agronomy

MDPI

Article

Population Genomics Unravels the Characteristic Relationship between Introgression and Geographical Distribution in Upland Cotton

Chao Shen [1,*], Zheng Cao [2], Zhiyong Xu [3], Lejun Ouyang [1], Xumin Zhang [1], Zhishan Guo [1], Jieli Yu [1], Rong Chen [1] and Wenxi Huang [1]

[1] College of Biological and Food Engineering, Guangdong University of Petrochemical Technology, Maoming 525000, China; ouyanglejun@gdupt.edu.cn (L.O.); xuminzcn@163.com (X.Z.); zhishan_guo33@163.com (Z.G.); yuli2101@163.com (J.Y.); cchrong@163.com (R.C.); wenxi_huang@163.com (W.H.)
[2] Maoming Agricultural Science and Technology Extension Center, Maoming 525000, China; caozheng5192@163.com
[3] National Key Laboratory of Crop Genetic Improvement, College of Plant Science & Technology, Huazhong Agricultural University, Wuhan 430070, China; xuzhiyong@webmail.hzau.edu.cn
* Correspondence: cshen@gdupt.edu.cn

Abstract: Introgression is an important driver of new genetic variation that increases species and genetic diversity. However, the relationship between introgression and geographic distribution of upland cotton is still unclear. Herein, we explored geographically comprehensive genomic data based on 890 upland cotton accessions to decipher the degree of introgression between different geographic distributions and its effect on selection and fiber quality. We found introgression intervals to exist across different geographic distributions. Introgression is affected by the similarity of the environment in which they live, and those with similar ecological environments tend to share the same introgression area. Introgression is affected by artificial selection. A genome-wide association study (GWAS) meta-analysis was performed with 6 fiber traits and identified 261 quantitative trait loci (QTLs). We found that 67 QTLs had introgression signals, and the genome interval size was 118.81 Mb, while 123 QTLs had selection signals, and the genomic interval was 28.38 Mb. These results provide insights into the population-scaled introgression landscape, suggesting that introgression contributed to the cotton genetic improvement, which provides a useful reference for studying intraspecific introgressions from different geographical distributions in other species.

Keywords: upland cotton; population genomics; introgression; geographical distribution

Citation: Shen, C.; Cao, Z.; Xu, Z.; Ouyang, L.; Zhang, X.; Guo, Z.; Yu, J.; Chen, R.; Huang, W. Population Genomics Unravels the Characteristic Relationship between Introgression and Geographical Distribution in Upland Cotton. *Agronomy* **2023**, *13*, 1781. https://doi.org/10.3390/agronomy13071781

Academic Editors: Jane K. Dever, Yang Zhu and Richard G. F. Visser

Received: 10 May 2023
Revised: 8 June 2023
Accepted: 29 June 2023
Published: 30 June 2023

1. Introduction

Introgression refers to the transfer of genetic material between or within species through hybridization and backcrossing in a manner that is very widespread in nature [1,2]. As such, introgression is one of the driving forces for increasing genetic diversity and promoting adaptive evolution [3]. At present, many known gene introgression events involve adaptation [2], which is now known as adaptive introgression and has been found in plants [3,4], humans [5,6] and animals [7,8]. Hybridization is a method commonly used to breed new cultivars with desirable traits between wild and domesticated species in plant breeding [9,10]. The genomic pattern of introgression is affected by recombination, drift and selection [11], so the beneficial and deleterious alleles may be introduced simultaneously [1]. In these cases, interspecific hybridization often results in genetic linkage drag and hybrid sterility, which is an obstacle to plant breeding in which deleterious alleles are introgressed alongside desired beneficial alleles [1]. As a result, introgression is more likely to occur between closely related species [12], especially if the allele for introgression is clearly beneficial in the donor species [2]. However, it is unrealistic to use traditional markers to characterize the fine−scale genomic landscape of introgression pattern. In contrast, population genomics based on advanced sequencing techniques can comprehensively and precisely describe genetic variation, which can be used to infer fine−scale characteristics genomic introgression patterns at the population level. There is little empirical data investigating the introgression patterns in intraspecific hybrid populations. Therefore, the hundreds of different sequenced cotton varieties were used to construct a cotton population level introgression map, which made it possible to analyze the potential introgression sites and potential improved trait sites in the domestication process of upland cotton [13–16].

Cotton genus is not only an excellent model system for the extraordinarily diverse, but also an important cash crop for its fiber, which has 52 species, including 45 diploids and 7 allopolyploids [17]. However, only four cotton species are independently domesticated cultivated cottons, which are two diploid A-genome species, *G. herbaceum* (A_1) and *G. arboreum* (A_2), and two allopolyploid species, *G. hirsutum* (AD_1) and *G. barbadense* (AD_2) [18]. *G. hirsutum* as the major germplasm pool, which has the characteristics of high fiber yield and wide adaptability, accounts for more than 95% of the world's cotton production [19], and exhibits the highest levels of genetic variation and gene flow [20,21]. The several major cultivation areas were formed in the history of China, including the main cultivation areas today, such as the northwest inland areas (mainly Xinjiang), the Yangtze River region and the Yellow River region [22,23]. However, some studies have shown that cultivated upland cotton has low genetic diversity and little signal of genetic introgression within and between species [24]. Compared with foxtail millet [25] and soybean [26], modern cultivated upland cotton is an admixed population with no geographical structure in China [15,23,27]. These may have something to do with the way it is bred, as cotton breeding is often cross-pollinated and re-selected within small sets of breeding materials [28–30]. Moreover, the extent of introgression in different geographic distributions of upland cotton remains unclear.

In the present study, based on 890 publicly re-sequenced upland cotton data sets, we explored the extent of intraspecific introgression between different geographic distributions. We characterized and analyzed population−scaled introgression of upland cotton cultivars at the genome−wide level, and compared the differences in introgression between different geographical distributions of upland cotton. We analyzed the relationship between local introgression region and QTLs related to fiber traits. This study aimed to provide valuable information for understanding the genetic bases of intraspecific introgression and geographical distribution of upland cotton, which will provide superior genetic resources for cotton fiber quality improvement.

2. Materials and Methods

2.1. Plant Materials and Genotyping

Here, 890 public upland cotton accessions were collected and downloaded from the National Center for Biotechnology Information (NCBI), which included 714 improved upland cotton cultivars from China and 176 improved upland cotton cultivars from the USA and other countries (FCM). Furthermore, the 714 improved upland cotton cultivars in China are divided into four groups by geographical distribution, 195 upland cotton cultivars from the Yangtze River region (YtRR), 390 upland cotton cultivars from the Yellow River region (YRR), 118 upland cotton cultivars from the northwestern inland region (NIR) and 11 upland cotton cultivars from other regions. A total of 12.84 Tb of raw genomic sequencing data was obtained, with an average sequencing depth of 6.3 ×(Table S1). First, the raw data were converted to fastq format with sratoolkit.2.4.3 software (NCBI, USA) [31]. Low-quality sequences were filtered with Trimmomatic (v.0.32) software (Jülich, Germany) [32]. Then, the filtered sequences were aligned against reference genome TM-1 [33] with BWA-MEM (v0.7.10-r789) software (Cambridge, UK) [34], and the duplicated mapping reads were filtered with Picard programme (http://broadinstitute.github.io/picard/, accessed on 26 February 2022). GATK (v4.0.1, MA, USA) [35] and SAMtools/bcftools (Cambridge, UK,) [36,37] were used for genotyping, respectively, as described by Shen et al. [23]. To further obtain high confidence variants, only variants on the chromosome shared by GATK [35] and SAMtools/bcftools were retained [36,37], and the variants on the scaffold and the minimum allele frequency less than 5 were filtered. Finally, 1,427,171 SNPs for all, 1,347,173 SNPs for YRR, 1,370,684 SNPs for NIR, 1,302,014 SNPs for YtRR and 1,250,515 SNPs for FCM, were obtained, respectively. The number of common SNP was 1,117,112. The software Beagle (v5.0, Seattle, Washington, DC, USA) [38] with a hidden Markov model was used for genotype phasing.

