PLncDB V2.0: a comprehensive encyclopedia of plant long noncoding RNAs

Jingjing Jin [1,†]**, Peng Lu**[1,†]**, Yalong Xu**[1]**, Zefeng Li**[1]**, Shizhou Yu**[2]**, Jun Liu**[3]**, Huan Wang**[4]**, Nam-Hai Chua**[5,6] **and Peijian Cao** [1,*]

[1]China Tobacco Gene Research Center, Zhengzhou Tobacco Research Institute of CNTC, Zhengzhou 450001, China, [2]Molecular Genetics Key Laboratory of China Tobacco, Guizhou Academy of Tobacco, Guiyang 550081, China, [3]National Key Facility for Crop Resources and Genetic Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, [4]Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China, [5]Temasek Life Sciences Laboratory, 1 Research Link, National University of Singapore, Singapore and [6]Laboratory of Plant Molecular Biology, Rockefeller University, New York, NY, USA

ABSTRACT

Long noncoding RNAs (lncRNAs) are transcripts longer than 200 nucleotides with little or no protein coding potential. The expanding list of lncRNAs and accumulating evidence of their functions in plants have necessitated the creation of a comprehensive database for lncRNA research. However, currently available plant lncRNA databases have some deficiencies, including the lack of lncRNA data from some model plants, uneven annotation standards, a lack of visualization for expression patterns, and the absence of epigenetic information. To overcome these problems, we upgraded our Plant Long noncoding RNA Database (PLncDB, http://plncdb. tobaccodb.org/), which was based on a uniform annotation pipeline. PLncDB V2.0 currently contains 1 246 372 lncRNAs for 80 plant species based on 13 834 RNA-Seq datasets, integrating lncRNA information from four other resources including EVLncRNAs, RNAcentral and etc. Expression patterns and epigenetic signals can be visualized using multiple tools (JBrowse, eFP Browser and EPexplorer). Targets and regulatory networks for lncRNAs are also provided for function exploration. In addition, PLncDB V2.0 is hierarchical and user-friendly and has five built-in search engines. We believe PLncDB V2.0 is useful for the plant lncRNA community and data mining studies and provides a comprehensive resource for data-driven lncRNA research in plants.

INTRODUCTION

Long noncoding RNAs (lncRNAs) are transcripts longer than 200 nucleotides with little or no protein coding potential (1–4). They account for a large portion of transcripts in cells; for instance, there are 189 901 lncRNAs annotated in RNAcentral (5) for human compared with 19 288 protein-coding genes in ENSEMBL (6). Mammalian lncRNAs have been shown to regulate gene expression and other cellular processes and are implicated in cancer and other diseases (7). In addition to functioning as miRNA sponges (8), plant lncRNAs are also known to interact with regulatory proteins to modulate gene expression (1,3,9–11). From the earliest investigations of important lncRNAs including *COLDAIR*/*COOLAIR* (12) and *IPS1* (13) almost two decades ago, recent years have seen rapidly growing interest in the functional exploration of individual plant lncRNA (14,15).

The expanding list of lncRNAs and accumulating functional evidence in plants have necessitated the development of a comprehensive database to act as a data repository and a platform for lncRNA analysis. Multiple databases have thus been developed in the past several years (16), such as lncRNAdb (17), NONCODE (18), EVLncRNAs (19), PLNlncRbase (20), CANTATAdb (21), GreeNC (22), RNAcentral (5) and PLncDB V1.0 (1), all of which are devoted to archiving plant-related lncRNA information (Supplementary Tables S1 and S2). These databases have greatly facilitated lncRNA studies, but they may have some deficiencies, such as uneven annotation standards, and they lack of some of the important features for lncRNAs, including expression information, potential target information and epigenetic signals. For instance, RNAcentral (5), which attempted to collect lncRNAs from several impor-

---

tant databases, lacks standard criteria for lncRNA annotation. Moreover, it only contains basic lncRNA features, such as genome coordinates and sequence information, without expression or target information. EVLncRNAs (19), lncRNAdb (17) and PLNlncRbase (20) mainly focused on experimental validated lncRNA candidates, which resulted in limited number of lncRNAs in these databases. As the most comprehensive databases for plant lncRNA research, although CANTATAdb (21) and GreeNC (22) attempted to annotate plant lncRNAs with standard criteria, the number of lncRNAs and annotated plant species is still far from sufficient due to the limited datasets they used. Moreover, these two databases are incompatible with known validated lncRNA information, for example, the well-known lncRNA *COLDAIR* cannot be found either by keyword search or by sequence blast. Thus, it is highly desirable to establish a plant lncRNA database with more uniform and strict annotation standards, greater data integrity and more comprehensive annotations.

