



# Deciphering extrachromosomal circular DNA in Arabidopsis

Keyi Wang<sup>a,1</sup>, Hui Tian<sup>b,1</sup>, Lequn Wang<sup>a</sup>, Lin Wang<sup>c</sup>, Yacong Tan<sup>a</sup>, Ziting Zhang<sup>a</sup>, Kai Sun<sup>a</sup>, Meng Yin<sup>a</sup>, Qinggong Wei<sup>a</sup>, Baohua Guo<sup>c</sup>, Jinyu Han<sup>a</sup>, Peiru Zhang<sup>a</sup>, Han Li<sup>a</sup>, Yulong Liu<sup>a</sup>, Hongwei Zhao<sup>c,\*</sup>, Xiaoyong Sun<sup>a,\*</sup>

<sup>a</sup>Agricultural Big-Data Research Center, College of Information Science and Engineering, Shandong Agricultural University, Taian 271018, China

<sup>b</sup>Shanghai ERA Medicine, Shanghai, China

<sup>c</sup>Department of Plant Pathology, Nanjing Agricultural University, 210095, China

## ARTICLE INFO

### Article history:

Received 26 November 2020

Received in revised form 27 January 2021

Accepted 30 January 2021

Available online 9 February 2021

### Keywords:

Extrachromosomal circular DNA

tRNA

Transposon

Arabidopsis

High-throughput sequencing

## ABSTRACT

Extrachromosomal circular DNA (eccDNA) is independent of the chromosome and exists in many eukaryotes. However, the nature and origin of eccDNA in plants remains unclear. In this study, we sequenced 12 samples from four tissues (leaf, flower, stem and root) with three biological replicates. In total, we found 743 eccDNAs found in at least two samples. Most of eccDNA have inverted repeats ranging from 4 to 12 bp in the boundaries. Interestingly, eccDNA is not only related to transposon activity, but also hosts tRNA genes, suggesting that the eccDNAs may be associated with tRNA abundance which controls protein synthesis under conditions of stress. Our results provide an unprecedented view of eccDNA, which is still naïve in scope.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Extrachromosomal circular DNA (eccDNA) is independent of the chromosome and widely exists in many eukaryotes [1–5]. eccDNA was first found in 1964 as a series of DNA circles, and later was reported as repetitive sequences which were homologous to genomic DNA. In 2012, Shibata et al. proved that eccDNA is also formed from non-repetitive sequences with direct repeats of 9–11 bp flanked on two sides [6]. eccDNA is reported to be tissue-specific in animals and may not only contribute to aging, but also participate in gene compensation and intercellular communication. In 2017, eccDNA was found, in normal tissues and cancer as the main driver with longer length [7,8]. As a result, eccDNA plays an important role in clinical practice by acting as a clear and significant biomarker for detecting cancer [9]. Recently, several studies extracted and examined eccDNA related to retrotransposons in plants [10,11]. eccDNA reflects the genome plasticity. Lanciano et al. [11] found that eccDNA is closely related to active transposon. Thieme et al. reported that transposon increases under stress in *Arabidopsis* and rice, and they suggested that the transposon mobility regulates the evolution speed to some extent. The huge diversity of transposon insertion is one of

the main drivers to the plant evolution. Thus, eccDNA may be involved in the plant evolution. However, the nature and origin of eccDNA in plants remains unclear.

With 12 samples from four tissues (leaf, flower, stem and root), we addressed this gap by systematically investigating and characterizing eccDNA in *Arabidopsis* using high-throughput sequencing. In this, we described the following properties: 1) 743 eccDNAs that were detected in *Arabidopsis* leaf, flower, stem and root tissues using high-throughput sequencing, 2) eccDNAs have inverted repeats at upstream and downstream of the boundaries, facilitating the formation of a circular structure, 3) A large number of tRNA genes and transposons were hosted by the eccDNAs, indicating the potential functions of these circular molecules. To support our findings, we used RT-PCR and southern blot to validate the related eccDNA. Our results provide an unprecedented view, which is still in naïve, of these enigmatic molecules in *Arabidopsis*.

## 2. Materials and methods

### 2.1. Sample collection

*Arabidopsis* Col-0 were grown in a controlled growth room at  $21 \pm 1$  °C in an 8 h light, 14 h dark photoperiod. At four-week-stage, one gram *Arabidopsis* leaves were clipped and stored in liquid nitrogen immediately. One gram stems from the same set of plants

\* Corresponding authors.

E-mail addresses: [johnsunx1@yahoo.com](mailto:johnsunx1@yahoo.com) (X. Sun), [hzhao@njau.edu.cn](mailto:hzhao@njau.edu.cn) (H. Zhao).

<sup>1</sup> These authors contributed equally.

were then collected and frozen in liquid nitrogen. At last, one gram roots were rinsed with water, soaked with paper towel briefly, then immediately frozen. Six-week-stage Arabidopsis with blooming flowers were held over a liquid nitrogen container with strong thrashing. Flowers dropped on the liquid nitrogen surface were collected. One gram was collected by deducing the weights between the empty tin foil and the whole package. Three biological replicates from each tissue were collected by using the same procedure.

