Genetics and Molecular Research

SSR-based association mapping of salt tolerance in cotton (*Gossypium hirsutum* L.)

Y.L. Zhao, H.M. Wang, B.X. Shao, W. Chen, Z.J. Guo, H.Y. Gong, X.H. Sang, J.J. Wang and W.W. Ye

State Key Laboratory of Cotton Biology, Institute of Cotton Research of CAAS, Anyang, Henan, China

Corresponding author: H.M. Wang
E-mail: aywhm@163.com / yunleizhao2002@126.com

Genet. Mol. Res. 15 (2): gmr.15027370
Received August 5, 2015
Accepted January 4, 2016
Published May 25, 2016
DOI http://dx.doi.org/10.4238/gmr.15027370

ABSTRACT. The identification of simple sequence repeat (SSR) markers associated with salt tolerance in cotton contributes to molecular assisted selection (MAS), which can improve the efficiency of traditional breeding. In this study, 134 samples of upland cotton cultivars were selected. The seedling emergence rates were tested under 0.3% NaCl stress. A total of 74 SSR markers were used to scan the genomes of these samples. To identify SSR markers associated with salt tolerance, an association analysis was performed between salt tolerance and SSR markers using TASSEL 2.1, based on the analysis of genetic structure using Structure 2.3.4. The results showed that the seedling emergence rates of 134 cultivars were significantly different, and 27 salt-sensitive and 10 salt-tolerant cultivars were identified. A total of 148 loci were found in 74 SSR markers involving 246 allelic variations, which ranged from 2 to 7 with an average of 3.32 per SSR marker. The gene diversity ranged from 0.0295 to 0.4959, with the average being 0.2897. The polymorphic information content ranged from0.0290 to 0.3729, with the average being 0.2381. This natural population was classified into two subgroups by Structure 2.3.4, containing 89 and 45 samples, respectively. Finally, eight SSR sites associated with salt tolerance ware

found through an association analysis, with the rate of explanation ranging from 2.91 to 7.82% and an average of 4.32%. These results provide reference data for the use MAS for salt tolerance in cotton.

Key words: Upland cotton; Associate analysis; SSR molecular markers; Salt tolerance

INTRODUCTION

Soil salinity and alkalinity are important abiotic stresses that affect crop productivity (Li et al., 2012). Globally, around 20% of irrigated agricultural land has been reported to be affected by salinity and alkalinity (Shi, 2004), with the proportion in China being higher than the global average (Lin and Dilbar, 2007). Utilization of soil salinity and alkalinity has become an important objective for scientists. As a naturally salt tolerant crop, cotton is regarded as an ideal pioneer for the utilization of soil salinity and alkalinity. Furthermore, the breeding of salt tolerant cultivars is important for improving cotton yields in fields with high soil salinity and alkalinity (Zhang et al., 2011).

With rapid developments in the fields of molecular biology and biotechnology, especially with the exploitation of molecular markers, exploring quantitative traits of plants by association mapping has become a focus of plant genome research. Association mapping, which is based on linkage disequilibrium (LD), has been used to investigate the relationships between target traits and molecular markers or candidate genes in maize, rice, soybean, and cotton (Zhu et al., 2008). Thirunavukkarasu et al. (2014) studied the functional mechanisms of drought tolerance in subtropical maize using genome-wide association mapping, and found that single nucleotide polymorphisms (SNPs) associated with ABA-dependent signaling pathways played a major role in the plant's response to stress by regulating a series of functions including flowering, root development, auxin metabolism, guard cell functions, and the scavenging of reactive oxygen species. Kumar et al. (2015) implemented Genome-Wide Association Mapping (GWAS) to identify loci controlling salinity tolerance in rice, and reported 60 SNPs (loci) to be significantly associated with the Na^+/K^+ ratio and other traits observed under stress conditions. In soybean, molecular markers associated with seed germination under salt stress were detected by genome-wide association analysis, resulting in the detection of eight SNPs or suggestive SNPs that were co-associated with two salt tolerance indices (Kan et al., 2015). Zhao et al. (2014) genotyped 158 elite cotton germplasms using 212 simple sequence repeat (SSR) markers and studied genetic structure and linkage disequilibrium (LD), before performing association mapping for Verticillium wilt resistance. This resulted in the detection of 42 loci that were significantly associated with Verticillium wilt resistance in cotton. Despite this, to date, no studies have identified genes associated with salinity tolerance using SSR marker-based association mapping in cotton. This study aimed to identify SSR markers associated with salt tolerance at the seedling stage in diverse cotton accessions using association analysis. The objectives of the study were i) to analyze the genetic variations of traits contributing to salt tolerance at the seedling stage, ii) to analyze the genetic structure of a diverse cotton collection, and, iii) to identify SSRs and possible candidate genes underlying salt tolerance at the seedling stage in cotton. The major SSR loci and possible candidate genes detected in this study may be useful for the genetic improvement of cotton and will provide a basis for marker assisted selection (MAS) for salt tolerance in cotton.

