



# A Comparative Transcriptional Landscape of Two Castor Cultivars Obtained by Single-Molecule Sequencing Comparative Analysis

Yaxing Zhou<sup>1</sup>, Guoli Zhu<sup>2</sup>, Yun Wang<sup>1</sup>, Zhibiao He<sup>2</sup> and Wei Zhou<sup>1\*</sup>

<sup>1</sup>Agricultural College of Inner Mongolia Minzu University, Tongliao, China, <sup>2</sup>Tongliao Academy of Agricultural Science, Tongliao, China

**Background and Objectives:** Castor (*Ricinus communis* L.) is an important non-edible oilseed crop. Lm-type female strains and normal amphiprotic strains are important castor cultivars, and are mainly different in their inflorescence structures and leaf shapes. To better understand the mechanisms underlying these differences at the molecular level, we performed a comparative transcriptional analysis.

**Materials and Methods:** Full-length transcriptome sequencing and short-read RNA sequencing were employed.

**Results:** A total of 76,068 and 44,223 non-redundant transcripts were obtained from high-quality transcripts of Lm-type female strains and normal amphiprotic strains, respectively. In Lm-type female strains and normal amphiprotic strains, 51,613 and 20,152 alternative splicing events were found, respectively. There were 13,239 transcription factors identified from the full-length transcriptomes. Comparative analysis showed a great variety of gene expression of common and unique transcription factors between the two cultivars. Meanwhile, a functional analysis of the isoforms was conducted. The full-length sequences were used as a reference genome, and a short-read RNA sequencing analysis was performed to conduct differential gene analysis. Furthermore, the function of DEGs were performed to annotation analysis.

**Conclusion:** The results revealed considerable differences and expression diversity between the two cultivars, well beyond what was reported in previous studies and likely reflecting the differences in architecture between these two cultivars.

**Keywords:** Lm type female strains, normal amphiprotic strains, pacbio, comparative transcriptome, isoform, full-length transcriptome

## INTRODUCTION

The castor plant (*Ricinus communis* L.), which originated in Africa, is an annual or perennial dicotyledonous. High ricinoleic acid content (80–90%) and high fatty acid content (more than 45%) in its seed oil make it one of the most important non-edible oilseed crops, and this has attracted much attention from chemists, biologists, and medical scientists (Fan et al., 2019). The inflorescence of common castor plants is gradient monoecious raceme, with male flowers on the lower portion and

## OPEN ACCESS

### Edited by:

Wei Xu,  
Texas A&M University Corpus Christi,  
United States

### Reviewed by:

Katarzyna Agata Knop,  
University of Dundee, United Kingdom  
Milind B. Ratnaparkhe,  
ICAR Indian Institute of Soybean  
Research, India

### \*Correspondence:

Wei Zhou  
ycyz958@yeah.net

### Specialty section:

This article was submitted to  
Genomic Assay Technology,  
a section of the journal  
Frontiers in Genetics

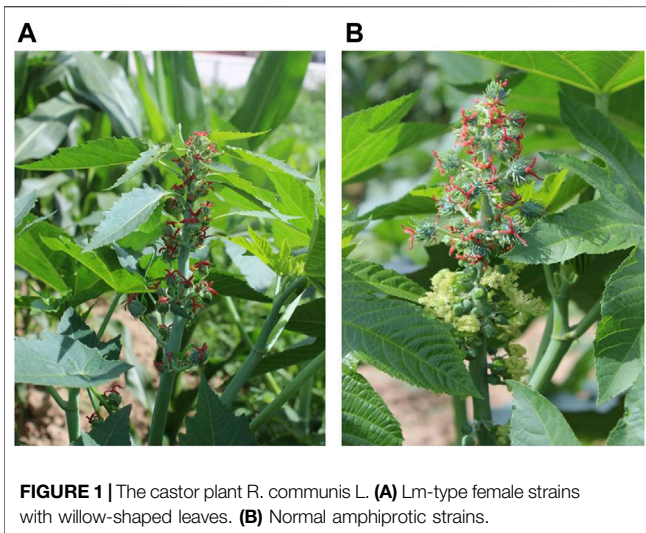
Received: 29 July 2021

Accepted: 30 September 2021

Published: 18 October 2021

### Citation:

Zhou Y, Zhu G, Wang Y, He Z and  
Zhou W (2021) A Comparative  
Transcriptional Landscape of Two  
Castor Cultivars Obtained by Single-  
Molecule Sequencing  
Comparative Analysis.  
Front. Genet. 12:749340.  
doi: 10.3389/fgene.2021.749340



**FIGURE 1** | The castor plant *R. communis* L. **(A)** Lm-type female strains with willow-shaped leaves. **(B)** Normal amphiprotic strains.

female flowers at the apex (Tan et al., 2015). Pistillate (bearing only female flowers) variations are bred to improve the seed yield. The Lm type castor is such a variety, obtained by exposing castor seeds to <sup>60</sup>Coγ.

With the development of sequencing technology, next-generation sequencing (NGS) has become an essential method for the study of genomes, epigenomes, and transcriptomes (Fan et al., 2019). The NGS method has been used in many model and non-model plant species, and large-scale genome sequences and

transcriptome data have been produced for deep analysis (Fan et al., 2019). However, the deficiencies of NGS, such as short reads, result in incompletely assembled transcripts that limit the better understanding of the transcriptomic data (Fan et al., 2019). The PacBio platform is based on the single-molecule real-time (SMRT) sequencing technology and provides longer and full-length transcripts without assembly, and can provide better information to understand the full-length transcriptome, such as alternative splicing, fusion transcripts, alternative polyadenylation, novel genes, and non-coding RNAs (Fan et al., 2019).