## 2.2. Genomic sequencing

Genomic DNA was extracted from the Arabidopsis tissue samples using the Qiagen Plant DNeasy mini kit (Qiagen), and DNA concentrations were measured using a NanoDrop Microvolume Spectrophotometer (Thermo Scientific). Qualified DNA samples (concentrations more than 10 ng/μl and OD260/280 values in between 1.8 and 2.2) were combined and used for further experiments. For each sample, 3.0 ug genomic DNA was used for genomic linear fragment removal by using a GeneClean kit (MPBio) as advised by the manufacturer's instructions. The GeneClean products were digested with the ATP-dependent PlasmidSafe DNase I kit (Epicentre) to remove the remaining linear DNA fragments by incubating the DNA samples with 15 units DNase I at 37 °C for 3 days. At every 24 h, 0.5 μl 10x reaction buffer, 3 μl ATP (25 mM) and 15 units of DNase I were added to maintain continuous enzymatic digestion. Complete removal of linear DNA was confirmed with polymerase chain reaction (PCR) using a pair of primers specific amplifying ACTIN2 (5'-GCACCCTGTTCTTCTTACCG-3' and 5'-AACCTCGTA GATTGGCACA-3'). After linear fragment removal, the DNA samples were precipitated by adding 0.1 vol of 3 M NaAc (pH5.2), 2.5 volumes of ethanol and 1 μl of glycogen (Fisher) and incubating overnight at -80 °C, and eccDNA were resuspended in 20 μl TE buffer (pH 8.0). The purified eccDNA samples were amplified by random RCA using the REPLI-g Mini Kits (Qiagen) as advised by the manufacturer. The RCA incubation was performed at 30 °C for 72 h. At the end of the reaction 1 μl of each amplified DNA sample was loaded on an agarose gel (1%) to confirm the successful RCA amplification.

The amplified eccDNA samples were sequenced by GENERGY BIO (Shanghai). In brief, from each sample 0.5 ug of the amplified eccDNA was used to prepare the libraries using the TruSeq® DNA LT Sample Prep Kit v2 (Illumina), and the protocol was according to the manufacturer's user guide. Each mobilome-seq library was amplified by 10 cycles of PCR using index primers. DNA qualities and concentrations were then determined by using Qubit picogreen (PicoGreen) and via 1% agarose gel electrophoresis (120 V, 15 min). Samples were pooled and loaded onto a flow cell and 2x150 nucleotides paired-end sequencing was performed using the HiSeq3000 platform (Illumina) using Hiseq3000SBS&Cluster kit (Illumina). Up to twelve mobilome-seq libraries were pooled into one run and an average of 1 million reads per library were obtained. Illumina reads were collected for the analysis as FASTQ files.

All raw sequence data can be found at NCBI BioProject database (BioProject ID: PRJNA572559, <https://www.ncbi.nlm.nih.gov/sra/PRJNA572559>). More details are available at <http://deepbiology.cn/circDNA/>.

## 2.3. Sequencing data analysis

To analyze the sequencing reads, we used Bioconductor packages: GenomicRanges, GenomicAlignments, Biostrings and R package: SplicingTypesAnno [12]. To visualize the related results, we chose R package: ggplot2 and lattice for figures. To compare the sequencing reads to the genome annotation, we used "countOverlap", "findOverlap" and "subsetByOverlap" function from GenomicRanges and GenomicAlignments package. To extract the intron sequences from

genome annotation, we used "translateGTF" function from SplicingTypesAnno package. To extract nucleotide sequences, we used "DNAStrngSet" and "substring" from Biostrings package.

## 2.4. eccDNA identification and boundary analysis

To identify the eccDNA, we used the following two strategies: Firstly, we used "segemehl.x" function from segemehl software [13] to align all sequencing reads to the Arabidopsis genome using the following parameters: "-s -b -S -A 95". Then we used "testrealigned.x" function to detect junction reads using default parameters. The results for different samples were combined for further analysis. For second strategy we analyzed the mapped reads (bam file) from "segemehl.x" with Bioconductor packages; GenomicRanges, GenomicAlignments; and R package, SplicingTypesAnno. The junctions (boundaries of eccDNA) from a previous step were mapped to the sequencing reads, and only those junctions with at least 2 mapped reads were kept for further analysis. The junctions confirmed from previous step were combined for all samples. These junctions were final targets for eccDNA. Single/double-strand eccDNAs were determined by the following methods: the criteria for classifying a circular DNA as single- or double-strand is dependent on the sequencing reads. Since we used strand-specific sequencing methods, we can determine the sequencing reads coming from either a positive strand or negative strand. If the eccDNAs were supported by reads coming from both strands, they were labelled as double-strand; otherwise, they were labelled as single-strand.

We extracted the 10 bp upstream and downstream of the eccDNA boundaries with the following steps: (i) firstly, we output the 10 bp upstream and downstream of the genomic coordinates of these eccDNA into a bed file. (ii) We used "bedtools getfasta" function to extract all the related sequences with the following parameters: "-tab -s -name". (iii) We then modified the sequences of 10 bp downstream of eccDNA with the following modes: "no change" and "reverse the sequence". The outputs from these modifications were translated into fasta format using R.

We used BLAST software to map the fasta files of the 10 bp upstream of eccDNA from the previous step to that of the 10 bp downstream of eccDNA with the following parameters: "-word\_size 4 -strand plus -outfmt 7". All the results were further analyzed with R. The eccDNA were extracted if they met the following criteria: the 10 bp upstream and downstream of the same eccDNA have "percent\_identity" more than 90.

## 2.5. Sequence composition analysis

compseq function in EMBOSS explorer (<http://emboss.bioinfo.ucr.edu/>) was utilized to analyze the sequence composition. The 10 bp upstream and downstream sequences of eccDNA were uploaded to the website. We set "word size to consider" as 2, 3, 4, 5, and 6 while other parameters were kept as default.

