



Article

Genetic Diversity Analysis of Cotton Cultivars Using a 40K Liquid Chip in Northern Xinjiang

Zhihong Zheng ^{1,†}, Ningshan Wang ^{1,†}, Shangkun Jin ², Kewei Ning ¹, Guoli Feng ¹, Haiqiang Gao ¹, Zhanfeng Si ², Tianzhen Zhang ² and Nijiang Ai ^{1,*}

¹ Shihezi Academy of Agricultural Sciences, Shihezi 832000, China

² Modern Seed Industry Research Institute, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China

* Correspondence: xjshznykxyjy@163.com

[†] These authors contributed equally to this work.

Abstract

Genetic diversity and kinship information of cotton germplasm resources are fundamental to breeding, providing a theoretical basis for the rational selection of hybrid parents and further breeding of new varieties with high yield, high quality, and multi-resistance. This study utilized cotton varieties that have been used for variety improvement or are widely planted in the Northern Xinjiang cotton region as materials. Genotyping was performed using the ZJU CottonSNP40K chip to analyze genetic diversity and kinship relationships. A total of 26,852 high-quality SNP markers were obtained, including 15,222 SNPs in subgenome A and 11,630 SNPs in subgenome D. The number of SNPs per chromosome ranged from 547 (A04) to 2168 (A08). Based on phylogenetic tree and principal component analysis, the 83 materials were clustered into 3 major subgroups. Group I contained varieties introduced from the former Soviet Union and the United States, which have become important parents for cotton breeding in Northern Xinjiang. Among them, as many as 27 varieties were derived and selected from the introduced US variety ‘Beiersinuo’ as a parent. While playing an important role in cotton breeding in Northern Xinjiang, this has also led to the current situation where the genetic base of Northern Xinjiang varieties is relatively narrow (average kinship coefficient 0.72). It clarifies the significant role of introduced American variety ‘Beiersinuo’ in the breeding of Northern Xinjiang cultivars and provides theoretical guidance for broadening the genetic base of Northern Xinjiang cotton varieties.

Keywords: ZJU CottonSNP40K; Northern Xinjiang cotton varieties; genetic relationship



Academic Editor: Nunzio D’Agostino

Received: 5 November 2025

Revised: 10 December 2025

Accepted: 10 December 2025

Published: 5 January 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Xinjiang is China’s most important high-quality commercial cotton production base. Based on climatic conditions, the Xinjiang cotton region is generally divided into the Northern Xinjiang early-maturing cotton region and the Southern Xinjiang early-to-medium-maturing cotton region. The Northern Xinjiang early-maturing cotton region holds an extremely important position in Xinjiang’s high-quality cotton production, accounting for 30% of the total planting area and over 32% of the total product in Xinjiang. The Northern Xinjiang cotton production area boasts superior farmland infrastructure, high mechanization levels, advanced irrigation techniques, and high cultivation management standards, with a favorable market for the cotton industry. Over the years, the “Xinluzao”

series varieties have become the dominant cultivars in this region across different periods due to their good early maturity, high yield, and stability, making outstanding contributions to cotton production development in Northern Xinjiang. Analyzing the genetic diversity and kinship relationships of cotton cultivars is of great significance for promoting the breeding of new cotton varieties suitable for the Northern Xinjiang cotton region with early maturity, high yield, superior quality, and multi-resistance.

In-depth research on the genetic diversity and kinship relationships of cotton germplasm resources is the foundation of crop breeding and variety improvement. Through systematic classification and genetic structure analysis of cotton germplasm resources, their genetic variation and gene flow can be effectively evaluated, thereby providing a scientific basis for germplasm innovation and new variety breeding [1–3]. Domestic and international scholars have conducted detailed research on cotton variety diversity using morphological traits, isozyme markers, and molecular marker technologies [4–9]. Single-nucleotide polymorphism (SNP), the most common form of variation in the genomes of many species with a high distribution rate in the genome, is widely used in genetic map construction, genetic diversity analysis, variety identification, and molecular marker-assisted breeding in crops [1,2,10–13]. With the continuous development of next-generation sequencing technologies and the gradual refinement of SNP marker technology, corresponding SNP detection techniques have also evolved. Among them, SNP chips, as a detection method integrating advantages such as high throughput, miniaturization, and automation, have been widely applied in population genetics, QTL mapping, and candidate gene screening in various plants [14–17]. For example, a genetic linkage map, constructed using the GBW16K array-based genotyping of a recombinant inbred line population derived from a cross of the CIMMYT wheat line Yaco “S” and the Chinese landrace Mingxian169, enables the identification of resistant candidate genes [18]. Zhejiang University developed the “ZJU CottonSNP40K” cotton chip that has been widely used in cotton biological breeding [19], for example, in constructing ultra-high-density genetic maps [19] and cotton diversity analysis [20].

This study utilizes the high-coverage “ZJU CottonSNP40K” cotton chip [19] to analyze the genetic diversity of 83 backbone materials for cotton variety improvement and main cultivars in the Northern Xinjiang cotton region. Combined with population genetics methods, it explores their kinship relationships, providing a theoretical basis for the rational selection of hybrid parents.

