RESEARCH ARTICLE

Open Access

CrossMark

Mapping and screening of the tomato *Stemphylium lycopersici* resistance gene, *Sm*, based on bulked segregant analysis in combination with genome resequencing

Huanhuan Yang[1†], Tingting Zhao[1†], Jingbin Jiang[1], Songbo Wang[2], Aoxue Wang[1], Jingfu Li[1*] and Xiangyang Xu[1*]

Abstract

Background: Tomato gray leaf spot disease caused by *Stemphylium lycopersici (S. lycopersici)* is considered one of the major diseases of cultivated tomatoes. The only *S. lycopersici* resistance gene, *Sm*, was derived from the wild tomato species *S. pimpinellifolium*. *Sm* has been identified as an effective source of gray leaf spot resistance in tomatoes and has been mapped to tomato chromosome 11. In this study, the first bulked segregant analysis (BSA) combined with genome resequencing for the mapping and screening of the *Sm* candidate gene was performed.

Results: Based on the resequencing results, we identified 50,968 Diff-markers, most of which were distributed on chromosome 11. A total of 37 genes were located in the interval of 0.26-Mb. The gene loci of resistant and susceptible lines were sequenced successfully using PCR products. The relative expression levels of candidate genes in resistant and susceptible lines were confirmed via qRT-PCR, Solyc11g011870.1.1 and Solyc11g011880.1.1 were identified through qRT-PCR. A marker, D5, which was cosegregated with the resistant locus, was identified according to the mutation of the Solyc11g011880.1.1 trait in the resistant line.

Conclusions: The *Sm* gene was mapped to the short arm of chromosome 11. The candidate genes Solyc11g011870.1. 1 and Solyc11g011880.1.1 displayed expression patterns related to the resistance response. This study will be valuable for *Sm* cloning and *Sm* gene breeding in tomato.

Keywords: *S. Lycopersici*, *Sm* gene, Tomato gray leaf spot, BSA, Genome resequencing

Background

Gray leaf spot disease is considered a common, devastating, and damaging disease of plants such as pepper [1], cotton [2], spinach [3] and eggplant [4]. In tomato, it is caused by four species of *Stemphylium*: *Stemphylium solani*, *Stemphylium floridanum* and *Stemphylium lycopersici* [5]. It is considered a major disease of cultivated tomatoes and has threatened tomato-growing areas worldwide [6]. In the early stages, tomato gray leaf spot disease symptoms appear as brownish-black specks that later expand into necrotic lesions with gray centers and dark brown borders. As the disease progresses, the affected leaves become chlorotic and the lesions develop perforated centers, ultimately causing the leaves to become dry and fall off. *S. lycopersici* has been shown to cause tomato gray leaf spot disease based on morphology and molecular identification [7]. However, it is increasingly difficult to control this disease in tomato-growing regions around the world. Therefore, breeding for gray leaf spot resistance will provide an attractive alternative to chemical control.

To date, only the resistance gene *Sm* has been identified as a dominant gene, which was mapped to chromosome 11 in tomatoes [8]. *Sm* was derived from the wild tomato species *S. pimpinellifolium,* which has been used to breed resistant tomato cultivars [9]. The *Sm* gene is considered an effective source of gray leaf spot resistance

* Correspondence: huanyaya0126@sina.com; xxy709@126.com
†Equal contributors
[1]College of Horticulture, Northeast Agricultural University, Mucai Street 59, Xiangfang District Harbin 150030, China
Full list of author information is available at the end of the article

Yang *et al. BMC Plant Biology* (2017) 17:266

Page 2 of 10

in tomatoes. To our knowledge, few studies have been conducted on the tomato gray leaf spot resistance gene.

Bulked segregant analysis (BSA) was first proposed by Michelmore et al. [10] and was considered an effective method for identifying markers linked to a target gene [11–13] based on the genotyping of only two bulked DNA samples from groups of individuals with distinct resistant and susceptible phenotypes [14]. Genome resequencing technology, which is based on high-throughput sequencing, is a recently developed high-resolution strategy for single nucleotide polymorphism (SNP), insertion-deletion (InDel) and structural variation (SV) marker discovery and genotyping [15].

Notably, this is the first time that bulked segregant analysis combined with genome resequencing technology has been used for the mapping and screening of candidate genes for the tomato *Stemphylium* resistance gene *Sm*. This study will be valuable for *Sm* cloning and resistance breeding in tomato.

Methods

Plant materials and *S. lycopersici* inoculation

Resistant female Motelle (P_1, kindly provided by the Chinese Academy of Agricultural Sciences), comprising the *Sm* gene, was crossed with the susceptible male Moneymaker (P_2, kindly provided by the Chinese Academy of Agricultural Sciences). The resulting F_1 plants were self-crossed to produce F_2 seeds, and the BC_1 plants were obtained from a backcross between Motelle and F_1. lines. These seedlings were bred in a greenhouse under favorable conditions. *S. lycopersici* was plated on potato dextrose agar (PDA) in Petri dishes. The isolated pathogen was incubated at 28 °C for 5–10 days with a 12-h photoperiod. The tomato seedlings were sprayed with a conidial suspension (1×10^4 conidia/ml). The mock-treated plants were sprayed with sterilized water. All seedlings were grown in the greenhouse at 28 °C with a relative humidity >85% [14]. After 3–5 days, symptoms of gray leaf spot disease appeared on the tomatoes.

At the 3–4 leaf stage, all P_1, P_2, F_1, F_2, and BC_1 plants were inoculated with *S. lycopersici.* The disease index was evaluated at 7 days post inoculation [16]. The plants were visually assessed for the severity of symptoms on a scale of 1–5 points: 1 point, no symptoms; 2 points, rare lesions; 3 points, few lesions; 4 points, numerous lesions; and 5 points, coalescence of lesions. Plants with a disease index of 0–1 were regarded as resistant, and those with a disease index ≥2 were regarded as susceptible.

After inoculation, the F_2 plants were used for genetic analysis and bulked segregate analysis (BSA). The parents (P_1 and P_2) and the F_2 lines were prepared for genome resequencing and the detection of molecular markers, respectively. The resistant pool (F_2R-pool, 25 resistant plants) and the susceptible pool (F_2S-pool, 25 susceptible plants) were built by screening resistant and susceptible plants from the F_2 populations [17]. The CTAB method was used for DNA extraction from young leaves, including the parents and the F_2 lines. Bulked DNA samples were also subjected to the CTAB method by mixing equal amounts of DNA at a final concentration of 200 mg [18].

