International Journal of Agriculture Sciences
ISSN: 0975-3710&E-ISSN: 0975-9107, Volume 10, Issue 4, 2018, pp.-5192-5195.
Available online at http://www.bioinfopublication.org/jouarchive.php?opt=&jouid=BPJ0000217

Research Article

ANALYSIS OF RICE GENOME LONG NON-CODING RNA SEQUENCES

SARANYA N.*, KANDAVELMANI A., SARAVANAN L.T., SHANMUGAPRIYA S., JAYAKANTHAN M., BHARATHI N. AND RAMALINGAM J.

Department of Plant Molecular Biology and Bioinformatics, Centre for Plant Molecular Biology and Biotechnology, Tamil Nadu Agricultural University, Coimbatore, 641003, Tamil Nadu
*Corresponding Author: Email-saranya.n@tnau.ac.in

Received: February 13, 2018; Revised: February 22, 2018; Accepted: February 23, 2018; Published: February 28, 2018

Abstract- LncRNA reserves meaningful biological information that has to be explored in order to understand the various regulatory mechanisms during stress conditions. Annotating the rice genomic lncRNA sequences based on the presence of motifs and repeats have been carried out in the present study. RNA C to U editing site, Polyadenylation sites, Rice splicing sites and untranslational region motifs are predominantly found in rice lncRNA sequences which shows their role in transcriptional and translational control of gene expression. Moreover, repeat analysis highlighted the presence of transposable elements in the lncRNAs. Studies involving lncRNA modifications with relevance to various biotic and abiotic stress conditions may provide various clues to reprogram transcriptional and translational events towards crop improvement and defence strategies.

Keywords – Rice long non-coding RNA, Computational analysis, Motifs, Repeats

Citation: Saranya N., *et al*., (2018) Analysis of Rice Genome Long Non-coding RNA sequences. International Journal of Agriculture Sciences, ISSN: 0975-3710 & E-ISSN: 0975-9107, Volume 10, Issue 4, pp.-5192-5195.

Introduction

Long non-coding RNAs (lncRNAs) are mRNA-like transcripts longer than 200 nucleotides that are not translated into protein. LncRNAs are mainly transcribed by RNA polymerase II (Pol II) and are polyadenylated, spliced, and mostly localized in the nucleus [1]. Unlike human and animal species, genome-wide identification of lncRNAs in plants is still in the infant stage [2]. Various studies have been carried out to understand the biological role of lncRNAs in plants [3]. LncRNAs play important roles in various biological processes including developmental regulations, plant reproductive developments [4], response to pathogen invasion [5, 6] and in stress responses [7]. LncRNAs have been found to regulate the expression levels of target genes via various molecular mechanisms [8, 9]. A number of lncRNAs have been identified in various plant genomes including Arabidopsis, wheat, maize and rice. However, functional mechanisms of lncRNAs in many plant species are not yet fully understood with only a few lncRNAs having been fully characterized.

In Arabidopsis, lncRNAs such as COLDAIR (cold-assisted intronic non-coding RNA) and COOLAIR (cold induced long antisense intragenic RNA) have been demonstrated to mediate chromatin modifying activities in transcriptional silencing of FLC during veRNAlization [10, 11]. Another antisense lncRNA, ASL, a non-polyadenylated transcript, has been discovered and is implicated in epigenetic silencing of FLC [12]. LincRNAs such as AT4 and IPS1 (INDUCED BY PHOSPHATE STARVATION1) act as decoys of miRNAs by a target mimicry mechanism. It has been suggested that AlteRNAtive Splicing Competitor long non-coding RNA (ASCO-lncRNA) also acts as a decoy, regulating gene expression in Arabidopsis during development [13]. 1113 long intergenic nocoding RNAs (lincRNAs) have been identified in potato (*Solanum tuberosum*) from stem tissues. Many of these lincRNAs have been suggested to have potential functional roles in potato defence responses [14]. Rice is one of the top leading food crops as well as an important model organism.

Systematic analysis of rice genome has identified 2224 lncRNAs which in turn are found to be highly tissue-specific [15]. In rice genome much of these lncRNAs have been observed to be involved in the regulation of reproductive development [16]. Although a large number of lncRNAs have been identified in plants, biological functions of many lncRNAs still remains unraveled. Biotechnological improvement of crops necessitates a deeper understanding of the structure and functions of lncRNAs [17]. In the present study, lncRNAs from rice genome have been collected from three important databases namely, GreeNC [18], PNRD [19] and plnLNCRBase [20], and functionally annotated based on bioinformatics approach.

