

RESEARCH

Open Access

Identification of novel microRNAs in the *Verticillium* wilt-resistant upland cotton variety KV-1 by high-throughput sequencing

Xiaohong He², Quan Sun², Huaizhong Jiang², Xiaoyan Zhu², Jianchuan Mo¹, Lu Long¹, Liuxin Xiang², Yongfang Xie², Yuzhen Shi³, Youlu Yuan^{3*} and Yingfan Cai^{1,2*}

Abstract

Plant microRNAs (miRNAs) play essential roles in the post-transcriptional regulation of gene expression during development, flowering, plant growth, metabolism, and stress responses. *Verticillium* wilt is one of the vascular disease in plants, which is caused by the *Verticillium dahliae* and leads to yellowing, wilting, lodging, damage to the vascular tissue, and death in cotton plants. Upland cotton varieties KV-1 have shown resistance to *Verticillium* wilt in multiple levels. However, the knowledge regarding the post-transcriptional regulation of the resistance is limited. Here two novel small RNA (sRNA) libraries were constructed from the seedlings of upland cotton variety KV-1, which is highly resistant to *Verticillium* wilts and inoculated with the V991 and D07038 *Verticillium dahliae* (*V. dahliae*) of different virulence strains. Thirty-seven novel miRNAs were identified after sequencing these two libraries by the Illumina Solexa system. According to sequence homology analysis, potential target genes of these miRNAs were predicted. With no more than three sequence mismatches between the novel miRNAs and the potential target mRNAs, we predicted 49 target mRNAs for 24 of the novel miRNAs.

These target mRNAs corresponded to genes were found to be involved in plant–pathogen interactions, endocytosis, the mitogen-activated protein kinase (MAPK) signaling pathway, and the biosynthesis of isoquinoline alkaloid, terpenoid backbone, primary bile acid and secondary metabolites. Our results showed that some of these miRNAs and their relative gene are involved in resistance to *Verticillium* wilts. The identification and characterization of miRNAs from upland cotton could help further studies on the miRNA regulatory mechanisms of resistance to *Verticillium* wilt.

Keywords: *Gossypium hirsutum*; KV-1; *Verticillium dahliae* Kleb; Deep sequencing; MicroRNAs

Introduction

Cotton is a fiber crop with worldwide economic importance. It is vulnerable to some diseases such as *Fusarium* wilt and *Verticillium* wilt, which caused by the phytopathogenic fungus *V. dahliae* Kleb and could have severe detrimental effects on the cotton in China, the Americas, and Mediterranean regions (Cai et al. 2009; Gao et al. 2010). The soil-borne fungal pathogen has displayed extraordinary

genetic plasticity and broad-hosts rang in diverse ecological niches (Fradin & Thomma 2006). They invade the vascular tissue by invading cortex of root via wounds or epidermal cells to cross the endodermis. The plant could show leaf vein browning, wilting, yellowing, vascular discoloration, defoliation and death after conidia are produced in large numbers and migrate via the xylem to the aerial part of the plants (Klosterman et al. 2009; BejaranoAlcazar et al. 1997). Consequently, *Verticillium* wilt is often referred to as the ‘cancer of cotton crops’. Therefore, the control against *Verticillium* wilt is necessary. However, it is also particularly difficult. The *V. dahliae* strains D07038 and V991 were used in this study and are moderately virulent and virulent, respectively. At the peak of infection, the defoliating type V991 strain has a disease incidence of 99.10% and a disease

* Correspondence: youluyuan@hotmail.com; yingfancai@gmail.com

³State Key Laboratory of Cotton Biology, Cotton Institute of the Chinese Academy of Agricultural Sciences, Key Laboratory of Cotton Genetic Improvement, Ministry of Agriculture, Anyang, Henan 455000, China

¹State Key Laboratory of Cotton Biology, Henan Key Laboratory of Plant Stress Biology, College of Life Sciences, Henan University, Kaifeng 475004, China

Full list of author information is available at the end of the article

index of 81.98 relative to uninfected control plants (Fan & Jiang 2006). The moderately toxic and defoliating wild-type D07038 strain was obtained from infected *Gossypium hirsutum*, plant and had a mean disease index of 46.1%.

Cotton varieties are assigned to five disease-resistance classes based on relative disease index scores. Varieties with relative disease index scores of 0.0 are designated immune (I); scores between 0.1 and 10.0 indicate high resistance (HR); varieties with scores between 10.1 and 20.0 are resistant (R); scores between 20.1 and 35.0 indicate tolerant (T) varieties; and varieties with scores greater than 35.0 are susceptible (S) (Zhang et al. 2012a). The cotton variety Zhongzhimian KV-1 (KV-1) was developed at Institute of Plant Protection, Chinese Academy of Agricultural Sciences (Beijing, China) and shows stable high resistance to Verticillium wilt (Zhao et al. 2009) with no significant differences between relative disease indices over three consecutive years. In this study we identify the novel miRNAs by high-throughput sequencing applying KV-1 highly resistant to Verticillium wilt.

MicroRNAs are small (~21 nt), single strand and non-coding RNAs which play key role in plant developmental and responses to stress. Precursor stem-loop secondary structures are characteristic features of miRNAs and are conserved across species (Bartel 2004; Carrington & Ambros 2003). miRNAs are transcribed from DNA. However, those miRNAs are not translated into protein. They regulate functions of other genes at the level of protein synthesis. To date, 4,933 miRNAs which are from 53 eudicotyledon species had been identified and submitted to miRBase (miRBase Release 20.0, <http://www.mirbase.org/>). A total of 78 precursors and 80 mature *Gossypium hirsutum* miRNAs have been deposited in miRBase. Previous studies have focused primarily on miRNA expression in ovules and during fiber development. Knowledge of miRNA regulatory mechanisms associated with *G. hirsutum* resistant to Verticillium wilt is very limited. The miRNAs, miR1321-1334, were predicted and identified in cotton using bioinformatics and the Solexa sequencing method (Yin et al. 2012a). Identification of novel miRNAs provides us an important genomic resource to investigate of the miRNAs regulatory mechanism of the resistance to wilt disease in upland cotton.

