New Phytologist

Research

Rapid report

Widespread noncoding circular RNAs in plants

Author for correspondence:
Longjiang Fan
Tel: +0086 571 88982730
Email: fanlj@zju.edu.cn

Received: *18 April 2015*
Accepted: *1 July 2015*

Chu-Yu Ye[1]*, Li Chen[2]*, Chen Liu[1], Qian-Hao Zhu[3] and Longjiang Fan[1,2]

[1]Institute of Crop Sciences, Zhejiang University, Hangzhou 310058, China; [2]Institute of Bioinformatics, Zhejiang University, Hangzhou 310058, China; [3]CSIRO Agriculture Flagship, Black Mountain Laboratories, Canberra ACT 2601, Australia

New Phytologist (2015)
doi: 10.1111/nph.13585

Key words: *Arabidopsis thaliana*, back splicing, circular RNA (circRNA), exonic circRNA, noncoding RNA, *Oryza sativa*.

Summary

- A large number of noncoding circular RNAs (circRNAs) with regulatory potency have been identified in animals, but little attention has been given to plant circRNAs.
- We performed genome-wide identification of circRNAs in *Oryza sativa* and *Arabidopsis thaliana* using publically available RNA-Seq data, analyzed and compared features of plant and animal circRNAs.
- circRNAs (12037 and 6012) were identified in *Oryza sativa* and *Arabidopsis thaliana*, respectively, with 56% (10/18) of the sampled rice exonic circRNAs validated experimentally. Parent genes of over 700 exonic circRNAs were orthologues between rice and *Arabidopsis*, suggesting conservation of circRNAs in plants. The introns flanking plant circRNAs were much longer than introns from linear genes, and possessed less repetitive elements and reverse complementary sequences than the flanking introns of animal circRNAs. Plant circRNAs showed diverse expression patterns, and 27 rice exonic circRNAs were found to be differentially expressed under phosphate-sufficient and -starvation conditions. A significantly positive correlation was observed for the expression profiles of some circRNAs and their parent genes.
- Our results demonstrated that circRNAs are widespread in plants, revealed the common and distinct features of circRNAs between plants and animals, and suggested that circRNAs could be a critical class of noncoding regulators in plants.

Introduction

In addition to messenger RNA (mRNA), cells contain diverse types of noncoding RNA such as microRNA (miRNA), small interfering RNA (siRNA), *trans*-acting siRNA (tasiRNA), and long noncoding RNA (lncRNA). Circular RNAs (circRNAs) are a distinct class of newly discovered endogenous noncoding RNAs (Salzman *et al.*, 2012; Memczak *et al.*, 2013; Wang *et al.*, 2014). Although observed for decades in eukaryotic cells, circRNAs have been previously perceived as splicing errors (Cocquerelle *et al.*, 1993). Due to the technological breakthroughs in high-throughput deep sequencing and functional genomics, recent works have not only uncovered large numbers of circRNAs that are abundantly and stably expressed in animal cells, but also demonstrated that circRNAs play an important role in a wide range of biological and developmental processes in animals (Salzman *et al.*, 2012, 2013; Jeck *et al.*, 2013; Memczak *et al.*, 2013; Guo *et al.*, 2014; Jeck & Sharpless, 2014). For example, nearly 100 000 circRNAs have been computationally predicted from the ENCODE transcriptome data derived from 15 cell lines (Gao *et al.*, 2015).

CircRNAs can arise from exons (exonic circRNA), introns (intronic circRNA), and intergenic region (Zhang *et al.*, 2013; Jeck & Sharpless, 2014). It was suggested that head-to-tail splicing (so-called back splicing), which is formed between the downstream genomic fragment (as a splicing donor site) and the upstream fragment (as a splicing acceptor site) (Fig. 1a), circularizes the precursor RNA with a poorly characterized mechanism (Ashwal-Fluss *et al.*, 2014; Suzuki & Tsukahara, 2014). It was reported that in humans, circularized exons are typically bracketed by unusually long introns that are highly enriched in ALU repeats (Jeck *et al.*, 2013). Competitive generation of circular and linear transcripts from the same pretranscript has been shown recently (Ashwal-Fluss *et al.*, 2014). Circularization of exon was mediated by complementary sequences in its flanking introns (X. O. Zhang *et al.*, 2014;

*These authors contributed equally to this work.

Wang & Wang, 2015). Using extensive mutagenesis of expression plasmids, it was showed that introns containing only the splice site sequences along with short inverted repeats (30–40 nt) are sufficient to allow the intervening exons to efficiently circularize (Liang & Wilusz, 2014). A more recent work showed that the presence of reverse complementary sequences in introns flanking circRNAs is a conserved feature of circRNA biogenesis in animals (Ivanov *et al.*, 2015). Furthermore, canonical splice signals were considered to be required for exon circularization (Starke *et al.*, 2015), and exon skipping was thought to be correlated with exon circularization (Kelly *et al.*, 2015). Another report showed that the RNA binding protein Quaking regulates formation of circRNAs (Conn *et al.*, 2015).

