BMC
Genomics

RESEARCH ARTICLE   **Open Access**

Prediction and identification of natural antisense transcripts and their small RNAs in soybean (*Glycine max*)

Hu Zheng, Jiang Qiyan, Ni Zhiyong and Zhang Hui[*]

Abstract

Background: Natural antisense transcripts (NATs) are a class of RNAs that contain a sequence complementary to other transcripts. NATs occur widely in eukaryotes and play critical roles in post-transcriptional regulation. Soybean NAT sequences are predicted in the PlantNATsDB, but detailed analyses of these NATs remain to be performed.

Results: A total of 26,216 NATs, including 994 *cis*-NATs and 25,222 *trans*-NATs, were predicted in soybean. Each sense transcript had 1–177 antisense transcripts. We identified 21 trans-NATs using RT-PCR amplification. Additionally, we identified 179 *cis*-NATs and 6,629 *trans*-NATs that gave rise to small RNAs; these were enriched in the NAT overlapping region. The most abundant small RNAs were 21, 22, and 24 nt in length. The generation of small RNAs was biased to one stand of the NATs, and the degradation of NATs was biased. High-throughput sequencing of the degradome allowed for the global identification of NAT small interfering RNAs (nat-siRNAs) targets. 446 target genes for 165 of these nat-siRNAs were identified. The nat-siRNA target could be one transcript of a given NAT, or from other gene transcripts. We identified five NAT transcripts containing a hairpin structure that is characteristic of pre-miRNA. We identified a total of 86 microRNA (miRNA) targets that had antisense transcripts in soybean.

Conclusions: We globally identified nat-siRNAs, and the targets of nat-siRNAs in soybean. It is likely that the *cis*-NATs, *trans*-NATs, nat-siRNAs, miRNAs, and miRNA targets form complex regulatory networks.
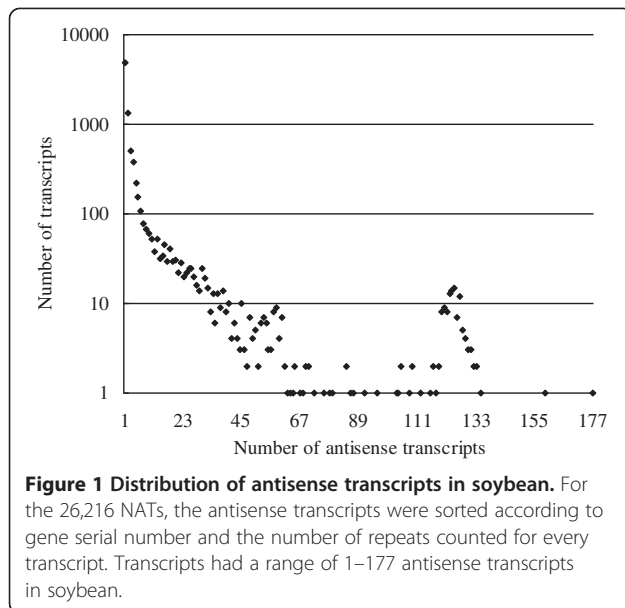
Background

Small RNAs play a crucial role in the regulation of gene expression in eukaryotes [1-5]. They are known to be involved in various aspects of genome stability, development, and response to biotic and abiotic stress [6]. Small RNAs regulate gene expression by modulating mRNA degradation, translational repression, and chromatin modification [1-6]. According to their origin or function in plants, these small RNAs are classified as microRNAs (miRNAs), natural antisense transcript (NAT) small interfering RNAs (nat-siRNAs), *trans*-acting short interfering RNAs, heterochromatic siRNAs, and long small interfering RNAs (lsiRNAs) [7-10].

NATs are a class of endogenous RNAs that have sequences partially, or completely, complementary to each other [11]. Based on their origin, NATs can be classified as either *cis* or *trans*. *cis*-NATs are formed from sense and antisense transcript that is transcribed from the same genomic loci, whereas *trans*-NATs have sense and antisense transcripts derived from different genomic loci [12-16]. NATs form double-stranded RNA (dsRNA) molecules with complementary sequences, and these dsRNAs are processed by Dicer-like proteins to generate nat-siRNAs [9]. These nat-siRNAs can be incorporated into the RNA-induced silencing complex (RISC) and act to guide the cleavage of complementary transcripts [9,17]. A transcript may form more than one trans-NAT with multiple antisense transcripts. These antisense transcripts can also form a trans-NAT with other transcripts. This process demonstrates the complexity of NAT involvement in the regulatory networks at the post-transcriptional level [15]. NATs are involved in numerous biological processes in plants. The expression of NAT genes can be tissue-specific, and many NATs are

* Correspondence: zhanghui06@caas.cn
The National Key Facilities for Crop Genetic Resources and Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China

Figure 1 Distribution of antisense transcripts in soybean. For the 26,216 NATs, the antisense transcripts were sorted according to gene serial number and the number of repeats counted for every transcript. Transcripts had a range of 1–177 antisense transcripts in soybean.

formed in response to environmental stimuli [11,15,18]. Several nat-siRNAs play roles in salt stress, bacterial resistance, cell wall biosynthesis, and fertilization in plants [9,17,19,20].

NATs are widespread in plant cells. In rice (*Oryza sativa*), 23.8% of genes exhibit antisense expression [21]. In *Arabidopsis*, more than 30% of the genome produces transcripts from both strands, and 25% of genes have antisense expression [22]. In bread wheat (*Triticum aestivum*), serial analysis of gene expression using tags revealed that 25.7% of unique genes exhibit antisense transcription [23]. Based on full-length cDNA and genomic data, 1,340 *cis*-NATs and 1,320 *trans*-NATs were predicted and identified in *Arabidopsis* [11,24]. In rice, 344 *cis*-NATs and 7,142 *trans*-NATs were identified to be formed by protein-coding genes [15]. The use of high-throughput sequencing data for small RNAs allowed the construction of a plant NAT database (PlantNATsDB) containing approximately two million NATs from 69

different plant species [25]. NATs and other small RNAs are annotated in the PlantNATsDB based on Gene Ontology categories (http://www.geneontology.org/). A total of 46,367 genes in the PlantNATsDB were used to predict 436 *cis*-NATs and 77,903 *trans*-NATs in soybean (*Glycine max*). However, the details for the soybean NATs remain to be determined.

