



Original article

Tissue-specific analysis of *Coffea arabica* L. transcriptome revealed potential regulatory roles of lncRNAs

Eslam M. Abdel-Salam^{a,*}, Ahmed A. Qahtan^a, Abdel-Rhman Z. Gaafar^a, Assem I. Zein El-Abedein^b, Aref M. Alshameri^a, Abdullah M. Alhamdan^{b,c}

^a Department of Botany & Microbiology, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia

^b Chair of Dates Industry and Technology, College of Food and Agricultural Sciences, King Saud University, PO Box 2460, Riyadh 11451, Saudi Arabia

^c Department of Agricultural Engineering, College of Food and Agricultural Sciences, King Saud University, PO Box 2460, Riyadh 11451, Saudi Arabia



ARTICLE INFO

Article history:

Received 29 May 2021

Revised 13 June 2021

Accepted 4 July 2021

Available online 10 July 2021

Keywords:

Coffee

Caffeine

Differential expression

Gene ontology

Non-coding RNA

ABSTRACT

Long non-coding RNAs (lncRNAs) play pivot roles in regulating mRNA expression in eukaryotic organisms without coding any proteins. In the current study, a comprehensive analysis of 260 published RNA-Seq datasets collected from different tissues (fruits, leaves, stems, and roots) of *Coffea arabica* L. was performed to discover potential lncRNAs. A total of 10,564 unique lncRNAs were identified. Our results showed that 77.14% of the lncRNAs were intergenic and 60.39% of them are located within 5 Kbp from the partner gene. In general, all the identified lncRNAs showed shorter lengths, fewer number of exons, and lower expression levels as compared to mRNAs in different studied tissues. Several lncRNAs were determined as differentially expressed (DE) in fruits as compared to leaves, stems, or roots. The functional characterization of the DE lncRNAs revealed their roles in regulating significant biological processes in different tissues of *C. arabica*. The current study provides a comprehensive analysis and dataset of lncRNAs in *C. arabica* that could be utilized in further studies concerning the roles of these molecules in plant cells.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

In eukaryotic genomes, non-coding RNA molecules (ncRNAs) constitute a major group of transcripts (Eddy, 2001). This group is typically classified into two sub-sections viz., regulatory ncRNAs and housekeeping ncRNAs. The housekeeping ncRNAs are essential to the basic functions of cells and include tRNAs, rRNAs, snRNAs, and snoRNAs (Zhao et al., 2019). On the other hand, regulatory ncRNAs are expressed in response to certain changes resulted from exposure to stress, in specific tissues, or during different developmental stages of the plant (Cuperus et al., 2011). This group of ncRNAs is further divided based on their length to small ncRNAs with length < 200 base pairs (bp) and long ncRNAs (lncRNAs) with length > 200 bp. Extensive research efforts were invested to exam-

ine the regulatory roles of small ncRNAs either in animals or plants; however, research concerning the roles of lncRNAs, especially in plants, is lacking (Ponting et al., 2009; Chen, 2012; Zhao et al., 2019). Recent reports indicated the regulatory roles of lncRNAs in different biological processes in the plant e.g., transcriptional regulation, cell growth and proliferation, and disease resistance pathways (Rinn and Chang, 2012; Yang et al., 2014; Liang et al., 2015).

Previous reports correlate the lncRNAs with different regulatory mechanisms of biological processes; however, these molecules are known to lack open reading frames (ORFs) (Rinn and Chang, 2012). The absence of ORFs is correlated with the role of lncRNAs in regulating their target genes (Sun et al., 2020). lncRNAs regulate their potential targets either via *trans* role or *cis* role, where the regulatory role of the lncRNA molecule is related to the adjacent coding gene. In this case, the lncRNA is either located downstream or upstream the regulated coding gene. The regulatory action of the lncRNA may happen at the transcriptional or post-transcriptional levels.

Data generated by high-throughput sequencing, especially RNA sequencing (RNA-Seq), techniques was successfully utilized to identify a large number of lncRNAs in different eukaryotic

* Corresponding author.

E-mail address: eabdel salam@ksu.edu.sa (E.M. Abdel-Salam).

Peer review under responsibility of King Saud University.



organisms. In all cases, lncRNAs were characterized by tissue-specific and low expression levels, in addition to less conservation between different species (Cabili et al., 2011; Wang et al., 2014). Roughly 6000 intergenic lncRNAs were identified and characterized in *Arabidopsis* (Di et al., 2014; Wang et al., 2014). Similarly, a massive database of lncRNAs in maize was developed and specific drought-responsive molecules were identified (Li et al., 2014; Zhang et al., 2014a). A comprehensive database of lncRNAs was also generated for cotton with the identification of potential molecules involved in the regulation of fiber initiation and elongation (Wang et al., 2015a). RNA-Seq data was utilized to identify sexual reproduction-related lncRNAs in rice (Zhang et al., 2014b). Similarly, RNA-Seq technique was used to detect 1565 lncRNAs in tomatoes with the identification of TYLCV infection-responsive lncRNAs (Wang et al., 2015b) and the potential regulatory roles of these molecules in fruit ripening and the origin of Lycopersicon (Wang et al., 2016).

