



## Development of EST-SSR markers related to salt tolerance and their application in genetic diversity and evolution analysis in *Gossypium*

B.H. Wang<sup>1,2</sup>, P. Zhu<sup>1</sup>, Y.L. Yuan<sup>2</sup>, C.B. Wang<sup>3</sup>, C.M. Yu<sup>1</sup>, H.H. Zhang<sup>1</sup>, X.Y. Zhu<sup>1</sup>, W. Wang<sup>4</sup>, C.B. Yao<sup>1</sup>, Z.M. Zhuang<sup>1</sup> and P. Li<sup>1,2</sup>

<sup>1</sup>School of Life Sciences, Nantong University, Nantong, Jiangsu Province, China

<sup>2</sup>State Key Laboratory of Cotton Biology, Anyang, Henan Province, China

<sup>3</sup>Agricultural Biotechnology Research Center,

Shanxi Academy of Agricultural Sciences, Taiyuan, Shanxi Province, China

<sup>4</sup>Agricultural Science Institute of Coastal Region of Jiangsu, Yancheng, Jiangsu Province, China

Corresponding author: P. Li

E-mail: pingli6@yahoo.com

Genet. Mol. Res. 13 (2): 3732-3746 (2014)

Received August 14, 2013

Accepted December 2, 2013

Published May 13, 2014

DOI <http://dx.doi.org/10.4238/2014.May.13.1>

**ABSTRACT.** Salt stress is becoming one of the major problems in global agriculture with the onset of global warming, an increasing scarcity of fresh water, and improper land irrigation and fertilization practices, which leads to reduction of crop output and even causes crop death. To speed up the exploitation of saline land, it is a good choice to grow plants with a high level of salt tolerance and economic benefits. As the leading fiber crop grown commercially worldwide, cotton is placed in the moderately salt-tolerant group of plant species, and there is promising potential to improve salt tolerance in cultivated cotton. To facilitate the mapping of salt-tolerant quantitative trait loci in cotton so as to serve the aims of salt-tolerant molecular breeding in cotton, it is necessary to develop salt-tolerant molecular markers.

The objective of this research was to develop simple sequence repeat (SSR) markers based on cotton salt-tolerant expressed sequence tags. To test the efficacy of these SSR markers, their polymorphism and cross-species transferability were evaluated, and their value was further investigated on the basis of genetic diversity and evolution analysis.

**Key words:** *Gossypium*; Salt-tolerant species; EST-SSR; Genetic diversity; Evolution analysis

## INTRODUCTION

Soil salinization is a global problem in agriculture and the ecological environment. Currently, there are 1 billion km<sup>2</sup> of saline-alkali soil in the world, covering 7.6% of the land area (Liu and Wang, 1998), which cannot be directly used in agriculture. Meanwhile, the salinization of land is becoming more and more serious with global warming and increasing scarcity of fresh water, as well as improper irrigation and fertilization of the soil (Flowers, 1999). Consequently, salt stress is becoming one of the major problems in global agriculture (Flowers, 1999). To advance the sound management of salinized land, an excellent approach is to grow plants with a high level of salt tolerance to enhance its economic benefits. As the leading fiber crop worldwide grown commercially in the temperate and tropical regions of more than 50 countries, cotton is placed in the moderately salt-tolerant group of plant species, and there is promising potential to improve salt tolerance in cultivated cotton (Ashraf, 2002). Soil with 0.2% or less salt content is beneficial to seedling emergence and growth of cotton, whereas 0.3% or higher salt content can be harmful to cotton (Dai et al., 2010). Consequently, it is important to improve salt tolerance in cotton to 0.3% or higher, so as to expedite the use of saline-alkali land and meet the requirements of modern agriculture development in coastal areas.

It is labor-intensive and time-consuming to improve the salt tolerance of cotton in traditional breeding. The development of DNA molecular markers will help enhance the progress of developing improved varieties. To breed salt-tolerant cotton varieties, it is necessary to develop salt-tolerant molecular markers for use in quantitative trait loci (QTL) mapping of salt-tolerance and marker-assisted selection of salt-tolerant cotton.

Simple sequence repeats (SSRs) are polymerase chain reaction (PCR)-based markers, and generally they have the distinguishing features of 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 source of PCR-based markers for targeting SSRs. EST-derived SSRs (EST-SSRs) from the transcribed regions of the DNA are generally more conserved across species than genomic SSRs from the untranscribed regions (Cuadrado and Schwarzacher, 1998), and EST-SSRs have higher cross-species transferability compared to genomic SSRs. Moreover, the functions of EST-SSRs can be presumed, and marker development is easy to perform at low cost. EST-SSRs have been widely developed and applied in plants such as barley (Thiel et al., 2003), pepper (Yi et al., 2006), rubber tree (Feng et al., 2009), chickpea (Choudhary et al., 2009), and sesame

(Wei et al., 2011). In recent years, an increasing number of ESTs of cotton species including *Gossypium hirsutum*, *G. raimondii*, *G. arboreum*, *G. barbadense*, and *G. herbaceum* have been included in GenBank (<http://www.ncbi.nlm.nih.gov/dbEST/>). Meanwhile, great efforts have also been made to develop EST-SSRs (Qureshi et al., 2004; Taliercio et al., 2006; Guo et al., 2006) for cotton. These SSR markers have been widely applied in genetic diversity analysis (Arunita et al., 2010), construction of linkage maps (Guo et al., 2007; Lin et al., 2009), QTL mapping and marker-assisted breeding (Zhang et al., 2009; Liu et al., 2012).

