

Construction of a genomewide RNAi mutant library in rice

Lei Wang^{1,*}, Jie Zheng¹, Yanzhong Luo¹, Tao Xu¹, Qiuxue Zhang¹, Lan Zhang¹, Miaoyun Xu¹, Jianmin Wan², Ming-Bo Wang³, Chunyi Zhang¹ and Yunliu Fan¹

¹Biotechnology Research Institute, The National Key Facility for Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing, China

²Crop Sciences Institute, The National Key Facility for Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing, China

³CSIRO Plant Industry, Canberra, NSW, Australia

Received 24 March 2013;

revised 18 May 2013;

accepted 24 May 2013.

*Correspondence (Tel/fax +86-10-

82106134;

email wanglei01@caas.cn)

Summary

Long hairpin RNA (hpRNA) transgenes are a powerful tool for gene function studies in plants, but a genomewide RNAi mutant library using hpRNA transgenes has not been reported for plants. Here, we report the construction of a hpRNA library for the genomewide identification of gene function in rice using an improved rolling circle amplification-mediated hpRNA (RMHR) method. Transformation of rice with the library resulted in thousands of transgenic lines containing hpRNAs targeting genes of various function. The target mRNA was down-regulated in the hpRNA lines, and this was correlated with the accumulation of siRNAs corresponding to the double-stranded arms of the hpRNA. Multiple members of a gene family were simultaneously silenced by hpRNAs derived from a single member, but the degree of such cross-silencing depended on the level of sequence homology between the members as well as the abundance of matching siRNAs. The silencing of key genes tended to cause a severe phenotype, but these transgenic lines usually survived in the field long enough for phenotypic and molecular analyses to be conducted. Deep sequencing analysis of small RNAs showed that the hpRNA-derived siRNAs were characteristic of Argonaute-binding small RNAs. Our results indicate that RNAi mutant library is a high-efficient approach for genomewide gene identification in plants.

Keywords: Rice, RNAi, gene silencing, siRNA, library, hairpin RNA.

Background

Forward and reverse genetics are powerful approaches for identifying the components involved in various biological processes in plants. With the completion of genomic sequencing for plants such as *Arabidopsis*, rice, maize and soybean, deciphering the function of predicted genes, identifying novel genes and understanding the molecular basis of important agronomic traits have become a major focus. This has driven the concurrent development of large libraries of mutant lines through chemical, physical or insertional mutagenesis (Bolle *et al.*, 2011; Chang *et al.*, 2012; Wang *et al.*, 2013). T-DNA insertion has been the most successful approach for generating mutagenized populations suitable for genetic studies. T-DNA insertion libraries with millions of mutants have been generated in *Arabidopsis* and rice, and these mutant lines are widely used for gene function studies (Ahn *et al.*, 2007; Bolle *et al.*, 2011; Wang *et al.*, 2013). Transposon insertion as well as physical (e.g. γ -radiation) or chemical (e.g. ethyl methanesulphonate and sodium azide) treatments has also been successfully used in mutagenesis studies (Kolesnik *et al.*, 2004; Piffanelli *et al.*, 2007; Till *et al.*, 2007). However, all of these techniques are limited by inherent imperfections. For example, a large mutagenized population must be generated to ensure sufficient genomewide coverage of genes, as mutations occur randomly and often at intergenic and non-coding regions (http://signal.salk.edu/Source/AtTOME_Data_Source.html). Many genes exist as multigene families, and obtaining mutants in which all members of a gene family are mutated is difficult; thus, scientists must commonly use mutant crossing to achieve multiple mutations, but obtaining mutations in essential genes is problematic as a loss of function is lethal to the plant. Additionally, the mapping of mutations induced by chemical or

physical means is tedious and expensive, although recent advances in high-throughput genome resequencing have made this easier than before (Austin *et al.*, 2011).

RNA silencing is a eukaryotic gene repression mechanism induced by double-stranded (ds) or hairpin RNA (hpRNA). dsRNAs or hpRNAs are processed by Dicer or Dicer-like (DCL) proteins, generating small interfering RNAs (siRNAs) 20–24 nt in length (Axtell, 2013; Jones-Rhoades *et al.*, 2006; Wang *et al.*, 2012). These siRNAs are loaded onto Argonaute (AGO) proteins to form RNA-induced silencing complexes (RISCs), which cleave the cognate target mRNA (Jones-Rhoades *et al.*, 2006). RNAi has been successfully exploited as a tool in gene function analyses. Short hairpin RNA (shRNA) constructs are particularly effective at inducing silencing in mammalian cells (Shirane *et al.*, 2004), promoting the construction of many shRNA libraries for genomewide gene function analyses. In plants, however, shRNA constructs have not proven to be highly effective; instead, long hpRNA transgenes have been widely used to induce gene silencing or confer virus resistance (Mao *et al.*, 2007; Wang and Waterhouse, 2002), with the frequency of silencing often reaching >70% in a transgenic population (Smith *et al.*, 2000). Therefore, the transformation of plants with a library of long hpRNA constructs targeting all transcripts is likely to generate populations of gene knockout or knockdown lines suited for genomewide gene function analyses.

The rolling circle amplification-mediated hpRNA (RMHR) construction technique permits the rapid, efficient and inexpensive preparation of genomewide long hpRNA expression libraries (Wang *et al.*, 2008). In the present study, we optimized our RMHR system and used it to generate a long hpRNA library targeting rice genes. Subsequently, this long hpRNA library was

transformed into rice to generate RNAi mutant library, resulting in a large number of transgenic lines containing hpRNAs against various types of genes and showing various phenotypic changes in comparison with wild-type (WT) plants. This RNAi mutant library therefore provides a high-efficiency tool for genomewide gene function analyses in rice.

Results

Construction of rice hpRNA libraries

Construction of intermediate rice cDNA libraries

Total RNA was extracted separately from various rice tissues (see Materials and Methods) and then mixed in equal amount for cDNA library construction, aiming to cover most rice transcripts. To minimize problems associated with the preferential ligation of shorter cDNA fragments over longer ones, cDNAs 200–400, 400–600 and 600–1000 bp in size were gel-fractionated (Figure 1a) and ligated separately into the vector pBsa2T.

pBsa2T is a modified version of pBsa, which was described previously for hpRNA library construction using the RMHR system (Wang *et al.*, 2008). In addition to two *Bsa*I sites for generating cDNA inserts with asymmetrical termini to prevent self-ligation, pBsa2T contains two *Ahd*I sites separated by the *ccdB* lethal gene sequence (Figure 1b). *Ahd*I digestion creates two 3'-T overhangs, converting pBsa2T into a 'T' vector, which is convenient for cDNA cloning. Using the T-vector ligation strategy, three cDNA libraries were generated: pOs2, pOs4 and pOs6. pOs2 contained 7.2×10^5 clones with inserts that were 200–400 bp in length, pOs4 contained 5.3×10^5 clones with inserts that were 400–600 bp in length, and pOs6 contained 2.0×10^5 clones with inserts that were 600–1000 bp in length. Sequencing of 108

randomly selected clones from the three libraries (36 clones per library) showed that they contained sequences of 92 different genes, indicating that the libraries were of high quality with relatively low sequence redundancy and suitable for hpRNA library construction.