Coffee is considered among the most important agricultural products in the world. The two main coffee species cultivated globally are the Arabian coffee (*Coffea arabica* L.) and the robusta coffee (*C. canephora* L.). *C. arabica* occupies 70% of the world coffee cultivation and production (Lashermes et al., 1999). Growing coffee represents the sole income source for more than 12.5 million households around the world (Montagnon et al., 2021). Therefore, coffee represents the most valuable tropical export crop worldwide, with an annual retail value of roughly 88 billion dollars (Valencia-Lozano et al., 2021). In the current study, we performed a large-scale analysis of the RNA-Seq published data via analyzing data from 260 published RNA-Seq experiments collected from different tissues of *Coffea arabica* plant to discover the potential tissue-specific lncRNAs. Moreover, the regulatory roles of the identified lncRNAs were examined via functional characterization of their potential mRNA gene targets using *cis* methods. The results of the current study will provide a comprehensive database of lncRNAs in *C. arabica* for further genome-wide studies, especially on transcriptional regulation.

2. Materials and methods

2.1. Data curation

A total of 260 experiments available in the Short Read Archive (SRA) on the National Center for Biotechnology Information (NCBI) website were downloaded and used for the analysis performed in the current study. The data was classified into four main groups based on the collected tissue, namely fruits (92 experiments), stems (108 experiments), leaves (52 experiments), and roots (8 experiments). The accession numbers of the obtained data and their phenotypic and technical description are shown in Supplementary Table (S1).

2.2. Data preprocessing and lncRNAs identification

FastQC v0.11.9 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was used to examine the quality of the obtained data. Trimmomatic v0.40 (Bolger et al., 2014) was used to remove low-quality reads. Low-quality reads for the purpose of our analysis were defined as follows: (1) reads containing adapter sequences, (2) reads containing more than 10% of N bases, and (3) reads with an average Phred quality score more than 30. Hisat2 v2.2.1 (Kim et al., 2019) was used to align all clean reads against the *C. arabica* reference genome (Cara_1.0, GCA_003713225.1) downloaded along with annotation files from the NCBI genome database. StringTie2 v2.1.5 (Kovaka et al., 2019) was used to assemble the transcripts of each experiment separately. After

merging all the assembled transcripts, GffCompare v0.12.2 (Pertea and Pertea, 2020) was used to compare them with the reference genome. The total number of assembled transcripts was 247,665 transcripts including 12,294 non-previously annotated transcripts and 85,596 exactly matching the known transcripts.

Potential lncRNAs in the assembled transcripts were identified using FEELnc v0.2 (Wucher et al., 2017). According to the analysis pipeline, short transcripts (200 bp or less) were removed using FEELncfilter. The coding potential score of the remaining transcripts was calculated using FEELncodpot. RNAs with a length ≥ 200 bp, coding potential score < 0.5 , and open reading frame (ORF) covers $< 50\%$ of length were defined as lncRNAs. FEELncclassifier was used to identify the potential regulatory roles of the identified lncRNAs via classifying these lncRNAs to the localization and the direction of transcription of proximal RNA transcripts.

2.3. Differential expression analysis

Ballgown v2.22.0 (Fu et al., 2021) on R programming language was used to identify differentially expressed (DE) transcripts in different studied pair-wise comparisons (fruits vs. stems, fruits vs. leaves, and fruits vs. roots). The abundance of assembled transcripts was expressed as fragments per kilobase of transcript per million mapped reads (FPKM). The transcript with P -value < 0.05 and $|\log_2(\text{fold change})|$ greater than 1 was considered as DE.

2.4. Annotation of DE lncRNAs

Gene ontology (GO) enrichment analysis of the gene targets of all DE lncRNAs was performed using Goseq v1.42.0 (Young et al., 2010) on R. The GO term with a corrected P -value < 0.05 was considered as significantly enriched.

3. Results

3.1. Genome-wide identification of *C. Arabica* lncRNAs

After mapping all the reads to the *C. arabica* reference genome, a total of 247,665 transcripts were assembled representing 62,833 unique transcripts among them 12,294 transcripts were non-previously annotated. The lengths of the assembled transcripts ranged from 23 to 60,411 bp with an average length of 2,117 bp (Fig. 1a). Approximately 50% of the assembled genes were represented by one transcript only. The maximum transcripts assembled for a single gene were 259 transcripts (Fig. 1b). FEELnc analysis led to the identification of 10,564 unique lncRNAs (length ≥ 200 bp, ORF cover $< 50\%$, potential coding score < 0.05). The length of the identified lncRNAs ranged from 201 to 20,208 bp with an average length of 573 bp.

3.2. Characterization of *C. arabica* lncRNAs

The locations of all the identified lncRNAs were analyzed relative to their partner genes. Firstly, lncRNAs were classified into either genic (overlaps partner gene) or intergenic (does not overlap partner gene). The results showed that 77.14% of the lncRNAs were intergenic (Fig. 2a). Our analysis revealed that 60.39% of the intergenic lncRNAs located within 5 Kbp from the partner gene either downstream or upstream (Fig. 2b). The genic lncRNAs showed almost even distribution with regard to their location either in exonic or intronic regions (Fig. 2c). Most of the identified lncRNAs (62.63%) had only two exons and $< 6.5\%$ had more than 5 exons (Fig. 2d).