This study aimed to identify novel miRNAs and their potential target genes, which specifically expressed in the Verticillium wilt-resistant upland cotton variety KV-1. So two independent small RNA (sRNA) libraries were constructed from whole KV-1 seedlings inoculated with the V991 and D07038 *V. dahliae* strains; the libraries were then sequenced using the Illumina Solexa system. Then the selected upland cotton miRNAs were confirmed by quantitative real time PCR (qRT-PCR).

The potential target genes of these novel miRNAs were predicted by transcriptome sequencing in *G. hirsutum*. We found some of these targets could function as new transcripts involved in plant-pathogen interactions, which will facilitate identification of candidate genes in *G. hirsutum* for resistance to *V. dahliae*.

Materials and methods

Fungal strains and inoculums preparation

The wild-type pathogenic *V. dahliae* V991 strain are used for inoculations, which was isolated from an infected upland cotton plant. Compared with the intermediately aggressive D07038 strain obtained from the Cotton Research Institute of the Chinese Academy of Agricultural Sciences, V991 is highly toxic, defoliant and more virulent. For conidial production, V991 and D07038 were subcultured from potato dextrose agar plates onto Czapek's medium and incubated at 26°C for 7 days. Fungal cultures were filtered through sterile gauze to removal the mycelium. The concentrations of the conidial suspensions were counted under a microscope using a hemocytometer and adjusted to 1.0×10^7 conidia/ml with sterile distilled water (Zhang et al. 2012b).

Plant culture and treatment

To investigate the mechanism of disease resistance in the cotton variety Zhongzhimian KV-1, we infected seedlings with the defoliating *V. dahliae* strains D07038 (moderate virulence) and V991 (high virulence). Seeds were treated with 98% H₂SO₄ to remove the surface fuzz, and then soaked in 70% ethanol for 5 min and in 10% H₂O₂ for 1 h to sterilize the surface, followed by three rinses with sterile water. Surface-sterilised seeds were germinated at 26°C on plates containing sterilised water and filter paper. The seedlings were grown under supplemental light outside the laboratory for 2 days and then transferred to an autoclaved mixture of vermiculite and peat (1:1 v/v). Cotton seedlings were inoculated using a root dip method. The *V. dahliae* strains D07038 and V991 were cultured for 7 days and used separately to infect roots of 21-day-old seedlings, at which point the seedlings had grown one to two euphylla. Roots were infected with *V. dahliae* inoculums of 10^7 CFU/mL (10 mL) fungal suspensions and incubated for ~1–2 days at 26°C and ~80% humidity. The seedlings were treated for 24 and 48 h with D07038 as the control group or with V991 as the experimental group. Total RNAs from seedlings were harvested 24 and 48 h after inoculation and pooled in a 1:1 ratio. The whole seedling was chosen as the experimental material because the pathogenic fungus directly infects cotton roots in soil and enters the vasculature through the cortical cells, resulting in disease in the stems and leaves (Cai et al. 2009).

Construction and sequencing of sRNA libraries

Two sRNA libraries were constructed using the procedure of Kwak et al. (Kwak et al. 2009). Total RNA was isolated from *G. hirsutum* seedlings using a Plant RNA Kit (Watson Biotech, Shanghai, China) according to the manufacturer's instructions. All the RNA samples were quantified and equalised to guarantee equal amounts of RNA from each treatment. The quality of RNA samples was evaluated using an Agilent 2100 Bioanalyzer (Agilent, Waldbronn, Germany). All the RNA samples were separated by 15% urea denaturing polyacrylamide gel electrophoresis and the 18–30 nucleotide sRNA bands were excised and extracted with 0.4 M NaCl over night at 4°C. The isolated sRNAs were converted into DNA by reverse transcription PCR (RT-PCR) after being ligated sequentially to 5' and 3' adapters. The DNAs were confirmed by sequencing with Solexa sequencing technology (BGI, Shenzhen, China).

Bioinformatics analysis

Raw sequences were processed using the SOAP 2.0 software (BGI, Shenzhen, China) as reported previously (Li et al. 2008). Without the vector sequences, the sequences longer than 17 nt were considered for further analyses. Excluding rRNA, scRNA, snoRNA, snRNA, tRNA and the sequences containing a polyA tail, the other sequences were compared with *G. hirsutum* non-coding RNA sequences in the NCBI GenBank and Rfam 10.1 databases. To identify conserved miRNAs in upland cotton, the unique small RNA sequences were selected to carry out BLASTN search in the miRBase. Conserved miRNAs require perfect match. The novel miRNAs are predicted according to the characteristic hairpin structure of microRNA precursors by the RNA-folding software Mireap (<http://sourceforge.net/projects/mireap/>), which could provide the information about the secondary structures, the minimum free energies and the Dicer cleavage sites of the unidentified sRNA tags. We used the *G. raimondii* genome sequences (ftp://ftp.ncbi.nih.gov/pub/TraceDB/gossypium_raitmondii/) as an miRNA positioning reference sequence. Only mature miRNA sequences were considered, which locate in the stem region of the stem-loop structure and range between 20–22 nt as well as a minimal folding free energy index (MFEI) greater than 0.85. After the selected sequences by the Mireap were folded into the secondary structure, a novel miRNA was selected when it was at 5' end arm of the stem of a perfect stem-loop structure, and another sequence was at 3' end arm and then the small RNA was consisted as a novel miRNA from cotton.

The targets of these novel microRNAs were predicted through BLAST analysis of transcriptome sequencing data from KV-1 plants infected with *V. dahliae* Kleb. Target gene predictions followed the methods reported by Allen et al. (Allen et al. 2005). Sequences less than 4 nt

mismatches relative to the query miRNA sequences were chosen manually. We allowed one mismatch in the region complementary to nucleotide positions 2–12 of the miRNA except the mismatch at a predicted cleavage site at positions 10 or 11. We also allowed three additional mismatches between positions 12 and 22 except more than two continuous mismatches. The G:U base pair was treated as a mismatch with a value of 0.5.

