



De Novo Transcriptome and Small RNA Analysis of Two Chinese Willow Cultivars Reveals Stress Response Genes in *Salix matsudana*

Guodong Rao^{2,3}, Jinkai Sui^{2,3}, Yanfei Zeng^{1,2}, Caiyun He^{1,2}, Aiguo Duan^{1,2}, Jianguo Zhang^{1,2*}

1 State Key Laboratory of Tree Genetics and Breeding, Research Institute of Forestry, Chinese Academy of Forestry, Beijing, Republic of China, **2** Key Laboratory of Tree Breeding and Cultivation, State Forestry Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing, Republic of China

Abstract

Salix matsudana Koidz. is a deciduous, rapidly growing, and drought resistant tree and is one of the most widely distributed and commonly cultivated willow species in China. Currently little transcriptomic and small RNAomic data are available to reveal the genes involved in the stress resistant in *S. matsudana*. Here, we report the RNA-seq analysis results of both transcriptome and small RNAome data using Illumina deep sequencing of shoot tips from two willow variants (*Salix matsudana* and *Salix matsudana* Koidz. cultivar 'Tortuosa'). *De novo* gene assembly was used to generate the consensus transcriptome and small RNAome, which contained 106,403 unique transcripts with an average length of 944 bp and a total length of 100.45 MB, and 166 known miRNAs representing 35 miRNA families. Comparison of transcriptomes and small RNAomes combined with quantitative real-time PCR from the two *Salix* libraries revealed a total of 292 different expressed genes (DEGs) and 36 different expressed miRNAs (DEMs). Among the DEGs and DEMs, 196 genes and 24 miRNAs were up regulated, 96 genes and 12 miRNA were down regulated in *S. matsudana*. Functional analysis of DEGs and miRNA targets showed that many genes were involved in stress resistance in *S. matsudana*. Our global gene expression profiling presents a comprehensive view of the transcriptome and small RNAome which provide valuable information and sequence resources for uncovering the stress response genes in *S. matsudana*. Moreover the transcriptome and small RNAome data provide a basis for future study of genetic resistance in *Salix*.

Citation: Rao G, Sui J, Zeng Y, He C, Duan A, et al. (2014) *De Novo* Transcriptome and Small RNA Analysis of Two Chinese Willow Cultivars Reveals Stress Response Genes in *Salix matsudana*. PLoS ONE 9(10): e109122. doi:10.1371/journal.pone.0109122

Editor: Matty Knight, George Washington University School of Medicine and Health Sciences, United States of America

Received: June 14, 2014; **Accepted:** September 8, 2014; **Published:** October 2, 2014

Copyright: © 2014 Rao et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: The research was sponsored by grants from the National Natural Science Foundation of China (31400569), the Fundamental Research Funds for the Central Non-profit Research Institution of CAF (RIF2013-11), Beijing Co-building Plan for Scientific Research and Postgraduate Education (2013K0140, 2013K0141, 2014K0151, 2014K0152), and Collaborative Innovation Plan of Jiangsu Higher Education. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: raoguodong02@163.com

† These authors contributed equally to this work.

Introduction

The increasing concern about climate change and energy security has resulted in the focusing on the economic importance of *Salix* species, given their utility for bioenergy production. However, plants have to tolerate many abiotic and biotic stresses, which are serious threats to agriculture and forestry. These common environmental stresses include salinity, extreme temperatures, drought, chemical toxicity, oxidative stress, pests, and pathogen infection. To exploit their potential for renewable energy, willows need to be kept free of pests and diseases and the yield should be improved without significantly increasing the requirement for fertilizers and water [1].

Previous studies have revealed many *Salix* species have the capacity to tolerate abiotic and biotic stress. The willow species *S. caprea* and *S. cinerea* are grown on nutrient-deficient and industrially-contaminated soils; thus, they serve as biological filters for waste water and sludge disposal [2]. Nine different clones of six species of *Salix* (*Salix cordata* Muhlenb. non Michaux, *S. fragilis*

L., *S. caprea* L., *S. cinerea* L., *S. burjatica* Nazarov. and *S. viminalis* L.) and one hybrid (*S. x calodendron* Wimm.) were exposed to heavy metals in solution culture showed reduced phytotoxicity and increased metal resistance [2]. Two Russian willow species, *Salix viminalis* and *Salix dasyclados* have shown strong genetic control of freezing resistance at during autumn [3]. *Salix caprea*, *Salix repens* and their F1 hybrids had the ability to tolerate different insects and pathogens [4].

Many willow species are characterized by their biological property and ability to adapt to stressful conditions, including high biomass, tolerance to heavy metals, rapid growth, and tolerance of flooding [5]. *Salix matsudana* Koidz. is a deciduous, rapidly growing tree that is native to northeastern China, and is one of the most widely distributed and commonly cultivated willow species in China [6]. The Chinese name for *S. matsudana* is 'Han Liu', where 'Liu' means willow and 'Han' means that the willow is drought tolerant. *Salix matsudana* Koidz. cultivar 'Tortuosa' is widely propagated and planted as an ornamental tree in China

because of its attractive twisted stem. This variant is also a water loving plant, which is often cultivated near the river.

The high-throughput next generation sequencing technology such as RNA-seq and Digital Gene Expression (DGE) is an efficient method to discover genes in an unbiased manner [7–10]. Measurement of mRNA and miRNA expression levels, and elucidation of the regulatory relationship between miRNA and their corresponding mRNA targets are crucial for understanding many pathways and biological systems including plant development, and biotic and abiotic stress responses [11].

However, there have been a few studies probing the ability of *S. matsudana* withstand drought stress. In this study, we performed a transcriptome and miRNA sequencing from one-year-old stem shoots of *Salix matsudana* Koidz. and *Salix matsudana* Koidz. cultivar ‘Tortuosa’ using deep sequencing technology to discover the differentially expressed genes and miRNAs in stem shoots of the two varieties of *Salix*. Our results showed that 196 transcripts, which are mainly related to plant abiotic and biotic stress responses, are commonly up-regulated in *S. matsudana*. These up-regulated genes in *S. matsudana* included a set of biotic and abiotic stress response genes, such as pathogen defense, insect resistance, antibiotics tolerance, antioxidant, and hormone related genes. Functional analysis of miRNA targets also showed that many genes were involved in stress resistance in *S. matsudana*. The transcriptome and small RNAome reported here on the two Chinese willow cultivars significantly enhance our knowledge of stress response genes in *Salix*.

Methods

Plant materials

Shoots from a natural population of *Salix matsudana* and *Salix matsudana* Koidz. cultivar ‘Tortuosa’ were harvested in spring 2013 at Beijing Botanical Garden of China. To ensure that there absolute minimal differences in the growth conditions, all shoots were sampled within the distribution of 30×30 m². Samples were frozen in liquid nitrogen immediately for the extraction of total RNA.

This study was carried out in strict accordance with the recommendations in the Guide for Observational and field studies (<http://www.plosone.org/static/publication>). All necessary permits were obtained for the described field studies. The sampling of all individuals of *Salix. matsudana* and *Salix matsudana* Koidz. cultivar ‘Tortuosa’ were approved by Xiping Zheng, director of Beijing Botanical Garden of China.

Libraries construction and deep sequencing

For each group of samples, equal numbers of shoots from five individual plants were pooled for the total RNA extraction with the TRIZOL reagent according to the manufacturer’ instructions (Invitrogen). Small RNA and transcriptome sequencing, using the total RNA from the different pooled samples. After eliminating traces of genomic DNA by treatment with DNase I, an automated capillary gel electrophoresis was used to assess RNA quality and quantity using a Bioanalyzer 2100 with RNA 6000 Nano Labchips (Agilent Technologies Ireland, Dublin, Ireland). sRNA and Transcriptome sequencing libraries were prepared using an Illumina Small RNA Sample Prep Kit and an Illumina TruSeq RNA Sample Prep Kit, respectively. After the two libraries had been prepared, raw reads generated by using the Illumina HiSeq 2000 were initially processed to get clean reads. Then, all the clean reads were assembled using a de novo assembly program Trinity [12,13].

Gene validation and quantitative real-time PCR analysis

Ten selected genes specific PCR primers (Table S2) were designed corresponding to the conserved region of willow cDNA sequences from the sequenced database. PCR was performed in a total volume of 50 µl containing 2.5 mmol/L Mg²⁺, 0.15 mmol/L dNTPs, 0.4 mmol/L of each primer, 15 ng cDNA and 0.5 U PFU DNA polymerase (NEB). The PCR products were purified and ligated into the PMD19-T vector (Takara Bio), and then transformed into *E. coli* Jm109. After sequencing the positive clones with ABI 3730 (Applied Biosystems, USA), the sequences were validated to the corresponding transcriptome sequences.

