PLANT SCIENCE TODAY
ISSN 2348-1900 (online)
Vol x(x): xx–xx
https://doi.org/10.14719/pst.3057

HORIZON
e-Publishing Group
HePG

RESEARCH ARTICLE

Elucidating role of long non-coding RNAs of *Tamarindus indica* Linn. in post-transcriptional gene regulation

Moumita Roy Chowdhury[1,2], Aman Kumar[3], Alfred Besra[3,4], Ranjit Prasad Bahadur[1] & Jolly Basak[3]*

[1]Department of Biotechnology, Indian Institute of Technology Kharagpur, Kharagpur 721 302, India

[2]Department of Biotechnology, Koneru Lakshmaiah Education Foundation, Guntur 522 302, India

[3]Department of Biotechnology, Visva-Bharati University, Santiniketan 731 235, India

[4]Department of Biotechnology, St. Xavier college, Pathalkudwa, Nayatoli, Ranchi, Jharkhand 834 001, India

*Email: jolly.basak@visva-bharati.ac.in

OPEN ACCESS

ARTICLE HISTORY
Received: 31 October 2023
Accepted: 29 September 2024
Available online
Version 1.0 : 19 January 2025

Check for updates

Additional information
Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

CITE THIS ARTICLE
Chowdhury MR, Kumar A, Besra A, Bahadur RP, Basak J. Elucidating role of long non-coding RNAs of *Tamarindus indica* Linn. in post-transcriptional gene regulation. Plant Science Today (Early Access). https:/doi.org/10.14719/pst.3057

Abstract

Tamarindus indica Linn., commonly known as tamarind, is a rich source of carbohydrates, proteins, lipids, fatty acids, vitamins, minerals and bioactive compounds. It is well established that long non-coding RNAs (lncRNAs) plays an important role in transcriptional, post-transcriptional and epigenetic regulation. Despite the availability of tamarind genome information, a handful of studies have been done on its non-coding genome, especially lncRNAs. In this study, we have computationally predicted lncRNAs from the coding DNA sequences of *T. indica* and analysed the sequences. We experimentally validated seven randomly chosen lncRNAs by performing quantitative Real Time Polymerase Chain Reaction (qRT-PCR). 320 lncRNAs have been predicted and sequence analysis of these predicted, lncRNAs reveals the presence of different motifs and tandem repeats. Along with the experimental validation of 7 randomly chosen lncRNAs, functional analysis of the predicted lncRNAs and their targets elucidated their roles in various biological pathways. We believe prediction and validation of the lncRNAs along with their interaction with mRNAs will enhance our knowledge about the non-coding genome of tamarind and their involvement in post transcriptional gene regulation, medicinal properties, metabolic engineering, stress tolerance and genome editing.

Keywords

Tamarindus indica Linn.; long non-coding RNAs (lncRNAs); Coding DNA sequence (CDS); Open Reading Frame (ORF); quantitative Real Time Polymerase Chain Reaction (qRT-PCR).

Introduction

Non-coding RNAs are found to be involved in various regulatory processes along with the maintenance and segregation of chromosomes (1). Long non-coding RNAs, transcribed by RNA polymerase I, II, or III, are > 200 nucleotides (nt) long transcripts with no coding potentiality (Open Reading Frame (ORF) < 100 amino acids) (2). Previously, lncRNAs were considered as transcriptional noises, but later it was found that lncRNAs are key players in numerous gene regulation processes, including transcriptional, post-transcriptional and epigenetic regulation (3). In plants, lncRNAs are involved in growth, reproduction, development, cell differentiation, chromatin modification, protein re-localization, phosphate homeostasis and responses to biotic and abiotic stresses (4-6). Plant lncRNAs can be classified based on their genomic location and functions (7). lincRNAs, transcribed from the intergenic regions, regulate gene expression epigenetically

and control protein levels (8). Intronic lncRNAs, transcribed from intronic regions, are found to be involved in epigenetic repression (9). Sense and anti-sense lncRNAs are transcribed from the sense and anti-sense strand of protein coding genes and anti-sense lncRNAs are more widespread in plants and well -studied than sense-lncRNAs (10, 11). *Cis*-acting lncRNAs can regulate genes transcribed from the same chromosomal locus (12). On the other hand, *trans*-acting lncRNAs can relieve the suppression of mRNAs by binding to the miRNA and by acting as a target-mimic for the specific miRNA (13).

