Plant Biotechnology Journal
Open Access

Plant Biotechnology Journal (2021) **19**, pp. 1492–1494 doi: 10.1111/pbi.13639

Brief Communication

RiceLncPedia: a comprehensive database of rice long non-coding RNAs

Zhengfeng Zhang[1,*,#], Yao Xu[2,#], Fei Yang[3], Benze Xiao[3] and Guoliang Li[2,4,*]

[1]*School of Life Sciences, Hubei Key Laboratory of Genetic Regulation and Integrative Biology, Central China Normal University, Wuhan, China*
[2]*National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China*
[3]*College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China*
[4]*Agricultural Bioinformatics Key Laboratory of Hubei Province, Hubei Engineering Technology Research Center of Agricultural Big Data, 3D Genomics Research Center, College of Informatics, Huazhong Agricultural University, Wuhan, China*

Received 30 January 2021;
revised 4 May 2021;
accepted 14 May 2021.
**Correspondence* (Tel +86 13627259418; email: zhengfeng@mail.ccnu.edu.cn (Z.Z.); Tel +86-027-87285078; email: guoliang.li@mail.hzau.edu.cn (G.L.))
[#] These authors could be regarded as Joint First Authors.

Keywords: rice, LncRNA, expression, multi-omics, phenotype, SNP, transposon, miRNA.

Long non-coding RNAs (lncRNAs) are referred as RNA molecules with length of at least 200 nucleotides (nt) and usually have low protein-coding potential (Chekanova, 2015). In plants, emerging evidence indicate that lncRNAs function as key modulators in development and stress response at the epigenetic, transcriptional and post-transcriptional levels (Chekanova, 2015; Lucero *et al*., 2021). Multiple comprehensive databases have been constructed for human or animals, such as LncBook (Ma *et al*., 2019). Comparatively, the small amount of lncRNAs, the small sample scale or less comprehensive were the main limitations for current plant lncRNA databases, especially for rice, one of the most widely staple food and model crops (Table 1). For example, GREENC provides lncRNAs in 45 plant species, but without their expression profile and genomic features (Paytuvi-Gallart *et al*., 2019). A rice genome re-annotation database IC4R 2.0 harbours 3215 lncRNA loci, 6259 transcripts but without the relevant multi-omic features of them (Sang *et al*., 2020). PLncDB V2.0 was updated very recently, containing information for 11565 rice lncRNAs identified from 98 RNA-Seq libraries (Jin *et al*., 2021). Here, we developed a database, RiceLncPedia (http://3dgenome.hzau.edu.cn/RiceLncPedia), to systematically characterize rice lncRNAs with expression profile and multi-omic features to facilitate the understanding and research of rice lncRNAs, including as follows: (i) lncRNA expression profiles in various tissues, development stages and stress treatments; (ii) lncRNA associations with genome variations; (iii) the linkage of lncRNAs with phenotypes; (iv) the overlap information of lncRNAs and transposon elements; and (v) the lncRNAs predicted as miRNA targets or miRNA precursors.

We identified high-confidence rice lncRNAs based on 2312 publicly available RNA-seq libraries following the unified pipeline (Figure 1a). Briefly, low-quality reads and adapter sequences were trimmed and the clean RNA-seq reads were mapped to rice reference genome Os-Nipponbare-Reference-IRGSP-1.0.41. Transcripts were assembled and merged to acquire comprehensive non-redundant transcripts for subsequent analysis: (i) filter out known protein-coding transcripts, rRNA and tRNA; (ii) transcripts with lengths less than 200nt and with FPKM scores smaller than 0.5 in all samples were discarded successively; (iii) protein-coding potential was predicted using Coding Potential Calculator (CPC2), Plant Long Non-Coding RNA Prediction by Random fOrests (PlncPRO) and PfamScan software. As a consequence, RiceLncPedia accommodates 6925 rice lncRNAs in 5812 gene loci (Figure 1b). In addition, 40 experimental validated rice lncRNAs were also integrated into RiceLncPedia (Table 1). In the transcript section, each lncRNA transcript is assigned to a unique accession number and shown with molecular features, including location, length, GC content (%), exon number and category. LncRNAs are classified into intergenic lncRNA (lincRNA), intronic lncRNA, sense lncRNA, antisense lncRNA and long non-coding isoforms of known genes according their positions relative to coding genes (Figure 1b).

