polymerase II, 3' polyadenylated, 5' capped and found to be multi-exonic [11].

Literature reports the discovery of lncRNAs in plants, such as Arabidopsis [12], rice [13], wheat [14], maize [15], rapeseed [16] and cassava [17], suggesting the role of lncRNAs in many biological processes contributing to the development of plants and their response to various stresses. Various research works have shown the evidence of the role of lncRNAs in abiotic stress, including drought stress but these are very few in number [17]. There have been instances where lncRNAs are reported to control drought stress tolerance, both in monocots and dicots. Drought-responsive lncRNAs are evident in monocots like rice [18], maize [19], foxtail millet [20], and switchgrass [21]. These lncRNAs mediate response to drought stress through mechanisms based on eTM, antisense transcription-mediated modulation, chromatin modulation, etc. They may directly regulate the transcription of drought-responsive genes as well [22–24].

Owing to the gaining importance of lncRNAs discovery, many lncRNA repositories and databases have been launched during the last 6–7 years. There exist many lncRNA databases, like PLncDB V2.0 which contains 1,246,372 lncRNAs for 80 plant species based on 13,834 RNA-Seq datasets [25]. A total of almost 200,000 putative lncRNAs from 50 species have been cataloged in Green Non-Coding Database (Gallart et al., 2016) [26]. Other existing lncRNA databases are EVLncRNAs [27], lncRNAdb [28], PLNlncRbase [29], CAN-TATAdb [30] and RNAcentral [31].

Pearl millet (*Pennisetum glaucum* L.) is the world's sixth most important cereal crop and is mostly grown by the poor and marginal farmers in arid and semi-arid tropics of Asia and Africa, due to its less intensive agronomic practices like less fertilizers and limited irrigation input [32]. It is cultivated mainly for its grains but also used as feed material for cattle, poultry, fish, etc. Pearl millet has multifarious applications as fodder for animals, feed for poultry, as bread/cookies ingredient (ICRISAT, 2003), as probiotic fermented food [33], vegetable salad (<http://sprigandvine.in/modern-millet-salad/>) (accessed on 4 August 2021), nutraceutical industry [34] and as a biofuel crop. The crop is able to perform well in drought conditions where most of the cereals like wheat, rice and maize fail. So, understanding the molecular mechanism of the responses of pearl millet to adverse conditions is important. Key candidate genes controlling drought response in Indian pearl millet have been discovered and reported [35] but lncRNA still remains uncovered. The expression of candidate genes is controlled by microRNA, TFs and lncRNA. The study aims at the identification of drought-responsive lncRNAs and the development of a web of genomic resources for user-friendly access to the investigation findings, which is otherwise lacking in this crop. The discovered lncRNAs can be used in the improvement of important traits as in the case of rice [36]. Such studies can further be used by integrating CRISPR-Cas technology in the editing of ncRNAs in plants for the improvement of their trait [37]. All of this information has been cataloged in the pearl millet drought-responsive long non-coding

RNA database (PMDlncRDB) accessible freely at <http://webtom.cabgrid.res.in/pmdlncrdb> (accessed on 10 January 2021). The study aims at the identification of drought-responsive lncRNAs and the development of a web of genomic resources for user-friendly access to the investigation findings, which is otherwise lacking for this crop. This information can be used for pearl millet improvement programs in endeavors of higher production and combating drought in pearl millet.

## 2. Material and Methods

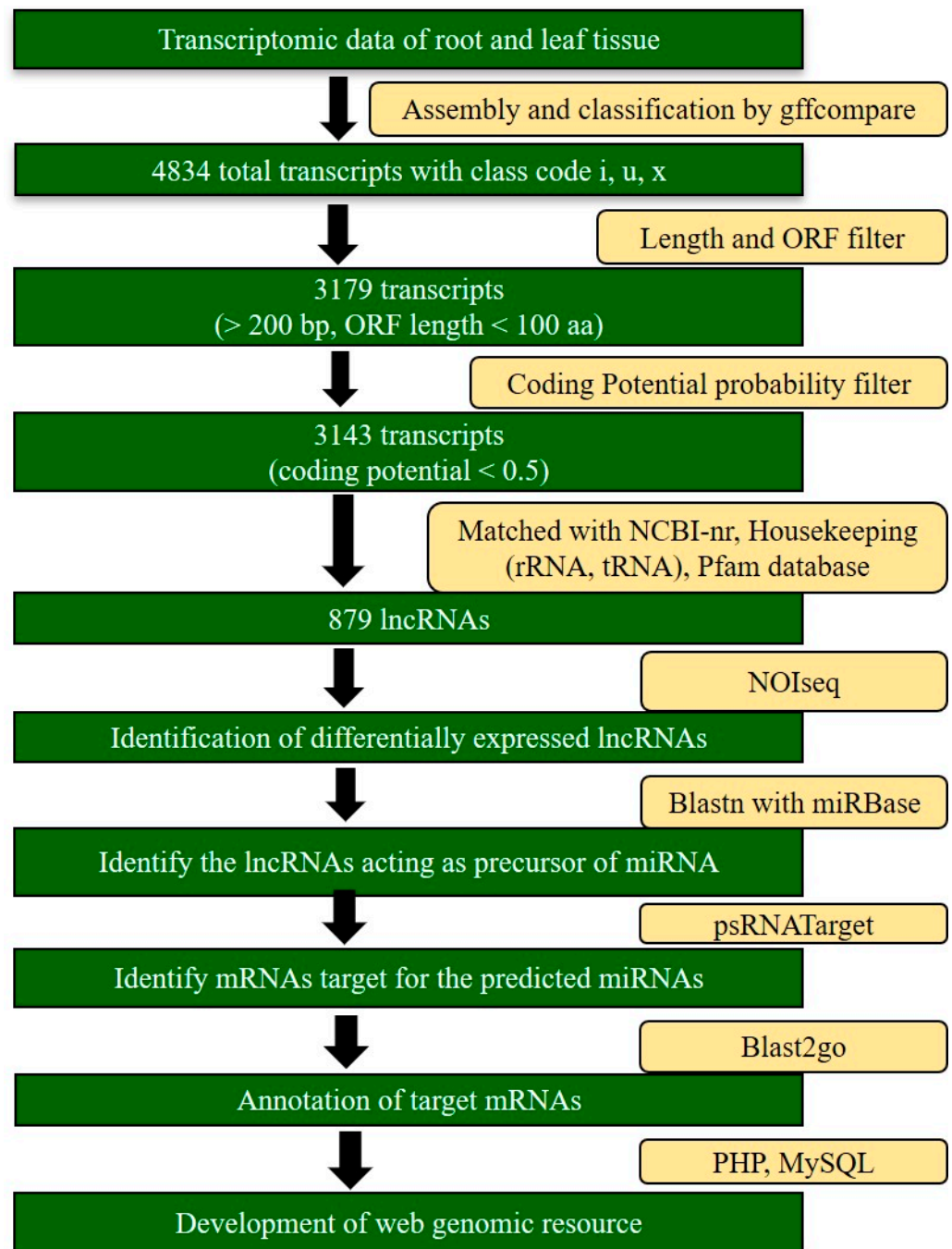
### 2.1. Data Pre-Preprocessing

*Pennisetum glaucum* J-2454 variety, which is drought tolerant, was grown in greenhouses during the summer season at Junagarh Agriculture University, Gujarat, India. About 25–30 seeds were sown per polythene bag in three replications with control and water stress. The tissues (leaf and root) were collected on the 29th day after sowing (withdrawal of irrigation for 6 days after the 23rd day of sowing) from control and drought-treated plants and stored in liquid nitrogen at  $-80\text{ }^{\circ}\text{C}$ . Tissues were frozen in liquid nitrogen and stored at  $-80\text{ }^{\circ}\text{C}$ . RNA was extracted using the standard protocol of TRIZOL RNA isolation in accordance with the manufacturer's protocol (Life Technologies, Grand Island, NY, USA) as described in Jaiswal et al., 2018. Experimental details and data generation is described by Jaiswal et al., 2018. Data are available at SRA NCBI with SRR5839373 (Leaf control), SRR5839374 (Leaf treated), SRR5839375 (Root control), SRR5839376 (Root treated). The reference genome of Pearl millet was downloaded from NCBI (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA294988>) (accessed on 10 January 2021). The raw reads were checked for quality using the FastQC tool [38]. Poor quality reads were trimmed using Trimmomatic [39] at default parameters and rechecked for quality by FastQC.

The clean and trimmed reads were aligned with the reference genome using HISAT2 aligner [40] and assembled by the StringTie v1.3.5 program, which is part of the “new Tuxedo” suit [41,42]. We used a reference annotation file to improvise the quality of the result and assembly of less abundant genes, which would have been missed otherwise. After getting all assembled reads, gffcompare tool (v0.11.2, <https://ccb.jhu.edu/software/stringtie/gffcompare.shtml>) (accessed on 15 February 2021) was employed to assemble all files in a single gtf file for further downstream analysis.

### 2.2. Candidate lncRNA Prediction

The pipeline used for the identification and development of lncRNA with annotation and their development of web genomic resources has been described in Figure 1. The gffcompare tool classifies all the four gtf (gene transfer format) assembly files into various class codes viz. *i* (transcript falling in the intronic region), *o* (exonic overlap), *u* (unknown/intergenic), *x* (overlapping an exon on opposite strand), *r* (repeats), etc. [43]. The transcripts corresponding to class codes *i*, *u* and *x* only were extracted. The transcript smaller than 200 nucleotides in length were removed using an in-house perl script. Then, transcripts with FPKM values  $< 0.5$  were discarded, and transcripts with a single exon having  $< 2$  FPKM values were also removed.