2.2. Population Genetic Diversity Analysis

STRUCTURE (v.2.3.4, Lausanne, Switzerland) is used to calculate the population structure with K values ranging from 1 to 10 for five repetitions using 1,626,626 SNPs [39]. GCTA software (v.1.25, Queensland, Australia) is used to calculate the principal component analysis (PCA) [40]. LD was calculated using PopLDdecay (v2.69, Xi′an, China) [41]. LD decay was estimated using pairwise r^2 values [27]. SNPhylo (Athens, GA, USA) [42] and iTOL (http://itol.embl.de/, accessed on 6 February 2023) were used to construct and visualize the neighbor-joining tree.

2.3. Genome-Wide Selective Sweeps Detection

To identify potential signals of selection between different geographic distributions, we analyzed the pairwise nucleotide variation (θπ)- and genetic differentiation (F_{ST})-based cross approaches. The θπ, Tajima's D value and F_{ST} were calculated for YRR, NIR, YtRR and FCM with a 100-kb window using VCFtools software (Cambridge, UK) [43]. We set 1% of the F_{ST} and θπ ratio empirical distributions as the threshold for scanning to identify selection signals between pairs of YRR, NIR, YtRR and FCM [44], which were merged into large intervals using BEDtools (v2.27) [45].

2.4. Detection of Intraspecific Introgressions

In order to analyze whether there is introgression between different geographical sources, the identity by descent (IBD) was detected with Refined IBD using default parameters [46] with phased SNPs from 890 upland cotton accessions, YRR accessions, NIR accessions, YtRR accessions and FCM accessions. The genome is divided into bins according to the window of 20 kb with the sliding window of 10 kb. Then, the IBD tracts were obtained each accession and between pairs of YRR, NIR, YtRR and FCM with a custom perl script as described by Fang et al. [47]. The relative frequency of IBD (rIBD) was used for statistics, mainly to infer the area of the introgression sliding window (window size, 20 kb; step size, 10 kb), where positive rIBD values indicate potential introgression [48]. We further considered results that deviated by more than two times standard deviations to be a significant criterion, which were merged into large intervals using BEDtools (v2.27, Charlottesville, VA, USA) [45].

2.5. Meta Genome Wide Association Study for Fiber Traits

We screened a total of 1089 cotton accessions from three previously published articles [14,15,27] based on sequencing depth and phenotype, and finally the remaining 890 cotton accessions. Furthermore, a large−scale Meta−GWAS was carried out with the GAPIT version 3 [49] using 1,427,171 high−quality SNPs ($MAF > 0.05$) using the basic scenario of compressed MLM model by Zhang et al. [50]. The distribution of total SNPS on chromosomes is shown in Figure S1. The distribution and correlation of the five fiber traits are shown in Figure S2. The threshold for significance for Meta-GWAS was evaluated as $-\log(0.05/1{,}427{,}171)$, and the threshold was ~7.39.

3. Results

3.1. Genetic Diversification and Population Properties

Through evolutionary relationship and population structure analysis, we found that 890 upland cotton germplasm materials were divided into four subpopulations (Figure 1a,b).

Each subpopulation contained the cotton materials from four geographically distributed sources (Figure 1c). PCA analysis showed no clear separation based on geographic distribution sources (Figure 1d). Genetic diversity was highest in NIR ($\theta\pi = 1.99 \times 10^{-4}$) and lowest in FCM ($\theta\pi = 1.52 \times 10^{-4}$), indicating a decreasing trend considering geographic distribution (Figure 1e). The difference in LD of different subgroups was small, but higher than that of the overall (Figure 1f). These results indicated that the genetic diversity of the upland cotton population decreased during the improvement process, and population structure cannot be determined based on geographic distribution.

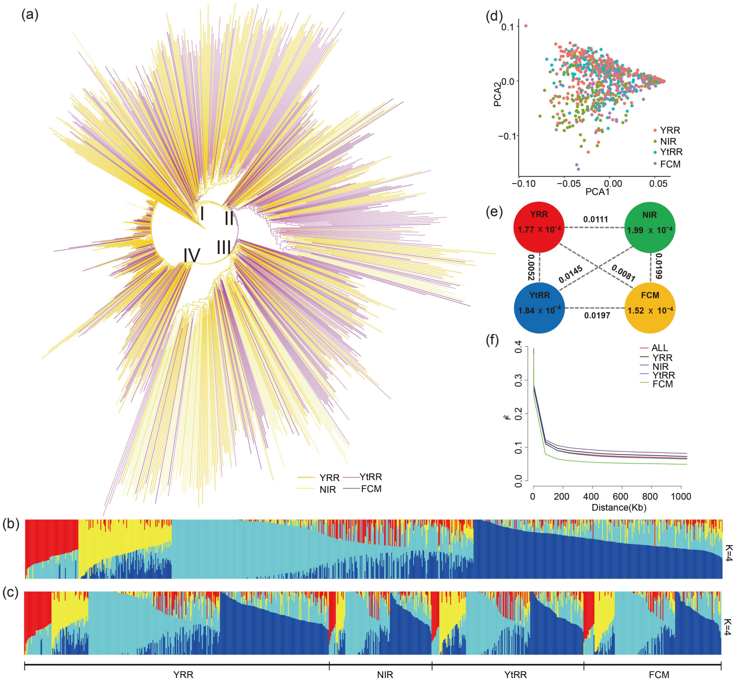

Figure 1. The population diversity of geographical distribution. (**a**) The neighbor-joining phylogenetic tree, which was divided into I, II, III and IV subgroups. (**b**) Population structure results for all samples (k = 4). (**c**) Population structure results for different geographical distribution (k = 4). (**d**) Principal component analysis plot for all accessions. (**e**) $\theta\pi$ and F_{ST} in different geographical distribution. (**f**) Linkage disequilibrium decay of different geographical distribution.