Here, we report the prediction of 994 *cis*-NATs and 25,222 *trans*-NATs based on 66,213 soybean transcripts downloaded from the Phytozome database (version 1.0; http://www.phytozome.net/index.php) [26]. A total of 21 trans-NATs were identified by RT-PCR amplification. In all, 189,348 small RNAs, 27,465 of which were unique, were derived from 6808 NATs. These small RNAs were found to be enriched in the overlapping regions of NATs. The use of deep sequencing of the degradome is broadly applicable for global identification of small RNA targets [27-30]. Analyses of the soybean degradome database [31,32] identified 446 genes as the targets of 165 nat-siRNAs in soybean. Furthermore, we detected five *trans*-NAT transcripts that can be folded into the stem-loop structures that are characteristic of pre-miRNAs, and identified 86 targets of soybean miRNA that contained antisense transcripts in soybean.

Results and discussion
Prediction of NATs in soybean

We analyzed 66,213 soybean transcripts downloaded from the Phytozome database (http://www.phytozome.net/index.php) [26]. Over 13% (8,634) of the transcripts had at least one antisense transcript in soybean. Among these transcripts, over 50% (4,788) had only one antisense transcript, while the others had from 2 to 177 antisense transcripts (Figure 1). A total of 26,216 NATs were identified in soybean. The NATs were categorized into *cis*-NATs and *trans*-NATs according to the transcript origin from the genomic loci. Mapping of the NAT transcripts to the soybean genome identified 994 *cis*-NATs and 25,222 *trans*-NATs (Additional files 1 and 2).
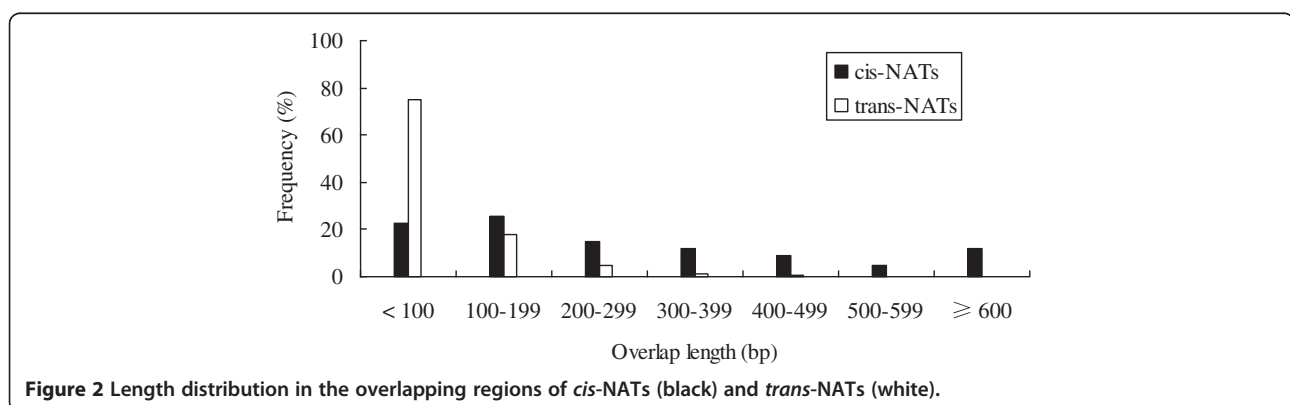

Figure 2 Length distribution in the overlapping regions of *cis*-NATs (black) and *trans*-NATs (white).

Table 1 Genes forming *cis*-NATs with multiple antisense transcripts

Gene 1	Gene 2	Overlap length (bp)	[1]Type
Glyma02g14410.1	Glyma02g14420.1	216	enclosed
	Glyma02g14430.1	333	convergent
Glyma02g36840.1	Glyma02g36830.1	149	divergent
	Glyma02g36850.1	53	convergent
Glyma04g34480.1	Glyma04g34490.1	62	enclosed
	Glyma04g34500.1	113	convergent
Glyma05g22920.1	Glyma05g22910.1	118	divergent
	Glyma05g22930.1	63	enclosed
Glyma08g01140.1	Glyma08g01150.1	69	enclosed
	Glyma08g01160.1	1932	convergent
Glyma13g11820.1	Glyma13g11830.1	428	divergent
	Glyma13g11840.1	593	enclosed
	Glyma13g11850.1	428	enclosed
	Glyma13g11860.1	230	enclosed
	Glyma13g11870.1	192	enclosed
	Glyma13g11880.1	186	enclosed
	Glyma13g11890.1	162	enclosed
	Glyma13g11900.1	271	enclosed
	Glyma13g11910.1	118	enclosed
	Glyma13g11920.1	72	enclosed
Glyma13g11940.1	Glyma13g11950.1	147	enclosed
	Glyma13g11960.1	118	enclosed
	Glyma13g11970.1	542	enclosed
Glyma15g37560.1	Glyma15g37570.1	69	enclosed
	Glyma15g37580.1	45	divergent
Glyma17g31520.1	Glyma17g31510.1	137	convergent
	Glyma17g31530.1	118	divergent
Glyma17g36410.1	Glyma17g36400.1	677	enclosed
	Glyma17g36420.1	304	divergent
Glyma18g03060.1	Glyma18g03050.1	58	convergent
	Glyma18g03070.1	453	divergent

[1]The types of *cis*-NAT were categorized as enclosed (full overlap), convergent (3'-ends overlap), or divergent (5'-ends overlap).

