Check for updates

Genome-wide identification, characterization, and functional analysis of lncRNAs in *Hevea brasiliensis*

Lingling Wang[1]*, Jingyi Wang[1], Hui Chen[1] and Bin Hu[2]*

[1]Ministry of Education Key Laboratory for Ecology of Tropical Islands, Key Laboratory of Tropical Animal and Plant Ecology of Hainan Province, College of Life Sciences, Hainan Normal University, Haikou, China, [2]Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China

Natural rubber (NR) is an essential industrial raw material widely used in our life. *Hevea brasiliensis* (Reyan7-33-97) is an economic plant producing natural rubber. Long non-coding RNAs (lncRNAs) are emerging as crucial regulators in numerous biological processes while the characterization and analysis of lncRNAs in *Hevea brasiliensis* are still largely unrevealed. We integrated the transcriptome datasets from multiple tissues to identify rubber lncRNAs. As a result, 12,029 lncRNAs were found and characterized with notably distinctive features such as longer exon, lower expression levels and GC content, and more tissue specificity in comparison with mRNAs. We discovered thousands of tissue-specific lncRNAs in rubber root, latex, bark, leaf, flower, and seed tissues. The functional enrichment result reveals that tissue-specific lncRNAs are potentially referred to particular functions of tissues, while the non-tissue specific is related to the translation and metabolic processes. In the present study, a comprehensive lncRNA dataset was identified and its functional profile in *Hevea brasiliensis* was explored, which provides an annotation resource and important clues to understand the biological functions of lncRNAs in *Hevea brasiliensis*.

KEYWORDS

long non-coding RNA, tissue specificity, rubber, RNA sequencing, biological function

Introduction

The eukaryotic genomes are pervasively transcribed (>90%). Except for the protein-coding RNAs (approximately 2%), a myriad of transcripts referred to as non-coding RNAs (ncRNAs) (Chekanova et al., 2007; Kapranov et al., 2007), including long non-coding RNAs (lncRNAs) that are >200 nt in length, possess little or no discernable protein-coding ability, have been reported to be involved in regulating various biological

processes (Rinn and Chang, 2012; Jin et al., 2013; Zhang et al., 2014). Based on their genomic localization relationship with the protein-coding genes, lncRNAs can be classified as intronic lncRNAs, antisense lncRNAs, intergenic lncRNAs (lincRNAs), etc. (Derrien et al., 2012). Differing from protein-coding genes, most lncRNAs share poor sequence conservation between species, present relatively low but tissue-specific expression patterns, raising their special functions in regulating tissue development, cellular and biological processes.

Animal lncRNAs have been proved to be implicated in modifying chromatin, regulating transcription, posttranscription, and other cellular processes (Geisler and Coller, 2013; Cech and Steitz, 2014). In the past decades, the genome-wide identification of plant lncRNAs is later and less comprehensive when compared with mammalians. While, with the rapid advancement of sequencing technology and bioinformatics analysis, thousands of lncRNAs have been identified and the functions of some lncRNAs have been demonstrated in several plant species. In *Arabidopsis*, more than 6000 lncRNAs have been identified, classified, and annotated under normal or stress conditions (Zhao et al., 2018). Among them, a lncRNA named as *MAS* is induced by cold, activates the transcription of *MAF4* and suppresses the precocious flowering (Zhao et al., 2018). In addition, lncRNAs have been reported to function in light response (Wang et al., 2014), stress responses (Di et al., 2014; Qin et al., 2017), seedling greening modulation (Wang et al., 2018b), and transcriptional modulation processes (Rigo et al., 2020). In rice, Wang et al. found that overexpression of *LAIR* (*LRK Antisense Intergenic RNA*) confers transgenic rice increased grain yield by upregulating the expression of neighboring *LRK* genes (Wang et al., 2018c). Additionally, 3488 high-confidence maize lncRNAs were identified under drought stress, and 1535 of them were drought responsive (Pang et al., 2019). Besides, several lncRNAs were identified to be mediated in pathogen/fungal stress responses/defenses in Grapevine, wheat, rubber tree, and other plants (Zhang et al., 2013; Xing et al., 2019; Yin et al., 2019; Zhang et al., 2020; Liu et al., 2022). Although some lncRNAs have been reported in the rubber tree, the previous study is solely conducted in a single tissue, leading to a limited lncRNAs in discovery (Yin et al., 2019; Li et al., 2021; Liu et al., 2022). Thus, the identification, characterization, and functional profile of rubber tree lncRNAs were still in its infancy.

The Para rubber tree, *Hevea brasiliensis*, is a kind of economically significant tropical tree that cultivated in South America, Malaysia, Indonesia, Thailand, and China. In China, *Hevea* Reyan7-33-97 was one of the elite cultivars (Tang et al., 2016). And the natural rubber (*cis*-1,4-polyisoprene) produced by *H. brasiliensis* accounts for >98% of the natural rubber production, which is a high-quality and indispensable source for numerous rubber products worldwide (van Beilen and Poirier, 2007). In the present study, we identified 12,029 lncRNAs from multiple rubber tissues. Following, we found that the rubber lncRNAs exhibited notably distinctive features when compared with mRNAs, such as longer exon, lower

expression, lower GC content, and higher tissue-specificity. Furthermore, thousands of tissue-specific lncRNAs in six tissues were identified depending on the tissue-specific score, revealing their potentially particular functions in tissue development and tissue-associated biological processes. The above results provide important clues to understand the functions of rubber lncRNAs and will lay a theoretical underpinning for further functional studies of lncRNA in *Hevea brasiliensis* and other plants.

Materials and methods

Transcriptome assembly

We obtained the RNA-seq data of *Hevea brasiliensis* (Reyan7-33-97) from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra) (Supplementary Table 1), and the reference genome of Reyan7-33-97 was downloaded from NCBI (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/654/055/) (Tang et al., 2016; Gong et al., 2018; Wang et al., 2021). We used the fastqc (v0.11.9, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to control the quality of RNA-seq data (Andrews, 2010) and AdapterRemoval (v2.3.1, https://github.com/MikkelSchubert/adapterremoval) to remove the adapter sequence residues from the reads (Lindgreen, 2012). After quality control, we mapped the RNA-seq data to *Hevea brasiliensis* (Reyan7-33-97) genome using hisat2 (v2.2.1, http://daehwankimlab.github.io/hisat2/) (Kim et al., 2019) and employed stringtie2 (v2.1.4, https://github.com/skovaka/stringtie2) to assemble the transcriptome of each sample. To remove the redundant transcripts, we merged the transcriptome from all samples and filtered the transcripts without strand information.