3.2. Whole Genome Characterization of Intraspecific Introgression

In order to estimate the potential pattern of introgression among NIR, YRR, YTRR and FCM, the rIBD statistic was used to infer introgression regions. Here, we found that a total of 42.73 Mb and 76 genomic regions containing 1470 protein−coding genes exhibited introgression signals that were introgressed from NIR to YRR (Figure 2 and Table S2). Noticeably, several large regions (>3 Mb) on chromosomes A10 and D06 exhibit very strong introgression signals (Figure 2). We found 462.49 Mb and 301 genomic intervals, containing 12,013 protein-coding genes exhibited introgression signals that were introgressed from YtRR to YRR (Figure S3 and Table S2). We obtained 25.42 Mb and 25 genomic intervals, containing 514 protein-coding genes that were introgressed from FCM to YRR (Figure S4 and Table S2). There were 45.91 Mb and 58 genomic intervals, containing 2888 protein-coding genes, introgressed from YtRR to NIR (Figure S5 and Table S2). No significant introgression signal was found between FCM and NIR (Figure S6 and Table S2). We obtained 88.29 Mb and 118 genomic regions, containing 2105 protein-coding genes, that were introgressed from FCM to YtRR (Figure S7 and Table S2). Genome-wide introgression pattern analysis showed that the introgression pattern was different among different geographical distributions.

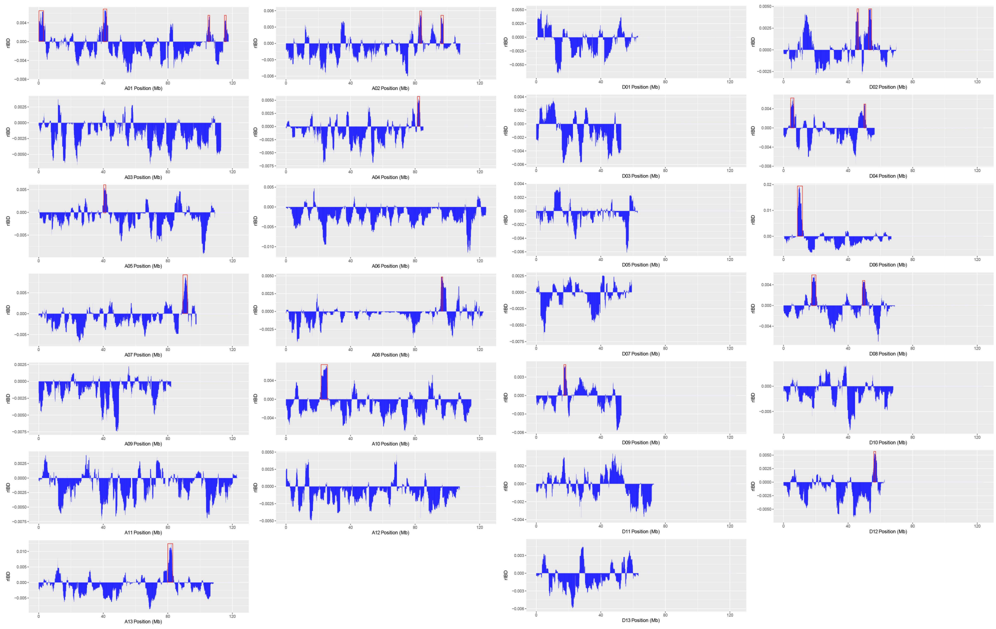

Figure 2. Genome−wide distribution of introgression signals between NIR and YRR. The red boxes indicate significant introgression intervals.

3.3. Selection Signals Reveal Genetic Improvement of Different Geographical Distribution of Upland Cotton Subpopulations

To reveal genetic patterns of genetic improvement in upland cotton across geographic distributions, we scanned the genome using sliding window (window size, 100 Kb; step size, 20 Kb) to identify regions with high differences in allele frequencies and regions with the largest differences in nuclear sequence. We set 1% as the threshold for scanning to identify positive selection intervals between pairwise subpopulations, which exhibited high population differentiation and low genetic diversity (Figures 3 and 4). Furthermore, we identified 134 selected regions (36.94 Mb) between the YRR and NIR subpopulations (Table S3). Between the YRR and YtRR subpopulations, we found 303 selected regions (58.66 Mb) (Table S3). We obtained 631 selected regions (102.44 Mb) between the YRR and FCM subpopulations (Table S3). Between the NIR and YtRR subpopulations, we identified 135 selected regions (36.52 Mb) (Table S3). Between the NIR and FCM subpopulations, we identified 274 selected regions (56.14 Mb) (Table S3). Between the YRR and FCM subpopulations, we identified 725 selected regions (120.14 Mb) (Table S3). In addition, we found that π and Tajima's D value showed significant positive correlation between different geographical sources, but Tajima's D was slightly lower than π. The distribution of YRR Tajima's D values showed more positive values than other geographic distributions, suggesting a large number of high-frequency segregating polymorphisms in YRR (Figure 5). Overall, these results suggested that there were differences in the process of breeding and genetic improvement of cotton from different geographical sources.

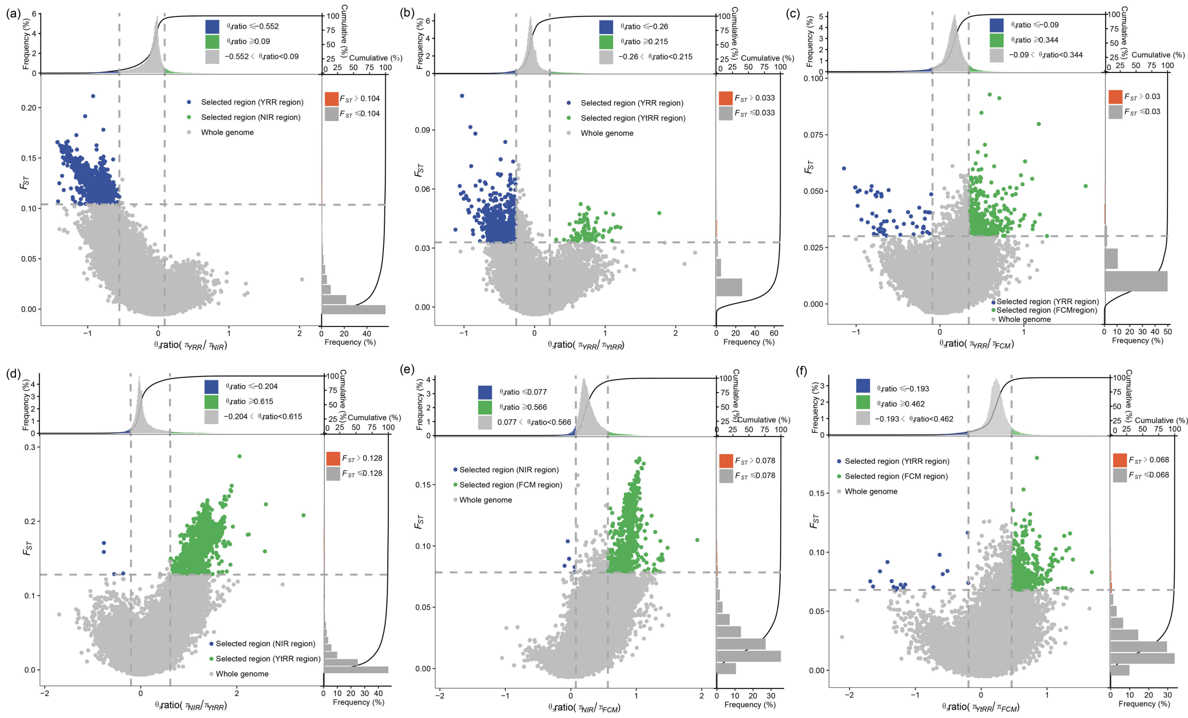

Figure 3. Genome−wide distribution of genomic regions with selective sweep signals of different geographical distribution. (**a**) Distribution of θπ ratios and F_{ST} values between YRR and NIR. (**b**) Distribution of θπ ratios and F_{ST} values between YRR and YtRR. (**c**) Distribution of θπ ratios and F_{ST} values between YRR and FCM. (**d**) Distribution of θπ ratios and F_{ST} values between NIR and YtRR. (**e**) Distribution of θπ ratios and F_{ST} values between NIR and FCM. (**f**) Distribution of θπ ratios and F_{ST} values between YtRR and FCM.