Quantitative real-time PCR of mature miRNA was performed following the high-stringency protocol in which a poly A polymerase was used to add a poly A tail. The selected genes real-time quantitative PCR were conducted using the SYBR Green Perfect (Takara) and StepOnePlus™ System (Applied Biosystems). All of the PCR products were sequenced and the dissociation curve was analyzed to verify the amplification specificity. The purified PCR products were used to make standard curve for establishing a quantitative correlation between the CT values and the transcript copy numbers [14]. Each qRT-PCR reaction was repeated at least three times, and each standard curve contains at least 5 points. 5.8 s rRNA and *Populus* actin gene were used for reference gene for miRNA and transcriptome genes expression validation, respectively.

Small RNA prediction and identification of new mi RNAs in willow

Small interfering RNA (siRNA) has their own structure feature, which is a 20–25 nt long double-stranded RNA, and each strand of which is 2 nt longer than the other at the 3’ end. According to this, reads were aligned with each other in order to find potential siRNA candidate by using the alignment software Bowtie combined with Perl scripts. To identify new miRNA, sRNAs with non-annotation information were subjected to a secondary structure prediction by using the software RNAfold [15]. Further identification of new miRNA was conducted by Mireap (<http://sourceforge.net/projects/mireap/>).

Result

Transcriptome sequencing by RNA-seq and de novo assembly

Two RNA-seq libraries were prepared from the total RNA extracted from *Salix matsudana* (‘SM’) and *Salix matsudana* Koidz. cultivar ‘Tortuosa’ (‘SMT’) and were subjected to pair-end read (PE) sequencing using Illumina platform. We removed reads with adaptors, reads with unknown nucleotides larger than 5%, and low quality reads to obtain clean PE reads in SM and SMT samples. A total of 60,521,772 clean PE reads consisting of 6,506,875,453 nucleotides (nt) with an average GC content of 45.11% were obtained in the SM sample while 58,648,772 clean PE reads consisting of 6,292,916,298 nt with an average GC content of 45.14% were obtained in the SMT sample (Table 1). An overview of the raw reads data is given in Figure S1. All high-quality clean reads were assembled into 106,430 contigs with an average length of 944 bp having a minimum length of 201 bp and a maximum length of 17,226 bp (Figure 1A). The contigs were further joined into 48,817 unigenes with an average length of 773 bp, a N50 length of 1,468 bp and a N90 length of 294 bp (Figure 1B).

Table 1. Statistics of transcriptome output sequencing.

Samples	SM	SMT
Total raw reads	61,958,422	60,131,098
Total clean reads	60,521,772	58,648,772
Total clean nucleotides (nt)	6,506,875,453	6,292,916,298
Q20 percentage	97.55%	98.14%
N percentage	0.06%	0.10%
GC percentage	45.11%	45.14%

doi:10.1371/journal.pone.0109122.t001

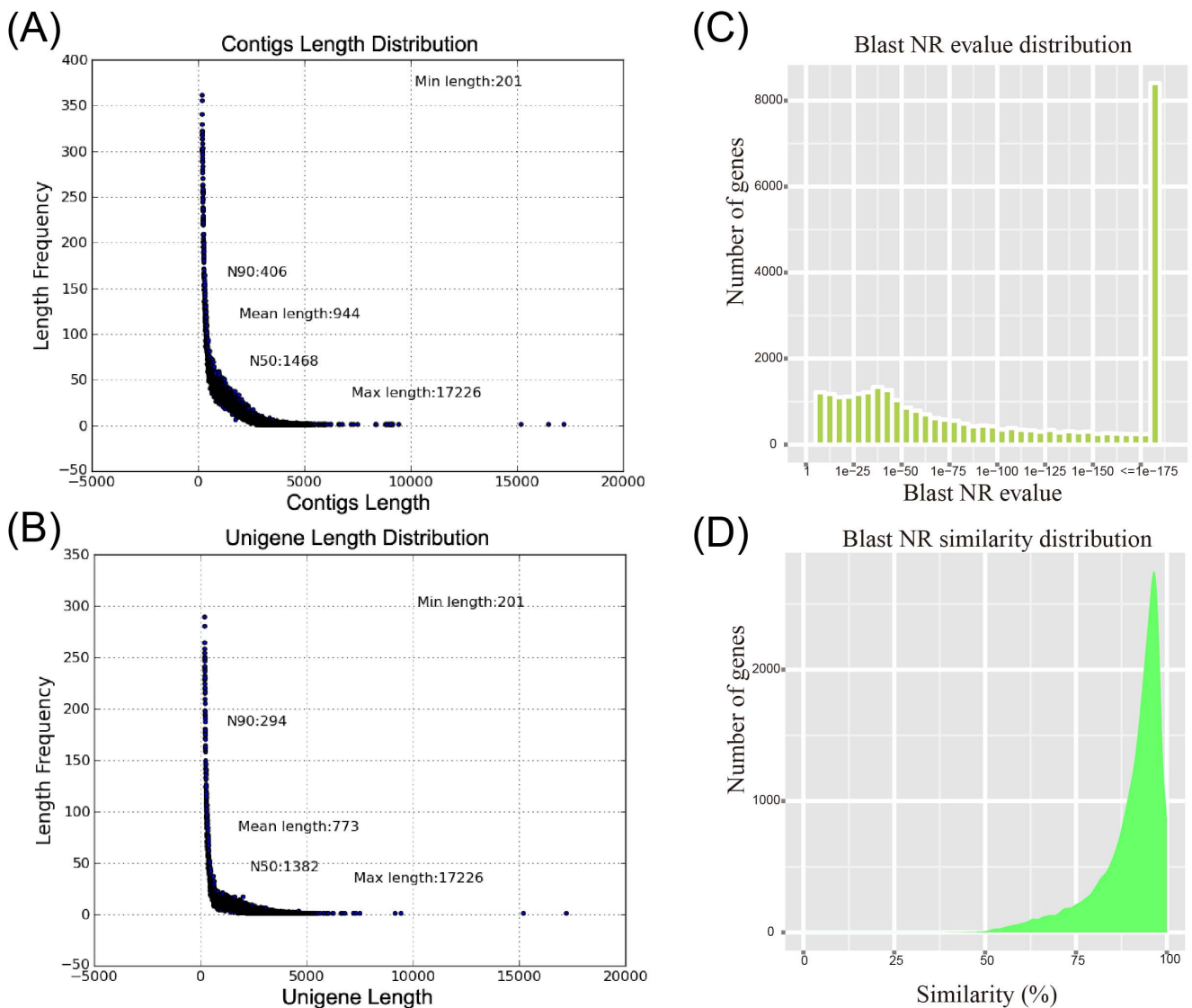


Figure 1. The length distribution of contigs and unigenes, and E-value distribution and similarity distribution of Blast NR hits. Minimum length, N90; mean length, N50; and maximum length of contigs and unigenes are shown in their corresponding positions. (A) Contigs length distribution. (B) Unigene length distribution. (C) E-value distribution of Blast NR hits for matched unigene sequences. (D) Similarity distribution of Blast NR hits for each unigene.

doi:10.1371/journal.pone.0109122.g001

Functional annotation and classification of the assembled unigenes

Gene annotation was performed based on sequence homologies to the databases of NCBI non-redundant protein (Nr), NCBI nucleotide sequence (Nt), Kyoto Encyclopedia of Genes and Genomes (KEGG), Swiss-Prot protein, Protein family, Gene Ontology (GO), and Cluster of Orthologous Groups (COGs) using BLAST2GO. In total, 2,592 (5.2% of all unigenes) unigenes annotated in all databases, and 34,049 (69.74% of all unigenes) significantly matched a sequence in at least one of the public databases (Table 2). For the nr annotations, 29,657 of unigenes (60.75%) were found to have matches in the database. Further analysis of the BLAST data showed that 64.11% of the top hits had strong homology with the E-value $<1.0e^{-50}$ while 35.89% of the matched sequences showed moderate homology with the E-value between $1.0e^{-5}$ and $1.0e^{-50}$ (Figure 1C).

The identity distribution pattern showed that 79.44% of the sequences had a similarity higher than 80% (Figure 1D). The majority of the annotated sequences corresponded to the known nucleotide sequences of plant species with 83.14%, 6.46%, 2.87%, 0.63%, and 0.36% matches with *Populus trichocarpa*, *Ricinus communis*, *Vitis vinifera*, *Glycine max*, and *Medicago truncatula* respectively.

