BMC
Genomics

RESEARCH ARTICLE
Open Access

Target mimics: an embedded layer of microRNA-involved gene regulatory networks in plants

Yijun Meng[1*†], Chaogang Shao[2†], Huizhong Wang[1] and Yongfeng Jin[3*]

Abstract

Background: MicroRNAs (miRNAs) play an essential role in gene regulation in plants. At the same time, the expression of miRNA genes is also tightly controlled. Recently, a novel mechanism called "target mimicry" was discovered, providing another layer for modulating miRNA activities. However, except for the artificial target mimics manipulated for functional studies on certain miRNA genes, only one example, IPS1 (Induced by Phosphate Starvation 1)—miR399 was experimentally confirmed in planta. To date, few analyses for comprehensive identification of natural target mimics have been performed in plants. Thus, limited evidences are available to provide detailed information for interrogating the questionable issue whether target mimicry was widespread in planta, and implicated in certain biological processes.

Results: In this study, genome-wide computational prediction of endogenous miRNA mimics was performed in Arabidopsis and rice, and dozens of target mimics were identified. In contrast to a recent report, the densities of target mimic sites were found to be much higher within the untranslated regions (UTRs) when compared to those within the coding sequences (CDSs) in both plants. Some novel sequence characteristics were observed for the miRNAs that were potentially regulated by the target mimics. GO (Gene Ontology) term enrichment analysis revealed some functional insights into the predicted mimics. After degradome sequencing data-based identification of miRNA targets, the regulatory networks constituted by target mimics, miRNAs and their downstream targets were constructed, and some intriguing subnetworks were further exploited.

Conclusions: These results together suggest that target mimicry may be widely implicated in regulating miRNA activities in planta, and we hope this study could expand the current understanding of miRNA-involved regulatory networks.

Keywords: Target mimic, Network, Subnetwork, MicroRNA, Degradome, Arabidopsis (*Arabidopsis thaliana*), Rice (*Oryza sativa*)

Background

MicroRNAs, the most sophisticatedly characterized small RNA (sRNA) species, were shown to play essential regulatory roles in gene expression in plants [1,2]. Based on the high complementary of the recognition sites on their targets, the plant miRNAs exert repressive roles mostly through target RNA cleavages at post-transcriptional level [2]. Similar to the protein-coding genes, a dominant portion of miRNA genes are transcribed by RNA polymerase II [3-5]. At the same time, the biogenesis and the activities of these critical small molecules themselves were under tight surveillance transcriptionally or post-transcriptionally [6].

One novel mechanism involved in modulating miRNA activities in plants was unraveled by Franco-Zorrilla et al. (2007) [7]. A 23-nt-long motif was observed to be highly conserved among the phosphate starvation-induced, noncoding RNAs transcribed from the TPSI family genes including IPS1 and At4. In Arabidopsis, this motif could be

---
* Correspondence: mengyijun@zju.edu.cn; jinyf@zju.edu.cn
†Equal contributors
[1]College of Life and Environmental Sciences, Hangzhou Normal University, Xuelin Street 16#, Xiasha, Hangzhou 310036, P. R. China
[3]Institute of Biochemistry, College of Life Sciences, Zhejiang University, Zijingang Campus, Yu Hang Tang Road 866#, Hangzhou 310058, P. R. China
Full list of author information is available at the end of the article

recognized by miR399, but could not serve as an effective target cleavage site due to a 3-nt bulge on the "target" RNA sequence opposite the position 10th to 11th nt of miR399 which is the canonical slicing site. Intriguingly, the non-cleavable transcript acts as a target mimic to sequester the corresponding miRNA, thus reducing the active level of miR399. Based on this result, the term "target mimicry" was coined to describe the target mimic—miRNA regulatory relationships. By generating artificial mimics, the authors demonstrated that "target mimicry" might be not only implicated in phosphate signaling, but also in other biological processes, and the mechanisms might be widespread in plants [7]. By using the IPS1 transcript as a scaffold, the subsequent research efforts generated a collection of target mimics in Arabidopsis [8,9], which were valuable for functional studies on certain miRNA genes.

To date, however, only IPS1—miR399 has been experimentally identified as an example of target mimicry that exists in planta naturally. Although dozens of manipulated target mimics have shown great potential for modulating the activities of specific miRNA genes, the widespread existence of the related mechanism in plants remains to be a pressing question. Only one study by Ivashuta et al. (2011) was performed to partially uncover the natural target mimics of the miRNAs in Arabidopsis [8]. However, no in-depth analysis was performed except for some basic statistical results. Besides, the old version of the miRNA registries (miRBase release 15 previously used vs. miRBase 17 currently available) [10] and the gene model annotations [TAIR (The Arabidopsis Information Resource) 9 vs. TAIR 10] [11] utilized in that study, and the exclusion of the currently available non-coding gene information may lead to insufficient exploration on this topic [8].

Here, by using the latest versions of the gene annotations from TAIR (release 10) [11] and TIGR rice (The Institute for Genome Research, release 6.1; currently named the J. Craig Venter institute) [12], genome-wide in silico prediction of potential target mimics was performed for all the registered miRNAs of Arabidopsis and rice in miRBase (release 17) [10]. The miRNAs predicted to be sequestered by certain transcripts were further included for degradome sequencing data-based identification of the downstream targets. Combining these two results, numerous target mimic—miRNA—target regulatory relationships were extracted for comprehensive network construction. Certain subnetworks were further analyzed, and some interesting findings were provided.
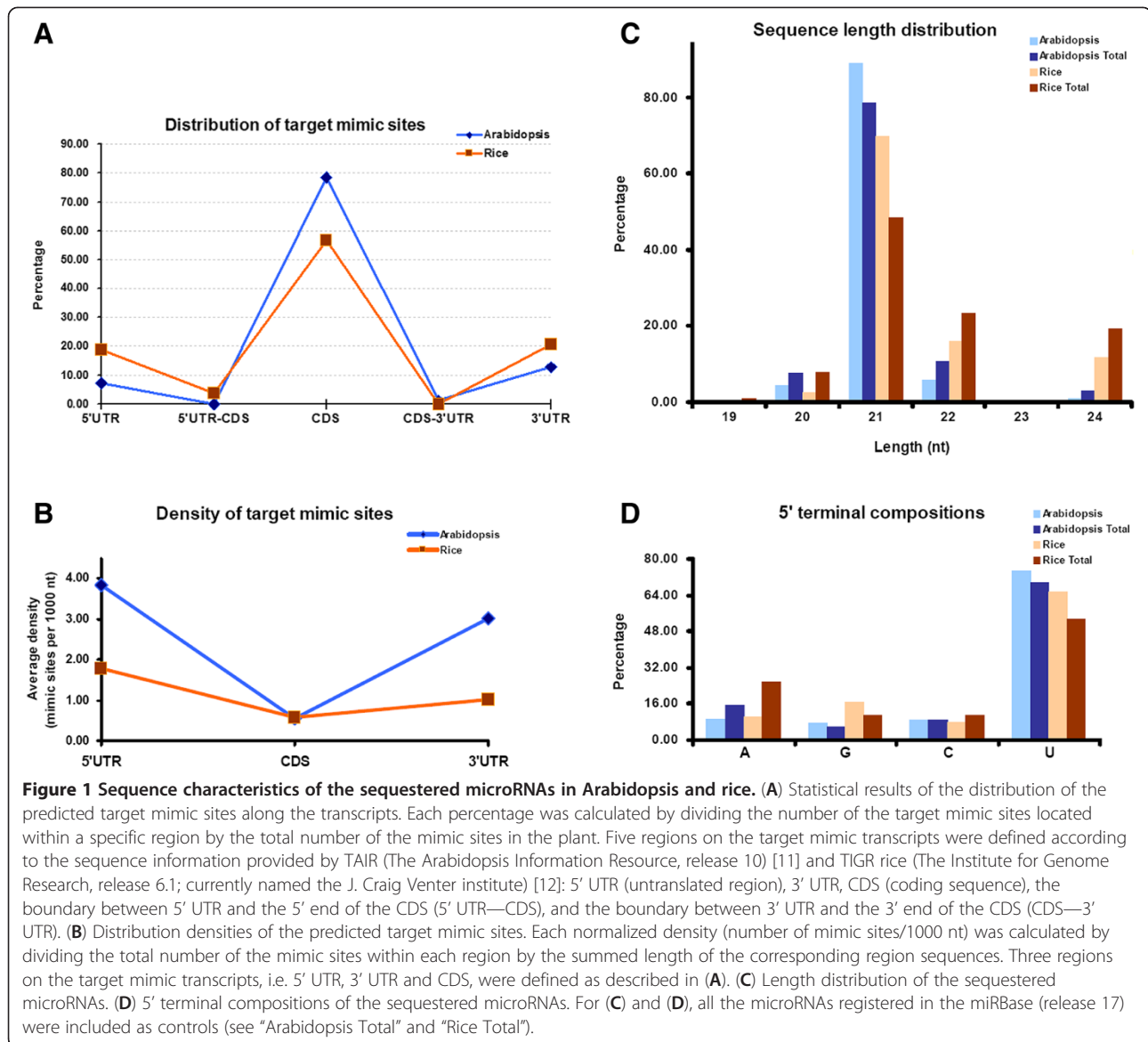
Results and discussion
Transcriptome-wide prediction of natural target mimics of plant miRNAs