2. Results

2.1. The SNP Distribution Characteristics in the Northern Xinjiang Cotton Population

After filtering, 83 *G. hirsutum* cultivars, including five varieties introduced from the USA and the former Soviet Union (108Φ, 611B, C1470, KK1543, and Beiersinuo), and 78 cotton varieties from Northern Xinjiang (mainly the Xinluzao series varieties), yielded a total of 26,852 high-quality SNP markers, with A subgenome containing 15,222 SNPs and D subgenome containing 11,630 SNPs. The number of SNPs per chromosome ranged from 547 (A04) to 2168 (A08) (Table 1). The SNP density was 12.4 per Mb, relatively evenly distributed across each chromosome (Table 1, Figure 1). The maximum density was 18.4 per Mb (D09), and the minimum density was 5.8 per Mb (A02).

Table 1. Distribution of SNPs across different chromosomes.

Chr	Length (bp)	SNP Number	Chr	Length (bp)	SNP Number
A01	118,174,371	1610	D01	64,698,102	1118
A02	108,272,889	625	D02	69,777,850	1003

Table 1. Cont.

Chr	Length (bp)	SNP Number	Chr	Length (bp)	SNP Number
A03	111,586,618	792	D03	53,896,199	785
A04	87,703,368	547	D04	56,935,404	555
A05	110,845,161	1123	D05	63,929,679	890
A06	126,488,190	1323	D06	65,459,843	1207
A07	96,598,283	1625	D07	58,417,686	837
A08	125,056,055	2168	D08	69,080,421	1052
A09	83,216,487	985	D09	52,000,373	959
A10	115,096,118	947	D10	66,881,427	877
A11	121,376,521	1372	D11	71,358,197	754
A12	107,588,319	867	D12	61,693,100	852
A13	110,367,549	1238	D13	64,447,585	741



Figure 1. The density of SNPs across the chromosomes. The data is shown in 1 Mb sliding windows.

2.2. Genetic Diversity in Northern Xinjiang Cotton Cultivars

To investigate the genetic relationships among these materials, genetic distances between every two of the 83 materials were calculated using Phylip software to construct a phylogenetic tree. The results showed that these materials could be divided into three subgroups (Figure 2A and Table S1). Group I contained 31 materials: the introduced early-maturing materials from the USSR (108Φ, 611B, C1470, KK1543), the US early-maturing material Beiersinuo, and varieties primarily bred by units such as the Agricultural Science Institute of the 7th Division, i.e., the first batch of early varieties bred using foreign varieties as foundational parents. Group II contained 27 materials, primarily bred by Shihezi Academy of Agricultural Sciences and Hexin Seed Industry. Group III contained 25 materials, primarily selected and bred by Xinjiang Academy of Agricultural and Reclamation Sciences, Huiyuan Seed Industry, etc. The results of principal component analysis (PCA) and population structure analysis were consistent with the phylogenetic tree results (Figure 2B,C), confirming the accuracy of this classification. Analysis of the kinship matrix revealed that the average kinship coefficient among these 83 materials reached 0.72. Specifically, the average kinship coefficient was 0.72 for Group I, 0.75 for Group II, and 0.77 for Group III. These results indicate that cotton varieties in Northern Xinjiang have a narrow genetic base and low genetic diversity.

