INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
17–0972/2018/20–5–1201–1210
DOI: 10.17957/IJAB/15.0574
http://www.fspublishers.org

Full Length Article

Identification of 12 lncRNAs as Potential miRNA Targets in Roots of *Rehmannia glutinosa* Responding to Replanting Disease

Yanhui Yang[1*], Mingjie Li[2], Yanjie Yi[1], Ruifang Li[1] and Chen Dong[1]
[1]*College of Bioengineering, Henan University of Technology, Lianhua Street, Zhengzhou High-technology Zero, Henan Province, China, 450001*
[2]*College of Crop Sciences, Fujian Agriculture and Forestry University, Jinshan Road, Cangshan District, Fuzhou, China, 350002*
[*]For correspondence: yyhui2004@126.com

Abstract

The productivity of the medicinal plant *Rehmannia glutinosa* is severely compromised by the so-called replanting disease, the molecular basis of which remains unclear. Long non-coding RNAs (lncRNAs) are known to regulate various post-transcriptional modifications in a wide range of eukaryotic genes, and the possibility that they represent a target for miRNA activity has recently been proposed. Here, two poly(A)$^-$ RNA libraries were generated, one based on RNA extracted from *R. glutinosa* roots harvested from a new planting (R1) and therefore free of replanting disease, and the other from a second crop (R2) affected by the disease. An *in silico* analysis of the resulting ~89,000 non-redundant transcripts detected 155 lncRNA candidates as miRNA targets. Based on the ceRNA hypothesis, a set of potential lncRNA candidate/mRNA modules was established, involving an interaction with the same miRNAs. Altogether, 32 of these lncRNA candidates differed in abundance between the R1 and R2 libraries, and a quantitative PCR-based assay verified differential transcription for 28 of them. Here, 12 lncRNAs were validated using RACE PCR. The abundance of the 12 lncRNA/mRNA modules was correlated with the presence of replanting disease. The indication was that the lncRNA candidates under continuous cultivation pressure, via their competition for specific mRNAs, could activate the specific stress signaling pathways, repress core mechanisms, and result in forming the disease in *R. glutinosa*. Our working hypothesis holds that a number of lncRNA candidates potentially as miRNA targets contribute either directly or indirectly to the disease phenomenon. © 2018 Friends Science Publishers

Keywords: Poly(A)$^-$ RNA; Replanting disease; ceRNA hypothesis; lncRNAs; Expression profile

Introduction

Eukaryotic cells produce a large number of non-coding RNA molecules (ncRNAs), which include transcribed microRNAs (miRNAs) and long ncRNAs (lncRNAs) (Wu *et al.*, 2013; Kwenda *et al.*, 2016). Some of these are known to have a regulatory role (Kumar, 2014; Joshi *et al.*, 2016). Especially, long noncoding RNAs (lncRNAs) are being rigorously studied for their identification and function (Tang *et al.*, 2016; Song and Zhang, 2017). The length of the lncRNAs must be longer than 200 nt, equivalent to about ten folds that of the miRNAs; however, they still lack both a coding sequence (CDS) and an open reading frame (ORF) (Ponjavic *et al.*, 2007; Ponting *et al.*, 2009). Some lncRNAs are polyadenylated (poly(A)$^+$), while others (at least 24%) are not (Cui *et al.*, 2010; Lee, 2012; Livyatan *et al.*, 2013). The involvement of lncRNAs in regulating various biological processes including plant development and stress responses has been demonstrated in a number of plant species (Zhu *et al.*, 2014; Liu *et al.*, 2015; Lu *et al.*, 2016). Moreover, many non-polyadenylated (poly(A)$^-$) lncRNAs appear to be selectively activated by environmental conditions, and are typically associated with a higher stress specificity than the poly(A)$^+$ lncRNAs (Ding *et al.*, 2013; Di *et al.*, 2014). Recently, the possibility that lncRNAs and mRNAs interact with the same miRNA has been demonstrated (Salmena *et al.*, 2011; Fan *et al.*, 2014). The proposed mechanistic basis of this interaction, so-called "competing endogenous RNA (ceRNA) hypothesis" is that they act as competitors for a common miRNA (Franco-Zorrilla *et al.*, 2007; Rubio-Somoza *et al.*, 2011; Salmena *et al.*, 2011; Song *et al.*, 2017). The nature of particular lncRNA/miRNA/mRNA networks has been used to predict the function of some lncRNAs (Li *et al.*, 2016; Su *et al.*, 2016).

The tuberous roots of *Rehmannia glutinosa* (*Orobanchaceae* family) are of commercial value as they harbor a number of medicinally active compounds (Wen *et al.*, 2002). Attempts to cultivate the species as a crop have been hindered by the so-called "replanting disease", a syndrome in which the plants develop fibrous rather than tuberous roots; the disease appears when the crop is planted into soil which has previously supported the same species (Zhang *et al.*, 2011). Some research effort has been devoted

To cite this paper: Yang, Y., M. Li, Y. Yi, R. Li and C. Dong, 2018. Identification of 12 lncRNAs as potential miRNA targets in roots of *Rehmannia glutinosa* responding to replanting disease. *Int. J. Agric. Biol.*, 20: 1201–1210

to identifying the causative factors for the disease, focusing on physiological aspects and the potential toxicity of root exudates (Li *et al.*, 2012; Gu *et al.*, 2013). Although evidence has recently emerged which suggests that specific miRNAs may make a contribution (Li *et al.*, 2013), the molecular basis of the species' sensitivity to the disease remains still exploring.

Compared with its stems, leaves and flowers, *R. glutinosa* roots are the most sensitive tissue-responsive to the replanting disease (Zhang and Lin, 2009), and so far, no poly(A)$^-$ lncRNAs as miRNA targets and their potential roles in the species' roots under continuous cultivation pressure are investigated. RNA-seq technology has been widely used to characterize the effect of external treatments on the transcriptome (Levin *et al.*, 2010), and RNA-seq experiments targeting the poly(A)$^-$ fraction of both the *Arabidopsis thaliana* and rice transcriptome have succeeded in identifying a number of new classes of lncRNAs (Liu *et al.*, 2013; Di *et al.*, 2014). Here, the RNA-seq platform was used to identify a set of lncRNA candidates present in *R. glutinosa* roots. Importantly, to investigate potential activities of poly(A)$^-$ lncRNAs as miRNA targets involved in replanting disease, some of these differed in abundance between the roots harvested from a new planting (R1) and those harvested from a replanted crop (R2). Some lncRNA candidates as miRNA targets will be explored in potential molecular regulation roles of the disease.

