Resequencing of 301 ramie accessions identifies genetic loci and breeding selection for fibre yield traits

Zheng Zeng\textsuperscript{1,1}, Siyuan Zhu\textsuperscript{1,3}, Yanzhou Wang\textsuperscript{1,4}, Xuehua Bai\textsuperscript{1,4}, Chan Liu\textsuperscript{1}, Jianrong Chen\textsuperscript{3}, Ting Zhang\textsuperscript{2}, Yiping Wei\textsuperscript{1}, Fu Li\textsuperscript{1}, Zhigui Bao\textsuperscript{2}, Li Yan\textsuperscript{1}, Hengyun Wang\textsuperscript{2} and Touming Liu\textsuperscript{1,2}

\textsuperscript{1}Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha, China
\textsuperscript{2}Shanghai OE Biotech. Co., Ltd, Shanghai, China
\textsuperscript{3}College of Biological and Environmental Engineering, Changsha University, Changsha, China

Received 23 April 2021; accepted 13 September 2021.
*Correspondence (Tel 86 731 88998580; fax 86 731 88998528; email liutoming@caas.cn)
\textsuperscript{*}These authors contributed equally to this article.

Keywords: ramie, fibre yield, genomic variation, whole genome association study, breeding selection.

Introduction

Bast fibres and silks are two of the most popular sources of textile fabric in Chinese history. Ramie, commonly known as China grass, is one of the most important bast fibre crops in ancient China, and its fibres have been used as textiles for over 4700 years, according to archaeological evidence (Liao \textit{et al.}, 2014). Presently, ramie is the second most important fibre crop in China, and its growth acreage and quantity of fibre production are second only to those of cotton (Liu \textit{et al.}, 2014a). Currently, ramie fibre is a popular natural fibre in the textile industry because of its ability to retain shape, reduce wrinkling and introduce a silky lustre to the appearance of a fabric (Kadolph and Langford, 2001).

Even though ramie fibres have been commonly used for thousands of years, the breeding history of ramie in China spans no more than 60 years as it began in the 1960s (Xiong, 2008). In recent years, despite the development of breeding studies, little progress has been made in improving fibre yield and quality. Focusing on the developmental process will improve ramie fibre traits; therefore, in the past ten years, hundreds of genes potentially associated with fibre development have been identified based on homology and/or expression analyses (Chen \textit{et al.}, 2014; Liu \textit{et al.}, 2013, 2014b; Tang \textit{et al.}, 2019). However, none of these genes were confirmed to be involved in fibre development. Recent genetic studies identified some genetic loci associated with fibre traits (Chen \textit{et al.}, 2019; Liu \textit{et al.}, 2014c, 2017; Zeng \textit{et al.}, 2019); however, these loci showed limited benefits for breeding improvement based on marker-assisted selection because the examined population originates from only a few parent individuals, which provided only a limited number of excellent alleles.

Comprehensive genomic variation maps are essential for breeding studies, because of its plentiful elite alleles. Genome-wide association studies (GWASs) based on whole-genomic variation maps are powerful tools for identifying genes or quantitative trait loci (QTLs) underlying complex traits, as was demonstrated in rice (Huang \textit{et al.}, 2010), maize (Tian \textit{et al.}, 2011), cotton (Du \textit{et al.}, 2018), peach (Li \textit{et al.}, 2019) and melon (Zhao \textit{et al.}, 2019). Here, we present a comprehensive genomic variation map of ramie by resequencing 301 landrace and cultivar accessions. Furthermore, to elucidate the genetic architecture of fibre yield traits, which is a major objective of this study, we performed a GWAS with respect to six fibre yield-related traits. We also characterized and compared population diversities of landraces and cultivars, thereby providing genomic insights into the breeding history of ramie.

Results

Genomic variation

In total, 301 ramie accessions including 254 landraces, 13 cultivated materials and 34 improved cultivars were used in this study (Table S1). Sequencing of these 301 accessions using an Illumina HiSeq platform (Illumina, San Diego, CA, USA) generated 11.4 billion paired-end reads of 150 bp in length (1.7 Tb of sequences), at 19.5-fold coverage. After aligning the produced reads to the ramie reference genome (downloaded from NCBI, accession no. PRJNA663427), we identified 5,951,826 single...
nucleotide polymorphisms (SNPs) and 781 484 small insertions or deletions (indels; ≤5 bp), with an average of 22.0 SNPs and 2.9 indels per kb (Tables S2 and S3). Among these variants, 522 587 SNPs and 18 237 indels were within coding regions, 245 951 of these SNPs caused a change in amino acids in the encoded protein, and 11 315 indels led to gene transcriptional frame shifts (Tables S2 and S3), suggesting that they may affect the function of the respective gene. Taken together, these genomic variations provide a valuable resource for studies on biology and breeding in ramie.

Association analyses

The genetic structure of a population can markedly influence the power of a GWAS (Korte and Farlow, 2013), thus, before performing the GWAS, we inferred the population structure using a subset of 1 433 616 SNPs without close linkages. A Bayesian clustering analysis and a principal component analysis (PCA) indicated no distinct grouping in the populations (Figures S1 and S2), suggesting that our population was suitable for a GWAS. We also analysed the linkage disequilibrium (LD), and discovered that the LD decay reached half its maximum average correlation coefficient ($r^2$) at a distance of 0.65 kb in the landrace and of 1.3 kb in the cultivar (Figure S3). The fibre yield of ramie is determined mainly by three stem bark traits (bark weight, bark thickness, and fibre output ratio) and two stem morphological traits (stem length and diameter). Phenotypic investigation indicated substantial variation among fibre yield traits and five associated traits in the population, with the coefficient of variation ranging from 16.1% to 45.3% (Table S4). These six traits showed significant correlations with each other in the three environments ($P<0.01$), except for fibre output ratio and bark thickness in 2019 (Table S5). We then performed an association analysis for fibre yield traits and five related traits in this population. In total, 129 significant association signals involved in 70 genomic regions were detected, nine of which were associated with two traits (Figures S4–S9, Tables S6 and S7). Twenty-one fibre yield-associated loci were identified, eight of which showed an overlap with loci of related traits (within 30 kb). Notably, the SNP on chromosome 10 (position 14,843,234) was strongly associated with the traits bark weight and fibre yield, and it showed the strongest signal associated with fibre yield ($P = 2.4 \times 10^{-5}$) among all association SNPs.