Construction of rice hpRNA libraries using rolling circle amplification

To create circular cDNA suitable for rolling circle amplification by Phi29 polymerase, the three cDNA libraries were digested with *Bsa*I, and the cDNA inserts were ligated with two DNA oligos (mini-hairpin 1 and mini-hairpin 2) that can self-anneal to form hairpin DNA with adhesive termini matching the two different *Bsa*I termini of pBsa2T, respectively (Figure 2a). Mini-hairpin 2 contains a 50-nt spliceable intron sequence from Os01g62100, which serves as a spacer between the sense and antisense sequences of the resulting inverted repeats to stabilize the hpRNA DNA clones in bacteria. The ligation product was amplified by rolling circle amplification to generate multimeric inverted repeat DNA (Figure 2b). This amplification product was digested with *Bam*HI and *Sac*I to release single inverted repeats (Figure 2c), which were inserted downstream of the Ubi promoter in p35S-Ubi, to form three hpRNA libraries, named OsHSP2, OsHSP4 and OsHSP6, respectively. p35S-Ubi contains a 35S-GUS:GFP cassette (Figure 2d), which drives GUS and GFP expression in transgenic cells, allowing for the quick screening of positive transgenic lines.

Quality check of the hpRNA libraries

In total, 72 clones from the three libraries (24 clones per library) were randomly selected and analysed by restriction digestion, and all of them contained inverted repeat DNA. In total, OsHSP2, OsHSP4 and OsHSP6 contained $\sim 1 \times 10^5$, 5×10^5 and 7×10^5 clones, respectively. We further examined the quality of the three hpRNA libraries by sequencing a subset of randomly selected clones. Of 833 clones from the OsHSP2 library, 793 (95.1%) had an inverted repeat sequence >100 bp in size (with an average size of 250 bp), covering sequences from 434 genes or gene families. For OsHSP4, 1367 of the 1500 clones (91.1%) sequenced had inverted repeat sequences >100 bp in length corresponding to 786 genes or gene families (with an average size of 450 bp). For OsHSP6, 1305 of the 1366 clones (95.5%) sequenced had inverted repeat sequences >200 bp in length from 662 genes or gene families (with an average size of 700 bp). The genomic locations of these target genes were distributed relatively evenly across the 12 rice chromosomes (Figure S1); these genes represent a diverse range of functional categories, including biological process, cellular component and molecular function. This finding indicates that the constructed OsHP libraries were of high quality with a good level of gene coverage.

Generation of the transgenic hpRNA population

For plant transformation, the three hpRNA libraries were introduced into *Agrobacterium tumefaciens* EHA105 by electroporation, generating three corresponding *Agrobacterium* libraries (Ag-OsHSP2, Ag-OsHSP4 and Ag-OsHSP6), each containing more than 10^6 clones. The Ag-OsHSP4 library was used to transform rice plants to produce a transgenic rice hpRNA population.

More than 6000 transgenic lines were obtained. Histochemical staining of leaves showed that about 3000 transgenic lines were GUS positive (Figure S2a). The hpRNA sequences in the GUS-

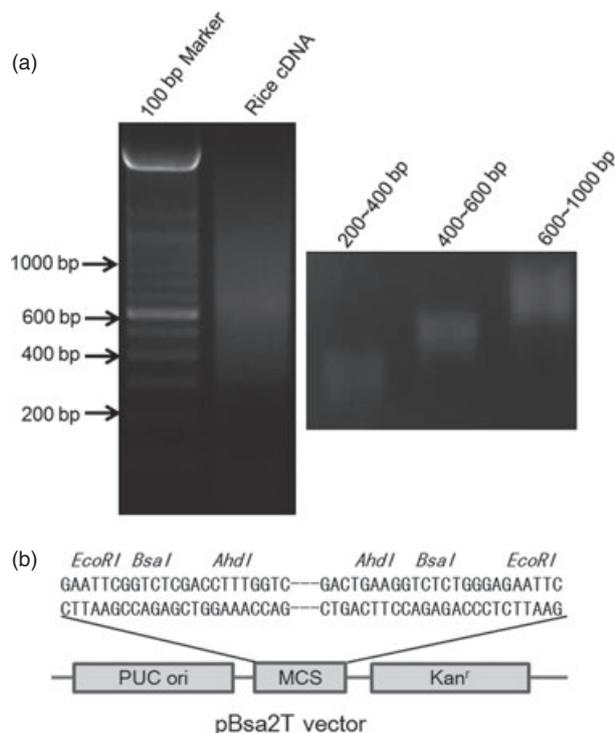


Figure 1 cDNA library construction. (a) Rice cDNAs were gel-purified and used for cDNA library construction. (b) pBsa2T can be converted to linearized, T-ended vector by *Ahd*I digestion.

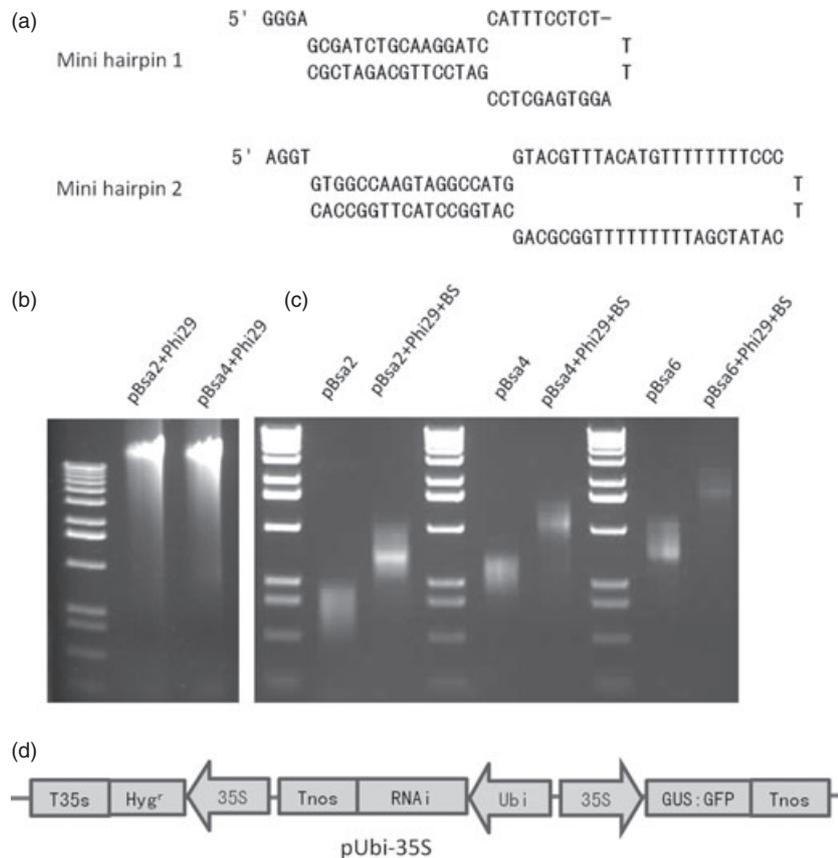


Figure 2 Construction of the hpRNA library. (a) The predicted secondary structures of the mini-hairpin 1 and mini-hairpin 2 oligonucleotides. (b) The closed circle ligation product was amplified with Phi29 DNA polymerase, producing large DNA fragments with multiple-unit inverted repeats of cDNA. (c) The amplified product was digested with *Bam*HI and *Sac*I (BS), producing a pool of single-unit inverted repeats; note that the size of the digested amplification product is approximately double that of the respective unamplified control, indicating the formation of inverted repeats. pBsa2, pBsa4 and pBsa6 represent the insert from the respective cDNA libraries. In pBsa2+Phi29+BS, the ligation product was amplified with Phi29 and digested with *Bam*HI and *Sac*I. (d) Map of the pUbi-35S vector used for hpRNA expression in rice.

positive plants were subsequently amplified by polymerase chain reaction (PCR) and sequenced, demonstrating that the hpRNA expression cassette was integrated into the rice plants (see Figure S2b for an example).

