



Genome-Wide Identification of Circular RNAs in *Arabidopsis* thaliana

Gang Chen¹, Jiawen Cui², Li Wang², Yingfang Zhu³, Zhaogeng Lu⁴ and Biao Jin^{2,5*}

¹ College of Bioscience and Biotechnology, Yangzhou University, Yangzhou, China, ² College of Horticulture and Plant Protection, Yangzhou University, Yangzhou, China, ³ Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN, United States, ⁴ Agricultural College, Yangzhou University, Yangzhou, China, ⁵ Key Laboratory of Plant Functional Genomics of the Ministry of Education, Yangzhou, China

Circular RNAs (circRNAs) are a family of transcripts with covalently closed circular structures and still largely unknown functions. Large numbers of circRNAs have been found in various biological processes in humans and animals, but fewer circRNAs have been identified in plants. We performed a genome-wide analysis of circRNAs in Arabidopsis thaliana via deep sequencing. We constructed 14 strand-specific libraries from 13 samples of plants from four developmental stages, four stress treatments, and five organs and a mixed sample across the lifespan. In total, we identified 5861 circRNAs, including 1275 novel ones, using the strict threshold of at least two unique back-spliced supporting reads. The circRNAs were non-randomly distributed in all chromosomes; most were exonic. Sequence similarity analysis of circRNAs between A. thaliana and four other species showed that some circRNAs are conserved in plants. Functional annotation indicated that many parent genes of circRNAs are involved in many fundamental processes including plant development, reproduction, and response to stimulus. In addition, a small proportion of circRNAs was shown to be potential targets of miRNAs, indicating that the circRNAs could interact with miRNAs to regulate gene expression. gRT-PCR analysis revealed that circRNAs displayed diverse expression patterns at different growth stages. Our results provide an important resource for continuing circRNA research in A. thaliana, and should enhance our understanding of circRNAs in plants.

OPEN ACCESS

Edited by:

Mario Alberto Arteaga-Vazquez, Universidad Veracruzana, Mexico

Reviewed by:

Turgay Unver, iBG-Izmir, International Biomedicine and Genome Institute, Turkey Alfredo Cruz-Ramirez, Laboratorio Nacional de Genómica para la Biodiversidad, Mexico

> ***Correspondence:** Biao Jin bjin@yzu.edu.cn

Specialty section:

This article was submitted to Plant Genetics and Genomics, a section of the journal Frontiers in Plant Science

Received: 27 January 2017 Accepted: 12 September 2017 Published: 27 September 2017

Citation:

Chen G, Cui J, Wang L, Zhu Y, Lu Z and Jin B (2017) Genome-Wide Identification of Circular RNAs in Arabidopsis thaliana. Front. Plant Sci. 8:1678. doi: 10.3389/fpls.2017.01678 Keywords: circular RNA, parent gene, Arabidopsis, RNA-seq, sequence conservation, microRNA

INTRODUCTION

There are two types of RNA in eukaryotic cells: messenger RNA (mRNA) and non-coding RNA. Non-coding RNA includes highly abundant and functionally important RNAs, such as transfer RNA (tRNA) and ribosomal RNA (rRNA), as well as microRNA (miRNA), small interfering RNA (siRNA), *trans*-acting siRNA (tasiRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), and long non-coding RNA (lncRNA). Circular RNAs (circRNAs), which are generally recognized as another group of non-coding RNAs, are characterized by the presence of a covalent bond linking the 3' and 5'-ends generated by back-splicing (Jeck et al., 2013). CircRNAs arise from back-spliced exons, introns as circular intronic RNAs (ciRNAs), intergenic regions, or both

1

exons and introns as exon-intron circRNAs (EIciRNAs) (Zhang et al., 2013; Li et al., 2015). In past decades, circRNAs have only been occasionally identified due to limitations of molecular techniques and bioinformatic tools (Sanger et al., 1976; Hsu and Coca-Prados, 1979; Grabowski et al., 1981; Kos et al., 1986; Nigro et al., 1991), and were thought to be functionless byproducts of transcription. In recent years, with the development of high-throughput sequencing technology and high-efficiency large-scale data analysis, many circRNAs have been identified in archaea, *Caenorhabditis elegans*, mouse, and human (Danan et al., 2012; Salzman et al., 2012; Memczak et al., 2013; Zhang et al., 2013); knowledge of circRNAs has been updated continuously and rapidly.

Recent research has suggested that circRNAs are widely expressed in a complex tissue-, cell-type- or developmental-stagespecific manner, and their sequences and expression patterns are often conserved (Salzman et al., 2012; Jeck et al., 2013; Memczak et al., 2013; Zhang et al., 2013, 2014). The conservation feature of circRNAs exist in various species (Wang et al., 2014; Zhao et al., 2017), and some conserved circRNAs were derived from important gene loci suggesting their potentially important functions (Vicens and Westhof, 2014; Gao et al., 2015). Many circRNA expression levels can be approximately 10-fold compared with their linear isoforms, suggesting that the formation of circRNAs may be finely regulated (Rybak-Wolf et al., 2015; Venø et al., 2015; Dang et al., 2016). Furthermore, circRNAs play important biological roles in transcriptional and post-transcriptional regulatory networks governing gene expression. Functional circRNAs have been shown to act as cytoplasmic miRNA sponges that sequester miRNAs (Hansen et al., 2013; Memczak et al., 2013) and RNA-binding protein sequestering agents (Zhang et al., 2013) as well as nuclear transcriptional regulators (Li et al., 2015). In addition, the production of circRNAs plays an important role in the regulation of alternative splicing (Zhang et al., 2014; Gao et al., 2016). Recent studies have shown that circRNAs serve as proteincoding sequences in vitro and in vivo (Granados-Riveron and Aquino-Jarquin, 2016; Legnini et al., 2017; Pamudurti et al., 2017; Yang et al., 2017). Studies have also demonstrated the important role of circRNAs as biomarkers of disease, including Alzheimer's disease (AD) and cancer (Lukiw, 2013; Li et al., 2015).