The latest versions of gene model annotations of Arabidopsis and rice were retrieved from TAIR [11] and TIGR

rice [12], respectively, serving as the transcript database for the following prediction. All the miRBase-registered miRNAs of both plants (release 17) were included to search for their complementary sites on the gene transcripts by using the tool Ssearch belonging to the FASTA3 package [13,14]. Then, the search results were filtered to identify the potential target mimics of certain miRNAs according to the rules established based on the previous experimental experiences [7-9] (see Methods for detailed rule-based filtering). As a result, 300 and 260 mimic—miRNA interactions were identified, involving 137 and 155 different mature miRNAs in Arabidopsis and rice, respectively (Additional file 1: Table S1 and Additional file 2: Table S2). In Ivashuta et al.'s study (2011), only a limited set of non-coding transcripts from TAIR were included for target mimic prediction [8]. Thus, the question whether the non-coding RNAs tend to be more or less likely to function as target mimics needs to be addressed. To interrogate this issue, most currently available non-coding RNA sequences were obtained from Genomic tRNA Database [15] and NONCODE [16,17]. A same Ssearch- and rule-based identification of target mimics was carried out. Surprisingly, only one mimic of ath-miR418, tRNA238-LysTTT on chromosome 1, was predicted to be a potential candidate in Arabidopsis (Additional file 1: Table S1).

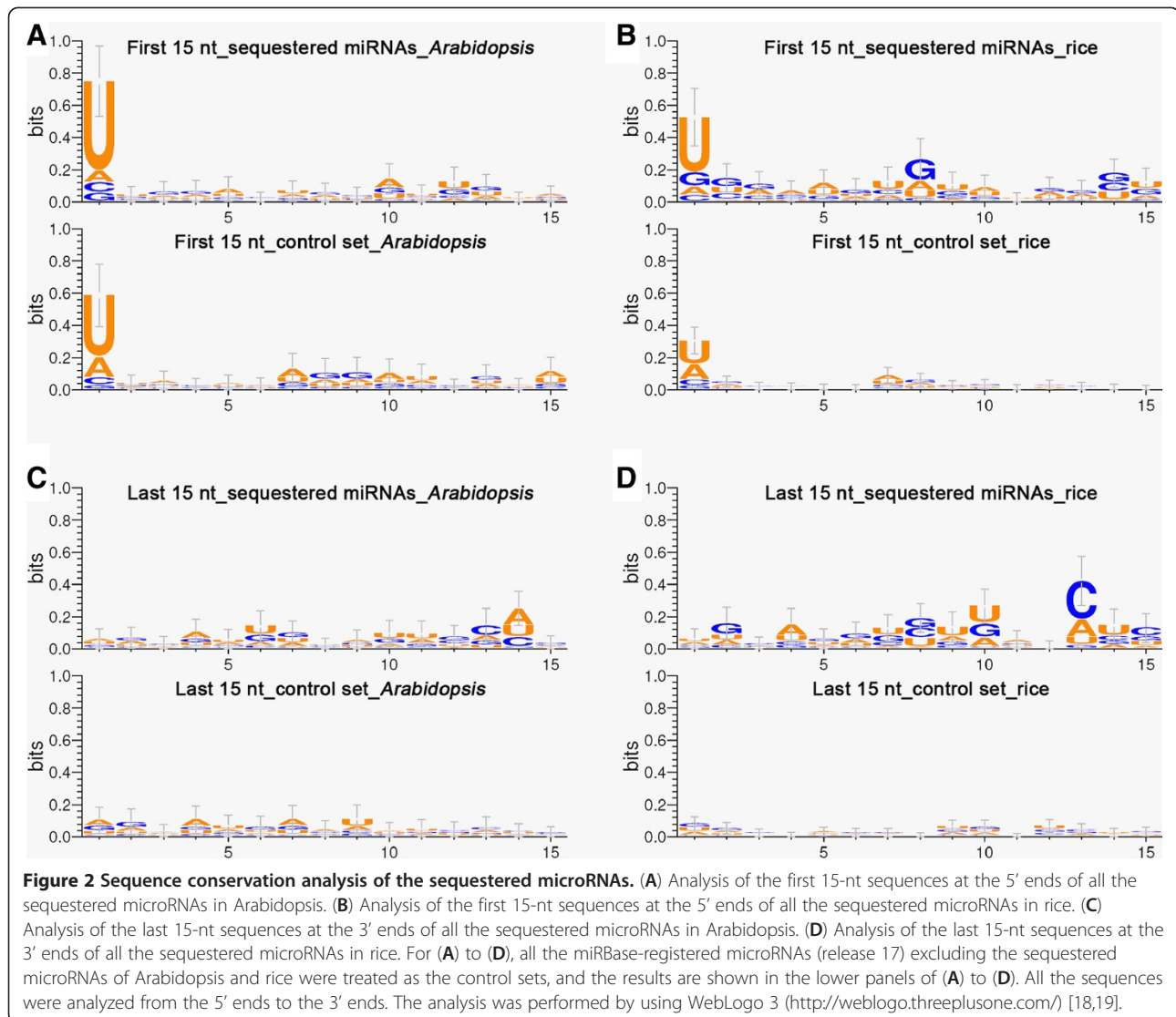### Sequence characteristics of the target mimic sites and the sequestered miRNAs

Then, the distribution patterns of the predicted target mimic sites on the corresponding transcripts were analyzed (see detail in Methods). Quite consistent in both plants, a dominant portion of the mimic sites reside within the CDS (coding sequence) regions (78.53% of the mimic sites in Arabidopsis, and 56.60% in rice), although large portion locate within the UTRs (untranslated regions) (20.25% in Arabidopsis, and 39.62% in rice) (Figure 1A). Notably, more target mimic sites tend to distribute within the 3' UTRs (12.88% in Arabidopsis, and 20.75% in rice) compared to the 5' UTRs (7.36% in Arabidopsis, and 18.87% in rice) in both plants. To calculate the distribution density of the mimic sites, the total number of the mimic sites belonging to each region category (i.e. 5' UTR, 3' UTR or CDS) was divided by the summed length of the corresponding region sequences. Thus, a normalized distribution density (number of mimic sites/ 1000 nt) was generated for each region category. Surprisingly, the result presented a quite different view when compared to the result in Figure 1A. The densities of the target mimic sites are much higher in UTRs (3.83 sites/ 1000 nt within 5' UTRs and 3.03 sites/1000 nt within 3' UTRs in Arabidopsis, and 1.77 sites/1000 nt within 5' UTRs and 1.02 sites/1000 nt within 3' UTRs in rice) when compared to those in CDS regions (0.54 sites/1000 nt in