More than 30% of the GUS-positive lines showed poor growth and/or sterility. We obtained seed from about 1000 T₀ lines in ~1500 GUS-positive survival lines. T₁ seed from 211 lines was planted in the field and their phenotypes assessed. A significant proportion (~50%) of the T₁ plants exhibited various abnormal phenotypes, including reduced or increased stature; discoloration or enhanced green colour of leaves; and changes in leaf shape, tiller number or fertility (Table S2). Other phenotypes, including alterations in grain size, spikes and flowering time, as well as an altered stress response, were observed. The severities of the phenotypes varied among the transgenic lines, with the majority showing a relatively mild phenotype (Figure 3a). Many of the hpRNA lines displayed more than one phenotype, consistent with a previous report (Larmande *et al.*, 2008), suggesting that the target genes function in more than one biological process. qRT-PCR corresponding to the target genes in 8 RNAi lines were investigated, which confirmed that the level of the target mRNA was dramatically reduced (Figure 3b).

The proportion of hpRNA lines with an observable phenotype (Table S2; 47.9%) was greater than that of a typical T-DNA insertion lines (3.5%), in which most mutations occur in

intergenic (35.4%) and 500-bp upstream (17.51%) regions (Wang *et al.*, 2013). This suggests that the hpRNA library effectively silenced the target genes. It also implies that this type of silencing often does not lead to the complete knockout of target genes, allowing for the recovery of phenotypes associated with essential genes that may not be recoverable using chemical, physical or insertional mutagenesis. Furthermore, compared with the mutants produced via these approaches, the target gene sequences of hpRNA lines can be easily identified by amplification of the hpRNA sequence through PCR (e.g. Figure S2b). Thus, our RNAi mutant library is potentially a powerful tool for gene identification in rice.

siRNAs are processed from hpRNAs to direct target gene silencing

hairpin RNAs are processed by DCLs to generate predominantly 21-nt siRNAs, which in turn guide the degradation of target mRNAs (Jones-Rhoades *et al.*, 2006; Watanabe, 2011). To determine whether siRNAs were produced in the rice hpRNA lines, we first analysed seven transgenic lines by small RNA Northern blot hybridization. Five of the plants showed a clear band ~21 nt in size, indicating that hpRNA was expressed and processed in these plants (Figure 4a). As expected, the abundance of the 21-nt band was variable among the different plants, reflecting different hpRNA expression levels among the transgenic lines.

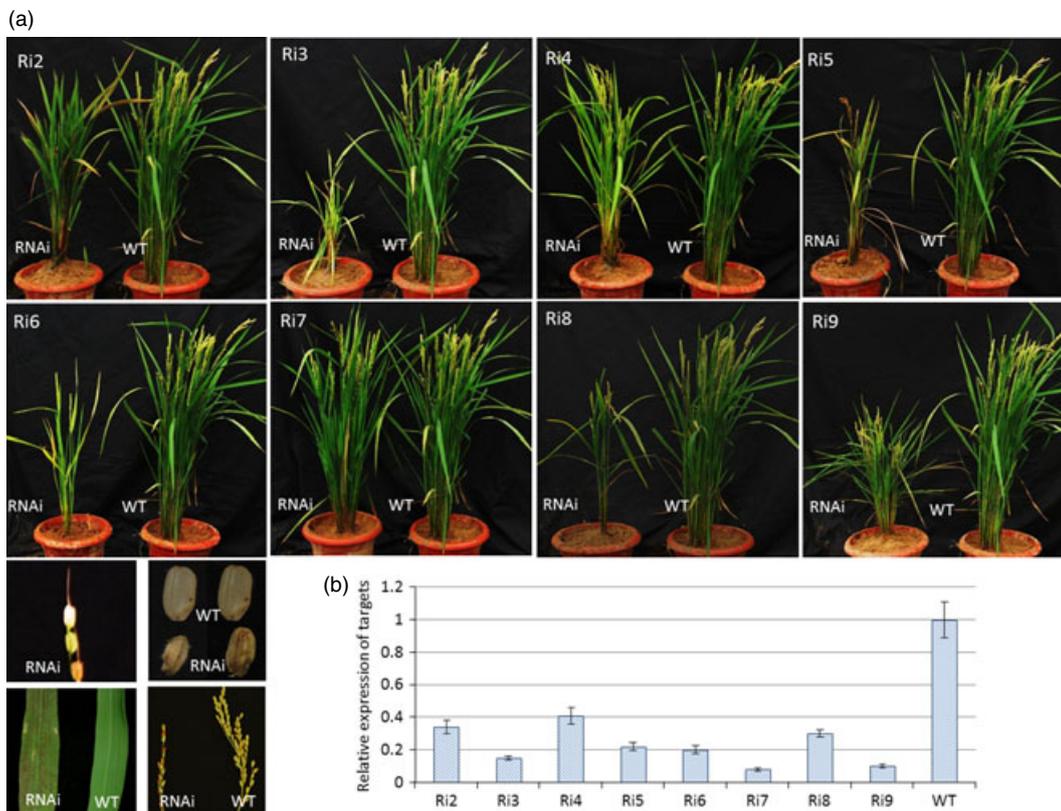


Figure 3 Representative phenotypes of the hpRNA transgenic plants. (a). Ri2–Ri9 are 3-month-old T1 hpRNA plants. Ri2, dark green and sterile; Ri3, pale green and dwarf; Ri4, pale green with reduced tillering; Ri5, dwarf with reduced tillering; Ri6, pale green dwarf with reduced tillering; Ri7, increased tillering; Ri8, dwarf with reduced tillering; Ri9, increased leaf angle and dwarf. The bottom left corner shows other observed phenotypes, including reduced seed number and spike size, leaf lesions and changed awn length. (b) qRT-PCR analysis of target gene mRNA level in the 8 hpRNA lines shown in (a).

To investigate whether the expression of the target genes was repressed, we analysed 12 hpRNA lines, including the seven plants in (Figure 4a), using real-time reverse transcription (RT)-PCR. As shown in Figure 4b, the mRNA levels were clearly down-regulated in 10 of the plants. The target genes in lines 7-436 and 8-454 were not down-regulated, and correspondingly, siRNA was not detected in these two plants (Figure 4a). Thus, siRNAs from the hpRNAs were directly responsible for the silencing of the genes.

Distribution pattern of hpRNA-derived siRNAs

We investigated the distribution pattern of the hpRNA-derived siRNAs via deep sequencing of small RNAs isolated and pooled from seven hpRNA lines (RNAi #22, #25, #29, #30, #33, #46 and #47). We obtained a total of ~51 million reads of 18- to 30-nt sequences, which contained ~1.7 million reads derived specifically from the hpRNA transgenes.