Confirmation of predicted miRNAs by qRT-PCR

To confirm miRNA expression, we analyzed seven conserved miRNAs by RNA-tailing and primer-extension qRT-PCR. Small RNA was separated from seedlings by the RNAiso for sRNA reagent (TaKaRa, Dalian, China). According to the manufacturer's protocols (New England Biolabs, Beijing, China), the treated sRNA (1.5 µg) was polyadenylated by polyA polymerase (PAP) at 37°C for 30 min of the Poly(A) Tailing Kit (New England Biolabs). The RNAs were dissolved in diethylpyrocarbonate (DEPC)-treated water and reverse-transcribed with 200 U M-MLV reverse transcriptase (RNase H⁻) and 2 µL 10 pmol poly(T) adapter after phenol-chloroform extraction and ethanol precipitation (Table 1) according to the reverse transcriptase M-MLV manufacturer's instructions (TaKaRa) (Shi & Chiang 2005). Here *G. hirsutum* 5S ribosomal RNA (rRNA, GenBank: U32085.1) worked as the internal reference gene of PCR quantization. A 3' adapter primer worked as the reverse primer for the miRNAs and the 5S rRNA. The sequence of the forward primer was selected by the entire test miRNA sequence. Reverse transcription reactions were performed at 50°C for 1 h, then inactivated at 75°C for 15 min. The qRT-PCR was carried out with SYBR Premix Ex Taq II (TaKaRa, Dalian, China) on a IQ⁵ real-time PCR detection system (Bio-Rad). Each PCR included 0.5 µL template cDNA, approximately 100 pg sRNA, 5 µL 2× SYBR

Table 1 Primers Used in this Study

Name	Sequence(5' → 3')
miR5266	CTGGGGGACTGTCTGGGGC
miR5562	TGTGGAGTCTTTGCATGAAG
miR3444a-5p	TTGGGAGCTCGATGAGATCG
miR2867-5p	TGTGCCATCCCACACATC
miR1148	CCAACGTGCAGGGGGACA
miR1423a-5p	AGGCAACTACACGTTGGGGC
miR952a	AACAGAGCATGCCGTTGGT
Ghr5S RU	GGATCCCATCAGAACTCCAC
Ghr5S RL	ACGAGGACTTCCCAGAAGGT
Poly(T) adapter	GCTGTCAACGATACGCTACGTAA CGGCATGACAGTG(T)22V
Reverse primer	GCTGTCAACGATACGCTACGTAACG

*V = A, G, C;

Green II PCR master mix, and 0.5 μ L 10 pmol the forward and reverse primers (Table 1) in 10 μ L reaction system. All reactions were performed in triplicate for each sample. All reactions used the following thermal cycling profile: an initial step at 95°C for 60 s, followed by 45 cycles of amplification (95°C for 20 s, 58°C for 20 s, 72°C for 30 s) with a termination reaction at 4°C. Relative miRNA expression levels were calculated by the method of Livak and Schmittgen (Livak & Schmittgen 2001).

Results

Analysis of sequences from libraries of small RNAs

Totally 2 small RNA, namely sRNA libraries were constructed and sequenced to identify sRNAs from *G. hirsutum* L. seedlings infected with either of two strains of *V. dahliae*. The *V. dahliae* strains D07038 and V991 were cultured separately for 7 days on infected 21-day-old KV-1 cotton seedlings. Then roots were infected with the cultured D07038 or V991 as control and experimental treatments, respectively. Two sRNA cDNA libraries were generated from pooled total RNA isolated 24 and 48 h after inoculation for each treatment. Each library was sequenced by a Solexa/Illumina analyser and generated 14,491,706 primary reads for KV-1 plants infected with D07038, and 14,543,279 primary reads for KV-1 plants infected with V991 (Table 2).

Approximately 75–90% of the sRNAs in the libraries were 20–24 nt in length, and 21 and 24 nt are the main size groups (Figure 1). This distribution of sizes was similar to the other plants species and indicates that cotton miRNAs are primarily processed by DCL1 (Wang et al. 2007). Most of the sRNAs in the libraries were 24 nt long and were 38.2% and 32.3% of the total number of sequences from the D07038 and V991 libraries, respectively. Small RNAs that were 21 nt long made up 32.3% and 25.2% of the D07038 and V991 libraries, respectively, and 22 nt sRNAs made up 9.8% and 9.5% of the D07038 and V991 libraries, respectively. This observation pointed out could reflect the complexity of the cotton genome because 24 nt

sRNAs are reported to regulate the heterochromatin modification of the genomes with many repetitive sequences (Vazquez 2006). The sRNAs identified by deep sequencing represented almost all RNA classes. By comparing with the sequences to those in databases, most of our sRNAs were assigned to the different classes. The sequences that could not be assigned were used to perform the prediction of novel miRNAs by Mireap software.

Thirty-four miRNAs were screened out from 31 families. We studied the miRNAs from known families and found they showed significant difference. The abundance of individual miRNAs in the libraries varied drastically in our dataset (Table 3). Some miRNAs were represented only a few times while others existed in hundreds of thousands of copies, which indicated that the cotton seedlings contained a large and diverse collection of sRNAs. Among them, ghr-miR394 and ghr-miR2950 miRNAs showed significant differences between treatment and control groups. These observations indicate that different members could have different expression levels from the same miRNA family. Probably expression is specific to different tissues or specific to different developmental stages (Chi et al. 2011). The differential expression of a few conserved miRNAs in the two samples is not more than two times, means that several families involved in the development regulation of upland cotton seedlings.