Since its launch, PLncDB has attracted wide interest from the plant lncRNA research community. As of August 2014, PLncDB was inducted into RNAcentral as a third party data specialist database. In the present study, with the goal of creating a one stop plant lncRNA database, we upgraded PLncDB V2.0 to use a standard annotation criteria to systematically annotate 1 246 372 lncRNAs in 80 phylogenetically representative plant species. Known and experimentally verified lncRNAs were collected from EVLncRNAs, and plant lncRNA candidates in RNAcentral were also integrated into our database. PLncDB V2.0 also provides an easy-to-use interface to browse, search, and download data, enabled by several search engines. In addition, PLncDB V2.0 deploys several tools (JBrowse (23), eFP Browser (24), EPexplorer and so on) to visualize expression patterns and epigenetic signals for lncRNAs. Taken together, PLncDB V2.0 is a comprehensive functional database amenable for data mining and database-driven research therefore providing a useful resource for the plant lncRNA research community.

MATERIALS AND METHODS

Data sources

The current version of PLncDB (version 2.0) consists of lncRNA entries from 80 plants from chlorophytes to embryophyta, including 16 monocotyledons and 53 eudicotyledons (Figure 1A, Supplementary Table S3). Whole genome references and gene annotations were downloaded from NCBI (25), Phytozome V12.1 (26), and other official websites for model plants (*Arabidopsis*: TAIR; Tobacco/tomato/potato: Sol Genomics Network (27); Maize: MaizeGDB (28); Rice: MSU RGAP (29)). More detailed information is in Supplementary Table S3. In total, 13 834 RNA-Seq datasets were obtained from the NCBI SRA Database (https://www.ncbi.nlm.nih.gov/sra/); details are in Supplementary Table S4. For model plants, we specially collected datasets from studies focusing on transcriptome landscape by mining literatures. Meanwhile, we downloaded RNA-Seq datasets from NCBI for each plant searching by scientific names. The RNA-Seq datasets missing information about tissue, development stage/age, or

treatment were discarded. In addition, methylation, histone and small RNA sequencing data were also collected for model plants (Supplementary Table S5). The known experimentally verified lncRNAs were collected from EVLncRNAs, and lncRNA candidates of plants in RNAcentral were also integrated into our database. For model plants, we also collected several well-studied lncRNAs from other studies (3,4,30).

Data analysis pipelines

Data pre-processing. Linux version 2.8 of the SRA toolkit was employed to convert original compressed sra files into Fastq format. Trim Galore (version 0.50) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used to trim adapter sequences with parameters '-q 30 –length 35' (Figure 1B).

Annotation of lncRNAs. Clean data were aligned to the reference genome using the read aligner HIAST2 (31). The transcriptome of each library was assembled separately using StringTie (32), and all gtf result files were then merged into one for each species with StringTie –merge (Figure 1B). Then, we compared the assembled transcript isoforms with the reference genome annotation information, which represents all protein coding gene models identified in each plant genome. Transcripts with a length shorter than 200nt and an open reading frame (ORF) length longer than 120 amino acids were discarded (https://www.ncbi.nlm.nih.gov/orffinder/). In order to remove transcripts that may encode short peptides, the Swiss-Prot database was searched using the blastx program with the parameters -e 1.0e-4 -S 1. Further, the transcripts overlapping with rRNA, tRNA, sRNA and miRNA in Rfam database were filtered. The CPC (33), PLEK (34), RNAplonc (35) and CREMA (36) programs were used to calculate the coding potential of the remaining transcripts. Only transcripts satisfied with at least two criteria were considered as lncRNA candidates in our database. High-confidence candidates were defined if more than three criteria were satisfied, and the left were defined as medium-confidence candidates.

Expression analysis. The expression values for lncRNAs were normalized by TPM as previously described (37). For each RNA-Seq dataset, reads were mapped to the reference genome by HIAST2 (31), and TPMs were calculated using StringTie (32). The tissue specific expressed lncRNA candidates were measured by ROKU (38) with cutoff 0.8.

Target prediction. The regulatory relationship between miRNAs and protein coding genes were predicted by psRobot (39). Mimic target prediction between lncRNAs and miRNAs was conducted by psMimic (40). The interactions between protein coding genes and lncRNAs were predicted by IntaRNA (41) based on secondary structures.