Materials and Methods

Plant Culture and RNA Extraction

R. glutinosa plants (cultivar "Wen 85-5") were obtained from the Wen Xian Agricultural Institute (Jiaozuo City, Henan Province, China). Two adjacent field plots were isolated from one another by a submerged brick partition, and each was filled with soil taken from a field where *R. glutinosa* had not been grown for at least ten years. In the first season (2015), *R. glutinosa* was planted in one plot and maize in the other; the following spring (2016), both plots were planted to *R. glutinosa*; the former plot is referred to hereafter as R2, and the latter as R1. The experimental pipeline was shown in Fig. 1. Roots were collected at six time points during 2016, spaced at 20 day intervals between May 22 and Aug 31; each sample represented a pool of roots produced by five plants. The samples used as a source of RNA for the construction and validation of an RNA library were harvested on July 22. The roots used to track the temporal variation in the abundance of specific lncRNA and mRNA molecules were those harvested at the six time points mentioned above. RNA was extracted from root material using the TRIzol reagent (Invitrogen Corporation, Carlsbad, USA). The concentration of RNA in the extract was estimated spectrophotometrically and its integrity was verified by electrophoresis through a 1.5% (w/v) agarose gel.

Poly(A)$^-$ RNA Sequencing and Assembly

The poly(A)$^+$ RNA content of each of the RNA extracts was eliminated by three treatments with an Oligotex mRNA Mini Kit (Qiagen, Hilden, Germany), while ribosomal RNA was removed by processing the sample three times with a RiboMinus™ Plant Kit (Invitrogen). After the addition of an adenosine to the 3' ends of the remaining RNA fraction, a SMART™ cDNA Library Construction Kit (Clontech Laboratories, Inc., California, USA) was used to convert it into cDNA. Illumina adaptor oligonucleotides were added and fragments in the length range 200–500 bp were isolated from an agarose gel, PCR-amplified and gel-purified. Raw sequence reads were generated by the HiSeq™ 2000 System (Illumina, Inc., San Diego, California). Adaptor sequences were trimmed and low-quality sequences removed. The assembly of the poly(A)$^-$ transcriptome was carried out using the Trinity tool (Grabherr *et al.*, 2011).

Construction of a lncRNA Candidate Library

The assembled poly(A)$^-$ lncRNA sequences were aligned via a BlastN search of the NCBI NR, SwissProt, KEGG and COG databases, applying a threshold E-value of 10^{-5} (Tatusov *et al.*, 2000; Ye *et al.*, 2006; Kanehisa *et al.*, 2008). Sequences lacking any functional annotation were subjected to ORF prediction using ESTScan software (Iseli *et al.*, 1999) based on default parameters. A cut-off of 100 amino acid (aa) residues was applied to distinguish lncRNAs from protein-encoding transcripts, as suggested elsewhere (MacIntosh *et al.*, 2001). High coding potential sequences were removed with the help of the coding potential calculator (CPC) program (Kong *et al.*, 2007), applying default parameters. Non-protein coding housekeeping sequences were identified by reference to the Rfam v12.0 database (rnacentral.org/expert-database/rfam), applying default parameters (Nawrocki *et al.*, 2015).

Identification of lncRNA Candidates or mRNAs as miRNA Targets

A set of mRNA sequences were collected from the assemblied 87,665 ploy (A)$^+$ transcripts in our previous reports (Li *et al.*, 2013) and the ploy (A)$^-$ CDS in the present research. A set of 660 *R. glutinosa* miRNAs identified previously (Li *et al.*, 2013) represented a foundation of lncRNA candidates and mRNAs as miRNA targets. Then we combined 367,290 from R1 plants with 769,336 from R2 ones and finally obtained 896,341 unique reads (Li *et al.*, 2013), which were used to verify the lncRNA and mRNA fragments cleaved by miRNAs targeted. The CleaveLand v4.0 program (Addo-Quaye *et al.*, 2009) was used in conjunction with the degradome data to identify lncRNA candidates and mRNAs as miRNA targets. At most one mismatch was allowed between positions 9–12 from the 5′ end of the miRNA sequence, the total number of bulges or mismatches in the other regions was limited to 4 nt and no

continuous mismatches were tolerated. To avoid false candidates, the p-values of the CleaveLand program were less than 0.05. The abundance of each distinct lncRNA/mRNA acting as an miRNA target was used to build a t-plot. Conserved lncRNAs were identified by searching the NONCODE v4.0 database (noncode.org/), using default parameters.

Prediction of lncRNA Candidate Function

Based on the ceRNA concept (Salmena *et al.*, 2011; Fan *et al.*, 2014), a set of the lncRNA/mRNA modules targeted by a given miRNA was established, and the function of the mRNA involved was taken as being also the function of the lncRNA. A BlastN search of the NCBI NR database was used to annotate the mRNAs (Ye *et al.*, 2006). The agriGO toolkit (bioinfo.cau.edu.cn/agriGO) was employed to classify mRNA function (Du *et al.*, 2010). The mRNA pathways were analyzed by searching aganist KEGG database (http://www.kegg.jp/kegg/tool/map_pathway).

Screening of Differentially Abundant lncRNAs

Firstly, the abundance of each lncRNA candidate was calculated from the number of reads per million mapped reads (RPKM). The frequency of each lncRNA candidate was first normalized, and was set to 0.01 for those which were not recovered in the library. Those lncRNAs which were present at a relative frequency of <10 in both the R1 and R2 libraries were excluded from the analysis. The p-value and false discovery rate (FDR) thresholds were set at 0.001 and 0.001, respectively, and a $|\log_2(\text{fold-change})| > 1$ was set as the threshold for significantly differential expression (Audic and Claverie, 1997; Benjamini *et al.*, 2001).

Secondly, to further screening the differentially abundant lncRNA candidates, the transcript abundance of the preliminarily differentially abundant lncRNA candidates were assessed by Quantitative real time PCR (qRT-PCR) with three replicates. In briefly, the poly(A)$^-$ RNA was first polyadenylated using poly(A) polymerase (Takara Bio Inc., Tokyo, Japan). A 0.5 μg aliquot of the added poly(A) RNA was transcribed with 5U M-MLV reverse transcriptase in a 20 μL reaction, formulated according to the manufacturer's instructions (Takara Bio Inc., Tokyo, Japan). For the subsequent PCR, primers recognizing the lncRNA sequences were designed using Beacon Designer v8.0 software (www.premierbiosoft.com/molecular_beacons). A fragment of the *Actin* gene (EU526396.1) was used as the reference sequence. Each 25 μL reaction contained 0.2 μM of each primer, 12.5 μL SYBR® Premix EX Taq™ and 100 ng cDNA, and was based on the SYBR Green method (Takara Bio Inc., Tokyo, Japan). Negative control reactions contained no cDNA. A five folds dilution of the cDNA template was tested under the same conditions. The PCR regime comprised an initial denaturing step (95°C/10s), followed by 40 cycles of 95°C/5s, 60°C/10s, 72°C/15s and a final ramping of 60–95°C at 88°C per s to determine the amplicon's dissociation behavior. Three technical replicates were included per sample. Relative transcript abundances were calculated using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001) and the data were normalized on the basis of the abundance of the reference gene transcript. The significance differences in relative expression of each sample were assessed by the LSD-t test method (Fisher, 1925). At last, the differentially abundant lncRNA candidates were confirmed by qRT-PCR methods.

