Scilight

*Annals of Agri-bio Research*

https://agribiop.com/



*Article*

Check for updates

# SSR-Based Genetic Diversity and Candidate Salinity-Response Associations in Uzbek Upland Cotton

Abdulahat Azimov, Jaloliddin Shavkiev *, Farkhod Abdurasulov and Mukhammad Alikulov

Institute of Genetics and Plant Experimental Biology, Academy of Sciences of the Republic of Uzbekistan, 266 Yuqori-Yuz Street, Kibray District, Tashkent Region 111208, Uzbekistan
* Correspondence: jaloliddinshavkiev1992@gmail.com

**How To Cite:** Azimov, A., Shavkiev, J., Abdurasulov, F., & Alikulov, M. (2026). SSR-Based Genetic Diversity and Candidate Salinity-Response Associations in Uzbek Upland Cotton. *Annals of Agri-bio Research*, *31*(2), 168–180. https://doi.org/10.53941/agrbio.2026.100019

Received: 14 July 2026
Revised: 17 September 2026
Accepted: 14 October 2026
Published: 13 November 2026

**Abstract:** Soil salinity constrains cotton productivity and fiber quality in irrigated arid regions. This study combined field phenotyping and a targeted simple sequence repeat (SSR) marker panel to characterize genetic diversity and identify preliminary allele–trait relationships in Uzbek upland cotton germplasm. Seventeen genotypes were evaluated under non-saline, moderately saline, and highly saline field conditions. Salinity significantly affected plant growth, boll development, yield, and fiber characteristics, with marked genotype-dependent responses. Several materials maintained comparatively favorable productivity or fiber properties under saline conditions. Most SSR loci were polymorphic and revealed moderate diversity and broad genetic differentiation within the evaluated panel, although clustering and ordination methods differed in exact group membership. Exploratory screening detected candidate allele–trait signals related to plant height, yield components, and fiber properties. These signals should not be considered validated diagnostic markers because of the limited germplasm and marker panels, binary allele scoring, incomplete genome coverage, and absence of multiple-testing-adjusted probabilities. The results identify phenotypically complementary materials and testable molecular hypotheses for cotton improvement under salinity. Validation using high-resolution codominant genotyping, larger representative populations, formal relatedness control, and multi-environment experiments is required before marker-assisted application.

**Keywords:** cotton germplasm; salinity response; SSR markers; genetic diversity; fiber quality; allele–trait association

## 1. Introduction

Upland cotton accounts for most global cotton production because of its productivity and broad adaptation. In Uzbekistan and other irrigated regions of Central Asia, however, salt accumulation reduces crop establishment, vegetative growth, boll development, yield, and technological fiber properties. Salinity imposes osmotic limitation, ion toxicity, nutrient imbalance, oxidative stress, and disturbed assimilate allocation. Although cotton is moderately salt tolerant as a species, the response varies substantially among cultivars and breeding lines, providing exploitable variation for genetic improvement (Sikder et al., 2020; Chaudhary et al., 2024).

Field phenotyping under naturally saline conditions is necessary for identifying locally adapted donors, but phenotypic selection alone is affected by environment and genotype × environment interaction.

**Copyright:** © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).
**Publisher's Note**: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

SSR markers remain practical for germplasm characterization because they are reproducible, codominant, multiallelic, and accessible to breeding laboratories. Their use can clarify genetic relationships, reduce parental redundancy, and provide candidate alleles for subsequent marker-assisted selection (Abdurakhmonov et al., 2008; Kuang et al., 2022).

Genetic-diversity analysis and marker-trait screening answer complementary questions. Distance-based clustering and ordination describe genetic relationships among the sampled materials, whereas exploratory allele-trait contrasts identify hypotheses for subsequent validation. These analyses are not equivalent to a model-based population-structure analysis. The objectives were to (i) quantify agronomic and fiber responses to natural soil salinity, (ii) characterize marker diversity and genetic differentiation using a targeted 20-SSR panel, (iii) identify preliminary allele-trait signals, and (iv) nominate phenotypically complementary materials for further regional breeding evaluation. Molecular-marker approaches are well established for cotton germplasm characterization (Abdurakhmonov et al., 2008; Kuang et al., 2022); the novelty claimed here is therefore not the SSR method itself, but the joint field and marker description of this specific, previously under-characterized Uzbek working panel under locally relevant saline environments.

## 2. Materials and Methods

### *2.1. Plant Material, Site, and Experimental Design*

Seventeen upland cotton genotypes maintained in the working collection of the Institute of Genetics and Plant Experimental Biology were evaluated: Yulduz-2, T-1001, T-1002, T-1003, T-1004, T-1005, T-1024, T-1050, T-1080, T-1090, T-2000, Orzu, Ishonch, Mehnat, L-217, L-218, and L-421. The panel intentionally combined four named Uzbek cultivars (Yulduz-2, Orzu, Ishonch, and Mehnat) with 13 institute breeding lines designated by T- or L-prefixes. It was a targeted working panel selected for contrasting field responses and availability, not a statistically sampled national core collection. Verified pedigree records and first-release/development years were not available in the analytical dataset and are therefore not reconstructed here. Consequently, the 17 entries cannot represent the full genetic diversity of Uzbek upland cotton; inference is restricted to this panel. Field trials were conducted in 2025 and 2026 at the experimental station of the Institute of Genetics and Plant Experimental Biology, Kibray District, Tashkent Region, Uzbekistan (41.389 N, 69.465 E), under a semiarid continental climate and irrigated sierozem soil. The identity and source categories of these materials are summarized in Table 1.

**Table 1.** Identity, source category, and within-study agronomic relevance of the 17 cotton materials.

| Genotype | Material Category | Source | Trait Noted in This Study |
|---|---|---|---|
| Yulduz-2 | Named Uzbek cultivar | IG&PEB working collection | Relatively divergent in marker analyses |
| Orzu | Named Uzbek cultivar | IG&PEB working collection | Higher yield maintenance |
| Ishonch | Named Uzbek cultivar | IG&PEB working collection | Evaluated comparator |
| Mehnat | Named Uzbek cultivar | IG&PEB working collection | Higher yield/boll-weight maintenance |
| T-1001 | Breeding line | IG&PEB working collection | Fiber stability |
| T-1002 | Breeding line | IG&PEB working collection | Isolated in NJ/PCA |
| T-1003 | Breeding line | IG&PEB working collection | Fiber stability |
| T-1004 | Breeding line | IG&PEB working collection | Relatively divergent in marker analyses |
| T-1005 | Breeding line | IG&PEB working collection | Closely grouped in UPGMA |
| T-1024 | Breeding line | IG&PEB working collection | UHML/strength stability |
| T-1050 | Breeding line | IG&PEB working collection | Height/UHML maintenance |
| T-1080 | Breeding line | IG&PEB working collection | Closely grouped in UPGMA |
| T-1090 | Breeding line | IG&PEB working collection | Higher yield maintenance |
| T-2000 | Breeding line | IG&PEB working collection | Higher yield maintenance |
| L-217 | Breeding line | IG&PEB working collection | Fiber stability |
| L-218 | Breeding line | IG&PEB working collection | Height maintenance |
| L-421 | Breeding line | IG&PEB working collection | Strong height reduction |