Although it has been demonstrated that circRNAs play important roles in a range of biological and developmental processes in animals (Memczak et al., 2013; Salzman et al., 2013), the knowledge of circRNAs in plants is limited (Wang et al., 2014; Lu et al., 2015; Ye et al., 2015; Darbani et al., 2016; Sablok et al., 2016; Wang et al., 2017). Wang et al. (2014) reported that circRNA is expressed in a wide range of eukaryotic species, including the model plant species *Arabidopsis thaliana*. Ye et al. (2015) used publicly available RNA sequencing data and anchor aligned the reads to the model plants *Oryza sativa* and *A. thaliana*. A study in *O. sativa* by Lu et al. (2015) identified 2,354 circRNAs, among which 1,356 were exonic circRNAs. Recently, a plant circRNA database (PlantcircBase) was created by collecting publicly available and newly identified circRNA sequences from O. sativa, A. thaliana, Zea mays, Solanum lycopersicum, and Hordeum vulgare (Chu et al., 2017).

Here, to uncover a more comprehensive profile of circRNAs in the plant model species A. thaliana, we identified and analyzed circRNAs from 14 ssRNA-seq (strand-specific sequencing of RNA) libraries from RNA samples corresponding to four different growth stages, five organs, and four stress treatments, as well as a mixed RNA sample, in A. thaliana. We then used deep sequencing by the Illumina ssRNA-seq approach: our average sequencing depth reached approximately 106× and we acquired high-quality sequence information for circRNAs. In particular, we used the strict threshold of at least two unique back-spliced supporting reads to positively identify circRNAs. In addition, functional annotation analysis and circRNA-originating target miRNA predictions were made to predict the function of circRNA in A. thaliana. Our data provide a genome-wide profiling of A. thaliana circRNAs and provide an important resource for future circRNA research in A. thaliana.

MATERIALS AND METHODS

Plant Materials

Seeds of *A. thaliana* [wild-type Columbia (Col-0)] were exposed to stratification for 2 days at 4°C, and sown in square surfacesterilized plastic pots (7 × 7 × 8 cm) containing sterile medium (1:1 v/v mixture of vermiculite and peat). Pots were arranged in a plastic pallet and placed in a growth chamber (23°C during the day and 18°C at night, with a 16-h photoperiod and 500 μ mol m⁻² s⁻¹ of photosynthetically active radiation (PAR); the plants were alternately watered to saturation with distilled water or 1/2 Murashige–Skoog solution (Jin et al., 2011).

Fourteen plant materials for RNA isolation were prepared as follows (Table 1). Plants at growth stages 1.04 (4 rosette leaves > 1 mm in length), 1.14 (14 rosette leaves > 1 mm in length), 3.90 (rosette growth complete), 5.10 (first flower buds visible), 6.50 (50% of flowers to be produced have opened), 8.00 (first silique shattered), and 9.70 (senescence complete; ready for seed harvest) (Boyes et al., 2001) were flash-frozen in liquid nitrogen. We used three plants in most of the growth stages, but five plants were sampled in the first stage for isolation of sufficient RNA. For organ sampling, we collected roots, stems, leaves, flowers, and siliques from three plants at stage 6.50. In addition, seedlings at stage 3.90 were exposed to the following stress environments: 300 mM mannitol (drought), 200 mM NaCl (salinity), and 38°C (heat). Plants grown in the same environment (3.90), without the additional stress component, were used as controls. The whole plants were collected 12 h after exposure to stress.

RNA Extraction and RNase R Treatment

Total RNAs from all samples above were extracted using the MiniBEST Plant RNA Extraction Kit (Takara, Dalian, China), in accordance with the manufacturer's protocol. We stored the total RNA aliquots at -80° C with 1 unit/µL RNaseOUT (Invitrogen), validated the RNA quality using an Agilent 2100 Bioanalyzer, and quantified the total RNA with Qubit 2.0 Fluorometer

Sample No.	Description			
Arab_mix	Mixed sample from whole plants at seven growth stages: 1.04 (4 rosette leaves > 1 mm in length), 1.14 (14 rosette leaves > 1 mm in length), 3.90 (rosette growth complete), 5.10 (first flower buds visible), 6.50 (50% of flowers to be produced have opened), 8.00 (first silique shattered), and 9.70 (senescence complete; ready for seed harvest).			
Arab_B673-T01	Three whole plants at growth stage 1.14			
Arab_B673-T02	Three whole plants at growth stage 3.90			
Arab_B673-T03	Three whole plants at growth stage 6.50			
Arab_B673-T04	Three whole plants at growth stage 8.00			
Arab_B673-T05	Roots from three plants at growth stage 6.50			
Arab_B673-T06	Stems from three plants at growth stage 6.50			
Arab_B673-T07	Leaves from three plants at growth stage 6.50			
Arab_B673-T08	Flowers from three plants at growth stage 6.50			
Arab_B673-T09	Siliques from three plants at growth stage 6.50			
Arab_B673-T10	Aerial part of three plants without any stress treatments at growth stage 3.90			
Arab_B673-T11	Aerial part of three plants with drought stress at growth stage 3.90			
Arab_B673-T12	Aerial part of three plants with salt stress at growth stage 3.90			
Arab_B673-T13	Aerial part of three plants with heat stress at growth stage 3.90			

(Invitrogen). The qualified total RNAs from the seven growth stages were equally mixed as a pool, while total RNAs from four different growth stages, five organs, and four stress treatments were separately prepared for library construction. Then, the 14 RNA samples were depleted of 18S and 28S rRNA using the Ribo-Zero Magnetic Kit (Epicentre, Madison, WI, United States) and treated with Ribonuclease R (Epicentre, Madison, WI, United States) to remove linear RNA.

