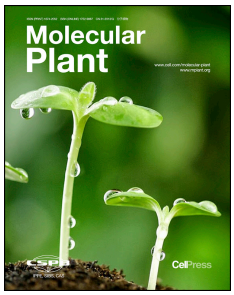
Accepted Manuscript



PlantcircBase: a database for plant circular RNAs

Qinjie Chu, Xingchen Zhang, Xintian Zhu, Chen Liu, Lingfeng Mao, Chuyu Ye, Qian-Hao Zhu, Longjiang Fan

PII: S1674-2052(17)30074-6
DOI: 10.1016/j.molp.2017.03.003
Reference: MOLP 442

To appear in: *MOLECULAR PLANT*
Accepted Date: 7 March 2017

Please cite this article as: **Chu Q., Zhang X., Zhu X., Liu C., Mao L., Ye C., Zhu Q.-H., and Fan L.** (2017). PlantcircBase: a database for plant circular RNAs. Mol. Plant. doi: 10.1016/j.molp.2017.03.003.

PlantcircBase: a database for plant circular RNAs

Dear Editor,

Circular RNAs (circRNAs) are a kind of non-coding RNAs which were first found over 35 years ago but become popular in recent years. They are covalently closed loops derived from back-splicing of precursor mRNAs. Based on their genomic location, circRNAs can be divided into exonic, intronic, intergenic and exon-intronic circRNAs (Chen, 2016). With the help of bioinformatics algorithms, a large number of circRNAs have been detected among diverse organisms, including plants like *Arabidopsis thaliana* and *Oryza sativa* (Memczak et al., 2013; Ye et al., 2015; Chen, 2016). CircRNAs have been shown to function as miRNA sponges to sequester miRNAs and play an essential role in transcriptional regulation (Memczak et al., 2013). A database of circRNAs is of great importance for studying their biogenesis and functions. Recently, several circRNA databases for humans and animals (e.g. circBase, circRNADb and circNet) have been established with functions of searching and browsing (Glažar et al., 2014; Chen et al., 2016; Liu et al., 2016). However, no such database is now available for plants.

Here, we created a database of plant circRNAs (termed as PlantcircBase; http://ibi.zju.edu.cn/plantcircbase/) for plant researchers. We have collected publicly available circRNAs identified in recent years by bioinformatics prediction and/or experimental validation from *O. sativa*, *A. thaliana*, *Zea mays*, *Solanum lycopersicum* and *Hordeum vulgare* (Wang et al, 2014; Ye et al., 2015; Lu et al., 2015; Ye et al., 2016; Supplemental Table 1), as well as unpublished circRNAs newly identified by this study. Based on the collected circRNAs, we further predicted those putatively acting as miRNA sponges and their potential networks involving circRNA-miRNA-mRNA in the corresponding organisms. Our database also provides other functions such as visualization of circRNA structures based on their genomic position and prediction of circRNAs from query sequences generated by Sanger sequencing or high-throughput sequencing. In order to provide the latest developments of circRNAs in plants, we will update PlantcircBase when newly identified plant circRNAs are available, and also plan to add new functionalities in the database in future, such as functions and biogenesis mechanisms of circRNAs.

First, PlantcircBase includes a total of 77,595 unique circRNAs from diverse tissues of five organisms (40,314, 35,884, 496, 854 and 47 for *O. sativa*, *A. thaliana*, *Z. mays*, *S. lycopersicum* and *H. vulgare*, respectively). Of the 77,595 circRNAs, 178 back-splicing sites and 70 full length sequences have been experimentally validated by Sanger sequencing of divergent PCR products (Figure 1A, Supplemental Table 1), 1,861 were predicted as putative miRNA sponges. In addition, 1,335

circRNA-miRNA-mRNA networks were constructed for the circRNAs included in the database.

