GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Development of EST-SSR markers related to disease resistance and their application in genetic diversity and evolution analysis in *Gossypium*

B.H. Wang[1], P. Rong[1], X.X. Cai[1], W. Wang[3], X.Y. Zhu[1], C.J. Chen[1], Y.Y. Xu[1], X.J. Huang[1], Z.M. Zhuang[1] and C.B. Wang[1,2]

[1]School of Life Sciences, Nantong University, Nantong, Jiangsu Province, China
[2]Agricultural Biotechnology Research Center, Shanxi Academy of Agricultural Sciences, Taiyuan, Shanxi Province, China
[3]Agricultural Science Institute of Coastal Region of Jiangsu, Yancheng, Jiangsu Province, China

Corresponding author: C.B. Wang
E-mail: 34987881@qq.com

Genet. Mol. Res. 14 (3): 10630-10644 (2015)
Received January 31, 2015
Accepted June 24, 2015
Published September 9, 2015
DOI http://dx.doi.org/10.4238/2015.September.9.4

ABSTRACT. Cotton (*Gossypium* spp) is one of the most economically important crops that provide the world's most widely used natural fiber. Diseases such as Fusarium wilt and particularly Verticillium wilt seriously affect cotton production, and thus breeding for disease resistance is one of the most important goals of cotton breeding programs. Currently, potential exists to improve disease resistance in cultivated cotton. Increasing the understanding of the distribution, structure, and organization of genes or quantitative trait loci for disease resistance will help the breeders improve crop yield even in the event of disease. To facilitate the mapping of disease-resistance quantitative

trait loci to achieve disease-resistant molecular breeding in cotton, it is necessary to develop polymorphic molecular markers. The objective of this study was to develop simple sequence repeat markers based on cotton expressed sequence tags for disease resistance. The efficacy of these simple sequence repeat markers, their polymorphisms, and cross-species transferability were evaluated. Their value was further investigated based on genetic diversity and evolution analysis. In this study, the unique sequences used to develop markers were compared with the *G. arboretum* and *G. raimondii* genome sequences to investigate their position, homology, and collinearity between *G. arboretum* and *G. raimondii*.

Key words: Disease resistance; Evolution; *Gossypium*; Expressed sequence tag-simple sequence repeat; Genetic diversity

INTRODUCTION

With the development of the cotton textile industry, it has become increasingly important to breed cotton varieties with very high fiber quality and disease resistance. Upland cotton (*Gossypium hirsutum* L.) is the source of most of the cotton production in the world. However, the narrow genetic diversity restricts cotton breeding by limiting the range of valuable traits available for use, and also leads to genetic vulnerability to biotic and abiotic stresses. Diseases, such as Fusarium wilt and particularly Verticillium wilt, are the main threats to cotton production, so breeding for disease resistance is one of the most important goals of cotton breeding programs. Verticillium wilt is a soil-borne vascular disease, which is mainly present in the catheter of cotton plants in the form of mycelium and conidia, causing abscission of cotton leaves and bolls. Verticillium wilt can survive in soil in the form of microsclerotium, making it difficult to prevent or cure. Thus, the most economical and effective solution is to plant disease-resistant varieties. Because most commercial cultivars of Upland cotton are susceptible or only slightly resistant to Verticillium wilt, it is necessary for cotton breeders to improve Verticillium wilt resistance in cotton by introgressing resistance genes from Sea Island cotton or pyramiding genes from other sources of resistance (Zhao et al., 2014). The most efficient and effective method for introgression of resistance genes or pyramiding is to use marker-assisted selection.

Simple sequence repeats (SSRs) are polymerase chain reaction (PCR)-based markers and generally show high information content, simplicity, maneuverability, codominance, even distribution throughout the genome, reproducibility, and locus specificity (Zhang et al., 2008). The development of a large number of expressed sequence tags (ESTs) provides a good resource for preparing PCR-based SSR markers (Buyyarapu et al., 2011). EST-derived SSRs (EST-SSRs) from the transcribed regions of the DNA are generally more conserved across species than are genomic SSRs (gSSRs) from the untranscribed regions and are frequently found in other related species (Scott et al., 2000). EST-SSRs show higher cross-species transferability compared to gSSRs. Moreover, EST-SSR markers can reveal putative functional genes and can aid in the map-based cloning of important genes; this method is also easy to perform and has a low cost. EST-SSRs have been widely developed and applied in many plants, including peanut (Koilkonda et al., 2012), barley (Zhang et al., 2014), wheat (Zhong et al.,

2009; Asadi and Rashidi Monfared, 2014), alfalfa (Wang et al., 2013), rubber tree (An et al., 2013), poplar (Du et al., 2013), and date palm (Zhao et al., 2013). In recent years, an increasing number of ESTs in cotton species, including *G. hirsutum*, *G. raimondii*, *G. arboreum*, *G. barbadense*, and *G. herbaceum*, have been included in GenBank (http://www.ncbi.nlm.nih.gov/dbEST/). Additionally, extensive effort has been exerted to develop EST-SSRs (Lü et al., 2010; Buyyarapu et al., 2011; Wang et al., 2014a,b) for cotton, and these SSR markers have been widely applied in genetic diversity analysis, construction of linkage maps, quantitative trait loci (QTL) mapping, and marker-assisted breeding (Adawy et al., 2013; Wang et al., 2014 a,b; Cao et al., 2014; He et al., 2014).

In this study, non-redundant disease-resistant EST-SSR primers were efficiently developed based on cotton disease-related ESTs and applied to study different cotton accessions. The polymorphisms of these markers across *Gossypium* species were evaluated and cross-species transferability was analyzed to support QTL mapping of disease-resistant traits, genetic diversity analysis, and construction of cotton maps.

MATERIAL AND METHODS

Cotton materials

Fifty cotton accessions were selected, including 7 *G. arboreum* and *G. herbaceum* diploids in the A genome, 7 *G. raimondii*, *G. aridum*, and *G. davidsonii*, diploids in the D genome and 24 AADD tetraploids covering all 5 tetraploid cotton species. The genomic groupings and abbreviations of specific accessions are shown in Table 1. The cultivated cotton species were collected from the National Cotton Medium-Term Germplasm Bank, Cotton Research Institute, Chinese Academy of Agricultural Sciences, whereas the wild species were collected from the Wild Cotton Species Research Team of Cotton Research Institute, Chinese Academy of Agricultural Sciences/National Wild Cotton Nursery, Sanya, China/National Crop Germplasm Resource Platform, China.