## 2.6. tRNA analysis

We used genomic annotation Araport11 from TAIR to extract all tRNA information. All those sequences annotated as "tRNA" in feature attributes of GTF file (Araport11) were selected for the following analysis. The genomic coordinates of those extracted tRNA were compared with those of the eccDNA. Those eccDNA hosting tRNA were selected, and kept for further analysis.

## 2.7. Transposon comparison

We selected "transposable\_element" (TE) records in feature attributes annotated in GTF file (genomic annotation Araport11).

All the selected TE were compared with eccDNA. All eccDNA which host TE were saved and analyzed with R.

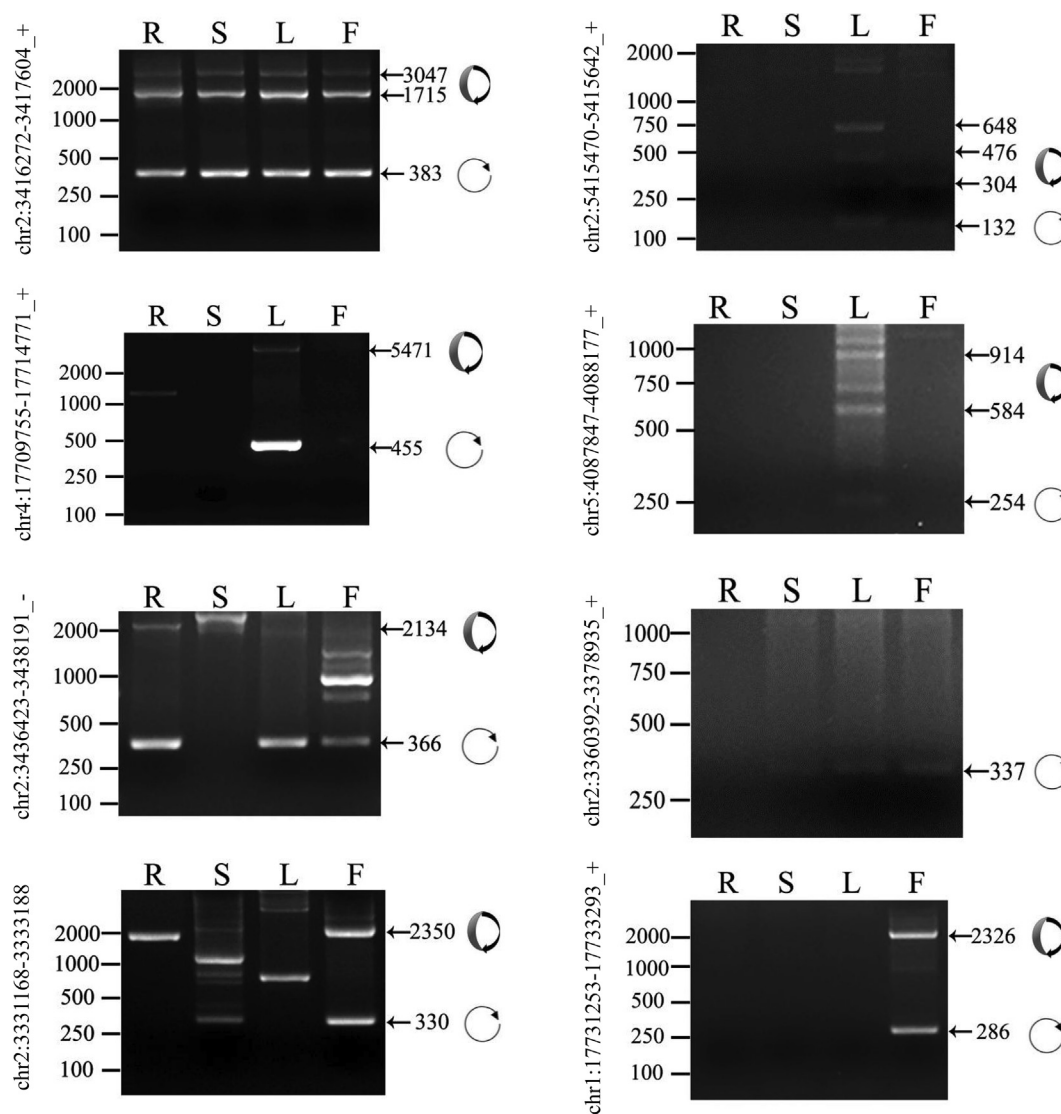
### 2.8. Comparison of our results with previous research

Supplementary Tables 5 and 6 from Lanciano et al.'work [11] were downloaded. All the potential eccDNA sequences for Arabidopsis identified in their work were compared with our results. Since all the eccDNA from Lanciano et al. do not have strand information, we used "ignore.strand = T" as the parameter in the GenomicRanges package of Bioconductor program.

All circRNAs for Arabidopsis were downloaded from AtCircDB (<http://deepbiology.cn/circRNA/>). Then we compared the genomic coordinates of circRNAs to those of eccDNA using R.

