Plant Communications
Resource Article

CellPress
Partner Journal

PlantcircBase 7.0: Full-length transcripts and conservation of plant circRNAs

Xiaoxu Xu[1,2], Tianyu Du[1,2], Weihua Mao[3], Xiaohan Li[1], Chu-Yu Ye[1], Qian-Hao Zhu[4], Longjiang Fan[1,2] and Qinjie Chu[1,*]

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

[2]Shandong (Linyi) Institute of Modern Agriculture of Zhejiang University, Linyi 310014, China

[3]Analysis Center of Agrobiology and Environmental Science, Zhejiang University, Hangzhou 310058, China

[4]CSIRO Agriculture and Food, Black Mountain Laboratories, Canberra, ACT 2601, Australia

***Correspondence: Qinjie Chu (qinjiechu@zju.edu.cn)**

https://doi.org/10.1016/j.xplc.2022.100343

ABSTRACT

Circular RNA (circRNA) is a special type of non-coding RNA that participates in diverse biological processes in both animals and plants. Five years ago, we developed a comprehensive plant circRNA database (PlantcircBase), which has attracted much attention from the plant circRNA community. Here, we report an updated PlantcircBase (v.7.0), which contains 171,118 circRNAs from 21 plant species. Over 31,000 of the circRNAs have full-length sequences constructed based on analysis of 749 bulk RNA sequencing (RNA-seq) datasets downloaded from the public domain and Nanopore long-read sequencing results of rice RNAs newly generated in this study. A plant multiple conservation score (PMCS), based on the conservation of both sequence and expression profiles, was calculated for each circRNA to quantify and compare the conservation of all circRNAs. A new parameter, plant circRNA confidence level (PCCL), is introduced to measure the identity reliability of each circRNA based on experimental validation results and the number of references that support the circRNA. All this information and other details of circRNAs can be browsed, searched, and downloaded from PlantcircBase 7.0, which also provides online bioinformatics tools for visualization and sequence alignment. PlantcircBase 7.0 is publicly and freely accessible at http://ibi.zju.edu.cn/plantcircbase/.

Key words: plant circRNA, PlantcircBase, conservation, full-length transcripts, long-read sequencing

Xu X., Du T., Mao W., Li X., Ye C.-Y., Zhu Q.-H., Fan L., and Chu Q. (2022). PlantcircBase 7.0: Full-length transcripts and conservation of plant circRNAs. Plant Comm. **3**, 100343.

INTRODUCTION

Circular RNAs (circRNAs) are single-stranded and covalently closed non-coding RNAs that have been documented in many animals and plants (Memczak et al., 2013; Ye et al., 2015). The majority of the currently available circRNAs were identified using RNA sequencing (RNA-seq) results generated by second-generation high-throughput short-read sequencing technology. Full-length sequences of circRNAs can be predicted based on reads that support back-splicing sites using bioinformatics tools such as CIRI-full (Zheng et al., 2019), CIRCexplorer (Zhang et al., 2016), and circseq_cup (Ye et al., 2017). In the past several years, third-generation long-read RNA-seq technology has been used in the identification of circRNAs (Rahimi et al., 2021a, 2021b). The major advantage of using this new technology is that it produces full-length circRNA transcripts without assembly, so that different isoforms of a circRNA derived from alternative splicing events can be accurately identified. For example, CIRI-long (Zhang et al., 2021b) and isoCirc (Xin et al., 2021) protocols have been developed based on Nanopore sequencing to identify full-length transcripts of circRNAs in human and mouse, respectively, but long-read sequencing has not yet been applied to plant circRNA identification.

Another recent development in RNA-seq is high-throughput single-cell RNA-seq (scRNA-seq) technology (Macosko et al., 2015), which can simultaneously sequence the transcriptomes of tens of thousands of cells and uncover transcripts with low expression and cell-specific transcripts. Many transcriptional atlases of mRNAs at a single-cell resolution have been constructed for both animals and plants (Han et al., 2018, 2020; Zhang et al., 2019,