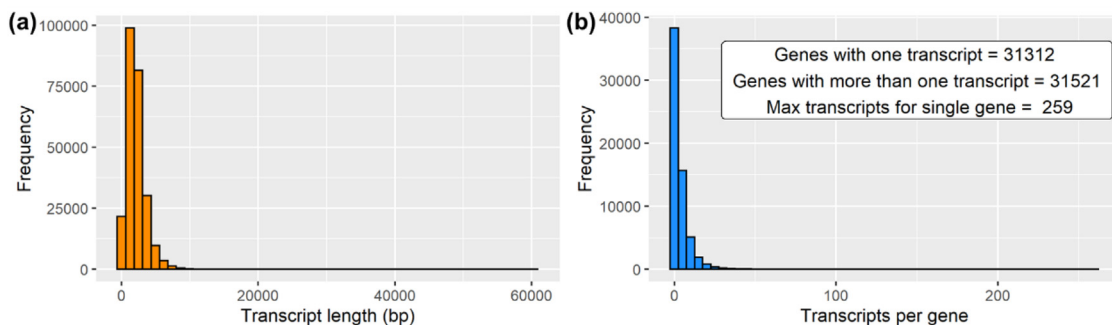


Fig. 1. Length distribution of the assembled transcripts (a) and distribution of number of exons in each assembled transcript (b).

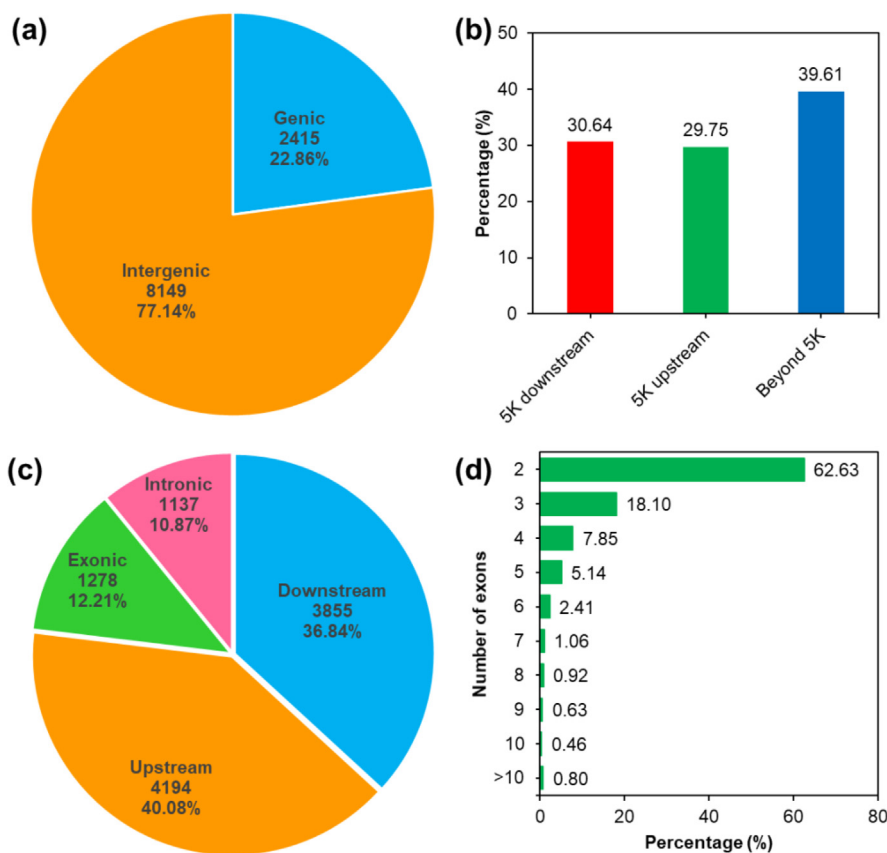


Fig. 2. Characterization of lncRNAs identified in different tissues of *Coffea arabica*. (a) Type of lncRNAs identified by FEELnc as either genic (overlaps partner gene) or intergenic (does not overlap partner gene), (b) the proportion of lncRNAs located within 5 K bp either downstream or upstream the partner gene and those which located beyond 5 K bp of the partner gene, (c) percentage of lncRNAs according to their location relative to the partner gene, and (d) number of exons in the identified lncRNAs.

3.3. Analysis of expression levels in different tissues

The overall genetic expression was examined in *C. arabica* via analyzing data from 260 different samples representing 4 major tissues, namely fruits (92 samples), stems (108 samples), leaves (52 samples), and roots (8 samples). The obtained results revealed slight variation in the expression levels between different tissues (Fig. 3a). It was evident that leaf and stem tissues have slightly higher expression levels than fruit and root tissues. The density blot showed that root tissues had the lowest levels of expression among the different tissues (Fig. 3b).

The average expression of mRNA and lncRNAs in all the studied tissues of *C. arabica* was examined. The obtained results clearly indicated that mRNA was expressed in higher levels as compared

to lncRNAs in all tissues. These results are depicted via violin (Fig. 3c) and density (Fig. 3d) plots of the log₁₀ transformation of the FPKM values.

