Genome-wide association analysis of fiber fineness and yield in ramie (*Boehmeria nivea*) using SLAF-seq

Kunyong Huang
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Yaliang Shi
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Yicheng Zhong
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Aiguo Zhu
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Zhimin Sun
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Juan Niu
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
Jing Chen
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops
M Luan (luanmingbao@caas.cn)
ibfc
Jianhua Chen
Chinese Academy of Agricultural Sciences Institute of Bast Fiber Crops

Research article

Keywords: Ramie, Genome-wide association study, Fiber fineness, Fiber yield, Significant SNPs

Posted Date: March 4th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-15219/v1

Version of Record: A version of this preprint was published at Euphytica on January 20th, 2021. See the published version at https://doi.org/10.1007/s10681-020-02757-w.

Abstract

Background Ramie is a major natural fiber crop cultivated in East Asia. The improvement of its fiber yield and fineness are important breeding goals. Fiber yield is a complex quantitative trait comprising ramet number, stem diameter, plant height, skin thickness, and fiber percentage. The fiber fineness is a crucial trait for ramie fiber quality. However, there are few association studies for fiber yield traits and fiber fineness on ramie, and lack high-density SNP maps in natural ramie population.

Results Here, a panel of 112 core ramie germplasms were genotyped by 215,376 consistent single-nucleotide polymorphisms (SNPs) from specific-locus amplified-fragment sequencing (SLAF-seq), and used for genome-wide association study of fiber fineness and yield. Subsequently, the genetic diversity, linkage disequilibrium (LD), and population structure was conducted based on 215,376 SNPs. Population cluster analysis disclosed five subpopulations. Neighbor-joining (NJ) analysis revealed three major clusters. No obvious relationships were identified between them and their geographic origins. The genome-wide linkage disequilibrium (LD) decayed to r 2 = 0.1 was ~11.75 kb in physical distance. One, seven, one, seven, and twenty-seven significant SNP marker associations were detected for fiber fineness (third season), stem diameter (third season), stem skin thickness (third season), fiber percentage (second season), and fiber percentage (third season), respectively. Two promising candidate genes, whole_GLEAN_10029622 and whole_GLEAN_10029638 resided in the significant trait-SNP association for fiber fineness (third season), which annotated as a cotton fiber-expressed protein and an Arabidopsis thaliana homeobox protein ATH1, respectively and validated by qPCR.

Conclusions These results provide information for understanding the population structure, LD decay and genetic variation in ramie natural population. The identified loci or genes for fiber fineness and yield may provide the basis for future research on fiber fineness and yield and marker-assisted selection breeding for ramie.

Background

Ramie (Boehmeria nivea L. Gaudich) is a perennial fiber crop and native to eastern Asia [1]. About 87,620 ha of ramie were cultivated in China in 2016 (http://data.stats.gov.cn). Ramie is the second largest fiber crop in China [2]. Fiber yield and quality limit ramie cultivation in China [3]. While, fiber yield is a complex quantitative trait comprising ramet number, stem diameter, plant height, skin thickness, and fiber percentage [2]. Qualitative fiber characteristics include length, strength, and fineness. The length and strength of both cotton and ramie fiber are suitable for textile manufacture. Unlike cotton, however, the fineness of ramie fiber is too lower for fabric production. Thus, it’s a crucial trait for fiber quality. Ramie fiber traits may be quantitative and regulated by numerous genes with small phenotypic effects [4, 5]. Thus, understanding the genetic basis of these traits were critical for molecular designing breeding of the traits.

Owing to the next-generation sequencing (NGS), large number of molecular markers detected and used for marker-assisted selection. Recently, 76 SSR markers were developed from 320 expressed sequence tags (ESTs) of ramie from the National Center for Biotechnology Information (NCBI), of these, twenty-seven SSR loci show polymorphisms among 62 ramie individuals [6]. Liu et al. [7] identified 1,827 EST-SSR markers. Chen et al. [3] identified 4,230 SSR loci and developed 2,431 SSR markers from the ramie transcriptome at different developmental stages. Using SSR markers, Liu et al. [8] and Luan et al. [2] identified several agronomic trait-related quantitative trait loci (QTLs) in ramie. Nevertheless, it is difficult to use them for cloning or breeding programs due to these QTLs with large intervals. Compared with SSR markers, SNPs are more abundant and stable in plant genomes and their frequency is 1 in 100–300 bp [9]. With the application of the NGS technologies, it is possible to generated thousands of genome-wide SNPs markers for high-density genetic maps [10–12]. Several sequencing-based technologies were developed for SNP discovery and high-throughput genotyping, such as two-enzyme genotyping-by-sequencing (GBS) [13], restriction site-associated sequencing (RADseq) [14], and double-digest RADseq [15]. Recently, another high-throughput DNA sequencing technology, specific-locus amplified-fragment sequencing (SLAF-seq) was developed Sun et al. [16] SLAF-seq had successfully create a high-density SNP maps for soybean (Glycine max L.) [17], cucumber (Cucumis sativus L.) [18], sesame (Sesamum indicum L.) [19] and upland cotton (Gossypium hirsutum L.) [20]. However, there are few reports of high-density SNP maps for ramie [21].

Using the genotyping-by-sequencing technique, Liu et al. [21] and Chen et al. [22] discovered numerous SNPs and detected QTLs associated with ramie fiber yield. However, the QTLs of Liu et al. [21] were based on the F2 agamous line population of the varieties Qingdaye and Zhongzhu 1. Chen et al. [22] focused only on the ramet number and made no reference to the ramie genome. There are still lack studies assess the genetic variation and GWAS for fiber yield and fineness in natural ramie population. Therefore, research is needed to assess the genetic variation and association analysis of fiber yield and fineness on natural ramie accessions using the Boehmeria nivea.

Genome-wide association studies (GWAS) are an efficient strategies for detect genomic regions controlling complex traits within unstructured crop germplasms [23]. In the present study, we used SLAF-seq technology perform a GWAS of 112 ramie accessions (107 core and 5 cultivar accessions). The aims of this study were to (1) develop high-density SNP map and genetic variants, (2) assess the genetic diversity of the B. nivea germplasm pool, (3) investigate the population structure and estimate the linkage disequilibrium (LD) across the accessions, and (4) identify single-nucleotide polymorphisms (SNPs) and candidate genes associated with fiber fineness and yield.

Results

Specific-locus, amplified-fragment sequencing and SNP Calling

The restriction enzymes RsaI and HaeIII were selected based on predictions from in silico digestion. There were 205,334 predicted SLAF tags 264–394 bp in length. The digestion efficiency was 93.30%. In this

experiment, 465.13 Gbp of sequences were generated from the 112 ramie accessions. There were 502,578 SLAF tags identified throughout the genome. The average depth per sample was 12.34×. There were 369,997 SLAF tags with polymorphisms. All SLAF tags aligned to the reference genome using BWA software [24]. The PLINK software [25] performed quality control on the data. A total of 2,955,337 SNP loci were identified by GATK and SAMtools. More than 85.9% of the SNP variant effects were in the intergenic, intron, downstream, and upstream areas of the coding region whereas only 13.58% were in the exons (Table 2). After filtration for integrity (> 80%) and low minor allele frequency (> 5%), 215,376 highly consistent SNPs were identified and used in the subsequent analyses.

Population genetics analysis

Genetic relationships among the 112 samples were calculated using the 215,376 highly consistent SNPs identified in the present study. Among these 12,432 pairwise combinations, 11,644 (93.662%) with genetic relationship coefficients < 0.05 (Additional files 1: Figure S1). Only four pairwise combinations (0.032%) (two pairs accessions: cd and br, ce and dr), which with higher kinship value between 0.6 to 0.65, suggesting the two pairs accessions may have closely relationship. Thus, most of accessions were very weak or no relationships in this study.