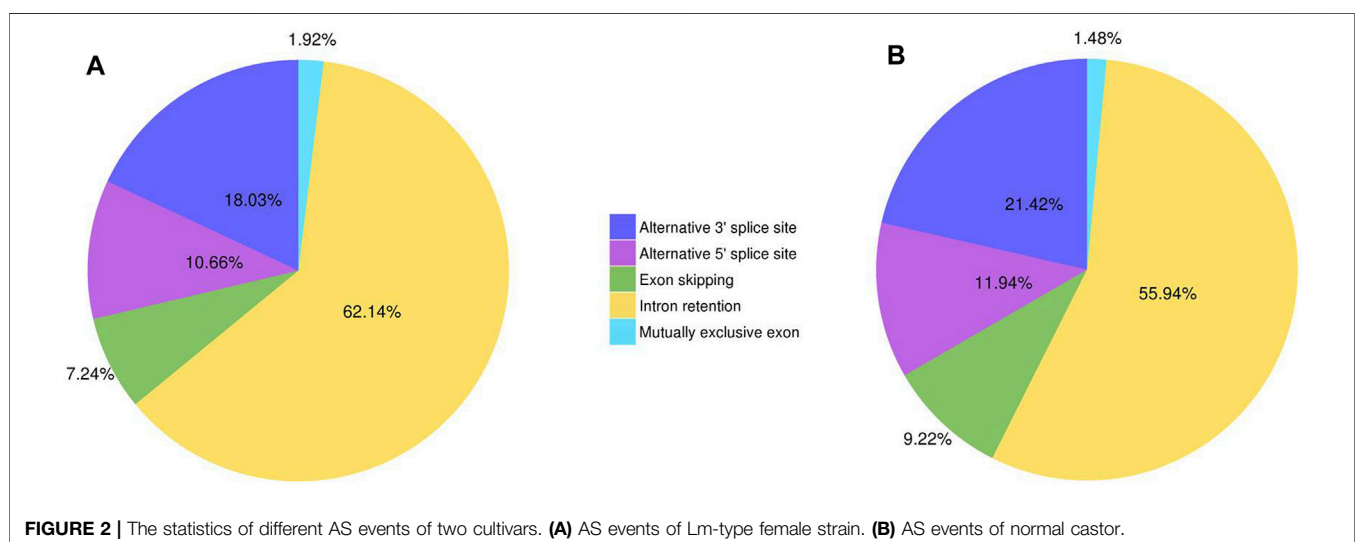
To gain an insight into how sex is differentially regulated at the molecular level, in the present study, full-length transcriptomes of Lm-type and normal castor cultivars were analyzed, and short-read RNA sequencing and single-molecule long-read sequencing were utilized to identify the differentially expressed genes and alternative splicing events between Lm-type female strains and normal amphiprotic strains. Furthermore, the study will provide valuable data for future studies of sex determination on castor plants.

## MATERIALS AND METHODS

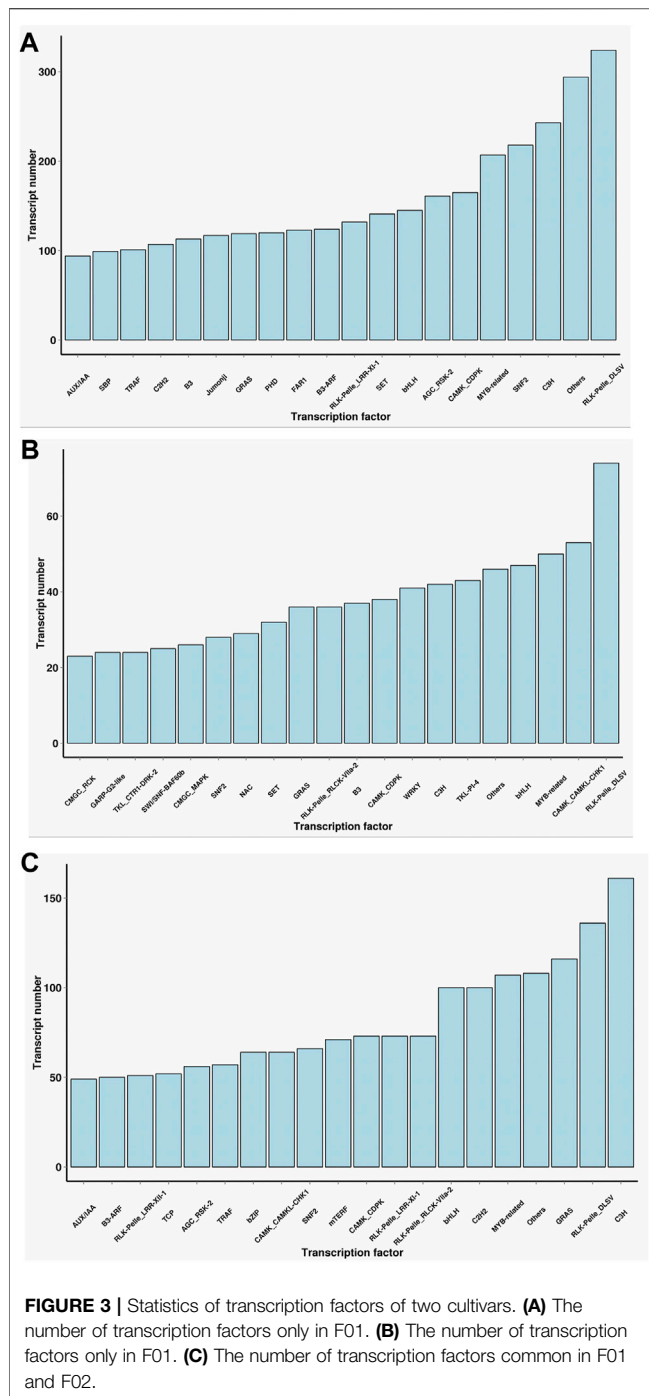
Two cultivars, Lm-type female plants with willow-shaped leaves and normal amphiprotic plants (Figure 1), were grown at the Experimental Base of the Agricultural College of INNER MONGOLIA MINZU UNIVERSITY, Tongliao City, Inner

**TABLE 1** | The statistic results of ICE clustering.

Samples	Number of consensus isoforms	Average consensus isoforms read length	Number of polished high-quality isoforms	Number of polished low-quality isoforms	Percent of polished high-quality isoforms (%)
F01	223,929	2,552	154,517	68,417	69.00
F02	138,066	2,056	105,536	32,086	76.44



**FIGURE 2** | The statistics of different AS events of two cultivars. **(A)** AS events of Lm-type female strain. **(B)** AS events of normal castor.



Mongolia Autonomous Region. The geographical position is between 42°15′–45°41′ north latitude and 119°15′–123°43′ east longitude. Ten plants of each cultivar were selected when the functional leaves grew to 1–6 cm (Lm-type) or 2–15 cm (normal type) on August 2nd, 2018. The Lm-type female plants and normal amphiprotic plants were designated as F01 and F02, respectively. For each cultivar around 10g of leaves and flowers were collected and frozen in liquid nitrogen and then stored at –80°C for subsequent RNA isolation. Using the Illumina

HiSeq X Ten platform, RNA was extracted from 5g of frozen leaves or flowers with two repeats, and an RNA-Seq library construction was performed following the instructions (Podnar et al., 2014).

## PacBio Library Construction and Sequencing

The total RNA was extracted from 5g mixtures of leaves and flowers (two repeats). Poly(T) oligo-attached magnetic beads (Dyna) were used to purify mRNA from about 3 μg total RNA. According to the protocols of the PacBio RS II platform, cDNA was synthesized using the SMART PCR cDNA Synthesis Kit (Clontech, CA, United States), and then fractionated with BluePippin® (Sage Science, Beverly, MA, United States). Then the final libraries were constructed using the Pacific Biosciences DNA Template Prep Kit (version 2.0). SMRT sequencing was performed with the Pacific Biosciences' real-time sequencer using C2 sequencing reagents.