The read counts for each transcript were calculated with software salmon (v1.4.0, https://salmon.readthedocs.io/en/latest/index.html) (Patro et al., 2017). After that the Trimmed Mean of the M-values (TMM) approach was used to normalize the raw counts, and the expression values quantified as FPKM (Fragments Per Kilobase of sequence per Million mapped reads) were calculated using the edgeR (v3.30.3, https://bioconductor.org/packages/release/bioc/html/edgeR.html) (Robinson et al., 2010).

Identification of long non-coding RNAs

To provide a landscape of lncRNAs in *Hevea brasiliensis* (Reyan7-33-97), we integrated a stringent pipeline to identify lncRNAs (Figure 1). Firstly, the multi-exon transcripts >= 200 nt in length were reserved, and the transcripts, of which the exon overlapped with protein-coding genes were removed (Li et al.,
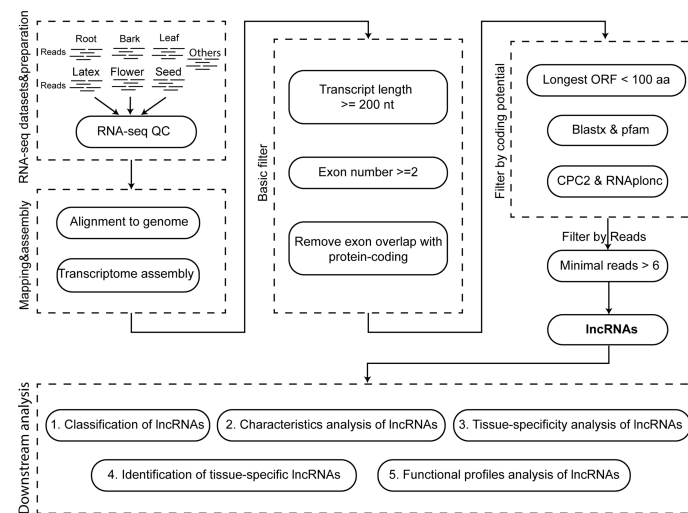
FIGURE 1

The flow chart for the identification and analysis of the long non-coding RNA. The first box indicates the RNA datasets and their quality control. The second box represents the read mapping and assemble. The third and fourth box indicate the processes for the identification and screen of the lncRNA. The fifth box is the downstream analysis of lncRNAs.

2014; Zhang et al., 2014; Li et al., 2021; Liu et al., 2022). Secondly, the transcripts were filtered based on their coding potential which were evaluated by the following indicators: longest ORF length, sequence similarity or domain with the known proteins, and coding score (Yin et al., 2019; Huang et al., 2021). The TransDecoder software (https://github.com/TransDecoder/TransDecoder) was chosen to calculate the length of the longest ORF, and only the transcripts containing the longest ORF< 100 aa in length were retained. Then, we used blastx to search against non-redundant proteins (nr) of NCBI with the parameter e-value=1e-3 and employed PfamScan to search against Pfam domain with -e_seq=1e-3 & -e_dom=1e-3 (Camacho et al., 2009; Mistry et al., 2021) to filter out the transcripts share significant homology with the known protein sequences or Pfam domains. After that, to further distinguish the non-coding from coding transcripts, we used Coding Potential Calculator 2 (CPC2) (http://cpc2.cbi.pku.edu.cn/) and RNAplonc software (v1.1, https://github.com/TatianneNegri/RNAplonc) to compute their coding scores, and eliminated the coding transcripts with default criterion (Kang et al., 2017; Negri et al., 2019). Finally, to guarantee the quality of lncRNAs, the transcripts supported by at least six reads were defined as the lncRNAs in *Hevea brasiliensis* (Reyan7-33-97).

Classification of the lncRNAs

The identified lncRNAs were classified into different groups, including intergenic lncRNAs, antisense lncRNAs, intronic lncRNAs, and other lncRNAs based on their relationships with the coding genes using FEELnc (v.0.1.1, https://github.com/tderrien/FEELnc) (Wucher et al., 2017). The intergenic lncRNAs are the lncRNAs locating in the intergenic range of coding genes, while the intronic lncRNAs are produced from intron. If a lncRNA is in the antisense strand of related coding genes, it will be defined as the antisense lncRNA.

Characteristic investigation for long non-coding RNAs

We investigated as well as compared the transcript length, exon number, exon length, intron length, GC content, and expression values of lncRNAs and mRNAs. Here, the R (v4.0.2, https://cran.r-project.org) and R package GenomicFeatures (v1.40.1, https://bioconductor.org/packages/release/bioc/html/GenomicFeatures.html) was utilized to compute the transcript length, exon number, exon length, and intron length of the lncRNAs and mRNAs (Lawrence et al., 2013). The GC content was calculated by using the LncFinder (v1.1.5, a R package) (Han et al., 2019).

Genome distribution of rubber lncRNAs

To examine the genome distribution of the rubber lncRNAs, we split the genome into different bins (size= 50000 nt) from the first position to the end and calculated the number of lncRNAs

in each bin. The genome distribution of mRNAs was calculated with the same approach followed by comparing the genome distribution between lncRNAs and mRNAs.

Tissue-specificity analysis

The expression values of six rubber tissues were used to compute the tissue specificity. The following formula was adopted to calculate the fraction for each lncRNA across these tissues:

$$F_{i,j} = \frac{E_{i,j}}{\sum_{j=1}^{n} E_{i,j}} \qquad n = 6 \ (n, \ Tissue \ number)$$

Here, i indicates the serial number of lncRNA, while j represents the serial number of the tissue. F is the fraction value; E represents the expression value. $E_{i,j}$ is the expression value of lncRNA i in the tissue j. $F_{i,j}$ is the fraction value of lncRNA i in the tissue j. For a given lncRNA i, the max fraction value among different tissues is defined as the tissue-specific score (Zhou et al., 2020). Then, the tissue-specific score of mRNAs was computed with the same formula. The tissue-specific score > 0.6 was set as the cutoff to screen the tissue-specific lncRNAs among all tested tissues.

Biological function analysis for lncRNAs

Functional analysis of the identified lncRNAs was conducted depending on the functional annotation of their co-expression mRNAs (Bhatia et al., 2020). To find the co-expression mRNAs for lncRNAs, we performed the Pearson correlation analysis of them using R package, stats (v4.0.2, http://www.R-project.org/) with the FPKM value, and the screen criteria was set as: Pearson correlation > 0.8 & adjusted P-value< 0.05. Then, we performed the Gene Ontology (biological process category) and KEGG (Kyoto Encyclopedia of Genes and Genomes) annotation for the mRNAs that co-expressed with lncRNAs using the software interproscan (v5.48-83.0, https://github.com/ebi-pf-team/interproscan) (Jones et al., 2014) and kobas (v3.0, http://kobas.cbi.pku.edu.cn/) (Bu et al., 2021), respectively. For each lncRNA, we conducted the BP (biological process) and KEGG enrichment analysis of its correlated mRNAs to predict its function. In the enrichment analysis, the adjusted P-value< 0.1 was set as the cutoff to determine the significance of enrichment, and the P-value was calculated by Hypergeometric test and adjusted by Benjamini & Hochberg.