3.4. Identification and functional characterization of DE lncRNAs

The transcriptomic data obtained from the 260 studied samples was utilized to examine the tissue-specificity of the identified DE lncRNAs in *C. arabica* different tissues. Fruit tissues were compared with all the other tissues. A total of 158 unique DE lncRNAs were identified in different pairwise comparisons (Fig. 4a). Only 35 lncRNA molecules were identified as common DE between all comparisons. Comparing fruits with leaves showed the highest number (50) of unique DE lncRNAs. DE lncRNAs identified in fruits as

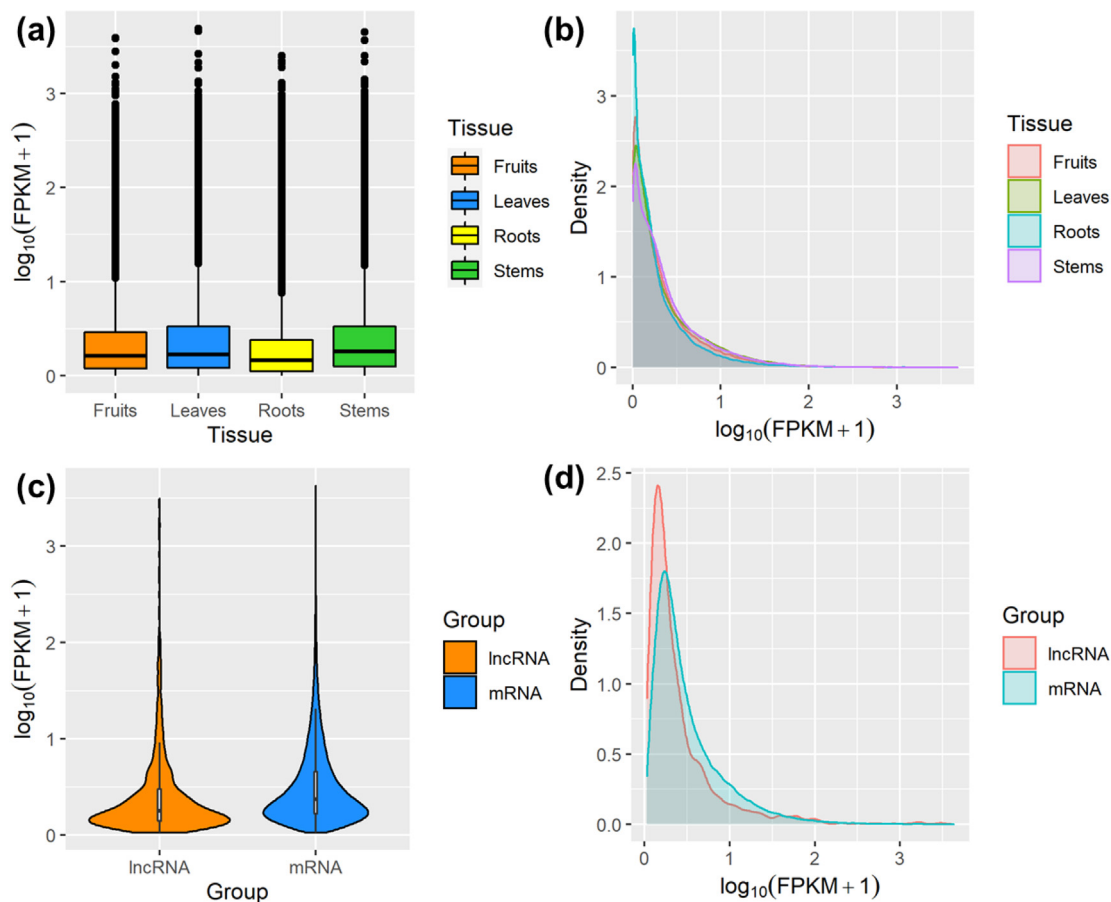


Fig. 3. The variation in FPKM values of all the expressed transcripts in fruits, stems, leaves, and roots was compared and visualized using boxplots (a) and density distribution profiles (b). The difference in the average expression levels between lncRNAs and mRNAs in all the studied tissues showed that mRNAs were expressed more than lncRNAs as depicted in the violin plots (c) and density distribution profiles (d).