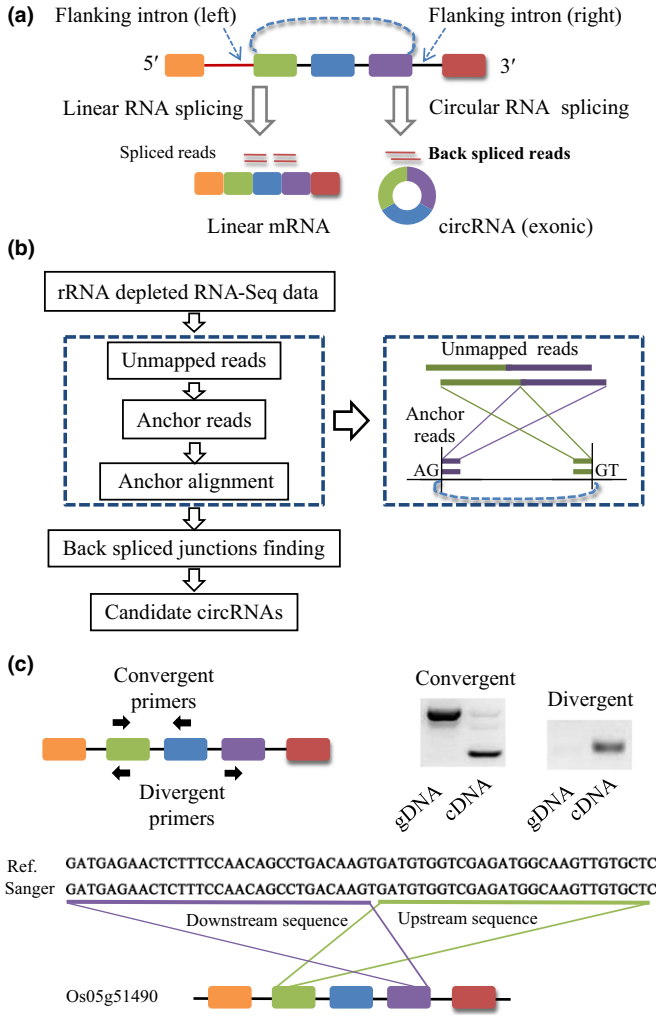

Fig. 1 Identification of circular RNAs (circRNAs) in plants. (a) A model showing linear and back splicing for generating linear mRNA and circRNA, respectively. (b) The pipeline for computational identification of circRNAs in plants following the method used in animals (Memczak *et al.*, 2013). (c) Validation of circRNAs by PCR with divergent primers. Upper left panel, a model showing convergent and divergent primers for amplification of linear RNA and circRNA, respectively. Upper right panel, an example showing that a set of divergent primers amplified circRNA in cDNA but not in genomic DNA; a set of convergent primers was used as a control. Lower panel, a Sanger sequencing example showing that Os_ciR182 was derived through exon back splicing from Os05g51490.

CircRNAs often show tissue/cell-type/developmental-stage specific expression patterns (Memczak *et al.*, 2013; Salzman *et al.*, 2013; Gao *et al.*, 2015). Although functions of circRNAs are still largely unknown, their high abundance, evolutionary conservation and derivation from important gene loci suggest that they play important roles in cells (Vicens & Westhof, 2014; Y. Zhang *et al.*, 2014; Gao *et al.*, 2015). For example, circRNAs were indicated to play regulatory roles in gene expression (Burd *et al.*, 2010; Y. Zhang *et al.*, 2014). It was found that the ratio of circular to linear RNA isoforms was always lower in tumor compared with normal colon samples and correlated negatively with the proliferation index (Bachmayr-Heyda *et al.*, 2015). A recent work found a regulatory role of circRNAs in transcriptional control via specific RNA–RNA interaction (Li *et al.*, 2015). Additionally, two studies revealed that circRNAs can act as miRNA sponges to sequester miRNAs (Hansen *et al.*, 2013; Memczak *et al.*, 2013).

It has been realized that circRNAs are a large class of noncoding RNAs with regulatory potency in animals. A recent study has further shown that circRNA is expressed in a wide range of eukaryotic species, including the model plant species *Arabidopsis thaliana* (Wang *et al.*, 2014), but little is known about the features of plant circRNAs. To examine and explore the scope of plant circRNAs, we first did genome-wide identification of circRNAs in *Oryza sativa* and *A. thaliana*, and found thousands of circRNAs in both species, suggesting that, as in animals, circRNAs may be also a class of noncoding regulators involved in a diverse of biological pathways in plants. We then investigated the properties of plant circRNAs, and revealed features similar to or distinct from that of animal circRNAs.

Materials and Methods

Computational identification of circular RNAs

Oryza sativa L. (Nipponbare) and *Arabidopsis thaliana* (L.) Heynh. (Col-0) genome sequences were downloaded from RGAP (v7, http://rice.plantbiology.msu.edu/) and TAIR (v10, https://www.arabidopsis.org/), respectively. Two sets of publicly available ribosomal RNA (rRNA) depleted RNA-Seq (RibominusSeq) data (Supporting Information Table S1) generated using *O. sativa* roots (GenBank accession PRJNA215013; read length 101 bp) and *A. thaliana* leaves (PRJNA218215; read length 100 bp) were used for genome-wide identification of circRNAs according to the method used in animals (Memczak *et al.*, 2013; Secco *et al.*, 2013). Briefly, RibominusSeq reads were first mapped to reference genomes using Bowtie2 (v2.0.5) (Langmead & Salzberg, 2012) with the same parameters used by Memczak *et al.* (2013). The reads that could be mapped to the genomes were discarded. For the remaining unmapped reads, we extracted 20-nt anchors from both ends and aligned them independently to the reference genomes to find unique anchor positions. The aligned anchors with a reversed orientation indicated circRNA splicing (Fig. 1b). The anchor alignments were then extended such that the complete read aligns and the breakpoints were flanked by GU/AG splice sites. A candidate circRNA was called if it was supported by at least two unique back spliced reads.