Results

Thousands of long non-coding RNAs were identified in *Hevea brasiliensis*

To build a comprehensive long non-coding RNA atlas for the Para rubber tree (*Hevea brasiliensis*, Reyan7-33-97), We collected the RNA sequencing (RNA-seq) data of root, bark, leaf, latex, flower, seed, etc. (Supplementary Table 1). And performed the transcriptome assembly, as well as a comprehensive analysis to identify, characterize and analyze the lncRNAs (Figure 1). After mapping the RNA-seq reads to the reference genome of *Hevea brasiliensis* and data assembling, we finally obtained 172,943 rubber transcripts.

Then, we employed a stringent criterion to identify the long non-coding RNAs (detail in Materials and methods). Only the long multi-exon transcripts without the evidence of protein-coding capacity were remained, as a result, 12,029 lncRNAs of *Hevea brasiliensis* were finally identified. The lncRNAs were divided into different categories based on their genomic position relationship with coding genes, and we found that most of them were intergenic (8442, about 70%) lncRNAs, only 19% antisense lncRNAs, about 5% intronic lncRNAs and 6% of them were other lncRNAs (Supplementary Table 2). Moreover, we noted that although the genomic distribution of lncRNAs and mRNAs presented weak correlations in the whole rubber genome, in some of the genomic regions such as 0.4-0.8 Mb of NW_018745697.1, the number of lncRNAs exhibited high correlations with mRNAs, indicating that, to some extent, the lncRNAs presented close transcriptional relationships with mRNAs (Figure 2 and Supplementary Figure 1).

The characteristics of lncRNAs are different from mRNAs in *Hevea brasiliensis*

To characterize the lncRNAs of *Hevea brasiliensis*, several features of lncRNAs and mRNAs, including exon number, transcript length, exon size, intron size, GC content, and expression values were inspected and compared. We found that the length of lncRNAs was generally shorter than that of mRNAs (Figure 3A), as about 80% lncRNA was 378-2668 nt in length while 80% mRNA was 805-3390 nt, which kept consistent with the features of soybean and *Camellia sinensis* lncRNAs (Lin et al., 2020; Wan et al., 2020). In addition, similarly to the lncRNAs of *Solanum lycopersicum* (Wang et al., 2018a), rubber
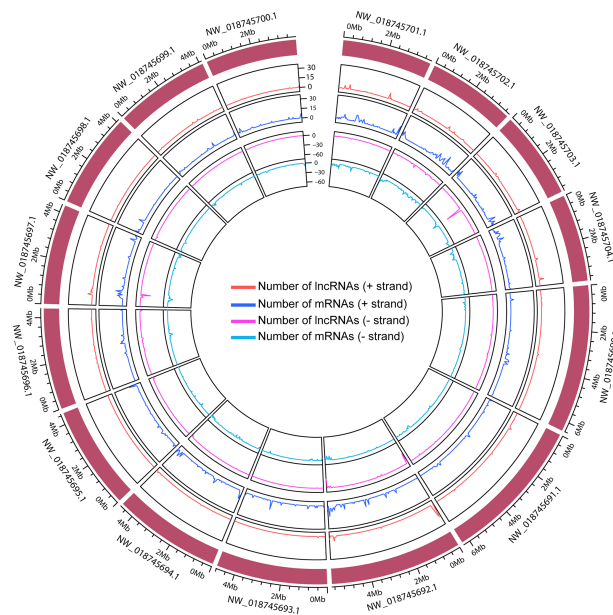
FIGURE 2
The distribution of lncRNAs and mRNAs across the top 15 length scaffolds. The out layer indicates the genomic scaffold. The orange, blue, pink, and turquoise colors represent the lncRNAs in + strand, mRNAs in + strand, lncRNA in - strand, and mRNA in - strand, respectively.

lncRNAs exhibited longer exon and intron length than mRNAs (Figures 3B, C). Meanwhile, we found that only about 2% lncRNAs owned >= 10 exons in number, while more than 20% mRNA possessed more than 10 exons (Figure 3D), suggesting that lncRNAs possess relatively fewer exon numbers than mRNA in the *Hevea brasiliensis* genome.

The stop codons (TAG, TAA, and TGA) are preferential to harbor the A+T base and present widely distribution in the lncRNA. To examine whether the lncRNAs of *Hevea brasiliensis* have a lower G+C contents, we calculated the G+C contents of both lncRNAs and mRNAs. We noted that the rubber lncRNAs exhibit significantly lower G+C content than mRNAs (Figure 3E). Interestingly, we also found that the rubber lncRNAs exhibited lower expression levels than mRNA (Figure 3F), suggesting that the G+C content may have certain relationships with the coding potential of genes and indicating the potential of DNA transcription.

Hevea brasiliensis lncRNAs exhibit higher tissue-specificity than mRNAs

To examine whether the rubber lncRNAs exhibit tissue specificity, we utilized the expression values of six tissues to evaluate their tissue-specific score, which was defined as the max fraction of the lncRNAs' tissue expression value to the cumulative expressions of all tested tissues. As a result, we found the lncRNAs displayed notably higher tissue-specific scores than mRNAs

(Figure 4A). In addition, the lncRNAs exhibited a higher percentage of tissue specificity than that of mRNAs in the *Hevea brasiliensis* (Figure 4B). Thus, the *Hevea brasiliensis* lncRNAs are preferred to exhibit tissue-specific expressions, suggesting that the tissue-specific lncRNAs are likely to play particular roles in the corresponding tissues. Besides, among the six tissues used in this research, we found more tissue-specific lncRNAs were identified in seed (more than 1,000), suggesting their important functions in rubber seed (Figures 4C, D). Meanwhile, except for bark (less than 500 lncRNAs were found), about 500-1,000 lncRNAs were respectively identified in root, leaf, flower, and latex of *Hevea brasiliensis* (Figures 4C, D).