Development of EST-SSR primers

EST-SSR primers were developed based on known disease-resistant ESTs. Disease-resistant genes were searched in the *Arabidopsis* Information Resource (http://www.arabidopsis.org/). *Arabidopsis* genes were used as queries for the blast search against the cotton EST bank (http://www.agcol.arizona.edu/cgi-bin/pave/Cotton/index.cgi), and the corresponding cotton transcript contigs were identified. Additionally, the contigs were BLASTed with existing sequences for primer development in the Cotton Marker Database (http://www.cottonmarker.org/), and a self-developed Perl scriptlet was then used to analyze the information and prevent redundancy. These non-redundant cotton transcript contigs were used to develop EST-SSR markers using the MISA (http://pgrc.ipk-gatersleben.de/misa/) and the Primer3 (http://frodo.wi.mit.edu/) software programs. The length of SSR primers was 18 to 25 bases with 20 bases as the optimum length. A product length of 100 to 500 base pairs was selected with annealing temperatures of 55 to 59°C; 57°C was found to be the optimum temperature. A total of 106 pairs of EST-SSR primers were developed and named "NTUXXX," where NTU indicates Nantong University and XXX indicates the serial number of the SSR primers, which ranged from 133 to 238.

Table 1. Fifty cotton accessions used for primer evaluation.

No.	Species name	Genome
1	*G. barbadense* L. 'Junhai1'	(AD)2
2	*G. barbadense* L. 'Tuhai2'	(AD)2
3	*G. barbadense* L. 'Yunnan3'	(AD)2
4	*G. barbadense* L. 'Jizha30'	(AD)2
5	*G. arboreum* L. 'Shandongxiaoguangxu'	A2
6	*G. arboreum* L. 'Yuanshizhongmian'	A2
7	*G. arboreum* L. 'Wanzizhongmian'	A2
8	*G. arboreum* L. 'Shanxianxiaozihua'	A2
9	*G. arboreum* L. 'Gebeimian'	A2
10	*G. hirsutum* L. 'Zhong142'	(AD)1
11	*G. hirsutum* L. 'Changkangmian'	(AD)1
12	*G. hirsutum* L. 'CCRI40'	(AD)1
13	*G. hirsutum* L. 'CCRI41'	(AD)1
14	*G. hirsutum* L. 'Yumian19'	(AD)1
15	*G. hirsutum* L. 'AuCS50/852	(AD)1
16	*G. hirsutum* L. 'DP99B'	(AD)1
17	*G. hirsutum* L. 'Yangmadatao'	(AD)1
18	*G. hirsutum* L. 'Hongyeai'	(AD)1
19	*G. hirsutum* L. 'CCRI27'	(AD)1
20	*G. hirsutum* L. 'Xinyan96-48'	(AD)1
21	*G. barbadense* L. 'Pima 3-79'	(AD)2
22	*G. mustelinum*	(AD)4
23	*G. aridum*	D4
24	*G. raimondii*	D5
25	*G. hirsutum* L. 'CCRI29'	(AD)1
26	*G. hirsutum* L 'P1'	(AD)1
27	*G. hirsutum* L 'RP4'	(AD)1
28	*G. hirsutum* L. 'CCRI12'	(AD)1
29	*G. hirsutum* L. 'CCRI35'	(AD)1
30	*G. hirsutum* L. 'TM-1'	(AD)1
31	*G. hirsutum* L. 'PD 94042'	(AD)1
32	*G. hirsutum* L. 'Hongyemian	(AD)1
33	*G. hirsutum* L. 'Jin6-8053'	(AD)1
34	*G. hirsutum* L. 'Jin3209'	(AD)1
35	*G. hirsutum* L. 'Sizimian29'	(AD)1
36	*G. hirsutum* L. 'Xin3-112'	(AD)1
37	*G. hirsutum* L. 'Shandong605'	(AD)1
38	*G. hirsutum* L. 'Shan3053'	(AD)1
39	*G. hirsutum* L. 'Shan1155'	(AD)1
40	*G. hirsutum* L. 'Qifengdaling'	(AD)1
41	*G. tomentosum*	(AD)3
42	*G. darwinii*	(AD)5
43	*G. hirsutum* L. 'Jiangsudatao'	(AD)1
44	*G. hirsutum* L. 'IL-2'	(AD)1
45	*G. hirsutum* L. 'IL-41'	(AD)1
46	*G. davidsonii*	D3-d
47	*G. hirsutum* L. 'Zhongzi4280'	(AD)1
48	*G. hirsutum* L. 'Zhong07'	(AD)1
49	*G. herbaceum* L. 'Jinta'	A1
50	*G. herbaceum* L. 'Gaotai'	A1

DNA extraction, PCR amplification, and electrophoresis

A modified cetyl trimethylammonium bromide method was utilized to extract genomic DNA from the cotton accessions (Paterson et al., 1993).

The 10-μL PCR mixture included 1X PCR buffer, 1.5 mM $MgCl_2$, 0.25 mM dNTPs, 5 pmol forward and reverse primers, 20 ng template DNA, and 0.15 U *Taq* polymerase. The PCR profile was as follows: pre-denaturation at 95°C for 3 min, followed by 30 cycles of 94°C

for 40 s, 57°C for 45 s, and 72°C for 1 min, and 72°C extension for 7 min. The PCR products were separated on 10% polyacrylamide gels and silver staining was used to develop DNA bands.

Data collection and analysis

Gel photos were used to score DNA bands. At each band position, the following values were assigned: band present "1" and band absent "0". The NTSYSpc ver 2.0 software (http://www.exetersoftware.com/) was used to calculate the genetic similarity of the 50 cotton accessions and unweighted pair group method with arithmetic mean was used to construct dendrograms.

The richness of alleles for each marker, which can also be regarded as a measure of usefulness of each marker in distinguishing one individual from another, was measured using the polymorphic information content (PIC) calculated as follows: PIC = 1 - ΣP_{ij}^2, where P_{ij} is the frequency of the *j*th allele for the *i*th locus, summed across all alleles of the locus (Nei, 1973).

The number of amplified fragments in each species was counted regardless of the presence of polymorphisms. Amplification was quantified according to the following equation: amplification (%) = number of amplified x 100/(total combinations of EST-SSRs x cotton species tested).