cis-NATs and *trans*-NATs in soybean

The *cis*-NATs can be classified into three types: convergent (with 3'-ends overlapping); divergent (with 5'-ends overlapping); and enclosed (with one transcript completely overlapping the other) [15]. Among the 994 soybean *cis*-NATs, 468 were arranged in the enclosed orientation; 291 were convergent; and 235 were divergent (Additional file 1). In contrast, most of the *cis*-NATs from *Arabidopsis* and rice are convergent [11,15].

cis-NAT overlaps length are usually longer than *trans*-NAT overlaps length [14], and this was also true for soybean NATs. The *cis*-NAT overlaps length ranged from 31–2,808 bp (308 bp average), whereas the *trans*-NAT overlaps length ranged from 31–1,716 bp (87 bp average). The overlapping length of the majority of *trans*-NATs (74.87%) was shorter than 100 bp, and only 7.31% were longer than 200 bp (Figure 2).

Many transcripts have multiple antisense transcripts in plant. For the *cis*-NATs, several genes are involved in two *cis*-NATs in *Arabidopsis* [11]. In soybean, we identified 11 transcripts that formed two or more *cis*-NATs with other transcripts (Table 1). *Glyma13g11820.1* and *Glyma13g11940.1* had ten and three antisense transcripts respectively. The large genomic sequence sizes of *Glyma13g11820.1* (78,178 bp) and *Glyma13g11940.1* (101,408 bp) may help to explain the reason they contained multiple antisense transcripts.

For the *trans*-NATs, one transcript commonly had many antisense transcripts [15,24]. The number of antisense transcripts ranged from 1 to 177 in soybean, possibly a consequence of the homologous genes in the gene families frequently having the same antisense transcript [24]. The soybean genome has gone through at least two rounds of polyploidy and subsequent diploidization events. Segmental duplications and chromosome-level homology are common in the soybean genome [33-36], and approximately 75% of genes have multiple copies [37]. Some transcripts can form both *cis*-NATs and *trans*-NATs [15]. Of the 8,634 transcripts in soybean, 1,200 transcripts were involved in both *cis*- and *trans*-NATs (Figure 3). These genes may be regulated by *cis*- and/or *trans*-NATs.

Identification of NATs in soybean

We identified 17 transcripts using RT-PCR amplification. These 17 transcripts can form 21 trans-NATs. One transcript may form NATs with multiple antisense transcripts [15]. We identified *Glyma01g09920.1*, *Glyma04g05850.1* and *Glyma08g42710.1* as having the same five antisense transcripts. The overlapping region in the sense transcripts had similar sequences (Additional file 3). *Glyma14g13230.1*
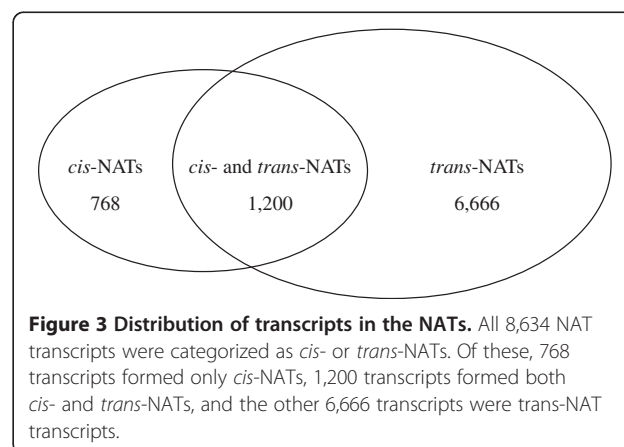

Figure 3 Distribution of transcripts in the NATs. All 8,634 NAT transcripts were categorized as *cis*- or *trans*-NATs. Of these, 768 transcripts formed only *cis*-NATs, 1,200 transcripts formed both *cis*- and *trans*-NATs, and the other 6,666 transcripts were trans-NAT transcripts.

can form NATs with *Glyma02g34100.1, Glyma10g23170.1, Glyma14g22790.1, Glyma18g16420.1, Glyma14g22790.1* and *Glyma20g06230*.1. The overlapping region of *Glyma02g34100.1, Glyma10g23170.1, Glyma18g16420.1* and *Glyma14g22790.1* had the same sequences, while *Glyma20g06230*.1 can form NAT at another location on the *Glyma14g13230.1* transcript.

Small RNAs originating from NATs

As NATs can generate small RNAs [14,15] we searched for the presence of small RNAs in our library for the 8,634 transcripts. We identified 2,286 transcripts able to give rise to small RNAs. 189,348 small RNA sequences, representing 27,465 unique small RNAs, were generated from these transcripts. The most abundant unique small RNAs were 21, 22, and 24 nt in length (Figure 4). In soybean, these 2,286 transcripts could form 179 *cis*-NATs and 6,629 *trans*-NATs (6,808 total; Additional file 4).

Most of the small RNAs were derived from one of the NAT transcripts in *Arabidopsis* [15]. Both *cis*- and *trans*-NATs mostly generated small RNAs from one strand of the NAT in soybean (Figure 5). Among the *cis*-NATs, 75.4% (135) generated small RNAs from only one strand of the NAT, and 9.5% (17) generated small RNAs equally from both transcripts. For the *trans*-NATs, 30.4% (2,019) generated small RNAs from only one strand, and 19.9% (1,321) generated small RNA equally from both strands.