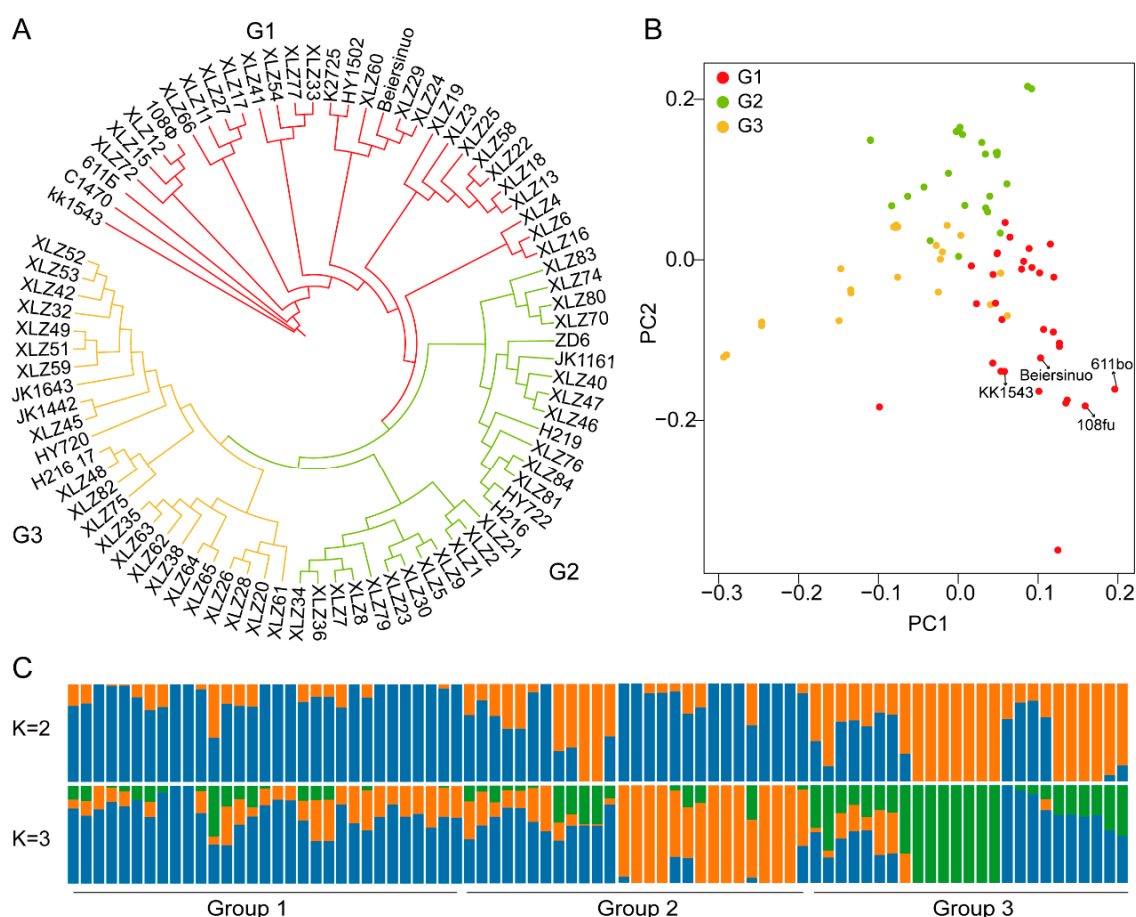


Figure 2. Clustering and population structure of cotton varieties in Northern Xinjiang. (A) Phylogenetic tree of early maturing varieties, different colors represent different groups (G1 to G3) (B) Principal component analysis. Different colors represent different groups. Arrows indicate important foreign-introduced parental lines. (C) Structure analysis with K = 2 and K = 3. The x-axis represents the different accessions. The orders and positions of accessions are consistent with those in the phylogenetic tree when K = 3.

2.3. Genetic Basis of Northern Xinjiang Cotton Cultivar Improvement

Annotation of these SNPs revealed that 18,363 SNP variants occurred in intergenic regions, 1988 SNPs occurred in exonic regions, 2147 SNPs occurred in introns, 2131 SNPs occurred in gene promoters, and 1854 SNPs occurred in downstream regulatory regions (Figure 3A). This indicated that variants occurring in gene regions and regulatory regions are relatively few, but these sites are highly likely to contribute to trait differences among different varieties. Further filtering of SNPs occurring in exons (filtering steps detailed in methods) yielded 27 non-synonymous mutation type SNPs. These SNPs exhibited significantly different distribution frequencies among the different clustered groups (Table S2), suggesting that these variant sites may represent points of population differentiation, breeding selection, or specific variation. For example, the SNPs carried by *GH_D05G2763* and *GH_D07G1031* both had higher distribution frequencies in Group III (Figure 3B), indicating that these variations arose during later breeding processes. *GH_D05G2763* harbored a non-synonymous mutation (C to T) in the first exon, causing an amino acid change from Leucine to Phenylalanine (Table S2). This gene encodes a Leucine-rich repeat receptor-like protein kinase (LRR-RLK), expressed throughout various developmental stages (Figure 3C). The *GH_D07G1031* gene harbored a non-synonymous mutation (T to A) in an exon, causing an amino acid change from Valine to Glutamic Acid (Figure 3D). This gene also encodes a receptor protein kinase (hercules receptor kinase), highly expressed in 1 DPA at ovules and 5 DPA and 10 DPA at fiber tissues (Figure 3D), suggesting it may play a key role in fiber initiation and elongation. Receptor kinases have been reported to play important roles in regulating plant reproductive growth, immune responses, and abiotic stress responses [21–23]. Therefore, these kinase proteins carrying variant sites may contribute to phenotypic variation in Xinjiang cotton varieties.