Library Preparation and Illumina Sequencing

The rRNA-depleted RNAs were first fragmented into short sequences in fragmentation buffer at 94°C for 15 min. Second, these fragments were used as templates for first-strand cDNA synthesis primed with random hexamers. The second-strand cDNA was synthesized using buffer, dNTPs, RNase H, and DNA polymerase I, followed by AMPure XP bead purification. Third, the short fragments were further resolved with NEBNext End Repair Reaction Buffer and NEBNext Pre Enzyme Mix for end repair and poly (A) ligation. Subsequently, the fragments were connected with adapters, and then the second strand containing "U" was degraded using uracil-specific excision reagent (USER). Finally, suitable fragments were selected as templates for the PCR amplification after agarose gel electrophoresis. The quantification and qualification of the libraries were further assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, United States) and Qubit 2.0 Fluorometer (Invitrogen). These libraries were paired-end (PE) sequenced using the Illumina HiSeq 2500 (read length, PE125, the mixed sample) and Illumina Hiseq X Ten (read length, PE150, 13 independent samples)

platforms. All raw sequence data have been deposited in the NCBI Sequence Read Archive (SRA, accession number SRP069764).

Read Alignment and Analysis

We used the NGS QC Toolkit v2.3.3 software to filter the reads with adapter sequences, reads with an unknown base (N, unknown bases in a read > 5%), and lowquality reads (reads with base quality \leq 20). Sequence alignments were performed using TopHat2 with BOWTIE2 (v2.0.5) (Langmead and Salzberg, 2012) software for mapping to the reference genome (Arabidopsis_thaliana TAIR10.30 ENSEMBL). The circRNAs were predicted using the find circ program with the same parameters as described previously (Memczak et al., 2013). Chimeric mapped reads were selected for circular ("back-spliced") junctions if the sequence reads mapped to one chromosome on the same strand, while the two sequence segments mapped to the genomic region with reverse order. CircRNA abundance was predicted on the basis of circular junction read counts. The overlapping region between predicted circRNAs and the gene region of the A. thaliana reference genome was identified using Bedtools software (Quinlan and Hall, 2010), and the circRNAs with overlapping regions were annotated. All expression data and other information (coordinates, strand, etc.) of identified A. thaliana circRNAs are in GSE77672. To evaluate the conservation of circRNAs, the circRNA sequences of A. thaliana, O. sativa, H. vulgare, S. lycopersicum, and Z. mays were first downloaded from PlantcircBase¹. Then, our identified circRNA sequences (Supplementary Table S6) from A. thaliana were used for a BLAST search (BlastN in BLAST+, v2.2.27, $E < 1e^{-5}$) against the circRNA sequences from A. thaliana, O. sativa, H. vulgare, S. lycopersicum, and Z. mays.

To evaluate the potential functions of the parent genes of circRNAs, the parent mRNAs were used for BLAST searches and functionally categorized according to gene ontology (GO) annotation by BLAST2GO software with the default parameters² (Conesa and Götz, 2008). To predict miRNAcircRNA interactions, TargetFinder³ was used to scan the miRNA-target sites of circRNA. The network of miRNAscircRNAs was generated using Cytoscape 3.5.1 software.

PCR Amplification, Sanger Sequencing, and Quantitative Real-Time PCR

Treated RNA solutions (10 μ L, without DNA contamination) from seven stages, leaves, and roots were subjected to reverse transcriptase reactions with the PrimeScript RT Reagent Kit (Takara, Dalian, China), in accordance with the manufacturer's protocol. Genomic DNA of *A. thaliana* was extracted using the Plant DNA Isolation Reagent (Takara, Dalian, China). To validate circRNAs identified in *A. thaliana*, polymerase chain reactions (PCRs) were performed using a set of divergent primers and a set of convergent primers that were used as a control (Supplementary)

¹http://ibi.zju.edu.cn/plantcircbase/index.php

²https://www.blast2go.com/

³https://github.com/carringtonlab/TargetFinder

Table S1). The divergent primers were designed using an "outfacing" strategy to guarantee that the amplifications were from a circular template (Shen et al., 2015). For each PCR amplification, 20 ng of cDNA or genomic DNA was used with rTaq DNA polymerase and $10 \times$ buffer (Takara, Dalian, China), and 35 cycles of PCR were performed. To confirm the PCR results, the PCR products were dissected from a gel and purified using MiniBEST Agarose Gel DNA Extraction Kit Ver. 4.0 (Takara, Dalian, China). Sanger sequencing was performed by Sangon Biotech Company (Shanghai, China).

A gRT-PCR experiment was performed to test the expression levels of circRNAs. The qRT-PCR reactions contained 1 μ L of diluted cDNA, 400 nM of each primer, 10 μ L of the 2× TransStart Tip Green qPCR SuperMix, 0.4 µL of the 50× passive reference dye (TransGen Biotech, Beijing, China), and 7.8 µL of ddH₂O, for a final volume of 20 µL. The following qRT-PCR program was used: denaturation at 94°C for 30 s, followed by 40 cycles of 94°C for 5 s, 55.5°C for 30 s, and 72°C for 10 s. The divergent primers used in the experiment are shown in Supplementary Table S1. Amplification results were analyzed using the comparative Ct method, which uses the formula $2^{-\Delta\Delta C_{\rm T}}$. Each qRT-PCR experiment was carried out in three independent biological replicates and the standard errors of the mean among the replicates were calculated. Statistical analysis was conducted using the SPSS Statistics 18.0 software (IBM Corporation, Armonk, NY, United States). The significance of differences between leaves and roots was analyzed using a *t*-test at the probability level of 0.05.