DNA sequencing and sequence analysis

The amplified fragments were excised from polyacrylamide gels. DNA recovery, vector ligation, and transformation of DH5α *Escherichia coli* competent cells were conducted as described by Guo et al. (2003). At least 3 positive clones from each amplified fragment were simultaneously sequenced by Invitrogen (Carlsbad, CA, USA). The DNAMAN program (Lynnon Biosoft, Quebec, Canada) was used to align the amplified SSR alleles. The consensus sequence was used as a reference in sequence comparisons and base mutation analyses.

Genome position of EST-SSRs and sequence homology between *G. arboretum* and *G. raimondii* genomes

The sequence of the diploid A2 genome of *G. arboreum* (http://cgp.genomics.org.cn/) and D5 genome of *G. raimondii* (http://www.phytozome.net/cotton.php; http://cgp.genomics.org.cn/) has been released. To investigate the position of these EST-SSRs and the homology and collinearity between *G. arboretum* and *G. raimondii*, we compared unique sequences used to develop markers with the *G. arboretum* and *G. raimondii* genome sequences using BLASTN with an E-value of 1e-3 and mismatch numbers of no more than 3.

RESULTS

Characteristics of ESTs and EST-SSR primers

A total of 106 pairs of EST-SSR primers were developed based on disease-resistant cotton ESTs. The primer sequences, serial number of the ESTs in the cotton EST database, and serial number of the corresponding genes in the *Arabidopsis* Information Resource are listed in (**Table S1**). Regarding the repeat motif type, the trinucleotide repeats showed the highest frequency of occurrence at 58.0%, followed by tetranucleotide repeats and pentanucleotide repeats with frequencies of 15.7 and 10.0%, respectively, whereas dinucleotide repeats and hexanucleotide repeats showed the lowest frequencies of 8.3 and 8%, respectively (Figure 1).

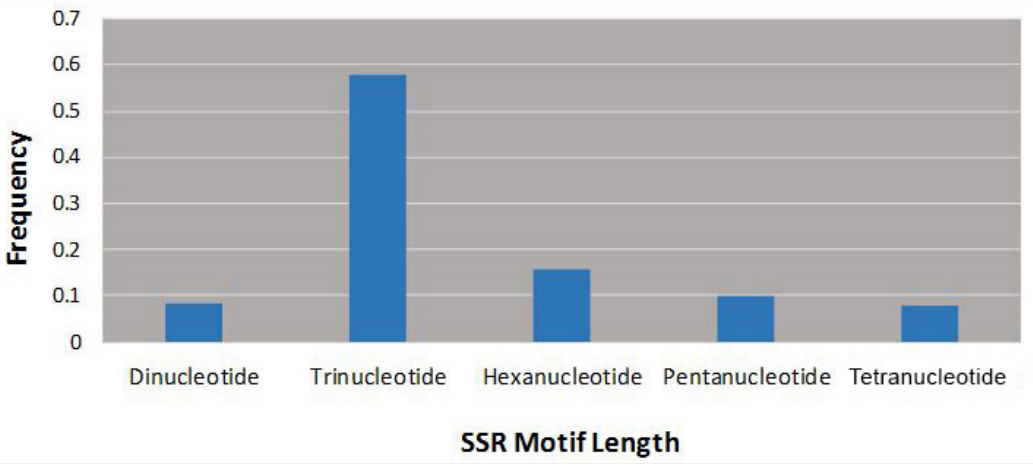

Figure 1. Frequency distribution of motif sizes among the 106 microsatellite loci.

Cross-species amplification of EST-SSR primers

The length of amplification products of the EST-SSR primers was from 100 to 500 base pairs (Figure 2). Of the 106 NTU primer pairs, clear bands were amplified in 84 pairs, with an effective amplification rate of 79.3%. The remaining 22 primer pairs produced no amplification products or specific bands. As an example, Figure 2 shows that both single-locus markers and multi-locus markers were developed in this study. The marker NTU147 with a single locus of 250 base pairs was amplified in all 50 accessions, whereas 8 loci were detected with the marker NTU150, which revealed polymorphisms between different accessions.

Forty-seven of the 84 effective markers (56.0%) were amplified and showed clear bands in all 50 cotton accessions with high universality. Thirty-one of the 84 markers amplified 47 to 49 accessions; 3 of these primer pairs were amplified in 43 to 46 accessions and another 3 primer pairs were only amplified in 34 to 40 accessions, of which 3 belonged to *G. arboreum* in the diploid A genome, 8 to *G. hirsutum*, and 1 to *G. tomentosum* in the tetraploid AD genome. None was amplified in any of the diploid D genome species, indicating the existence of genome-specific amplification. The average cross-species transferability rate was 97.4% for the 84 effective markers, whereas the 84 markers had different amplification efficiencies and cross-species transferability within different cotton species (Table 2). They showed higher cross-species transferability across the A genome (96.59%) than across the D genome (92.86%), whereas the cross-species transferability within tetraploid species was 97.86% (Table 2).

The EST-SSR markers also showed different amplification efficiencies in different cotton accessions, even within the same species. The highest amplification efficiency was observed in 23 tetraploid cotton accessions with an amplification ratio of 100%, and stable bands were amplified in all 23 accessions. For *G. klotzschianum* in the D genome, these primer pairs showed the lowest amplification efficiency of 85.71%, where only approximately 72 bands from the 84 primer pairs were amplified.

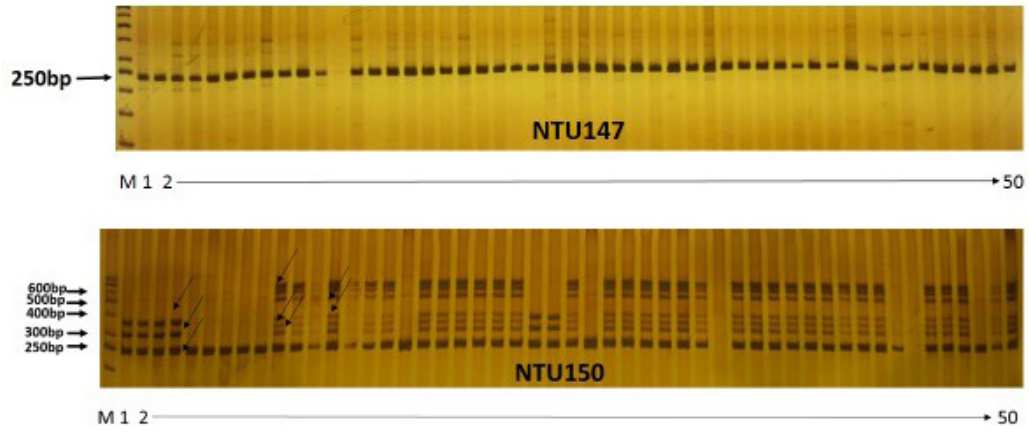

Figure 2. Electrophoretograms of 2 primer pairs designated NTU147 and NTU150 amplifying specific bands in 50 cotton lines.