Tamarind (*Tamarindus indica* Linn.) is a very important cultivar of the Fabaceae family and is widely cultivated in Africa and parts of southern Asia (14). Tamarind is a rich source of carbohydrates, protein, lipids and other nutrients like vitamins (thiamine, riboflavin, niacin and Vitamin C) and minerals (Ca, Mg, Fe, K, Cd, Mn, p, Cr and Zn) (15, 16). Tamarind seeds are also a rich source of different fatty acids like linoleic, oleic and palmitic acids (17). Along with that, various bioactive compounds like alkaloids, flavonoids, steroids, tannins, saponins, phenols and glycosides are present in the fruit pulp (18, 19). The presence of these bioactive components is the probable reason for the anti-inflammatory and analgesic properties of tamarind, which is used for the treatment of inflammatory diseases like arthritis and body pain (18, 20). It has been found that tamarind has an anti-hyperglycaemic activity, which may be attributed to glycosides, saponins and anthraquinones present in the fruit pulp (19). Methanol and acetone extracts of tamarind have shown potent antibacterial activity against *Klebsiella pneumoniae*, while concentrated extracts showed significant antibacterial activity against *Bacillus subtilis, Salmonella typhi*, *S. paratyphi and Staphylococcus aureus* (21). Aqueous-ethanol extract of tamarind has shown protective effects against nephrotoxic effects of gentamicin (22). It has been seen in a recent study that the methanol extract of tamarind seed coat contains phenolic compounds which are responsible for the antioxidant activities of tamarind (23).

Recent studies have identified lncRNAs from medicinal plants like *Salvia miltiorrhiza*, *Panax ginseng* and *Digitalis purpurea* (24). mRNA like non-coding RNAs possessing similar characteristics with lncRNAs identified in *S. miltiorrhiza* were found to elicit a positive response to treatments with Methyl Jasmonate, yeast extract and Ag+ (25). The tissue-specific expression of lncRNAs in *S. miltiorrhiza*, *P. ginseng* and *D. purpurea* underscores the significance of these non-coding RNAs in the growth and development of these medicinal plants. However, only a handful number of studies have been done to identify lncRNAs from medicinal plants. Although a large number of plants lncRNAs are identified and characterized through the blending of genome-wide computational and experimental approaches (6, 7, 10, 26-29), no lncRNAs have been reported in tamarind so far, which prompts us to perform this study. As lncRNAs might play a major role in the gene regulation of tamarind, as well as in its medicinal properties, gaining insight by initially identifying the lncRNAs would be valuable.

In this study, we have computationally predicted 320 lncRNAs by considering coding potentiality and ORF length as significant parameters. Along with that, we analysed the lncRNA sequences in search of the presence of functional motifs and tandem repeats. Five diverse motifs have been found and 5 lncRNAs were found, to have simple repeats, whereas the other 2 lncRNAs have low complexity regions and LINE/CR1 repeats. Additionally, interaction between lncRNAs and mRNAs has been studied, revealing that eight lncRNAs target 5 protein-coding genes. Finally, we experimentally validated randomly chosen seven lncRNAs in this study through quantitative Real Time Polymerase Chain Reaction (qRT-PCR).

Materials and Methods

Computational prediction of lncRNAs and their characterization

Coding DNA sequences (CDSs) of *T. indica* having > 200 nt length were first selected. EMBOSS get orf was used to get the ORFs with lengths less than 100 amino acids (30). Coding Potential Calculator 2.0 (CPC2) (31) and Coding Potential Assessment Tool (CPAT) (32) were used to check the coding potentiality of these sequences. BLASTX was performed with the SWISS-PROT database as the subject and the remaining non-coding sequences as a query and with 0.001 e-value cut-off. Sequences with 40 % resemblance or above were removed and the rest of the non-coding sequences were considered as lncRNAs. The flowchart of predicting lncRNAs is shown in Fig. 1. The presence of the motifs in the lncRNAs was identified by using DREME (33), while Tandem Repeats Finder was used to identify the presence of the tandem repeats (TR) (34).