Figure 1 Sequence characteristics of the sequestered microRNAs in Arabidopsis and rice. (**A**) Statistical results of the distribution of the predicted target mimic sites along the transcripts. Each percentage was calculated by dividing the number of the target mimic sites located within a specific region by the total number of the mimic sites in the plant. Five regions on the target mimic transcripts were defined according to the sequence information provided by TAIR (The Arabidopsis Information Resource, release 10) [11] and TIGR rice (The Institute for Genome Research, release 6.1; currently named the J. Craig Venter institute) [12]: 5′ UTR (untranslated region), 3′ UTR, CDS (coding sequence), the boundary between 5′ UTR and the 5′ end of the CDS (5′ UTR—CDS), and the boundary between 3′ UTR and the 3′ end of the CDS (CDS—3′ UTR). (**B**) Distribution densities of the predicted target mimic sites. Each normalized density (number of mimic sites/1000 nt) was calculated by dividing the total number of the mimic sites within each region by the summed length of the corresponding region sequences. Three regions on the target mimic transcripts, i.e. 5′ UTR, 3′ UTR and CDS, were defined as described in (**A**). (**C**) Length distribution of the sequestered microRNAs. (**D**) 5′ terminal compositions of the sequestered microRNAs. For (**C**) and (**D**), all the microRNAs registered in the miRBase (release 17) were included as controls (see "Arabidopsis Total" and "Rice Total").

Arabidopsis, and 0.56 sites/1000 nt in rice) in both plants (Figure 1B). It is also quite different from the previous observation that no significant difference was observed among the three categorized regions in Arabidopsis (~0.6 sites/1000 nt for 5′ UTRs, ~0.5 sites/1000 nt for CDSs, and ~0.4 sites/1000 nt for 3′ UTRs) [8]. We attributed this discrepancy to the different rules employed for target mimic site identification, since our search criteria were more stringent than the ones used in the previous study. Our statistical result raised the presumption that the non-coding regions of the gene transcripts, i.e. the UTRs, might be preferentially selected to serve as target mimic site-containing regions in plants. However, this possibility still needs verification.

Next, the sequence features of the sequestered miRNAs were characterized. Compared to all the miRBase-annotated miRNAs of Arabidopsis and rice, the sequence length of the sequestered miRNAs tends to be more enriched in 21 nt, and less enriched in 20, 22, and 24 nt (Figure 1C). Moreover, the 5′ terminal compositions of the sequestered miRNAs tend to be more enriched in U (uridine), and less in A (adenine) (Figure 1D). The first 15-nt sequences at both 5′ and 3′ ends of the sequestered miRNAs were collected for sequence conservation analysis. When compared to the control sets (all the miRBase-registered mature miRNAs of Arabidopsis and rice, excluding the sequestered miRNAs, were treated as the control sets), in both plants, the sequestered miRNAs preferentially start with 5′ U (Figure 2A and 2B), which was consistent with the above statistical results. More interestingly, the third nucleotides at the 3′ ends of the sequestered miRNAs were found to be dominantly occupied by C (cytosine) in

Figure 2 Sequence conservation analysis of the sequestered microRNAs. (**A**) Analysis of the first 15-nt sequences at the 5′ ends of all the sequestered microRNAs in Arabidopsis. (**B**) Analysis of the first 15-nt sequences at the 5′ ends of all the sequestered microRNAs in rice. (**C**) Analysis of the last 15-nt sequences at the 3′ ends of all the sequestered microRNAs in Arabidopsis. (**D**) Analysis of the last 15-nt sequences at the 3′ ends of all the sequestered microRNAs in rice. For (**A**) to (**D**), all the miRBase-registered microRNAs (release 17) excluding the sequestered microRNAs of Arabidopsis and rice were treated as the control sets, and the results are shown in the lower panels of (**A**) to (**D**). All the sequences were analyzed from the 5′ ends to the 3′ ends. The analysis was performed by using WebLogo 3 (http://weblogo.threeplusone.com/) [18,19].

rice, although it was less prominent in Arabidopsis (Figure 2C and 2D). Further investigation of the canonical example, IPS1—miR399, revealed that all the miR399 family members of both Arabidopsis and rice were 21 nt in length, and started with 5′ U, and the third nucleotides at their 3′ ends were C, only one exception was observed that the third nucleotide at the 3′ end of ath-miR399e was U. In this regard, whether these features, such as the third nucleotide at the 3′ end, could determine the probability that a specific miRNA is regulated by certain target mimic (s) needs to be further interrogated. Together, all the evidences point to the fact that the miRNAs tending to be sequestered by specific target mimics are more likely to possess the classic sequence features of plant miRNAs [6].

Functional analysis of the target mimics
All the target mimics were included for GO term enrichment analysis by using agriGO [20]. Intriguingly, many

identified enriched GO terms were highly conserved between the two plants analyzed. For example, within the "Biological Process" category, the phosphorus metabolism-related processes such as "phosphorylation" were indicated to be enriched processes that the mimics were involved in (Figure 3A and 3B). Supporting this observation, within the "Molecular Function" category, "ATP binding" and "kinase activity" which were implicated in phosphorus metabolism, were highly enriched functions possessed by the sets of target mimics in both plants (Figure 3C and 3D). Besides, the GO terms "transcription factor activity" and "zinc ion binding" were also highly enriched in the target mimic gene set of Arabidopsis. Based on the gene annotations provided by TAIR and TIGR, we found that a portion of the target mimics were encoded by transposable element (TE) genes (4.59% in Arabidopsis and 12.59% in rice), indicating a novel functional activity of the TEs for regulating the miR-NAs in plants. Based on the TAIR annotations, five mimic
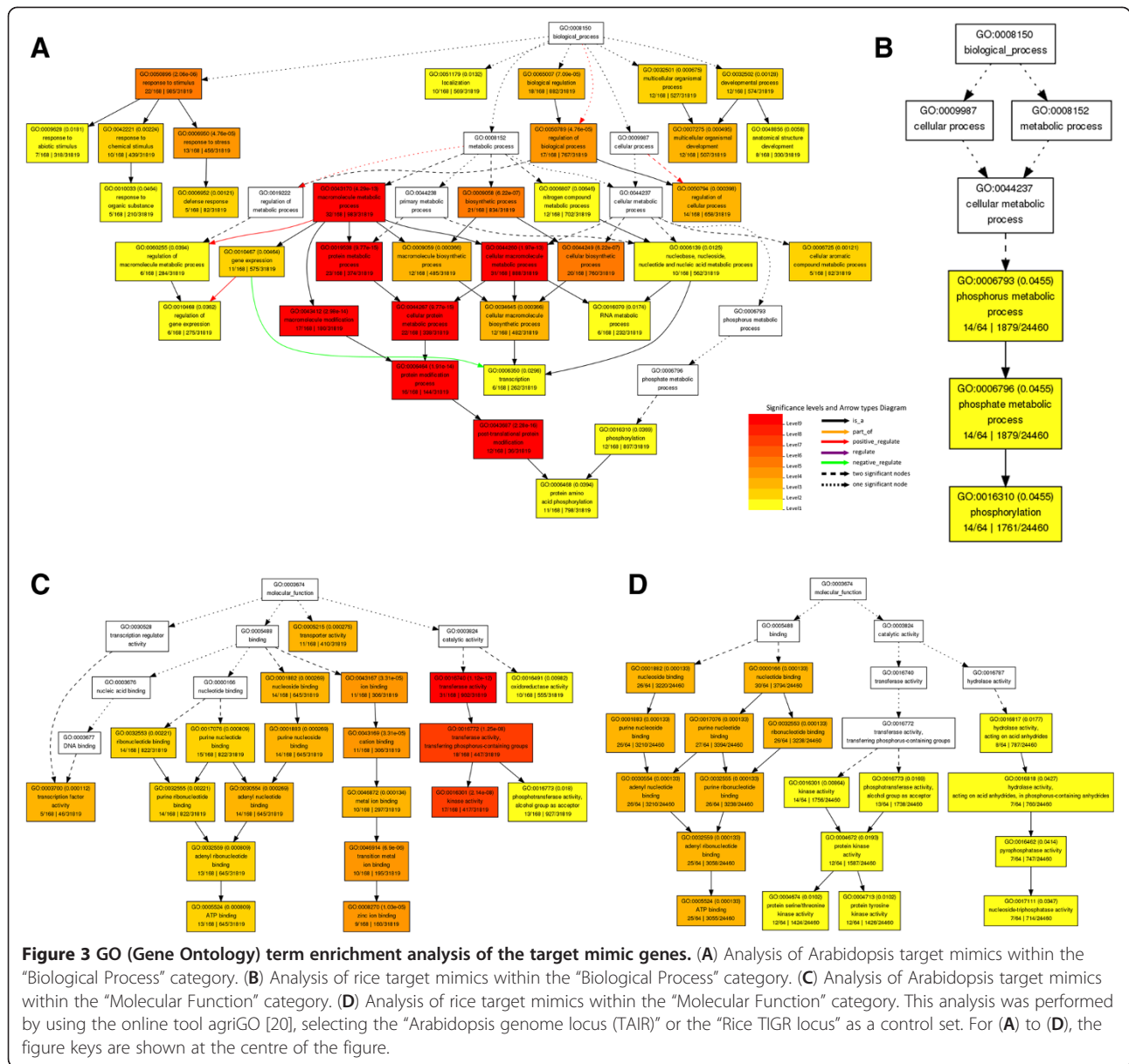
Figure 3 GO (Gene Ontology) term enrichment analysis of the target mimic genes. (**A**) Analysis of Arabidopsis target mimics within the "Biological Process" category. (**B**) Analysis of rice target mimics within the "Biological Process" category. (**C**) Analysis of Arabidopsis target mimics within the "Molecular Function" category. (**D**) Analysis of rice target mimics within the "Molecular Function" category. This analysis was performed by using the online tool agriGO [20], selecting the "Arabidopsis genome locus (TAIR)" or the "Rice TIGR locus" as a control set. For (**A**) to (**D**), the figure keys are shown at the centre of the figure.

