THE CROP JOURNAL 2 (2014) 315–328

Available online at www.sciencedirect.com

ScienceDirect

Exp2 polymorphisms associated with variation for fiber quality properties in cotton (*Gossypium* spp.)

Daohua He[a,*,1], Zhongping Lei[b,1], Hongyi Xing[a], Baoshan Tang[a], Junxing Zhao[a], Bixia Lu[a]

[a]College of Agronomy, Northwest A&F University, Yangling, 712100, China
[b]College of Life Sciences, Northwest A&F University, Yangling, 712100, China

ARTICLE INFO

Article history:
Received 30 December 2013
Received in revised form
17 April 2014
Accepted 9 June 2014
Available online 12 July 2014

Keywords:
Cotton
Expansin
Fiber quality
Linkage disequilibrium
Association mapping

ABSTRACT

Plant expansins are a group of extracellular proteins thought to affect the quality of cotton fibers. Previous expression profile analysis revealed that six *Expansin A* genes are present in cotton, of which two (*GhExp1* and *GhExp2*) produce transcripts that are specific to the developing cotton fiber. To identify the phenotypic function of *Exp2*, and to determine whether nucleotide variation among alleles of *Exp2* affects fiber quality, candidate gene association mapping was conducted. Gene-specific primers were designed to amplify the *Exp2* gene. By amplicon sequencing, the nucleotide diversity of *Exp2* was investigated across 92 accessions (including 7 *Gossypium arboreum*, 74 *Gossypium hirsutum*, and 11 *Gossypium barbadense* accessions) with different fiber qualities. Twenty-six SNPs and seven InDels including 14 from the coding region of *Exp2* were detected, forming twelve distinct haplotypes in the cotton collection. Among the 14 SNPs in the coding region, five were missense mutations and nine were synonymous nucleotide changes. The average SNP/InDel per nucleotide ratio was 2.61% (one SNP per 39 bp), with 1.81 and 3.87% occurring in coding and non-coding regions, respectively. Nucleotide and haplotype diversity across the entire *Exp2* region was 0.00603 (π) and 0.844, respectively, and diversity in non-coding regions was higher than that in coding regions. For linkage disequilibrium (LD), the mean r^2 value for all polymorphism loci pairs was 0.48, and LD did not decay over 748 bp. Based on 132 simple sequence repeat (SSR) loci evenly covering 26 chromosomes, the population structure was estimated, and the accessions were divided into seven groups that agreed well with their genomic origin and evolutionary history. A general linear model was used to calculate the *Exp2*-wide diversity–trait associations of 5 fiber quality traits, considering population structure (Q). Four SNPs in *Exp2* were associated with at least one of the fiber quality traits, but not with fiber elongation. The highest positive effect on UHML and STR was observed for haplotype Hap_6 of *Exp2*. There was a significant association of *Exp2* with fiber quality traits. There were many haplotypes in the *Exp2* region, of which the most favorable was Hap_6. The association between nucleotide diversity and these fiber traits

---

* *Corresponding author*.
E-mail address: daohuahe@nwafu.edu.cn (D. He).

[1] These authors contributed equally to the article.
Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.cj.2014.06.002

sheds light on the gene's potential contribution to the improvement of fiber quality, and should be useful to facilitate MAS programs in cotton.

1. Introduction

Cotton, which provides the most popular natural textile fiber, is one of the most important crops in the world. The genus *Gossypium* comprises about 45 diploid and 5 allotetraploid species. Four species are cultivated; *Gossypium hirsutum* and *Gossypium barbadense* account for 90% and 5% of the world cotton production, respectively, and *Gossypium arboreum* and *Gossypium herbaceum* are grown in a few areas. Fiber length and fiber strength are the primary quality properties that influence textile processing [1]. After fiber yield, improving fiber quality is a goal of breeders. To develop cultivars with further improved fiber quality, it is critical to characterize and dissect the molecular genetic bases of fiber quality.

Hitherto, advances in molecular genetics have increased genetic knowledge in fiber quality, such as by QTL mapping and gene expression profile analysis. Unfortunately, low resolution, lack of knowledge of phenotypic functions of candidate genes in natural populations, and other factors have prevented these advances from facilitating genetic design and selection for breeding.