RESULTS

Identification of CircRNAs in A. thaliana

RNAs were extracted from 14 samples (13 separate samples and 1 mixed sample from seven growth stages) of *A. thaliana*. rRNA and other linear RNA were removed using the Ribo-ZeroTM Magnetic Kit (Epicentre, Madison, WI, United States) and Ribonuclease R (Epicentre, Madison, WI, United States), respectively. The treated RNA samples were deep-sequenced, resulting in a total of 181.97 Gb of data (151.37 Gb from 13 samples and 30.6 Gb from a mixed sample; average sequencing depth reached approximately $106 \times$), and yielding 795.62 million raw reads that uniquely mapped to the reference genome (*Arabidopsis_thaliana* TAIR10.30 ENSEMBL). After filtering out the contaminated and low-quality reads, 763.81 million clean reads (96.00% of the total raw reads) mapped to the reference genome, indicating the relatively high quality of the sequenced samples (Supplementary Table S2).

Distribution of CircRNAs in the *A. thaliana* Genome

CircRNAs are recognizable based on back-spliced reads in rRNAdepleted RNA sequencing (RibominusSeq) data. RibominusSeq reads were first mapped to the reference and the circRNAs were then predicted using the find_circ program. In total, 5,861 circRNAs were identified from 14 samples of *A. thaliana*, using the strict threshold of at least two unique back-spliced supporting reads, and quantified. Upon comparison with all A. thaliana circRNAs from the plant circbase, 4,586 (78.25%) (Supplementary Table S3) of our identified circRNAs were found. Accordingly, the other 1,275 (21.75 %) circRNAs were regarded as candidate novel circRNAs (Supplementary Table S4). The identified circRNAs were generated from all of the chromosomes, as well as from the mitochondrial and chloroplast genomes (Figure 1). However, the circos plot revealed a non-random distribution of circRNAs in the chromosomes. Some chromosomal regions lacked circRNAs (Figure 1, black arrows), and some regions had a high density of circRNAs (Figure 1, red arrows). The regions with black arrows were around the centromere (the pericentromeric region), where there were low levels of gene density and gene expression (The Arabidopsis Genome Initiative, 2000). A total of 5,861 circRNAs and 54,873 back-spliced reads were identified. In addition, the numbers of circRNAs and back-spliced reads differed among the chromosomes (Table 2). There were only 93 circRNAs and 2,560 back-spliced reads on mitochondrial DNA, accounting for about 1.6 and 4.7% of the total numbers of circRNAs and backspliced reads, respectively. On chromosome 1 and chromosome 5, 1,260 circRNAs (21.5%) and 1,152 circRNAs (19.6%) were found, respectively. The circRNA numbers on chromosomes 2, 3, and 4, and chloroplast DNA ranged from 702 to 936. In addition, compared with other chromosomes, chromosome 5 had more back-spliced reads (15,420, 28.1%). This indicated that circRNAs are the most abundant on chromosomes 1 and 5, less abundant on chromosomes 2, 3, and 4 and chloroplast DNA, and sparsely distributed on mitochondrial DNA.

Classification and Conservation of CircRNAs in *A. thaliana*

Among the identified circRNAs, 4,990 (85.1%) were generated from exons of a protein-coding gene, that is, they were exonic circRNAs, meaning that both back-splice sites coincided with known exonic boundaries. A total of 650 circRNAs (11.1%) were derived from intergenic regions. Furthermore, only 221 (3.8%) of the circRNAs were generated by introns (Supplementary Table S5). These results indicated that circRNAs in *A. thaliana* were generated from different genomic regions (**Figure 1**) and mainly from coding regions. Additionally, we also found that most parent genes (2,546) could produce more than one circRNA, while other parent genes (770) produced only one circRNA (Supplementary Table S6).

We further performed reciprocal BLAST analysis to evaluate the conservation of circRNAs, and found that 2,364 and 494 identified circRNAs of *A. thaliana* had sequence similarity with 1,231 and 188 identified circRNAs from *O. sativa* and *S. lycopersicum*, respectively (BlastN, word_size 5, $E < 1e^{-5}$). Moreover, 23 and 51 circRNAs also shared sequence similarity with 8 and 37 identified circRNAs from *H. vulgare* and *Z. mays*. In addition, there were 404 circRNAs possessing similar sequences in *A. thaliana*, *O. sativa*, and *S. lycopersicum* (Supplementary Table S7). Interestingly, six circRNAs (Ath_circ_FC1780, Ath_circ_FC3573, Ath_circ_FC3780, Ath_circ_FC1779, Ath_circ_FC1980, and Ath_circ_FC3745) were found in



three other species (*O. sativa*, *S. lycopersicum*, and *Z. mays*). These results suggested that the sequences of some circRNAs are conserved among different plant species.

Functional Annotation Analysis of Parent Genes of CircRNAs in *A. thaliana*

Functional annotation analysis was performed to evaluate the potential functions of the parent genes of circRNAs. GO categories were assigned to the parent genes of circRNAs, and the genes were classified into three GO categories: cellular component, biological process, and molecular function (**Figure 2A**). In the biological process category,

TABLE 2 Distribution of circRNAs and their reads on different chromosomes	in
A. thaliana.	

Chromosome	CircRNAs	Percentage (%)	Reads	Percentage (%)
1	1,260	21.5	10,578	19.3
2	809	13.8	7,078	12.9
3	936	16.0	7,386	13.4
4	702	12.0	5,425	9.9
5	1,152	19.6	15,420	28.1
Mt	93	1.6	2,560	4.7
Pt	909	15.5	6,426	11.7
Total	5,861	100.0	54,873	100.0

Mt, mitochondria; Pt, chloroplasts.

some important GO subcategories such as "developmental process," "reproduction," and "response to stimulus" were annotated. In the molecular function category, the "antioxidant activity," "electron carrier activity," and "signal transducer activity" subcategories were also annotated. The biological interpretations of the circRNA parent genes were further analyzed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database. As shown in **Figure 2B**, the most highly represented pathways included biosynthesis of amino acids, ribosome, spliceosome, protein processing in endoplasmic reticulum, photosynthesis, and carbon metabolism, implying that many parent genes of circRNAs in *A. thaliana* are involved in protein synthesis and processing, photosynthesis, and carbon metabolism.