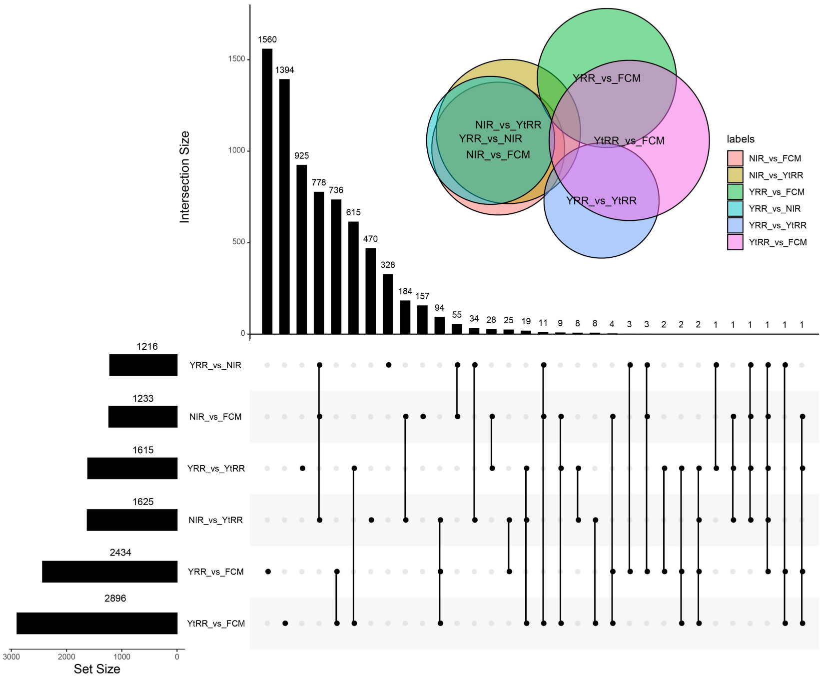

Figure 4. Statistical distribution of pairwise geographic area selection signal intervals across different geographic sources.

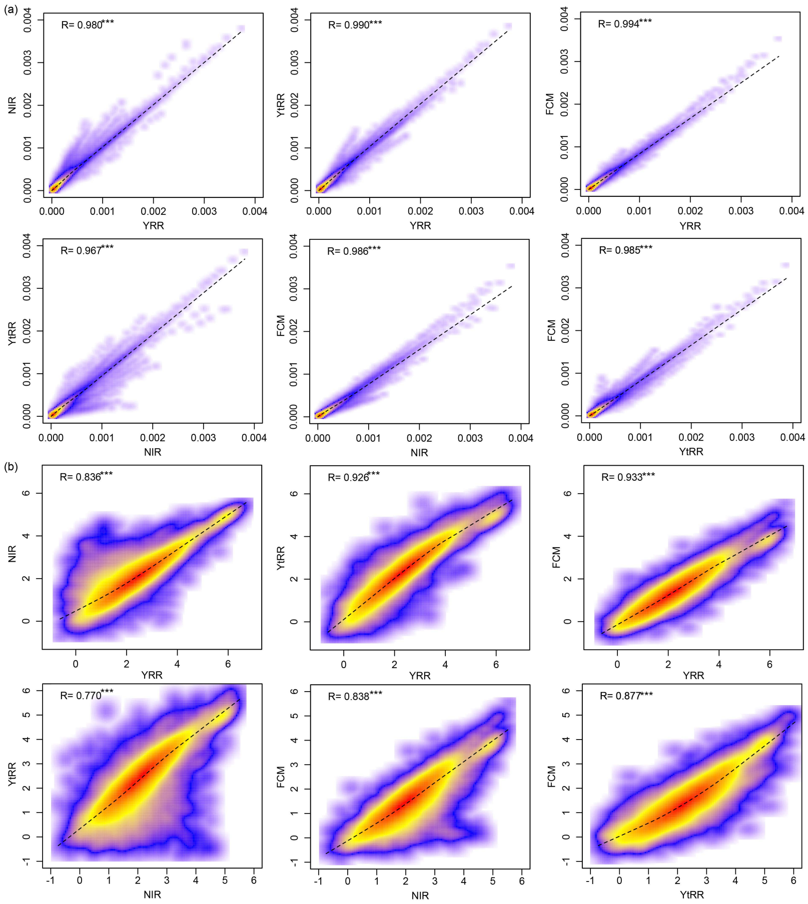

Figure 5. Distribution and pairwise correlation of nucleotide diversity (**a**) and Tajima's D (**b**) of geographic sources. The red−yellow−blue color gradient means decreased density in the graph. *** means p-value $\leq 2.2 \times 10^{-16}$.

3.4. Meta Genome Wide Association Study for Fiber Traits

To further analyze the relationship between introgression and cotton genetic improvement, we conducted a meta-GWAS using 1,427,171 high−quality SNPs (−log (0.05/1,427,171); 7.39) and identified 512 significantly SNPs loci that were significantly associated with fiber traits (Figure 6). Finally, a total of 261 fiber-related QTLs were located, including 184 for fiber elongation (FE), 1 for FL, 2 for FM, 2 for FU and 113 for FS. We found that some QTLs had fewer candidate genes (Table S4). Further comparisons showed that there were 42 QTLs with FS and FE co-located on A01, A03, A04, A05, A06, A07, A08, A09, A10, A11, A12, D01, D02, D04, D07, D08, D10 and D11, 1 QTLs with FL and FE co-located on D11 and 1 QTLs with FS and FU co-located on A02 (Table S4). The tissue expression patterns and functional annotations of 4,423 candidate genes located in 261 QTLs regions were shown in Table S5. The QTL region for FL and FE has 19 candidate coding genes, in which the *Ghir_D11G020360* encoding AP2/B3-like transcriptional factor was involved in cotton fiber development as a candidate gene (Tables S4 and S6). A candidate gene related to fiber elongation, *Ghir_D05G013680*, encoding indeterminate (ID)-domain 7, showed different expression patterns in different fiber development stages (Tables S4 and S6). These results provide new QTL candidates resources that may be applied to cultivate desirable agronomic traits in future genomics—guided cotton breeding.

