

ORIGINAL ARTICLE

Long- and short-read sequencing methods discover distinct circular RNA pools in *Lotus japonicus*

Asa Budnick | Megan J. Franklin | Delecia Utley | Brianne Edwards |
Melodi Charles | Eli D. Hornstein | Heike Sederoff 

Department of Plant and Microbial Biology,
North Carolina State University, Raleigh,
North Carolina, USA

Correspondence

Heike Sederoff, Department of Plant and
Microbial Biology, Systems and Synthetic
Biology Cluster, North Carolina State
University, Raleigh, NC, USA.

Email: hwsederoff@ncsu.edu

Present address Eli D. Hornstein, Elysia
Creative Biology, 840 Oval Drive, Raleigh,
NC, USA.

Assigned to Associate Editor Luigi
Cattivelli.

Funding information

Directorate for Biological Sciences,
Grant/Award Number: DGE-1828820;
North Carolina State University; U.S.
Department of Education, Grant/Award
Number: P200A210002; Novo Nordisk
Fonden, Grant/Award Number:
NNF19SA0059362; Biological and
Environmental Research, Grant/Award
Number: DE-SC0018269; Southern
Regional Educational Board Fellowship;
National Institute of Health, Grant/Award
Number: 1T32GM133366-01

Abstract

Circular RNAs (circRNAs) are covalently closed single-stranded RNAs, generated through a back-splicing process that links a downstream 5' site to an upstream 3' end. The only distinction in the sequence between circRNA and their linear cognate RNA is the back splice junction. Their low abundance and sequence similarity with their linear origin RNA have made the discovery and identification of circRNA challenging. We have identified almost 6000 novel circRNAs from *Lotus japonicus* leaf tissue using different enrichment, amplification, and sequencing methods as well as alternative bioinformatics pipelines. The different methodologies identified different pools of circRNA with little overlap. We validated circRNA identified by the different methods using reverse transcription polymerase chain reaction and characterized sequence variations using nanopore sequencing. We compared validated circRNA identified in *L. japonicus* to other plant species and showed conservation of high-confidence circRNA-expressing genes. This is the first identification of *L. japonicus* circRNA and provides a resource for further characterization of their function in gene regulation. CircRNAs identified in this study originated from genes involved in all biological functions of eukaryotic cells. The comparison of methodologies and technologies to sequence, identify, analyze, and validate circRNA from plant tissues will enable further research to characterize the function and biogenesis of circRNA in *L. japonicus*.

Abbreviations: AS, alternative splice; BSJ, back splice junction; CEnLR, circRNA enriched long read; circPanel-LRS, circular RNA panel long-read sequencing; circRNA, circular RNA; CiRNA, circular intronic RNA; FP, false positive; FPB, fragments per billion mapped bases; IRES, internal ribosome entry sites; RIN, RNA Integrity Number; RT-PCR, reverse transcription polymerase chain reaction; RUBISCO, ribulose biphosphate carboxylase/oxygenase; SPRI, solid phase reversible immobilization; SR, short read.

Asa Budnick and Megan J. Franklin are co-first authors.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Authors. *The Plant Genome* published by Wiley Periodicals LLC on behalf of Crop Science Society of America.

Plain Language Summary

All living organisms contain genes that are expressed to fulfill all the functions an organism needs to survive. Understanding when and how genes are expressed is a major scientific goal. Circular RNAs are a type of RNA involved in plant gene expression, but we don't know much about what circular RNAs do in plants, how they work, or which ones are important. We used two different methods to find circRNAs in the plant *Lotus japonicus* because this plant is used by many research groups to understand gene expression in legumes, which are plants that can form symbioses with microbes. We found that both methods worked well, but neither method found all circular RNAs. We were able to identify almost 6000 circular RNAs. Now that we know some of the circRNAs in *Lotus japonicus* the research community can come up with hypotheses and design experiments to learn more about how circRNAs are involved in plant gene expression.

1 | INTRODUCTION

Circular RNAs (circRNAs) are covalently closed single-stranded RNA molecules that originate from transcribed nuclear or organellar DNA (Nielsen et al., 2022). While RNA molecules with a covalently closed circular topology were originally observed in viroids in the 1970s (Sanger et al., 1976), recent discoveries have shown that all eukaryotic organisms generate circRNA with a wide range of molecular and biological functions (Chen, 2020; Staff, 2014). Formation of a circRNA occurs through the ligation of the 3' and 5' ends of linear RNA, forming the back splice junction (BSJ), the only sequence feature that distinguishes a circRNA from the sequence of its linear cognate RNA. The lack of a 5' or 3' end makes circRNA resistant to exonucleases, which could increase its stability.

Relatively little is known about the biogenesis of circRNAs in plants. The formation of circRNA requires a spliceosomal machinery. CircRNA can originate from exonic, intronic, and intergenic regions. A single gene can produce single or multiple circRNAs, and their occurrence, abundance, and composition can be tissue specific, under developmental control, or respond to biotic or abiotic stress (H. Liu et al., 2020; Philips et al., 2020; Z. Wang et al., 2017; Xia et al., 2023; Yin et al., 2022; P. Zhang et al., 2019). In animal systems, the base-pairing of reverse complementary sequences in the upstream and downstream regions flanking circularized RNA has been shown to be sufficient for the formation of circRNA (Liang & Wilusz, 2014; X.-O. Zhang et al., 2014). However, only a small fraction of plant circRNAs have short inverted repeats in their flanking sequences, suggesting that this is not a major mechanism of RNA circularization in plants (R. Liu et al., 2022). RNA-binding proteins as well as DNA and RNA methylation have been shown to be factors in the biogenesis of some circRNAs in other systems but have not yet specifically

been shown to be involved in circRNA biogenesis in plants (Y. Wang et al., 2020; Z. Zhang, Wang, et al., 2021).

Our understanding of how circRNAs contribute to biological processes in plants is evolving. CircRNAs have been shown to have a variety of functions and mechanisms including regulating alternative splicing through the formation of R-loops (V. M. Conn et al., 2017), interactions with microRNAs (Kleaveland et al., 2018; Zhou et al., 2021), and diverse interactions with RNA-binding proteins (Chen, 2020). Some circRNAs have been shown to be translated into short peptides in animal systems through internal ribosome entry sites (IRES) and N6-methyladenosine (m6A) modifications. These IRES and m6A features have been identified in plant circRNAs, but the translation of plant circRNAs has not yet been shown (R. Liu et al., 2022; Y. Wang et al., 2020). There are relatively few studies that have concretely demonstrated functions of specific plant circRNAs in overexpression or knockdown experiments (V. M. Conn et al., 2017; Z. Gao et al., 2019; R. Liu et al., 2022; Xia et al., 2023; Zhou et al., 2021). One bottleneck to functional studies of plant circRNAs is the difficulty involved in creating circRNA overexpression and knockout lines.

Most known circRNAs have been discovered using short-read (SR) sequencing. SR sequencing with random primers allows both linear and circular RNAs to be sequenced, but circular RNAs can only be identified by the reads that contain the BSJ, meaning most circular RNAs will not have their full internal sequence covered. Recently, several publications have also reported long-read sequencing (LRS) strategies for sequencing full-length circular RNAs (Rahimi, Venø, et al., 2021; Xin et al., 2021; J. Zhang, Hou, et al., 2021).

Lotus japonicus is a legume plant that is widely used as a genetics model. As a model *L. japonicus* offers several advantages for the study of circular RNAs including a relatively small genome; extensive genetic, genomic and

proteomic resources for different ecotypes and accessions; and transformability (Kamal et al., 2020; Udvardi et al., 2005). *L. japonicus* is also a good plant species to study circRNAs due to the availability of mutant lines with LORE1 transposon insertions which nearly saturate the genome with knockout mutants (Małolepszy et al., 2016; Urbański et al., 2012). The LORE1 mutant lines may prove useful in future functional studies of circRNAs for examination of circRNA knockouts. Identification of circRNAs in *L. japonicus* will also further comparative genetic study of circRNAs across legumes and other plant species.

We chose to identify *L. japonicus* (ecotype Gifu) circRNAs from leaf tissue using two different methods to capture a broader landscape of circRNAs than may be produced by either method on its own. The SR sequencing approach included rRNA depletion of total RNA for Illumina deep sequencing using random primers. For LRS, the enrichment of circRNAs is necessary prior to library preparation. The circRNA enrichment with long-read nanopore sequencing (CEnLR) involved the enrichment of circRNAs with an extensive protocol and used the isoCirc method for library preparation and nanopore sequencing (Xin et al., 2021). In this study, we identify a combined total of 5926 putative *L. japonicus* circular RNAs and examine their characteristics. We compare SR and CEnLR methods as well as different computational pipelines for circRNA identification. Select circRNAs identified by both or either sequencing method were validated using reverse transcription polymerase chain reaction (RT-PCR) and an LRS method to characterize sequence variations of specific circRNAs.

2 | MATERIALS AND METHODS

A more detailed description of the methods used in this study is provided as Supporting Information (“Materials and Methods” in Supporting Information). PCR library preparation kit PCB-109 (Oxford Nanopore Technologies) was used for part of these methods and is now discontinued PCB-111 (Oxford Nanopore Technologies) is a suitable replacement.

2.1 | Plant material

Lotus japonicus Gifu B-129 seeds were sterilized, germinated on filter paper for 1–2 weeks, and plated on one-fourth strength Broughton & Dilworth media plates supplemented with 3 mM KNO₃ (Broughton & Dilworth, 1971). Plants were grown vertically at 16 h light 8 h dark, 21–23°C with light intensity of 50–120 μmol/m²/s for 63–65 days from germination to harvest. Roots were protected from light by placing foil bars at the root–shoot interface and wrapping black poster board sleeves around the root portion of the plates. Leaf tissue was harvested and flash frozen in liquid Nitrogen.

Core Ideas

- Different amplification, sequencing, and analysis methods result in the identification of different pools of circRNA.
- The authors offer the first discovery of circRNAs in *L. japonicus*.
- Validation methods can distinguish true and false positive circRNAs.
- Variants of specific circRNAs can be identified with nanopore sequencing.

2.2 | SR Illumina sequencing and identification of circRNA

2.2.1 | RNA extraction

Total RNA from three biological replicates was extracted using the Purelink RNA Mini Kit (Invitrogen) according to the manufacturer’s protocol. Extractions were performed so that each biological replicate contained RNA from a pool of leaf material from 20 plants. Total RNA was treated with DNase I (TURBO DNase; Invitrogen) according to the manufacturer’s protocol. RNA concentration was measured fluorometrically (Qubit Broad Range RNA; Invitrogen). RNA quality was confirmed via capillary electrophoresis using the Bioanalyzer 2100 Plant RNA Nano protocol, and samples with an RNA Integrity Number (RIN) of at least 7 were used for library preparation and Illumina sequencing.

2.2.2 | Library preparation

Total RNA was depleted of ribosomal RNA using QIAseq FastSelect–rRNA Plant Kits (Qiagen). Library preparation was performed using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs) with NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1 NEB). Libraries were sequenced with a single lane of an S4 chip on an Illumina NovaSeq platform resulting in 100 million 150-bp paired reads per sample.

2.2.3 | Sequence analysis and circRNA identification

FastQC was used to confirm the quality of the raw reads with default settings (Andrews, 2010). Poor quality reads and Illumina adapters (Table S1) were removed with Trimmomatic (minimum read length: 30; average phred score of 20 within sliding window of 4 bp) (Bolger et al., 2014). Unpaired reads

were removed after trimming. This set of trimmed fastq files was used as input for circRNA analyses. The identification of circRNA from SR sequencing was done through two independent detection pipelines: CIRI2 (Y. Gao et al., 2015, 2018) and CLEAR (Ma et al., 2019).

CIRI2 analysis was performed as instructed in usage documentation (J. Zhang, 2021). Trimmed read pairs were aligned to the *Lotus* genome (Gifu v1.2) using BWA-MEM (version 0.7.17) (Li & Durbin, 2009) with an alignment score cutoff of -T 19 and otherwise default parameter settings. The *L. japonicus* genome and annotation file used for CIRI2 Gifu v1.2 was downloaded from Lotus Base (Mun et al., 2016). The CLEAR pipeline was used to identify circRNAs and calculate circRNA:linear RNA transcript ratios. The CLEAR pipeline was performed with recommended parameter settings. Reads were aligned with splice-aware alignment tool, HISAT2 (Kim et al., 2019), to the *L. japonicus* genome (Gifu v1.2). CLEAR normalizes linear and circular transcripts by read depth and transcript length and quantifies them in fragments per billion mapped bases (FPB). CLEAR assigns a CIRCscore to each circRNA which is the ratio of circRNA FPB to linear FPB. Outputs of CIRI2 and CLEAR were analyzed with custom R scripts to filter, cluster, and summarize the circRNA results. These are available in our Github repository.

2.3 | CircRNA enrichment and nanopore long-read sequencing

2.3.1 | RNA extraction

Total RNA was extracted from ~6 g of leaf tissue (65 days old) ground with a mortar and pestle under liquid nitrogen. RNA was extracted from tissues using the Purelink RNA mini kit (Invitrogen) according to the manufacturer's protocol. Total RNA was post-treated as described above in SR sequencing. RNA samples with RINs of at least 8 were used for enrichment and sequencing.