A randomized complete block design (Gomez & Gomez 1984) with two field replications was used within each of three naturally occurring salinity environments: non-saline control, moderate salinity (0.3–0.6% total soluble salts), and high salinity (0.6–0.9%). Each plot contained four rows spaced 60 cm apart, with 10–15 cm between plants. Ten plants were sampled from the two central rows, and plot means were treated as experimental units. Soil cores (0–30 cm) collected before planting were composited by plot for determination of total soluble salts. Crop management followed local recommendations. For classification, the 0.6% boundary is assigned to the high-salinity category; moderate salinity is defined as $0.3\% \leq TSS < 0.6\%$, and high salinity as $0.6\% \leq TSS \leq 0.9\%$.

*2.2. Agronomic and Fiber Measurements*

Plant height was measured at physiological maturity. Boll weight was calculated from 20 mature bolls per experimental unit, and seed cotton yield was recorded from the sampled plants and expressed as g plant$^{-1}$. Lint samples were conditioned for 24 h at 21 ± 1 °C and 65 ± 2% relative humidity. UHML (inches), fiber strength (g tex$^{-1}$), and SFI (%) were measured using an HVI 1000 system (Uster Technologies, Switzerland).

*2.3. DNA Extraction, SSR Amplification, and Scoring*

Genomic DNA was isolated from young leaves using a PureLink Genomic DNA Mini Kit. DNA concentration and purity were assessed with a NanoDrop 2000, and integrity was checked by agarose electrophoresis. Twenty SSR loci were chosen from published cotton marker resources (Abdurakhmonov et al., 2008; Lacape et al., 2006) because primer sequences were available and the loci had previously produced scorable polymorphism in cultivated cotton. The panel was not selected by a chromosome-stratified sampling design, and verified chromosome coordinates were not available for every locus in the source dataset. It must therefore be regarded as a targeted marker panel rather than genome-wide coverage; clustering and diversity estimates may be sensitive to marker number and genomic distribution. PCR reactions (20 µL) contained approximately 50 ng template DNA, 10 µL commercial 2× PCR master mix (1× final), each primer at 0.5 µM final concentration, and nuclease-free water. The proprietary master-mix formulation, including exact $MgCl_2$ and dNTP concentrations, was not recorded in the analytical dataset and cannot be reported retrospectively. Cycling comprised 95 °C for 5 min; 35 cycles of 94 °C for 30 s, locus-specific annealing at 55–60 °C for 30 s, and 72 °C for 45 s; and 72 °C for 10 min. Products were separated on 3% agarose gels. Because agarose cannot reliably resolve all small SSR length differences, only clearly separated, reproducible bands were called; ambiguous or closely spaced products were not forced into distinct allele classes. PAGE or capillary electrophoresis was not used, so allele counts may be underestimated and size labels are approximate. Each distinct, unambiguous size class at a locus was converted to a separate binary feature (present = 1, absent = 0). This allele-presence recoding treats the derived features as dominant variables and does not retain diploid genotype phase or distinguish all homozygous from heterozygous states.

Primer sequences were compiled from Abdurakhmonov et al. (2008) and Lacape et al. (2006). The retained laboratory sheet recorded only a 55–60 °C locus-optimization range, not an exact annealing temperature or complete allele-size range for each locus. These unavailable values are identified explicitly in Table 2B rather than reconstructed retrospectively.

**Table 2.** (**A**). Primer sequences of the 20 SSR loci screened across 17 Uzbek upland cotton genotypes. (**B**) Available amplification metadata for the SSR panel. * The archived protocol retained a locus-optimization range only; exact locus-specific temperatures and complete observed size ranges must be recovered from original laboratory records before independent replication.

**(A)**

| Locus | Forward Primer (5′–3′) | Reverse Primer (5′–3′) |
|---|---|---|
| BNL1421 | TGAAGATTTGGAGGCAATTG | GAAATCAAGCCTCAATTCGG |
| BNL1604 | AGAGGGAGTAAAGATTTGGGG | TCCAGTTCTTTTTGCCTTGG |
| BNL1672 | TGGATTTGTCCCTCTGTGTG | AACCAACTTTTCCAACACCG |
| BNL2634 | AACAACATTGAAAGTCGGGG | CCCAGCTGCTTATTGGTTTC |
| BNL3280 | GCAGAACTGCCACTTGTTTG | AGAAAATGGGTTGTGCTTGG |
| BNL3545 | AGTCAGTTTTTTGTTAGCAATATGC | AACCATTAATTCCCTATTTAACCG |
| BNL4028 | CTCGAATCACTGTTCCGACA | TTGTTGCGGATCATTGGATA |
| DPL0530 | AGACTTACTTAAAGGCACCATTCG | GCAGACTCTTCTGGTGTAACAGTG |
| HAU0058 | CCGAAGCATATTATGGAAGC | TGGCAGTAACACAACAGAGC |
| JESPR65 | CCACCCAATTTAAGAAGAAATTG | GGTTAGTTGTATTAGGGTCGTTG |
| MUCS223 | ACCTCCACTTTGTTTGTGGC | CACCTTGAATTTAGCACCCC |
| NAU2176 | TGGCACTTGCTAAACCATAA | GGTTTGAGCAGAAGGGTATG |
| NAU2277 | GAACTAGCCACATGATGCAC | TTGTTGAGGCATTAGTTTGC |
| NAU3011 | CCTTCACCATTTGCTTACCT | CTCCGTTTCTATTCCCAAAA |
| NAU3377 | TGCAAGGAATCAAGTTCACA | CTGATTGTACTTTGCGGGTA |
| NAU3588 | CCCCATAGGGCATACTTCTA | GCCAACAAGAACAACAACAC |

**Table 2.** *Cont.*

| (A) | | |
|---|---|---|
| **Locus** | **Forward Primer (5′–3′)** | **Reverse Primer (5′–3′)** |
| NAU5434 | AAAAGAACTTACGGCACAGG | AAATCACTGGCACTGGAATC |
| NAU6315 | CACTCACAAAATGTCCCTCA | CGGTTGATTCAAGCCTTATC |
| STV023 | GACAAACTACAAAAGGATGAACATGG | AGGCTAGGAAGCTTCTTGTTTTCC |
| TMB1295 | CTGATCCAAACACCCATGC | CGTGGAATTTGGTCATTTTG |

| (B) | | | |
|---|---|---|---|
| **SSR Locus** | **Annealing Temperature** | **Observed Allele-Size Range** | **Source** |
| BNL1421 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| BNL1604 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| BNL1672 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| BNL2634 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| BNL3280 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| BNL3545 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| BNL4028 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| DPL0530 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| HAU0058 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| JESPR65 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| MUCS223 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| NAU2176 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| NAU2277 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| NAU3011 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| NAU3377 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| NAU3588 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| NAU5434 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| NAU6315 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| STV023 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |
| TMB1295 | 55–60 °C * | Not retained | (Abdurakhmonov et al., 2008; Lacape et al., 2006) |