Validation of circular RNAs

To validate the circRNAs identified in rice, we isolated total RNA from rice plants grown under the same conditions as previously reported (Secco *et al.*, 2013) to generate the transcriptome data (PRJNA215013) used in this study. Briefly, *Oryza sativa* (cv Nipponbare) seedlings were first grown hydroponically for 2 wk (30/22°C of day/night temperature and a 12-h photoperiod regime) on the phosphate (Pi)-sufficient medium (0.32 mM Pi), and then a portion of the plants were transferred to Pi-deficient solution (0 mM Pi) and grown for 1 and 6 h (Secco *et al.*, 2013). Total RNAs were extracted from roots of the rice seedlings under Pi-sufficient (control) and Pi-deficient treatments using RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions. First-strand cDNA was synthesized from *c.* 1 μg of total RNA with random hexamer using the First Strand cDNA Synthesis kit (TaKaRa, Dalian, China). Genomic DNA of rice was extracted using the conventional cetyltrimethylammonium bromide (CTAB) method. Polymerase chain reactions (PCRs) were performed using a set of divergent primers (Table S2) and a set of convergent primers that were used as a control. PCR conditions were as follows: an initial 3 min step at 94°C followed by 35 cycles of 45 s denaturing at 94°C, 35 s annealing at the appropriate annealing temperature (depending on the primer set used) and 30 s extension at 72°C. The final step was conducted at 72°C for 10 min. PCR products were separated using agarose gel and bands were individually excised, purified and directly Sanger-sequenced.

Sequence analysis

Reciprocal BLAST search (BlastP in BLAST+, v2.2.27, $E < 1e\text{-}10$) was first performed for all annotated *Oryza sativa* and *Arabidopsis thaliana* genes. Those genes that were best hits to each other were defined as orthologous genes. Gene ontology (GO) enrichment analysis was performed using agriGO (http://bioinfo.cau.edu.cn/agriGO/) with Fisher's exact test (Du *et al.*, 2010) and the whole set of the annotated genes as the reference. The repetitive sequences of *O. sativa* and *A. thaliana* were downloaded from RGAP (ftp://ftp.plantbiology.msu.edu/pub/data/TIGR_Plant_Repeats/) and TAIR (ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR10_genome_release/TAIR10_transposable_elements/), respectively. The intronic sequences from *O. sativa* and *A. thaliana* were used to Blast (BlastN in BLAST+, v2.2.27, $E < 1e\text{-}5$) against the repetitive sequences of *O. sativa* and *A. thaliana*, respectively. For identification of reverse complementary sequence, the sequences of two introns flanking a circRNA were used to Blast (BlastN in BLAST+, v2.2.27, -strand minus, -word_size 5, $E < 1e\text{-}5$) against each other. MEME was used to identify motifs with a width ranging from 5 to 60 nt among the introns flanking circRNAs (Bailey *et al.*, 2009).

Expression analysis

Mapping of the transcriptome reads was performed by TopHat v2.0.9 (Trapnell *et al.*, 2009) with the default settings. Quantification of gene expression was done using the FPKM algorithm. The expression levels of circRNAs were determined by the number of back spliced reads normalized by the total number of sequence reads in each RNA-Seq data set (Gao *et al.*, 2015). Coexpression of circRNAs and their corresponding parent genes was determined by R^2 value of Pearson correlation coefficient.

Results

Identification of circular RNAs in plants

CircRNAs are recognizable based on back spliced reads in rRNA depleted RNA-Seq (RibominusSeq) data (Fig. 1a). Back spliced reads refer to the RNA-Seq reads that contain splice junction formed by joining a splice donor to an upstream splice acceptor. These reads usually cannot be mapped onto reference genome (Fig. 1b; Table S1). We adopted the method used in animals (Memczak *et al.*, 2013) to identify circRNAs in plants (Fig. 1b, and see the Materials and Methods section). Two sets of publicly available RibominusSeq data from *Oryza sativa* and *Arabidopsis thaliana* (Table S1) were collected for genome-wide identification of circRNAs. From these two datasets, our mapping exercise identified 1355 322 273 (39.1% of the total reads) and 96 864 928 (20.3%) unmapped reads in *O. sativa* and *A. thaliana*, respectively. Using these unmapped RibominusSeq reads and the in-house developed bioinformatic pipeline, we found 12 037 and 6012 circRNAs in *O. sativa*, and *A. thaliana*, respectively (Tables 1, S3, S4).

Among the identified circRNAs, 6074 (50.5%) of the *Oryza sativa* ones and 5152 (85.7%) of the *Arabidopsis thaliana* ones were generated from exons of a single protein-coding gene, that is, exonic circRNAs (Tables 1, S3, S4). Among these exonic circRNAs, the majority of them possessed only a few (1–3) parent-gene-derived-exons in both species. By contrast, most linear protein coding genes harbors ≥7 exons (Fig. S1). Of the remaining circRNAs, besides those generated from introns (intronic circRNAs), untranslated regions (UTR circRNAs) and intergenic region (intergenic circRNAs), some were found to contain back spliced reads that aligned to two or more different genes, that is, generated from trans-backsplicing. These were classified as other circRNAs in this study (Table 1). A relatively higher rate (34.4%) of such other circRNAs was found in *O. sativa* than in *A. thaliana*

Table 1 Genome-wide identification of circular RNAs (circRNAs) in plants based on RibominusSeq data

	Oryza sativa		*Arabidopsis thaliana*	
Type of circRNA	Number	Percentage	Number	Percentage
Exonic circRNA	6074	50.5	5152	85.7
Intronic circRNA	485	4.0	1	0
UTR circRNA	636	5.3	19	0.3
Intergenic circRNA	705	5.9	32	0.5
Other circRNA[a]	4137	34.4	808	13.4
Total	12 037	100	6012	100

UTR, untranslated region.
[a]Other circRNAs refer to those derived from two or more genes.

