New Phytologist

Research

Circular RNAs mediated by transposons are associated with transcriptomic and phenotypic variation in maize

Lu Chen[1]*, Pei Zhang[1]*, Yuan Fan[1]*, Qiong Lu[1], Qing Li[1], Jianbing Yan[1], Gary J. Muehlbauer[2,3], Patrick S. Schnable[4], Mingqiu Dai[1] and Lin Li[1]

[1]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China; [2]Department of Agronomy and Plant Genetics, University of Minnesota, St Paul, MN 55108, USA; [3]Department of Plant and Microbial Biology, University of Minnesota, St Paul, MN 55108, USA; [4]Department of Agronomy, Iowa State University, Ames, IA 50011, USA

Author for correspondence:
Lin Li
Tel: +86 027 87280110
Email: hzaulilin@mail.hzau.edu.cn

Received: *21 July 2017*
Accepted: *18 October 2017*

New Phytologist (2017)
doi: 10.1111/nph.14901

Key words: circular RNAs (circRNAs), LINE1, maize, phenotypic variation, transposons.

Summary

- Circular RNAs (circRNAs) are covalently closed RNA molecules. Recent studies have shown that circRNAs can arise from the transcripts of transposons. Given the prevalence of transposons in the maize genome and dramatic genomic variation driven by transposons, we hypothesize that transposons in maize may be involved in the formation of circRNAs and further modulate phenotypic variation.
- We performed circRNA-Seq on B73 seedling leaves and uncovered 2804 high-confidence maize circRNAs, which show distinct genomic features.
- Comprehensive analyses demonstrated that sequences related to LINE1-like elements (LLEs) and their Reverse Complementary Pairs (LLERCPs) are significantly enriched in the flanking regions of circRNAs. Interestingly, as the number of LLERCPs increase, the accumulation of circRNAs varies, whereas that of linear transcripts decreases. Furthermore, genes with LLERCP-mediated circRNAs are enriched among loci that are associated with phenotypic variation. These results suggest that circRNAs are likely to be involved in the modulation of phenotypic variation by LLERCPs.
- Further, we showed that the presence/absence variation of LLERCPs was associated with expression variation of circRNA-circ1690 and was related to ear height, potentially through the interplay between circRNAs and functional linear transcripts. Our first study of maize circRNAs uncovers a potential new way for transposons to modulate transcriptomic and phenotypic variations.

Introduction

RNAs have been defined primarily by the central dogma as messenger molecules in the process of gene expression. However, ample evidence suggests that RNAs function not only to encode protein, but also as noncoding regulatory molecules (Rinn & Chang, 2012; Morris & Mattrick, 2014). Noncoding RNAs include microRNAs, small interfering RNAs, natural *cis*-acting siRNAs and other long noncoding RNAs. One rising star of these noncoding RNA species is circular RNAs (circRNAs), which are covalently closed, single-stranded RNA molecules with a 3′, 5′ – phosphodiester or a 2′, 5′ – phosphodiester bond at the junction site (Chen & Yang, 2015; Chen *et al.*, 2015). Approximately 30 yr ago a handful of circRNAs were identified and thought to be transcriptional noise due to aberrant splicing and lack of an identified functional role (Hsu & Coca-Prados, 1979; Nigro *et al.*, 1991; Capel *et al.*, 1993; Pasman *et al.*, 1996; Zaphiropoulos, 1996). Recently, however, circRNAs have been reported to influence the expression of parental linear transcripts and other genes (Chen *et al.*, 2015).

With the advent of genomic sequencing techniques and high-efficiency bioinformatic analysis, genome-wide scans for circRNAs in different cell types or tissues have been conducted in a wide range of species (Chen *et al.*, 2015). Although most circRNAs are expressed at low concentrations, thousands of circRNAs have been identified in mice, *Caenorhabditis elegans*, Drosophila and humans (Salzman *et al.*, 2012; Memczak *et al.*, 2013; Ashwal-Fluss *et al.*, 2014; Guo *et al.*, 2014; Westholm *et al.*, 2014; Bachmayr-Heyda *et al.*, 2015; Gao *et al.*, 2015) and dozens of circRNAs have been shown to be highly expressed in a cell type- or tissue-specific manner, which is suggestive of functionality. CircRNAs have been identified across the eukaryotic tree of life, which suggests that they are an evolutionarily conserved feature of the eukaryotic gene expression (Jeck *et al.*, 2013; Wang *et al.*, 2014). More evidence has accumulated to indicate that circRNAs have the potential to shape gene expression as micro (mi)RNA sponges, regulating transcription and interfering with splicing (Hansen *et al.*, 2013; Memczak *et al.*, 2013;

*These authors contributed equally to this work.

Kulcheski *et al.*, 2016). More recent mammalian studies have shown that transposons are enriched in the flanking regions of circRNAs and have the potential to mediate the formation of circRNAs via reverse complementary pairing (Jeck *et al.*, 2013; Zhang *et al.*, 2014). Further studies have shown that circRNAs are also derived from pre-mRNAs transcribed by RNA polymerase II, and are produced through back-splicing circularization, which is controlled by *cis*-regulatory elements and *trans*-acting factors (Ashwal-Fluss *et al.*, 2014; Liang & Wilusz, 2014; Zhang *et al.*, 2014; Starke *et al.*, 2015).

In plants, Wang *et al.* (2014) performed the first genome-wide scan of circRNAs in Arabidopsis and identified three circRNAs. In rice, 2354 circRNAs were identified via deep sequencing and computational analysis of strand-specific RNA-Seq data (Lu *et al.*, 2015). Meanwhile, Ye *et al.* (2015) performed a comparative transcriptomic analysis of circRNAs, identifying 12 037 and 6012 circRNAs in rice and Arabidopsis, respectively, indicating widespread existence of circRNAs in plants. Plant circRNAs have been shown to be conserved, expressed at low concentrations and in a tissue-specific manner, and have been found to be expressed differentially under phosphate-sufficient and starvation conditions (Ye *et al.*, 2015). A robust software pipeline PcircRNA_finder, which integrates the circRNA predication results from multiple popular *de novo* circRNA tools and optimizes the results especially for plants, was devised and identified a substantial number of circRNAs in crops (Chen *et al.*, 2016; Chu *et al.*, 2017). In contrast to what has been observed in mammals, repetitive elements and reverse complementary sequences are not significantly enriched in the flanking regions of circRNAs in rice and Arabidopsis (Lu *et al.*, 2015; Ye *et al.*, 2015).

Maize is one of the most widely grown crops in the world. Although the first version of a maize reference genome was released in 2009, the genome-wide annotation of functional elements is ongoing (Schnable *et al.*, 2009). For example, Li *et al.* (2014) identified and characterized over 20 000 long noncoding RNAs. Increased complexity of the maize transcriptome has been revealed through third generation sequencing (B. Wang *et al.*, 2016). Until recently, transcriptomic studies were focused on genes with linear transcripts (both protein coding and noncoding), ignoring the widely existing noncoding circRNAs in maize. Meanwhile, > 85% of the maize genome is repetitive transposable elements (Schnable *et al.*, 2009; Jiao *et al.*, 2017). Transposons have been reported to be involved in the regulation of functional gene expression and phenotypic variation by either insertion in coding, regulatory and intergenic regions, by alteration of epigenomic variation, or large-scale genomic transposition (Lisch, 2009, 2013; Wei & Cao, 2016). A recent study showed that human reverse complementary pairs of transposons can mediate the formation of circRNAs (Liang & Wilusz, 2014). Given the prevalence of transposons in the maize genome (Schnable *et al.*, 2009; Jiao *et al.*, 2017) and the pilot discovery of circular DNA elements mediated by Mu transpositions in maize (Sundaresan & Freeling, 1987), we would expect ample circRNAs in maize.

Additionally, maize pan-genomes exhibit enormous genomic structure variations, such as copy number variations (CNVs) and presence/absence variations (PAVs), dispensable and nondispensable regions, as well as tens of millions of single nucleotide polymorphisms (SNPs) and small insertion and deletions (InDels) (Gore *et al.*, 2009; Springer *et al.*, 2009; Beló *et al.*, 2010; Lai *et al.*, 2010; Swanson-Wagner *et al.*, 2010; Chia *et al.*, 2012; Hansey *et al.*, 2012; Hirsch *et al.*, 2014, 2016). As one of the major contributors to the dynamic maize pan-genome, transposons vary dramatically among haplotypes (Fu & Dooner, 2002; Bennetzen, 2005; Dooner & He, 2008), which may lead to the extensive diversity of circRNAs among haploypes. Together with the evidence that circRNAs can function as miRNA sponges, and transcriptional regulators of their parental functional linear transcript isoforms (Chen *et al.*, 2015), we predict that transposons may also function to modulate phenotypic variation via the formation of circRNAs.

In order to test our hypothesis, we collected a comprehensive transcriptome dataset including circRNA-Seq data (Jeck *et al.*, 2013) and hundreds of public available mRNA-Seq data, and identified 2804 high-confidence circRNAs in maize, which showed distinct genomic features around the junction loci. Interestingly, unlike rice and Arabidopsis, transposons, especially retrotransposon LINE1-like elements (LLEs) and their reverse complementary pairs (LLERCPs) are likely to affect the formation of circRNAs. The dramatic genomic variation of LLEs among diverse inbreds is significantly associated with the expression–concentration variation of circRNAs and parental linear RNAs, which is further related to agronomic trait variation. Our study provided a first study of maize circRNAs, and uncovered a potential novel functional role of transposons in the regulation of transcriptomic and phenotypic variation in maize probably via the formation of circRNAs.

Materials and Methods

CircRNA-Seq on B73 seedlings

The third leaves of B73 V3 stage seedlings were sampled for the isolation of total RNA and genomic DNA. The total RNA was subjected to RNase R treatment followed by circular RNA (circRNA)-Seq (Jeck *et al.*, 2013). Detailed information about circRNA-Seq can be found in the Supporting Information Methods S1.

Public datasets used for the identification of circular RNAs

Public maize RNA-Seq datasets were obtained from Sequence Read Archive (SRA) database. These include 368 lines of an association panel (kernels; SRP026161), 503 diverse inbred lines (seedings; SRP018753) and various tissues of B73 (leaf, seed, embryo, seeding, endosperm, embryo sac, ovule, pollen, root and shoot apical meristem) (Table S1).