The size distribution of the BLAST-aligned coding sequence (CDS) and predicted proteins are shown in Figure S2A, B, respectively. The remaining 41.49% (20,256/48,817) of unigenes that did not match sequences in the databases were analyzed by ESTScan to predict coding regions. An additional 12,740 unigenes (62.89%) also showed orientation in the transcriptome coding sequence (Figure S2C, D). The sequences without a homologous hit may represent novel genes specifically expressed in willow stem shoots, or they could be attributed to other technical or biological biases such as assembly parameters. Furthermore, some cDNA are non-coding, lineage-specific or highly variable and need to be further verified.

GO annotation is an international classification system that can provide standardized vocabulary for assigning functions of the uncharacterized sequences [16]. A total of 23,249 unigenes (47.62% of all the assembled unigenes) were assigned at least one GO term. GO annotation assignments classified 21,577 unique contigs into 22 subcategories of the biological process category, 11 subcategories of the molecular function category, and 14 subcategories of the cellular component category at level 2 (Figure 2). Among biological processes, transcript sequences

assigned to biological regulation, cellular process, metabolic process and single-organism were the most abundant. The subcategories with the most highly abundant transcripts in the cellular components category included cell, cell part, macromolecular complex, membrane, and organelle. Within the molecular function category, the majority of the GO terms were predominantly assigned to the binding and catalytic activity.

The database of COGs is a phylogenetic classification of protein encoded in completely sequenced genomes [17]. Overall, 10,073 of 48,817 (20.63%) unigenes were assigned to the COG classification (Table 2). Among the 26 COG categories, the cluster for 'general function prediction only' (1,663, 14.66%) associated with the basic physiological and metabolic functions represented the largest group, followed by post-translational modification, protein turnover, chaperons (1,376, 12.13%), signal transduction (840, 7.40%), 'translation (812, 7.15%), and intracellular trafficking and secretion (577, 5.08%). However, only a few unigenes were assigned to cell motility (5, 0.04%), extracellular structures (37, 0.33%), and nuclear structure (49, 0.43%). One unigene was annotated as 'unnamed protein' (Figure 3A).

KEGG is a database resource that integrates genomic, chemical and systemic functional information that can facilitate to understand the biological functions of genes in terms of networks [18]. The assembled unigenes were annotated with KEGG Orthology (KO) numbers using BLASTx alignments against KEGG with a cut-off E value of 10^{-5} . Overall, 5,444 of 48,817 (11.15%) unigenes were matched in the database and were assigned to 242 KEGG pathways. Among the five categories of pathway hierarchy 1, the 'metabolism' represented the largest group, which were sorted into 11 subcategories including carbohydrate metabolism, energy metabolism, amino acid metabolism, lipid metabolism, nucleotide metabolism, and some others (Figure 3B).

Detection of differentially expressed genes between two transcription libraries

A quality control test for the data assembled from each cDNA library confirmed that they were suitable for statistical analysis for differentially expressed genes identification (Figure S3). A total of 292 differentially expressed genes were detected between SM and SMT groups when \log_2 (fold change) >1 and $P < 0.005$ were used as cutoff values. Further analysis showed that 196 genes involved in stress resistance, pathogen defense, antibiotics tolerance, insect resistance, antioxidant, hormone related and ribosomal RNAs

Table 2. statistics of functional annotation of unigenes in public databases.

Public database	Number of Unigenes	Percentage (%)
Annotated in NR	29,657	60.75
Annotated in NT	30,551	62.58
Annotated in KEGG	5,444	11.15
Annotated in SwissProt	19,354	39.64
Annotated in PFAM	17,869	36.6
Annotated in GO	23,249	47.62
Annotated in KOG/COG	10,073	20.63
Annotated in all Databases	2,592	5.3
Annotated in at least one Database	34,049	69.74
Total Unigenes	48,817	100

doi:10.1371/journal.pone.0109122.t002

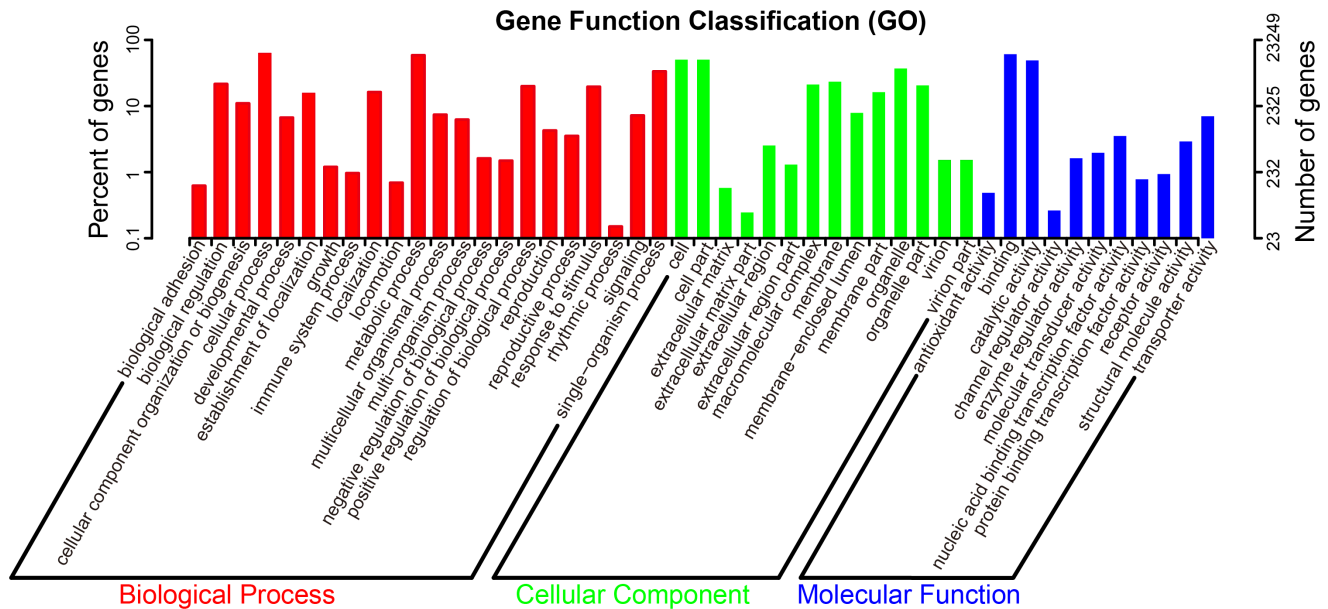


Figure 2. Number of unigenes in each functional category. Unigenes were classified into different functional groups based on a set of plant specific GO terms within biological process, cellular component, and molecular function. doi:10.1371/journal.pone.0109122.g002

were up regulated (Figure 4A), whereas 96 genes involved in cell wall synthesis, photosynthesis, stress resistance and antibiotics tolerance were down regulated in SM as compared to SMT. Most of the differentially expressed genes are related to plant adaption stress which includes drought, cold, salt, pathogen, insect, and antibiotics. Among the 196 up regulated genes, 133 genes (67.86%) are related to stress adaption indicating that *S. matsudana* has a better ability to tolerant harsh environment.

Deep sequencing of small RNA

Two libraries were constructed for small RNA deep sequencing, and we identified 12,532,326 and 20,113,588 sequence reads respectively from SM and SMT. Among the two data sets, the most abundant classes of small RNA found were 21 and 24 nt, which concurs with most reports of deep sequencing of small RNA in plants (Figure 4B) [19–21]. The sRNA were classified into eight categories (miRNA, tRNA, rRNA, snRNA, snoRNA, ta-siRNA, repeats, and others) by comparing the acquired to miRBase (<http://microrna.sanger.ac.uk/sequences>), Rfam (<http://rfam.sanger.ac.uk/>) and Genbank (<http://www.ncbi.nih.gov/Genbank/>) (Figure S4). Besides the unknown sRNA, miRNA made up 16.25% and 15.94% of the total acquired sequences respectively, which were represented in the main by 21 nt small RNAs. There was little variation in the different sRNA categories between SM and SMT (SD = 2.31%) indicating that sRNA categories are relatively stable in the two willow species.

Known miRNA and comparison of miRNA expression levels between the two libraries

A total of 166 known miRNAs belonging to 35 families were obtained in SM and SMT. These miRNA exhibited variable abundance. Deep sequencing analysis is a powerful strategy that allows the identification and quantification of differences in the expression of sRNA. In the 35 families, miR159 was the most accumulated miRNA with a total of 410,956 reads detected between the two libraries (87,687 miRNAs in SM and 323,269 in SMT) (Table 3). There was considerable expression level diversity

between the families probably because the willow one-year-old stem is a highly differentiated organ and the genes are in the process of dynamic expression during development. For instance, miR159, miR166, miR319, and miR396 were sequenced more than one hundred thousand times while miRNAs, miR399, miR408, miR481, miR1446, and miR6459 were detected less than 10 times. Members of the same family were also not expressed equally. The two libraries, SM and SMT, had 36 miRNAs that were differentially expressed, among which 24 were up regulated, and 12 were down regulated in SM compared to SMT.