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

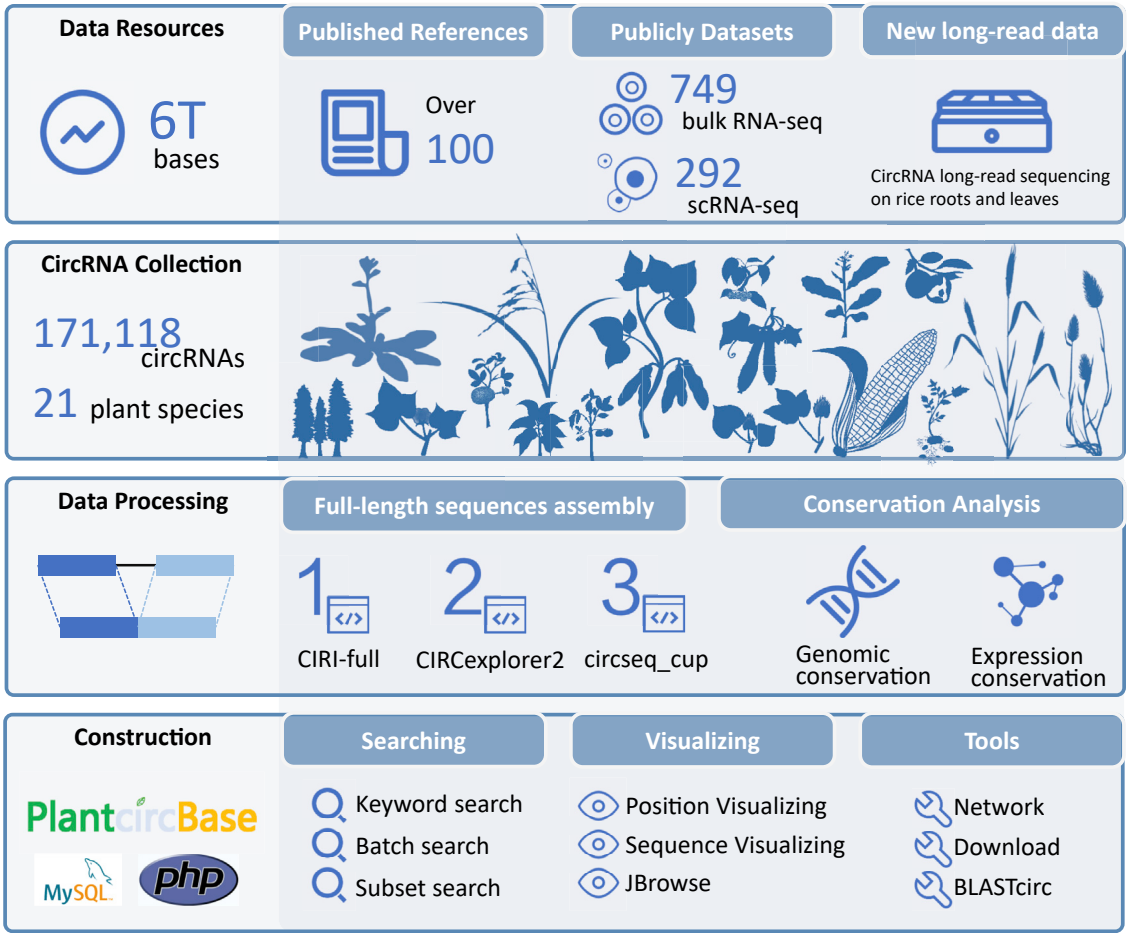

Figure 1. Data resources, construction, and features of PlantcircBase 7.0.

2021a). By contrast, a comprehensive atlas of non-coding RNAs at a single-cell resolution is still rarely seen, owing to difficulties in the enrichment of specific RNAs and library construction, despite the use of SUPeR-seq (Fan et al., 2015) for identification of linear RNAs and circRNAs at the single-cell level.

Several plant circRNA databases have been developed, such as PlantcircBase (Chu et al., 2017), PlantCircNet (Zhang et al., 2017), CircFunBase (Meng et al., 2019), GreenCircRNA (Zhang et al., 2020), CropCircDB (Wang et al., 2019), and AtCircDB (Ye et al., 2019). Each database has its own pros and cons; the major issues are the limited number of plant species and/or collected circRNAs and the fact that the databases are rarely updated after being released. We developed PlantcircBase in 2017 (Chu et al., 2017) and have updated the database six times by adding newly reported plant circRNAs, as well as new features and tools.

In this study, we further updated PlantcircBase (v.7.0) by adding three new features: full-length sequences of plant circRNAs, quantification of circRNA conservation, and expression patterns of circRNAs in different tissues or at a single-cell resolution. Full-length circRNA transcripts and expression patterns were generated by reanalyzing over 6 terabytes of data from 749 paired-end bulk RNA-seq datasets from 12 plant species and 292 scRNA-seq datasets from *Arabidopsis thaliana*. Full-length rice circRNAs were also obtained from newly generated Nanopore long-read sequencing data. The new information will be valuable for further functional investigation of plant circRNAs.

RESULTS

Construction and overall features of the updated PlantcircBase 7.0

To obtain a comprehensive set of plant circRNAs, almost all publicly available circRNAs, including 171,118 circRNAs from 21 plant species (as of December 31, 2021), were collected and included in PlantcircBase 7.0 (Figure 1). Back-splicing sites of these circRNAs, which were identified by CIRI (Gao et al., 2015, 2018), CIRCexplorer (Zhang et al., 2016), find_circ (Memczak et al., 2013), or other well-known tools, were collected from publications. The internal structures of these circRNAs were newly identified in this study using CIRI-full (Zheng et al., 2019), CIRCexplorer (Zhang et al., 2016), or circseq_cup (Ye et al., 2017).

To investigate the expression patterns and assemble full-length transcripts of plant circRNAs, a total of 1,041 transcriptomic datasets (749 bulk RNA-seq and 292 scRNA-seq datasets; Supplemental Table 1) from different plant tissues were re-analyzed. In addition, Nanopore long-read sequencing was

performed to directly sequence full-length rice circRNA transcripts according to the CIRI-long sequencing protocol (Zhang et al., 2021b). Based on the genome locations of the collected circRNAs, each circRNA was assigned an ID in PlantcircBase so that its detailed information, such as type, parental gene, junction sequences, genomic sequences, alternative back-splicing events, splicing signals, expression pattern, microRNA (miRNA)-binding information, coding potential, full-length sequence structure, references, and so forth, could be obtained by ID searching. Conservation of each circRNA was quantified by calculating the plant multiple conservation score (PMCS) (Chu et al., 2022) based on its expression conservation and genomic conservation.

To make PlantcircBase a user-friendly and comprehensive online database, it was established based on MySQL and PHP. Online visualization tools such as Visualize, BLASTcirc,Jbrowse, and Network were used to visualize plant circRNAs by position, by sequence, at the whole-genome level, and by potential interactions among circRNAs, miRNAs, and mRNAs.