Size distribution pattern

The size distribution of the total small RNA population was as described previously (Sunkar *et al.*, 2005), with the 24-nt class being the most abundant followed by the 21-, 22- and 23-nt classes (Figure 5a). The size distribution pattern of hpRNA-derived siRNAs was distinct from that of the total small RNAs (Figure 5b), with 21-nt siRNAs being the most dominant (~57%), followed by the 22-nt (~27%) and 24-nt (~9.7%) classes. This is consistent with previous reports showing that hpRNAs are processed primarily by DCL4, but also by DCL2 and DCL3 (Fusaro *et al.*,

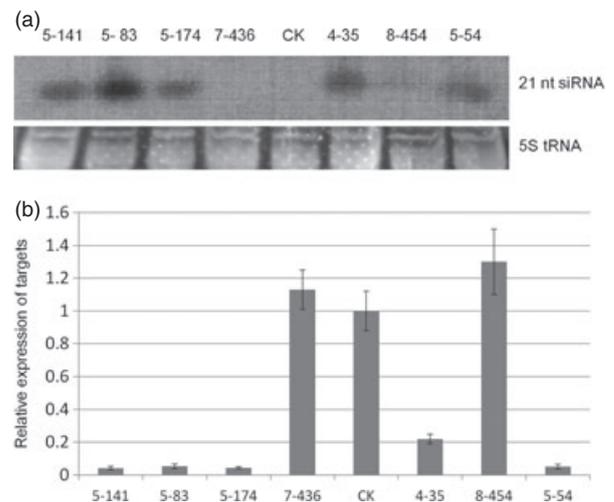


Figure 4 Northern blot analysis of small RNA accumulation (a) and qRT-PCR analysis of target gene expression (b) in transgenic hpRNA plants. CK, plant transformed with empty vector.

2006). In contrast to the total small RNA population, which contained substantial amounts of 18- to 20-, 23- and >25-nt RNA species, the abundance of hpRNA-derived small RNAs belonging to these size classes was low. This indicates that these size classes of small RNAs were not the product of *in vitro* RNA degradation, which may occur during RNA extraction or deep sequencing;

instead, they were produced by *in vivo* RNase processing. Examining whether these small RNA size classes are functionally important would be interesting, particularly the 20- and 23-nt classes, which are always present in significant amounts in deep sequencing data.

Distribution of siRNAs along the hpRNAs

siRNAs derived from all seven hpRNA lines overlapped with the dsRNA arm of the hpRNA, with few siRNAs corresponding to the loop region (Figure 6a). The siRNAs were unevenly distributed along the dsRNA arm, with some regions showing high abundance (siRNA hot spot) and others with low levels of siRNAs (Figure 6a). The sense siRNA hot spots often did not overlap with the antisense siRNA hot spots (Figure 6a), suggesting that these hot spots were not created by the preferential Dicer processing of specific dsRNA regions, which would otherwise generate hot spots with equal amounts of sense

and antisense siRNAs (Figure 6b). The nucleotide composition analysis showed the 20- to 22-nt siRNAs were enriched for uridine (U) at the 5' terminus, while the 24-nt siRNAs were enriched for adenine (A) at the 5' terminus, which is typical of small RNAs associated with AGO1 and AGO4, respectively (Wei *et al.*, 2012). This suggests that the siRNA hot spots resulted from preferential binding with AGO proteins, which may have stabilized the siRNAs.

For hpRNA lines 46, 29 and 30, the numbers of siRNAs with sense and antisense polarities were close in number. However, for the other four lines, siRNAs with a sense polarity were more abundant than those with an antisense polarity (Figure 6c). This was not due to the sequence orientation of the inverted repeats in the hpRNA constructs, of which some had a sense orientation in the front near the promoter, while others had an antisense orientation in the front (Figure 6b). A possible explanation for this is that the presence of the target mRNA increased the turnover

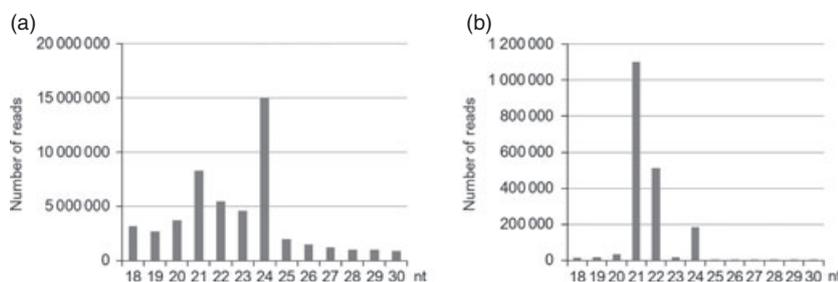


Figure 5 Size distribution of the 18- to 30-nt small RNA population (a) and hpRNA-derived small RNAs (b) in transgenic hpRNA lines obtained using deep sequencing.

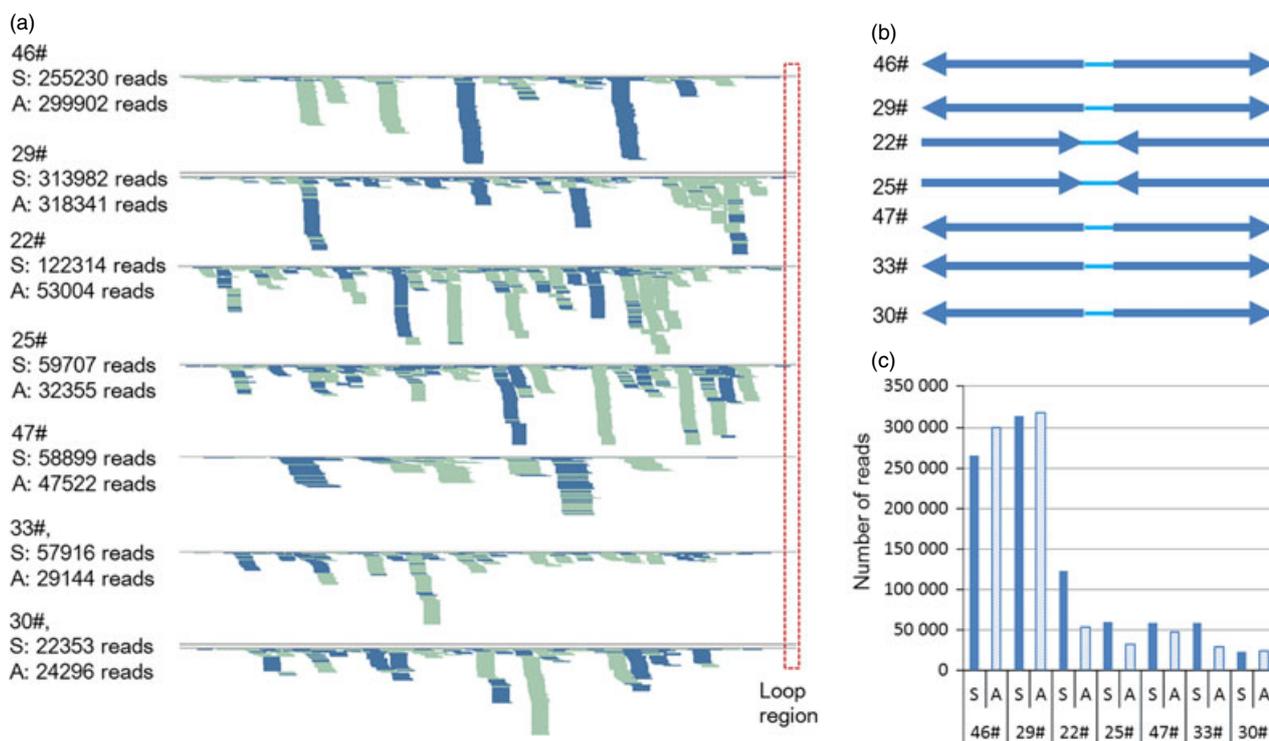


Figure 6 The siRNAs were unevenly distributed along the dsRNA arm of the hpRNA in seven independent hpRNA lines (a). The green lines represent siRNAs with a sense polarity, while the blue lines represent siRNAs with an antisense polarity that have the potential to silence target genes. The loop region, which is framed by a red dashed line, contains a small number of siRNAs. The sequence orientation of the inverted repeats in the seven hpRNA lines analysed was investigated (b). The read numbers of the sense and antisense siRNAs were calculated (c). S, sense siRNA; A, antisense siRNA.

rate of the corresponding small RNAs (i.e. the antisense siRNAs), resulting in the overrepresentation of sense siRNAs. Consistent with this, a recent study on human cells showed that the miRNA decay rate was dramatically enhanced by the presence of highly expressed target genes (Baccarini *et al.*, 2011).