(13.4%). These results suggested that circRNAs in plants were generated from diverse genomic regions, but mainly from coding regions, and that circRNAs exist in both monocot and dicot plant species. Additionally, we found that some parent genes produced more than one circRNA although most genes produced only one circRNA in both species (Fig. S2). These circRNAs were generated by alternative back splicing or from different exons/introns in a single gene.

To confirm our identification of circRNAs, 18 exonic circRNAs randomly selected from the top 100 highly expressed rice ones were used in experimental validation using reverse transcription (RT)-PCR. A set of divergent primers (Table S2), that is, the forward primer being located downstream of the reverse primer when they are aligned to genomic sequence (Fig. 1c), were designed for each circRNA and used to amplify both cDNA and genomic DNA. The expected results would be positive and negative amplification in cDNA and genomic DNA, respectively. As a control, convergent primers that should amplify the linear mRNAs were also designed for each circRNA used in verification. The amplified PCR products using divergent primers were sequenced to confirm the presence of the back spliced junctions, As a result, 10 of the 18 circRNAs were validated (Table S2; see an example in Fig. 1c). Considering the spatial and temporal specific expression patterns of circRNAs, it is reasonable for the 56% (10/18) verification rate. Additionally, we also randomly selected 10 highly expressed circRNAs from each of the other categories (intergenic, intronic and other categories) for validation; however, none of these circRNAs was successfully validated (Table S2). Therefore, our further analyses were focused on the exonic circRNAs.

Conservation of circular RNAs in plants

According to reciprocal BLAST search analysis, 10 768 gene pairs were identified as orthologs in *Oryza sativa* and *Arabidopsis thaliana*. Among the 6074 and 5152 parent genes generating exonic circRNAs in *O. sativa* and *A. thaliana*, respectively, over 700 gene pairs are orthologs (Table S5; see e.g. in Fig. 2a), accounting for 12.2% and 14.5% of rice and *Arabidopsis* genes generating exonic circRNAs, respectively. Considering the data we used, which were from different tissues (root and leaf of rice and *Arabidopsis*, respectively) and stress treatments (Pi-starvation and high-light for rice and *Arabidopsis*, respectively), these ratios were quite high. Furthermore, over 300 orthologous parent genes generated circRNAs from a similar position (e.g. shown in Fig. 3a). These results demonstrated the conservation feature of plant circRNAs. We further examined conservation of the flanking introns of these conserved circRNAs; however, we did not find any sequence similarity (BlastN, word_size 5, $E < 1e\text{-}5$). Additionally, no shared motif was found in these flanking intronic sequences between *O. sativa* and *A. thaliana*. GO enrichment analysis for the orthologous parent genes generating circRNAs in *O. sativa* and *A. thaliana* showed that they were involved in diverse biological processes such as developmental process and response to abiotic stimulus (FDR < 0.05, Table S6), suggesting that the circRNAs generated from these orthologous genes might also play a diverse role in plants.

Sequence structure of introns flanking circular RNAs in plants

We compared the length of introns flanking exonic circRNAs (Fig. 1a) with the intron length of linear genes, and found that circRNAs had significantly longer flanking introns ($P < 2.2e\text{-}16$, Wilcoxon rank sum test) in both *Oryza sativa* and *Arabidopsis thaliana* (Fig. 2b). This is consistent with the result discovered in animals (Jeck *et al.*, 2013).

Repetitive elements (such as ALU repeats) or reverse complementary sequences are enriched in introns bracketing circRNAs in animals, and 30–40 nt short inverted repeats are sufficient for circularization (Jeck *et al.*, 2013; Ivanov *et al.*, 2015). Through blast (BlastN, $E < 1e\text{-}5$) against the annotated repetitive sequences, we found that only a very low proportion of intronic sequences bracketing circRNAs could be aligned to repetitive sequences in both *Oryza sativa* (6.2%) and *Arabidopsis thaliana* (0.3%). These ratios were significantly lower ($P < 1e\text{-}6$, Exact Binomial test) than those observed in introns from linear genes (13.2% and 2.3% for *O. sativa* and *A. thaliana*, respectively; Fig. 2c). Among the 6074 rice and 5152 *Arabidopsis* exonic circRNAs, we found only 33 and zero circRNAs possessing sequence similarity (BlastN, word_size 5, $E < 1e\text{-}5$) in their flanking introns. To further confirm the BLAST results, we manually investigated reverse complementary sequences (perfect match) in their flanking introns (Table S7) and found that reverse complementary sequences with a length of > 15 nt only exited in 46 (1.12%) rice and in only one (0.02%) *Arabidopsis* circRNAs. In the 13 validated plant circRNAs (ten rice ones verified in this study and three *Arabidopsis* ones validated by Wang *et al.* (2014)), we found only two circRNAs (Os_ciR158 and Os_ciR133) containing > 15 nt reverse complementary repetitive sequences in their flanking introns. This ratio (2 : 13 or 15.4%) was still much lower than that reported in animals (e.g. 8 : 15% or 53.3%; Li *et al.*, 2015). Taken together, our results showed limited repetitive and reverse complementary sequences in intronic sequences flanking exonic circRNAs in plants compared with that in animals.