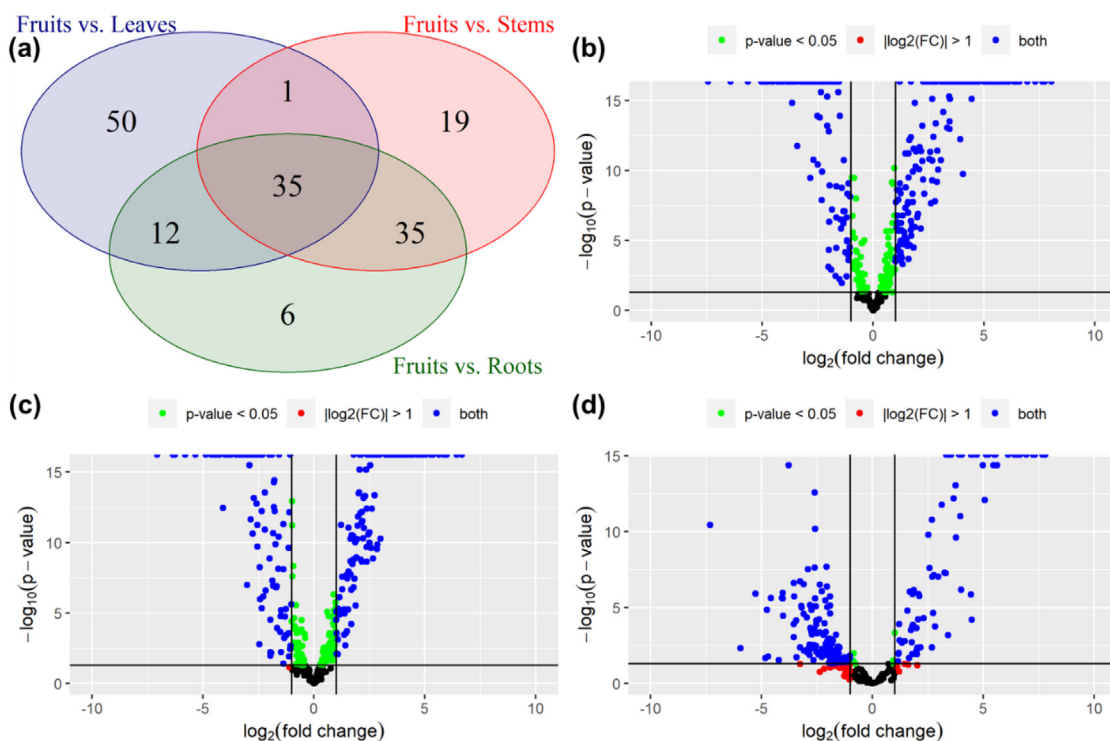


Fig. 4. (a) Number of DE lncRNAs in fruits as compared to leaves, stems, and roots. The expression landscape of lncRNAs in different pairwise comparisons is depicted by the volcano plots between (b) fruits vs. leaves, (c) fruits vs. stems, and (d) fruits vs. roots.

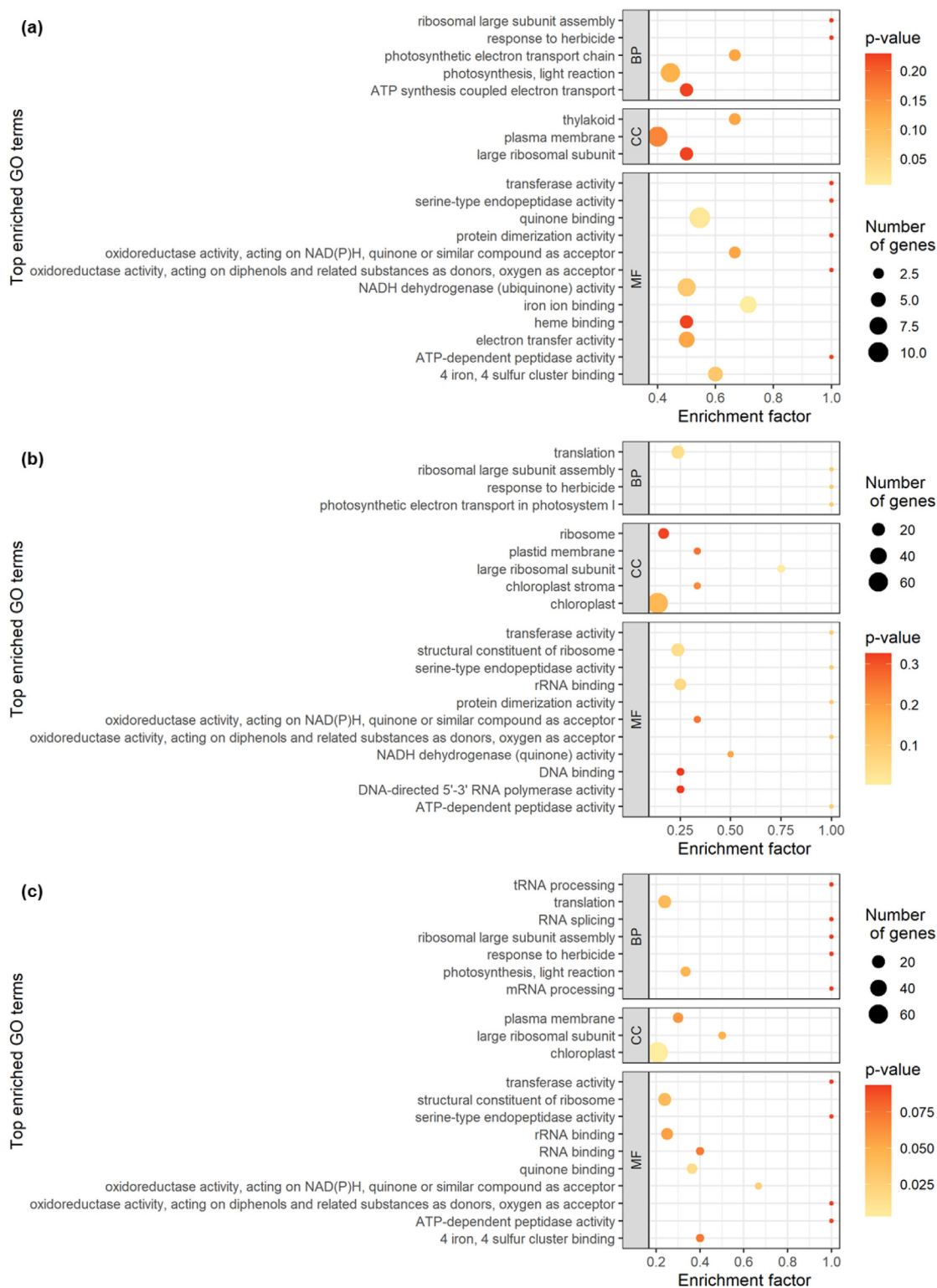


Fig. 5. The most enriched GO terms in *C. arabica* fruit tissues as compared to (a) leaves, (b) stems, and (c) roots. BP: biological process, CC: cellular component, MF: molecular function.