genes, AT1G55860 regulating ath-miR159a, AT2G29070 regulating ath-miR172, AT1G77870 modulating ath-miR775, and AT4G02950 and AT4G03360 modulating ath-miR862-3p, were implicated in post-translational protein modification through a ubiquitin-related pathway (Figure 3A and Additional file 3: Table S3).

Uncovering the functions of the sequestered miRNAs through target identification

Most plant miRNAs exert their regulatory roles in various biological processes through direct target cleavages [6]. To gain deeper functional insights of the miRNAs regulated by the target mimics, large-scale target prediction was performed by using miRU algorithm [21,22]. The predicted results were filtered by using degradome sequencing data-

based approach in order to gain highly reliable miRNA—target regulatory relationships (Additional file 4: Figure S1 and Additional file 5: Figure S2). Among the identified miRNA—target pairs, some interesting regulations were observed. In Arabidopsis, the precursors of ath-miR172b and ath-miR400 could be recognized by their own mature miRNAs at the miRNA*-coding regions, and significant cleavage signals were observed in the middle of the target regions (Figure 4A and 4B). This observation further supports the "self-regulation" notion proposed previously [23,24]. In rice, both LOC_Os10g33700.1 and LOC_Os10-g39970.1 were indicated to be co-regulated by osa-miR809 and osa-miR819 family members (Figure 4C and 4D). Additionally, alternative splicing employed by numerous genes may enable certain transcript variants to escape
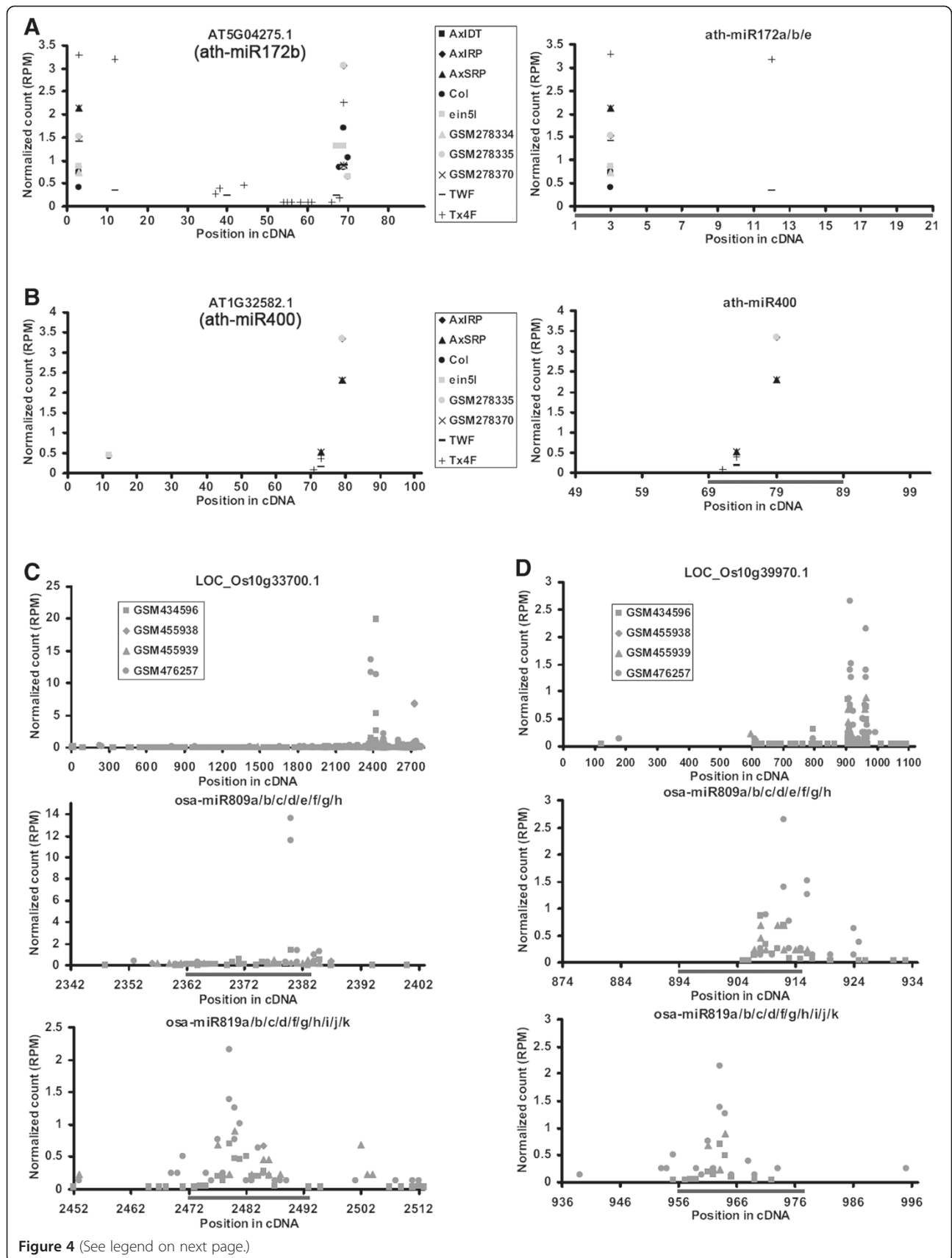
Figure 4 (See legend on next page.)

(See figure on previous page.)

Figure 4 Degradome sequencing data-based identification of the targets regulated by the sequestered microRNAs revealed the novel self-regulation and co-regulation mechanisms. (**A**) Self-regulation of ath-miR172b. (**B**) Self-regulation of ath-miR400. (**C**) Co-regulation of the transcript LOC_Os10g33700.1 by osa-miR809 and osa-miR819. (**D**) Co-regulation of the transcript LOC_Os10g39970.1 by osa-miR809 and osa-miR819. For all the sub-figures [(**A**) to (**D**)], the first panels depict the degradome signals all along the target transcripts, and the other panels provide detailed views of the cleavage signals within the regions surrounding the target recognition sites (denoted by gray horizontal lines). The transcript IDs are shown in the first panels, and the miRNA IDs are listed in the other panels. The x axes measure the positions of the signals along the transcripts, and the y axes measure the signal intensities based on normalized counts (in RPM, reads per million), allowing cross-library comparison. See "Data sets used in this study" in the METHODS section for the degradome data sets used in this analysis.

from miRNA-mediated regulation. For instance, the splicing variant AT5G63260.2 was found to be cleaved by ath-miR415, while AT5G63260.1 could not be targeted by this miRNA due to the lack of the target recognition site.