The population structure analysis indicated that the 112 association panel derived from five ancestors based on cross-validation (CV) errors (Figure 1). Of these 112 accessions, 57 were group one, 11 were group two, 15 were group three, 18 were group four, and the remaining 11 were group five (Figure 1). Group one accessions came from Hunan (15) and Jiangxi (13). The remaining 29 originated from Hubei (6), Guangxi (5), Guizhou (5), Sichuan (5), Yunnan (2), Hainan (2), Zhejiang (1), Chongqing (1), India (1), and Shanxi (1). Group two accessions were collected from Sichuan (2), Guangxi (2), Shanxi (2), Indonesia (2), Zhejiang (1), Jiangxi (1), and Hunan (1). Group three accessions were derived from Guizhou (4), Chongqing (3), Hunan (2), Jiangxi (2), Sichuan (1), Hubei (1), Shanxi (1), and Brazil (1). Group four accessions were gathered from Hunan (4), Guizhou (3), Sichuan (3), Hubei (2), Jiangxi (2), Yunnan (1), Hainan (1), Guangxi (1), and Cuba (1). Group five accessions were acquired from Hunan (3), Jiangxi (2), Sichuan (2), Guangxi (1), Shanxi (1), Guizhou (1), and Indonesia (1).

Neighbor-joining (NJ) clustering show this set of germplasm formed three major groups (Figure 2). Based on the Admixture cross-validation (CV) figure (Figure 1), the second choice consisted of three subpopulations (K) when the cross-validation error was the second lowest. Another selection comprised three subgroups for the 112 core ramie germplasms. Principal component analysis (PCA) showed genotype dispersal among the various components, and the first three PCs collectively explained only about 11.13% of the total variance, suggesting that the population structure of the 112 accessions was very low (Additional files 1: Figure S2). No significant correlation was detected between genetic relationships and geographic origins.

Linkage disequilibrium across the whole ramie genome

LD decayed was estimated between SNP pairs along each scaffolds using the 215,376 highly consistent SNPs. The genome-wide LD decayed to r^2 = 0.1 was ~11.75 kb in physical distance (Figure 3), which is lower than that reported for most other crops. There are major difference among the various scaffolds. LD ranged from 1.09 kb (Scaffold11 (PHNS01006314.1)) to 20.95 kb (Scaffold2) (Additional files 2: Table S1).

GWAS of phenotypic traits and quantitative real-Time PCR analysis of candidate genes

The GWAS models tested included GLM-Q, compressed MLM-Q+K, and EMMAX. According to the quantile-quantile (Q–Q) plot results, the GLM-Q model was strongly skewed towards significance for most traits (Additional files 1:Figure S3). Therefore, the Q matrix did not adequately account for the population structure or cryptic relatedness. The MLM-Q+K used both the Q and kinship matrices overcorrected these confounders (Additional files 1:Figure S4). The EMMAX model use the identity-by-state (IBS) relationship matrix and effectively reduced the false positive rate. It had sufficient statistical power to identify significant marker-trait associations (Additional files 1:Figure S5). Therefore, the EMMAX-Q model was applied for GWAS in this study.

Variations in all 13 traits were detected among the genotypes (Table 1). We tested the associations between the 215,376 SNPs and each trait using the EMMAX model. Based on the chosen $P < 4.64 \times 10^{-7}$, we identified 44 trait-SNP associations for five traits, namely, FF (third season), RSD (third season), SKT (third season), FP (second season), and FP (third season). These were designated the top QTL candidates (Additional files 3: Table S2). The association pattern across the genome scaffolds visualized with Manhattan plots showed that association trait-SNPs were scattered throughout the genome (Figures 4; Additional files 1: Figure S6–S8). We arbitrarily searched 100 kb upstream and downstream of these trait-SNP association loci already known for candidate genes in the *Boehmeria nivea* cultivar Zhongzhu No. 1 genome.

We predicted the number of candidate genes associated with the five aforementioned traits. We found 30 candidate genes from the significant SNP regions (PHNS01005842.1-2705313) associated with FF of the third season (Additional files 4: Table S3). Whole_GLEAN_10029631 and whole_GLEAN_10029638 located at 66.7 kb and 60.4 kb from the trait-SNP, respectively, were promising candidate genes encoding a cotton fiber-expressed protein of unknown function and an Arabidopsis thaliana homeobox protein ATH, respectively.

Seven significant SNP markers were associated with RSD of the third season (Additional files 5: Table S4). The most significantly associated SNP marker was at PHNS01002615.1:385693 ($-\log_{10}P$ = 7.47), which corresponded to the transcripts of whole_GLEAN_10015232 annotated as AT-rich interactive domain-containing protein 2 whose function is unknown. A subsequent survey identified the following promising candidate genes: a zinc finger protein CONSTANS-LIKE 3 gene (whole_GLEAN_10015241) located 63.3 kb from the significant SNP (PHNS01002615.1:385693), a cotton fiber-expressed protein of unknown function (whole_GLEAN_10025327) located 9.2 kb from the significant SNP

(PHNS01003751.1:2205139), an *Arabidopsis thaliana* trihelix transcription factor ASIL2 (whole_GLEAN_10026480) located 97 kb from the significant SNP (PHNS01007343.1:2569037), and an *Arabidopsis thaliana* E3 ubiquitin-protein ligase RHB1A (whole_GLEAN_10015233) located 5.2 kb from the significant SNP (PHNS01002615.1:385693).

One SNP marker (PHNS01006314.1:1911874) was associated with SKT of the third season (Additional files 6: Table S5). The transcript of whole_GLEAN_10021907 was annotated as *Arabidopsis thaliana* G-box-binding factor 1 and was located 31.5 kb from the marker. The transcript of whole_GLEAN_10021909 was annotated as *Arabidopsis thaliana* homeobox-leucine zipper protein ATHB-40 and was located 13 kb from the marker. The transcript of whole_GLEAN_10021913 was annotated as *Arabidopsis thaliana* BES1/BZR1 homolog protein 2 and was located 22.4 kb from the marker. All of these were promising candidate genes.

For FP of the second season, seven SNP markers were detected above the threshold (P = 4.64 × 10-7) in the scaffolds PHNS01004650.1, PHNS01011408.1, PHNS01001065.1, PHNS01005672.1, PHNS01004146.1, PHNS01003683.1, and PHNS01007343.1 (Additional files 7: Table S6). One *Arabidopsis thaliana* homeobox-leucine zipper protein HDG5 homolog (whole_GLEAN_10016513) and one *Arabidopsis thaliana* transcription factor KAN2 homolog (whole_GLEAN_10016511) were located 22.1 kb and 81.2 kb, respectively, from the trait-SNP PHNS01003683.1:866079. One *Arabidopsis thaliana* WAT1-related protein At3g18200 homolog (Whole_GLEAN_10007759) was located 71.9 kb from SNP PHNS01011408.1:355775 and three *Arabidopsis thaliana* transcription factor ORG2 homologs (whole_GLEAN_10011656, whole_GLEAN_10011657, and whole_GLEAN_10011658) were located at 74.8 kb, 83.8 kb, and 88.8 kb, respectively, from the significant SNP PHNS01004146.1:755398. These were promising candidate genes.

Twenty-seven significant GWAS loci among 17 scaffolds were identified for FP of the third season (Additional files 8: Table S7). There were five SNP markers in scaffold PHNS01000838.1 within whole_GLEAN_10010742 whose function is unknown. The whole_GLEAN_10006587 homolog of *Trema orientalis* GRAS transcription factor was located 30.4 kb from the marker PHNS01001992.1:258552. The whole_GLEAN_10025406 homolog of *Arabidopsis thaliana* Auxin-responsive protein IAA13 was located 0.7 kb from PHNS01003751.1:2871833. The whole_GLEAN_10029892 ortholog of *Trema orientalis* GRAS transcription factor was located 48.7 kb from PHNS01005842.1:4650339. Whole_GLEAN_10017958 was annotated as the NAC domain-containing protein (NAC29) and was located 97 kb from PHNS01010418.1:933907.

The candidate genes of FF (third season), RSD (third season), SKT (third season), and FP (second season) were validated by qRT-PCR. Whole_GLEAN_10029631 and whole_GLEAN_10029638 were comparatively upregulated in varieties with relatively higher fiber fineness (Figure 5). Whole_GLEAN_10021913 was comparatively downregulated in thicker SKT varieties. No other candidate genes were correlated with the RSD (third season), SKT (third season), or FP (second season) for six accessions by qRT-PCR (Figures S9–S11).

Discussion

Population structure and linkage disequilibrium

The phylogenetic tree constructed using neighbor-joining algorithm based on the genetic similarity coefficients among individuals to reflects genetic relationships. In contrast, the population structure analysis using Admixture software, which estimates individual ancestries by efficiently computing maximum likelihood in a parametric model. In the present study, three major clusters were defined in a dendrogram. The second lowest cross-validation error was obtained for K = 3. The 112 accessions may be partitioned into three subgroups. No obvious relationship was detected between the accessions and their geographic origins. Indicate these lines might have undergone introgression or gene flow during ramie breeding in China. This finding was consistent with those in previous reports on *Boehmeria nivea* [1, 26, 27].