### *2.4. Statistical and Molecular Analyses*

Two-way ANOVA was performed in R 4.3.0 to test genotype (G), salinity (S), and G × S effects. Shapiro-Wilk and Levene tests were used to assess residual normality and variance homogeneity, and Tukey HSD was applied at $p < 0.05$. For each scored allele feature, carriers and non-carriers were compared for phenotypic traits. These were exploratory contrasts; because the original output did not retain Benjamini-Hochberg-adjusted probabilities, unadjusted $p < 0.05$ signals are reported only as candidates and not as validated associations.

PowerMarker 3.25 (Liu & Muse 2005) was used for Na, Ne, He, and PIC summaries. DARwin 6.0 generated Jaccard dissimilarities and exploratory UPGMA and Neighbor-Joining trees (Saitou & Nei 1987) from the binary allele-feature matrix; no bootstrap resampling was performed, so branch support cannot be inferred. R was used for molecular PCA (Jolliffe 2002), and GenAlEx 6.5 (Peakall & Smouse 2012) for AMOVA (Excoffier et al., 1992) with 999 permutations. MEGA X and OriginPro 2024 were used for tree display and graphics, respectively. NAU2176 was retained in the descriptive locus table to document screening failure but, as a monomorphic locus, was excluded from distance, clustering, ordination, AMOVA, and allele-trait analyses. The three AMOVA groups were post hoc groups defined from the broad UPGMA partition, not independent prior groups based on origin. AMOVA is therefore descriptive and partly circular; PhiST and the Nm approximation (Wright 1978; Slatkin 1985) are not estimates of national population differentiation or contemporary breeding gene flow.

Observed allele number (Na), effective allele number (Ne), gene-diversity/expected-heterozygosity index (He), PIC, and locus polymorphism status were summarized from the allele-feature data (Botstein et al., 1980). Because the available 0/1 matrix did not retain diploid codominant genotypes, observed heterozygosity (Ho), inbreeding coefficients, and Hardy-Weinberg equilibrium could not be calculated validly. He and PIC are retained as diversity summaries under the original allele-frequency treatment and should not be interpreted as a complete codominant population-genetic analysis. Jaccard dissimilarity (Jaccard 1908) was used because the analytical matrix represented allele-feature presence/absence and shared absence was not treated as evidence of common ancestry. Nei distance would be preferable for fully resolved codominant genotype calls; that alternative could not be reconstructed from the retained binary matrix.

## 3. Results

### *3.1. Agronomic Response to Salinity*

Genotype, salinity, and their interaction significantly affected plant height, boll weight, and seed cotton yield ($p < 0.001$; Table 3). Mean plant height declined from 77.26 cm in the control to 62.84 cm under moderate salinity and 49.04 cm under high salinity, corresponding to reductions of 18.7% and 36.5%. T-1050, Orzu, L-218, and T-1090 maintained comparatively greater height, whereas L-421 showed the strongest reduction. Plant height did not consistently predict reproductive performance.

Mean boll weight declined from 5.28 to 4.23 and 3.88 g across control, moderate, and high salinity, respectively. Mehnat, Orzu, and T-2000 retained relatively high boll weight. Seed cotton yield was most sensitive, decreasing from 26.74 to 18.47 and 8.72 g plant$^{-1}$, a 67.4% loss under high salinity. Orzu, Mehnat, T-1090, and T-2000 displayed comparatively better yield maintenance.

**Table 3.** Two-way ANOVA for agronomic traits under natural soil salinity.

| Source | df | Plant Height | Boll Weight | Seed Cotton Yield |
|---|---|---|---|---|
| Genotype (G) | 16 | F = 21.63 *** | F = 24.83 *** | F = 16.74 *** |
| Salinity (S) | 2 | F = 24.12 *** | F = 452.61 *** | F = 534.62 *** |
| G × S | 32 | F = 10.29 *** | F = 4.32 *** | F = 4.39 *** |
| Error | 51 | — | — | — |

*** $p < 0.001$.

### *3.2. Fiber-Quality Response*

Salinity significantly affected UHML, fiber strength, and SFI, with significant G × S interaction for each trait (Table 4). SFI increased with salinity; T-1003, T-1001, and T-1024 showed relatively stable values, whereas Mehnat, L-218, T-1002, and T-1005 showed larger increases. UHML decreased, but T-1024, T-1050, T-1001, and L-217 maintained length more effectively. Fiber strength also declined; T-2000, T-1002, L-217, and T-1003 were comparatively stable. Trait-specific responses demonstrate that productivity and multiple fiber attributes must be selected jointly.

**Table 4.** Two-way ANOVA for fiber-quality traits under natural soil salinity.

| Source | df | UHML | SFI | Strength |
|---|---|---|---|---|
| Genotype (G) | 16 | F = 28.74 *** | F = 41.58 *** | F = 32.19 *** |
| Salinity (S) | 2 | F = 395.62 *** | F = 512.41 *** | F = 476.83 *** |
| G × S | 32 | F = 4.81 *** | F = 6.14 *** | F = 5.47 *** |
| Error | 51 | — | — | — |

UHML, upper half mean length; SFI, short fiber index. *** $p < 0.001$.

### *3.3. Integrated Phenotypic Variation*

Principal component analysis of the combined agronomic and fiber dataset separated productivity-related and fiber-related response patterns. PC1 explained 34.5% and PC2 explained 22.2% of the phenotypic variation (56.7% jointly). Plant height, boll weight, and seed cotton yield loaded in broadly similar directions, whereas UHML and fiber strength were opposed to SFI. T-2000 was positioned toward greater plant height, while T-1001 and T-1003 were aligned more closely with UHML and fiber strength. Mehnat and L-421 were displaced toward higher SFI, illustrating that productive or vigorous genotypes do not necessarily retain favorable fiber profiles under salinity. The genotype distribution and trait loadings are presented in Figure 1.