Resequencing and association analysis

The genomic DNA of each individual plant was extracted and then fragmented randomly. Adapter ligation and DNA cluster preparation were then performed, followed by Hiseq2000 sequencing. SOAP2 software (http://soap.genomics.org.cn/soapaligner.html) was used for sequencing reads, which were mapped to the reference genome sequence [19]. The sequencing depth and coverage compared with the reference genome were calculated based on the alignments. SNPs and InDels in the sequenced genome were detected using SOAP snp (http://soap.genomics.org.cn/soapsnp.html) and SOAP indel, respectively. SNP annotation (including synonymous and non-synonymous SNP mutations) and InDel annotation were performed according to the tomato reference genome (estimated size of 0.9 Gb) [20].

In this study, an average of at least 20× were generated from each sample to detect all SNPs and lx for each progeny (25 for each pool) [15]. Raw sequence reads (150 bp in length) were filtered and trimmed for quality control and adaptor removal [15]. A library was constructed for each sample, and all clean Q20 (%) values for the four libraries (P_1, P_2, F_2R-pool and F_2S-pool) were greater than 96%, indicating that the data quality was very high. Primary sequencing data were cleaned by removing reads with adapters, and a low-quality read was defined based on the number of low-quality reads (<20). Low-quality reads were removed, and the remaining high-quality data were used for mapping [21]. An SNP was defined according to its presence in the two parent lines and markers with a depth no less than 5 and a base quality >20. A homozygous and variant SNP marker, which originates from two parent lines, was used for further analysis [22, 23].

The resistant phenotype is dominant, and the ratios of resistant plants should be consistent with those of the susceptible plants for the unrelated markers [24]. Therefore, only the genotype of the P_2 plants could be found in terms of related markers. The SNP index is a method of marker association analysis, which is used to judge differences in genotype frequencies between pooled samples [25]. The SNP indexes of the resistant and susceptible pools were 0.25 and 0.75, respectively, and those SNPs were used to analyze the associated

Yang *et al. BMC Plant Biology* (2017) 17:266

Page 3 of 10

regions of candidate genes [24]. The resequencing and association analysis were performed by BGI Tech (Shenzhen, China).

Parental genome sequencing and candidate gene screening

The alignment results were used to calculate the average sequencing depth and coverage [26, 27]. All SNP and InDel polymorphisms were detected using the mapped reads. In this part of the study, DNA from the parents was used for sequencing, read mapping, and analysis of SNPs and InDels [28]. The Heinz 1706 tomato [20] was used as the reference genome for read mapping. All the SNPs in the association region were used for BSA combined with genome resequencing. Genes in the association region were annotated based on the NCBI and SGN (http://solgenomics.net/) websites. SNPs that were consistent with the susceptible genotype in the F_2S-pool were used to narrow the candidate region.

Quantitative real-time PCR analysis of candidate genes

Candidate gene expression analysis was performed using qRT-PCR. P_1 and P_2 were inoculated at the 3–4 true leaf stage. Young leaves were collected at 0, 3, 5 and 8 days after inoculation. Total RNA was extracted using the TRIzol reagent method, with three biological repeats. Reverse transcription was performed using the TaKaRa M-MLV Reverse Transcriptase (RNase H-) reverse transcription kit according to the operating instructions.

The qRT-PCR mixture contained 10 μL of 2xTrans Start Top Green q PCR Super Mix (Trans Gen, China), 1 μL of each primer, 2 μL of the cDNA template (1:5 dilution) and the appropriate volume of sterile distilled water for a total volume of 20 μL. The thermal conditions were as follows: 95 °C for 10 min, followed by 40 cycles of 95 °C for 5 s, 59 °C for 15 s, and 72 °C for 30 s. To detect primer dimerization or other artifacts of amplification, a melting-curve analysis was performed immediately after the completion of qRT-PCR (95 °C for 15 s, 55 °C 15 s, then slowly increasing the temperature by 0.5 °C per cycle to 95 °C with continuous measurements of fluorescence). The data analysis was performed using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001) with EFa1 (R: 5'-CCACCAATCTTGTACACATCC-3' S: 5'-AGACCACCAAGTACTACTGCAC-3') as a reference gene for normalization.

Sequencing of candidate gene loci and DNA sequence analysis

Primer 5.0 software was used to produce the sequences of candidate genes, and the primers of candidate genes are shown in Table 4. Reference genome sequences of candidate genes originated from the SGN.

A PCR purification kit (Takara) was used to purify parental PCR products. These purified products were then cloned into a pMD18-T vector (Takara) for sequencing. All fragments that came from sequencing were uploaded to GenBank.

Marker development and linkage analysis

To develop more polymorphic markers between Motelle and Moneymaker based on sequence variations in or near the resistant gene trait, the sequences of candidate genes were obtained from the SGN, and Primer 5.0 software was used to develop the marker primers. After PCR and sequencing, eight markers, including CAPS markers and SCAR markers, were developed to screen an F2 population of 519 plants for linkage analysis.

Results

Identification of the pathogen and genetic analysis of the *Sm* gene

In this study, genomic DNA of the pathogen was extracted, and the internal transcribed spacer (ITS) regions were amplified and sequenced with the primers ITS1 and ITS4. All sequences were submitted to GenBank (accession nos. KX858848, KX858849). BLAST search results revealed that all sequences exhibited 99% identity with *S. lycopersici* [29]. Figure 1a shows the symptoms of gray leaf spot disease on tomato leaves. Figure 1b

Fig. 1 a Symptoms of gray leaf spot disease on tomato leaves; (**b**) Conidia and conidiophores of *Stemphylium lycopersici* as seen under 400X magnification

Yang *et al. BMC Plant Biology* (2017) 17:266

Page 4 of 10

shows the conidia and conidiophores of *S. lycopersici* as observed under 400X magnification.

Finally, we confirmed that gray leaf spot disease was caused by *S. lycopersici* based on morphological characteristics and molecular identification. Motelle and F_1 plants were resistant to *S. lycopersici*, while Moneymaker plants were susceptible. The segregation ratio between resistant and susceptible plants in the F_2 population was 3:1. The segregation ratio between resistant and susceptible plants in BC_1 plants was 1:1 (Table 1).