For each given lncRNA transcript, a specific page was linked to incorporate sequence, coding score, genome browser view, expression profile, variation, overlapped transposons, small RNA associations, QTL and GWAS information relevant to the lncRNA (Figure 1c). Because the specific expression in a specific tissue or under a specific condition indicates the function association (Yanai *et al*., 2005), the expression profiles of any given lncRNA can be visualized in bar charts, for a few represented projects, covering diverse tissues such as leaf, shoot, root, seed, glume and panicle callus, samples from phosphate starvation, salt, cadmium, drought, cold, osmotic and flood stresses as well as samples grown under JA and ABA treatments. An interactive graphic was presented for further visualization of each lncRNA expression in genome browser (Figure 1c). To explore the relationship of samples, we clustered the represented 339 libraries mentioned above according to all 6925 lncRNAs expression values (Figure 1d). The resultant clusters were well matched between the *indica* and *japonica* groups, basically indicating the reliability of lncRNA expression profiles in RiceLncPedia.

The multi-omics page provides different molecular features for all lncRNAs. The lncRNA expression profiles were provided across all 2312 collected RNA-seq libraries, which are available for download in RiceLncPedia. We calculated the maximum, average and median (FPKM) as well as expression breadth, coefficient of variance (CV), tissue-specificity index and stress-responsive index (τ-Value) for each lncRNA transcript, which were harboured in expression section. The specific expressed lncRNAs in given tissues or growth conditions can be screened by selecting a dataset such as 'ABA treatment' and defining a specified range of CV, τ-value or expression breadth (Figure 1c, Help section).

Genome variation section contained 50441 SNPs in 4883 lncRNA transcripts with an average of about 10 SNPs per lncRNA transcript (Figure 1e,f), by comparing the position of lncRNAs with SNPs based on 3000 genome projects (http://snpseek.irri.

Table 1. Key differences between existing rice lncRNA databases and RiceLncPedia

Data Item	RiceLncPedia	PLncDB V2.0	GREENC	CANTATAdb 2.0	PNRD	PLNincRbase	PLncPRO	EVLncRNAs	IC4R-2.0	NONCODE
Rice RNA-seq libraries	**2312**	98	NA	9	NA	NA	9	NA	1503	NA
Species	**Rice**	80 species	45 species	39 species	150 species	43 species	Rice, chickpea	77	Rice	Animals & Plants
LncRNA entries	**6965**	1246372	120,000	239,631	25739[a]	1187	12314	1543	6259	Many
Rice LncRNA entries	**6965**	11565	5237	2788	790	unknown[b]	7345	40	6259	1190
Experimentally identified lncRNAs	**40[c]**	39[c]	NA	NA	Partial	1060	NA	√	NA	NA
Sequences	√	√	√	√	√	√	√	√	√	√
Coding potential	√	√	√	√	√	NA	NA	NA	√	√
Genomic information	√	√	√	√	√	√	NA	√	√	√
Expression	√	√	NA	√	NA	√	NA	√	√	√
Small RNA associations	√	NA	√	NA	NA	NA	NA	NA	NA	NA
LncRNA genome variation	√	NA	NA	NA	NA	NA	NA	NA	NA	NA
Phenotype association	√	NA	NA	NA	NA	√	NA	√	NA	NA
TE-lncRNAs	√	NA	√	NA	NA	NA	NA	NA	NA	NA
Conservation analysis	**NA**	NA	NA	NA	NA	NA	NA	NA	NA	√
Accessible	√	√	√	√	√	NA	√	√	√	√

Note: a indicates the number include both microRNAs and lncRNAs. b represents the unknown number of rice lncRNAs since the database is not accessible following the link in the reference. c means the experimentally validated lncRNAs were integrated from EVLncRNAs database. All the data item features in RiceLncPedia are shown in bold font.