Materials and Methods

In this study, 6064 rice genome lncRNA sequences mainly from three databases GreeNC (http://GreeNC.sciencedesigners.com/wiki/Main_Page), Plant Non-Coding RNA Database (http://structuralbiology.cau.edu.cn/PNRD/) and PLNlncR Base (http://bioinformatics.ahau.edu.cn/PLNlncRBase) databases were analyzed for the presence of motifs and repeats [Fig-1]. Experimentally identified sequences were retained from PLNlncRBase and the same sequences in other databases were removed. RegRNA 2.0 server was used to predict 14 types of motifs such as A-U rich elements, CIS-regulatory elements, Functional RNA sequences (fRNA), long stems, micro RNA target sites, nc hybrdization, ORF, Polyadenylation sites (PAS), Rho independent terminator, Ribosome binding site, Rice splicing sites (RSS), RNA C to U editing site, Transcriptional regulatory motif and untranslational region(UTR) motifs. Repeat masker is a computational tool used for identifying, classifying and masking repeat elements. Using repeat masker the repetitive sequences along with their type, class/family, position and length of these sequences were identified. Statistical analysis of the motifs and repeats were carried out using Microsoft excel software.

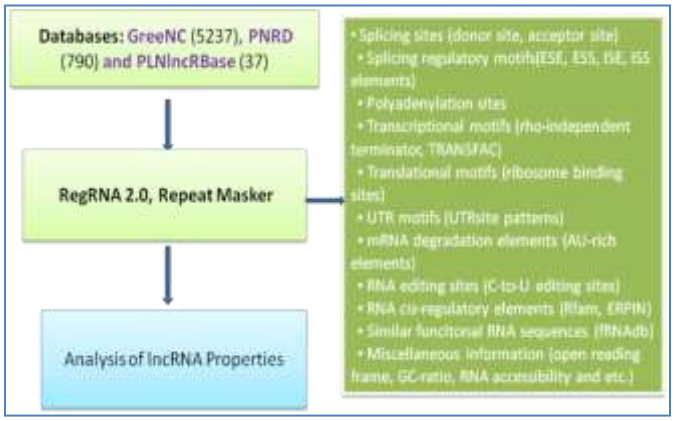

Fig-1 Flowchart describes the methodology involved in the analysis of rice LncRNA genome sequences.

Results and Discussion

LncRNA sequences from GreeNC database was least annotated when compared to other two databases [Fig-2]. Percentage of lncRNA sequences annotated in GreeNC, PNRD and PLNlncRBase were observed to be 33%, 40% and 92% respectively.

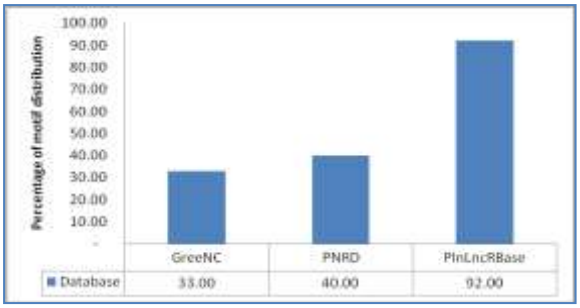

Fig-2 Percentage of lncRNA sequences annotated in GreeNC, PNRD and PlnLncRBase databases.

Among the 5237 number of sequences obtained from the GreeNC database, 6 types of motifs has shown its presence in ≥5% of the sequences which includes miRNA target sites (5%), polyadenylation sites (5%), RNA C to U editing sites (8%), ORF (14%), Rice splicing sites (17%) and UTR (18%) (Number in parenthesis indicates the percentage of sequences occupied by the particular motif). Other motifs were observed in less than 5% or present in very few numbers. In experimentally identified lncRNAs from PLNlncRBase database most of the sequences were annotated. Among the 37 sequences, 81% of the sequences contained UTR motifs. RNA C to U editing site and rice splicing site motifs were observed in 57% and 68% of the sequences respectively. Both the ORF and polyadenylation site motif were observed in 30% of the sequences. No transcription region motif was predicted in experimentally identified lncRNA sequences. From the 790 lncRNA sequences retrieved from the PNRD database, motifs were predicted successfully for 318 sequences, which accounts for 40% approximately. Details of motifs observed in these sequences were given in percentage as following: Rice splicing sites (36%), untranslational region motifs (36%), Polyadenylation sites (33%), RNA C to U editing site (32%), CIS-regulatory elements (31%), micro RNA target sites (30%), long stems (28%), A-U rich elements (24%), Rho independent terminator (19%), nc hybrdization (18%), Functional RNA sequences (17%), Ribosome binding site (15%), ORF (7%) and Transcriptional regulatory motif (2%) respectively. Higher number of ORF was observed in GreeNC database. ORF motif predicted in PNRD database is comparatively less with other databases. Higher numbers of Functional RNA motifs were observed in PNRD database. No Transcriptional regulatory motif was observed in PLNlncRbase.