MATERIAL AND METHODS

Plant materials

A total of 134 cotton (*Gossypium hirsutum* L.) cultivars were selected for this study, including intermediate breeding materials (62), main cultivars of different cotton-growing regions in China (48), transgenic materials (22), and cultivars from America (3). A salt-tolerant cultivar (Zhong 07) and a salt-sensitive cultivar (Zhong S9612) were used as controls for the estimation of salt tolerance.

Phenotypic evaluation

Salt tolerance was evaluated using the method of sand culture. Culture medium was created using the following procedure: sodium chloride (NaCl), high temperature-sterilized sand and sterile water were mixed thoroughly, with a NaCl content of 0.3% by weight of the sand and a water content of 22% by the total weight of NaCl and sand. Seeds of different cultivars were planted in germinating boxes containing aliquots of culture medium under conditions of 14-h light/day, and day and night temperatures of 28° and 25°C, respectively. The experimental design was a randomized block with three replicates. After two weeks, the emergence rate was estimated for each sample. ANOVA was performed to compare the emergence rate of 134 cultivars by SAS 9.2 software.

Isolation of plant DNA

DNA was isolated from cotton seedling leaves by means of a cetyltrimethylammonium bromide (CTAB) procedure (Paterson et al., 1993). Each PCR reaction was carried out in a volume of 10 μL containing 1X PCR buffer, 1.5 mM $MgCl_2$, 0.20 mM dNTPs, 2.5 μM each of upstream and downstream primers, and 0.5U Taq polymerase. PCR conditions were as follows: 95°C initial denaturation for 3 min; 94°C for 30 s, 56°C for 90 s, and 72°C for 1 min for 30 cycles followed by an extension at 72°C for 5 min. PCR products were analyzed by 10% polyacrylamide gel electrophoresis. The gels were silver-stained to visual differential bands among different samples (Zhang et al., 2000).

Statistical analysis

The population structure of 134 cultivars was estimated using the model-based (Bayesian) cluster software STRUCTURE 2.3.4 (Pritchard and Wen, 2007). STRUCTURE was run under the 'admixture model' with a burn-in period of 10,000 followed by 100,000 replications of Markov Chain Monte Carlo. Five independent runs were performed with the number of clusters (K) varying from 1 to 10. An ad hoc measure, *Dk*, based on the relative rate of change in the likelihood of the data between successive K values was used to determine the optimal number of clusters (Evanno et al., 2005). Marker-trait association was estimated using the general linear model (GLM) of the TASSEL 2.1 software package (Bradbury et al., 2007), in which the percentage of admixture of each accession (Q matrix) was used as a covariate to conduct the regression between phenotypic variation and markers.

RESULTS

Phenotypic analysis of salt tolerance

Some salt-tolerant and salt-sensitive cultivars were detected by estimating the emergence rate of 134 cultivars under 0.3% NaCl stress (Table 1). The frequency distribution histogram was produced according to the emergence rate of 134 cultivars (Figure 1). The emergence rate of most cultivars was in the range of 60 to 80%. The salt-sensitive cultivar Zhong S9612 had an emergence rate of 37.67%, and the salt-tolerant cultivar Zhong 07 had an emergence rate of 84.34%. A total of 27 cultivars were regarded as salt-sensitive and had lower emergence rates than the salt-sensitive cultivar Zhong S9612, and 10 cultivars were regarded as salt-tolerant and had higher emergence rates than the salt-tolerant cultivar Zhong 07 (Table 1).

Table 1. Salt tolerant and salt sensitive materials screened during the preliminary selection.