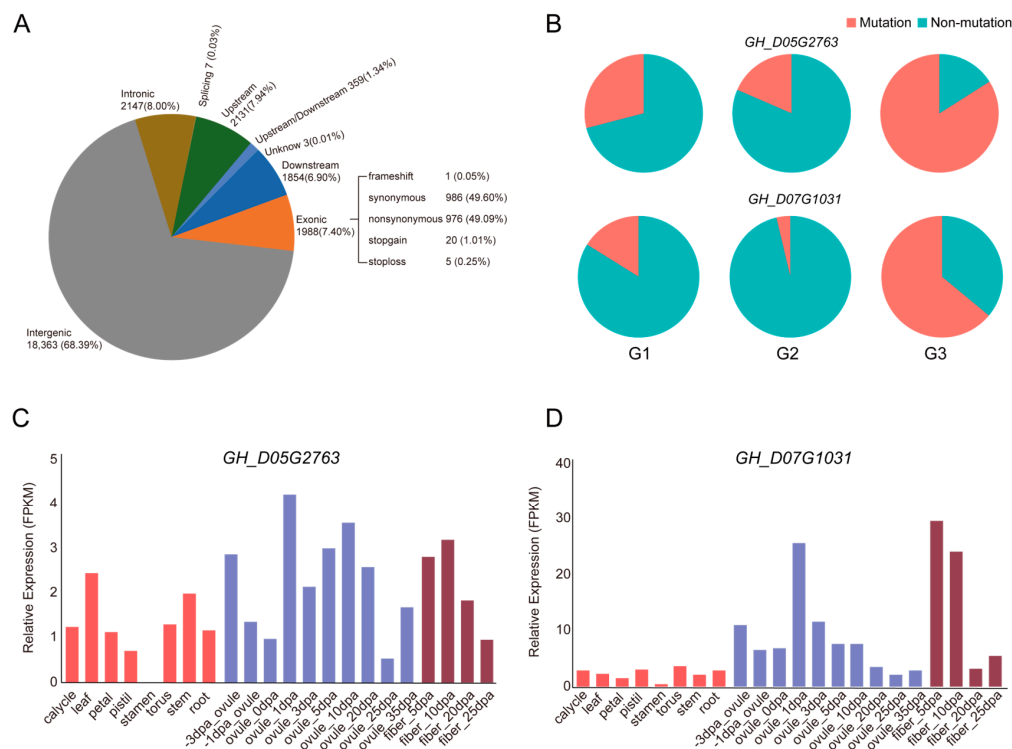


Figure 3. Genetic variation and expression profiles of *GH_D05G2763* and *GH_D07G1031*. (A) Statistics of SNP annotation information (B) Proportional distribution of SNPs occurred in *GH_D05G2763* and *GH_D07G1031* across different cotton populations. The pie chart shows the percentage of SNPs among G1, G2 and G3 group. (C) Expression level of *GH_D05G2763* in various tissues (D) Expression level of *GH_D07G1031* in various tissues.

3. Discussion

Molecular marker technology, such as RFLP and SRR, is widely used in previous genetic diversity analysis of cotton varieties [4,6,9]. However, the practical limitations of these molecular markers, including their finite number and operational complexity, restrict their use in large-scale studies. With the publication of the *Gossypium hirsutum* reference genome [24], SNP is approaching the ultimate standard for variation detection at the molecular level [1,25–27]. Due to the high-cost of re-sequencing of large population, efficient and low-cost SNP genotyping technology has become the optimal choice for SNP detection. Currently, low-cost liquid-phase SNP arrays have been developed in multiple species for breeding research [16,18,28,29]. The “ZJU CottonSNP40K” chip [19], developed using SNP-based targeted sequencing genotyping technology based on the cotton genome, was used in this study for genotyping target sites in Northern Xinjiang cotton cultivars. This cotton chip that has been widely used in cotton biological breeding, such as QTL mapping [19,20].

Pedigree analysis indicates that the parental sources of Northern Xinjiang cotton cultivars mainly originate from three varieties: Beiersinuo, KK1543, and 611B. The first lineage uses Beiersinuo as a parent to breed Xinluzao 6, Xinluzao 16, Xinluzao 27, and Xinluzao 28 through hybridization and selection. Subsequently, using Xinluzao 6, Xinluzao 16, and Xinluzao 28 as backbone parents, a series of varieties were developed. For example, Xinluzao 22 was bred from the elite line 451 and Xinluzao 6. Xinluzao 29 was selected from lines of Xinluzao 16. The second lineage uses KK1543 as a parent, breeding Xinluzao 3 and Xinluzao 4 through multiple rounds of hybridization. The third lineage uses 611B and Sining SA as parents to breed Xinluzao 1. Shihezi Academy of Agricultural Sciences further used Xinluzao 1 as a backbone parent to breed Xinluzao 2, Xinluzao 8, and Xinluzao 36. These pedigrees provide valuable genetic information and references for parental selection in Northern Xinjiang cotton breeding, aiding in optimizing breeding strategies. This study, based on 40K liquid SNP chip technology, systematically analyzed the genetic diversity and kinship relationships of 83 *G. hirsutum* main cultivars in the Northern Xinjiang cotton region, revealing the characteristics of their genetic structure and the genetic bottleneck in parental utilization. The results indicate that Northern Xinjiang cotton cultivars could be divided into three major genetic groups, corresponding to different parental sources and breeding pedigrees. Group I was mainly composed of varieties introduced from the former Soviet Union and the USA (such as Beiersinuo, KK1543, etc.) and their derivatives. Among them, Beiersinuo, as a core parent, directly or indirectly gave rise to 27 bred cultivars. Its excessive use resulted in a high average kinship coefficient of 0.72 within the population, significantly exacerbating genetic background homogenization. This finding aligns with previous conclusion [1,2,30,31] about the narrow genetic base of Chinese self-bred *G. hirsutum* varieties, further confirming at the molecular level that the singular utilization of exotic germplasm resources is a key factor limiting the genetic diversity of cotton in Northern Xinjiang. In contrast, Groups II and III are mainly composed of varieties bred by local research units in Xinjiang (Figure 2). Their parental sources rely more heavily on early-introduced former Soviet germplasm (e.g., 611B), but genetic diversity remains at a low level, indicating that the scope of parental selection in current breeding strategies needs to be expanded.