2.3.2 | CircRNA enrichment

To enrich circRNA from total RNA, linear RNA was reduced using a combination of processes established in the literature (Panda et al., 2017; Xiao & Wilusz, 2019). Briefly, 90 µg of total RNA was depleted of ribosomal RNA (rRNA) using the RiboMinus Plant Kit (Thermo Fisher) followed by purification via solid phase reverse immobilization (SPRI) (CleanNGS Beads, Bulldog Bio) using a 2× volumetric ratio of beads to sample. The remaining linear RNA was polyadenylated with five units of *E. coli* Poly(A) polymerase (New England Biolabs) in samples of 1.2-µg rRNA depleted RNA. The now polyadenylated RNA was depleted using the NEBNext Poly(A) mRNA Magnetic Isolation Module (New

England Biolabs) according to the manufacturer's instructions. Supernatant containing non-polyadenylated RNA was purified with 1.8× SPRI. Remaining RNA (~2 µg) was treated with 20 U RNase R (Lucigen) in a modified LiCl buffer according to Xiao and Wilusz (2019). The now enriched circular RNA was purified with 1.8× SPRI and quantified using the Qubit High Sensitivity RNA kit (Invitrogen). Yield was ~700 ng of enriched RNA.

2.3.3 | Nanopore library preparation

CircRNA libraries were prepared according to the isoCirc protocol (Xin et al., 2021). Enriched RNA was denatured to reduce secondary structure; then, cDNA was generated using 1 µg RNA, random hexamers, and 10 U ProtoScript II Reverse Transcriptase (New England Biolabs). cDNA was SPRI purified at a 1.8× ratio and then treated with 10 U Mung Bean Exonuclease (New England Biolabs) as directed by the manufacturer to remove 5' overhangs. The sample was again 1.8× SPRI purified and entered into a SplintR ligase (New England Biolabs) ligation reaction with 2 µL (50 U) enzyme to re-circularize the circRNA-originating cDNA. The resulting circular cDNAs were ethanol precipitated, and cDNA concentration was measured with the Qubit High Sensitivity DNA kit (Invitrogen).

Purified circular cDNA was denatured and then amplified using φ29 polymerase from the TempliPhi kit (Cytiva) to facilitate rolling circle amplification and produce a long linear cDNA molecule theoretically containing the full sequence of the circRNA multiple times. Amplified DNA was purified, and size was selected either with BluePippin with a size gate set >3 kb or with a modified SPRI protocol to target size selection to retain species above 1.5 kb (Stortchevoi et al., 2020). Purified and size-selected DNA was quantified via the Qubit High Sensitivity DNA kit (Invitrogen), and size distribution was evaluated with capillary electrophoresis (2200 TapeStation System, Agilent Technologies).

DNA libraries were prepared for nanopore sequencing using the ligation sequencing method (SQK-LSK109, Oxford Nanopore Technologies). This was done according to the manufacturer's protocol with the exception that the initial input was 1 µg of material. All SPRI purification steps were included, and the short fragment buffer was used in the final washing steps. The final material was quantified with the Qubit DNA High Sensitivity kit (Invitrogen).

2.3.4 | Nanopore sequencing

Nanopore libraries were sequenced on the MinION Mk1B sequencer using an R9 Spot-ON flow cell (Oxford Nanopore Technologies). Flow cells were primed and loaded according to the manufacturer's instructions with between 50 and 100

fmol of constructed libraries. Bases were called using Guppy (v6.1.5), and a quality threshold of 8 was used to filter passing reads.

2.3.5 | CIRI-long analysis

Sequencing data were trimmed with porechop (v0.2.4) and then analyzed with the CIRI-long pipeline (J. Zhang, Hou, et al., 2021) (v1.1.0). The *L. japonicus* Gifu genome and annotations (v1.2) were used. The output of CIRI-long was analyzed with custom R scripts to filter, cluster, and summarize the circRNA results.

2.4 | PCR confirmation and Sanger sequencing

Putative circRNAs detected by SR or CEnLR sequencing were validated using RT-PCR. DNase-treated total RNA from 65-day-old leaf tissue was used as a template. Divergent primers were designed (Table S1) to generate PCR products spanning the BSJ. Two primer sets were designed per circRNA, with the goal of minimizing homo- and heterodimer formation while avoiding repetitive sequences in the circRNA template. Primer sets varied in location on the circRNA by 0–160 nt from the 5' end of the primer in the same direction. When possible, primers from set one were located on different exons than those from primer set two. Reactions were constructed with 500 ng RNA input following the OneTaq One-Step RT-PCR Kit (New England Biolabs) manufacturer's protocol, including the denaturation step. Primers for LjUbiquitin were used for a linear positive control reaction. RNase A-treated gDNA template was used for false positive controls. For gDNA control reactions, OneTaq Hot Start 2X Master Mix (New England Biolabs) with Standard Buffer was used, with 100 ng of gDNA template added with the same divergent primers used for RT-PCR of the RNA template. Cycling conditions were the same as RT-PCR, with the reverse transcription step excluded. PCR products were visualized with agarose gel electrophoresis, and divergent products of expected length were excised and purified for confirmation of BSJ by Sanger sequencing.

2.5 | Circular RNA panel long-read sequencing

A variation of circular RNA panel long-read sequencing (circPanel-LRS) was run to validate circRNAs discovered by sequencing and to discover additional circRNAs with shared regions (Rahimi, Færch Nielsen, et al., 2021). Divergent circPanel-LRS primers were designed targeting seven genes

with at least one predicted circRNA. Primers were designed with the 3' end between 30 and 50 nt away from the end of the exon or predicted BSJ. Primers had a binding temperature between 60 and 66°C and included adapters for the nanopore PCR sequencing primers (Table S1).

2.5.1 | cDNA synthesis

First-strand cDNAs were synthesized using 1 µg DNase I-treated total RNA and Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific). Two equimolar pools of seven primers were made, one containing forward primers, and the other containing the reverse primers—this was done to capture circRNAs from either sense or antisense strands. Each pool was at a final total concentration of 10 pmol. RNA and primers underwent denaturation prior to reverse transcription. Completed first-strand synthesis reactions were treated with 1 µL of RNase H (Thermo Fisher Scientific) and 1 µL RNaseCocktailEnzyme Mix (Thermo Fisher Scientific).

2.5.2 | Amplification

The two RNase-treated first-strand synthesis reactions (forward and reverse) were pooled for PCR amplification with LongAmp Hot Start Taq Mastermix (New England Biolabs). Reactions were treated with 1 µL Exonuclease I (New England Biolabs) then pooled and purified with SPRI beads using a 0.8× volumetric ratio and eluted in 40 µL of nuclease free water. The concentration of the first round amplicons was measured via Qubit DNA HS kit (Invitrogen) at 1.39 ng/µL. A second round of PCR was performed with the PCB Nanopore primers from the PCB-109 kit (Oxford Nanopore Technologies, discontinued PCB-111 is a suitable replacement) and LongAmp Hot Start Taq (New England Biolabs). Reactions were again treated with exonuclease 1 and purified by 0.8× SPRI. The final concentration was 250 ng/µL measured by nanodrop, and the size distribution of the library was measured via capillary electrophoresis as described above.

2.5.3 | Final library preparation and nanopore sequencing

The stoichiometry of the library was 110 fmol/µL. From this library, 0.77 µL (85 fmol) was entered into the adapter addition step of the PCB-109 protocol (Oxford Nanopore Technologies) with 1 µL of Rapid Adapter (RAP) and elution buffer to bring the total volume up to 12 µL. This was incubated at room temperature for 5 min and loaded onto an R9 Spot-On flow cell (Oxford Nanopore Technologies) and sequenced on a MinION Mk1B (Oxford

Nanopore Technologies) according to the manufacturer's protocol.

2.5.4 | CircPanel-LRS analysis with CIRI-long

Analysis of circPanel-LRS sequencing data was performed similarly to the above CIRI-long analysis of isoCirc data as both protocols generate cyclic concatemer sequences from a circular RNA template. Custom R scripts were used to analyze the output from CIRI-long and are available on Github.

3 | RESULTS

3.1 | Experimental design

The identification of circRNA requires sequencing the BSJ to identify the 5' and 3' borders that were covalently linked to form the circRNA and their location in the genome. Both SR sequencing (150 nt) and CEnLR sequencing (>150 nt) were used for the genome-wide identification of circRNA from *L. japonicus* leaf tissue. These methods require different inputs, enrichment methods, and analyses and have been shown to identify different subpools of circRNA (Rahimi, Venø, et al., 2021; Xin et al., 2021; J. Zhang, Hou, et al., 2021) (Figure 1). To compare outcomes from SR and CEnLR, we used computational tools to identify the circRNAs from sequence data (CIRI-long, CIRI2, and CLEAR) and validated select circRNAs identified by the different methods. We also quantified the abundance of circRNAs to their linear cognate RNAs and characterized putative functional characteristics and involvement in biological processes (Y. Gao et al., 2018; Ma et al., 2019; J. Zhang, Hou, et al., 2021).

3.2 | CEnLR and SR methods require different inputs

CEnLR sequencing of *L. japonicus* circular RNAs used an extensive circRNA enrichment process followed by library preparation according to the isoCirc method (Xin et al., 2021). We performed two nanopore sequencing runs from a single enrichment of *L. japonicus* leaf RNA. The enrichment consisted of ribosomal RNA depletion, poly-A tailing, poly-A pulldown, and finally RNase R treatment in order to remove as many linear RNA species as possible (Panda et al., 2017; Xiao & Wilusz, 2019). In total, 3.54 Gb of passing reads were produced. Reads were analyzed using the CIRI-long pipeline (J. Zhang, Hou, et al., 2021). In total, before filtering, 4655 unique putative circRNAs were identified with CEnLR (Table S2).

SR sequencing of *L. japonicus* leaf tissue was performed without circRNA enrichment to capture both linear and cir-

cular RNAs. RNA was extracted from three bioreplicates of 65-day-old *L. japonicus* leaf tissue. Ribosomal RNA was depleted, and reverse transcription and library preparation were done with random hexamer primers. SR sequences were analyzed with two circRNA identification pipelines, CIRI2 (Y. Gao et al., 2018) and CLEAR (Ma et al., 2019). From 79.5Gb of total reads, CIRI2 identified 719 putative circRNAs, and CLEAR identified 1345 putative circRNAs prior to any filtering for reliable circRNAs. The initial unfiltered results from individual sequencing runs and analyses are available in Table S2.

The low abundance of circRNA as a fraction of total RNA requires either relatively large amounts of total RNA as input for enrichment and LRS or very deep SR sequencing of libraries generated with random primers. The enrichment of circRNA for LRS requires at least 15 times more total RNA input compared to SR because circRNA enrichment depletes ~99% of total RNA. A comparison of the tissue and RNA input, sequencing data produced, and the number of circRNAs discovered is shown in Table 1.

3.3 | CEnLR and SR methods identify different circRNA pools

Repetitive sequences that can appear identical to BSJs in alignments leads to false positives identification of circRNAs (Dodbele et al., 2021). Therefore, we performed additional filtering steps to increase confidence in putative circRNAs. Putative circRNAs whose BSJ coordinates were within 5 bp of previously annotated repetitive regions of the *L. japonicus* Gifu genome were removed (Kamal et al., 2020). CEnLR raw data contained putative circRNAs with homopolymer and oligomeric repeat sequences which were likely false positives. Due to their repetitive structure, CIRI-long identified these artifacts as circRNAs since a tandem repeat can appear as a BSJ. To remove these, CEnLR circRNAs were filtered based on the circRNA sequence repetitiveness and complexity using the index of repetitiveness ($I_r < 0.2$) and Trifonov's complexity score (>0.1), respectively (Haubold & Wiehe, 2006; Tremblay & Nystrom, 2023). CircRNAs from the SR datasets were additionally filtered by size with circRNAs larger than 10 kb being filtered out as likely false positives. This filtering removed 448 (9.6%) circRNAs from the CEnLR dataset and 142 (6.9%) of circRNAs from the SR dataset (Figure 2A).