Salt-sensitive material	Seedling emergence rate (%)	Salt-tolerant material	Seedling emergence rate (%)
2010yc-2	0.47	ZhongZZS-31	84.44
ZongZZS-29	0.67	Yongmian2	84.54
YNB-1	3.08	ZhongZZS-30	87.92
Jinke178	3.33	T-1	88.00
ZhongZZS-36	3.79	ZhongZZS-34	88.26
ZhongZZS-45	4.50	ZhongZZS-42	89.05
ZhongZZS-22	5.50	ZhongZZS-4	89.74
80619	10.00	ZhongZZS-15	92.50
50217	11.23	Hebeikanghuang	92.86
Wuzhuan5	12.75	Zhong3538	93.98
ZhongZZS-14	13.00		
Wuzhuan4	14.23		
ZhongZZS-18	14.35		
Xinluzhong36	17.00		
Han109	17.00		
2010BP4	17.50		
Arcot-1	21.11		
Nongdamian8	29.00		
Xinqiumian1	30.38		
ZhongZZS-19	30.50		
Difen765078	30.67		
ZhongZZS-47	34.00		
GK99-1	35.00		
2010BP5	35.90		
Zhong990659	36.50		
Wuzhuan2	37.00		
Jimian228	37.00		

ANOVA showed that there were significant differences ($P < 0.01$) for the estimated emergence rate among the 134 cultivars (Table 2), implying that sufficient phenotypic polymorphism existed in this group to make it suitable for association mapping.

Genetic diversity

Allele number, gene diversity, and polymorphism information content (PIC) were calculated to estimate genetic diversity within the group (Table 3). The gene diversity and PIC

were 0.2897 (0.0295-0.4959) and 0.2381 (0.0290-0.3729), respectively. The number of alleles for 74 SSR markers was in the range of 2-7, of which the markers BNL3823and TMB1296 were the most abundant. The total number of polymorphic SSR loci in 134 cultivars was 148, involved in 246 alleles, with an average of 3.32 alleles per locus.

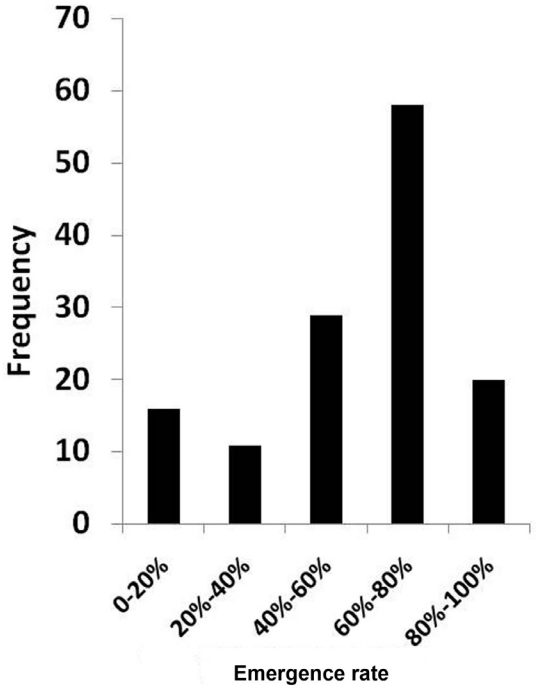

Figure 1. Frequency distribution diagram of the emergence rate of seedlings of 134 upland cotton cultivars under the 0.3% NaCl stress.

Table 2. ANOVA of the seedling emergence rate of 134 upland cotton cultivars under 0.3% NaCl stress.

Source	d.f.	Squares	Mean square	F value	Pr> F
Model	133	18.03	0.136	10.46	<0.0001
Error	225	2.915	0.013		
Corrected total	358	20.95			

Population structure of 134 cotton cultivars

Inference of the population structure of the 134 cotton cultivars was performed using the model-based software STRUCTURE. The LnP(D) value increased continuously with K from 1 to 10 (Figure 2A), and the highest ΔK value was observed at K = 2 followed by a drastic decline of ΔK from K= 3 (Figure 2B). Therefore, this group could be divided into two subgroups (Figure 3). Using a probability of membership threshold of 50%, 89 and 45 lines were assigned into the two subgroups, respectively.

Table 3. Diversity of 74 simple sequence repeat (SSR) markers.