## Preprocessing of SMRT Reads

The subreads were filtered using the standard protocol of the SMRT Analysis software suite (<http://www.pacificbiosciences.com>), and the reads of insert (ROIs) were obtained. After examining the poly(A) signals and 5' and 3' adaptors, full-length (FL) and non-full-length (nFL) reads were identified.

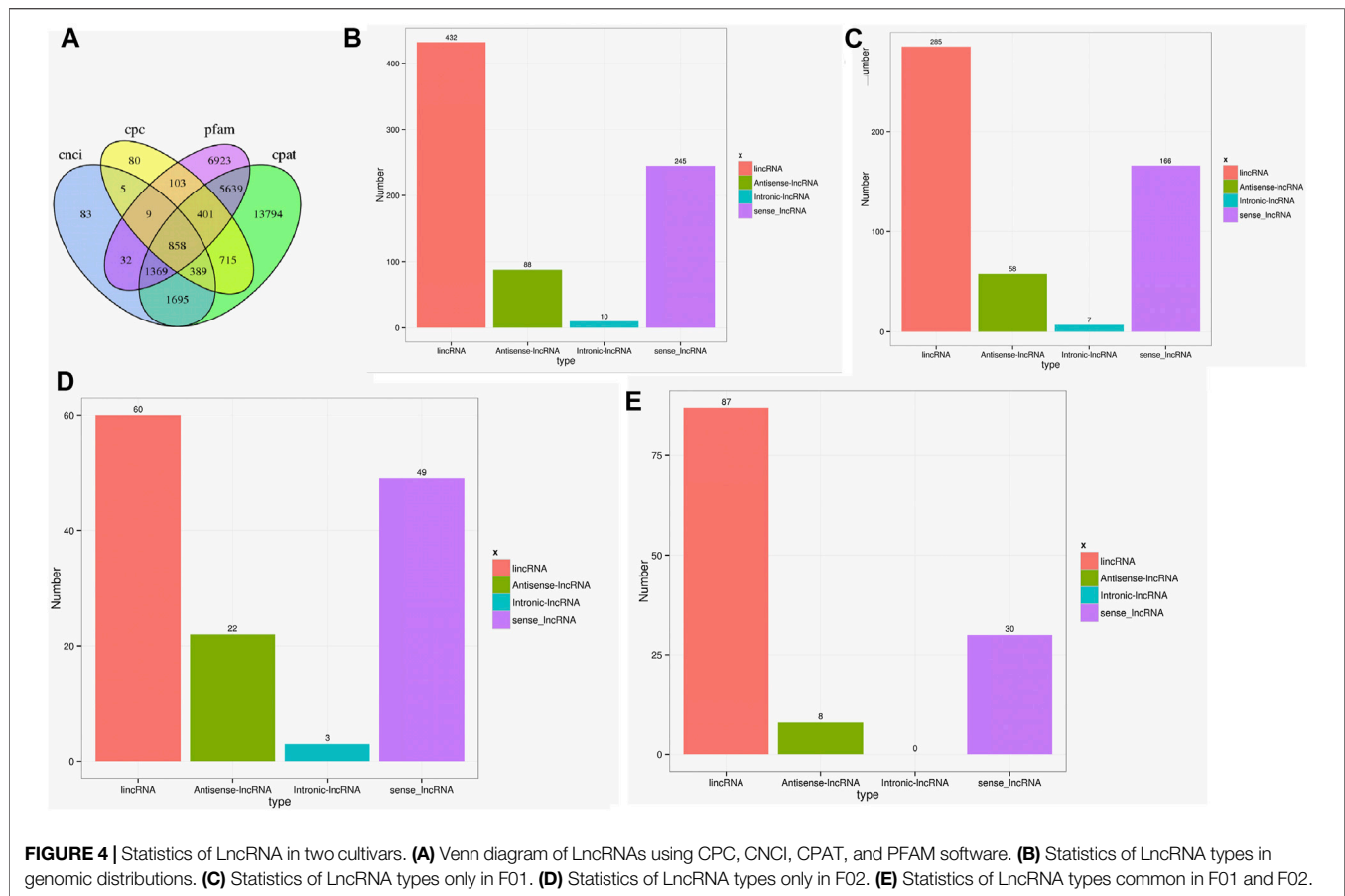
Consensus sequences were obtained from high-quality isoform sequences. The final transcriptome isoform sequences were filtered by removing the redundant sequences with the CD-HIT package ([http://weizhong-lab.ucsd.edu/cdhit\\_suite/cgi-bin/index.cgi?cmd=cd-hit](http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi?cmd=cd-hit)) to cluster and compare protein or nucleotide sequences.

## Alternative Splicing Analysis of Transcriptomes

To identify alternative splicing (AS) events, SpliceGrapher (Rogers et al., 2012) was used to analyze the transcriptome-wide AS events. AS events were predicted from non-redundant transcripts. The prediction criterion is as following: the sequence should be greater than 1,000 bp, the AS gap should be greater than 100 bp and at least 100 bp from the 3′/5′-end, and there should be a 5-bp overlap in the spliced transcript. Compared with the reference castor genome (<http://castorbean.jcvi.org/>), the full-length transcripts can be classified as derivations from the known genes and novel genetic loci.

Candidate coding regions were identified by TransDecoder (Broad Institute, Cambridge, MA, United States) from the final transcriptome isoform sequence. Sequences were searched using BLASTX (Buchfink et al., 2015) against the NCBI non-redundant protein and the UniProt with E-value cutoff at  $1 \times 10^{-6}$ . To further distinguish protein-coding and non-coding RNAs, the dbHT-Trans tool (v1.0) (Deng and Chen 2016) was used for all PacBio transcripts.

The gene ontology (GO) enrichments were analyzed using the GOseq (Young et al., 2010). The KEGG (<http://www>.



**TABLE 2 |** Statistics of annotations for genes of castor.

Anno database	Annotated number
COG Annotation	35,743
GO Annotation	61,390
KEGG Annotation	39,598
KOG Annotation	57,228
Pfam Annotation	63,089
Swissprot Annotation	62,336
eggNOG Annotation	83,826
Nr Annotation	85,286
All Annotated	85,322

genome.jp/kegg/) pathway analysis was implemented as reported (Kanehisa et al., 2017).

### Short-Read RNA Sequencing Analysis and Quantification of Gene Expression

The clean reads were screened from raw sequencing reads by removing low-quality reads and reads containing adaptors or ploy-Ns. Sequences of clean reads were aligned to the full-length sequences. Differential expression analysis was performed with EBSeq package (Leng et al., 2013), with FDR <0.05 and  $|\log_2(\text{fold-change})| \geq 1$ .