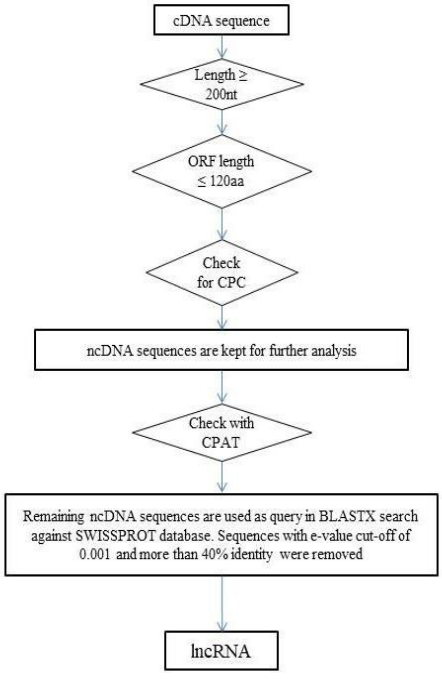

Fig. 1. Schematic diagram of the prediction method of lncRNAs. cDNA sequences of *B. oleracea* were downloaded from NCBI. lncRNAs are predicted after filtering out all the criteria discussed in the materials and methods section.

Prediction of lncRNA targets on mRNAs and functional annotation of the target

To predict the targets of lncRNAs on mRNAs, LncTar (35) has been used. LncTar is based on the calculation of the approximate binding free energy, delta G (dG), of each pairing which indicates the stability of complementarity between

lncRNA and mRNA. This binding free energy calculation is based on the default parameter of LncTar software. Further, the predicted mRNA targets were analyzed for their functional annotation.

Plant materials for validation of the lncRNAs

Tamarind seeds (Obtained from Bolpur, West Bengal) were surface sterilized by using 0.1% $HgCl_2$ for 2-5 min and then washed with distilled water. Seeds were sowed in soil-rite and grew in the dark. Seedlings were grown for 10 days in a plant growth chamber (Labtech model no. LGC-5101/5201/5301) with a photoperiod of 16 h day and 8 h night with 70-80% relative humidity at 24 °C.

Total RNA Extraction

Total RNA was extracted by using a plant RNA isolation kit (Macherey-Nagel nucleospin) from the leaves of 10 days old tamarind seedlings according to the manufacturer's instructions. Briefly, 100 mg plant leaves were ground and lysed. The lysate was cleared by filtration (NucleoSpin® Filter) and the homogenized lysate was mixed with 350 µL of 70 % ethanol. For the binding of RNA, NucleoSpin® RNA Plant Column is used, followed by desalting of the membrane. This salt removal step is used to make the following rDNase digest more effective. After the removal of DNA, the membrane is dried and washed, followed by the elution of RNA. The total extracted RNA was quantified by spectrophotometer (model: Eppendorf) and stored at -20 °C for further use.

cDNA synthesis and primer design

First-strand cDNA synthesis kit (Thermo) was used to transcribe cDNA from the isolated RNA following the manufacturer's instructions. The extracted total RNA is mixed with the Oligo (dT)18 primer and nuclease free water and mixed. After that, 5X Reaction Buffer, RNase Inhibitor, dNTP Mix and reverse transcriptase were added to the mixture and centrifuged briefly and the mixture was first incubated at 42 °C for 1 h. The reaction was terminated by heating at 70 °C for 5 min. Primers (forward and reverse) were designed using Primer3 (https://primer3.org/) software and are listed in Table 1.

Quantitative Real Time Polymerase Chain Reaction (qRT-PCR)

For qRT-PCR, a Bio-Rad CFX96 RT-PCR system along with SYBR green supermix (Bio-Rad iQ) was used. Each reaction mixture (10 µL) contained cDNA, SYBR green supermix and 300 nM of forward and reverse primers. The PCR cycle protocol was followed as mentioned (36-38). Amplicon specificity was verified by analyzing the melting curves. All reactions were executed in triplicates and verified by the presence of single peaks of the melt curves. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as an endogenous control.

Statistical analysis

All the data have been tested separately for statistical analysis. Data is denoted as mean ± standard error of means (SEM). A student's t-test was performed to identify significant changes by considering $p < 0.05$ as a threshold by using Microsoft Excel (Office 2019; Version 16). All the tests are carried out on biological triplicates independently.