Motif Analysis

Motifs present in biological sequences define the function and their role in biological system. AU rich element motifs, which have pronounced role in regulation of gene expression, were predicted in lncRNA sequences of all the three databases. Further, most of the CIS regulatory element motifs were predicted as polyadenylated sites. Other CIS regulatory motifs like Hammer head_form_III, Hammer head_form_I, Tbox, IRE_storage, Rho_independent_ terminatorv CRE_type_3_poliovirus, TAR_HIV-1, Rfam RF00037 (IRE family) and Rfam RF01382 (HIV-1_SL4 family) were of less importance in plants as well they are present in very small numbers that can occur by chance. Analysis of splicing site motifs in rice lncRNA sequences has shown the presence of donor and acceptor splice sites. Presence of rice splicing site motifs is considered important for translation of proteins. Research on alternative splicing is gaining importance due to its role in various stress conditions [21]. Thus the study of RSS motifs in lncRNA may provide valuable insight on the role of lncRNA in stress conditions. Long stems of 32 types were predicted in rice lncRNA sequences, of which 96% was observed to be stem_1_3 prime (43%) and stem_1_5 prime (43%) respectively. Other types were observed in minimal or unaccountable numbers. Length of the long stem ranged from 43-481 bases of which 77% of sequences were of length less than 100 bases. Many precursor microRNA motifs and tRNA motifs were predicted as Functional RNA sequence motifs in rice lncRNA sequences. Complete information of all the fRNA motifs are provided as supplementary information. RNA editing motifs observed at the rate of occurrence of c (44%), a (20%), t (19%) and g (17%) in the RNA bases, respectively. As RNA editing has the power to reprogram genetic information and also perform transcript repair [22] it becomes important to know the RNA editing events in lncRNA to understand the genetic makeup of the plants. Most of UTR motifs predicted were observed to be PAS (25%) and Mushashi binding element (MBE) (44%). MBE is of less importance for plants as it is mostly observed in mammals. Apart from the all the above discussed motifs, Ribosome binding site, PAS (which is also predicted in UTR motif and CIS regulatory element motif) and Rho dependent terminator motifs were also predicted which remains important for the initiation, termination of translational events in the biological system.

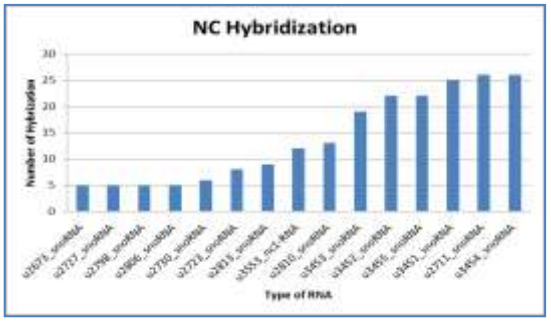

Fig-3 Number of small nucleolar RNAs and non-coding RNA hybridized with lncRNA

Analysis of Non coding RNA hybridization motifs showed hydridization with small nucleolar RNAs (snoRNA) in all the database sequences. To the very negligible level, hybridization was reported with small nuclear RNA and non codingRNA (u3553_nc1-RNA). [Fig-3] provides information of various small nucleolar RNAs involved in lncRNA hybridization.

Fig-4 Percentage distributions of transcriptional regulatory motifs

Transcriptional regulatory motifs were mostly predicted in the sequences obtained from GreeNC and PNRD database that is also in very low numbers. [Fig-4] illustrates the details of regulatory motifs that has occurred more than 1% rice lncRNA sequences. SMAD3 (7%), LEF1 (6%), cMyc:Max (6%), SMAD (5%), PEA3 (5%) and SMAD4 (5%) were the regulatory motifs observed in rice lncRNAsequences which mostly studied in mammals/vertebrates.