Table 2. Amplification results and cross-species transferability of 106 EST-SSRs among different genomes.

Genome	No. of primers amplified wholly	No. of primers amplified partially	No. null amplified	Total bands	Total combinations	Amplification (%)
A1	78	3	3	568	588	96.59%
A2	79	5	0			
D3	80	0	4	235	252	92.86%
D4	72	0	12			
D5	82	0	2			
AD1	67	17	0	3288	3360	97.86%
AD2	80	7	0			
AD3	79	0	5			
AD4	72	0	12			
AD5	78	0	6			

Polymorphism of EST-SSR markers

Clear bands were amplified using 84 of the 106 (79.3%) EST-SSR primer pairs in the 50 cotton accessions; the 46 primer pairs that were amplified with obvious polymorphism and stable band types are listed in Table 3. PIC values were calculated based on the alleles detected by the 46 primer pairs in the 50 cotton accessions. A total of 147 polymorphic alleles were detected by the 46 primer pairs, with an average of 3.20 per primer pair, ranging from 2 to 8 (NTU150, Figure 2). PIC values ranged from 0.3902 for NTU209 to 0.9109 for NTU150 (Table 3), with an average value of 0.6458.

Genetic similarity index and cluster analysis of the 50 cotton accessions

As shown in Figure 3, the 50 accessions were divided into 2 groups at a similarity coefficient of 0.68, wherein all 3 accessions from the D genome were in one group; the other group consisted of accessions from the A genome and the tetraploid cotton accessions. With a similarity coefficient of 0.706, all accessions from the A genome were distinguished from the tetraploid cotton accessions. Within the tetraploid cotton accessions with a similarity coef-

ficient of 0.756, *G. mustelinum* was differentiated from other tetraploid cotton accessions; for 0.792, *G. tomentosum* and *G. darwinii* were differentiated from other tetraploid cotton accessions, and these 2 tetraploid species were further differentiated with a similarity of 0.872. At 0.826, *G. barbadense* accessions were differentiated from *G. hirsutum*; *G. hirsutum* Cotton Research Institute, Chinese Academy of Agricultural Sciences 27 and Jin3209 still could not be differentiated with a similarity coefficient of 0.99.

Table 3. PIC value of the EST-SSR primers.

SSR primer	No. of alleles	PIC	SSR primer	No. of alleles	PIC
NTU209	3	0.3902	NTU236	5	0.6744
NTU143	3	0.4315	NTU172	5	0.6821
NTU196	3	0.4555	NTU210	6	0.6859
NTU212	3	0.4584	NTU215	5	0.6874
NTU163	3	0.468	NTU208	5	0.6894
NTU138	4	0.4713	NTU197	5	0.6936
NTU231	4	0.4852	NTU190	6	0.6995
NTU140	3	0.4899	NTU182	6	0.7119
NTU188	4	0.4909	NTU159	6	0.7133
NTU202	4	0.4913	NTU216	5	0.7293
NTU193	4	0.5041	NTU185	5	0.7341
NTU165	4	0.5063	NTU148	6	0.7452
NTU233	4	0.5084	NTU225	6	0.7468
NTU151	3	0.532	NTU237	7	0.7809
NTU224	3	0.5414	NTU195	10	0.7831
NTU168	3	0.5534	NTU166	7	0.8021
NTU146	4	0.5623	NTU229	9	0.8049
NTU213	5	0.6122	NTU174	10	0.8051
NTU191	6	0.6444	NTU178	9	0.8369
NTU204	4	0.6456	NTU235	12	0.8449
NTU219	5	0.6464	NTU175	14	0.8604
NTU171	6	0.648	NTU214	16	0.8965
NTU226	5	0.6501	NTU150	15	0.9109

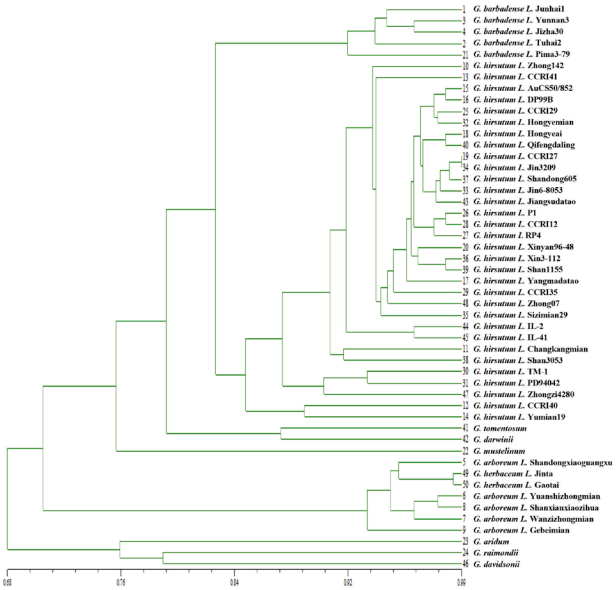

Figure 3. Dendrogram showing 50 cotton accessions based on the EST-SSRs markers used in this study.

DNA sequencing analysis of EST-SSR markers

To study how DNA polymorphisms are produced, the amplified products from 2 of the 6 markers related to *Verticillium* wilt resistance, NTU189 and NTU216, were sequenced across 7 accessions covering 6 species, including *G. arboreum* L. 'Yuanshizhongmian' from the A2 genome, *G. raimondii* from the D5 genome, *G. hirsutum* L. TM-1 and Sumian22, *G. barbadense* L. Pima 3-79, *G. tomentosum*, and *G. mustelinum*. A total of 42 clones were sequenced and the alignment results are shown in Figure 4.