### Candidate genes of fibre yield and stem bark traits

We examined genes containing SNPs associated with traits and found that 58 fibre yield-associated SNPs were located in 21 genes (including the genic and upstream/downstream regions within 2 kb), and 47 SNPs associated with three stem bark traits (bark weight, bark thickness and fibre output ratio) occurred in 27 genes (Table S6). Because fibre yield and three stem bark traits are mainly determined by fibre growth in the bast bark, we subsequently investigated the expression profile of bark at different stages of fibre development to identify candidate genes. Our results revealed that, among those genes containing SNPs associated with these four traits, seven genes showed distinct expression differences in fibre-developing bast bark of stems, compared to stem bark that did not show initiation of fibre secondary wall growth (Table 1).

A peak strongly associated with bark thickness, identified on chromosome 9, was located in Bnt09G014694 (Figure S10), which is an SBP-box transcription factor-encoding gene, and its expression was considerably downregulated in fibre-developing bast bark (Figure S10). Similarly, expression of the cotton SBP transcript TC253516 was gradually reduced from the elongation stage to the stage of secondary wall thickening; TC253516 is a target of small RNA Gb-miRNA156/157, and suppression of the activity of miRNA156/157 not only caused SBP transcript accumulation but also resulted in inhibition of fibre elongation, indicating that the SBP transcript TC253516 negatively modulates initiation of production and elongation of cotton beard (Liu et al., 2014). Therefore, our results suggest that Bnt09G014694 is a candidate gene responsible for bark thickness.

Among the 14 signals associated with fibre output ratio, one was located in the promoter region of Bnt03G004997 (Figure 1a). Bnt03G004997 is the homologue of Arabidopsis VND4 and VND5, which are two master switches eliciting secondary wall biosynthesis (Figure 1b; Zhou et al., 2014), and overexpression of Bnt03G004997 caused an increase in the number of fibres in transgenic Arabidopsis (Figure 1c). Additionally, the rice COBRA-Like gene Brittle Culm1 can modulate cellulose assembly by interacting with cellulose and affecting microfibril crystallinity; thus, this gene is critical for secondary wall cellulose biosynthesis (Liu et al., 2013a, b). We here identified an associated signal-located 265-bp downstream of the COBRA-like gene Bnt10G014913; expression analysis of Bnt10G014913 revealed

### Table 1 Association signal-located genes with differential expression in the fibre developmental barks of stems

| Gene            | Trait            | SNP position | $P$-value  | Signal position in gene | Annotation                  |
|-----------------|------------------|--------------|------------|-------------------------|----------------------------|
| Bnt09G014694   | Bark thickness  | 17 322 051   | $6.63 \times 10^{-7}$ | Coding region           | SBP-box transcription factor |
| Bnt03G004293   | Fibre output ratio | 11 733 967   | $5.11 \times 10^{-7}$ | Coding region           | BTB/POZ domain-containing protein |
| Bnt03G004997   | Fibre output ratio | 18 319 491   | $4.39 \times 10^{-7}$ | Promoter                | NAC domain-containing protein |
| Bnt10G014913   | Fibre output ratio | 5 882 131    | $7.89 \times 10^{-7}$ | Downstream              | COBRA-like protein          |
| Bnt10G015830   | Fibre yield      | 15 536 000   | $6.94 \times 10^{-10}$ | Coding region           | Peroxidase                 |
| Bnt10G015830   | Bark weight      | 15 536 000   | $2.55 \times 10^{-7}$ | Coding region           | Peroxidase                 |
| Bnt10G014876   | Fibre yield      | 5 097 702    | $6.94 \times 10^{-7}$ | Coding region           | –                          |
| Bnt14G019616   | Bark weight      | 6 934 389    | $3.87 \times 10^{-7}$ | Coding region           | Pectin methyl esterase     |
| Bnt14G019616   | Fibre yield      | 6 934 389    | $7.72 \times 10^{-7}$ | Coding region           | Pectin methyl esterase     |
a distinct expression difference between fibre-developing bast bark and stem bark whose fibres did not show initiation of secondary wall thickening (Figure S11). Taken together, our results indicate that Bnt03G004997 and Bnt10G014913 are two candidate genes associated with fibre output ratio.

One peak on chromosome 10 was strongly associated with both bark weight and fibre yield. This peak was produced by the peroxidase-encoding gene Bnt10G015830 (Figure 2a). Peroxidases play key roles in the biosynthesis of secondary walls and are involved in the polymerization of lignin (Herrero et al., 2013; Shigeto et al., 2013, 2014). Our expression analysis showed significant downregulation in the expression of Bnt10G015830 in fibre-developing bast bark (Figure 2b). Notably, we identified a single-base insertion in the fourth exon of Bnt10G015830 in six accessions, which caused a frameshift change in the encoded protein (Figure 2c). Phenotypic investigation showed that more stem bark and fibre could be harvested from these six mutants than from the others (Figure 2d). Therefore, our results suggested that Bnt10G015830 is a pleiotropic candidate gene responsible for bark weight and fibre yield traits.

Natural variation in Bnt14G019616 associated with extremely low fibre content

We identified a germplasm, ‘1380’, which showed extremely thin bark (~0.33 mm), low bark weight (~5.2 g/stem), and low fibre content (~0.64 g/stem), thereby causing brittleness of the stem (Figure 3a). To identify QTLs responsible for thin bark and low fibre content in ‘1380’, we developed an F2 segregation population derived from a crossing of ‘1380’ and the elite cultivar Xiangzhu 1 and then produced a high-density genetic map (Figure S12). Genetic mapping for bark thickness and fibre yield traits identified a major-effect QTL on chromosome 14 (Figure 3b; Figure S13). We further mapped this QTL by resequencing 40 F2 progenies with extremely thin stem bark (<0.28 mm), and delimited the QTL into a 168-kb region (from position 6 783 728 to 6 952 260 of chromosome 14; Figure 3c).
in two recombinant individuals. At this interval, we identified an SNP (position 6,934,389) that was strongly associated with the two traits bark weight and fibre yield (Figure 3d). This associated SNP was found in two recombinants of the ‘1380’ genome, including four deletions and 148 SNPs (within ~25 kb; Figure 3g; Figure S15). After comparing the two parents’ expression levels of Bnt14G019616, we observed markedly more transcripts in the bark of ‘1380’ than in that of Xiangzhu 1 (Figure 3h). Furthermore, Bnt14G019616 expression levels showed a significant positive correlation with the pectin content of bark ($P = 0.0262$, $r^2 = 0.689$), but a negative correlation with fibre yield ($P = 0.0277$, $r^2 = -0.590$), among ten investigated accessions (Figure 3h). Therefore, our result indicated that natural variations from promoter region of Bnt14G019616 were likely associated with the extremely low fibre content in the bark of germplasm ‘1380.’