Parental genome sequencing and candidate gene screening

In this study, a total of 67 Gb of data, including 447 M reads, and 89 Gb of data, including 593 M reads, were obtained through parental genome resequencing and F_2 bulked segregant analysis, respectively (Table 2).

Based on the resequencing results, 50,968 Diff-markers in parent lines were obtained, and 46,941 of these markers were distributed on chromosome 11. A distribution diagram of polymorphic markers (green lines) on 12 chromosomes was drawn according to the results of the resequencing positioning on the genome (Fig. 2). A region was considered an association region based on three or more consecutive Diff-markers. We found that 37 genes were distributed in the association region on chromosome 11 (Table 3). SNPs that were consistent with the susceptible genotype in the F_2S-pool were used to narrow the candidate region.

To further narrow the candidate region, SNPs in the association region were screened by resequencing combined with the F_2 BSA data and were analyzed. All SNPs, including 8 with an SNP index close to 0.33, were located in the interval of 0.26-Mb region (Table 4) (Additional file 1: Table S1). Based on the results of the association analysis and gene function annotation, six genes were screened out: Solyc11g011520.1.1, Solyc11g011530.1.1, Solyc11g011690.1.1, Solyc11g011750.1.1, Solyc11g011870.1.1 and Solyc11g011880.1.1 (Fig. 3).

Quantitative real-time PCR analysis

The relative expression levels of candidate genes in Motelle and Moneymaker were confirmed using qRT-PCR. The results showed that Solyc11g011870.1.1 and Solyc11g011880.1.1 presented expression patterns related to the resistance response (Fig. 4). The primer sequences for all candidate genes are reported in Table 4. These two candidate genes were expressed at a low level before inoculation, which increased slightly after inoculation. Their expression levels then increased rapidly 5 days after inoculation and continued to increase during the following days. In particular, compared with 0 days, the expression levels of the two resistance-related genes increased at 5 and 8 days after inoculation. Nevertheless, the expression levels of *the other four genes* were incompatible with a relationship with resistance. In conclusion, the qRT-PCR results indicated that the expression levels of Solyc11g011870.1.1 and Solyc11g011880.1.1 were compatible with a relationship with resistance. The other four genes all showed expression patterns unrelated to the resistance response.

In a previous study, we performed a transcriptome analysis of the *Sm*-mediated resistance response to *S. lycopersici* in tomatoes. Our RNA-Seq results showed that Solyc11g011870.1.1 and Solyc11g011880.1.1 were specifically up-regulated (Log2Fold-Change ≥2) in the Motelle line after 5 days of inoculation compared with the Moneymaker line. Therefore, the expression profiles of Solyc11g011870.1.1 and Solyc11g011880.1.1 were verified in our RNA-Seq analysis. The clean data from all samples have been submitted to the NCBI (http://www.ncbi.nlm.nih.gov/biosample) under accession number SRP097450.

Candidate gene sequencing and sequence analysis

The Solyc11g011870.1.1 and Solyc11g011880.1.1 loci of Motelle and Moneymaker were successfully sequenced using the PCR products (GenBank: MF059094, MF059095, MF059096, MF059097). All other association region gene loci of Motelle and Moneymaker were also sequenced successfully using the PCR products. Sequence alignment showed that the DNA sequences of thirty-two gene loci contained SNPs between Motelle and Moneymaker. There were several SNPs located in coding regions at some candidate

Table 1 Genetic analysis of *Sm* disease-resistance in different generations

Generation	Total	No. of plants		Expected segregation	
		Resistant (R)	Susceptible (S)	ratio R:S	χ2
P_1 (Motelle)	80	80	0		
P_2 (Moneymaker)	80	0	80		
F_1	50	50	0		
F_2	519	379	140	1:3	0.93
BC_1P_1	185	89	96	1:1	0.20

P_1: Resistant female, Motelle; P_2: Susceptible male, Moneymaker. The F_1 plants were self-crossed to produce F_2 seeds, and the BC_1 plants were obtained from a backcross between P_1 and F_1 lines

Yang *et al. BMC Plant Biology* (2017) 17:266

Page 5 of 10

Table 2 Summary of the sequencing data

Sample	GC_rate (%)	Q20_rate (%)	Q30_rate (%)	Bases (G)	Average depth of sequencing	Genome coverage_ rate (%)	Unique_map_ read_rate (%)	Map_read_rate (%)
P2	39.29	96.68	92.66	33.68	40.87	88.61	78.34	81.81
P1	38.92	96.35	92.06	33.38	40.51	88.5	79.77	83.63
F2S	38.93	96.42	92.13	46.05	55.89	88.89	78.29	81.94
F2R	38.65	96.56	92.43	42.98	52.17	88.76	79.9	83.62

Average depth of sequencing = total bases/genome size; Genome coverage_rate = mapped genome bases / total genome size; Unique_map_read_rate: the rate of reads in only one place is aligned on the genome; Map_reads_rate: the rate of reads is aligned on the genome

gene loci, but no difference in the distribution of conserved domains was found between Motelle and Moneymaker.