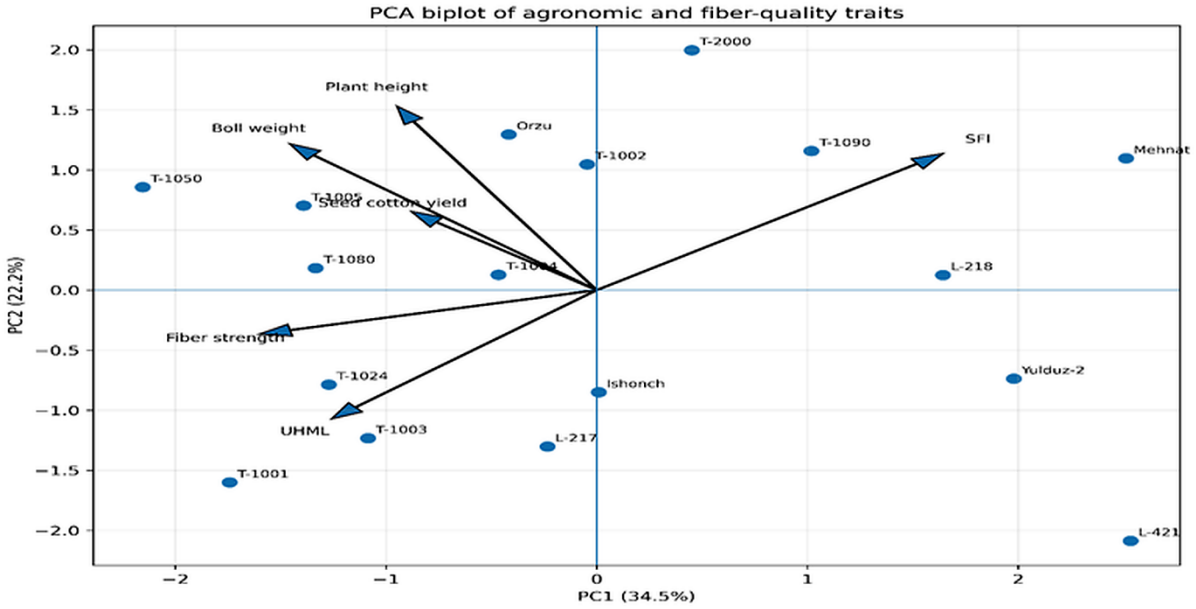



**Figure 1.** PCA biplot integrating agronomic and fiber-quality traits across the 17 genotypes. PC1 and PC2 explain 34.5% and 22.2% of phenotypic variation, respectively.

### *3.4. SSR Polymorphism and Marker Informativeness*

The 20 loci produced 53 alleles, averaging 2.65 alleles per locus. Nineteen loci were polymorphic (95%); NAU2176 was monomorphic. Across all loci, mean He was 0.488 and mean PIC was 0.436. NAU2277 showed the greatest allelic richness (Na = 5) and informativeness (PIC = 0.761), followed by MUCS223, BNL3545, JESPR65, NAU6315, and DPL0530. These loci are useful for discriminating among the present genotypes, although marker informativeness itself does not establish a functional association with salinity tolerance. Detailed locus-level diversity statistics are provided in Supplementary Table S1. The distribution of PIC values is illustrated in Supplementary Figure S1.

### *3.5. Genetic Relationships and Marker-Defined Differentiation*

After excluding monomorphic NAU2176, pairwise Jaccard genetic distance averaged 0.487 and ranged from 0.214 to 0.762 in the retained analysis (Table 5). T-2000 × T-1005 was the most similar pair, followed by T-1080 × T-2000 and T-2000 × T-1003. T-1004 x Mehnat was the most divergent pair, followed by T-1001 × Orzu. UPGMA produced three broad groups (5, 7, and 5 entries), with T-1002 in Group II. Neighbor-Joining produced a different configuration (5, 9, and a singleton T-1002), and molecular PCA likewise positioned T-1002 separately. Thus, the methods agree only on broad differentiation and several close relationships; they do not consistently recover identical membership. Because no bootstrap support was calculated and PC1 + PC2 captured only 34.7% of marker variation, none of the three displays is designated definitive. The genotype membership of the Neighbor-Joining groups is listed in Supplementary Table S2. The corresponding UPGMA and Neighbor-Joining relationships are illustrated in Figures 2 and 3, respectively.