Identification of new miRNAs

Plant miRNA research has significantly progressed because of deep sequencing technology; however, stricter requirements are needed for identifying a miRNA candidate as a reliable new miRNA. Altogether, we obtained 31 predicted novel miRNAs by using two additional criteria [22]. Among these novel miRNAs, 28 and 30 were expressed in SM and SMT, respectively. The majority of the miRNAs (18 of the 31, 58.06%) have a uridine at 5' terminal, showing a preference of AGO1 [23]. The novel miRNAs also displayed unequal expression levels as known miRNAs between the two libraries (Table S1). As none of the novel miRNAs had homologs in miRBase, we presumed that they were *S. matsudana*-specific, or were restricted to closely related species.

miRNA Targets predicting and sequences validating

Transcriptome of SM and SMT was used to identify miRNA targets. The 35 known miRNA families had 108 affiliated target genes while the 31 novel miRNAs did not target any genes. Functional annotation of the targets was found to be diverse and included the response to stimulus, signal transduction factors, transcription factors, and genes involved in other biological processes. Among the different expressed targets, 24 miRNAs-target pairs showed reverse expression changing pattern when the results from miRNA profiling and the transcriptome profiling were

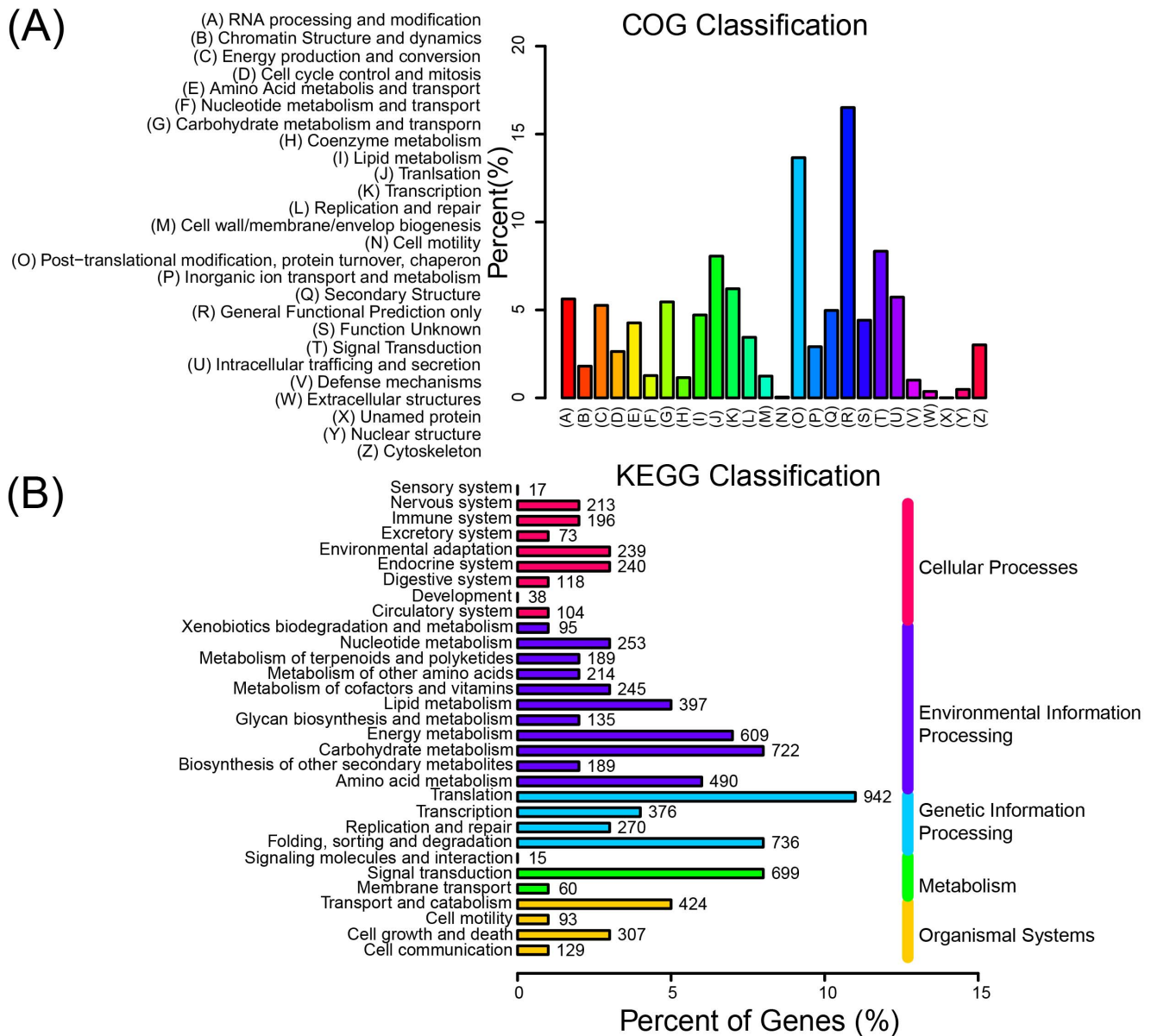


Figure 3. COG and KEGG function classification of unigenes. (A) COG function classification of unigenes shows 26 different functional groups. (B) Five subcategories (cellular processes, environmental information processes, genetic information processes, metabolism, and organismal systems) were classified by KEGG function classification.
doi:10.1371/journal.pone.0109122.g003

compared (Figure 5). Functional annotations showed that the 24 miRNA-targets pairs were involved in auxin signaling, stress tolerance, signal transduction factors, and other proteins were involved in various biological processes.

To validate the reliability of RNA-seq, full-length cDNA sequences of twenty selected genes were isolated by T-A cloning and compared with the assembled sequences. The length of these genes ranged from 192 to 1,960 bp, and the sequence variation was minimal. All of the genes had the sequence pairwise identity bigger than 98% (Table S2), which validated the reliability of RNA-seq. To partially validate the miRNA and its target genes expression levels, nine miRNA-target pairs were randomly selected for a real-time qPCR experiment. The results showed that the expression of miRNA target pairs changed trends as compared with the sequencing results (Figure 6).

Discussion

Overview of the deep sequencing datasets

Although many studies have reported the transcriptome analysis of willow, here have been no studies on mRNA and miRNA expression profiles in *S. matsudana*. In this study, the RNA-seq technology was used to generate transcriptome and small RNAome data and examined the global gene expression profiling of shoot tips from *S. matsudana*. This work demonstrates that RNA-seq is a useful and effective tool for *de novo* transcriptome and small RNA assembly. Moreover, it facilitates the identification of differentially expressed miRNAs and their target genes between the two willow cultivars, even in the absence of a genomic database. We generated both mRNA and siRNA libraries from the shoot tips of the two willow species to substantially increase the

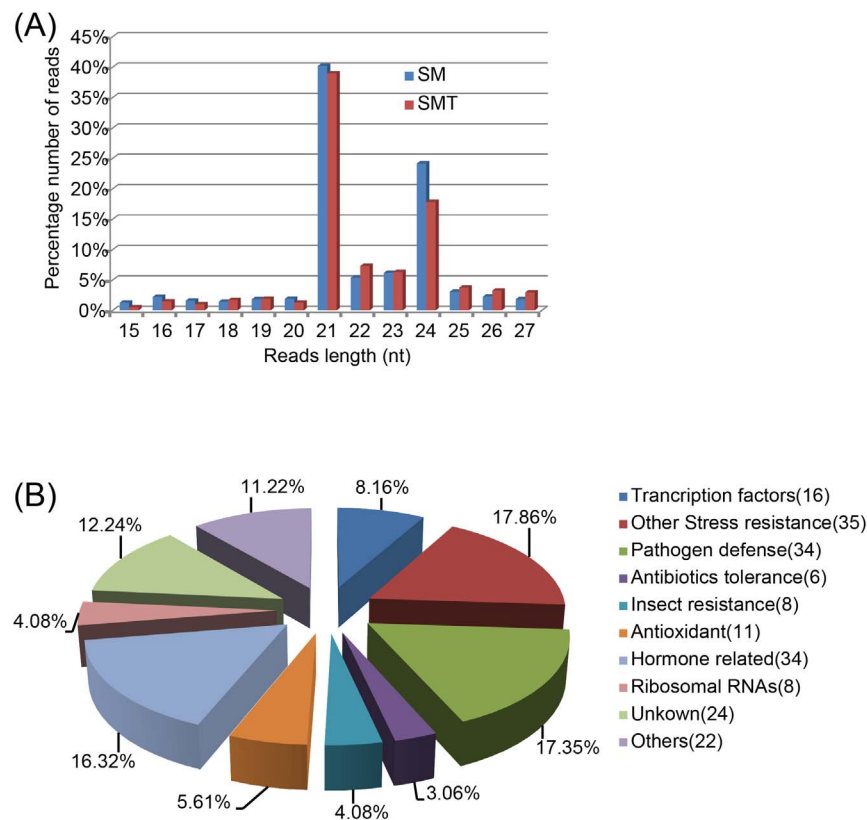


Figure 4. Classification of DEGs and sequence length distribution of small RNA libraries of SM and SMT. (A) Up-regulated genes in SM compared to SMT. (B) Sequence length distribution shows 21 and 24-nucleotides reads were of significant greater proportion than others. doi:10.1371/journal.pone.0109122.g004