compared to stems or roots shared 35 similar lncRNAs. Comparing fruits with leaves showed a higher number of up-regulated DE lncRNAs (Fig. 4b), and a similar pattern was observed in the comparison of fruits with stems (Fig. 4c). Contrarily, the number of down-regulated DE lncRNAs was higher than up-regulated ones when comparing fruits with roots (Fig. 4d). A full list of the identi-

fied lncRNAs and their potential target(s) could be found in Supplementary Table (S2).

The target mRNA genes of all identified lncRNAs were identified in each pairwise comparison and examined for their potential functional role based on GO classification. Enriched GO terms were classified into 3 major categories viz., biological process (BP),

cellular component (CC), and molecular function (MF). DE lncRNAs, that were identified in comparison of fruits with leaves, showed a higher number of enriched GO terms in the MF category (Fig. 5a). GO:0005506 (iron ion binding), GO:0048038 (quinone binding), and GO:0008137 (NADH dehydrogenase (ubiquinone) activity) were the highest enriched terms in fruits as compared to leaves. Targets of DE lncRNAs, that were identified in fruits as compared to stems, enriched several GO terms especially in the MF category (Fig. 5b) including GO:0004176 (ATP-dependent peptidase activity) and GO:0003735 (structural constituent of ribosome). Interestingly, oxidoreductase activity (GO:0016682 and GO:0016655) was enriched by the target genes identified when comparing fruits with any other tissue. GO:0048038 (quinone binding) was also enriched by targets of DE lncRNAs in fruits as compared to roots (Fig. 5c).

4. Discussion

In eukaryotic organisms, the transcriptomic landscape consists mainly of protein-coding genes and ncRNAs. Deeper biotechnological analyses and studies revealed that the genetic expression process in eukaryotes is extensively complex. Moreover, plants characterize by bigger genome sizes and higher ploidy levels compared to other eukaryotes and possess increasing the genetic expression complexity (Berretta and Morillon, 2009; Dinger et al., 2009). Non-coding genes are generally classified into short non-coding genes (e.g., small RNAs and micro RNAs) and long non-coding genes. The regulatory roles of short non-coding RNA groups were extensively studied and fully elaborated (Ghildiyal and Zamore, 2009). On the other hand, information regarding the potential roles of lncRNAs in different biological processes is still lacking and needs more research efforts. Previous studies indicated the pivot roles of lncRNAs in organisms' growth and development (Wang and Chang, 2011).

In the current study, we analyzed 260 published transcriptomic RNA-Seq datasets generated from different tissues of *C. arabica* obtained from the NCBI SRA. The analysis led to the assembly of 247,665 transcripts. Using FEELnc pipeline, 10,564 unique lncRNAs were identified. Our results showed that lncRNAs are shorter and have lower expression levels than mRNAs. These results agree with the previous findings indicated that in different organisms, lncRNAs generally have fewer exons, shorter lengths, and lower expression levels as compared to mRNAs (Cabali et al., 2011; Pauli et al., 2012; Zhao et al., 2019).

Caffeine is a purine alkaloid that is naturally biosynthesized in both leaves and fruits or seeds of *C. arabica*. In leaves, caffeine has insecticidal effects (Nathanson, 1984), while in fruits, it inhibits the germination of the seeds of other competing species (Pacheco et al., 2008). The main precursor of caffeine is xanthosine resulted from the purine metabolism pathway. Xanthosine conversion into caffeine happens via several steps that is regulated by different coding genes including xanthosine methyltransferase, theobromine synthase, and caffeine synthase. In *C. arabica*, 7-methylxanthosine synthase 1-like (LOC113710464) regulates the conversion of xanthosine into 7-methylxanthosine which in turn converted into 7-methylxansine. The latter is converted into theobromine by the action of theobromine synthase (LOC113710465). In the final step, caffeine synthase (LOC113709599) catalyzes the conversion of theobromine into caffeine (Denoëud et al., 2014). Our differential expression analysis showed several DE lncRNA. Comparing fruits with other tissues revealed the presence of 3 DE lncRNA molecules that may regulate the expression of genes involved in caffeine biosynthesis. MSTRG.42643 may regulate the expression of 7-methylxanthosine synthase 1-like gene via *cis* role as it is located 3159 bp upstream this gene and may regulate its expression. Similarly, MSTRG.42638 may regulate the expression

of theobromine synthase. The lncRNA MSTRG.42638 was found to be located 2972 bp upstream the probable caffeine synthase 3 gene indicating its potential *cis* regulatory role.

5. Conclusions

In the current study, 10,564 unique lncRNAs were identified using 260 published transcriptomic datasets from 4 different tissues of *C. arabica*. Our results revealed that lncRNAs are characterized by shorter length and lower expression levels as compared to mRNAs in *C. arabica*. The majority of the identified lncRNAs were classified as intergenic and located within 5 Kbp from the target gene. Several lncRNAs were differentially expressed in fruits as compared to leaves, stems, or roots. The roles of these molecules related to important biological processes as revealed by GO enrichment analysis. The current study lays the foundation for future studies examining the roles of lncRNAs in regulating mRNA expression in different tissues of *C. arabica* via generating a comprehensive database of lncRNAs.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Saud University for funding this work through the research group Project no. RG-1441-472.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.sjbs.2021.07.006>.