Cloning and Sequence Analysis of the lncRNA Complete Transcripts

The lncRNA sequences were extended using 5' and 3'-RACE PCR. The relevant primers were designed using Oligo v6.0 software (www.oligo.net) based on the assembled lncRNA sequences. For 3'-RACE PCR, the poly(A)$^-$ RNAs were first poly(A) tailed using poly(A) polymerase (Takara Bio Inc., Tokyo, Japan), and then a 0.5 μg aliquot of the added poly(A) RNA and a 0.5 μg aliquot of poly(A)$^-$ RNA were, respectively, transcribed by SMARTScribeTM reverse transcriptase (Takara Bio Inc., Tokyo, Japan) with the 3' or the 5'-RACE CDS Primer A. A touchdown PCR was implemented to amplify the first cDNA with each lncRNA fragment-specific primer of 5' or 3'-RACE, using the Universal Primer A Mix supplied by the kit. Nested PCRs were performed with the gene-specific primer 5' or 3'-RACE, along with the short Nested Universal Primer A provided within the kit. PCR products were purified, cloned and sequenced. The cloned lncRNA sequences were assembled using CAP3 (seq.cs.iastate.edu/cap3), and any ORFs harbored within the derived lncRNAs were predicted using the ORF Finder tool (www.ncbi.nlm.nih.gov/orffinder).

Expression Patterns Analysis of the lncRNAs and mRNAs

The abundance of a selection of the lncRNAs and mRNAs was derived by qRT-PCR (three replicates) in order to validate the *in silico*-derived estimates. To assay the lncRNAs, the poly(A)$^-$ RNA was first polyadenylated (Takara Bio Inc., Tokyo, Japan). A 0.5 μg aliquot of the added poly(A) RNA and a 0.5 μg aliquot of poly(A)$^+$ RNA were, respectively, transcribed with 5U M-MLV reverse transcriptase in a 20 μL reaction, formulated according to the manufacturer's instructions (Takara Bio Inc., Tokyo, Japan). For the subsequent PCR, primers recognizing the lncRNA and mRNA sequences were designed using Beacon Designer v8.0 software. QRT-PCR was performed as the mentioned above.

Results

The R1 and R2 lncRNA Libraries

A total of 42,010,624 raw RNA-Seq reads was derived from

the R1 root poly(A)$^-$ RNA library and 58,688,784 from the R2 library. Adapter trimming and quality filtering reduced these numbers to, respectively, 40,936,179 and 57,124,179. The pooled reads resolved into 88,951 non-redundant transcripts of length >200 nt (hereafter termed "unigenes") (Table 1). Subsequent analysis revealed that 25,944 (CDS) of the unigene sequences encoded a protein, so these were discarded. Application of the CPC led to the deletion of a further 5,738 sequences, along with 24 identified via the Rfam database as housekeeping npcRNAs. The remaining 56,705 unigenes constituted the final lncRNA candidate library. The length of about 84% of the unigene sequences layed in the range 200–500 nt (the characteristic length of lncRNAs) (Liu *et al.*, 2015). Their length was generally below than that of the poly(A)$^-$ mRNAs as a whole (Table 1).

Identification and Predicted Function of lncRNA Candidates as miRNA Targets

The CleaveLand pipeline output suggested that 155 lncRNA candidates were formed by lncRNA-miRNA duplexes. When the lncRNA sequences were used to query the ncRNA database, only three aligned with *A. thaliana* lncRNAs, supporting the suggestion that this class of RNA sequence is less well conserved than other non-coding RNAs (Liu *et al.*, 2015). This left a total of 152 lncRNA candidates which appeared to be specific to *R. glutinosa.*

To predict the function of the 152 specific and three conserved lncRNA candidates, based on the ceRNA hypothesis, a set of modules was established involving mRNAs and lncRNAs interacting with the same miRNAs. The 25,944 discarded coding poly(A)$^-$ mRNAs were combined with the 87,665 previously assembled poly(A)$^+$ mRNAs, and the degradome data was used to identify potential lncRNA-miRNA-mRNA duplexes. Thus the final set of lncRNA candidate/mRNA modules comprised 114 lncRNA candidates and 214 poly(A)$^+$ mRNAs as the same miRNA targets. In one example, the abundance of degradome reads and cleavage sites in lncR31 and unigene9575_All, suggested that the targeting miRNA was miR171b (Fig. 2).

Assuming that the 114 lncRNA candidates were co-transcribed with the 214 mRNAs, the function of these lncRNAs could be inferred from the function of the latter. NR annoation showed that, the set of mRNAs encoded a range of products, for example, several mRNAs of the modules encoded chaperone protein dnaJ 8 (dnaJ 8), cation transport ATPase (CTP), MADS-box transcription factor MSM1 (MSM1), peroxisomal membrane protein PEX14-like (PEX14), a wall-associated receptor kinase-like 14 (WAKL14), etc. (Table 2). After using the agriGO toolkit to perform GO analysis of mRNAs, we found that these lncRNA candidates were involved in multiple biological processes, participated in the formation of many cellular components, and influenced the activities of molecular functions (Fig. 3A). We performed enrichment analysis again, and found that these lncRNA candidates mainly participated in cellular, metabolic and response to stimulus processes, and the moleculuar of these lncRNAs were focused on catalytic activity and binding functions. They could be mostly involved in the formation of cell parts, cells and organelles. Based on KEGG analysis, these lncRNA candidates could be involved in 12 KEGG pathways and mainly participated in Primary metabolism, Folding, sorting and degradation, Transcription and Translation pathways (Fig. 3B).

Differentially Abundant lncRNA Candidates

Evidence for differential abundance of particular lncRNA candidates in the R1 and R2 roots was sought by comparing the frequency of the 155 lncRNA candidates identified as miRNA targets. In all, 32 of the lncRNAs initially showed evidence of differential abundance (Fig. 4). And then the differential abundance of 28 (17 more abundant in R2 than in R1, and 11 less abundant) of these was confirmed using qRT-PCR (Fig. 4).