In this research, a total of 132 pairs of non-redundant EST-SSR primers were effectively developed according to the salt-resistant cotton ESTs. The primers were applied to 38 cotton accessions, including 7 diploids in the A genome, 7 diploids in the D genome, and 24 accessions covering the 5 tetraploids. The markers were evaluated, and cross-species transferability was analyzed to support QTL mapping of salt-tolerant traits, genetic diversity analysis, and construction of cotton maps.

## MATERIAL AND METHODS

### Cotton materials

Totally, 38 cotton accessions were selected including 7 diploids in the A genome including *G. arboreum* and *G. herbaceum*, 7 diploids in the D genome including *G. raimondii*, and 24 AADD tetraploids covering all the 5 tetraploid cotton species. Their genomic groups and abbreviations are given 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 (CCRI), whereas the wild species were collected from the Wild Cotton Species Research Team of CCRI. Four of the 38 cotton materials have known salt tolerance of which 2 accessions *G. hirsutum* L. cv. CCRI35 and *G. hirsutum* L. Zhong07 are salt tolerant and two other accessions *G. hirsutum* L. cv. CCRI12 and *G. hirsutum* L. Xinyan96-48 are salt sensitive (Zhang et al., 2010).

### Development of EST-SSR primers

EST-SSR primers were developed on the basis of the known salt-tolerant ESTs as follows. Salt-tolerant genes were searched in the *Arabidopsis* Information Resource (TAIR, <http://www.arabidopsis.org/>). The *Arabidopsis* genes were used as queries for BLAST against cotton EST bank (<http://www.agcol.arizona.edu/cgi-bin/pave/Cotton/index.cgi>), and the corresponding cotton transcript contigs were found. These cotton transcript contigs were used to develop EST-SSR markers by using MISA (<http://pgrc.ipk-gatersleben.de/misa/>) and 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 100-500 bp was selected with annealing temperatures of 55°-59°C; 57°C was desired as the optimum temperature. Meanwhile, the SSR primers were BLASTed with existing primers 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. A total of 132 pairs of EST-SSR primers were developed and named "NTUXXX", in which NTU means Nantong University and XXX the serial number of the SSR primers.

**Table 1.** List of cotton accessions.

No.	Species name	Abbreviation	Genome
1	<i>G. herbaceum</i> L. cv. Jinta	<i>G. her.</i> Jt	A1
2	<i>G. herbaceum</i> L. cv. Gaotai	<i>G. her.</i> Gt	A1
3	<i>G. arboreum</i> L. cv. Yunnanfuning	<i>G. arb.</i> Ynfn	A2
4	<i>G. arboreum</i> L. cv. Sihua	<i>G. arb.</i> Sh	A2
5	<i>G. arboreum</i> L. cv. Jiangsuhongjingjijiao	<i>G. arb.</i> Jshjjj	A2
6	<i>G. arboreum</i> L. cv. Jiangling	<i>G. arb.</i> Jl	A2
7	<i>G. arboreum</i> L. cv. Haimenxiaobaihua	<i>G. arb.</i> Hmxbh	A2
8	<i>G. thurberi</i>	<i>G. thu</i>	D1
9	<i>G. davidsonii</i>	<i>G. dav</i>	D3-d
10	<i>G. klotzschianum</i>	<i>G. klo</i>	D3-k
11	<i>G. aridum</i>	<i>G. ari</i>	D4
12	<i>G. raimondii</i>	<i>G. rai</i>	D5
13	<i>G. gossypoides</i>	<i>G. gos</i>	D6
14	<i>G. trilobum</i>	<i>G. tri</i>	D8
15	<i>G. hirsutum</i> L. Zhong4133	<i>G. hir.</i> Z4133	(AD)1
16	<i>G. hirsutum</i> L. cv. Simian3	<i>G. hir.</i> Sm3	(AD)1
17	<i>G. hirsutum</i> L. 87-28	<i>G. hir.</i> 87-28	(AD)1
18	<i>G. hirsutum</i> L. Zhongyuan9111	<i>G. hir.</i> Zy9111	(AD)1
19	<i>G. hirsutum</i> L. ZhongyuanHAS-1	<i>G. hir.</i> ZyHAS-1	(AD)1
20	<i>G. hirsutum</i> L. cv. CCRI35	<i>G. hir.</i> CCRI35	(AD)1
21	<i>G. hirsutum</i> L. cv. CCRI12	<i>G. hir.</i> CCRI12	(AD)1
22	<i>G. hirsutum</i> L. Zhong2201	<i>G. hir.</i> Z2201	(AD)1
23	<i>G. hirsutum</i> L. Yahuang9103	<i>G. hir.</i> Yh9103	(AD)1
24	<i>G. hirsutum</i> L. Zhongyuan9112	<i>G. hir.</i> Zy9112	(AD)1
25	<i>G. hirsutum</i> L. ZhongAR40772	<i>G. hir.</i> Z40772	(AD)1
26	<i>G. hirsutum</i> L. Zhong07	<i>G. hir.</i> Z07	(AD)1
27	<i>G. hirsutum</i> L. Xuzhou261	<i>G. hir.</i> Xz261	(AD)1
28	<i>G. hirsutum</i> L. Jiangsudatao	<i>G. hir.</i> Jsdt	(AD)1
29	<i>G. hirsutum</i> L. TM-1	<i>G. hir.</i> TM-1	(AD)1
30	<i>G. hirsutum</i> L. PD94042	<i>G. hir.</i> PD94042	(AD)1
31	<i>G. hirsutum</i> L. Xinyan96-48	<i>G. hir.</i> Xy96-48	(AD)1
32	<i>G. hirsutum</i> L. cv. Sumian15	<i>G. hir.</i> Sm15	(AD)1
33	<i>G. hirsutum</i> L. Zhongzi4280	<i>G. hir.</i> Zz4280	(AD)1
34	<i>G. barbadense</i> L. Pima3-79	<i>G. bar.</i> P3-79	(AD)2
35	<i>G. barbadense</i> L. Hai7124	<i>G. bar.</i> H7124	(AD)2
36	<i>G. tomentosum</i>	<i>G. tom</i>	(AD)3
37	<i>G. mustelinum</i>	<i>G. mus</i>	(AD)4
38	<i>G. darwinii</i>	<i>G. dar</i>	(AD)5