**Figure 1.** Schematic diagram for identification of lncRNAs, annotation of target mRNAs and development of web genomic resource.

OrfPredictor tool [44] was used to predict the ORFs. The transcripts with ORF length > 100 amino acids were discarded. The coding potential calculator (CPC2) tool (<http://cpc2.cbi.pku.edu.cn/>) (accessed on 10 April 2021) was used to predict the coding potential of the transcripts [45] and the sequences with those with coding potential probability values >0.5 were discarded on the account of being coding. The house-keeping genes (including tRNAs, rRNAs, snRNAs and snoRNAs) were removed by aligning the lncRNAs at different lncRNA databases (<http://noncode.org>, <http://gtrna>, <http://silva>) [46] (accessed on 27 April 2021). While searching novel transcripts, BLAST [47] was performed with NCBI non-redundant database and homologous pairs were removed. The protein-coding transcript was eliminated by scanning against the Pfam protein family database [48]

using HMMER [49] to exclude any of the possibility of known protein domains. The unique transcripts were considered as potential lncRNAs.

### 2.3. Identification of Differentially Expressed lncRNAs, Their Characterization and Annotation

For the differential expression of lncRNAs, analysis of four pairs of combinations *viz.*, leaf control root control (LC:RC), leaf treated root treated (LT:RT), leaf control leaf treated (LC:LT) and root control root treated (RC:RT) was done. We used the “new Tuxedo” protocol [42]. The reads were aligned by HISAT, followed by transcripts assembly using *StringTie*. This was followed by *StringTie*—merge to merge GTF files. The read count matrix was generated through python script (prepDE.py). To identify the differentially expressed lncRNA, NOISeq [50] of the R Bioconductor package was used with an adjusted  $p$ -value  $< 0.05$ .

Homology search of the differentially expressed lncRNAs of pearl millet was performed against the known plant lncRNA database, that is, CANTATAdb [51] with the threshold  $e$ -value of 0.001 and sequence coverage and identity  $> 90\%$ . To know if the predicted lncRNAs act as a precursor of miRNA (pre-miRNA), they were subjected to a BLAST search against the known miRNAs available in miRbase (release 22.1) [52]. The secondary structures were predicted and plotted using the Vienna RNA package RNAfold web [53]. Characterization of the identified mRNAs was performed using the Blast2GO tool (<https://www.blast2go.com/>) (accessed on 12 May 2021) thereby classifying them into subcategories like biological process, molecular function and cellular component.

### 2.4. Identification of miRNAs Targeting the lncRNAs and Their Target mRNAs

lncRNAs are known to be very poorly conserved among species, but miRNAs on the other hand are known to be conserved throughout the plant species. To identify the miRNAs targeting lncRNAs, the lncRNAs transcripts were BLAST against the miRBase database using the blastn program with stringent parameters like  $e$ -value  $< 0.001$ , percent identity  $> 90\%$  and coverage  $> 80\%$ .

lncRNAs can also function as target mimics of the miRNAs imitating the corresponding mRNAs. Potential targets of miRNAs against mRNAs were predicted by psRNATarget [54], with an expectation of  $< 3$  and unpaired energy (UPE)  $< 25$ .

### 2.5. Development of Pearl Millet lncRNAs Database (PMDlncRDB)

A web genomic resource, the Pearl millet lncRNAs database (PMDlncRDB) was developed which contains the information about identified lncRNAs in response to drought stress in pearl millet. The database also contains the publicly available lncRNAs of the pearl millet crop. The developed web transcriptome database is based upon the “three-tier architecture” of a database system consisting of the client tier, the middle tier and the database tier. The client layer was developed using HTML and JavaScript for browsing the database and defining queries. The middle tier was made using PHP, which functions in database connectivity, executing queries and fetching data from the database. The last tier, database tier was made using MySQL, which will be having the data about the identified lncRNAs, publicly available lncRNAs in pearl millet, GO (gene ontology) terms, etc.

## 3. Results

### 3.1. Data Pre-Preprocessing

The single-end RNASeq data used in this study comprised a total of 12.14 million raw reads. After trimming and adapter removal, the clean reads were mapped over the pearl millet reference genome using HISAT2 software to get the overall aligned rate of 73.46%, 73.53%, 69.81% and 72.95% for LC, LT, RC and RT samples (Table 1). and recovered 3.35 million mapped reads Table 1. A total of 39,978 transcripts were generated assembly by *StringTie*. Table 2 delineates the summary statistics of transcriptome assembly.



**Table 1.** Summary statistics of read alignment over the reference pearl millet genome.

Sample	Total Raw Reads	Unique Mapped Reads	Multiple Mapped Reads	Overall Aligned Rate
Leaf-Control	3,406,324	346,223 (68.58%)	24,342 (4.82%)	73.46%
Leaf-Treated	3,103,247	1,011,944 (67.86%)	84,531 (5.67%)	73.53%
Root-Control	2,272,632	762,761 (65.64%)	48,489 (4.17%)	69.81%
Root-Treated	3,360,164	775,850 (52.66%)	298,994 (20.29%)	72.95%
<b>TOTAL</b>	<b>12,142,367</b>	<b>2,896,778 (62.54%)</b>	<b>456,356 (9.85%)</b>	<b>72.39%</b>

**Table 2.** Assembly statistics.

Sample	Transcripts	Multi Exon	Multi Exon/Transcripts
Leaf-Control	4872	2159	44.31%
Leaf-Treated	12,076	6855	56.76%
Root-Control	13,506	7507	55.58%
Root-Treated	9524	4988	52.37%
<b>TOTAL</b>	<b>39,978</b>	<b>21,509</b>	<b>53.80%</b>

### 3.2. Candidate lncRNA Prediction

All the transcripts were subjected to the Gffcompare tool, which, on further filtering and differentiating into various class codes, leaves 4834 transcripts having 288 intronic, 4458 intergenic and 88 transcripts with exonic overlap on the opposite strand. After discarding the sequences smaller than 200 nucleotides using perl scripts and predicting the ORFs in the transcripts using the OrfPredictor tool, a total of 3170 sequences were having ORF lengths less than 100 amino acids and 9 sequences with no ORFs were retained for further filtering. CPC2 online tool classified the 3179 input sequences into 36 coding and 3143 non-coding sequences, keeping 0.5 as the threshold value of the coding probability. Blast search against Housekeeping RNAs (rRNA, tRNA, etc.), Pfam, NCBI-nr helps in retaining 879 sequences that did not match with any sequences in the database. Expression of 57 lncRNAs (6 from the intronic region and 51 from the intergenic region) was observed under leaf control condition, 285 lncRNAs (33 from intronic region, 246 from intergenic region and 6 from exonic overlap of opposite strand) was observed under leaf treated, 494 lncRNAs (67 from intronic region, 408 from intergenic region and 19 from exonic overlap of opposite strand) was observed under root control and 210 lncRNAs (30 from intronic region, 179 from intergenic region and 1 from exonic overlap of opposite strand) was observed under root treated.

It was observed that 195 lncRNAs (22.2%), 393 (44.7%), 34 (3.9%) and 126 (14.3%) lncRNAs were unique for leaf treated, root control, leaf control and root treated, respectively. It was also found that 3 (0.3%) lncRNAs were common in all the conditions, which indicates that these are expressed under all the conditions.

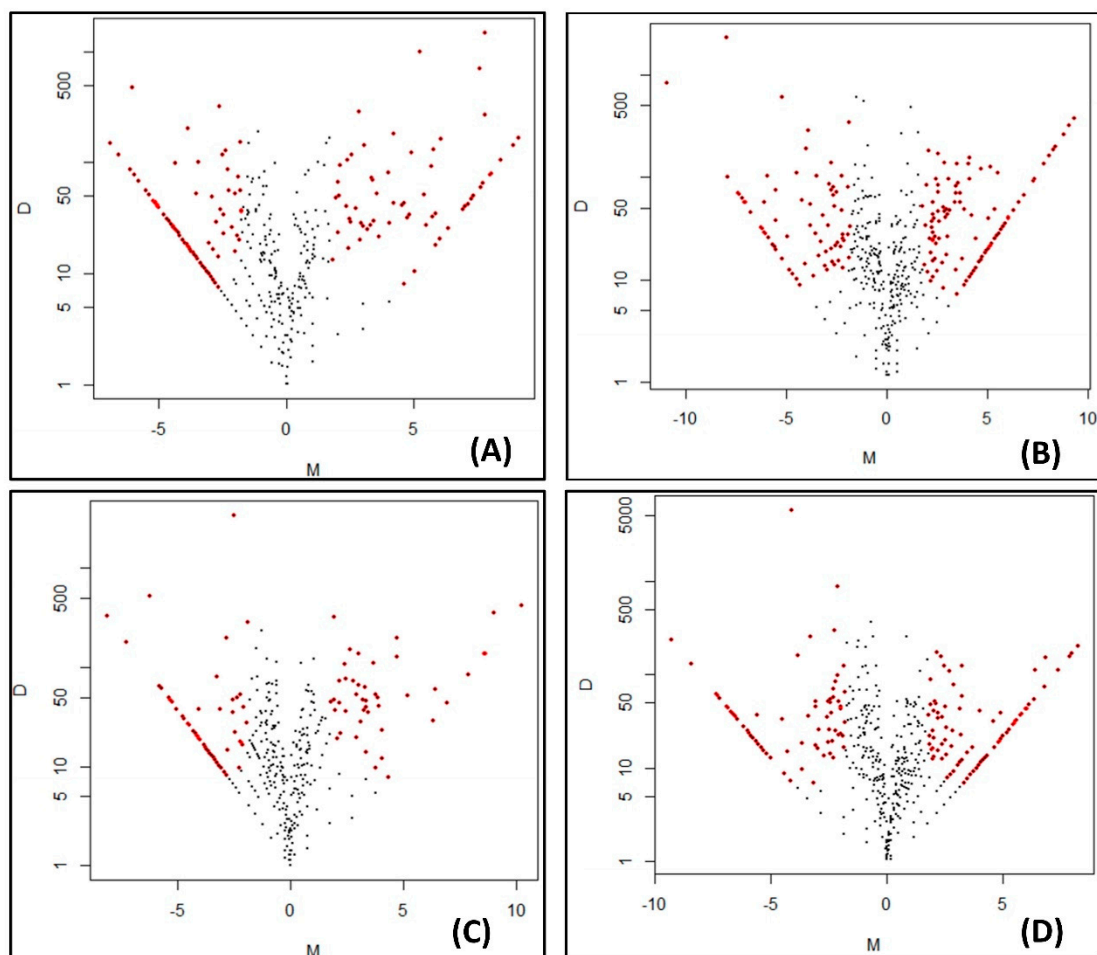
Low exon number and the short transcript length were the typical features of lncRNA [55]. The maximum number of lncRNAs belonged to the length range of 200–400 bp, followed by 400–600 bp under all four conditions. Most of the identified lncRNAs showed the involvement of 1 to 4 exons, as also supported by Li et al., 2014 [15]; Shuai et al., 2014 [56]; and Pauli et al., 2012 [57].