Identification of conserved miRNAs from upland cotton

We raked through sequence homology analysis for known microRNAs in the two sRNA libraries with a requirement of at least 18 nt in length with a maximum of three mismatches compared to miRNAs in miRBase17. The abundance of specific miRNAs varied greatly between different miRNA families and between members within families. The depth of sequencing in our study was successfully identified 284 significant genes out of 443 differences expressed miRNA families of conservative miRNAs (Additional file 1). Among the miRNAs identified by comparing resistance to V991 and D07038 in KV-1 seedlings, 119 miRNAs were differentially up-regulated and 165 miRNAs were differentially down-regulated in cotton seedlings. As shown in Additional file 1, miR156, miR2911, miR2916, and miR172 were high expression in two sRNA libraries but a few conserved miRNAs (for example, miR161, miR163, and miR416) and most of the non-conserved miRNAs were found to have either very low expression. We also found some miRNAs such as miR5266, miR2867, miR3437, miR2086, and miR5652 were expressed preferentially corresponding to resistance to Verticillium wilt in KV-1 cotton. The results indicate that the expression of miRNAs was significantly affected by infection with *V. dahliae*, which suggested that miRNAs could be participate in the regulation of gene

Table 2 Statistics of small RNA sequences from two upland cotton libraries generated from KV-1 seedlings infected with the D07038 or V991 strain

Category	KV-1_D07038		KV-1_V991	
	Sequences	Unique sequences	Sequences	Unique sequences
Clean reads	14,156,228	4,348,363	14,186,947	4,597,927
miRNA	3,689,747	34,177	2,956,044	27,783
rRNA	745,970	79,074	1,549,145	145,979
snRNA	8,674	2,586	12,067	3,710
snoRNA	2,518	953	4,612	1,429
tRNA	232,045	13,306	778,699	40,648
Unannotated	9,477,274	4,608,248	8,886,380	4,378,378

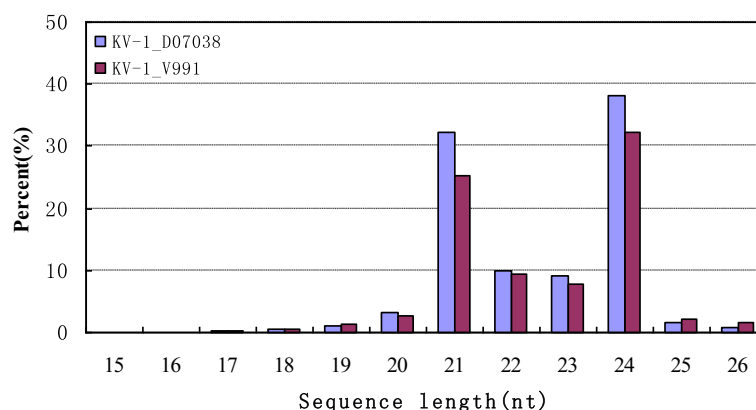


Figure 1 The size distribution of the small RNAs in upland cotton seedling KV-1_D07038 and KV-1_V991 libraries.

expression as part of the resistance to *Verticillium* wilt infection in the KV-1 variety.

Identification of novel miRNAs in upland cotton

Many researchers have identified novel miRNAs in numerous plant species by high-throughput sequencing means based on small RNA. They have found sRNA to a genomic location based on genomic DNA sequences of the species to predict secondary structure characteristic of an miRNA precursor. As described by Reinhart et al. (Reinhart et al. 2002), the lowest-energy structures sRNA is identified as new miRNAs. A total of 78 miRNA precursors and 80 mature miRNAs from *G. hirsutum* are listed in miRbase release 20 and they belong to 51 miRNA families respectively. We screened the sRNA sequences generated in this study against the Rfam database to remove all known non-coding RNAs, which including scRNA, rRNAs, snoRNAs, tRNAs,

piRNA, et al. For the identification of novel cotton miRNAs using the Mireap software, we used the *G. raimondii* genome sequence as an miRNA positioning reference sequence. We predicted 1141 miRNA candidate genes from the two sRNA libraries.

For 37 of the genes (Table 4), complementary miRNA* species (Meyers et al. 2008) with MFEI values higher than 0.85 were present in the libraries providing an indication of precise excision from the stem-loop precursor, which has recently been proposed as a primary criterion for the confident identification of an miRNA. The presence of microRNA* sequences is very important criterion because it indicates the release of the microRNA duplex from the predicted fold-back structure (Lu et al. 2006; Rajagopalan et al. 2006; Fahlgren et al. 2007), which were predicted using the genome sequence flanking the cloned sRNAs for all 37 of the novel microRNAs (Additional file 2) and microRNA precursors varied in length from 69 to 345 nt.

Table 3 Expression levels of upland cotton miRNA families showed by Solexa sequencing

Family	KV-1_D07038 expressed	KV-1_V991 expressed	fold-change*	Family	KV-1_D07038 expressed	KV-1_V991 expressed	Fold-change*
ghr-miR156	678465	372468	-0.87	ghr-miR398	6	2	-1.59
ghr-miR162	587	727	0.31	ghr-miR399	23	26	0.17
ghr-miR164	17299	26429	0.61	ghr-miR479	135	195	0.53
ghr-miR166	147030	125066	-0.24	ghr-miR482	1093	1101	0.01
ghr-miR167	967	2049	1.08	ghr-miR827	4191	3246	-0.37
ghr-miR172	1	2	1.00	ghr-miR2948	6453	2260	-1.52
ghr-miR390	35883	18183	-0.98	ghr-miR2949	2528	3049	0.27
ghr-miR393	18	30	0.73	ghr-miR2950	6806	1028	-2.73
ghr-miR394	121	10	-3.60	ghr-miR3476	20464	16352	-0.33
ghr-miR396	3666	3444	-0.09				

*Compare the known miRNA expression level in two samples. The procedures are shown as following:

(1) Calibration of the transcript expression per million (TPM) by normalizing the expression of microRNA in control and treatment samples.

Normalization formula: Normalized expression = Actual miRNA count/Total count of clean reads*1000000.

(2) Calculation of fold-change based on the normalized expression: Fold_change = log2 (KV-1_V991-std /KV-1_D07038 -std).