Bioinformatic pipeline for the identification of circRNAs and routine RNA-Seq analysis

For each RNA-seq sample, the pseudo-reference-based strategy, which reshuffles the genome to construct a pseudo scramble

reference genome for the mapping of scramble RNA-Seq reads, was implemented by KNIFE (Szabo *et al.*, 2015) and used for the genome-wide identification of circRNAs. For the datasets collected by circRNA-Seq, multiple types of bioinformatic software were employed for the identification of circRNAs in maize. Detailed description of the bioinformatic analyses could be found in the Methods S1.

Validation of circRNAs

In order to validate the maize circRNAs that we identified, seeds of B73 were germinated under the same conditions as samples that had been subjected to circRNA-Seq and mRNA-Seq. The third leaves were harvested at the V3 stage. Total RNA was isolated from the pooled sample using a similar protocol. The PCR primers were designed for the validation of 30 randomly selected circRNAs from the unique candidates identified by CIRI, CIRCEXPLORER2 and CIRC_FINDER (10 per tool). A further 15 randomly selected circRNAs identified by CIRCEXPLORER2 were subject to wet experiment validation. cDNA (RNase R+), genomic DNA (gDNA) and total RNA were used as templates for the PCR validation of circRNAs; this is because circRNAs are covalently closed RNA molecules and such a circular structure can allow several rounds of amplification, which might generate cDNAs with different lengths. The reagent 2× Taq Master Mix (Vazyme, Nanjing, CN) was used for cDNA and gDNA amplification with touchdown PCR to detect the circRNA templates. Touchdown PCR was carried out using Thermocycler T100 with our customized program (95°C × 3′ for one cycle; 95°C × 30″, 60°C × 30″, 72°C × 30″, for 9 cycles by decreasing 0.5°C per cycle; 95°C × 30″, 55°C × 20″, 70°C × 30″, for 30 cycles, 72°C × 5′ for one cycle, 12°C × 1″ for one cycle). Then, Sanger sequencing was performed on all PCR products. All primers could be obtained in Table S2.

Characteristics analysis of circular RNAs

The gene length of circular/linear genes was calculated based on the B73 annotations (AGPv3) (Schnable *et al.*, 2009; Jiao *et al.*, 2017). As a control, the random sampling simulation method was used for linear transcripts. Briefly, 2009 (same as circular genes) genes without detectable circRNAs were selected randomly. The average gene length of each simulation were summarized and kept as one specific simulation value. Then, we repeated the process 1000 times to get 1000 simulation values for the specific feature, which were used for the comparison with the average gene length of circular RNAs for the significance test of specific feature enrichment. A similar method was used for the comparison of other genomic features between circRNAs and linear transcripts. For the conservation analysis of circRNAs, we first collected the circRNA information from rice and Arabidopsis (Lu *et al.*, 2015; Ye *et al.*, 2015), and then used BLAT (default parameters) (Kent, 2002) to align the circRNAs especially the junction sequences (upstream 50 bp and downstream 50 bp of the splicing site) identified in our study against the ones from rice and Arabidopsis. If maize circRNA had significant alignment with rice or Arabidopsis circRNAs (identity ≥ 0.9), then we carefully checked if the junction sites were conserved in both species. If the junction sequences were identified in both related species for specific back-splicing sites of homologous genes (both upstream 10 bp and downstream 10 bp of splicing site were successfully aligned), then the circRNAs were considered to be conserved circRNAs.

Bioinformatic analyses for microRNAs target predication, transposon enrichment and small RNA enrichment in maize circRNAs

In order to check if microRNAs targets are significantly enriched in circRNAs, psRNATarget was employed for microRNA target prediction with default parameters (Dai & Zhao, 2011). To detect the relationship between circular RNAs and transposons, repetitive information was obtained using REPEATMASKER (v.2.1; -species maize; http://www.repeatmasker.org/). A small RNA dataset on B73 seedlings was collected and mapped to maize reference genome (Zuo *et al.*, 2016). The small RNAs mapped on genome-wide LINE1-like elements (LLEs) were extracted and summarized using SAMTOOLS v.0.1.19 (Li *et al.*, 2009) for the enrichment analysis in maize circRNAs. A sampling simulation was conducted for the significant enrichment analyses of microRNAs target, transposon and small RNA reads around maize circRNAs. Detailed information about all the bioinformatic pipelines can be found in the Methods S1.

Discovery of potential functional roles for circRNAs

In order to test if circRNAs play a role in phenotypic variation, genes with trait-associated sites were collected from five different genome-wide association mapping (GWAS) studies (Li *et al.*, 2013; Peiffer *et al.*, 2014; Wallace *et al.*, 2014; Wen *et al.*, 2014; Yang *et al.*, 2014). Genes with significantly trait-associated sites are more likely to be associated with phenotypic variation. Thus, trait-associated genes could be a good resource for the functional enrichment analysis of circRNA genes. Likewise, the randomly sampling simulation method was employed for the comparison of appearance frequency in the trait-associated gene list between linear and circular genes. To rule out the spurious enrichment of circRNA genes in the trait-associated gene list, we performed functional enrichment analysis for genes without detectable circRNAs, the longest genes in B73 annotations, randomly selected genes with comparable lengths of flanking introns, randomly selected genes with the highest expression concentration, and randomly selected genes at the same time. Additionally, gene ontology (GO) enrichment analyses of circular genes were performed using AGRIGO (Du *et al.*, 2010) with default parameters.

Association analysis between presence/absence of LLERCPs and ear height variation in maize

A previous study has identified a functional gene *GRMZM2G089149*, that was associated with ear height variation in a US association panel (Peiffer *et al.*, 2014). Here, we found

that a cirRNA, circ1690, is derived from this ear-height associated gene and has intact LLEs and their reverse complementary pairs (LLERCPs). To test the relationship between LLERCPs and phenotypic variation, we randomly selected 43 diverse inbreds from a Chinese association panel (Yang *et al.*, 2014), designed primers to amplify the intact LLERCPs of circ1690, and conducted quantitative real-time PCRs (qRT-PCRs) to quantify the relative expression concentrations of circ1690 across different inbreds in the 3rd leaves of seedlings. For qRT-PCR, cDNA samples of linear and circular transcripts were amplified using the SYBR Premix Ex Taq™ II (Tli RNaseH plus) (TaKaRa) on the CFX96 Real-Time PCR detection system (Bio-Rad). Each PCR reaction contained 10 µl of reagent, consisting of 0.8 µl cDNA, 5 µl of the SYBR Premix Ex Taq™ II (Tli RNaseH plus), 3.4 µl of nuclease-free water, and 0.8 µl of the forward and reverse primers (10 µM stock). The qRT-PCR conditions included an initial incubation at 95°C for 30 s, followed by 40 cycles of 95°C for 5 s, 60°C for 30 s. The presence/absence of LLERCPs across 43 diverse inbreds was detected using PCR amplification and Sanger sequencing. Previous studies showed that the closest reverse complementary pair structure is predominantly associated with the formation of circRNAs (Jeck *et al.*, 2013; Zhang *et al.*, 2014). For this reason, we validated only those LLERCPs nearby circ1690. The amplification of LLEs in the 3rd and 11th introns could largely verify the presence of LLERCP of circ1690. A student's *t*-test and association mapping of mixed linear model (Tassel v3.0) were employed to test the ear height difference between inbreds with LLERCPs and without LLERCPs.

Data accessibility

The datasets of CircRNA-Seq (PRJNA356366) and poly (A) selected mRNA-Seq (PRJNA356498) were generated and deposited in the Sequence Read Archive database (https://www.ncbi.nlm.nih.gov/sra).

Customized bioinformatic scripts accessibility

All of the circRNA detection commands and bioinformatic scripts have been deposited in github (https://github.com/conniecl/maize_circRNA) and released for public accessibility free of charge. The results demonstrated in this study could be reproduced using these bioinformatic scripts and the same dataset.

Results

CircRNAs exist widely in maize

CircRNAs can be classified into intergenic and genic groups according to their origin. To detect circRNAs, we performed circRNA-Seq on 2-wk-old seedlings of the maize inbred line B73. We also collected a comprehensive transcriptome dataset, including 977 publicly available RNA-Seq datasets derived from diverse B73 tissues and from a diverse panel of maize inbreds (Hirsch *et al.*, 2014; Fu *et al.*, 2013; Table S1). A series of circRNA identification and quality-control steps were implemented to ensure only the inclusion of high-confidence circRNAs for further analysis (Fig. 1a; see Methods S1).

For circRNAs identified from 977 publicly available RNA-Seq datasets, only 256 circRNAs were maintained after a stringent filtering criterion. In the bench work, randomly selected 14 circRNA were elected for validation, four of 10 circRNAs derived from two genes were successfully validated and all four circRNAs from single genea were successfully validated (Table S2). Therefore, we excluded circRNAs from two genes in the downstream analyses. For circRNAs identified from the circRNA-Seq experiment, preliminary wet experiments with PCR and sequencing only validated ~10% of randomly selected circRNAs with only one junction read, suggestive of low confidence of circRNAs with only one junction read. However, the rate of validated circRNAs reached up to 85% when the number of junction reads went up to two. Therefore, circRNAs with at least two junction reads, were extracted in maize. In total, 5329 circRNAs were identified to have at least two junction reads covering the back splicing sites. Nearly half (2235) were derived from unknown/unannotated splicing sites detected by CIRI (Gao *et al.*, 2015). Of these circRNAs derived from known/annotated splicing sites, a small fraction (267) were identified by circ_finder (Westholm *et al.*, 2014), whereas the rest were identified by CIRCexplorer2 (Zhang *et al.*, 2016) and CIRI (Gao *et al.*, 2015) (Fig. S1). Surprisingly, < 21% of 5244 circRNAs could be detected by different tools. Based on the detection tools and whether the splicing sites are known or unknown, we classified all 5329 circRNAs into five groups: KNIFE, circ_finder, CIRCexplorer2, CIRI_annotated and CIRI_unannotated groups. Because the KNIFE group had been validated before, we only randomly selected 10 nonoverlapped circRNAs per group from other four circRNAs groups for the validation. Preliminary wet experiments with PCR and sequencing suggested that a majority of circRNAs could be validated for CIRCexplorer2 group (nine of 10) and CIRI_annotated group (seven of 10), whereas a few of circRNAs could be validated for CIRI_unannotated group (two of 10) and circ_finder group (three of 10) (Figs S1, S2; Table S2). These results confirmed that different software has different accuracy and sensitivity for the identification of maize circRNAs, consistent with the previous study (Hansen *et al.*, 2016). Together, we extracted circRNAs with at least two junction reads and derived from the CIRCexplorer2 and CIRI_annotated groups as high-confidence circRNAs in the further analyses.