### 3.3. Identification of Differentially Expressed lncRNAs, Their Characterization and Annotation

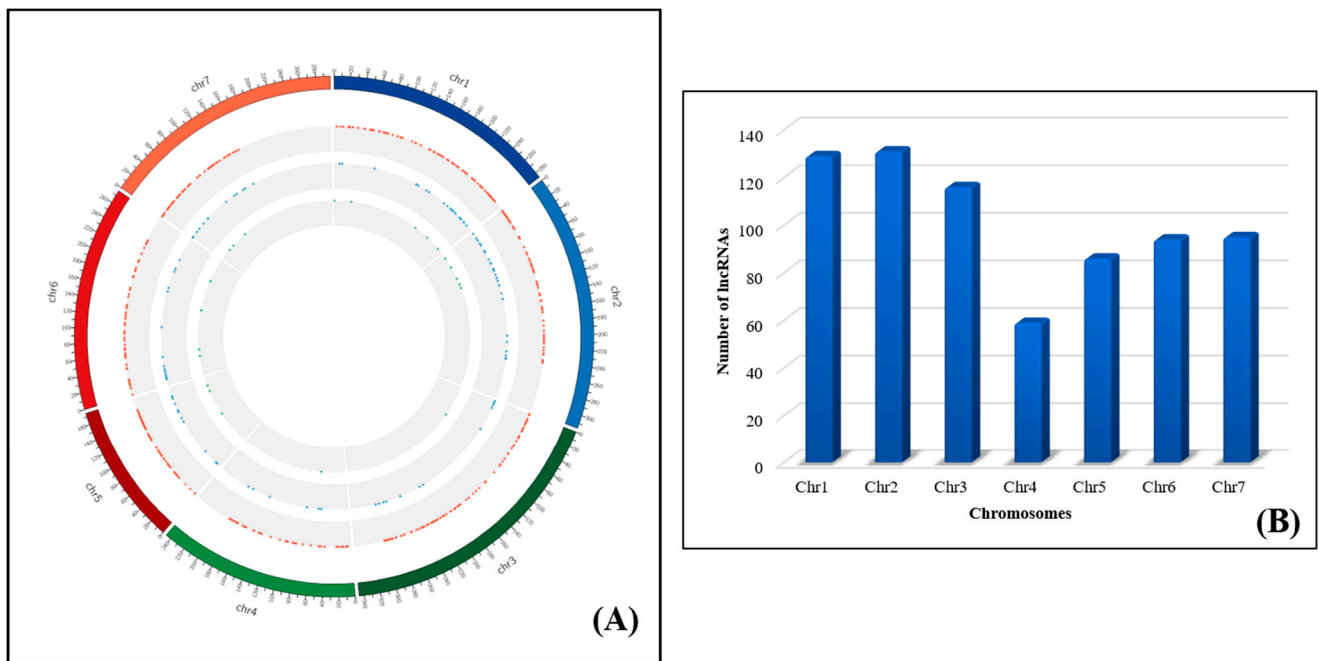
Differential expression of the common lncRNAs was found using NOISeq [50], which uses data-adaptive and non-parametric approaches to calculate the differential expressing transcripts. NOISeq can control the rate of the false discoveries by combining the data of fold change and real expression differences and making a null distribution and finally

comparing this null distribution to the observed or actual data. NOISeq is preferably used under low expression values and a lesser number of replicates.

A total of 209, 198, 115 and 194 differentially expressed lncRNAs were discovered for LC:RC, LT:RT, LC:LT and RC:RT, respectively. Out of these 67, 123, 46 and 116 lncRNAs were upregulated and 142, 75, 69 and 78 lncRNAs were down-regulated for LC:RC, LT:RT, LC:LT and RC:RT, respectively. An adjusted  $p$ -value cutoff of 0.05 was set to filter out differential lncRNAs. The up and down-regulated differentially expressed lncRNAs are represented with Volcano plots for all four comparisons (Figure 2). The dots show the up and down-regulated lncRNAs under different combinations. Figure 3A shows the distribution of identified lncRNAs along the pearl millet chromosomes where the red dot represents intergenic lncRNA, the green dot is intronic lncRNA and the blue dot is exonic overlap on the opposite strand lncRNA respectively. Figure 3B shows the graphical representation of the distribution of lncRNAs over the pearl millet chromosomes which can be used for designing different types of chemistry, multiplexing, etc. There is much use for developed lncRNAs as markers for molecular-assisted breeding.

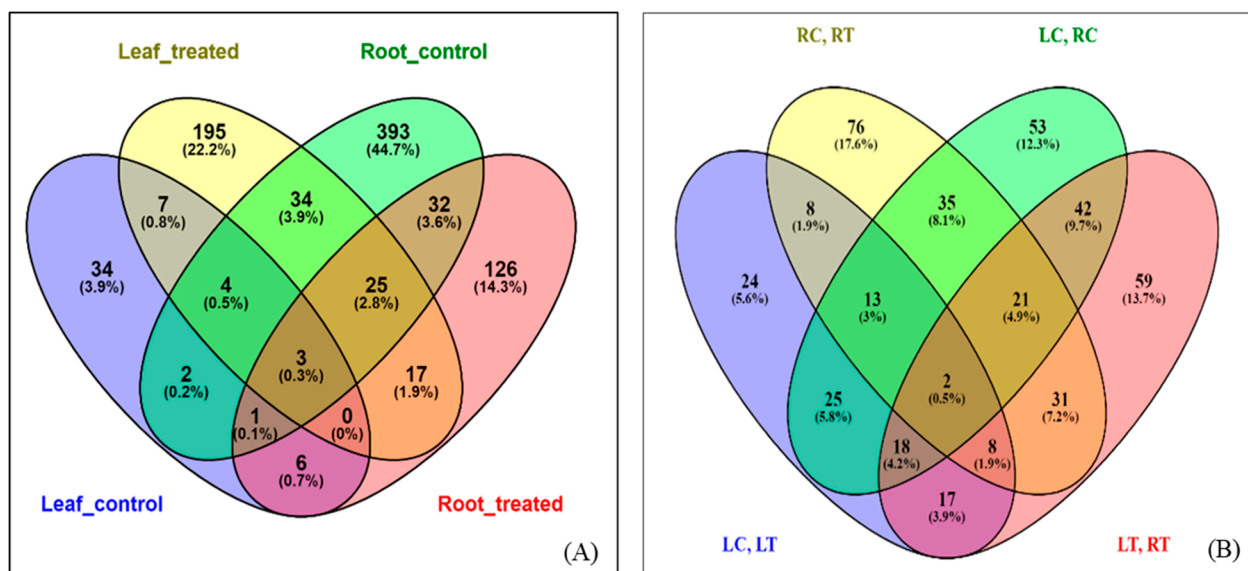


**Figure 2.** Volcano plots (A) LC:RC (B) LT:RT (C) LC:LT and (D) RC:RT.



**Figure 3.** (A) Chromosome-wise distribution of lncRNAs in pearl millet. Red dots represent intergenic lncRNA, blue dots as intronic lncRNA and green dots as exonic overlap on the opposite strand lncRNA (B) Graphical representation of distribution of lncRNAs over the pearl millet chromosomes on the opposite strand lncRNA.

The analysis showed 24 (5.6%), 76 (17.6%), 53 (12.3%), and 59 (13.7%) unique lncRNAs expressed under LC:LT, RC:RT, LC:RC and LT:RT, respectively, while 2 (0.5%) lncRNAs were commonly expressed in all the combinations (Figure 4). The top 50 differentially expressed lncRNAs were represented as heatmap under LC:LT and RC:RT comparison (Figure 5).



**Figure 4.** Venn diagram representing shared and unique (A) lncRNAs under four conditions and (B) differentially expressed lncRNAs in four comparisons.