Prediction of miRNA Target Sites in CircRNAs

To determine whether circRNAs in A. thaliana could affect post-transcriptional regulation by binding to miRNAs and preventing them from regulating their target mRNAs, we identified miRNA target sites in circRNAs in A. thaliana, and found that 39 of 5,861 (0.67%) circRNAs had putative miRNA-binding sites (Supplementary Table S8). Of these 39 circRNAs, 9 had more than one different miRNA-binding site, and the greatest number of miRNA-binding sites (188) was found in Ath_circ_FC3355 (Figure 3A). Three circRNAs (Ath circ FC0322, Ath circ FC2235, and Ath circ FC5684) with the miRNA target sites and their pairing miRNA sequences are shown in Figures 3B-D. Most of the sequences of circRNAs and their miRNA target sites completely matched. In addition, we analyzed the numbers of circRNA with potential target sites for each miRNA. The results showed ath-miR5021 binding to 14 circRNAs, ath-miR414 binding to 13 circRNAs, and ath-miR5658 binding to 9 circRNAs (Supplementary Table S8). This indicated that some circRNAs in A. thaliana have many potential miRNA binding sites, and may be able to affect the post-transcriptional regulation of various genes.

Validation of CircRNAs in A. thaliana

To confirm our identification of circRNAs, five randomly selected circRNAs were used in experimental validation using reverse transcription PCR and Sanger sequencing. A set of divergent primers (Supplementary Table S1) was designed for each circRNA and used to amplify both cDNA and genomic DNA. It was expected that positive and negative results of amplification would be obtained for cDNA and genomic DNA, respectively (**Figure 4A**). As a control, convergent primers that should amplify the linear mRNAs were also designed for each circRNA used in verification. The amplified PCR products using divergent primers were sequenced to confirm the presence of the back-spliced junctions. As a result, three circRNAs were validated (Ath_circ_FC4757, Ath_circ_FC4468, and Ath_circ_FC2295) (**Figure 4B**).

In addition, we tested the expression levels of 10 selected circRNAs at different growth stages using qRT-PCR. The results showed that these circRNAs at different growth



stages had diverse expression patterns. For example, the expression level of Ath_circ_FC1408 increased from growth stages 1.04 to 3.90 and then decreased at growth stage 5.10. In addition, the peak value occurred at growth stage 6.50 whereas the lowest value was at growth stage 8.00 (**Figure 5A**). Moreover, we selected two circRNAs (Ath_circ_FC4468 and Ath_circ_FC5838) with parent genes related to photosynthesis to test their expression levels in leaves and roots. As shown in **Figures 5B,C**, both of the two circRNAs, as well as their parent genes, were expressed more highly in leaves than in roots.

DISCUSSION

The expression of some circRNAs is often cell-type, tissue, and developmental-stage-specific, suggesting that circRNAs are important functional molecules (Dang et al., 2016; Gao et al., 2016). Thousands of circRNAs have been identified in mammals, fish, worms, insects, fungi, protists, and human (Danan et al., 2012; Salzman et al., 2012; Memczak et al., 2013; Zhang et al., 2013; Wang et al., 2014; Dang et al., 2016), and this number is rapidly increasing. In plants, circRNAs in *A. thaliana*, *O. sativa*, tomato, and wheat were predicted from public RNA



sequencing data or RNA extracted from a plant at a particular stage of development (Lu et al., 2015; Ye et al., 2015; Zuo et al., 2016; Wang et al., 2017). Here, we report genome-wide identification and analysis of circRNAs from 14 ssRNA-seq libraries for RNA samples from four different growth stages, five organs, and four stress treatments, as well as a mixed RNA sample in *A. thaliana*. We obtained 5,861 circRNAs including 1,275 novel ones. Our data demonstrate the presence of circRNAs over the whole of the genome of *A. thaliana* and provide important complementary plant circRNAs for further classification, analysis of their properties, and functional research. In addition, sequence conservation analysis among four species

suggested that the conservation of circRNAs in plants is probably widespread.

CircRNAs are primarily derived from the exons of proteincoding genes, although they can also arise from intronic, intergenic, and untranslated regions, ncRNA loci, and from locations antisense to known transcripts (Jeck et al., 2013; Memczak et al., 2013; Ye et al., 2015). In our study, the predicted circRNAs were predominantly exonic circRNAs. As the exonic circRNAs and their parent genes exhibit significant positive correlations in their expression levels, this could be a result of *cis*-transcriptional promotion of circRNAs on their parent genes (Li et al., 2015). This suggests



that the non-random distribution of circRNAs on different chromosomes is associated with gene activity, especially in the pericentromeric region. In addition, exonic circRNAs may be subject to splicing in order to remove the intervening introns either before or after circularization. However, new research has demonstrated that EIciRNAs are specified by the presence of a retained intron, and can interact with U1 small nuclear ribonucleoproteins and enhance the expression of their parent genes (Li et al., 2015). Studies in rice have shown that multiple circRNAs can originate from a single gene (Lu et al., 2015). In this study, there were also multiple circRNAs from one parent gene, indicating that alternative back-splicing also occurs in A. thaliana. In addition, the emergence of different types of circRNA is also associated with alternative splicing (Zhang et al., 2014; Gao et al., 2016). As the alternative splicing events in circRNAs are not consistent with the corresponding mRNAs (Gao et al., 2016); therefore, the biogenesis, regulation, and function of these alternatively spliced circRNAs in plants are worthy of further study.