Interestingly, at the Solyc11g011880.1.1 locus, a difference in the domain distribution was detected between Motelle and Moneymaker. In the Motelle genotype, there was a 56-bp insertion in the CDS region compared with the Moneymaker genotype, and this insertion changed the ORF. Based on the analysis of the conserved domain and its characteristics, this insertion in the coding region alters the conserved SCOP domain to S_TKc (Fig. 5). The conserved protein structure of Solyc11g011880.1.1 from Moneymaker included a signal peptide, a Stress-antifung domain, two transmembrane regions and an SCOP domain. The conserved protein structure of Solyc11g011880.1.1from Motelle in the CDS region included a signal peptide, a Stress-antifung domain, two transmembrane regions and an S_TKc domain.
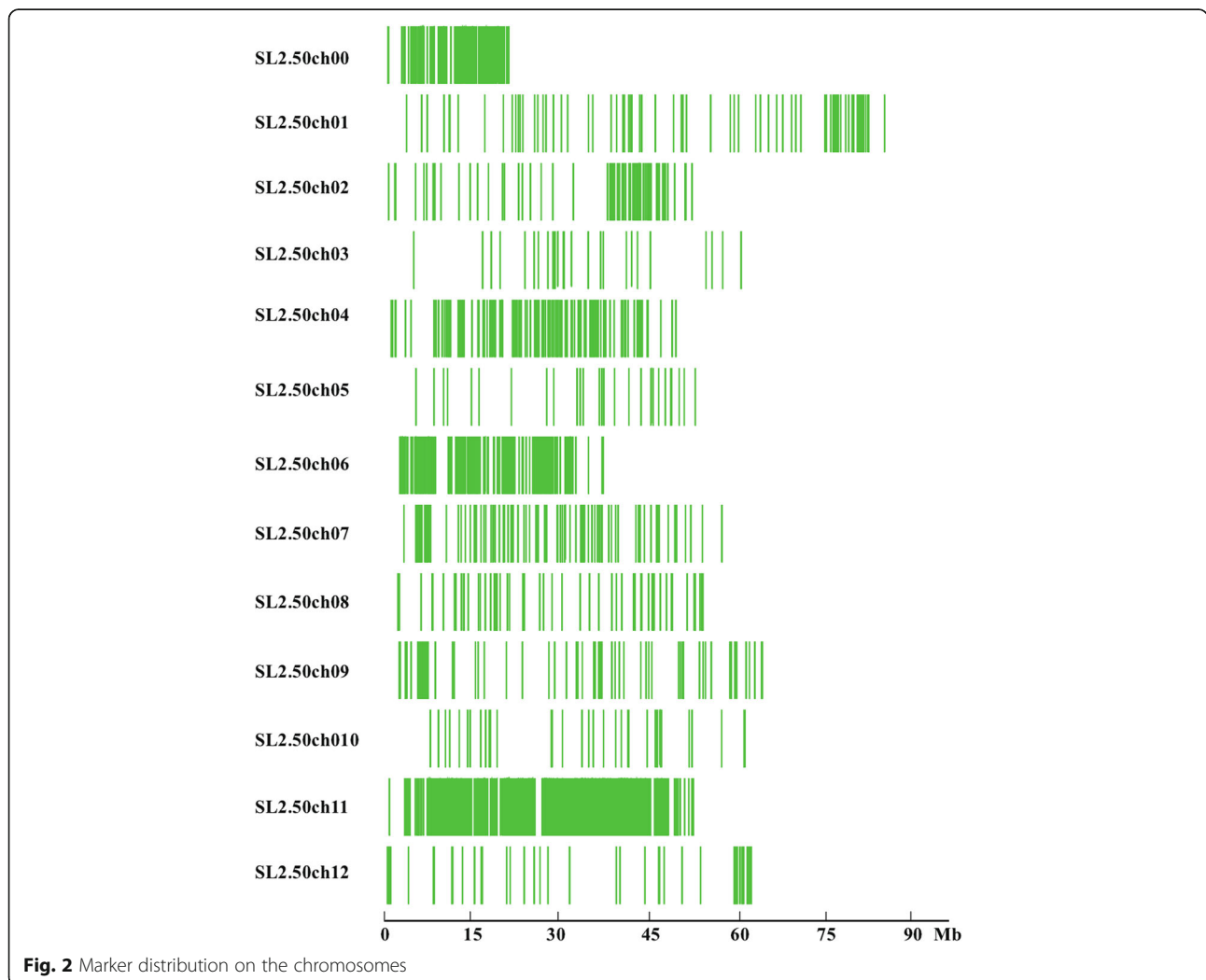

Fig. 2 Marker distribution on the chromosomes

Yang *et al. BMC Plant Biology* (2017) 17:266

Page 6 of 10

Table 3 Statistics of the candidate region

Start position	End position	Region Size (Mb)	Gene number
4,570,194	4,830,659	0.26	37

PCR validation and marker development

The SCAR marker D5 (forward primer: 5′-CCCGTGGCACTACAACTCTT-3′; reverse primer: 5′-TCTGCTTTCGCTCTGCTTGA-3′) was cosegregated with the resistance locus. The D5 was designed according to a 56-bp insertion in the resistance trait of Motelle. An 876-bp sequence was amplified from resistant plants, while an 820-bp sequence was amplified from susceptible plants, and the two sequences were amplified in the F_1 plants (Fig. 6). Among the total F_2 populations, the susceptible samples displayed the Moneymaker genotype, resistant samples displayed the Motelle or F_1 genotype, and five susceptible plants showed the Motelle genotype. The results revealed that the segregation ratio of the Solyc11g011880.1.1 genotype between resistant (876 bp) and susceptible (820 bp) individuals of the F_2 population was 3:1. The phenotype identification and molecular marker identification results for all F_2-line plants were nearly consistent.