Table 1. Primers used for RT-qPCR assay

Primers	Sequence (5'-3')
GAPDH F	ATGGCATTCCGTGTTCCTAC
GAPDH R	CCTTCAACTTGCCCTCTGAC
tin-lnc-43088 F	TGGGACGATGGGTGAGTATG
tin-lnc-43088 R	CAGATCAGTTGCCCGAATCC
tin-lnc-46734 F	GGCGAAATCCTTCAGTGGTAC
tin-lnc-46734 R	CCCAGACCGGCTTACTAATG
tin-lnc-57967 F	GGCGAAATCCTTCAGTGGTAC
tin-lnc-57967 R	CCCAGACCGGCTTACTAATG
tin-lnc-587877 F	GGCGAAATCCTTCAGTGGTAC
tin-lnc-587877 R	CCCAGACCGGCTTACTAATG
tin-lnc-498934 F	GAACGGCCTTGACTACCTCT
tin-lnc-498934 R	ATGTCGCATTTTGGGCTTGT
tin-lnc-045038 F	CCAAGGGGTCTTTCTGTGGA
tin-lnc-045038 R	TTCCCTCATTCCACTCGTCC
tin-lnc-737446 F	GGCGAAATCCTTCAGTGGTAC
tin-lnc-737446 R	CCCAGACCGGCTTACTAATG
tin-lnc-026685 F	GAGCCCTTATGTGTCTTGAACT
tin-lnc-026685 R	TGAAACAATAGCTGGTCCGATAG

Results

Computational prediction of lncRNAs and their characterization

In this study, we have predicted 320 lncRNAs (supplementary Table S1). Most of the lncRNAs are greater than 400 nt in length (17%). The length of the remaining lncRNAs ranges between 200-400 nt (83 %). The distribution of lncRNA length is shown in Fig. 2. We have found 5 diverse motifs in the computationally predicted lncRNAs that are involved in different biological processes, molecular functions and being part of cellular components. Detailed description of the motifs and their functional significances are described in Table 2. In this study, we have predicted TRs and we found that 7 lncRNAs have simple repeats, LINE/CR1 repeats as well as low complexity regions (Table 3).

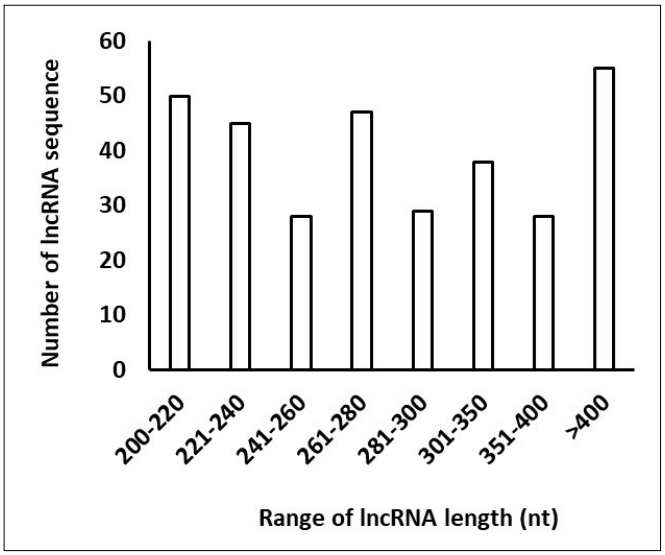

Fig. 2. Distribution of lncRNA based on their length.

Table 2. Motifs and their functional significance

Motif	Functional significance
CTTCTW	CC chloroplast thylakoid membrane
	BP transmembrane receptor protein tyrosine kinase signaling pathway
	BP protein amino acid phosphorylation
	MF protein serine/threonine kinase activity
	MF ATP binding
	CC nucleus
	CC plasma membrane
AAAAGRAA	BP regulation of transcription, DNA-dependent
	MF transcription factor activity
	MF kinase activity
CTTKTTCA	MF kinase activity
ACGTWGG	Nil
AAAACSAT	Nil

Table 3. Repetitive elements in the lncRNAs

Repeat class/family	lncRNAs
Simple repeats	tin-lnc-0000023c, tin-lnc-0000026a, tin-lnc-0000036d
Low complexity regions	tin-lnc-026685f
LINE/CR1	tin-lnc-0000029f

Prediction of lncRNA targets on mRNAs and functional annotation of the targets

Among the lncRNAs, 8 lncRNAs are found to have targets on 5 protein coding genes (Table 4). Seven lncRNAs namely, tin-lnc-026685c, tin-lnc-026685l, tin-lnc-0000037, tin-lnc-0000034, tin-lnc-0000024e, tin-lnc-000009f and tin-lnc-000009i are found to target a single gene UDP-sulfoquinovose synthase (SQD1) protein coding gene while tin-lnc-498934 targets 4 different FLO/LFY-like protein coding genes (A4, A5, A6 and A7). UDP-sulfoquinovose synthase is involved in nucleotide sugars metabolism and glycerolipid metabolism. Floricaula/leafy-like transcription factor is involved in the regulation of transcription and DNA binding.