Small RNAs originated from both the overlapping and non-overlapping regions of NATs [15]. The distribution of small RNAs in these two regions varies in different plants [38]. In soybean, the average densities (the number of small RNA loci per kilobase) of the unique and total small RNAs in the overlapping regions were 103.84 and 517.80, respectively, and 48.72 and 344.24 for the entire NATs. T-tests for the unique ($P < 0.0001$) and total ($P < 0.0001$) small RNAs suggested that both were enriched in the overlapping region.
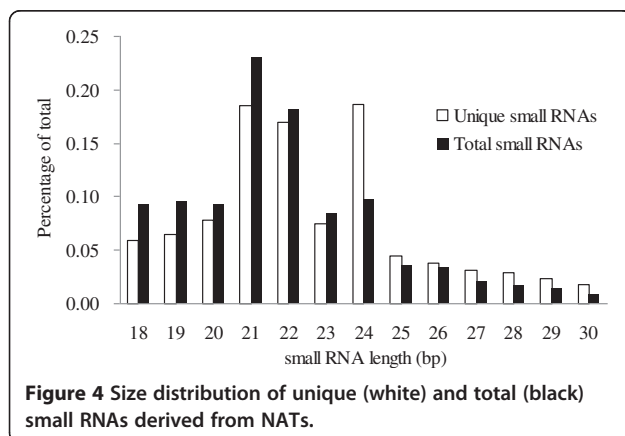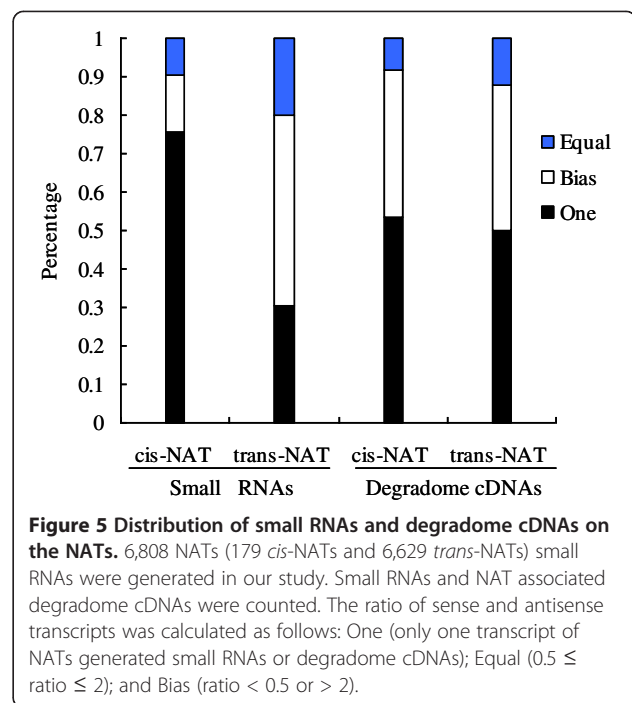

Figure 5 Distribution of small RNAs and degradome cDNAs on the NATs. 6,808 NATs (179 *cis*-NATs and 6,629 *trans*-NATs) small RNAs were generated in our study. Small RNAs and NAT associated degradome cDNAs were counted. The ratio of sense and antisense transcripts was calculated as follows: One (only one transcript of NATs generated small RNAs or degradome cDNAs); Equal (0.5 ≤ ratio ≤ 2); and Bias (ratio < 0.5 or > 2).

The NATs degradome in soybean

NATs can produce small RNAs, which suggests that these transcripts are excised by Dicer-like proteins. We searched for the degradome tags of the 6,808 NATs that could produce small RNAs. A total of 122 *cis*-NAT and 4,425 *trans*-NAT transcripts were identified as having degradomes (Additional file 4). Most degradome tags were derived from one NAT transcript (Figure 5): 53.3% (65) *cis*-NATs, and 50.2% (2,222) *trans*-NATs, generated tags from only one transcript. This was consistent with the small RNA bias towards one strand of NATs.

Identification of NAT-derived small RNA targets in soybean

nat-siRNA can regulate gene expression by guiding target mRNA degradation at the posttranscriptional level [9,19]. The targets of siRNAs can be globally identified by analyzing the degradome [27-32]. We searched the nat-siRNA targets by analyzing the soybean degradome and identified 446 target genes for the 165 nat-siRNAs (Additional file 5). Of these 165 nat-siRNAs, 83 were derived from *trans*-NATs, 81 from *cis*- or *trans*-NATs, and only one was generated from a *cis*-NAT. Regarding the 446 target genes, 203 were targeted by a nat-siRNA derived from the corresponding NAT sense strand, and 75 were targeted by a nat-siRNA produced from the corresponding antisense strand. The nat-siRNAs targets not only the transcript of their own NATs but also that of other transcripts. A total of 168 genes were identified as targets of nat-siRNAs, these nat-siRNAs were not produced from target sense or antisense transcripts.

Figure 4 Size distribution of unique (white) and total (black) small RNAs derived from NATs.

miRNAs may be involved in the formation of NATs in soybean

Some NATs can form stem-loop structures and generate mature miRNAs. In rice, some miRNAs are derived from the overlapping transcript antisense of MADS box transcripts, and act to guide MADS transcript cleavage [39]. We used the UNAfold program to simulate folding of 2,286 transcripts identified as being able to produce small RNAs [40]. Five transcripts were predicted to contain a stem-loop structure characteristic of pre-miRNA (Additional file 6). These transcripts were *Glyma02-g02440.1*, *Glyma04g38430.1*, *Glyma05g03670.1*, *Glyma05-g32980.1*, and *Glyma05g37200.1*. Further analysis revealed that *Glyma04g38430.1* and *Glyma05g32980.1* were miR166 genes; *Glyma05g37200.1* produced miR319; and *Glyma02-g02440.1* and *Glyma05g03670.1* generated small RNAs randomly from both sense and antisense strands (Additional file 7). These five genes may be involved in the biogenesis of both miRNAs and NATs. There are two possible pathways by which miRNAs could be generated from these transcripts. One pathway occurs when the sense and antisense transcripts are co-expressed in the same cell, form a double RNA duplex, and produce nat-siRNAs. This then guides the generation of small RNAs from their sense or antisense transcripts [9]. Another pathway occurs when the sense and antisense transcripts are not co-expressed in the same cell; these transcripts can fold into a hairpin and produce miRNAs.