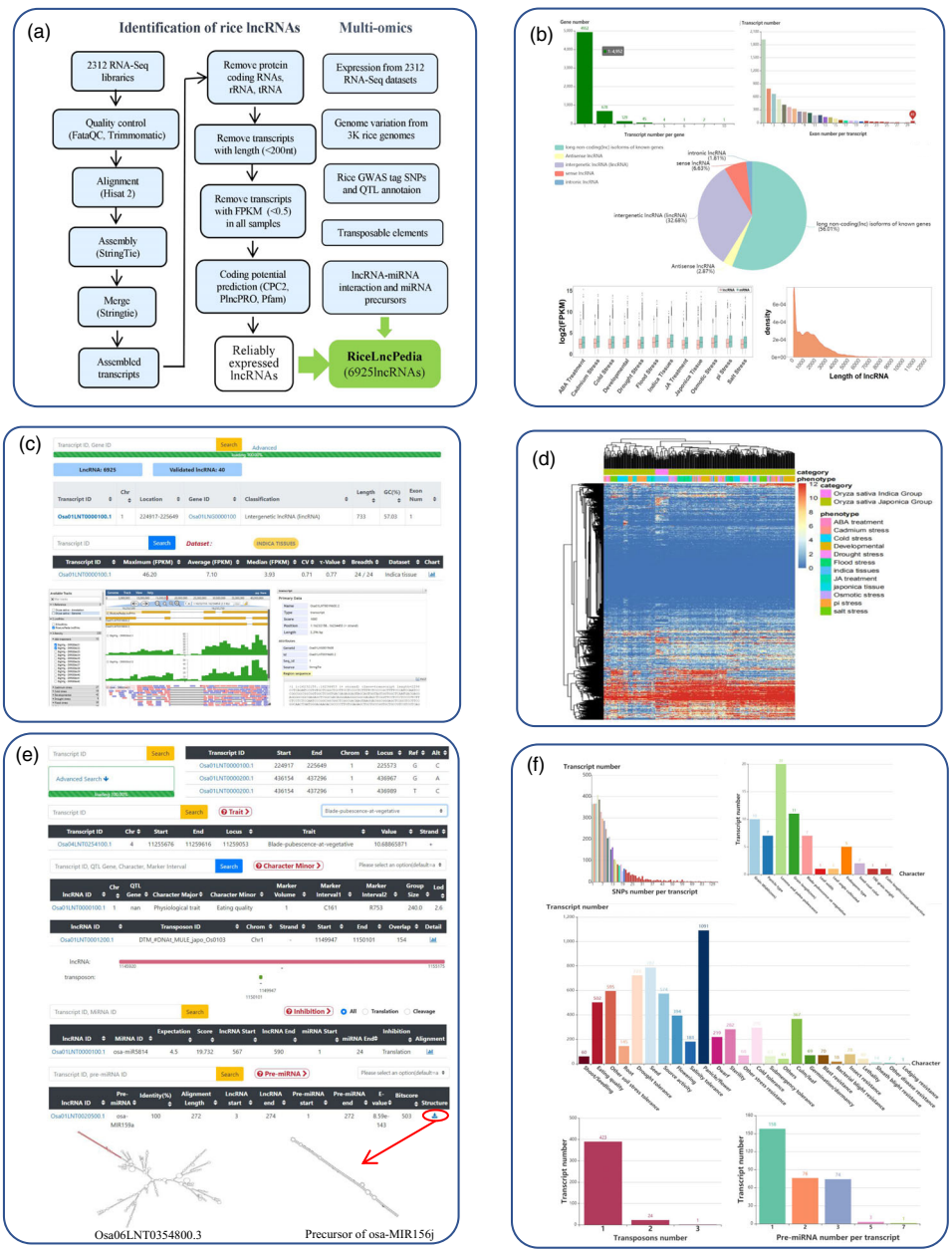

Figure 1 The architecture of RiceLncPedia. (a) Pipeline for rice lncRNAs identification and multi-omics data integration. (b) The distribution of transcript numbers, exon numbers, classification, expression and length of rice lncRNAs. (c) Snapshots of basic characterization, expression profile and Gbrowser view of lncRNAs. (d) Hierarchical clustering heatmap in selected RNA-seq libraries based on lncRNA expression values. (e) Snapshots for lncRNAs associated with SNP, GWAS, QTL, transposons, microRNA targets and precursors of small RNAs with an example of lncRNA and the microRNA precursor structures. (f) The distribution of multi-omics features associated with lncRNAs.

org/_download.zul). This information will facilitate the research of lncRNA variation association with their structures, expressions, interactions and functions.

In plants, some lncRNA-SNPs were implicated to play potential roles in regulating agricultural traits through GWAS or QTL analysis. We, therefore, predicted the lncRNA-SNP-phenotype association if any rice agricultural GWAS tag SNP co-located with a specific lncRNA. Similarly, a lncRNA resided in any rice QTL was also thought of being associated with the relevant trait. The QTL section shows 6684 rice lncRNAs co-located with 513 QTLs, such as 1000 grain weight, drought tolerance and so on, belonging to 25 tissues, development stages or stress tolerance. The GWAS section presents 384 GWAS SNPs residing in 66 lncRNAs transcripts, which refers to 11 agricultural traits (Figure 1e,f). A specified trait-related lncRNAs can be retrieved by selecting the trait in left menu.

A number of lncRNAs were reported to be originated from transposons in plants, and it was demonstrated that TE-associated lncRNAs show tissue-specific transcription and play vital roles in plant abiotic stress responses (Wang *et al*., 2017). We overall identified 82 transposons overlapped with 448 lncRNA transcripts, involving 474 transposon and lncRNA transcript relations (Figure 1e,f) by comparing the positions of lncRNAs with transposons (Genomic coordinates of Japonica transposon elements, https://www.genome.arizona.edu/cgi-bin/rite/index.cgi). All lncRNAs overlapping with TE were contained in RiceLncPedia as TE-lncRNAs associations in transposon section.