Experimental validation of the lncRNAs

In this study, we have randomly selected seven lncRNAs for experimental validation. Expression profiles of the 7 lncRNAs from qRT-PCR analysis are shown in Fig. 3. Relative quantity of lncRNA expression is gained by $2^{-\Delta C_T}$, where ΔC_T is depicted as C_T lncRNA- C_T control (GAPDH). Amongst the seven lncRNAs, the C_T value of tin-lnc-587877 is the lowest (17.385), followed by tin-lnc-737446 (C_T = 18.8), tin-lnc-57967 (C_T= 19.26) and tin-lnc-498934 (C_T = 19.49). For the remaining three lncRNAs, C_T values of 2 lncRNAs namely tin-lnc-045038 (C_T = 23.6) and tin-lnc-026685 (C_T = 25.4) are in the medium range, while tin-lnc-43088 (C_T= 30.6) has the highest C_T value.

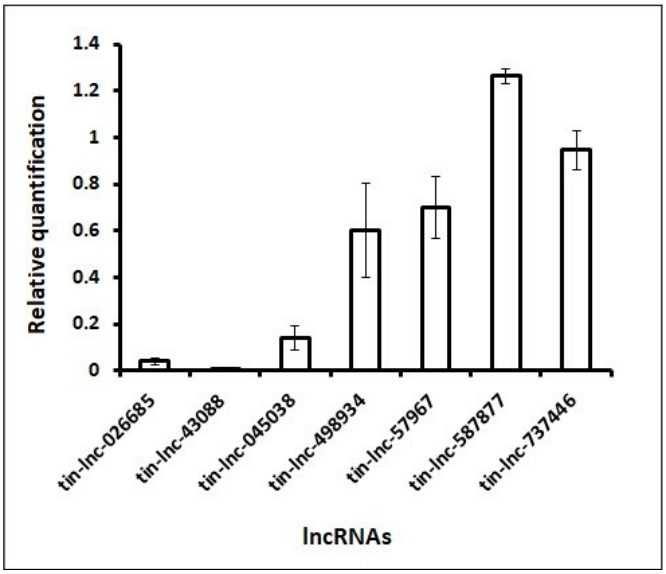

Fig. 3. Expression profile of 10 lncRNAs from qRT-PCR analysis. Relative quantification was measured by $2^{-\Delta C_T}$ where ΔC_T = (C_T lncRNA- C_T GAPDH).

Discussion

The presence of very less information on tamarind lncRNAs has prompted us to perform this study to gather more information on tamarind lncRNAs. Being a rich source of various secondary metabolites including polyphenols and flavonoids, tamarind possesses antioxidant, antimicrobial, antidiabetic, anti-inflammatory, antihyperlipidemic and anticancer activities (38). Although there is information available about the coding genome of tamarind, very less information is available on the non-coding genome, especially on lncRNAs. In recent years, it has been found that plant lncRNAs play important roles in various biological processes like growth, development, cell differentiation, protein-re-localization, response to different biotic and abiotic stresses, phosphate homeostasis and chromatin modification (4-6). A handful of studies have been conducted to predict lncRNAs from medicinal plants like

Table 4. lncRNA targets on mRNAs and functional annotation of the targets

lncRNA	Target	Biological processes	Cellular components	Molecular function
tin-lnc-498934	A5 FLO/LFY-like protein gene	Regulation of transcription, DNA-templated	Nucleus	DNA binding
tin-lnc-498934	A6 FLO/LFY-like protein gene			
tin-lnc-498934	A4 FLO/LFY-like protein gene			
tin-lnc-498934	A7 FLO/LFY-like protein gene			
tin-lnc-026685c	UDP-sulfoquinovose synthase (SQD1) gene	Nucleotide sugars metabolism and glycerolipid metabolism		
tin-lnc-026685l	UDP-sulfoquinovose synthase (SQD1) gene			
tin-lnc-0000037	UDP-sulfoquinovose synthase (SQD1) gene			
tin-lnc-0000034	UDP-sulfoquinovose synthase (SQD1) gene			
tin-lnc-0000024e	UDP-sulfoquinovose synthase (SQD1) gene			
tin-lnc-000009f	UDP-sulfoquinovose synthase (SQD1) gene			
tin-lnc-000009i	UDP-sulfoquinovose synthase (SQD1) gene			