Breeding selection for increasing fibre yield

To acquire insights with respect to selection for increasing fibre yield during ramie breeding history, we scanned genomic regions that showed markedly decreased nucleotide diversity by comparing the cultivar group with the landrace group using 50-kb windows. We discovered 173 putative selection sweeps in ramie (i.e. with $\pi/\theta$ value in the top 5%), covering 6.48% of the ramie genome (17.5 Mb) and comprising 1,128 predicted genes (Figure 4a; Tables S8 and S9). The largest region reflecting selection (925 kb) was identified on chromosome 9. Notably, ~40.5% (7.08 Mb) of the selection sweeps were identified on chromosome 9. Additionally, substantial variations were identified in the promoter region of Bnt14G019616 in the ‘1380’ genome, including four deletions and 148 SNPs (within ~25 kb; Figure 3g; Figure S15). After comparing the two parents’ expression levels of Bnt14G019616, we observed markedly more transcripts in the bark of ‘1380’ than in that of Xiangzhu 1 (Figure 3h). Furthermore, Bnt14G019616 expression levels showed a significant positive correlation with the pectin content of bark ($P = 0.0262$, $r^2 = 0.689$), but a negative correlation with fibre yield ($P = 0.0277$, $r^2 = -0.590$), among ten investigated accessions (Figure 3h). Therefore, our result indicated that natural variations from promoter region of Bnt14G019616 were likely associated with the extremely low fibre content in the bark of germplasm ‘1380.’

Breeding selection for increasing fibre yield

To acquire insights with respect to selection for increasing fibre yield during ramie breeding history, we scanned genomic regions that showed markedly decreased nucleotide diversity by comparing the cultivar group with the landrace group using 50-kb windows. We discovered 173 putative selection sweeps in ramie (i.e. with $\pi/\theta$ value in the top 5%), covering 6.48% of the ramie genome (17.5 Mb) and comprising 1,128 predicted genes (Figure 4a; Tables S8 and S9). The largest region reflecting selection (925 kb) was identified on chromosome 9. Notably, ~40.5% (7.08 Mb) of the selection sweeps were identified on
chromosome 9, accounting for 39.6% of the length of this chromosome; this indicates that chromosome 9 was frequently selected for the improvement of ramie during its breeding history.

Stem length is one of the most important factors determining fibre yield and has received substantial attention from breeders. We identified one stem length-associated signal that overlapped with the selective sweep on chromosome 9 (Figure S16). Gibberellins (GAs) play key roles in promoting stem elongation (Yamaguchi, 2008), and we identified four gibberellin-related genes from the selection sweeps, including two gibberellin 2-beta-dioxygenase genes (Bnt13G019022 and Bnt13G019023), one gibberellin receptor GID1C gene (Bnt02G003447), and one DELLA gene (Bnt09G013717). DELLA protein is a key regulator of plant growth because of its role in GA-signalling repression (Daviere et al., 2008). We investigated the allele frequency of the DELLA gene Bnt09G013717 in 301 germplasms and identified a total of eight haplotypes with respect to this gene (Figure 4b). Our results revealed that four haplotypes appeared only in...
landraces, including Hap3 and Hap7, and two haplotypes (Hap4 and Hap6) were present mainly in the cultivars. Only one variation (a C/T substitution at the 7,774,070th nucleotide site on chromosome 9) was identified between Hap7 and Hap4, and one (a C/G substitution at the 7,774,544th nucleotide site on chromosome 9) between Hap3 and Hap4, indicating that these two variations of \textit{Bnt09G013717} were frequently the focus of selection during ramie breeding history, thereby resulting in two selection signatures in this DELLA gene of the breeding cultivars.

The improved cultivar showed better performance regarding fibre yield and its related traits, compared to the landraces (Figure 4c; \( P < 0.0001 \)). To understand the selection of genes underlying these agronomic traits during ramie improvement, expression levels of genes from the sweeps were investigated. Our results revealed that 156 genes, which had undergone putative selection, showed significant differences in the abundance of transcripts and/or peptides in fibre-developing bast bark (Tables S10 and S11). Fibber development mainly involves in the biosynthesis of the secondary wall, which is mediated by an NAC-MYB-based regulatory network (Zhong and Ye, 2015). Among these differentially expressed genes, one NAC gene (\textit{Bnt05G0007257}) and four MYB genes (\textit{Bnt04G006700}, \textit{Bnt07G011845}, \textit{Bnt09G014274} and \textit{Bnt09G014294}) were identified (Figure 4d). \textit{Bnt05G0007257} is an orthologue of \textit{Arabidopsis} fibre developmental regulator AtMYB103, and \textit{Bnt07G011845} is an orthologue of AtMYB103 that is involved in fine-tuning the transcriptional regulation of secondary wall biosynthesis (Figure 4e; Zhong et al., 2008). Moreover, among the signals associated with fibre yield and three bast bark traits, four were located in the sweeps (Figure 4a), indicating that these associated loci were subjected to selection during the improvement history. \textit{Bnt05G008345} is a peroxidase-encoding gene, near a signal associated with bark weight, which was located in a sweep on chromosome 5 (Figure S17). Allele frequency investigation identified six haplotypes of \textit{Bnt05G008345} in the 301 germplasms, and only two of them
occurred in cultivars (Figure 5), further suggesting that this gene had undergone selection for improvement. Expression analysis showed a distinct decrease in the abundance of transcripts and peptides in fibre-developing bark (Figure S18). Because peroxidases play a role in the biosynthesis of secondary walls (Herrero et al., 2013; Shigeto et al., 2013, 2014), we deduced that Bnt05G008345 is a candidate gene for bark weight and that it had undergone selection for ramie improvement.