Diverse expression patterns of exonic circular RNAs in plants

Based on the number of back spliced reads identified in the datasets used in this study and the CIRI approach (Gao *et al.*, 2015), we compared the expression profiles of rice exonic circRNAs under Pi-sufficient and Pi-starvation stress (18 samples) and *Arabidopsis* exonic circRNAs under low- and high-light stresses (six samples), and found that exonic circRNAs often expressed specifically in different developmental stages and/or stress treatments (Tables S8, S9). In total, only 21 rice and 25 *Arabidopsis* exonic circRNAs were commonly expressed in all the 18 rice and six *Arabidopsis* samples, respectively. Many exonic circRNAs (4248 rice and 1919 *Arabidopsis* ones, accounting for 69.9% and 37.2%, respectively) were only expressed in a single sample. Some exonic circRNAs showed time point specific expression pattern, for example, Os_ciR1837 only expressed at 21 d and Os_ciR1569 at 6 h under both Pi-sufficient and Pi-starvation conditions. In *Arabidopsis*, many circRNAs were uniquely expressed in the *tnr-1* mutant, particularly after 2 h of high-light treatment. Diverse expression

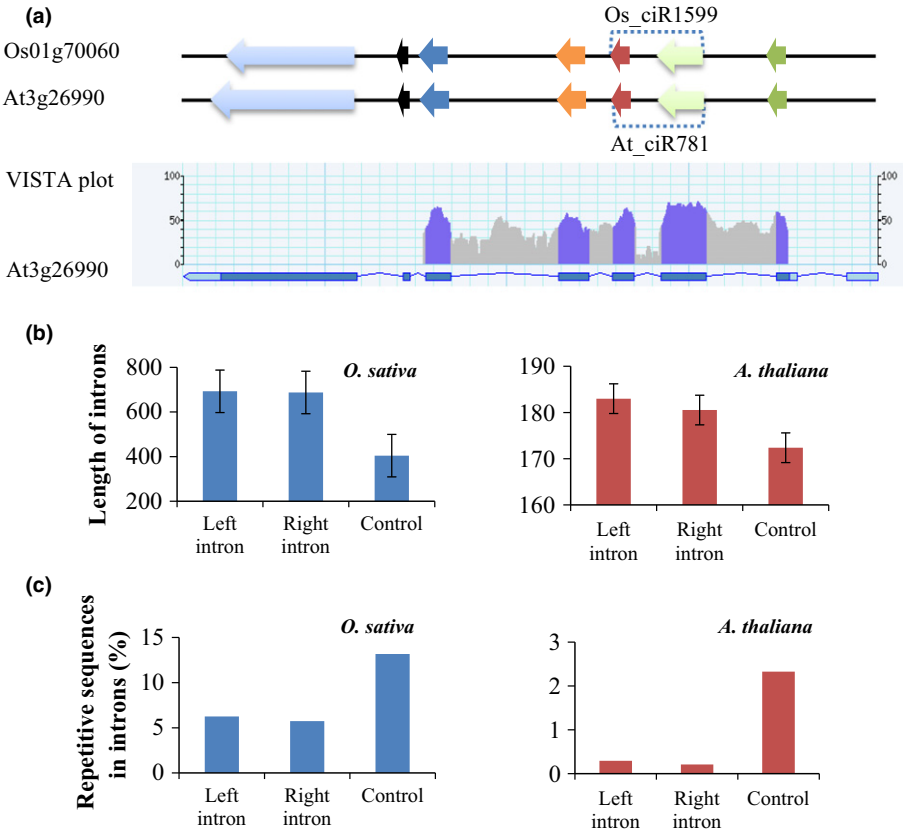

Fig. 2 Features of circular RNAs (circRNAs) in plants. (a) An example showing conservation of exonic circRNAs. Upper panel, circRNAs are generated from the same region of orthologous gene pairs in *Oryza sativa* and *Arabidopsis thaliana*. Lower panel, VISTA plot of Os01g70060 with At3g26990 as the reference. The vertical bar shows sequence identity of the orthologous gene pair. Exons and introns were indicated by blue and gray, respectively. (b) CircRNAs have significantly long flanking introns in *O. sativa* and *A. thaliana*. Error bars indicate ± SD. (c) A lower percentage of introns bracketing circRNAs belong to repetitive sequences in *O. sativa* and *A. thaliana*.

patterns of 100 circRNAs randomly selected from the top 200 highly expressed circRNAs in *O. sativa* and *A. thaliana* were showed in Figs 3(a) and Fig. S3, respectively.

Based on the average expression levels under Pi-sufficient (nine samples) and Pi-starvation (nine samples) conditions, we identified 27 rice exonic circRNAs differentially expressed ($P < 0.05$, *t*-test; Fig. 3b) under these two conditions, with six and 21 showing significant upregulation and downregulation after Pi-starvation, respectively (Table S10). We further examined the expression profiles of the parent genes of these 27 Pi-starvation responsive circRNAs and found that 11 of them were also differentially expressed. Interestingly, 10 of the 11 differentially expressed parent genes shared a common expression trend with their corresponding circRNAs (eight of them shown in Fig. 3b), suggesting coregulation of circRNAs and their parent genes during Pi-starvation stress.