Database development. PLncDB V2.0 was constructed using the Python language (https://www.python.org/), Vue.js (https://vuejs.org/), ElementUI (https://element.eleme.io/#/), and Django (https://www.djangoproject.com/) framework. Network proxy services were provided through nginx
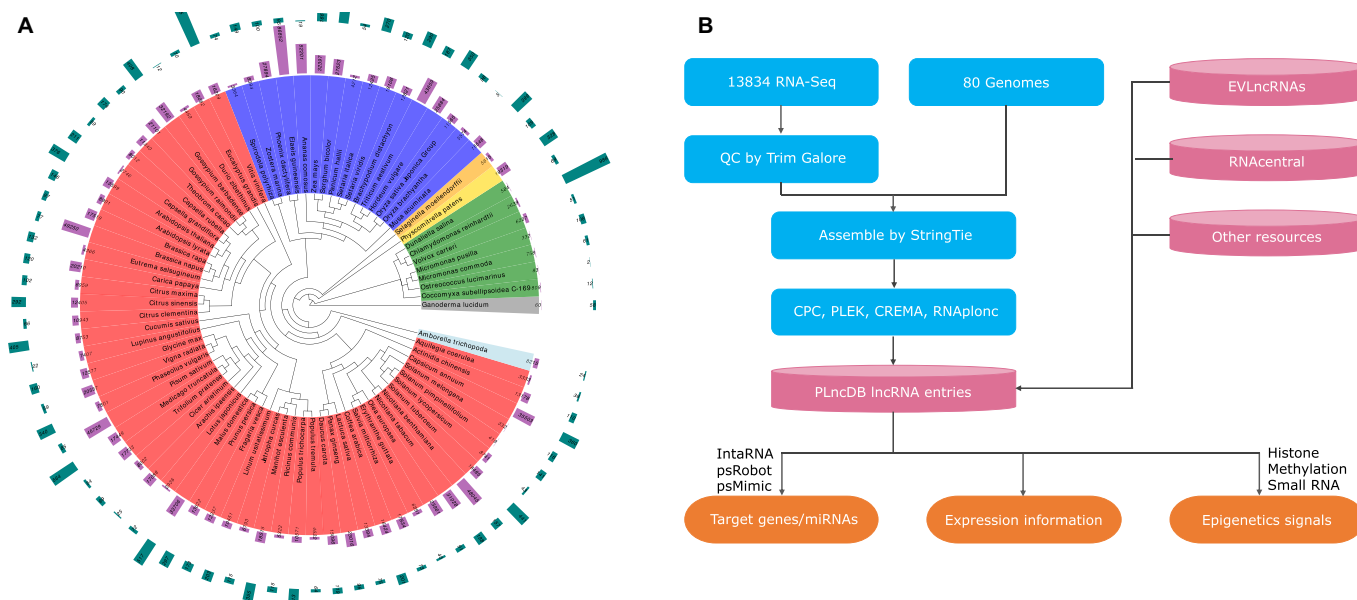
Figure 1. Phylogenetic tree of 80 species included in our analysis and data processing workflow of PLncDB V2.0. (**A**) Phylogenetic tree of 80 species in PLncDB V2.0. From the inside to the outside: phylogenetic tree of 80 species; numbers of identified lncRNAs and numbers of RNA-Seq datasets used in the analysis pipeline. (**B**) Data processing workflow and outcomes of PLncDB V2.0.

(https://www.nginx.com/). The graph and network were visualized using Echarts (https://echarts.apache.org/zh/index.html) and Cytoscape (42).

RESULTS

Data content of PLncDB V2.0

Data in the current version of PLncDB (version 2.0) were based on 13 834 RNA-Seq datasets from 80 plant species ranging from chlorophytes to angiosperms (Figure 1A and Supplementary Table S3). Using a standard method with strict annotation criteria (Figure 1B), 1 246 372 lncRNA transcripts were discovered. Moreover, experimentally verified lncRNAs from EVLncRNAs and plant lncRNAs collected in RNAcentral were also integrated into PLncDB. Due to the difficulty of combining lncRNAs with overlapped regions from different sources, PLncDB currently considers these candidates independently and provides the corresponding source information for users. In addition, PLncDB provides information on epigenetic modification signals on lncRNA-encoding loci and their flanking genomic regions by analyzing epigenetic sequencing data. Compared with other plant lncRNA databases, the coverage of lncRNA annotation was greatly increased in PLncDB V2.0 (Figure 1A, Supplementary Tables S1 and S2, Supplementary Figure S1). In addition, PLncDB V2.0 provides expression patterns for lncRNAs across organ types and developmental stages for all species and potential targets were also predicted by IntaRNA (41) and psMimic (40).

lncRNA features of PLncDB V2.0

Similar to other databases, PLncDB includes all basic information of a given lncRNA, including nucleotide sequence and longest ORF sequence, predicted secondary structure and genome coordinates. Moreover, lncRNA annotation in PLncDB has several unique features. First, since all annotated lncRNAs were predicted from RNA-Seq datasets, lncRNA expression patterns were directly obtained from normalized TPM of these RNA-Seq datasets. Hence, expression of these lncRNAs can be directly used to cluster lncRNA candidates with similar functions (Figure 2A–C). Using heatmaps, we could easily identify some lncRNAs with tissue-specific expression (Figure 2A). Second, except for the known targets of several lncRNAs, the prediction of other targets in PLncDB V2.0 was done using psRobot (39), psMimic (40) and IntaRNA (41) (Figure 2D). Hence, users can predict the putative functions for lncRNAs by their potential targets. Third, expression of thousands of lncRNAs is associated with epigenetic regulation. Yet, little is known about epigenetic regulation of lncRNA expression itself in plants. Hence, in PLncDB V2.0, we analyzed ChIP-chip/ChIP-Seq data of the following histone modifications [H3K27me3, H3K4me3, H3K36me3 and H3K9me3] (Figure 2B). Moreover, we profiled expression changes of lncRNAs in multiple RdRM-related mutants [RDD, DCL1/2/3/4, AGO4, RDR2 and DMS1].

Functions of PLncDB V2.0

PLncDB V2.0 provides convenient access, functional search engines and powerful analytical tools (Figure 3). Users can browse all data by shortcuts and multiple layers of webpages. All data can be downloaded in FTP bulk and customized manners (Figure 3).