A large portion of miRNA targets have been demonstrated to be transcription factors in plants, suggesting their important role in gene regulatory cascades [2]. To gain a global functional view of these identified targets of the sequestered miRNAs, GO term enrichment analysis was performed again. As expected, "transcription factor activity" is a highly enriched function possessed by the target sets in both plants (Additional file 6: Figure S3A and S3B). Interestingly, the GO term "hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides" belonging to the "Molecular Function" category was found to be enriched in the rice target set (Additional file 6: Figure S3B). Considering the functional enrichment of the target mimic genes in phosphorus metabolism-related processes in both Arabidopsis and rice, the embedded implication is worth investigating. Moreover, according to the GO annotations, a large portion of the miRNA targets in Arabidopsis were suggested to be involved in the biological processes "RNA interference", "cell differentiation", "vegetative (leaf) and reproductive (flower, fruit, and seed) organ development", and "meristem initiation" (Additional file 6: Figure S3C).

Also based on target prediction and degradome data-based validation, certain target mimics were identified to be regulated by specific miRNAs. For instance, the mimic transcripts AT1G69440.1 and AT5G03545.1 were indicated to be regulated by ath-miR5021 and ath-miR414 respectively (Additional file 7: Figure S4), and LOC_Os02g36880.3 was cleaved by osa-miR164a-f in rice (Additional file 8: Figure S5).

Construction of the networks constituted by "miRNA—mimic—miRNA—target" regulatory cascades and subnetwork characterization

Through target mimic prediction and degradome data-based miRNA target identification, the basic data for establishing the "miRNA—mimic—miRNA—target" regulatory relationships were obtained. Thus, we set out to construct comprehensive networks involving target mimic—miRNA regulations in both Arabidopsis and rice

by using Cytoscape [25]. At first glance, 465 nodes (including miRNAs, miRNA targets, and target mimics) were found to be connected by 559 edges in Arabidopsis, and 441 nodes connected by 1048 edges in rice (Additional file 9: Figure S6 and Additional file 10: Figure S7). To demonstrate the biological meanings of the constructed networks, certain subnetworks were further investigated.

In Arabidopsis, the genes IPS1 (AT3G09922) and AT4 (AT5G03545) belonging to TPSI/Mt4 family were identified as the target mimics of ath-miR399 (Figure 5A and Additional file 3: Table S3), which was consistent with the previous experimental report [7]. It also indicates the high reliability of our criteria set for the prediction of target mimics. Based on the same prediction criteria, the IPS1 homologous gene LOC_Os03g05334 was identified in rice (Figure 5B and Additional file 11: Table S4). Besides, several novel genes (AT1G21930, AT2G19950, LOC_Os02g43840, LOC_Os06g03690, and LOC_Os09-g33510) not belonging to the TPSI/Mt4 family were also indicated to have great potential to modulate the activities of miR399 in both plants. Whether these potential mimics are involved in phosphate signaling needs to be further addressed. One example is that miR399 was predicted to be under surveillance of the mimic gene LOC_Os02g43840 encoding an ethylene-responsive element-binding protein, suggesting a potential interplay between phosphorus and ethylene signaling pathways in rice. This scenario is supported by the recent finding that ethylene response factors are potentially implicated in the regulatory responses to phosphate deprivation in Arabidopsis [26]. More interestingly, within the ath-miR399-involved subnetwork, one mimic transcript encoded by AT5G03545 (annotated as AT4 belonging to the TPSI/Mt4 family) was found to be regulated by ath-miR414 through cleavage (Additional file 6: Figure S3). To our best knowledge, ath-miR414 has never been uncovered to participate in the phosphorus-related signaling pathways. Thus, the regulatory cascade ath-miR414—AT4—ath-miR399—PHO2 [the reported target PHO2 of miR399 [27] was not identified based on our degradome-based search, which might be attributed to the stringent search criteria that we used or the limited degradome sequencing data sets that available for this analysis] is worth further investigating.
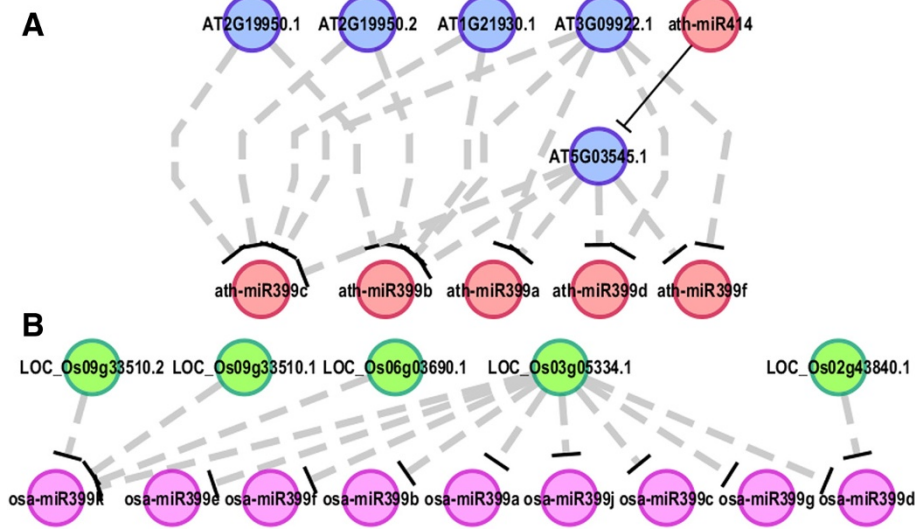
Figure 5 miR399-involved subnetworks in Arabidopsis and rice. (**A**) ath-miR399- and ath-miR414-involved phosphorous signaling-related subnetwork. (**B**) osa-miR399-involved phosphorous signaling-related subnetwork. The predicted regulatory relationships between target mimics and certain microRNAs were depicted by gray dashed lines, and the microRNA—target regulatory relationship was denoted by black line. All the networks were drawn by Cytoscape [25].

Another novel finding was exploited from the miR159/ 319-, miR164- and miR395-involved subnetwork in rice (Figure 6). osa-miR164 was demonstrated to cleave the transcript LOC_Os02g36880.3 [belonging to NAC family, a well-characterized target gene family regulated by miR164 in plants [2]], which in turn regulated osa-miR395t through target mimicry. Since miR164 and miR395 were reported to be involved in auxin signaling [28] and sulfate metabolism [29,30] respectively, the possibility that the auxin—sulfur signal interplay could be mediated by the cascade miR164—NAC—miR395 needs in-depth investigations. However, we observed that only osa-miR395t was predicted to be regulated by the mimic transcript LOC_Os02g36880.3, and no downstream target was identified for this miRNA based on degradome sequencing data. Thus, whether the regulatory relationship, LOC_Os02g36880.3—osa-miR395t specifically exists in rice needs to be interrogated carefully. But, we suggest that the "target mimicry" relationship between NAC family gene transcript and miR395 identified in this subnetwork may not be a false positive. Another NAC—miR395 relationship was identified between LOC_Os10g33760.1 and most of the miR395 family members in rice (Figure 6 and Additional file 11: Table S4). More complicatedly, the target genes of osa-miR319, LOC_Os03g57190 and LOC_Os07g05720, encode transcription factors belonging to TCP family, which was demonstrated to positively regulate the expression of miR164 at the transcriptional level in Arabidopsis [31]. Thus, the regulatory cascade TCP— miR164—NAC—miR395 may be at the nexus of the auxin and the sulfur signaling pathways. In Arabidopsis, a largely