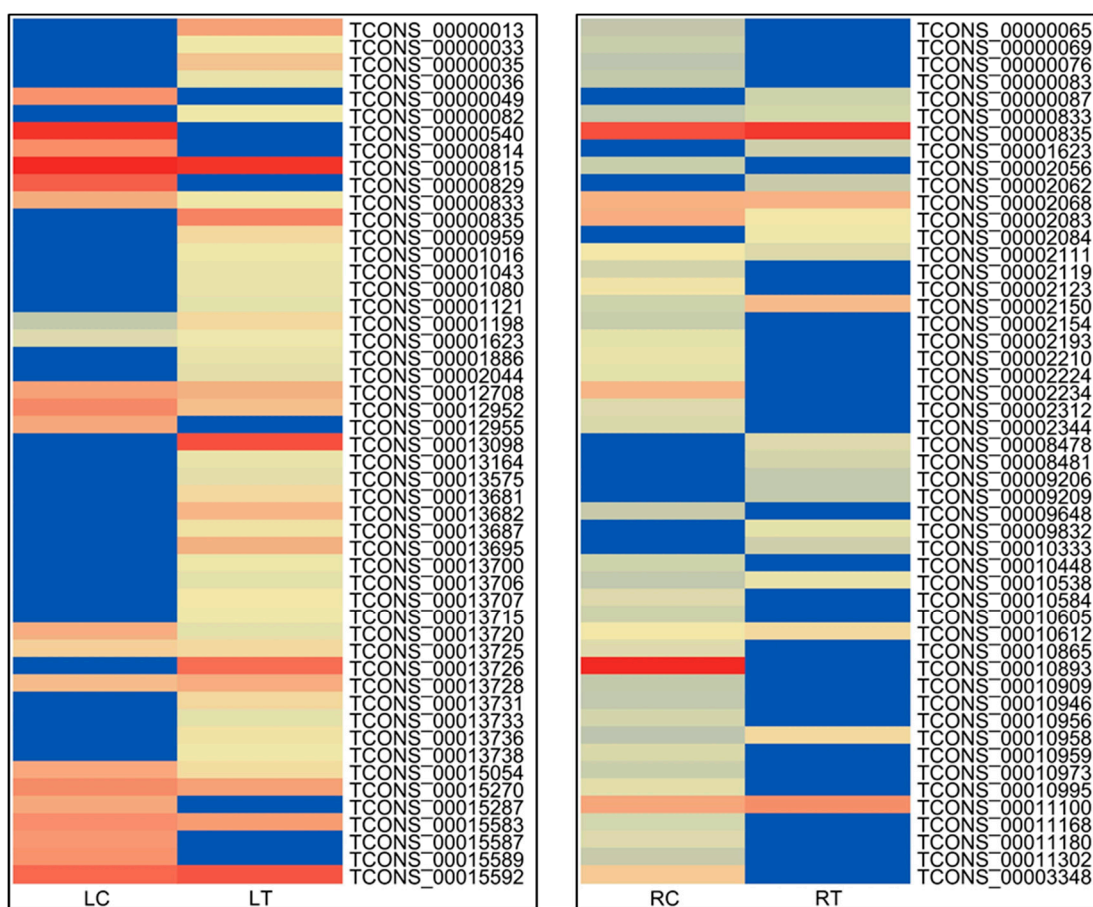


Figure 5. Heat map of the differentially expressed lncRNAs under LC:LT and RC: RT.

To test if these lncRNAs could regulate the expression of protein-coding genes as long molecules, lncRNAs were tested for homology to CDS sequences by BLAST with >87% percent identity and >83% query coverage. The result shows that one lncRNA could pair with eight CDS sequences with a very good match, which suggests that this one lncRNA might regulate the expression of eight proteins by inducing transcriptional or post-transcriptional gene silencing. The detailed information of these lncRNAs when mapped with the CDS sequences of pearl millet is presented in Table 3.

Table 3. Blast result of lncRNAs with the CDS sequences of pearl millet.

Query ID	Seq ID	Q Coverage	P-Identity	Length	QStart	QEnd	SeqStart	SeqEnd	E-Value
TCONS_00024950	Pgl_GLEAN_10035860	99	87.793	213	1	212	540	751	$1.02 \times 10^{-65}$
TCONS_00024950	Pgl_GLEAN_10005667	84	91.16	181	1	181	711	891	$3.67 \times 10^{-65}$
TCONS_00024950	Pgl_GLEAN_10005566	84	91.11	180	2	181	688	867	$1.32 \times 10^{-64}$
TCONS_00024950	Pgl_GLEAN_10016655	84	90.608	181	1	181	273	453	$1.71 \times 10^{-63}$
TCONS_00024950	Pgl_GLEAN_10032756	84	88.333	180	2	181	805	984	$2.88 \times 10^{-56}$
TCONS_00024950	Pgl_GLEAN_10030411	84	87.845	181	1	181	981	1161	$3.72 \times 10^{-55}$
TCONS_00024950	Pgl_GLEAN_10008530	83	88.202	178	2	179	529	706	$3.72 \times 10^{-55}$
TCONS_00024950	Pgl_GLEAN_10036781	83	87.079	178	2	179	55	232	$8.06 \times 10^{-52}$

Similarity search of predicted lncRNAs with the known plant lncRNA database CANATAdb keeping e-value less than e-10 (<http://cantata.amu.edu.pl/index.html>) (accessed on 1 June 2021), showed 8 out of 879 (~0.001%) lncRNAs identified matched with

entries in the database which shows lncRNAs are poorly conserved among species. *Oryza nivara* lncRNAs were found to be more closely related to pearl millet lncRNAs (Table 4).

**Table 4.** Significant blast results from CANATAdb.

Sequence ID	Cantana DB Seq ID	Length	QStart	QEnd	SStart	SEnd	Organism
TCONS_00000829	CNT2084344	236	1	232	524	758	<i>Oryza nivara</i>
TCONS_00000829	CNT2084343	236	1	232	524	758	<i>Oryza nivara</i>
TCONS_00012492	CNT20135897	207	7	213	14	220	<i>Oryza rufipogon</i>
TCONS_00012492	CNT2081201	207	7	213	1936	2142	<i>Oryza nivara</i>
TCONS_00012492	CNT20187813	207	7	213	1837	2043	<i>Oryza sativa</i>
TCONS_00012492	CNT20187804	207	7	213	434	228	<i>Oryza sativa</i>
TCONS_00012492	CNT2020493	195	7	201	188	6	<i>Brachypodium distachyon</i>
TCONS_00009216	CNT2099321	220	7	224	229	11	<i>Setaria italica</i>
TCONS_00013098	CNT2099766	408	13	416	40	446	<i>Setaria italica</i>

### 3.4. Identification of miRNAs Targeting the lncRNAs

Blastn of the lncRNA transcripts against the miRBase database was done to find out the miRNAs with lncRNAs as potential targets. A total of three miRNAs targeting the lncRNAs were found, out of which two were unique. Among the identified differentially expressed lncRNAs, only two lncRNAs (TCONS\_00024337 and TCONS\_00026046) were targeted by the two miRNAs (dps-mir-2526 and oni-mir-10840). The total number of miRNAs target sites was found to be eight (Table 5).

**Table 5.** MiRNAs targeting identified lncRNAs.

Mirna	Lncrna	Length	Mismatch	MiRNA Start	MiRNA End	LncRNA Start	LncRNA End
dps-mir-2526	TCONS_00024337	27	3	72	98	395	371
dps-mir-2526	TCONS_00024337	27	3	49	75	395	371
dps-mir-2526	TCONS_00024337	27	3	26	52	395	371
dps-mir-2526	TCONS_00024337	27	3	3	29	395	371
oni-mir-10840-1	TCONS_00026046	29	1	27	53	192	219
oni-mir-10840-1	TCONS_00026046	25	0	1	23	218	195
oni-mir-10840-2	TCONS_00026046	26	0	30	53	195	219
oni-mir-10840-2	TCONS_00026046	28	1	1	26	218	192

Identification of target mRNAs for the predicted miRNAs was also performed. It was found that two of the three predicted miRNAs were found to target different mRNAs of pearl millet. The number of unique mRNAs targeted was 14. The number of interactions between predicted miRNAs and mRNAs was 16 and all the interactions were found to be cleavage inhibitive in nature (Table 6).

**Table 6.** Identified mRNAs for the miRNAs.

MiRNA_Acc.	Target_Acc.	UPE\$	MiRNA Start	MiRNA End	mRNA Start	mRNA End	Aligned mRNA Sequence	Inhibition
dps-miR-2526-3p	JT845468.1	-1	1	20	121	140	GCCGUCGGCGUGUUGCGGUG	Cleavage
dps-miR-2526-3p	JT845783.1	-1	1	20	308	327	GCCGUCGGCGUGUUGCGGUG	Cleavage
dps-miR-2526-3p	KR908724.1	-1	1	20	770	789	GCCGUCGGCGUGUUGCGGUG	Cleavage
dps-miR-2526-3p	CD725410.1	-1	1	20	187	206	GCCGUCGGCGUGUUGCGGUG	Cleavage
dps-miR-2526-3p	JT844189.1	-1	1	20	150	169	GCCGUCGGCGUGUUGCGGUG	Cleavage
dps-miR-2526-3p	CD725449.1	-1	1	20	20	40	GCCGUCGGCGUGUUGCGGUG	Cleavage
dps-miR-2526-3p	CD725265.1	-1	1	20	346	365	GCCGUCGGCGUGUUGCGGUG	Cleavage
dps-miR-2526-3p	JZ681148.1	-1	1	20	359	378	GCCGUCGGCGUGUUGCGGUG	Cleavage
dps-miR-2526-3p	JZ681188.1	-1	1	20	375	394	GCCGUCGGCGUGUUGCGGUG	Cleavage
dps-miR-2526-3p	EB410974.1	-1	1	20	379	398	GCCGUCGGCGUGUUGCGGUG	Cleavage
dps-miR-2526-3p	JZ681109.1	-1	1	20	102	121	GCCGUCGGCGUGUUGCGGUG	Cleavage
dps-miR-2526-3p	CD724914.1	-1	1	20	47	66	GCCGUCGGCGUGUUGCGGUG	Cleavage
oni-miR-10840	CD725654.1	-1	1	22	199	220	AAUGAGCUCAUUUUGAGAACUU	Cleavage
oni-miR-10840	HQ214676.1	-1	1	22	1582	1603	AAUGAGCUCAUUUUGAGAACUU	Cleavage
oni-miR-10840	CD725654.1	-1	1	22	199	220	AAUGAGCUCAUUUUGAGAACUU	Cleavage
oni-miR-10840	HQ214676.1	-1	1	22	1582	1603	AAUGAGCUCAUUUUGAGAACUU	Cleavage