Discussion

In this study, we characterized the population structure and elucidated the genetic architecture associated with fibre yield and its related traits in ramie based on whole genome variations. Our population analysis revealed that the LD of ramie decayed to half-maximum within only 0.65–1.3 kb, which is lower than that in other small LD-extent species reported previously, such as tea (5 kb; Xia et al., 2020) and wild European grapevines (2.9 kb; Liang et al., 2019). Therefore, a comprehensive map of genomic variations is essential to identify more markers/genes associated with traits from the GWAS analysis in ramie. Two previous studies performed association analyses of agronomic traits using limited single sequence repeats and SNPs from specific-locus amplified fragment sequencing technology in ramie (Chen et al., 2018; Luan et al., 2017), and only few markers have been identified to be associated with traits at a very low statistical level (P < 0.01 and P < 0.05, respectively). The current study showed large-scale variations in the ramie genome (~24.9 markers/kb), which constitutes a comprehensive genomic variation map for this crop. Identification of these variations facilitates comprehensive

Figure 5 Haplotypes and their effects of five candidate genes. Haplotypes were analysed based on the SNPs of coding region of each gene. The violin plot was produced using fibre yield data of accessions with the respective haplotype. The y axis indicates fibre yield (g) that was investigated in 2018. The orange and blue part in each circle under the haplotype ID indicate the ratio of landrace and cultivar with this haplotype.
association analyses of agronomic traits in ramie, we, therefore, identified hundreds of SNPs associated with fibre yield-related traits and identified several genes as candidates for respective traits. Fibre yield was significantly correlated with its related traits, especially with bark weight ($r^2 = 0.875$–0.911). Notably, genetic analysis showed that 8 of 21 fibre yield-associated genomic regions showed a pleiotropy. This result indicated that these pleotropic regions represent an important genetic basis for the correlation of fibre yield and its related traits. Taken together, the results of the present study provide new insights into the genetic architecture of fibre yield of ramie, and genome-wide variations are valuable resources for future breeding studies on this fibre crop.

Pectin is one of the main components of the primary cell wall of plants, and pectin secreted from the Golgi complex occurs in highly methylsterified forms. For cell wall biosynthesis, these forms must be further modified by pectinases such as pectin methyltransferases, which catalyse demethylsterification of acidic pectin and methanol (Micheli, 2001). Therefore, pectin methylsterases are essential for cell wall growth. Fibres are cells with a thickened secondary cellular wall, and the role of pectin methylsterase in the biosynthesis of secondary cellular walls is largely unknown. The results of the current study strongly suggest that the pectin methylsterase-encoding gene Bnt14G019616 was negatively associated with fibre growth in bast bark of ramie, thereby providing the first evidence for the role of pectin methyltransferase in bast fibre growth in plants. However, due to the lack of a practical and highly efficient method for genetic transformation of ramie, it is challenging to verify the function of Bnt14G019616 in ramie. The exact mechanism of fibre growth-inhibition by Bnt14G019616 thus requires further clarification.

A large number of landraces has been generated in China throughout the cultivation history over thousands of years (Ni et al., 2018). Since the 1960s, ramie breeding studies have been performed in China using elite landraces as breeding materials, which yielded dozens of cultivars (Xiong, 2008). The genomic basis underlying the selection of traits during the breeding history of ramie is poorly understood. This study discovered evidence for substantially focussed selection on chromosome 9 (including a DELLA gene) by breeders. DELLA protein acts as a negative regulator of GA responses and plays a crucial role in determining plant height (Sun, 2010). There were two DELLA genes in the ramie genome (Bnt04G005980 and Bnt09G013717), one of which, Bnt09G013717, had undergone selection for improvement of ramie. It is likely that the focus of Bnt09G013717 is based on its contribution to stem length and that this gene was, therefore, selected by breeders. Although five genomic regions associated with fibre yield-related traits were identified, most of the fibre yield-related loci had not undergone selection in ramie breeding. This is likely due to its short breeding history.

Exploring and utilizing favourable alleles is an effective method for improving crop traits (Guo et al., 2020). Substantial allelic variation was identified in ramie germplasms with respect to the fibre yield-associated candidate genes identified in the current study, and different haplotypes showed considerable differences in fibre yield (Figure 5). However, our study revealed that most haplotypes of these genes commonly occurred in cultivated accessions, indicating that they were rarely selected during breeding. This is probably due to the short history of ramie breeding. Notably, some favourable alleles such as Hap_6 rarely occurred in the cultivars, suggesting large potential for the improvement of fibre yield in this crop. Therefore, pyramiding these favourable alleles may, in theory, substantially improve the fibre yield of ramie. Taken together, this study not only provides insights into the genomic basis underlying fibre yield improvement in ramie breeding history but also provides an important basis for further improvement of traits in the future breeding of this fibre crop.

**Methods**

**Experimental materials and phenotype collecting**

In total, 301 ramie accessions were collected and grown on the experimental farm of the Institute of Bast Fiber Crops (IBFC) in 2016. Information on the accessions, including names, collection sites, categories, and resequencing data summary, is provided in Table S1. For each accession, two replications with six individuals were planted, according to Zeng et al. (2019). Fibre yield-related traits were investigated each year in June, from 2018 to 2020. Stem length was measured from the bottom of the stem to the shoot. Stem diameter and bark thickness were measured in the middle of the stem using a Vernier calliper. Subsequently, stem bark was harvested individually and was weighed; bark weight was calculated as mean bark weight per stem. Fibres were extracted from these stem barks and were dried, and fibre yield was calculated as the mean weight of fibre per stem. The fibre output ratio was estimated for each accession as the ratio of fibre yield to bark weight. Pectin content in stem bark of ten accessions was assessed according to the following method: 5 g dry sample was placed in 150 mL ammonium oxalate solution with a concentration of 5 g/L, which was boiled for 3 h. The sample was then removed from the solution and was dried and weighed; the dry weight divided by total weight (i.e. 5 g) was considered the pectin content.