Discussion

The *Sm* gene maps to chromosome 11

The *Sm* gene was previously assigned to the long arm of chromosome 11 between TG110 and T10 in 1991. In this study, the *Sm* gene was mapped to the short arm of chromosome 11, and one association region of the *Sm* gene was identified based on the results of F_2 bulked segregant analysis in combination with genome resequencing technology. Although some Diff-markers were found on the long arm of chromosome 11, they did not form association regions in the association analysis. To the best of our knowledge, few studies of this gene have been performed in recent years.
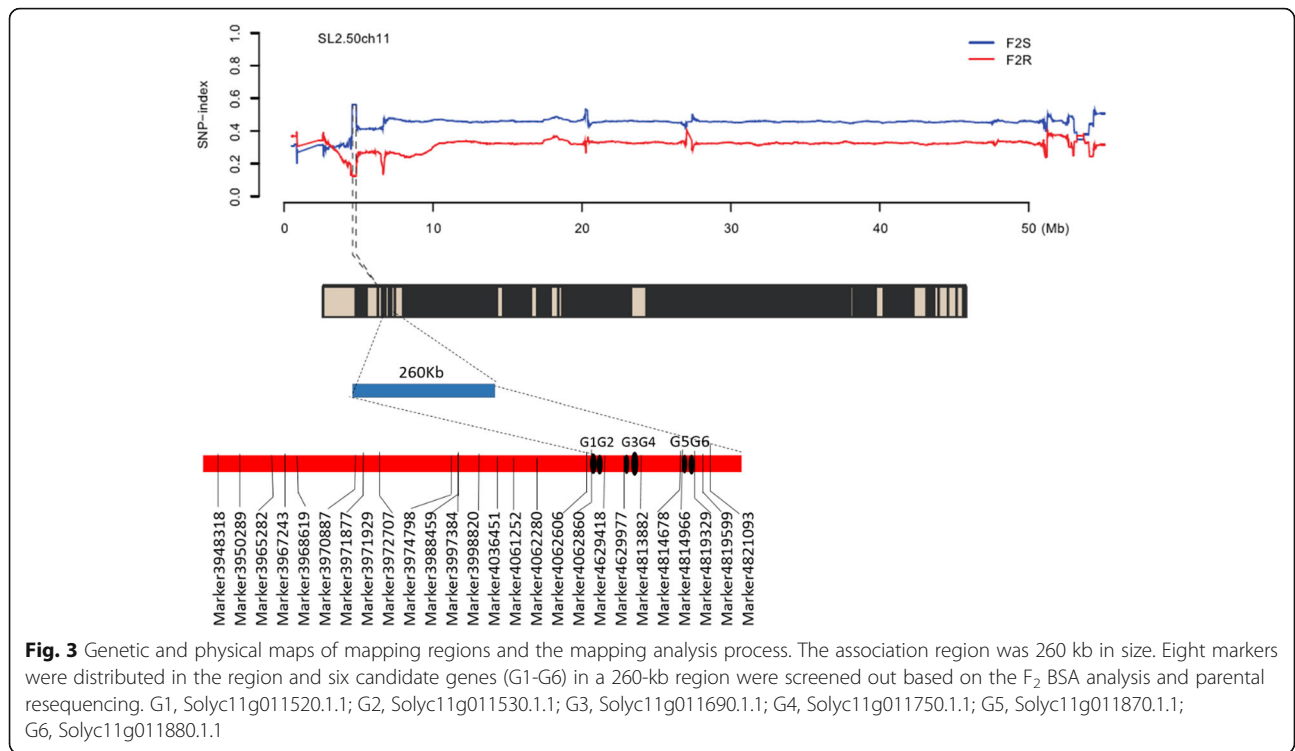
The 56-bp insertion of Solyc11g011880.1.1 in Motelle in the coding region changes the conserved domain

Pathogens harbor toxic and avirulent (AVR) genes, and host plants exhibit resistance (R) and susceptible (S) genes. A plant can develop induced resistance only when a pathogen carrying an avirulent gene infects a host plant carrying a corresponding R gene; otherwise, the plant will be infected [30]. However, it is also theoretically suggested that there is a potential for resistance in plants carrying either R or S genes. The specialization between plants and pathogens is based on the level of receptor recognition. Therefore, all plant R genes encode receptor proteins. Thus, whether induced disease resistance occurs depends on the properties of mutual plant-for-pathogen and gene-for-gene recognition. Plant disease resistance is generally defined by the gene-for-gene hypothesis [31], which states that when an avirulent (AVR) gene product of a pathogen is specifically recognized, an R gene encoding a receptor protein induces the process of plant resistance.

Stress-antifung proteins are considered effective in resistance to fungal diseases. Ginkbilobin-2 (Gnk2) is an antifungal protein, which was identified in the endosperm of Ginkgo seeds, that was found to play a vital role in the development of phytopathogenic fungi (e. g., *Fusarium oxysporum)*. Previous studies indicated that Gnk2 was very similar to the extracellular domain of cysteine-rich receptor-like kinases (CRK) in *Arabidopsis*. These findings also demonstrated that CRKs could be induced by pathogen infection, a series of responses to reactive oxygen species or salicylic acid as a component
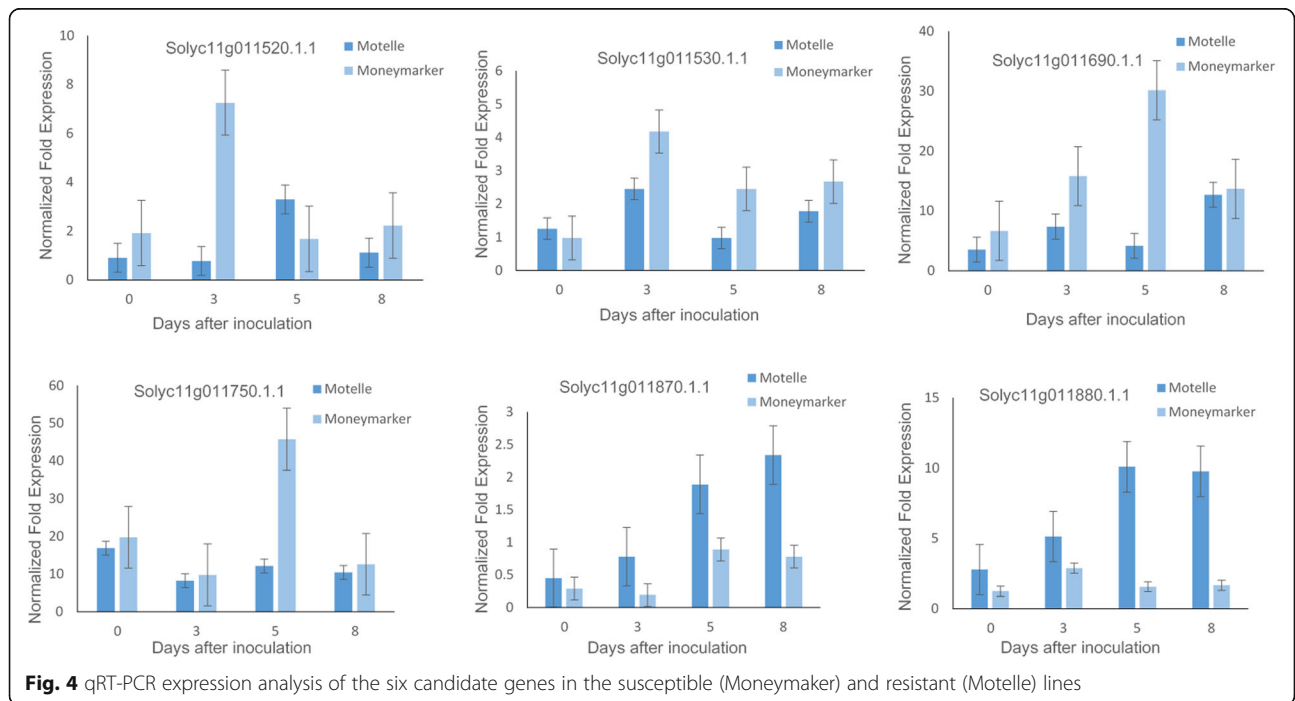
Table 4 Primers used for qRT-PCR analysis and candidate loci sequencing