In the present study, we used the reference genome of *Boehmeria nivea* zhongzhu1 based on SLAF-seq and identified 215,376 highly consistent SNPs which was ~2× the number of SNPs (108,888) reported by Chen et al. [22]. This abundance of SNPs will be useful for constructed a precise haplotype map and LD studies [28, 29].

It is crucial to the LD decay of the genome for identifying the number and density of molecular markers for GWAS and their analysis methods [30]. Ramie is predominantly a cross-pollinated or clonally propagated species. Thus, it should present with lower LD levels than those of self-pollinated crops. Despite several earlier association studies on ramie [2, 21, 22], none of them reported any LD decay values. Here, the genome-wide LD decayed to r^2 = 0.1 was 11.75 kb in physical distance. This value was substantially lower than those for most other crops such as rice (*Oryza sativa* L.) (167 kb) [31], soybean (420 kb) [32], and cultivated maize (*Zea mays* L.) (30 kb) [33], but higher than that of cultivated cassava (*Manihot esculenta* L.) (8 kb) [34], inbred maize (*Zea mays* L.) (1.5 kb) [35], and inbred potato (*Solanum tuberosum* L.) (1 kb) [36]. Therefore, the 112 ramie accessions of the present study should have abundant allelic diversity based on the rapid LD decay in most of the scaffolds. In this way, the search for candidate genes is facilitated through efficient narrowing of the putative QTL regions.

Genome-wide association study (GWAS)

The selection of best statistical models can accurately evaluate the associations between markers and phenotypes. As the availability of genotype data increases, comprehensive statistical models are required to distinguish true biological associations from the false positives arising from population structure and LD [37]. Numerous statistical models are available to calculate the significance of the associations between SNPs and traits [38]. Luan et al. [2] detected 16 stable molecular markers from an association analysis of fiber yield traits in ramie based on the MLM-Q+K model. However, this model considers the effects of Q for the population structure, analyzes K for the genetic relationships, and Bonferroni correction is required for multiple tests . When marker number is very large, this test is too strict [39]. In terms of the Q-Q for our results, MLM-Q+K overcorrected the confounders while GLM-Q was strongly

skewed towards significance for every trait. EMMAX used the identity-by-state (IBS) relationship matrix, effectively controlled the false positive rate, and provided suitable statistical power to identify significant marker-trait associations. Therefore, the identity-by-state (IBS) relationship matrix was appropriate for ramie GWAS in natural populations.

The discovery of QTL for agronomically and economically important crop traits was of vital importance for marker-assisted breeding. To our best knowledge, this is the first report on the QTL for fiber yield and fineness based on the reference genome of *Boehmeria nivea* [40]. We found one, seven, one, seven, and twenty-seven significant SNP marker associations for FF (third season), RSD (third season), RST (third season), FP (second season), and FP (third season), respectively. The identified trait-SNP associations serve as unique tools for molecular-assisted breeding.

Two candidate genes in the significant trait-SNP associations were identified for FF (third season). The whole_GLEAN_10029622, Pfam annotation encodes a cotton fiber-expressed protein of unknown function. The head of the protein includes the domain DUF4408 and the tail is the cotton fiber-expressed protein domain DUF761. The whole_GLEAN_10029638 is an ortholog of the *Arabidopsis thaliana* homeobox protein ATH1. In Arabidopsis, the TALE homeobox gene ATH1 positive regulation of the floral repressor FLC, controls floral competency [41]. The qPCR validation disclosed that these genes were upregulated in three fineness accessions compared to their expression levels in three fiber thickness accessions. Therefore, these two may be candidates for fiber fineness. The transcript of the whole_GLEAN_10021913 homolog of *Arabidopsis thaliana* BES1/BZR1 protein 2 was associated with SKT of the third season. It was a downstream target gene of the brassinosteroid (BR) signal. BRs regulate multiple biological processes including vasculature differentiation, cell elongation, senescence, photomorphogenesis, and stress response [42]. In Arabidopsis, *AtMYB30* is a direct target of BES1 and co-regulates brassinosteroid-induced gene expression with BES1 [43]. Chai et al. [44] reported that *PdMYB10* and *PdMYB128* in Arabidopsis increased stem fiber cell wall thickness and delayed flowering. Overexpression of PdMYB90, *PdMYB167*, *PdMYB92*, and *PdMYB125* in Arabidopsis decreased stem fiber and vessel cell wall thickness and promoted flowering. The qPCR validation showed that whole_GLEAN_10021913 expression was relatively lower in thick SKT varieties. Thus, it is a promising candidate gene associated with SKT of the third season. Another candidate gene, whole_GLEAN_10015241, was associated with RSD of the third season. It was annotated as a zinc finger protein CONSTANS-LIKE 3 gene. The Arabidopsis CONSTANS gene promotes flowering and its overexpression results in earlier flowering than the wild type [45]. Early flowering reduces the time for vegetative growth and results in comparatively smaller plants. In contrast, late flowering prolongs vegetative growth which culminates in relatively larger plants [46]. The candidate gene whole_GLEAN_10016511 was associated with PC of the second season. It is a homolog of the *Arabidopsis thaliana* transcription factor KAN2 gene. In Arabidopsis thaliana, the transcription factor KAN2 regulates lateral organ polarity and determines abaxial cell fate during lateral organ formation [47]. *KAN* genes participate in ovule development. The growth of the outer integument observed is retarded in certain *kan* mutant combinations [48]. The qPCR indicates that the relative expression levels of these two genes did not correlate with the RSD or PC phenotypes for six varieties (Additional files 1:Figure S9-S11).

This anomaly merits further investigation. In the present study, the GWAS results were limited duo to the small population size. Comprehensive phenotypic studies are currently in progress , and the future studies on association mapping in a wide range of samples may be performed.

Conclusions

This study provide information for understanding the population structure, LD decay and genetic variation in ramie natural population. The identified SNPs or genes for fiber fineness and yield may provide the basis for future research on fiber fineness and yield and marker-assisted selection breeding for ramie.

Materials And Methods

Plant materials and phenotyping

A set of 112 ramie accessions from the National Infrastructure for Ramie Germplasm Resources of China were selected for this study (Additional files 9: Table S8), including 106 accessions originally collected from 11 provinces of China (96 landrace accessions and 10 cultivated varieties), one landrace accession from Brazil, one landrace accession from India, one landrace accession from Cuba , and three landrace accessions from Indonesia. The germplasm was selected to represent variations in fiber fineness and yield. Briefly, the accessions were vegetatively propagated from the National Infrastructure for Ramie Germplasm Resources of China and planted in our scientific research field (Wangcheng Experimental Station of the Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha, China) with a two replicates, randomized block design. Each plot consisted of one accession and five clones planted 1 m apart in a single row (3 m long × 1 m wide). Standard cultural practices were performed according to Luan et al. [2] for three successive seasons in 2016 from March to June (environment 1), June to August (environment 2), and August to October (environment 3). All ramie plants were planted under the same field conditions and were under conventional agricultural management.

The three harvest periods represented different phases of phenological development, so they were analyzed separately. Thirteen plant traits were measured and the data were applied to the association mapping (Table 1). Measurements were made according to the methods of Luan et al. [2]. Briefly, 15 ramets per plot were randomly selected to measure plant height, stem diameter, skin thickness, and fiber percentage using tape measure, Vernier caliper, and screw micrometer, respectively. Fiber fineness was examined with optical fiber diameter analyzer systems (OFDA 100 and OFDA 2000; HORNIK FIBERTECH, Wald, Switzerland) according to Wang et al. [49]. Fiber percentage was calculated as follows:

$$\text{Fiber \%} = \text{FW (after bast decortication; dry bast fiber weight)} / \text{BW(fresh bast weight)} \times 100 \quad (1)$$

SLAF sequencing and data evaluation

Young healthy leaves were excised from plants growing in the Wangcheng Experimental Station in Changsha, China (Wangcheng) for DNA extraction with a DNeasy plant mini kit (Qiagen, Germany). DNA quality and quantity were assessed with a NanoDropTM 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

Ramie genomic DNA was analyzed by SLAF-seq [16] as described in Zhou et al. [50]. Briefly, genomic DNA was digested with restriction enzyme RsaI and HaeIII (New England Biolabs Inc., Ipswich, MA, USA), followed by fragment-end repair, duplex tag-labeled adapter ligation, PCR amplification, and size-selected for SLAF libraries. Finally, the SLAF libraries were sequencing using an Illumina HiseqTM 2500 (Illumina, San Diego, CA, USA) at Biomarker Technologies Corporation in Beijing, China. The average sequencing depth was 10.89×. After sequencing, the low-quality reads (quality score < 20) were filtered out using dual-index software [51]. All high-quality reads were mapped onto the *Boehmeria nivea* reference genome [40] using the Burrows-Wheeler alignment tool (BWA) [24] with default parameters.