CircRNA BSJ coordinates may not be entirely accurate as they are dependent on the sequence at the splice junctions. If small portions of identical sequence exist on both sides of the BSJ, its coordinates can be assigned differently depending on the software and exact supporting reads. To account for this, circRNAs with very similar BSJs (Manhattan distance ≤ 10 across start and end coordinates) were clustered into identical circRNAs (Supplementary Table S3). After clustering,

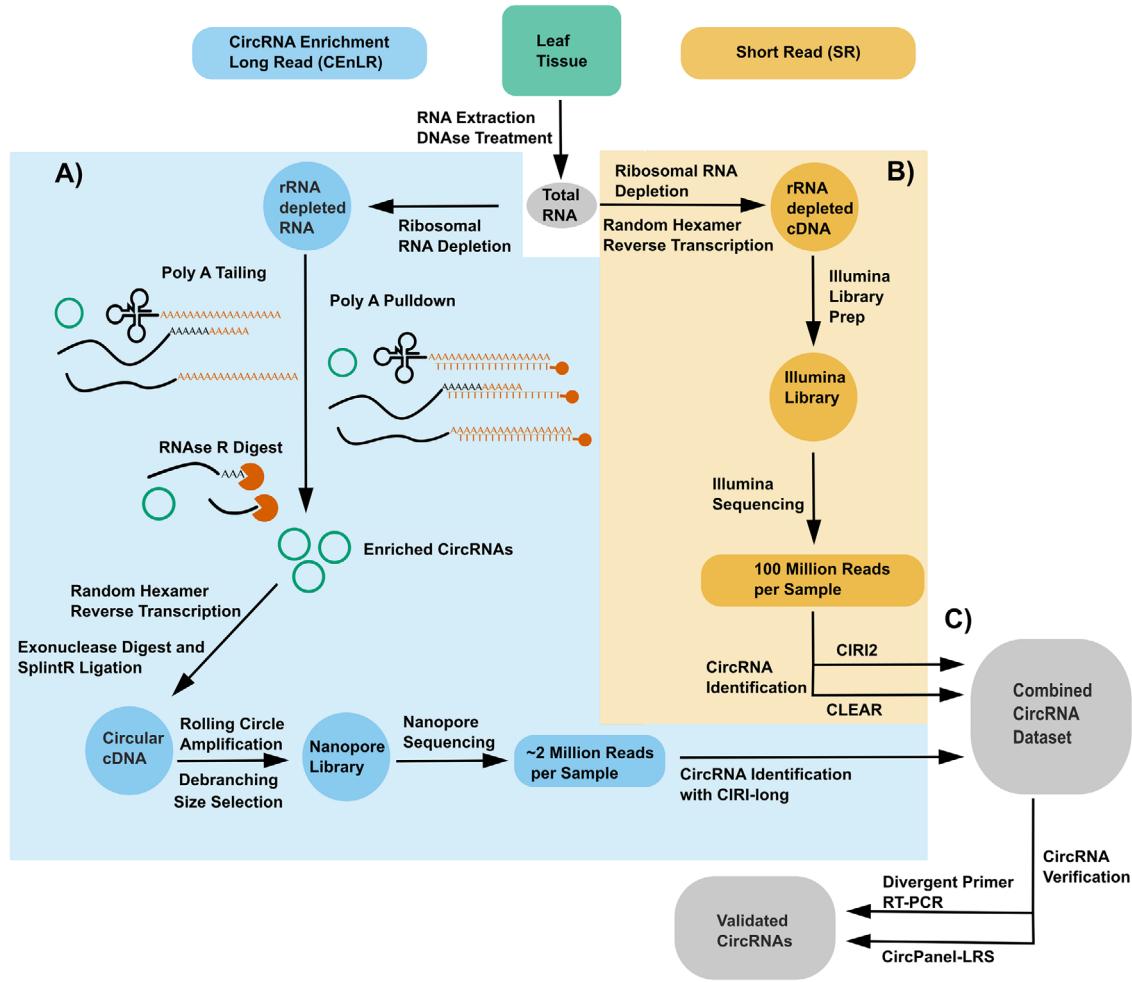


FIGURE 1 Overview of the experimental design and analyses of circular RNA (circRNA) identification from *L. japonicus* leaf tissue. Total RNA extracted from the tissues was either (A) processed for circRNA enrichment, isoCirc library preparation (random hexamer RT, exonuclease digest and SplintR ligation, as well as rolling circle amplification and debranching), and subsequent long-read nanopore sequencing (CEnLR) or (B) amplified using random primers and rRNA depletion followed by short-read Illumina sequencing (SR). (C) Select circRNAs identified through either or both methods were validated using reverse transcription polymerase chain reaction (RT-PCR) and circular RNA panel long-read sequencing (circPanel-LRS) and characterized for their potential functions and quantitative ratios.

TABLE 1 Comparison of inputs and circular RNA (circRNA) sequencing yields for *L. japonicus* leaf tissue.

	CircRNA enrichment long-read sequencing	Short read sequencing
FW tissue (g)	2	0.2–0.6
Total RNA input (μg)	15	0.74–1 (1 μg limit)
RNA remaining after enrichment (ng)	100	N/A
Sequenced nt (Gb) BluePippin/SPRI	0.65/2.88	~27 Gb per replicate
Total reads (million reads/run) BluePippin/SPRI	0.19/1.63	~90 million reads/ replicate
Median circRNA length (all runs and replicates) (bp)	208	814 ^a
Total unique circRNAs after clustering and filtering	4327	1614
Runs/replicates combined		

Note: The same input pool was used for the two nanopore runs. The reported number is the amount of material for an individual run.

Abbreviations: FW, fresh weight; N/A, not available; SPRI, solid phase reversible immobilization.

^aDistance from start and end coordinates.

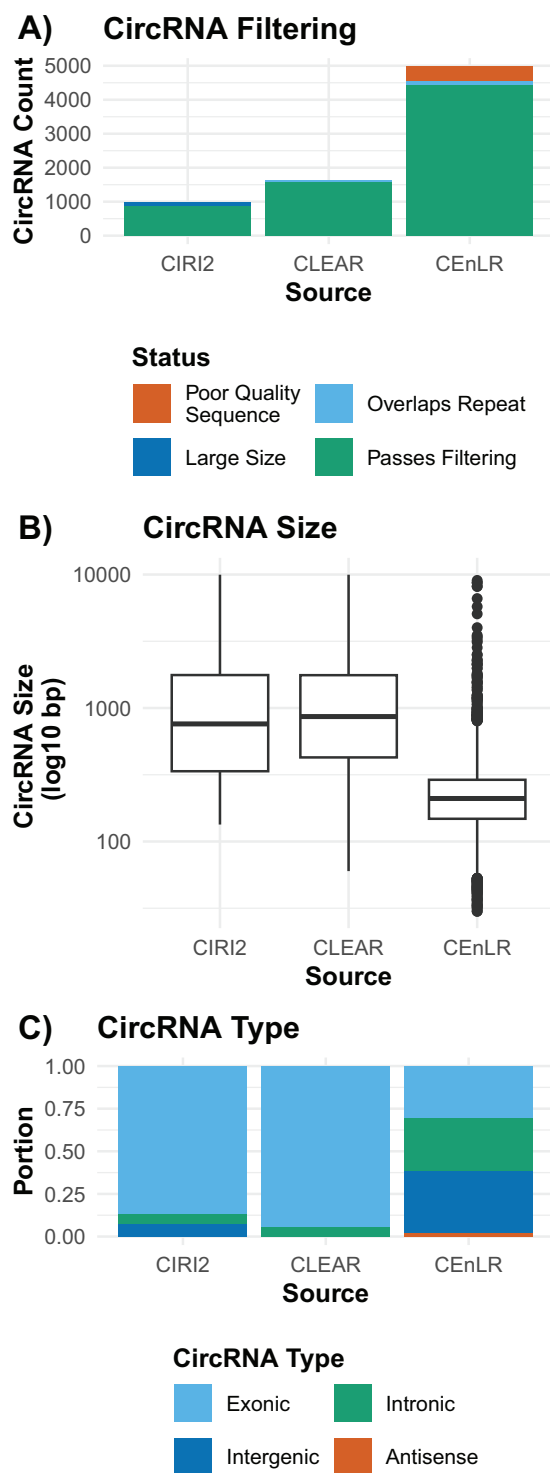


FIGURE 2 Circular RNA (circRNA) characteristics across circRNA enrichment with long-read nanopore sequencing (CEnLR) and short-read (SR) methods. (A) Plot showing the status of circRNAs after different filtering steps across CIRI2 (SR), CLEAR (SR), and CEnLR datasets. CircRNAs shown by the green bar pass through to final clustering and analysis. (B) Boxplot of the sizes of circRNAs identified across datasets. (C) Portion of types of circRNAs identified across datasets.

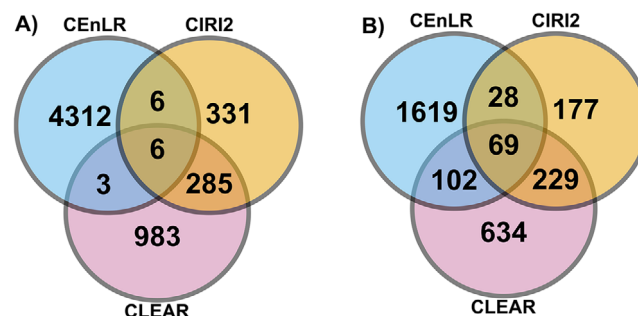


FIGURE 3 Comparison of filtered and clustered circular RNAs (circRNAs) identified by circRNA enrichment with long-read nanopore sequencing (CEnLR) and short-read (SR) methods. (A) Venn diagram of circRNAs identified across CEnLR, CIRI2, and CLEAR datasets. (B) Venn diagram of *L. japonicus* gene IDs that contain circRNAs across CEnLR, CIRI2, and CLEAR datasets.

348 circRNAs were found nearly identical leaving a total of 5926 unique putative circRNAs across datasets.

The populations of circRNAs discovered by the two different methods showed differences in size and type (Figure 2B,C). The median reported size of circRNAs identified with SR sequencing methods was 814 nt, and the median size of CEnLR circRNAs was 208 nt. Importantly, the size of circRNAs from the SR dataset is only based on the distance between BSJ coordinates and likely is not equal to the real sequence length of the circRNA whenever the circRNA contains more than one exon. CEnLR circRNA sizes are calculated based on the full consensus sequence and therefore are accurate. SR circRNAs were also more likely to be assigned as exonic (91.7%) compared to CEnLR circRNAs (31.1%).

3.4 | Comparison between CEnLR and SR circRNAs

It has been observed in animal cells that CEnLR and SR methods capture distinct pools of circular RNAs (Rahimi, Venø, et al., 2021; J. Zhang, Hou, et al., 2021). The mature leaf sequencing runs were from analogous, but not identical tissues, between the two sequencing approaches. Fifteen circRNAs are identified by both CEnLR and SR (Figure 3A) (Supplementary Table S4). This very small portion of shared circRNAs is low compared to other publications, but the methods used here are also more disparate due to the presence or absence of circRNA enrichment. Even within methods, the majority of circRNAs were seen only once. Only 8% (336) of circRNAs were found in both CEnLR sequencing runs. In the SR datasets, CIRI2 found 79 and CLEAR found 80 circRNAs in all three bioreps (Figure S1). Note that 26.6% of CIRI2 and 15.8% of CLEAR circRNAs were identified in two or more bioreps. Taken together, these results show that

neither CEnLR nor SR methods reached saturation for detection of *Lotus* circRNAs.

The circRNAs identified by different methods also originated from distinct *L. japonicus* genes, with only 199 gene IDs shared between CEnLR and SR methods (Figure 3B). The large difference of circRNAs identified (Figure 3A) and circRNA producing gene IDs (Figure 3B) is due to the large number of genes that contained more than one circRNA locus as well as the >30% of putative circRNA identified by CEnLR that originated from intergenic loci (Figure 2C).

3.5 | CircRNAs were validated with RT-PCR and circPanel-LRS

Subsets of circRNAs identified by CEnLR or SR methods were selected for validation by reverse transcription PCR (RT-PCR) or nanopore sequencing of amplicons using circRNA specific primers (circPanel-LRS) (Table S1).

3.5.1 | Reverse transcription polymerase chain reaction

RT-PCR was used to confirm the BSJ of 27 genic circRNAs identified in the discovery sequencing round. Of these 27 circRNAs, 13 were chosen from circRNAs only discovered through CEnLR, five from only SR, and nine from the 15 overlapping circRNAs discovered by both CEnLR and SR (CIRI2 and/or CLEAR) methods (Figure 3A). Divergent primers were used to produce an RT-PCR product spanning the BSJ (Figure 4A) (Table S1). Divergent primers annealing to a linear RNA face away from each other, producing no PCR product. On a circular template, divergent primers face toward each other, producing a product which spans and includes the BSJ sequence. At least two primer pairs that anneal to different locations on the circRNA are used for validation. Twelve targets were successfully amplified via RT-PCR, and Sanger sequencing confirmed the BSJ sequence (Figure 4B–D) (Figure S2). CircRNAs were confirmed for 2/13 CEnLR-only set, 3/5 SR-only set, and 7/9 of the CEnLR + SR overlapping set. Of these, one circRNA was intronic, and 11 circRNAs were exonic. A change in Sanger sequencing quality across the BSJ was observed for several circRNAs, including *circHislys*(5,6,7,8) (Figure 4B). This quality change could be caused by multiple unique circRNAs from the same gene, multiple circRNA isoforms, or concatemers of the same circRNA. To rule out false positives, a gDNA control was included using the same divergent primers to ensure there are no genomic repeats causing a false BSJ. In some cases, primers produce off-target products which are assessed by Sanger sequencing to determine if a gDNA product is identical in sequence to the putative circRNA. From our 27 circRNAs that underwent validation, six are likely false posi-

tives based on products from gDNA PCR which span the BSJ (Figure 4E). This demonstrates the importance of having false positive controls. Two separate primer sets were designed to confirm each circRNA. When possible, primer set one was located on different exons than primer set two. Primer quality is essential for detecting circRNAs and false positives. As shown in Figure 4E, the false positive circRNA from *Insp* would not have been detected using only the second primer set. When validating a circRNA for further functional characterization, it is essential to confidently confirm it is a true circular RNA, and one primer pair may not be sufficient to do so.