Browse. Users can browse lncRNAs using species shortcuts on the home page or through the Browse tag in the toolbar. The Browse tag will guide user to the species
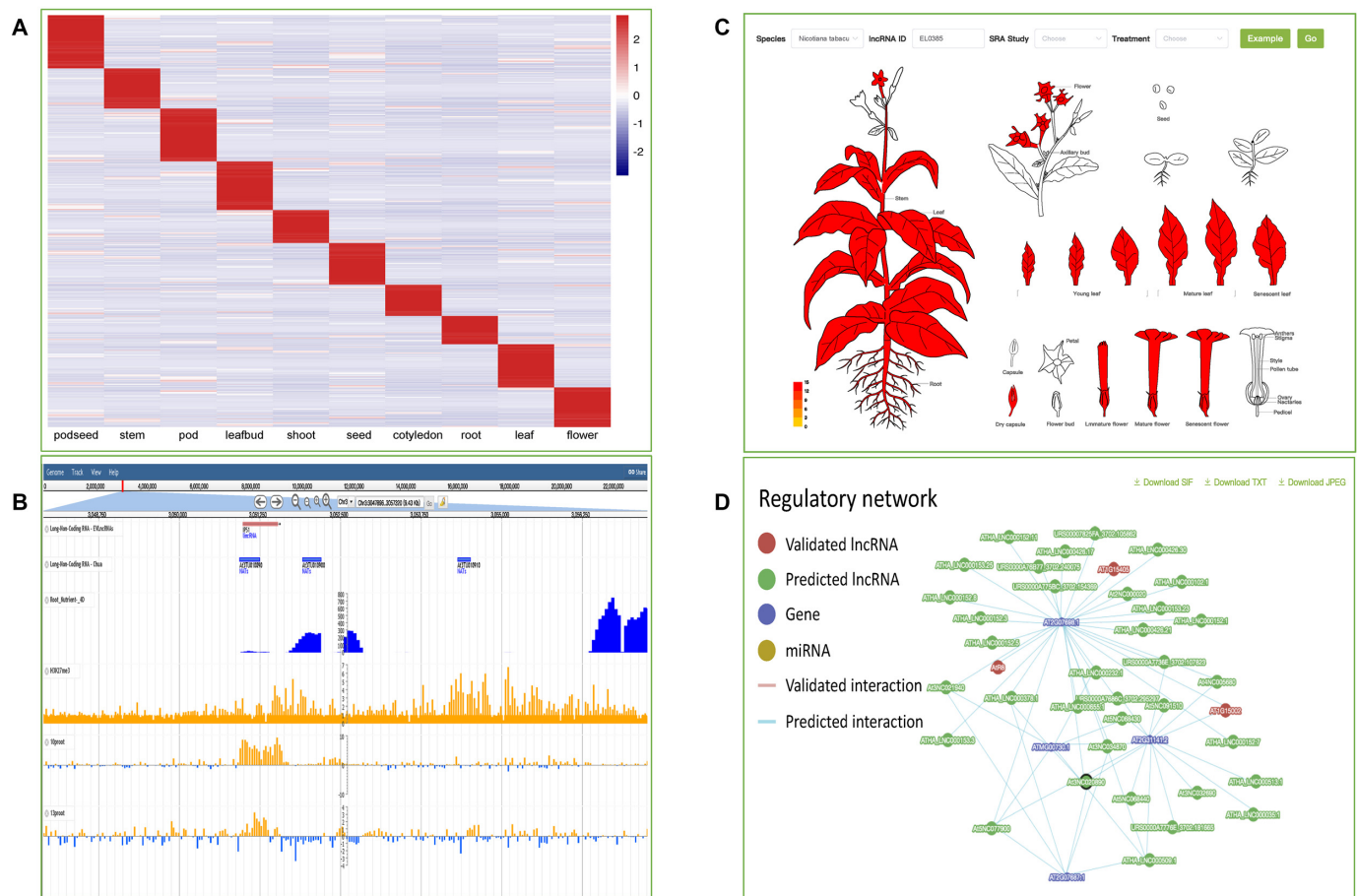
Figure 2. Unique features of PLncDB V2.0. (**A**) An example of lncRNA expression patterns by EPexplorer. Expression values (TPM) of all lncRNAs in soybean were obtained from different tissues including leaf, root, seed, stem, etc (**B**) Visualization of expression patterns and epigenetic modification for lncRNAs (*IPS1*) by JBrowse. (**C**) Visualization of expression patterns for one lncRNA in tobacco using eFP Browser. (**D**) Example of validated and predicted lncRNA targets for one lncRNA in *Arabidopsis*.

lists where the species phylogenetic information is included. Upon clicking the species name, the summary information including the reference genome version, the numbers of predicted and validated lncRNAs and the number of RNA-Seq datasets will be displayed. The next level of Browse is a summary of all lncRNAs and genes in a specific species where users can check detailed information. Further, when clicking each lncRNA candidate, the link will direct users to a detailed information page, consisting of the following six sections: basic type information, sequence information, secondary structure, genomic map information, regulatory network and expression pattern information. For validated lncRNA candidates, it also contains the corresponding experimental information extracted from EVLncRNAs, and the related PubMed literature.