Primer name	Forward primer sequence (5′-3′)	Reverse primer sequence (5′-3′)
qRT-PCR primers		
qRT-520	GAACTCCGCCTGAGATGCAA	TTGTGCTTTAGGCGATGCAC
qRT-530	GACTCCAACTCGACACTCGT	TCGATCTCGGGACAACAAGG
qRT-690	AAGAATTTGGGCTGCCGAGT	TGGATTCCACTTTCTTGGTGC
qRT-750	AGGCCATGGGGAAAGTATGC	GCGATTCGCGCACTTTATCT
qRT-870	TTGGACGCATGCTTTGAACC	CCGCATGTTTAGGCGAACAG
qRT-880	GATTCGCCACGACACAACAG	AAGAGTTGTAGTGCCACGGG
EFα1	CCACCAATCTTGTACACATCC	AGACCACCAAGTACTACTGCAC
Sequencing primers		
880–1	CCCGTGGCACTACAACTCTT	TCTGCTTTCGCTCTGCTTGA
880–2	CGAATGCTTTAAGCCCCGTG	CCGGATGCTTCCCTCTTTGT
870–1	TTGGACGCATGCTTTGAACC	CCGCATGTTTAGGCGAACAG
870–2	TGTTCGCCTAAACATGCGGA	CCCAAGTTCCATCAAGGCCA

Fig. 3 Genetic and physical maps of mapping regions and the mapping analysis process. The association region was 260 kb in size. Eight markers were distributed in the region and six candidate genes (G1-G6) in a 260-kb region were screened out based on the F₂ BSA analysis and parental resequencing. G1, Solyc11g011520.1.1; G2, Solyc11g011530.1.1; G3, Solyc11g011690.1.1; G4, Solyc11g011750.1.1; G5, Solyc11g011870.1.1; G6, Solyc11g011880.1.1

of hypersensitivity reactions, which is a typical system of programmed cell death [32]. A similar result was obtained in the present study, in which the sequence of Solyc11g011880.1.1 was found to be identical to the extracellular domain of cysteine-rich receptor-like kinases from *Nicotiana tomentosiformis*. Therefore, Solyc11g011880.1.1

is a CRK that may be involved in hypersensitivity reactions. According to the reference genome, the functions of the candidate gene Solyc11g011880.1.1 were annotated as follows: receptor-like kinase (RLK), receptor-like protein (RLP) and putative resistance protein (PRR) with an antifungal domain. Interestingly, at the Solyc11g011880.1.1

Fig. 4 qRT-PCR expression analysis of the six candidate genes in the susceptible (Moneymaker) and resistant (Motelle) lines

Yang *et al. BMC Plant Biology* (2017) 17:266

Page 8 of 10
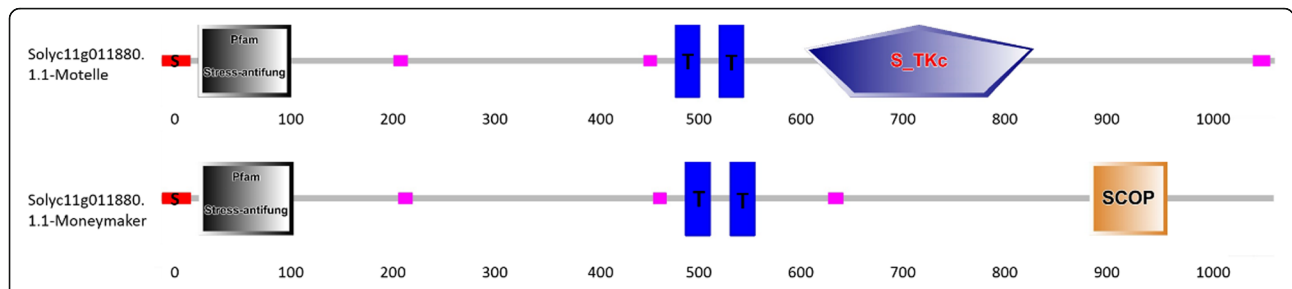
Fig. 5 The protein structure of the candidate loci in Motelle and Moneymaker. The Solyc11g011880.1.1 conserved protein structures in Motelle and Moneymaker were different; Solyc11g011880.1.1-Moneymaker included a signal peptide, a Stress-antifung, two transmembrane regions and an SCOP domain, and Solyc11g011880.1.1-Motelle included a signal peptide, a Stress-antifung, two transmembrane regions and an S_TKc domain. The "S" inside shows the signal peptide domain and "T" shows the transmembrane region

locus, a difference in the domain distribution was detected between Motelle and Moneymaker. In the Motelle genotype, there was a 56-bp insertion in the fifth exon of the CDS region compared with the Moneymaker genotype, and this insertion altered the ORF. Based on the analysis of conserved domains and characteristics, this insertion in the coding region changes the SCOP conserved domain to S_TKc. STKs catalyze the transfer of the gamma-phosphoryl group from ATP to serine/threonine residues on protein substrates. The subfamily of IRAKs includes plant RLKs, and the RLKs include *Arabidopsis thaliana* BAK1. Research has shown that BAK1 functions in pathways involved in plant resistance to pathogen infection [33].

Our results indicated that conserved domain changes in Solyc11g011880.1.1 of Motelle may result in an R gene encoding a receptor protein, causing the avirulent (AVR) gene products of a pathogen to be specifically recognized. Furthermore, the entire process of plant disease resistance is probably induced by the R gene Solyc11g011880.1.1 of Motelle, indicating that the Solyc11g011880.1.1 candidate gene of Motelle may be the target gene, *Sm*. Transcriptome analysis of the *Sm*-mediated resistance response to *S. lycopersici* in tomatoes was carried out within the scope of our research. The RNA-Seq results showed that Solyc11g011870.1.1 and Solyc11g011880.1.1 were specifically up-regulated (Log_2Fold-Change ≥2) in the Motelle line after inoculating the plants for 5 days compared with the Moneymaker line.