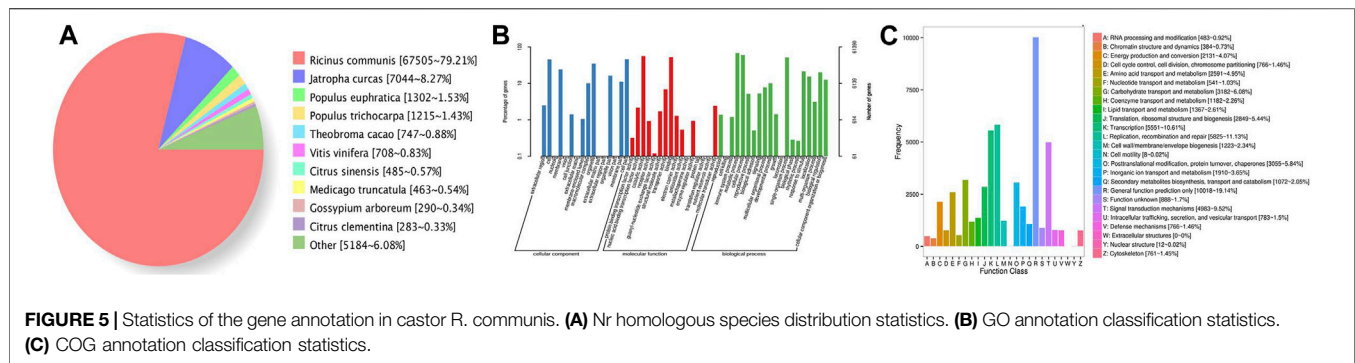
## RESULTS

### PacBio Iso-Seq Sequencing

The SMRT sequencing generated 456,994 polymerase reads in total, and 26.25 Gb and 16.38 Gb clean reads were obtained from Lm-type female and normal castor cultivars, respectively. Under the conditions of full passes of  $\geq 0$  and quality of  $>0.80$ , 647,205 and 328,497 ROIs were obtained from two cultivars, respectively (**Supplementary Table S1**). In addition, 448,217 and 258,645 full-length non-chimeric sequences were identified from Lm-type and normal castors, respectively (**Supplementary Table S2**).

The SMRT sequencing generated 456,994 polymerase reads in total, and 26.25 Gb and 16.38 Gb of clean reads were obtained from Lm-type female and normal castor cultivars, respectively. Under the conditions of full passes of  $\geq 0$  and quality of  $>0.80$ , 647,205 and 328,497 ROIs were obtained from two cultivars, respectively (**Supplementary Table S1**). In addition, 448,217 and 258,645 full-length non-chimeric sequences were identified from Lm-type and normal castors, respectively (**Supplementary Table S2**).

The lengths of full-length cDNA in the Lm-type female strain ranged from 281 to 11,430 bp with an average length of 2,702 bp. For the normal castor strain, the full-length cDNA showed an average length of 2,192 bp, and ranged from 303 to 9,681 bp. The



N50 values of that cDNA were 3,093 and 2,408 bp in Lm-type and normal castor cultivars, respectively. Then, from 223,929 (Lm-type female cultivar) and 138,066 (normal castor) full-length consensus cDNA, 76,068 out of 154,517 (49%) and 44,223 out of 105,536 (42%) high quality full-length consensus were obtained, respectively. The ICE clustering results are shown in Table 1.

### Alternative Splicing and Polyadenylation

A total of 51,613 and 20,152 AS events were found in Lm-type and normal castor cultivars, respectively, including exon skipping (ES), intron retention (IR), alternative 3' sites (Alt. 3'), alternative 5' sites (Alt. 5'), and mutually exclusive exons. The results showed that intron retention (IR) was the foremost AS event, with 62.14% and 55.94% in Lm-type and normal castor cultivars, respectively. The results of statistical analysis of different AS events in Lm-type female strain and normal castor were showed in Figure 2 and Supplementary Table S3.

### Comparative Analysis of lncRNA and Transcription Factors

Transcription factors which need to specifically bind to certain genes are essential for the regulation of gene expression. A total of 13,239 encoded transcription factors were identified from the full-length transcriptome in the two cultivars. Furthermore, we performed a comparative analysis of the common and unique transcription factors in the two cultivars. The main transcription factor types in Lm-type female castor include Rlk-Pelle-Dlsv, C3H, SNF2, and MYB-related families. In normal castor, the dominant transcription factors were Rlk-pelle-dlsv, camk-camkl-chk1, and MYB-related bHLH types. Although the two cultivars shared some types of transcription factors, the expression of corresponding genes was completely different (Figure 3).

As the key regulators in biological processes, long non-coding RNA (lncRNA) is a type of RNA that does not encode proteins (Jathar et al., 2017). A total of 858 lncRNAs were found in the two cultivars using CPC, CNCL, CPAT, and PFAM software (Gong et al., 2017). The genomic distributions of lncRNAs were classified into four types, namely lincRNA, antisense-lncRNA, intronic-lncRNA, and sense\_lncRNA. The ratios of different