To understand the functional role of mRNAs involved in the network, the GO analysis was performed by employing the BLAST2GO software. The GO functional categorization generated 10 annotations from the 14 predicted mRNAs that were targeted by miRNAs. In that, a total of three, four, and three mRNAs were classified as the first level classification of biological processes, molecular functions, and cellular components, respectively. Among the genes involved in the biological process, three mRNAs each were classified into the categories of “metabolic process (GO: 0008152)” and “cellular process (GO: 0009987)”. In the classification of molecular functions, two main classes were “binding (GO: 0005488)” and “catalytic activity (GO: 0003824)”, which had three and two predicted mRNAs, respectively. When the predicted mRNAs were classified according to the cellular component classification, categories “cell (GO: 0005623)” and “cell part (GO: 0044464)” both made up the largest proportion of three predicted mRNAs, followed by “organelle (GO: 0043226)” that had two predicted mRNAs. The GO analysis on predicted mRNAs showed that the targets of lncRNAs under drought stress were associated with various functions involving different cellular components, biological processes and molecular functions.

The potential functions of the mRNAs categorized in the GO terms were also found. GO-based functions of mRNAs in pearl millet-drought were identified as Orcinol O-methyl transferase activity, Lignin biosynthetic process, Melatonin biosynthetic process, flavone and flavonoid biosynthesis process, phenylpropanoid biosynthesis pathway, Integral components of membrane and Structural constituent of ribosome. These functions carried out by the mRNAs are involved in the regulation of drought stress responsive mechanisms in soybean. So, when the miRNAs that are responsible for silencing these mRNAs get bind to the circRNAs, they remain no more available to regulate the mRNAs and so the genes become free to carry out their functions and help the plants in tolerating the drought stress conditions.

### 3.5. Development of Pearl Millet lncRNAs Database (PMDlncRDB)

The PMDlncRDB (URL: <http://webtom.cabgrid.res.in/pmdlncrdb/> (accessed on 10 January 2021)) is a three-tier architecture database containing information about the drought-responsive lncRNAs of pearl millet crop. It contains information on 879 lncRNAs from four samples, namely, leaf control, leaf treated, root control and root treated, miRNAs targeting it, target mRNAs of the miRNAs, etc. Users can browse through the submitted lncRNAs on several criteria like length, number of exons, FPKM values and

their position with respect to the coding genes. The submitted lncRNAs have length values between 200 and 1013 nucleotides. Figure 6 is the web interface of PMDlncRDB for various search options.

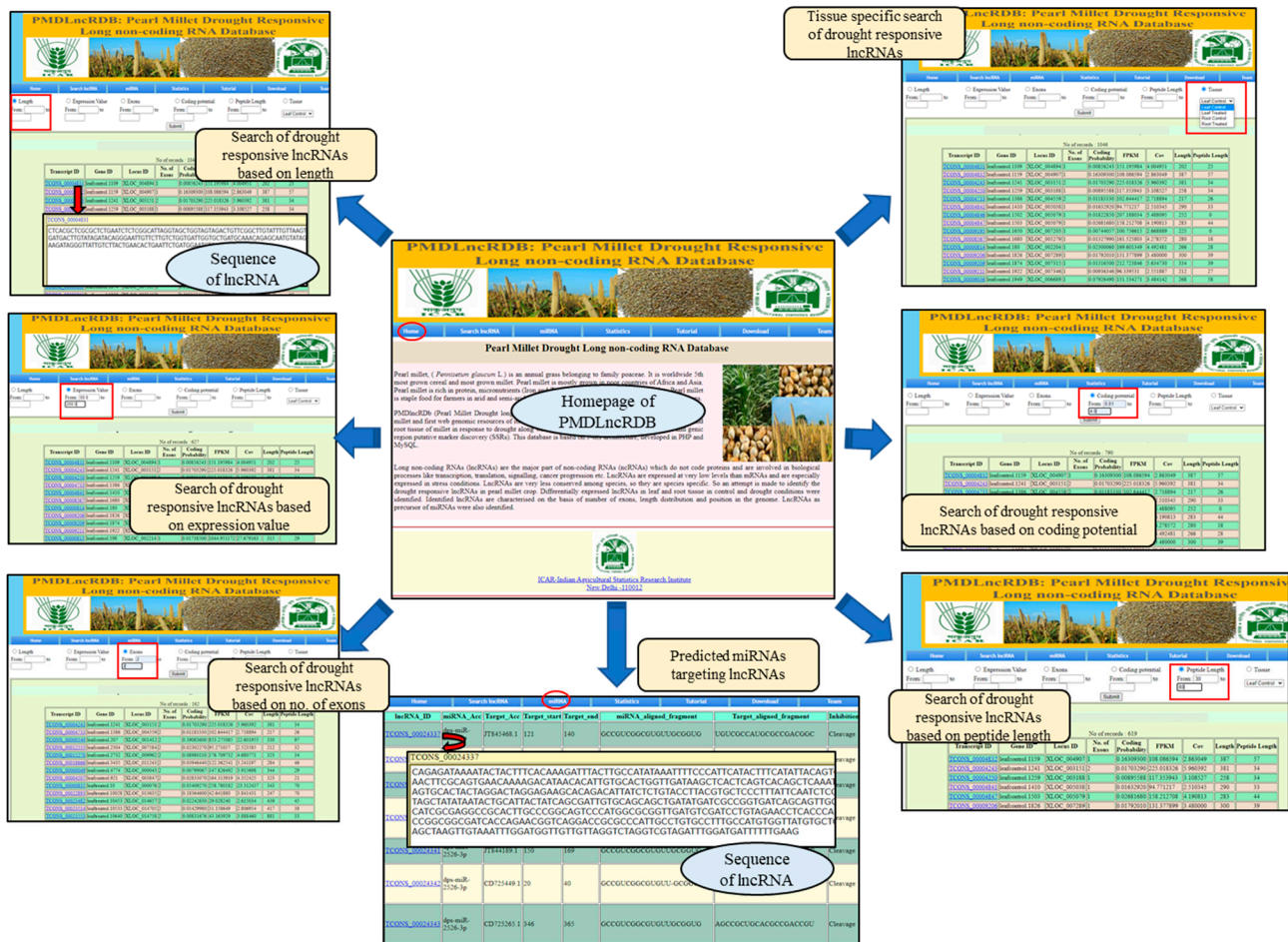


Figure 6. Web genomic resource PMDlncRDB.

## 4. Discussion

### 4.1. Data Pre-Processing

Pre-processing of the Next Generation sequencing data before going down further in the RNA-seq pipeline is very important as the raw data may contain low-quality reads, noise, adaptor sequences, etc. which deteriorates the quality of assembly. RNA-seq data for the four conditions of the pearl millet crop was downloaded from the SRA database of NCBI and visualized using the FastQC tool [38] which gives out graphs and statistical figures for the assessment of the sequence reads. Trimmomatic [39] software with single-end read parameters was used to filter out the bad quality reads to finally give the cleaned trimmed reads, which were further used in the analysis.

### 4.2. Candidate lncRNA Prediction

Although the role of lncRNAs in drought stress has been reported in plants like wheat [46], cassava [17], etc. its role in pearl millet drought tolerance is still under cover. Though transcriptome studies on proteome [58] and transcriptome signature [35] of pearl millet against the drought stress have been conducted, a comprehensive study on pearl millet’s lncRNAs contributing to its drought tolerance needs to be investigated knowing the role of lncRNAs in stress conditions already identified in Brassica napus [59], rice [13], mulberry [10], tomato [60], Populus [61], wheat [62], etc. In our study for the steps of



mapping the reads to the reference genome to assembly and estimating the abundance of the assembled transcripts, “new tuxedo” package [42] programs have been used which includes HISAT2 for reads alignment, StringTie for transcripts assembly and abundance estimation. In our study of the identification of drought-responsive lncRNAs, we found only 879 lncRNAs. The reason for the lesser number of lncRNAs is the expression of lncRNAs at very low levels as in the case of tomato [63]. Most of the lncRNAs identified have only one exon in them, which may be due to their smaller size. We also observed a few lncRNAs having 3–4 exons, but they are lesser in number. The identified lncRNAs were divided into three groups on the basis of their position with respect to the protein-coding genes in the reference genome of pearl millet. Intergenic lncRNAs are present in between the coding genes, intronic lncRNAs which are completely in intronic regions and the lncRNAs having exonic overlaps on the opposite strand. Most of the lncRNAs were found to be intergenic in nature. Our study identified 57, 285, 594 and 210 lncRNAs in leaf control, leaf treated, root control and root treated conditions, respectively. The highest number of lncRNAs in root control conditions is potentially due to the much role of root-related mechanisms in the growth and development of plants. Under normal conditions, we expect a higher number of lncRNAs in the roots of plants, especially during the stress condition, but in our study, a limited number of lncRNAs were identified, which may be because of certain growth and development related pathways being slowed down or switched off. Under water stress conditions, the first response of plants is the closing of stomata, which in turn controls the CO<sub>2</sub> uptake, photosynthesis and transpiration of water [58]. We got a higher number of lncRNAs in leaf-treated conditions in our study in response to self-defense against drought.