Targets of miRNAs may be involved in the formation of NATs. We collected 596 candidate targets of miRNAs and searched for targets that could form NATs. 86 miRNA targets were identified as having antisense transcripts (Additional file 8). These targets could form *cis*- and *trans*-NATs. Analysis of the soybean degradome of these 86 targets validated 28 as being miRNA targets [31,32].

NATs may form complex regulatory networks in soybean

It has been suggested that NATs form complex regulatory networks in plants [15]. One transcript often has many antisense transcripts, and these can form NATs with other transcripts. In soybean, 1,200 transcripts were predicted to form both *cis*- and *trans*-NATs (Figure 3). 11 transcripts had multiple *cis*-NATs. Soybean commonly has one transcript that has many antisense transcripts forming the *trans*-NATs. Of the 8,634 transcripts that form NATs, 3,846 contain multiple (2–177) antisense transcripts (Figure 1).

The nat-siRNAs play important roles in plant development. NATs produce nat-siRNAs via a process mediated by Dicer-like RNA-dependent RNA polymerase and Suppressor of Gene Silencing 3. The nat-siRNA is then incorporated into the RISC and directs the cleavage of a complementary mRNA [9,17]. With high-throughput

sequences of small RNAs from different soybean tissues, we detected 6,808 NATs that produced at least one small RNA (Additional file 4). These small RNAs potentially regulate gene expression at the posttranscriptional level. In recent years, deep sequencing of the degradome has been used extensively to globally identify small RNA targets. Analysis of the soybean degradome database enabled identification of 446 genes as targets of 165 nat-siRNAs. These nat-siRNAs targets included NAT sense or antisense transcripts, and other transcripts (Additional file 5).

miRNAs and their targets may be involved in NAT regulatory networks. Five transcripts with pre-miRNA stem-loop structures had antisense transcripts. These transcripts may generate nat-siRNAs or miRNAs; this is dependent on whether the transcripts are co-expressed with antisense transcripts in the same cell. Furthermore, we detected 86 miRNA targets that had antisense transcripts in soybean. These miRNA targets might be regulated by their antisense transcripts.

NATs may form complex regulatory networks in soybean (Figure 6). In these networks, gene expression is regulated by other genes forming *cis*- or *trans*-NATs. NATs can produce nat-siRNAs that self-target their NAT transcripts and other gene transcripts. Some NATs produce miRNAs to regulate expression of other genes, and some miRNAs guide the cleavage of NATs.

Conclusions

We globally predicted NATs in soybean and confirmed the identity of 21 trans-NATs by RT-PCR. The use of high-throughput sequencing of the small RNAs and degradome in soybean enabled the identification of 27,465 unique NAT-derived small RNAs, and 446 targets of 165 nat-siRNAs. The identification of these nat-siRNA targets can help to determine the function of nat-siRNAs in soybean. Furthermore, we identified five pre-miRNAs, and 86 miRNA targets that contained antisense transcripts. NATs, NAT-derived small RNAs, nat-siRNA targets, NAT-related pre-miRNAs, and NAT-related miRNA targets, may form complex regulatory networks. It follows that an understanding of these networks will further our understanding of the roles that NATs play in soybean development.

Methods

Plant material and RNA isolation

Soybean (Glycine max) seeds of the cultivar Williams82 were planted in the experimental station of the Institute of Crop Sciences at the Chinese Academy of Agricultural Sciences, in Beijing in May. Flowers were collected and quickly frozen in liquid nitrogen and then stored at −70°C for use in future RNA isolation. Leaves and roots were collected from 12 days old soybean seedlings. Total RNA from different tissues was isolated separately using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) according to the
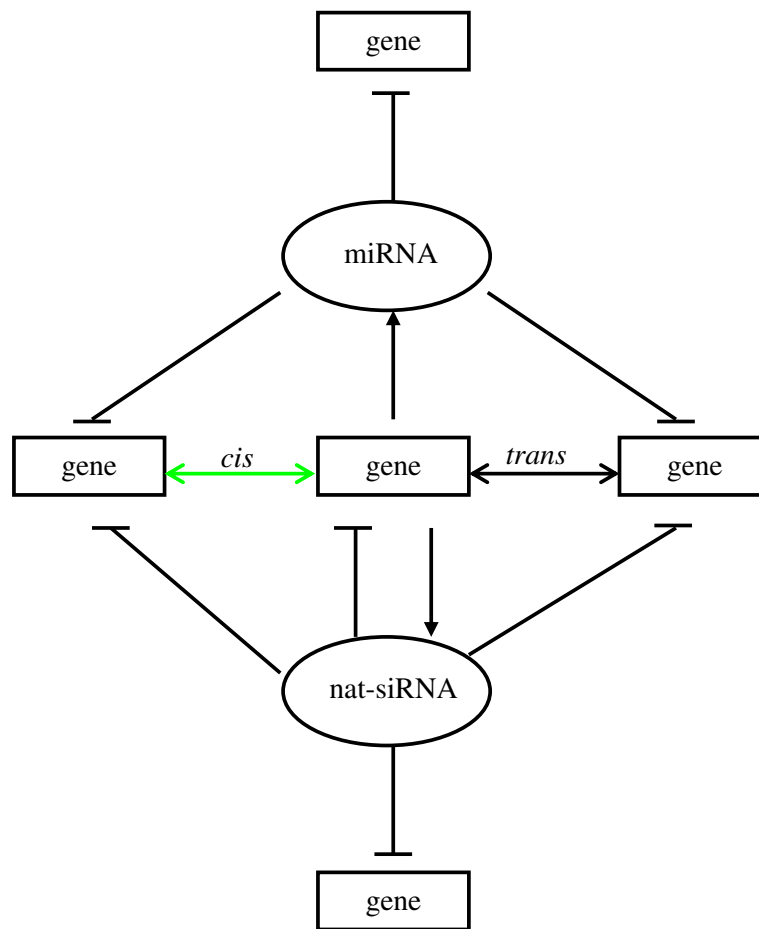
Figure 6 The complex regulatory networks of NATs. In the NAT regulatory networks genes may form *cis*- and *trans*-NATs. Some NATs may fold into the hairpin structure characteristic of pre-miRNAs, and generate miRNAs; some NATs may give rise to nat-siRNAs. The nat-siRNAs can self-regulate the expression of NAT sense or antisense transcripts, and they can target other genes. Additionally, many miRNA targets may be involved in the formation of NATs.

manufacturer's instructions. RNA samples were evaluated by electrophoresis on a 1% agarose gel.