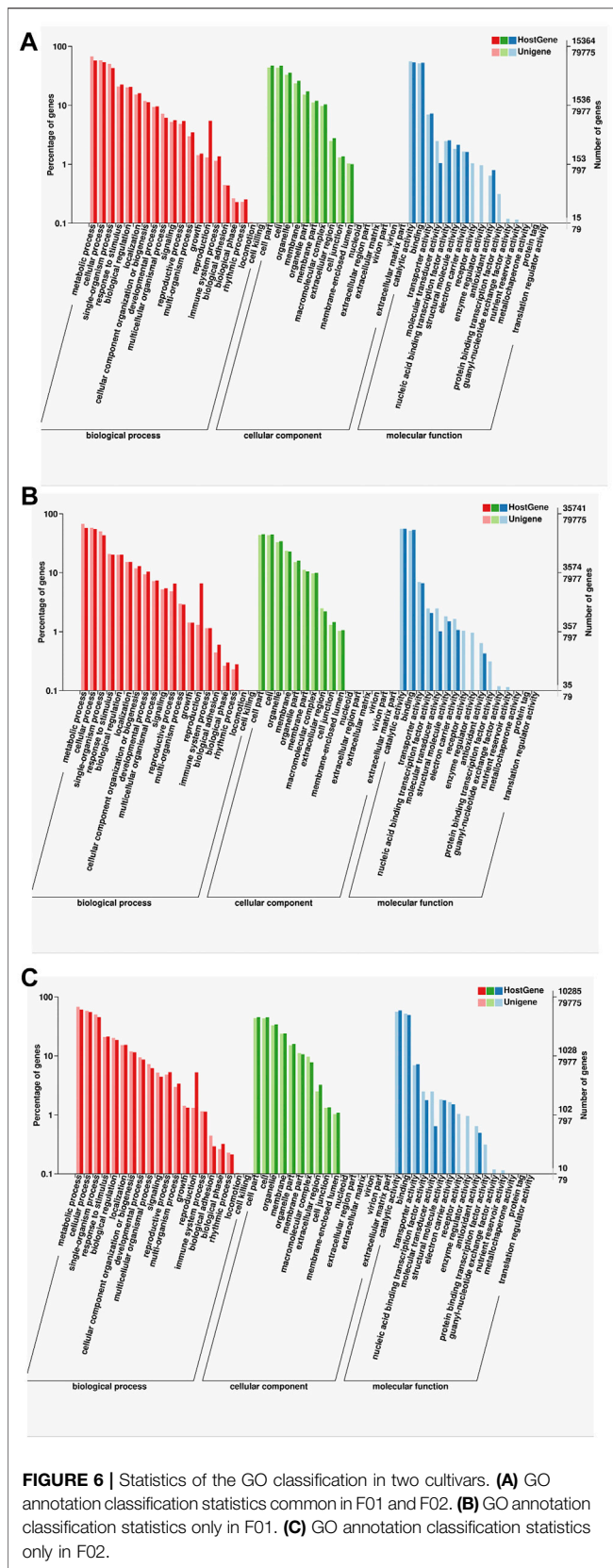
types varied greatly, with 285 lincRNA, 58 antisense-lncRNA, 7 intronic-lncRNA, and 166 sense\_lncRNA in the Lm-type, and 60, 22, 3, and 49 in the normal castor cultivar, respectively (Figure 4).

### Functional Annotation and Analysis of Isoform

For the functional annotation of gene isoforms, these genes were searched against the Genbank NR, Swissprot, GO, COG, KOG, Pfam, and KEGG databases, and a total of 85,322 genes were annotated by those seven databases. Among them, 85,286 genes (99.96%) were aligned to the NR, and 62,336 genes were matched to the SWISS-PROT (Table 2). Approximately 79.21% of genes were aligned to *R. communis*, followed by *Jatropha curcas* (8.27%) (Figure 5A).

The GO annotation system is a directed acyclic graph, including three categories: biological process (BP), molecular function (MF), and cellular component (CC). In this study, GO analysis was conducted using Blast2GO, and detailed GO distributions in GO categories are shown in Figure 5B. The vast majority of the genes were in cells or cell parts in the cellular component. In the molecular function class, most of the genes were classified as catalytic activity and binding. In the biological process class, genes classified as metabolic processes and cellular processes were the most common. "COG" refers to clusters of orthologous groups for eukaryotic complete genomes, and every protein in the database is assumed to be evolved from a common ancestor protein. In total, 35,743 out of 85,286 genes were classified into 25 different COG categories (Figure 5C), and the genes with general functions were the largest category, followed by replication, recombination and repair, and transcription.

The KEGG database is used to determine whether the genes are involved in specific metabolic or signal transduction pathways. In this study, a total of 125 KEGG pathways were identified. Several enriched pathways were involved in plant hormone signal transduction (ko04075), starch and sucrose metabolism (ko00500), and protein processing in the endoplasmic reticulum (ko04141) (Supplementary list S1).



### Functional Comparative Analysis of Isoform

For further annotation analysis of gene functionality, the functions of specific and common isoforms in the two samples were analyzed systematically, indicating that although many of the isoforms in the two samples were different, the corresponding gene functions were similar. The GO analysis showed that many of the isoforms enriched in the following items: metabolic process, cellular single-organism, cell part, catalytic activity, and binding (Figure 6).

The results of the COG functional annotation analysis on the specific and common isoforms in Lm-type and normal castors were similar to that of the GO analysis, and the function of the isoforms remained consistent. The COG analysis indicated the most common gene functions in L (replication, recombination and repair) and R (general function prediction) (Figure 7).

### Genes Associated With Sex Expression and Reproduction in Castor

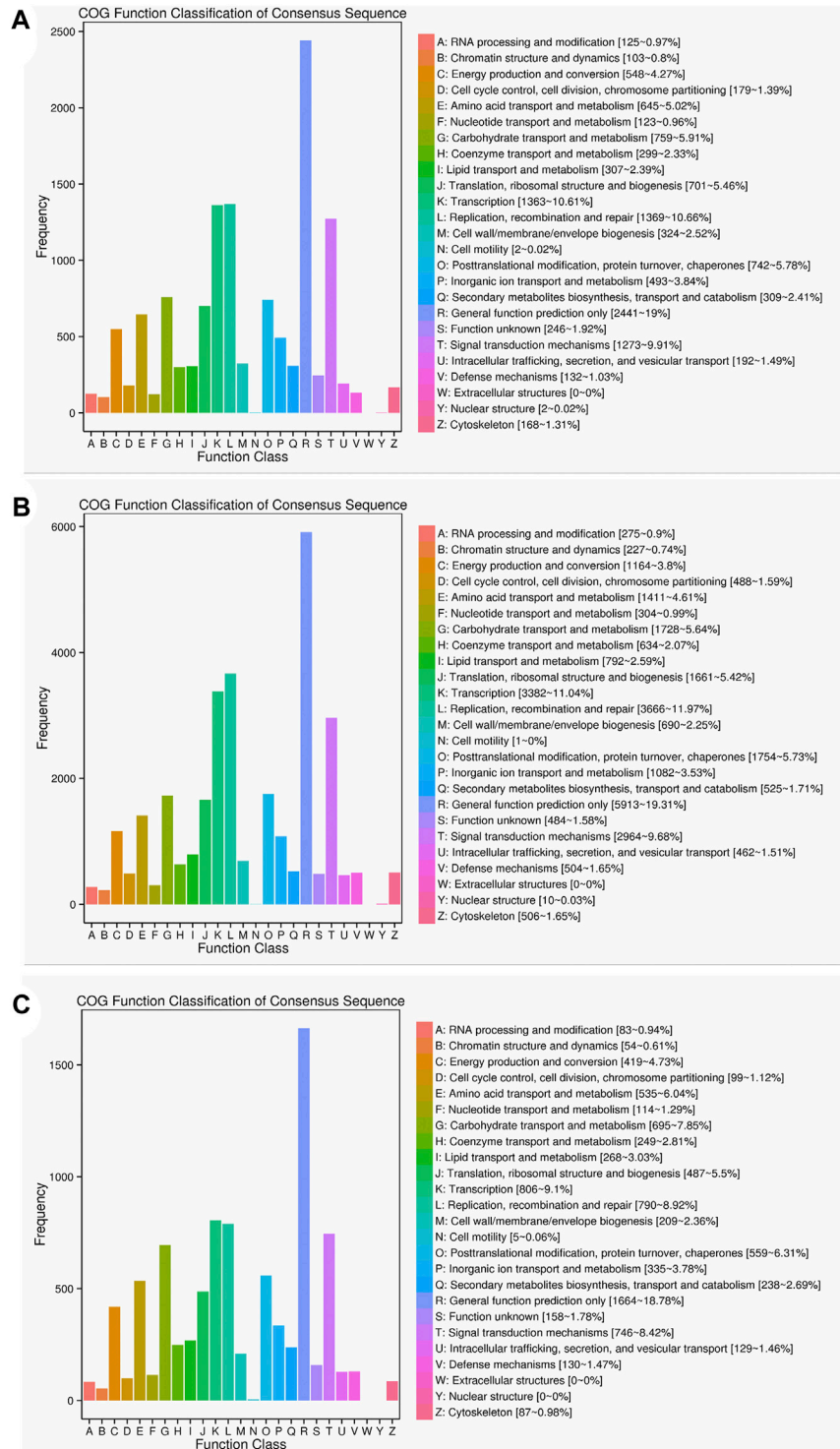
According to previous studies of sex determination between the monoecious and female *R. communis*, several subgroups genes were assumed to be putatively related to sex determination, such as auxin response factor, dynamin-2A, PCI domain containing protein, Xaa-Pro amino peptidase, ATP-binding protein, set domain protein, spermidine synthase, arginine/serine-rich splicing factor, eukaryotic translation initiation factor 2c, DNA (cytosine-5)-methyltransferase, s-adenosyl-methyltransferase, and acid phosphatase. In this study, many of these subgroups were identified in novel unigenes (Table 3). These genes were associated with hormone stimulus and participate in hormone-mediated signaling pathways, and also play a role in tissue and organ developmental processes.