### 2.9. Pathway analysis

We used DAVID [14] to perform nuclear gene ontology analysis and agriGo [15] for organellar gene ontology analysis. First, we extracted all the TAIR ID hosting the eccDNA using R/Bioconductor GenomicRanges package. Then we uploaded all TAIR ID and chose "TAIR ID" as the identifiers in the "start analysis" section. We kept all the other parameters as default, and DAVID utilized Fisher's



**Fig. 1.** Validation of eight extrachromosomal circular DNAs from Arabidopsis root (R), stem (S), leaf (L) and flower (F) samples by divergent PCR analyses. Primers were designed go away from each other so that only circular molecules are amplified. Eight predicted eccDNA structures were selected and tested with the method. Outward-directed PCR primers were designed for each predicted eccDNA structure, and can specifically amplify the junction sites. Results are visualized on 2% agarose gel electrophoresis. The ladders were listed on the left side of each electrophoresis gel picture. The original location size of the amplified fragments eccDNAs were indicated on the left right side of the gel pictures. The ladders were listed on the left side of each electrophoresis gel picture. The expected product sizes were listed on the right side as indicated by arrows and numbers, and the circles indicated amplification products in different forms- single fragment or rolling-over fragments. 8 eccDNAs were validated by PCR analysis, and 7 of these eccDNAs were further validated by SANGER sequencing except one eccDNA (chr2:5415470-5415642\_+). A. chr2:3416272-3417604\_+; B. chr2:5415470-5415642\_+; C. chr4:17709755-17714771\_+; D. chr5:4087847-4088177\_+; E. chr2:3436423-3438191\_-; F. chr2:3360392-3378935\_+; G. chr2:3331168-3333188\_+; H. chr1:17731253-17733293\_+.

Exact test for statistical analysis. The gene functional groups with FDR adjusted at a significance level of  $p < 0.05$  were selected for further exploration using R.

### 2.10. Experimental validation

PCR validation: 8 eccDNAs were randomly selected for PCR validation. All primers were designed using the online primer design tool from the NCBI website ([www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)). The DNA samples (before and after the RCA amplification) were diluted 10 times and used as PCR templates. Seven  $\mu\text{L}$  of PCR reaction products were loaded on the 2% agarose gel for electrophoresis (100 V, 30 min). To further confirm the eccDNA structures, PCR products were purified using a gel purification kit (Axygen) and analyzed by SANGER sequencing from both sides.

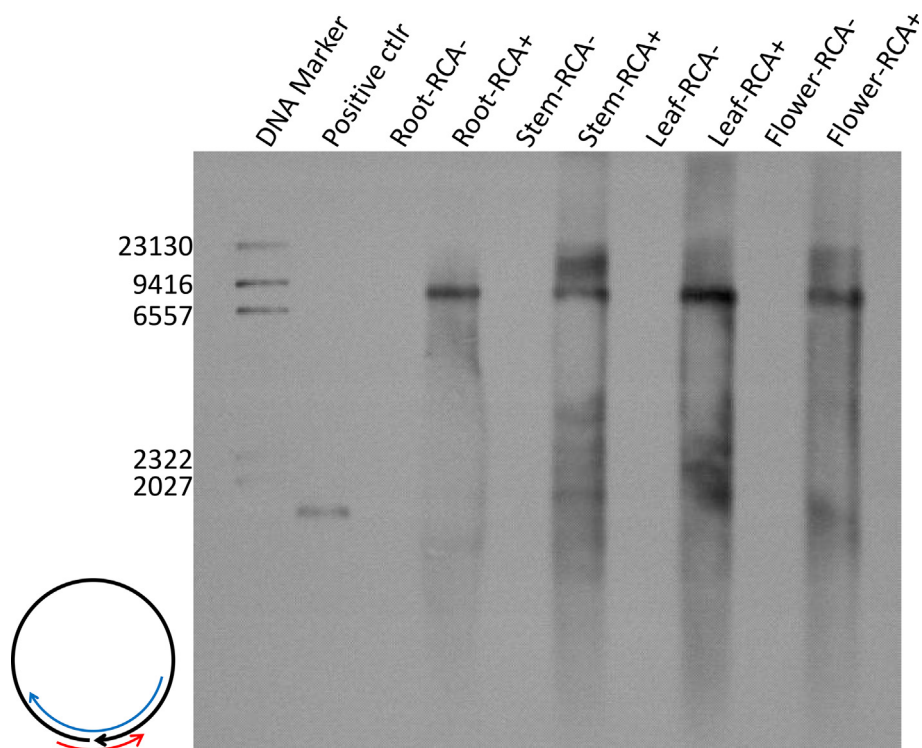
Southern blot analysis: Total genomic DNA was extracted from Arabidopsis Col-0 wild type tissues (leaf, stem, leaf and root) using the CTAB method. Linear genomic DNA was digested with the ATP-dependent PlasmidSafe DNase I (Epicentre) and the remaining eccDNA molecules were then amplified by rolling circle amplification (RCA) using the REPLI-g Mini Kit (Qiagen). Southern blot analysis was performed on the putative eccDNA structure chr2:3360392-3378935. DNA samples before and after RCA were used for hybridization. A positive control was produced by PCR amplification, using the flower RCA + DNA sample as a template and the primer pair (F: GCACCAACGAGAGATTGAAGTACG; R: AAAGTACAGGCAGAGCCAAACC). As a result a 1.6 kb product corresponding to the junction region of this circular structure was obtained (Supplemental Data 1). DNA samples were digested with the enzyme FspI and loaded onto 0.8% agarose gel and transferred onto Hybond-N + nylon membrane (GE healthcare). A 700 bp DNA strand corresponding to the junction site of this eccDNA structure was synthesized, labeled with Digoxin, and used as a hybridization

probe (Supplemental Data 5). Stringency washes were performed at 65 °C in 0.5x SSC. The hybridization signal was detected with the DIG system (Roche) following the manufacturer's instructions.