In this study we focus mainly on exonic circRNAs, because they might be directly associated with gene expression and phenotypic variation (Fig. 1b). To further validate these high-confidence circRNAs, we amplified B73 leaf cDNAs from total RNAs or from RNAs, of which the linear mRNAs have been treated by RNase R, as well as genomic DNA using pairs of divergent and convergent primers for 15 randomly selected circRNAs. All convergent primers successfully amplified transcribed fragments with the expected length. Meanwhile, all 15 pairs of divergent primers yielded amplification products from

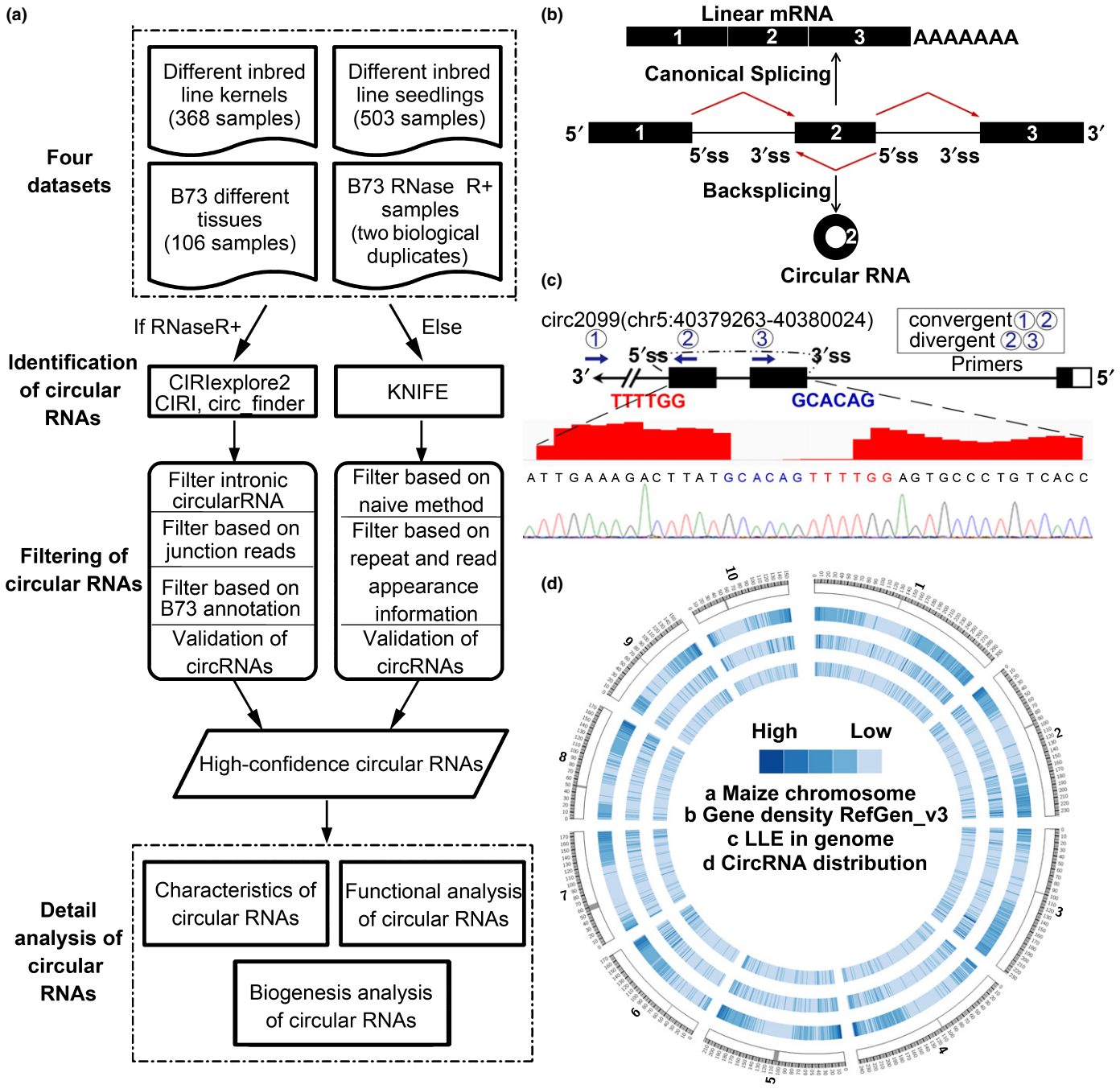

Fig. 1 Genome-wide identification of circular RNAs (circRNAs) in maize (a) The flowchart of bioinformatic analysis pipeline for the genome-wide identification of circRNAs in maize. (b) Origins of exonic circRNAs. Reverse red arrow, backsplicing junction site of the exonic circRNA; forward red arrow, normal splicing site; black rectangles, exons. (c) An example of circRNA that was validated via amplification and sequencing. The order of sequence of junction site was flipped from the order in the Sanger trace. Arrows (①, ②, ③) designate PCR primers. (d) Genome-wide distribution of circRNAs identified in maize leaf.

either cDNAs that had or had not been treated with RNase R, but not from the genomic DNAs. All of the amplification products from the divergent primers were shown to be derived from the target regions of circRNAs (i.e. spanning the junction sites) via Sanger sequencing (Figs 1c, S2; Table S2; see Methods S1). These results confirmed the reliability of our high-confidence circRNA discovery pipeline. After removing redundant circRNAs from the KNIFE (118; one from two transcripts were excluded; 32 were redundant), CIRCexplorer2 (1355) and CIRI_annotated groups (2448; 27 circRNAs from two transcripts were excluded; 1084 were redundant), we uncovered in total 2804 exonic circRNAs from 2009 gene loci (Table S3; Fig. S1). These high-confidence exonic circRNAs were distributed across the whole genome. However, as with genes, circRNA genes were more commonly found at both ends of chromosomes (Fig. 1d).

CircRNAs and their parental gene loci show distinct characteristics as compared to genes without detectable circRNAs

In order to characterize circRNAs and their genomic loci, we compared circRNAs with the linear transcripts from which they were derived and randomly selected genes without detectable circRNAs. As in previous reports, circRNAs exhibit distinct features in maize. Genes with detectable circRNAs are significantly longer than randomly selected genes (Fig. 2a). Overall, the average exon length of circRNAs is significantly shorter than that of randomly selected linear transcripts (Fig. 2b). Notably, the flanking intron length of the junctions of circRNAs is significantly larger than that of linear transcripts of randomly selected genes (Fig. 2c). Although circRNAs themselves are likely to be expressed at low concentrations, most of their parental genes are likely to be highly expressed (Fig. 2d). Notably, there is no obvious correlation between the accumulation of circRNAs and their corresponding parental genes even excluding circRNAs with low expression abundance (Fig. S3), suggesting complicated relationships between these two RNA types in maize. Additionally, a significantly higher proportion of circRNAs could be aligned with micro (mi)RNAs than that of the genome-wide linear transcripts (15/2804 for circRNAs vs 5/2804 for randomly selected linear transcripts. These circRNAs have an average of 1.33 miRNA binding sites ranging from 1 to 9, indicative of potential miRNA mimic capability. Furthermore, circRNAs are more likely to be expressed in a tissue-specific manner than are genome-wide linear transcripts (Fig. S4). The conservation analysis of circRNAs between related plant species showed that a small proportion (47

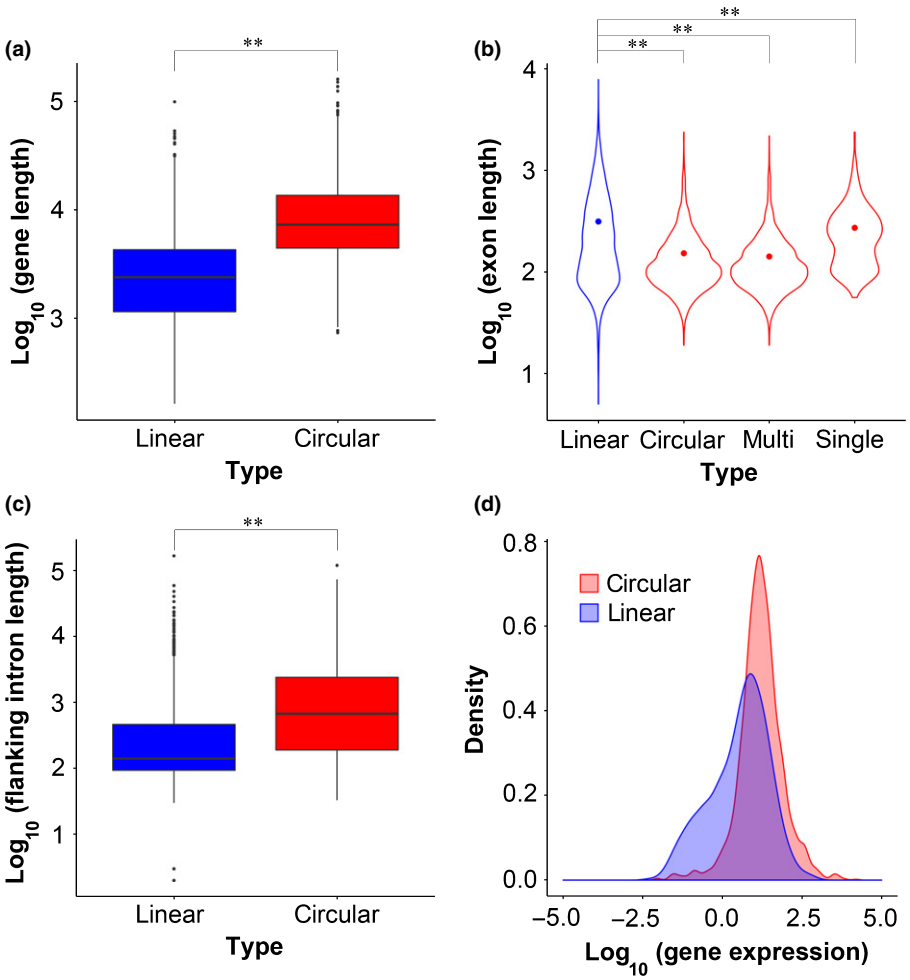

Fig. 2 Distinct genomic features of circular RNAs (circRNAs) in maize. (a) Comparison of gene lengths between genes with and without detected circRNAs. Linear and circular represent randomly selected genes without detectable circRNAs and genes with detectable circRNAs in B73 v3.26 annotation, respectively. (b) Comparison of exon length between exons with and without detectable circRNAs. Linear, random selected exons without detectable circRNAs; circular, exons with detectable circRNAs; multi and single, exons from circRNAs with multiple exons and single exon, respectively. (c) Comparison of flanking-intron length of junction sites between linear transcripts and circRNAs. *Y*-axis in (b, c) indicates the $\log_{10}$ of exon or intron length (bp). *X*-axis in (a–c) represents different types of circRNAs. The horizontal lines within box plots in (a) and (c) represent the median values. The black 'dots' in each boxplot are outliers, the values of which are either higher than the upper inner fence or lower than the lower inner fence of the boxplot. (a–c) Significant difference between groups: **, $P \leq 0.01$. (d) Comparison of expression concentrations (FPKM) between genes with and without detectable circRNAs. Linear and circular represent randomly-selected genes without detectable circRNAs and genes with detectable circRNAs in B73 v3.26 annotation, respectively.

of 2804) of maize circRNAs were conserved in rice, and three conserved circRNAs were found in Arabidopsis, suggesting either that there is low conservation of circRNAs or that the wrong tissues were sampled for the identification of circRNAs between different species. However, most features of maize circRNAs are similar to those described in other species, suggesting that fundamental features and regulation of circRNAs exhibit evolutionary conservation.