Tissue-specific lncRNAs presented several potential functions in corresponding tissues

Thousands of tissue-specific lncRNAs have been identified in *Hevea brasiliensis*, while the potential functions of them were still obscure. Gene expression is temporal and spatial, the tissue-specific expression patterns of lncRNAs also provide important clues for exploring their specific biological functions. The functional annotation of co-expressed mRNAs were used to predict the biological functions and build a functional profile for the tissue-specific lncRNAs. To obtain the co-expression relationships between these lncRNAs and mRNAs, the co-expression analysis was performed and the co-expressed
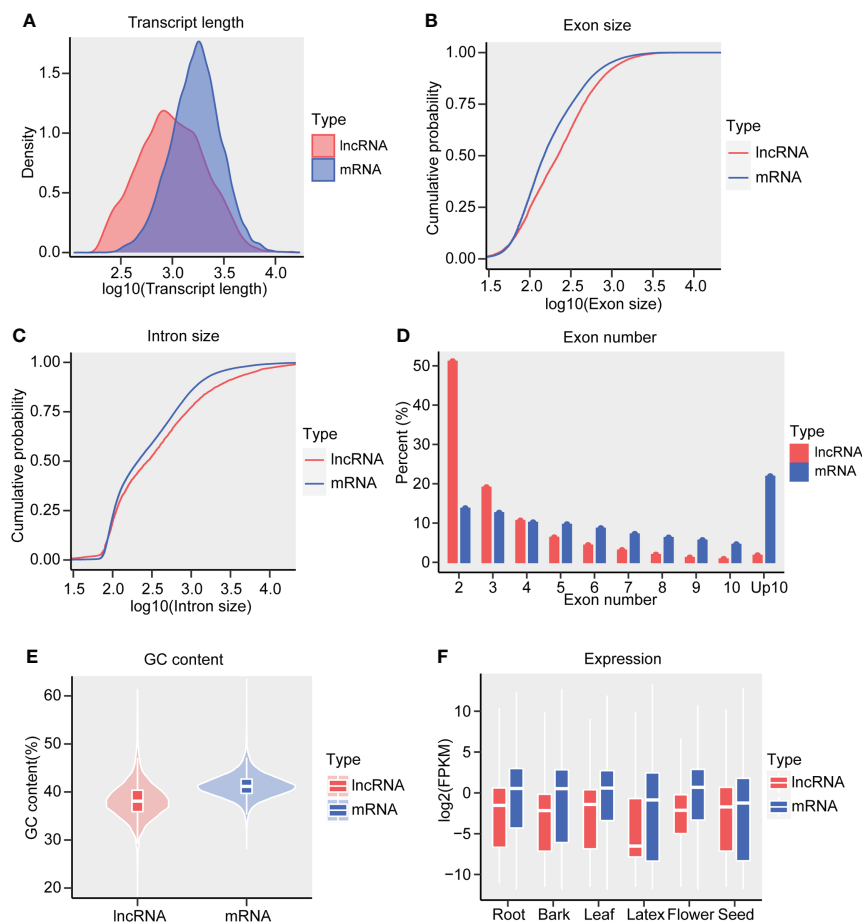
FIGURE 3
Characteristics of *Hevea brasiliensis* lncRNAs. **(A)** Density distribution of transcript length between lncRNA and mRNAs. The cumulative probability of **(B)** exon size and **(C)** intron size in lncRNAs and mRNAs. **(D)** Percent of lncRNAs and mRNAs in different exon numbers. **(E)** The GC content (%) of the lncRNAs and mRNAs. **(F)** The expression values of lncRNAs and mRNAs, Y axis indicate the expression value (log2(FPKM)).

mRNAs of lncRNAs were identified. Further, the Gene Ontology (GO) and KEGG enrichment analysis were conducted for the co-expressed mRNAs of each lncRNAs (Supplementary Tables 3, 4).

The above analyses showed that most of the seed-specific lncRNAs were enriched in the biological processes including endoplasmic reticulum to Golgi vesicle–mediated transport, intracellular protein transport, rRNA processing, mRNA methylation, mRNA splicing, transcription, translation, etc., suggesting that the seed-specific lncRNAs may be involved in transcription and protein synthesis and transport processes to prepare the storage-product accumulation for seed development (Figure 5A and Supplementary Figure 2). The leaf-specific lncRNAs were also preferentially enriched in translation, blue light signaling pathway, and photosynthesis, implying their relative activities in the protein synthesis and photosynthesis of the leaf (Figure 5B and Supplementary Figure 3A), demonstrating that the leaf-specific lncRNAs were likely to be involved in

regulating the process of photosynthesis. Differently, the root-specific lncRNAs were enriched in oxidation–reduction, hydrogen peroxide catabolic, and response to oxidative stress processes, etc. (Figure 5C and Supplementary Figure 3B), indicating that these lncRNAs may function in the process of reactive oxygen species (ROS) formation (Grene, 2002). Interestingly, the bark-specific lncRNAs presented high enrichment not only in the oxidation–reduction process, but also in the protein phosphorylation, plant organ development, cell differentiation, etc. (Figure 5D). In the flower, the lncRNAs were enriched in the carbohydrate metabolic process, cell wall modification, response to auxin, etc., implying their crucial roles in the regulation of the above pathways (Figure 5E and Supplementary Figure 4A). Unlike the flower-specific lncRNAs, the latex-specific lncRNAs are preferentially enriched in the biological processes including protein phosphorylation, transmembrane transport, protein refolding, etc., suggesting that latex-specific lncRNAs exhibited a vital role
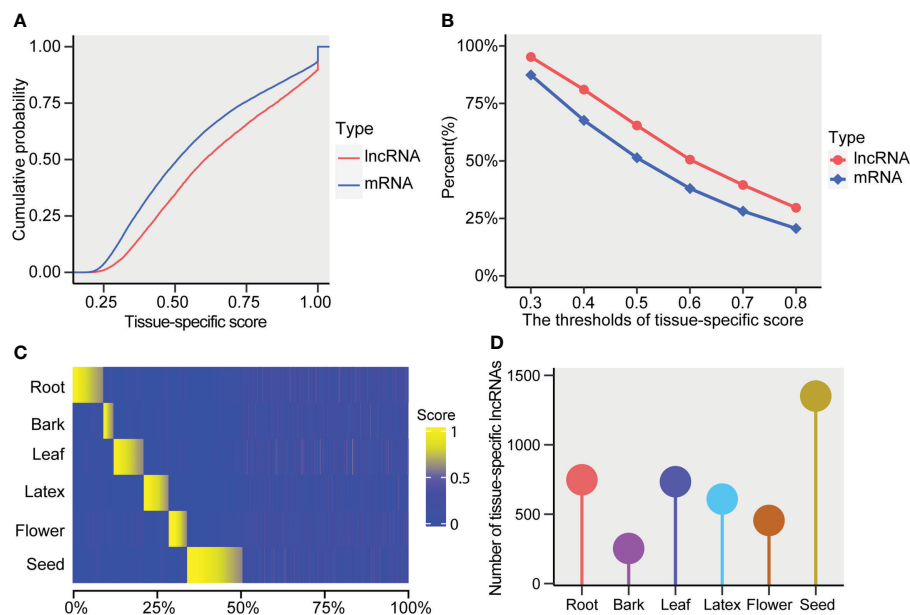
FIGURE 4
Tissue-specificity of the lncRNAs and mRNAs. **(A)** The cumulative probability of tissue-specific score in lncRNAs and mRNAs. **(B)** Percent of tissue-specific lncRNAs and mRNAs under varying thresholds of the tissue-specific score. **(C)** Heatmap of tissue-specific scores for the lncRNAs. **(D)** The number of tissue-specific lncRNAs in the tested tissues (root, bark, leaf, latex, flower, and seed).

in the protein modification and the solute transport (Figure 5F and Supplementary Figure 4B). Protein modification, as protein phosphorylation, was reported to be an important regulation form of latex biosynthetic pathway, implying the non-ignorable roles of lncRNAs in regulating latex biosynthesis. In line with GO analysis, KEGG enrichment also reveals that the tissue-specific lncRNAs are preferentially exhibited particular functions in corresponding tissues (Supplementary Figure 5).