## DNA extraction, PCR amplification and electrophoresis

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

The PCR mixture of 10 µL included 1X PCR buffer, 1.5 mM MgCl<sub>2</sub>, 0.25 mM dNTP, 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 for an extension of 7 min. The PCR products were separated on 10% polyacrylamide gels with silver staining 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 given: band present “1”, band absent “0”. The NTSYSpc ver 2.0 software (<http://www.exetersoftware.com/>) was used to calculate genetic similarity of the 38 cotton accessions and UPGMA was used to construct dendrograms.

The richness of alleles for each marker, which can also be regarded as the measure of usefulness of each marker in distinguishing one individual from another, was measured using the polymorphic information content (PIC) calculated as follows:  $PIC = 1 - \sum P_{ij}^2$ , given that  $P_{ij}$  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 polymorphism. 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 DH5 $\alpha$  *Escherichia coli* competent cells were conducted as described by Guo et al. (2003). At least three positive clones from each amplified fragment were simultaneously sequenced by Invitrogen Biotech (Shanghai), China. The DNAMAN program (Lynnon Biosoft, Quebec, Canada) was used to align the amplified SSR alleles. The consensus sequence was used as the reference in sequence comparisons and base mutation analyses.

## RESULTS

### Characteristics of the ESTs and EST-SSR primers

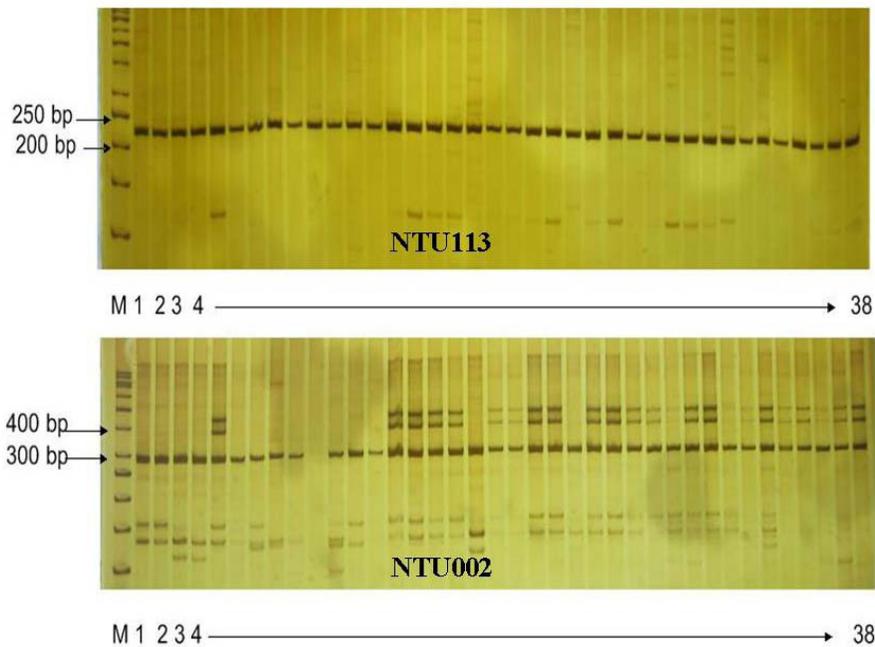
In total, 132 pairs of EST-SSR primers were developed on the basis of the salt-tolerant cotton ESTs. The primer sequences, the serial No. of the ESTs in cotton EST database, and the serial No. of the corresponding genes in TAIR are listed in [Table S1](#). Regarding the repeat motif type, the trinucleotide repeats had the highest frequency of 28.79%, followed by hexanucleotide repeats, pentanucleotide repeats, and tetranucleotide repeats with the frequencies of 26.51, 24.24 and 16.67%, respectively, whereas dinucleotide repeats had the lowest frequency of 3.79%.

### Cross-species amplification of the EST-SSR primers

The amplification products of the EST-SSR primers were 100 to 500 bp (Figure 1). Of the 132 NTU primer pairs, clear bands were amplified in 106 pairs with an effective amplification rate of 80.3%. The rest of the primers, 26 pairs had no amplification product or specific bands.

As shown in Figure 1, both single-locus markers and multilocus markers were developed in this research. The marker NTU113 with a single locus of 230 bp was amplified in all 38 accessions, whereas 3 loci were detected with the marker NTU002 that showed polymorphisms between different accessions.