## 3. Results and discussion

### 3.1. The landscape of eccDNA

We collected 12 samples from four tissues (leaf, flower, stem and root) with three biological replicates and followed the enriched eccDNA protocol [6] to isolate and further sequence the eccDNA. We then performed the analysis described in the Materials and Methods section to identify all the related eccDNAs. Two methods were utilized to detect the eccDNA: the first one used segemehl software [13] to detect the junction reads of eccDNA and the second one used the Bioconductor GenomicRanges package to confirm the reads overlapping with junction reads. In total, we found 289,529 eccDNAs in 12 samples from all chromosomes. These eccDNAs were highly dynamic and to evaluate these molecules accurately, we only selected 743 eccDNAs found in at least two samples for the following analysis (Supplemental Data 1). To check the results, we randomly selected 8 eccDNAs for experimental validation, and seven of these eccDNAs were confirmed by both PCR analysis and SANGER sequencing, except one eccDNA (chr2:5415470-5415642\_+, i.e., chromosome 2, start: 5415470, end: 5415642, strand: +) was only validated by PCR analysis, suggesting a high accuracy of the detection algorithm (Fig. 1). The divergent PCR not only amplified DNA fragments with the expected size (e.g. 383 bp for chr2:3416272-3417604\_+), but also amplified DNA fragments with one or more rolling-over (e.g.  $383 + 1332 = 1715$ ;  $383 + 1332 * 2 = 3047$ ), demonstrating the circular nature of the PCR templates. Our results clearly showed that some eccDNAs (chr2:3416272-3417604\_+) had a ubiquitous pres-



**Fig. 2.** Validation of extrachromosomal circular DNAs (chr2:3360392-3378935\_+) using Southern blot. 10  $\mu\text{g}$  DNA extracted from Arabidopsis root, stem, leaf, and flower tissues were digested with FspI and separated by a 0.7% agarose gel. After being transferred to a nylon membrane, samples were detected by using a 700 bp DNA fragment that is reverse-complementary to the junction of this eccDNA and its flanking regions. A 1690 bp PCR fragment encompassing the junction region was used as a positive control. Black: eccDNA; Blue: positive control; Red: probe. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

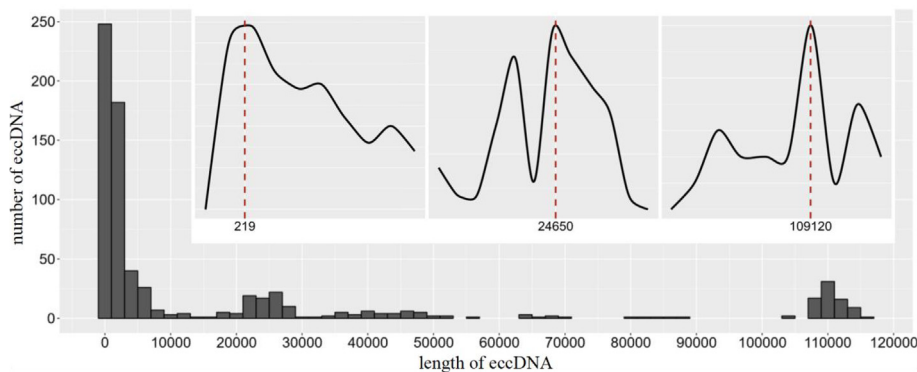


Fig. 3. The length distribution of eccDNAs. There are three peaks at 219 bp, 24,650 bp, and 109,120 bp, respectively.

ence in each plant tissue, while other eccDNA are only present in specific tissues (chr4:17705755-17714711-+ was found only in leaf). We further validated one eccDNA using southern blot assay. By using a probe reverse-complementary to the junction region, we detected one eccDNA (chr2:3360392-3378935\_+) in all the four samples, indicating a ubiquitous presence of this eccDNA in Arabidopsis (Fig. 2). Interestingly, we found that the length of eccDNA varied widely, ranging from 102 bp to 1,931 kbp, and showed three peaks: the first at 219 bp, the second at 24,650 bp and the third at 109,120 bp (Fig. 3). In addition, the eccDNAs included both single-strand and double-strand types. For example, 271 single-strand eccDNAs and 472 double-strand eccDNAs were detected in 12 samples. Surprisingly, chromosome 2 (chr2) hosted 207 eccDNAs, which is more than any other chromosome (Fig. 4). The high number of eccDNAs may result from the large proportion of sequence similarity between chromosome 2 and mitochondrial [16]. Richmond et al. (2000) [17] reported that chromosome 2 has “unusual feature”: some sequences are incorporated into chromosome 2 directly from mitochondrial genome. Also, it is worth mentioning that 152 and 184 eccDNAs were derived from chloroplasts and mitochondria, respectively.

Next, we explored the source of eccDNA using genome annotation, Araport11 from TAIR (<https://www.arabidopsis.org>). 62% of eccDNAs (457) overlapped with genes and 38% of eccDNAs (286) came from intergenic regions. In addition, the starting/ending boundaries of 6/12, 17/18, 39/72, 85/116 and 552/493 eccDNA originated from 5' UTR, 3' UTR, intron region, CDS region and intergenic regions, respectively (Fig. 5). These eccDNAs overlapped with 577 genes and a large number of eccDNAs were enriched together in certain genomic regions, i.e., super eccDNA regions. For example, chr2:3239460-3364069\_+ region was one of these super eccDNA regions, hosting 108 eccDNAs and 35 genes (Supplemental Fig. 5). Strikingly, 77% of the overlapped regions between the genes and the eccDNA contained the first exon.

### 3.2. Upstream and downstream of the eccDNA boundaries

To understand how eccDNA is formed, we utilized BLAST software to analyze 10 bp upstream and downstream of the eccDNA boundaries. Interestingly, we found 421 (57%) of the eccDNAs had inverted repeats ranging from 4 to 12 bp both upstream and downstream of the boundaries, suggesting that inverted repeats may mediate in the process of circular formation. For example, the eccDNA (chr4:11037854-11038022\_+) had a length of 168 bp, and 7 bp upstream and downstream of its boundaries were inverted repeats (Fig. 6). Also, we calculated the sequence composition using CompSeq software, hosted at EMBOSS explorer (<http://emboss.bioinformatics.nl/cgi-bin/emboss/compseq>). AA or TT

occurred more than any other combination at both 10 bp upstream and downstream of the eccDNA boundaries (Fig. 7). This result is consistent with previous human data [6].