#### 4.3. Identification of Differentially Expressed lncRNAs and Their Annotation

To compare whether the lncRNAs express differently in roots and leaves in control and treated (drought) conditions, four comparisons were made to compare the difference in expressions in roots and leaves in control and treated conditions individually and also the difference in lncRNAs expression in root and leaves in control and treated conditions separately. We found 209 (67 upregulated, 142 downregulated), 198 (123 upregulated, 75 downregulated), 115 (42 upregulated, 69 downregulated) and 194 (116 upregulated, 78 downregulated) differentially expressed lncRNAs in LC:RC (leaf control, root control), LT:RT (leaf treated, root treated), LC:LT (leaf control, leaf treated) and RC:RT (root control, root treated) conditions, respectively. Differential expression of lncRNAs clearly depicts its regulatory roles in drought response in pearl millet.

A similarity search was performed against the known plant database CANTATAdb which resulted in only 9 instances of our identified lncRNAs matching with any previously existing lncRNAs in the database. This result is in concordance with the studies of Ma et al. (2013) [64] and Sahu et al. (2018) [65], which stated that as compared to other RNA transcripts like mRNAs, miRNAs, snoRNAs, etc. lncRNAs are poorly conserved. Ma et al. (2013) [64] explain this as lncRNAs being functional in a species-specific manner. Most of the matches were found to be with *Oryza nivara*, which is also a member of the poaceae family, this also restricts the conserved nature of lncRNAs to species other than the same family. Our analysis shows that most of the lncRNAs of pearl millet are specific to the pearl millet itself. Studies on lncRNAs with well-defined functions are reported, for example, LDMAR, cis-NAT, PHO1;2, TL, LAIR in rice [66–69], Enod4 in rice, maize, legumes and soybean [70–72], HvCesa6 in barley [73] and WSGAR in wheat [74]. lncRNAs have also been seen to regulate the expression of protein-coding genes as long molecules [75]. Our lncRNAs when blasted with the CDS sequences of pearl millet showed only one result of lncRNA corresponding to eight protein-coding genes of pearl millet, which supports that the lncRNA can regulate the function of those proteins by transcriptional or post-transcriptional silencing [76]. Also, studies show lncRNAs to be important modulators of drought tolerance in plants as reported in various crops like rice, wheat, maize, sorghum, tomato, coffee, cassava and peanut, etc. [77]. The identified lncRNAs in this study can be of immense use for future studies of their association with gene expression.



#### 4.4. Identification of miRNAs Targeting the lncRNAs

To evaluate whether lncRNAs in pearl millet could affect post-transcriptional regulation of functional genes by binding to miRNAs, the bioinformatics methods were employed to identify the lncRNA-originating target mimics in pearl millet based on the differentially expressed lncRNAs. In our study, we found three miRNAs targeting the lncRNAs [78]. A total of eight interactions between the lncRNAs and miRNAs were found.

#### 4.5. lncRNA-miRNA-mRNA Interaction

A total of 14 mRNAs were found to be potential targets for the identified miRNAs. The functions of the predicted mRNAs as reported by BLAST2GO in pearl millet under drought showed that the mRNAs were involved in plant hormone signal transduction, response to stress, defense response mechanism, transcription factor activity, Orcinol O-methyl transferase activity, Lignin biosynthetic process, Metalonin biosynthetic process, Flavanol biosynthetic process, Integral components of membrane, Structural constituent of ribosome, etc. [78]. Flavone and flavonoid biosynthesis pathways (Pathway ID: ko00944; Ec:2.1.1.42) found in our study are reported to have key significance in plants under drought stress [79], specifically in *Ligustrum vulgare* [80] and peanut [81]. Similarly, the phenylpropanoid biosynthesis (Pathway ID: ko00940, Ec:2.1.1.68) reported in our study is in concordance with the previous studies related to drought stress in crops like apple [82], foxtail millet [83] and tomato [84].

#### 4.6. Development of Pearl Millet lncRNAs Database (PMDlncRDB)

There exists a drought-associated web genomic resource of pearl millet, *PMDDb* but there is no information on lncRNAs (Jaiswal et al., 2018). The developed web genomic resource, *PMDlncRDB* catalogs the identified drought-responsive lncRNAs. It is the first such resource for pearl millet lncRNAs. It contains information like lncRNA sequence, ID, length, peptide length, number of exons, miRNAs targeting the lncRNAs, etc. This will help researchers as a foundation for understanding the role of lncRNAs in pearl millet crop and further studies in improving crop performance in drought stress through selective breeding approaches. The developed genomic resource of lncRNAs in this study can be of immense use for future studies of their association with gene expression.

## 5. Conclusions

The study deals with the drought-responsive long non-coding RNAs in pearl millet, which is the staple food for >90 million poverty-driven farmers. The drought-associated genome-wide lncRNAs were identified in pearl millet through the transcriptome data generated from root and leaf tissues. A total of 1046 candidate lncRNAs were predicted with reference-based mapping. We report 879 differentially expressed lncRNAs under the four comparison sets (LC:RC, LT:RT; LC:LT and RC:RT). Three miRNAs were found targeting these lncRNAs and a total of 14 mRNAs had potential target mimics of these miRNAs. The developed web-genomic resource of pearl millet lncRNAs (*PMDlncRDB*) cataloging these drought-responsive lncRNAs is the first such resource that can be used in trait improvement as well as CRISPR-Cas technology in the editing of ncRNAs. This will help in the further investigation of lncRNA-related studies in pearl millet improvement programs for drought tolerance in endeavors of higher production.

**Author Contributions:** D.K. and S.J. conceptualized the study. B.K., A.K., S.J., M.A.I., U.B.A. performed the data curation, data analysis and development of web-resource. B.K., S.J., M.A.I. and D.K. drafted the paper. S.J., D.K., A.R. and R.S.T. reviewed and edited the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by CABIn Grant [Grant number: F. No. Agril. Edn.4-1/2013-A&P] by Indian Council of Agricultural Research, Ministry of Agriculture and Farmers' Welfare, Govt. of India. The grant of fellowship (National Talent Scholarship) to BK by the Indian Council of Agricultural Research, New Delhi, is duly acknowledged.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data used in the study are from the public domain, which has been mentioned along with the source in the manuscript.

**Acknowledgments:** Authors are thankful to the Indian Council of Agricultural Research, Ministry of Agriculture and Farmers' Welfare, Govt. of India for the creation of the Advanced Super Computing Hub for Omics Knowledge in Agriculture (ASHOKA) facility where the work was carried out.