Characterization of 12 Differentially Abundant lncRNA Candidate Sequences

12 of the differentially abundant lncRNA candidates were followed up by obtaining a more complete sequence, using the RACE PCR approach. Their full lengths were found to range from 511-3,072 nt (Table 2). According to a BlastN search, none had a match with any sequence in the NCBI NR database. Each of the sequences harbored an ORF smaller than 100 aa residues. They encoded proteins typical for lncRNA products (Rymarquis *et al.*, 2008). Thus, the 12 sequences were considered as being specific to *R. glutinosa.*

Temporal Behavior of the Potential lncRNA/mRNA Modules

The qRT-PCR approach was used to characterize variation over time in the abundance of the 12 potential lncRNA/mRNA modules (Fig. 5). In the R2 roots, the seven lncRNAs showing enhanced abundance (along with their competitive mRNAs) were both present in high abundance in at least three of the sampling times, while most of the five reduced abundance ones were associated with a low abundance mRNA. The temporal profiling of the potential lncRNA/mRNA modules showed that their abundance was induced by replanting disease during the development of the *R. glutinosa* root system. Their abundance was particularly divergent between R1 and R2 roots during the tuberous roots' late expansion phase (July 2 through Aug 12). For example, the presence of lncR37 was higher in the R2 than the R1 root, along with that of *AP2 TF*, at four of the six sampling times; that of both lncR149 and *RNase H2* was lower at five of the six sample times.

Table 1: Length distribution of the poly(A)$^-$ transcripts in *R. glutinosa* roots

Size (bp)	Unigenes (percentage)	CDS (percentage)	lncRNA candidates (percentage)	Potential coding sequences and npcRNAs (percentage)
200-300	35,595 (40.02)	271 (1.04)	31,911 (56.28)	3,413 (54.16)
300-400	14,241 (16.01)	3,173 (12.23)	10,177 (17.95)	891 (14.14)
400-500	8,486 (9.54)	2,453 (9.45)	5,518 (9.73)	515 (8.17)
500-600	5,590 (6.28)	2,005 (7.73)	3,192 (5.63)	393 (6.24)
600-700	3,941 (4.43)	1,728 (6.66)	1,930 (3.40)	283 (4.49)
700-800	3,040 (3.42)	1,631 (6.29)	1,210 (2.13)	199 (3.16)
800-900	2,450 (2.75)	1,516 (5.84)	805 (1.42)	129 (2.05)
900-1000	1,883 (2.12)	1,282 (4.94)	509 (0.90)	92 (1.46)
1000-1500	6,578 (7.40)	5,278 (20.34)	1,049 (1.85)	251 (3.98)
1500-2000	3,518 (3.95)	3,189 (12.29)	267 (0.47)	62 (0.98)
2000-2500	1,770 (1.99)	1,654 (6.38)	61 (0.11)	55 (0.87)
2500-3000	907 (1.02)	860 (3.31)	34 (0.06)	13 (0.21)
>3000	952 (1.07)	904 (3.48)	42 (0.07)	6 (0.10)
Total	88,951 (100.00)	25,944 (100.00)	56,705 (100.00)	6,302 (100.00)

Table 2: The 12 validated lncRNA sequence characteristics and their competitive mRNA annoation

lncRNA (ID)	Assembled length (nt)	Complete length (nt)	ORF maximal length (nt)	ORF amount	Competetive mRNA annoation, abbreviation
lncR17	569	2010	225	14	transcript variant 1, TV1
lncR25	239	491	254	3	MADS-box transcription factor MSM1, MSM1
lncR37	525	798	225	4	AP2 transcription factor, AP2
lncR75	348	621	240	5	Translocon at the outer membrane of chloroplasts 64, TOC64
lncR90	828	1425	138	6	cation transport ATPase, CTP
lncR91	904	1636	198	9	chaperone protein dnaJ 8, dnaJ 8
lncR94	879	1456	237	10	anaphase-promoting complex subunit 11, APC11
lncR103	1491	3068	198	9	wall-associated receptor kinase-like 14, WAKL14
lncR111	1311	2493	267	17	Pleiotropic drug resistance protein 2, PDR2
lncR118	1132	2257	231	18	peroxisomal membrane protein PEX14-like, PEX14
lncR122	1099	1720	204	16	transcript variant 2, TV2
lncR149	619	1561	177	13	putative ribonuclease H2, RNase H2

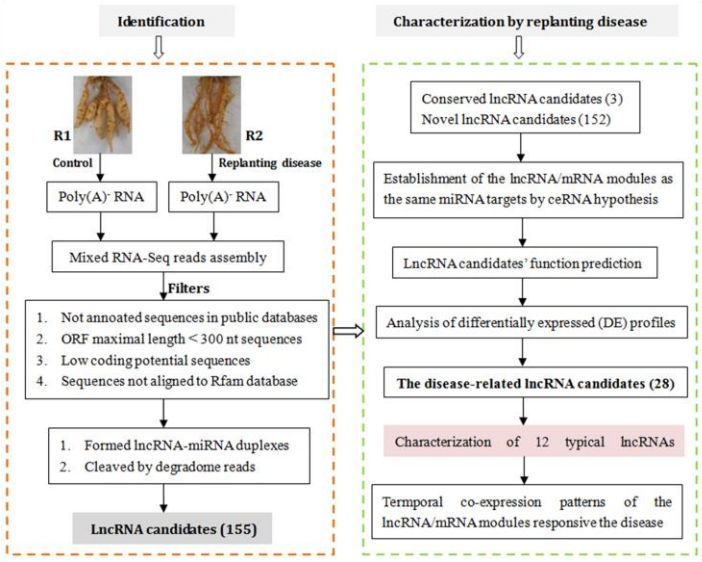

Fig. 1: Overview of the experimental methodology

Discussion

Current high capacity sequencing platforms have facilitated the acquisition of a number of both poly(A)$^+$ and poly(A)$^-$ transcriptomes from a range of plant species (Di *et al.*, 2014; Liu *et al.*, 2015). The *R. glutinosa* poly(A)$^+$ transcriptome has already been assembled (Li *et al.*, 2013; Yang *et al.*, 2014, 2015), but its poly(A)$^-$ transcriptome has only now been acquired in the course of the present study.