**Resequencing and variation calling**

Young leaves of all 301 accessions were individually sampled, and DNA was extracted. The DNA extracts were separately used to construct Illumina sequencing libraries with ~ 350-bp inserts, according to the manufacturer’s instructions. Paired-end sequencing of the libraries was performed using an Illumina HiSeq X Ten platform (Illumina). Raw sequencing reads were filtered to produce clean reads, which were aligned to the ramie genome (downloaded from NCBI; accession no. PRJNA663427) using BWA (version 0.7.15; Li and Durbin, 2010) and allowing no more than 4% mismatches and one gap. The alignment results were converted to bam format using SAMtools (version: 1.9; Li et al., 2009), and duplicated reads were removed using the Picard package. GATK (version 3.8.1; McKenna et al., 2010) was used to identify SNPs and small indels. Briefly, GATK local realignment was performed to refine the read mapping in the presence of the variants, thereby generating a gVCF file. Thereafter, SNP calling was performed using GenotypeGVCFs software, and SNPs were filtered out using the following parameters: call quality divided by depth (QD), 2.0; mapping quality (MQ), 40.0; Fisher’s exact test (FS), 60.0; minor allele frequency (MAF), 0.05; missing genotype rate (MG), 0.2. Indel variations in ramie accessions were identified and filtered out using GATK, with the following parameters: QD, 2.0; MQ, 40.0; FS, 200.0; MGR = 0.2. Variations were annotated using SnpEff software (version: 4.3T; Cingolani et al., 2012).

**Population structure and LD analysis**

All SNPs from ramie accessions were filtered using the PLINK software (version: 1.9; parameters –indep-pairwise 50 5 0.5;
Purcell et al., 2007), generating a subset of SNPs without close linkages. Investigation of population structures was carried out using ADMIXTURE software (version: 1.3.0; Alexander et al., 2009), and 10 times for each K value was run with varying random seeds; the Q-matrices were aligned using pong software (Behr et al., 2016) and were clustered based on similarity. Thereafter, the matrices from the largest cluster were averaged to produce a final matrix of admixture proportions. A PCA was performed using the smartPCA program of EIGENSOFT software (version: 6.1.4; Price et al., 2006). To measure LD levels in groups of landraces and the bred cultivars, $r^2$ of each allele was estimated using PopLDDecay (version: 3.34; Zhang et al., 2019) and the following parameters: -MAF 0.05 –Miss 0.2 –MaxDist 300. Average $r^2$ values were calculated for each distance.

**GWAS**

All high-quality SNPs (MAF ≥ 0.05, missing rate ≤0.2) identified from 301 accessions were used to perform the GWAS. The efficient mixed model association expedited (EMMAX) program (Kang et al., 2010) was used to carry out the GWAS analysis for the six traits in the three environments. Population stratification and hidden relatedness were determined using a kinship (K) matrix generated using the FaST-LMM program (Lippert et al., 2011). The P-value thresholds for suggestive associations were set to 10^{-6}.

**Identification of domestication sweeps**

To detect genomic regions that had potentially undergone selection during ramie breeding, we calculated the nucleotide diversity ($\pi$) in the landrace and bred cultivar groups and then compared them, based on 50-kb sliding windows in 25-kb steps, using VCFTools (version: 0.1.7; Danecek et al., 2011). Windows with the top 5% of the $\pi$ ratios ($\pi_0/\pi_1 > 1.318$) were selected and were merged as candidate selective-sweep regions.

**QTL analysis from the F2 population**

To develop a segregated population, a crossing between germplasm ‘1380’ and the elite cultivar Xiangzhu 1 was produced, resulting in a population comprising 528 F2 progenies. All F2 progenies and two parents were grown in the field of IBFC in April, 2019, and bark thickness and weight of all individuals were investigated in June 2020, as described above. A subpopulation comprising 88 F2 progenies was randomly chosen, and 40 F2 individuals with extremely thin bark (<0.28 mm) were selected from the subpopulation. These 128 progenies and two parents underwent resequencing using genotyping-by-sequencing (GBS) technology. Briefly, GBS libraries of each plant were constructed using the restriction enzymes PstI-HF andMspI, according to the manual of the GBS Kit, and the GBS libraries were subjected to Illumina sequencing. After filtering, clean reads of F2 progenies and two parents were aligned with the ramie genome (NCBI accession number PRJNA663427) using bowtie2 (Langmead and Salzberg, 2012). Each plant’s SNPs were called using HaplotypeCaller in GATK (version 3.8.1; McKenna et al., 2010), and those of different individuals were merged using the GATK GenotypeGVCFs program. Low-quality SNPs were removed using the following parameters: MAF > 0.01, DP > 4, and Miss < 20.0. In total, 22,321 high-quality SNPs were obtained from these F2 progenies, 6035 of which were homozygote in two parents (aa × bb), which were used for constructing a genetic map of the subpopulation (Figure S12) using joinmap5 (Van Ooijen, 2006). QTL analysis of traits in the population consisting of 88 F2 lines was performed using MapQTL6 (Van Ooijen, 2011), and the experimental LOD threshold significance level was determined by computing 1,000 permutations ($P < 0.05$). To delimit QTL to an exact genome region, recombinant individuals in the QTL interval were screened from the 40 progenies with extremely thin bark, and the region in which SNPs with the ‘1380’ allele showed cosegregation with genes responsible for thin bark traits was determined based on the genotypes of recombinant individuals.

**Expression analysis**

Bark tissue samples from the top and middle parts of the stems, where the secondary walls of fibres do not initiate growth or thickening (Chen et al., 2014), were collected from the elite variety Zhongzhu 1 and were used to analyse expression differences of genes/proteins, thereby identifying fibre growth-related genes and proteins. Transcriptome comparison between the two tissue types was performed in our previous study, and 1758 differentially expressed genes were identified (Wang et al., 2021). To detect differentially expressed proteins between them, the two tissue types (using three replicates of each) were subjected to proteome analysis. Briefly, protein extraction was carried out using a Tris-saturated phenol method, followed by digestion and labelling using tandem mass tags (TMT). After fractionation, the peptides were subjected to proteomic analysis using an LC-MS/MS system. The MS/MS data of the proteome analysis were processed using MaxQuant software (v.1.5.2.8; Tyanova et al., 2016). Tandem mass spectra were searched against the protein sequences annotated according to the ramie genome. Subsequently, normalization of protein abundance was performed by TMT, and proteins with more than 1.5-fold change were considered significantly differentially expressed ($P < 0.001$). To compare the expression levels of Bnt14G019616 in stem barks of ten accessions, bark samples from stem of 50-day-old ramie were collected, and RNA was extracted. After reverse transcription, qRT-PCR of Bnt14G019616 was performed using the 18S ribosomal RNA gene as an internal control (primer sequences are listed in Table S12). The relative expression levels were determined as previously described (Livak and Schmittgen, 2001).