### Comparative Analysis of Differential Gene Expression Profiling

The full-length sequences were used as a reference genome, and the sequences from the short-read RNA sequencing were used to conduct a differential gene analysis. A total of 2,461 genes were found to be differentially expressed, with 655 up-regulated and 1806 down-regulated. These differentially expressed genes (both up-regulated genes and down-regulated genes) were classified according to their KEGG pathway (Figure 8). The results showed that up-regulated genes were classified as pathways of ribosome, carbon metabolism, pentose phosphate pathway, and biosynthesis of amino acids. The down-regulated genes were involved in plant hormone signal transduction, phenylpropanoid biosynthesis, carbon metabolism, and plant-pathogen interaction.

### DISCUSSION

Alternative splicing was involved in phenotypic differences of Lm-type and normal castor cultivars.



**FIGURE 7 |** Statistics of the COG classification in two cultivars. **(A)** COG annotation classification statistics common in F01 and F02. **(B)** COG annotation classification statistics only in F01. **(C)** COG annotation classification statistics only in F02.

In recent years, comparative transcriptome analyses have successfully revealed specific genes responsible for C4 photosynthesis in many grasses, including maize and

switchgrass (Fan et al., 2019). Furthermore, recent studies of castor transcriptomes are mainly focused on gene expression from short-read RNA sequencing (Sood et al., 2014; Sturtevant

et al., 2019) which cannot identify alternative gene splice forms (Fan et al., 2019). The development of full-length sequencing technology provides a span-new approach to study full-length sequences, alternative splicing, gene structures, and APA of RNA (Grabherr et al., 2011; Rhoads and Au 2015). We thus conducted a comprehensive comparative analysis for two cultivars using this method. In this work, a total of 76,068 and 44,223 non-redundant transcripts were obtained from the high-quality transcripts of Lm-type female strain and normal castor cultivars, respectively. Among these genes, 51,613 and 20,152 AS events were found in Lm-type female strain and normal castor, respectively, of which intron retention (IR) was the foremost AS event, with 62.14% and

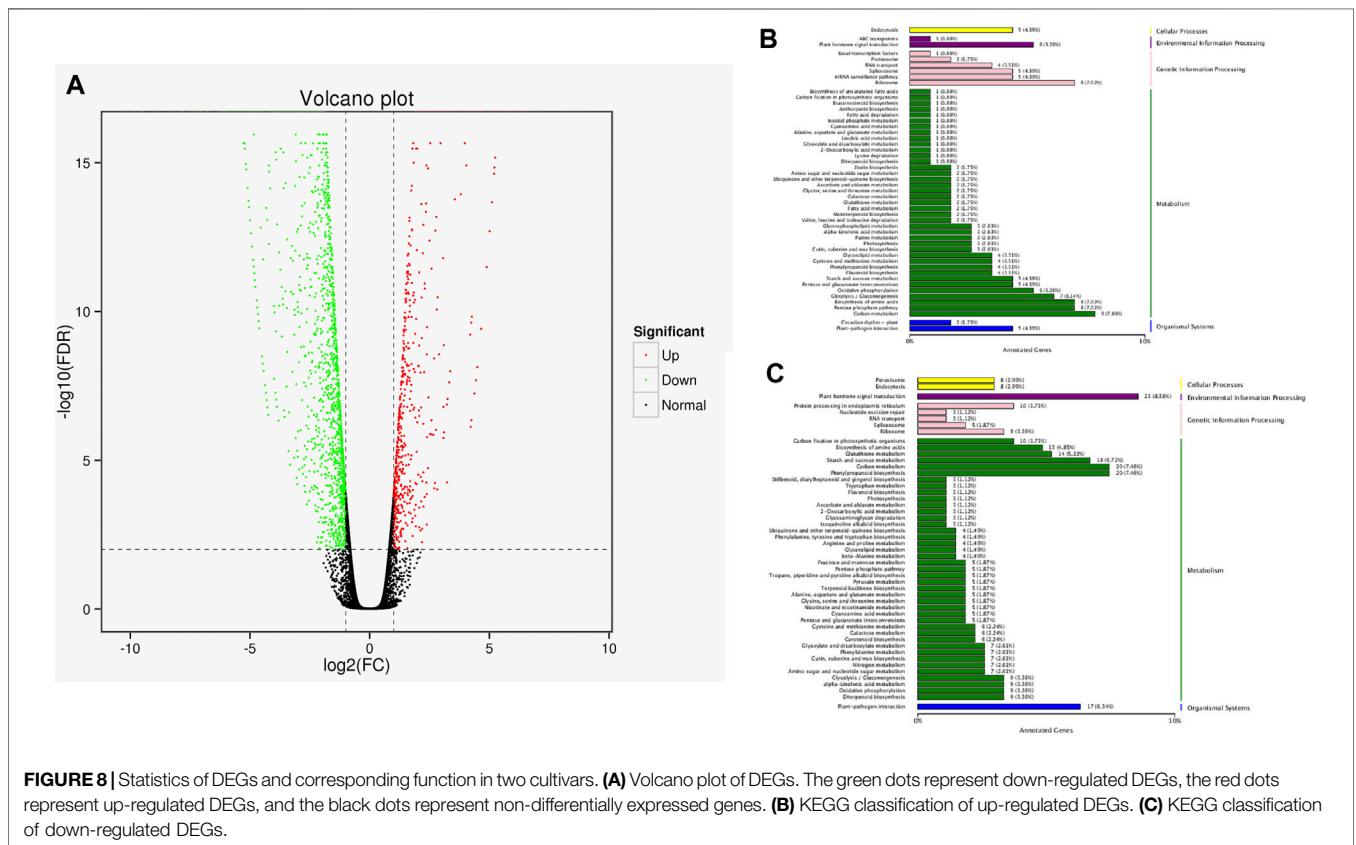
55.94% in the two cultivars, respectively. Its confirmed gene expression and splicing levels may have a significant impact on the morphological and other phenotypic differences between the two cultivars (Wang et al., 2018). The alternative splicing of eukaryotic transcripts is a mechanism that enables cells to generate vast protein diversity from a limited number of genes (Baralle and Giudice 2017; Bush et al., 2017). The mechanism and outcomes of the alternative splicing of individual transcripts are well understood (Fan et al., 2019). Some studies find that AS regulation is independent or partially independent of transcriptional regulation (Siam et al., 2019) and implements great function at the early stage of the heat response (Keller et al., 2017; Siam et al., 2019), useful for future heat sensing and signaling studies (Fan et al., 2019). Our new findings about AS provide important information for facilitating castor genome annotation, and the full characterization of the castor transcriptome.