Eight of the 106 effective markers were amplified with clear bands in all 38 cotton accessions, namely NTU023, NTU056, NTU097, NTU103, NTU113, NTU121, NTU125, and NTU132. Fifty-six of the 106 markers were amplified in 30-37 accessions; forty-one of these primers were amplified in 17-29 accessions. One primer NTU031 was only amplified in 12 accessions, of which 3 belonged to *G. arboreum* in the diploid A genome, 8 belonged to *G. hirsutum*, and 1 *G. tomentosum* in the tetraploid AD genome, whereas 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 80.3%.



**Figure 1.** Polyacrylamide gel electrophoresis patterns of microsatellite alleles amplified using two EST-SSR markers. PCR products could be amplified from all 38 accessions by NTU113 and with only a single locus amplified in each accession; null alleles were detected in *Gossypium klotzschianum* for NTU002 and more than one locus was amplified in each accession. Lane M = molecular weight marker; lanes 1-38 = *G. herbaceum* L. cv. Jinta, *G. herbaceum* L. cv. Gaotai, *G. arboreum* L. cv. Yunnanfuning, *G. arboreum* L. cv. Sihua, *G. arboreum* L. cv. JiangsuHongjingjijiao, *G. arboreum* L. cv. Jiangling, *G. arboreum* L. cv. Haimenxiaobaihua, *G. thurberi*, *G. davidsonii*, *G. klotzschianum*, *G. aridum*, *G. raimondii*, *G. gossypoides*, *G. trilobum*, *G. hirsutum* L. Zhong4133, *G. hirsutum* L. cv. Simian3, *G. hirsutum* L.87-28, *G. hirsutum* L. Zhongyuan9111, *G. hirsutum* L. ZhongyuanHAS-1, *G. hirsutum* L. cv. CCR135, *G. hirsutum* L. cv. CCR112, *G. hirsutum* L. Zhong2201, *G. hirsutum* L. Yahuang9103, *G. hirsutum* L. Zhongyuan9112, *G. hirsutum* L. ZhongAR40772, *G. hirsutum* L. Zhong07, *G. hirsutum* L. Xuzhou261, *G. hirsutum* L. Jiangsudatao, *G. hirsutum* L. TM-1, *G. hirsutum* L. PD94042, *G. hirsutum* L. Xinyan96-48, *G. hirsutum* L. cv. Sumian15, *G. hirsutum* L. Zhongzi4280, *G. barbadense* L. Pima3-79, *G. barbadense* L. Hai7124, *G. tomentosum*, *G. mustelinum*, and *G. darwinii*, respectively.

The 106 effective markers had different amplification efficiency, and cross-species transferability, within different cotton species (Table 2). They had much higher cross-species transferability across the A genome (86.52%) than that across the D genome (64.50%), whereas the cross-species transferability within tetraploid species was 82.78% (Table 2). Only 19 pairs of the primers were amplified in all the tetraploid cotton materials, which was a small portion of the total primers, whereas most primers were amplified only in part of the tetraploid cotton materials.

The EST-SSR markers also had different amplification efficiencies in different cotton accessions even in the same species. The highest amplification efficiency was observed in two tetraploid cotton *G. hirsutum* L. Zhong2201 and *G. hirsutum* L. Yahuang9103 with an amplification ratio of 100%, and stable bands were amplified in both accessions. For *G. klotzschianum* in the D genome, these primers had the lowest amplification efficiency of 43.3%, where only about 47 of the 106 primers were amplified.

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

Genome	No. of primers			Total bands	Total combinations	Amplification (%)
	Whole amplification	Partial amplification	Null amplification			
A1	95	7	4	642	742	86.52%
A2	66	40	0			
D1	77	0	29	486	742	64.50%
D3	44	46	16			
D4	73	0	33	2106	2544	82.78%
D5	61	0	45			
D6	81	0	25	60	0	46
D8	60	0	46			
AD1	19	87	0	44	45	17
AD2	44	45	17			
AD3	78	0	28	83	0	23
AD4	83	0	23			
AD5	102	0	4			

Whole amplification = all SSR primer pairs could produce PCR products in a given genome; partial amplification = there were no amplicons in some species in a given genome; null amplification = no amplicon of expected size was produced in any species of a given genome; total combinations = total combination of EST-SSRs x tested cotton species.

### DNA sequencing analysis of EST-SSR markers

To study how DNA polymorphisms are produced, the amplified products from NTU060 and NTU121 were sequenced across 10 accessions covering 8 species. A total of 60 clones were sequenced and the alignment results are shown in Figure 2.

For amplification products of the primer NTU121, most of the flanking sequences of the 10 cotton accessions were the same, whereas two transitions (A to G and T to C) and two transversions (A to C and G to T) were observed in the flanking sequences. For amplification products of the primer NTU060, more differences were found between the 10 cotton accessions although most of the flanking sequences were still the same. Eight transitions, 5 instances of A to G and 3 instances of T to C, and one transversion of G to C were observed. Moreover, there were deletions observed in 4 of the 10 accessions; *G. raimondii*, *G. hirsutum* L. cv. CCRI35, *G. hirsutum* L. cv. CCRI12, and *G. mustelinum*.

### Polymorphism of EST-SSR markers

Clear bands were amplified with 106 of the 132 (80.3%) EST-SSR primers in the 38 cotton accessions, and 50 primers that were amplified with obvious polymorphism and stable band types are listed in Table 3. PIC values were calculated on the basis of the alleles detected by the 50 markers in the 38 cotton accessions. A total of 133 polymorphic alleles were detected by the 50 markers with an average of 2.66 per marker, ranging between 2 and 7. The PIC values ranged between 0.2163 for NTU032 and 0.8370 for NTU082 (Table 3), with an average value of 0.4810.