3.5.2 | CircPanel-LRS

To characterize specific circRNAs and their sequence variations with high throughput, we used the circPanel-LRS method. CircPanel-LRS is a method developed in Rahimi, Venø, et al. (2021) which uses pools of divergent primers to reverse transcribe and amplify target circRNAs for LRS with nanopore. We designed primers targeting seven genes of the *L. japonicus* genome with putative circRNAs. Total RNA from *L. japonicus* leaf tissue was sequenced according to the circPanel-LRS method. CircRNAs were identified from long-read sequences using CIRI-long. A total of 53 circRNAs were identified from the circPanel-LRS run (Table S2). Five of these circRNAs had been previously found in the CEnLR and SR datasets and 48 of them were new. The confirmed circRNA *circCCR4*(4,5) and false positive FP:Ser/Thr(10,11) were both re-identified in circPanel-LRS. Both loci had additional putative circRNAs that were not seen in the previous divergent RT-PCR validation (Figure 5, Figure S4). RT-PCR and Sanger sequencing showed that FP:Ser/Thr(10,11) is likely a false positive due to genomic duplication. It is expected that reverse transcription of a circRNA will produce a cDNA that contains concatemeric repeats of the circRNA sequence. We measured each repeat as a loop and report the maximum number of loops and average loops for each circRNA identified in circPanel-LRS (Figure 5). FP:Ser/Thr(10,11) has a very low number of average loops indicating that the majority of reads that crossed the back splice junction did not contain multiple repeats and may not have originated from circular topology which aligns with the gDNA PCR evidence of a potential tandem duplication false BSJ (Figure S4). New circRNAs identified from the Ser/thr locus are not necessarily implicated as FP by this duplication since they have different BSJs, but it is prudent to confirm these through RT-PCR and gDNA PCR before considering them valid. A low number of average loops may also be caused by PCR biases toward the smallest amplicons which would contain only one copy of the sequence between the divergent primers. CircPanel-LRS can be a useful method

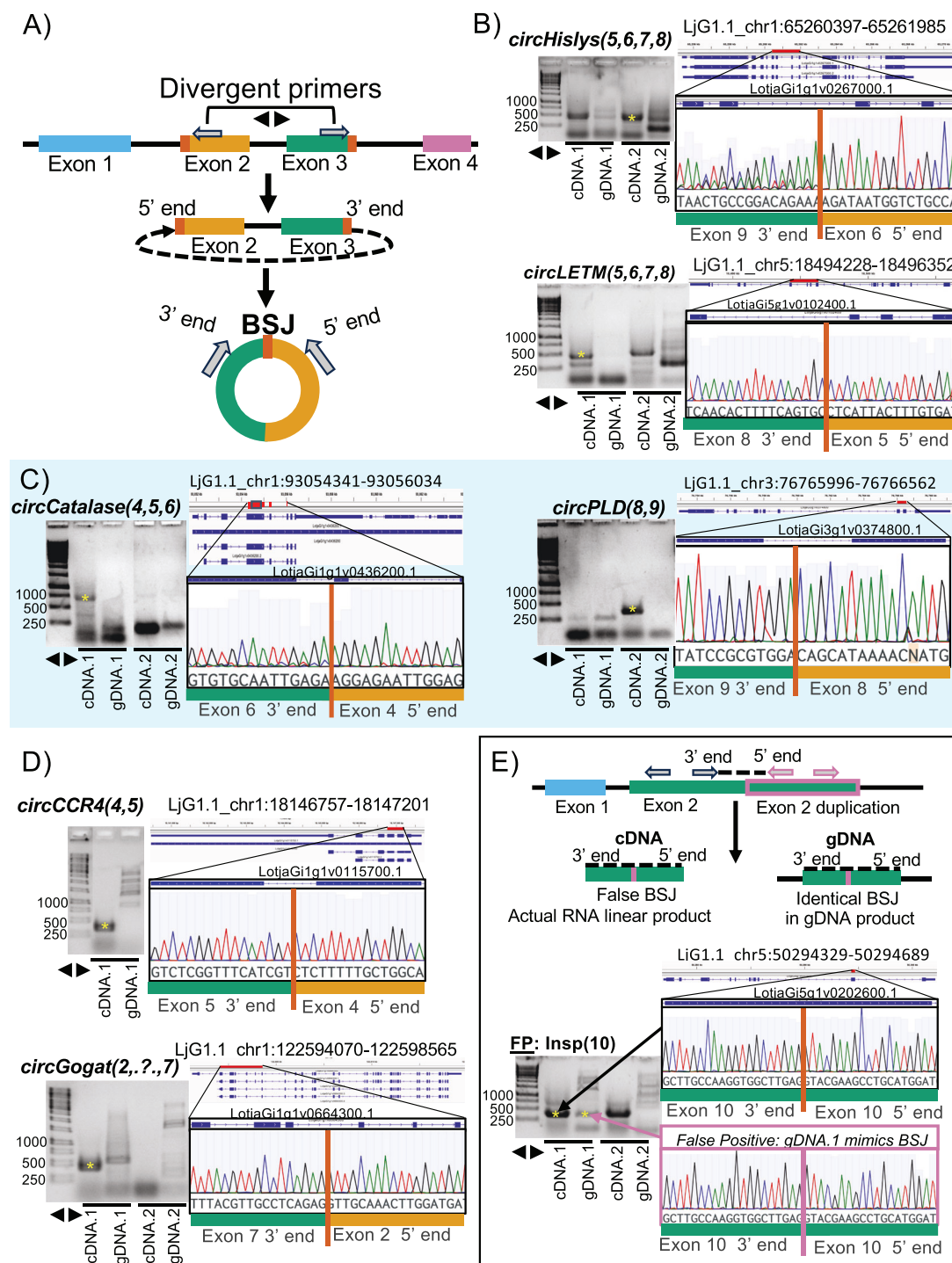


FIGURE 4 Validation of circular RNAs (circRNAs) using reverse transcription polymerase chain reaction (RT-PCR) and Sanger sequencing across back splice junction (BSJ). (A) Illustration of RT-PCR strategy using divergent primers to confirm BSJ. (B) Subset of circRNAs confirmed from the overlap group using two primer pairs (cDNA.1/gDNA.1 = first primer pair; cDNA.2/gDNA.2 = second primer pair). (C) A subset of circRNAs identified only by circRNA enrichment with long-read nanopore sequencing were confirmed using two primer pairs. (D) circRNAs confirmed from short-read group using either one or two primer pairs. (E) Illustration and example of circRNA determined to be a false positive. Asterisk indicates sequenced gel products. A complete list of RT-PCR-confirmed circRNAs, corresponding primers, and original gel electrophoresis images is shown in Table S1 and Figure S3.

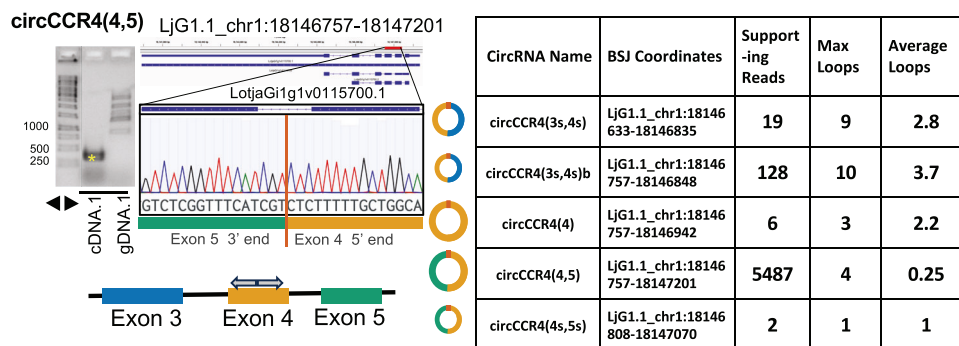


FIGURE 5 Representative results from reverse transcription polymerase chain reaction (RT-PCR) and circular RNA panel long-read sequencing (circPanel-LRS) for *circCCR4(4,5)*. RT-PCR results are shown as in Figure 4 at the left, circPanel-LRS results are summarized in the table at the right, and illustrations of the various circRNA compositions are displayed in the center. CircRNA representations are downsized when there are multiple circular RNAs (circRNAs) that contain different subsets of the same exons.

for validating circRNAs while discovering additional loci-specific circRNAs, but it can be vulnerable to the same false positive discoveries as genome-wide sequencing methods and should therefore be used in combination with other validation strategies, namely divergent primer RT-PCR.

3.6 | Conservation of validated circRNAs

To examine the sequence conservation of the validated *L. japonicus* circRNAs to circRNAs identified in other plant species, we used circRNAs from PlantcircBase for *Arabidopsis thaliana*, *Glycine max*, and *Oryza sativa* (Xu et al., 2022) as well as *Phaseolus vulgaris* circRNAs discovered by Wu et al. (2020). First, gene orthologs of *L. japonicus* parent genes of the 12 genic confirmed circRNAs were analyzed to see if orthologous genes also had known circRNAs (Table 2, Table S5). In *Arabidopsis*, there were 11 orthologous genes with expressed circRNAs, each with multiple (2–47) circRNAs per gene. *Proteus vulgaris* and *O. sativa* also had 11, and *G. max* had eight circRNA-producing orthologous genes. Five *L. japonicus*-confirmed circRNA genes had circRNA-producing orthologs in all four other plants. In addition to the gene level analysis, confirmed circRNA sequences were blasted against the reported circRNA sequences for *At*, *Gm*, and *Os*. This blast produced significant hits ($e < 1.06 \times 10^{-40}$) between soybean and three validated *Lotus* circRNAs: *circHislys(5,6,7,8)* (LjG1.1_chr1:65260397-65261985) to *gma_circ_000998*; *circLETM(5,6,7,8)* (LjG1.1_chr5:18494228-18496352) to *gma_circ_002602*; and *circGOGAT(2,?,7)* (LjG1.1_chr1:122594069-122598565) to *gma_circ_004953* and *gma_circ_009653*. Only validated circRNAs were used for conservation analysis leaving a small sample size; nevertheless, these observations indicate that circRNA gene level conservation appears to be higher than circRNA sequence level conservation between *L. japonicus*, *A. thaliana*, *P. vulgaris*, and *G. max*.

3.7 | CircRNA ratios, GOterm distribution, and potential functions

CircRNA has been shown to originate from genes in all cellular functions. To evaluate the functional classification of circRNA-producing genes in our dataset, we performed gene ontology analysis using the PANTHER database (Mi et al., 2013). To classify *L. japonicus* circRNA parent genes into ontological categories, we referred to their *A. thaliana* homologs provided through Lotus Base (<https://lotus.au.dk/>) (Mun et al., 2016). Because the two different approaches to identify circRNA, CEnLR and SR, resulted in different pools of circRNA (see Figure 3), we estimated the overrepresentation of cognate genes for GOterms separately. Of the unique clustered circRNAs that passed filtering, we compare the amount of circRNAs found by CEnLR versus SR that are nuclear and found in genic regions only. The CEnLR method identified 1747 unique nuclear circRNAs that originated from 1513 unique *L. japonicus* gene IDs, homologous to 1399 *Arabidopsis* genes. From the SR method, we identified 1538 unique nuclear circRNAs from 1187 unique *L. japonicus* gene IDs homologous to 1112 *Arabidopsis* genes. The difference of unique gene IDs with circRNA from CEnLR and SR of 1399 and 1112, respectively, requires that we compare fold-change overrepresentation to evaluate enrichment in GOterms between the two approaches. Significantly overrepresented GOterms (FDR < 0.05; > twofold) are shown in Figure 6 (Table S6).