Table 4 Identified of novel miRNAs from upland cotton

miRNA_id	Mature miRNA sequence(5'→3')	LM(nt)	MFEIs
ghr-miR8156-5p	AAACUAUUCCUGGCUGAUUUCG	21	1.00
ghr-miR7513-3p	AAUCAGCCAGGAAUCGUUUGA	21	0.97
ghr-miR8157-5p	AAGGCAAAGGAAGAAAAAGAGUG	23	1.08
ghr-miR8158-3p	AAGGGAGAACCUAGAUUCAUU	21	1.21
ghr-miR8159-5p	AAUGGAGGAGUUGGAAAGAUU	21	1.23
ghr-miR8160-5p	ACAGCUUUAGAAAUCAUCCCU	21	1.64
ghr-miR8161-3p	GUGGAUUAAAAUUUGGUUGG	21	1.05
ghr-miR8162-5p	ACUUGCCUGCAUCUUUCAAGA	22	0.87
ghr-miR8163-3p	GGUUGCUUACUUCUCUUCUGU	21	1.10
ghr-miR8164-3p	CCUAAUAAGGAUGAUGUCUCA	21	1.03
ghr-miR8165-3p	UCCAUAUUUCACUAUCUCUUA	21	1.31
ghr-miR8166-3p	CACAGGGACAUAACCUUCUAC	21	1.04
ghr-miR8167-5p	AGCUUUAGAAUCAUCCCUU	20	1.54
ghr-miR7495a-3p	UUACUUUAGAUUCUCCUUA	21	0.87
ghr-miR8168-5p	AUUCAAACACAACACAGUGCA	21	0.97
ghr-miR7508-5p	CAAGAAAAGAAGUCGGGAGAG	21	0.94
ghr-miR8169-5p	CGGACUCUCAACAGUGGAGGUA	23	0.90
ghr-miR8170-3p	CAAUAGUUAGGCGAGAGGU	21	1.12
ghr-miR8171-3p	UCGGGGCUUUAGCGGCGUUUUA	23	1.12
ghr-miR8172-5p	GACGGGUGAUGGAAGUUUUUGG	22	0.96
ghr-miR8173-5p	GGAAUGGAGAGUUGGAAAGA	21	1.40
ghr-miR8174-3p	ACAGCUUUAGAAAUCAUCCCU	21	1.00
ghr-miR8175-3p	AUGAGCUAGAAGUUGGAACUC	21	1.21
ghr-miR8176-5p	UAAGUGAAGAAAGAGGUAGGUU	22	1.16
ghr-miR8177-5p	UCAUGGUCUUUAGCGGUGUUU	21	1.34
ghr-miR8178-3p	AUGAGCUAGAAGUUGGAACUC	21	1.24
ghr-miR8179-5p	UCGGACUGGAUUUGUUGACAA	21	0.91
ghr-miR8180-5p	UGAACUUGGUAACUAUCCAC	22	1.16
ghr-miR8181-5p	UGAGUGGAGUUAGGAGACAAA	21	0.92
ghr-miR8182-3p	UCGCUUCCUAAUUUGGACGA	21	0.86
ghr-miR8183-3p	GCAUCAGAGGACUCAGGCAGGU	22	1.07
ghr-miR8184-5p	UGGCACGGCUCAAUCAAUUA	21	0.98
ghr-miR8185-5p	UUAGAAGUUAGAUUGCAUUUUG	22	1.43
ghr-miR8186-3p	UCUAACGUGUAGGGACUAAUU	21	1.22
ghr-miR8187-5p	UUUAUGAUCUUUAGCGGCGUUU	21	1.38
ghr-miR8188-5p	UUGCAUGACACUACUUAAAAU	21	1.06
ghr-miR8189-3p	GUGUUUCGCGGUGGACGACG	21	1.26

Putative hairpin structures are shown in Additional file 3 with the novel miRNA candidates highlighted in blue. The predicted hairpins had MFEI ranging from 0.86 to 1.64 and negative folding free energies ranging from -117.6 to -33.8 kcal·mol⁻¹. These values were much lower than the reported folding free energies of tRNA (-27.5 kcal·mol⁻¹) and rRNA (-33 kcal·mol⁻¹) (Bonnet et al. 2004).

Identification of miRNA targets in *G. hirsutum*

Identification of putative miRNA targets is an effective approach for identifying the functions of miRNAs. Plant miRNAs typically have higher sequence complementarity with their target mRNAs than do animal miRNAs (Wang et al. 2004). We search for potential silenced target genes by the novel miRNAs isolated from cotton base on transcriptome sequencing data generated for KV-1 cotton infected with the D07038 and V991 *V. dahliae* strains as described in the Materials and Methods section. We were predicting 49 putative targets (Additional file 4) for 24 novel microRNA candidates. No potential targets were identified for 13 of the novel miRNAs.

To have a deeper understand functions of microRNA, we subjected the identified target genes to gene ontology (GO) analysis (<http://www.geneontology.org>), a method for identifying relevant miRNA–gene regulatory networks based on biological processes and molecular functions (Ashburner & Bergman 2005). KEGG is the major public pathway-related database (Elliott et al. 2008) and it was also used to estimate the potential functions of predicted target gene candidates of the novel miRNAs. Analysis of regulatory networks and biochemical pathways help understand the biological functions of the miRNA targets. Some miRNA families had multiple target genes, which were usually different function with homologous sequences. For these target genes, binding sites of the putative miRNA were often located in the highly conserved regions. It is notable that two miRNAs, ghr-miR8156-3p and ghr-miR8170-3p, potentially target disease resistance genes such as mitogen-activated protein kinase kinase 17, Nbs-Irr resistance protein, and plant viral-response family proteins. These targets include genes involved in plant–pathogen interactions, endocytosis, terpenoid backbone biosynthesis, isoquinoline alkaloid biosynthesis, primary bile acid biosynthesis, the MAPK signalling pathway, and biosynthesis of secondary metabolites. Our results may help advance understanding of the mechanism of resistance to Verticillium wilts in the upland cotton variety KV-1.