### Genetic similarity index and cluster analysis of the 38 cotton accessions

As shown in Figure 3, the 38 accessions were divided into 2 groups at a similarity coefficient of 0.57.

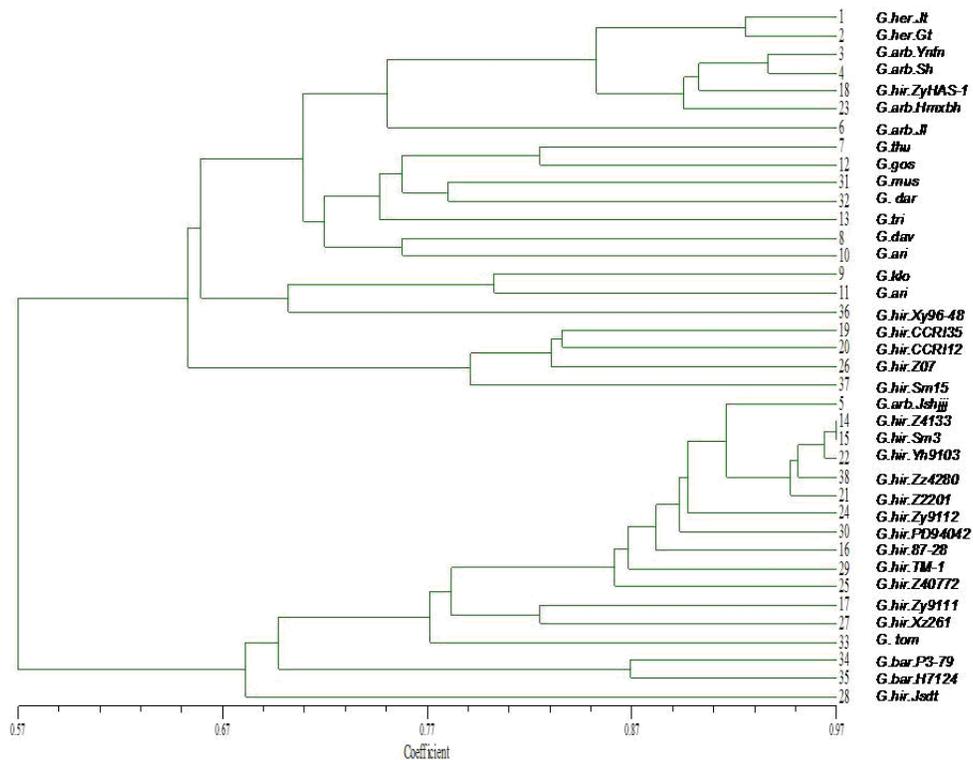


**Table 3.** Allele number and PIC value of the 50 EST-SSR primers.

EST-SSR primer	No. of alleles	PIC
NTU032	3	0.2163
NTU104	2	0.2188
NTU108	2	0.2235
NTU025	2	0.2688
NTU077	3	0.3317
NTU003	2	0.3628
NTU067	2	0.3829
NTU081	2	0.3911
NTU112	2	0.3911
NTU022	2	0.3950
NTU026	2	0.3950
NTU014	3	0.3972
NTU015	2	0.3995
NTU078	2	0.4012
NTU020	2	0.4132
NTU094	3	0.4192
NTU091	2	0.4200
NTU016	2	0.4248
NTU130	2	0.4260
NTU053	2	0.4297
NTU034	3	0.4335
NTU118	2	0.4383
NTU002	2	0.4486
NTU120	2	0.4486
NTU079	2	0.4494
NTU128	2	0.4614
NTU119	2	0.4664
NTU093	2	0.4728
NTU100	2	0.4764
NTU124	2	0.4837
NTU070	3	0.4886
NTU041	2	0.4956
NTU055	2	0.4959
NTU125	3	0.4995
NTU042	2	0.4998
NTU111	2	0.4998
NTU038	3	0.5130
NTU123	3	0.5240
NTU098	3	0.5544
NTU009	2	0.6208
NTU028	3	0.6389
NTU023	3	0.6750
NTU036	4	0.7120
NTU062	4	0.7148
NTU008	4	0.7149
NTU060	5	0.7275
NTU068	4	0.7340
NTU107	4	0.7423
NTU004	5	0.7889
NTU082	7	0.8370

All the accessions from the D genome, most of the accessions from the A genome except for *G. arboreum* L. cv. Jiangsuhongjingjijiao, and part of tetraploid cotton accessions were in one group; the other group consisted of *G. arboreum* L. cv. Jiangsuhongjingjijiao from the A genome and most of the tetraploid cotton accessions. At a similarity coefficient of 0.754, *G. arboreum* L. cv. Jiangling was distinguished from other accessions in the A genome. At 0.94, all the accessions in the A genome were differentiated from each other. *G. hirsutum* L. Jiangsudatao was distinguished from other tetraploid cotton accessions at a similarity coefficient of 0.682.

The two *G. barbadense* L. accessions *G. barbadense* L. Pima3-79 and *G. barbadense* L. Hai7124 were differentiated from other tetraploid cotton accessions at a similarity coefficient of 0.70, and they were differentiated from each other at 0.87. *G. tomentosum* was differentiated from other tetraploids at a similarity coefficient of 0.77. *G. arboreum* L. cv. Jiansuhongjingjijiao could not be differentiated until a similarity coefficient of 0.915 was reached. *G. hirsutum* Zhong4133 and Simian3 still could not be differentiated at a similarity coefficient of 0.97.