available data on willow mRNAs and sRNAs to obtain genomic resources for further investigating candidate genes especially the stress resistance genes in various metabolic pathways in willow species. Comparison of transcriptome sequence data of the two libraries revealed 292 differentially expressed genes with over 67% involved in stress response. Differential expression of miRNAs also showed many miRNAs and their target genes were stress related. The real-time PCR analysis of the differentially expressed miRNAs and genes further verified the miRNA and the transcript expression level revealed by small RNA and transcriptome comparison from RNA-seq data.

Transcription factors related to stress tolerance in *S. matsudana*

Transcription factors play a key role in modulating the acclimation response of plants to various internal or external cues. They potentially control downstream gene expression in stress signal transduction pathways through activation and repression of the genes after certain stress treatments. Plant genomes contain a large number of transcription factors (TFs). It has been estimated that the *Arabidopsis* genome codes for at least 1,533 transcription factors and over 5.9% of its total estimated genes. Among these, about 800 genes encode AP2/EREBP, bZIP, zinc finger proteins, Myb proteins, and AUX/IAA types of transcription factors. Sixteen transcription factors have higher transcription levels in SM compared to SMT, which include C-repeat binding factors (CBF), Myb proteins, TINY proteins, AP2/EREBP, WRKY type, and other types of zinc finger proteins.

CBF proteins bind to the cold-responsive *cis*-acting elements that contain the same motif (CCGAC), are encoded by AP2/EREBP multigene families, and mediate the transcription of genes in response to stress. Overexpression of two *Arabidopsis* AP2/EREBP genes, CBF1 and DREB1A, resulted in stronger tolerance to drought, salt, and freezing [24–26]. Transgenic *Arabidopsis* plants overexpressing CBF3 under the control of cauliflower mosaic virus (CaMV) 35S promoter also showed a high tolerance to drought, high-salinity, and freezing stress [25,27,28]. Five CBF and AP2/EREBP genes were highly expressed in SM as compared to SMT. The sequenced CBF and AP2/EREBP gene reads in SM are 2 to 30 times over that in SMT indicating that the SM may have a high tolerance to abiotic stress.

WRKY family is defined by a DNA binding domain that contains the strictly conserved amino acid sequence WRKY. Members of WRKY family transcription factors have been implicated in the response to biotic and abiotic stresses. Overexpression of the GhWRKY39-1 gene enhances the resistance to pathogen infection and tolerance to high salt and oxidative stress in transgenic *Nicotiana benthamiana* [29]. The same results were obtained from the research of *Tamarix hispida*, overexpression of ThWRKY4 conferred tolerance to salt, oxidative and ABA treatment in transgenic plants [30]. The overexpression of a subgroup IIe WRKY transcription factor CaWRKY27 of *Capsicum annuum* also positively regulates the resistance of tobacco to the pathogen *Ralstonia solanacearum* infection [31]. In the present study of the two transcriptome libraries, three WRKY showed higher expression levels in SM as compared to SMT that may enhance the tolerance of SM to biotic and abiotic stress.

Table 3. Conserved miRNA families expression in SM and SMT.

Family	Mature sequence	Total	SM		SMT	
			Raw	TPM	Raw	TPM
miR156	UUGACAGAAGAUAGAGAGCAC	23	5	18	18	42
miR159	UUUGGAUUGAAGGGAGCUCUA	410,956	87,687	323,269	175,659	406,596
miR160	UGCCUGGCUCUUGAUGCCA	905	193	712	181	419
miR162	UCGAUAAACCUUGCAUCCAG	7,920	1,690	6,230	909	2,104
miR164	UGGAGAAGCAGGGCAGUGCA	1,237	264	973	112	259
miR166	UCGGACCAGGCUUCAUCCCC	140,711	30,024	110,687	28,821	66,712
miR167	UGAAGCUGCCAGCAUGAUCU	1,219	260	959	92	213
miR168	UCGCUUGGUCAGGUCGGAA	4,312	920	3,392	374	866
miR171	UGAUUGAGCCGUGCCAAUUC	366	78	288	37	86
miR172	AGAAUCCUGAUGAUGCUGCAG	984	210	774	127	294
miR319	UUGGACUGAAGGGAGCUCCC	335,098	71,501	263,597	154,043	356,562
miR390	AAGCUCAGGAGGGAUAGCGCC	1,875	400	1,475	201	465
miR393	UCCAAAGGGAUCGCAUUGAUC	347	74	273	15	35
miR394	UUGGCAUUCUGUCCACCUCC	7,217	1,540	5,677	144	333
miR396	UUCCACAGCUUUCUUGAACUU	153,993	32,858	121,135	14,009	32,427
miR398	UUCUCAGGUCACCCUUUGGG	951	203	748	983	2,275
miR399	UGCCAAAGGAGAUUUGCCCGG	5	1	4	1	2
miR403	UUAGAUUCACGCACAACUCG	17,261	3,683	13,578	233	539
miR408	AUGCACUGCCUUCUCCUGGC	5	1	4	0	0
miR472	UUUCCUACUCCACCCAUCCC	22,304	4,759	17,545	2,568	5,944
miR475	UUACAAUCCAUUGAUUAAACCGU	19	4	15	11	25
miR481	AGGACCUCACCUAACAGCUUAAGC	9	2	7	0	0
miR482	UUGCCUACUCCACCAUGCCAC	731	156	575	102	236
miR530	UGCAUUGGACCUUGCACCUUC	506	108	398	53	123
miR828	UCUUGUCUAAAUGAGAUUCCA	37	8	29	5	12
miR1444	UCCACAUUCGGUCAUUGUUC	42	9	33	3	7
miR1446	UUCUGAACUCUCUCCUCAAA	5	1	4	0	0
miR2111	UAAUCUGCAUCCUGAGGUUUG	23	5	18	4	9
miR6421	UAGAGCAGAUUGUAAGGGAAG	28	6	22	24	56
miR6427	UCGUAAUGCUUCAUUCACAAA	173	37	136	40	93
miR6445	UUCAUUCUUCUUCUAAAUGG	8,586	1,832	6,754	1,215	2,812
miR6457	UAAUCUCUCUGCAGAAUGCUG	0	0	0	28	65
miR6459	UCGAAUUUGGGCUUGAGAUUG	9	2	7	2	5
miR6474	UGUUCAGAUAGAUAGCA	80	17	63	21	49
miR6476	UCAGUGGAGAUAAACAUGA	0	0	0	1	2

doi:10.1371/journal.pone.0109122.t003

Plant zinc finger transcription factors play an important role in tolerance to various environmental stresses such as drought, cold, osmotic stress, wounding and mechanical loading. The zinc finger proteins characterized by zinc finger domains enable protein interaction with DNA. A large number of studies have shown that zinc finger proteins may not be specific to one particular stress, but may regulate responses to several stresses. In *A. thaliana*, the zinc finger proteins have been shown to be involved in cold, dehydration, reactive oxygen, and salt stress [32,33]. In *Populus euphratica*, the overexpression of a zinc finger protein gene *PSTZ* showed enhanced salt tolerance [34]. In rice, seven zinc finger transcription factors were involved in the response to different abiotic stresses such as cold, drought, and mechanical stress [35]. In *Capsicum annuum*, a zinc finger transcription factor gene

functions as a pathogen induced early-defense gene [36]. Six zinc finger transcription factors have a higher transcription level in SM than SMT indicating that the SM has a stronger ability to tolerate abiotic stresses.