The SNP annotation results showed that only about 7.4% of the variants were located in gene coding or regulatory regions, that are strong candidates for mediating the phenotypic differences observed. Further filtering of SNPs occurring in exons, yielded 27 non-synonymous mutation type SNPs. These SNPs exhibited significantly different distribution frequencies among the different clustered groups. This pattern is consistent with genetic differentiation between populations, which could arise from various processes

including genetic drift, direct or indirect selection, or localized mutation. We also found two kinase proteins carrying variant sites that may contribute to phenotypic variation in Xinjiang cotton varieties. This provides important clues for subsequent functional gene mining and molecular marker-assisted breeding.

This study achieved more efficient genotyping through high-density SNP chips, and the classification results highly coincided with the subgroup division reported previously, further validating the profound influence of former Soviet germplasm on Xinjiang cotton breeding. However, this study found that the genetic contribution of Beiersinuo far exceeded expectations. Its derivative varieties accounted for 32.5% of the main cultivars in Northern Xinjiang, higher than the contribution ratio of former Soviet varieties (e.g., KK1543 only derived 2 varieties). This phenomenon highlights the risk in commercial breeding of over-reliance on a single high-yielding parent, which may lead to decreased stress resistance and adaptability. In similar international studies, the 63K cotton SNP chip also revealed similar issues, namely that the genetic base of modern varieties has significantly narrowed due to the heavy use of core parents [8]. Therefore, the findings of this study not only provide implications for breeding in Northern Xinjiang but also offer a reference for the sustainable utilization of global cotton genetic resources.

4. Materials and Methods

4.1. Materials

Eighty-three *G. hirsutum* varieties were collected from cotton breeding units including the Xinjiang Academy of Agricultural Sciences. Variety names and breeding organization information are shown in Table S1. The varieties were uniformly planted in an artificial climate chamber at 28 °C with a photoperiod of 16 h light/8 h dark. True leaf tissue was sampled separately after cotton seedlings reached the two-true-leaves stage. Genomic DNA was extracted using the modified CTAB method [32].

4.2. Genotyping and SNP Analysis

Genotyping was performed using the 40K chip [19], based on liquid probe hybridization targeted genotyping technology (Genotyping By Target Sequencing, GBTS), jointly designed and developed by Zhejiang University and Borui Di Biotechnology Co., Ltd. Genotyping was completed by Borui Di Biotechnology Co., Ltd. Quality-controlled clean reads were aligned to the reference genome TM-1 [33] using the BWA-mem alignment method [34]. SNPs with minor allele frequency (MAF) < 0.05, missing rate > 0.8, or identical to the TM-1 reference genome were removed. Ultimately, 26,852 high-quality SNP markers were determined for subsequent analysis.

4.3. SNP Variant Annotation

Based on TM-1 reference genome information and annotation files, ANNOVAR software (Version: 2.1.1) was used to annotate and classify genome-wide SNP variants [35]. Variants were mainly classified into the following types: SNPs located in intergenic regions, upstream regions, downstream regions, intronic regions, and exonic regions. SNPs located in the coding sequence (CDS) region were further categorized as stop gain, stop loss, synonymous mutation, non-synonymous mutation, frameshift deletion, and splicing.

4.4. Kinship Analysis

Based on the SNP variant information of this population, a phylogenetic tree was constructed using the neighbor-joining (NJ) method in Phyip software (Version: 3.696) [36]. Population structure analysis was performed using ADMIXTURE software (version 1.3.0, <https://dalexander.github.io/admixture/> (accessed on 10 January 2025)). Principal com-

ponent analysis (PCA) was conducted using GCTA software (Version 1.26.0) [37]. Kinship coefficients were estimated using the SNPRelate (version 1.40.0) [38] package in R. After quality control (excluding samples and SNPs with >20% missing data and imputing remaining missing values), a robust kinship matrix was generated by calculating Identity-by-State (IBS) similarities. To ensure reliable computation, genotypes were managed in a GDS file created via the gdsfmt package, and IBS values were computed using all QC-passed SNPs without linkage disequilibrium pruning. The resulting matrix was used to assess genetic relatedness among all pairwise combinations of individuals.

4.5. Transcriptome Analysis

The clean RNA-seq reads were mapped to the reference genome of TM-1 [33] using HISAT2 [39] (v 2.1.0) with default settings. High-quality mapping reads were used to calculate gene expression levels using StringTie [40] (v2.1.4) with parameter settings (--fr -e -G).