For amplification products of the primer NTU189, most flanking sequences of the 7 cotton accessions were the same, whereas 2 types of transitions, including 1 instance of an A to G and 2 instances of a T to C transition were observed in the flanking sequences. Moreover, 5 deletions were observed, including deletion of a single nucleotide of A/T, and deletion of multiple nucleotides of ACA/CCACCACC. For amplification products of the primer NTU216, fewer differences were observed between the 7 cotton accessions, although most of the flanking sequences were still the same. Two transitions, 1 instance of A to G and 1 instance of T to C, were observed. Moreover, 3 instances of deletions were observed, including deletion of a single nucleotide of T and deletion of multiple nucleotides of AAG.

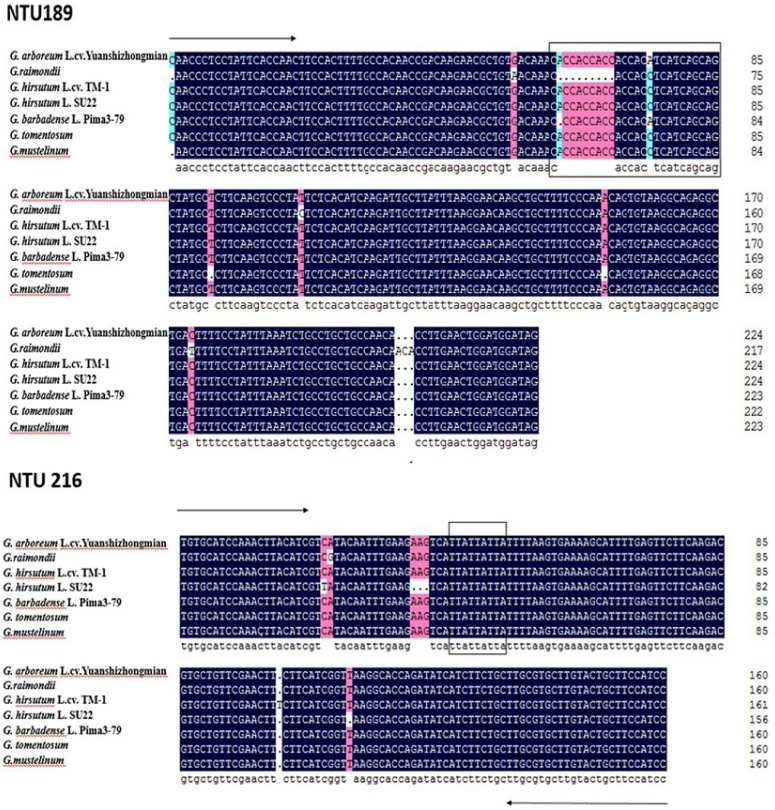

Figure 4. CLUSTAL alignment of amplified products in 7 cotton accessions by primers NTU189 and NTU216. Primer-binding sites are represented by arrows. Repetitive sequences are indicated in box. Consensus nucleotides are listed under the sequences.

Genome position of the EST-SSRs and homology between *G. arboretum* and *G. raimondii* genome

The 106 unique sequences were compared with the *G. arboretum* and *G. raimondii* genomes. The genome position of these sequences as well as the homology and collinearity between *G. arboretum* and *G. raimondii* are shown in **Table S2**.

Ninety-four of the 106 unique sequences were matched to the *G. arboretum* and *G. raimondii* genome, of which 84 were matched on both chromosomes of the *G. arboretum* and *G. raimondii* genome; 8 sequences, including NTU141, NTU142, NTU171, NTU191, NTU201, NTU203, NTU222, and NTU233, were only matched in the *G. raimondii* genome, whereas 2 sequences-NTU210 and NTU226 were only matched to the *G. arboretum* genome. Additionally, 12 of the 106 unique sequences were not matched on either genome. A total of 165 homologous sequences were found in the A2 genome, whereas 168 homologous sequences were found in the D5 genome. For the 84 EST-SSRs matched to both the A2 and D5 genomes, the order in the *G. arboretum* physical map was in accordance with the position of the hits on the *G. raimondii* physical map, such as CA_chr3 (GR_chr03), CA_chr8 (GR_chr04), CA_chr10 (GR_chr11), CA_chr11 (GR_chr07), and CA_chr13 (GR_chr13). However, dislocation and inversion of the markers' order occurred on some chromosomes.

DISCUSSION

Development, amplification, and characteristics of disease-resistant EST-SSR primer pairs

ESTs are parts of genes with conserved sequences. With the release of an increasing number of ESTs, the development of EST-SSR primers has become easier and more efficient. Compared to gSSRs, EST-SSRs show higher cross-species transferability and universality because EST-SSRs are derived from expressed sequences that are more conserved than the non-genic sequences and are easily found in other relative species (Scott et al., 2000). Moreover, because EST-SSRs originate from a coding region of the genome, they are closely related to functional genes, as well as show advantages such as good repeatability, stability, and high polymorphism. Consequently, EST-SSR markers have been widely used to construct genetic maps, for genetic diversity analysis, and for marker-assisted selection in crop disease-resistance improvement. To identify and evaluate chickpea wilt-resistant lines against *Fusarium oxysporum f sp ciceris* (Schlechtends), Ahmad et al. (2014) identified an SSR marker "TA194" showing linkage to the wilt-resistant locus, which can be useful for marker-assisted selection breeding for yield improvement of chickpea. Maruthachalam et al. (2010) employed molecular analyses including SSR markers to characterize the genetic variability and race structure of 101 isolates of *Verticillium dahliae* from a variety of hosts, suggesting that the PCR assay can be applied to differentiate the 2 races of *V. dahliae* to support the aims of breeding host resistance. This further reveals insights into the distribution of these races in tomato and lettuce cropping systems in California. Qin et al. (2014) conducted a survey to examine the presence of SSRs in viroid genomes. They found that the distribution of SSRs in viroids may influence secondary structure, and that SSRs may play a role in generating genetic diversity. Bansal et al. (2014) identified molecular markers linked with adult plant stem rust resistance gene Sr56 in winter wheat. Saturation mapping of the Sr56 region resulted in the flanking of Sr56 by sun209 (SSR) and sun320 (STS)

at 2.6 and 1.2 cM on the proximal and distal ends, respectively.