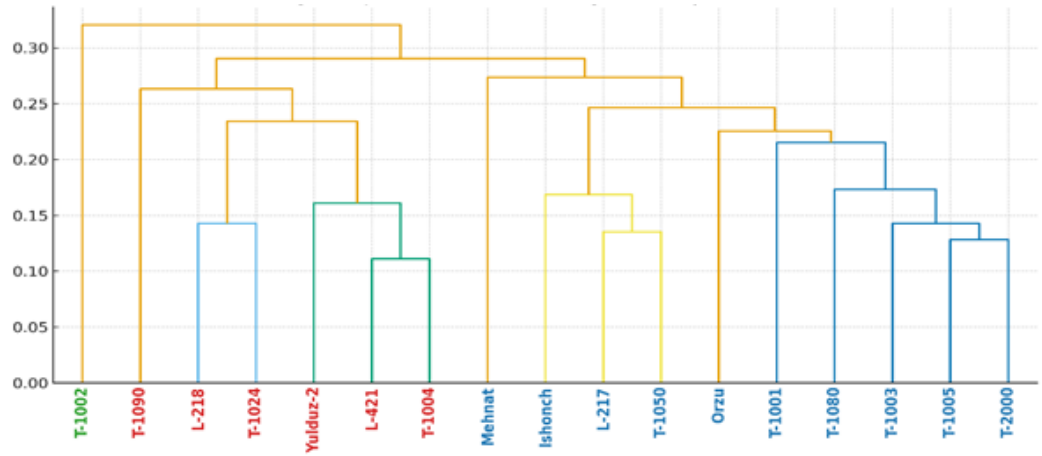



**Figure 2.** UPGMA dendrogram of 17 upland cotton genotypes based on Jaccard distances from binary SSR profiles.

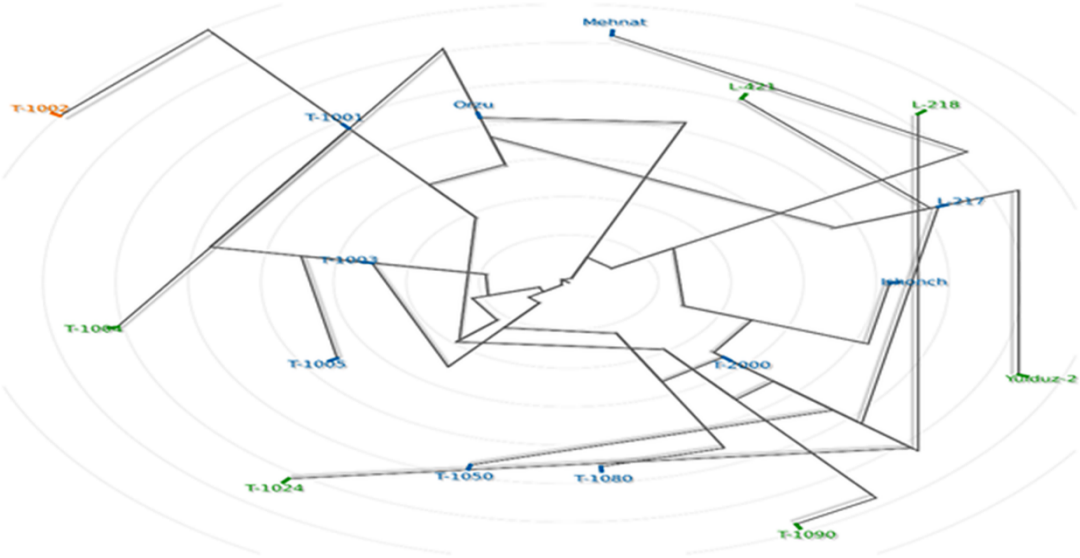



**Figure 3.** Radial Neighbor-Joining representation of genetic relationships among the 17 genotypes.

The first two molecular principal components explained 34.7% of allelic-feature variation (PC1 = 20.4%; PC2 = 14.3%). This modest two-axis representation is insufficient to display the full relationship structure and may reflect both limited marker number and moderate polymorphism; PC3 and PC4 increased cumulative representation to 53.8%. T-1002 occupied an isolated position in the first two axes, while several core genotypes clustered more closely. Using the post hoc UPGMA groups, AMOVA assigned 42% of variation among groups and 58% within groups (PhiST = 0.42, permutation $p < 0.001$; 999 permutations; Table 6). The high PhiST and corresponding low Nm approximation (0.35) largely reflect marker-defined, non-independent groups in a small breeding panel. They should not be interpreted as evidence of restricted contemporary gene exchange among Uzbek breeding programs. The molecular PCA is illustrated in Figure 4.

**Table 5.** Summary of genetic distance and molecular differentiation.

| Parameter | Value |
|---|---|
| Minimum genetic distance | 0.214 |
| Maximum genetic distance | 0.762 |
| Mean genetic distance | 0.487 |
| Broad marker-defined groups | 3 |
| Most similar pair | T-2000 × T-1005 |
| Most divergent pair | T-1004 × Mehnat |

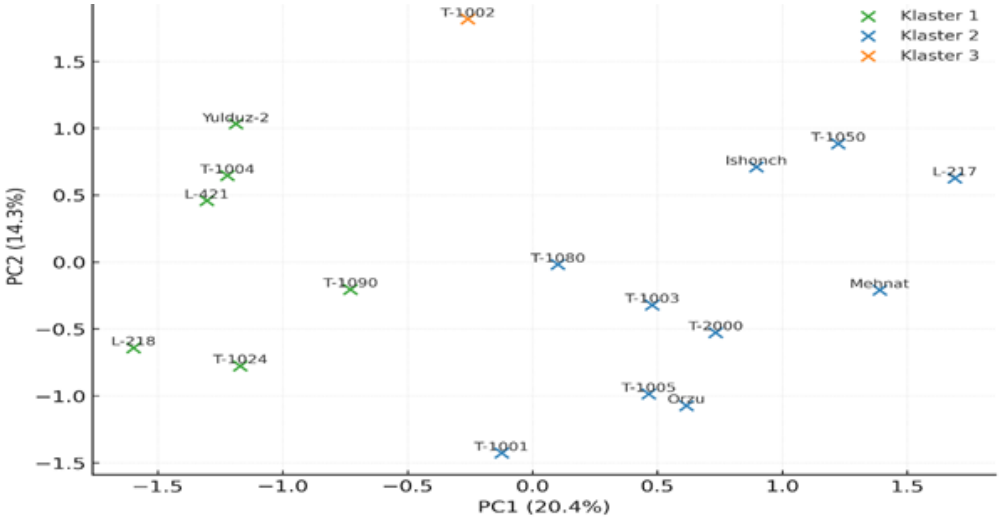



**Figure 4.** Molecular PCA of 17 upland cotton genotypes; PC1 and PC2 explain 20.4% and 14.3% of variation, respectively.

**Table 6.** Analysis of molecular variance among SSR-defined groups.

| Source | df | Sum of Squares | Variance Component | Variation (%) | Φ-Statistic |
|---|---|---|---|---|---|
| Among groups | 2 | 86.74 | 4.21 | 42.0 | ΦST = 0.42*** |
| Within groups | 14 | 81.57 | 5.81 | 58.0 | — |
| Total | 16 | 168.31 | 10.02 | 100 | — |

*** $p < 0.001$ based on permutation testing.

A complementary principal coordinate analysis based on the same Jaccard matrix showed Mehnat, T-1002, and several L-series materials at relatively isolated positions, while T-2000, T-1003, T-1005, and Orzu occupied a more central region. Hierarchical analysis of the allele-by-genotype matrix resolved four finer subclusters within the three broad groups. This distinction reflects analytical resolution: the broad UPGMA/NJ structure comprised three major groups, whereas the heatmap subdivided the collection at a finer cut level. Allele-presence patterns at BNL3545, NAU3588, BNL1421, NAU5434, and TMB1295 contributed visibly to genotype separation. The corresponding ordination is presented in Figure 5, and the molecular principal-component statistics are summarized in Table 7. The allele-presence patterns and four finer subclusters are shown in Figure 6.

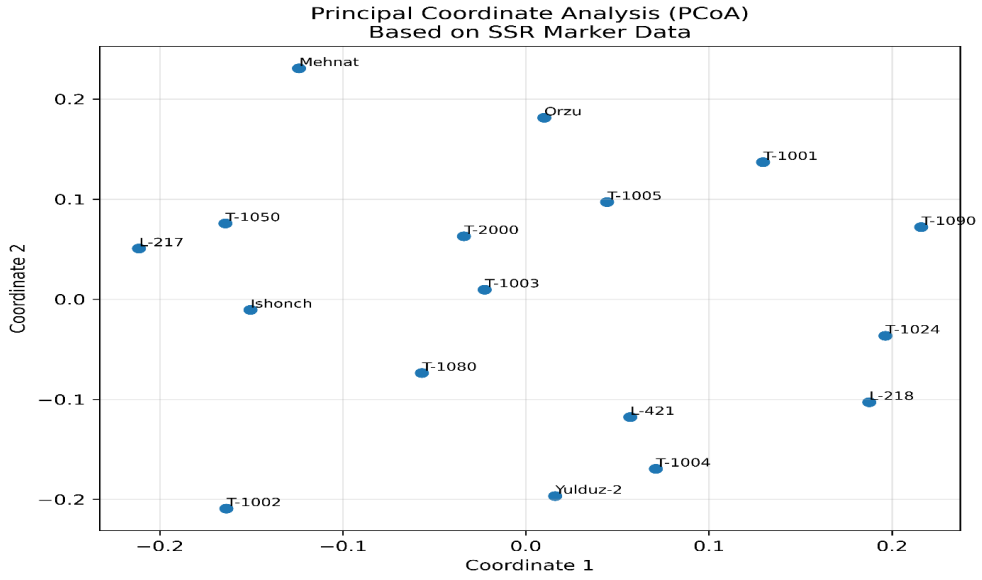



**Figure 5.** Principal coordinate analysis of the 17 genotypes based on Jaccard distances calculated from binary SSR allele profiles.