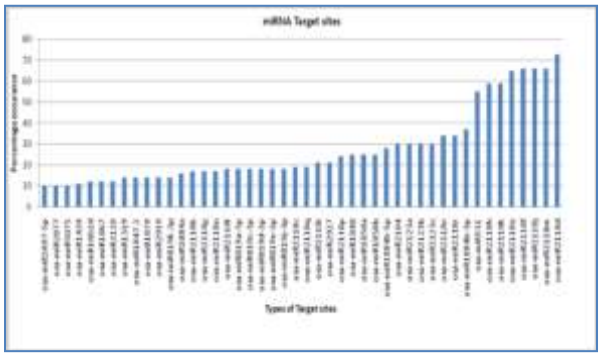

Fig-5 Percentage distributions of miRNA target sites

Among the 6064 lncRNA sequences analyzed, 412 types of miRNA target sites were predicted from *Oryza sativa japonica sps*. Experimentally identified rice lncRNAs obtainedfrom PLNlncRBase have 23 microRNA target sites. osa-miR399a, osa-miR399b, osa-miR399c, osa-miR399d, osa-miR399e, osa-miR399f, osa-miR399g, osa-miR399h, osa-miR399i, osa-miR399j, osa-miR399k, osa-miR2927, osa-miR5517, osa-miR172a, osa-miR172b, osa-miR172c, osa-miR172d, osa-miR2103, osa-miR5500, osa-miR3979-5p, osa-miR5082, osa-miR5539 and osa-miR1862e were miRNA target sites observed in PLNlncRbase. Information regarding all the miRNA target sites is given in the supplementary information. miRNA 2118d is the target site that had maximum occurrence of 73 (3% approx.) among the total number of 2299 miRNA target sites and this site was observed in PNRD (72 sites) and GreeNC (1 site) database sequences [Fig-5]. Many target sites were observed only once, in which 392 types of miRNA target sites occurred less than 1%. Different target sites were predicted other than that observed for the experimentally identified lncRNAs. This prediction provides the knowledge of other possible target sites that exists in the rice lncRNA sequences. Information regarding the percentage (>10%) of occurrence of various miRNA target sites is given in figure.

Repeats analysis

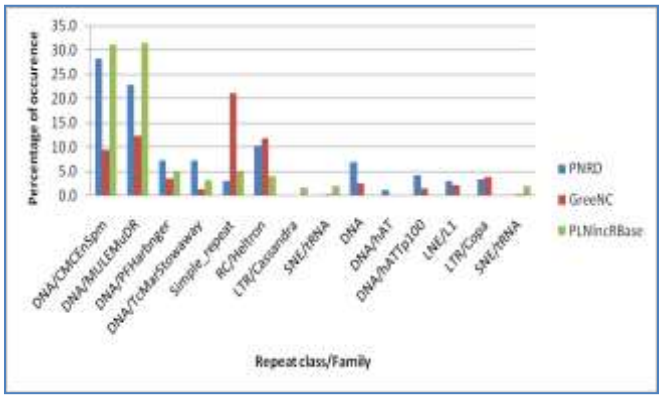

Fig-6 Types of repeat class/family observed in rice LncRNA sequences.

In rice, 82 repeat class/family were predicted for all the lncRNAs from GreeNC, PNRD and PLNlncRBase databases. Maximum number of 80 repeat classes was observed in GreeNC database [Fig-6]. 14 and 10 types of repeats were predicted in PNRD and PlnLNCRBase databases respectively. Majority of repeats that occurred belong to the DNA/TcMarStowaway, RC/Heltron, DNA/PFHarbnger, Simple repeat, DNA/CMCEnSpm and DNA/MULEMuDRclass/family. These repeats mainly constitute class II type transposable elements which move directly by a 'cut and paste' mechanism [23].

Conclusion

From the comparison of 14 types of motifs present in 6064 rice lncRNA sequences from the GreeNC, PNRD and PLNlncRBase databases, untranslated region (UTR) and rice splicing site motifs occurred maximum in all the databases. lncRNA sequences mostly contained UTR, Rice splicing sites, RNA C to U editing site, Polyadenylation site, ORF, miRNA target sites, CIS regulatory element motifs. 210 types of snoRNA were found to have ncRNA hybridization Transcriptional regulatory motifs were least or not present in the lncRNA sequences. Repeat analysis showed the presence of transposable repeat elements in lncRNA sequences. This study provided general insight on the possible motifs and repeats present in rice lncRNA sequences. Motifs and repeat analysis have showed the prevalence of lncRNAs regulatory role in transcriptional and translational events. Results obtained from this study have provided insight on occurrence of various motifs in the rice lncRNA sequences. It remains important to increase the rice productivity and also to reduce crop and grain loss due to various abiotic and biotic stress conditions. With the development of advanced next generation sequencing technologies, future research on disease relevance of the sequence modifications undergone in lncRNA may enlighten the knowledge of genetic changes at various stress conditions.