In cotton research, Ulloa et al. (2013) comprehensively analyzed gene action in cotton governing *Fusarium* wilt (*Fusarium oxysporum f sp vasinfectum*, FOV) race 4 resistance by combining conventional inheritance and QTL mapping with molecular markers. They found that a single resistance gene (Fov4) had a major dominant gene action and conferred resistance to the FOV race 4 in Pima S6, which appears to be located on chromosome 14. Additional genetic and QTL analyses also identified a set of 11 SSR markers, indicating the involvement of more than one gene and gene interactions in the inheritance of FOV race 4 resistance. Zhao et al. (2014) performed association mapping of *Verticillium* wilt resistance in cotton. Forty-two marker loci associated with *Verticillium* wilt resistance were identified through association mapping, among which 10 marker loci were found to be consistent with previously identified QTLs and 32 were new, unreported marker loci.

Although numerous EST-SSR markers have been developed from *G. hirsutum*, *G. arboreum*, *G. raimondii*, *G. barbadense*, and *G. herbaceum*, few EST-SSR primers have been developed specifically using disease-resistance ESTs in cotton. In this study, 106 pairs of EST-SSR primers were developed based on cotton disease-resistance ESTs obtained from BLAST using *Arabidopsis* disease-resistant genes for cotton disease-resistance molecular breeding efforts. Previous studies indicated that the effective amplification rate of EST-SSRs should be 60 to 90% (Saha et al., 2004), and that some primers may not be amplified once the primer covers an mRNA editing site or the amplification product contains large introns. In our previous study, the effective amplification rate was 80.3% (Wang et al., 2014a). In the current study, clear bands were amplified in 84 of the 106 NTU primer pairs, and the effective amplification rate was 79.3%, which is a high value compared to values reported in previous studies. There are 3 reasons why the remaining 22 primer pairs may have produced no amplification product or specific bands. First, we used the same annealing temperature (57°C) for PCR using all primers for simplicity and convenience, which may be suitable for some primers but not for others; optimization of PCR conditions may improve amplification efficiency. Second, the primers were designed based on EST sequences, whereas introns or other structures in the genome may have prevented amplification; when the introns were too long or if the primers were designed from the adjoining regions of an intron, no amplification would occur (Zhong et al., 2009). Third, the SSR loci and the markers were within transposons or repetitive DNA, thus resulting in multiple products deviating from the expected bands.

PIC values can be used to evaluate the allele variation of each primer, which is determined by the allele number and distribution frequency (Botstein et al., 1980). Accordingly, PIC indicates the degree of the markers' polymorphism. A higher PIC value gives a greater increase in the ability to reveal allelic variation. Generally, EST-SSRs have a lower likelihood of revealing allelic variation in crops compared to gSSRs (Blair et al., 2006). As proposed by Botstein et al. (1980), the criteria of PIC measuring genetic diversity were as follows: PIC < 0.25, low polymorphism; 0.25 < PIC < 0.5, medium polymorphism; PIC > 0.5, high polymorphism. In this study, the PIC values of the 46 primers ranged from 0.3902 to 0.9109 (Table 3), with an average value of 0.6458. Ten of the 46 markers used in this study showed a medium level of polymorphism; 36 markers were highly polymorphic, whereas no markers showed a low level of polymorphism, thus indicating that these EST-SSR markers can provide abundant genetic information and that there was wide genetic diversity between the cotton accessions adopted in this study. The average PIC value of these disease-resistant EST-SSR markers (0.6458) were even higher than that of salt-tolerant EST-SSR markers reported in our previous study (0.4810) (Wang et al., 2014a).

Evaluation of genetic diversity and cotton evolution analysis using the developed EST-SSR markers

The collection and conservation of wild germplasm resources are important for expanding the narrow genetic base of cultivated cotton. In this study, cluster analysis was performed on the 50 cotton accessions based on the amplification results of the SSR primers to reveal the genetic similarity of these materials. The results showed that these disease-resistance markers were efficiently amplified and that they could differentiate the 50 cotton accessions and help reveal their evolution relationships.

There are 5 tetraploid species of *Gossypium*, designated (AD)1 through (AD)5 based on their genome constitutions (Wendel and Cronn, 2003) that evolved 1 to 2 million years ago from a tetraploid originating through the hybridization of *G. herbaceum* from the A genome with *G. raimondii* from the D genome (Wendel and Cronn, 2003). In our study, the high transferability of the EST-SSRs among the 50 tested accessions strongly confirmed that the different diploid and tetraploid species evolved from one common ancestor, despite their subsequent distinctly different and monophyletic evolutionary paths produced by geographical isolation and variations in ecological conditions (Fryxell, 1992).

As shown in Figure 3, the 50 accessions were divided into 2 groups with a similarity coefficient of 0.68, wherein all 3 accessions from the D genome were in one group; the other group consisted of accessions from the A genome and the tetraploid cotton accessions, which is consistent with the perspective that the A-genome of allopolyploid cotton is more similar to that of the A-genome diploids than the D-genome of the allopolyploid is to that of the D-genome diploids (Phillips, 1964). Within the A genome, the 2 *G. herbaceum* accessions were closely related to *G. arboreum* with a genetic similarity of 0.95. As indicated by Gerstel (1953), *G. arboreum* was derived from that of *G. herbaceum*, and these 2 species were set apart by a reciprocal translocation, and thus, both species are genetically closely related.

Within the tetraploid cotton accessions, at a similarity coefficient of 0.756, *G. mustelinum* was first differentiated from other tetraploid cotton accessions. Recent phylogenetic analyses have demonstrated that since formation, allopolyploid cottons have radiated into 3 lineages comprised of 5 species, and the only living descendant of one branch of the first cladogenetic event in the allopolyploids is *G. mustelinum* (Wendel et al., 1994).

Interestingly, uncommon to the perspective that *G. barbadense* is the sister-species of *G. darwinii* from the Galapagos Islands (Wendel and Percy, 1990), and *G. hirsutum* is sister to *G. tomentosum* from the Hawaiian Islands (DeJoode and Wendel, 1992), the 2 island endemics species *G. darwinii* and *G. tomentosum* were differentiated from other tetraploid cotton accessions with a similarity coefficient of 0.792, and they were separated until a similarity coefficient of 0.872; *G. barbadense* accessions were differentiated from *G. hirsutum* with a similarity coefficient of 0.825. Our EST-SSR markers may have disclosed some common aspects of these 2 island endemics species and the 2 cultivated species. *Gossypium hirsutum* L. IL2 and IL41 were clustered in a group with a similarity coefficient of 0.96; pedigree analysis showed that they were a pair of near-isogenic lines.

There were 32 *G. hirsutum* accessions evaluated in this study, which were clustered in a group with a similarity coefficient of 0.85. The narrow genetic diversity among the *G. hirsutum* cultivars was a result of using only a few underlying parents in modern cotton breeding.