Quantitative real time-PCR and data analyses

Sequence counts were normalized between the datasets to account of different magnitudes raw counts. To accurate validate the deep sequencing miRNA data, we examined the correlation between normalized sequencing counts and gene expression as determined by poly (A) qRT-PCR. The expression patterns of some genes in the KV-1 plants infected by D07038 or V991 were conforming to their sequencing results. Some of the miRNA families, such as miR5266, miR5562, miR3444, and miR1148, were highly expressed, whereas miR2867, miR1423, and miR952 had relatively low level expression. A number of miRNAs

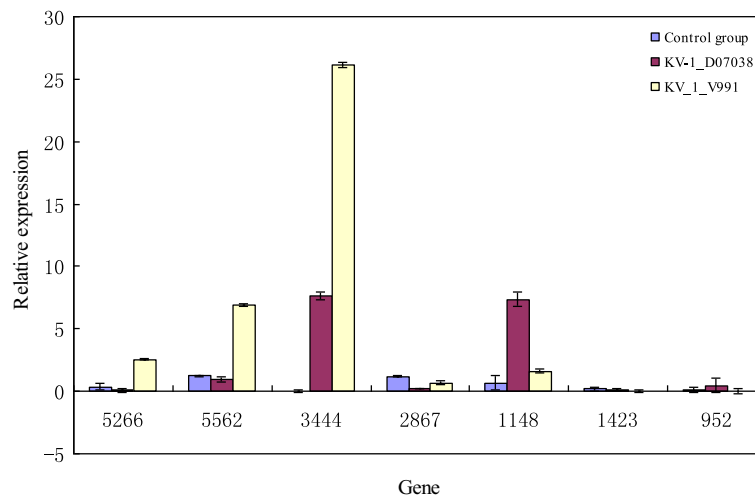


Figure 2 Relative expression of seven miRNAs in different virulence *Verticillium* wilts of Upland cotton.

were preferentially expressed in KV-1 plants infected by D07038 (e.g., miR1148) and others were preferentially expressed in KV-1 plants infected by V991 (such as miR3444) (Figure 2). There was no significant difference in miRNA levels between the Solexa and qRT-PCR data sets except for miR952 (Figure 3). We concluded that this difference was primarily due to the limitations associated with using qRT-PCR technology. The miRNA expression profiling by deep sequencing and qRT-PCR provided valuable information, however, the additional effort required to validate the data makes this approach less attractive for quantifying the effects of miRNA expression on target genes.

Discussions

The development of high-throughput sequencing methods have greatly accelerated the identification of miRNAs

in species without or with fully sequenced genomes (Peláez et al. 2012). The newly identified cotton miRNAs belonging to known miRNA families exhibited a wide range of characteristics. Previous studies have focused primarily on miRNA expression in ovules, during fiber development, and with respect to stress responses (Khan Barozai et al. 2008; Pang et al. 2009; Yin et al. 2012b). For example, miR172, miR2948, miR2949, miR2950, miR3476, miR399, miR479, and miR827 were identified during the development of ovule and fiber in upland cotton (Pang et al. 2009). In addition to miRNAs related to cotton fiber development, some miRNAs have been showed to play most vital role in response to growth stages and growth conditions. For example, ghr-miR482 regulates NBS-LRR defence genes by suppressing the miRNA-mediated gene-silencing pathway in cotton during fungal pathogen infection in cotton (Zhu et al. 2013).

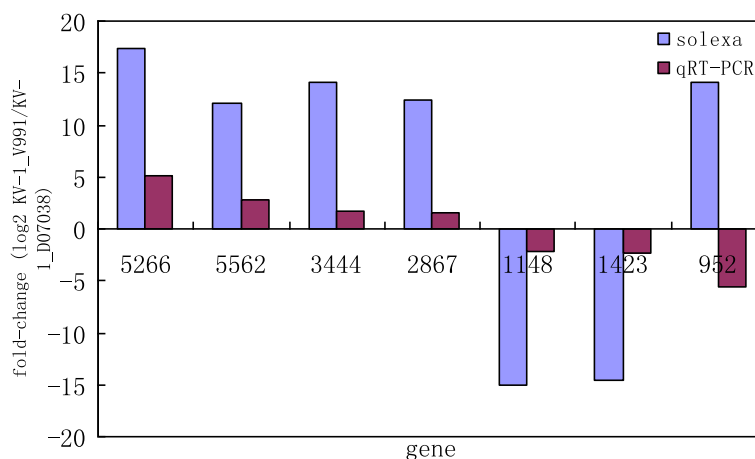


Figure 3 Expression confirmation of miRNAs in *Gossypium hirsutum* derived from high throughput sequencing. Differentially expressed miRNAs expression detected by poly(A) qRT-PCR.

Liu Yang also reported that bra-miR1885 targets protein-coding disease-resistance genes of the TIR-NBS-LRR class (Yang et al. 2013). 14 novel cotton miRNAs (miR1321 to miR1334) were identified from the *G. barbadense* L. variety 'Hai-7124', a Verticillium-tolerant cultivar, and the *G. hirsutum* L. variety 'Yimian-11', a Verticillium-sensitive cultivar, after roots were mock-infected and infected with Verticillium (Yin et al. 2012a). There have been no reports on the potential involvement of miRNAs in mechanisms of resistance to different Verticillium wilts in upland cotton varieties. Plenty of plant miRNAs have been submitted to miRBase and their biological functions have been investigated using the Illumina Genome Analyzer. Our deep sequencing data includes conserved and novel miRNAs from upland cotton based on the *G. raimondii* genome sequence. We identified 443 conserved and 37 non-conserved miRNAs from our cotton sRNA data, and also analysed the miRNAs which differentially expressed against the two *V. dahliae* strains used. Moreover, we predicted 49 potential targets of novel miRNAs to take part in plant-pathogen interactions. Our results show that miR8163, miR8165, miR8170, miR8175, and miR8178 are involved in defence responses through relevant regulatory networks, but the underlying mechanisms remain to be elucidated. Consistent with the control group, miR8163, miR8165, miR8170 and miR8175 showed significantly low level expression in our study. Similarly, because the expression level of NBS-LRR defence genes and plant viral-response family proteins were up-regulated, we can conclude that the new miRNA may involve in the regulation of the cotton in response to Verticillium wilt. The identification of these novel microRNAs and their potential target genes improves our understanding of the mechanisms regulating defense responses. We used qRT-PCR analysis to confirm the predicted known miRNAs and found that most of the expressed miRNA have Verticillium wilt resistance characteristics. Their specific expression could provide important information about how they function. Our future work will focus on their roles in upland KV-1 cotton resistance to Verticillium wilts. This study also provides a glimpse of the abundance and diversity of sRNAs in upland cotton.

Additional files

Additional file 1: Differential miRNA expression between the defoliating *V. dahliae* strains D07038 (moderate virulence) and V991 (high virulence) at the existence in the Verticillium wilt-resistant upland cotton variety KV-1.