Association mapping (AM) can be used to relate natural variation in candidate genes to agronomic phenotypes. AM provides a high-resolution alternative for the characterization of candidate genes and has the potential to allow exploring and evaluating a wide range of alleles [2]. Recently, AM has been successfully applied to plant populations [3–5]. In an attempt to validate the function of the *Dwarf8* locus, a large AM population of maize inbred lines was genotyped for *Dwarf8* polymorphism and phenotyped for flowering time, and an association of a *Dwarf8* polymorphism with flowering time was detected [6,7]. Later studies associated the candidate gene *su1* with sweetness [8]; *bt2*, *sh1*, and *sh2* with kernel composition; and *ae1* and *sh2* with starch pasting properties [9]. *DREB1A* showed associations with vegetation index, heading date, biomass, and spikelet number. Both *ERA1-A* and *ERA1-B* were associated with harvest index, flag leaf width, and leaf senescence. Gene *1-FEH-A* was associated with grain yield, and *1-FEH-B* was associated with thousand kernel weight and test weight [10]. In sunflower, *HaCOI1-1* and *HaCOI1-2* were found to be strongly associated with *Sclerotinia* stalk rot resistance [11]. In waxy rice, Xu et al. [12] associated *starch synthase IIa* (*SSIIa* or *SSII-3*) and *SSI* with starch properties. As these examples illustrate, AM is useful for dissecting candidate genes underlying complex traits. In cotton, some AM studies have been reported [5,13–16], but these were all genome wide association studies (GWAS) rather than candidate gene association studies.

Expansin refers to a family of closely related non-enzymatic proteins found in the plant cell wall, with important roles in plant cell growth, emergence of root hairs, meristem function, and other developmental processes in which cell wall loosening occurs. The elongation of cotton fiber is associated with the expression of many genes, among which Expansin is one of the most highly expressed [17–19]. That Expansin may control fiber development is of interest in strategies aimed at improving fiber quality, because final fiber length and strength largely determine the quality of commercial cotton thread. Given that Expansins play a pivotal role in cell wall extension, they are attractive targets for strategies designed to alter cell shape and size, and this consideration led us to characterize some of the genes that encode Expansins in *Gossypium*.

Six cDNAs encoding α-expansins were identified in a previous study of cotton fiber development [18]. RT-PCR expression analysis showed that the mRNA from *GhExp2* was specific to cotton fibers, where it was the second most abundant transcript (at a low level) during the elongation phase of fiber development [18]. Intron and exon sizes of *GhExp2* were all different from those of the other five genes. In GhExp2, a Cys → Arg substitution at the first Cys [a residue conserved in most α-expansins previously described [20]] was found, and the Phe [which commonly is contained in a His-Phe-Asp (HFD) domain] residue had been replaced by Lys. But, the amino acid sequence derived from *GhExp2* was most closely related (with 97% sequence identity) to that from *GhExp1*, which may play an important role in cell wall extension during fiber development [18]. It is still unclear whether the nucleotide diversity of *GhExp2* is associated with phenotypic variation.

After sequence alignment of six genes (*GhExp1*–*GhExp6*) and AY189969 (*expansin* mRNA), gene-specific primers were designed to amplify only *Exp2*. The objectives of this study were to investigate the nucleotide and haplotype diversity and the extent and pattern of linkage disequilibrium (LD) in the *Exp2* gene, and then to validate the association between *Exp2* and fiber quality by AM, and identify the most favorable allele of *Exp2*, with the aim of providing knowledge for future fiber quality breeding efforts in cotton.

2. Materials and methods

2.1. Plant material

A collection of 92 cultivated cotton accessions (presented in Table 1) consisting of 7 *G. arboreum*, 74 *G. hirsutum*, and 11 *G. barbadense* accessions collected worldwide was used. This population of accessions covered multiple ecological regions and periods of cultivar development, revealing a wide range of phenotypic variation among fiber quality traits. Seeds of all accessions were obtained from the germplasm storage of Chinese National Center for Cotton Improvement (Institute of Cotton Research of CAAS, Anyang, China).

2.2. Field trials and trait evaluation

The germplasm collection was sown in two growing seasons (2008 and 2009) at Sanyuan (34° 36′ N; 108° 56′ E; elevation 416.25 m.a.s.l.) Experimental Station of Northwest A&F