**Functional characterization of candidate genes**

To characterize the function of candidate genes, amplification of its full-length cDNA was carried out using a high-fidelity thermostable DNA polymerase (primer pair listed in Table S12). The amplified cDNA was ligated to the downstream CaMV 35S promoter in a PB121 vector. The produced vector was introduced into an Agrobacterium tumefaciens GV3101 strain using the heat shock method, and the resulting Agrobacterium strain was infiltrated into Arabidopsis using through floral dipping (Zhang et al., 2006). The middle parts of stems of transgenic and wild-type plants were used for cell observation. Briefly, transverse sections (~13 μm thick) of stems were prepared from 40-days-old plants, were stained using Safranin O-Fast Green resulting in red coloration of lignified cell walls, and were then observed using light microscopy.

**Gene sequence analysis and comparison**

To compare the sequences of Bnt14G019616 between ‘1380’ and Xiangzhu 1, the genome sequences of the genic and promoter regions were amplified from these two accessions using six primer pairs (Table S12). All amplified fragments were ligated into plasmids through TA cloning, and Sanger sequencing was
performed. Sequences were compared using the SnapGene program. The haplotype type and statistics of candidate genes in different populations were reconstructed using the R packages pegas (Paradis, 2010) and VcF (Knaus and Grünwald, 2017), and a haplotype network of Bnt09G013717 was constructed using popart (Leigh and Bryant, 2015). In total, 13 NAC proteins and 16 MYB proteins have previously been identified to be involved in the biosynthesis of secondary cellular walls in Arabidopsis (Zhong and Ye, 2015). To discover the homologs of our candidates from Arabidopsis fibre growth-related MYB/NAC proteins, a protein alignment was carried out using the Clustal program (Thompson et al., 1997), and an unrooted phylogenetic tree was constructed using MEGA 5 with the neighbour-joining method and a bootstrap test with 1000 replicates (Tamura et al., 2011).

Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (31871678, 32001512), Natural Science Foundation of Hunan Province (2021JJ30769), the Agricultural Science and Technology Innovation Program of China (CAAS-ASTIP-IBFC), Central Public-interest Scientific Institution Basal Research Fund (1610242021002), and the National Modern Agroindustry Technology Research System of China (nycxt-19-E16).

Conflict of interest

The authors declare no competing financial interests.

Author contributions

Z.Z., S.Z., YA.W. and X.B. contributed equally to this work. C.L. and L.Y. managed the project. YA.W. and H.W. performed the sequencing. Z.B. and T.Z conducted the data analyses. Y.L., X.B. and J.C. carried out field experiment, phenotype investigation and data availability statement. The variation data reported in this paper have been deposited in the Genome Variation Map (GVM) in Big Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Science, under accession numbers GVM000103 at http://bigd.big.ac.cn/gvm/getProjectDetail?project=GVM000103.

References

Alexander, D.H., Novembre, J. and Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664.
Behr, A.A., Liu, K.Z., Liu, F.G., Nakka, P. and Ramachandran, S. (2016) pong: Fast analysis and visualization of latent clusters in population genetic data. Bioinformatics, 32, 2817–2823.
Chen, J., Pei, Z., Dai, L., Wang, B., Liu, L., An, X. and Peng, D. (2014) Transcriptome profiling using pyrosequencing shows genes associated with bast fiber development in ramie (Boehmeria nivea L.). BMC Genom. 15, 919.
Chen, J., Ruo, J., Wang, Y., Zeng, Z., Liu, F., Tang, Y., Chen, X. et al. (2019) Integration of quantitative trait loci mapping and expression profiling analysis to identify genes potentially involved in ramie fiber lignin biosynthesis. Genes, 10, 842.
Chen, K., Luan, M., Xiong, H., Chen, P., Chen, J., Gao, G., Huang, K. et al. (2018) Genome-wide association study discovered favorable nucleotide polymorphisms and candidate genes associated with ramet number in ramie (Boehmeria nivea L.). BMC Plant Biol. 18, 345.
Cingolani, P., Platts, A., Wang, I.L., Coon, M., Nuyten, T., Wang, L., Land, S.J. et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly, 6, 80–92.
Danecék, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E. et al. (2011) The variant call format and VCFtools. Bioinformatics, 27, 2156–2158.
Davore, I.M., de Lucass, M. and Prat, S. (2008) Transcriptional factor interaction: a central step in Della function. Curr. Opin. Genet. Dev. 18, 295–303.
Du, X., Huang, G., He, S., Yang, Z., Sun, G., Ma, X., Li, N. et al. (2018) Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. Nat. Genet., 50, 796–802.
Guo, H., Zeng, Y., Li, J., Ma, X., Zhang, Z., Lou, Q., Li, J. et al. (2020) Differentiation, evolution and utilization of natural alleles for cold adaptability at the reproductive stage in rice. Plant Biotechnol. J., 18, 2491–2503.
Herrero, J., Fernandez-Perez, F., Yebra, T., Novo-Uzal, E., Pomar, F., Pedrero, M.A., Cuello, J. et al. (2013) Bioinformatic and functional characterization of the basic peroxidase 72 from Arabidopsis thaliana involved in lignin biosynthesis. Planta, 237, 1599–1612.
Huang, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C. et al. (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. 42, 961–967.
Jolie, R.P., Duvetter, T., Van Loey, A.M. and Hendrickx, M.E. (2010) Pectin methylesterase and its proteinaceous inhibitor: a review. Carbohydr. Res. 345, 2583–2595.
Kadolph, S.J. and Langford, A. (2001) Textiles, 9th ed. Upper Saddle River, NJ: Prentice Hall.
Kang, H.M., Sul, J.H., Saerwicke, S.K., Zaitlen, N., Kong, S., Freimer, N., Sabatti, C. et al. (2010) Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. 42, 348–354.
Knaus, B.J. and Grünwald, N.J. (2017) vcfR: A package to manipulate and visualize variant call format data in R. Mol. Ecol. Resour. 17, 44–53.
Korte, A. and Farlow, A. (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods, 9, 29.
Langmead, B. and Salzberg, S.L. (2011) Fast gapped-read alignment with Bowtie 2. Nat Methods, 9, 357–359.
Leigh, J.W. and Bryant, D. (2015) POPART: full-feature software for haplotype network construction. Methods Ecol. Evol. 6, 1110–1116.
Li, H. and Durbin, R. (2010) Fast and accurate read alignment with Burrows-Wheeler transform. Bioinformatics, 26, 589–595.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Horn, M., Marth, G. et al. (2009) The sequence alignment/map format and SAMtools. Bioinformatics, 25, 2078–2079.
Li, Y., Cao, K.E., Zhu, G., Fang, W., Chen, C., Wang, X., Zhao, P. et al. (2019) Genomic analyses of an extensive collection of wild and cultivated accessions of Ramie (Boehmeria nivea L.) Gaud. in Freyc. determined by nuclear SSR marker analysis. Genet. Resour. Crop Evol. 61, 55–67.
Lippert, C., Listgarten, J., Liu, Y., Kadie, C.M., Davidson, R.I. and Heckerman, D. (2011) FaST linear mixed models for genome-wide association studies. Nat. Methods, 8, 833–835.
Li, C., Zhu, S., Tang, S., Wang, H., Zheng, X., Chen, X., Dai, Q. et al. (2017) QTL analysis of four main stem bark traits using a GBS-SNP-based high-density genetic map in ramie. Sci. Rep. 7, 13458.
Liu, L., Shang-Guan, K., Zhang, B., Liu, X., Yan, M., Zhang, L., Shi, Y. et al. (2013) Brittle Culm1, a COBRA-like protein, functions in cellulose assembly through binding cellulose microfibrils. PLoS Genet. 9, e1003704.
Liu, T., Tang, S., Zhu, S., Tang, Q. and Zheng, X. (2014a) Transcriptome comparison reveals the patterns of selection in domesticated and wild ramie (Boehmeria nivea L. Gaud). Plant Biol. 96, 85–92.