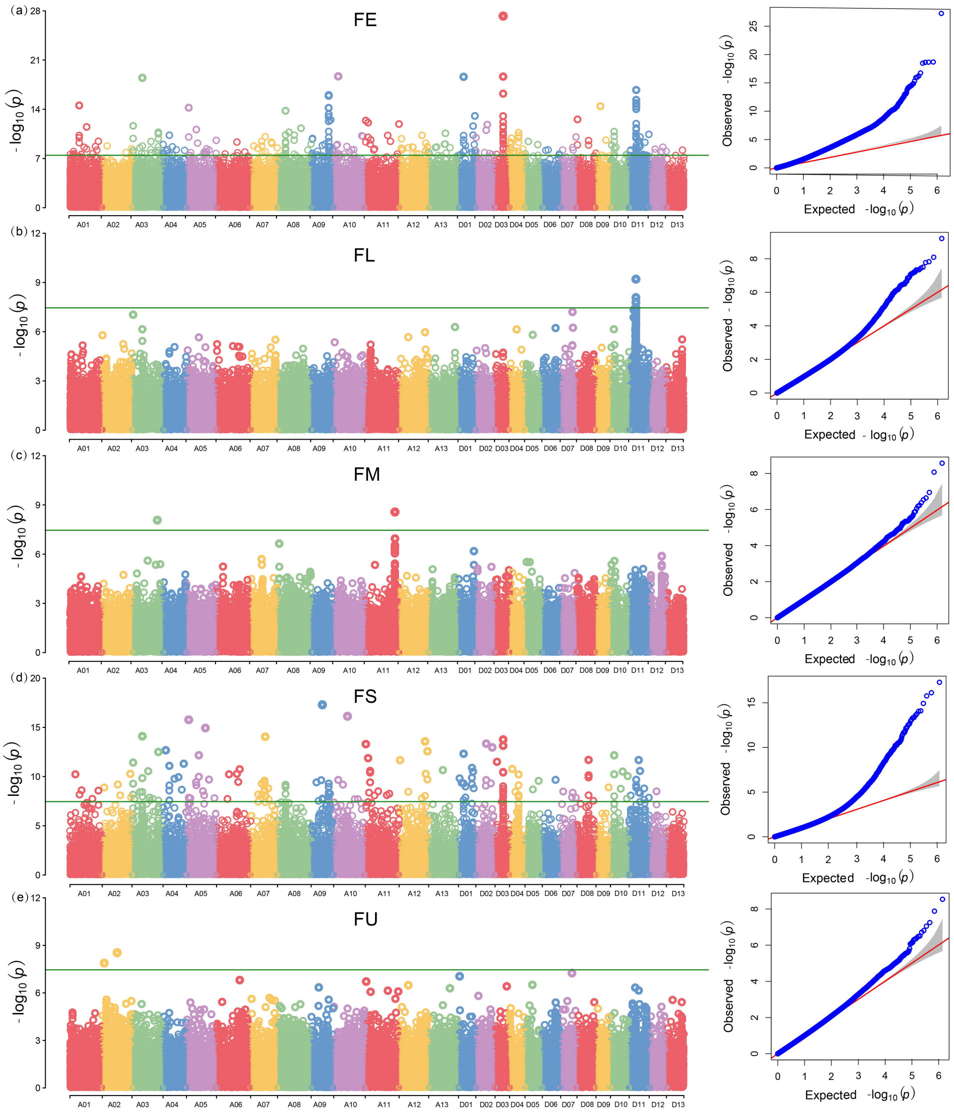

Figure 6. The results of genome-wide association study for fiber trait. Manhattan plots and quantile-quantile (Q-Q) plot for FE (**a**), FL (**b**), FM (**c**), FS (**d**) and FU (**e**).

3.5. Intraspecific Introgression and Selection Accelerate Genetic Improvement of Cotton

Among the 261 QTLs, 67 intersected with introgression signals, totaling 118.81 Mb, containing 2939 candidate genes (Figure 7 and Table S6), and 123 intersected with selection signals, totaling 28.38 Mb, containing 879 candidate genes (Figure 7 and Table S7). For example, a FS association with introgression signal was identified on A01, a candidate gene *Ghir_A01G011950*, encoding UDP-Glycosyltransferase superfamily protein. Another example shows the QTL23 related to FS association with selection signal was located on A02 among the YRR_vs_YtRR, YRR_vs_FCM and YtRR_vs_FCM, and a gene *Ghir_A02G015550*, encoding xyloglucan endotransglycosylase 6 was identified as the candidate gene. These results indicated that intraspecific introgression and selection contribute the genetic improvement of cotton.

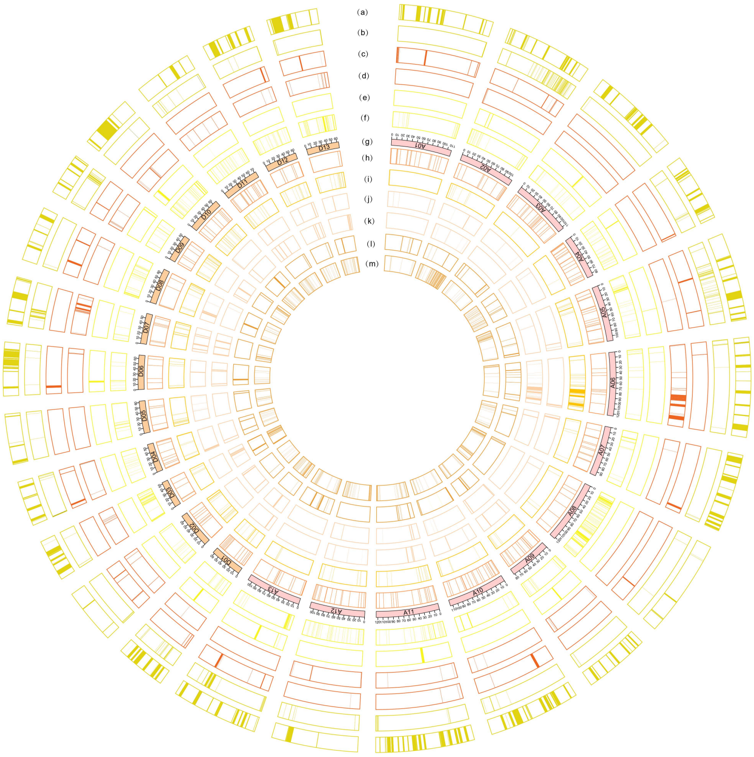

Figure 7. Genome-wide landscape of significant introgression signals, selection signal intervals and QTLs. Distribution of significant introgression signals (**a**) and selection signal intervals (**b**) between YRR and YtRR. Distribution of significant introgression signals (**c**) and selection signal intervals (**d**) between YRR and NIR. Distribution of significant introgression signals (**e**) and selection signal intervals (**f**) between YRR and FCM. (**g**) The chromosome. (**h**) Distribution of QTLs. (**i**) Distribution of selection signal intervals between NIR and FCM. Distribution of significant introgression signals (**j**) and selection signal intervals (**k**) between NIR and YtRR. Distribution of significant introgression signals (**l**) and selection signal intervals (**m**) between YtRR and FCM.

4. Discussion

Upland cotton produces 95% of the world's cotton fiber [19]. Here, we compared and characterized genome wide patterns of genetic diversity, population divergence and phylogenetic relationship in different geographical distribution of upland cotton. The NIR region demonstrated the highest genetic diversity compared with other regions (Figure 1e), which differs from previous studies [22] and may be due to the increased amount of material. In addition, it was found that there were strong correlations of genetic diversity among groups (Figure 5a), which is consistent with previous findings [22]. Compared to genetic diversity, we observed a weaker correlations of Tajima's D values (Figure 5b). The correlation coefficient is higher than that of the previous article [22], which may be caused by the difference in the interval of 100 Kb and the number of materials. Compared to the *Populus*, the correlation coefficient R was higher than that in three species of *Populus* [51]. These findings indicated that the modern cultivars showed a low level of genetic differentiation among different geographical distribution. Additionally, phylogenetic analysis reveals that these materials do not differ strictly according to geographical distribution (Figure 1a,c). The reason for this is that breeders developed new varieties based on introduced materials from America in the early days of cotton breeding in China, and later on, breeders combined excellent materials from different regions to produce new varieties.