conserved subnetwork involving miR159/319, miR164 and miR395 was also extracted from the comprehensive network (Additional file 12: Figure S8). However, no NAC-related gene was discovered as a potential target mimic of miR395. Thus, whether the NAC—miR395-mediated auxin—sulfur signal interaction is specifically existed in rice needs further interpretation. Another interesting finding is that different from rice, in which miR159 and miR319 regulate distinct sets of target genes separately (MYB and TCP genes, respectively), miR159 and miR319 in Arabidopsis share an overlapping set of targets. Previous study indicates that although the sequences of miR159 and miR319 show high identity with each other, these two miRNA species possess specialized functions by regulating distinct sets of targets, i.e. MYB and TCP genes, separately [32]. Our observation in Arabidopsis indicated the partial functional redundancy between the two homologous miRNA families. Additionally, the mimic gene of ath-miR164, AT4G03280, is suggested to play a role in photosynthetic electron transfer based on TAIR annotation. On the other hand, the involvement of miR164 in plant photosynthesis is supported by several pieces of recent evidences [33-35]. Notably, based on the previous study, the expression of AT1G60710, a potential mimic of ath-miR395a/day/e, showed specific response to sulfur depletion treatment [36]. Thus, the target mimicry relationship between AT1G60710 and ath-miR395 could add a potential layer of miR395-mediated sulfur signaling pathways.

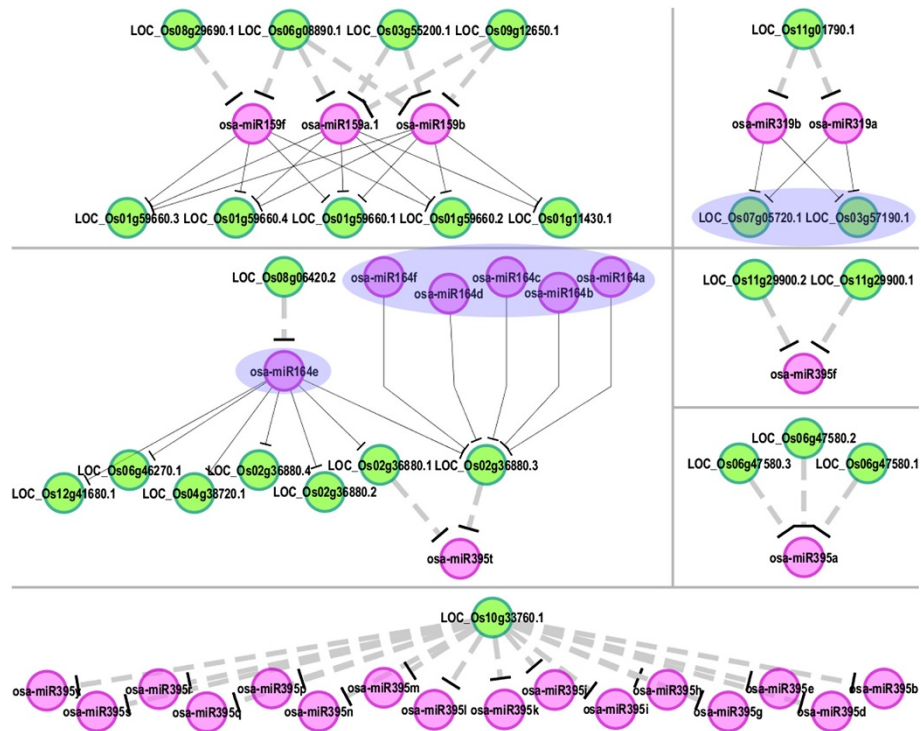Several other interesting subnetworks were also extracted for characterization. For instances, only one

Figure 6 osa-miR159-, osa-miR319-, osa-miR164-, and osa-miR395-involved subnetworks. The predicted regulatory relationships between target mimics and certain microRNAs were depicted by gray dashed lines, and the microRNA—target regulatory relationships were denoted by black lines. The transcripts LOC_Os03g57190.1 and LOC_Os07g05720.1 belonging to TCP gene family, and the osa-miR164 family members were highlighted by translucent blue ovals, indicating that the expression of osa-miR164 was positively regulated by the TCP family genes. The network was drawn by Cytoscape [25].

member of miR156 families in both plants, i.e. ath-miR156h and osa-miR156k, was identified within the established comprehensive networks (Additional file 13: Figure S9). Both mimics and downstream targets were discovered for the two miR156 genes. Thus, whether these two miR156 family members play a dominant role in specific regulatory pathway in both plants requires further investigations. Based on our computational approach, the target genes of miR172 belonging to the AP2 family were identified based on the significant cleavage signals within the target sites in Arabidopsis (Additional file 4: Figure S1 and Additional file 13: Figure S9). Thus, in addition to the previously reported translational repressive effect of miR172 on the AP2 genes [37,38], target cleavages of the AP2 transcripts may also play an indispensable role in floral organ development. Within the ath-miR169-mediated subnetwork, the miRNA star species, ath-miR169g*, was found to be potentially regulated by three mimic transcripts, AT1G52060.1, AT4G16070.1 and AT4G16070.2 (Additional file 13: Figure S9). Considering the widespread regulatory activities of miRNA*s unraveled in recent years [39-42], it is reasonable that the active levels of certain miRNA*s should be under strict surveillance through target

mimicry. In rice, a similar miR169-invovel subnetwork was identified, although no such mimic—miR169* regulatory relationship existed (Additional file 13: Figure S9). However, based on the TIGR rice annotations, two mimic genes of osa-miR169, LOC_Os05g24010 and LOC_Os09g37800, were suggested to be responsive to stress. Considering the reported involvement of miR169 in drought [43] and nitrogen starvation [44] response in rice, the characterized subnetwork may play an essential role in multi-stress-induced response. Moreover, the largest subnetwork involving miR446, miR809 and miR819 was identified in rice (Additional file 13: Figure S9). These three miRNA families seem to be rice-specific according to the current miRBase registries (release 17), and their biological functions have not been characterized. Unfortunately, we could not gain any informative hints from the current annotations of the mimic and target genes within this subnetwork. Hence, it will be interesting to gain functional insights from the established subnetwork through experimental approaches.

Conclusions

Taken together, comprehensive networks constituted by numerous target mimic—miRNA—target regulatory

cascades have been constructed in Arabidopsis and rice through in silico target mimic site prediction and degradome data-based target identification. By in-depth characterization of certain interesting subnetworks, the established networks were demonstrated to be relatively reliable and biologically meaningful. Several subnetworks were observed to be conserved between Arabidopsis and rice to some extent (Figure 5, Figure 6, Additional file 12: Figure S8 and Additional file 13: Figure S9). And some might be species-specific. We hope this study could expand the current view of miRNA-mediated regulatory networks in plants, and will inspire more research efforts on the novel regulatory mechanisms for modulating miRNA activities, such as the target mimicry characterized in our analysis.

Methods

Data sets used in this study

The degradome sequencing data sets were retrieved from GEO (Gene Expression Omnibus; http://www.ncbi. nlm.nih.gov/geo/) [45] and NGSDBs (Next-Gen Sequence Databases; http://mpss.udel.edu/) [46]. The accession numbers of these data sets are: (1) Arabidopsis degradome data: GSM278333, GSM278334, GSM278335, and GSM278370 from GEO; and AxIDT, AxIRP, AxSRP, Col, ein5l, TWF, and Tx4F from NGSDBs. (2) rice degradome data: GSM434596, GSM455938, GSM455939, and GSM476257 from GEO. The gene annotation and sequence information of Arabidopsis and rice were retrieved from the FTP sites of TAIR (release 10; ftp:// ftp.arabidopsis.org/home/tair/Sequences/blast_datasets/) [11] and TIGR rice (release 6.1; ftp://ftp.plantbiology. msu.edu/pub/data/Eukaryotic_Projects/o_sativa/anno tation_dbs/pseudomolecules/) [12], respectively. The sequences of the miRNA precursors and the mature miRNAs were downloaded from miRBase (release 17; http://www.mirbase.org/) [10]. The other non-protein-coding RNA sequences were retrieved from Genomic tRNA Database (http://gtrnadb.ucsc.edu/download. html) [15] and NONCODE (http://www.noncode.org/ NONCODERv3/download.htm) [16,17].