CircRNAs have been reported to regulate expression of their parent genes in *cis* in humans (Li *et al.*, 2015). To further investigate the expression relationship between the 6074 exonic circRNAs and their parent genes in *Oryza sativa*, we calculated the Pearson correlation coefficient R^2 values based on their expression levels in 18 samples, and found a significantly positive correlation in 349 pairs (Table S11). We did not find any negatively correlated pair.

Discussion

Thousands of circRNAs have been identified in humans, mouse and other animals, and the numbers are likely to grow further. The large number and unique features of circRNAs make them an important class of noncoding regulators that are worthy of further research (Glazar *et al.*, 2014). In plants, circRNAs identification has been performed in *Arabidopsis thaliana*, and three were verified (Wang *et al.*, 2014), demonstrating the presence of circRNAs in plants, but genome-wide and systematic identification of plant circRNAs has not been reported so far and their features are not clear. In this study we conducted a genome-wide identification of circRNAs in *A. thaliana* and *Oryza sativa*, the model species of dicot and monocot plants, respectively, and compared the features of plant circRNAs with those of animal ones. Our results demonstrate the presence of a large number of circRNAs in plants, provide the resources of plant circRNAs for further functional characterization, and first reveal the common and distinct properties of plant and animal circRNAs.

Several reports have revealed regulatory functions of circRNAs in animals. We found that most plant circRNAs showed developmental-/stress-specific expression patterns (Figs 3a, S3), similar to that that has been reported in animals (Gao *et al.*, 2015). For example, 27 differentially expressed rice exonic circRNAs under Pi-sufficient and Pi-starvation conditions suggested a potential role of these circRNAs in responsive to Pi-starvation stress in rice (Fig. 3b); circRNAs specifically expressed at a specific developmental time-point could have their roles related to certain development events at that time-point (Tables S8, S9). These results suggested that circRNAs are also important functional regulators involved in developmental-/stress-specific biological processes in plants. In