**TABLE3 |** Putative genes involved in sex determination of *R. communis*.

Gene name	Numbers in novel unigenes
Dynamain-2A	17
Auxin response factor	232
ATP-binding protein	3
Spermidine synthase	20
Arginine/serine-rich splicing factor	72
Acid phosphatase	123
Eukaryotic translation initiation factor 2c	124
Set domain protein	30
DNA (cytosine-5)-methyltransferase	23
S-adenosyl-methyltransferase	4

### Transcription Factors and lncRNAs Played Important Role in Phenotypic Differences of Lm-Type and Normal Castor Cultivars

As the key regulators of transcription, TFs play an important role in the physiological regulation of plants (Olsen et al., 2005; Zhou and Memelink 2016). Our results (Figure 3) suggested Rlk-pelle-dsv, C3H, SNF2, and MYB-related transcription factors were the main types in the Lm-type cultivar.





Transcription factors of *rlk-pelle-dlsv*, *camk-camkl-chk1*, MYB-related BHLH, and other types were mainly expressed in the normal castor cultivar. We speculate that the difference in TFs has a significant effect on the difference in morphology. Similarly, as the important regulator, the number of lncRNA was very different: there were 285 lincRNA, 58 antisense-lincRNA, 7 intronic-lincRNA, and 166 sense\_lincRNA in the Lm-type cultivar, while 60, 22, 3, and 49 in the normal castor cultivar, respectively. Emerging work has revealed that many types of lncRNA regulate gene expression and have a great influence on genome stability in plants (Wang and Chekanova 2017; Sun et al., 2018). Studies on *Arabidopsis* show that lncRNA can serve as a molecular sponge and as a decoy, functioning in the regulation of transcription and silencing, particularly in RNA-directed DNA methylation, and in epigenetic regulation of flowering time (Zhao et al., 2018; Liu et al., 2019). Many plants reduce the expression of some lncRNAs to affect developmental phenotypes or molecular changes (Wang et al., 2014). We speculate that these regulators also played an important role in the growth and development of castor, and contribute significantly to phenotypic differences of Lm-type and normal cultivars.

## DEGs Implement a Significant Function in the Morphological Differences of the Two Cultivars

Using the full-length sequences as a reference genome, 2,461 differentially expressed genes were found, including 655 up-regulated genes and 1806 down-regulated genes, which was far more than in our previous RNA-seq transcriptome analysis (Fan et al., 2019). In this study, the functional analysis showed that proteins encoded by up-regulated genes (655) were classified to ribosome, carbon metabolism, pentose phosphate pathway, and biosynthesis of amino acids. Proteins encoded by down-regulated genes (1806) were attributed to plant hormone signal transduction, phenyl propanoid biosynthesis, carbon metabolism, and plant-pathogen interaction. We speculate that the differentially expressed genes were the main reason for the differences between the two castors, while the specific regulation mechanisms remain unclear.

## CONCLUSION

To the best of our knowledge, this study is the first large-scale comparative analysis of the transcriptome Lm-type and normal castor cultivars by single-molecule long-read sequencing.

## REFERENCES

- Baralle, F. E., and Giudice, J. (2017). Alternative Splicing as a Regulator of Development and Tissue Identity. *Nat. Rev. Mol. Cell Biol.* 18, 437–451. doi:10.1038/nrm.2017.27
- Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and Sensitive Protein Alignment Using DIAMOND. *Nat. Methods* 12, 59–60. doi:10.1038/nmeth.3176

Comparative analysis of the isoforms, transcription factors, lncRNAs, and AS in the two cultivars was performed systematically. The gene annotation analysis showed that although the isoforms were diverse in the two cultivars, the implemented functions were similar. Many species-specific differences are mainly attributed to small effects at multiple loci, probably. However, differences in the expression of genes and alternative splicing events have a profound effect on the evolution of major morphological diversification for different individuals in the developmental processes. The new findings of this study provided invaluable information for facilitating genome annotation and the full characterization of the transcriptome of these two cultivars.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/>, SRR8662424; <https://www.ncbi.nlm.nih.gov/>, SRR8662425.

## AUTHOR CONTRIBUTIONS

WZ performed the study, analyzed data, involved in the writing of the manuscript. YZ involved in data analyses, helped in writing the manuscript. GZ involved in sample collection and preparation, and helped in writing the manuscript. YW and ZH helped perform the analysis, and provided constructive discussions. All authors have read and approved the final manuscript.