Hormones related genes involved in stress tolerance in *S. matsudana*

Plants hormones such as auxins, abscisic acid (ABA), gibberellins (GA), salicylic acid (SA), ethylene (ET), cytokinins (CK), jasmonates (JA), and brassinosteroids (BR) play crucial roles in various biotic and abiotic responses. A vast array of pathogenic microorganisms which deliver virulence factors into plant cell to promote virulence and cause disease and plant can change the

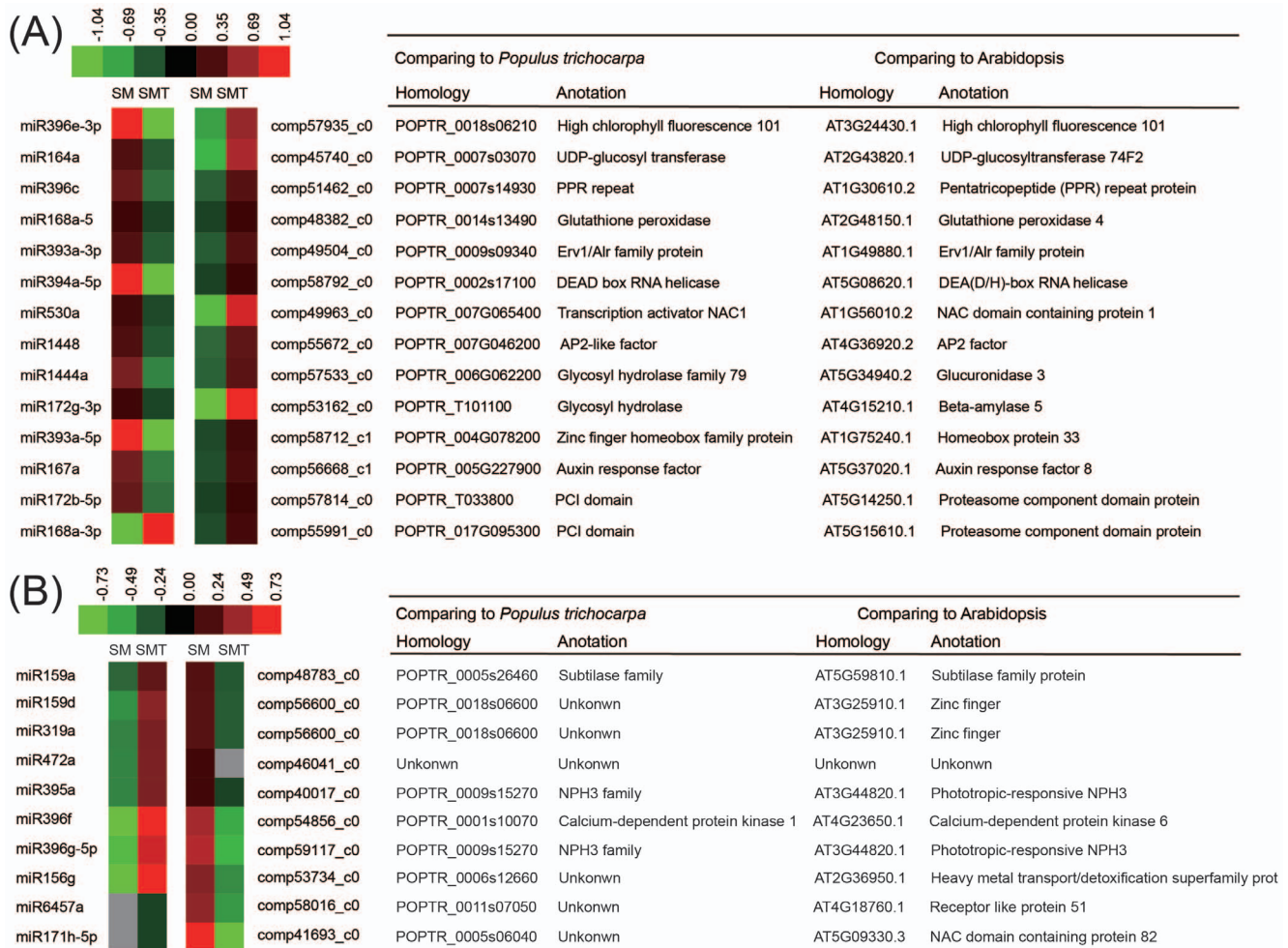


Figure 5. A combined view of reverse expressions between miRNA and its target. The left side of hotspot figure show miRNA expression level, and the right side show corresponding target expression levels of both SM and SMT. Up and down regulation in the expression were based on normalize data (color bar at the top) generated by Cluster 3.0 software. The tables show homologous genes of miRNA target genes by comparing to *Populus trichocarpa* and Arabidopsis. **(A)** Reverse expression of Up-regulated miRNA in SM and its targets. **(B)** Reverse expression of down-regulated miRNA in SM and its targets.

doi:10.1371/journal.pone.0109122.g005

level of various phytohormones to protect the cell from infection when plants were attacked by these diverse pathogens [37,38]. Plants hormones also have the ability to adapt changing environments by mediating a wide range of adaptive responses [39,40]. It is notable that xx hormone-related transcripts have significantly higher expression levels in SM than in SMT, which include eighteen ABA related genes, ten auxin response genes, four JA, and two ET related genes.

Under abiotic stresses such as drought, cold, and salinity, ABA levels increased in vegetative tissues, which are adaptive responses essential for the plant survival and productivity [41]. In Arabidopsis, increasing the transcription of a key ABA biosynthesis gene, *NCED3*, resulted in the increased production of ABA and enhanced tolerance to drought-shock and osmotic stress [42]. Overexpression of *AITRE1* gene that encodes the Arabidopsis trehalase conferred the ability to tolerate drought stress through increased sensitivity toward ABA-dependent stomatal closure [43]. A maize calcium-dependent protein kinase gene, *ZmCPK4*, positively regulates ABA signaling and enhanced drought stress tolerance in transgenic Arabidopsis [44]. Apart from the vital function for abiotic stress tolerance, ABA plays a modulating role

in plant pathogen infections [45–47]. Many key components of ABA signal transduction have been identified including protein kinases, RNA processing factors, phosphatases, proteasome components, and chromatin remodeling proteins [48,49]. In *S. matsudana*, thirteen genes involved in the response to ABA transduction had a higher transcription level than SMT. Moreover, five miRNAs, which target five ABA transduction genes were down regulated as compared to SMT. These genes could increase the ABA levels and enhance the ability of *S. matsudana* to tolerate abiotic and biotic stresses.

Auxin responsive genes also regulate plant stress response. Overexpression of an auxin responsive gene *GH3-8* resulted in stronger pathogen resistance in rice and this resistance was shown to be independent of SA and JA signaling [50]. A rice auxin responsive gene *OsGH3.13*, encoding IAA-amido synthetase was shown to enhance the expression of genes that confer drought tolerance [51]. Ten genes related to auxin response had higher transcription levels in SM than SMT, and one miRNA had a lower expression level in SM than SMT. These genes could change the auxin levels and confer increased stress tolerance ability upon *S. matsudana*. Auxin plays a major role of in