**Figure 3.** Dendrogram of 38 cotton accessions based on EST-SSR markers.

## DISCUSSION

### Development, amplification, and characteristics of the salt-tolerant EST-SSR primers

Regarding crop salt-tolerance improvement, previous research has focused mainly on gene mapping, QTL analysis, and genetic diversity analysis, but little effort has been made to develop SSR markers for salt-tolerance. Shan et al. (2006) detected SSR markers linked to salt tolerance in wheat by using the bulked segregation analysis method combined with SSR markers and then located the salt tolerance-related genes. Cheng et al. (2012) mapped 47 salt-tolerance QTLs in rice using two sets of reciprocal introgression lines. Islam et al. (2011) detected three major QTLs on chromosomes 1, 8 and 10 with phenotypic variances of 12.50,

29.0, and 20.20%, respectively, at the seedling stage of rice. Lee et al. (2004) mapped a major QTL by using SSR markers and a set of soybean near-isogenic lines. In cotton research, Zhang et al. (2010) screened 2 salt-tolerant cotton accessions, *G. hirsutum* L. cv. CCRI35 and *G. hirsutum* L. Zhong07, and 2 salt-sensitive cotton accessions, *G. hirsutum* L. cv. CCRI12 and *G. hirsutum* L. Xinyan96-48, with 274 SSR markers, and they identified 10 primer pairs that could be useful in detecting salt tolerance in cotton.

Although many EST-SSR markers have been developed from *G. hirsutum*, *G. arboreum*, *G. raimondii*, *G. barbadense*, and *G. herbaceum*, no EST-SSR primers have been developed specifically using salt-tolerance ESTs. In this research, 132 pairs of EST-SSR primers were developed based on cotton salt-tolerance ESTs obtained from BLAST with *Arabidopsis* salt-tolerance genes, so as to serve cotton salt-tolerance molecular breeding efforts. Previous research has indicated that the effective amplification rate of EST-SSRs should be between 60 and 90% (Thiel et al., 2003; Saha et al., 2004), and that some primers may have no amplification once the primer covers an mRNA editing site or the amplification product contains large introns. In our research, clear bands could be amplified in 106 of the 132 NTU primer pairs, and the effective amplification rate was 80.3%, which is a high amount within the range reported in previous studies. There may be two reasons as to why the remaining 26 primer pairs have no amplification product or specific bands. The first is that large introns exist within the amplification region, which may make the product much bigger than expected; the second is that the SSR loci and the markers are within transposons or repetitive DNA, resulting in multiple products deviating from the expected bands. Regarding no amplification in any of the 38 accessions for some primers, the reason may be that the forward primer or reverse primer covers the editing site of mRNA or that large introns exist in the genomic DNA (Vendramin et al., 2007).

PIC values can be used to evaluate the allele variation of each primer, and it is determined by the allele number and their distribution frequency (Botstein et al., 1980). Accordingly, PIC is a measure of 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 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 taken as follows:  $PIC < 0.25$ , low polymorphism;  $0.25 < PIC < 0.5$ , medium polymorphism;  $PIC > 0.5$ , high polymorphism. In this research, the PIC values of the 50 primers ranged between 0.2163 and 0.8370 (Table 3) with an average value of 0.4810. Consequently, 33 of 50 markers used in this research had a medium level of polymorphism, 14 markers were highly polymorphic, whereas only 3 had a low level of polymorphism, thus indicating that these EST-SSR markers can provide abundant genetic information, and also that there is wide genetic diversity between the cotton accessions adopted in this research.

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

The collection and conservation of wild germplasm resources are important to expand the narrow genetic base of cultivated cotton. In this research, cluster analysis was performed on the 38 cotton accessions on the basis of the amplification results of the SSR primers to reveal the genetic similarity of these materials. The results showed that these salt-tolerance

markers were efficiently amplified and that they could differentiate the 38 cotton accessions and help reveal their evolution relationship. As shown in Figure 3, all 38 accessions were divided into 2 groups at a similarity coefficient of 0.57.

In the first group, all the accessions from the D genome and most of the accessions from the A genome were grouped together, indicating that they have a close genetic relationship, which is consistent with the conclusion suggested by DNA-sequence phylogenetic data that 6-7 million years ago, following a trans-oceanic dispersal event, the D genome, a separate lineage in the new world, diverged from an African lineage that also eventually gave rise to the A genome within the old world (Wendel and Cronn, 2003). Within the A genome, the two *G. herbaceum* accessions were closely related to *G. arboreum* with a genetic similarity of 0.85. As indicated by Gerstel (1953), *G. arboreum* is derived from that of *G. herbaceum*, and these two species are set apart by a reciprocal translocation, so genetically both species are closely related. *G. hirsutum* L. ZhongyuanHAS-1 was grouped together with cotton accessions in the A genome, and it was differentiated at a similarity coefficient of 0.905, coinciding with the pedigree analysis that *G. hirsutum* L. ZhongyuanHAS-1 was derived from crosses between *G. arboreum* (A genome) and other cotton accessions.