Validation of cross-species transferability of EST-SST markers

There have been a few reports of studies in cotton (Buyyarapu et al., 2011; Wang et al., 2014a,b) and many reports in other plants (Koilkonda et al., 2012; Wang et al., 2013; An et al., 2013; Du et al., 2013; Zhao et al., 2013; Zhang et al., 2014; Asadi and Rashidi Monfared, 2014) examining cross-species transferability using EST-SSR markers.

In our study, the high cross-species transferability of EST-SSRs from the disease-resistant ESTs implies that there are some common genes related to disease resistance in different cotton genomes. Sequence alignment analysis of amplified fragments from the NTU189 and NTU216 loci revealed that the primer binding regions were highly conserved, except for some minor indels or base substitutions, and that microsatellite repeats were present in all loci within all 7 accessions (Figure 4). Thus, DNA sequencing showed that the high transferability could be attributed to a higher-level conservation in the flanking regions among these *Gossypium* species. Different cotton species showed significant differences in disease resistance, and the different disease-resistant phenotypes may be the manifestation of complex mutational events involving indels or base substitutions in SSR regions.

Interestingly, some differences were found between *G. arboreum* and *G. raimondii* (Figure 2), 2 diploid species that diverged from a common ancestor approximately 2 to 3 million years ago (Wikström et al., 2001). Additionally, *G. raimondii* is nearly immune to the pathogen, whereas *G. arboreum* is susceptible (Khadi et al., 2010). It is reasonable to hypothesize that these variations in the EST-SSR regions may have regulatory functions during disease resistance or may even be critical genetic factors in the disease resistance process. Li et al. (2014) suggested that tandem duplications play a significant role in expanding the nucleotide-binding site-encoding gene family in *G. raimondii* after its divergence from *G. arboreum*, segmental loss contributed to its contraction in *G. arboretum*, and the expansion and contraction in the numbers of nucleotide-binding site-encoding genes in different cotton species may have altered their resistance to *V. dahliae*.

EST-SSR markers are very useful for characterizing species relationships and introgression as well as for searching for desirable alleles from wild germplasm pools in *Gossypium*, which can be employed because of their high cross-species transferability. The transferability between the different species presented here can be used to increase the efficiency of transferring genetic information across species by the molecular tagging of important genes, particularly disease-resistance genes using the EST-SSR markers.

Genome position of EST-SSRs and homology and collinearity between *G. arboretum* and *G. raimondii* genome

In our study, 94 of the 106 markers were matched to the *G. arboretum* and *G. raimondii* genomes. The known position of these EST-SSRs will facilitate the application of these newly developed EST-SSR markers in QTL mapping, identification of candidate genes, and map-based cloning. Eighty-four of the 106 markers were matched to chromosomes of both the *G. arboretum* and *G. raimondii* genomes; 165 homologous sequences were found in the A2 genome, whereas 168 homologous sequences were found in the D5 genome. Thus, there was clearly a high degree of homology between the genomes of *G. arboretum* and *G. raimondii*, which is consistent with the origin and evolution of cotton that began 6 to 7 million years ago following a trans-oceanic dispersal event, and a D genome diverged from the African lineage

that eventually gave rise to the A genome and became a separate lineage in the Americas (Wendel and Cronn, 2003).

Among the 84 EST-SSRs, the order of most markers in the A2 genome agreed well with the order of corresponding homologous sequences in the D5 genome. As shown in **Table S2**, we found hits between GR_chr03 with both CA_chr02 and CA_chr03, which is consistent with the results of Rong et al. (2004) who described that chromosome 14 had homeologous relationships with 2 At chromosomes, chromosomes 2 and 3. The number of hits between CA_chr02 and GR_chr03 was lower than that between CA_chr03 and GR_chr03 (1 *vs* 4). However, dislocation and inversion of the markers' order occurred on some chromosomes, and regions of the nonhomologous chromosomes in the A2 and D5 genomes showed some homology.

ACKNOWLEDGMENTS

Research supported by the Natural Science Foundation of Jiangsu Province of China (#BK20131204, #BK20130429), the National Natural Science Foundation of China (#31000729), the Natural Science Foundation of Shanxi Province of China (#2009021032-1), the Qing Lan Project of Colleges and Universities in Jiangsu Province, and the Practice Innovation Training Program Projects for the Jiangsu College Students (#201410304082X).

Supplementary material

REFERENCES

Adawy SS, Diab AA, Atia MA and Hussein EH (2013). Construction of genetic linkage map with chromosomal assignment and quantitative trait loci associated with some important agronomic traits in cotton. *GM Crops Food* 4: 36-49.

Ahmad Z, Mumtaz AS, Ghafoor A, Ali A, et al. (2014). Marker Assisted Selection (MAS) for chickpea *Fusarium oxysporum* wilt resistant genotypes using PCR based molecular markers. *Mol. Biol. Rep.* 41: 6755-6762.

An ZW, Li YC, Zhai QL, Xie LL, et al. (2013). Development and characterization of novel expressed sequence tag-derived simple sequence repeat markers in *Hevea brasiliensis* (rubber tree). *Genet. Mol. Res.* 12: 5905-5910.

Asadi AA and Rashidi Monfared S (2014). Characterization of EST-SSR markers in durum wheat EST library and functional analysis of SSR-containing EST fragments. *Mol. Genet. Genomics* 289: 625-640.

Bansal U, Bariana H, Wong D, Randhawa M, et al. (2014). Molecular mapping of an adult plant stem rust resistance gene Sr56 in winter wheat cultivar Arina. *Theor. Appl. Genet.* 127: 1441-1448.

Blair MW, Giraldo MC, Buendía HF, Tovar E, et al. (2006). Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 113: 100-109.

Botstein B, White RL, Skolnick M and Davis RW (1980). Construction of a genetic linkage map using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32: 314-331.

Buyyarapu R, Kantety RV, Yu JZ, Saha S, et al. (2011). Development of new candidate gene and EST-based molecular markers for *Gossypium* species. *Int. J. Plant Genomics* 2011: ID894598.

Cao Z, Wang P, Zhu X, Chen H, et al. (2014). SSR marker-assisted improvement of fiber qualities in *Gossypium hirsutum* using *G. barbadense* introgression lines. *Theor. Appl. Genet*. 127: 587-594.