Among the latex-specific lncRNAs, we found some lncRNAs such as Reyan73397.1363.1, Reyan73397.550.6, XR_002495983.1 are enriched in a lot of biological functions, including isoprenoid biosynthetic process, farnesyl diphosphate biosynthetic process (mevalonate pathway), isopentenyl diphosphate biosynthetic process (mevalonate pathway), jasmonic acid biosynthetic process, etc. (Figure 6). Meanwhile, in the latex, the KEGG enrichment of these tissue-specific lncRNAs are involved in terpenoid backbone biosynthesis, biosynthesis of secondary metabolites, and other pathways, suggesting their potential functions in the synthetic process of latex (Supplementary Figure 5).

The functional profile of non-tissue-specific lncRNAs

To provide a comprehensive functional profile of lncRNAs, we also conducted a functional analysis for the non-tissue-specific lncRNAs. Among them, there were 2,181 and 1,477 lncRNAs significantly enriched in the biological processes and KEGG pathways, respectively (Supplementary Tables 5, 6), and most of them were enriched in the translation, protein phosphorylation, transmembrane transport, glucose metabolic process, carbohydrate metabolic processes, etc. (Figure 7A). Consistently, among KEGG enrichment analysis, we found that ribosome pathway is in the top enrichment list for the non-tissue-specific lncRNAs (Figure 7B). The above results suggested that the non-tissue-specific lncRNAs are likely to play vital roles in the translation, transmembrane transport, and metabolic processes in multiple tissues of *Hevea brasiliensis*.