Full-length transcripts of plant circRNAs

Most circRNAs were identified based on reads that supported the back-splicing sites, and their internal splicing events were largely unknown. Full-length transcripts of circRNAs are crucial for deeply understanding their internal structures and alternative splicing events, which may be important for their biological functions. Bioinformatics tools (Zhang et al., 2016; Ye et al., 2017; Zheng et al., 2019) can be used to reconstruct the internal sequence structures of circRNAs using paired-end reads that support the back-splicing sites of the circRNAs. We thus reconstructed about 30,000 full-length circRNA transcripts (about 17.5% of all circRNAs collected in PlantcircBase; Supplemental Table 2) using 749 bulk RNA-seq datasets with three tools (Figure 2A; details in materials and methods). Recent studies have reported that full-length transcripts of human and mouse circRNAs could be generated by long-read sequencing technology such as CIRI-long (Zhang et al., 2021b) and isoCirc (Xin et al., 2021). To investigate the effectiveness of long-read sequencing technology for obtaining plant circRNAs, we performed Nanopore sequencing using rice roots and leaves (details in materials and methods) according to the CIRI-long sequencing protocol (Zhang et al., 2021b). We generated a total of 3,643 full-length rice circRNAs, which can easily be accessed in the "Download" section of PlantcircBase.

By comparing the assembled/sequenced full-length transcripts of plant circRNAs with their corresponding genomic sequences, the circRNAs were grouped into three different types (Figure 2B): "same", "bases_skip", and "alternative". Interestingly, a certain number of circRNAs had alternative splicing events supported by full-length circRNA evidence (i.e., type 3, "alternative" circRNAs; Supplemental Figure 1). For example, 681 (5.21%) and 992 (7.33%) circRNAs had at least two transcripts in *O. sativa* and *A. thaliana*, respectively, and the number of transcripts ranged from 2 to 60 in rice and from 2 to 51 in *Arabidopsis*.

Of the *O. sativa* circRNAs that had full-length sequence information, 9,440 (72.15%), 3,603 (27.54%), and 40 (0.31%) were assembled by bioinformatics tools, uncovered by long-read sequencing, or both, respectively (Figure 2C). In other plant species, the percentage of circRNAs with full-length information ranged from about 1% to 37% (Supplemental Table 2). The transcript length of most circRNAs ranged consistently from 100 to 300 bp in different plant species (about 60% of all circRNA transcripts in *O. sativa* and *A. thaliana*; Figures 2D and Supplemental Figure 2), in line with previously reported results (Ye et al., 2017). There was a significant difference between the length distribution of circRNA transcripts and the length of their genomic sequences in rice (Wilcoxon rank sum test, *P*-Value < 0.01; Figure 2D, blue and red areas).

Conservation and expression patterns of plant circRNAs

Given the association between sequence conservation and biological function conservation, we quantified the conservation of plant circRNAs using the PMCS (Chu et al., 2022) based on 749 bulk RNA-seq datasets and comparative genomic analysis results of 12 plant species with well-assembled reference genomes. The PMCS is an integrated measurement of genomic and expression conservation with a value up to two; the higher the value, the more conserved the circRNA of interest (Chu et al., 2022). Most circRNAs deposited in PlantcircBase 7.0 have a PMCS of 0.2 to 1.0 (Figure 3A), with 75 circRNAs having a PMCS larger than 1.5. Thus, only a very small proportion of circRNAs (75/171,118) are relatively highly conserved in different plant species, consistent with the notion of the recent origin of plant circRNAs (Chu et al., 2022). As examples, two highly conserved circRNAs (with a PMCS >1.4), one from *O. sativa* (osa_circ_038222, chr9: 7545–7862) and another from *A. thaliana* (ath_circ_029811, chr4: 2719063–2719290), are shown in Figure 3B. Both circRNAs have orthologous genomic sequences in almost all investigated plant species and were detected in almost all collected RNA-seq datasets. Alternative back-splicing events were found in both circRNAs, and at least one alternatively spliced circRNA (e.g., osa_circ_038214, chr9: 7446–7764 in *O. sativa* and ath_circ_029812, chr4: 2719063–2719462 in *A. thaliana*) was experimentally validated. These results showed that the PMCS is a reliable measurement of circRNA conservation and can be used to shortlist conserved circRNAs for functional investigations.

Expression patterns of the circRNAs were further explored in both bulk and scRNA-seq datasets. Expression levels of circRNAs were based on the number of reads that supported the back-splicing sites of the circRNAs (details in materials and methods). Some circRNAs were found to be tissue specific in both *A. thaliana* (Figure 3C) and other plants (Supplemental Figures 3 and 4). Although scRNA-seq investigation is still very limited in plants, cell-type-specific expression of some circRNAs was discovered in the *A. thaliana scRNA-seq datasets collected in this study* (Figure 3D).

Among the large number of circRNAs collected in PlantcircBase, the majority lack experimental supporting evidence, and their circRNA identities are thus still in question. Here, we introduced a confidence level score, the plant circRNA confidence level (PCCL), based on the number of references that supported a circRNA and whether or not it is supported by experimental evidence, to evaluate the confidence of the circRNA identity. In general, plant circRNAs with high PCCL scores are relatively highly conserved, i.e., they have a higher PMCS value (Supplemental Figure 5).