3.7.1 | Biological process

CircRNAs identified from both methods, CEnLR and SR, showed enrichment in some of the same GOterm categories but with significant differences in other GOterms (Figure 6A). The regulation of stomatal movement (GO:0010119) and especially the subcategory for stomatal closure (GO:0090333)

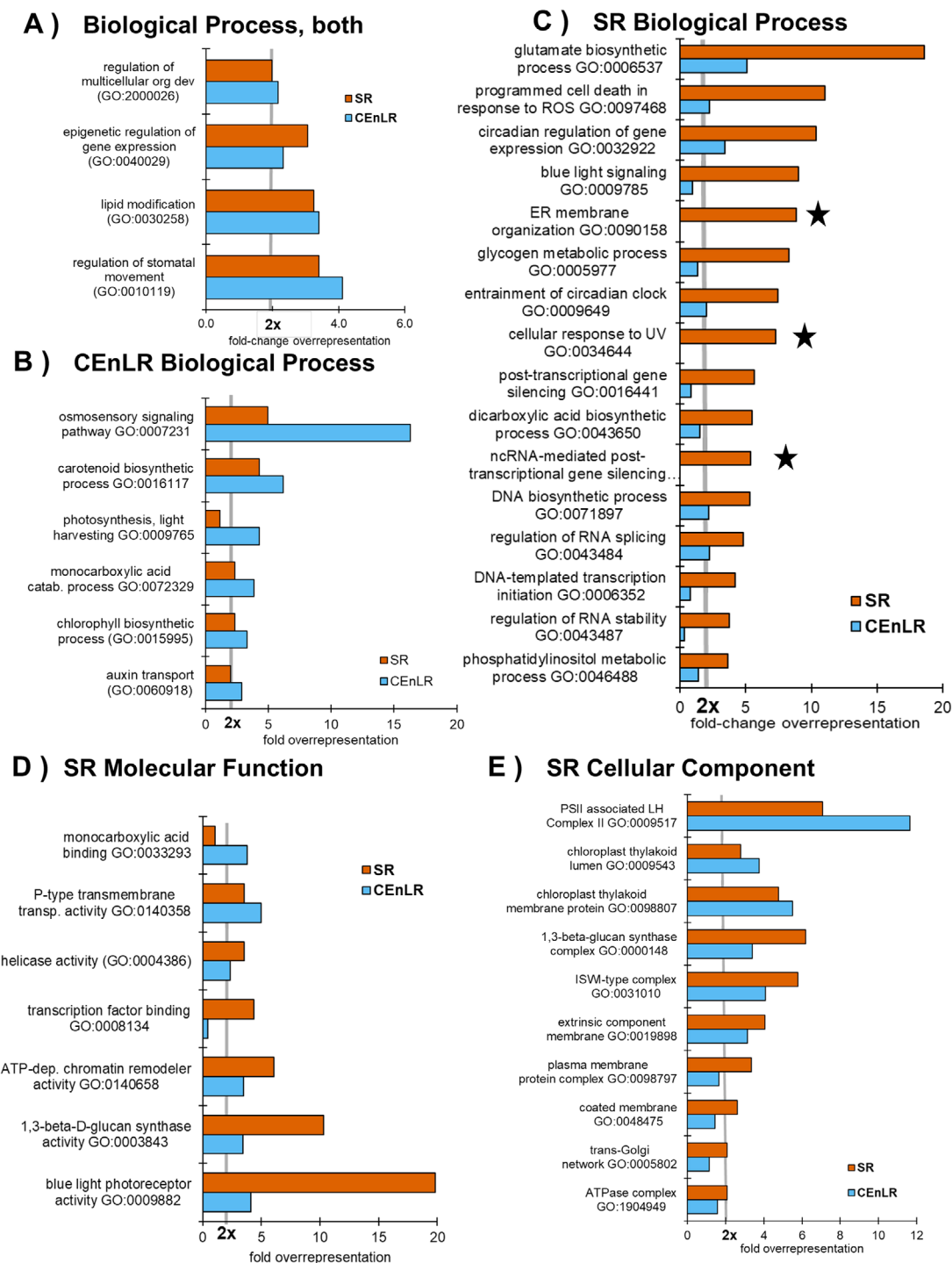


FIGURE 6 Gene ontology of circular RNA (circRNA) genes overrepresented in categories of biological processes, molecular function, and cellular component. The different approaches (circRNA enrichment with long-read nanopore sequencing [CEnLR] and short-read [SR]) to identify circRNA resulted in different pools as defined by GOterm analysis. Both approaches identified circRNA from genes significantly overrepresented in select biological process GOterms (A). Some GOterm categories were significantly enriched only in CEnLR-sequenced circRNA (B) or only in SR-sequenced circRNAs (C), molecular function (D), and cellular component (E). The gray bar represents twofold overrepresentation (FDR < 0.05; fold-change > 2). GOterms exclusively represented in one of the methods are identified by ★. The full GOterm lists of genes in overrepresented GOterms are in Table S6.

TABLE 2 Circular RNAs (circRNAs) expressed by gene orthologs of confirmed *L. japonicus* circRNA genes.

Lotus CircRNA ID	Lotus gene ID	Lj gene description	Number of circRNAs in ortholog genes			
			Ath	Gma	Oryza	Phvul
LjG1.1_chr1:34194372-34194761	LotjaGi1g1v0178300	MORC family CW-type zinc finger protein 4	9	1	6	12
LjG1.1_chr1:18146756-18147201	LotjaGi1g1v0115700	Cinnamoyl-CoA reductase 4	3	2		1
LjG1.1_chr1:122594069-122598565	LotjaGi1g1v0664300	Glutamate synthase	47	7	40	13
LjG1.1_chr1:65260397-65261985	LotjaGi1g1v0267000	Histone-lysine N-methyltransferase	11	1	4	1
LjG1.1_chr5:18494228-18496352	LotjaGi5g1v0102400	LETM1-like protein	7	5	5	5
LjG1.1_chr1:16045704-16046857	LotjaGi1g1v0103800	KH domain-containing protein	2	1	1	5
LjG1.1_chr1:93054412-93054854	LotjaGi1g1v0436200	Catalase	12	2	11	0
LjG1.1_chr2:92355020-92356270	LotjaGi2g1v0425700 and LotjaGi2g1v0425800	Ankyrin repeat-containing protein/acid beta-fructofuranosidase	0	0	2	1
LjG1.1_chr3:76765996-76766562	LotjaGi3g1v0374800	Phospholipase D	2	0	4	0
LjG1.1_chr1:134081453-134082655	LotjaGi1g1v0768100	BTB/POZ domain-containing protein family	0	0	1	1
LjG1.1_chr6:38896647-38897426	LotjaGi3g1v0509000	Basic HLH transcription factor	0	0	1	0
LjG1.1_chr6:38896647-38897426	LotjaGi6g1v0134600_LC	Beta Amylase/BES1/BZR1 plant transcription factor	0	0	0	0

were significantly (three- to fivefold) enriched GOterms. Key genes with circRNA in these categories included the chloroplast localized Calcium Sensing Receptor (*CaS*) which regulates stomatal movement in response to extracellular calcium (Weinl et al., 2008) and the G-protein alpha SU1 (*GPA1*), shown to be a positive regulator of abscisic acid-mediated inhibition of stomatal opening (Nilson & Assmann, 2010). Another overrepresented GOterm for both methods is lipid modification (GO:0030258) and especially lipid oxidation (GO:0034440). Many of the genes in the lipid-modifying category were either lipid/fatty acid degradation enzymes in the peroxisomes and mitochondria, including lipoxigenases (*LOX1*, 2, and 4), or enzymes involved in phosphatidylinositol and phosphoinositide metabolism. Lipid metabolism is key in photosynthetically active tissue because the constant turnover of membrane lipids in the thylakoid membrane reduces damage caused by reactive oxygen species produced by water oxidation.

Some biological process GOterms were significantly overrepresented in CEnLR compared to SR (Figure 6B). Osmosensory signaling (GO:0007231) was represented with four circRNA gene IDs out of the total five genes in that

GOterm in the CEnLR pool compared to only one in the SR pool. Most of the genes in this category have multiple alternative splice (AS) forms, including the Cytokinin receptor *CRE1* with six AS versions in Arabidopsis and at least three known AS versions in lotus (Inoue et al., 2001). *CRE1* circRNA was only identified in CEnLR, not SR pools. In contrast, SR circRNA gene IDs were overall more highly enriched than the CEnLR identified circRNA genes in a large number of GOterm categories involved in glutamate biosynthesis, light responses, and circadian rhythm as well as nuclear transcription, and transcript stability and modifications (Figure 6C).

Three GOterms were uniquely represented and significant in SR but not present in CEnLR, while none were unique to CEnLR and absent from SR. Those three SR specific GOterms were ER membrane organization (GO:0090158), cellular response to UV (GO:0034644), and ncRNA-mediated post-transcriptional silencing (GO:0035194). The ncRNA-mediated post-transcriptional silencing category contains mostly genes involved with miRNA transcription, processing, and miRNA-mediated mRNA degradation machinery.

3.7.2 | Molecular function

Most of the molecular function categories with circRNA producing genes are involved in the regulation of DNA replication, gene expression, and organization (Figure 6D). The only GOterm that was significantly overrepresented in CEnLR (13 circRNA genes) versus SR (3 circRNA genes) pools was monocarboxylic acid binding (GO:0033293). The genes in this GOterm (GO:0033293) are functionally involved in beta-oxidation of fatty acids (Lipid oxygenase, Acyl-CoA oxidase, and mitochondrial Pyruvate Dehydrogenase), the methylgroup transfer (*SHM4*), the tricarboxylic acid cycle and glycolysis as well as photosynthesis and CO₂ assimilation like Photosystem II subunit P1 and Ribulose Bisphosphate Carboxylase/Oxygenase (RUBISCO) small subunit, the primary CO₂ fixation enzyme in plants. CircRNA involved in the regulation of RUBISCO SU has been shown to possibly function through antisense regulation of RNA stability or translation (J. Zhang, Hou, et al., 2021). Similar to the enrichment of GOterms for biological functions, SR circRNA showed stronger overrepresentation in select molecular function categories. Here, especially the GOterm blue light receptor activity (GO:0009882) was overrepresented with the presence of Cryptochromes 1 and 2 (*CRY1*; *CRY2*) as well as the phototropism receptors *PHOT1* and *PHOT2* (Cashmore et al., 1999; Christie, 2007). *CRY2* and both phototropins have been shown to have alternative splice variants. The enriched GOterm 1,3-beta-D-glucan synthase activity is overrepresented in SR with circRNAs from five of the 12 Callose synthases, a cell wall glucan polymerase gene family involved in many functional aspects of permeability and plasmodesmal transport as well as cell wall modifications (Zavaliev et al., 2011).

3.7.3 | Cellular component

CircRNA gene IDs were enriched in photosynthetically active components in the leaf tissue (Figure 6E). Enrichment in components of light-harvesting complexes of Photosystem II (GO:0009517) are an integral part of the thylakoid membrane (GO:0009543). The overrepresented GOterm category ISWI-type complex (GO:0031010) is a nuclear complex that contains an ATPase subunit of the imitation switch (ISWI). These ATPases are involved in assembling chromatin and spacing nucleosomes to regulate transcription of nuclear RNA polymerases, DNA replication, and recombination. Half of the 15 genes in the ISWI complex generate circRNA that was identified by SR, while only three were identified through CEnLR.

Overall, the consensus between GOterms between CEnLR and SR was higher than the consensus in gene IDs or circRNA IDs. This indicates that while the meth-

ods detect different pools of circRNAs originating from different genes, they are both sampling from a related landscape of biological processes and molecular function categories.

3.7.4 | Quantification of circRNA

Quantification is based on the number of BSJ reads per sample from SR sequencing (Table S3, “CIRI2_#junction_reads” and “CLEAR_readNumber”). Total BSJ read numbers per sample varied from 110 reads to a single read identified in the CLEAR pipeline and 168 to 2 reads in the CIRI2 pipeline. CIRI2 has a cutoff of BSJ reads at 2, discarding all single read BSJs. Overall, circRNA read number distribution from both pipelines was similar at and above 2 reads per BSJ. Unique BSJs that were identified in 2 or 3 reads per sample comprised the majority of 61% and 69% of all BSJ reads from CIRI2 and CLEAR, respectively. Abundances of 4 to 10 BSJ reads per sample accounted for 27% and 21%, respectively, and unique BSJs counted between 10 and 100 times in samples accounted for only 10% and 12%, respectively. Only three circRNAs identified by the CIRI2 pipeline and 1 circRNA identified by the CLEAR pipeline had more than 100 reads. These high (>100) read numbers of putative circRNA identified by CIRI2 came from intronic sequence of the *BES1/BZR1* plant transcription factor (LotjaGi6g1v0134600) (validated, Figure S2), exonic circRNA from an overlap of Proline-rich protein 2 (LotjaGi3g1v0531300), and the Eukaryotic translation initiation factor 3 subunit A (LotjaGi3g1v0531400), as well as an exon that is encoded in the overlap of extension 3 (LotjaGi2g1v0134300) and Carbamoyl-phosphate synthase LSU (LotjaGi2g1v0134400) genes. The putative circRNA with the highest number of BSJ containing read counts identified by CLEAR originated from intronic sequence of the Polyubiquitin gene (LotjaGi5g1v0317900), a gene with eight known alternative splice forms.

Interpreting quantitative information of circRNA abundance from the CEnLR pipeline is more challenging because the enrichment process and lack of internal standard prevent correlation to the linear transcripts or any other molecular component. It is also unknown how biases in rolling circle amplification of circular cDNAs may alter the estimation of quantities of circRNAs. With those caveats, we found that circRNAs identified by CEnLR varied between 27,141 reads to 2 reads (CIRILong_Score; Table S3). About 40% of all CEnLR sequenced circRNAs were identified with only 2 reads, and 44% with 3–10 reads. Only 13% of all CEnLR circRNAs were identified by 11–100 reads and 2% with reads between 100 and 1000. The most abundant genic reads (27,141) originated from an exon of LotjaGi4g1v0068400, a Serine/Threonine-kinase and LotjaGi6g1v0296700, a Peroxiredoxin (7939 reads), located in the mitochondrial matrix.

Both high read count circRNAs were only detected by CEnLR, not by SR.

CircRNA that was identified by both methods could serve to compare relative abundances between methods in theory. However, most of the 15 circRNAs identified by both methods showed read counts of 2–5 for either method, and quantitative analysis is not informative since all methods are close to the threshold of detection.

In addition to relative abundance of a putative circRNA in a sample, the relative abundance of that circRNA compared to its cognate RNA can be informative for potential functions. For example, functions that rely on multiple interactions with other molecules such as miRNA sponging or protein sequestration would necessitate a higher abundance of circular RNA. Alternatively, R-loop formation would require only two copies of a circRNA to bind each chromosome in a diploid cell. These R-loops could result in many alternatively spliced linear isoforms of the cognate transcript, leading to a low CIRCscore. CLEAR and CIRI2 use different baselines to estimate relative abundance of BSJ reads to reads of cognate linear RNAs. CIRI2 counts only the reads mapped to the BSJ region but which are consistent with linear RNA as a linear transcript to calculate the CircRNA:linear RNA ratio (“CIRI2_junction_reads_ratio” in Table S3).