SNP calling

The software GATK [52] and SAMtools [53] were used to discover SNPs, and SNPs detected by both methods were treated as reliable. After SNPs with minor allele frequency (MAF) of less than 5% and missing data of more than 20% were filtered out, a total of 215,376 consistent SNPs remained and were used for analysis.

Phylogenetic tree and principal component analysis

Based on the 215,376 consistent SNPs identified from the 112 ramie samples, a phylogenetic tree was constructed in MEGA 5 [54] using the neighbor-joining algorithm [55]. The genetic structure of the panel was assessment using principal components analysis (PCA) with GAPIT software [56].

Deducing population structure and estimating linkage disequilibrium (LD)

Population structure was investigated using Admixture software [57] based on those consistent SNPs. The number of subgroups (K) was varied from 1 to 10, which represented the assumed groups of simulated population in ancient times. The best value of K was determined by the cross-validation (CV). The PLINK 1.9 BETA [58] was used to obtain the required data files for ADMIXTURE. LD was estimated between SNP pairs located on the same scaffold based on the alignment of reference genome *Boehmeria nivea* [40]. The ramie chromosome sequence was not publicly available at that time. The 215,376 consistent SNPs were used to infer LD in TASSEL [59]. The r^2 were obtained for all SNP pairs using Plink [25].

Association mapping

Three Genome-wide association studies (GWAS) models were used to association analysis. The models GLM-Q and compressed MLM-Q+K were analysis in TASSEL [59], kinship (K) matrix was estimated using SPAGe Di v. 1.4b [60] and the best admixture results (Q) representing population membership served as

covariates in the model. The EMMAX model were analysis using identify-by-state (IBS) relationship matrix according to Kang et al. [23].

Based on the number of SNPs analyzed (n = 215,376), the significance threshold was estimated use the Bonferroni correction method, it was P = 4.64 × 10-7 (0.1/215,376). SAMtools [53] was used for the manual verification of regions of the aligned sequencing reads with significant association relative to the reference genome of *Boehmeria nivea* [40].

Identification of candidate genes and quantitative PCR (qRT-PCR) analysis

Functionally characterized genes harboring the detected SNPs or within a 100-kb interval were sought in the reference genome of *Boehmeria nivea*. Then qRT-PCR was used to determine differential expression of the genes putatively controlling fiber fineness, stem diameter, and skin thickness. Based on our phenotypic data of field experiment, we selected six accessions (1–26 (kk), 2–34 (ct), 2–35 (bo), 1–43 (aw), 2–41 (kr), and 2–7 (ba)) for RT-qPCR. Bast samples were collected from those six accessions at the vegetative stage (~30 cm height) and freezing in liquid nitrogen immediately. Total RNA was extracted with a TaKaRa MinBEST plant RNA extraction kit (TaKaRa Bio Inc., Shiga, Japan) and reverse-transcribed with a Thermo Scientific RevertAid first-strand cDNA synthesis kit (Thermo Scientific, Vilnius, Lithuania) according to the manufacturer's instructions. The qRT-PCR was performed in a Monad selected q225 real-time PCR system (Monad Biotech Co. Ltd., Shanghai, China) in a 10 μL reaction volume containing 0.5 μL cDNA sample, 5 μL MonAmpTM SYBR Green qPCR mix (Monad Biotech Co. Ltd., Shanghai, China), 0.5 μL each forward and reverse primers (10 μM each), and 3.5 μL nuclease-free water. The PCR parameters were 95 °C for 1 min followed by 40 cycles of 95 °C for 5 s and 55 °C for 30 s. Each sample was analyzed in triplicate. Relative expression were calculated as previously described [61]. All qRT-PCR primers (Additional files 10: Table S9) were designed by the Primer3 online tool (http://primer3.ut.ee/) and adjusted with Oligo v. 7.56 [62]. The internal control of choice was 18S based on preliminary experiments (data not shown). The qRT-PCR data were analyzed in Microsoft Excel and reported as means ± SD.

Additional File Information

Additional files 1: Figure S1. Kinship hist of the 112 accessions. Figure S2. PCA analysis in 112 accession of ramie. Figure S3 Quantile-quantile (Q–Q) plot of GLM-Q model. Figure S4. Quantile-quantile (Q–Q) plot of MLM-Q+K model. Figure S5. Quantile-quantile (Q–Q) plot of EMMAX model. Figure S6. Manhattan plots and Q-Q plots of SNP markers for fiber fineness of third season. Figure S7. Manhattan plots and Q-Q plots of SNP markers for fiber percentage of second season. Figure S8. Manhattan plots and Q-Q plots of SNP markers for fiber percentage of third season. Figure S9. Relative expression of candidate genes of stem diameter in third season. Figure S10. Relative expression of candidate genes of SKT in third season. Figure S11. Relative expression of candidate genes of PC of second season.

Additional files 2: Table S1. LD decay value among scaffolds.

Additional files 3: Table S2. Associated significant SNP-traits by EMMAX model

Additional files 4: Table S3. Function annotation of candidate genes from the significant SNP-trait of FF of third season.

Additional files 5: Table S4. Candidate genes of significant SNP-trait of RSD of third season.

Additional files 6: Table S5. Function annotation of candidate genes from the significant SNP-trait of SKT of third season.

Additional files 7: Table S6. The significant SNP markers association with FP of second season

Additional files 8: Table S7. The significant SNP markers association with FP of third season.

Additional files 9: Table S8. ramie germplasms using this study.

Additional files 10: Table S9. Primers for qRT-PCR.

Abbreviation

GWAS: Genome-wide association studies; SNPs: Single nucleotide polymorphisms; LD: Linkage disequilibrium; NGS: Next-generation sequencing; FF: fiber fineness; RH: plant height; RSD: stem diameter; SKT: ramie skin thickness; FP: fiber percentage; SD: standard deviation; CV: coefficient of variance.

Declarations

Funding

This study was financially supported by the National Natural Science Foundation of China (Grant Nos. 31671744 and 31471547), the Hunan Provincial Natural Science Foundation of China (Grant No. 2018JJ2465),and the National Modern Agro-industry Technology Research System (Grant No. CARS-19-E04).

Availability of data and materials

We have provided detailed information about materials and method in our manuscript, so we will not provide data and supporting materials in this section.

Authors' contributions

KH, YS, ML, and JC conceived and designed the experiments; KH, YS, YZ, AZ, ZS, JN, JC, ML, and JC performed the experiments; and KH, YS, ML, and JC analyzed the data and wrote the paper.

Competing interests

The authors declared no conflict of interest in the authorship or publication of this document.