**Table 7.** Molecular principal components.

| Component | Eigenvalue | Variance (%) | Cumulative (%) |
|---|---|---|---|
| PC1 | 4.08 | 20.4 | 20.4 |
| PC2 | 2.86 | 14.3 | 34.7 |
| PC3 | 2.14 | 10.7 | 45.4 |
| PC4 | 1.68 | 8.4 | 53.8 |

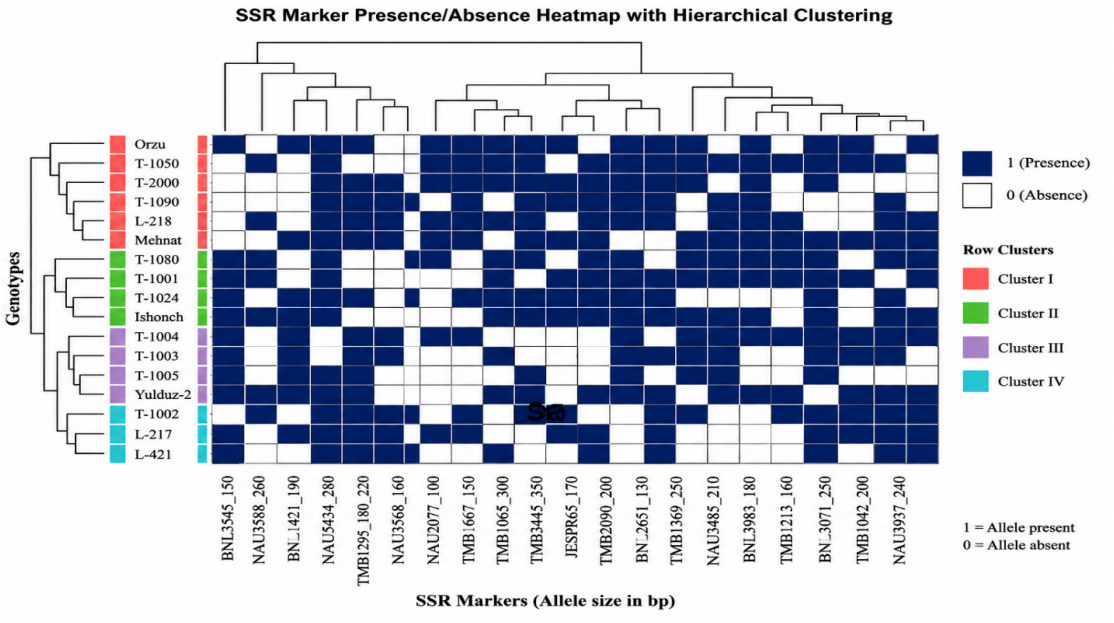



**Figure 6.** Hierarchical heatmap of SSR allele presence (blue) and absence (white), showing four fine-scale genotype subclusters.

### *3.6. Candidate SSR Allele–Trait Relationships*

Seven nominal relationships involving six loci connected SSR allelic states with agronomic or fiber measurements under saline conditions (Table 8). The NAU3588 260-bp allele was associated with both greater plant height (65.0 vs. 47.0 cm) and fiber strength (31.0 vs. 28.0 g tex$^{-1}$). BNL3545 150 bp was associated with greater boll weight, while BNL1421 190 bp and NAU5434 280 bp were associated with greater seed cotton yield. The TMB1295 180/220-bp profile was associated with a small increase in UHML. Conversely, JESPR65 170 bp was associated with a higher SFI and was unfavorable in this panel.

**Table 8.** Candidate allele–trait relationships detected in the 17-genotype panel.

| Locus | Allelic State | Trait | Carrier vs. Non-Carrier Mean | Direction |
|---|---|---|---|---|
| NAU3588 | 260 bp | Plant height | 65.0 vs. 47.0 cm | Favorable |
| BNL3545 | 150 bp | Boll weight | 3.90 vs. 3.67 g | Favorable |
| BNL1421 | 190 bp | Seed cotton yield | 9.3 vs. 6.9 g plant$^{-1}$ | Favorable |
| NAU5434 | 280 bp | Seed cotton yield | 10.5 vs. 8.3 g plant$^{-1}$ | Favorable |
| JESPR65 | 170 bp | Short fiber index | 9.4 vs. 8.0% | Unfavorable |
| NAU3588 | 260 bp | Fiber strength | 31.0 vs. 28.0 g tex$^{-1}$ | Favorable |
| TMB1295 | 180/220 bp | UHML | 1.07 vs. 1.03 inches | Favorable |

Relationships were detected at the nominal threshold (unadjusted $p < 0.05$). FDR-adjusted $p$-values were not available in the source analyses; therefore, these signals are candidates, not validated diagnostic markers.

### *3.7. NAU3588 Associations with Plant Height and Fiber Strength*

Carriers of the NAU3588 260-bp allele averaged 65.0 cm in plant height under salinity, compared with 47.0 cm in non-carriers, an absolute difference of 18.0 cm (38.3%). The same allelic state was associated with mean fiber strength of 31.0 g tex$^{-1}$ versus 28.0 g tex$^{-1}$. The parallel favorable pattern makes NAU3588 a priority candidate, but the two responses must be validated independently because the signal may reflect linkage or genetic background rather than pleiotropy. The corresponding comparisons are shown in Supplementary Figures S2 and S3.

### *3.8. Yield-Component Associations*

The BNL3545 150-bp allele was associated with a modest increase in boll weight: carriers averaged 3.90 g and non-carriers 3.67 g (6.3% difference). For seed cotton yield, BNL1421 190-bp carriers averaged 9.3 g plant$^{-1}$ compared with 6.9 g plant$^{-1}$ in non-carriers. NAU5434 280-bp carriers averaged 10.5 g plant$^{-1}$ compared with 8.3 g plant$^{-1}$. These effects identify loci for targeted validation in segregating populations rather than markers ready for direct deployment. The distributions and mean comparisons are provided in Supplementary Figures S4–S6.

### *3.9. Fiber-Length and Short-Fiber Associations*

The TMB1295 180/220-bp profile was associated with UHML of 1.07 inches in carriers and 1.03 inches in non-carriers (approximately 3.9% difference). Although small, this difference may be technologically relevant if stable across environments. In contrast, carriers of JESPR65 170 bp had an SFI of 9.4%, compared with 8.0% among non-carriers, an unfavorable increase of 17.5%. Absence of this JESPR65 allele was therefore the favorable state in the current panel. The corresponding fiber-trait comparisons are presented in Supplementary Figures S7 and S8.