Application of Research

Rice lncRNA motifs predicted can be used for comparative genome analysis to study about lncRNA genome variation in other plant species. Biologically important motifs identified can be used by plant biologists to perform genetic engineering studies in order to regulate the expression of genes.

Research Category: comparative genome analysis

Abbreviations

LncRNA – Long non-coding RNA
ORF - Open Reading Frame
PAS – Poly Adenylation Site
UTR - Untranslated Region
fRNA – Functional RNA
MBE - Mushashi binding element
IRE – Iron Response Element
RSS – Rice Splicing Sites

Acknowledgement / Funding: Author are thankful to Tamil Nadu Agricultural University, Coimbatore, 641003, Tamil Nadu, Coimbatore and DBT-BTIS facility provided by DBT, Govt. of India.

***Principal Investigator: Dr N. Saranya**
University: Tamil Nadu Agricultural University, Coimbatore, 641003, Tamil Nadu
Research project name or number: Nil

Author Contributions

NS designed and executed the work. LTS and SS involved in data collection. NS, NB, AK and MJ analyzed the results. NS, AK and JR prepared and revised the manuscript.

Author statement: All authors read, reviewed, agree and approved the final manuscript

Conflict of Interest: None declared

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Wierzbicki A.T. (2012) *Curr.Opin. Plant Biol*, 15, 517–522.

[2] Kim E.D., Sung S. (2012) *Trends Plant Sci*, 17, 16–21.
[3] Liu X., Hao L., Li D., Zhu L. and Hu S. (2015) *Genomics Proteomics Bioinformatics*, 13(3),137–47.
[4] Heo J.B. and Sung S. (2011) *Science,* 331(6013), 76–9.
[5] Liu J., Jung C., Xu J., Wang H., Deng S., Bernad L., Arenas-Huertero C. and Chua N.H. (2012). *Plant Cell* 24, 4333–4345.
[6] Xin M., Wang Y., Yao Y., Song N., Hu Z., Qin D., Xie C., Peng H., Ni Z. and Sun Q. (2011) *BMC Plant Biol.,* 11, 61.
[7] Kim E.D. and Sung S. (2012) *Trends Plant Sci*., 17(1), 16–21.
[8] Quan M., Chen J. and Zhang D. (2015). *Int J Mol.Sci*., 16, 5467–5496.
[9] Zhu Q.H. and Wang M.B. (2012) *Genes*. 3,176–190.
[10] Swiezewski S., Liu F., Magusin A. and Dean C. (2009) *Nature*, 462(7274) ,799–802.
[11] Heo J.B. and Sung S. (2011) *Science*, 331, 76–79.
[12] Shin J.H. and Chekanova J.A. (2014) *PLoS Genet*., 10(9), e1004612.
[13] Bardou F., Ariel F., Simpson C.G., Romero-Barrios N., Laporte P., Balzergue S., Brown J.W. and Crespi M. (2014) *Dev Cell*. 30(2), 166–76.
[14] Kwenda S, Birch PR and Moleleki LN (2016) *BMC genomics*, 17(1), 614.
[15] Zhang YC, Liao JY, Li ZY, Yu Y, Zhang JP, Li QF, et al (2014) *Genome* Biol., 15, 512.
[16] Zhang YC and Chen YQ (2013) *Biochem Biophys Res Commun.,* 436(2), 111-114.
[17] Shin SY and Shin C (2016) *Plant Biotechnol Rep.,* 11. 10(2), 35-47.
[18] Paytuví Gallart A., Hermoso Pulido A., AnzarMartínez de Lagrán I., Sanseverino W., Aiese Cigliano R. (2016) *Nucleic Acids Res.,* 44(D1), D1161-D1166.
[19] Yi X., Zhang Z., Ling Y., Xu W. and Su Z. (2014) *Nucleic Acids Res.,* 43(D1), D982-D989.
[20] Xuan H., Zhang L., Liu X., Han G., Li J., Li X., Liu A., Liao M. and Zhang S. (2015) *Gene*, 573, 328–332.
[21] Zhiguo E., Wang L. and Zhou J. (2013) *BMB Rep*., 46(9), 439.
[22] Vu L.T. and Tsukahara T. (2017). *Biosci trends*., 11(3), 243-253.
[23] Docking T.R., Saadé F.E., Elliott M.C. and Schoen D.J. (2006) *J Mol Evol*, 62(4), 375-87.