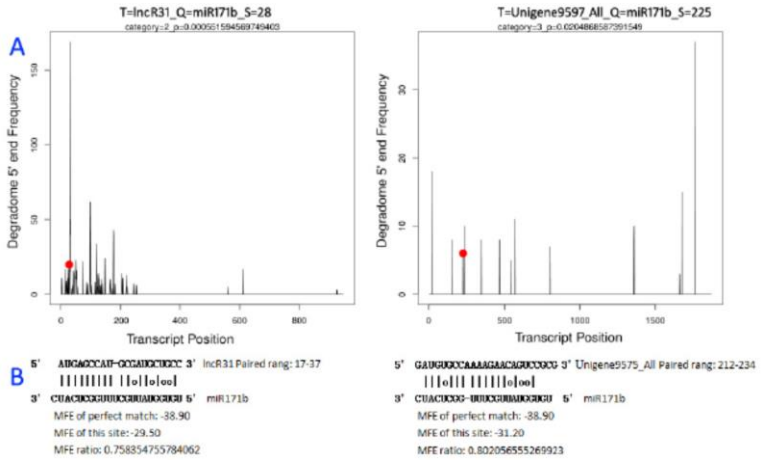

Fig. 2: The targeting of miR171b to lncR37 and Unigene9597_All, as suggested by the degradome. (A) Target plots (t-plots) of validated miR171b targets. The distribution of degradome reads between lncR31 and Unigene9597_All. Peaks at validated cleavage sites of the corresponding miRNAs are shown as red dots. (B) Alignment of the miRNA/lncRNA and miRNA/mRNA duplexes

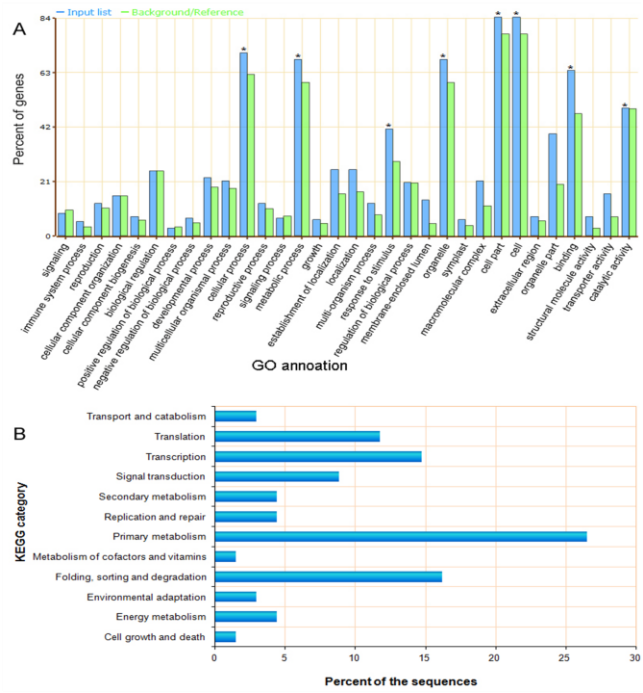

Fig. 3: GO and KEGG annoation of potential lncRNA candidate/mRNA modules. (A) GO category and enrichment: "*" represents significant categories; (B) KEGG category

Poly(A)$^-$ transcriptomes have been shown to be dominated by lncRNA candidates in certain species (Djebali *et al.*, 2012), and this was also the case for *R. glutinosa*, where about 64% of the library was represented by this class of molecule. This facilitates the further lncRNA discovery research in *R. glutinosa*. In addition, we have also gained a set of poly(A)$^-$ mRNA transcripts, filling in *R. glutinosa* mRNA transcriptome contents.

Although the genome sequence of *R. glutinosa* has not been acquired, combining its poly(A)$^-$ transcriptome with previously documented miRNA and degradome data allowed for the identification of a set of 155 lncRNA candidates as miRNA targets. The implication is that, along with mRNAs, lncRNA candidates too potentially act as miRNA targets, and as a result contribute to post-transcriptional modifications in *R. glutinosa*.

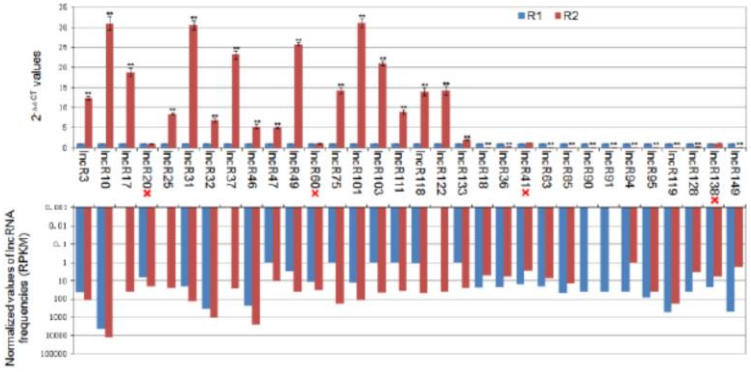

Fig. 4: Differential abundance of selected lncRNAs in the roots of R1 and R2 plants. Relative abundances were inferred from (A) qRT-PCR data, (B) RNA-Seq data. Error bars represented standard error (SE). "*" and "**" represent significant differences of tested indexes between R2 and R2 roots at p-value <0.05 and p-value < 0.01, respectively

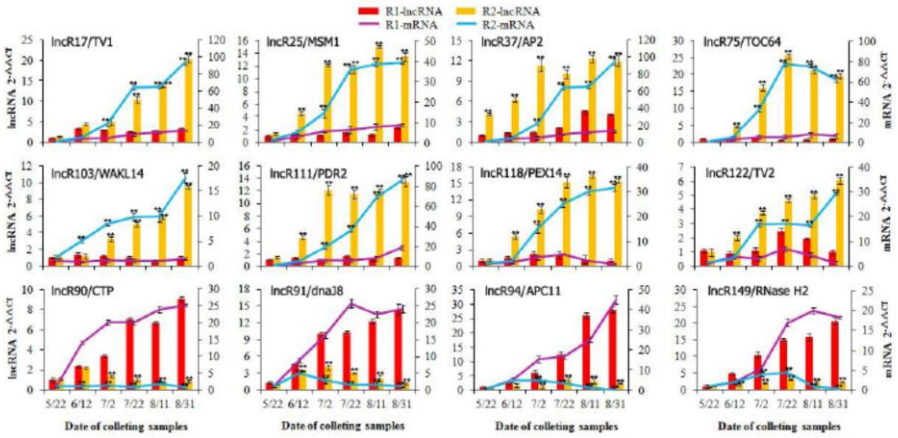

Fig. 5: Temporal behavior of 12 selected potential lncRNA/mRNA modules in R1 and R2 roots. Error bars represent standard error (SE); "*" and "**" represent significant differences of tested indexes between R2 and R2 roots at p-value <0.05 and p-value < 0.01, respectively

The ceRNA hypothesis proposes that lncRNAs compete with mRNAs as targets for miRNA activity (Salmena *et al.*, 2011; Fan *et al.*, 2014). Here, it was possible to establish some interactive modules involving some miRNAs which simultaneously target one (or more) members of the set of lncRNA candidates and one (or more) members of the set of mRNAs. At the same time, it was found that the poly(A)$^-$ mRNAs were not directly cleavable by miRNAs, which implied that none of the lncRNAs acted as a ceRNA for the poly(A)$^-$ mRNAs in *R. glutinosa* roots. Based on the ceRNA concept, the predictive functional analysis of the lncRNA candidates revealed that some could be involved in the response to stress biological processes and several key metabolism pathways. Plant lncRNAs have been shown to be a component of stress-responsive signaling (Fan *et al.*, 2014; Di *et al.*, 2014). If, as has been suggested, miRNA-directed mRNA degradation is a component of the *R. glutinosa* response to replanting disease (Li *et al.*, 2013), then the likelihood is that lncRNA candidates associated with some of the critical mRNAs may also be involved in, and perhaps even be causative of, the disease response.