Additional file 2: Novel miRNAs detected in the Verticillium wilt-resistant upland cotton variety KV-1.

Additional file 3: Predicted fold-back structures using precursor sequences of newly identified miRNAs in the Verticillium wilt-resistant upland cotton variety KV-1.

Additional file 4: Predicted targets for the newly identified putative miRNAs in the Verticillium wilt-resistant upland cotton variety KV-1.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

YFC conceived the study. XHH, QS, HZJ, XYZ, YLY and YZS in experiment materials preparation. XYZ, LXX, JCM, LL participated extraction. XHH, YFC analyzed data and performed qRT-PCR. YFC, XHH paper. All authors read and approved the final manuscript.

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (No. 31071461), Chongqing Municipal Commission of Education (No. KJ130520, KJ130502, KJ130510, and KJ050511), Natural Sciences Foundation of Chongqing (cstc2012jjA80037).

Author details

¹State Key Laboratory of Cotton Biology, Henan Key Laboratory of Plant Stress Biology, College of Life Sciences, Henan University, Kaifeng 475004, China. ²College of Bioinformation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China. ³State Key Laboratory of Cotton Biology, Cotton Institute of the Chinese Academy of Agricultural Sciences, Key Laboratory of Cotton Genetic Improvement, Ministry of Agriculture, Anyang, Henan 455000, China.

Received: 6 September 2014 Accepted: 16 September 2014

Published: 27 September 2014

References

- Allen E, Xie Z, Gustafson AM, Carrington JC (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* 121:207–221, doi: 10.1016/j.cell.2005.04.004 [PubMed] [Cross Ref]
- Ashburner M, Bergman CM (2005) Drosophila melanogaster: a case study of a model genomic sequence and its consequences. *Genome Res* 15:1661–1667, doi: 10.1101/gr.3726705 [PubMed] [Cross Ref]
- Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116:281–297, doi: 10.1016/S0092-8674(04)00045-5 [PubMed] [Cross Ref]
- BejaranoAlcazar J, BlancoLopez MA, MeleroVara JM, JimenezDiaz RM (1997) The influence of verticillium wilt epidemics on cotton yield in southern Spain. *Plant Pathol* 46:168–178, doi: 10.1046/j.1365-3059.1997.d01-221.x [Cross Ref]
- Bonnet E, Wuys J, Rouze P, Van de Peer Y (2004) Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics* 20:2911–2917, doi: 10.1093/bioinformatics/bth374 [PubMed] [Cross Ref]
- Cai YF, He XH, Mo JC, Sun Q, Yang JP, Liu JG (2009) Molecular research and genetic engineering of resistance to Verticillium wilt in cotton: A review. *Afr J Biotechnol* 8:7363–7372, doi: 10.5897/AJB2009.000-9571 [Cross Ref]
- Carrington JC, Ambros V (2003) Role of microRNAs in plant and animal development. *Science* 301:336–338, doi: 10.1126/science.1085242 [PubMed] [Cross Ref]
- Chi XY, Yang QL, Chen XP, Wang JY, Pan LJ, Chen MN, Yang Z, He YA, Liang XQ, Yu SL (2011) Identification and characterization of microRNAs from peanut (*Arachis hypogaea* L.) by high-throughput sequencing. *PLoS One* 6:e27530, doi: 10.1371/journal.pone.0027530 [PMC free article] [PubMed] [Cross Ref]
- Elliott B, Kirac M, Cakmak A, Yavas G, Mayes S, Cheng E, Wang Y, Gupta C, Ozsoyoglu G, Meral Ozsoyoglu Z (2008) PathCase: pathways database system. *Bioinformatics* 24:2526–2533, doi: 10.1093/bioinformatics/btn459 [PubMed] [Cross Ref]
- Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangl JL, Carrington JC (2007) High-throughput sequencing of arabidopsis microRNAs: evidence for frequent birth and death of Mirna genes. *PLoS One* 2:e219, doi: 10.1371/journal.pone.0000219 [PMC free article] [PubMed] [Cross Ref]
- Fan J, Jiang JK (2006) A new strain of China's first transgenic cotton of resistance to verticillium wilts. *Sci Times*, http://www.most.gov.cn/gnwkjdt/200608/t20060831_35734.htm [Cross Ref] (In Chinese)