Liu, T., Zhu, S., Tang, Q., Chen, P., Yu, Y. and Tang, S. (2013) De novo assembly and characterization of transcriptome using illumina paired-end sequencing and identification of CsEs gene in ramie (Boehmeria nivea L. Gaud). BMC Genom. 14, 125.

Liu, T., Zhu, S., Tang, Q. and Tang, S. (2014b) Identification of 32 full-length NAC transcription factors in ramie (Boehmeria nivea L. Gaud) and characterization of the expression pattern of these genes. Mol. Genet. Genomics. 289, 675–684.

Liu, T., Zhu, S., Tang, Q. and Tang, S. (2014c) QTL mapping for fiber yield-related traits by constructing the first genetic linkage map in ramie (Boehmeria nivea L. Gaud). Mol. Breeding. 34, 883–892.

Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. Methods, 25, 402–408.

Luan, M., Liu, C., Wang, X., Xu, Y., Sun, Z. and Chen, J. (2017) SSR markers associated with fiber yield traits in ramie (Boehmeria nivea L. Gaudich). Ind. Crops Prod. 107, 439–445.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K. et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303.

Micheli, F. (2001) Pectin methyltransferases: cell wall enzymes with important roles in plant physiology. Trends Plant Sci. 6, 414–419.

Ni, J., Zhu, A., Wang, X., Xu, Y., Sun, Z. and Chen, J. (2018) Phylogenetic diversity and population structure of ramie (Boehmeria nivea L.). Ind. Crops Prod. 115, 340–347.

Paradis, E. (2010) pegas: an R package for population genetics with an integrated-multipopulation approach. Bioinformatics. 26, 419–420.

Price, A., Patterson, N., Plenge, R., Weinblatt, M., Shadick, N. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904–909.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J. et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575.

Shigeto, J., Kiyonaga, Y., Fujita, K., Kondo, R. and Tutsunami, Y. (2013) Putative cationic cell-wall-bound peroxidase homologues in Arabidopsis, AtPpx2, AtPpx25, and AtPpx71, are involved in lignification. J. Agric. Food Chem. 61, 3781–3788.

Shigeto, J., Nagano, M., Fujita, K. and Tutsunami, Y. (2014) Catalytic profile of Arabidopsis peroxidases, AtPpx2, 25 and 71, contributing to stem lignification. PLoS One, 9, e105332.

Sun, T.P. (2010) Gibberellin-GID1/DELLA: a pivotal regulatory module for plant growth and development. Plant Physiol. 54, 1567–570.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGAS: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol., 28, 2731–2739.

Tang, Y., Liu, F., Xing, H., Mao, K., Chen, G., Guo, Q. and Chen, J. (2019) Correlation analysis of lignin accumulation and expression of key genes involved in lignin biosynthesis of ramie (Boehmeria nivea). Genes, 10, 389.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res., 25, 4876–4882.

Tian, F., Bradbury, P., Brown, P., Hung, H., Sun, Q., Flint-Garcia, S., Rocheford, T.R. et al. (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat. Genet. 43, 159–162.

Tyunova, S., Temu, T. and Cox, J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat. Protoc. 11, 2301–2319.

Van Ooijen, J.W. (2006) JoinMap 4.0, Software for the Calculation of Genetic Linkage Maps in Experimental Populations. Wageningen: Kyazma B.V.

Van Ooijen, J.W. (2011) Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. Genet. Res. 93, 343–349.

Wang, Y., Li, F., Bao, Z., Zeng, Z., An, D., Zhang, T., Rao, J., Ma, Y., Wang, M., Gao, X. et al. (2021) Genomic analyses provide comprehensive insights into the domestication of bast fiber crop ramie (Boehmeria nivea). Under review.

Xia, E., Tong, W., Hou, Y., An, Y., Chen, L., Wu, Q., Liu, Y. et al. (2020) The reference genome of tea plant and resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation. Molecular Plant, 13, 1013–1026.

Xiong, H.P. (2008) Breeding of Bast Fiber Crops, 1st ed. China Agricultural Science and Technology Press.

Yamaguchi, S. (2008) Gibberellin metabolism and its regulation. Annu. Rev. Plant Biol. 59, 225–251.