Search. Users can search the whole database with five search engines. Using keyword or lncRNA identifier search engine located on the webpage banner, users can search all fields of the whole database, and the result will be a summarized of a list of all hit lncRNAs. The second search engine is by sequence, and here a BLAST web interface is deployed. For the convenience of quickly searching for lncRNA-related information such as genome coordinates

and expression values, PLncDB V2.0 provides the respective search engines. Meanwhile, PLncDB also supports tissue specific expressed lncRNA candidates search for convenience. All the search results can be downloaded in FASTA, CSV and GFF format.

Tools. We have constructed three useful tools including EPexplorer (Figure 2A), JBrowse (Figure 2B) and eFP Browser (Figure 2C) for PLncDB V2.0. By EPexplorer, we could obtain expression landscape for batch lncRNA candidates and identify tissue specific expressed candidates easily (Figure 2A). The online genome browser with a large set of transcriptome data will be useful for biologists to further investigate functional roles of lncRNAs. With the eFP Browser (Figure 2C), users can quickly obtain the expression landscape for a specific lncRNA, and then design appropriate experiments to verify their hypothesis. In our study, we analyzed Chip-chip and Chip-Seq data for histone modifications including H3K27me3, H3K4me3, H3K36me3 and H3K9me3 and so on (Figure 2B). We also profiled expression changes for some lncRNAs in the context of several RdRM-related mutants, such as RDD, DCL1/2/3/4, AGO and RDR2. These results can be easily accessed using JBrowse for each plant. For instance, for
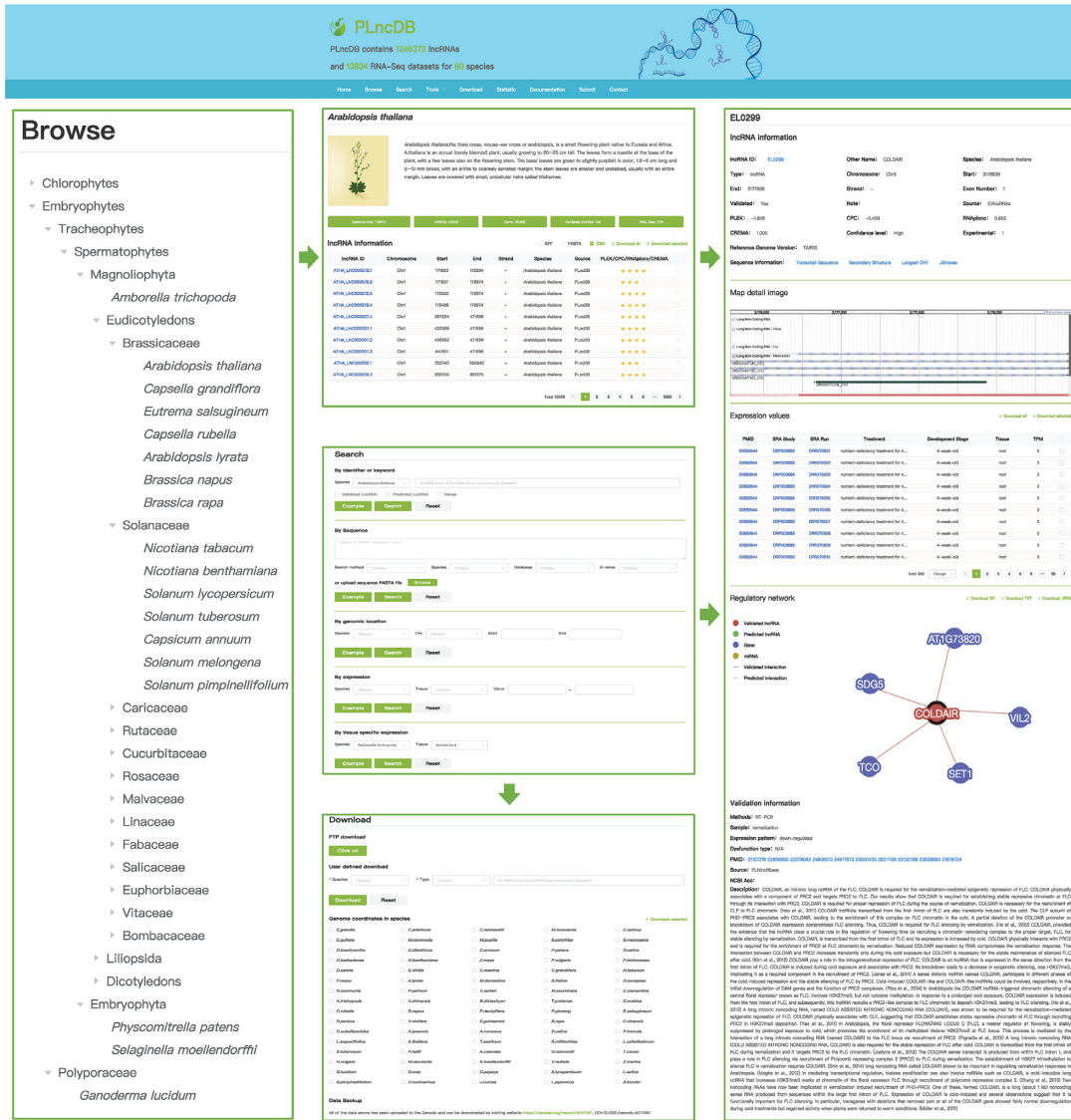
Figure 3. A schematic view of PLncDB V2.0 features.