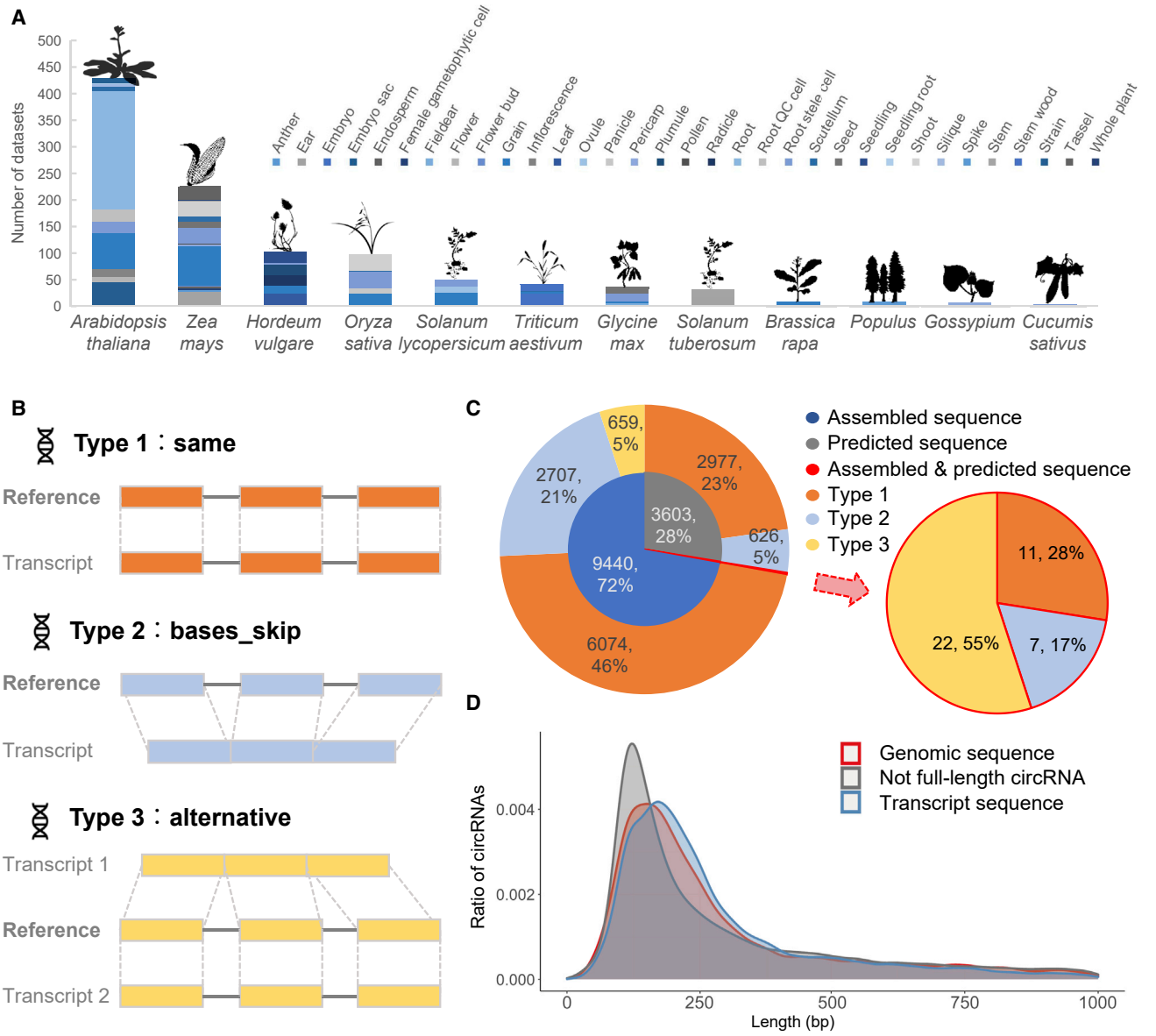

Figure 2. Full-length transcripts of the circRNAs in PlantcircBase 7.0.

(A) Number of datasets used in this study from different tissues of different plant species. Different colors represent different tissues.

(B) Plant circRNAs were separated into three different types (“same”, “bases_skip”, and “alternative”) by comparing the transcript sequence with the genomic sequence. Type 1 (same) represents circRNAs with identical transcript and genomic sequences, type 2 (bases_skip) represents circRNAs that have spliced introns or gaps in their transcript sequences compared with their genomic sequences, and type 3 (alternative) represents circRNAs with a combination of type 1 and type 2 features. Blocks and lines represent exons and introns, respectively.

(C) Percentage of rice circRNAs assembled by bioinformatics tools (blue), sequenced by long-read sequencing (gray), and obtained by both methods (red). Orange, light blue, and yellow represent types 1, 2, and 3 mentioned in (B).

(D) Transcript length distribution (blue) and genomic length distribution (red) of the assembled rice circRNAs and genomic length distribution of the non-assembled (gray) rice circRNAs.

DISCUSSION

The most comprehensive collection of plant circRNAs

PlantcircBase 7.0 contains 171,118 circRNAs reported from 2015 to the end of 2021 in 21 plant species. Many of these species have multiple reference genomes that have been used by different researchers in the identification of circRNAs. To unify the circRNAs from different publications, for each plant species, we selected a well-assembled and annotated version of the reference genome to align all circRNAs from the species using BLASTN with 100% sequence identity so that all circRNAs from the same species had comparable genomic coordinates. Furthermore, for 12 plant species, RNA-seq datasets from different tissues were collected and used to assemble full-length circRNA transcripts. Therefore, the circRNAs deposited in PlantcircBase 7.0 are not a simple collection of reported circRNAs but are instead a unified resource of plant circRNAs.