The CLEAR pipeline instead compares BSJ supporting reads to all reads from the linear transcript which contain a splice junction other than the BSJ and normalizes both relative to the total fragments per billion mapped bases (FPB). These gene-specific, normalized circRNA reads (FPBcirc) and linear RNA reads (FPBlinear) can be used to estimate the level of circular RNA abundance compared to the linear cognate which is assigned by CLEAR as a CIRCscore (FPBcirc/FPBlinear). We examined the correlation of unambiguously linear RNA FPB to their CircRNA FPB and found a weak positive correlation (Spearman's Rho = 0.16; Figure S5). CIRCscores vary widely from ratios above 1, indicating more circRNA than linear RNA, to less than five circRNAs per 100,000 linear cognate RNAs. About 50% of the circRNAs have a CIRCscore that estimates their abundance to be between 2.5% and 0.5% of its linear RNA abundance (Figure 7).

CircRNA with high CIRCscores indicate their abundance is equal to or higher than their linear RNA (>1). The two highest scoring transcripts in this category were involved with auxin transport and function. The highest CIRCscore of 2.1 was identified for LotjaGi1g1v0099700, a Myb transcription factor homologous to the Arabidopsis REVEILLE 1 (*RVE1*) Myb transcription factor. Revielle 1 produces two alternative splice versions and has been shown to be the mechanistic link to the circadian control of auxin levels (Rawat et al., 2009). CircRNA from the zinc-induced facilitator-1 like (LotjaGi3g1v0102200) transcript was more abundant than its cognate RNA (CIRCscore 1.5), indicating a potential function

CIRCscore distribution

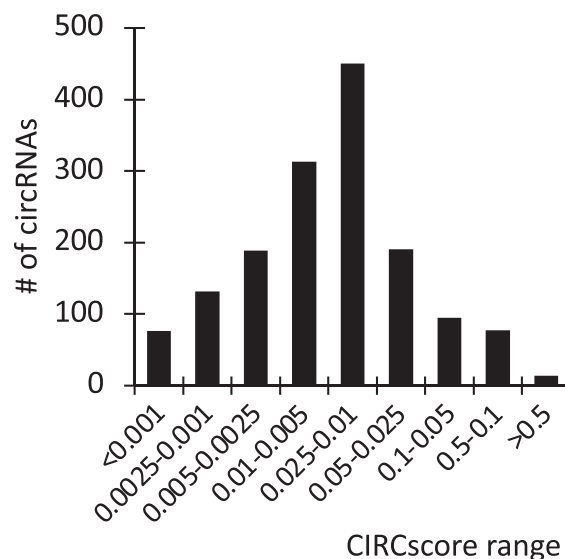


FIGURE 7 Distribution of CIRCscores from CLEAR pipeline. CIRCscores estimate the relative ratio of circular RNA (circRNA) to its origin RNA based on gene specific splice junction reads.

in protein or RNA sequestration. The Arabidopsis homolog AT5G13750 has been shown to produce two different splice forms, *ZIFL1.1*, and a truncated splice isoform, *ZIFL1.3*, which localize to the tonoplast of root cells or the plasma membrane of leaf stomatal guard cells, respectively. The *ZIFL1.1* transporter regulates various root auxin-related processes, while the *ZIFL1.3* isoform mediates drought tolerance by regulating stomatal closure (Remy et al., 2013).

4 | DISCUSSION

Two of the challenges to high-throughput sequencing of circRNA are their low abundance and high sequence identity with cognate linear RNAs. We have shown here that two very different approaches to circRNA identification from *L. japonicus* leaf tissue can each yield large numbers of unique circRNAs, albeit with little overlap between the identified pools. The first approach—enriching circRNA from total RNA, amplification, and nanopore sequencing (CEnLR)—has the potential to identify the actual composition of the full length circRNA, rather than just the BSJ read. The cDNA synthesis and rolling circle amplification in CEnLR appears to also generate artifacts from highly repetitive regions which can be filtered out (Figure 2A). The second approach using SR sequencing does not enrich for circRNA and therefore requires deep sequencing of rRNA-depleted total RNA using random primers for cDNA synthesis and relies entirely on the detection and identification of BSJs in short (150 nt) reads.

This method does not provide the complete internal sequence of the circRNA but offers qualitative and semi-quantitative information about the linear origin RNA. The CEnLR method was more labor and material intensive than the SR approach. Combinations of both methods, for example, SR sequencing of circRNA enriched libraries, are possible alternatives, but also have some of the drawbacks described above.

The CEnLR method identified more than twice as many circRNAs compared to SR, and only 15 circRNAs were identical between both methods. Similar results have been described by other researchers using both methods in animal circRNA research (Ma et al., 2023). One of the major differences is the presence and large numbers of nuclear intergenic (1210 unique), chloroplast (930), and mitochondrial (395) circRNAs identified by CEnLR versus those intergenic (39) circRNA and only few chloroplast (11) and mitochondrial (13) circRNAs in SR.

One of the major differences between CEnLR and SR approaches is the circRNA enrichment process required for long-read library preparation. This enrichment process utilizes RNase R, a 3'-to-5' exoribonuclease to degrade linear RNAs. In addition to the removal of linear RNAs, the RNase R treatment step has been shown to remove false positives that result from trans-splicing events in human cells (Chuang et al., 2018; Yu et al., 2014) but can also lead to the degradation of larger circRNAs during prolonged incubation (Y. Zhang et al., 2016).

4.1 | Computational pipelines to identify CircRNA

A further difference between the CEnLR and SR methods is the computational pipelines used to analyze the sequencing data. There are many software options to identify circRNAs from short reads, some have gone through multiple versions and improvements or have been wrapped into convenient pipelines, and empirical comparisons are available (Gaffo et al., 2017; Y. Gao et al., 2018; Ma et al., 2019; Vromman et al., 2023). CIRI2 and CLEAR are both established softwares with similar precision (Vromman et al., 2023). In this study, CLEAR identified roughly twice the circRNAs than CIRI2, in part due to having no thresholds for the number of reads supporting a BSJ or the size of circRNAs.

Relatively few circRNA identification software options exist for LRS data, and in general, they have been created for specific wet-lab procedures which limit their ability to be used in conjunction or empirically compared (Rahimi, Færch Nielsen, et al., 2021). We found CIRI-long compatible with our organism and sequencing strategy. Unable to improve confidence in CEnLR circRNAs with multiple detection softwares, we performed filtering based on sequence repetitiveness and complexity, which has not been

described previously. Further development of long-read circRNA identification softwares, which are method and organism agnostic, would help expand the options available to researchers and likely improve the confidence of putative circRNAs.

4.2 | Distinction of true and false positive circRNAs

Repetitive genomic sequences can mimic a BSJ and thereby be identified as circRNA by the applied bioinformatic tools, causing false positive circRNAs. Successful validation of a portion of circRNAs from each method (2/13 CEnLR-only, 3/5 SR-only, and 7/12 by both CEnLR+SR) shows that both CEnLR and SR can identify true circRNAs. Considering each method individually, there were in total 10 validated from SR discovery and nine validated from CEnLR. The small proportion of validated circRNAs identified through CEnLR-only (2/13) suggests that the CEnLR method may be more prone to producing false positive circRNAs than the SR method. Another explanation could be that the CEnLR method might have greater sensitivity for scarce circRNAs that were more difficult to validate via RT-PCR and circPanel-LRS. This is due to low circRNA abundance that may require an enrichment step prior to validation, which was not performed here. However, since we sampled such small subsets of circRNAs, these proportions of validated circRNAs from CEnLR and SR are not representative of the amount of true circRNAs we expect within each sequence-identified pool of putative circRNAs.

CircPanel-LRS enables higher throughput than RT-PCR and has the advantage of being able to discover the full sequence and variations of new circRNAs, although full-length circRNA sequences have also been characterized and confirmed from SR sequencing of RNase R-treated samples through computational methods (Ye et al., 2017). However, CircPanel is unable to discredit false positive circRNAs as can be seen in Figure S4 where the FP:Ser/Thr(10,11) was shown to be a false positive due to genomic duplication by PCR and Sanger sequencing but appeared legitimate in circPanel-LRS. We address genomic duplications via PCR of gDNA templates with circRNA specific divergent primers. gDNA products that contain what appears to be a BSJ suggest the presence of tandem genomic duplication. It is recommended to test each circRNA with at least two pairs of divergent primers. This not only improves the circRNA detection chance but also gives more opportunities to detect false positives. Out of the six false positives identified during circRNA validations, only four were clearly false positives for both primer sets. Since circPanel-LRS is not capable of identifying false positives, it is necessary to pair it with at least a gDNA template PCR to check for genomic duplications.

False positives may occur from multiple scenarios, including genomic duplications, trans-splicing of pre-mRNA, and enzymatic artifacts such as ligation and template switching during cDNA synthesis and PCR (Dodbele et al., 2021; Jeck & Sharpless, 2014; Szabo & Salzman, 2016). Though we only focused on detecting possible genomic duplications, there are strategies for detecting other false positives which can be important to investigate especially when selecting circRNAs for further functional analysis. False positives caused by trans-splicing of pre-mRNA and RT enzyme template switching can be removed via RNase R treatment (Chuang et al., 2018; Yu et al., 2014). This in theory depletes the linear RNA which contains a false BSJ from in vivo trans-splicing events as well as in vitro cDNA synthesis artifacts where RT dissociates from the RNA template. However, true circRNAs are also susceptible to RNase R (Y. Zhang et al., 2016). Enzyme artifacts from template switching can also be distinguished by using different RT enzymes (such as from avian myeloblastosis virus vs. Moloney murine leukemia virus), as they are unlikely to jump at the same sequences (Dodbele et al., 2021; Tang et al., 2018; Yu et al., 2014). Alternative validation methods which do not include an RT step can be used, such as hybridization methods with probes like Northern blotting, nanoString, nCounter, and SplintQuant, among other strategies (V. Conn & Conn, 2019; Geiss et al., 2008; Nielsen et al., 2022).

4.3 | Intronic circRNA

The removal of intronic sequences from transcripts is a well-described mechanism in animal and plant genomes. Intron excision involves the synthesis of a lariat RNA which is circularized by a non-colinear 2'-5' junction at the 5' and branchpoint nucleotides (R. Liu et al., 2022). These lariats are formed during pre-mRNA splicing but are linearized by the RNA-debranching enzyme and subsequently degraded. Some lariats have been shown to escape degradation and become stable circular intronic RNAs (ciRNAs), resistant to RNase R degradation (Y. Zhang et al., 2013). While ciRNA has been identified in several plant species in high numbers, putative function of intronic circRNA has only been shown by one group to possibly be involved in miRNA sponging (Zhou et al., 2021). In human cells, it has recently been shown that intronic circRNAs can regulate expression of their cognate genes, including the insulin gene (Stoll et al., 2020).

4.4 | Organellar circRNA—Potential challenges

Putative circRNAs originating from transcripts of the chloroplast and mitochondrial genomes were mostly identified by

CEnLR rather than SR sequencing (Liao et al., 2022). The chloroplast genome of *L. japonicus* (var. MG-20) has been fully sequenced and shown to encode 84 proteins, four rRNAs, and 37 tRNAs (Kato et al., 2000). A large section of 25,156 bp is duplicated as an inverted repeat in the chloroplast genome and contains seven genes, including the NADH dehydrogenase SU (*NDHB*); ribosomal rRNAs 5S, 23S, and 16S, as well as small and large subunits of ribosomal proteins (*RPS*, *RPL*); and tRNA genes.

As a remnant of their prokaryotic ancestors, most chloroplast and mitochondrial genes are organized into multicistronic operons. However, in contrast to prokaryotes, many of those organellar genes contain one or more group I and II introns and a splicing machinery (Barbrook et al., 2010). Part of that splicing machinery in chloroplasts and mitochondria is a large RNA-protein supercomplex that catalyzes trans-splicing, the ligation of exons from different transcripts while removing introns (Kück & Schmitt, 2021). This mechanism and the high potential for false-positive circRNA identification from the large numbers of circular chromosomal DNA that comprises the mitochondrial and plastid genomes make the confirmation of real organellar circRNA, should it exist, difficult. While other researchers have reported putative organellar circRNAs (S. Liu et al., 2019; Philips et al., 2020), no validation or function has been shown. It has recently been shown that plant mitochondrial circRNA is translatable, and peptides derived from these circRNAs have been sequenced (Liao et al., 2022). The challenge will be proving the function of those circRNAs originating from organellar chromosomes as genetic engineering, and the regeneration of stable transgenic homoplasmic plants is challenging.

Both intronic and organellar circRNAs are not fully confirmed or understood. Further study of both may be complex, but it is interesting that the CEnLR method produces more putative circRNAs for both categories, and future long-read methods may be well suited to the study of these circRNAs.