IPS1 in *Arabidopsis*, using JBrowse (Figure 2B), we can easily find that *IPS1* is expressed in response to phosphate starvation. In addition, there are some epigenetic signals (H3K27me3) around its flanking region. For *enod40* candidate of tobacco, we find that in addition to expression in stem and leaf (Figure 2C), this lncRNA is also expressed in flower tissue, implying that it may also regulate flower development.

Download. Through the Download tag located in the toolbar, all basic information on lncRNAs in each species could be downloaded in bulk or for specific species. Users can download data using three ports on the Download page. First, by FTP link, users can browse all available information in bulk. Second, via the port of user-defined Download, users can choose the information they are interested in by selecting species and data types. Third, quick access is provided to users so that the genome coordinates of lncR-

NAs could be readily downloaded. Meanwhile, a backup copy of all data in PLncDB was also uploaded to Zenodo (https://zenodo.org/) with ID of 4017591.

Submit. In an effort to make PLncDB a one stop community resource, we have begun to accept submissions of plant lncRNAs with required information alone or in batch. All submitted lncRNA items will be processed by our standard procedure described in the Material and Methods. By this process, the submitted lncRNAs will be grouped into categories.

DISCUSSION

In the past two decades, great efforts have been made toward the identification, characterization and functional analysis of plant lncRNAs. In addition, abundant data for RNA-Seq and Chip-Seq have accumulated as a result of rapid

development of next generation sequencing techniques. Together, they have provided a solid foundation for annotating plant lncRNAs and, therefore, there is an urgent need to construct a comprehensive database to store and organize all knowledge relating to plant lncRNAs. Currently, CANTATAdb (21) and GreeNC (22) are the most comprehensive databases for plant lncRNA research, but the number of lncRNAs and annotated plant species is still far from sufficient due to the limited datasets they used. EVLncRNAs mainly focused on validated lncRNA candidates thereby limiting its wide-spread usage. Therefore, we set out to use a standardized pipeline with the latest plant lncRNA annotation criteria to annotate lncRNAs for all plant species with genome references and available RNA-Seq datasets (Figure 1B and Supplementary Table S4).

Compared to CANTATAdb/GreeNC/EVLncRNAs and other lncRNA databases, not only more plant species were annotated, but the RNA-Seq datasets and epigenetic information have also been improved in PLncDB V2.0 (Supplementary Tables S1 and S2). As shown in Figure 1, the number of annotated lncRNAs was improved substantially. One of the immediate consequences of this improvement is that users can now conduct comparative analysis of lncRNAs among different tissues and species (Figure 2).

In addition, several features for plant lncRNAs were provided in PLncDB V2.0. By normalizing expression levels across different tissues, development stages, or stress treatments, lncRNA expression patterns in species can be easily established (Figure 2A and C). Analysis of the epigenetic dataset clearly showed that lncRNA genes are regulated by methylation and histone modification (Figure 2B). Target collection and prediction provides information on the regulatory network for specific lncRNA and increases our understanding of plant lncRNAs based on available information (Figure 2D). From the tissue specific expression analysis function provided in PLncDB V2.0, we could find a lot of lncRNAs with tissue specific expression pattern for most of plant species (Figure 2A), which is consistent with other findings (43–46). Compared to protein coding genes, the tissue specific expression of lncRNAs provides important clues about their specific functions within tissues (2).

Meanwhile, PLncDB V2.0 provides a user-friendly interface to browse and access all data via multi-layer webpages, powerful search engines and download ports (Figure 3). At the same time, we plan to update and improve PLncDB semi-annually, by integrating data from other databases and researchers. For example, there are still some plants not covered by PLncDB V2.0 due to the huge amount of computation needed, which will be integrated in the next update. We will use the same standard pipeline for annotating new entries and add them into PLncDB. Hence, we believe PLncDB V2.0 will be a comprehensive functional database for data mining and database-driven research, and therefore a useful platform for the plant lncRNA community.