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

## References

1. Wang, K.C.; Chang, H.Y. Molecular mechanisms of long noncoding RNAs. *Mol. Cell* **2011**, *43*, 904–914. [[CrossRef](#)] [[PubMed](#)]
2. Mattick, J.S. RNA regulation: A new genetics? *Nat. Rev. Genet.* **2004**, *5*, 316. [[CrossRef](#)] [[PubMed](#)]
3. Mercer, T.R.; Dinger, M.E.; Mattick, J.S. Long non-coding RNAs: Insights into functions. *Nat. Rev. Genet.* **2009**, *10*, 155–159. [[CrossRef](#)] [[PubMed](#)]
4. Tay, Y.; Rinn, J.; Pandolfi, P.P. The multilayered complexity of ceRNA crosstalk and competition. *Nature* **2014**, *505*, 344–352. [[CrossRef](#)]
5. Amaral, P.P.; Clark, M.B.; Gascoigne, D.K.; Dinger, M.E.; Mattick, J.S. lncRNADB: A reference database for long noncoding RNAs. *Nucleic Acids Res.* **2010**, *39* (Suppl. S1), D146–D151. [[CrossRef](#)]
6. Quinn, J.J.; Chang, H.Y. Unique features of long non-coding RNA biogenesis and function. *Nat. Rev. Genet.* **2016**, *17*, 47–62. [[CrossRef](#)]
7. Wilusz, J.E.; Sunwoo, H.; Spector, D.L. Long noncoding RNAs: Functional surprises from the RNA world. *Genes Dev.* **2009**, *23*, 1494–1504. [[CrossRef](#)]
8. Amor, B.B.; Wirth, S.; Merchan, F.; Laporte, P.; d'Aubenton-Carafa, Y.; Hirsch, J.; Maizel, A.; Mallory, A.; Lucas, A.; Deragon, J.M.; et al. Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses. *Genome Res.* **2009**, *19*, 57–69. [[CrossRef](#)]
9. De Lucia, F.; Dean, C. Long non-coding RNAs and chromatin regulation. *Curr. Opin. Plant Biol.* **2011**, *14*, 168–173. [[CrossRef](#)]
10. Song, X.; Sun, L.; Luo, H.; Ma, Q.; Zhao, Y.; Pei, D. Genome-wide identification and characterization of long non-coding RNAs from mulberry (*Morus notabilis*) RNA-seq Data. *Genes* **2016**, *7*, 11. [[CrossRef](#)]
11. Chen, L.L.; Carmichael, G.G. Decoding the function of nuclear long non-coding RNAs. *Curr. Opin. Cell Biol.* **2010**, *22*, 357–364. [[CrossRef](#)] [[PubMed](#)]
12. Liu, J.; Jung, C.; Xu, J.; Wang, H.; Deng, S.; Bernad, L.; Chua, N.H. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* **2012**, *24*, 4333–4345. [[CrossRef](#)]
13. Zhang, Y.C.; Liao, J.Y.; Li, Z.Y.; Yu, Y.; Zhang, J.P.; Li, Q.F.; Chen, Y.Q. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol.* **2014**, *15*, 512. [[CrossRef](#)] [[PubMed](#)]
14. Zhao, X.; Liu, X.; Guo, C.; Gu, J.; Xiao, K. Identification and characterization of microRNAs from wheat (*Triticum aestivum* L.) under phosphorus deprivation. *Biotechnology* **2013**, *22*, 113–123. [[CrossRef](#)]
15. Li, L.; Eichten, S.R.; Shimizu, R.; Petsch, K.; Yeh, C.T.; Wu, W.; Muehlbauer, G.J. Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol.* **2014**, *15*, R40. [[CrossRef](#)] [[PubMed](#)]
16. Tan, X.; Li, S.; Hu, L.; Zhang, C. Genome-wide analysis of long non-coding RNAs (lncRNAs) in two contrasting rapeseed (*Brassica napus* L.) genotypes subjected to drought stress and re-watering. *BMC Plant Biol.* **2020**, *20*, 81. [[CrossRef](#)] [[PubMed](#)]
17. Li, S.; Yu, X.; Lei, N.; Cheng, Z.; Zhao, P.; He, Y.; Peng, M. Genome-wide identification and functional prediction of cold and/or drought-responsive lncRNAs in cassava. *Sci. Rep.* **2017**, *7*, 45981. [[CrossRef](#)] [[PubMed](#)]
18. Chung, P.J.; Jung, H.; Jeong, D.H.; Ha, S.H.; Choi, Y.D.; Kim, J.K. Transcriptome profiling of drought responsive noncoding RNAs and their target genes in rice. *BMC Genom.* **2016**, *17*, 563. [[CrossRef](#)]
19. Zhang, W.; Han, Z.; Guo, Q.; Liu, Y.; Zheng, Y.; Wu, F.; Jin, W. Identification of maize long non-coding RNAs responsive to drought stress. *PLoS ONE* **2014**, *9*, e98958. [[CrossRef](#)]
20. Qi, X.; Xie, S.; Liu, Y.; Yi, F.; Yu, J. Genome-wide annotation of genes and noncoding RNAs of foxtail millet in response to simulated drought stress by deep sequencing. *Plant Mol. Biol.* **2013**, *83*, 459–473. [[CrossRef](#)]
21. Zhang, C.; Tang, G.; Peng, X.; Sun, F.; Liu, S.; Xi, Y. Long non-coding RNAs of switchgrass (*Panicum virgatum* L.) in multiple dehydration stresses. *BMC Plant Biol.* **2018**, *18*, 79. [[CrossRef](#)] [[PubMed](#)]
22. Yan, Q.; Wu, F.; Yan, Z.; Li, J.; Ma, T.; Zhang, Y.; Zhang, J. Differential co-expression networks of long non-coding RNAs and mRNAs in *Cleistogenes songorica* under water stress and during recovery. *BMC Plant Biol.* **2019**, *19*, 23. [[CrossRef](#)] [[PubMed](#)]
23. Weidong, Q.I.; Hongping, C.; Zuozhen, Y.; Biaolin, H.U.; Xiangdong, L.; Bing, A.; Fantao, Z. Systematic characterization of long non-coding RNAs and their responses to drought stress in Dongxiang wild rice. *Rice Sci.* **2020**, *27*, 21–31. [[CrossRef](#)]

24. Jha, U.C.; Nayyar, H.; Jha, R.; Khurshid, M.; Zhou, M.; Mantri, N.; Siddique, K.H. Long non-coding RNAs: Emerging players regulating plant abiotic stress response and adaptation. *BMC Plant Biol.* **2020**, *20*, 466. [CrossRef] [PubMed]
25. Jin, J.; Lu, P.; Xu, Y.; Li, Z.; Yu, S.; Liu, J.; Cao, P. PLncDB V2. 0: A comprehensive encyclopedia of plant long noncoding RNAs. *Nucleic Acids Res.* **2021**, *49*, D1489–D1495. [CrossRef] [PubMed]
26. Gallart, A.P.; Pulido, A.H.; De Lagrán, I.A.M.; Sanseverino, W.; Cigliano, R.A. GREENC: A Wiki-based database of plant lncRNAs. *Nucleic Acids Res.* **2016**, *44*, D1161.
27. Zhou, B.; Zhao, H.; Yu, J.; Guo, C.; Dou, X.; Song, F.; Wang, J. EVLncRNAs: A manually curated database for long non-coding RNAs validated by low-throughput experiments. *Nucleic Acids Res.* **2018**, *46*, D100–D105. [CrossRef]
28. Quek, X.C.; Thomson, D.W.; Maag, J.L.; Bartonicek, N.; Signal, B.; Clark, M.B.; Dinger, M.E. lncRNAdb v2.0: Expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.* **2015**, *43*, D168–D173. [CrossRef]
29. Xuan, H.; Zhang, L.; Liu, X.; Han, G.; Li, J.; Li, X.; Zhang, S. PLNlncRbase: A resource for experimentally identified lncRNAs in plants. *Gene* **2015**, *573*, 328–332. [CrossRef]
30. Szcześniak, M.W.; Bryzghalov, O.; Ciomborowska-Basheer, J.; Makałowska, I. CANTATAdb 2.0: Expanding the collection of plant long noncoding RNAs. *Methods Mol. Biol.* **2019**, *1933*, 415–429.
31. Sweeney, B.A.; Tagmazian, A.A.; Ribas, C.E.; Finn, R.D.; Bateman, A.; Petrov, A.I. Exploring Non-Coding RNAs in RNAcentral. *Curr. Protoc. Bioinform.* **2020**, *71*, e104. [CrossRef] [PubMed]
32. Satyavathi, C.T.; Ambawat, S.; Khandelwal, V.; Srivastava, R.K. Pearl Millet: A Climate-Resilient Nutricereal for Mitigating Hidden Hunger and Provide Nutritional Security. *Front. Plant Sci.* **2021**, *12*, 1828. [CrossRef] [PubMed]
33. Di Stefano, E.; White, J.; Seney, S.; Hekmat, S.; McDowell, T.; Sumarah, M.; Reid, G. A Novel Millet-Based Probiotic Fermented Food for the Developing World. *Nutrients* **2017**, *22*, 529. [CrossRef]
34. Sloan, A.E. Positive eating and problem treating: Nutraceuticals and cereal-based foods in the 21st century. *Cereal Foods World* **1999**, *44*, 746–750.
35. Jaiswal, S.; Antala, T.J.; Mandavia, M.K.; Chopra, M.; Jasrotia, R.S.; Tomar, R.S.; Kheni, J.; Angadi, U.B.; Iquebal, M.A.; Golakia, B.A.; et al. Transcriptomic signature of drought response in pearl millet (*Pennisetum glaucum* (L.) and development of web-genomic resources. *Sci. Rep.* **2018**, *8*, 3382. [CrossRef] [PubMed]
36. Zhou, R.; Sanz-Jimenez, P.; Zhu, X.T.; Feng, J.W.; Shao, L.; Song, J.M.; Chen, L.L. Analysis of rice transcriptome reveals the lncRNA/CircRNA regulation in tissue development. *Rice* **2021**, *14*, 14. [CrossRef] [PubMed]
37. Basak, J.; Nithin, C. Targeting non-coding RNAs in plants with the CRISPR-Cas technology is a challenge yet worth accepting. *Front. Plant Sci.* **2015**, *6*, 1001. [CrossRef]
38. Andrews, S. FastQC, a Quality Control Tool for High throughput Sequence Data. 2010. Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 12 May 2021).
39. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]
40. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [CrossRef]
41. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **2015**, *33*, 290–295. [CrossRef]
42. Pertea, M.; Kim, D.; Pertea, G.M.; Leek, J.T.; Salzberg, S.L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **2016**, *11*, 1650–1667. [CrossRef] [PubMed]
43. Roberts, A.; Trapnell, C.; Donaghey, J.; Rinn, J.L.; Pachter, L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol.* **2011**, *12*, R22. [CrossRef] [PubMed]
44. Min, X.J.; Butler, G.; Storms, R.; Tsang, A. OrfPredictor: Predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Res.* **2005**, *33* (Suppl. S2), W677–W680. [CrossRef] [PubMed]
45. Kang, Y.J.; Yang, D.C.; Kong, L.; Hou, M.; Meng, Y.Q.; Wei, L.; Gao, G. CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **2017**, *45*, W12–W16. [CrossRef] [PubMed]
46. Shumayla, S.S.; Taneja, M.; Tyagi, S.; Singh, K.; Upadhyay, S.K. Survey of High Throughput RNA-Seq Data Reveals Potential Roles for lncRNAs during Development and Stress Response in Bread Wheat. *Front. Plant Sci.* **2017**, *8*, 1019. [CrossRef]
47. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]
48. Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The protein families database. *Nucleic Acids Res.* **2013**, *42*, D222–D230. [CrossRef]
49. Finn, R.D.; Clements, J.; Arndt, W.; Miller, B.L.; Wheeler, T.J.; Schreiber, F.; Bateman, A.; Eddy, S.R. HMMER web server: 2015 update. *Nucleic Acids Res.* **2015**, *43*, W30–W38. [CrossRef]
50. Tarazona, S.; Furió-Tari, P.; Turra, D.; Pietro, A.D.; Nueda, M.J.; Ferrer, A.; Conesa, A. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res.* **2015**, *43*, e140.
51. Szcześniak, M.W.; Rosikiewicz, W.; Makałowska, I. CANTATAdb: A collection of plant long non-coding RNAs. *Plant Cell Physiol.* **2015**, *57*, e8. [CrossRef]
52. Griffiths-Jones, S. miRBase: The microRNA sequence database. *MicroRNA Protoc.* **2006**, *342*, 129–138.
53. Hofacker, I.L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **2003**, *31*, 3429–3431. [CrossRef] [PubMed]