5 | CONCLUSIONS

High-throughput identification of real circRNA is still challenging with existing methods. We compared two very different methodologies to enrich and sequence circRNA from lotus leaf tissue and found that the two approaches resulted in the identification of distinct pools of circRNA with less than 0.4% overlap. Select circRNAs from each pool were validated, confirming that both pools contain “real” circRNA, but that verification is required due to possible artifacts produced by each method. It has been suggested that some circRNAs are produced as splicing byproducts and are not functional molecules in their own right. While our methods can identify true circRNAs from false positive artifacts, they cannot make assertions about the biological relevance of these circRNAs. It

is tempting to assume that high total or relative abundance of a circRNA is indicative of function; it is not necessarily the case as different mechanisms would require different amounts of circRNAs. We have included abundance metrics and GO analysis to help contextualize our annotated circRNAs, but these do not give adequate functional insight. Rather, differential expression of circRNAs across different conditions is more likely to give functional information. In our dataset, lacking different tissues or timepoints, we can only annotate the circRNAs and leave further functional characterization to future studies.

Our results suggest that plant studies focusing on how circRNAs are differentially expressed, especially in relation to linear RNA species, will benefit more from using the SR strategy. Studies interested in uncovering the full population of circRNAs in an organism of interest are more likely to benefit from combining SR and LRS methods. The development of improved enrichment methods for circRNA and bioinformatic pipelines is necessary to arrive at a high-confidence set of circRNAs from LRS. The results provide a first set of 5924 circRNAs from *L. japonicus* for functional analysis.

AUTHOR CONTRIBUTIONS

Asa Budnick: Data curation; formal analysis; investigation; methodology; software; validation; visualization; writing—original draft; writing—review and editing. **Megan Franklin:** Data curation; formal analysis; investigation; methodology; software; validation; visualization; writing—original draft; writing—review and editing. **Delecia Utley:** Data curation; investigation; methodology; validation; writing—review and editing. **Brianne Edwards:** Data curation; methodology; software; writing—review and editing. **Melodi Charles:** Data curation; software; writing—review and editing. **Eli Hornstein:** Methodology; writing—review and editing. **Heike I Sederoff:** Conceptualization; formal analysis; funding acquisition; investigation; project administration; supervision; visualization; writing—original draft; writing—review and editing.

ACKNOWLEDGMENTS

We thank Swathi Barampuram and Pooja Narasimhan for their technical support in the lab. Funding for this work was provided by the Novo Nordisk Foundation (InRoot NNF19SA0059362); the Department of Energy, Office of Science (DE-SC0018269); a National Science Foundation NRT Fellowship to Asa Budnick, Delecia Utley, and Eli D. Hornstein (DGE-1828820); a USDoEd GAANN fellowship (P200A210002) to Melodi Charles; an NIH Molecular Biotechnology Training Grant # 1T32GM133366-01 Fellowship to Melodi Charles; Southern Regional Educational Board Fellowship to Delecia Utley; and North Carolina State University Genetics and Genomics Scholar fellowship to Asa Budnick.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Raw sequencing data from this experiment are available via NCBI SRA: (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1024973>). Additional sequencing data from younger leaf tissue CEnLR were uploaded to the SRA but are not included in this analysis. Custom R scripts used for the formatting and analysis of data are available via Github (https://github.com/SederoffLab/Lj_Leaf_circRNA_Code). Other data are available in Supporting Information, Data Dryad (<https://doi.org/10.5061/dryad.hqbzkh1q2>), or via contacting the authors.

ORCID

Heike Sederoff  <https://orcid.org/0000-0002-0960-9678>

REFERENCES

- Andrews, S. (2010). *Babraham bioinformatics—FastQC: A quality control tool for high throughput sequence data*. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Barbrook, A. C., Howe, C. J., Kurniawan, D. P., & Tarr, S. J. (2010). Organization and expression of organellar genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1541), 785–797. <https://doi.org/10.1098/rstb.2009.0250>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Broughton, W. J., & Dilworth, M. J. (1971). Control of leghaemoglobin synthesis in snake beans. *Biochemical Journal*, 125(4), 1075–1080. <https://doi.org/10.1042/bj1251075>
- Cashmore, A. R., Jarillo, J. A., Wu, Y. J., & Liu, D. (1999). Cryptochromes: Blue light receptors for plants and animals. *Science*, 284(5415), 760–765. <https://doi.org/10.1126/science.284.5415.760>
- Chen, L.-L. (2020). The expanding regulatory mechanisms and cellular functions of circular RNAs. *Nature Reviews Molecular Cell Biology*, 21(8), 475–490. <https://doi.org/10.1038/s41580-020-0243-y>
- Christie, J. M. (2007). Phototropin blue-light receptors. *Annual Review of Plant Biology*, 58, 21–45. <https://doi.org/10.1146/annurev.arplant.58.032806.103951>
- Chuang, T.-J., Chen, Y.-J., Chen, C.-Y., Mai, T.-L., Wang, Y.-D., Yeh, C.-S., Yang, M.-Y., Hsiao, Y.-T., Chang, T.-H., Kuo, T.-C., Cho, H.-H., Shen, C.-N., Kuo, H.-C., Lu, M.-Y., Chen, Y.-H., Hsieh, S.-C., & Chiang, T.-W. (2018). Integrative transcriptome sequencing reveals extensive alternative trans-splicing and cis-backsplicing in human cells. *Nucleic Acids Research*, 46(7), 3671–3691. <https://doi.org/10.1093/nar/gky032>
- Conn, V., & Conn, S. J. (2019). SplintQuant: A method for accurately quantifying circular RNA transcript abundance without reverse transcription bias. *RNA*, 25(9), 1202–1210. <https://doi.org/10.1261/rna.070953.119>
- Conn, V. M., Hugouvieux, V., Nayak, A., Conos, S. A., Capovilla, G., Cildir, G., Jourdain, A., Tergaonkar, V., Schmid, M., Zubietta, C., & Conn, S. J. (2017). A circRNA from SEPALLATA3 regulates splicing of its cognate mRNA through R-loop formation. *Nature Plants*, 3(5), 1–5. <https://doi.org/10.1038/nplants.2017.53>

- Dodbele, S., Mutlu, N., & Wilusz, J. E. (2021). Best practices to ensure robust investigation of circular RNAs: Pitfalls and tips. *EMBO Reports*, 22(3), e52072. <https://doi.org/10.15252/embr.202052072>
- Gaffo, E., Bonizzato, A., te Kronnie, G., & Bortoluzzi, S. (2017). CirComPara: A multi-method comparative bioinformatics pipeline to detect and study circRNAs from RNA-seq data. *Non-Coding RNA*, 3(1), 8. <https://doi.org/10.3390/ncrna3010008>
- Gao, Y., Wang, J., & Zhao, F. (2015). CIRI: An efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biology*, 16(1), 4. <https://doi.org/10.1186/s13059-014-0571-3>
- Gao, Y., Zhang, J., & Zhao, F. (2018). Circular RNA identification based on multiple seed matching. *Briefings in Bioinformatics*, 19(5), 803–810. <https://doi.org/10.1093/bib/bbx014>
- Gao, Z., Li, J., Luo, M., Li, H., Chen, Q., Wang, L., Song, S., Zhao, L., Xu, W., Zhang, C., Wang, S., & Ma, C. (2019). Characterization and cloning of grape circular RNAs identified the cold resistance-related Vv-circATS11. *Plant Physiology*, 180(2), 966–985. <https://doi.org/10.1104/pp.18.01331>
- Geiss, G. K., Bumgarner, R. E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D. L., Fell, H. P., Ferree, S., George, R. D., Grogan, T., James, J. J., Maysuria, M., Mitton, J. D., Oliveri, P., Osborn, J. L., Peng, T., Ratcliffe, A. L., Webster, P. J., Davidson, E. H., ... Dimitrov, K. (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotechnology*, 26(3), 317–325. <https://doi.org/10.1038/nbt1385>
- Haubold, B., & Wiehe, T. (2006). How repetitive are genomes? *BMC Bioinformatics*, 7(1), 541. <https://doi.org/10.1186/1471-2105-7-541>
- Inoue, T., Higuchi, M., Hashimoto, Y., Seki, M., Kobayashi, M., Kato, T., Tabata, S., Shinozaki, K., & Kakimoto, T. (2001). Identification of CRE1 as a cytokinin receptor from Arabidopsis. *Nature*, 409(6823), 1060–1063. <https://doi.org/10.1038/35059117>
- Jeck, W. R., & Sharpless, N. E. (2014). Detecting and characterizing circular RNAs. *Nature Biotechnology*, 32(5), 453–461. <https://doi.org/10.1038/nbt.2890>
- Kamal, N., Mun, T., Reid, D., Lin, J., Akyol, T. Y., Sandal, N., Asp, T., Hirakawa, H., Stougaard, J., Mayer, K. F. X., Sato, S., & Andersen, S. U. (2020). Insights into the evolution of symbiosis gene copy number and distribution from a chromosome-scale *Lotus japonicus* Gifu genome sequence. *DNA Research*, 27(3), dsaa015. <https://doi.org/10.1101/2020.04.17.042473>
- Kato, T., Kaneko, T., Sato, S., Nakamura, Y., & Tabata, S. (2000). Complete structure of the chloroplast genome of a legume, *Lotus japonicus*. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, 7(6), 323–330. <https://doi.org/10.1093/dnares/7.6.323>
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 37(8), 907–915. <https://doi.org/10.1038/s41587-019-0201-4>
- Kleaveland, B., Shi, C. Y., Stefano, J., & Bartel, D. P. (2018). A network of noncoding regulatory RNAs acts in the mammalian brain. *Cell*, 174(2), 350–362.e17. <https://doi.org/10.1016/j.cell.2018.05.022>
- Kück, U., & Schmitt, O. (2021). The chloroplast trans-splicing RNA-protein supercomplex from the green alga *Chlamydomonas reinhardtii*. *Cells*, 10(2), 290. <https://doi.org/10.3390/cells10020290>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Liang, D., & Wilusz, J. E. (2014). Short intronic repeat sequences facilitate circular RNA production. *Genes & Development*, 28(20), 2233–2247. <https://doi.org/10.1101/gad.251926.114>
- Liao, X., Li, X.-J., Zheng, G.-T., Chang, F.-R., Fang, L., Yu, H., Huang, J., & Zhang, Y.-F. (2022). Mitochondrion-encoded circular RNAs are widespread and translatable in plants. *Plant Physiology*, 189(3), 1482–1500. <https://doi.org/10.1093/plphys/kiac143>
- Liu, H., Yu, W., Wu, J., Li, Z., Li, H., Zhou, J., Hu, J., & Lu, Y. (2020). Identification and characterization of circular RNAs during wood formation of poplars in acclimation to low nitrogen availability. *Planta*, 251(2), 47. <https://doi.org/10.1007/s00425-020-03338-w>
- Liu, R., Ma, Y., Guo, T., & Li, G. (2022). Identification, biogenesis, function, and mechanism of action of circular RNAs in plants. *Plant Communications*, 4(1), 100430. <https://doi.org/10.1016/j.xplc.2022.100430>
- Liu, S., Wang, Q., Li, X., Wang, G., & Wan, Y. (2019). Detecting of chloroplast circular RNAs in Arabidopsis thaliana. *Plant Signaling & Behavior*, 14(8), 1621088. <https://doi.org/10.1080/15592324.2019.1621088>
- Ma, X.-K., Wang, M.-R., Liu, C.-X., Dong, R., Carmichael, G. G., Chen, L.-L., & Yang, L. (2019). CIRCexplorer3: A CLEAR pipeline for direct comparison of circular and linear RNA expression. *Genomics, Proteomics & Bioinformatics*, 17(5), 511–521. <https://doi.org/10.1016/j.gpb.2019.11.004>
- Ma, X.-K., Zhai, S.-N., & Yang, L. (2023). Approaches and challenges in genome-wide circular RNA identification and quantification. *Trends in Genetics*, 39(12), 897–907. <https://doi.org/10.1016/j.tig.2023.09.006>
- Małolepszy, A., Mun, T., Sandal, N., Gupta, V., Dubin, M., Urbański, D., Shah, N., Bachmann, A., Fukai, E., Hirakawa, H., Tabata, S., Nadzieja, M., Markmann, K., Su, J., Umehara, Y., Soyano, T., Miyahara, A., Sato, S., Hayashi, M., ... Andersen, S. U. (2016). The LORE1 insertion mutant resource. *The Plant Journal*, 88(2), 306–317. <https://doi.org/10.1111/tjp.13243>
- Mi, H., Muruganujan, A., & Thomas, P. D. (2013). PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research*, 41(Database issue), D377–386. <https://doi.org/10.1093/nar/gks1118>
- Mun, T., Bachmann, A., Gupta, V., Stougaard, J., & Andersen, S. U. (2016). Lotus Base: An integrated information portal for the model legume *Lotus japonicus*. *Scientific Reports*, 6(1), 39447. <https://doi.org/10.1038/srep39447>
- Nielsen, A. F., Bindereif, A., Bozzoni, I., Hanan, M., Hansen, T. B., Irimia, M., Kadener, S., Kristensen, L. S., Legnini, I., Morlando, M., Jarlstad Olesen, M. T., Pasterkamp, R. J., Preibisch, S., Rajewsky, N., Suenkel, C., & Kjems, J. (2022). Best practice standards for circular RNA research. *Nature Methods*, 19(10), 1208–1220. <https://doi.org/10.1038/s41592-022-01487-2>
- Nilson, S. E., & Assmann, S. M. (2010). The α -subunit of the Arabidopsis heterotrimeric G protein, GPA1, is a regulator of transpiration efficiency 1[C][W][OA]. *Plant Physiology*, 152(4), 2067–2077. <https://doi.org/10.1104/pp.109.148262>
- Panda, A. C., De, S., Grammatikakis, I., Munk, R., Yang, X., Piao, Y., Dudekula, D. B., Abdelmohsen, K., & Gorospe, M. (2017). High-purity circular RNA isolation method (RPAD) reveals vast collection of intronic circRNAs. *Nucleic Acids Research*, 45(12), e116. <https://doi.org/10.1093/nar/gkx297>
- Philips, A., Nowis, K., Stelmaszczuk, M., Jackowiak, P., Podkowiński, J., Handschuh, L., & Figlerowicz, M. (2020). Expression landscape of