Many studies have reported that some circRNAs are conserved across species. For example, Jeck et al. (2013) identified 69 murine circRNAs in precisely orthologous locations to human cirRNAs. Similarly, recent study reported almost 600 circRNAs in which genomic location in human were overlapped with the syntenic region in mouse (Legnini et al., 2017). In addition, the exonic sequences known to circularize seem to be more conserved at the third codon position (Ebbesen et al., 2016). These results indicated that some circRNAs in mammal are stable and conserved. In plants, Ye et al. (2015) firstly explored the conservation of circRNAs in two model plants, O. sativa and A. thaliana, and identified more than 300 orthologous parent genes generating circRNAs from a similar position, implying that plant circRNAs have the conservation feature as in animals. More recently, a study on soybean circRNAs found that 551 parent gene pairs producing exonic circRNAs among O. sativa, A. thaliana, and soybean were orthologs (Zhao et al., 2017). In our study, we found 404 identified circRNAs derived from 190 parent genes in A. thaliana have sequence similarity with those in O. sativa and S. lycopersicum. Also, six identified circRNAs in A. thaliana, O. sativa, S. lycopersicum, and Z. mays showed sequence similarity. These results suggested the sequence conservation of circRNAs may exist widely across plant species. However, the orthologous parent genes producing circRNAs from orthologous splice donor and splice acceptor sites of identical exons in a broader plant species, and particularly the potential function of these conserved circRNAs need further study in the plant kingdom.



Previous research has indicated that circRNAs can regulate transcription and circRNA production may occur post-transcriptionally (Liang and Wilusz, 2014). Chao et al. (1998) first

proposed that the formation of circRNAs traps the transcripts arising from the linear gene in a nonfunctional form and prevents the existence of certain normal linear transcripts that could be translated. Thus, circRNAs can act as an "mRNA trap" by sequestering the translation start site, leaving a non-coding linear transcript and thereby reducing the expression level of the parent genes (Jeck and Sharpless, 2014). These results suggest a potential role for plant circRNAs in the regulation of their parent genes. To understand further the regulation of circRNA transcription in A. thaliana, GO and KEGG analyses were performed to annotate the biological functions of the parent genes of circRNAs. We noted that some important functions related to developmental processes, reproduction, responses to stimuli, protein synthesis, photosynthesis, and carbon metabolism were annotated in the GO and KEGG databases. These results suggest that these circRNAs generated from these parent genes might be also involved in many fundamental metabolic processes in plant growth, development, and reproduction. This is reasonable as gene expression is readily affected by developmental processes and the environment, and plant circRNAs display diverse expression patterns (Ye et al., 2015). According to previous reports (Lu et al., 2015; Ye et al., 2015; Wang et al., 2017) and the present study (Table 2), although a large number of back-splicing sites could be identified, far fewer circRNAs were obtained. In addition, the expression of circRNAs at different developmental stages or different organs displayed rapid changes and variations, and the co-expression of circRNAs at different stages and in different tissues is very limited (Supplementary Table S9). Therefore, it is difficult to determine the developmental stage- and tissue-specific circRNAs, even if biological replicates are used.

There is a regulatory mechanism called endogenous target mimicry (eTM), which involves binding to the corresponding miRNAs to block the binding of a specific transcript, leading to an increase in mRNA expression (Karakülah et al., 2016). Previous studies in plants have reported that lncRNAs could bind to miRNAs as target mimics (Franco-Zorrilla et al., 2007). Although circRNAs are not linear like lncRNAs, a small proportion of plant circRNAs was potential target mimics of miRNAs because the genomic regions that produce circRNAs contain miRNA binding sites (Ye et al., 2015; Wang et al., 2017). This mechanism is similar to miRNA sponges in animals in that circRNAs can act as decoys for miRNAs to regulate gene expression at the epigenetic level (Hansen et al., 2013; Memczak et al., 2013). In mouse, Sry circRNA harboring 16 putative binding sites acted as an inhibitor of miR138, and ciRS-7 could suppress miR-7 activity, suggesting circRNAs as miRNA sponges (Hansen et al., 2013). This sponge effect was also found in Drosophila that highly conserved miRNA binding sites could overlap with circRNA production (Westholm et al., 2014). Although in plants many miRNA binding sites within circRNAs have been predicted, no direct evidence has proved that circRNAs act as miRNA sponges (Lu et al., 2015; Ye et al., 2015; Wang et al., 2017). For example, in soybeans, 2,134 circRNAs contained predicted binding sites for 92 miRNAs (Zhao et al., 2017), and 1,861 circRNAs from PlantcircBase were predicted as putative miRNA sponges (Chu et al., 2017). In this study, we found that 39 circRNAs had putative miRNA-binding sites, and the number of miRNA-binding sites among circRNAs varies considerably, suggesting that some circRNAs are potential miRNA targets in

A. thaliana. This indicated that plant circRNAs might play roles in gene expression regulation by interacting with miRNAs. For example, miR172 plays a crucial role in regulating the transitions between developmental stages and in specifying floral organ identity, as well as in regulating legume–rhizobia nitrogen-fixing symbiosis (Zhu et al., 2009; Zhu and Helliwell, 2011; Íñiguez et al., 2015). Furthermore, a circRNA has been predicted to be a target mimic of miR172 in rice and tomato (Lu et al., 2015; Zuo et al., 2016). In our results, three circRNAs were also found to have miR172 binding sites, indicating that the circRNAs interacting with miRNAs to regulate gene expression may be conserved in plants.

Most circRNAs were considered to be non-coding RNAs. Recently, Yang et al. (2017) reported that m6A-driven translation of circRNAs is widespread in human cells, with hundreds of endogenous circRNAs having translation potential. Moreover, a circRNA (circ-ZNF609) containing an open-reading frame can be translated into a protein, providing an example of a protein-coding circRNA in murine and human myoblasts (Legnini et al., 2017). It was also reported in fly that a group of circRNAs is associated with translating ribosomes and that a circRNA generated from the muscleblind locus encodes a protein (Pamudurti et al., 2017). Thus, it appears that circRNA translation is widespread in eukaryotes. In our study, we found that most of the identified circRNAs were derived from exons (85.16%), raising the intriguing possibility of circRNA translation to produce proteins. Therefore, the issues of how circRNAs are translated and which functions circRNA-coding proteins have in plants might be promising research areas warranting further investigation.

AUTHOR CONTRIBUTIONS

GC and BJ designed the research. GC, JC, ZL, and LW performed the research; GC, YZ, and BJ analyzed the data. GC and BJ wrote the manuscript.