A number of the lncRNA candidate/mRNA modules proved to be either more or less abundant in R2 (affected by replanting disease) than in R1 (unaffected) roots. Eight modules fitted the former category and four the latter. In the first category, two mRNAs – one (*MSM1*) encoding a MADS-box transcription factor (TF) and the other a member of the AP2 TF family – were identified. TFs are known to be the master regulators of gene expression in response to environmental stresses (Alter *et al.*, 2016; Dossa *et al.*, 2016). Both lncR24 and lncR37, the alternative targets for the critical miRNAs, were particularly abundant, thereby favoring the transcription and expression of both TF genes. The two TFs could be involved in lncRNA-miRNA-mRNA post-transcriptional networks, which formed part of the regulatory machinery underpinning the development of replanting disease. The next module involved lncR103 and the mRNA encoding WAKL14, an enzyme-linked

receptor responding to extracellular signals originating from the binding of respective ligands or growth factors (Meier *et al.*, 2010). A working hypothesis to describe the regulation of replanting disease holds that the exudation (so-called allelochemicals) of autotoxins from the root system is inductive (Zhang and Lin, 2009). In the replanted *R. glutinosa* root, these allelochemicals could act as the stress signals to induce the transcription of both lncR103 and *WAKL14*, the protein of WAKL14 would transduct the signals to the root intracellular and be involved in the stress signal pathways. A third lncRNA/mRNA module implicated in the disease response included lncR32 with its competing mRNA, which encodes a translocon formed at the outer membrane of the chloroplasts (TOC64), the product of this gene, may act as a docking protein for cytosolic cofactors required for the import of protein into the chloroplast (Sohrt and Soll, 2000). Overexpression of the lncR75/*TOC64* module could reshape the replanted *R. glutinosa* root cell protein components. A further module involved lncR118/*PEX14*. PEX14 is the only component of the peroxisomal translocon required for pexophagy (Zutphen *et al.*, 2008; Galiani *et al.*, 2016). Under stressful growing conditions, the up-regulation of *PEX14* in plants may increase the volume of the peroxisome, suggesting a possible mechanism for the plant to deal with enhanced oxidative stress (León, 2008; Cross *et al.*, 2016). An increase in the abundance of *PEX14* may result in a rise in the peroxisome level, implying a means for *R. glutinosa* responding to replanting disease stress. Of the other lncRNA candidates, which were more abundant in R2 roots, one interacted with the product of a homolog of *PDR2* (encoding the pleiotropic drug resistance protein 2), and two with the transcript variants *TV1* and *TV2*, which are known to be involved in the plant response to various stresses (Stukkens *et al.*, 2005; Dubrovina *et al.*, 2013).

Among the group of four modules present in reduced amount in the R2 roots, one involved lncR149, which competes with RNase H2 mRNA. This gene makes an important contribution to suppressing genome instability in *A. thaliana*, since its absence triggers an increase in the incidence of homologous recombination (Kalhorzadeh *et al.*, 2014). Thus its absence in the R2 roots could have an impact on genome stability, with far-reaching consequences for gene expression. A second module involved lncR90 and *CTP* mRNA, which encodes a cation transport ATPase, a key enzyme determining the uptake of nutrients from the soil (Mills *et al.*, 2012). The reduction in lncR90 abundance would have been associated with a fall in the abundance of *CTP*, which would be predicted to compromise the supply of nutrients. The other two lncRNAs implicated competed with mRNAs encoding the chaperone protein dnaJ8 and an anaphase-promoting complex subunit 11 (APC11). Both of these gene products are involved in crucial roles of protein synthesis (Zhu *et al.*, 2015; Guo *et al.*, 2016), so that a reduction in the modules' abundance would likely also inhibit inhibit much of the plant's core metabolism.

Conclusion

The interaction between lncRNA candidates and mRNAs ensures that the abundance of certain lncRNAs in the *R. glutinosa* root potentially correlates with the expression of replanting disease. The indication was that the lncRNA candidates under continuous cultivation pressure, via their competition for specific mRNAs, could activate the specific stress signaling pathways, repress core mechanisms, and result in forming the disease in *R. glutinosa*. Our working hypothesis holds that a number of lncRNA candidates potentially as miRNA targets contribute either directly or indirectly to the disease phenomenon. Uncovering the mechanistic basis of this contribution still requires further experimentation to validate the molecular functions of the lncRNAs in response to replanting disease.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (Nos. 81403037 and 81503193), the Project Supported by Science Foundation of Henan University of Technology (No. 2017RCJH05) and the Science and Technology Research Key Project of Henan Educational Committee (No. 12A180006).

References

Addo-Quaye, C., W. Miller and M.J. Axtell, 2009. CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. *Bioinformatics*, 25: 130–131

Alter, P., S. Bircheneder, L.Z. Zhou, U. Schlüter, M. Gahrtz, U. Sonnewald and T. Dresselhaus, 2016. Flowering time-regulated genes in maize include the transcription factor ZmMADS1. *Plant Physiol.*, 172: 389–404

Audic, S. and J.M. Claverie, 1997. The significance of digital gene expression profiles. *Genome Res*., 7: 986–995

Benjamini, Y., D. Drai, G. Elmer, N. Kafkafi and I. Golani, 2001. Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.*, 125: 279–284

Cross, L.L., H.T. Ebeed and A. Baker, 2016. Peroxisome biogenesis, protein targeting mechanisms and *PEX* gene functions in plants. *Biochem. Biophys. Acta*, 1863: 850–862

Cui, P., Q. Lin, F. Ding, C. Xin, W. Gong, L. Zhang, J. Geng, B. Zhang, X. Yu, J. Yang, S. Hu and J. Yu, 2010. A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. *Genomics*, 96: 259–265

Di, C., J. Yuan, Y. Wu, J. Li, H. Lin, L. Hu, T. Zhang, Y. Qi, M.B. Gerstein, Y. Guo and Z.J. Lu, 2014. Characterization of stress-responsive lncRNAs in *Arabidopsis thaliana* by integrating expression, epigenetic and structural features. *Plant J*., 80: 848–861