## 4. Discussion

The sharp reduction in yield relative to plant height shows that reproductive processes were more sensitive to salinity than vegetative stature. Osmotic restriction and ion imbalance can limit assimilate supply to developing bolls and reduce both boll filling and final yield (Sikder et al., 2020; Chaudhary et al., 2024; Guo et al., 2021). Significant G × S effects confirm that genotype rankings depend on salinity intensity. Orzu, Mehnat, T-1090, and T-2000 merit further evaluation as productivity donors, whereas T-1001, L-217, and T-1024 contribute fiber stability. Crosses between these groups may combine yield retention with acceptable technological quality.

The detection of 53 alleles and mean PIC of 0.436 indicates moderate SSR diversity comparable to cultivated upland cotton panels reported elsewhere (Abdurakhmonov et al., 2008; Kuang et al., 2022; Tyagi et al., 2014; Fatima et al., 2025). NAU2277 and MUCS223 were especially informative for genotype discrimination. However, informativeness should not be confused with trait linkage: NAU2277 was the best diversity marker but was not among the candidate trait-associated loci. Additional cotton genetic-diversity and marker–trait studies provide relevant comparative context (Chapara et al., 2025; Iqbal et al., 2023; Normamatov et al., 2026).

The relationship analyses were only partially concordant. UPGMA, Neighbor-Joining, and PCA recovered broad differentiation and some similar close relationships, but differed notably in treatment of T-1002 and exact group membership. Such sensitivity is expected with 17 entries, 19 polymorphic loci, binary recoding, possible uneven genome distribution, and absence of branch-support resampling. The AMOVA among-group component is inflated as independent evidence because its groups were derived from the same marker matrix. Accordingly, the trees and ordination are used to generate crossing hypotheses, not to define stable populations. Phenotypically complementary crosses may include a yield-maintaining parent such as T-2000 or Mehnat and a fiber-stable parent such as T-1001, L-217, or T-1024, but divergence alone does not guarantee favorable salinity response.

The candidate allele features provide hypotheses for validation rather than ready-to-use selection assays (Abdurakhmonov et al., 2007; Sun et al., 2018). NAU3588 is notable because the 260-bp feature was related to two distinct traits, yet pleiotropy cannot be inferred from this small panel. The pattern could reflect linkage disequilibrium, unmodeled genetic relatedness, or shared ancestry. Similarly, the yield effects at BNL1421 and NAU5434 and the fiber effects at TMB1295 and JESPR65 require confirmation in segregating populations and independent environments. Multiple-testing correction is particularly important: nominal $p < 0.05$ associations can include false positives when many alleles and traits are examined. Marker–trait discovery for lint-yield components has likewise been explored by Kumar et al. (2021).

A practical validation pipeline should genotype the candidate allele features in larger biparental or association populations, fit mixed models that account for kinship and any independently estimated structure, report raw and false-discovery-rate-adjusted probabilities, and test the direction and magnitude of effects across locations and years. Only reproducible signals should subsequently be converted into robust PCR or KASP assays and evaluated for selection efficiency.

The study used only 17 deliberately selected materials and two field replications; it neither constitutes nor statistically represents the national Uzbek upland-cotton gene pool. Verified pedigrees and development/

release years were unavailable in the analytical dataset. Natural field salinity improves agronomic realism but may include spatial heterogeneity and correlated soil factors. The targeted SSR panel had only 19 polymorphic loci, was not chromosome-balanced, and was resolved on 3% agarose rather than PAGE or capillary electrophoresis. Binary allele-feature recoding discarded codominant genotype information, preventing valid Ho and Hardy-Weinberg analyses and making Jaccard-based results exploratory. No bootstrap values were available for the trees. The first two molecular PCs explained only 34.7% of variation. AMOVA groups were inferred from the same marker data, so PhiST and Nm are descriptive and not independent population-genetic evidence. Finally, FDR-adjusted association statistics were unavailable. These limitations constrain representativeness, causal inference, and predictive use; independent high-resolution genotyping in a larger, pedigree-documented, genome-wide panel is required.

## 5. Conclusions

Integrated phenotypic and targeted SSR analyses revealed variation within the 17-material working panel under natural soil salinity. Orzu, Mehnat, T-1090, and T-2000 retained comparatively higher productivity, while T-1001, L-217, and T-1024 contributed stability for selected fiber characteristics. The marker panel detected 53 allele-sized bands, with NAU2277, MUCS223, BNL3545, JESPR65, NAU6315, and DPL0530 providing high discrimination within this panel. The three post hoc marker-defined groups, PhiST = 0.42, and Nm = 0.35 are exploratory summaries and do not establish national population differentiation or contemporary gene flow.

Preliminary favorable signals were observed at NAU3588, BNL3545, BNL1421, NAU5434, and TMB1295, whereas JESPR65 170 bp corresponded to an unfavorable increase in SFI. None of these features should yet be used diagnostically. Validation requires high-resolution codominant genotyping, larger structured populations, independent environments, correction for multiple testing, and models controlling relatedness. At present, the defensible breeding use of the study is to prioritize phenotypically complementary parents and test the candidate loci in segregating progeny.

## Supplementary Materials

The additional data and information can be downloaded at: https://media.sciltp.com/articles/others/2611021442362484/AGRBIO-26070138-SM.pdf. Table S1. Genetic diversity parameters of SSR markers in *G. hirsutum* genotypes. Table S2. Summary of Neighbor-Joining phylogenetic groups identified among *Gossypium hirsutum* genotypes. Figure S1. Distribution of PIC values among SSR loci. Figure S2. Distribution and mean comparison of plant height for genotypes carrying (C1) or lacking (C2) the NAU3588 260-bp allele. Figure S3. Distribution and mean comparison of fiber strength for genotypes carrying (C1) or lacking (C2) the NAU3588 260-bp allele. Figure S4. Distribution and mean comparison of boll weight for carriers (C1) and non-carriers (C2) of the BNL3545 150-bp allele. Figure S5. Distribution and mean comparison of seed cotton yield for carriers (C1) and non-carriers (C2) of the BNL1421 190-bp allele. Figure S6. Distribution and mean comparison of seed cotton yield for carriers (C1) and non-carriers (C2) of the NAU5434 280-bp allele. Figure S7. Distribution and mean comparison of UHML for genotypes with (C1) or without (C2) the TMB1295 180/220-bp profile. Figure S8. Distribution and mean comparison of SFI for genotypes carrying (C1) or lacking (C2) the JESPR65 170-bp allele.

## Author Contributions

J.S.: conceptualization, methodology, formal analysis, data curation, writing—original draft preparation, and supervision; A.A.: investigation and field experimentation; F.A.: validation and writing—reviewing and editing; M.A.: field investigation, soil analysis, data validation, and writing—reviewing and editing. All authors have read and agreed to the published version of the manuscript.