In the second group, it is interesting that *G. arboreum* L. cv. Jiangsuhongjingjijiao from the A genome was clustered here where most of the tetraploid cotton accessions existed; moreover, it could not be differentiated until a similarity coefficient of 0.915 was reached, indicating that it has close relationship with tetraploid cottons especially *G. hirsutum*, which is consistent with the fact that quite a few tetraploid cotton accessions in our research including *G. hirsutum* L. Zhongyuan9112 and *G. hirsutum* L. Yahuang9103 were bred by crossing *G. arboreum* with other cotton accessions. *G. hirsutum* L. Zhong07, *G. hirsutum* L. cv. CCRI12, and *G. hirsutum* L. cv. CCRI35 were clustered in a group with a similarity coefficient of 0.83. Pedigree analysis showed that all three tetraploid cotton accessions have Uganda cotton ancestors and that *G. hirsutum* L. cv. CCRI12 is even one of the parents of *G. hirsutum* L. cv. CCRI35.

There are five tetraploid species of *Gossypium* designated (AD)1 through (AD)5 for their genome constitutions (Wendel and Cronn, 2003) that evolved 1-2 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 research, the high transferability of the EST-SSRs among the 38 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 et al., 1992). As pointed out by Wendel et al. (1994), the A and D genomes of the South American tetraploid *G. mustelinum* are genetically most similar to the ancestral type that differentiated into the five present-day widely dispersed tetraploids. In our research, *G. mustelinum* was clustered with most of the accessions of the diploid A genome and D genome and further supported this conclusion.

There were totally 19 *G. hirsutum* accessions studied in this research, which were separated into two groups at a similarity coefficient of 0.57; 6 of them were clustered with most of the diploid species in the first group, whereas the other 13 were clustered in the second group with narrow genetic diversity except for *G. hirsutum* Jiangsudatao being differentiated from others at a similarity coefficient of 0.68. The narrow genetic diversity between the *G. hirsutum* is a result of only a few underlying parents being used in modern cotton breeding.

## Validation of cross-species transferability of EST-SSR markers

There have been a few reports of studies in cotton (Guo et al., 2006; Yu et al., 2008) and many reports in other plants (Eujayl et al., 2004; Saha et al., 2004; Feng et al., 2009; Choudhary et al., 2009) examining cross-species transferability using EST-SSR markers.

In our research, the high cross-species transferability of EST-SSRs from the salt-tolerant ESTs may imply that there are some common genes related to salt tolerance in different cotton genomes. Sequence alignment analysis of amplified fragments from NTU060 and NTU121 loci revealed that the primer binding regions were highly conserved except for some minor indels or base substitutions, and that microsatellite repeats were present at all loci within all of the 10 accessions (Figure 2). Consequently, DNA sequencing showed that the high transferability can be attributed to a higher level conservation in the flanking regions among these *Gossypium* species. As indicated by Pang et al. (2006), different cotton species had significant differences in salt tolerance, and *G. arboreum* contributed to salt tolerance more than other cotton species. Different salt-tolerant phenotypes may be the manifestation of complex mutational events involving indels or base substitutions in SSR regions. It is interesting that some point mutations were found between salt-tolerant *G. hirsutum* L. cv. CCRI35 and salt-sensitive *G. hirsutum* L. cv. CCRI12 (Figure 2). For amplification products of the primer NTU121, a transition of A to G was observed in the flanking sequences. For amplification products of the primer NTU060, besides the transition of A to G, a transition of C to T and a deletion of G were observed. It is reasonable to suppose that these variations in EST-SSR regions may have regulatory functions during salt tolerance or may even be critical genetic factors in the salt tolerance process.

As shown in Figure 2, the flanking regions of SSR loci changed whereas the number of repeats remained unchanged between the different species. Previous research showed that the variations in the sequences are mainly due to the variation in the number of repeat motifs in the SSR region combined with insertions and base substitutions (Gutierrez et al., 2005; Feng et al., 2009; Choudhary et al., 2009). In our study, point mutations, insertions, and deletions were found only in flanking regions, suggesting the presence of evolutionary constraints within transcribed regions that limit the mutational events and increase sequence similarity.

EST-SSR markers are very useful in characterizing species relationships and in the introgression and search for desirable alleles from wild germplasm pools in *Gossypium*, which may be employed due to their high cross-species transferability. The transferability between these different species presented here can increase the efficiency of transferring genetic information across species by the molecular tagging of important genes, especially salt-tolerance genes using the EST-SSR markers.

## ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (#31000729, #30900911), the State Key Laboratory of Cotton Biology Open Fund (#CB2013A12), the Natural Science Foundation of Jiangsu Province of China (#BK20131204, #BK20130429), the Qing Lan Project of Colleges and Universities in Jiangsu Province, the Project of Foreign Experts Invitation for Cultivation of Distinguished Young Teachers in Colleges and Universities of Jiangsu Province, the National Practice Innovation Training Program Projects for College

Students (#201210304020), the Construction Plan for Science and Technology Platform of Nantong City (#CP12012002), and the Scientific Research Foundation for Introduced Scholars of Nantong University (#03080472).