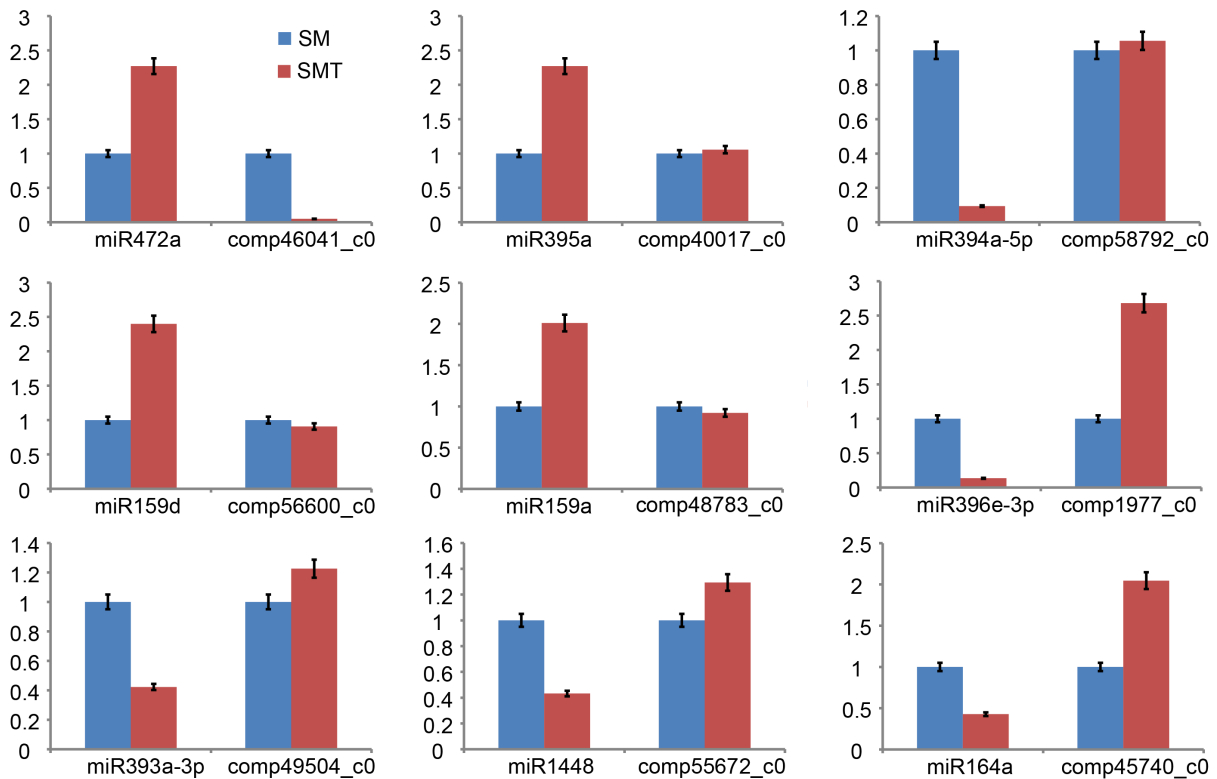


Figure 6. Real-time quantitative PCR validate expressions of 9 selected miRNA and its targets. The amount of miRNA expression was normalized to the level of 5.8rRNA, and the mRNA expression was normalized to the level of *Populus* actin gene. The normalized miRNA and its target gene levels of SM were arbitrarily set to 1. doi:10.1371/journal.pone.0109122.g006

integrating the activities of multiple phytohormones to control plant growth in response to environmental signals [52]. Except ABA and auxin, JA and ET related genes were also detected that had different transcription in SM and SMT, which in turn could result in variable stress response activity. Four JA related genes and two ET related genes had higher transcript levels in SM than SMT, which could enhance the ability of *S. matsudana* to tolerance stress.

Antioxidants and detoxification genes related to stress tolerance in *S. matsudana*

Biotic and abiotic stresses such as pathogen infection, salt, drought, oxidation and heat stress are accompanied by the formation of reactive oxygen species (ROS). The ROS such as H_2O_2 , O_2 , and OH^\cdot are the main damages of membranes and macromolecules [53]. Plants have developed many antioxidation strategies to scavenge ROS. Enzymes such as thioredoxin, ascorbate peroxidase (APX), superoxide dismutase (SOD), glutathione peroxidase, glutathione transferases, phospholipid-hydroperoxide glutathione peroxidase, galactose oxidase, and glutathione reductase form a part of the main ROS scavenging strategy. The increased transcript levels of these enzyme genes can result in enhanced tolerance to oxidative stress. In potato, overexpression of an APX gene resulted in enhanced tolerance to methyl viologen-induced oxidative stress and high temperature in transgenic potato plants [54]. Overexpression of a tobacco glutathione S-transferase with glutathione peroxidase activity in transgenic tobacco enhanced seedling growth under a variety of stress conditions [55]. In *Solanum tuberosum*, substantial accumulation of a plant thioredoxin named CDSP 32 mRNA and protein were revealed

upon oxidative treatments [56]. In *S. matsudana*, eleven genes involved in response to oxidative stress had higher transcription levels than SMT. These genes encode the abovementioned enzymes including one APX, three glutathione transferases, four thioredoxin, one galactose oxidase, and two NADP reductase. The higher transcription levels of these oxidative stress response genes in SM compared to SMT suggest that SM has stronger stress tolerance.

Supporting Information

Figure S1 Classification of raw reads of SM and SMT transcriptome sequences. (TIF)

Figure S2 The length distribution of the coding sequence (CDS) and predicted proteins. (A) Aligned CDS by BLASTX. (B) predicted proteins by BLASTX. (C) Aligned CDS by ESTScan. (D) predicted proteins by ESTScan. (TIF)

Figure S3 Saturation test and RPKM distribution of transcriptome sequences. (TIF)

Figure S4 Summary of sequence classifications of sequenced sRNA. (TIF)

Table S1 Mature sequences and expression of novel miRNA. (XLSX)

Table S2 Primers, sequence length and sequence identity of unigenes validated by real-time PCR. (XLSX)