Marker	Chromosome	Gene diversity	PIC*	No. of alleles
NAU5120	Chr16	0.4368	0.3410	4
NAU5152	Chr7/16	0.1752	0.1599	2
NAU5428	Chr11	0.4889	0.3694	2
TMB1767	Chr13/18	0.1056	0.1000	4
BNL2733	Chr7	0.1988	0.1791	3
NAU2277	Chr2	0.4469	0.3465	6
HAU0119	Chr17	0.3427	0.2837	4
DPL0209	Chr11	0.3017	0.2555	2
DPL0212	Chr19	0.4959	0.3729	2
NAU2627	Chr16	0.0927	0.0882	3
DPL0249	Chr18	0.2439	0.2136	2
NAU2820	Chr7/16	0.4318	0.3381	4
NAU2894	Chr19	0.2647	0.2238	3
NAU2954	Chr23/25	0.1000	0.0946	2
NAU3017	Chr13/18	0.3516	0.2887	4
NAU3096	Chr5/19	0.3420	0.2831	3
NAU3100	Chr23	0.4839	0.3667	4
NAU3201	Chr8/24	0.0508	0.0495	4
NAU3212	Chr5	0.3168	0.2658	4
NAU3414	Chr9/23	0.2902	0.2478	2
NAU3419	Chr2	0.4500	0.3483	3
NAU3468	Chr13	0.3516	0.2887	4
NAU3486	Chr16	0.0437	0.0426	4
NAU3609	Chr19	0.3177	0.2671	4
NAU3820	Chr14	0.3669	0.2850	4
NAU4031	Chr5	0.2275	0.2015	4
NAU4042	Chr19	0.0295	0.0290	3
NAU5064	Chr11	0.2633	0.2282	4
NAU5107	Chr1	0.1315	0.1227	3
NAU5163	Chr1	0.1118	0.1052	3
NAU5233	Chr3	0.1759	0.1526	3
NAU5351	Chr10/11/13	0.4219	0.3325	3
NAU5433	Chr6	0.3661	0.2987	4
NAU5463	Non	0.3288	0.2741	3
NAU5467	Chr14	0.3618	0.2958	2
BNL1026	Chr7/16	0.1898	0.1712	2
BNL1040	Chr18	0.1056	0.1000	4
BNL2766	Chr7/16	0.0854	0.0817	2
CIR0253	Chr5/22	0.2270	0.2009	2
CIR0328	Chr5	0.4930	0.3715	4
JESPR204	Chr5/13/18	0.0724	0.0697	4
NAU1043	Chr7	0.3513	0.2701	3
NAU2121	Chr5	0.4918	0.3709	4
NAU2274	Chr5/19	0.3420	0.2831	3
NAU2980	Chr18	0.1506	0.1392	3
CIR0412	Chr7	0.1062	0.0980	4
NAU2265	Chr2	0.4485	0.4485	4
NAU3424	Chr16	0.1087	0.1087	3
NAU3703	Chr11	0.2687	0.2687	4
JESPR292	Chr16	0.1752	0.1599	2
NAU3839	Chr3	0.3410	0.2829	2
BNL3295	Chr6	0.0718	0.0693	4
BNL3414	Chr12	0.1253	0.1175	2
NAU1369	Chr24/25	0.3082	0.2599	4
NAU2723	Chr9	0.4554	0.3517	2
NAU2580	Chr6/25	0.4627	0.3556	4
BNL3469	Non	0.1877	0.1642	6
BNL3823	Chr23	0.0448	0.0438	7
BNL3971	Chr2/17	0.4224	0.3332	4

Continued on next page

Table 3. Contributed.

Marker	Chromosome	Gene diversity	PIC*	No. of alleles
BNL3988	Chr4	0.2726	0.2354	2
BNL3255	Chr8/4	0.2777	0.2376	2
BNL3452	Chr19	0.2293	0.2026	3
BNL1053	Chr11/21	0.4890	0.3694	4
TMB1296	Chr19/25	0.3979	0.3127	7
NAU0980	Chr11	0.2617	0.2271	3
NAU1047	Chr23	0.4462	0.3464	3
NAU1102	Chr19	0.4178	0.3301	4
SWU10056	Chr16	0.2894	0.2282	4
SWU10161	Chr16	0.4298	0.3369	2
SWU10214	Chr16	0.4264	0.3349	2
SWU10054	Chr16	0.2844	0.2253	4
SWU10163	Chr16	0.4351	0.3400	4
JESPR274	Chr9/23/26	0.3407	0.2606	2
SWU10218	Chr16	0.4945	0.3722	2

PIC = polymorphism information content.

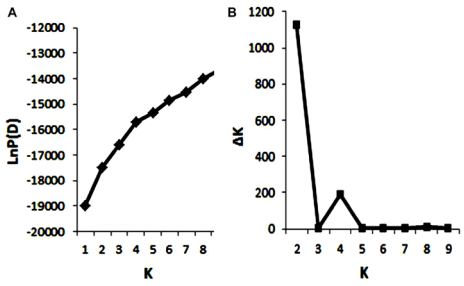

Figure 2. Line chart of K (the number of clusters) with LnP(D) (the log likelihood of the data) and Δ*K* [an ad hoc statistic based on the second-order rate of change in LnP(D) between successive K values].

