

RESEARCH ARTICLE

Genetic diversity, linkage disequilibrium, and association mapping analyses of *Gossypium barbadense* L. germplasm

Alisher A. Abdullaev^{1,2}✉, Ilkhom B. Salakhutdinov¹, Sharof S. Egamberdiev¹, Ernest E. Khurshut¹, Sofiya M. Rizaeva², Mauricio Ulloa³✉, Ibrokhim Y. Abdurakhmonov¹✉*

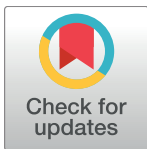
1 Center of Genomics and Bioinformatics, Academy of Sciences of Uzbekistan, Tashkent, Uzbekistan,

2 Institute of Genetics and Plant Experimental Biology, Academy of Sciences of Uzbekistan, Tashkent, Uzbekistan, **3** Cropping Systems Research Laboratory, United States Department of Agriculture -

Agricultural Research Services, Lubbock, Texas, United States of America

✉ These authors contributed equally to this work.

* genomics@uzsci.net



OPEN ACCESS

Citation: Abdullaev AA, Salakhutdinov IB, Egamberdiev SS, Khurshut EE, Rizaeva SM, Ulloa M, et al. (2017) Genetic diversity, linkage disequilibrium, and association mapping analyses of *Gossypium barbadense* L. germplasm. PLoS ONE 12(11): e0188125. <https://doi.org/10.1371/journal.pone.0188125>

Editor: Harsh Raman, New South Wales Department of Primary Industries, AUSTRALIA

Received: June 23, 2017

Accepted: November 1, 2017

Published: November 14, 2017

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the Academy of Sciences of Uzbekistan, and Committee for Coordination Science and Technology Development of Uzbekistan have provided research grant fund under the project No. FA-F5-T030 to IYA and the Office of International Research Programs (OIRP) of the United States

Abstract

Limited polymorphism and narrow genetic base, due to genetic bottleneck through historic domestication, highlight a need for comprehensive characterization and utilization of existing genetic diversity in cotton germplasm collections. In this study, 288 worldwide *Gossypium barbadense* L. cotton germplasm accessions were evaluated in two diverse environments (Uzbekistan and USA). These accessions were assessed for genetic diversity, population structure, linkage disequilibrium (LD), and LD-based association mapping (AM) of fiber quality traits using 108 genome-wide simple sequence repeat (SSR) markers. Analyses revealed structured population characteristics and a high level of intra-variability (67.2%) and moderate interpopulation differentiation (32.8%). Eight percent and 4.3% of markers revealed LD in the genome of the *G. barbadense* at critical values of $r^2 \geq 0.1$ and $r^2 \geq 0.2$, respectively. The LD decay was on average 24.8 cM at the threshold of $r^2 \geq 0.05$. LD retained on average distance of 3.36 cM at the threshold of $r^2 \geq 0.1$. Based on the phenotypic evaluations in the two diverse environments, 100 marker loci revealed a strong association with major fiber quality traits using mixed linear model (MLM) based association mapping approach. Fourteen marker loci were found to be consistent with previously identified quantitative trait loci (QTLs), and 86 were found to be new unreported marker loci. Our results provide insights into the breeding history and genetic relationship of *G. barbadense* germplasm and should be helpful for the improvement of cotton cultivars using molecular breeding and omics-based technologies.

Introduction

Cultivated cotton (*Gossypium* spp.) is the most important natural fiber worldwide. Fiber quality is a key factor for determining price and quality of cotton textile products, and is significantly affected by different environmental factors [1]. In addition, genetic improvement of

Department of Agriculture (USDA)—Agricultural Research Service (ARS) and U.S. Civilian Research and Development Foundation (CRDF) provided an international cooperative grant No. UZB-TA-31017 to IYA and MU. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

fiber quality is a challenge due to the narrow genetic base of modern cotton cultivars and the existence of negative correlations between major fiber quality traits and key agronomic characteristics [2, 3]. All of the above highlight a great need to study genetic resources preserved and maintained in world cotton germplasm collections [4], and to use these resources in breeding of superior cotton genotypes.

There are several major cotton germplasm collections in the world. One of the biggest and richest germplasm collections is housed in Uzbekistan with extensive genetic diversity [4, 5]. In the *Gossypium* genus, genetic diversity exists in its genome with unique traits or sometimes hidden elements, or genes that can have a positive impact on the expression of agronomic traits and resistance to biotic and abiotic factors. Introduction of valuable traits in modern cotton germplasm enriches and improves the diversity of cultivated cotton [4–8]. Genetic studies and evaluations of cotton germplasm resources provide specific information on the degree of phylogenetic relatedness of accessions in these collections and its/their representation. In addition, evaluations shed light on many questions of complex genetic traits that will eventually allow the use of the genetic potential of cotton germplasm for introduction of important and useful features/traits in modern cotton cultivars. For the introduction of important traits, marker-assisted selection (MAS) is one of the key and valuable tools for the introduction and introgression of useful traits.