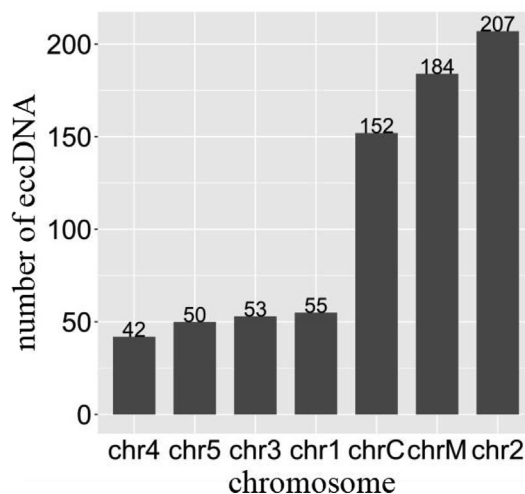


Fig. 4. The number of eccDNAs in different chromosomes.

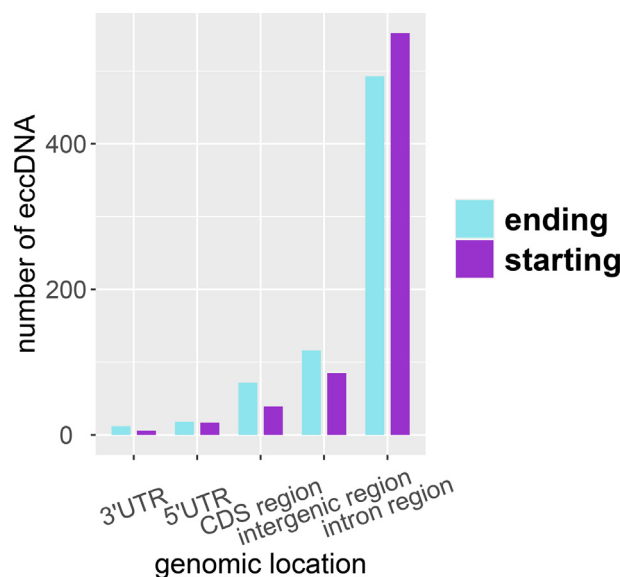
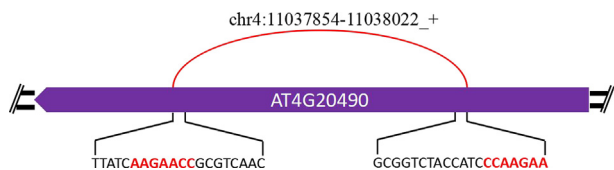


Fig. 5. The genomic locations of starting and ending boundaries of eccDNAs.



**Fig. 6.** Inverted repeats from the 10 bp upstream and downstream of an eccDNA (chr2:3398614-3399073\_+).

### 3.3. Biological function of eccDNA

To determine the functions of these eccDNAs, we utilized R/Bioconductor GenomicRanges package and SplicingTypesAnno package to compare the genomic coordinates of eccDNA with those of lncRNA, tRNA, snoRNA, microRNA and others. Surprisingly, we found 27% (199) of eccDNA host genes encoding tRNA (Supplemental Data 2), mainly in chr2, chrM and chrC. According to genome annotation (Araport11), the length of tRNA genes ranged from 69 to 2,631 bp and some of them were clustered closely together in the chromosome. The eccDNAs hosted a different number of tRNA genes, ranging from 1 to 11. Most of the eccDNAs (73) contained four tRNA genes, and three eccDNAs even hosted 11 tRNA genes in a string. For example, chr2:3360392-3378935\_+ hosted one tRNA gene, AT2G07793 (Fig. 8). This gene is related to pre-tRNA-tRNA-Ser (anticodon: TGA) and the genes encoding all 20 amino

acids are fully detected in the eccDNAs (Fig. 9), suggesting that the eccDNAs may be associated with tRNA abundance, which is related to protein synthesis under conditions of stress[18].

Lanciano et al. reported that eccDNA is related to transposon activity [11]. Our findings prove that high proportion of eccDNAs is closely related to the genomic location of transposon. Among 743 eccDNAs, 168 (23%) eccDNAs overlapped with transposable elements (TE) (Supplemental Data 3). Notably, among the eccDNAs hosting TE, 42% eccDNAs (71) hosted 21 TEs and 31% of eccDNAs (52) hosted only one TE. Totally, we found 168 eccDNAs host 449 transposable elements from 98 transposon families. We further compared the results from this research to the work of Lanciano et al. Around 16% of reported eccDNAs (10) from 40 families overlapped with our results (Supplemental Fig. 6). Furthermore, we compared these reported eccDNAs with the unfiltered eccDNAs (289,529) from 12 samples, and 89% (56) of the reported eccDNAs were found to overlap with the unfiltered eccDNAs. Finally, we mapped the eccDNAs to the circular RNAs (circRNA) downloaded from AtCircDB [19]. Interestingly, the genomic coordinate of two eccDNAs (chr2: 9128960-9129170\_-, chr5:4002358-4014861\_+) exactly matched the genomic location of two circRNAs (ath-circ26178-AT2G21330, ath-circ70870-AT5G12400). Iparraguirre et al. [20] proposed that “Transcription events across the junction of circular DNAs would result in a transcript with a junction similar to those present in circRNAs. There in this report, we want to draw attention to transcripts from such circular DNAs both as an inter-