DATA AVAILABILITY

PLncDB V2.0 is freely available at http://plncdb.tobaccodb.org/.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Jin,J., Liu,J., Wang,H., Wong,L. and Chua,N.H. (2013) PLncDB: plant long non-coding RNA database. *Bioinformatics*, **29**, 1068–1071.
2. Sun,H.X. and Chua,N.H. (2019) Bioinformatics approaches to studying plant long noncoding RNAs (lncRNAs): Identification and functional interpretation of lncRNAs from RNA-Seq data sets. *Methods Mol. Biol.*, **1933**, 197–205.
3. Liu,J., Jung,C., Xu,J., Wang,H., Deng,S., Bernad,L., Arenas-Huertero,C. and Chua,N.H. (2012) Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. *Plant Cell*, **24**, 4333–4345.
4. Wang,H., Chung,P.J., Liu,J., Jang,I.C., Kean,M.J., Xu,J. and 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.
5. The,R.C. (2019) RNAcentral: a hub of information for non-coding RNA sequences. *Nucleic Acids Res.*, **47**, D1250–D1251.
6. Yates,A.D., Achuthan,P., Akanni,W., Allen,J., Allen,J., Alvarez-Jarreta,J., Amode,M.R., Armean,I.M., Azov,A.G., Bennett,R. *et al.* (2020) Ensembl 2020. *Nucleic Acids Res.*, **48**, D682–D688.
7. Huarte,M. (2015) The emerging role of lncRNAs in cancer. *Nat. Med.*, **21**, 1253–1261.
8. Paschoal,A.R., Lozada-Chavez,I., Domingues,D.S. and Stadler,P.F. (2018) ceRNAs in plants: computational approaches and associated challenges for target mimic research. *Brief. Bioinform.*, **19**, 1273–1289.
9. Seo,J.S., Sun,H.X., Park,B.S., Huang,C.H., Yeh,S.D., Jung,C. and Chua,N.H. (2017) ELF18-induced long-noncoding RNA associates with mediator to enhance expression of innate immune response genes in arabidopsis. *Plant Cell*, **29**, 1024–1038.
10. Seo,J.S., Diloknawarit,P., Park,B.S. and Chua,N.H. (2019) ELF18-induced long noncoding RNA 1 evicts fibrillarin from mediator subunit to enhance pathogenesis-related gene 1 (PR1) expression. *New Phytologist*, **221**, 2067–2079.
11. Sun,H.X., Li,Y., Niu,Q.W. and Chua,N.H. (2017) Dehydration stress extends mRNA 3′ untranslated regions with noncoding RNA functions in Arabidopsis. *Genome Res.*, **27**, 1427–1436.
12. Heo,J.B. and Sung,S. (2011) Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science*, **331**, 76–79.
13. Franco-Zorrilla,J.M., Valli,A., Todesco,M., Mateos,I., Puga,M.I., Rubio-Somoza,I., Leyva,A., Weigel,D., Garcia,J.A. and Paz-Ares,J. (2007) Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.*, **39**, 1033–1037.
14. Wu,H.W., Deng,S., Xu,H., Mao,H.Z., Liu,J., Niu,Q.W., Wang,H. and Chua,N.H. (2018) A noncoding RNA transcribed from the AGAMOUS (AG) second intron binds to CURLY LEAF and represses AG expression in leaves. *New Phytologist*, **219**, 1480–1491.
15. Henriques,R., Wang,H., Liu,J., Boix,M., Huang,L.F. and Chua,N.H. (2017) The antiphasic regulatory module comprising CDF5 and its

antisense RNA FLORE links the circadian clock to photoperiodic flowering. *New Phytologist*, **216**, 854–867.

16. Maracaja-Coutinho,V., Paschoal,A.R., Caris-Maldonado,J.C., Borges,P.V., Ferreira,A.J. and Durham,A.M. (2019) Noncoding RNAs databases: current status and trends. *Methods Mol. Biol.*, **1912**, 251–285.

17. Quek,X.C., Thomson,D.W., Maag,J.L., Bartonicek,N., Signal,B., Clark,M.B., Gloss,B.S. and Dinger,M.E. (2015) lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.*, **43**, D168–D173.

18. Fang,S., Zhang,L., Guo,J., Niu,Y., Wu,Y., Li,H., Zhao,L., Li,X., Teng,X., Sun,X. *et al.* (2018) NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.*, **46**, D308–D314.

19. Zhou,B., Zhao,H., Yu,J., Guo,C., Dou,X., Song,F., Hu,G., Cao,Z., Qu,Y., Yang,Y. *et al.* (2018) EVLncRNAs: a manually curated database for long non-coding RNAs validated by low-throughput experiments. *Nucleic Acids Res.*, **46**, D100–D105.

20. Xuan,H., Zhang,L., Liu,X., Han,G., Li,J., Li,X., Liu,A., Liao,M. and Zhang,S. (2015) PLNlncRbase: a resource for experimentally identified lncRNAs in plants. *Gene*, **573**, 328–332.

21. Szczesniak,M.W., Bryzghalov,O., Ciomborowska-Basheer,J. and Makalowska,I. (2019) CANTATAdb 2.0: expanding the collection of plant long noncoding RNAs. *Methods Mol. Biol.*, **1933**, 415–429.

22. Paytuvi Gallart,A., Hermoso Pulido,A., Anzar Martinez de Lagran,I., Sanseverino,W. and Aiese Cigliano,R. (2016) GREENC: a Wiki-based database of plant lncRNAs. *Nucleic Acids Res.*, **44**, D1161–D1166.