Sequence datasets

Soybean gene sequences and annotations were downloaded from the Phytozome database (version 1.0; http://www.phytozome.net/index.php) [26]. The small RNAs and the degradome were previously identified with deep sequencing in our laboratory. Information for the soybean small RNAs and degradome is from the NCBI-GEO database (accession no. GSE33380). The soybean miRNAs were downloaded from miRBase (version Release 18; http://microrna.org/) [41].

Prediction of NATs in soybean

NATs were detected by aligning predicted *Glycine max* cDNA sequences to each other. If a pair of overlapping genes were matched at opposite strands with an E-value ≤ $1e-9$[19], then they were defined as a NAT pair. The NAT pair was located on the soybean genome to identify *cis*- and *trans*-NATs. If a pair of NATs was located at the same genome locus, they were considered a *cis*-NAT pair. If they were located at different genomic loci, they were considered a *trans*-NAT pair. Based on the overlap between the two transcripts, the *cis*-NATs were categorized into three types: convergent (3'-ends overlap); divergent (5'-ends overlap); and enclosed (full overlap).

Identification of NATs by RT-PCR

Several NATs were identified by use of RT-PCR. We designed gene-specific primers to amplify cDNAs based on their NAT transcript sequence (Additional file 9). 50 μg leaf RNA, 25 μg root RNA, and 25 μg flower RNA were added to a tube and mixed gently, these RNAs were treated with DNase I (Fermentas, Harrington, Ontario, Canada) for 30 minutes at 37°C, and then purified with phenol-chloroform. A total of 4 μg purified RNA was used in a 20 μl RT reaction containing 2 μl gene-specific RT

primer (10 μM) (Additional file 9), 4 μl 5× reaction buffer, 1 μl RiboLock RNase inhibitor (20 u/μl), 2 μl dNTP (10 mM), and 1 μl (200 u/μl) RevertAid M-MuLV reverse transcriptase; this was carried out using the RevertAid First Strand cDNA Synthesis kit (Fermentas, Harrington, Ontario, Canada) according to the manufacturer's instructions. 1 μl of the first strand cDNA sample was used as template for subsequent PCR reactions in 25 μl reactions using gene-specific primers with the following cycle conditions: 95°C, 30 s; 55°C, 30 s; 72°C, 1 min; the run was for 35 cycles. The RT-PCR products were evaluated by electrophoresis on 2% agarose gel. 1 μl of RT-PCR product was ligated into the pGEM-T vector using the pGEM-T easy vector system (Progema, Madison, WI, USA) according to the manufacturer's instructions; next, 2 μl ligation reaction was transformed into TOP10 competent cells. Five clones of each gene were sequenced (Additional file 10).

Analysis of small RNAs
The small RNAs were screened against the Sanger Noncoding RNA Database (http://www.sanger.ac.uk/resources/databases/rfam.html) to eliminate rRNAs, tRNAs, and snoRNAs [42]. Small RNAs that were identical to the transposable elements identified in the *G. max* genome, downloaded from SoyTE (http://www.soybase.org/soytedb/), were also removed. Small RNAs were aligned to NAT transcripts using SOAP [43]. Sequences that were identical to NAT transcripts were considered as NAT-derived siRNAs. The significance of the enrichment of small RNAs in the overlapping regions of NATs was calculated according to the method previously described by Chen et al. [25]. Briefly, the number of unique small RNAs generated from the overlapping region (No) and NAT transcripts (Nt), and their corresponding lengths (Lo and Lt) were determined. The ratios No/Nt and Lo/Lt were used to calculate the density of small RNAs in the overlapping (Do) and entire (Dt) regions of the NAT. The ratio of Do/Dt was considered to be the enrichment score, and a standard χ^2 test was performed to test the significance of the enrichment of small RNAs in the overlapping regions of NATs.

Identification of nat-siRNA targets
Degradomes were mapped onto NAT transcripts using SOAP. The sequences that identically matched NAT transcripts were considered to be the NAT degradome. The locus sequence, containing the 20 bp upstream and downstream regions of the NAT degradome, was extracted as the long degradome from the transcript. Next, a search for small RNA targets was performed as described by Schwab et al. [44]. Total NAT-derived small RNAs were used to query the long degradome sequences, and small RNAs and complementary cDNA pairs for potential target sites were obtained using Patscan set at the default parameters: three mismatches, zero insertions, and zero deletions were permitted [45]. Only hits with fewer than two mismatches in positions 1–9, no mismatches in positions 10 and 11, and fewer than three mismatches after position 11 in the small RNAs were considered good target sequences.

Identification of pre-miRNAs and miRNA targets involved in the NAT network
Flanking sequences of the small RNAs that matched identically with the NAT transcripts were obtained as described by Sunkar and Zhu [46]. Fragment sequences 200 bp upstream and downstream of NAT-derived small RNAs were extracted from the NAT transcripts. Simulation of folding was then performed using UNAfold [40]. Identified secondary structures were checked for miRNA features using MirCheck [47]. The miRNA targets in soybean were predicted using the psRNATarget server (http://plantgrn.noble.org/psRNATarget/) [48]. Validated targets of miRNAs in soybean were obtained from Song et al. and Hu et al. [31,32]. The targets containing the antisense transcripts were considered to be NAT-related miRNA targets.