Ding, Y., Y. Tang, C.K. Kwok, Y. Zhang, P.C. Bevilacqua and S.M. Assmann, 2013. *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*, 505: 696–700

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

Dossa, K., X. Wei, D. Li, D. Fonceka, Y. Zhang, L. Wang, J. Yu, L. Boshou, D. Diouf, N. Cissé and X. Zhang, 2016. Insight into the AP2/ERF transcription factor superfamily in sesame and expression profiling of DREB subfamily under drought stress. *BMC Plant Biol*., 16: 171

Du, Z., X. Zhou, Y. Ling, Z. Zhang and Z. Su, 2010. AgriGO: a GO analysis toolkit for the agricultural community. *Nucl. Acids Res.*, 38: 64–70

Dubrovina, A.S., K.V. Kiselev and Y.N. Zhuravlev, 2013. The role of canonical and noncanonical pre-mRNA splicing in plant stress responses. *Biomed. Res. Int.*, 2013: 264314

Fan, C., Q. Wei, Z. Hao and G. Li, 2014. Prediction and functional analysis of lincRNAs targeted by miRNAs. *Yi Chuan*, 36: 1226–1234

Fisher, R.A., 1925. The resemblance between twins, a statistical examination of Lauterbach's measurements. *Genetics*, 10: 569–579

Franco-Zorrilla, J.M., A. Valli, M. Todesco, I. Mateos, M.I. Puga, I. Rubio-Somoza, A. Leyva, D. Weigel, J.A. García and J. Paz-Ares, 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.*, 39: 1033–1037

Galiani, S., D. Waithe, K. Reglinski, L.D. Cruz-Zaragoza, E. Garcia, M.P. Clausen, W. Schliebs, R. Erdmann and C. Eggeling, 2016. Super-resolution microscopy reveals compartmentalization of perxisomal membrane proteins. *J. Biol. Chem.*, 291: 16948–16962

Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. Di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman and A. Regev, 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.*, 29: 644–652

Gu, L., M.M. Niu, H.Y. Zheng, J.M. Wang, L.K. Wu, Z.F. Li and Z.Y. Zhang, 2013. Effect of continuous cropping of *Rehmannia* on its morphological and physiological characteristics. *Zhong Yao Cai*, 36: 691–695

Guo, L., L. Jiang, Y. Zhang, X.L. Lu, Q. Xie, D. Weijers and C.M. Liu, 2016. The anaphase-promoting complex initiates zygote division in *Arabidopsis* through degradation of cyclin B1. *Plant J.*, 86: 161–174

Iseli, C., C.V. Jongeneel and P. Bucher, 1999. ESTS can: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, 99: 138–148

Joshi, R.K., S. Megha, U. Basu, M.H. Rahman and N.N. Kav, 2016. Genome wide identification and functional prediction of long non-coding RNAs responsive to *Sclerotinia sclerotiorum* infection in *Brassica napus. PLoS One*, 11: e0158784

Kalhorzadeh, P., Z. Hu, T. Cools, S. Amiard, E.M. Willing, N. De Winne, K. Gevaert, G. De Jaeger, K. Schneeberger, C.I. White and L. De Veylder, 2014. *Arabidopsis thaliana* RNase H2 deficiency counteracts the needs for the WEE1 checkpoint kinase but triggers genome instability. *Plant Cell*, 26: 3680–3692

Kanehisa, M., M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu and Y. Yamanishi, 2008. KEGG for linking genomes to life and the environment. *Nucl. Acids Res.*, 36: 480–484

Kong, L., Y. Zhang, Z.Q. Ye, X.Q. Liu, S.Q. Zhao, L. Wei and G. Gao, 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucl. Acids Res.*, 35: 345–349

Kumar, R., 2014. Role of microRNAs in biotic and abiotic stress responses in crop plants. *Appl. Biochem. Biotechnol.*, 174: 93–115

Kwenda, S., P.R. Birch and L.N. Moleleki, 2016. Genome-wide identification of potato long intergenic noncoding RNAs responsive to *Pectobacterium carotovorum* subspecies brasiliense infection. *BMC Genom.*, 17: 614

Lee, J.T., 2012. Epigenetic regulation by long noncoding RNAs. *Science*, 338: 1435–1439

León, J., 2008. Peroxisome proliferation in *Arabidopsis*: The challenging identification of ligand perception and downstream signaling is closer. *Plant Signal. Behav.*, 3: 671–673

Levin, J.Z., M. Yassour, X. Adiconis, C. Nusbaum, D.A. Thompson, N. Friedman, A. Gnirke and A. Regev, 2010. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods*, 7: 709–715

Li, D.S., J.L. Ainiwaer, I. Sheyhiding, Z. Zhang and L.W. Zhang, 2016. Identification of key long non-coding RNAs as competing endogenous RNAs for miRNA-mRNA in lung adenocarcinoma. *Eur. Rev. Med. Pharmacol. Sci.*, 20: 2285–2295

Li, M.J., Y.H. Yang, X.J. Chen, F.Q. Wang, W.X. Lin, Y.J. Yi, L. Zeng, S.Y. Yang and Z.Y. Zhang, 2013. Transcriptome-wide identification of *R. glutinosa* miRNAs and their targets: the role of miRNA activity in the replanting disease. *PloS One*, 8: e68531

Li, Z.F., Y.Q. Yang, D.F. Xie, L.F. Zhu, Z.G. Zhang and W.X. Lin, 2012. Identification of autotoxic compounds in fibrous roots of *Rehmannia* (*Rehmannia glutinosa Libosch.*). *PLoS One*, 7: e28806

Liu, J., H. Wang and N.H. Chua, 2015. Long noncoding RNA transcriptome of plants. *Plant Biotechnol. J.*, 13: 319–328

Liu, T.T., D. Zhu, W. Chen, W. Deng, H. He, G. He, B. Bai, Y. Qi, R. Chen and X.W. Deng, 2013. A global identification and analysis of small nucleolar RNAs and possible intermediate-sized non-coding RNAs in *Oryza sativa. Mol. Plant*, 6: 830–846

Livak, K.J. and T.D. Schmittgen, 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2 (-Delta Delta C (T)). *Method*, 25: 402–408

Livyatan, I., A. Harikumar, M. Nissim-Rafinia, R. Duttagupta, T.R. Gingeras and E. Meshorer, 2013. Non-polyadenylated transcription in embryonic stem cells reveals novel non-coding RNA related to pluripotency and differentiation. *Nucl. Acids Res.*, 41: 6300–6315