- Fradin EF, Thomma BP (2006) Physiology and molecular aspects of Verticillium wilt diseases caused by *V. dahliae* and *V. albo-atrum*. *Mol Plant Pathol* 7:71–86, doi: 10.1111/j.1364-3703.2006.00323.x [PubMed] [Cross Ref]
- Gao F, Zhou BJ, Li GY, Jia PS, Li H, Zhao YL, Zhao P, Xia GX, Guo HS (2010) A Glutamic Acid-Rich Protein Identified in Verticillium dahliae from an Insertional Mutagenesis Affects Microsclerotial Formation and Pathogenicity. *PLoS One* 5:e15319, doi: 10.1371/journal.pone.0015319 [PMC free article] [PubMed] [Cross Ref]
- Khan Barozai MY, Irfan M, Yousaf R, Ali I, Qaisar U, Maqbool A, Zahoor M, Rashid B, Hussain T, Riazuddin S (2008) Identification of micro-RNAs in cotton. *Plant Physiol Biochem* 46:739–751, doi: 10.1016/j.plaphy.2008.05.009 [PubMed] [Cross Ref]
- Klosterman SJ, Atallah ZK, Vallad GE, Subbarao KV (2009) Diversity, pathogenicity, and management of Verticillium species. *Annu Rev Phytopathol* 47:39–62, doi: 10.1146/annurev-phyto-080508-081748 [PubMed] [Cross Ref]
- Kwak PB, Wang QQ, Chen XS, Qiu CX, Yang ZM (2009) Enrichment of a set of microRNAs during the cotton fiber development. *BMC Genomics* 10:457, doi: 10.1186/1471-2164-10-457 [PMC free article] [PubMed] [Cross Ref]
- Li RQ, Li YR, Kristiansen K, Wang J (2008) SOAP: short oligonucleotide alignment program. *Bioinformatics* 24:713–714, doi: 10.1093/bioinformatics/btn025 [PubMed] [Cross Ref]
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2^{−(Delta Delta C(T))} Method. *Methods* 25(4):402–408, doi: 10.1006/meth.2001.1262 [PubMed] [Cross Ref]
- Lu C, Kulkarni K, Souret FF, Muthuvallappan R, Tej SS, Poethig RS, Henderson IR, Jacobsen SE, Wang WZ, Green PJ, Meyers BC (2006) Micro-RNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant. *Genome Res* 16:1276–1288, doi: 10.1101/gr.5530106 [PMC free article] [PubMed] [Cross Ref]
- Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, Cao XF, Carrington JC, Chen XM, Green PJ, Jones SG, Jacobsen SE, Mallory AC, Martienssen RA, Poethig RS, Qi YJ, Vaucheret H, Voinnet O, Watanabe YC, Weigel D, Zhu J (2008) Criteria for annotation of plant MicroRNAs. *Plant Cell* 20:3186–3190, doi: 10.1105/tpc.108.064311 [PMC free article] [PubMed] [Cross Ref]
- Pang M, Woodward AW, Agarwal V, Guan X, Ha M, Ramachandran V, Chen X, Triplett BA, Stelly DM, Chen ZJ (2009) Genome-wide analysis reveals rapid and dynamic changes in miRNA and siRNA sequence and expression during ovule and fiber development in allotetraploid cotton (*Gossypium hirsutum* L.). *Genome Biol* 10:R122, doi: 10.1186/gb-2009-10-11-r122 [PMC free article] [PubMed] [Cross Ref]
- Peláez P, Trejo MS, Iñiguez LP, Estrada-Navarrete G, Covarrubias AA, Reyes JL, Sanchez F (2012) Identification and characterization of microRNAs in phaseolus vulgaris by high-throughput sequencing. *BMC Genomics* 13:83, doi: 10.1186/1471-2164-13-83 [PMC free article] [PubMed] [Cross Ref]
- Rajagopalan R, Vaucheret H, Trejo J, Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes Dev* 20:3407–3425, doi: 10.1101/gad.1476406 [PMC free article] [PubMed] [Cross Ref]
- Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP (2002) MicroRNAs in plants. *Genes Dev* 16:1616–1626, doi: 10.1101/gad.1004402 [PMC free article] [PubMed] [Cross Ref]
- Shi R, Chiang VL (2005) Facile means for quantifying microRNA expression by real-time PCR. *Biotechniques* 39:519–25, doi: 10.2144/000112010 [PubMed] [Cross Ref]
- Vazquez F (2006) Arabidopsis endogenous small RNAs: highways and byways. *Trends Plant Sci* 11:460–468, doi: 10.1016/j.tplants.2006.07.006 [PubMed] [Cross Ref]
- Wang XJ, Reyes JL, Chua NH, Gaasterland T (2004) Prediction and identification of Arabidopsis thaliana microRNAs and their mRNA targets. *Genome Biol* 5:R65, doi: 10.1186/gb-2004-5-9-r65 [PMC free article] [PubMed] [Cross Ref]
- Wang L, Wang MB, Tu JX, Helliwell CA, Waterhouse PM, Dennis ES, Fu TD, Fan YL (2007) Cloning and characterization of microRNAs from Brassica napus. *FEBS Lett* 581(20):3848–3856, doi: 10.1016/j.febslet.2007.07.010 [PubMed] [Cross Ref]
- Yang L, Jue D, Li W, Zhang R, Chen M, Yang Q (2013) Identification of MiRNA from eggplant (*Solanum melongena* L.) by small RNA deep sequencing and their response to Verticillium dahliae infection. *PLoS One* 8(8):e72840, doi: 10.1371/journal.pone.0072840 [PMC free article] [PubMed] [Cross Ref]
- Yin ZJ, Li Y, Han XL, Shen FF (2012a) Genome-wide profiling of mirnas and other small non-coding rnas in the Verticillium dahliae-inoculated cotton roots. *PLoS One* 7:e35765, doi: 10.1371/journal.pone.0035765 [PMC free article] [PubMed] [Cross Ref]
- Yin Z, Li Y, Yu J, Liu Y, Li C, Han X, Shen F (2012b) Difference in miRNA expression profiles between two cotton cultivars with distinct salt sensitivity. *Mol Biol Rep* 39:4961–4970, doi: 10.1007/s11033-011-1292-2 [PubMed] [Cross Ref]
- Zhang WW, Zhao L, Wang SZ, Jiang TF, Qi FJ, Jian GL (2012a) Stability analysis on Verticillium Wilt resistance of ten cotton cultivars in disease nursery. *China Cotton* 39:20–22, doi: 10.3969/j.issn.1000-632X.2012.05.006 [Cross Ref] (In Chinese)
- Zhang WW, Jian GL, Jiang TF, Wang SZ, Qi FJ, Xu SC (2012b) Cotton gene expression profiles in resistant *Gossypium hirsutum* cv. Zhongzhimian KV1 responding to Verticillium dahliae strain V991 infection. *Mol Biol Rep* 39:9765–74, doi: 10.1007/s11033-012-1842-2 [PubMed] [Cross Ref]
- Zhao L, Wang SZ, Qi FJ, Zhang WW, Jian GL (2009) Isolation, characterization and phylogenetic analysis of the resistance genealogues (RGAs) in upland cotton (*Gossypium hirsutum* L.) Zhongzhimian KV-1. *Cotton Science* 21:462–467, [Cross Ref] (In Chinese)
- Zhu QH, Fan L, Liu Y, Xu H, Llewellyn D, Wilson I (2013) miR482 regulation of NBS-LRR defense genes during fungal pathogen infection in cotton. *PLoS One* 8(12):e84390, doi: 10.1371/journal.pone.0084390 [PMC free article] [PubMed] [Cross Ref]

doi:10.1186/2193-1801-3-564

Cite this article as: He et al.: Identification of novel microRNAs in the Verticillium wilt-resistant upland cotton variety KV-1 by high-throughput sequencing. *SpringerPlus* 2014 3:564.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com