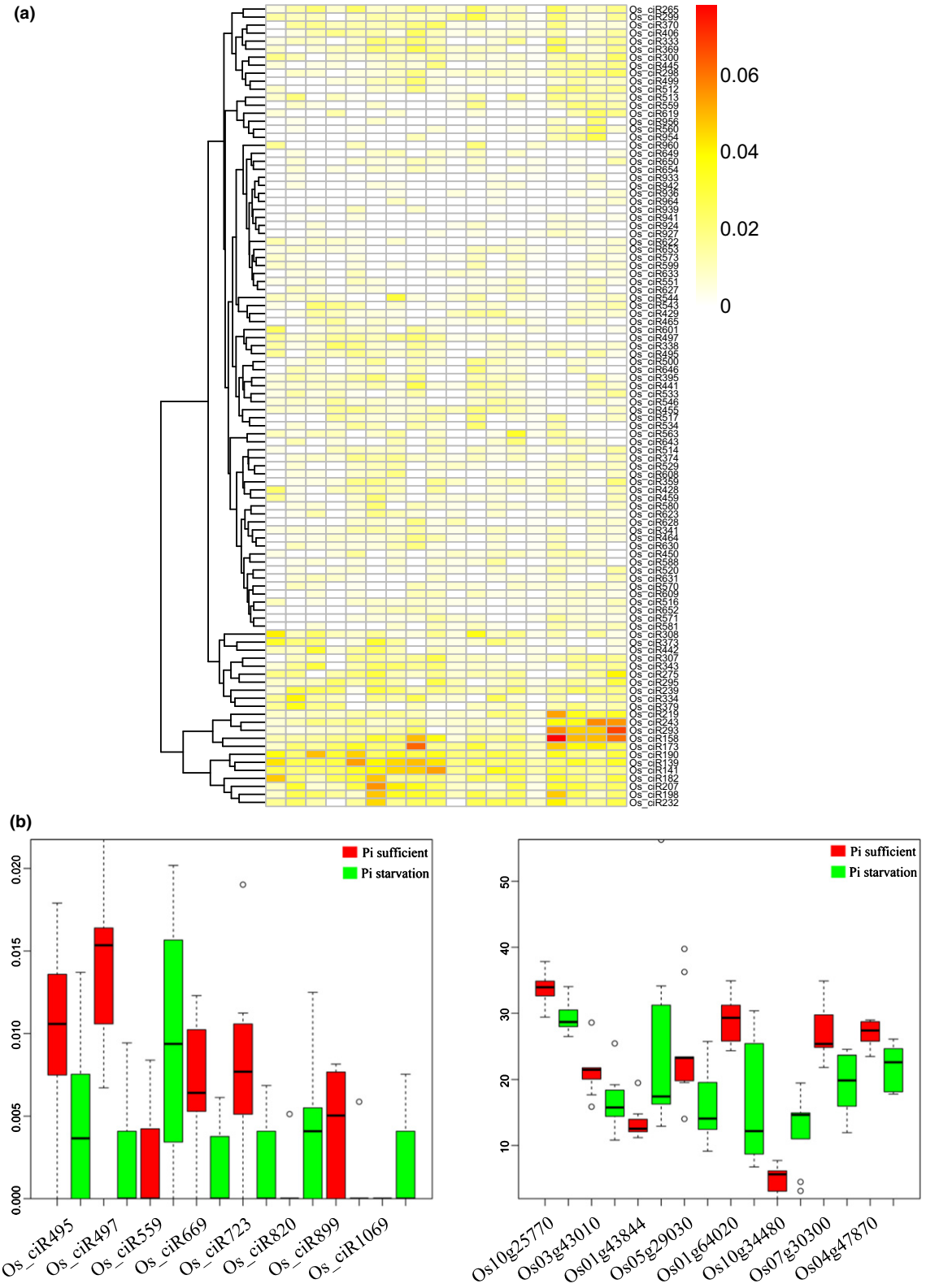

Fig. 3 Expression patterns of circular RNAs (circRNAs) in *Oryza sativa* and *Arabidopsis thaliana*. (a) Heat maps showing the expression patterns of 100 circRNAs (horizontal rows) randomly selected from the top 200 highly expressed circRNAs among different treatments (vertical columns) in *O. sativa*. Columns 1–9 (from left to right) represent Pi-sufficient treatment for 1 h, 6 h, 24 h, 3 d, 7 d, 21 d, 21 d + 1 h, 21 d + 6 h and 21 d + 24 h, and columns 10–18 represent Pi-starvation treatment for 1 h, 6 h, 24 h, 3 d, 7 d, 21 d, 21 d + 1 h, 21 d + 6 h and 21 d + 24 h. (b) Expression changes of eight differentially expressed circRNAs (left panel) and their parent genes (right panel) in *O. sativa* roots under Pi-sufficient and Pi-starvation conditions. The relative expression values of circRNA were calculated based on the counts of back spliced reads for each circRNA normalized by the total number of sequence reads in each RNA-Seq data set (Gao *et al.*, 2015), while the expression levels of parent genes were determined by fragment per kilobase of exon per million fragments mapped (Secco *et al.*, 2013). Error bars indicate ± SD.

contrast to the findings in animals, where little or no correlation was found for the expression levels of circRNAs and their parent genes (Salzman *et al.*, 2013), our results showed that some exonic circRNAs and their parent genes had their expression levels significantly positively correlated (Table S11). Whether or not this was a result of *cis*-transcriptional promotion of circRNAs on their parent genes as reported in animal (Li *et al.*, 2015) requires further investigation; nevertheless, this result suggested a role of plant circRNAs in regulation of their parent genes or *vice versa*. It has been shown in animals that circRNAs can function as miRNA sponge (Hansen *et al.*, 2013; Memczak *et al.*, 2013). Using the rules for target mimics identification in plants (Wu *et al.*, 2013; Ye *et al.*, 2014), we found that a small proportion of circRNAs (6.6% and 5.0% in rice and *Arabidopsis*, respectively) were potential target mimics of miRNAs because the genomic regions producing circRNAs contain miRNA binding site (data not shown). However, it is hard to distinguish whether the target mimic site belongs to the linear or circRNA transcripts until having the full length of cDNAs of both the circRNAs and their parent linear mRNAs. Therefore, the function of circRNAs as target mimics remains tentative and needs further experimental validation.

Several mechanisms have been proposed to address the question of circRNA biogenesis in animals. It has been shown that circRNA-forming exons are often bracketed by unusually long introns, in which splicing is thought to be less efficient (Ashwal-Fluss *et al.*, 2014). We found that, as in animals, the flanking introns of circRNAs were much longer than introns of linear genes in plants (Fig. 2b). Two recent works reported that introns bracketing circRNAs are highly enriched for ALU repeats and reverse complementary sequences, and that short inverted repeats (30–40 nt) in flanking introns are sufficient to allow exons to efficiently circularize in animals (Jeck *et al.*, 2013; Ivanov *et al.*, 2015). We found a significantly lower proportion of both repetitive and reverse complementary sequences in introns flanking exonic circRNAs (Fig. 2c). This suggested that formation of only a few plant circRNAs could be related to repetitive and reverse complementary sequences present in flanking introns and that biogenesis of plant circRNAs could use mechanism(s) different from those proposed for animals. In addition, we did not find any sequence conservation in the flanking introns for the circRNAs conserved between *Oryza sativa* and *Arabidopsis thaliana*. This might suggest that exonic sequences themselves could be also important for circRNA biogenesis besides their flanking introns.

In addition to exonic circRNAs, we also predicted a considerable number of circRNAs from noncoding regions (UTR, introns and intergenic regions) and those covering two or more genes, however, none of the 30 sampled circRNAs of these types was successfully validated (Table S2). The difficulty in validation may be caused by their low and/or temporal specific expression patterns. Compared with the verified exonic circRNAs, only a few of these types of circRNAs were validated although a large number of them were also predicted in animals (Gao *et al.*, 2015). Additionally, it is possible that differences in genomic features, such as repetitive sequence, between plants and animals make the prediction algorithm that was originally developed for application in animals less sensitive when employed in plants or cause incorrect predictions for plant circRNAs. Alternatively, the filtration strategy used by this algorithm may be insufficient for removal of false positives. Whatever the possible reason, it is necessary to develop algorithm specific for prediction of plant circRNAs according to the features of plant genome and transcriptome.

In summary, genome-wide identification of circRNAs and analysis of their conserved features suggested that circRNAs are present in both monocot and dicot plants and are a critical class of noncoding regulators in plants. Our results revealed the common as well as distinct features of plant and animal circRNAs.