- circRNAs in *Arabidopsis thaliana* seedlings and adult tissues. *Frontiers in Plant Science*, 11, 576581. <https://www.frontiersin.org/article/10.3389/fpls.2020.576581>
- Rahimi, K., Færch Nielsen, A., Venø, M. T., & Kjems, J. (2021). 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., & Kjems, J. (2021). Nanopore sequencing of brain-derived full-length circRNAs reveals circRNA-specific exon usage, intron retention and microexons. *Nature Communications*, 12(1), 4825. <https://doi.org/10.1038/s41467-021-24975-z>
- Rawat, R., Schwartz, J., Jones, M. A., Sairanen, I., Cheng, Y., Andersson, C. R., Zhao, Y., Ljung, K., & Harmer, S. L. (2009). REVEILLE1, a Myb-like transcription factor, integrates the circadian clock and auxin pathways. *Proceedings of the National Academy of Sciences of the United States of America*, 106(39), 16883–16888. <https://doi.org/10.1073/pnas.0813035106>
- Remy, E., Cabrito, T. R., Baster, P., Batista, R. A., Teixeira, M. C., Friml, J., Sá-Correia, I., & Duque, P. (2013). A major facilitator superfamily transporter plays a dual role in polar auxin transport and drought stress tolerance in *Arabidopsis*. *The Plant Cell*, 25(3), 901–926. <https://doi.org/10.1105/tpc.113.110353>
- Sanger, H. L., Klotz, G., Riesner, D., Gross, H. J., & Kleinschmidt, A. K. (1976). Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proceedings of the National Academy of Sciences of the United States of America*, 73(11), 3852–3856.
- Staff, T. P. O. (2014). Correction: Circular RNA is expressed across the eukaryotic tree of life. *PLoS One*, 9(4), e95116. <https://doi.org/10.1371/journal.pone.0095116>
- Stoll, L., Rodríguez-Trejo, A., Guay, C., Brozzi, F., Bayazit, M. B., Gattesco, S., Menoud, V., Sobel, J., Marques, A. C., Venø, M. T., Esguerra, J. L. S., Barghouth, M., Suleiman, M., Marselli, L., Kjems, J., Eliasson, L., Renström, E., Bouzakri, K., Pinget, M., ... Regazzi, R. (2020). A circular RNA generated from an intron of the insulin gene controls insulin secretion. *Nature Communications*, 11(1), 5611. <https://doi.org/10.1038/s41467-020-19381-w>
- Stortchevoi, A., Kamelamel, N., & Levine, S. S. (2020). SPRI beads-based size selection in the range of 2–10 kb. *Journal of Biomolecular Techniques*, 31(1), 7–10. <https://doi.org/10.7171/jbt.20-3101-002>
- Szabo, L., & Salzman, J. (2016). Detecting circular RNAs: Bioinformatic and experimental challenges. *Nature Reviews Genetics*, 17(11), 679–692. <https://doi.org/10.1038/nrg.2016.114>
- Tang, C., Yu, T., Xie, Y., Wang, Z., McSwiggin, H., Zhang, Y., Zheng, H., & Yan, W. (2018). Template switching causes artificial junction formation and false identification of circular RNAs. *BioRxiv*. <https://doi.org/10.1101/259556>
- Tremblay, B. J.-M., & Nystrom, S. (2023). universalmotif: Import, modify, and export motifs with R (1.18.1) [Computer software]. Bioconductor version: Release (3.17). <https://doi.org/10.18129/B9.bioc.universalmotif>
- Udvardi, M. K., Tabata, S., Parniske, M., & Stougaard, J. (2005). Lotus japonicus: Legume research in the fast lane. *Trends in Plant Science*, 10(5), 222–228. <https://doi.org/10.1016/j.tplants.2005.03.008>
- Urbański, D. F., Małolepszy, A., Stougaard, J., & Andersen, S. U. (2012). Genome-wide LORE1 retrotransposon mutagenesis and high-throughput insertion detection in *Lotus japonicus*. *The Plant Journal: For Cell and Molecular Biology*, 69(4), 731–741. <https://doi.org/10.1111/j.1365-3113.2011.04827.x>
- Vromman, M., Anckaert, J., Bortoluzzi, S., Buratin, A., Chen, C.-Y., Chu, Q., Chuang, T.-J., Dehghannasiri, R., Dieterich, C., Dong, X., Flicek, P., Gaffo, E., Gu, W., He, C., Hoffmann, S., Izuogu, O., Jackson, M. S., Jakobi, T., Lai, E. C., ... Volders, P.-J. (2023). Large-scale benchmarking of circRNA detection tools reveals large differences in sensitivity but not in precision. *Nature Methods*, 20(8), 1159–1169. <https://doi.org/10.1038/s41592-023-01944-6>
- Wang, Y., Wang, H., Xi, F., Wang, H., Han, X., Wei, W., Zhang, H., Zhang, Q., Zheng, Y., Zhu, Q., Kohlen, M. V., Reddy, A. S. N., & Gu, L. (2020). Profiling of circular RNA N6-methyladenosine in moso bamboo (*Phyllostachys edulis*) using nanopore-based direct RNA sequencing. *Journal of Integrative Plant Biology*, 62(12), 1823–1838. <https://doi.org/10.1111/jipb.13002>
- Wang, Z., Liu, Y., Li, D., Li, L., Zhang, Q., Wang, S., & Huang, H. (2017). Identification of circular RNAs in kiwifruit and their species-specific response to bacterial canker pathogen invasion. *Frontiers in Plant Science*, 8, 413. <https://www.frontiersin.org/articles/10.3389/fpls.2017.00413>
- Weinl, S., Held, K., Schlücking, K., Steinhörst, L., Kuhlert, S., Hippler, M., & Kudla, J. (2008). A plastid protein crucial for Ca^{2+} -regulated stomatal responses. *The New Phytologist*, 179(3), 675–686. <https://doi.org/10.1111/j.1469-8137.2008.02492.x>
- Wu, Z., Huang, W., Qin, E., Liu, S., Liu, H., Grennan, A. K., Liu, H., & Qin, R. (2020). Comprehensive identification and expression profiling of circular RNAs during nodule development in *Phaseolus vulgaris*. *Frontiers in Plant Science*, 11, 587185. <https://doi.org/10.3389/fpls.2020.587185>
- Xia, K., Pan, X., Zeng, X., & Zhang, M. (2023). Xoo-responsive transcriptome reveals the role of the circular RNA133 in disease resistance by regulating expression of OsARAB in rice. *Phytopathology Research*, 5(1), 33. <https://doi.org/10.1186/s42483-023-00188-8>
- Xiao, M.-S., & Wilusz, J. E. (2019). An improved method for circular RNA purification using RNase R that efficiently removes linear RNAs containing G-quadruplexes or structured 3' ends. *Nucleic Acids Research*, 47(16), 8755–8769. <https://doi.org/10.1093/nar/gkz576>
- Xin, R., Gao, Y., Gao, Y., Wang, R., Kadash-Edmondson, K. E., Liu, B., Wang, Y., Lin, L., & Xing, Y. (2021). isoCirc catalogs full-length circular RNA isoforms in human transcriptomes. *Nature Communications*, 12(1), 266. <https://doi.org/10.1038/s41467-020-20459-8>
- Xu, X., Du, T., Mao, W., Li, X., Ye, C.-Y., Zhu, Q.-H., Fan, L., & Chu, Q. (2022). PlantcircBase 7.0: Full-length transcripts and conservation of plant circRNAs. *Plant Communications*, 3(4), 100343. <https://doi.org/10.1016/j.xplc.2022.100343>
- Ye, C. Y., Zhang, X., Chu, Q., Liu, C., Yu, Y., Jiang, W., Zhu, Q. H., Fan, L., & Guo, L. (2017). Full-length sequence assembly reveals circular RNAs with diverse non-GT/AG splicing signals in rice. *RNA Biology*, 14(8), 1055–1063.
- Yin, J., Liu, Y., Lu, L., Zhang, J., Chen, S., & Wang, B. (2022). Comparison of tolerant and susceptible cultivars revealed the roles of circular RNAs in rice responding to salt stress. *Plant Growth Regulation*, 96(2), 243–254. <https://doi.org/10.1007/s10725-021-00772-y>
- Yu, C.-Y., Liu, H.-J., Hung, L.-Y., Kuo, H.-C., & Chuang, T.-J. (2014). Is an observed non-co-linear RNA product spliced in trans, in cis or just in vitro? *Nucleic Acids Research*, 42(14), 9410–9423. <https://doi.org/10.1093/nar/gku643>
- Zavaliev, R., Ueki, S., Epel, B. L., & Citovsky, V. (2011). Biology of callose (β -1,3-glucan) turnover at plasmodesmata. *Protoplasma*, 248(1), 117–130. <https://doi.org/10.1007/s00709-010-0247-0>

- Zhang, H., Liu, S., Li, X., Yao, L., Wu, H., Baluška, F., & Wan, Y. (2021). An antisense circular RNA regulates expression of RuBisCO small subunit genes in Arabidopsis. *Frontiers in Plant Science*, 12, 665014. <https://www.frontiersin.org/article/10.3389/fpls.2021.665014>
- Zhang, J. (2021). Welcome to CIRI Toolkit's documentation!—CIRI-cookbook 1.1.0 documentation. <https://ciri-cookbook.readthedocs.io/en/latest/index.html>
- Zhang, J., Hou, L., Zuo, Z., Ji, P., Zhang, X., Xue, Y., & Zhao, F. (2021). Comprehensive profiling of circular RNAs with nanopore sequencing and CIRI-long. *Nature Biotechnology*, 39(7), 836–845. <https://doi.org/10.1038/s41587-021-00842-6>
- Zhang, P., Fan, Y., Sun, X., Chen, L., Terzaghi, W., Bucher, E., Li, L., & Dai, M. (2019). A large-scale circular RNA profiling reveals universal molecular mechanisms responsive to drought stress in maize and Arabidopsis. *The Plant Journal*, 98(4), 697–713. <https://doi.org/10.1111/tpj.14267>
- Zhang, X.-O., Wang, H.-B., Zhang, Y., Lu, X., Chen, L.-L., & Yang, L. (2014). Complementary sequence-mediated exon circularization. *Cell*, 159(1), 134–147. <https://doi.org/10.1016/j.cell.2014.09.001>
- Zhang, Y., Yang, L., & Chen, L.-L. (2016). Characterization of circular RNAs. In Y. Feng & L. Zhang (Eds.), *Long non-coding RNAs: Methods and protocols* (pp. 215–227). Springer. https://doi.org/10.1007/978-1-4939-3378-5_17
- Zhang, Y., Zhang, X.-O., Chen, T., Xiang, J.-F., Yin, Q.-F., Xing, Y.-H., Zhu, S., Yang, L., & Chen, L.-L. (2013). Circular intronic long noncoding RNAs. *Molecular Cell*, 51(6), 792–806. <https://doi.org/10.1016/j.molcel.2013.08.017>
- Zhang, Z., Wang, H., Wang, Y., Xi, F., Wang, H., Kohnen, M. V., Gao, P., Wei, W., Chen, K., Liu, X., Gao, Y., Han, X., Hu, K., Zhang, H., Zhu, Q., Zheng, Y., Liu, B., Ahmad, A., Hsu, Y.-H., ... Gu, L. (2021). Whole-genome characterization of chronological age-associated changes in methylome and circular RNAs in moso bamboo (*Phyllostachys edulis*) from vegetative to floral growth. *The Plant Journal*, 106(2), 435–453. <https://doi.org/10.1111/tpj.15174>
- Zhou, J., Yuan, M., Zhao, Y., Quan, Q., Yu, D., Yang, H., Tang, X., Xin, X., Cai, G., Qian, Q., Qi, Y., & Zhang, Y. (2021). Efficient deletion of multiple circle RNA loci by CRISPR-Cas9 reveals Os06circ02797 as a putative sponge for OsMIR408 in rice. *Plant Biotechnology Journal*, 19(6), 1240–1252. <https://doi.org/10.1111/pbi.13544>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Budnick, A., Franklin, M. J., Utley, D., Edwards, B., Charles, M., Hornstein, E. D., & Sederoff, H. (2024). Long- and short-read sequencing methods discover distinct circular RNA pools in *Lotus japonicus*. *The Plant Genome*, 17, e20429. <https://doi.org/10.1002/tpg2.20429>