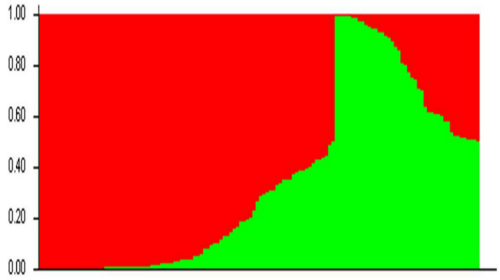

Figure 3. Population structure of 134 upland cotton cultivars. Numbers beside the vertical axis represent the probability of membership, and the two subgroups are represented by red and green, respectively.

Association analysis of SSR markers with salt tolerance

Associations between 74 marker loci and salt tolerance were determined by the GLM method. Eight marker loci with the phenotypic variation range of 2.91 to 7.82% were found to be significantly associated with salt tolerance in cotton (Table 4). Among these eight marker loci, two (NAU2580-1 and NAU2580-2) were highly significantly associated with salt tolerance and explained more phenotypic variation (7.82 and 6.26%, respectively) than other loci.

Table 4. Loci associated with traits and their explained phenotypic variation.

Marker	Chromosome	P value	Explained phenotypic variation (%)
NAU2580-1**	Chr6/25	0.001	7.82
NAU2580-2**	Chr6/25	0.004	6.26
BNL3414*	Chr9/23	0.014	4.50
HAU0119-1*	Chr17	0.029	3.56
SWU10054-1*	Chr16	0.038	3.23
TMB1296-1*	Chr19	0.039	3.20
NAU3100-1*	Chr23	0.043	3.10
NAU3486-2*	Chr16	0.049	2.91

*and **Indicate a significant correlation at $P < 0.05$ and $P < 0.01$, respectively.

DISCUSSION

Salt tolerance in cotton is very complex, as shown by Shen et al. (2001), who studied the inheritance of salt tolerance using Hayman's diallel cross analysis. Those authors suggested that both additive and dominant effects appeared to be important for the expression of variation under salt stress, and that the effect of genes with additive properties was more pronounced. For a long time, the focus of cotton breeding has been on improving the yield, fiber quality, and resistance to disease and pests, and few studies have investigated the improvement of salt tolerance in cotton. It is important to collect salt tolerant germplasm resources in order to enrich the genetic variation in cultivated salt tolerance cotton varieties. In this study, 134 cotton cultivars were used to estimate salt tolerance in an incubator under stable test conditions. The results showed that 27 cultivars could be regarded as salt-sensitive and 10 could be regarded as salt-tolerant, which provides a reference for cotton salt tolerance breeding. ANOVA showed that significant differences exist in phenotypes of salt tolerance among the 134 cultivars, implying that sufficient phenotypic variation exists in this group, thus making it suitable for association mapping.

In traditional breeding, it is very inefficient to breed varieties with salt tolerance by hybridizing different salt-tolerance lines. By screening molecular markers that are related to salt tolerance, and combining molecular and traditional breeding, the efficiency of salt tolerant breeding in cotton can be greatly improved. Recently, Zhang et al. (2010) screened two salt-tolerant cotton accessions, *G. hirsutum* L. cv. CCRI35 and *G. hirsutum* L. Zhong07, and two salt-sensitive cotton accessions, *G. hirsutum* L. cv. CCRI12 and *G. hirsutum* L. Xinyan96-48, with 274 SSR markers, and subsequently identified 10 primer pairs that could be used to detect salt-tolerance in cotton. Wang et al. (2014) developed a total of 132 pairs of non-redundant expressed sequence tag-simple sequence repeat (EST-SSR) primers related to salt-tolerance according to salt-resistant cotton ESTs and performed genetic diversity and evolution analyses on cotton. To some extent, these studies have accelerated the application of molecular markers to the breeding of salt-tolerance in cotton.

Two measures, quantitative trait loci (QTL) mapping and association mapping, are usually used to detect molecular markers associated with yield, quality, and resistance. In contrast to QTL mapping, which is based on bi-parental populations, association mapping is based on LD and uses a sample of lines from the broader breeding population that are unrelated to any specific crossing design (Zhu et al., 2008). Therefore, a higher number of historical recombination events can be explored in natural populations than in the bi-parental segregating populations, resulting in a higher resolution of QTL mapping (Ersoz et al., 2007).