Starting boundaries					Ending boundaries				
W	O	Obs	Exp	Obs/	W	O	Obs	Exp	Obs/
o	bs	Fre	Fre	Exp	o	bs	Fre	Fre	Exp
r	Co	que	que	Freq	r	Co	que	que	Freq
d	unt	ncy	ncy	uenc	d	unt	ncy	ncy	uenc
				y					y
AA	54330	0.1	0.06	1.53	TT	53341	0.09	0.06	1.5
TT	52982	0.09	0.06	1.49	AA	53044	0.09	0.06	1.49
AT	43623	0.08	0.06	1.23	AT	43363	0.08	0.06	1.22
GA	40012	0.07	0.06	1.13	GA	39969	0.07	0.06	1.13
TC	38676	0.07	0.06	1.09	TC	38496	0.07	0.06	1.08
AG	36136	0.06	0.06	1.02	AG	36417	0.06	0.06	1.03
TG	35996	0.06	0.06	1.01	TG	36083	0.06	0.06	1.02
CT	35417	0.06	0.06	1	CT	35909	0.06	0.06	1.01
TA	34939	0.06	0.06	0.98	TA	35182	0.06	0.06	0.99
CA	33751	0.06	0.06	0.95	CA	33548	0.06	0.06	0.94
GT	30491	0.05	0.06	0.86	GT	30563	0.05	0.06	0.86
AC	28983	0.05	0.06	0.82	GG	29391	0.05	0.06	0.83
GG	28601	0.05	0.06	0.81	AC	28907	0.05	0.06	0.81
CC	27304	0.05	0.06	0.77	CC	27154	0.05	0.06	0.76
GC	24267	0.04	0.06	0.68	GC	24486	0.04	0.06	0.69
CG	22720	0.04	0.06	0.64	CG	22474	0.04	0.06	0.63
Other	131	0.00	0.00	inf	Other	32	0.00	0.00	inf

**Fig. 7.** Sequence composition in the starting and ending boundaries of eccDNAs (chi-square test, *p* value < 2.2e–16).



Fig. 8. Three eccDNAs hosting genes encoding tRNA.

tRNA_ID	Tyr	Gln	Trp	Leu	Asn	Arg	Ile	Ala	Val	Ser	Asp
eccDNA_no	101	79	78	69	68	66	51	47	35	32	17
tRNA_ID	Glu	Lys	Met-i	Cys	Gly	Met-e	Phe	Pro	Thr	His	
eccDNA_no	16	8	6	5	5	4	3	3	3	1	

Fig. 9. The number of eccDNAs hosting genes encoding tRNA. Note: the symbol of amino acids represents the related tRNA genes. For example, “Tyr” represents the gene encoding tRNA<sup>Tyr</sup>.

esting new player in the transcriptome and also as a confounding factor that must be taken into account when studying circRNAs”. Our finding supports this hypothesis: some circular RNA may come directly from the eccDNAs instead from the back splicing mechanism.

### 3.4. Tissue-specific eccDNA

To investigate eccDNA variation in different tissues, we chose four tissues (leaf, flower, stem and root) for extraction of eccDNA. In total, we found 212 eccDNAs were only detected in one tissue, suggesting the existence of tissue-specific eccDNA (Supplemental Data 4). For example, 97, 69, 36 and 10 unique eccDNAs were found only in the stem, leaf, flower or root, respectively (Fig. 10). Next, we utilized DAVID (<https://david.ncifcrf.gov/>) to perform gene ontology analysis of three tissues (stem, leaf and flower).

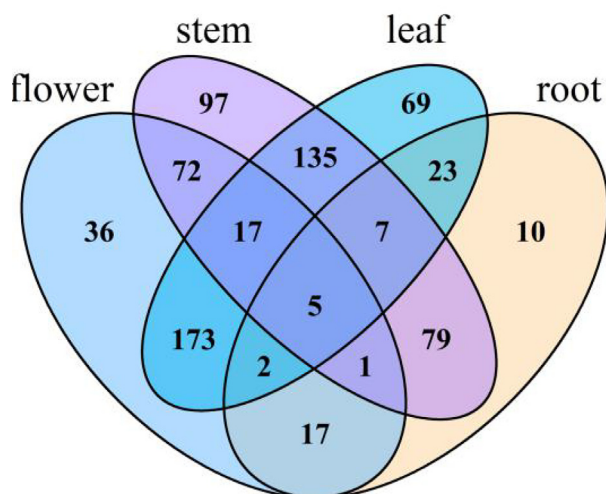


Fig. 10. Venn diagram for the number of eccDNAs in different tissues.

Since there are few genes in root, which is not appropriate for enrichment analysis, we did not include root tissue to avoid ambiguity. Importantly, the statistically significant pathways for all tissues (stem, leaf, and flower) were highly consistent. Seven pathways were enriched for nuclear genes, including mitochondrion, oxidative phosphorylation, triplet codon-amino acid adaptor activity, translation, translational elongation and mitochondrial membrane (Supplemental Fig. 7).

### 4. Conclusion

In conclusion, we report 743 eccDNAs that were detected in Arabidopsis leaf, flower, stem and root tissues using high-throughput sequencing. We found both double-strand and single-strand eccDNAs with a distribution of lengths which peaked at 219 bp, 24,650 bp and 109,120 bp. Some eccDNAs came from the same genomic region, forming a super eccDNA region. In addition, these eccDNAs had inverted repeats upstream and downstream of the boundaries, facilitating the formation of a circular structure. A large number of tRNAs and transposons were hosted by the eccDNAs, indicating the potential functions of these circular molecules. These results provide an unprecedented view of eccDNA.