Retrotransposon LLEs and LLERCPs are significantly enriched in the flanking regions of circRNAs

In order to test if transposons play a role in the formation and expression of circRNAs, we annotated 35 kb of genomic sequences upstream and downstream of the circularization junctions (Fig. 3a) for repetitive elements using RepeatMasker (see Methods S1). There were six types of transposons (DNA/hAT-Tag1, DNA/PIF-Harbinger?, LINE/RTE-Bov, LTR/Copia, SINE/tRNA and LINE/L1) significantly enriched around either the upstream or the downstream flanking regions of the circRNA junction sites (Fig. S5). Interestingly, three transposons – retrotransposon LLE, LINE/RTE-Bov (LRE) and DNA/hAT-Tag1 like elements (DLE) were significantly enriched on both sides of these flanking genomic regions (two-fold more; Fig. S6), spreading up to 35 kb upstream and downstream of the circRNA junctions (Figs 3a, S7A). Like Alu sequences in humans (Zhang *et al.*, 2014), LLEs, LRE and DLEs have divergent or convergent distribution patterns around circRNAs, ensuring the formation of reverse complementary pairs of LLEs (LLERCPs), LREs (LRERCPs) and DLEs (DLERCPs). LLERCPs, LRERCPs and DLERCPs may form a Watson–Crick complementary stem-loop structure, of which the stem might be excised for the formation of circRNAs (Fig. 3b). However, LRERCPs and DLERCPs were significantly enriched for the flanking regions of circRNAs, of which the number was pretty small (4.14% together; Fig. S8). Herein, we focused on LLERCPs in the further analyses.

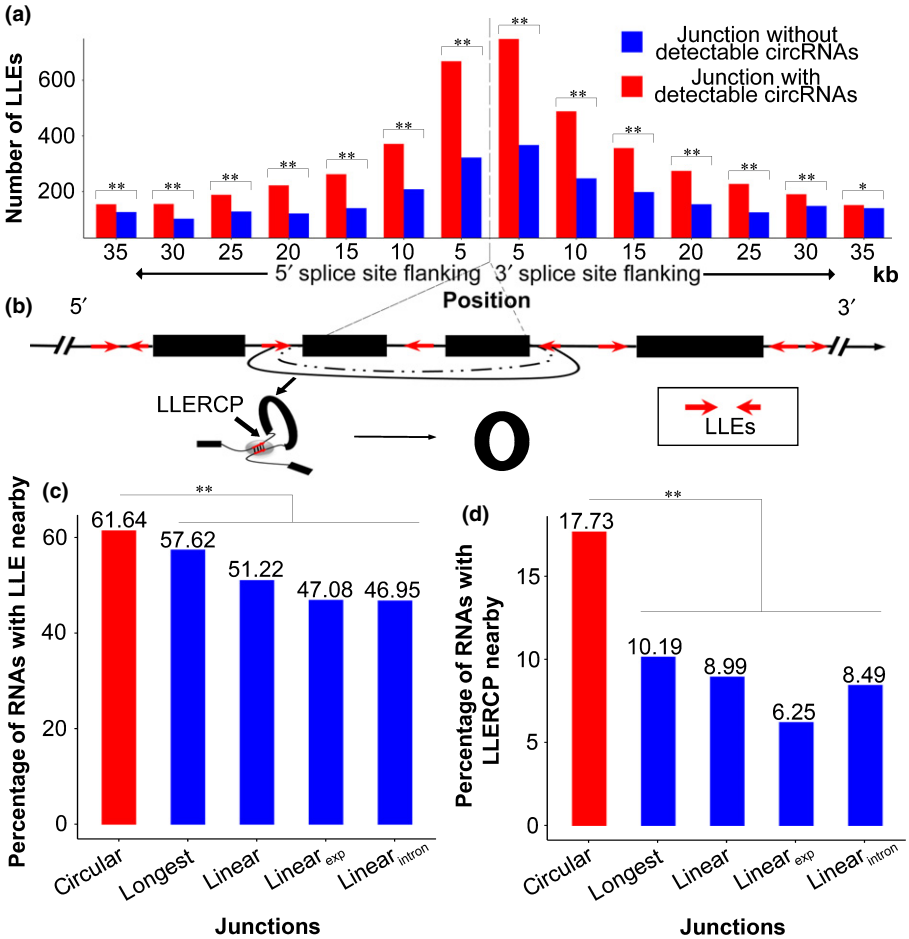

Fig. 3 Enrichment of retrotransposon LINE1-like elements (LLEs) and their reverse complementary pairs (LLERCPs) around circular RNAs (circRNAs) in maize. (a) Distribution of LLEs in the flanking regions of circular RNAs and randomly selected linear RNAs. **, $P < 0.01$; *, $P < 0.05$, according to the 1000 samplings. (b) Schematic diagrams showing LLERCPs and formation of circRNAs. LLERCPs may form stem-loop structure during the formation of circRNAs. Solid black line linking two red arrows, two LLEs have a reverse complementary relationship; dashed black line, back-splicing site of the exonic circRNA; grey oval in the stem-loop structure, reverse complementary region of LLERCPs; back arrow in the stem-loop structure, back-splicing process required for formation of circRNAs. (c, d) Enrichments of LLEs and LLERCPs in the flanking regions of genes with or without circRNAs. Linear, randomly selected genes without detectable circRNAs; $linear_{intron}$, genes with comparable length of flanking introns; $linear_{exp}$, genes with comparable expression concentrations to the ones with detectable circRNAs; longest, genes with longest gene length annotated in the reference genome.

Compared to randomly selected genes without detectable circRNAs, genes with longest gene length annotated by the maize sequencing project (Schnable *et al.*, 2009; Jiao *et al.*, 2017), genes with comparable length of flanking introns, and genes with higher comparable expression concentrations, LLEs are significantly enriched in genes with detectable circRNAs (Fig. 3c). Notably, there was a weak significant difference (62% vs 58%) for the enrichment of LLEs between genes with detectable circRNAs and genes with longest gene length, suggesting that the existence of LLEs *per se* might not be the direct reason for the formation of circRNAs. However, LLERCPs are significantly more likely to be located in the flanking regions of genes with detectable circRNAs compared to genes with comparable long introns, genes with longest gene length, genes without detectable circRNAs and genes with higher comparable expression concentrations ($P < 0.01$; Fig. 3d). In addition, LLERCPs exhibit greater enrichment in the upstream and downstream of circRNAs than do LLEs *per se* (300% change vs 20% more; Fig. 3d), suggesting that LLERCPs, rather than solely LLEs, play a role in the process of RNA circularization (Fig. 3b), which accords with that in humans and animals (Zhang *et al.*, 2014).

LLERCPs are associated with the expression of circular and linear RNAs

In order to determine if LLERCPs affect the accumulation of circRNAs, we identified 51 genes for which both circRNAs with LLERCPs and circRNAs without LLERCPs have been detected (Fig. 4a). A paired *t*-test of circRNAs with and without LLERCPs for these 51 gene loci showed that circRNAs with LLERCPs accumulated to significantly higher concentrations than did circRNAs without LLERCPs nearby, indicating that LLERCPs could reinforce the expression of circRNAs (Fig. 4b). A similar significant difference between circRNAs with LLERCPs and circRNAs without LLERCPs nearby was also observed if we only accounted for the very closest flanking introns for the

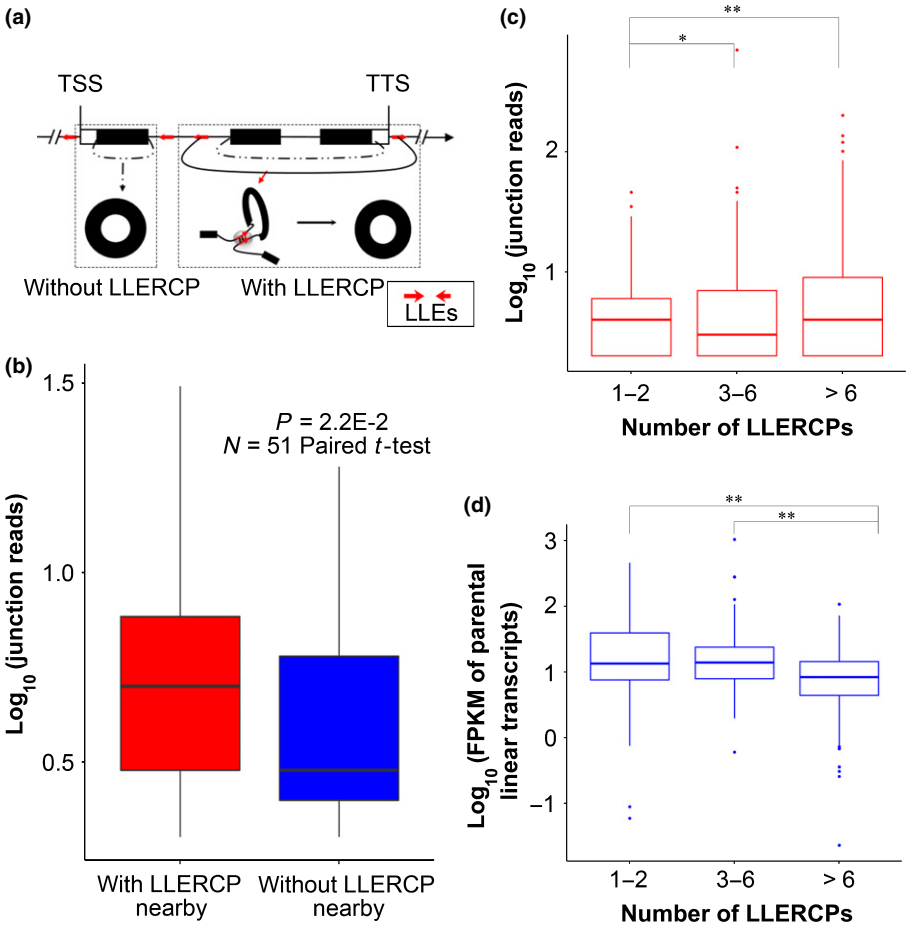

Fig. 4 The number of reverse complementary pairs of LINE1-like elements (LLERCPs) is associated with the expression of circular RNAs (circRNAs) and linear transcripts in maize. (a) A schematic diagram of the origination of circRNAs. There are some gene loci, where both circRNAs without intact LLERCPs and circRNAs with LLERCPs are observed. LLERCPs may form stem-loop structures before recruitment of splicesomes during the formation of circRNAs. (b) Expression concentrations of circRNAs with LLERCPs nearby are significantly higher than those of circRNAs without intact LLERCPs nearby for specific gene loci. (c) Expression-concentration variation of circRNAs along with the increase of LLERCPs. Red line links the average value of each group. The number under each box plot shows the number of circRNAs in each LLERCP group. (d) Expression-concentration variation of linear transcripts along with the increase of LLERCPs. The horizontal lines within box plots in (b–d) represent the median values. The 'dots' in each boxplot are outliers, the values of which are either higher than the upper inner fence or lower than the lower inner fence of the boxplot. Significance concentration: *, $P \leq 0.05$; **, $P \leq 0.01$.