4.6. Published Data Download

The Illumina RNA-seq data of TM-1 was retrieved from the NCBI Sequence Read Archive (BioProject: PRJNA490626) [33] and the published cotton genome sequences was downloaded from COTTONOMICS database [41] (<http://cotton.zju.edu.cn/> (accessed on 10 January 2025)), respectively.

5. Conclusions

This study utilized the “ZJU CottonSNP40K” chip to conduct a genome-wide genetic analysis of 83 *G. hirsutum* cultivars from the Northern Xinjiang cotton region, obtaining 26,852 high-quality SNP markers. It systematically elucidated their genetic structure characteristics and kinship relationships. The results show that the tested materials can be divided into three major genetic groups. Group I, centered around the introduced US variety “Beiersinuo” as the core parent, accounted for 32.5% of the derivative varieties (Figure S1). Its excessive use led to genetic homogenization within the population (average kinship coefficient 0.72). Groups II and III are dominated by locally bred varieties; their genetic diversity levels also remain low, with average kinship coefficients of 0.75 and 0.77, respectively. From the perspective of genomic variation patterns, the number of SNPs in subgenome A (15,222) was significantly higher than in subgenome D (11,630), and their distribution was uneven across chromosomes (chromosome A08 had the highest marker density). Only 7.4% of SNPs were located in functional regions, which are strong candidates for mediating the phenotypic differences observed. The root cause of the narrow genetic base of current Northern Xinjiang cotton cultivars lies in the singular utilization of exotic germplasm, especially “Beiersinuo”. Future efforts require a diversified parental selection system, prioritizing the integration of wild or local germplasm with stress resistance and high-quality traits, and breaking genetic bottlenecks through molecular marker-assisted selection. This study provides a molecular basis for the efficient utilization of genetic resources in Northern Xinjiang cotton and holds significant guiding importance for promoting the breeding of new varieties with high yield, superior quality, and multi-resistance.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms27010545/s1>. References [42,43] are cited in the Supplementary Materials.

Author Contributions: N.A. and T.Z. designed and supervised the research. Z.Z., N.W., K.N., G.F., H.G. and Z.S. selected sample leaves. S.J. performed the data analysis. Z.Z. and S.J. wrote the manuscript. N.A. and T.Z. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Eighth Division Shihezi City Mid-Aged and Young Scientific and Technological Innovation Backbone Talent Program Project (2024RC01); Eighth Division Shihezi City Key Areas Science and Technology Tackling Plan Project (2024NY01); Corps Key Tackling Project in Science and Technology Field (2024DA003); and the Shicheng Outstanding Talent-Zhihong Zheng.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets presented in this study can be found in the figshare database (<https://doi.org/10.6084/m9.figshare.30817382> (accessed on 9 December 2025)).