Prediction of endogenous target mimics in Arabidopsis and rice

First, Ssearch from the FASTA3 program package [downloaded from the FTP site of EBI (European Bioinformatics Institute), ftp://ftp.ebi.ac.uk/pub/software/ unix/fasta/] [13,14] was used to search for the sites in cDNA sequences that were reverse complementary to the miRNAs. Each miRNA of Arabidopsis and rice was included to search against the cDNA sequence library of the corresponding plant species. The cDNA sequences were retrieved from TAIR and TIGR rice, and the miRNAs from miRBase as mentioned above. To retain the

predicted site with low complementary to the miRNAs (to allow the identification of target mimics with big bulges within the complementary sites), the first 5,000 Ssearch results for each miRNA were obtained for further identification. To discover the miRNA mimics, we applied the following set of rules referring to the previous experimental results [7-9]: (1) The 3- to 5-nt bulges must exist within the complementary sites of the cDNAs, and the bulges should located in the middle of the corresponding miRNAs (definition of the middle positions: 9th to 11th nt of the 19-nt-long miRNAs; 10th to 11th nt of the 20-nt miRNAs; 10th to 12th nt of the 21-nt ones; 11th to 12th nt of the 22-nt ones; 11th to 13th nt of the 23-nt ones; 12th to 13th nt of the 24-nt ones). (2) For the total mismatches within the non-middle region of each miRNA, no more than 4 were allowed, and the consecutive mismatches should not exceed 2 nt. (3) No bulge was permitted within the non-middle regions of the miRNAs. A Perl script was developed to perform this rule-based screening. The cDNAs satisfied the above criteria were considered to be target mimic candidates.

Distribution pattern analysis of the target mimic sites on the gene transcripts

Based on the TAIR and TIGR rice annotations, only the target mimic sites located on the gene transcripts with 5' UTR—CDS—3' UTR structure were included in this analysis. In some cases, one mimic site might be recognized by two or more different miRNAs (especially for the members of the same miRNA families). These sites were considered only once.

Prediction and degradome-based validation of miRNA targets

Target prediction was performed by using miRU algorithm [21,22] with default parameters. The degradome sequencing data were utilized to validate the predicted miRNA—target pairs. First, the read counts of all the degradome reads from each library were normalized in order to allow cross-library comparison. The normalized read count (in RPM, reads per million) of a short read from a specific library was calculated by dividing the raw count of this read by the total counts of the library, and then multiplied by 10^6. Second, all the degradome short reads were mapped to the predicted target transcripts by using BLAST algorithm [47], and only the perfectly matched reads were retained. Then, two-step filtering was performed to extract the most likely miRNA—target pairs. During the first step, the predicted targets were retained for further validation only if there were three or more degradome reads with identical 5' ends located within the predicted target binding sites. For this filtering step, all the degradome data sets were utilized at the same time to do a comprehensive screening. It was

based on the scenario that a miRNA—target pair was considered to be the candidate once the cleavage signal(s) existed in any data set(s). After the first filtering, the degradome signals along each retained transcript were obtained from the BLAST results to provide a global view of the signal noise when compared to the signal intensity within a specific target binding site. Referring to our previous study [48], both the global and the local t-plots (target plots) [49,50] were drawn. Finally, exhaustive manual filtering was performed, and only the transcripts with cleavage signals easy to be recognized were extracted as the miRNA—target pairs.

Additional files

Additional file 1: Table S1. Prediction results of the target mimics of the Arabidopsis microRNAs.

Additional file 2: Table S2. Prediction results of the target mimics of the rice microRNAs.

Additional file 3: Table S3. Annotation information of the target mimics in Arabidopsis.

Additional file 4: Figure S1. Degradome sequencing data-based identification of the targets of sequestered microRNAs in Arabidopsis.

Additional file 5: Figure S2. Degradome sequencing data-based identification of the targets of sequestered microRNAs in rice.

Additional file 6: Figure S3. GO (Gene Ontology) term enrichment analysis of the targets of sequestered microRNAs in Arabidopsis and rice.

Additional file 7: Figure S4. Degradome sequencing data-based identification of the microRNAs regulating the target mimics in Arabidopsis.

Additional file 8: Figure S5. Degradome sequencing data-based identification of the microRNAs regulating the target mimics in rice.

Additional file 9: Figure S6. The whole network constructed in Arabidopsis.

Additional file 10: Figure S7. The whole network constructed in rice.

Additional file 11: Table S4. Annotation information of the target mimics in rice.

Additional file 12: Figure S8. Ath-miR159/319-, ath-miR164-, and ath-miR395-involved subnetworks.

Additional file 13: Figure S9. Certain subnetworks in Arabidopsis and rice.

Abbreviations

miRNA: MicroRNA; IPS1: Induced by Phosphate Starvation 1; UTR: Untranslated region; CDS: Coding sequence; GO: Gene Ontology; sRNA: Small RNA; TAIR: The Arabidopsis Information Resource; TIGR: The Institute for Genome Research; U: Uridine; A: Adenine; C: Cytosine; TE: Transposable element; GEO: Gene Expression Omnibus; NGSDBs: Next-Gen Sequence Databases; RPM: Reads per million; t-plots: Target plots.

Competing interests

The authors declare that they have no competing interest.

Author details

[1]College of Life and Environmental Sciences, Hangzhou Normal University, Xuelin Street 16#, Xiasha, Hangzhou 310036, P. R. China. [2]College of Life Sciences, Huzhou Teachers College, Huzhou 313000, P. R. China. [3]Institute of Biochemistry, College of Life Sciences, Zhejiang University, Zijingang Campus, Yu Hang Tang Road 866#, Hangzhou 310058, P. R. China.

Authors' contributions

Conceived and designed the experiments: YM, HW, YJ. Performed the experiments: CS. Analyzed the data: YM CS. Contributed reagents/materials/analysis tools: YM CS. Wrote the paper: YM, HW, YJ. All authors read and approved the final manuscript.