In conclusion, according to the RNA-Seq results and the bulked segregant analysis performed in combination with genome resequencing technology, the candidate gene Solyc11g011880.1.1 in the Motelle line may be our target gene, *Sm*. As two candidate genes, Solyc11g011870.1.1 and Solyc11g011880.1.1 were found in this target region, and the function of these two genes will be verified through virus-induced gene silencing (VIGS). Functional verification of the candidate genes in the resistant tomato line Motelle is currently ongoing in our laboratory.

The marker D5 can be used in marker-assisted selection (MAS) breeding

The marker D5 was cosegregated with the resistance locus. Interestingly, an 876-bp sequence was amplified in resistant plants, while an 820-bp sequence was amplified in susceptible plants, and the two sequences were amplified in the F_1 plants. This marker was tested in F_2 individuals. However, five F_2 individuals showed an inconsistent genotype. Considering the possible explanations for this finding, we assumed that the resistance was controlled by an incomplete dominant gene.

Although several plants were inconsistent with D5 tests, verification of the D5 marker during genotype identification is required for tomato MAS breeding. The results of the research will provide a basis for future MAS breeding and studies on the mechanism of tomato gray leaf spot disease resistance.

Conclusions

In this study, the F_2 bulked segregant analysis combined with genome resequencing were used to locate the *Sm*
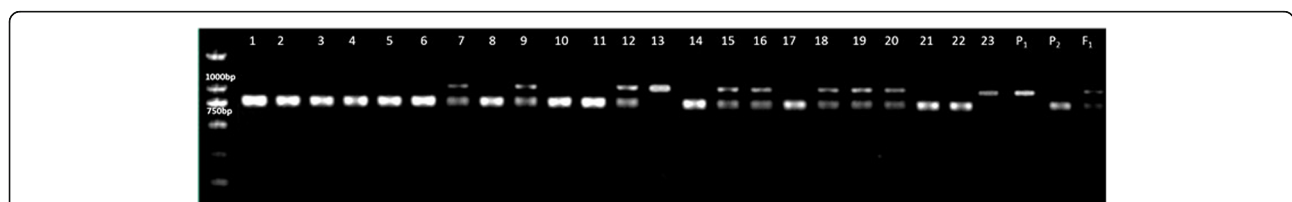

Fig. 6 Marker D5 amplification in different generations. P_1: Motelle; P_2: Moneymaker; F_1: Motelle × Moneymaker; 1–23: F_2 individuals

Yang et al. BMC Plant Biology (2017) 17:266

Page 9 of 10

gene. A total of 50,968 Diff-markers were obtained, most of which were distributed on chromosome 11. A 0.26-Mb region with 37 genes was obtained the resequencing results. The candidate genes Solyc11g011870.1.1 and Solyc11g011880.1.1 were identified through qRT-PCR, which showed related expression patterns. This study will provide a basis for *Sm* cloning and application of the *Sm* gene in breeding.

Additional file

Additional file 1: Table S1. Results from bulked segregant analysis in combination with genome resequencing. (DOCX 15 kb)

Abbreviations

Avr: AvIrulence gene; BSA: Bulked segregate analysis; CAPS: Cleaved amplified polymorphic sequences; CTAB: Cetyl trimethyl ammonium bromide; HR: Hypersensitive response; InDel: Insertion/deletion; MAS: Marker assistant selection; NCBI: National center for biotechnology information; ORF: Open reading frame finder; PRR: Putative resistance protein.; qRT-PCR: Quantitative real-time PCR; R gene: Resistance gene; RLK: Receptor-like kinase; RLP: Receptor-like protein; *S. lycopersici*: *Stemphylium lycopersici*; SCAR: Sequence characterized amplified region; SGN: SOL Genomics network; SNP: Single nucleotide polymorphism; VIGS: Virus-induced gene silencing

Availability of data and materials

The datasets supporting the results of this study are included within the article and its additional files.
The materials are available upon request by contacting the corresponding author. The clean data of the transcriptome from samples have been submitted to the NCBI (http://www.ncbi.nlm.nih.gov/sra) under accession number SRP097450. The clean data of parental resequencing from samples were also submitted to the NCBI under the accession numbers SRS2321649 and SRS2316052.
Data regarding the symptoms of gray leaf spot disease on tomato leaves and on the conidia and conidiophores of *Stemphylium lycopersici* are in shown Fig. 1.
Data regarding marker distribution on the chromosomes are shown in Fig. 2.
Data regarding genetic and physical maps of mapping regions are shown in Fig. 3.
Data regarding the qRT-PCR expression analysis are shown in Fig. 4.
Data regarding the protein structures of the candidate loci are shown in Fig. 5.
Data regarding marker D5 amplification in different generations are shown in Fig. 6.
Data regarding the genetic analysis of *Sm* disease-resistance are shown in Table 1.
Data regarding the summary of the sequencing data are shown in Table 2.
Data regarding the statistics of the candidate regions are shown in Table 3.
Data regarding the primers for the qRT-PCR analysis are shown in Table 4.
Data regarding the results from genome resequencing are shown in Additional file 1: Table S1.

Authors' contributions

XX, AW and JL conceived and designed the experiments; JJ and SW performed the RNA isolation and qRT-PCR experiments; HY and TZ performed the data analysis and wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]College of Horticulture, Northeast Agricultural University, Mucai Street 59, Xiangfang District Harbin 150030, China. [2]BGI, Shenzhen, China.