Interestingly, we found large quantities of eccDNA (289,529 eccDNAs) in 12 samples, and only 743 eccDNAs in at least two samples simultaneously, reflecting their highly dynamic properties. The active elements hosted in the eccDNAs, including transposons and tRNA genes, may help the eccDNA play a role in evolution, aging, gene expression, immune response, etc. Many studies have been done on eccDNA amplifications and revealed that the number of eccDNAs changed under stress, but the functions of these mystery molecules need to be further investigated.

### CRedit authorship contribution statement

**Keyi Wang:** Data curation, Formal analysis, Resources, Software, Visualization. **Hui Tian:** Methodology, Investigation, Validation, Writing - original draft. **Lequn Wang:** Data curation,

Resources, Visualization. **Lin Wang:** Methodology, Investigation. **Yacong Tan:** Data Curation, Visualization. **Ziting Zhang:** Visualization, Writing - review & editing. **Kai Sun:** Visualization. **Meng Yin:** Validation, Data Curation, Visualization. **Qinggong Wei:** Validation, Data Curation, Visualization. **Baohua Guo:** Methodology, Investigation. **Jinyu Han:** Validation, Data Curation, Visualization. **Peiru Zhang:** Validation, Data Curation, Visualization. **Han Li:** Visualization, Writing - review & editing. **Yulong Liu:** Validation, Data Curation, Visualization. **Hongwei Zhao:** Conceptualization, Methodology, Investigation, Writing - review & editing, Supervision. **Xiaoyong Sun:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Visualization, Writing - original draft, Writing - review & editing.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

This work was supported by the National Natural Science Foundation of China (grant number 32070684,31571306 to X.S.) and Key R&D Program of Shandong Province (2020LZGC009). We thank National Supercomputer Center in Jinan and Supercomputing Center in Shandong Agricultural University for technical support. No conflict of interest declared.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2021.01.043>.

### References:

- [1] Cohen S, Segal D. Extrachromosomal circular DNA in eukaryotes: possible involvement in the plasticity of tandem repeats. *Cytogenet Genome Res* 2009;124:327–38.
- [2] Dillon LW, Kumar P, Shibata Y, Yuh-Hwa W, Willcox S, et al. Production of extrachromosomal MicroDNAs is linked to mismatch repair pathways and transcriptional activity. *Cell Rep* 2015;11:1749–59.
- [3] Møller, HD., Larsen, CE., Parsons, L., Hansen AJ., Regenberg B., and Mourier T. 2015a, Formation of Extrachromosomal Circular DNA from Long Terminal Repeats of Retrotransposons in *Saccharomyces cerevisiae*, G3 (Bethesda) , 6, 453–462.
- [4] Pennisi E. Circular DNA throws biologists for a loop. *Science* 2017;356:996.
- [5] Paulsen T, Kumar P, Koseoglu MM, Dutta A. Discoveries of extrachromosomal circles of DNA in normal and tumor cells. *Trends Genet* 2018;S0168–9525:30230–5.
- [6] Shibata Y, Kumar P, Layer R, et al. Extrachromosomal microDNAs and chromosomal microdeletions in normal tissues. *Science* 2012;336:82–6.
- [7] Turner KM, Deshpande V, Beyter D, et al. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* 2017;543:122–5.
- [8] Kumar P, Dillon LW, Shibata Y, et al. Normal and cancerous tissues release extrachromosomal circular DNA (eccDNA) into the circulation. *Mol Cancer Res* 2017;15:1197–205.
- [9] Wu S, Turner KM, Nguyen N, et al. Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature* 2019;575:699–703.
- [10] Thieme M, Lanciano S, Balzergue S, et al. Inhibition of RNA polymerase II allows controlled mobilisation of retrotransposons for plant breeding. *Genome Biol* 2017;18:134.
- [11] Lanciano S, Carpentier MC, Llauro C, et al. Sequencing the extrachromosomal circular mobilome reveals retrotransposon activity in plants. *PLoS Genet* 2017;13:e1006630.
- [12] Sun X, Zuo F, Ru Y, et al. SplicingTypesAnno: annotating and quantifying alternative splicing events for RNA-Seq data. *Comput Methods Programs Biomed* 2015;119:53–62.
- [13] Hoffmann S, Otto C, Doose G, Tanzer A, Langenberger D, Christ S, Kunz M, Holdt LM, Teupser D, Hackermüller J, Stadler PF. A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. *Genome Biol* 2014;15(2):R34. <https://doi.org/10.1186/gb-2014-15-2-r34>.
- [14] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4:44–57.
- [15] Tian T, Liu Y, Yan H, et al. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res* 2017;45:W122–9.
- [16] Lin X, Kaul S, Rounsley S, et al. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 1999;402:761–8.
- [17] Richmond T. *Arabidopsis* chromosome 2 sequence. *Genome Biol* 2000;1:reports029.
- [18] Torrent M, Chalancon G, de Groot NS, Wuster A, Madan BM. Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions. *Sci Signal* 2018;11:eaat6409.
- [19] Ye J, Wang L, Li S, et al. AtCircDB: a tissue-specific database for *Arabidopsis* circular RNAs. *Brief Bioinform* 2019;20:58–65.
- [20] Iparraquirre L, Prada-Luengo I, Regenberg B, et al. To be or not to be: circular RNAs or mRNAs from circular DNAs?. *Front Genet* 2019;10:940.