References

- Karp A, Hanley SJ, Trybush SO, Macalpine W, Pei M, et al. (2011) Genetic improvement of willow for bioenergy and biofuels. *Journal of integrative plant biology* 53: 151–165.
- Punshon T, Dickinson NM (1997) Acclimation of *Salix* to metal stress. *New Phytologist* 137: 303–314.
- Tsarouhas V, Gullberg U, Lagercrantz U (2004) Mapping of quantitative trait loci (QTLs) affecting autumn freezing resistance and phenology in *Salix*. *TAG Theoretical and applied genetics Theoretische und angewandte Genetik* 108: 1335–1342.
- Hjältén J (1998) An experimental test of hybrid resistance to insects and pathogens using *Salix caprea*, *S. repens* and their F1 hybrids. *Oecologia* 117: 127–132.
- Greger M, Landberg T (1999) Use of Willow in Phytoextraction. *International journal of phytoremediation* 1: 115–123.
- Yang J, Yi J, Yang C, Li C (2013) *Agrobacterium tumefaciens*-mediated genetic transformation of *Salix matsudana* Koidz. using mature seeds. *Tree physiology* 33: 628–639.
- Wong MM, Cannon CH, Wickneswari R (2011) Identification of lignin genes and regulatory sequences involved in secondary cell wall formation in *Acacia auriculiformis* and *Acacia mangium* via *de novo* transcriptome sequencing. *BMC genomics* 12: 342.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29: 644–652.
- Li Q, Jin X, Zhu YX (2012) Identification and analyses of miRNA genes in allotetraploid *Gossypium hirsutum* fiber cells based on the sequenced diploid *G. raimondii* genome. *Journal of genetics and genomics* 39: 351–360.
- Brunskill EW, Potter SS (2012) RNA-Seq defines novel genes, RNA processing patterns and enhancer maps for the early stages of nephrogenesis: Hox supergenes. *Developmental biology* 368: 4–17.
- He CY, Cui K, Zhang JG, Duan AG, Zeng YF (2013) Next-generation sequencing-based mRNA and microRNA expression profiling analysis revealed pathways involved in the rapid growth of developing culms in *Moso bamboo*. *BMC plant biology* 13: 119.
- Almeida R, Gonçalves S, Romano A (2005) In vitro micropropagation of endangered *Rhododendron ponticum* L. subsp. *bacticum* (Boissier & Reuter) Handel-Mazzetti. *Biodiversity & Conservation* 14: 1059–1069.
- Rao G, Wang Y, Zhang D, Liu D, Li F, et al. (2012) Isolation and characterisation of an HsSHP gene from *Hosta plantaginea*. *Molecular biology reports* 39: 6887–6894.
- Suzuki S, Li L, Sun YH, Chiang VL (2006) The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in *Populus trichocarpa*. *Plant physiology* 142: 1233–1245.
- Lv X, Song X, Rao G, Pan X, Guan L, et al. (2009) Construction vascular-specific expression bi-directional promoters in plants. *Journal of biotechnology* 141: 104–108.
- Chang CH, Wang HI, Lu HC, Chen CE, Chen HH, et al. (2012) An efficient RNA interference screening strategy for gene functional analysis. *BMC genomics* 13: 491.
- Muoki RC, Paul A, Kumari A, Singh K, Kumar S (2012) An improved protocol for the isolation of RNA from roots of tea (*Camellia sinensis* (L.) O. Kuntze). *Molecular biotechnology* 52: 82–88.
- Zhang C, Galbraith DW (2012) RNA interference-mediated gene knockdown within specific cell types. *Plant molecular biology* 80: 169–176.
- Ivanova D, Milev I, Vachev T, Baev V, Yahubyan G, et al. (2014) Small RNA analysis of Potato spindle tuber viroid infected *Phelipanche ramosa*. *Plant physiology and biochemistry: PPB/Societe francaise de physiologie vegetale* 74: 276–282.
- Markakis MN, Boron AK, Van Loock B, Saini K, Cirera S, et al. (2013) Characterization of a small auxin-up RNA (SAUR)-like gene involved in *Arabidopsis thaliana* development. *PLoS one* 8: e82596.
- Carnavale Bottino M, Rosario S, Gratiol C, Thiebaut F, Rojas CA, et al. (2013) High-throughput sequencing of small RNA transcriptome reveals salt stress regulated microRNAs in sugarcane. *PLoS one* 8: e59423.
- Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, et al. (2008) Criteria for annotation of plant MicroRNAs. *The Plant cell* 20: 3186–3190.
- Morgan T, Craven C, Ward K (1998) Human spiral artery renin-angiotensin system. *Hypertension* 32: 683–687.
- Zhang DX, Lu HL, Liao X, St Leger RJ, Nuss DL (2013) Simple and efficient recycling of fungal selectable marker genes with the Cre-loxP recombination system via anastomosis. *Fungal genetics and biology: FG & B* 61: 1–8.
- Ye ZY, Lu HL, Geng Y, Gu YZ, Xie ZY, et al. (2013) Structural, electrical, and optical properties of Ti-doped ZnO films fabricated by atomic layer deposition. *Nanoscale research letters* 8: 108.
- Zheng S, Sun QQ, Yang W, Zhou P, Lu HL, et al. (2013) Modulation in current density of metal/n-SiC contact by inserting Al₂O₃ interfacial layer. *Nanoscale research letters* 8: 116.
- Ding SJ, Chen HB, Cui XM, Chen S, Sun QQ, et al. (2013) Atomic layer deposition of high-density Pt nanodots on Al₂O₃ film using (MeCp)Pt(Me)3 and O₂ precursors for nonvolatile memory applications. *Nanoscale research letters* 8: 80.
- Liu Z, Zhang D, Liu D, Li F, Lu H (2013) Exon skipping of AGAMOUS homolog PrSeAG in developing double flowers of *Prunus lannesiana* (Rosaceae). *Plant cell reports* 32: 227–237.
- Shi W, Hao L, Li J, Liu D, Guo X, et al. (2014) The *Gossypium hirsutum* WRKY gene GhWRKY39-1 promotes pathogen infection defense responses and mediates salt stress tolerance in transgenic *Nicotiana benthamiana*. *Plant cell reports* 33: 483–498.
- Zheng L, Liu G, Meng X, Liu Y, Ji X, et al. (2013) A WRKY gene from *Tamarix hispida*, ThWRKY4, mediates abiotic stress responses by modulating reactive oxygen species and expression of stress-responsive genes. *Plant molecular biology* 82: 303–320.
- Dang F, Wang Y, She J, Lei Y, Liu Z, et al. (2014) Overexpression of CaWRKY27, a subgroup IIc WRKY transcription factor of *Capsicum annuum*, positively regulates tobacco resistance to *Ralstonia solanacearum* infection. *Physiologia plantarum* 150: 397–411.
- Sakamoto H, Maruyama K, Sakuma Y, Meshi T, Iwabuchi M, et al. (2004) *Arabidopsis* Cys2/His2-type zinc-finger proteins function as transcription repressors under drought, cold, and high-salinity stress conditions. *Plant physiology* 136: 2734–2746.
- Davletova S, Schlauch K, Couto J, Mittler R (2005) The zinc-finger protein Zat12 plays a central role in reactive oxygen and abiotic stress signaling in *Arabidopsis*. *Plant physiology* 139: 847–856.
- Wang JY, Xia XL, Wang JP, Yin WL (2008) Stress responsive zinc-finger protein gene of *Populus euphratica* in tobacco enhances salt tolerance. *Journal of integrative plant biology* 50: 56–61.
- Figueiredo DD, Barros PM, Cordeiro AM, Serra TS, Lourenco T, et al. (2012) Seven zinc-finger transcription factors are novel regulators of the stress responsive gene *OsDREB1B*. *Journal of experimental botany* 63: 3643–3656.
- Kim SH, Hong JK, Lee SC, Sohn KH, Jung HW, et al. (2004) CAZFP1, Cys2/His2-type zinc-finger transcription factor gene functions as a pathogen-induced early-defense gene in *Capsicum annuum*. *Plant molecular biology* 55: 883–904.
- Adie BA, Perez-Perez J, Perez-Perez MM, Godoy M, Sanchez-Serrano JJ, et al. (2007) ABA is an essential signal for plant resistance to pathogens affecting JA biosynthesis and the activation of defenses in *Arabidopsis*. *The Plant cell* 19: 1665–1681.
- Robert-Seilaniantz A, Navarro L, Bari R, Jones JD (2007) Pathological hormone imbalances. *Current opinion in plant biology* 10: 372–379.
- Santner A, Estelle M (2009) Recent advances and emerging trends in plant hormone signalling. *Nature* 459: 1071–1078.
- Argueso CT, Ferreira EJ, Kieber JJ (2009) Environmental perception avenues: the interaction of cytokinin and environmental response pathways. *Plant, cell & environment* 32: 1147–1160.
- Leung J, Giraudat J (1998) Abscisic Acid Signal Transduction. *Annu Rev Plant Physiol Plant Mol Biol* 49: 199–222.
- Watanabe S, Matsumoto M, Hakomori Y, Takagi H, Shimada H, et al. (2014) The purine metabolite allantoin enhances abiotic stress tolerance through synergistic activation of abscisic acid metabolism. *Plant, cell & environment* 37: 1022–1036.
- Van Houtte H, Vandesteene L, Lopez-Galvis L, Lemmens L, Kissel E, et al. (2013) Overexpression of the trehalase gene AtTRE1 leads to increased drought stress tolerance in *Arabidopsis* and is involved in abscisic acid-induced stomatal closure. *Plant physiology* 161: 1158–1171.
- Jiang S, Zhang D, Wang L, Pan J, Liu Y, et al. (2013) A maize calcium-dependent protein kinase gene, ZmCPK4, positively regulated abscisic acid signaling and enhanced drought stress tolerance in transgenic *Arabidopsis*. *Plant physiology and biochemistry: PPB/Societe francaise de physiologie vegetale* 71: 112–120.
- Fan J, Hill L, Crooks C, Doerner P, Lamb C (2009) Abscisic acid has a key role in modulating diverse plant-pathogen interactions. *Plant physiology* 150: 1750–1761.
- Oide S, Bejai S, Staal J, Guan N, Kaliff M, et al. (2013) A novel role of PR2 in abscisic acid (ABA) mediated, pathogen-induced callose deposition in *Arabidopsis thaliana*. *The New phytologist* 200: 1187–1199.
- Mauch-Mani B, Mauch F (2005) The role of abscisic acid in plant-pathogen interactions. *Current opinion in plant biology* 8: 409–414.
- Chinnusamy V, Zhu JK (2009) Epigenetic regulation of stress responses in plants. *Current opinion in plant biology* 12: 133–139.

Author Contributions

Conceived and designed the experiments: GR JZ. Performed the experiments: GR JS YZ. Analyzed the data: GR CH AD. Contributed reagents/materials/analysis tools: GR CH. Wrote the paper: GR JZ.

49. Cutler SR, Rodriguez PL, Finkelstein RR, Abrams SR (2010) Abscisic acid: emergence of a core signaling network. *Annual review of plant biology* 61: 651–679.
50. Ding X, Cao Y, Huang L, Zhao J, Xu C, et al. (2008) Activation of the indole-3-acetic acid-amido synthetase GH3-8 suppresses expansin expression and promotes salicylate- and jasmonate-independent basal immunity in rice. *The Plant cell* 20: 228–240.
51. Zhang SW, Li CH, Cao J, Zhang YC, Zhang SQ, et al. (2009) Altered architecture and enhanced drought tolerance in rice via the down-regulation of indole-3-acetic acid by TLD1/OsGH3.13 activation. *Plant physiology* 151: 1889–1901.
52. Jaillais Y, Chory J (2010) Unraveling the paradoxes of plant hormone signaling integration. *Nature structural & molecular biology* 17: 642–645.
53. Mittler R (2002) Oxidative stress, antioxidants and stress tolerance. *Trends in plant science* 7: 405–410.
54. Kim MD, Kim YH, Kwon SY, Yun DJ, Kwak SS, et al. (2010) Enhanced tolerance to methyl viologen-induced oxidative stress and high temperature in transgenic potato plants overexpressing the *CuZnSOD*, *APX* and *NDPK2* genes. *Physiologia plantarum* 140: 153–162.
55. Roxas VP, Lodhi SA, Garrett DK, Mahan JR, Allen RD (2000) Stress tolerance in transgenic tobacco seedlings that overexpress glutathione S-transferase/ glutathione peroxidase. *Plant & cell physiology* 41: 1229–1234.
56. Broin M, Guine S, Peltier G, Rey P (2000) Involvement of CDSP 32, a drought-induced thioredoxin, in the response to oxidative stress in potato plants. *FEBS letters* 467: 245–248.