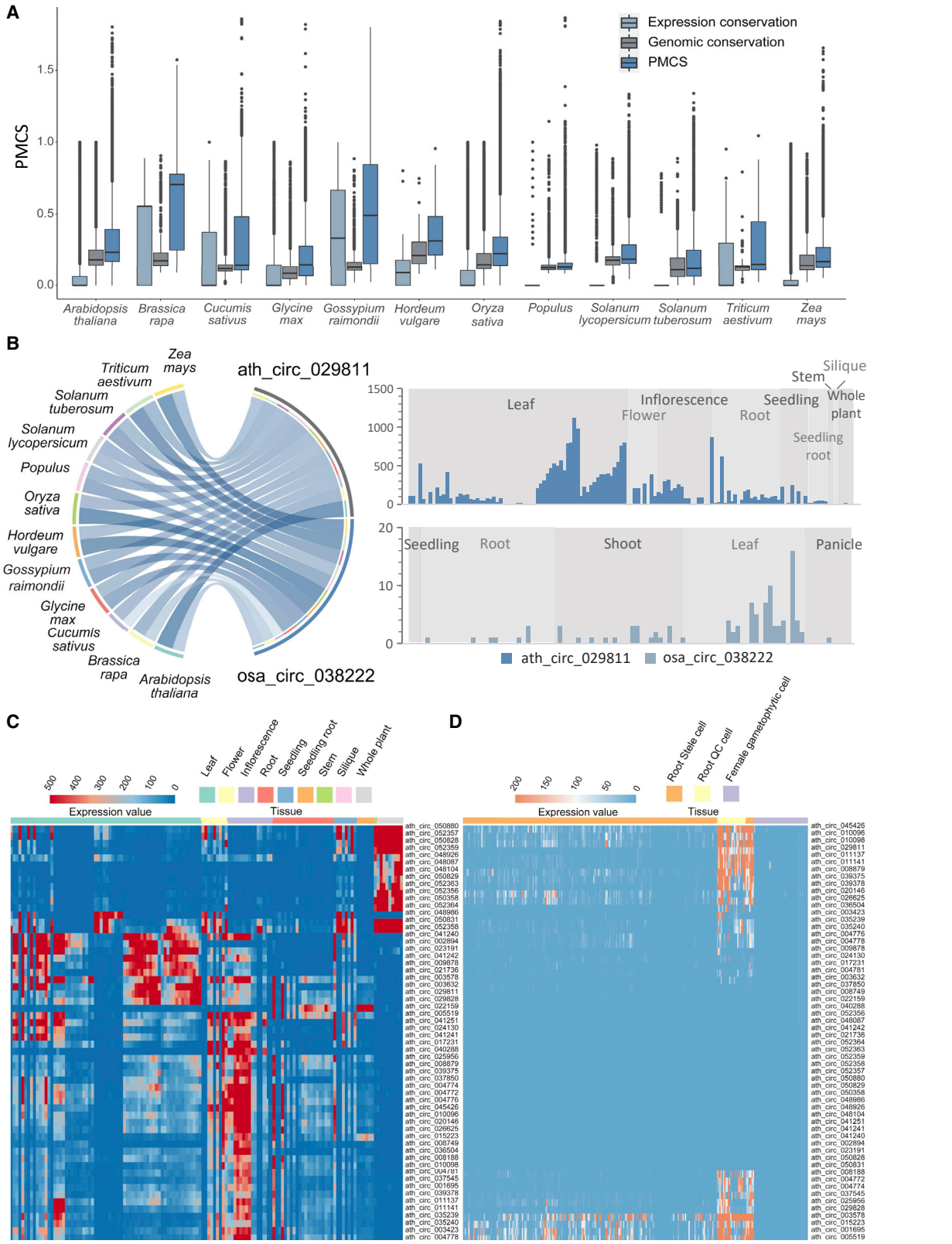

Figure 3. Conservation and expression patterns of plant circRNAs in different tissues and cell types.

(A) PMCS of circRNAs from 12 plant species. Light blue, gray, and blue boxes represent expression conservation, genomic conservation, and PMCS, respectively.

(legend continued on next page)

PlantcircBase 7.0 provides a user-friendly interface that enables users to search via keywords (e.g., ID assigned by PlantcircBase 7.0, parental gene name, targeted miRNA name, etc.), browse by plant species, and download circRNA information. For each circRNA, information such as genomic position, genomic sequence, full-length transcripts, parental gene, experimental evidences, alternative splicing events, splicing signals, original references, and so forth, can be displayed online. A genome browser is provided to show the back-splicing sites of the circRNAs with a zoom-in/-out window. Other tools such as visualization and sequence alignment are also available. By providing a resource and analytical tools that are as comprehensive as possible, we hope PlantcircBase can serve as a one-stop site where the plant circRNA community can find all information on plant circRNAs.

Investigation of plant circRNAs at a single-cell resolution

Most circRNAs are expressed at low levels, and their expression may be cell specific. Traditional bulk RNA-seq analysis has difficulty in accurately characterizing these circRNAs. Even for relatively highly expressed cell-specific circRNAs, elucidation of their expression patterns is compromised when highly heterogeneous tissue is used for RNA-seq. scRNA-seq can reveal the transcriptomic landscape at a single-cell resolution and quantify cell-to-cell variation in different forms of transcripts, including circRNAs, from the same gene. It is thus an ideal tool for the identification of cell-specific circRNAs. For instance, using scRNA-seq, circRNA generated from the *ASXL1* locus in human was found to show significant variation among different cells and was the predominant transcript in only some cells (Koh et al., 2016). We attempted to use single-cell transcriptomic datasets generated in *A. thaliana* (Efroni et al., 2015; Song et al., 2020) to explore the expression profiles of circRNAs in different cell populations. The limitation of using published scRNA-seq data is that they were usually generated using poly(A)-enriched RNA and sequenced from only the 5′ or 3′ end of the captured transcripts. Such transcripts are theoretically not suitable for the identification of circRNAs. However, with the rapid development of scRNA-seq technology, we envision that an scRNA-seq pipeline specific for sequencing and identifying circRNAs will be developed soon and that it will bring the study of plant circRNAs into a new era.