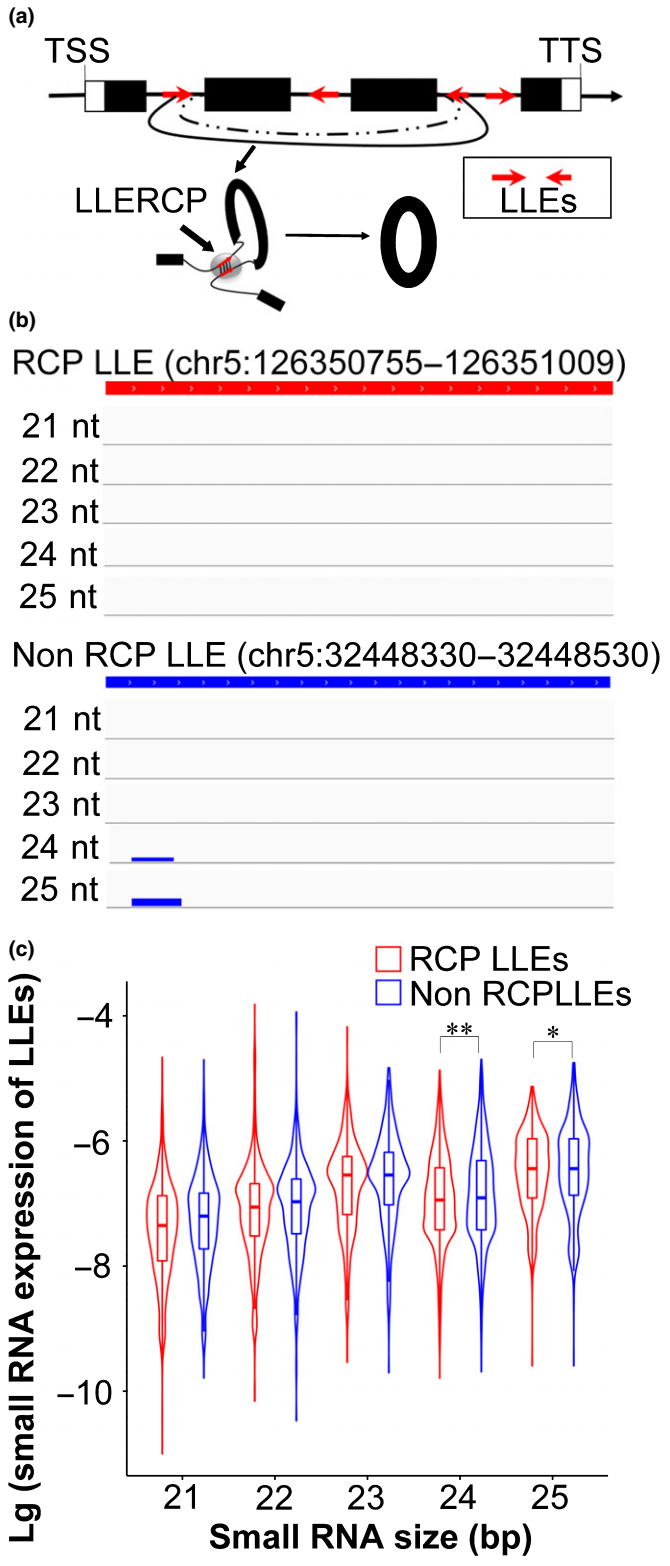

Fig. 5 Depletion of small RNAs from reverse complementary pairs of LINE1-like elements (LLEs) in maize (LLERCPs). (a) A schematic diagram of LLERCPs that might be involved in the formation of circular RNAs (circRNAs) in maize. LLERCPs might form a stem-loop structure, which may either be excised for the biogenesis of small RNAs, or provide a stable structure to recruit the splicesome for formation of circRNAs. (b) An example of alignment of small RNAs in genic LLEs with reverse complementary pairs and randomly selected genic LLEs. Randomly selected genic LLEs had more small RNAs aligned than genic LLE with reverse complementary pairs, indicative of depletion of small RNAs (especially 24 nt species) in RCP LLEs. (c) Normalized accumulation of genic RCP LLE-derived 24 nt small RNAs is significantly lower than that of genic non-RCP LLE-derived small RNAs. Significance concentration: *, $P \leq 0.05$

identification of LLERCPs (Fig. S9A,B). Different genes are controlled by different transcription regulatory mechanisms, and are usually expressed in different amounts, which makes direct comparison among genes difficult. Most interestingly, although the expression–concentration of different genes varies, the expression concentrations of corresponding circRNAs significantly varied along with the increase of the number of LLERCPs, confirming the potential regulatory role of LLERCPs in circRNAs ($P < 0.05$; Fig. 4c). In contrast to circRNAs, the expression–concentrations of linear transcripts significantly decreased as the number of LLERCPs increased ($P < 0.001$; Fig. 4d). The complicated pattern of expression–concentration variation for circRNAs and linear transcripts along with the increase of LLERCPs may indicate a complex interaction between circRNAs and linear RNAs. However, we did not observe such relationships between the number of LLERCPs and the expression–concentrations of linear and circular transcripts if we only accounted for the LLERCPs in the flanking introns (Fig. S9C,D). This may be due to either the small number of gene loci for the analysis or dramatic expression–concentration variation among different genes at a specific tissue, which makes comparison difficult. Taken together, these results indicate that the number of LLERCPs may play a role in the regulation of the expression of circRNAs and linear RNAs.

LLERCPs are less likely to be associated with small RNAs in maize

LLERCPs may form stem-loop structures during the formation of circRNAs (Fig. 5a). These stem-loop structures are similar to the hairpin structures associated with the formation of small RNAs. To test the relationship between small RNAs and LLERCPs, we downloaded the small RNA-Seq dataset from maize B73 2-wk-old seedlings, mapped small RNAs against the whole maize genome, and quantified the accumulation concentrations of 21 nt to 25 nt species of small RNAs within all LLEs (Zuo *et al.*, 2016). Overall, the normalized number of small RNA reads that were aligned to circRNA flanking LLERCPs was less than the number aligned to genic randomly selected LLEs (Fig. 5b,c). Specifically, the accumulation concentrations of 24 nt and 25 nt small RNAs on genic LLERCPs are significantly lower than those of randomly selected genic LLEs. These results indicate that LLERCPs may stabilize precursor transcripts for the subsequent formation of circRNAs.

Variation in LLERCP content is associated with phenotypic variation

A series of analyses was conducted to test for potential functional roles of circRNAs. First, a GO enrichment analysis showed that

genes that give rise to circRNAs are more likely to be associated with a cellular process ($P = 4.40E-14$), binding ($P = 8.20E-09$) or an organelle ($P = 3.4E-27$) than the reference genes in the B73 genome (Fig. S10). Additionally, as compared with genes without detectable circRNAs, genes with detectable circRNAs were more likely to overlap a set of 7500 nonredundant genes (Table S4) that have been associated with phenotypic variation based on GWAS (see Methods S1), as tested via the random sampling simulation ($P \leq 0.001$; see Methods S1). The proportion of genes with circRNAs that harbour trait-association sites is significantly higher than the randomly selected genes without detectable circRNAs ($P = 1.37E-10$), longest genes in B73 annotation but without detectable circRNAs ($P = 7.97E-04$), randomly selected genes with similar flanking intron length as circRNAs genes but without detectable circRNAs ($P = 9.00E-33$), and randomly selected genes with similar parent gene expression concentration as circRNAs genes but without detectable circRNAs ($P = 1.81E-06$; Fig. 6a). This enrichment suggests that genes with circRNAs are more likely to be involved in the phenotypic variation. More intriguingly, *c.* 17% of genes with LLERCPs and GWAS signals are the ones with circRNAs (Fig. 6b). Such an enrichment of LLERCPs in genes with circRNAs and GWAS signals is much higher than in the randomly selected genes without detectable circRNAs ($P = 9.73E-12$). Meanwhile, an enrichment was also observed when compared to genes without detectable circRNA but with longest gene length in the B73 annotation ($P = 8.81E-09$), randomly selected genes with similar flanking intron length as circRNAs genes but without detectable circRNAs ($P = 2.12E-17$), and randomly selected genes with similar parent gene expression distribution as circRNAs genes but without detectable circRNAs ($P = 3.90E-31$; Fig. 6b). The enrichment of LLERCPs suggests that LLERCPs may be the causal reason for the phenotypic variations.