Zeng, Z., Wang, Y., Liu, C., Yang, X., Wang, H., Li, F. and Liu, T. (2019) Linkage mapping of quantitative trait loci for fiber yield and its related traits in the population derived from cultivated ramie and wild B. nivea var. tenacissima. Sci. Rep. 9, 16855.

Zhang, C., Dong, S.S., Xu, J.Y., He, W.M. and Yang, T.L. (2019) PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics. 35, 1786–1788.

Zhao, G., Lian, Q., Zhang, Z., Fu, Q., He, Y., Ma, S., Ruggieri, V. et al. (2019) A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. Nat. Genet. 51, 1607–1615.

Zhong, R., Lee, C., Zhou, J., McCarthy, R.L. and Ye, Z.H. (2008) A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis. Plant Cell, 20, 2763–2782.

Zong, R. and Ye, Z. (2015) Secondary cell walls: biosynthesis, patterned deposition and transcriptional regulation. Plant Cell Physiol. 56, 195–214.

Zhou, J., Zhong, R. and Ye, Z.H. (2014) Arabidopsis NAC domain proteins, VN1D to VNDS, are transcriptional regulators of secondary wall biosynthesis in vessels. PLoS One, 9, e105726.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 Principal component analysis plots of the first three components of 301 landrace and cultivar accessions.

Figure S2 Model-based clustering analysis with different cluster numbers (k = 2–7). The y axis quantifies cluster membership, and the x axis lists the different accessions.

Figure S3 Decay of linkage disequilibrium measured by r2 for the landrace and cultivar group.

Figure S4 Whole genome association analysis of the fibre yield trait in 301 accessions in three environments.

Figure S5 Whole genome association analysis of the bark weight trait in 301 accessions in three environments.

Figure S6 Whole genome association analysis of the fibre output ratio trait in 301 accessions in three environments.

Figure S7 Whole genome association analysis of the bark thickness trait in 301 accessions in three environments.

Figure S8 Whole genome association analysis of the stem length trait in 301 accessions in three environments.

Figure S9 Whole genome association analysis of the stem diameter trait in 301 accessions in three environments.

Figure S10 The significant signal was associated with bark thickness on chromosome 9 using the EMMAX, and it was in Bnl09G014694. The graph shows the genomic position (x axis) against its significance expressed as −log10 P-value (y axis). The genomic position spans ~40 kb on either side of the peak SNP, indicated by a grey dashed vertical line. The red diamond
indicates the associated SNP, and the lighter colours of the remaining markers indicate their successively lower $r^2$ values (indicated in the legend on the top right). The yellow and purple boxes under the association graph indicate the coding and untranslated region of annotated candidate gene.

**Figure S11** (a) Significant signal was associated with fibre output ratio on chromosome 10 using the EMMAX, and it was located into the downstream of Bnt10G014913, with a 265-bp distance. The graph plots genomic position ($x$ axis) against its significance expressed as $-\log_{10} P$-value ($y$ axis). Genomic position spans ~20 kb on either side of the peak SNP, indicated by a grey dashed vertical line. The red diamond indicates the associated SNP, and the lighter colours of the remaining markers reflect their successively lower $r^2$ values (indicated in the legend on the top right). The yellow boxes under the association graph indicated the coding region of annotated candidate gene. (b) Expression of Bnt10G014913 in the bark from the top part of the stem (TPS) and the middle part of the stem (MPS), where the fibres do not initiate growth or are growing, respectively. FPKM, fragments per kilobase per million reads.

**Figure S12** Genotyping-by-sequencing-based high-density genetic map based on the ramie F$_2$ population derived from the crossing of ‘1380’ and Xiangzhu 1.

**Figure S13** Quantitative trait locus (QTL) analysis for fibre yield trait using a F$_2$ population derived from the cross of ‘1380’ and Xiangzhu 1. A major-effect QTL was identified on chromosome 14. The broken line indicates the genome-wide significance LOD threshold.

**Figure S14** Microscopic examination of the transected stems of Bnt14G019616-overexpressing (OE) and wild Arabidopsis (WT), at 100-fold magnification. The red arrow indicates fibre cells, and the scale bar indicates 200 μm.

**Figure S15** Three deletion in the promoter region of Bnt14G019616 in ‘1380’ were identified at the 561 bp, 681 bp and 778 bp site of gene upstream, from Sanger sequencing.

**Figure S16** Association signal for stem length trait identified in this study that overlapped with selective sweeps on chromosome 9.

**Figure S17** Significant signal was associated with the bark weight trait on chromosome 5 using the EMMAX, and it was located into the upstream of Bnt05G008345, with a 5466-bp distance. The graph plots genomic position ($x$ axis) against its significance expressed as $-\log_{10} P$-value ($y$ axis). Genomic position spans ~50 kb on either side of the peak SNP, indicated by a grey dashed vertical line. The red diamond indicates the associated SNP, and the lighter colours of the remaining markers reflect their successively lower $r^2$ values (indicated in the legend on the top right). The purple boxes under the association graph represented the coding region of annotated candidate gene.

**Figure S18** Expression of mRNA (left) and protein (right) for Bnt05G008345 in the bark from the top part of the stem (TPS) and the middle part of the stem (MPS), where the fibres do not initiate growth or are growing, respectively. The $y$ axis in the left and right graph represented the FPKM value (i.e. fragments per kilobase per million reads) and normalized protein abundance.

Table S1 Basic information for 301 ramie accessions
Table S2 Distribution of SNPs in ramie genome
Table S3 Distribution of small indels within genomic region of ramie
Table S4 Statistical description of five yield traits and five related traits in three environments
Table S5 Correlation coefficient of fibre yields and related traits
Table S6 SNPs associated with six traits in ramie population
Table S7 Genomic regions associated with fibre yield and its related traits in ramie
Table S8 Genomic region with top $5\% \pi_p/\pi_c$ value
Table S9 Genes located in the sweep region
Table S10 Genes from the sweeps with expressed difference between the barks of the top part of the stem (TPS) and the middle part of the stem (MPS)
Table S11 Genes from the sweeps with differential protein expression between bark of the top part of the stem (TPS) and the middle part of the stem (MPS)
Table S12 Primer sequences used in the experiment