Consent for publication

Not applicable

Ethics approval and consent to participate

Not applicable

Author details

Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences/Key Laboratory of Stem-Fiber Biomass and Engineering Microbiology, Ministry of Agriculture, Changsha 410205, China

References

1. Luan M, Zou Z, Zhu J, Wang X, Xu Y, Sun Z, Chen J: **Genetic diversity assessment using simple sequence repeats (SSR) and sequence-related amplified polymorphism (SRAP) markers in ramie**. *BIOTECHNOL BIOTEC EQ* 2015, **29**(4):624-630.
2. Luan M, Liu C, Wang X, Xu Y, Sun Z, Chen J: **SSR markers associated with fiber yield traits in ramie (Boehmeria nivea L. Gaudich)**. *IND CROP PROD* 2017, **107**:439-445.
3. Chen J, Yu R, Liu L, Wang B, Peng D: **Large-scale developing of simple sequence repeat markers and probing its correlation with ramie (Boehmeria nivea L.) fiber quality**. *MOL GENET GENOMICS* 2016, **291**(2):753-761.
4. Xiong H, Jiang J, Yu C, Guo Y: **Relation between yield-related traits and yield in ramie**. *Acta Agronomica Sinica* 1998, 24: 155–160.
5. Lu K, Wang T, Xu X: **Genome-wide association analysis of height of podding and thickness of pod canopy in Brassica napus**. *Acta Agronomica Sinica* 2016, 42: 344–352.
6. Chen J, Luan M, Song S, Zou Z, Wang X, Xu Y, Sun Z: **Isolation and characterization of EST-SSRs in the Ramie**. *AFRICAN JOURNAL OF MICROBIOLOGY RESEARCH* 2011, **5**(21):3504-3508.
7. Liu T, Zhu S, Fu L, Tang Q, Yu Y, Chen P, Luan M, Wang C, Tang S: **Development and Characterization of 1,827 Expressed Sequence Tag-Derived Simple Sequence Repeat Markers for Ramie (Boehmeria nivea L. Gaud)**. *PLOS ONE* 2013, **8**(e603464).
8. Liu T, Tang S, Zhu S, Tang Q: **QTL mapping for fiber yield-related traits by constructing the first genetic linkage map in ramie (Boehmeria nivea L. Gaud)**. *MOL BREEDING* 2014, **34**(3):883-892.
9. Edwards D, Batley J, Cogan N, Forster J, Chagne D: **Single nucleotide polymorphism discovery**. In Oraguzie N, Rikkerink E, Gardiner S, DeSilva H, editor, Association Mapping in Plants. *Springer* 2007, New York. p. 53-76.
10. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML: **Genome-wide genetic marker discovery and genotyping using next-generation sequencing**. *NAT REV GENET* 2011, **12**(7):499-510.

11. Qi Z, Huang L, Zhu R, Xin D, Liu C, Han X, Jiang H, Hong W, Hu G, Zheng H *et al*: **A high-density genetic map for soybean based on specific length amplified fragment sequencing**. *PLOS ONE* 2014, **9**(8):e104871.
12. Zhang Z, Wei T, Zhong Y, Li X, Huang J: **Construction of a high-density genetic map of Ziziphus jujube Mill. using genotyping by sequencing technology**. *Tree Genetics & Genomes* 2016, 12: 76.
13. Poland JA, Brown PJ, Sorrells ME, Jannink JL: **Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach**. *PLOS ONE* 2012, **7**(2):e32253.
14. Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA: **Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers**. *GENOME RES* 2007, **17**(2):240-248.
15. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE: **Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species**. *PLOS ONE* 2012, **7**(5):e37135.
16. Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, Jiang C, Guan N, Ma C, Zeng H *et al*: **SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing**. *PLOS ONE* 2013, **8**(3):e58700.
17. Li B, Tian L, Zhang J, Huang L, Han F, Yan S, Wang L, Zheng H, Sun J: **Construction of a high-density genetic map based on large-scale markers developed by specific length amplified fragment sequencing (SLAF-seq) and its application to QTL analysis for isoflavone content in Glycine max**. *BMC GENOMICS* 2014, **15**:1086.
18. Wei Q, Wang Y, Qin X, Zhang Y, Zhang Z, Wang J, Li J, Lou Q, Chen J: **An SNP-based saturated genetic map and QTL analysis of fruit-related traits in cucumber using specific-length amplified fragment (SLAF) sequencing**. *BMC GENOMICS* 2014, **15**:1158.
19. Zhang Y, Wang L, Xin H, Li D, Ma C, Ding X, Hong W, Zhang X: **Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing**. *BMC PLANT BIOL* 2013, **13**:141.
20. Zhang Z, Shang H, Shi Y, Huang L, Li J, Ge Q, Gong J, Liu A, Chen T, Wang D *et al*: **Construction of a high-density genetic map by specific locus amplified fragment sequencing (SLAF-seq) and its application to Quantitative Trait Loci (QTL) analysis for boll weight in upland cotton (Gossypium hirsutum.)**. *BMC PLANT BIOL* 2016, **16**:79.
21. Liu C, Zhu S, Tang S, Wang H, Zheng X, Chen X, Dai Q, Liu T: **QTL analysis of four main stem bark traits using a GBS-SNP-based high-density genetic map in ramie**. *Sci Rep* 2017, **7**(1):13458.
22. Chen K, Luan M, Xiong H, Chen P, Chen J, Gao G, Huang K, Zhu A, Yu C: **Genome-wide association study discovered favorable single nucleotide polymorphisms and candidate genes associated with ramet number in ramie (Boehmeria nivea L.)**. *BMC PLANT BIOL* 2018, **18**(1):345.
23. Kang Y, Sakiroglu M, Krom N, Stanton-Geddes J, Wang M, Lee YC, Young ND, Udvardi M: **Genome-wide association of drought-related and biomass traits with HapMap SNPs in Medicago truncatula**.

PLANT CELL ENVIRON 2015, **38**(10):1997-2011.

24. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform**. *BIOINFORMATICS* 2009, **25**(14):1754-1760.
25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ *et al*: **PLINK: a tool set for whole-genome association and population-based linkage analyses**. *AM J HUM GENET* 2007, **81**(3):559-575.
26. Luan MB, Zou ZZ, Zhu JJ, Wang XF, Xu Y, Ma QH, Sun ZM, Chen JH: **Development of a core collection for ramie by heuristic search based on SSR markers**. *Biotechnol Biotechnol Equip* 2014, **28**(5):798-804.
27. Zhu S, Liu T, Dai Q, Wu D, Zheng X, Tang S, Chen J: **Genetic structure and relationships of an associated population in ramie (Boehmeria nivea L. Gaud) evaluated by SSR markers**. *BIOTECHNOL BIOTEC EQ* 2017, **31**(1):36-44.
28. Buckler E, Gore M: **An Arabidopsis haplotype map takes root**. *NAT GENET* 2007, **39**(9):1056-1057.
29. Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J *et al*: **A first-generation haplotype map of maize**. *SCIENCE* 2009, **326**(5956):1115-1117.
30. Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES: **Association mapping: critical considerations shift from genotyping to experimental design**. *PLANT CELL* 2009, **21**(8):2194-2202.
31. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z *et al*: **Genome-wide association studies of 14 agronomic traits in rice landraces**. *NAT GENET* 2010, **42**(11):961-967.
32. Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, Yu Y, Shu L, Zhao Y, Ma Y *et al*: **Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean**. *NAT BIOTECHNOL* 2015, **33**(4):125-408.
33. Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia JM, Cartwright RA, et al: **Comparative population genomics of maize domestication and improvement**. *Nat. Genet.* 2012, 44: 808–811.
34. Zhang S, Chen X, Lu C, Ye J, Zou M, Lu K, Feng S, Pei J, Liu C, Zhou X *et al*: **Genome-Wide Association Studies of 11 Agronomic Traits in Cassava (Manihot esculenta Crantz)**. *FRONT PLANT SCI* 2018, **9**:503.
35. Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ET: **Structure of linkage disequilibrium and phenotypic associations in the maize genome**. *Proc Natl Acad Sci U S A* 2001, **98**(20):11479-11484.
36. Simko I: **One potato, two potato: haplotype association mapping in autotetraploids**. *TRENDS PLANT SCI* 2004, **9**(9):441-448.
37. Wang H, Qin F: **Genome-Wide Association Study Reveals Natural Variations Contributing to Drought Resistance in Crops**. *FRONT PLANT SCI* 2017, **8**:1110.
38. Ogura T, Busch W: **From phenotypes to causal sequences: using genome wide association studies to dissect the sequence basis for variation of plant development**. *CURR OPIN PLANT BIOL* 2015, **23**:98-

108.