### [Supplementary material](#)

## REFERENCES

- Arunita R, Rakshit S, Santhy V, Gotmare VP, et al. (2010). Evaluation of SSR markers for the assessment of genetic diversity and fingerprinting of *Gossypium hirsutum* accessions. *J. Plant Biochem. Biot.* 19: 153-160.
- Ashraf M (2002). Salt tolerance of cotton: some new advances. *Crit. Rev. Plant Sci.* 21: 1-30.
- Blair MW, Giraldo MC, Buendia HF, Tovar E, et al. (2006). Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 113: 100-109.
- Botstein D, White RL, Skolnick M and Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32: 314-331.
- Cheng L, Wang Y, Meng L, Hu X, et al. (2012). Identification of salt-tolerant QTLs with strong genetic background effect using two sets of reciprocal introgression lines in rice. *Genome* 55: 45-55.
- Choudhary S, Sethy NK, Shokeen B and Bhatia S (2009). Development of chickpea EST-SSR markers and analysis of allelic variation across related species. *Theor. Appl. Genet.* 118: 591-608.
- Cuadrado A and Schwarzacher T (1998). The chromosomal organization of simple sequence repeats in wheat and rye genomes. *Chromosoma* 107: 587-594.
- Dai JL, Dong HZ and Duan LS (2010). Technology and mechanism in control of salt injury in cotton. *Cotton Sci.* 22: 486-494.
- Eujayl I, Sledge MK, Wang L, May GD, et al. (2004). *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. *Theor. Appl. Genet.* 108: 414-422.
- Feng SP, Li WG, Huang HS, Wang JY, et al. (2009). Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). *Mol. Breed.* 23: 85-97.
- Flowers TJ (1999). Salinisation and horticultural production. *Sci. Hortic.* 78: 1-4.
- 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). Chromosome translocations in interspecific hybrids of the genus *Gossypium*. *Evolution* 7: 234-244.
- Guo WZ, Zhang TZ, Sheng XL, John Y, 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.
- Guo W, Wang W, Zhou B and Zhang T (2006). Cross-species transferability of *G. arboreum*-derived EST-SSRs in the diploid species of *Gossypium*. *Theor. Appl. Genet.* 112: 1573-1581.
- Guo W, Cai C, Wang C, Han Z, et al. (2007). A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in *Gossypium*. *Genetics* 176: 527-541.
- Gutierrez MV, Vaz Patto MC, Huguet T, Cubero JI, et al. (2005). Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. *Theor. Appl. Genet.* 110: 1210-1217.
- Islam MR, Salam MA, Hassan L and Collard BCY (2011). QTL mapping for salinity tolerance at seedling stage in rice. *Emir. J. Food Agric.* 23: 137-146.
- Lee GJ, Carter TE Jr, Villagarcia MR, Li Z, et al. (2004). A major QTL conditioning salt tolerance in S-100 soybean and descendent cultivars. *Theor. Appl. Genet.* 109: 1610-1619.
- Lin ZX, Zhang YX, Zhang XL and Guo XP (2009). A high-density integrative linkage map for *Gossypium hirsutum*. *Euphytica* 166: 35-45.
- Liu YL and Wang LJ (1998). Plant Response to Salt Stress and Salt Tolerance. In: *Plant Physiology and Molecular Biology* (Yu SW and Tang ZC, eds.). 2nd edn. Science Press, Beijing, 752-769.
- Liu RZ, Wang BH, Guo WZ and Qin YS (2012). Quantitative trait loci mapping for yield and its components by using two immortalized populations of a heterotic hybrid in *Gossypium hirsutum* L. *Mol. Breed.* 29: 297-311.
- Nei M (1973). Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. U. S. A.* 70: 3321-3323.
- Pang CY, Du XM and Ma ZY (2006). Evaluation of the introgressed lines and screening for elite germplasm in *Gossypium*. *Chin. Sci. Bull.* 51: 304-312.
- Paterson AH, Brubaker C 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.

- Qureshi SN, Saha S, Kantety RV and Jenkins JN (2004). EST-SSR: a new class of genetic markers in cotton. *J. Cotton Sci.* 8: 112-123.
- 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.
- Shan L, Zhao SY, Chen F and Xia GM (2006). Screening and localization of SSR markers related to salt tolerance of somatic hybrid wheat Shanrong No. 3. *Sci. Agr. Sin.* 39: 225-230.
- Taliercio E, Allen RD, Essenberg M, Klueva N, et al. (2006). Analysis of ESTs from multiple *Gossypium hirsutum* tissues and identification of SSRs. *Genome* 49: 306-319.
- Thiel T, Michalek W, Varshney RK and Graner A (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 106: 411-422.
- Vendramin E, Dettori MT, Giovinazzi J, Micali S, et al. (2007). A set of EST-SSRs isolated from peach fruit transcriptome and their transportability across *Prunus* species. *Mol. Ecol. Notes* 7: 307-310.
- Wei WL, Qi XQ, Wang LH and Zhang YX (2011). Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC Genomics* 12: 451.
- Wendel JF and Cronn RC (2003). Polyploidy and the evolutionary history of cotton. *Adv. Agron.* 78: 139-186.
- Wendel JF, Rowley R and Stewart J (1994). Genetic diversity in and phylogenetic relationship of the Brazilian endemic cotton, *Gossypium mustelinum* (Malvaceae). *Plant Syst. Evol.* 192: 49-59.
- Yi G, Lee JM, Lee S, Choi D, et al. (2006). Exploitation of pepper EST-SSRs and an SSR-based linkage map. *Theor. Appl. Genet.* 114: 113-130.
- Yu Y, Wang ZW, Feng CH and Zhang YX (2008). Genetic evaluation of EST-SSRs derived from *Gossypium herbaceum*. *Acta Agron. Sin.* 34: 2085-2091.
- Zhang HB, Li Y, Wang B and Chee PW (2008). Recent advances in cotton genomics. *Int. J. Plant Genomics* 2008: 742304.
- Zhang ZS, Hu MC, Zhang J and Liu DJ (2009). Construction of a comprehensive PCR-based marker linkage map and QTL mapping for fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *Mol. Breed.* 24: 49-61.
- Zhang LN, Ye WW, Wang JJ and Fan BX (2010). Studies of salinity-tolerance with SSR markers on *G. hirsutum* L. *Cotton Sci.* 22: 175-180.