To facilitate the function prediction of rice lncRNAs, we predicted lncRNA targets of microRNAs with psRNATarget software (Dai *et al*., 2018) and screened the precursors of miRNAs by comparing lncRNAs sequences with rice miRNA precursor (pre-miRNA) hairpin sequences (http://www.mirbase.org/). Blast 2.7 was used with the threshold e-value $\leq 10^{-5}$, coverage per cent bigger than 90% and -max_hsps as 1. The secondary structures of lncRNAs and relevant pre-miRNAs were built with the localized RNAfold program with default parameters (Gruber *et al*., 2008). In the end, the Small RNA targets section contains 6060 lncRNAs targets of 713 Osa-miRNAs, building up 64153 lncRNA and Osa-miRNA interactions (Figure 1e). Pre-miRNA section harbours 312 lncRNAs with high homology with 48 pre-miRNAs, involving 554 relations of lncRNAs and miRNAs (Figure 1e,f). The homology information of lncRNAs with rice pre-miRNAs, the optimal secondary structures in both dot-bracket notation and graphical style with the minimum free energy were available to be downloaded in RiceLncPedia.

RiceLncPedia database was constructed using Django as back-end Web framework and PostgreSQL (https://www.postgresql.org/) as the database engine. JQuery and AJAX (Asynchronous JavaScript and XML) were used to develop Web interfaces. As for the front-end framework, we employed Bootstrap (https://getbootstrap.com) to supply a series of templates to design Web pages with consistent interface components. We adopted the icon in Font Awesome in the RiceLncPedia website (http://www.fontawesome.com.cn/). Data visualization was powered by Pyecharts (https://github.com/pyecharts/pyecharts) to add interactive diagrams to our website. All the data and methods can be downloaded in download page.

In summary, RiceLncPedia houses a comprehensive collection of rice lncRNAs from the widest samples and with systematic annotation through integrating multi-omics data, covering molecular features, expression profiles, sequences variations, lncRNA-miRNA association, lncRNAs-transposon association and agricultural traits association. All the methods and data are available in help or download page. Future development of RiceLncPedia will refer to regular updates of newly discovered rice lncRNAs, integration of differentially expressed lncRNAs in more diverse tissues and environments, epigenetic features of lncRNAs and the association of lncRNAs with protein-coding genes, experimentally validated lncRNAs and more lncRNA-phenotype associations. We are looking forward to any reasonable suggestions from worldwide scientists, with the aim to provide a continually updated and rich knowledge reservoir of rice lncRNAs and serve as a valuable resource for rice research communities.

Acknowledgements

This work was supported by the self-determined research fund of Central China Normal University (CCNU18QN027) and the National Special Key Project of China on Transgenic Research (2016ZX 08001-003).

Conflict of interest

The authors have no conflict of interest to declare.

Author contributions

Z.Z. and G.L. designed the project and wrote the manuscript. Z.Z., Y.X., F.Y. and B.X. analysed the data. Y.X. constructed the database.

References

Chekanova, J.A. (2015) Long non-coding RNAs and their functions in plants. *Curr. Opin. Plant. Biol.* **27**, 207–216.

Dai, X., Zhuang, Z. and Zhao, P.X. (2018) psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Res.* **46**, W49–W54.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R. and Hofacker, I.L. (2008) The Vienna RNA websuite. *Nucleic Acids Res.* **36**, W70–W74.

Jin, J., Lu, P., Xu, Y., Li, Z., Yu, S., Liu, J., Wang, H. *et al*. (2021) PLncDB V2.0: a comprehensive encyclopedia of plant long noncoding RNAs. *Nucleic Acids Res.* **49**(D1), D1489–D1495.

Lucero, L., Ferrero, L., Fonouni-Farde, C. and Ariel, F. (2021) Functional classification of plant long noncoding RNAs: a transcript is known by the company it keeps. *New Phytol.* **229**, 1251–1260.

Ma, L., Cao, J., Liu, L., Du, Q., Li, Z., Zou, D., Bajic, V.B. *et al*. (2019) LncBook: a curated knowledgebase of human long non-coding RNAs. *Nucleic Acids Res.* **47**, 2699.

Paytuvi-Gallart, A., Sanseverino, W. and Aiese Cigliano, R. (2019) A Walkthrough to the Use of GreeNC: The Plant lncRNA Database. *Methods Mol. Biol.* **1933**, 397–414.

Sang, J., Zou, D., Wang, Z., Wang, F., Zhang, Y., Xia, L., Li, Z. *et al*. (2020) IC4R-2.0: Rice Genome Reannotation Using Massive RNA-seq Data. *Genomics Proteomics Bioinformatics* **18**, 161–172.

Wang, D., Qu, Z., Yang, L., Zhang, Q., Liu, Z., Do, T., Adelson, D. *et al*. (2017) Transposable elements (TEs) contribute to stress-related long intergenic noncoding RNAs in plants. *Plant J.* **90**, 133–146.

Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A. *et al*. (2005) Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659.