39. Hou S, Zhu G, Li Y, Li W, Fu J, Niu E, Li L, Zhang D, Guo W: **Genome-Wide Association Studies Reveal Genetic Variation and Candidate Genes of Drought Stress Related Traits in Cotton (Gossypium hirsutum L.)**. *FRONT PLANT SCI* 2018, **9**(1276).
40. Luan MB, Jian JB, Chen P, Chen JH, Chen JH, Gao Q, Gao G, Zhou JH, Chen KM, Guang XM *et al*: **Draft genome sequence of ramie, Boehmeria nivea (L.) Gaudich**. *MOL ECOL RESOUR* 2018, **18**(3):639-645.
41. Proveniers M, Rutjens B, Brand M, Smeekens S: **The Arabidopsis TALE homeobox gene ATH1 controls floral competency through positive regulation of FLC**. *PLANT J* 2007, **52**(5):899-913.
42. Yu X, Li L, Zola J, Aluru M, Ye H, Foudree A, Guo H, Anderson S, Aluru S, Liu P *et al*: **A brassinosteroid transcriptional network revealed by genome-wide identification of BESI target genes in Arabidopsis thaliana**. *PLANT J* 2011, **65**(4):634-646.
43. Li L, Yu X, Thompson A, Guo M, Yoshida S, Asami T, Chory J, Yin Y: **Arabidopsis MYB30 is a direct target of BES1 and cooperates with BES1 to regulate brassinosteroid-induced gene expression**. *PLANT J* 2009, **58**(2):275-286.
44. Chai G, Wang Z, Tang X, Yu L, Qi G, Wang D, Yan X, Kong Y, Zhou G: **R2R3-MYB gene pairs in Populus: evolution and contribution to secondary wall formation and flowering time**. *J EXP BOT* 2014, **65**(15):4255-4269.
45. Putterill J, Robson F, Lee K, Simon R, Coupland G: **The CONSTANS gene of Arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors**. *CELL* 1995, **80**(6):847-857.
46. Lenser T, Theissen G: **Molecular mechanisms involved in convergent crop domestication**. *TRENDS PLANT SCI* 2013, **18**(12):704-714.
47. McAbee JM, Hill TA, Skinner DJ, Izhaki A, Hauser BA, Meister RJ, Venugopala RG, Meyerowitz EM, Bowman JL, Gasser CS: **ABERRANT TESTA SHAPE encodes a KANADI family member, linking polarity determination to separation and growth of Arabidopsis ovule integuments**. *PLANT J* 2006, **46**(3):522-531.
48. Eshed Y, Baum SF, Perea JV, Bowman JL: **Establishment of polarity in lateral organs of plants**. *CURR BIOL* 2001, **11**(16):1251-1260.
49. Wang HM, Wang X: **Evaluation of the fineness of degummed bast fibers**. *FIBER POLYM* 2004, **5**(3):171-176.
50. Zhou Q, Zhou C, Zheng W, Mason AS, Fan S, Wu C, Fu D, Huang Y: **Genome-Wide SNP Markers Based on SLAF-Seq Uncover Breeding Traces in Rapeseed (Brassica napus L.)**. *FRONT PLANT SCI* 2017, **8**:648.
51. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD: **Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform**. *APPL ENVIRON MICROB* 2013, **79**(17):5112-5120.

52. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M *et al*: **The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data**. *GENOME RES* 2010, **20**(9):1297-1303.
53. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The Sequence Alignment/Map format and SAMtools**. *BIOINFORMATICS* 2009, **25**(16):2078-2079.
54. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees**. *MOL BIOL EVOL* 1987, **4**(4):406-425.
55. Eickmeyer K, Huggins P, Pachter L, Yoshida R: **On the optimality of the neighbor-joining algorithm**. *ALGORITHM MOL BIOL* 2008, **3**(5).
56. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z: **GAPIT: genome association and prediction integrated tool**. *BIOINFORMATICS* 2012, **28**(18):2397-2399.
57. Alexander DH, Novembre J, Lange K: **Fast model-based estimation of ancestry in unrelated individuals**. *GENOME RES* 2009, **19**(9):1655-1664.
58. Shaun S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, et al.: **PLINK: a tool set for whole-genome association and population-based linkage analyses**. *The American journal of human genetics* 2007, 81: 559–575.
59. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES: **TASSEL: software for association mapping of complex traits in diverse samples**. *BIOINFORMATICS* 2007, **23**(19):2633-2635.
60. Hardy OJ, Vekemans X: **SPAGeDi: a versatile computer program to analyses spatial genetic structure at the individual or population levels**. *Molecular ecology notes* 2002, 2: 618-620.
61. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method**. *METHODS* 2001, **25**(4):402-408.
62. Rychlik W: **OLIGO 7 primer analysis software**. *Methods Mol Biol* 2007, **402**:35-60.

Tables

Table 1. Summary statistics of phenotypes used in GWAS.

Trait	Unit	Min	Max	Mean	SD	CV
FF (first season)	m g^{-1}	1,062.00	2,869.00	1,616.00	324.16	0.20
FF (second season)	m g^{-1}	916.00	2,412.50	1,429.00	292.50	0.20
FF (third season)	m g^{-1}	965.50	2,815.50	1,796.00	390.13	0.22
RH (first season)	cm	93.83	213.83	152.30	19.53	0.13
RH (second season)	cm	61.82	145.55	109.80	14.92	0.14
RH (third season)	cm	46.00	129.50	96.00	16.75	0.17
RSD (first season)	mm	6.68	11.78	9.20	1.25	0.14
RSD (second season)	mm	5.03	10.59	7.83	1.08	0.14
RSD (third season)	mm	3.54	21.57	7.02	2.35	0.33
SKT (second season)	mm	0.48	1.19	0.77	0.13	0.18
SKT (third season)	mm	0.30	0.73	0.49	0.09	0.18
FP (second season)	g g^{-1}	0.03	0.25	0.12	0.03	0.29
FP (third season)	g g^{-1}	0.02	0.25	0.11	0.04	0.35

* *FF* fiber fineness, *RH* plant height, *RSD* stem diameter, *SKT* ramie skin thickness, *FP* fiber percentage, *SD* standard deviation, *CV* coefficient of variance.

Table 2. Number of SNP effects by genomic region.

Type	Count	Percent (%)
Intergenic	773,264	26.17
Intron	438,233	14.83
Upstream	686,213	23.22
Downstream	640,610	21.68
Splice_site_acceptor	741	0.025
Splice_site_donor	661	0.022
Splice_site_region	12,155	0.41
Exon (start lost)	440	0.015
Exon (synonymous_coding)	195,519	6.62
Exon (nonsynonymous coding)	195,860	6.63
Exon (synonymous_stop)	309	0.010
Exon (stop_gained)	5,892	0.20
Exon (stop_lost)	402	0.014
Other	5,038	0.17
Total	2,955,337	100