54. Dai, X.; Zhuang, Z.; Zhao, P.X. psRNATarget: A plant small RNA target analysis server (2017 release). *Nucleic Acids Res.* **2018**, *46*, W49–W54. [[CrossRef](#)]
55. Wang, Y.; Gao, L.; Zhu, B.; Zhu, H.; Luo, Y.; Wang, Q.; Zuo, J. Integrative analysis of long non-coding RNA acting as ceRNAs involved in chilling injury in tomato fruit. *Gene* **2018**, *667*, 25–33. [[CrossRef](#)] [[PubMed](#)]
56. Shuai, P.; Liang, D.; Tang, S.; Zhang, Z.; Ye, C.Y.; Su, Y. Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. *J. Exp. Bot.* **2014**, *65*, 4975–4983. [[CrossRef](#)] [[PubMed](#)]
57. Pauli, A.; Valen, E.; Lin, M.F.; Garber, M.; Vastenhouw, N.L.; Levin, J.Z. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* **2012**, *22*, 577–591. [[CrossRef](#)]
58. Ghatak, A.; Chaturvedi, P.; Nagler, M.; Roustan, V.; Lyon, D.; Bachmann, G.; Weckwerth, W. Comprehensive tissue-specific proteome analysis of drought stress responses in *Pennisetum glaucum* (L.) R. Br. (Pearl millet). *J. Proteom.* **2016**, *143*, 122–135. [[CrossRef](#)]
59. Joshi, R.K.; Megha, S.; Basu, U.; Rahman, M.H.; Kav, N.N. Genome wide identification and functional prediction of long non-coding RNAs responsive to *Sclerotinia sclerotiorum* infection in Brassica napus. *PLoS ONE* **2016**, *11*, e0158784. [[CrossRef](#)]
60. Wang, J.; Yu, W.; Yang, Y.; Li, X.; Chen, T.; Liu, T.; Ma, N.; Yang, X.; Liu, R.; Zhang, B. Genome-wide analysis of tomato long non-coding RNAs and identification as endogenous target mimic for microRNA in response to TYLCV infection. *Sci. Rep.* **2015**, *5*, 16946. [[CrossRef](#)]
61. Chen, M.; Wang, C.; Bao, H.; Chen, H.; Wang, Y. Genome-wide identification and characterization of novel lincRNAs in Populus under nitrogen deficiency. *Mol. Genet. Genom.* **2016**, *291*, 1663–1680. [[CrossRef](#)]
62. Zhang, H.; Chen, X.; Wang, C.; Xu, Z.; Wang, Y.; Liu, X.; Ji, W. Long non-coding genes implicated in response to stripe rust pathogen stress in wheat (*Triticum aestivum* L.). *Mol. Biol. Rep.* **2013**, *40*, 6245–6253. [[CrossRef](#)] [[PubMed](#)]
63. Zhou, Y.; Cho, W.K.; Byun, H.S.; Chavan, V.; Kil, E.J.; Lee, S.; Hong, S.W. Genome-wide identification of long non-coding RNAs in tomato plants irradiated by neutrons followed by infection with Tomato yellow leaf curl virus. *PeerJ* **2019**, *7*, 6286–7013. [[CrossRef](#)] [[PubMed](#)]
64. Ma, L.; Bajic, V.B.; Zhang, Z. On the classification of long non-coding RNAs. *RNA Biol.* **2013**, *10*, 924–933. [[CrossRef](#)] [[PubMed](#)]
65. Sahu, S.; Rao, A.R.; Pandey, J.; Gaikwad, K.; Ghoshal, S.; Mohapatra, T. Genome-wide identification and characterization of lincRNAs and miRNAs in cluster bean (*Cyamopsis tetragonoloba*). *Gene* **2018**, *667*, 112–121. [[CrossRef](#)]
66. Ding, J.; Lu, Q.; Ouyang, Y.; Mao, H.; Zhang, P.; Yao, J.; Zhang, Q. A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 2654–2659. [[CrossRef](#)]
67. Jabnourne, M.; Secco, D.; Lecampion, C.; Robaglia, C.; Shu, Q.; Poirier, Y. A rice cis-natural antisense RNA acts as a translational enhancer for its cognate mRNA and contributes to phosphate homeostasis and plant fitness. *Plant Cell* **2013**, *25*, 4166–4182. [[CrossRef](#)]
68. Liu, X.; Li, D.; Zhang, D.; Yin, D.; Zhao, Y.; Ji, C.; Zhu, L. A novel antisense long noncoding RNA, TWISTED LEAF, maintains leaf blade flattening by regulating its associated sense R2R3-MYB gene in rice. *New Phytol.* **2018**, *218*, 774–788. [[CrossRef](#)]
69. Wang, Y.; Luo, X.; Sun, F.; Hu, J.; Zha, X.; Su, W.; Yang, J. Overexpressing lincRNA LAIR increases grain yield and regulates neighbouring gene cluster expression in rice. *Nat. Commun.* **2018**, *9*, 3516. [[CrossRef](#)]
70. Röhrig, H.; Schmidt, J.; Miklashevichs, E.; Schell, J.; John, M. Soybean ENOD40 encodes two peptides that bind to sucrose synthase. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 1915–1920. [[CrossRef](#)]
71. Campalans, A.; Kondorosi, A.; Crespi, M. Enod40, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. *Plant Cell* **2004**, *16*, 1047–1059. [[CrossRef](#)]
72. Gulyaev, A.P.; Roussis, A. Identification of conserved secondary structures and expansion segments in enod40 RNAs reveals new enod40 homologues in plants. *Nucleic Acids Res.* **2007**, *35*, 3144–3152. [[CrossRef](#)] [[PubMed](#)]
73. Held, M.A.; Penning, B.; Brandt, A.S.; Kessans, S.A.; Yong, W.; Scofield, S.R.; Carpita, N.C. Small-interfering RNAs from natural antisense transcripts derived from a cellulose synthase gene modulate cell wall biosynthesis in barley. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 20534–20539. [[CrossRef](#)]
74. Guo, G.; Liu, X.; Sun, F.; Cao, J.; Huo, N.; Wuda, B.; Yao, Y. Wheat miR9678 affects seed germination by generating phased siRNAs and modulating abscisic acid/gibberellin signaling. *Plant Cell* **2018**, *30*, 796–814. [[CrossRef](#)]
75. Wang, Y.; Xu, T.; He, W.; Shen, X.; Zhao, Q.; Bai, J.; You, M. Genome-wide identification and characterization of putative lincRNAs in the diamondback moth, *Plutella xylostella* (L.). *Genomics* **2017**, *110*, 35–42. [[CrossRef](#)] [[PubMed](#)]
76. Yoon, J.H.; Abdelmohsen, K.; Gorospe, M. Posttranscriptional gene regulation by long noncoding RNA. *J. Mol. Biol.* **2013**, *425*, 3723–3730. [[CrossRef](#)]
77. Gelaw, T.A.; Sanan-Mishra, N. Non-Coding RNAs in Response to Drought Stress. *Int. J. Mol. Sci.* **2021**, *22*, 12519. [[CrossRef](#)] [[PubMed](#)]
78. Wu, P.; Zuo, X.; Deng, H.; Liu, X.; Liu, L.; Ji, A. Roles of long noncoding RNAs in brain development, functional diversification and neurodegenerative diseases. *Brain Res. Bull.* **2013**, *97*, 69–80. [[CrossRef](#)] [[PubMed](#)]
79. Kumar, S.; Pandey, A.K. Chemistry and biological activities of flavonoids: An overview. *Sci. World J.* **2013**, *2013*, 162750. [[CrossRef](#)]
80. Tattini, M.; Galardi, C.; Pinelli, P.; Massai, R.; Remorini, D.; Agati, G. Differential accumulation of flavonoids and hydroxycinnamates in leaves of *Ligustrum vulgare* under excess light and drought stress. *New Phytol.* **2004**, *163*, 547–561. [[CrossRef](#)]
81. Kubra, G.; Khan, M.; Munir, F.; Gul, A.; Shah, T.; Hussain, A.; Amir, R. Expression characterization of flavonoid biosynthetic pathway genes and transcription factors in peanut under water deficit conditions. *Front. Plant Sci.* **2021**, *12*, 1140. [[CrossRef](#)]



82. Geng, D.; Shen, X.; Xie, Y.; Yang, Y.; Bian, R.; Gao, Y.; Guan, Q. Regulation of phenylpropanoid biosynthesis by MdMYB88 and MdMYB124 contributes to pathogen and drought resistance in apple. *Hortic. Res.* **2020**, *7*, 102. [[CrossRef](#)] [[PubMed](#)]
83. Yu, A.; Zhao, J.; Wang, Z.; Cheng, K.; Zhang, P.; Tian, G.; Wang, Y. Transcriptome and metabolite analysis reveal the drought tolerance of foxtail millet significantly correlated with phenylpropanoids-related pathways during germination process under PEG stress. *BMC Plant Biol.* **2020**, *20*, 274. [[CrossRef](#)] [[PubMed](#)]
84. Zhao, T.; Wu, T.; Pei, T.; Wang, Z.; Yang, H.; Jiang, J.; Xu, X. Overexpression of SIGATA17 promotes drought tolerance in transgenic tomato plants by enhancing activation of the phenylpropanoid biosynthetic pathway. *Front. Plant Sci.* **2021**, *12*, 634888. [[CrossRef](#)] [[PubMed](#)]