Second, detailed and comprehensive information are displayed for each circRNA entry in PlantcircBase (Figure 1C). Except for general information, such as the circRNA ID, circRNA aliases used in other publications, organism, position in genome, parent gene and its annotation, and other useful information that are specific for plant circRNAs have also been presented. For example, our previous study showed that circRNAs in plants are generated with diverse non GT-AG splicing signals (Ye et al., 2016), while most exonic circRNAs in humans contain the canonical splicing signals. Therefore, in addition to the information of whether or not the splicing sites are on the exonic boundaries, splicing signals of each circRNA are also provided. For each circRNA entry, additional circRNAs overlapping or sharing the parent gene with it are considered as products of alternative splicing, and presented under the category “Alternative splicing”. Furthermore, PlantcircBase provides the back-splicing junction sequences of circRNAs, of which the uppercase and lowercase sequences represent the two sides of a back-splicing site, respectively. PlantcircBase also provides the total number of RNA-seq reads in different tissues that support the back-splicing site. Importantly, two pieces of evidences for the presence of a circRNA, i.e. experimental validation results of the back-splicing site (together with the PCR primers) and the full length circRNA sequence, are presented in PlantcircBase if such information is available. Additionally, the conservation of circRNAs among the five organisms has been determined based on sequence similarity and orthologous relationship of their parent genes (for details see Supplemental Information). Finally, we have predicted miRNA sponges of circRNAs using eTM_finder (Ye et al., 2014), and targets of the miRNA acting as sponges using psRNATarget (http://plantgrn.noble.org/psRNATarget/). Such information was used to predict potential networks involving circRNA, miRNAs and their target mRNAs.

Third, diverse bioinformatics tools are available at PlantcircBase: (1) basic tools including browsing, searching (Figure 1D) and downloading: the browsing tool provides a glance over the whole database; the searching function provides tool to find circRNAs using keywords (e.g. ID, parent gene name, sponge miRNAs), sequence comparison, or conditional filtering; the detailed information of a circRNA or all entries in PlantcircBase can be downloaded using the download tool. (2) advanced tools, including visualizing structures of circRNAs and predicting circRNAs: all circRNAs in the database can be visualized in a “circular format”, a “linear format” or their relationship with the parent gene (Figure 1E, 1F). Circular format shows a circRNA with the two back-splicing sites being stick together and gene structures color coded. The exact position of the circRNA in its parent gene (Figure 1F, the grey area) and its flanking genes are also shown. The predicting tool, which is based on a

new bioinformatics algorithm (our unpublished data), can be used to predict whether or not a query sequence forms a circRNA. If the answer is yes, it will further tell users whether or not the circRNA has a match in PlantcircBase.

In summary, we aim to track and assemble all plant circRNAs documented in the latest publications and provide researchers with the most comprehensive information about plant circRNAs, especially the experimentally validated ones, in a user friendly platform. We believe that PlantcircBase will be a useful tool set facilitating functional studies of circRNAs in plants.

Figure 1. Summary and functions of PlantcircBase.

(A) Summary of circRNAs from the five plant organisms in PlantcircBase.

(B) Homepage of PlantcircBase.

(C) An example entry of circRNA (ath_circ_025680) in PlantcircBase (Supplemental Figure 1).

(D) Searching page of PlantcircBase, which can be searched using keyword, subset information and sequences.

(E) Visualization of a circRNA (osa_circ_026780, whose position in genome is Chr5: 3768053-3768717). The top and bottom panels show the circular and the linear format of a circRNA, respectively. Different genomic components (CDS, intron, exon and UTR) are color coded. The arrow line represents the direction of transcripts. The vertical bar with two flanking numbers represents the back-splicing site of the circRNA.

(F) Visualization of the parent gene of a circRNA (osa_circ_026780). The grey area represents the exact position of the circRNA within its parent gene.

SUPPLEMENTAL INFORMATION

Supplemental Information is available at **.