References

- Berretta, J., Morillon, A., 2009. Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Rep.* 10, 973–982. <https://doi.org/10.1038/embor.2009.181>.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Cabali, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., Rinn, J.L., 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927. <https://doi.org/10.1101/gad.17446611>.
- Chen, X., 2012. Small RNAs in development – insights from plants. *Curr. Opin. Genet. Dev.* 22, 361–367. <https://doi.org/10.1016/j.gde.2012.04.004>.
- Cuperus, J.T., Fahlgren, N., Carrington, J.C., 2011. Evolution and Functional Diversification of *MIRNA* Genes. *Plant Cell* 23, 431–442. <https://doi.org/10.1105/tpc.110.082784>.
- Denoëud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M., Zheng, C., Alberti, A., Anthony, F., Aprea, G., Aury, J.-M., Bento, P., Bernard, M., Bocs, S., Campa, C., Cenci, A., Combes, M.-C., Crouzillat, D., Da Silva, C., Daddiego, L., De Bellis, F., Dussert, S., Garsmeur, O., Gayraud, T., Guignon, V., Jahn, K., Jamilloux, V., Joët, T., Labadie, K., Lan, T., Leclercq, J., Lepelletier, M., Leroy, T., Li, L.-T., Librado, P., Lopez, L., Muñoz, A., Noel, B., Pallavicini, A., Perrotta, G., Poncet, V., Pot, D., Rigoreau, Priyono M., Rouard, M., Rozas, J., Tranchant-Dubreuil, C., VanBuren, R., Zhang, Q., Andrade, A.C., Argout, X., Bertrand, B., de Kochko, A., Graziosi, G., Henry, R.J., Ming, Jayarama R., Nagai, C., Rounsley, S., Sankoff, D., Giuliano, G., Albert, V.A., Wincker, P., Lashermes, P., 2014. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* 345, 1181. <https://doi.org/10.1126/science.1255274>.
- Di, C., Yuan, J., Wu, Y., Li, J., Lin, H., Hu, L., Zhang, T., Qi, Y., Gerstein, M.B., Guo, Y., Lu, Z.J., 2014. Characterization of stress-responsive lncRNAs in *Arabidopsis thaliana* by integrating expression, epigenetic and structural features. *Plant J.* 80, 848–861. <https://doi.org/10.1111/tpj.12679>.
- Dinger, M.E., Amaral, P.P., Mercer, T.R., Mattick, J.S., 2009. Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications.