References

1. Cho HJ, Kim BS, Hwang HS. Resistance to gray leaf spot in capsicum peppers. Hortscience. 2001;36(4):752–4.
2. Fracovig PC, Mehta YR, Fonseca NS. Sources of resistance to *stemphylium solani* in cotton cultivars. Summa Phytopathol. 1999;25(3):217–22.
3. Koike ST, Henderson DM, Butlerre E. Leaf spot disease of spinach in California caused by *Stemphylium botryosum*. Plant Dis. 2001;85:126–30.
4. Yang HH, Xu XY, Zhao, TT, and Li JF. First Report of Stemphylium lycopersici Causing gray Leaf Spot on Eggplant in China. Plant Dis. 2017; https://doi.org/10.1094/PDIS-09-16-1343-PDN
5. Miranda BEC, Bolteux LS, Identification REISA. Of Solanum (section *lycopersicon*) accessions with resistance to *Stemphylium solani* and S. Lycopersici. Hortic Bras. 2010;28:178–84.
6. Simmons EG. Perfect states of *Stemphylium* IV. Harv Pap Bot. 2001;6:199–208.
7. Graham JH, Zeiders KE. Pathogenicity and morphology of some leguminicolous and related species of *Stemphylium*. Phytopathology. 1960; 50:757–60.
8. Behare J, Laterrot H, Safatti M. Restriction fragment length polymorphism mapping of *Stemphylium* resisitance gene in tomato. Mol Plant-Microbe Interact. 1991;4(5):489–92.
9. Dennett RK. The association leaf spot disease of resistance to Fusarium wilt and Stemphylium leaf spot disease in tomato, *Lycopersicon esculentum*. Proc. am. Soc. Hortic Sci. 1950;56:353–7.
10. Michelmore RW, Paran I, Kesseli RV. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci. 1991;21:9828–32.
11. Chague V, Mercier JC, Guenard M, de Courcel A, Vedel F. Identification of RAPD markers linked to a locus involved in quantitative resistance to TYLCV in tomato by bulked segregant analysis. Theor Appl Genet. 1997;4:671–7.
12. Huang Q, Li X, Chen WQ, Xiang ZP, Zhong SF, Chang ZJ. Genetic mapping of a putative Thinopyrum Intermedium-derived stripe rust resistance gene on wheat chromosome 1 B. Theor Appl Genet. 2014;4:843–53.
13. Yi B, Chen YN, Lei SL, JX T, Fu TD. Fine mapping of the recessive genic male-sterile gene (Bnmsl) in Brassica Napus L. Theor Appl Genet. 2006;4:643–50.
14. Sun XT, Zhang L, Zhang JZ. First report of tomato gray leaf spot caused by *Stemphylium lycopersici* in ZheJiang Province, China. Plant Dis. 2016;100:227.
15. Xu X, Chao J, Cheng X, et al. Mapping of a novel race specific resistance gene to Phytophthora root rot of pepper (*Capsicum annuum*) using bulked Segregant analysis combined with specific length amplified fragment sequencing strategy. PLoS One. 2016;11(3):e0151401.
16. Blansard D, Laterrat H. Les *Stemphylia* rencontres surtomate. Phytopathol Medit. 1986;25:140–4.

Yang *et al. BMC Plant Biology* (2017) 17:266

Page 10 of 10

17. Zhang H, Yi H, Wu M, et al. Mapping the flavor contributing traits on "Fengwei melon" (*Cucumis melo L.*) chromosomes using parent Resequencing and super bulked-Segregant analysis. PLoS One. 2016;11(2): e0148150.

18. Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nat Protoc. 2006;5:232 0–2325.

19. Pacurar DI, Pacurar ML, Pacurar AM, et al. A novel viable allele of Arabidopsis CULLIN1 identified in a screen for Superroot2 suppressors by next generation sequencing-assisted mapping. PLoS One. 2014;9(6):e100846.

20. Consortium TG. The tomato genome sequence provides insights into fleshy fruit evolution. Nature. 2012;485(7400):635–41.

21. Zhang JZ, Liu SR, Identifying HCG. The genome-wide genetic variation between precocious trifoliate orange and its wild type and developing new markers for genetics research. DNA Res. 2016;23(4):403–14.

22. Chen W, Yao J, Li C, et al. Genetic mapping of the nulliplex-branch gene (gb_nb1) in cotton using next-generation sequencing. Tag theoretical & Applied Genetics theoretische Und Angewandte Genetik. 2015;128(3):539.

23. Kang YJ, Yul-Kyun A, Ki-Taek K, et al. Resequencing of capsicum annuumparental lines (YCM334 and Taean) for the genetic analysis of bacterial wilt resistance [J]. BMC Plant Biol. 2016;16(1):235.

24. Zhao T, Jiang J, Liu G, et al. Mapping and candidate gene screening of tomato Cladosporium Fulvum-resistant gene Cf-19, based on high-throughput sequencing technology. BMC Plant Biol. 2016;16(1):51.

25. Abe A, Kosugi S, Yoshida K, et al. Genome sequencing reveals agronomically important loci in rice using MutMap. Nat Biotechnol. 2012;30(2):174.

26. Subbaiyan GK, Waters DL, Katiyar SK, Sadananda AR, Vaddadi S, Genome-wide DNA. Polymorphisms in elite indica rice inbreds discovered by whole-genome sequencing. Plant Biotechnol J. 2012;10:623–34. https://doi.org/10.1111/j.1467-7652.2011.00676.x. PubMed: 22222031.

27. Li S, Wang S, Deng O, Zheng A, Zhu J. Identification of genome-wide variations among three elite restorer lines for hybrid-rice. PLoS One 7: e30952. 2012; doi:10.1371fournaLpone.0030952. PubMed: 22383984.

28. Bai H, Cao Y, Quan J, Dong L, Li Z, Zhu Y, Li D. Ident fying the genome wide sequence variations and developing new molecular markers for genetics research by resequencing a landrace cultivar of foxtail millet. PLoS One. 2013;8:e73514.

29. Hong, SK. Mycobiology 40:268. https://doi.org/10.5941/MYCO.2012; 40 (4): 268.

30. Hoorn RALVD, Wulff BBH, Rivas S. Structure–function analysis of Cf-9, a receptor-like protein with Extracytoplasmic Leucine-rich repeats. Plant Cell. 2005;17(3):1000–15.

31. Dangl JL, Jones JD. Plant pathogens and integrated defence responses to infection. Nature. 2001;411(6839):26–833.

32. Miyakawa T, Sawano Y, Miyazono K, et al. Crystallization and preliminary X-ray analysis of ginkbilobin-2 from Ginkgo Biloba seeds: a novel antifungal protein with homology to the extracellular domain of plant cysteine-rich receptor-like kinases. Acta Crystallogr. 2007;63(10):737.

33. Liu J, Chen S, Chen L, et al. BIK1 cooperates with BAK1 to regulate constitutive immunity and cell death in Arabidopsis. J Integr Plant Biol. 2017;59(4):234–9.