Discussion

A growing number of lncRNAs have been reported to play important roles in various biological processes (Statello et al., 2021). Although some studies on the lncRNAs of rubber have been published (Yin et al., 2019; Li et al., 2021; Liu et al., 2022), the previously published researches mainly focused on lncRNAs in one tissue (respectively focused on the leaf, latex, and bark) rather than multiple tissues. To conduct a comprehensive identification and characterization of rubber lncRNAs, we collected and integrated the datasets from multiple tissues in the present study (Figure 1 and Supplementary Table 1). As a
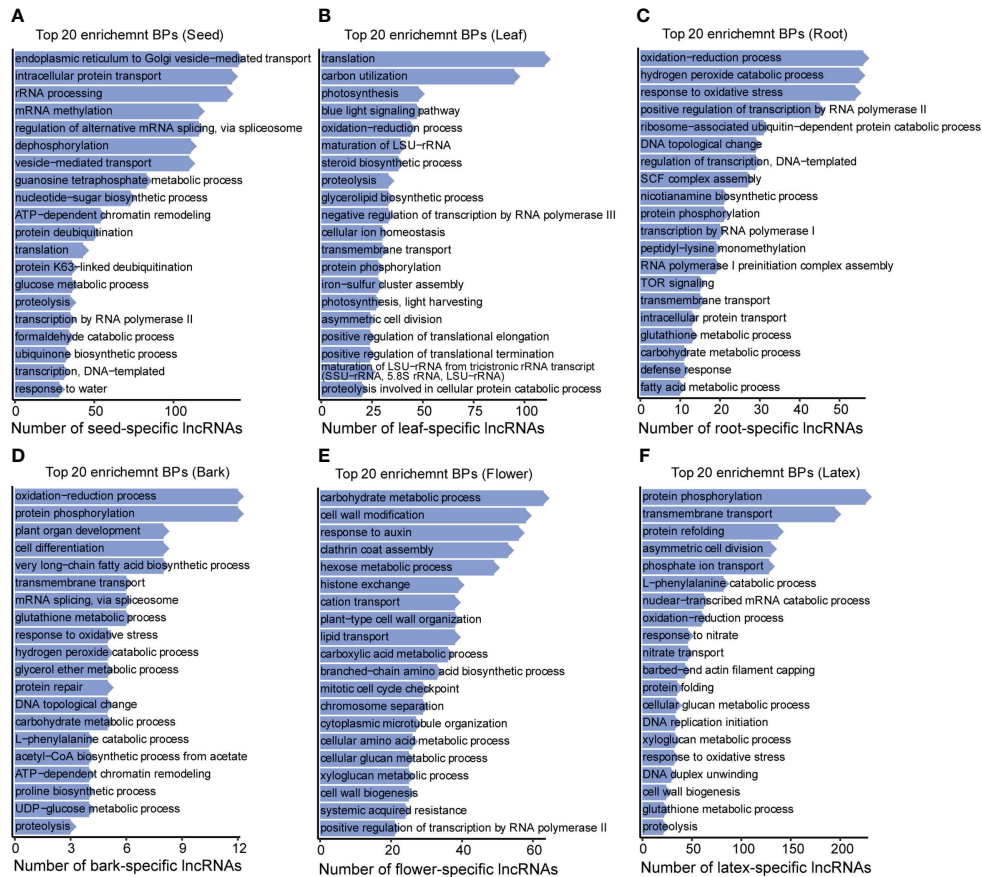
FIGURE 5
Top 20 enrichment BPs for the tissue-specific lncRNAs. Top 20 enrichment biological processes (BPs) for the tissue-specific lncRNAs in
(A) seed, **(B)** leaf, **(C)** root, **(D)** bark, **(E)** flower, and **(F)** latex, separately. Y axis is the term of biological processes, while X axis represents the
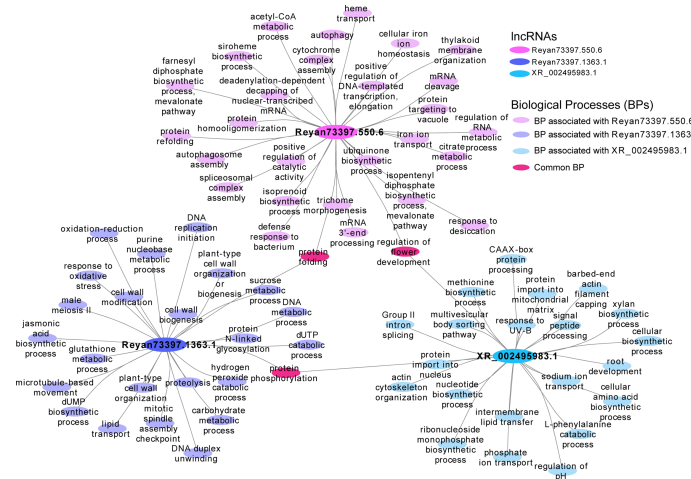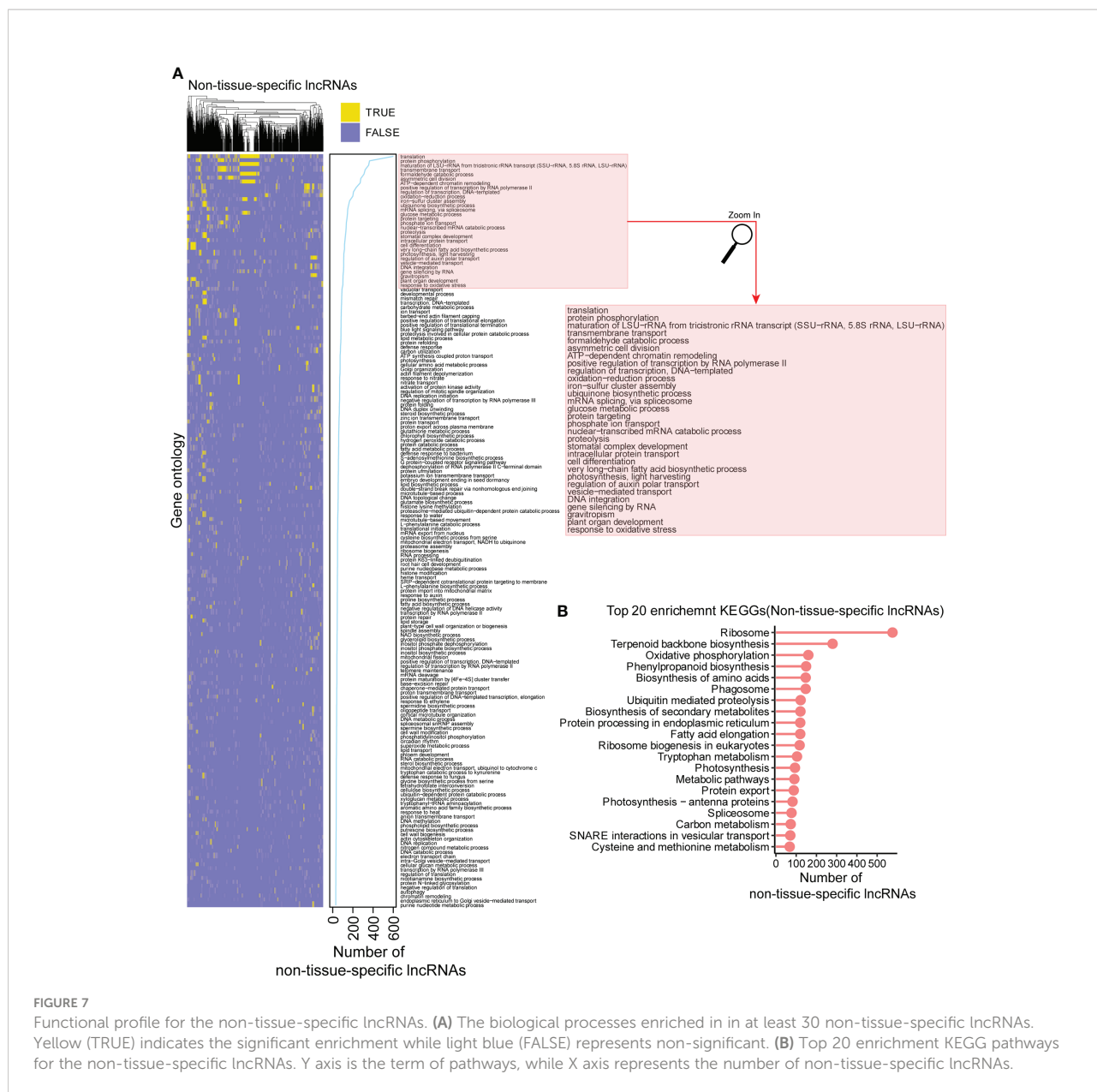number of tissue-specific lncRNAs.

FIGURE 6
Functional network of three latex-specific lncRNAs. The biological processes in association with the latex-specific lncRNAs, Reyan73397.1363.1,
Reyan73397.550.6, and XR_002495983.1. The big ellipse represents the lncRNAs, while the small ellipse indicates the biological processes.

FIGURE 7

Functional profile for the non-tissue-specific lncRNAs. **(A)** The biological processes enriched in in at least 30 non-tissue-specific lncRNAs. Yellow (TRUE) indicates the significant enrichment while light blue (FALSE) represents non-significant. **(B)** Top 20 enrichment KEGG pathways for the non-tissue-specific lncRNAs. Y axis is the term of pathways, while X axis represents the number of non-tissue-specific lncRNAs.

result, we identified about 12,029 lncRNAs and found that they exhibited distinctive features when compared with the mRNAs, such as the lncRNAs possessing shorter transcript length, less exon numbers, lower expression levels and GC content, etc. (Figure 3), which is consistent with the previous lncRNAs reported in *Hevea brasiliensis*, maize, *Solanum lycopersicum*, and *Camellia sinensis* (Li et al., 2014; Wang et al., 2018a; Yin et al., 2019; Wan et al., 2020; Li et al., 2021; Liu et al., 2022). These results demonstrated the reliability of our results and consistency of lncRNA characteristics between rubber and other plant species. In addition, we found *Hevea brasiliensis* lncRNAs exhibited higher tissue specificity than mRNAs (Figures 4A, B).

In this study, thousands of tissue-specific lncRNAs from six tissues were identified, more than those in previous researches (Yin et al., 2019; Li et al., 2021; Liu et al., 2022), which is mainly because of the number of tissues used for lncRNAs identification The results of functional analyses for these tissue-specific lncRNAs suggested that the tissue-specific lncRNAs could be involved in the growth and development regulation, as well as tissue-related biological processes in corresponding tissues. Which is consistent with the previous research that the lncRNAs can act in the regulation of development and stress responses, because of their significantly different expression patterns among tissues (Liu et al., 2012).