- Briefings in Functional Genomics 8, 407–423. <https://doi.org/10.1093/bfgp/elp038>.
- Eddy, S.R., 2001. Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genetics* 2, 919–929. <https://doi.org/10.1038/35103511>.
- Fu, J., Frazee, A.C., Collado-Torres, L., Jaffe, A.E., Leek, J.T., 2021. ballgown: Flexible, isoform-level differential expression analysis. R package version 2 (24).
- Ghildiyal, M., Zamore, P.D., 2009. Small silencing RNAs: an expanding universe. *Nat. Rev. Genetics* 10, 94–108. <https://doi.org/10.1038/nrg2504>.
- Kim, D., Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. <https://doi.org/10.1038/s41587-019-0201-4>.
- Kovaka, S., Zimin, A.V., Pertea, G.M., Razaghi, R., Salzberg, S.L., Pertea, M., 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 20, 278. <https://doi.org/10.1186/s13059-019-1910-1>.
- Lashermes, P., Combes, M.C., Robert, J., Trouslot, P., D'Hont, A., Anthony, F., Charrier, A., 1999. Molecular characterisation and origin of the *Coffea arabica* L. genome. *Mol. General Genetics MGG* 261, 259–266. <https://doi.org/10.1007/s004380050965>.
- Li, L., Eichten, S.R., Shimizu, R., Petsch, K., Yeh, C.-T., Wu, W., Chettoor, A.M., Givan, S. A., Cole, R.A., Fowler, J.E., Evans, M.M.S., Scanlon, M.J., Yu, J., Schnable, P.S., Timmermans, M.C.P., Springer, N.M., Muehlbauer, G.J., 2014. Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol.* 15, R40. <https://doi.org/10.1186/gb-2014-15-2-r40>.
- Liang, L., Ai, L., Qian, J., Fang, J.-Y., Xu, J., 2015. Long noncoding RNA expression profiles in gut tissues constitute molecular signatures that reflect the types of microbes. *Sci. Rep.* 5, 11763. <https://doi.org/10.1038/srep11763>.
- Montagnon, C., Mahyoub, A., Solano, W., Sheibani, F., 2021. Unveiling a unique genetic diversity of cultivated *Coffea arabica* L. in its main domestication center: Yemen. *Genet. Resour. Crop. Evol.* <https://doi.org/10.1007/s10722-021-01139-y>.
- Nathanson, J.A., 1984. Caffeine and related methylxanthines: possible naturally occurring pesticides. *Science* 226, 184. <https://doi.org/10.1126/science.6207592>.
- Pacheco, A., Pohlan, J., Schulz, M., 2008. Allelopathic effects of aromatic species intercropped with coffee: Investigation of their growth stimulation capacity and potential of caffeine uptake in Puebla, Mexico. *Allelopathy J.* 21, 39–57.
- Pauli, A., Valen, E., Lin, M.F., Garber, M., Vastenhouw, N.L., Levin, J.Z., Fan, L., Sandelin, A., Rinn, J.L., Regev, A., Schier, A.F., 2012. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* 22, 577–591. <https://doi.org/10.1101/gr.133009.111>.
- Pertea, G., Pertea, M., 2020. GFF Utilities: GffRead and GffCompare. *F1000Research* 9. <https://doi.org/10.12688/f1000research.23297.1>
- Ponting, C.P., Oliver, P.L., Reik, W., 2009. Evolution and Functions of Long Noncoding RNAs. *Cell* 136, 629–641. <https://doi.org/10.1016/j.cell.2009.02.006>.
- Rinn, J.L., Chang, H.Y., 2012. Genome Regulation by Long Noncoding RNAs. *Annu. Rev. Biochem.* 81, 145–166. <https://doi.org/10.1146/annurev-biochem-051410-092902>.
- Sun, Z., Huang, K., Han, Z., Wang, P., Fang, Y., 2020. Genome-wide identification of *Arabidopsis* long noncoding RNAs in response to the blue light. *Sci. Rep.* 10, 6229. <https://doi.org/10.1038/s41598-020-63187-1>.
- Valencia-Lozano, E., Ibarra, J.E., Herrera-Ubaldo, H., De Folter, S., Cabrera-Ponce, J.L., 2021. Osmotic stress-induced somatic embryo maturation of coffee *Coffea arabica* L., shoot and root apical meristems development and robustness. *Sci. Rep.* 11, 9661. <https://doi.org/10.1038/s41598-021-88834-z>.
- Wang, H., Chung, P.J., Liu, J., Jang, I.-C., Kean, M.J., Xu, J., Chua, N.-H., 2014. Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. *Genome Res.* 24, 444–453. <https://doi.org/10.1101/gr.165555.113>.
- Wang, J., Yu, W., Yang, Y., Li, X., Chen, T., Liu, T., Ma, N., Yang, X., Liu, R., Zhang, B., 2015a. Genome-wide analysis of tomato long non-coding RNAs and identification as endogenous target mimic for microRNA in response to TYLCV infection. *Sci. Rep.* 5, 16946. <https://doi.org/10.1038/srep16946>.
- Wang, Kevin C., Chang, Howard Y., 2011. Molecular Mechanisms of Long Noncoding RNAs. *Mol. Cell* 43, 904–914. <https://doi.org/10.1016/j.molcel.2011.08.018>.
- Wang, M., Yuan, D., Tu, L., Gao, W., He, Y., Hu, H., Wang, P., Liu, N., Lindsey, K., Zhang, X., 2015b. Long noncoding RNAs and their proposed functions in fibre development of cotton (*Gossypium* spp.). *New Phytol.* 207, 1181–1197. <https://doi.org/10.1111/nph.13429>.
- Wang, X., Ai, G., Zhang, C., Cui, L., Wang, J., Li, H., Zhang, J., Ye, Z., 2016. Expression and diversification analysis reveals transposable elements play important roles in the origin of Lycopersicon-specific lncRNAs in tomato. *New Phytol.* 209, 1442–1455. <https://doi.org/10.1111/nph.13718>.
- Wucher, V., Legeai, F., Hédan, B., Rizk, G., Lagoutte, L., Leeb, T., Jagannathan, V., Cadieu, E., David, A., Lohi, H., Cirera, S., Fredholm, M., Botherel, N., Leegwater, P. A.J., Le Béguec, C., Fieten, H., Johnson, J., Alfoldi, J., André, C., Lindblad-Toh, K., Hitte, C., Derrien, T., 2017. FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res.* 45. <https://doi.org/10.1093/nar/gkw1306> e57.
- Yang, G., Lu, X., Yuan, L., 2014. lncRNA: A link between RNA and cancer. *Biochim. Biophys. Acta* 1839, 1097–1109. <https://doi.org/10.1016/j.bbagr.2014.08.012>.
- Young, M.D., Wakefield, M.J., Smyth, G.K., Oshlack, A., 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11, R14. <https://doi.org/10.1186/gb-2010-11-2-r14>.
- Zhang, W., Han, Z., Guo, Q., Liu, Y., Zheng, Y., Wu, F., Jin, W., 2014a. Identification of Maize Long Non-Coding RNAs Responsive to Drought Stress. *PLoS One* 9. <https://doi.org/10.1371/journal.pone.0098958> e98958.
- Zhang, Y.-C., Liao, J.-Y., Li, Z.-Y., Yu, Y., Zhang, J.-P., Li, Q.-F., Qu, L.-H., Shu, W.-S., Chen, Y.-Q., 2014b. 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. <https://doi.org/10.1186/s13059-014-0512-1>.
- Zhao, X., Gan, L., Yan, C., Li, C., Sun, Q., Wang, J., Yuan, C., Zhang, H., Shan, S., Liu, J.N., 2019. Genome-Wide Identification and Characterization of Long Non-Coding RNAs in Peanut. *Genes* 10, 536. <https://doi.org/10.3390/genes10070536>.