23. Buels,R., Yao,E., Diesh,C.M., Hayes,R.D., Munoz-Torres,M., Helt,G., Goodstein,D.M., Elsik,C.G., Lewis,S.E., Stein,L. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.

24. Winter,D., Vinegar,B., Nahal,H., Ammar,R., Wilson,G.V. and Provart,N.J. (2007) An 'Electronic Fluorescent Pictograph' browser for exploring and analyzing large-scale biological data sets. *PLoS One*, **2**, e718.

25. Benson,D.A., Cavanaugh,M., Clark,K., Karsch-Mizrachi,I., Ostell,J., Pruitt,K.D. and Sayers,E.W. (2018) GenBank. *Nucleic Acids Res.*, **46**, D41–D47.

26. Goodstein,D.M., Shu,S., Howson,R., Neupane,R., Hayes,R.D., Fazo,J., Mitros,T., Dirks,W., Hellsten,U., Putnam,N. *et al.* (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.*, **40**, D1178–D1186.

27. Fernandez-Pozo,N., Menda,N., Edwards,J.D., Saha,S., Tecle,I.Y., Strickler,S.R., Bombarely,A., Fisher-York,T., Pujar,A., Foerster,H. *et al.* (2015) The sol genomics network (SGN)–from genotype to phenotype to breeding. *Nucleic Acids Res.*, **43**, D1036–D1041.

28. Portwood,J.L. 2nd, Woodhouse,M.R., Cannon,E.K., Gardiner,J.M., Harper,L.C., Schaeffer,M.L., Walsh,J.R., Sen,T.Z., Cho,K.T., Schott,D.A. *et al.* (2019) MaizeGDB 2018: the maize multi-genome genetics and genomics database. *Nucleic Acids Res.*, **47**, D1146–D1154.

29. Kawahara,Y., de la Bastide,M., Hamilton,J.P., Kanamori,H., McCombie,W.R., Ouyang,S., Schwartz,D.C., Tanaka,T., Wu,J., Zhou,S. *et al.* (2013) Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice*, **6**, 4.

30. Johnson,C., Conrad,L.J., Patel,R., Anderson,S., Li,C., Pereira,A. and Sundaresan,V. (2018) Reproductive long intergenic noncoding RNAs exhibit male gamete specificity and polycomb repressive complex 2-Mediated repression. *Plant Physiol.*, **177**, 1198–1217.

31. Kim,D., Langmead,B. and Salzberg,S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*, **12**, 357–360.

32. Pertea,M., Pertea,G.M., Antonescu,C.M., Chang,T.C., Mendell,J.T. and Salzberg,S.L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.*, **33**, 290–295.

33. Kong,L., Zhang,Y., Ye,Z.Q., Liu,X.Q., Zhao,S.Q., Wei,L. and Gao,G. (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic. Acids Res.*, **35**, W345–W349.

34. Li,A., Zhang,J. and Zhou,Z. (2014) PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics*, **15**, 311.

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

36. Simopoulos,C.M.A., Weretilnyk,E.A. and Golding,G.B. (2018) Prediction of plant lncRNA by ensemble machine learning classifiers. *BMC Genomics*, **19**, 316.

37. Jin,J., Zhang,H., Zhang,J., Liu,P., Chen,X., Li,Z., Xu,Y., Lu,P. and Cao,P. (2017) Integrated transcriptomics and metabolomics analysis to characterize cold stress responses in Nicotiana tabacum. *BMC Genomics*, **18**, 496.

38. Kadota,K., Ye,J., Nakai,Y., Terada,T. and Shimizu,K. (2006) ROKU: a novel method for identification of tissue-specific genes. *BMC Bioinformatics*, **7**, 294.

39. Wu,H.J., Ma,Y.K., Chen,T., Wang,M. and Wang,X.J. (2012) PsRobot: a web-based plant small RNA meta-analysis toolbox. *Nucleic Acids Res.*, **40**, W22–W28.

40. Wu,H.J., Wang,Z.M., Wang,M. and Wang,X.J. (2013) Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants. *Plant Physiol.*, **161**, 1875–1884.

41. Mann,M., Wright,P.R. and Backofen,R. (2017) IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. *Nucleic Acids Res.*, **45**, W435–W439.

42. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.

43. Bhatia,G., Goyal,N., Sharma,S., Upadhyay,S.K. and Singh,K. (2017) Present scenario of long non-coding RNAs in plants. *Non-coding RNA*, **3**, 16.

44. Wang,H., Niu,Q.W., Wu,H.W., Liu,J., Ye,J., Yu,N. and Chua,N.H. (2015) Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved lncRNAs associated with agriculture traits. *Plant J.*, **84**, 404–416.

45. Liu,J., Wang,H. and Chua,N.H. (2015) Long noncoding RNA transcriptome of plants. *Plant Biotechnol. J.*, **13**, 319–328.

46. Lv,Y., Liang,Z., Ge,M., Qi,W., Zhang,T., Lin,F., Peng,Z. and Zhao,H. (2016) Genome-wide identification and functional prediction of nitrogen-responsive intergenic and intronic long non-coding RNAs in maize (Zea mays L.). *BMC Genomics*, **17**, 350.