As a secondary metabolite of the rubber tree, natural rubber is synthesized from isopentenyl diphosphate which is synthesized *via* two different approaches: the plastidial 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway and the cytosolic mevalonate (MVA) pathway (Tang et al., 2016). Interestingly, three latex-specific lncRNAs (Reyan73397.1363.1, Reyan73397.550.6, and XR_002495983.1) were found to be involved in the latex synthetic pathway (the mevalonate pathway). These findings will provide new perspectives to understand the particular functions of tissue-specific lncRNAs and the regulation processes of natural rubber biosynthesis. Taking together, our genome-wide identification, characterization, and functional analysis of the lncRNAs will lay a theoretical foundation for further functional studies of lncRNA in *Hevea brasiliensis* and other plants.

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 in the article/Supplementary Material Table 1.

Author contributions

LW and BH planned, designed, and wrote the manuscript, JW and HC collected the data. LW, JW, and HC analyzed the data. All authors contributed to the article and approved the submitted version.

Funding

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1012576/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The genomic distribution correlation between lncRNAs and mRNAs. The genomic distribution correlation between lncRNAs and mRNAs in **(A)** + strand and **(B)** - strand. One point represents one bin in the genome. The correlation calculation (R) is based on the method of Pearson.

SUPPLEMENTARY FIGURE 2
The biological processes enriched in at least 10 seed-specific lncRNAs. Yellow (TRUE) indicates the significant enrichment while light blue (FALSE) is not.

SUPPLEMENTARY FIGURE 3
The biological processes enriched in at least 10 leaf/root-specific lncRNAs. The biological processes enriched in at least 10 **(A)** leaf-specific lncRNAs or **(B)** root-specific lncRNAs. Yellow (TRUE) indicates the significant enrichment while the light blue (FALSE) is not.

SUPPLEMENTARY FIGURE 4
The biological processes enriched in at least 10 flower/latex-specific lncRNAs. The biological processes enriched in at least 10 **(A)** flower-specific lncRNAs or **(B)** latex-specific lncRNAs. Yellow (TRUE) indicates the significant enrichment while the light blue (FALSE) is not.

SUPPLEMENTARY FIGURE 5
Top 20 enrichment KEGG pathways for the tissue-specific lncRNAs. Top 20 enrichment KEGG pathways for the tissue-specific lncRNAs in **(A)** seed, **(B)** leaf, **(C)** root, **(D)** bark, **(E)** flower, and **(F)** latex, separately. Y axis is the term of KEGG pathways, while X axis represents the number of tissue-specific lncRNAs.

SUPPLEMENTARY TABLE 1
Sample information of this study.

SUPPLEMENTARY TABLE 2
The classification of long non-coding RNAs in Hevea brasiliensis.

SUPPLEMENTARY TABLE 3
The functional profile of tissue-specific lncRNAs.

SUPPLEMENTARY TABLE 4
The KEGG pathway profile of tissue-specific lncRNAs.

SUPPLEMENTARY TABLE 5
The functional profile of non-tissue-specific lncRNAs.

SUPPLEMENTARY TABLE 6
The KEGG pathway profile of non-tissue-specific lncRNAs.

References

Andrews, S. (2010) *FastQC: a quality control tool for high throughput sequence data*. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Bhatia, G., Singh, A., Verma, D., Sharma, S., and Singh, K. (2020). Genome-wide investigation of regulatory roles of lncRNAs in response to heat and drought stress in *Brassica juncea*. *Environ. Exp. Bot.* 171, 103922. doi: 10.1016/j.envexpbot.2019.103922

Bu, D., Luo, H., Huo, P., Wang, Z., Zhang, S., He, Z., et al. (2021). KOBAS-i: Intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res.* 49, W317–W325. doi: 10.1093/nar/gkab447

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST plus: Architecture and applications. *BMC Bioinform.* 10, 421. doi: 10.1186/1471-2105-10-421

Cech, T. R., and Steitz, J. A. (2014). The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* 157, 77–94. doi: 10.1016/j.cell.2014.03.008

Chekanova, J. A., Gregory, B. D., Reverdatto, S. V., Chen, H., Kumar, R., Hooker, T., et al. (2007). Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the *Arabidopsis* transcriptome. *Cell* 131, 1340–1353. doi: 10.1016/j.cell.2007.10.056

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775–1789. doi: 10.1101/gr.132159.111

Di, C., Yuan, J., Wu, Y., Li, J., Lin, H., Hu, L., et al. (2014). Characterization of stress-responsive lncRNAs in *Arabidopsis thaliana* by integrating expression, epigenetic and structural features. *Plant J.* 80, 848–861. doi: 10.1111/tpj.12679

Geisler, S., and Coller, J. (2013). RNA In unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell Biol.* 14, 699–712. doi: 10.1038/nrm3679

Gong, X. X., Yan, B. Y., Hu, J., Yang, C. P., Li, Y. J., Liu, J. P., et al. (2018). Transcriptome profiling of rubber tree (*Hevea brasiliensis*) discovers candidate regulators of the cold stress response. *Genes Genom.* 40, 1181–1197. doi: 10.1007/s13258-018-0681-5

Grene, R. (2002). Oxidative stress and acclimation mechanisms in plants. *Arabidopsis Book* 1, e0036. doi: 10.1199/tab.0036.1

Han, S., Liang, Y., Ma, Q., Xu, Y., Zhang, Y., Du, W., et al. (2019). LncFinder: an integrated platform for long non-coding RNA identification utilizing sequence intrinsic composition, structural information and physicochemical property. *Brief Bioinform.* 20, 2009–2027. doi: 10.1093/bib/bby065

Huang, X., Zhang, H., Wang, Q., Guo, R., Wei, L., Song, H., et al. (2021). Genome-wide identification and characterization of long non-coding RNAs involved in flag leaf senescence of rice. *Plant Mol. Biol.* 105, 655–684. doi: 10.1007/s11103-021-01121-3

Jin, J., Liu, J., Wang, H., Wong, L., and Chua, N. H. (2013). PLncDB: Plant long non-coding RNA database. *Bioinformatics* 29, 1068–1071. doi: 10.1093/bioinformatics/btt107

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W. Z., McAnulla, C., et al. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Kang, Y. J., Yang, D. C., Kong, L., Hou, M., Meng, Y. Q., Wei, L., et al. (2017). CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* 45, W12–W16. doi: 10.1093/nar/gkx428

Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T., et al. (2007). RNA Maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 316, 1484–1488. doi: 10.1126/science.1138341

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., et al. (2013). Software for computing and annotating genomic ranges. *PloS Comput. Biol.* 9, e1003118. doi: 10.1371/journal.pcbi.1003118

Li, L., Eichten, S. R., Shimizu, R., Petsch, K., Yeh, C. T., Wu, W., et al. (2014). Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol.* 15, R40. doi: 10.1186/gb-2014-15-2-r40

Lindgreen, S. (2012). AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res. Notes* 5, 337. doi: 10.1186/1756-0500-5-337