FUNDING

This work was funded by the National Science Foundation of China (No. 31670695) and Graduate Student Innovation Foundation of Jiangsu Province (No. KYLX16_1399).

ACKNOWLEDGMENT

We thank Biomarker Technologies and CapitalBio for Illumina ssRNA-sequencing and annotation of the circRNAs.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2017.01678/ full#supplementary-material Chen et al

- Boyes, D. C., Zayed, A. M., Ascenzi, R., McCaskill, A. J., Hoffman, N. E., Davis, K. R., et al. (2001). Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants. *Plant Cell* 13, 1499–1510. doi: 10.1105/tpc.13.7.1499
- Chao, C. W., Chan, D. C., Kuo, A., and Leder, P. (1998). The mouse formin (Fmn) gene: abundant circular RNA transcripts and gene-targeted deletion analysis. *Mol. Med.* 4, 614–628.
- Chu, Q., Zhang, X., Zhu, X., Liu, C., Mao, L., Ye, C., et al. (2017). PlantcircBase: a database for plant circular RNAs. *Mol. Plant* 10, 1126–1128. doi: 10.1016/j. molp.2017.03.003
- Conesa, A., and Götz, S. (2008). Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Genomics* 2008:619832. doi: 10.1155/2008/619832
- Danan, M., Schwartz, S., Edelheit, S., and Sorek, R. (2012). Transcriptome-wide discovery of circular RNAs in Archaea. Nucleic Acids Res. 40, 3131–3142. doi: 10.1093/nar/gkr1009
- Dang, Y., Yan, L., Hu, B., Fan, X., Ren, Y., Li, R., et al. (2016). Tracing the expression of circular RNAs in human pre-implantation embryos. *Genome Biol.* 17:130. doi: 10.1186/s13059-016-0991-3
- Darbani, B., Noeparvar, S., and Borg, S. (2016). Identification of circular RNAs from the parental genes involved in multiple aspects of cellular metabolism in barley. *Front. Plant Sci.* 7:776. doi: 10.3389/fpls.2016.00776
- Ebbesen, K. K., Kjems, J., and Hansen, T. B. (2016). Circular RNAs: identification, biogenesis and function. *Biochim. Biophys. Acta* 1859, 163–168. doi: 10.1016/j. bbagrm.2015.07.007
- Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., et al. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* 39, 1033–1037. doi: 10.1038/ ng2079
- Gao, Y., Wang, J., and Zhao, F. (2015). CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biol.* 16:4. doi: 10.1186/s13059-014-0571-3
- Gao, Y., Wang, J., Zheng, Y., Zhang, J., Chen, S., and Zhao, F. (2016). Comprehensive identification of internal structure and alternative splicing events in circular RNAs. *Nat. Commun.* 7:12060. doi: 10.1038/ncomms 12060
- Grabowski, P. J., Zaug, A. J., and Cech, T. R. (1981). The intervening sequence of the ribosomal RNA precursor is converted to a circular RNA in isolated nuclei of *Tetrahymena. Cell* 23, 467–476. doi: 10.1016/0092-8674(81)90142-2
- Granados-Riveron, J. T., and Aquino-Jarquin, G. (2016). The complexity of the translation ability of circRNAs. *Biochim. Biophys. Acta* 1859, 1245–1251. doi: 10.1016/j.bbagrm.2016.07.009
- Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., et al. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384–388. doi: 10.1038/nature11993
- Hsu, M. T., and Coca-Prados, M. (1979). Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. *Nature* 280, 339–340. doi: 10.1038/280339a0
- Íñiguez, L. P., Nova-Franco, B., and Hernández, G. (2015). Novel players in the AP2-miR172 regulatory network for common bean nodulation. *Plant Signal. Behav.* 10:e1062957. doi: 10.1080/15592324.2015.1062957
- Jeck, W. R., and Sharpless, N. E. (2014). Detecting and characterizing circular RNAs. *Nat. Biotechnol.* 5, 453–461. doi: 10.1038/nbt.2890
- Jeck, W. R., Sorrentino, J. A., Wang, K., Slevin, M. K., Burd, C. E., Liu, J., et al. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 19, 141–157. doi: 10.1261/rna.035667.112
- Jin, B., Wang, L., Wang, J., Jiang, K. Z., Wang, Y., Jiang, X. X., et al. (2011). The effect of experimental warming on leaf functional traits, leaf structure and leaf biochemistry in *Arabidopsis thaliana*. *BMC Plant Biol*. 11:35. doi: 10.1186/1471-2229-11-35
- Karakülah, G., Yücebilgili Kurtoğlu, K., and Unver, T. (2016). PeTMbase: a database of plant endogenous target mimics (eTMs). *PLOS ONE* 11:e0167698. doi: 10.1371/journal.pone.0167698
- Kos, A., Dijkema, R., Arnberg, A. C., van der Meide, P. H., and Schellekens, H. (1986). The hepatitis delta (δ) virus possesses a circular RNA. *Nature* 323, 558–560. doi: 10.1038/323558a0

- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Legnini, I., Timoteo, G. D., Rossi, F., Morlando, M., Briganti, F., Sthandier, O., et al. (2017). Circ-ZNF609 Is a circular RNA that can be translated and functions in myogenesis. *Mol. Cell* 66, 22–37. doi: 10.1016/j.molcel.2017.02.017
- Li, P., Chen, S., Chen, H., Mo, X., Li, T., Shao, Y., et al. (2015). Using circular RNA as a novel type of biomarker in the screening of gastric cancer. *Clin. Chim. Acta* 444, 132–136. doi: 10.1016/j.cca.2015.02.018
- Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X., et al. (2015). Exon-intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol. Biol.* 22, 256–264. doi: 10.1038/nsmb.2959
- Liang, D., and Wilusz, J. E. (2014). Short intronic repeat sequences facilitate circular RNA production. *Genes Dev.* 28, 2233–2247. doi: 10.1101/gad.251926.114
- Lu, T., Cui, L., Zhou, Y., Zhu, C., Fan, D., Gong, H., et al. (2015). Transcriptomewide investigation of circular RNAs in rice. *RNA* 21, 2076–2087. doi: 10.1261/ rna.052282.115
- Lukiw, W. J. (2013). Circular RNA (circRNA) in Alzheimer's disease (AD). Front. Genet. 4:307. doi: 10.3389/fgene.2013.00307
- Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495, 333–338. doi: 10.1038/nature11928
- Nigro, J. M., Cho, K. R., Fearon, E. R., Kern, S. E., Ruppert, J. M., Oliner, J. D., et al. (1991). Scrambled exons. *Cell* 64, 607–613. doi: 10.1016/0092-8674(91)90244-S
- Pamudurti, N. R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., et al. (2017). Translation of circRNAs. *Mol. Cell* 66, 9–21. doi: 10.1016/j.molcel. 2017.02.021
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/ bioinformatics/btq033
- Rybak-Wolf, A., Stottmeister, C., Glažar, P., Jens, M., Pino, N., Giusti, S., et al. (2015). Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol. Cell* 58, 870–885. doi: 10.1016/j.molcel.2015. 03.027
- Sablok, G., Zhao, H., and Sun, X. (2016). Plant circular RNAs (circRNAs): transcriptional regulation beyond miRNAs in plants. *Mol. Plant* 9, 192–194. doi: 10.1016/j.molp.2015.12.021
- Salzman, J., Chen, R. E., Olsen, M. N., Wang, P. L., and Brown, P. O. (2013). Cell-type specific features of circular RNA expression. *PLOS Genet*. 9:e1003777. doi: 10.1371/journal.pgen.1003777
- Salzman, J., Gawad, C., Wang, P. L., Lacayo, N., and Brown, P. O. (2012). Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLOS ONE* 7:e30733. doi: 10.1371/journal.pone.0030733
- Sanger, H. L., Klotz, G., Riesner, D., Gross, H. J., and Kleinschmidt, A. K. (1976). Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc. Natl. Acad. Sci. U.S.A.* 73, 3852–3856. doi: 10.1073/pnas.73.11.3852
- Shen, T., Han, M., Wei, G., and Ni, T. (2015). An intriguing RNA speciesperspectives of circularized RNA. *Protein Cell* 6, 871–880. doi: 10.1007/s13238-015-0202-0
- The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi: 10.1038/35048692
- Venø, M. T., Hansen, T. B., Venø, S. T., Clausen, B. H., Grebing, M., Finsen, B., et al. (2015). Spatio-temporal regulation of circular RNA expression during porcine embryonic brain development. *Genome Biol.* 16:245. doi: 10.1186/s13059-015-0801-3
- Vicens, Q., and Westhof, E. (2014). Biogenesis of circular RNAs. *Cell* 159, 13–14. doi: 10.1016/j.cell.2014.09.005
- Wang, P. L., Bao, Y., Yee, M. C., Barrett, S. P., Hogan, G. J., Olsen, M. N., et al. (2014). Circular RNA is expressed across the eukaryotic tree of life. *PLOS ONE* 9:e90859. doi: 10.1371/journal.pone.0090859
- Wang, Y., Yang, M., Wei, S., Qin, F., Zhao, H., and Suo, B. (2017). Identification of circular RNAs and their targets in leaves of *Triticum aestivum* L. under dehydration stress. *Front. Plant Sci.* 7:2024. doi: 10.3389/fpls.2016.02024
- Westholm, J. O., Miura, P., Olson, S., Shenker, S., Joseph, B., Sanfilippo, P., et al. (2014). Genome-wide analysis of *Drosophila* circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Rep.* 9, 1966–1980. doi: 10.1016/j.celrep.2014.10.062

- Yang, Y., Fan, X., Mao, M., Song, X., Wu, P., Zhang, Y., et al. (2017). Extensive translation of circular RNAs driven by N6-methyladenosine. *Cell Res.* 27, 626–641. doi: 10.1038/cr.2017.31
- Ye, C. Y., Chen, L., Liu, C., Zhu, Q. H., and Fan, L. (2015). Widespread noncoding circular RNAs in plants. *New Phytol.* 208, 88–95. doi: 10.1111/nph. 13585
- Zhang, X. O., Wang, H. B., Zhang, Y., Lu, X., Chen, L. L., and Yang, L. (2014). Complementary sequence-mediated exon circularization. *Cell* 159, 134–147. doi: 10.1016/j.cell.2014.09.001
- Zhang, Y., Zhang, X. O., Chen, T., Xiang, J. F., Yin, Q. F., Xing, Y. H., et al. (2013). Circular intronic long noncoding RNAs. *Mol. Cell* 51, 792–806. doi: 10.1016/j.molcel.2013.08.017
- Zhao, W., Cheng, Y., Zhang, C., You, Q., Shen, X., Guo, W., et al. (2017). Genome-wide identification and characterization of circular RNAs by high throughput sequencing in soybean. *Sci. Rep.* 7:5636. doi: 10.1038/s41598-017-05922-9
- Zhu, Q. H., and Helliwell, C. A. (2011). Regulation of flowering time and floral patterning by miR172. J. Exp. Bot. 62, 487–495. doi: 10.1093/jxb/erq295

- Zhu, Q. H., Upadhyaya, N. M., Gubler, F., and Helliwell, C. A. (2009). Overexpression of miR172 causes loss of spikelet determinacy and floral organ abnormalities in rice (*Oryza sativa*). *BMC Plant Biol.* 9:149. doi: 10.1186/1471-2229-9-149
- Zuo, J., Wang, Q., Zhu, B., Luo, Y., and Gao, L. (2016). Deciphering the roles of circRNAs on chilling injury in tomato. *Biochem. Biophys. Res. Commun.* 479, 132–138. doi: 10.1016/j.bbrc.2016.07.032

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Chen, Cui, Wang, Zhu, Lu and Jin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.