If circRNAs contribute to phenotypic variation, we would expect that LLERCPs would exhibit presence/absence polymorphism in the trait-associated genes among the lines used in the GWAS and that this variation would itself be associated to phenotypic variation. To test this hypothesis, we profiled a gene, *GRMZM2G089149* (Fig. 7a), that was significantly associated with ear height in a US association panel (Peiffer *et al.*, 2014), and that contains LLERCPs and accumulates a circRNA, circ1690, for variation in LLERCPs among 43 diverse inbreds in a Chinese association panel (Fig. 7b). *GRMZM2G089149* has 14 exons and 13 introns, of which the 1st and 14th are the biggest exons. Circ1690, a circRNA, spans the 3rd to 11th exons, and is associated with the LLERCP within the 2rd and 11th introns. Although there are two LLEs in the 3rd intron, both LLEs are co-segregated in the 43 diverse inbreds. Therefore, only one LLERCP between 3′ UTR and 11th introns can be formed. *GRMZM2G089149* encodes an unknown protein with a UVB photosynthesis domain, which may be involved in plant vegetative growth and further related to ear height (Tong *et al.*, 2008; Peiffer *et al.*, 2014). Genomic amplification and qRT-PCR analysis showed that nearly all inbreds (except one) having LLERCPs in *GRMZM2G089149* had detectable amounts of circ1690 (Table S5), and the accumulated detectable concentrations of circ1690 in 11 inbreds with LLERCPs are significant higher than those inbreds without RCPLLEs ($P = 3.2E-03$; Fig. 7c). These results provide evidence that presence/absence of LLERCPs is related to the formation and accumulation of circRNAs. Furthermore, the expression of circ1690 is negatively associated with the concentration of the linear transcript of *GRMZM2G089149* (PCC = −0.68; $P = 1.87E-04$; Fig. 7d), suggesting that the expression of circRNAs may have functional consequences on the linear transcripts. More interestingly, inbreds with detectable LLERCPs are significantly taller than those without detectable LLERCPs ($P = 3.51E-02$; Fig. 7e), consistent with the hypothesis that the presence/absence of LLERCPs affects variation in ear height. Notably, mixed linear model with controlled population structure and inbred kinship showed a weak significant association ($P = 5.11E-02$) between LLERCPs and ear height variation,

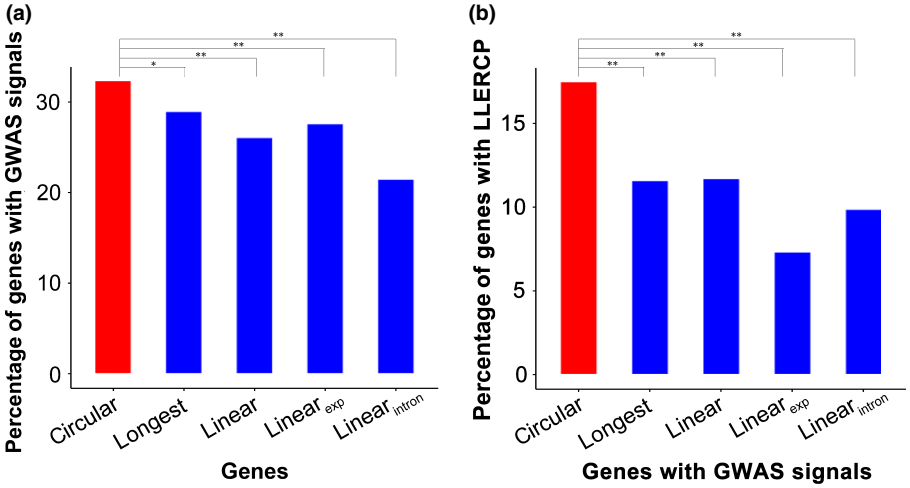

Fig. 6 Genes associated with phenotypic variation are enriched among genes with detectable circular RNAs (circRNAs) and reverse complementary pairs of LINE1-like elements (LLERCPs) in maize. (a) Enrichment of genome-wide association mapping (GWAS) signals in genes with detectable circRNAs. (b) Enrichment of LLERCPs in genes with both detectable circRNAs and GWAS signals. Linear, randomly selected genes without detectable circRNAs; $linear_{intron}$, genes with comparable length of flanking introns; $linear_{exp}$, genes with comparable expression concentrations to the ones with detectable circRNAs; longest, genes with longest gene length annotated in the reference genome. Significance concentration: **, $P < 0.01$; *, $P < 0.05$.

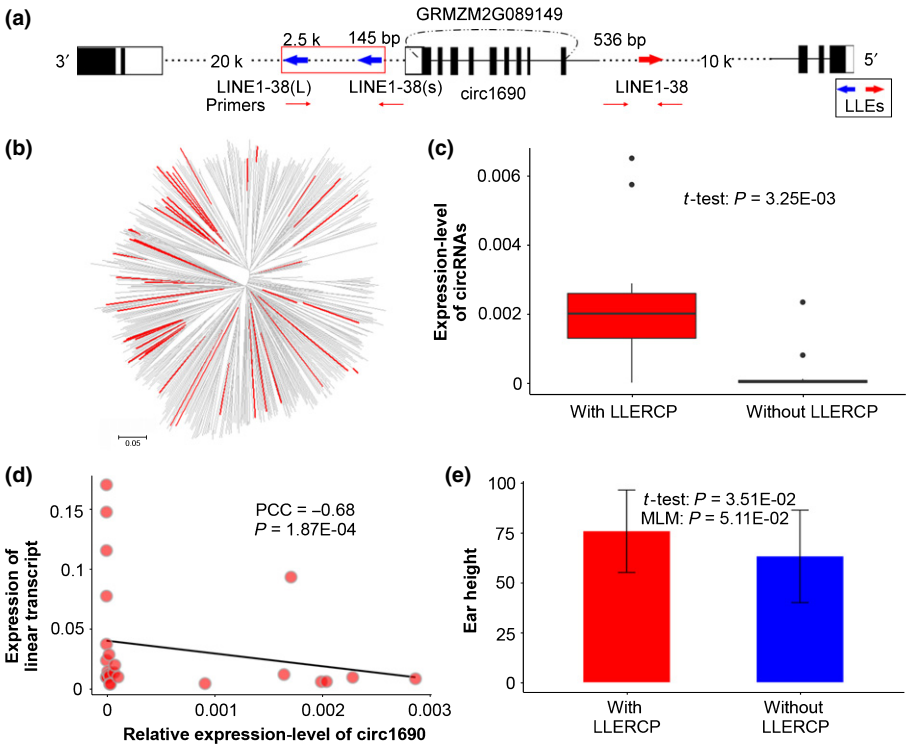

Fig. 7 The presence/absence of reverse complementary pairs of LINE1-like elements (LLERCPs) is associated with variation in ear height among maize inbreds. (a) circ1690 is derived from the gene *GRMZM2G089149*. The B73 allele of *GRMZM2G089149* contains an intact LLERCP. Red box indicates that two LLEs were amplified by one pair of primers together. (b) Thirty-eight diverse inbreds (in red) were randomly selected for analysis from a larger association mapping panel in maize (Yang *et al.* 2011). (c) The presence/absence of LLERCPs in *GRMZM2G089149* is associated with the accumulation of detectable concentrations of circ1690 among inbreds. (d) The accumulation of circ1690 is associated with that of the linear transcript from *GRMZM2G089149*. PCC indicates Pearson Correlation Coefficient. According to the results in (c), there were four inbreds with exceptionally higher expression–concentrations of circular RNAs (circRNAs) as outliers, which may be due to either the other molecular mechanism for the formation of circRNAs or stochastic error. We excluded these four inbreds for the expression analysis between circ1690 and its counterpart linear transcript. The horizontal lines within box plots represent the median values of expression–concentration of circRNAs. The black 'dots' in each boxplot are outliers, the values of which are either higher than the upper inner fence or lower than the lower inner fence of the boxplot. (e) Inbreds with LLERCP are significantly higher than those that do not contain LLERCPs. *t*-test, Student's *t*-test statistical analysis; MLM, mixed linear model analysis by controlling population structure and kinships among inbreds. Error bars represent deviation of ± 1 SE from the mean value.

which might be due to either the small analysed population size or the existence of a rare allele in the Chinese association panel. We did not, however, observe a correlation between ear height and the accumulation of linear or circular transcripts across inbreds (Table S5). This might be due to the fact that the tissue analysed (topmost leaf of 2-wk-old seedlings) was not associated with ear height. Even so, these results provide intriguing clues about a mechanism by which transposons may play a functional role in the phenotypic variation via the formation of circRNAs.

Discussion

Widespread existence of circRNAs in Eukaryotes

With the advent of next generation sequencing, ENCODE Projects of humans and mice have been completed, which have dramatically revolutionized our understanding of the eukaryotic transcriptome (Djebali *et al.*, 2012; Dunham *et al.*, 2012; Yue *et al.*, 2014). Many lines of evidence have shown the prevalence of transcription from the whole genome, leading to the discovery of hundreds of thousands of long noncoding RNAs (lncRNAs). LncRNAs have been shown to function as super-regulators at the genome, transcriptome, protein and metabolic levels (Rinn & Chang, 2012). Although one type of lncRNA, circular (circ) RNAs, were identified over 30 yr ago (Hsu & Coca-Prados, 1979), only a handful of circRNAs had been identified until recently due to the limitation of molecular techniques and were considered to be transcriptional noise as a consequence of their structure lacking a 3′ poly(A) tail and 5′ end caps (Hsu & Coca-Prados, 1979; Nigro *et al.*, 1991; Capel *et al.*, 1993; Pasman *et al.*, 1996; Zaphiropoulos, 1996). Recently, with the development of circRNA-Seq, thousands of circRNAs have been identified in animals and plants (Jeck *et al.*, 2013; Chen, 2016). These circRNAs are much more stable than linear transcripts, are conserved among species, and their flanking sequences are associated with Alu elements in animals (Jeck *et al.*, 2013). In Rice and Arabidopsis, over 2000 circRNAs have been identified, showing distinct genomic features relative to linear genes and transcripts (Ye *et al.*, 2015). In maize, we collected a comprehensive dataset from different Next Generation Sequencing (NGS) techniques, including traditional linear mRNA-Seq and circRNA-Seq with circRNA enrichments, and identified over 5000 circRNAs in

maize, of which 2804 were high-confidence circRNAs. Taken together, detection of circRNAs in mammals and plants suggests that circRNAs are ubiquitous, and thus are not likely to be transcriptional noise or a by-product of transcription of functional linear transcripts, but instead function as micro (mi)RNA mimics and regulators of gene expression (Chen, 2016).

Our first study of maize circRNAs not only detected thousands of circRNAs, but also found many interesting aspects of circRNAs in maize. Maize circRNAs show distinct features, similar to those of circRNAs in other species. Most strikingly, although transposons and reverse complementary pairs of repetitive elements have not been shown to be enriched in the flanking regions of the circRNAs of rice and Arabidopsis (Lu *et al.*, 2015; Ye *et al.*, 2015), maize circRNAs are more likely to be surrounded by retrotransposon LINE1-like elements (LLEs) and their reverse complementary pairs (LLERCPs), which is similar to what has been observed in humans (Jeck *et al.*, 2013; Zhang *et al.*, 2014) and also might be related to the biogenesis of circRNAs in maize.