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

References

1. Fang, L.; Wang, Q.; Hu, Y.; Jia, Y.H.; Chen, J.D.; Liu, B.L.; Zhang, Z.Y.; Guan, X.Y.; Chen, S.Q.; Zhou, B.L.; 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]
2. 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]
3. Ma, Z.Y.; He, S.P.; Wang, X.F.; Sun, J.L.; Zhang, Y.; Zhang, G.Y.; Wu, L.Q.; Li, Z.K.; Liu, Z.H.; Sun, G.F.; 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]
4. Percy, R.G.; Wendel, J.F. Allozyme evidence for the origin and diversification of *Gossypium barbadense* L. *Theor. Appl. Genet.* **1990**, *79*, 529–542. [CrossRef]
5. Wendel, J.F.; Percy, R.G. Allozyme diversity and introgression in the Galapagos Islands endemic *Gossypium darwinii* and its relationship to continental *G. barbadense*. *Biochem. Syst. Ecol.* **1990**, *18*, 517–528. [CrossRef]
6. 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]
7. Brubaker, C.L.; Wendel, J.F. Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*; Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). *Am. J. Bot.* **1994**, *81*, 1309–1326. [CrossRef]
8. Hulse-Kemp, A.M.; Lemm, J.; Plieske, J.; Ashrafi, H.; Buyyarapu, R.; Fang, D.D.; Frelichowski, J.; Giband, M.; Hague, S.; Hinze, L.L.; et al. Development of a 63K SNP array for cotton and high-density mapping of intraspecific and interspecific populations of *Gossypium* spp. *G3 (Bethesda)* **2015**, *5*, 1187–1209. [CrossRef] [PubMed]
9. Malik, W.; Ashraf, J.; Iqbal, M.Z.; Khan, A.A.; Qayyum, A.; Ali Abid, M.; Noor, E.; Ahmad, M.Q.; Abbasi, G.H. Molecular markers and cotton genetic improvement: Current status and future prospects. *Sci. World J.* **2014**, *2014*, 607091. [CrossRef]
10. Jin, S.; Han, Z.; Hu, Y.; Si, Z.; Dai, F.; He, L.; Cheng, Y.; Li, Y.; Zhao, T.; Fang, L. Structural variation (SV)-based pan-genome and GWAS reveal the impacts of SVs on the speciation and diversification of allotetraploid cottons. *Mol. Plant* **2023**, *16*, 678–693. [CrossRef]
11. Wang, S.; Chen, J.D.; Zhang, W.P.; Hu, Y.; Chang, L.J.; Fang, L.; Wang, Q.; Lv, F.N.; Wu, H.T.; Si, Z.F.; et al. Sequence-based ultra-dense genetic and physical maps reveal structural variations of allopolyploid cotton genomes. *Genome Biol.* **2015**, *16*, 108–125. [CrossRef]
12. Zhou, Z.; Jiang, Y.; Wang, Z.; Gou, Z.; Lyu, J.; Li, W.; Yu, Y.; Shu, L.; Zhao, Y.; Ma, Y. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* **2015**, *33*, 408–414. Erratum in *Nat. Biotechnol.* **2016**, *34*, 441. [CrossRef]
13. Huang, X.; Sang, T.; Zhao, Q.; Feng, Q.; Zhao, Y.; Li, C.; Zhu, C.; Lu, T.; Zhang, Z.; Li, M. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **2010**, *42*, 961–967. [CrossRef]
14. Zhang, J.; Yang, J.J.; Zhang, L.K.; Luo, J.; Zhao, H.; Zhang, J.; Wen, C.L. A new SNP genotyping technology Target SNP-seq and its application in genetic analysis of cucumber varieties. *Sci. Rep.* **2020**, *10*, 5623–5633. Correction in *Sci. Rep.* **2021**, *11*, 8010. [CrossRef] [PubMed]
15. Guo, Z.F.; Yang, Q.; Huang, F.F.; Zheng, H.J.; Sang, Z.Q.; Xu, Y.F.; Zhang, C.; Wu, K.S.; Tao, J.J.; Prasanna, B.M.; et al. Development of high-resolution multiple-SNP arrays for genetic analyses and molecular breeding through genotyping by target sequencing and liquid chip. *Plant Commun.* **2021**, *2*, 100230. [CrossRef] [PubMed]
16. Liu, Y.C.; Liu, S.L.; Zhang, Z.F.; Ni, L.B.; Chen, X.M.; Ge, Y.X.; Zhou, G.A.; Tian, Z.X. GenoBaits Soy40K: A highly flexible and low-cost SNP array for soybean studies. *Sci. China Life Sci.* **2020**, *65*, 359–362. [CrossRef]