Figures

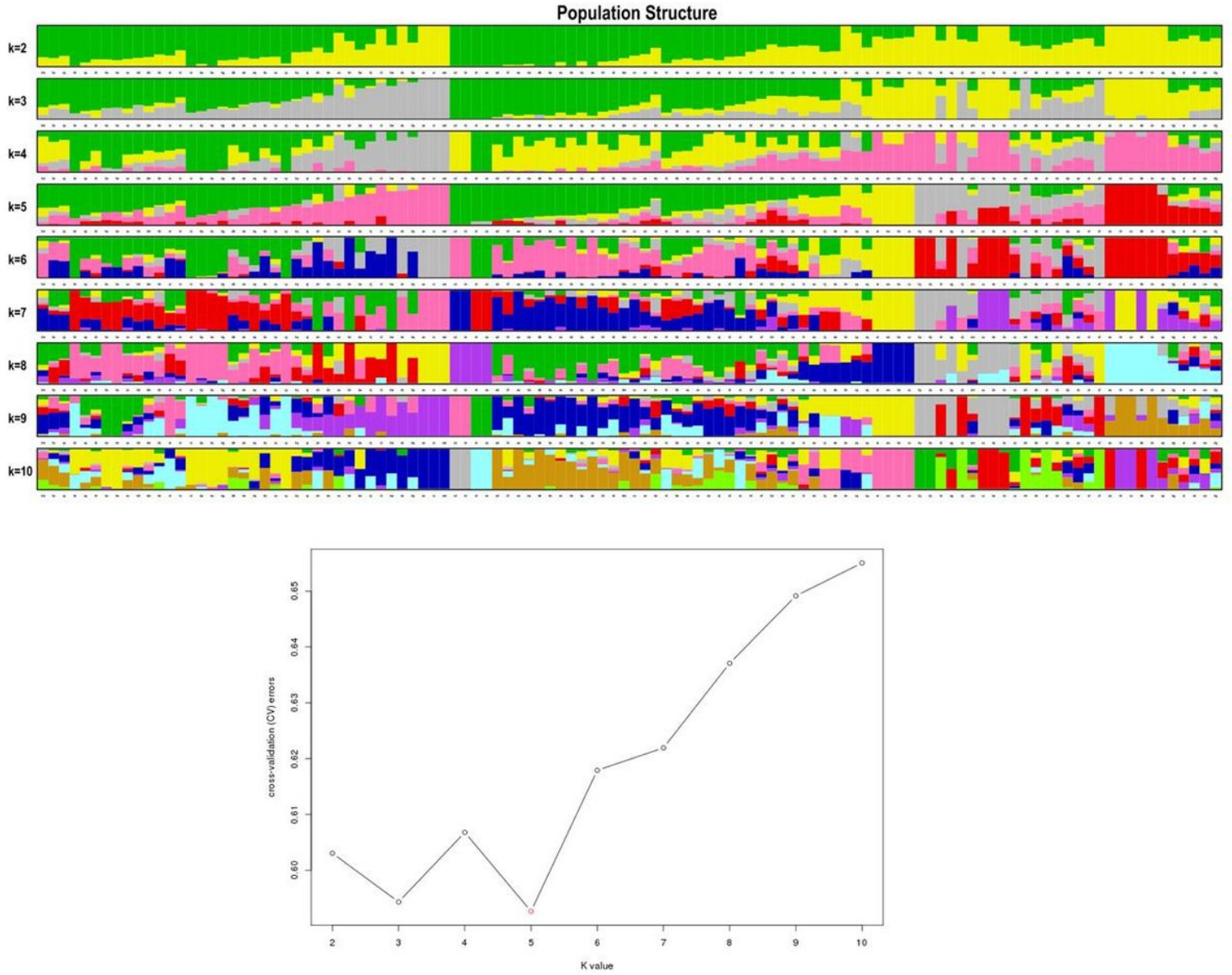

Figure 1

Population structure inferred by Admixture analysis. The accessions were divided into five subgroups. There were minimum cross-validation errors when K =5. Within each subgroup, the accessions were ordered according to their genetic component. Each line indicates the subgroup value. Each accession is represented by a vertical line partitioned into K colored components which represent the inferred memberships in K genetic clusters.

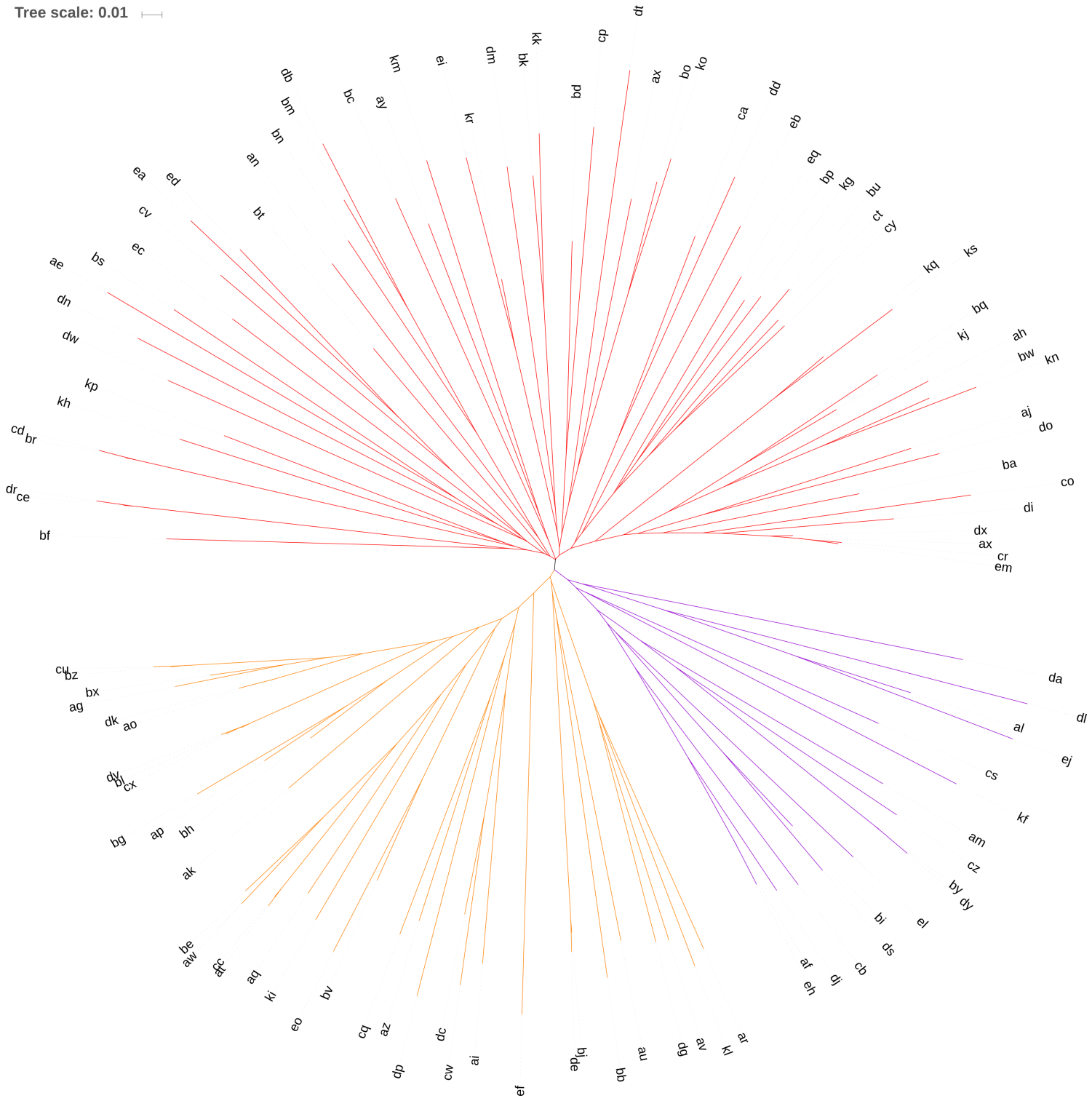

Figure 2

Neighbor-joining (NJ) tree of 112 ramie core germplasms.