Silencing of genes in multigene families

hpRNA lines 4-68 and 7-120 contained the same hpRNA construct with a 698-bp sequence derived from the open reading frame (ORF) of Os08g0430500, a member of a four-membered gene family that includes Os0280580300, Os11g0546900 and Os03g0710800. The relative nucleotide sequence identities of the ORFs between Os08g0430500 and the other three members are 78.3%, 70.5% and 76.3%, respectively. A close look at the sequence identities revealed perfectly matched sequences between Os08g0430500 and Os0280580300 (two sequences, >21 bp in length [22 and 32 bp]), Os03g0710800 (one sequence, 23 bp in length) and Os11g0546900 (one sequence, 29 bp in length; Figure S3).

The expression level of all four members was investigated in the two hpRNA lines. Our results show that Os08g0430500, Os0280580300 and Os11g0546900 were down-regulated at different levels in the two hpRNA lines (Figure 7a), indicating that the hpRNA derived from Os08g0430500 induced silencing of itself and the related family members Os0280580300 and Os11g0546900. However, the remaining member, Os03g0710800, was not down-regulated. Instead, it was significantly up-regulated (Figure 7a). A possible explanation for this is that the hpRNA did not target Os11g0546900 mRNA effectively due to the existence of only a short stretch (23 bp) of a perfectly matched sequence and that its transcription may be up-regulated in the absence of the other three members.

The hpRNA in line 22 (Figure 8) matched perfectly with the sequence of Os04g44924 from 1 to 748 nt (Figure S4), which is conserved in the four-membered short-chain dehydrogenase/reductase family. The relative nucleotide identities in this region between Os04g44924 and the other three members of the gene family (Os04g44950, Os04g45000 and Os04g44980) are 88.6%, 82% and 66%, respectively. Although many perfectly matched sequences >21 nt in length were identified (Figure S4), only

Os04g44924 was significantly down-regulated. An examination of the small RNA deep sequencing data showed that a total of 157 700 reads mapped to the Os04g44924 transcript, with 20% of these siRNAs (~33 500) being antisense siRNAs with the potential to direct mRNA silencing. In contrast, a much smaller number of antisense siRNAs matched with the Os04g44950 (~6500 reads), Os04g45000 (~900 reads), but this gene showed no expression in leaves) and Os04g44980 (15 reads) transcripts. Thus, for the Os04g44924 gene family, effective silencing may require a threshold level of siRNAs matching the respective transcripts.

Silencing of key genes

Obtaining knockout mutants of some key genes using chemical, physical and insertional mutagenesis is problematic due to the strong deleterious and often lethal phenotypes of the homozygous mutants. In analysing our hpRNA transgenic population, we identified some lines in which key genes for development were targeted. As expected, these plants tended to show severe phenotypes; however, they survived, allowing phenotypic analyses that provided clues as to their function.

hpRNA lines 5-141 and 5-174 contained the same hpRNA construct targeting Os08g0558600, which is homologous to a protein essential for seed maturation (vesicle-associated membrane protein 727) (Ebine *et al.*, 2008). The expression of the gene was dramatically down-regulated in both of the lines (Figure 4b). Correspondingly, the plants showed similarly strong phenotypes, including a reduced tiller number with no heading. Although the plants yielded no seed, they survived in the greenhouse for more than 6 months and could be propagated clonally, allowing for phenotypic and molecular analyses.

The hpRNA in lines 4-68 and 7-120 targeted the four-membered 14-3-3 gene family, resulting in strong repression of three family members (Figure 7a). These hpRNA lines showed strong phenotypes, including the inability to sprout a head, diminished tiller number and infertility. In plants, 14-3-3 proteins have been compared with spiders in a web of phosphorylation (De Boer *et al.*, 2013). These phenotypes suggest that the 14-3-3 gene family is required for rice growth and development.

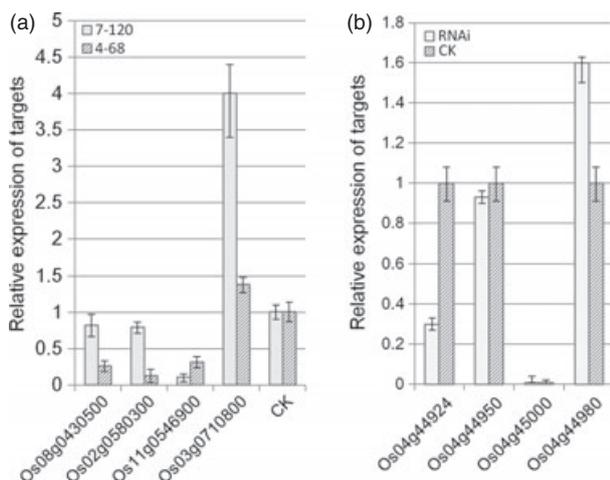


Figure 7 qRT-PCR analysis of target gene expression in the hpRNA lines. (a) The expression of a four-membered gene family was investigated in two RNAi plants. (b) Expression of the short-chain dehydrogenase/reductase family, which consists of four members, was analysed.

Discussion

Long hpRNA technology is effective at inducing gene silencing in plants and hence is useful in gene function studies; however, genomewide gene function analyses require a large number of hpRNA constructs targeting >25 000 genes in a typical plant species. Such a large number is difficult to produce using conventional cloning methods involving the PCR amplification of each specific gene sequence followed by single or multistep DNA ligation (Smith *et al.*, 2000). We therefore developed the RMHR method, which can be used to produce a large number of hpRNA constructs simultaneously from cDNA libraries. Using RMHR, we successfully generated a long hpRNA library from an *Arabidopsis* cDNA population containing known and unknown genes (Wang *et al.*, 2008). To make the RMHR system more efficient, in this study, we introduced two *AhdI* restriction sites into the intermediate cDNA cloning vector pBsa2T, turning it into a T-ended vector, and a *ccdB* lethal gene between the two *AhdI* sites to prevent the recovery of insert-free *Escherichia coli* clones. This improved vector, called pBsa2T, was highly efficient for cDNA ligation. We also found that the use of loop-specific primers (20 nt) for the rolling circle amplification reaction