Biogenesis of circRNAs in maize

CircRNAs have been uncovered for a long time (Hsu & Coca-Prados, 1979). However, besides circular single-stranded RNA genomes of viroids and the hepatitis delta virus, and structural circRNAs of tRNAs and rRNAs by ribozymes (Group I and Group II introns) (Grabowski *et al.*, 1981; Lehmann & Schmidt, 2003), most other endogenous mRNAs were usually thought to be by-products of pre-mRNA processing, and thus were considered to be formed by splicing errors (Capel *et al.*, 1993; Pasman *et al.*, 1996; Zaphiropoulos, 1996). Nevertheless, deep high-throughput sequencing and robust bioinformatic analysis methods have detected tens of thousands of exonic and intronic endogenous circRNAs (Jeck *et al.*, 2013; Zhang *et al.*, 2014; Lu *et al.*, 2015; Ye *et al.*, 2015), which are abundant and even more stable than their parental linear RNAs. Ample evidence has implied that endogenous circRNAs may not solely be formed by splicing error, but also by certain biological mechanisms.

It has been reported that circRNAs are surrounded by long introns, and that intronic elements are sufficient to promote circularization in the case of *sex-determining region Y* (*Sry*) circRNA (Dubin *et al.*, 1995). Recently, Jeck *et al.* (2013) showed that introns flanking circRNAs are enriched in Alu repeats in humans. Furthermore, circRNAs are generated co-transcriptionally with linear transcripts and their production rate is determined mainly by intronic sequences (Ashwal-Fluss *et al.*, 2014). Moreover, Zhang *et al.* (2014) analysed the enrichment of transposon Alus and their reverse complementary structures, and validated that the complementary sequence could mediate RNA circularization. Finally, lariat-driven circularization and intron pairing-driven circularization were proposed for the biogenesis of circRNAs (Chen *et al.*, 2015, 2016).

Here, we focused primarily on the exonic endogenous circRNAs and bioinformatic analyses of flanking genomic sequences of circRNAs and we uncovered that retrotransposon LLEs and LLERCPs were significantly enriched around circRNAs, which is concordant with the intron-driven circularization model. Notably, not all of the genes with LLERCPs have detectable circRNAs in our study. This may be due to the fact that these genes might not have been expressed in our sample, given the strong tissue-specific expression feature of circRNAs. We noticed that a substantial number of circRNAs in our study do not have LLERCPs either. These circRNAs may be caused either by reverse complementary sequences of other repetitive elements (Fig. S8) or other biological mechanisms. Of particular note, there is also an enrichment of reverse complementary pairs in genes with detectable circRNAs, which are over 10 kb from circRNA junction sites. This indicates that genomic structure, especially the 3D structure, may play a role in the formation of circRNAs. Because maize lacks 3D genome data, we analysed 3D data and circRNAs in humans. By taking the human GM12878 1-kb resolution intrachromosomal contact matrix (Rao *et al.*, 2014) and human GM12878 circRNAs (CIRCpedia, http://www.picb.ac.cn/rnomics/circpedia/) together, we found that 90.83% of circRNA splice sites had spatial interaction, suggestive of a potential relationship between circRNAs and genomic 3D structure. Small RNAs exist widely and are usually derived from double-stranded RNAs (Vazquez *et al.*, 2010). In plants, small RNAs can silence active transposable elements (Fultz *et al.*, 2015). The depletion of small RNA reads within the flanking LLERCPs of circRNAs implies the importance of LLERCP structure for the formation of circRNAs, and the link between small RNAs and the biogenesis of circRNAs. Our results not only provide further evidence that reverse complementary sequences can mediate the cotranscription of circRNAs and linear RNAs, but also highlight the important roles of the 5′ and 3′ regions of genes in the formation or regulation of circRNAs and linear transcripts in maize.

A new functional role in phenotypic variation for transposons in maize

Transposons or transposon-like repetitive elements constitute >85% of the maize genome (Schnable *et al.*, 2009; Jiao *et al.*, 2017). These elements dramatically diversified the maize genome, creating or reversing mutations, causing presence/absence variation or copy number variation of genomic regions, and even altering genome size among maize inbreds. As a result, the maize genome exhibits dramatic diversity among inbreds (Lai *et al.*, 2010). Accordingly, maize phenotypic variation could be caused by the insertion of transposons either in the protein coding regions with amino acid alteration, or in flanking regions of functional genes with epigenomic variation, which can further alter nearby gene expression (Wei & Cao, 2016). Besides, phenotypic variation could result from the small interfering RNAs, which are generated by homologous genomic sequences captured and moved by transposons from the original functional genes (Lisch, 2009). It has been widely reported that the phenotypic variations of maize plant architecture (*tb1* – tillering; Wang *et al.*, 1999), drought resistance (*ZmVPP1*; X. Wang *et al.*, 2016) and pathogen resistance (Yang *et al.*, 2013; Zuo *et al.*, 2014) were associated with transposons, which either altered the chromatin status or resulted in coding changes of the functional genes.

In the present study, we observed that genes with circRNAs are more likely to be associated with trait-associated sites (TAS), which

were identified by whole genome association mapping. Most intriguingly, *c.* 17% of these circRNA genes with TAS are associated with LLERCPs. Given the dramatic genomic structural variation among maize inbreds, which is caused primarily by transposons, we propose a potential functional role for transposons in modulating phenotypic variation through the formation of circRNAs and regulation of gene expression in maize. For a gene affecting ear height, absence of LLERCP will predominantly produce a linear transcript and lead to a certain amount of ear height. When there are LLERCPs, both circRNA and linear transcript could be produced, and the circRNA could potentially interact with the linear transcript either reducing or increasing its expression, resulting in variations in ear height. The association mapping between ear height variation and the accumulation of circRNA-circ1690, which is derived from an ear height-associated gene, supports the plausibility of our proposed model. Taken together, our study suggests a new functional role for transposons to form circRNAs that modulate phenotypic variation in maize, a role which may also function in other plants or animals.

Conclusions

We performed circRNA-Seq on 2-wk-old seedling leaves of maize reference inbred B73, carried out genome-wide discovery and characterization of circRNAs in maize, and explored the biogenesis mechanism of circRNAs and the potential functional roles in maize. Two key findings can be obtained from our work: (1) LLEs and their LLERCPs are significantly enriched in the flanking regions of circRNAs, suggesting that transposons are critical to the formation of circRNAs in maize; (2) the accumulation of circRNA transcripts and linear parental transcripts is associated with the number of LLERCPs, and genes with transposon-derived circRNAs are likely to be associated with phenotypic variation.

Acknowledgements

This research was supported by the National Key Research and Development Program of China to M.D. and L.L. (2016YFD0100600; 2016YFD0100802), and by Huazhong Agricultural University Scientific & Technological Self-innovation Foundation to L.L. (program no. 2015RC016).

Author contributions

L.L. and M.D. designed and supervised this study; L.C., P.Z., Y.F. and Q.Lu performed the data analysis and circular RNA validation; L.C. and L.L. wrote the manuscript; and P.S.S., G.J.M., Q.Li and J.Y. reviewed and edited the manuscript. L.C., P.Z. and Y.F. contributed equally to this work.

References

Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, Evantal N, Memczak S, Rajewsky N, Kadener S. 2014. circRNA biogenesis competes with pre-mRNA splicing. *Molecular Cell* **56**: 55–66.

Bachmayr-Heyda A, Reiner AT, Auer K, Sukhbaatar N, Aust S, Bachleitner-Hofmann T, Mesteri I, Grunt TW, Zeillinger R, Pils D. 2015. Correlation of circular RNA abundance with proliferation – exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. *Scientific Reports* **5**: 8057.

Beló A, Beatty MK, Hondred D, Fengler KA, Li B, Rafalski A. 2010. Allelic genome structural variations in maize detected by array comparative genome hybridization. *Theoretical and Applied Genetics* **120**: 355–367.

Bennetzen JL. 2005. Transposable elements, gene creation and genome rearrangement in flowering plants. *Current Opinion in Genetics & Development* **15**: 621–627.

Capel B, Swain A, Nicolis S, Hacker A, Walter M, Koopman P, Goodfellow P, Lovell-Badge R. 1993. Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell* **73**: 1019–1030.

Chen LL. 2016. The biogenesis and emerging roles of circular RNAs. *Nature Reviews Molecular Cell Biology* **17**: 205–211.

Chen I, Chen CY, Chuang TJ. 2015. Biogenesis, identification, and function of exonic circular RNAs: biogenesis, identification, and function of exonic circular RNAs. *Wires RNA* **6**: 563–579.

Chen LL, Yang L. 2015. Regulation of circRNA biogenesis. *RNA Biology* **12**: 381–388.

Chen L, Yu Y, Zhang X, Liu C, Ye C, Fan L. 2016. PcircRNA_finder: a software for circRNA prediction in plants. *Bioinformatics* **32**: 3528–3529.

Chen LL, Yang L. 2015. Regulation of circRNA biogenesis. *RNA Biology* **12**: 381–388.

Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC *et al.* 2012. Maize HapMap2 identifies extant variation from a genome in flux. *Nature Genetics* **44**: 803–807.

Chu Q, Zhang X, Zhu X, Liu C, Mao L, Ye C, Zhu QH, Fan L. 2017. PlantcircBase: a database for plant circular RNAs. *Molecular Plant* **S1674–2052**: 30074–30076.

Dai X, Zhao PX. 2011. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Research* **39**: W155–W159.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F *et al.* 2012. Landscape of transcription in human cells. *Nature* **489**: 101–108.

Dooner HK, He L. 2008. Maize genome structure variation: interplay between retrotransposon polymorphisms and genic recombination. *Plant Cell* **20**: 249–258.

Du Z, Zhou X, Ling Y, Zhang Z, Su Z. 2010. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Research* **38**: W64–W70.

Dubin RA, Kazmi MA, Ostrer H. 1995. Inverted repeats are necessary for circularization of the mouse testis *Sry* transcript. *Gene* **167**: 245–248.

Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R *et al.* 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.

Fu H, Dooner HK. 2002. Intraspecific violation of genetic colinearity and its implications in maize. *Proceedings of the National Academy of Sciences, USA* **99**: 9573–9578.

Fu J, Cheng Y, Linghu J, Yang X, Kang L, Zhang Z, Zhang J, He C, Du X, Peng Z *et al.* 2013. RNA sequencing reveals the complex regulatory network in the maize kernel. *Nature Communications* **4**: 3832.