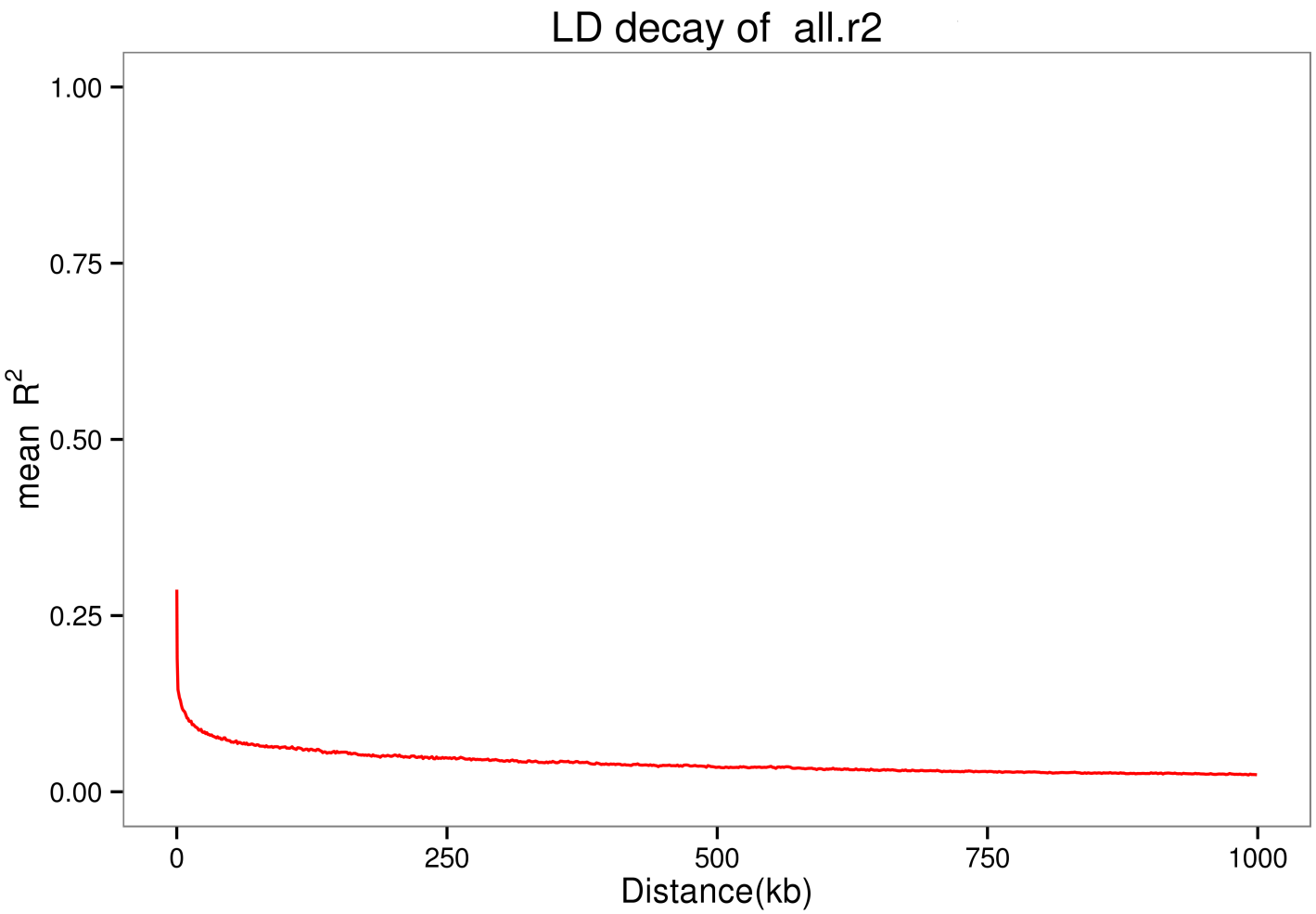

Figure 3

Genome-wide linkage disequilibrium (LD) decay for all ramie accessions.

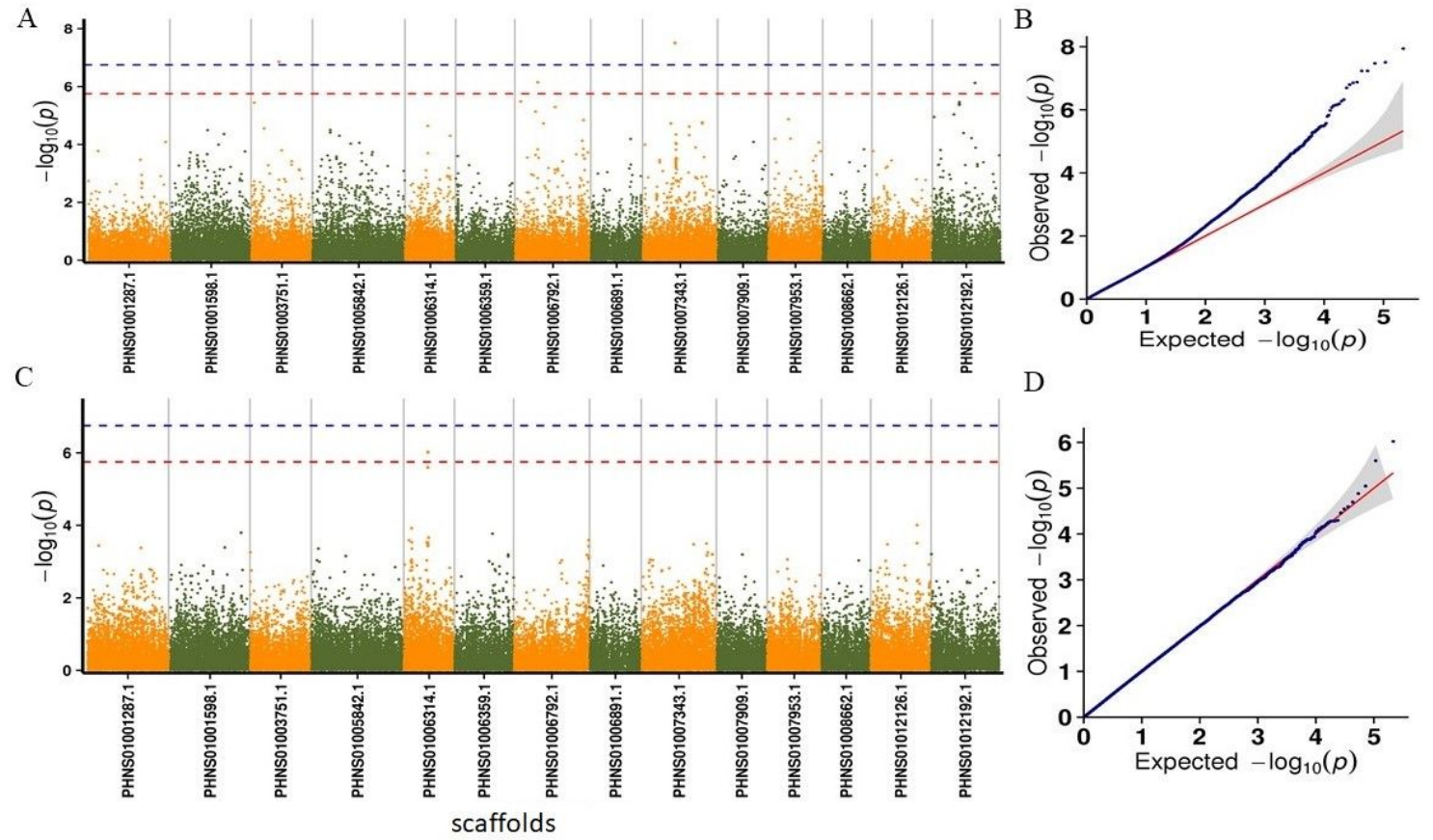

Figure 4

Manhattan and Q-Q plots of the SNP markers for stem diameter and skin thickness in the third season.

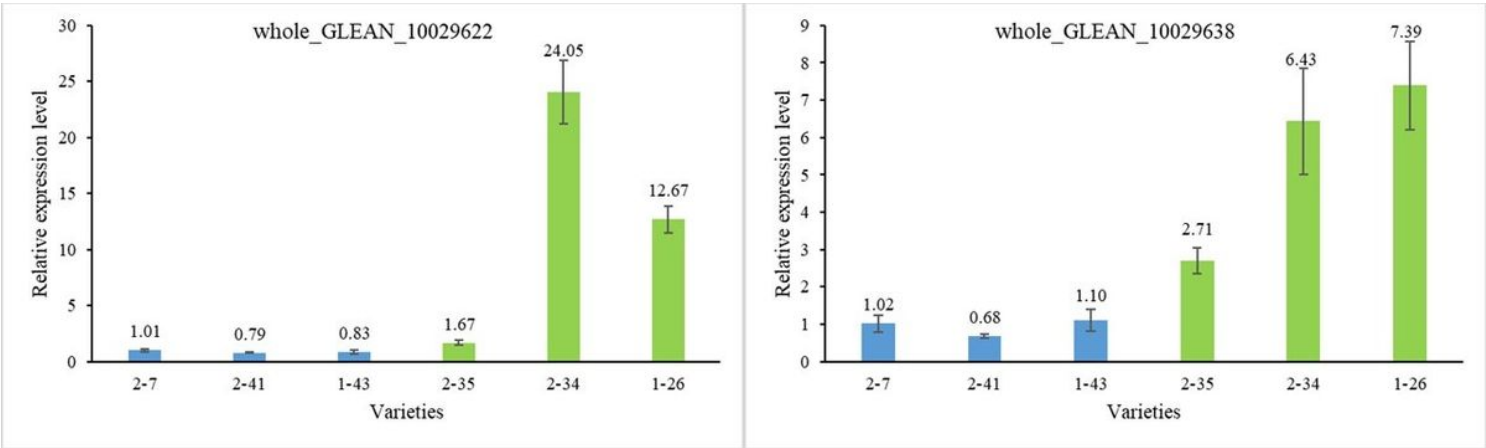

Figure 5

Validation of the expression patterns of the candidate gene encoding fiber fineness. The average fiber fineness of the six varieties were 2–7 (1,387 m g-1), 2–41 (1,455 m g-1), 1–43 (1,526 m g-1), 2–35 (2,068 m g-1), 2–34 (2,199 m g-1), and 1–26 (2,237.5 m g-1), respectively

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfiles4.xls
- Additionalfiles7.xls
- Additionalfiles8.xls
- Additionalfiles1.docx
- Additionalfiles6.xls
- Additionalfiles10TableS9.PrimersforqRTPCR.xlsx
- Additionalfiles5.xls
- Additionalfiles2TableS1.LDdecay.xlsx
- Additionalfiles3.xlsx
- Additionalfiles9TableS8.ramiegermplasms.xlsx