References

1. Chen X: **Small RNAs and their roles in plant development.** *Annu Rev Cell Dev Biol* 2009, **25**:21–44.
2. Jones-Rhoades MW, Bartel DP, Bartel B: **MicroRNAs and their regulatory roles in plants.** *Annu Rev Plant Biol* 2006, **57**:19–53.
3. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN: **MicroRNA genes are transcribed by RNA polymerase II.** *EMBO J* 2004, **23**:4051–4060.
4. Meng Y, Huang F, Shi Q, Cao J, Chen D, Zhang J, Ni J, Wu P, Chen M: **Genome-wide survey of rice microRNAs and microRNA—target pairs in the root of a novel auxin-resistant mutant.** *Planta* 2009, **230**:883–898.
5. Xie Z, Allen E, Fahlgren N, Calamar A, Givan SA, Carrington JC: **Expression of Arabidopsis MIRNA genes.** *Plant Physiol* 2005, **138**:2145–2154.
6. Voinnet O: **Origin, biogenesis, and activity of plant microRNAs.** *Cell* 2009, **136**:669–687.
7. Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-Somoza I, Leyva A, Weigel D, Garcia JA, Paz-Ares J: **Target mimicry provides a new mechanism for regulation of microRNA activity.** *Nat Genet* 2007, **39**: 1033–1037.
8. Ivashuta S, Banks IR, Wiggins BE, Zhang Y, Ziegler TE, Roberts JK, Heck GR: **Regulation of gene expression in plants through miRNA inactivation.** *PLoS One* 2011, **6**:e21330.
9. Todesco M, Rubio-Somoza I, Paz-Ares J, Weigel D: **A collection of target mimics for comprehensive analysis of microRNA function in Arabidopsis thaliana.** *PLoS Genet* 2010, **6**:e1001031.
10. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ: **miRBase: microRNA sequences, targets and gene nomenclature.** *Nucleic Acids Res* 2006, **34**:D140–144.
11. Huala E, Dickerman AW, Garcia-Hernandez M, Weems D, Reiser L, LaFond F, Hanley D, Kiphart D, Zhuang M, Huang W, *et al*: **The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant.** *Nucleic Acids Res* 2001, **29**:102–105.
12. Yuan Q, Ouyang S, Liu J, Suh B, Cheung F, Sultana R, Lee D, Quackenbush J, Buell CR: **The TIGR rice genome annotation resource: annotating the genome and creating resources for plant biologists.** *Nucleic Acids Res* 2003, **31**:229–233.
13. Pearson WR: **Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms.** *Genomics* 1991, **11**:635–650.
14. Pearson WR: **Flexible sequence similarity searching with the FASTA3 program package.** *Methods Mol Biol* 2000, **132**:185–219.
15. Chan PP, Lowe TM: **GtRNAdb: a database of transfer RNA genes detected in genomic sequence.** *Nucleic Acids Res* 2009, **37**:D93–97.
16. He S, Liu C, Skogerbo G, Zhao H, Wang J, Liu T, Bai B, Zhao Y, Chen R: **NONCODE v2.0: decoding the non-coding.** *Nucleic Acids Res* 2008, **36**: D170–172.
17. Liu C, Bai B, Skogerbo G, Cai L, Deng W, Zhang Y, Bu D, Zhao Y, Chen R: **NONCODE: an integrated knowledge database of non-coding RNAs.** *Nucleic Acids Res* 2005, **33**:D112–115.
18. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator.** *Genome Res* 2004, **14**:1188–1190.
19. Schneider TD, Stephens RM: **Sequence logos: a new way to display consensus sequences.** *Nucleic Acids Res* 1990, **18**:6097–6100.
20. Du Z, Zhou X, Ling Y, Zhang Z, Su Z: **agriGO: a GO analysis toolkit for the agricultural community.** *Nucleic Acids Res* 2010, **38**:W64–70.

21. Dai X, Zhao PX: **psRNATarget: a plant small RNA target analysis server.** *Nucleic Acids Res* 2011, **39**:W155–159.

22. Zhang Y: **miRU: an automated plant miRNA target prediction server.** *Nucleic Acids Res* 2005, **33**:W701–704.

23. Meng Y, Gou L, Chen D, Mao C, Jin Y, Wu P, Chen M: **PmiRKB: a plant microRNA knowledge base.** *Nucleic Acids Res* 2011, **39**:D181–187.

24. Meng Y, Gou L, Chen D, Wu P, Chen M: **High-throughput degradome sequencing can be used to gain insights into microRNA precursor metabolism.** *J Exp Bot* 2010, **61**:3833–3837.

25. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res* 2003, **13**:2498–2504.

26. Nilsson L, Muller R, Nielsen TH: **Dissecting the plant transcriptome and the regulatory responses to phosphate deprivation.** *Physiol Plant* 2010, **139**:129–143.

27. Bari R, Datt Pant B, Stitt M, Scheible WR: **PHO2, microRNA399, and PHR1 define a phosphate-signaling pathway in plants.** *Plant Physiol* 2006, **141**:988–999.

28. Guo HS, Xie Q, Fei JF, Chua NH: **MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for Arabidopsis lateral root development.** *Plant Cell* 2005, **17**:1376–1386.

29. Kawashima CG, Matthewman CA, Huang S, Lee BR, Yoshimoto N, Koprivova A, Rubio-Somoza I, Todesco M, Rathjen T, Saito K, *et al*: **Interplay of SLIM1 and miR395 in the regulation of sulfate assimilation in Arabidopsis.** *Plant J* 2011, **66**:863–876.

30. Kawashima CG, Yoshimoto N, Maruyama-Nakashita A, Tsuchiya YN, Saito K, Takahashi H, Dalmay T: **Sulphur starvation induces the expression of microRNA-395 and one of its target genes but in different cell types.** *Plant J* 2009, **57**:313–321.

31. Koyama T, Mitsuda N, Seki M, Shinozaki K, Ohme-Takagi M: **TCP transcription factors regulate the activities of ASYMMETRIC LEAVES1 and miR164, as well as the auxin response, during differentiation of leaves in Arabidopsis.** *Plant Cell* 2010, **22**:3574–3588.

32. Palatnik JF, Wollmann H, Schommer C, Schwab R, Boisbouvier J, Rodriguez R, Warthmann N, Allen E, Dezulian T, Huson D, *et al*: **Sequence and expression differences underlie functional specialization of Arabidopsis microRNAs miR159 and miR319.** *Dev Cell* 2007, **13**:115–125.

33. Jia X, Ren L, Chen QJ, Li R, Tang G: **UV-B-responsive microRNAs in Populus tremula.** *J Plant Physiol* 2009, **166**:2046–2057.

34. Nikovics K, Blein T, Peaucelle A, Ishida T, Morin H, Aida M, Laufs P: **The balance between the MIR164A and CUC2 genes controls leaf margin serration in Arabidopsis.** *Plant Cell* 2006, **18**:2929–2945.

35. Pulido A, Laufs P: **Co-ordination of developmental processes by small RNAs during leaf development.** *J Exp Bot* 2010, **61**:1277–1291.

36. Nikiforova V, Freitag J, Kempa S, Adamik M, Hesse H, Hoefgen R: **Transcriptome analysis of sulfur depletion in Arabidopsis thaliana: interlacing of biosynthetic pathways provides response specificity.** *Plant J* 2003, **33**:633–650.

37. Aukerman MJ, Sakai H: **Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes.** *Plant Cell* 2003, **15**:2730–2741.

38. Chen X: **A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development.** *Science* 2004, **303**:2022–2025.

39. Ghildiyal M, Xu J, Seitz H, Weng Z, Zamore PD: **Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway.** *RNA* 2010, **16**:43–56.

40. Meng Y, Shao C, Gou L, Jin Y, Chen M: **Construction of microRNA- and microRNA*-mediated regulatory networks in plants.** *RNA Biol* 2011, **8**:1124–1148.

41. Okamura K, Phillips MD, Tyler DM, Duan H, Chou YT, Lai EC: **The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution.** *Nat Struct Mol Biol* 2008, **15**:354–363.

42. Yang JS, Phillips MD, Betel D, Mu P, Ventura A, Siepel AC, Chen KC, Lai EC: **Widespread regulatory activity of vertebrate microRNA* species.** *RNA* 2011, **17**:312–326.

43. Zhao B, Liang R, Ge L, Li W, Xiao H, Lin H, Ruan K, Jin Y: **Identification of drought-induced microRNAs in rice.** *Biochem Biophys Res Commun* 2007, **354**:585–590.

44. Zhao M, Ding H, Zhu JK, Zhang F, Li WX: **Involvement of miR169 in the nitrogen-starvation responses in Arabidopsis.** *New Phytol* 2011, **190**:906–915.

45. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Marshall KA, *et al*: **NCBI GEO: archive for high-throughput functional genomic data.** *Nucleic Acids Res* 2009, **37**:D885–890.

46. Nakano M, Nobuta K, Vemaraju K, Tej SS, Skogen JW, Meyers BC: **Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA.** *Nucleic Acids Res* 2006, **34**:D731–735.

47. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403–410.

48. Meng Y, Shao C, Chen M: **Toward microRNA-mediated gene regulatory networks in plants.** *Brief Bioinform* 2011, **12**:645–659.

49. German MA, Luo S, Schroth G, Meyers BC, Green PJ: **Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome.** *Nat Protoc* 2009, **4**:356–362.

50. German MA, Pillay M, Jeong DH, Hetawal A, Luo S, Janardhanan P, Kannan V, Rymarquis LA, Nobuta K, German R, *et al*: **Global identification of microRNA—target RNA pairs by parallel analysis of RNA ends.** *Nat Biotechnol* 2008, **26**:941–946.