17. Cui, F.; Zhang, N.; Fan, X.; Zhang, W.; Zhao, C.; Yang, L.; Pan, R.; Chen, M.; Han, J.; Zhao, X. Utilization of a Wheat660K SNP array-derived high-density genetic map for high-resolution mapping of a major QTL for kernel number. *Sci. Rep.* **2017**, *7*, 3788–3799. [\[CrossRef\]](#)
18. Liu, S.; Xiang, M.; Wang, X.; Li, J.; Cheng, X.; Li, H.; Singh, R.P.; Bhavani, S.; Huang, S.; Zheng, W.; et al. Development and application of the GenoBaits WheatSNP16K array to accelerate wheat genetic research and breeding. *Plant Commun.* **2025**, *6*, 101138. [\[CrossRef\]](#)
19. Si, Z.; Jin, S.; Li, J.; Han, Z.; Li, Y.; Wu, X.; Ge, Y.; Fang, L.; Zhang, T.; Hu, Y. The design, validation, and utility of the “ZJU CottonSNP40K” liquid chip through genotyping by target sequencing. *Ind. Crop. Prod.* **2022**, *188*, 115629–115636.
20. Chen, H.; Han, Z.; Ma, Q.; Dong, C.; Ning, X.; Li, J.; Lin, H.; Xu, S.; Li, Y.; Hu, Y.; et al. Identification of elite fiber quality loci in upland cotton based on the genotyping-by-target-sequencing technology. *Front. Plant Sci.* **2022**, *13*, 1027806. [\[CrossRef\]](#)
21. Fan, M.; Wang, M.; Bai, M.-Y. Diverse roles of SERK family genes in plant growth, development and defense response. *Sci. China Life Sci.* **2016**, *59*, 889–896. [\[CrossRef\]](#)
22. Lan, Z.; Song, Z.; Wang, Z.; Li, L.; Liu, Y.; Zhi, S.; Wang, R.; Wang, J.; Li, Q.; Bleckmann, A.; et al. Antagonistic RALF peptides control an intergeneric hybridization barrier on *Brassicaceae* stigmas. *Cell* **2023**, *186*, 4773–4787.e4712. [\[CrossRef\]](#)
23. Zhang, Y.; Tian, H.; Chen, D.; Zhang, H.; Sun, M.; Chen, S.; Qin, Z.; Ding, Z.; Dai, S. Cysteine-rich receptor-like protein kinases: Emerging regulators of plant stress responses. *Trends Plant Sci.* **2023**, *28*, 776–794. [\[CrossRef\]](#)
24. Zhang, T.Z.; Hu, Y.; Jiang, W.K.; Fang, L.; Guan, X.Y.; Chen, J.D.; Zhang, J.B.; Saski, C.A.; Scheffler, B.E.; Stelly, D.M.; et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* **2015**, *33*, 531–537. [\[CrossRef\]](#) [\[PubMed\]](#)
25. He, S.P.; Sun, G.F.; Geng, X.L.; Gong, W.F.; Dai, P.H.; Jia, Y.H.; Shi, W.J.; Pan, Z.E.; Wang, J.D.; Wang, L.Y.; et al. The genomic basis of geographic differentiation and fiber improvement in cultivated cotton. *Nat. Genet.* **2021**, *53*, 916–924. [\[CrossRef\]](#)
26. Li, X.; Wang, Y.; Cai, C.; Ji, J.; Han, F.; Zhang, L.; Chen, S.; Zhang, L.; Yang, Y.; Tang, Q. Large-scale gene expression alterations introduced by structural variation drive morphotype diversification in *Brassica oleracea*. *Nat. Genet.* **2024**, *56*, 517–529. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Zhang, C.; Shao, Z.; Kong, Y.; Du, H.; Li, W.; Yang, Z.; Li, X.; Ke, H.; Sun, Z.; Shao, J. High-quality genome of a modern soybean cultivar and resequencing of 547 accessions provide insights into the role of structural variation. *Nat. Genet.* **2024**, *56*, 2247–2258. [\[CrossRef\]](#)
28. Guan, H.; Lu, Y.; Li, X.; Liu, B.; Li, Y.; Zhang, D.; Liu, X.; He, G.; Li, Y.; Wang, H.; et al. Development of a MaizeGerm50K array and application to maize genetic studies and breeding. *Crop J.* **2024**, *12*, 1686–1696. [\[CrossRef\]](#)
29. Li, Z.; Wang, L.; Liu, Y.; Ma, X.; Zhang, A.; Luo, Z.; Yan, M.; Zhou, L.; Chen, L.; Luo, L.; et al. WDR6K, a designed SNP array for the research and improvement of rice drought-resistance. *Plant Stress.* **2025**, *15*, 100800. [\[CrossRef\]](#)
30. 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*, 33–45. [\[CrossRef\]](#)
31. Li, Y.; Si, Z.; Wang, G.; Shi, Z.; Chen, J.; Qi, G.; Jin, S.; Han, Z.; Gao, W.; Tian, Y. Genomic insights into the genetic basis of cotton breeding in China. *Mol. Plant* **2023**, *16*, 662–677. [\[CrossRef\]](#)
32. Paterson, A.H.; Brubaker, C.L.; Wendel, J.F. A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. *Plant Mol. Biol. Rep.* **1993**, *11*, 122–127. [\[CrossRef\]](#)
33. Hu, Y.; Chen, J.D.; Fang, L.; Zhang, Z.Y.; Ma, W.; Niu, Y.C.; Ju, L.Z.; Deng, J.Q.; Zhao, T.; Lian, J.M.; et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat. Genet.* **2019**, *51*, 739–748. [\[CrossRef\]](#)
34. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [\[CrossRef\]](#)
35. Wang, K.; Li, M.Y.; Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **2010**, *38*, e164. [\[CrossRef\]](#)
36. Felsenstein, J. *PHYLIP (Phylogeny Inference Package)*, version 3.6.; Department of Genome Sciences, University of Washington: Seattle, WA, USA, 2005.
37. 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. [\[CrossRef\]](#)
38. Zheng, X.; Gogarten, S.M.; Lawrence, M.; Stilp, A.; Conomos, M.P.; Weir, B.S.; Laurie, C.; Levine, D. SeqArray—A storage-efficient high-performance data format for WGS variant calls. *Bioinformatics* **2017**, *33*, 2251–2257. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.-C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **2015**, *33*, 290–295. [\[CrossRef\]](#) [\[PubMed\]](#)

41. Dai, F.; Chen, J.; Zhang, Z.; Liu, F.; Li, J.; Zhao, T.; Hu, Y.; Zhang, T.; Fang, L. COTTONOMICS: A comprehensive cotton multi-omics database. *Database* **2022**, 2022, baac080. [[CrossRef](#)]
42. Han, Z.G.; Chen, H.; Cao, Y.W.; He, L.; Si, Z.F.; Hu, Y.; Lin, H.; Ning, X.Z.; Li, J.L.; Ma, Q.; et al. Genomic insights into genetic improvement of upland cotton in the world's largest growing region. *Ind. Crop. Prod.* **2022**, 183, 114929–114938. [[CrossRef](#)]
43. Ma, Z.; Zhang, Y.; Wu, L.; Zhang, G.; Sun, Z.; Li, Z.; Jiang, Y.; Ke, H.; Chen, B.; Liu, Z.; et al. High-quality genome assembly and resequencing of modern cotton cultivars provide resources for crop improvement. *Nat. Genet.* **2021**, 53, 1385–1391. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.