Determination of full-length circRNA transcripts by long-read sequencing

Full-length transcripts from many plant species have been generated using single-molecule long-read sequencing technologies such as PacBio and Nanopore. One of the advantages of long-read sequencing is the ability to accurately identify alternative splicing events or fusion transcripts. For circRNAs, two protocols, CIRI-long (Zhang et al., 2021b) and isoCirc (Xin et al., 2021), have been developed to sequence full-length circRNA transcripts in mouse and human by Nanopore sequencing and have identified many circRNAs with a low expression level and alternative splicing events. To explore the feasibility of long-read sequencing for the identification of plant circRNAs, we applied the CIRI-long protocol to rice leaves and roots and identified 3,643 circRNAs, of which 3,572 (~98%) were newly found and only 71 (~2%) have previously been identified with circRNA prediction tools. This result clearly demonstrates the power of long-read sequencing for the identification of plant circRNAs and shows that the circRNA repertoire of rice, and probably of most other plants as well, is far from saturation. However, the low percentage of circular consensus reads that can be used to identify and quantify circRNAs and the poor repeatability of identified circRNAs from the same samples are among the issues that need to be addressed when applying the CIRI-long protocol to plant species. In addition, although the protocol developed for animals is suitable for plants, development of a specific algorithm for plant circRNA identification is essential.

A recent study in *A. thaliana* reported the identification of full-length transcripts by combining scRNA-seq and long-read sequencing (Long et al., 2021). We expect that a similar approach will be adopted to uncover full-length circRNA transcripts at the single-cell level.

Toward functional annotation of plant circRNAs

Although a large number of circRNAs have been identified in many plant species, their biological functions are far from clear. Databases with a focus on the functionality of human circRNAs have been built, such as Circ2Disease (Yao et al., 2018), circRNADisease (Zhao et al., 2018), and circad (Rophina et al., 2020), but no similar database is currently available for plants. In PlantcircBase 7.0, we have predicted the potential of circRNAs to function as endogenous target mimics of miRNAs using eTM_finder (Ye et al., 2014) and their coding potential with cORF_pipeline (Pamudurti et al., 2017). As a result, 6,951 and 30,064 circRNAs were predicted to have the potential to interact with miRNAs and coding short peptides, respectively (Supplemental Table 3).

Given the relationship between conservation and functionality, we have collected supporting evidence for as many circRNAs as possible, including 526 plant circRNAs whose identity was supported by polymerase chain reaction using divergent primers (Supplemental Table 4) and the potential functions of 8,187 circRNAs predicted by studies reporting the circRNAs (Supplemental Table 5). Importantly, we calculated the PMCS for every circRNA included in PlantcircBase, which can be used to select candidate circRNAs with potential biological functions for experimental verification. In the future, we will further update PlantcircBase by adding functional features of circRNAs once such data are available.

(B) Genomic conservation (left) and expression patterns (right) of two representative circRNAs, osa_circ_038222 and ath_circ_029811. In the left plot, the depth of color represents the degree of sequence similarity. In the right plot, the horizontal axis represents different RNA-seq samples, and the vertical axis represents read counts supporting the back-splicing sites of circRNAs.
(C) Expression patterns of *Arabidopsis* circRNAs in different tissues.
(D) Expression patterns of *Arabidopsis* circRNAs in different cell types. PMCS, plant multiple conservation score.

MATERIALS AND METHODS

Collection of RNA-seq datasets and reference genomes

A total of 749 bulk RNA-seq datasets from different plant tissues (including root, leaf, flower, embryo, and so forth) and 292 scRNA-seq datasets from the *A. thaliana* root stele, quiescent center cells, and female gametophytic cells (Supplemental Table 1) were downloaded from the NCBI SRA.

The genome sequences and annotation documents for the 12 plant species used to analyze conservation and expression patterns of circRNAs and to assemble full-length circRNAs were downloaded from Ensembl Plants v.38. The 12 plant species were *Arabidopsis thaliana*, *Brassica rapa*, *Cucumis sativus*, *Glycine max*, *Gossypium raimondii*, *Hordeum vulgare*, *Oryza sativa* ssp. *japonica*, *Populus trichocarpa*, *Solanum lycopersicum*, *Solanum tuberosum*, *Triticum aestivum*, and *Zea mays*. Nine plant species (*Camellia sinensis*, *Echinochloa crus-galli*, *Gossypium arboretum*, *Gossypium hirsutum*, *Nicotiana benthamiana*, *Oryza sativa* ssp. *indica*, *Pyrus betulifolia*, *Poncirus trifoliata*, and *Panax ginseng*) were used only for the collection of reported circRNAs. Their genome sequences and annotation documents were downloaded using the information provided in the corresponding publications from which the circRNAs were collected.

Construction of full-length circRNA sequences

The full-length sequences of circRNAs were assembled using a previously described approach (Chu et al., 2022). In brief, genomic sequences of circRNAs collected by PlantcircBase 7.0 with back-splicing sites were extracted. For each circRNA, two copies of the corresponding genomic sequence were tandemly assembled to form a pseudo-reference. Paired-end reads were mapped to the pseudo-references, and those that aligned across the back-splicing sites of circRNAs were used to assemble full-length transcripts with CIRI-full (Zheng et al., 2019), CIRCexplorer (Zhang et al., 2016), or circseq_cup (Ye et al., 2017). Full-length transcripts assembled by at least one of the three tools were retained.

Nanopore sequencing and identification of circRNAs in rice roots and leaves

Leaves and roots from 2-week-old seedlings of *O. sativa* ssp. *japonica* (Nipponbare) were used for Nanopore sequencing of circRNAs with two repetitions. Tissues were harvested directly into liquid nitrogen and stored at −80°C until use to minimize RNA degradation. Total RNA was extracted using TRIzol reagent (Invitrogen) following the manufacturer's protocol. The Ribo-off rRNA depletion kit (Plant) (Vazyme) was used to remove rRNA according to the manufacturer's instructions. Construction of the Nanopore sequencing libraries for circRNA identification was performed according to the CIRI-long sequencing protocol (Zhang et al., 2021b). Nanopore sequencing was performed using the GridION Mk1 platform. The CIRI-long algorithm (Zhang et al., 2021b) was used for full-length circRNA identification and quantification. circRNAs that were expressed in at least two samples were selected for differential expression analysis.