## ACKNOWLEDGMENTS

The authors appreciate the great help from GZ of the Tongliao Academy of Agricultural Sciences and ZS of the Shenyang Agricultural University.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.749340/full#supplementary-material>

- Bush, S. J., Chen, L., Tovar-Corona, J. M., and Urrutia, A. O. (2017). Alternative Splicing and the Evolution of Phenotypic Novelty. *Phil. Trans. R. Soc. B*, 372, 20150474. doi:10.1098/rstb.2015.0474
- Deng, F., and Chen, S.-Y. (2016). dbHT-Trans: An Efficient Tool for Filtering the Protein-Encoding Transcripts Assembled by RNA-Seq According to Search for Homologous Proteins. *J. Comput. Biol.* 23, 1–9. doi:10.1089/cmb.2015.0137

- Fan, W., Lu, J., Pan, C., Tan, M., Lin, Q., Liu, W., et al. (2019). Sequencing of Chinese castor Lines Reveals Genetic Signatures of Selection and Yield-Associated Loci. *Nat. Commun.* 10, 3418. doi:10.1038/s41467-019-11228-3
- Gong, Y., Huang, H.-T., Liang, Y., Trimarchi, T., Aifantis, I., and Tsiganos, A. (2017). lncRNA-Screen: an Interactive Platform for Computationally Screening Long Non-coding RNAs in Large Genomics Datasets. *BMC genomics* 18, 434. doi:10.1186/s12864-017-3817-0
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length Transcriptome Assembly from RNA-Seq Data without a Reference Genome. *Nat. Biotechnol.* 29, 644–652. doi:10.1038/nbt.1883
- Jathar, S., Kumar, V., Srivastava, J., and Tripathi, V. (2017). *Technological Developments in lncRNA Biology. Adv. Exp. Med. Biol.* 1008, 283–323. doi:10.1007/978-981-10-5203-3\_10
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs. *Nucleic Acids Res.* 45, D353–d361. doi:10.1093/nar/gkw1092
- Keller, M., Hu, Y., Mesihovic, A., Fragkostefanakis, S., Schleiff, E., and Simm, S. (2017). Alternative Splicing in Tomato Pollen in Response to Heat Stress. *DNA Res.* 24, dsw051–217. doi:10.1093/dnares/dsw051
- Leng, N., Dawson, J. A., Thomson, J. A., Ruotti, V., Rissman, A. I., Smits, B. M. G., et al. (2013). *EBSeq: an Empirical Bayes Hierarchical Model for Inference in RNA-Seq Experiments.* Oxford, England: Bioinformatics, 1035–1043. doi:10.1093/bioinformatics/btt087
- Liu, F., Xu, Y., Chang, K., Li, S., Liu, Z., Qi, S., et al. (2019). The Long Noncoding RNA T5120 Regulates Nitrate Response and Assimilation in Arabidopsis. *New Phytol.* 224, 117–131. doi:10.1111/nph.16038
- Olsen, A. N., Ernst, H. A., Leggio, L. L., and Skriver, K. (2005). NAC Transcription Factors: Structurally Distinct, Functionally Diverse. *Trends Plant Sci.* 10, 79–87. doi:10.1016/j.tplants.2004.12.010
- Podnar, J., Deiderick, H., Huerta, G., and Hunicke-Smith, S. (2014). Next-Generation Sequencing RNA-Seq Library Construction. *Curr. Protoc. Mol. Biol.* 106, 4–19. doi:10.1002/0471142727.mb0421s106
- Rhoads, A., and Au, K. F. (2015). PacBio Sequencing and its Applications. *Genomics, Proteomics & Bioinformatics* 13, 278–289. doi:10.1016/j.gpb.2015.08.002
- Rogers, M. F., Thomas, J., Reddy, A. S., and Ben-Hur, A. (2012). SpliceGrapher: Detecting Patterns of Alternative Splicing from RNA-Seq Data in the Context of Gene Models and EST Data. *Genome Biol.* 13, R4. doi:10.1186/gb-2012-13-1-r4
- Siam, A., Baker, M., Amit, L., Regev, G., Rabner, A., Najjar, R. A., et al. (2019). Regulation of Alternative Splicing by P300-Mediated Acetylation of Splicing Factors. *Rna*, 25, 813–824. doi:10.1261/rna.069856.118
- Sood, A., Jaiswal, V., Chanumolu, S. K., Malhotra, N., Pal, T., and Chauhan, R. S. (2014). Mining Whole Genomes and Transcriptomes of *Jatropha* (*Jatropha Curcas*) and Castor Bean (*Ricinus communis*) for NBS-LRR Genes and Defense Response Associated Transcription Factors. *Mol. Biol. Rep.* 41, 7683–7695. doi:10.1007/s11033-014-3661-0
- Sturtevant, D., Romsdahl, T. B., Yu, X.-H., Burks, D. J., Azad, R. K., Shanklin, J., et al. (2019). Tissue-specific Differences in Metabolites and Transcripts Contribute to the Heterogeneity of Ricinoleic Acid Accumulation in *Ricinus communis* L. (castor) Seeds. *Metabolomics* 15, 6. doi:10.1007/s11306-018-1464-3
- Sun, X., Zheng, H., and Sui, N. (2018). Regulation Mechanism of Long Non-coding RNA in Plant Response to Stress. *Biochem. biophysical Res. Commun.* 503, 402–407. doi:10.1016/j.bbrc.2018.07.072
- Tan, M., Xue, J., Wang, L., Huang, J., Fu, C., and Yan, X. (2015). Transcriptomic Analysis for Different Sex Types of *Ricinus communis* L. During Development from Apical Buds to Inflorescences by Digital Gene Expression Profiling. *Front. Plant Sci.* 6, 1208. doi:10.3389/fpls.2015.01208
- Wang, B., Regulski, M., Tseng, E., Olson, A., Goodwin, S., McCombie, W. R., et al. (2018). A Comparative Transcriptional Landscape of maize and Sorghum Obtained by Single-Molecule Sequencing. *Genome Res.* 28, 921–932. doi:10.1101/gr.227462.117
- Wang, H.-L. V., and Chekanova, J. A. (2017). Long Noncoding RNAs in Plants. *Adv. in Exp. Med. Biol.* 1008, 133–154. doi:10.1007/978-981-10-5203-3\_5
- Wang, Y., Wang, X., Deng, W., Fan, X., Liu, T.-T., He, G., et al. (2014). Genomic Features and Regulatory Roles of Intermediate-Sized Non-coding RNAs in Arabidopsis. *Mol. Plant* 7, 514–527. doi:10.1093/mp/sst177
- Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene Ontology Analysis for RNA-Seq: Accounting for Selection Bias. *Genome Biol.* 11, R14. doi:10.1186/gb-2010-11-2-r14
- Zhao, X., Li, J., Lian, B., Gu, H., Li, Y., and Qi, Y. (2018). Global Identification of Arabidopsis lncRNAs Reveals the Regulation of MAF4 by a Natural Antisense RNA. *Nat. Commun.* 9, 5056. doi:10.1038/s41467-018-07500-7
- Zhou, M., and Memelink, J. (2016). Jasmonate-responsive Transcription Factors Regulating Plant Secondary Metabolism. *Biotechnol. Adv.* 34, 441–449. doi:10.1016/j.biotechadv.2016.02.004

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

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

Copyright © 2021 Zhou, Zhu, Wang, He and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.