S. miltiorrhiza, *P. ginseng* and *D. purpurea* (24); however, no lncRNA has been predicted in tamarind. Hence, through this study, we have computationally predicted new lncRNAs and tried to find out their functional significance. Besides predicting novel lncRNAs of tamarind, we also predicted their interaction with mRNAs and the involvement of these mRNAs in various biological processes. Through this study, we have seen single lncRNA targeting multiple protein coding genes as well as multiple lncRNAs targeting single protein coding genes. We found SQD1, an enzyme involved in the biosynthesis of sulfoquinovosyl headgroup of plant sulfolipids and catalyzing the transfer of SO3- to UDP-glucose (39), is targeted by 7 different lncRNAs. On the other hand, tin-lnc-498934 targets 4 different FLO/LFY-like protein coding genes (A4, A5, A6 and A7), which regulate first cell division after zygote formation (40). Among the five motifs present in these lncRNAs, functional annotation is carried out for 3 motifs in this study. Motif "CTTCTW" is involved in biological processes like trans-membrane receptor protein tyrosine kinase signaling pathway and protein amino acid phosphorylation. It also takes part in molecular functions like protein serine/threonine kinase activity and ATP binding. Motif "AAAAGRAA" is part of the biological process like regulation of transcription along with transcription factor activity like molecular functions. Motif "CTTKTTCA" is involved in kinase activity. For the remaining 2 motifs, functional annotation did not predict any function, neither biological process nor molecular function. We have also predicted TRs that are important driving factors in the emergence of lncRNAs (41). Three lncRNAs namely, tin-lnc-0000023c, tin-lnc-0000026a and tin-lnc-0000036d consist of simple repeats "ATT", "GAAT", "CTTTTT" respectively. Two lncRNAs namely tin-lnc-026685f and tin-lnc-0000029f consist of low complexity regions and LINE/CR1 repeats respectively. Randomly chosen 7 lncRNAs were experimentally validated. tin-lnc-587877 showed the highest expression. Other highly expressed lncRNAs are tin-lnc-737446, tin-lnc-57967 and tin-lnc-498934. Two moderately expressed lncRNAs are tin-lnc-045038 and tin-lnc-026685, whereas tin-lnc-43088 showed the lowest expression.

Conclusion

In summary, we have identified 320 lncRNAs from the coding DNA sequence of *Tamarindus indica* Linn. and experimentally validated randomly chosen 7 lncRNAs. The possible roles were investigated for lncRNAs, including the prediction of target genes, TRs and functional motifs. We predicted five mRNA targets of four lncRNAs. We found 7 lncRNAs containing simple repeats, LINE/CR1 repeats and low complexity regions. We found 5motifs, among which 3 motifs are involved in various biological processes and molecular functions like enzyme activity, DNA binding properties, signalling activity and transcriptional regulation. However, we acknowledge our limitations since we were unable to predict any direct link between the predicted lncRNAs and the medicinal properties of tamarind. However, this complete study is the first to provide more insights into the existing knowledge of the tamarind non-coding genome, which is still in its infancy. We strongly believe that predicting the lncRNAs and analysing their role in various biological processes will shed light on many unknown functions related to the medicinal and therapeutical properties of tamarind.

Acknowledgements

MRC acknowledges IIT Kharagpur for the research fellowship and Visva-Bharati university for providing the research facilities for wet-lab experiments. JB acknowledges CSIR (38/1446/17/EMR-II) for their support.

Authors' contributions

MRC, AK and AB performed the study. MRC wrote the manuscript. JB and RPB conceived of the study and participated in its design and coordination. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interests to declare.