Quantifying conservation of circRNAs

The PMCS, as previously described (Chu et al., 2022), was used to quantify the conservation of circRNAs. The PMCS measures both genomic conservation and expression conservation. The genomic conservation value of a given circRNA was calculated with MUMMER (Marçais et al., 2018) based on the length and consistency of its sequence among plant species. The expression conservation value of a given circRNA was calculated based on the number of samples that expressed the circRNA, i.e., the number that had reads supporting the back-splicing sites of the circRNA based on mapping with Bowtie (Langmead et al., 2009) and the count of reads in each sample (Wu et al., 2020). Quantification of circRNAs based on RNA-seq datasets was performed using previously described methods (Chu et al., 2022).

Calculation of PCCL scores

The PCCL score of a given circRNA was calculated based on the number of references supporting the circRNA and the experimental validation results using the following formula:

$$PCCL = 10P + R,$$

in which P represents the experimental validation results, with one and zero being validated and not validated, respectively, and R represents the number of references that support the circRNA. Given the significance of experimental validation, which is the gold standard for the identity of circRNA, the weights for experimental validation and reference support were assigned values of ten and one, respectively.

DATA AND CODE AVAILABILITY

All detailed information for plant circRNAs are freely accessible at PlantcircBase (http://ibi.zju.edu.cn/plantcircbase/). The newly sequenced RNA-seq data from this study are available at the NCBI SRA under accession number PRJNA837470.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

FUNDING

This work was supported by the National Natural Science Foundation of China (32101729, 31871589, and 91740108) and the National Postdoctoral Program for Innovative Talents (BX20200301).

AUTHOR CONTRIBUTIONS

Q.C. and L.F. conceived the project. X.X. and Q.C. carried out the RNA-seq dataset collection and analysis, analyzed the findings, prepared the figures and tables, and wrote the manuscript. W.M. carried out plant circRNA library construction and Nanopore sequencing. X.L. and T.D. collected datasets and publicly available references. Q.C. wrote the manuscript. L.F., Q.-H.Z., and C.-Y.Y. supervised the manuscript.

ACKNOWLEDGMENTS

No conflict of interest declared.

Received: March 15, 2022
Revised: May 13, 2022
Accepted: May 26, 2022
Published: May 30, 2022

REFERENCES

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 **10**:1126–1128. https://doi.org/10.1016/j.molp.2017.03.003.

Chu, Q., Ding, Y., Xu, X., Ye, C., Zhu, Q., Guo, L., and Fan, L. (2022). Recent origination of circular RNAs in plants. New Phytol. **233**:515–525. https://doi.org/10.1111/nph.17798.

Efroni, I., Ip, P.L., Nawy, T., Mello, A., and Birnbaum, K.D. (2015). Quantification of cell identity from single-cell gene expression profiles. Genome Biol. **16**:9. https://doi.org/10.1186/s13059-015-0580-x.

Fan, X., Zhang, X., Wu, X., Guo, H., Hu, Y., Tang, F., and Huang, Y. (2015). Single-cell RNA-seq transcriptome analysis of linear and circular RNAs in mouse preimplantation embryos. Genome Biol. **16**:148. https://doi.org/10.1186/s13059-015-0706-1.

Gao, Y., Wang, J., and Zhao, F. (2015). CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. Genome Biol. **16**:4. https://doi.org/10.1186/s13059-014-0571-3.

Gao, Y., Zhang, J., and Zhao, F. (2018). Circular RNA identification based on multiple seed matching. Brief. Bioinform. **19**:803–810. https://doi.org/10.1093/bib/bbx014.

Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F., et al. (2018). Mapping the mouse cell atlas by Microwell-seq. Cell **173**:1307. https://doi.org/10.1016/j.cell.2018.05.012.

Han, X., Zhou, Z., Fei, L., Sun, H., Wang, R., Chen, Y., Chen, H., Wang, J., Tang, H., Ge, W., et al. (2020). Construction of a human cell landscape at single-cell level. Nature **581**:303–309. https://doi.org/10.1038/s41586-020-2157-4.

Koh, W., Gonzalez, V., Natarajan, S., Carter, R., Brown, P.O., and Gawad, C. (2016). Dynamic ASXL1 exon skipping and alternative circular splicing in single human cells. PLoS One **11**, e0164085. https://doi.org/10.1371/journal.pone.0164085.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. **10**:R25. https://doi.org/10.1186/gb-2009-10-3-r25.

Long, Y., Liu, Z., Jia, J., Mo, W., Fang, L., Lu, D., Liu, B., Zhang, H., Chen, W., and Zhai, J. (2021). FlsnRNA-seq: protoplasting-free full-length single-nucleus RNA profiling in plants. Genome Biol. **22**:66. https://doi.org/10.1186/s13059-021-02288-0.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell **161**:1202–1214. https://doi.org/10.1016/j.cell.2015.05.002.

Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., and Zimin, A. (2018). MUMmer4: a fast and versatile genome alignment system. PLoS Comput. Biol. **14**:e1005944. https://doi.org/10.1371/journal.pcbi.1005944.

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. https://doi.org/10.1038/nature11928.

Meng, X., Hu, D., Zhang, P., Chen, Q., and Chen, M. (2019). CircFunBase: a database for functional circular RNAs. Database (Oxford) **2019**:baz003. https://doi.org/10.1093/database/baz003.

Pamudurti, N.R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., Hanan, M., Wyler, E., Perez-Hernandez, D., Ramberger, E., et al. (2017). Translation of CircRNAs. Mol. Cell **66**:9–21.e7. https://doi.org/10.1016/j.molcel.2017.02.021.