Lin, X., Lin, W., Ku, Y. S., Wong, F. L., Li, M. W., Lam, H. M., et al. (2020). Analysis of soybean long non-coding RNAs reveals a subset of small peptide-coding transcripts. *Plant Physiol.* 182, 1359–1374. doi: 10.1104/pp.19.01324

Liu, H., Hu, Y., Yuan, K., Feng, C., He, Q., Sun, L., et al. (2022). Genome-wide identification of lncRNAs, miRNAs, mRNAs and their regulatory networks involved in tapping panel dryness in rubber tree (*Hevea brasiliensis*). *Tree Physiol.* 42, 629–645. doi: 10.1093/treephys/tpab120

Liu, J., Jung, C., Xu, J., Wang, H., Deng, S., Bernad, L., et al. (2012). Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* 24, 4333–4345. doi: 10.1105/tpc.112.102855

Li, H. L., Wang, Y., Guo, D., Zhu, J. H., and Peng, S. Q. (2021). Differential expression of lncRNAs and miRNAs between self-rooting juvenile and donor clones unveils novel insight into the molecular regulation of rubber biosynthesis in *Hevea brasiliensis*. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.740597

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913

Negri, T. D. C., Alves, W. A. L., Bugatti, P. H., Saito, P. T. M., Domingues, D. S., and Paschoal, A. R. (2019). Pattern recognition analysis on long noncoding RNAs: a tool for prediction in plants. *Brief Bioinform.* 20, 682–689. doi: 10.1093/bib/bby034

Pang, J., Zhang, X., Ma, X., and Zhao, J. (2019). Spatio-temporal transcriptional dynamics of maize long non-coding RNAs responsive to drought stress. *Genes (Basel)* 10, 138. doi: 10.3390/genes10020138

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197

Qin, T., Zhao, H., Cui, P., Albesher, N., and Xiong, L. (2017). A nucleus-localized long non-coding RNA enhances drought and salt stress tolerance. *Plant Physiol.* 175, 1321–1336. doi: 10.1104/pp.17.00574

Rigo, R., Bazin, J., Romero-Barrios, N., Moison, M., Lucero, L., Christ, A., et al. (2020). The *Arabidopsis* lncRNA *ASCO* modulates the transcriptome through interaction with splicing factors. *EMBO Rep.* 21, e48977. doi: 10.15252/embr.201948977

Rinn, J. L., and Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* 81, 145–166. doi: 10.1146/annurev-biochem-051410-092902

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Statello, L., Guo, C. J., Chen, L. L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* 22, 96–118. doi: 10.1038/s41580-020-00315-9

Tang, C., Yang, M., Fang, Y., Luo, Y., Gao, S., Xiao, X., et al. (2016). The rubber tree genome reveals new insights into rubber production and species adaptation. *Nat. Plants* 2, 16073. doi: 10.1038/nplants.2016.73

van Beilen, J. B., and Poirier, Y. (2007). Establishment of new crops for the production of natural rubber. *Trends Biotechnol.* 25, 522–529. doi: 10.1016/j.tibtech.2007.08.009

Wang, H., Chung, P. J., Liu, J., Jang, I. C., Kean, M. J., Xu, J., et al. (2014). Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. *Genome Res.* 24, 444–453. doi: 10.1101/gr.165555.113

Wang, Y., Li, J., Deng, X. W., and Zhu, D. (2018b). *Arabidopsis* noncoding RNA modulates seedling greening during deetiolation. *Sci. China Life Sci.* 61, 199–203. doi: 10.1007/s11427-017-9187-9

Wang, Y., Li, H. L., Zhou, Y. K., Guo, D., Zhu, J. H., and Peng, S. Q. (2021). Transcriptomes analysis reveals novel insight into the molecular mechanisms of somatic embryogenesis in *Hevea brasiliensis*. *BMC Genom.* 22, 183. doi: 10.1186/s12864-021-07501-9

Wang, Y., Luo, X., Sun, F., Hu, J., Zha, X., Su, W., et al. (2018c). Overexpressing lncRNA *LAIR* increases grain yield and regulates neighbouring gene cluster expression in rice. *Nat. Commun.* 9, 3516. doi: 10.1038/s41467-018-05829-7

Wang, M., Zhao, W., Gao, L., and Zhao, L. (2018a). Genome-wide profiling of long non-coding RNAs from tomato and a comparison with mRNAs associated with the regulation of fruit ripening. *BMC Plant Biol.* 18, 75. doi: 10.1186/s12870-018-1300-y

Wan, S., Zhang, Y., Duan, M., Huang, L., Wang, W., Xu, Q., et al. (2020). Integrated analysis of long non-coding RNAs (lncRNAs) and mRNAs reveals the regulatory role of lncRNAs associated with salt resistance in *Camellia sinensis*. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00218

Wucher, V., Legeai, F., Hedan, B., Rizk, G., Lagoutte, L., Leeb, T., et al. (2017). FEELnc: A tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res.* 45, e57. doi: 10.1093/nar/gkw1306

Xing, Q., Zhang, W., Liu, M., Li, L., Li, X., and Yan, J. (2019). Genome-wide identification of long non-coding RNAs responsive to *Lasiodiplodia theobromae* infection in grapevine. *Evol. Bioinform. Online* 15, 1176934319841362. doi: 10.1177/1176934319841362

Yin, H., Zhang, X., Zhang, B., Luo, H., and He, C. (2019). Revealing the dominant long noncoding RNAs responding to the infection with *Colletotrichum gloeosporioides* in *Hevea brasiliensis. Biol. Direct.* 14, 7. doi: 10.1186/s13062-019-0235-z

Zhang, H., Chen, X., Wang, C., Xu, Z., Wang, Y., Liu, X., et al. (2013). Long non-coding genes implicated in response to stripe rust pathogen stress in wheat (*Triticum aestivum* l.). *Mol. Biol. Rep.* 40, 6245–6253. doi: 10.1007/s11033-013-2736-7

Zhang, H., Guo, H., Hu, W., and Ji, W. (2020). The emerging role of long non-coding RNAs in plant defense against fungal stress. *Int. J. Mol. Sci.* 21, 2659. doi: 10.3390/ijms21082659

Zhang, Y. C., Liao, J. Y., Li, Z. Y., Yu, Y., Zhang, J. P., Li, Q. F., et al. (2014). Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol.* 15, 512. doi: 10.1186/s13059-014-0512-1

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, Q., Wan, Q., Jiang, Y., Liu, J., Qiang, L., and Sun, L. (2020). A landscape of murine long non-coding RNAs reveals the leading transcriptome alterations in adipose tissue during aging. *Cell Rep.* 31, 107694. doi: 10.1016/j.celrep.2020.107694