In this study, association mapping was conducted to estimate marker-trait association using the GLM in the TASSEL 2.1 software package. False association was avoided by estimating population structure and proceeding the regression between the phenotypic variation and markers with the percentages of admixture of each accession (Q matrix) as covariates. A total of eight marker loci associated with salt tolerance in cotton were identified, among which two loci (NAU2580-1and NAU2580-2) explained the most phenotypic variation of salt tolerance, implying a strong association between the marker NAU2580 and salt tolerance genes in cotton. This paper tentatively explored the associations between SSR markers and salt tolerance in cotton, and the results provide references for MAS of salt tolerance in cotton.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the National Hi-Tech Research and Development Program of China (Grant #2012AA101108-02-02) and by the Basic R&D Operation Special Fund for the Central Level, Non-Profit, Scientific Research Institutes, China (#1610162015A02). The funders had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

REFERENCES

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, et al. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633-2635. http://dx.doi.org/10.1093/bioinformatics/btm308

Ersoz ES, Yu J and Buckler ES (2007). Applications of linkage disequilibrium and association mapping in crop plants. In: Genomics-assisted crop improvement-Vol 1: Genomics approaches and platforms (Varshney RK and Tuberosa R, eds.).Springer, Dordrecht, 97-120.

Evanno G, Regnaut S and Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-2620. http://dx.doi.org/10.1111/j.1365-294X.2005.02553.x

Kan G, Zhang W, Yang W, Ma D, et al. (2015). Association mapping of soybean seed germination under salt stress. *Mol. Genet. Genomics* 290: 2147-2162. http://dx.doi.org/10.1007/s00438-015-1066-y

Kumar V, Singh A, Mithra SVA, Krishnamurthy SL, et al. (2015). Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa*). *DNA Res.* 22: 133-145. http://dx.doi.org/10.1093/dnares/dsu046

Li J, Pu L, Zhu M and Zhang R (2012). The status of soil salinization and future research focus. *Acta Geogr. Sin.* 67: 1233-1245.

Lin J and Dilbar S (2007). Progress in the study on soil salinization. *J. Xinjiang Univ.* (*Nat. Sci. Ed.*) 24: 318-323.

Paterson AH, Brubaker CL and Wendel JF (1993). A rapid method for extraction of cotton (*Gossypium*spp) genomic DNA suitable for RFLP and PCR analysis. *Plant Mol. Biol. Rep.* 11: 122-127. http://dx.doi.org/10.1007/BF02670470

Pritchard JK and Wen W (2007). Documentation for STRUCTURE Software: Version 2.2. The University of Chicago Press, Chicago.

Shen F, Yu Y, Bi J, Liu F, et al. (2001). A diallel analysis of salt tolerance in upland cotton. *Acta Agron. Sin.* 27: 50-54.

Shi Y (2004). Research of land desertification and soil and water resources in the northwest area. China Sciences Press, Peking.

Thirunavukkarasu N, Hossain F, Arora K, Sharma R, et al. (2014). Functional mechanisms of drought tolerance in subtropical maize (*Zea mays* L.) identified using genome-wide association mapping. *BMC Genomics* 15: 1182. http://dx.doi.org/10.1186/1471-2164-15-1182

Wang BH, Zhu P, Yuan YL, Wang CB, et al. (2014). 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. http://dx.doi.org/10.4238/2014.May.13.1

Zhang GW, Lu HL, Zhang L, Chen BL, et al. (2011). Salt tolerance evaluation of cotton (Gossypium hirsutum) at its germinating and seedling stages and selection of related indices. *Ying Yong Sheng Tai Xue Bao* 22: 2045-2053.

Zhang J, Wu Y, Guo W and Zhang T (2000). Fast screening of microsatellite markers in cotton with PAGE/silver staining. *J. Cotton Sci.* 12: 267-269.

Zhang L, Ye W, Wang J and Fang B (2010). Studies of Salinity-tolerance with SSR Markers on *G. hirsutum* L. *J. Cotton Sci.* 22: 175-180.

Zhao Y, Wang H, Chen W and Li Y (2014). Genetic structure, linkage disequilibrium and association mapping of *Verticillium* wilt resistance in elite cotton *(Gossypium hirsutum* L.) germplasm population. *PLoS One* 9: e86308. http://dx.doi.org/10.1371/journal.pone.0086308

Zhu C, Gore M, Buckler ES and Yu J (2008). Status and prospects of association mapping in plants. *Plant Genome* 1: 5-20. http://dx.doi.org/10.3835/plantgenome2008.02.0089