Additional files

Additional file 1: The *cis*-NATs in soybean. The overlap length and types of *cis*-NATs are shown.

Additional file 2: The *trans*-NATs in soybean. The overlap length is shown.

Additional file 3: Identification of NATs in soybean. The transcripts were amplified by RT-PCR and sequenced. The overlapping regions are shown at the location of the transcript sequence.

Additional file 4: The small RNAs and degradome cDNAs of NATs. Unique and total small RNAs and unique and total degradome cDNAs are shown.

Additional file 5: Identification of nat-siRNA targets in soybean. The degradome, their abundance, and the genes from which they were derived are shown. nat-siRNAs, nat-siRNA target sites, nat-siRNA abundance, and the origin of the nat-siRNAs are indicated. nat-siRNAs can target the sense or antisense transcript of a given NAT and other transcripts. The origins of nat-siRNAs are shown as a, s, and o. a: the nat-siRNA derived from the antisense transcript of a given NAT targets the sense transcript; s: the nat-siRNA derived from the sense transcript of a given NAT targets the sense transcript; o: the nat-siRNA derived from a given NAT guides the expression of another gene. Searches for genes that generate the nat-siRNAs identified nat-siRNAs derived from different genes. Nat-siRNAs that identically matched more than five sense or antisense transcripts are denoted as multiple.

Additional file 6: Secondary structures of five NAT-related pre-miRNAs. Five NAT transcripts had the stem-loop structure that is characteristic of pre-miRNAs. Of these five transcripts, two were miR166 pre-miRNA and one was miR319 pre-miRNA. The miRNAs are indicated in red.

Additional file 7: The small RNAs were matched to the five NAT-related pre-miRNAs. The small RNAs are shown along with information on their length, abundance, and location in the pre-miRNAs.

Additional file 8: The miRNA targets containing antisense transcripts. The miRNA targets could form *cis*- and *trans*-NATs with other genes. a: These targets were validated previously by the analysis of the degradome in soybean [31,32].

Additional file 9: RT-PCR primers used for the amplification of NATs.

Additional file 10: The sequences of transcripts identified by RT-PCR.

Competing interest
The authors declare that they have no competing interests.

Authors' contributions
ZH drafted the initial manuscript. ZH and QJ performed the bioinformatics analysis. QJ performed the small RNA library construction. ZN carried out the tissue collection, RNA extraction and RT-PCR. HZ contributed to the design of the study, obtained the funding, and finalized the manuscript. All authors read and approved the final manuscript.