Acknowledgements

This work was supported by the National Basic Research Program of China (grant no. 2011CB109306/2015CB150200), the National Science Foundation of China (grant no. 91435111) and Shanghai Municipal Commission of Agriculture (grant no. 2014-7-1-4). The authors also thank Jiangsu Collaborative Innovation Center for Modern Crop Production (JCIC-MCP).

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.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* **37**: W202–W208.

Burd CE, Jeck WR, Liu Y, Sanoff HK, Wang Z, Sharpless NE. 2010. Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genetics* **6**: e1001233.

Cocquerelle C, Mascrez B, Hetuin D, Bailleul B. 1993. Mis-splicing yields circular RNA molecules. *FASEB Journal* **7**: 155–160.

Conn SJ, Pillman KA, Toubia J, Conn VM, Salmanidis M, Phillips CA, Roslan S, Schreiber AW, Gregory PA, Goodall GJ. 2015. The RNA binding protein Quaking regulates formation of circRNAs. *Cell* **160**: 1125–1134.

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

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

Glazar P, Papavasileiou P, Rajewsky N. 2014. circBase: a database for circular RNAs. *RNA* **20**: 1666–1670.

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.

Ivanov A, Memczak S, Wyler E, Torti F, Porath HT, Orejuela MR, Piechotta M, Levanon EY, Landthaler M, Dieterich C *et al.* 2015. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Reports* **10**: 170–177.

Jeck WR, Sharpless NE. 2014. Detecting and characterizing circular RNAs. *Nature Biotechnology* **32**: 453–461.

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.

Kelly S, Greenman C, Cook PR, Papantonis A. 2015. Exon skipping is correlated with exon circularization. *Journal of Molecular Biology*. doi: 10.1016/j.jmb.2015.1002.1018.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.
Li Z, Huang C, Bao C, Chen L, Lin M, Wang X, Zhong G, Yu B, Hu W, Dai L *et al.* 2015. Exon–intron circular RNAs regulate transcription in the nucleus. *Nature Structural & Molecular Biology* **22**: 256–264.
Liang D, Wilusz JE. 2014. Short intronic repeat sequences facilitate circular RNA production. *Genes & Development* **28**: 2233–2247.
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.
Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. 2013. Cell-type specific features of circular RNA expression. *PLoS Genetics* **9**: e1003777.
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.
Secco D, Jabnoune M, Walker H, Shou HX, Wu P, Poirier Y, Whelan J. 2013. Spatio-temporal transcript profiling of rice roots and shoots in response to phosphate starvation and recovery. *Plant Cell* **25**: 4285–4304.
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.
Suzuki H, Tsukahara T. 2014. A view of pre-mRNA splicing from RNase R resistant RNAs. *International Journal of Molecular Sciences* **15**: 9331–9342.
Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105–1111.
Vicens Q, Westhof E. 2014. Biogenesis of circular RNAs. *Cell* **159**: 13–14.
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 Y, Wang Z. 2015. Efficient backsplicing produces translatable circular mRNAs. *RNA* **21**: 172–179.
Wu HJ, Wang ZM, Wang M, Wang XJ. 2013. Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants. *Plant Physiology* **161**: 1875–1884.
Ye CY, Xu H, Shen EH, Liu Y, Wang Y, Shen YF, Qiu J, Zhu QH, Fan LJ. 2014. Genome-wide identification of non-coding RNAs interacted with microRNAs in soybean. *Frontiers in Plant Science* **5**: 743.
Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L. 2014. Complementary sequence-mediated exon circularization. *Cell* **159**: 134–147.
Zhang Y, Yang L, Chen LL. 2014. Life without a tail: new formats of long noncoding RNAs. *International Journal of Biochemistry & Cell Biology* **54**: 338–349.
Zhang Y, Zhang XO, Chen T, Xiang JF, Yin QF, Xing YH, Zhu SS, Yang L, Chen LL. 2013. Circular intronic long noncoding RNAs. *Molecular Cell* **51**: 792–806.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 The distribution of the number of circRNAs per gene.

Fig. S2 The distribution of the number of exons per circRNA.

Fig. S3 Heat maps showing the expression patterns of 100 circRNAs randomly selected from the top 200 highly expressed circRNAs among different treatments in *Arabidopsis thaliana*.

Table S1 RibominusSeq data used in this study

Table S2 Divergent primers for validation of candidate circRNAs

Table S3 Genome-wide identification of circRNAs in *Oryza sativa* based on RibominusSeq data

Table S4 Genome-wide identification of circRNAs in *Arabidopsis thaliana* based on RibominusSeq data

Table S5 Orthologous genes generating circRNAs in *Oryza sativa* and *Arabidopsis thaliana*

Table S6 Gene ontology enrichment analysis for orthologous genes generating circRNAs in *Oryza sativa* and *Arabidopsis thaliana*

Table S7 Short reverse complementary sequences in the inrons flanking circRNAs in *Oryza sativa* and *Arabidopsis thaliana*

Table S8 The expression profiles of exonic circRNAs in *Oryza sativa* under Pi-sufficient and Pi-starvation stress

Table S9 The expression profiles of exonic circRNAs in *Arabidopsis thaliana* under low- and high-light stress

Table S10 Differentially expressed circRNAs in *Oryza sativa* with and without Pi-starvation treatments

Table S11 Significantly positive correlation for expression profiles between exonic circRNAs and their parent genes in *Oryza sativa*

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.