Ethical issues: None

References

1. Mattick JS, Makunin IV. Non-coding RNA. Hum Mol Genet. 2006;15:17-29. https://doi.org/10.1093/hmg/ddl046
2. Wang HLV, Chekanova JA. Long noncoding RNAs in plants. Adv Exp Med Biol. 2017;1008:133-54. https://doi.org/10.1007/978-981-10-5203-3_5
3. Chen X, Sun Y, Cai R, Wang G, Shu X, Pang W. Long noncoding RNA: multiple players in gene expression. BMB Rep. 2018;51(6):280-89. https://doi.org/10.5483/BMBRep.2018.51.6.025
4. Yu T, Zhu H. Long non-coding RNAs: Rising regulators of plant reproductive development. Agronomy. 2019;9(2):53. https://doi.org/10.3390/agronomy9020053
5. Nejat N, Mantri N. Emerging roles of long non-coding RNAs in plant response to biotic and abiotic stresses. Crit Rev Biotechnol. 2018;38(1):93-105. https://doi.org/10.1080/07388551.2017.1312270
6. Nithin C, Thomas A, Basak J, Bahadur RP. Genome-wide identification of miRNAs and lncRNAs in *Cajanus cajan*. BMC Genomics. 2017;18(1):878. https://doi.org/10.1186/s12864-017-4232-2
7. Rai MI, Alam M, Lightfoot DA, Gurha P, Afzal AJ. Classification and experimental identification of plant long non-coding RNAs. Genomics. 2019;111(5):997-1005. https://doi.org/10.1016/j.ygeno.2018.04.014
8. Yamada M. Functions of long intergenic non-coding (linc) RNAs in plants. J Plant Res. 2017;130(1):67-73. https://doi.org/10.1007/s10265-016-0894-0.
9. Heo JB, Sung S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. Science. 2011;331(6013):76-79. https://doi.org/10.1126/science.1197349.
10. Wang H, Chung PJ, Liu J, Jang IC, Kean MJ, Xu J, et al. Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. Genome Res. 2014;24(3):444-53. https://doi.org/10.1101/gr.165555.113.
11. Ietswaart R, Wu Z, Dean C. Flowering time control: another window to the connection between antisense RNA and chromatin. Trends Genet TIG. 2012;28(9):445-53. https://doi.org/10.1016/j.tig.2012.06.002.
12. Fedak H, Palusinska M, Krzyczmonik K, Brzezniak L, Yatusevich R, Pietras Z, et al. Control of seed dormancy in *Arabidopsis* by a cis-acting noncoding antisense transcript. Proc Natl Acad Sci. 2016;113(48):E7846-55. https://doi.org/10.1073/pnas.1608827113.

13. Wu HJ, Wang ZM, Wang M, Wang XJ. Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants. Plant Physiol. 2013;161(4):1875-84. https://doi.org/10.1104/pp.113.215962.

14. Doughari JH. Antimicrobial activity of *Tamarindus indica* Linn. Trop J Pharm Res. 2006;5(2):597-603. https://doi.org/10.4314/tjpr.v5i2.14637.

15. A CF, G N, A-H A. A review of explored uses and study of nutritional potential of tamarind (*Tamarindus indica* L.) in Northern Ghana. Afr J Food Sci. 2020;14(9):285-94. https://doi.org/10.5897/AJFS2018.1744.

16. Caluwé ED, Halamová K, Damme PV. *Tamarindus indica* L. - A review of traditional uses, phytochemistry and pharmacology. Afr Focus. 2010;23(1). https://doi.org/10.21825/af.v23i1.5039.

17. Kumar CS, Bhattacharya S. Tamarind seed: Properties, processing and utilization. Crit Rev Food Sci Nutr. 2008;48(1):1-20. https://doi.org/10.1080/10408390600948600.

18. Komakech R, Kim YG, Matsabisa GM, Kang Y. Anti-inflammatory and analgesic potential of *Tamarindus indica* Linn. (Fabaceae): a narrative review. Integr Med Res. 2019;8(3):181-86. doi:10.1016/j.imr.2019.07.002.

19. Krishna RN, Anitha R, Ezhilarasan D. Aqueous extract of *Tamarindus indica* fruit pulp exhibits anti-hyperglycaemic activity. Avicenna J Phytomedicine. 2020;10(5):440-47. PMCID: PMC7508317.

20. Bhadoriya SS, Mishra V, Raut S, Ganeshpurkar A, Jain SK. Anti-Inflammatory and antinociceptive activities of a hydroethanolic extract of *Tamarindus indica* leaves. Sci Pharm. 2012;80(3):685-700. https://doi.org/10.3797/scipharm.1110-09.

21. Bhadoriya SS, Ganeshpurkar A, Narwaria J, Rai G, Jain AP. *Tamarindus indica*: Extent of explored potential. Pharmacogn Rev. 2011;5(9):73-81. https://doi.org/10.4103/0973-7847.79102.

22. Ullah N, Azam Khan M, Khan T, Ahmad W. Protective potential of *Tamarindus indica* against gentamicin-induced nephrotoxicity. Pharm Biol. 2014; https://doi.org/10.3109/13880209.2013.840318.

23. Sandesh P, Velu V, Singh RP. Antioxidant activities of tamarind (*Tamarindus indica*) seed coat extracts using *in vitro* and *in vivo* models. J Food Sci Technol. 2014;51(9):1965-73. https://doi.org/10.1007/s13197-013-1210-9.