References
1. Axtell MJ, Snyder JA, Bartel DP: Common functions for diverse small RNAs of land plants. *Plant Cell* 2007, **19**(6):1750–1769.
2. Brodersen P, Voinnet O: The diversity of RNA silencing pathways in plants. *Trends Genet* 2006, **22**(5):268–280.
3. Chapman EJ, Carrington JC: Specialization and evolution of endogenous small RNA pathways. *Nat Rev Genet* 2007, **8**(11):884–896.
4. Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC: Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol* 2004, **2**(5):E104.
5. Vaucheret H: Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev* 2006, **20**(7):759–771.
6. Ruiz-Ferrer V, Voinnet O: Roles of plant small RNAs in biotic stress responses. *Annu Rev Plant Biol* 2009, **60**:485–510.
7. Allen E, Xie Z, Gustafson AM, Carrington JC: MicroRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* 2005, **121**(2):207–221.
8. Lippman Z, Martienssen R: The role of RNA interference in heterochromatic silencing. *Nature* 2004, **431**(7006):364–370.
9. Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK: Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell* 2005, **123**(7):1279–1291.
10. Katiyar-Agarwal S, Gao S, Vivian-Smith A, Jin H: A novel class of bacteria-induced small RNAs in Arabidopsis. *Genes Dev* 2007, **21**(23):3123–3134.
11. Wang XJ, Gaasterland T, Chua NH: Genome-wide prediction and identification of *cis*-natural antisense transcripts in Arabidopsis thaliana. *Genome Biol* 2005, **6**(4):R30.
12. Lapidot M, Pilpel Y: Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO Rep* 2006, **7**(12):1216–1222.
13. Henz SR, Cumbie JS, Kasschau KD, Lohmann JU, Carrington JC, Weigel D, Schmid M: Distinct expression patterns of natural antisense transcripts in Arabidopsis. *Plant Physiol* 2007, **144**(3):1247–1255.
14. Jin H, Vacic V, Girke T, Lonardi S, Zhu JK: Small RNAs and the regulation of *cis*-natural antisense transcripts in Arabidopsis. *BMC Mol Biol* 2008, **9**:6.
15. Zhou X, Sunkar R, Jin H, Zhu JK, Zhang W: Genome-wide identification and analysis of small RNAs originated from natural antisense transcripts in Oryza sativa. *Genome Res* 2009, **19**(1):70–78.
16. Gan Q, Li D, Liu G, Zhu L: Identification of potential antisense transcripts in rice using conventional microarray. *Mol Biotechnol* 2011, **51**(1):37–43.
17. Ron M, Alandete Saez M, Eshed Williams L, Fletcher JC, McCormick S: Proper regulation of a sperm-specific *cis*-nat-siRNA is essential for double fertilization in Arabidopsis. *Genes Dev* 2010, **24**(10):1010–1021.
18. Smith CA, Robertson D, Yates B, Nielsen DM, Brown D, Dean RA, Payne GA: The effect of temperature on natural antisense transcript (NAT) expression in aspergillus flavus. *Curr Genet* 2008, **54**(5):241–269.
19. Katiyar-Agarwal S, Morgan R, Dahlbeck D, Borsani O, Villegas A Jr, Zhu JK, Staskawicz BJ, Jin H: A pathogen-inducible endogenous siRNA in plant immunity. *Proc Natl Acad Sci U S A* 2006, **103**(47):18002–18007.
20. Held MA, Penning B, Brandt AS, Kessans SA, Yong W, Scofield SR, Carpita NC: Small-interfering RNAs from natural antisense transcripts derived from a cellulose synthase gene modulate cell wall biosynthesis in barley. *Proc Natl Acad Sci U S A* 2008, **105**(51):20534–20539.
21. Li L, Wang X, Stolc V, Li X, Zhang D, Su N, Tongprasit W, Li S, Cheng Z, Wang J, *et al*: Genome-wide transcription analyses in rice using tiling microarrays. *Nat Genet* 2006, **38**(1):124–129.
22. Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, *et al*: Empirical analysis of transcriptional activity in the Arabidopsis genome. *Science* 2003, **302**(5646):842–846.
23. Poole RL, Barker GL, Werner K, Biggi GF, Coghill J, Gibbings JG, Berry S, Dunwell JM, Edwards KJ: Analysis of wheat SAGE tags reveals evidence for widespread antisense transcription. *BMC Genomics* 2008, **9**:475.
24. Wang H, Chua NH, Wang XJ: Prediction of trans-antisense transcripts in Arabidopsis thaliana. *Genome Biol* 2006, **7**(10):R92.
25. Chen D, Yuan C, Zhang J, Zhang Z, Bai L, Meng Y, Chen LL, Chen M: PlantNATsDB: a comprehensive database of plant natural antisense transcripts. *Nucleic Acids Res* 2012, **40**:D1181–D1193.
26. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N: Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* 2012, **40**(D1):D1178–D1186.
27. Addo-Quaye C, Eshoo TW, Bartel DP, Axtell MJ: Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. *Curr Biol* 2008, **18**(10):758–762.
28. Addo-Quaye C, Snyder JA, Park YB, Li YF, Sunkar R, Axtell MJ: Sliced microRNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the Physcomitrella patens degradome. *RNA* 2009, **15**(12):2112–2121.
29. German MA, Pillay M, Jeong DH, Hetawal A, Luo S, Janardhanan P, Kannan V, Rymarquis LA, Nobuta K, German R, De Paoli E, Lu C, Schroth G, Meyers BC, Green PJ: Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat Biotechnol* 2008, **26**(8):941–946.
30. Wu L, Zhang Q, Zhou H, Ni F, Wu X, Qi Y: Rice MicroRNA effector complexes and targets. *Plant Cell* 2009, **21**(11):3421–3435.
31. Song QX, Liu YF, Hu XY, Zhang WK, Ma B, Chen SY, Zhang JS: Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. *BMC Plant Biol* 2011, **11**:5.
32. Hu Z, Jiang Q, Ni Z, Chen R, Xu S, Zhang H: Analyses of a Glycine max degradome library identify microRNA targets and microRNA that trigger secondary siRNA biogenesis. *J Integr Plant Biol*. in press. Published abstract.
33. Shoemaker RCP K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP, Kochert G, Boerma HR: Genome duplication in soybean (Glycine subgenus soja). *Genetics* 1996, **144**:329–338.
34. Pagel J, Walling JG, Young ND, Shoemaker RC, Jackson SA: Segmental duplications within the Glycine max genome revealed by fluorescence in situ hybridization of bacterial artificial chromosomes. *Genome* 2004, **47**(4):764–768.
35. Walling JG, Shoemaker R, Young N, Mudge J, Jackson S: Chromosome-level homeology in paleopolyploid soybean (Glycine max) revealed through integration of genetic and chromosome maps. *Genetics* 2006, **172**(3):1893–1900.
36. Van K, Kim DH, Cai CM, Kim MY, Shin JH, Graham MA, Shoemaker RC, Choi BS, Yang TJ, Lee SH: Sequence level analysis of recently duplicated regions in soybean [Glycine max (L.) Merr.] genome. *DNA Res* 2008, **15**(2):93–102.
37. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, *et al*: Genome sequence of the palaeopolyploid soybean. *Nature* 2010, **463**(7278):178–183.
38. Chen D, Meng Y, Ma X, Mao C, Bai Y, Cao J, Gu H, Wu P, Chen M: Small RNAs in angiosperms: sequence characteristics, distribution and generation. *Bioinformatics* 2010, **26**(11):1391–1394.
39. Lu C, Jeong DH, Kulkarni K, Pillay M, Nobuta K, German R, Thatcher SR, Maher C, Zhang L, Ware D, *et al*: Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). *Proc Natl Acad Sci U S A* 2008, **105**(12):4951–4956.

40. Markham NR, Zuker M: **UNAFold: software for nucleic acid folding and hybridization.** *Methods Mol Biol* 2008, **453**:3–31.
41. Kozomara A, Griffiths-Jones S: **MiRBase: integrating microRNA annotation and deep-sequencing data.** *Nucleic Acids Res* 2011, **39**(suppl 1):D152–D157.
42. Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A: **Rfam: annotating non-coding RNAs in complete genomes.** *Nucleic Acids Res* 2005, **33**(suppl 1):D121–D124.
43. Li R, Li Y, Kristiansen K, Wang J: **SOAP: short oligonucleotide alignment program.** *Bioinformatics* 2008, **24**(5):713–714.
44. Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, Weigel D: **Specific effects of microRNAs on the plant transcriptome.** *Dev Cell* 2005, **8**(4):517–527.
45. Dsouza M, Larsen N, Overbeek R: **Searching for patterns in genomic data.** *Trends Genet* 2007, **13**:497–498.
46. Sunkar R, Zhu JK: **Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis.** *Plant Cell* 2004, **16**(8):2001–2019.
47. Jones-Rhoades MW, Bartel DP: **Computational identification of plant microRNAs and their targets, including a stress-induced miRNA.** *Mol Cell* 2004, **14**(6):787–799.
48. Dai X, Zhao PX: **psRNATarget: a plant small RNA target analysis server.** *Nucleic Acids Res* 2011, **39**:W155–W159.