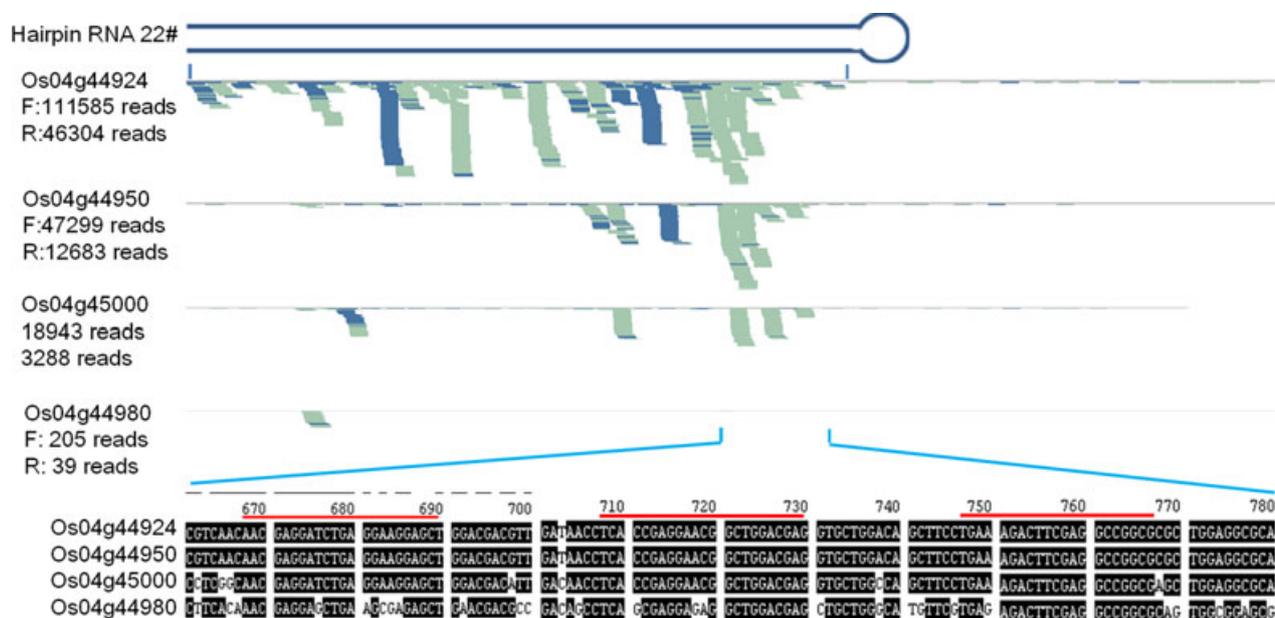


Figure 8 siRNAs derived from the hpRNA of the Os04g44924 sequence in line #22 partially match the other three members of the gene family. The red lines above the sequence alignment on the bottom indicate the three siRNA-rich regions in the four members.

eliminated the production of noninverted repeats, which sometimes occurs when 6-nt random primers are used.

Using this RMHR system, we generated a long hpRNA library with good coverage of rice genes and produced thousands of transgenic hpRNA lines. More than 50% of these transgenic lines displayed visible phenotypes, which is a much higher proportion than that reported for a typical T-DNA insertion or EMS-mutagenized population, of which only ~3.5% usually show visible phenotypes (Hirochika *et al.*, 2004). This result is likely due to several factors. First, all of the hpRNA constructs targeted exon sequences because they were derived from cDNA libraries; this is in contrast to T-DNA insertions or EMS mutations, of which the vast majority occurs in nonexon sequences. Second, a hpRNA derived from a single member of a multigene family may direct the silencing of all or most members, as was the case for the Os08g0430500 family; this resulted in strong phenotypic changes, but such multigene repression is unlikely to occur through T-DNA or EMS mutagenesis. Furthermore, hpRNA-induced gene silencing may be partial in some members of the transgenic population, allowing for the recovery of hpRNA lines in which housekeeping or lethal genes are targeted. The hpRNA constructs used in this study were driven by the maize ubiquitin promoter, which is a strong constitutive promoter. A tissue-specific or inducible promoter could be used in future studies to further improve the recovery of hpRNA lines targeting housekeeping genes.

However, it is worth noting that although hpRNA-induced RNAi is effective for knocking down gene expression, a small amount of targeted mRNA is expected to exist in transgenic cells due to the nature of post-transcriptional gene silencing. Therefore, RNAi may not be suitable for disrupting the function of genes, which requires only a low level of mRNA accumulation. Furthermore, as the use of hpRNA libraries on tissue culture and plant transformation, care is needed to distinguish hpRNA-induced phenotypes from those of tissue culture-induced somaclonal variation (Jeong *et al.*, 2002). In the current study,

we observed dominant mutant phenotypes not only in transgenic hpRNA lines but sometimes also in untransformed lines regenerated through tissue culture. However, somaclonal variations can be differentiated from hpRNA-induced phenotypes through examination of cosegregation between a specific phenotype and a hpRNA transgene in a segregating transgenic population.

Experimental procedures

Construction of pBsa2T and pUbi-35S

To produce pBsa2T, two oligonucleotides were synthesized: Bsa-Ahd Fw (5'-GGGAATTCGGTCTCGACCTTTGGTCCAAGCTTAAATGGTTACTAAAAGCCAGA) and Bsa-Ahd Rv (5'-CCGAATTCTCCAGAGACCTTCAGTCTATATCCCCAGAACAT). Polymerase chain reaction was performed in a total volume of 50 μ L consisting of 1 \times PCR buffer, 0.2 mM dNTPs, 0.25 μ M each primer, 2 U of Taq DNA polymerase and 1 ng of pENTR1A, using pENTR1A (Qiagen, Hilden, Germany) as template. The amplification programme consisted of 1 min at 95 $^{\circ}$ C, 25 cycles of 30 s at 94 $^{\circ}$ C, 30 s at 56 $^{\circ}$ C and 30 s at 72 $^{\circ}$ C, followed by 10 min at 72 $^{\circ}$ C. The product was digested with *Eco*RI and cloned into the *Eco*RI site of pENTR1A to form the intermediate vector pBsa2T. For cDNA cloning, pBsa2T was digested with *Ahd*I, and the larger fragment was gel-purified, generating linearized pBsa2T with T overhangs.

To prepare pUbi-35S, pBI121 was digested with *Hind*III and *Eco*RI, and the 35S-GUS-Nos cassette was gel-purified and inserted into the *Hind*III and *Eco*RI sites in pCambia1303, generating pC-35S. The Ubi-GUS expression cassette in pAHC25 was excised by *Hind*III and *Sac*I digestion and inserted into the same sites in pC-35S, forming pUbi-35S.

Rice cDNA library construction

Rice plants (*Oryza sativa* cv. Nipponbare) were grown in the field. Various tissues were collected, including whole seedlings, roots,

stems and leaves at the tillering stage; young panicles and stems at the heading stage; panicles at the flowering stage; and anthers and panicles at the filling stage and at the yellow ripening stage. Total RNA was extracted separately from the tissues using TRIzol reagent (Invitrogen, Carlsbad, CA) and mixed in equal amounts. mRNA was isolated from the total RNA using PolyATtract mRNA Isolation System III (Promega, Madison, WI).