Fultz D, Choudury SG, Slotkin RK. 2015. Silencing of active transposable elements in plants. *Current Opinion in Plant Biology* **27**: 67–76.

Gao Y, Wang J, Zhao F. 2015. CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome Biology* **16**: 4.

Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J *et al.* 2009. A first-generation haplotype map of maize. *Science* **326**: 1115–1117.

Grabowski PJ, Zaug AJ, Cech TR. 1981. The intervening sequence of the ribosomal RNA precursor is converted to a circular RNA in isolated nuclei of tetrahymena. *Cell* **23**: 467–476.

Guo JU, Agarwal V, Guo H, Bartel DP. 2014. Expanded identification and characterization of mammalian circular RNAs. *Genome Biology* **15**: 409.

Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. 2013. Natural RNA circles function as efficient microRNA sponges. *Nature* **495**: 384–388.

Hansen TB, Venø MT, Damgaard CK, Kjems J. 2016. Comparison of circular RNA prediction tools. *Nucleic Acids Research* **44**: e58.

Hansey CN, Vaillancourt B, Sekhon RS, de Leon N, Kaeppler SM, Buell CR. 2012. Maize (*Zea mays* L.) genome diversity as revealed by RNA-Sequencing. *PLoS ONE* **7**: e33071.

Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, Penagaricano F, Lindquist E, Pedraza MA, Barry K *et al.* 2014. Insights into the maize pan-genome and pan-transcriptome. *Plant Cell* **26**: 121–135.

Hirsch CN, Hirsch CD, Brohammer AB, Bowan MJ, Soifer I, Barad O, Shem-Tov D, Baruch K, Lu F, Hernandez AG *et al.* 2016. Draft assembly of elite inbred line PH207 provides insights into genomic and transcriptome diversity in maize. *Plant Cell* **28**: 2700–2714.

Hsu MT, Coca-Prados M. 1979. Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. *Nature* **280**: 339–340.

Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, Sharpless NE. 2013. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**: 141–157.

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS *et al.* 2017. Improved maize reference genome with single-molecule technologies. *Nature* **546**: 524–527.

Kent WJ. 2002. BLAT – the BLAST-like alignment tool. *Genome Research* **12**: 656–664.

Kulcheski FR, Christoff AP, Margis R. 2016. Circular RNAs are miRNA sponges and can be used as a new class of biomarker. *Journal of Biotechnology* **238**: 42–51.

Lai J, Li R, Xu X, Jin W, Xu M, Zhao H, Xiang Z, Song W, Ying K, Zhang M *et al.* 2010. Genome-wide patterns of genetic variation among elite maize inbred lines. *Nature Genetics* **42**: 1027–1030.

Lehmann K, Schmidt U. 2003. Group II introns: structure and catalytic versatility of large natural ribozymes. *Critical Reviews in Biochemistry and Molecular Biology* **38**: 249–303.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, Han Y, Chai Y, Guo T, Yang N *et al.* 2013. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nature Genetics* **45**: 43–50.

Li L, Eichten SR, Shimizu R, Petsch K, Yeh CT, Wu W, Chettoor AM, Givan SA, Cole RA, Fowler JE *et al.* 2014. Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biology* **15**: R40.

Liang D, Wilusz JE. 2014. Short intronic repeat sequences facilitate circular RNA production. *Genes & Development* **28**: 2233–2247.

Lisch D. 2009. Epigenetic regulation of transposable elements in plants. *Annual Review of Plant Biology* **60**: 43–66.

Lisch D. 2013. How important are transposons for plant evolution? *Nature Reviews Genetics* **14**: 49–61.

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

Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M *et al.* 2013. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**: 333–338.

Morris KV, Mattrick JS. 2014. The rise of regulatory RNA. *Nature Reviews Genetics* **15**: 423–437.

Nigro JM, Cho KR, Fearon ER, Kern SE, Ruppert JM, Oliner JD, Kinzler KW, Vogelstein B. 1991. Scrambled exons. *Cell* **64**: 607–613.

Pasman Z, Been MD, Garcia-Blanco MA. 1996. Exon circularization in mammalian nuclear extracts. *RNA* **2**: 603–610.

Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z, Millard MJ, Gardner CA, McMullen MD, Holland JB, Bradbury PJ *et al.* 2014. The genetic architecture of maize height. *Genetics* **196**: 1337–1356.

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES *et al.* 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**: 1665–1680.

Rinn JL, Chang HY. 2012. Genome regulation by long noncoding RNAs. *Annual Review of Biochemistry* **81**: 145–166.

Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. 2012. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS ONE* **7**: e30733.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA *et al.* 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**: 1112–1115.

Springer NM, Ying K, Fu Y, Ji T, Yeh CT, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H *et al.* 2009. Maize inbreds exhibit high concentrations of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genetics* **5**: e1000734.

Starke S, Jost I, Rossbach O, Schneider T, Schreiner S, Hung LH, Bindereif A. 2015. Exon circularization requires canonical splice signals. *Cell Reports* **10**: 103–111.

Sundaresan V, Freeling M. 1987. An extrachromosomal form of the Mu transposons in maize. *Proceedings of the National Academy of Sciences, USA* **84**: 4924–4928.

Swanson-Wagner RA, Eichten SR, Kumari S, Tiffin P, Stein JC, Ware D, Springer NM. 2010. Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Research* **20**: 1689–1699.

Szabo L, Morey R, Palpant NJ, Wang PL, Afari N, Jiang C, Parast MM, Murry CE, Laurent LC, Salzman J. 2015. Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development. *Genome Biology* **16**: 126.

Tong H, Leasure CD, Hou X, Yuen G, Briggs W, He ZH. 2008. Role of root UV-B sensing in *Arabidopsis* early seedling development. *Proceedings of the National Academy of Sciences, USA* **105**: 21 039–21 044.

Vazquez F, Legrand S, Windels D. 2010. The biosynthetic pathways and biological scopes of plant small RNAs. *Trends in Plant Science* **15**: 337–345.

Wallace JG, Bradbury PJ, Zhang N, Gibon Y, Stitt M, Buckler ES. 2014. Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS Genetics* **10**: e1004845.

Wang PL, Bao Y, Yee MC, Barrett SP, Hogan GJ, Olsen MN, Dinneny JR, Brown PO, Salzman J. 2014. Circular RNA is expressed across the eukaryotic tree of life. *PLoS ONE* **9**: e90859.

Wang RL, Stec A, Hey J, Lukens L, Doebley J. 1999. The limits of selection during maize domestication. *Nature* **398**: 236–239.

Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, Lu Z, Olson A, Stein JC, Ware D. 2016. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nature Communications* **7**: 11 708.

Wang X, Wang H, Liu S, Ferjani A, Li J, Yan J, Yang X, Qin F. 2016. Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. *Nature Genetics* **48**: 1233–1241.

Wei L, Cao X. 2016. The effect of transposable elements on phenotypic variation: insights from plants to humans. *Science China Life Sciences* **59**: 24–37.

Wen W, Li D, Li X, Gao Y, Li W, Li H, Liu J, Liu H, Chen W, Luo J *et al.* 2014. Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nature Communications* **5**: 3438.

Westholm JO, Miura P, Olson S, Shenker S, Joseph B, Sanfilippo P, Celniker SE, Graveley BR, Lai EC. 2014. Genome-wide analysis of Drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Reports* **9**: 1966–1980.

Yang XH, Gao SB, Xu ST, Zhan ZX, Prasanna BM, Lin L, Li JS, Yan JB. 2011. Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Molecular Breeding* **28**: 511–526.

Yang Q, Li Z, Li W, Ku L, Wang C, Ye J, Li K, Yang N, Li Y, Zhong T *et al.* 2013. CACTA-like transposable element in ZmCCT attenuated photoperiod

sensitivity and accelerated the postdomestication spread of maize. *Proceedings of the National Academy of Sciences, USA* **110**: 16969–16974.

Yang N, Lu Y, Yang X, Huang J, Zhou Y, Ali F, Wen W, Liu J, Li J, Yan J. 2014. Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic fraits in an enlarged maize association panel. *PLoS Genetics* **10**: e1004573.

Ye CY, Chen L, Liu C, Zhu QH, Fan L. 2015. Widespread noncoding circular RNAs in plants. *New Phytologist* **208**: 88–95.

Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, Sandstrom R, Ma Z, Davis C, Pope BD *et al.* 2014. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**: 355–364.

Zaphiropoulos PG. 1996. Circular RNAs from transcripts of the rat cytochrome P450 2C24 gene: correlation with exon skipping. *Proceedings of the National Academy of Sciences, USA* **93**: 6536–6541.

Zhang XO, Dong R, Zhang Y, Zhang JL, Luo Z, Zhang J, Chen LL, Yang L. 2016. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Research* **26**: 1277–1287.

Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L. 2014. Complementary sequence-mediated exon circularization. *Cell* **159**: 134–147.

Zuo T, Zhang J, Lithio A, Dash S, Wise D, Weber D, Nettleton R. 2016. Genes and small RNA transcripts exhibit dosage-dependent expression pattern in maize copy-number alterations. *Genetics* **203**: 1133–1147.

Zuo W, Chao Q, Zhang N, Ye J, Tan G, Li B, Xing Y, Zhang B, Liu H, Fengler KA *et al.* 2014. A maize wall-associated kinase confers quantitative resistance to head smut. *Nature Genetics* **47**: 151–157.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Comparison and bench work validation of circRNAs identified through different tools.

Fig. S2 CircRNAs that were successfully amplified and sequenced for the validation.

Fig. S3 Expression–concentration relationships between circRNAs (RPM) and their parental genes (FPKM).

Fig. S4 Expression concentration variation of circRNAs across B73 different tissues.

Fig. S5 Schematic diagram of the *P*-value of enrichment test of different transposons in the flanking genomic regions of circRNAs.

Fig. S6 Fold change of significant repeat between circRNA and linear control.

Fig. S7 Enrichments of different transposons and their reverse complementary pairs in the significant enriched flanking regions of genes with or without circRNA.

Fig. S8 The number of circRNAs with similar reverse complementary pairs of other repetitive elements in the flanking regions.

Fig. S9 The number of reverse complementary pairs of LINE1-like elements (LLERCPs) is associated with the expression of circRNAs and linear transcripts.

Fig. S10 Gene ontology analysis of genes with detectable circRNAs.

Table S1 Datasets used in our study

Table S2 Detailed information on the primers used in our study

Table S3 High-confidence circRNAs

Table S4 Number of trait-associated genes identified by different studies

Table S5 Validation of circ1690 and its potential association with ear height variation in a diverse inbred panel

Methods S1 Detailed materials and methods.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.