DeJoode DR and Wendel JF (1992). Genetic diversity and origin of the Hawaiian Islands cotton, *Gossypium tomentosum*. *Am. J. Bot.* 79: 1311-1319.

Du FK, Xu F, Qu H, Feng SS, et al. (2013). Exploiting the transcriptome of Euphrates poplar, *Populus euphratica* (Salicaceae), to develop and characterize new EST-SSR markers and construct an EST-SSR database. *PLoS One* 8: e61337.

Fryxell PA, Craven LA and Stewart JM (1992). A revision of *Gossypium* sect. *Grandicalyx* (Malvaceae), including the description of six new species. *Syst. Bot.* 17: 91-114.

Gerstel DU (1953). Chromosomal translocations in interspecific hybrids of the genus *Gossypium*. *Evolution* 7: 234-244.

Guo WZ, Zhang TZ, Shen XL, Yu JZ, et al. (2003). Development of SCAR marker linked to a major QTL for high fiber strength and its molecular marker assisted selection in Upland cotton. *Crop Sci.* 6: 2252-2256.

He Y, Kumar P, Shen X, Davis RF, et al. (2014). Re-evaluation of the inheritance for root-knot nematode resistance in the

Upland cotton germplasm line M-120 RNR revealed two epistatic QTLs conferring resistance. *Theor. Appl. Genet.* 127: 1343-1351.

Khadi BM, Santhy V and Yadav MS (2010). Cotton: Biotechnological Advances (Zehr UB, eds.). Springer, New York, 15-44.

Koilkonda P, Sato S, Tabata S, Shirasawa K, et al. (2012). Large-scale development of expressed sequence tag-derived simple sequence repeat markers and diversity analysis in *Arachis* spp. *Mol. Breed*. 30: 125-138.

Li FG, Fan GY, Wang KB, Sun FM, et al. (2014). Genome sequence of the cultivated cotton *Gossypium arboretum*. *Nat. Genet.* 46: 567-574.

Lü YD, Cai CP, Wang L, Lin SY, et al. (2010). Mining, characterization, and exploitation of EST-derived microsatellites in *Gossypium barbadense*. *Chin. Sci. Bull.* 55: 1889-1893.

Maruthachalam K, Atallah ZK, Vallad GE, Klosterman SJ, et al. (2010). Molecular variation among isolates of *Verticillium dahlia* and polymerase chain reaction-based differentiation of races. *Phytopathology* 100: 1222-1230.

Nei M (1973). Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. U. S. A.* 70: 3321-3323.

Paterson AH, Brubaker CL and Wendel JF (1993). A rapid method for extraction of cotton (*Gossypium* spp) genomic DNA suitable for RFLP or PCR analysis. *Plant Mol. Biol. Rep.* 11: 122-127.

Phillips LL (1964). Segregation in new allopolyploids of *Gossypium*. V. Multivalent formation in New World x Asiatic and New World x wild American hexaploids. *Am. J. Bot.* 51: 324-329.

Qin L, Zhang Z, Zhao X, Wu X, et al. (2014). Survey and analysis of simple sequence repeats (SSRs) present in the genomes of plant viroids. *FEBS Open Bio.* 4: 185-189.

Rong JK, Abbey C, Bowers JE, Brubaker CL, et al. (2004). A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics* 166: 389-417.

Saha MC, Mian MA, Eujayl I, Zwonitzer JC, et al. (2004). Tall fescue EST-SSR markers with transferability across several grass species. *Theor. Appl. Genet.* 109: 783-791.

Scott KD, Eggler P, Seaton G, Rossetto M, et al. (2000). Analysis of SSRs derived from grape ESTs. *Theor. Appl. Genet.* 100: 723-726.

Ulloa M, Hutmacher RB, Roberts PA, Wright SD, et al. (2013). Inheritance and QTL mapping of Fusarium wilt race 4 resistance in cotton. *Theor. Appl. Genet.* 126: 1405-1418.

Wang BH, Zhu P, Yuan YL, Wang CB, et al. (2014a). Development of EST-SSR markers related to salt tolerance and their application in genetic diversity and evolution analysis in *Gossypium*. *Genet. Mol. Res.* 13: 3732-3746.

Wang HT, Li XM, Gao WH, Jin X, et al. (2014b). Comparison and development of EST-SSRs from two 454 sequencing libraries of *Gossypium barbadense*. *Euphytica* 198: 277-288.

Wang Z, Yan H, Fu X, Li X, et al. (2013). Development of simple sequence repeat markers and diversity analysis in alfalfa (*Medicago sativa* L.). *Mol. Biol. Rep*. 40: 3291-3298.

Wendel JF and Percy RG (1990). Allozyme diversity and introgression in the Galapagos endemic *Gossypium darwinii* and its relationship to continental *G. barbadense*. *Biochem. Syst. Ecol.* 18: 517-528.

Wendel JF and Cronn RC (2003). Polyploidy and the evolutionary history of cotton. *Adv. Agron.* 78: 139-186.

Wendel JF, Rowley R and Stewart JM (1994). Genetic diversity in and phylogenetic relationships of the Brazilian endemic cotton, *Gossypium mustelinum* (Malvaceae). *Plant Syst. Evol.* 192: 49-59.

Wikström N, Savolainen V and Chase MW (2001). Evolution of the angiosperms: calibrating the family tree. *Proc. Biol. Sci.* 268: 2211-2220.

Zhang HB, Li YN, Wang BH and Chee PW (2008). Recent advances in cotton genomics. *Int. J. Plant Genomics* 2008: 742304.

Zhang M, Mao WH, Zhang GP and Wu FB (2014). Development and characterization of polymorphic EST SSR and genomic SSR markers for Tibetan annual wild barley. *PLoS One* 9: e94881.

Zhao YL, Williams R, Prakash CS, He GH (2013). Identification and characterization of gene-based SSR markers in date palm (*Phoenix dactylifera* L.). *BMC Plant Biol.* 12: 237.

Zhao YL, Wang HM, Chen W and Li YH (2014). Genetic structure, linkage disequilibrium and association mapping of Verticillium wilt resistance in elite cotton (*Gossypium hirsutum* L.) germplasm population. *PLoS One* 9: e86308.

Zhong S, Leng Y, Friesen TL, Faris JD, et al. (2009). Development and characterization of expressed sequence tag-derived microsatellite markers for the wheat stem rust fungus *Puccinia graminis* f. sp. *Tritici*. *Phytopathology* 99: 282-289.