Genomic big data reveals that introgression is universal and frequently facilitates adaptation and diversification in nature [52–55]. Previous introgression studies have typically focused on large-scale pairs of individual species [56] and sympatric related species [57], while studies of distinct geographic distributions within species are rarely studied. Recently, several studies have mainly focused on the identification of introgression between sea island cotton and upland cotton to improve upland cotton [47,58–60]. Here we analyzed and characterized the genome-wide introgression characteristics between different geographical sources of upland cotton, and found that there were more introgression intervals between YRR and YtRR, indicating that there were more introductions of YRR and YtRR in the process of upland cotton breeding. No significant introgression signal was found between FCM and NIR, suggesting that foreign-introduced materials may be mainly introduced and bred new varieties in the YRR and YtRR. Domestication and genetic improvement are important processes in cotton breeding. Thousands of years of domestication and artificial selection have altered cotton morphology, including growth cycle, photoperiod and fiber quality [61,62]. We detected genomic regions that have been subject to selection as inferred from different geographic distribution. We found that the largest genomic regions were under selection before.

The improvement of cotton fiber quality has always been one of the most important cotton breeding goals. In recent years, GWAS based on the high-throughput sequencing technology has been become a powerful method to identify the selection signals that control the complex agriculture traits [63]. GWASs were recently applied to 267, 258 and 419 diverse accessions of upland cotton and 230 *G. arboreum* (A_2) accessions, and these results demonstrated that the combination of snp-based GWAS and functional genomics is an integrated approach to map complex agriculture traits to underlying causal genetic variation [14,15,27,64]. Until now, favorable QTL alleles for agronomic traits such as yield and fiber quality have been identified via GWAS in upland cotton [14,15,22,27,65]. We overlapped introgression, selection sweep signals with the locations of 261 QTLs and found that 67 QTLs had introgression signals, and the genome interval size was 118.81 Mb, including 2939 candidate genes (Table S6), 123 QTLs having selection signals and the genomic interval being 28.38 Mb (Table S7), containing 879 candidate genes. For example, the gene *Ghir_D05G013680* obtained in this study has been confirmed to be related to fiber development, which also shows the reliability of our results [66]. These results imply that environmental divergence may be the major determinant of cotton introgression.

5. Conclusions

Introgression plays an important role in adaptation to various environmental changes during plant domestication and variety formation. Here, we examined a detailed investigation of the extent of introgression that existed among four geographic distributions of upland cotton and its effect on selection and fiber quality, using whole-genome sequencing data form large-scale individuals. We found that introgression exists in different geographic distributions. Meta-GWAS and selective sweeps revealed that introgression was subject to artificial selection. In addition, introgression is affected by the similarity of their living environment, and materials with similar ecological environments tend to share the same introgression area. These results provide a useful reference for studying intraspecific introgressions from different geographical distributions in other species.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy13071781/s1, Figure S1: The Distribution of SNPs on all chromosomes with 1 Mb window. Figure S2. The distribution and correlation of the five fiber traits. Figure S3. Genome-wide distribution of introgression signals between YtRR and YRR. Figure S4. Genome-wide distribution of introgression signals between FCM and YRR. Figure S5. Genome-wide distribution of introgression signals between YtRR and NIR. Figure S6. Genome-wide distribution of introgression signals between FCM and NIR. Figure S7. Genome-wide distribution of introgression signals between FCM and YtRR. Table S1. Summary of genomic sequencing data of 890 cottons in this study. Table S2. The significant introgression signals containing genes in different geographical distributions. Table S3. The genetic improvement signals and covering genes in different geographical distributions. Table S4. GWAS QTL regions information. Table S5. Expression and functional annotation of GWAS candidate genes. Table S6. Summary of QTLs that overlapped with significance introgression signals. Table S7. Summary of QTLs that overlapped with genetic development regions.