A cDNA library was constructed following the SMART cDNA library manual (Clontech Laboratories Inc., Mountain View, CA) with the following exceptions: LD PCR products were digested with *Bam*HI, *Sac*I, *Bsa*I and *Xba*I and then treated with T4 DNA polymerase to produce blunt ends. An 'A' tail was added to the 3' terminus using Taq DNA polymerase. The DNA was fractionated using preparative DNA gel electrophoresis after digestion, and fragments 200–400, 400–600 and 600–1000 bp in length were purified and ligated separately into linearized pBsa2T. Ligations with pBsa2T were performed immediately following cDNA size fractionation. Three libraries, pOs2, pOs4 and pOs6, were generated with 1.45×10^6 clones in total.

hpRNA library construction

To generate closed circle cDNA, the pOs2, pOs4 and pOs6 cDNA libraries were digested with *Bsa*I, and the insert was gel-purified and ligated separately with the mini-hairpin 1 (5'-GGGAGCGATCTGCAAGGATCCATTTCTTCTTTAGGTGAGCTCCGATCCTTGCCAGATCGC) and mini-hairpin 2 (5'-GTGGCCAAGTAGGCCATGCTGCCCAAAAAAAAAATCGATATGAAGGGAAAAAACATGTAACGTACCATGGCTACTTGGCCACACCT) oligonucleotides, which were preannealed and phosphorylated. The circular DNAs were amplified by rolling circle amplification using loop-specific primers (F: 5'-GGAAAAAACATGTAAACG and R: 5'-TTCATATCGATTTTTTTGG), yielding linear concatemers of inverted repeat DNA (Wang *et al.*, 2008). Following digestion with *Bam*HI and *Sac*I, purified fragments from the three libraries were inserted into pUbi-35S, generating three hpRNA libraries, OsHP2, OsHP4 and OsHP6.

Rice transformation

Rice (*O. sativa* L. cv. Kitaake) was transformed using seed-derived callus tissue based on the protocol of Hiei *et al.* (Hiei and Komari, 2008) with modifications. Surface-sterilized seeds were germinated and cultured on NBI medium (NB + 2 mg/L 2,4-D; PhytoTechnology Laboratories, Shawnee Mission, KS). After 7 days of incubation at 30 °C in the dark, the calli were transferred to fresh NBI medium and cultured for 25–30 days at 30 °C. The calli were then harvested and cultured on fresh NBI medium for another 3–4 days. The calli were then incubated for 20 min in *A. tumefaciens* resuspended ($OD_{600} = 0.3–0.4$) in liquid NBASS medium (NBI + 200 μ M acetosyringone + 10 g/L sucrose), dried on sterile filter paper and cocultivated with *A. tumefaciens* for 3 days at 25 °C on NBASS medium. After cocultivation, the rice calli were rinsed thoroughly (four times) in sterile water containing 50 mg/L timentin and transferred to selective medium (NB + 40 mg/L hygromycin-B, 500 mg/L cefotaxime and 2 mg/L 2,4-D) for 4 weeks at 28 °C. Vigorously growing callus pieces 0.5–1 mm in size were transferred to regeneration medium (NB + 0.5 mg/L NAA, 4 mg/L 6-BA and 40 mg/L hygromycin-B) for 2–3 weeks at 28 °C under 5000 Lux of light. Shoots arising from the calli on regeneration medium were transferred to rooting media ($\frac{1}{2}$ MS salt + MS vitamin solution + 30 g/L sucrose and 40 mg/L hygromycin-B). Once roots formed, the plants were transferred to soil.

Identification of hpRNA transformants

Putative transgenic plants were screened first by GUS histochemical staining (Gao *et al.*, 2010). GUS-positive plants were further analysed by PCR and sequencing using primers to amplify the Nos terminator and Ubi promoter sequences, respectively: OsPin fw1 (5'-TGTAACGTACCATGGCCTAC) and Nos60 (5'-CAACAG GAT TCA ATC TTA AGA AAC, to amplify the Nos terminator sequence) and OsPin rv1 (5'-CATATCGATTTTTTTTGGCG) and OsUbiFw (5'-TATGCAGCAGCTATATGTGG for the ubiquitin promoter sequence).

Small RNA Northern blot hybridization

Northern blot hybridization was performed based on a published protocol (Wang *et al.*, 2007) with slight modifications. Total RNA was extracted from leaves (~400 mg) using TRIzol reagent (Invitrogen) and dissolved in 20 μ L of sterile water; 15 μ g was separated in a 17% denaturing polyacrylamide gel and then transferred to a Hybond+ membrane. About 30 ng of DNA for each probe was labelled with 32 P-dATP for ~15 h at room temperature using a Prime-a-Gene Kit (Promega). Church buffer was used for hybridization at 40 °C overnight.

Real-time PCR

Total RNA was treated with DNase I, and cDNA was synthesized using a reverse transcription system (Promega) following the manufacturer's instructions. qRT-PCR was performed using an Applied Biosystems Prism 7500 analyzer and GoTaq qPCR Master Mix (Promega). For each plant, three independent biological replicates were analysed. Sample comparisons were performed using the $2^{-\Delta\Delta CT}$ method (Xu *et al.*, 2012). The primers used are listed in Table S1.

Small RNA deep sequencing and bioinformatic analysis

Total RNAs were extracted from the leaves of transgenic plants. Small RNAs were then enriched as described previously (Wang *et al.*, 2007). Small RNA sequence reads were generated using an Illumina Genome Analyzer II (Illumina Inc., San Diego, CA). The raw data were preprocessed using the NGSQC Toolkit (<http://59.163.192.90:8080/ngsqctoolkit/>) pipeline to remove low-quality (Patel and Jain, 2012) and contaminated reads and to clip adapter sequences. The clean reads were then filtered for siRNA analysis (Hardcastle *et al.*, 2012). Small RNAs ranging from 18 to 30 nt in length were collected and mapped to the hpRNA sequences using Bowtie 2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>) (Langmead and Salzberg, 2012). The analysed data are presented using Tablet (<http://bioinf.scri.ac.uk/tablet/>) (Milne *et al.*, 2010).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (grant number 31271801) and National High Technology Research and Development Program of China (grant numbers 2012AA10A306) and the National Special Program for Genetically Modified Organisms Development (grant number 2009ZX08009-16B).

References

- Ahn, J.H., Kim, J., Yoo, S.J., Yoo, S.Y., Roh, H., Choi, J.H., Choi, M.S., Chung, K.S., Han, E.J., Hong, S.M., Jung, S.H., Kang, H.J., Kim, B.K., Kim, M.D., Kim, Y.K., Kim, Y.H., Lee, H., Park, S.H., Yang, J.H., Yang, J.W., Yoo, D.H., Yoo,