24. Li C, Wang M, Qiu X, Zhou H, Lu S. Noncoding RNAs in medicinal plants and their regulatory roles in bioactive compound production. Curr Pharm Biotechnol. 2020; 22(3):341-59. https://doi.org/10.2174/1389201021666200529101942.

25. Li D, Shao F, Lu S. Identification and characterization of mRNA-like noncoding RNAs in *Salvia miltiorrhiza*. Planta. 2015;241(5):1131-43. https://doi.org/10.1007/s00425-015-2246-z.

26. Zhao X, Li J, Lian B, Gu H, Li Y, Qi Y. Global identification of *Arabidopsis* lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. Nat Commun. 2018;9(1):1-12. https://doi.org/10.1038/s41467-018-07500-7.

27. Boerner S, McGinnis KM. Computational identification and functional predictions of long noncoding RNA in *Zea mays*. PLOS ONE. 2012;7(8):e43047. https://doi.org/10.1371/journal.pone.0043047.

28. Lv Y, Liang Z, Ge M, Qi W, Zhang T, Lin F, et al. Genome-wide identification and functional prediction of nitrogen-responsive intergenic and intronic long non-coding RNAs in maize (*Zea mays* L.). BMC Genomics. 2016;17(1):350. https://doi.org/10.1186/s12864-016-2650-1.

29. Golicz AA, Singh MB, Bhalla PL. The long intergenic noncoding RNA (LincRNA) landscape of the soybean genome. Plant Physiol. 2018;176(3):2133-47. https://doi.org/10.1104/pp.17.01657.

30. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. Trends Genet TIG. 2000;16(6):276-77. https://doi.org/10.1016/s0168-9525(00)02024-2.

31. Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. Nucleic Acids Res. 2017;45(W1):W12-6. https://doi.org/10.1093/nar/gkx428.

32. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-potential assessment tool using an alignment-free logistic regression model. Nucleic Acids Res. 2013;41(6):e74-e74. https://doi.org/10.1093/nar/gkt006.

33. Bailey TL. DREME: motif discovery in transcription factor ChIP-seq data. Bioinforma Oxf Engl. 2011 ;27(12):1653-59. https://doi.org/10.1093/bioinformatics/btr261.

34. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573-80. https://doi.org/10.1093/nar/27.2.573.

35. Li J, Ma W, Zeng P, Wang J, Geng B, Yang J, et al. LncTar: a tool for predicting the RNA targets of long noncoding RNAs. Brief Bioinform. 2015;16(5):806-12. https://doi.org/10.1093/bib/bbu048.

36. Chowdhury MR, Chatterjee C, Ghosh D, Mukherjee J, Shaw S, Basak J. Deciphering miRNA-lncRNA-mRNA interaction through experimental validation of miRNAs, lncRNAs and miRNA targets on mRNAs in *Cajanus cajan*. Plant Biol. 2024;26(4):560-67. https://doi.org/10.1111/plb.13639.

37. Roy Chowdhury M, Bahadur RP, Basak J. Genome-wide prediction of cauliflower miRNAs and lncRNAs and their roles in post-transcriptional gene regulation. Planta. 2021;254(4):72. https://doi.org/10.1007/s00425-021-03689-y.

38. Devi B, Boruah T. Tamarind (*Tamarindus indica*). In: Nayik GA, Gull A, editors. Antioxidants in Fruits: Properties and Health Benefits. Singapore: Springer; 2020. p. 317-32. https://doi.org/10.1007/978-981-15-7285-2_16.

39. Mulichak AM, Theisen MJ, Essigmann B, Benning C, Garavito RM. Crystal structure of SQD1, an enzyme involved in the biosynthesis of the plant sulfolipid headgroup donor UDP-sulfoquinovose. Proc Natl Acad Sci USA. 1999;96(23):13097-102. doi:10.1073/pnas.96.23.13097.

40. Tanahashi T, Sumikawa N, Kato M, Hasebe M. Diversification of gene function: homologs of the floral regulator FLO/LFY control the first zygotic cell division in the moss *Physcomitrella patens*. Development. 2005;132(7):1727-36. https://doi.org/10.1242/dev.01709.

41. Nelson ADL, Forsythe ES, Devisetty UK, Clausen DS, Haug-Batzell AK, Meldrum AMR, et al. A genomic analysis of factors driving lincRNA diversification: Lessons from plants. G3 Genes Genomes Genetics. 2016;6(9):2881-91. https://doi.org/10.1534/g3.116.030338.