Author Contributions: Conceptualization, C.S.; methodology, C.S. and Z.C.; investigation, Z.X., L.O., X.Z. and W.H.; data curation, C.S., Z.G., J.Y. and R.C.; writing—original draft preparation, C.S.; writing—review and editing, C.S. and L.O.; and funding acquisition, C.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the National Natural Science Foundation of China (32201873), the Guangdong Basic and Applied Basic Research Foundation (2019A1515110288) and the Projects of Talents Recruitment of Guangdong University of Petrochemical Technology (2019rc112).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dreissig, S.; Maurer, A.; Sharma, R.; Milne, L.; Flavell, A.J.; Schmutzer, T.; Pillen, K. Natural variation in meiotic recombination rate shapes introgression patterns in intraspecific hybrids between wild and domesticated barley. *New Phytol.* **2020**, *228*, 1852–1863. [CrossRef] [PubMed]
2. Mallet, J.; Besansky, N.; Hahn, M.W. How reticulated are species? *BioEssays* **2016**, *38*, 140–149. [CrossRef] [PubMed]
3. Jiang, N.; Li, Z.; Dai, Y.; Liu, Z.; Han, X.; Li, Y.; Li, Y.; Xiong, H.; Xu, J.; Zhang, G.; et al. Massive genome investigations reveal insights of prevalent introgression for environmental adaptation and triterpene biosynthesis in *Ganoderma*. *Mol. Ecol. Resour.* **2022**, 1–18. [CrossRef] [PubMed]
4. Imaizumi, T.; Kawahara, Y.; Auge, G. Hybrid-derived weedy rice maintains adaptive combinations of alleles associated with seed dormancy. *Mol. Ecol.* **2022**, *31*, 6556–6569. [CrossRef] [PubMed]
5. Huerta-Sánchez, E.; Jin, X.; Asan; Bianba, Z.; Peter, B.M.; Vinckenbosch, N.; Liang, Y.; Yi, X.; He, M.; Somel, M.; et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **2014**, *512*, 194–197. [CrossRef] [PubMed]
6. Zhang, X.; Kim, B.; Singh, A.; Sankararaman, S.; Durvasula, A.; Lohmueller, K.E. MaLAdapt Reveals novel targets of adaptive introgression from neanderthals and denisovans in worldwide human populations. *Mol. Biol. Evol.* **2023**, *40*, msad001. [CrossRef]
7. Dasmahapatra, K.K.; Walters, J.R.; Briscoe, A.D.; Davey, J.W.; Whibley, A.; Nadeau, N.J.; Zimin, A.V.; Hughes, D.S.T.; Ferguson, L.C.; Martin, S.H.; et al. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **2012**, *487*, 94–98.
8. Lamichhaney, S.; Berglund, J.; Almén, M.S.; Maqbool, K.; Grabherr, M.; Martinez-Barrio, A.; Promerová, M.; Rubin, C.-J.; Wang, C.; Zamani, N.; et al. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **2015**, *518*, 371–375. [CrossRef]
9. Hufford, M.B.; Xu, X.; van Heerwaarden, J.; Pyhäjärvi, T.; Chia, J.-M.; Cartwright, R.A.; Elshire, R.J.; Glaubitz, J.C.; Guill, K.E.; Kaeppler, S.M.; et al. Comparative population genomics of maize domestication and improvement. *Nat. Genet.* **2012**, *44*, 808–811. [CrossRef]
10. He, F.; Pasam, R.; Shi, F.; Kant, S.; Keeble-Gagnere, G.; Kay, P.; Forrest, K.; Fritz, A.; Hucl, P.; Wiebe, K.; et al. Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. *Nat. Genet.* **2019**, *51*, 896–904. [CrossRef]
11. Martin, S.H.; Jiggins, C.D. Interpreting the genomic landscape of introgression. *Curr. Opin. Genet. Dev.* **2017**, *47*, 69–74. [CrossRef] [PubMed]
12. Fu, R.; Zhu, Y.; Liu, Y.; Feng, Y.; Lu, R.S.; Li, Y.; Li, P.; Kremer, A.; Lascoux, M.; Chen, J. Genome-wide analyses of introgression between two sympatric Asian oak species. *Nat. Ecol. Evol.* **2022**, *6*, 924–935. [CrossRef] [PubMed]
13. Fang, L.; Gong, H.; Hu, Y.; Liu, C.; Zhou, B.; Huang, T.; Wang, Y.; Chen, S.; Fang, D.D.; Du, X.; et al. Genomic insights into divergence and dual domestication of cultivated allotetraploid cottons. *Genome Biol.* **2017**, *18*, 33. [CrossRef] [PubMed]
14. Fang, L.; Wang, Q.; Hu, Y.; Jia, Y.; Chen, J.; Liu, B.; Zhang, Z.; Guan, X.; Chen, S.; Zhou, B.; et al. Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat. Genet.* **2017**, *49*, 1089–1098. [CrossRef] [PubMed]
15. Ma, Z.; He, S.; Wang, X.; Sun, J.; Zhang, Y.; Zhang, G.; Wu, L.; Li, Z.; Liu, Z.; Sun, G.; et al. Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. *Nat. Genet.* **2018**, *50*, 803–813. [CrossRef] [PubMed]
16. He, S.; Sun, G.; Geng, X.; Gong, W.; Dai, P.; Jia, Y.; Shi, W.; Pan, Z.; Wang, J.; Wang, L.; et al. The genomic basis of geographic differentiation and fiber improvement in cultivated cotton. *Nat. Genet.* **2021**, *53*, 916–924. [CrossRef]
17. Hu, G.; Grover, C.E.; Jareczek, J.; Yuan, D.; Dong, Y.; Miller, E.; Conover, J.L.; Wendel, J.F. Evolution and Diversity of the Cotton Genome. In *Cotton Precision Breeding*; Rahman, M.-U., Zafar, Y., Zhang, T., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 25–78.
18. Wang, M.; Li, J.; Qi, Z.; Long, Y.; Pei, L.; Huang, X.; Grover, C.E.; Du, X.; Xia, C.; Wang, P.; et al. Genomic innovation and regulatory rewiring during evolution of the cotton genus *Gossypium*. *Nat. Genet.* **2022**, *54*, 1959–1971. [CrossRef]

19. Tyagi, P.; Gore, M.A.; Bowman, D.T.; Campbell, B.T.; Udall, J.A.; Kuraparthy, V. Genetic diversity and population structure in the US Upland cotton (*Gossypium hirsutum* L.). *Theor. Appl. Genet.* **2014**, *127*, 283–295. [CrossRef]
20. Wendel, J.F.; Brubaker, C.L.; Percival, A.E. Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. *Am. J. Bot.* **1992**, *79*, 1291–1310. [CrossRef]
21. Abdurakhmonov, I.Y.; Saha, S.; Jenkins, J.N.; Buriev, Z.T.; Shermatov, S.E.; Scheffler, B.E.; Pepper, A.E.; Yu, J.Z.; Kohel, R.J.; Abdukarimov, A. Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* L. variety germplasm. *Genetica* **2009**, *136*, 401–417. [CrossRef]
22. Huang, C.; Nie, X.; Shen, C.; You, C.; Li, W.; Zhao, W.; Zhang, X.; Lin, Z. Population structure and genetic basis of the agronomic traits of upland cotton in China revealed by a genome-wide association study using high-density SNPs. *Plant Biotechnol. J.* **2017**, *15*, 1374–1386. [CrossRef] [PubMed]
23. Shen, C.; Wang, N.; Huang, C.; Wang, M.; Zhang, X.; Lin, Z. Population genomics reveals a fine-scale recombination landscape for genetic improvement of cotton. *Plant J.* **2019**, *99*, 494–505. [CrossRef]
24. Campbell, B.T.; Saha, S.; Percy, R.; Frelichowski, J.; Jenkins, J.N.; Park, W.; Mayee, C.D.; Gotmare, V.; Dessauw, D.; Giband, M.; et al. Status of the global cotton germplasm resources. *Crop Sci.* **2010**, *50*, 1161–1179. [CrossRef]
25. Jia, G.; Huang, X.; Zhi, H.; Zhao, Y.; Zhao, Q.; Li, W.; Chai, Y.; Yang, L.; Liu, K.; Lu, H.; et al. A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nat. Genet.* **2013**, *45*, 957–961. [CrossRef]
26. 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. *Nature Biotechnol.* **2015**, *33*, 408–414. [CrossRef]
27. Wang, M.; Tu, L.; Lin, M.; Lin, Z.; Wang, P.; Yang, Q.; Ye, Z.; Shen, C.; Li, J.; Zhang, L.; et al. Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nat. Genet.* **2017**, *49*, 579–587. [CrossRef]
28. Zhang, T.-T.; Zhang, N.-Y.; Li, W.; Zhou, X.-J.; Pei, X.-Y.; Liu, Y.-G.; Ren, Z.-Y.; He, K.-L.; Zhang, W.-S.; Zhou, K.-H.; et al. Genetic structure, gene flow pattern, and association analysis of superior germplasm resources in domesticated upland cotton (*Gossypium hirsutum* L.). *Plant Divers.* **2020**, *42*, 189–197. [CrossRef]
29. May, O.L.; Bowman, D.T.; Calhoun, D.S. Genetic Diversity of U.S. Upland cotton cultivars released between 1980 and 1990. *Crop Sci.* **1995**, *35*, 1570–1574. [CrossRef]
30. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [CrossRef]
31. Leinonen, R.; Sugawara, H.; Shumway, M.; International Nucleotide Sequence Database Collaboration. The Sequence Read Archive. *Nucleic Acids Res.* **2010**, *39*, 19–21. [CrossRef]
32. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
33. Wang, M.; Tu, L.; Yuan, D.; Zhu, D.; Shen, C.; Li, J.; Liu, F.; Pei, L.; Wang, P.; Zhao, G.; et al. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nat. Genet.* **2019**, *51*, 224–229. [CrossRef] [PubMed]
34. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef] [PubMed]
35. 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*, 1297–1303. [CrossRef]
36. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Subgroup, G.P.D.P. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]
37. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **2011**, *27*, 2987–2993. [CrossRef]
38. Browning, B.L.; Zhou, Y.; Browning, S.R. A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* **2018**, *103*, 338–348. [CrossRef]
39. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [CrossRef]
40. Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **2011**, *88*, 76–82. [CrossRef]
41. Zhang, C.; Dong, S.S.; Xu, J.Y.; He, W.M.; Yang, T.L. PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **2019**, *35*, 1786–1788. [CrossRef]
42. Lee, T.-H.; Guo, H.; Wang, X.; Kim, C.; Paterson, A.H. SNPhylo: A pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genom.* **2014**, *15*, 162. [CrossRef]
43. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef]
44. Li, M.; Tian, S.; Jin, L.; Zhou, G.; Li, Y.; Zhang, Y.; Wang, T.; Yeung, C.K.; Chen, L.; Ma, J.; et al. Genomic analyses identify distinct patterns of selection in domesticated pigs and *Tibetan* wild boars. *Nat. Genet.* **2013**, *45*, 1431–1438. [CrossRef]

45. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [CrossRef] [PubMed]
46. Browning, B.L.; Browning, S.R. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **2013**, *194*, 459–471. [CrossRef] [PubMed]
47. Fang, L.; Zhao, T.; Hu, Y.; Si, Z.; Zhu, X.; Han, Z.; Liu, G.; Wang, S.; Ju, L.; Guo, M.; et al. Divergent improvement of two cultivated allotetraploid cotton species. *Plant Biotechnol. J.* **2021**, *19*, 1325–1336. [CrossRef] [PubMed]
48. Wu, D.D.; Ding, X.D.; Wang, S.; Wójcik, J.M.; Zhang, Y.; Tokarska, M.; Li, Y.; Wang, M.S.; Faruque, O.; Nielsen, R.; et al. Pervasive introgression facilitated domestication and adaptation in the Bos species complex. *Nat. Ecol. Evol.* **2018**, *2*, 1139–1145. [CrossRef]
49. Wang, J.; Zhang, Z. GAPIT Version 3: Boosting power and accuracy for genomic association and prediction. *Genom. Proteom. Bioinf.* **2021**, *19*, 629–640. [CrossRef]
50. Zhang, Z.; Ersoz, E.; Lai, C.-Q.; Todhunter, R.J.; Tiwari, H.K.; Gore, M.A.; Bradbury, P.J.; Yu, J.; Arnett, D.K.; Ordovas, J.M.; et al. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **2010**, *42*, 355–360. [CrossRef] [PubMed]
51. Wang, J.; Street, N.R.; Scofield, D.G.; Ingvarsson, P.K. Natural selection and recombination rate variation shape nucleotide polymorphism across the genomes of three related populus species. *Genetics* **2015**, *202*, 1185–1200. [CrossRef]
52. Arnold, M.L.; Kunte, K. Adaptive Genetic Exchange: A tangled history of admixture and evolutionary innovation. *Trends Ecol. Evol.* **2017**, *32*, 601–611. [CrossRef]
53. Taylor, S.A.; Larson, E.L. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nat. Ecol. Evol.* **2019**, *3*, 170–177. [CrossRef]
54. Edelman, N.B.; Mallet, J. Prevalence and adaptive impact of introgression. *Ann. Rev. Genet.* **2021**, *55*, 265–283. [CrossRef]
55. Wang, Y.; Wang, Y.; Cheng, X.; Ding, Y.; Wang, C.; Merilä, J.; Guo, B. Prevalent introgression underlies convergent evolution in the diversification of pungitius sticklebacks. *Mol. Biol. Evol.* **2023**, *40*, msad026. [CrossRef] [PubMed]
56. Leroy, T.; Roux, C.; Villate, L.; Bodénès, C.; Romiguier, J.; Paiva, J.A.; Dossat, C.; Aury, J.M.; Plomion, C.; Kremer, A. Extensive recent secondary contacts between four European white oak species. *New Phytol.* **2017**, *214*, 865–878. [CrossRef]
57. Beatty, G.E.; Montgomery, W.I.; Spaans, F.; Tosh, D.G.; Provan, J. Pure species in a continuum of genetic and morphological variation: Sympatric oaks at the edge of their range. *Ann. Bot.* **2016**, *117*, 541–549. [CrossRef] [PubMed]
58. Wang, P.; Dong, N.; Wang, M.; Sun, G.; Jia, Y.; Geng, X.; Liu, M.; Wang, W.; Pan, Z.; Yang, Q.; et al. Introgression from *Gossypium hirsutum* is a driver for population divergence and genetic diversity in *Gossypium barbadense*. *Plant J.* **2022**, *110*, 764–780. [CrossRef] [PubMed]
59. He, S.; Wang, P.; Zhang, Y.-M.; Dai, P.; Nazir, M.F.; Jia, Y.; Peng, Z.; Pan, Z.; Sun, J.; Wang, L.; et al. Introgression leads to genomic divergence and responsible for important traits in upland cotton. *Front. Plant Sci.* **2020**, *11*, 929. [CrossRef]
60. Nie, X.; Wen, T.; Shao, P.; Tang, B.; Nuriman-guli, A.; Yu, Y.; Du, X.; You, C.; Lin, Z. High-density genetic variation maps reveal the correlation between asymmetric interspecific introgressions and improvement of agronomic traits in Upland and Pima cotton varieties developed in Xinjiang, China. *Plant J.* **2020**, *103*, 677–689. [CrossRef]
61. Rapp, R.A.; Haigler, C.H.; Flagel, L.; Hovav, R.H.; Udall, J.A.; Wendel, J.F. Gene expression in developing fibres of upland cotton (*Gossypium hirsutum* L.) was massively altered by domestication. *BMC Biol.* **2010**, *8*, 139. [CrossRef]
62. Yoo, M.-J.; Wendel, J.F. Comparative evolutionary and developmental dynamics of the cotton (*Gossypium hirsutum*) fiber transcriptome. *PLOS Genet.* **2014**, *10*, e1004073. [CrossRef]
63. Zhou, H.; Li, P.; Xie, W.; Hussain, S.; Li, Y.; Xia, D.; Zhao, H.; Sun, S.; Chen, J.; Ye, H.; et al. Genome-wide association analyses reveal the genetic basis of stigma exsertion in Rice. *Mol. Plant.* **2017**, *10*, 634–644. [CrossRef]
64. Du, X.; Huang, G.; He, S.; Yang, Z.; Sun, G.; Ma, X.; Li, N.; Zhang, X.; Sun, J.; Liu, M.; et al. Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nat. Genet.* **2018**, *50*, 796–802. [CrossRef] [PubMed]
65. Sun, Z.; Wang, X.; Liu, Z.; Gu, Q.; Zhang, Y.; Li, Z.; Ke, H.; Yang, J.; Wu, J.; Wu, L.; et al. Genome-wide association study discovered genetic variation and candidate genes of fibre quality traits in *Gossypium hirsutum* L. *Plant Biotechnol. J.* **2017**, *15*, 982–996. [CrossRef] [PubMed]
66. Li, J.; Yuan, D.; Wang, P.; Wang, Q.; Sun, M.; Liu, Z.; Si, H.; Xu, Z.; Ma, Y.; Zhang, B.; et al. Cotton pan-genome retrieves the lost sequences and genes during domestication and selection. *Genome Biol.* **2021**, *22*, 119. [CrossRef] [PubMed]