Rahimi, K., Færch Nielsen, A., Venø, M.T., and Kjems, J. (2021a). Nanopore long-read sequencing of circRNAs. Methods **196**:23–29. https://doi.org/10.1016/j.ymeth.2021.09.010.

Rahimi, K., Venø, M.T., Dupont, D.M., and Kjems, J. (2021b). Nanopore sequencing of brain-derived full-length circRNAs reveals circRNA-specific exon usage, intron retention and microexons. Nat. Commun. **12**:4825. https://doi.org/10.1038/s41467-021-24975-z.

Rophina, M., Sharma, D., Poojary, M., and Scaria, V. (2020). Circad: a comprehensive manually curated resource of circular RNA associated with diseases. Database **2020**:baaa019. https://doi.org/10.1093/database/baaa019.

Song, Q., Ando, A., Jiang, N., Ikeda, Y., and Chen, Z.J. (2020). Single-cell RNA-seq analysis reveals ploidy-dependent and cell-specific transcriptome changes in Arabidopsis female gametophytes. Genome Biol. **21**:178. https://doi.org/10.1186/s13059-020-02094-0.

Wang, K., Wang, C., Guo, B., et al. (2019). CropCircDB: a comprehensive circular rna resource for crops in response to abiotic stress. Database (Oxford) **2019**:baz053. https://doi.org/10.1093/database/baz053.

Wu, W., Ji, P., and Zhao, F. (2020). CircAtlas: an integrated resource of one million highly accurate circular RNAs from 1070 vertebrate transcriptomes. Genome Biol. **21**:101. https://doi.org/10.1186/s13059-020-02018-y.

Xin, R., Gao, Y., Gao, Y., Wang, R., Kadash-Edmondson, K.E., Liu, B., Wang, Y., Lin, L., and Xing, Y. (2021). isoCirc catalogs full-length circular RNA isoforms in human transcriptomes. Nat. Commun. **12**:266. https://doi.org/10.1038/s41467-020-20459-8.

Yao, D., Zhang, L., Zheng, M., Sun, X., Lu, Y., and Liu, P. (2018). Circ2Disease: a manually curated database of experimentally validated circRNAs in human disease. Sci. Rep. **8**:11018. https://doi.org/10.1038/s41598-018-29360-3.

Ye, C.-Y., Xu, H., Shen, E., Liu, Y., Wang, Y., Shen, Y., Qiu, J., Zhu, Q.-H., and Fan, L. (2014). Genome-wide identification of non-coding RNAs interacted with microRNAs in soybean. Front. Plant Sci. **5**:743. https://doi.org/10.3389/fpls.2014.00743.

Ye, C.Y., Chen, L., Liu, C., Zhu, Q.H., and Fan, L. (2015). Widespread noncoding circular RNAs in plants. New Phytol. **208**:88–95. https://doi.org/10.1111/nph.13585.

Ye, C.-Y., Zhang, X., Chu, Q., Liu, C., Yu, Y., Jiang, W., Zhu, Q.-H., Fan, L., and Guo, L. (2017). Full-length sequence assembly reveals circular RNAs with diverse non-GT/AG splicing signals in rice. RNA Biol. **14**:1055–1063. https://doi.org/10.1080/15476286.2016.1245268.

Ye, J., Wang, L., Li, S., Zhang, Q., Zhang, Q., Tang, W., Wang, K., Song, K., Sablok, G., Sun, X., et al. (2019). AtCircDB: a tissue-specific database for Arabidopsis circular RNAs. Brief. Bioinform. **20**:58–65. https://doi.org/10.1093/bib/bbx089.

Zhang, X.O., Dong, R., Zhang, Y., Zhang, J.L., Luo, Z., Zhang, J., Chen, L.L., and Yang, L. (2016). Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. Genome Res. **26**:1277–1287. https://doi.org/10.1101/gr.202895.115.

Zhang, P., Meng, X., Chen, H., et al. (2017). PlantCircNet: a database for plant circrna-mirna-mrna regulatory networks. Database (Oxford) **2017**:bax089. https://doi.org/10.1093/database/bax089.

Zhang, T.Q., Xu, Z.G., Shang, G.D., and Wang, J.W. (2019). A single-cell RNA sequencing profiles the developmental landscape of Arabidopsis root. Mol. Plant **12**:648–660. https://doi.org/10.1016/j.molp.2019.04.004.

Zhang, J., Hao, Z., Yin, S., and Li, G. (2020). GreenCircRNA: a database for plant circrnas that act as mirna decoys. Database **2020**:baaa039. https://doi.org/10.1093/database/baaa039.

Zhang, T.Q., Chen, Y., Liu, Y., Lin, W.H., and Wang, J.W. (2021a). Single-cell transcriptome atlas and chromatin accessibility landscape reveal differentiation trajectories in the rice root. Nat. Commun. **12**:2053. https://doi.org/10.1038/s41467-021-22352-4.

Zhang, J., Hou, L., Zuo, Z., Ji, P., Zhang, X., Xue, Y., and Zhao, F. (2021b). Comprehensive profiling of circular RNAs with nanopore sequencing and CIRI-long. Nat. Biotechnol. **39**:836–845. https://doi.org/10.1038/s41587-021-00842-6.

Zhao, Z., Wang, K., Wu, F., Wang, W., Zhang, K., Hu, H., Liu, Y., and Jiang, T. (2018). circRNA disease: a manually curated database of experimentally supported circRNA-disease associations. Cell Death Dis. **9**:475. https://doi.org/10.1038/s41419-018-0503-3.

Zheng, Y., Ji, P., Chen, S., Hou, L., and Zhao, F. (2019). Reconstruction of full-length circular RNAs enables isoform-level quantification. Genome Med. **11**:2. https://doi.org/10.1186/s13073-019-0614-1.