AUTHOR CONTRIBUTIONS

Conceptualization, Q.C., C.Y. and L.F.; Methodology, Q.C. and X.C.Z.; Validation, C.L. and C.Y.; Formal Analysis, Q.C. and X.C.Z.; Investigation, L.M. and X.T.Z.; Resources, C.L., C.Y., L.M. and X.T.Z.; Visualization, Q.C. and X.C.Z.; Writing - Original Draft, Q.C.; Writing - Review & Editing, C.Y., Q.H.Z. and L.F.; Supervision, C.Y., Q.H.Z. and L.F.

FUNDING

This work was supported by the National Science Foundation of China (91435111), the National Basic Research Program of China (2015CB150200), Shanghai Municipal Commission of Agriculture (2014-7-1-4), and Jiangsu Collaborative Innovation Center for Modern Crop Production (JCIC-MCP).

Qinjie Chu[1,*], Xingchen Zhang[1,*], Xintian Zhu[1], Chen Liu[1], Lingfeng Mao[1], Chuyu Ye[1], Qian-Hao Zhu[2], Longjiang Fan[1]

[1]Institute of Crop Science and Institute of Bioinformatics, Zhejiang University, Hangzhou 310058, China

[2]CSIRO Agriculture and Food, Canberra, ACT 2601, Australia

[*]These authors contributed equally to this work

Corresponding author:
Prof. Dr. Longjiang Fan
Phone: +86-571-88982730
E-mail: fanlj@zju.edu.cn

REFERENCES

Chen, L.L. (2016). The biogenesis and emerging roles of circular RNAs. Nat. Rev. Mol. Cell Biol. 17, 205-211.

Chen, X.P., Han, P., Zhou, T., Guo, X.J., Song, X.F. and Li, Y. (2016). circRNADb: A comprehensive database for human circular RNAs with protein-coding annotations. Sci. Rep. 6, 34985.

Glažar, P., Papavasileiou, P., Rajewsky, N. (2014). circBase: a database for circular RNAs. RNA 20, 1666-1670

Liu, Y.C., Li, J.R., Sun, C.H., Andrews, E., Chao, R.F., Lin, F.M., Weng, S.L., Hsu, S. Da, Huang, C.C., Cheng, C., et al. (2016). CircNet: A database of circular RNAs derived from transcriptome sequencing data. Nucleic Acids Res. 44, D209-D215.

Lu, T.T., Cui, L.L., Zhou, Y., Zhu, C.R., Fan, D.L., Gong, H., Zhao, Q., Zhou, C.C., Zhao, Y., Lu, D.F., et al. (2015). Transcriptome-wide investigation of circular RNAs in rice. RNA 21(12), 2076-2087

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L.,

Mackowiak, S.D., Gregersen, L.H., Munschauer, M., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495, 333–338.

Wang, P.L., Bao, Y., Yee, M.C., Barrett, S.P., Hogan, G.J., Olsen, M.N., Dinneny, J.R., Brown, P.O., and Salzman, J. (2014). Circular RNA is expressed across the eukaryotic tree of life. PLoS One 9(6). e90859.

Ye, C.Y., Chen, L., Liu, C., Zhu, Q.H., and Fan, L.J. (2015). Widespread noncoding circular RNAs in plants. New Phytol. 208, 88-95.

Ye, C.Y., Xu, H., Shen, E.H., Liu, Y., Wang, Y., Shen, Y.F., Qiu, J., Zhu, Q.H. and Fan, L.J. (2014). Genome-wide identification of non-coding RNAs interacted with microRNAs in soybean. Front. Plant Sci. 5, 743.

Ye, C.Y., Zhang, X.C., Chu, Q.J., Liu, C., Yu, Y.Y., Jiang, W.Q., Zhu, Q.H., Fan, L.J. and Guo, L.B. (2016). Full-length sequence assembly reveals circular RNAs with diverse non-GT/AG splicing signals in rice. RNA Biol. 0, 1-9.