Lu, X., X. Chen, M. Mu, J. Wang, X. Wang, D. Wang, Z. Yin, W. Fan, S. Wang, L. Guo and W. Ye, 2016. Genome-wide analysis of long noncoding RNAs and their responses to drought stress in cotton (*Gossypium hirsutum* L.). *PLoS One*, 11: e0156723

MacIntosh, G.C., C. Wilkerson and P.J. Green, 2001. Identification and analysis of *Arabidopsis* expressed sequence tags characteristic of non-coding RNAs. *Plant Physiol.*, 127: 765–776

Meier, S., O. Ruzvidzo, M. Morse, L. Donaldson, L. Kwezi and C. Gehring, 2010. The *Arabidopsis* wall associated kinase-like 10 gene encodes a functional guanylyl cyclase and is co-expressed with pathogen defense related genes. *Plos One*, 5: e8904

Mills, R.F., K.A. Peaston, J. Runions and L.E. Williams, 2012. HvHMA2, a P(1B)-ATPase from barley, is highly conserved among cereals and functions in Zn and Cd transport. *Plos One*, 7: e42640

Nawrocki, E.P., S.W. Burge, A. Bateman, J. Daub, R.Y. Eberhardt, S.R. Eddy, E.W. Floden, P.P. Gardner, T.A. Jones, J. Tate and R.D. Finn, 2015. Rfam 12.0: updates to the RNA families database. *Nucl. Acids Res.*, 43: 130–137

Ponjavic, J., C.P. Ponting and G. Lunter, 2007. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genom. Res.*, 17: 556–565

Ponting, C.P., P.L. Oliver and W. Reik, 2009. Evolution and functions of long noncoding RNAs. *Cell*, 136: 629–641

Rubio-Somoza, I., D. Weigel, J.M. Franco-Zorilla, J.A. Garcia and J. Paz-Ares, 2011. ceRNAs: miRNA target mimic mimics. *Cell*, 147: 1431–1432

Rymarquis, L.A., J.P. Kastenmayer, A.G. Huttenhofer and P.J. Green, 2008. Diamonds in the rough: mRNA-like non-coding RNAs. *Trends Plant Sci.*, 13: 329–334

Salmena, L., L. Poliseno, Y. Tay, L. Kats and P.P. Pandolfi, 2011. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*, 146: 353–358

Sohrt, K. and J. Soll, 2000. Toc64, a new component of the protein translocon of chloroplasts. *J. Cell Biol.*, 148: 1213–1222

Song, Y. and D. Zhang, 2017. The role of long noncoding RNAs in in plant stress tolerance. *Methods Mol. Biol.*, 1631: 41–68

Song, Y.X., J.X. Sun, J.H. Zhao, Y.C. Yang, J.X. Shi, Z.H. Wu, X.W. Chen, P. Gao, Z.F. Miao and Z.N. Wan, 2017. Non-coding RNAs participate in the regulatory network of CLDN4 via ceRNA mediated miRNA evasion. *Nat. Commun.*, 8: 289

Stukkens, Y., A. Bultreys, S. Grec, T. Trombik, D. Vanham and M. Boutry, 2005. NpPDR1, a pleiotropic drug resistance-type ATP-binding cassette transporter from *Nicotiana plumbaginifolia*, plays a major role in plant pathogen defense. *Plant Physiol.*, 139: 341–352

Su, Z., X. Zhi, Q. Zhang, L. Yang, H. Xu and Z. Xu, 2016. LncRNA H19 functions as competing endogenous RNA to regulate AQP3 expression by sponging miR-874 in the intestinal barrier. *FEBS Lett.*, 590: 1354–1364

Tang, W., Y. Zheng, J. Dong, J. Yu, J. Yue, F. Liu, X. Guo, S. Huang, M. Wisniewski, J. Sun, X. Niu, J. Ding, J. Liu, Z. Fei and Y. Liu, 2016. Comprehensive transcriptome profiling reveals long noncoding RNA expression and alternative splicing regulation during fruit development and ripening in Kiwifruit (*Actinidia chinensis*). *Front. Plant Sci*., 7: 335

Tatusov, R.L., M.Y. Galperin, D.A. Natale and E.V. Koonin, 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucl. Acids Res*., 28: 33–36

Wen, X.S., S.L. Yang, J.H. Wei and J.H. Zheng, 2002. Textual research on planting history of *Rehmannia glutinosa* and its cultivated varieties. *Chin. Trad. Herb. Drugs*, 33: 946–949

Wu, H.J., Z.M. Wang, M. Wang and X.J. Wang, 2013. Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants. *Plant Physiol*., 161: 1875–1884

Yang, Y.H., M.J. Li, X.Y. Li, X.J. Chen, W.X. Lin and Z.Y. Zhang, 2015. Transcriptome-wide identification of the genes responding to replanting disease in *Rehmannia glutinosa* L. roots. *Mol. Biol. Rep*., 42: 881–892

Yang, Y.H., M.J. Li, X.J. Chen, P.F. Wang, F.Q. Wang, W.X. Lin, Y.J. Yi and Z.Y. Zhang, 2014. De novo characterization of the *Rehmannia glutinosa* leaf transcriptome and analysis of gene expression associated with replanting disease. *Mol. Breed*., 34: 905–915

Ye, J., S. McGinnis and T.L. Madden, 2006. BLAST: improvements for better sequence analysis. *Nucl. Acids Res.*, 34: 6–9

Zhang, Z. and W. Lin, 2009. Continuous cropping obstacle and allelopathic autotoxicity of medicinal plants. *Chin. J. Eco-Agric*., 17: 189–196

Zhang, Z., W. Lin, Y. Yang, H. Chen and X. Chen, 2011. Effects of continuous cropping *Rehmannia glutinosa* L. on diversity of fungal community in rhizospheric soil. *Agric. Sci. Chin.*, 10: 1374–1384

Zhu, Q.H., S. Stephen, J. Taylor, C.A. Helliwell and M.B. Wang, 2014. Long noncoding RNAs responsive to *Fusarium oxysporum* infection in *Arabidopsis thaliana*. *New Phytol*., 201: 574–584

Zhu, X., S. Liang, J. Yin, C. Yuan, J. Wang, W. Li, M. He, J. Wang, W. Chen, B. Ma, Y. Wang, P. Qin, S. Li and X. Chen, 2015. The DnaJ OsDjA7/8 is essential for chloroplast development in rice (*Oryza sativa*). *Gene*, 574: 11–19

Zutphen, T.V., M. Veenhuis and I.J. van der Klei, 2008. Pex14 is the sole component of the peroxisomal translocon that is required for pexophagy. *Autophagy*, 4: 63–66

(Received 28 August 2017; Accepted 28 December 2017)