- S.K. and Lee, J.S. (2007) Isolation of 151 mutants that have developmental defects from T-DNA tagging. *Plant Cell Physiol.* **48**, 169–178.
- Austin, R.S., Vidaurre, D., Stamatiou, G., Breit, R., Provart, N.J., Bonetta, D., Zhang, J., Fung, P., Gong, Y., Wang, P.W., McCourt, P. and Guttman, D.S. (2011) Next-generation mapping of *Arabidopsis* genes. *Plant J.* **67**, 715–725.
- Axtell, M.J. (2013) Classification and Comparison of Small RNAs from Plants. *Annu. Rev. Plant Biol.* **64**, 137–159.
- Baccarini, A., Chauhan, H., Gardner, T.J., Jayaprakash, A.D., Sachidanandam, R. and Brown, B.D. (2011) Kinetic analysis reveals the fate of a microRNA following target regulation in mammalian cells. *Curr. Biol.* **21**, 369–376.
- Bolle, C., Schneider, A. and Leister, D. (2011) Perspectives on systematic analyses of gene function in *Arabidopsis thaliana*: new tools, topics and trends. *Curr. Genomics*, **12**, 1–14.
- Chang, Y., Long, T. and Wu, C. (2012) Effort and contribution of T-DNA insertion mutant library for rice functional genomics research in China: review and perspective. *J. Integr. Plant Biol.* **54**, 953–966.
- De Boer, A.H., van Kleeff, P.J. and Gao, J. (2013) Plant 14-3-3 proteins as spiders in a web of phosphorylation. *Protoplasma*, **250**, 425–440.
- Ebine, K., Okatani, Y., Uemura, T., Goh, T., Shoda, K., Niihama, M., Morita, M.T., Spitzer, C., Otegui, M.S., Nakano, A. and Ueda, T. (2008) A SNARE complex unique to seed plants is required for protein storage vacuole biogenesis and seed development of *Arabidopsis thaliana*. *Plant Cell*, **20**, 3006–3021.
- Fusaro, A.F., Matthew, L., Smith, N.A., Curtin, S.J., Dedic-Hagan, J., Ellacott, G.A., Watson, J.M., Wang, M.B., Brosnan, C., Carroll, B.J. and Waterhouse, P.M. (2006) RNA interference-inducing hairpin RNAs in plants act through the viral defence pathway. *EMBO Rep.* **7**, 1168–1175.
- Gao, F., Su, Q., Fan, Y. and Wang, L. (2010) Expression pattern and core region analysis of AtMPK3 promoter in response to environmental stresses. *Sci. China Life Sci.* **53**, 1315–1321.
- Hardcastle, T.J., Kelly, K.A. and Baulcombe, D.C. (2012) Identifying small interfering RNA loci from high-throughput sequencing data. *Bioinformatics*, **28**, 457–463.
- Hiei, Y. and Komari, T. (2008) *Agrobacterium*-mediated transformation of rice using immature embryos or calli induced from mature seed. *Nat. Protoc.* **3**, 824–834.
- Hirochika, H., Guiderdoni, E., An, G., Hsing, Y.I., Eun, M.Y., Han, C.D., Upadhyaya, N., Ramachandran, S., Zhang, Q., Pereira, A., Sundaresan, V. and Leung, H. (2004) Rice mutant resources for gene discovery. *Plant Mol. Biol.* **54**, 325–334.
- Jeong, D.H., An, S., Kang, H.G., Moon, S., Han, J.J., Park, S., Lee, H.S., An, K. and An, G. (2002) T-DNA insertional mutagenesis for activation tagging in rice. *Plant Physiol.* **130**, 1636–1644.
- Jones-Rhoades, M.W., Bartel, D.P. and Bartel, B. (2006) MicroRNAs and their regulatory roles in plants. *Annu. Rev. Plant Biol.* **57**, 19–53.
- Kolesnik, T., Szeverenyi, I., Bachmann, D., Kumar, C.S., Jiang, S., Ramamoorthy, R., Cai, M., Ma, Z.G., Sundaresan, V. and Ramachandran, S. (2004) Establishing an efficient Ac/Ds tagging system in rice: large-scale analysis of Ds flanking sequences. *Plant J.* **37**, 301–314.
- Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
- Larmande, P., Gay, C., Lorieux, M., Perin, C., Bouniol, M., Droc, G., Sallaud, C., Perez, P., Barnola, I., Biderre-Petit, C., Martin, J., Morel, J.B., Johnson, A.A., Bourgis, F., Ghesquiere, A., Ruiz, M., Courtois, B. and Guiderdoni, E. (2008) Oryza Tag Line, a phenotypic mutant database for the Genoplante rice insertion line library. *Nucleic Acids Res.* **36**, D1022–D1027.
- Mao, Y.B., Cai, W.J., Wang, J.W., Hong, G.J., Tao, X.Y., Wang, L.J., Huang, Y.P. and Chen, X.Y. (2007) Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. *Nat. Biotechnol.* **25**, 1307–1313.
- Milne, I., Bayer, M., Cardle, L., Shaw, P., Stephen, G., Wright, F. and Marshall, D. (2010) Tablet–next generation sequence assembly visualization. *Bioinformatics*, **26**, 401–402.
- Patel, R.K. and Jain, M. (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE*, **7**, e30619.
- Piffanelli, P., Droc, G., Mieulet, D., Lanau, N., Bes, M., Bourgeois, E., Rouviere, C., Gavory, F., Cruaud, C., Ghesquiere, A. and Guiderdoni, E. (2007) Large-scale characterization of Tos17 insertion sites in a rice T-DNA mutant library. *Plant Mol. Biol.* **65**, 587–601.
- Shirane, D., Sugao, K., Namiki, S., Tanabe, M., Iino, M. and Hirose, K. (2004) Enzymatic production of RNAi libraries from cDNAs. *Nat. Genet.* **36**, 190–196.
- Smith, N.A., Singh, S.P., Wang, M.B., Stoutjesdijk, P.A., Green, A.G. and Waterhouse, P.M. (2000) Total silencing by intron-spliced hairpin RNAs. *Nature*, **407**, 319–320.
- Sunkar, R., Girke, T. and Zhu, J.K. (2005) Identification and characterization of endogenous small interfering RNAs from rice. *Nucleic Acids Res.* **33**, 4443–4454.
- Till, B.J., Cooper, J., Tai, T.H., Colowit, P., Greene, E.A., Henikoff, S. and Comai, L. (2007) Discovery of chemically induced mutations in rice by TILLING. *BMC Plant Biol.* **7**, 19.
- Wang, M.B. and Waterhouse, P.M. (2002) Application of gene silencing in plants. *Curr. Opin. Plant Biol.* **5**, 146–150.
- Wang, L., Wang, M.B., Tu, J.X., Helliwell, C.A., Waterhouse, P.M., Dennis, E.S., Fu, T.D. and Fan, Y.L. (2007) Cloning and characterization of microRNAs from *Brassica napus*. *FEBS Lett.* **581**, 3848–3856.
- Wang, L., Luo, Y.Z., Zhang, L., Jiao, X.M., Wang, M.B. and Fan, Y.L. (2008) Rolling circle amplification-mediated hairpin RNA (RMHR) library construction in plants. *Nucleic Acids Res.* **36**, e149.
- Wang, M.B., Masuta, C., Smith, N.A. and Shimura, H. (2012) RNA silencing and plant viral diseases. *Mol. Plant Microbe Interact.* **25**, 1275–1285.
- Wang, N., Long, T., Yao, W., Xiong, L., Zhang, Q. and Wu, C. (2013) Mutant resources for the functional analysis of the rice genome. *Mol. Plant*, **6**, 596–604.
- Watanabe, Y. (2011) Overview of plant RNAi. *Methods Mol. Biol.* **744**, 1–11.
- Wei, W., Ba, Z., Gao, M., Wu, Y., Ma, Y., Amiard, S., White, C.I., Rendtlew, D.J., Yang, Y.G. and Qi, Y. (2012) A role for small RNAs in DNA double-strand break repair. *Cell*, **149**, 101–112.
- Xu, M.Y., Dong, Y., Zhang, Q.X., Zhang, L., Luo, Y.Z., Sun, J., Fan, Y.L. and Wang, L. (2012) Identification of miRNAs and their targets from *Brassica napus* by high-throughput sequencing and degradome analysis. *BMC Genomics*, **13**, 421.

Supporting information

Additional Supporting information may be found in the online version of this article:

Figure S1 Locations of the sequenced clones from the three hpRNA libraries in the rice genome.

Figure S2 Screening of hpRNA transgenic plants.

Figure S3 Alignment of the nucleotide sequences from Os08g0430500, Os0280580300, Os11g0546900, and Os03g0710800.

Figure S4 Alignment of the nucleotide sequences from Os04g44924, Os04g44950, Os04g45000, and Os04g44980.

Table S1 Primers used for real-time PCR.

Table S2 Number of phenotypes observed in 211 RNAi mutants.