## Funding

This research received no external funding.

## Institutional Review Board Statement

Not applicable. This study involved only plant materials (*Gossypium hirsutum* L.) and did not involve human participants or animals. Therefore, institutional ethical review and approval were not required.

## Informed Consent Statement

Not applicable. This study did not involve human participants or patients; therefore, informed consent was not required.

## Data Availability Statement

The phenotypic and SSR datasets and the statistical outputs supporting this study are available from the corresponding author upon reasonable request and will be retained for at least 10 years. Statistical software, analytical procedures, permutation settings, and the handling of multiple testing are described in the Materials and Methods section. No observations were intentionally omitted from the reported analyses.

## Acknowledgments

The authors acknowledge the technical and scientific support of the Institute of Genetics and Plant Experimental Biology, Academy of Sciences of the Republic of Uzbekistan, and the assistance of the experimental-station staff in field management, sampling, and laboratory analyses.

## Conflicts of Interest

The authors declare no conflict of interest.

## Use of AI and AI-Assisted Technologies

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) to assist with English language editing, grammatical corrections, and improving the clarity and readability of the text. All scientific content, experimental data, statistical analyses, interpretations, and conclusions were reviewed and verified by the authors. The authors take full responsibility for the accuracy, originality, and integrity of the published work.

## References

Abdurakhmonov, I. Y., Buriev, Z. T., Saha, S., Pepper, A. E., Musaev, J. A., Almatov, A., Shermatov, S. E., Kushanov, F. N., Mavlonov, G. T., Reddy, U. K., Yu, J. Z., Jenkins, J. N., Kohel, R. J., & Abdukarimov, A. (2007). Microsatellite markers associated with lint percentage trait in cotton, *Gossypium hirsutum*. *Euphytica*, *156*(1–2), 141–156. https://doi.org/10.1007/s10681-007-9361-2.

Abdurakhmonov, I. Y., Kohel, R. J., Yu, J. Z., Pepper, A. E., Abdullaev, A. A., Kushanov, F. N., Salakhutdinov, I. B., Buriev, Z. T., Saha, S., Scheffler, B. E., Jenkins, J. N., & Abdukarimov, A. (2008). Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. *Genomics*, *92*(6), 478–487. https://doi.org/10.1016/j.ygeno.2008.07.013.

Botstein, D., White, R. L., Skolnick, M. H., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics, 32*(3), 314–331.

Chaudhary, M. T., Majeed, S., Rana, I. A., Ali, Z., Jia, Y., Du, X., Hinze, L., & Azhar, M. T. (2024). Impact of salinity stress on cotton and opportunities for improvement through conventional and biotechnological approaches. *BMC Plant Biology*, *24*(1), 97. https://doi.org/10.1186/s12870-023-04558-4.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, *131*(2), 479–491. https://doi.org/10.1093/genetics/131.2.479.

Fatima, S. Y., Luqman, T., Hussain, S., Nadeem, S., Khan, M. K. R., Ditta, A., & Asif, M. (2025). Genetic diversity and phylogenetic relationship through the use of microsatellites in *Gossypium hirsutum* L. *BMC Plant Biology*, *25*(1), 418. https://doi.org/10.1186/s12870-025-06363-7.

Gomez, K. A., & Gomez, A. A. (1984). *Statistical Procedures for Agricultural Research* (2nd ed.). John Wiley & Sons.

Guo, A. H., Su, Y., Huang, Y., Wang, Y. M., Nie, H. S., Zhao, N., & Hua, J. P. (2021). QTL controlling fiber quality traits under salt stress in upland cotton (*Gossypium hirsutum* L.). *Theoretical and Applied Genetics*, *134*(2), 661–685. https://doi.org/10.1007/s00122-020-03721-x.

Jaccard, P. (1908). Nouvelles recherches sur la distribution florale. *Bulletin de la Société vaudoise des sciences naturelles*, *44*(163), 223–270. https://doi.org/10.5169/seals-268384.

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). Springer-Verlag. https://doi.org/10.1007/b98835.

Kuang, Z., Xiao, C., Ilyas, M. K., Ibrar, D., Khan, S., Guo, L., Wang, W., Wang, B., Huang, H., Li, Y., Li, Y., Zheng, J., Saleem, S., Tahir, A., Ghafoor, A., & Chen, H. (2022). Use of SSR Markers for the Exploration of Genetic Diversity and DNA Finger-Printing in Early-Maturing Upland Cotton (*Gossypium hirsutum* L.) for Future Breeding Program. *Agronomy*, *12*(7), 1513. https://doi.org/10.3390/agronomy12071513.

Lacape, J. M., Dessauw, D., Rajab, M., Noyer, J. L., & Hau, B. (2006). Microsatellite diversity in tetraploid Gossypium germplasm: Assembling a highly informative genotyping set of cotton SSRs. *Molecular Breeding*, *19*(1), 45–58. https://doi.org/10.1007/s11032-006-9042-1.

Liu, K., & Muse, S. V. (2005). PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics*, *21*(9), 2128–2129. https://doi.org/10.1093/bioinformatics/bti282.

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—An update. *Bioinformatics*, *28*(19), 2537–2539. https://doi.org/10.1093/bioinformatics/bts460.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454.

Sikder, R. K., Wang, X., Jin, D., Zhang, H., Gui, H., Dong, Q., Pang, N., Zhang, X., & Song, M. (2020). Screening and evaluation of reliable traits of upland cotton (*Gossypium hirsutum* L.) genotypes for salt tolerance at the seedling growth stage. *Journal of Cotton Research*, *3*(1), 11. https://doi.org/10.1186/s42397-020-00049-1.

Slatkin, M. (1985). Gene Flow in Natural Populations. *Annual Review of Ecology and Systematics*, *16*(1), 393–430. https://doi.org/10.1146/annurev.ecolsys.16.1.393.

Sun, Z., Li, H., Zhang, Y., Li, Z., Ke, H., Wu, L., Zhang, G., Wang, X., & Ma, Z. (2018). Identification of SNPs and Candidate Genes Associated with Salt Tolerance at the Seedling Stage in Cotton (*Gossypium hirsutum* L.). *Frontiers in Plant Science*, *9*, 1011. https://doi.org/10.3389/fpls.2018.01011.

Tyagi, P., Gore, M. A., Bowman, D. T., Campbell, B. T., Udall, J. A., & Kuraparthy, V. (2014). Genetic diversity and population structure in the US Upland cotton (*Gossypium hirsutum* L.). *Theoretical and Applied Genetics*, *127*(2), 283–295. https://doi.org/10.1007/s00122-013-2217-3.

Wright, S. (1978). *Evolution and the Genetics of Populations, Volume 4: Variability Within and Among Natural Populations* (Vol. 4). University of Chicago Press.