Unfortunately, MAS in cotton lags behind other crops due to limited genetic polymorphism of cultivated cotton germplasm as a result of the historical process of domestication [9–11]. This also complicates the process of genetic mapping of quantitative trait loci (QTLs) associations with traits of interest using DNA markers. Moreover, much of the molecular-genomic researches including association studies and MAS focused on members of the species *G. hirsutum* [12–20]. This species supplies around 95.5% of the cotton production worldwide. Studies in other species unfortunately are limited as is the case with *G. barbadense* germplasm resources [21].

The *G. hirsutum* (also known as Upland cotton) and *G. barbadense* [also known as Sea Island, Egyptian, or extra-long staple (ELS) cotton] are the two main widely grown cotton species. Although *G. barbadense* only accounts for around 4.5% of the cotton production worldwide, this species is known for its superior fiber quality (length, fineness and strength). Its fiber is highly valued in the premium textile market [22, 23]. *G. barbadense* is indigenous to the northern part of South America and extends into Mesoamerica and the Caribbean [24]. In the United States, modern elite *G. barbadense* cultivars trace their origins to the Sea Island cottons developed on the coastal islands of Georgia and South Carolina that probably originated from west Andean Peruvian germplasm [24, 25]. The Sea Island cotton production collapsed in the USA under boll weevil (*Anthonomus grandis* Boheman) pressure in 1920 [7]. This Sea Island pool contributed to the development of the Egyptian cottons which in the 1940s were reintroduced into the USA as a part of the genetic base of the Pima gene pool by USDA-ARS [22, 24].

Molecular marker technology and QTL-mapping approach using bi-parental mapping populations resulted in a number of potential DNA markers for future breeding programs of cotton through MAS [11]. A classic QTL-mapping using bi-parental population exploits the short history of recombination and consequently QTL can only be localized to large chromosomal regions. By contrast, mapping approaches that exploit linkage disequilibrium make use of all recombination that have occurred during the breeding history resulting in much higher mapping resolution [26, 27]. The extent of genome-wide LD or allelic association is the key starting point for association mapping (AM). Quantification of the LD extent and AM have been successfully applied for many plant species [28, 29] including cotton [9, 30–33]. Recent LD-based

studies of *G. hirsutum* germplasm resulted in association mapping of *Verticillium* wilt resistance [18], salinity tolerance [31] and seed oil and protein contents [32].

Here, we for the first time report SSR marker-based genetic analyses of 288 *G. barbadense* germplasm accessions of Uzbekistan germplasm collection, grown and phenotypically evaluated in two diverse environments, Uzbekistan and USA. The molecular genetic characteristics and diversity, population structure, the extent of LD, and association mapping of main fiber quality traits of *G. barbadense* germplasm are reported based on genotypic data of a core set of 108 microsatellite markers, evenly distributed in the cotton genome. SSR marker loci were statistically associated with fiber quality traits specific to Uzbekistan and USA environments. Mixed linear model (MLM) analysis, considering confounding effects of structured population, detected 100 reliable SSRs, which were common and statistically significant in the two distinct environments, Uzbekistan and USA. The results of this study are, to the best of our knowledge, the first report on a genome-wide LD analysis and LD-based AM of fiber quality traits using SSR markers in *G. barbadense* germplasm resources of Uzbekistan. In addition, these findings are very useful for the application of association study in cotton and should accelerate the development of superior cotton cultivars through MAS programs.

Materials and methods

Plant materials

The Uzbek cotton germplasm collection houses and preserves more than 7500 accessions of different cotton species at the Institute of Genetics and Plant Experimental Biology (IG&PEB), Academy of Sciences of Uzbekistan. Out of 7500 cotton germplasm accessions, *G. barbadense* comprises approximately 13% with broad geographic and breeding coverage. A total of 288 *G. barbadense* germplasm accessions from Central Asian (237), African (35) and American (16) origin were selected from the Uzbek collection and used in this study, including for genome-wide LD and association mapping analyses.

Phenotypic analyses in the Uzbekistan environment

Analyses of morpho-biological characteristics of these selected *G. barbadense* germplasm and cultivars were performed in the field stations of the IG&PEB, Tashkent, Uzbekistan in 2010. Standard field plots, irrigation and agronomic technologies were used for growing cultivars in the Tashkent cultivation environment. Detailed information about weather data for specific years can be obtained from the archive of the meteorology center (<http://www.wunderground.com>). Ten plants of each cultivar were grown and self-pollinated by sealing flowers with florist wire just before the flowers opened. Cotton fiber samples from self-pollinated cotton bolls were harvested from field-grown plants in the beginning of October. At least 25 fully opened self-pollinated cotton bolls were harvested from each group of cultivars (pooled from ten plants per accession). Fiber quality traits of cultivars grown in Uzbekistan environment, such as fiber length (FL), fiber strength (FS), fiber micronaire (FM), and fiber uniformity (FU), were measured for 247 accessions (Table 1; S1 Data) by High Volume Instrument (HVI) of the certified “SIFAT” agency, Tashkent, Uzbekistan.

Phenotypic analyses in the California environment

In 2010, the germplasm and cultivars were evaluated under the San Joaquin Valley environment of California, Shafter Research Station (35°31'52" N 119°16'41" W), Kern County, USA. Accessions were grown in one-row plot, 5 by 1 meter, in a complete randomized design with a plant density ranging from 40 to 50 plants per plot. To examine fiber quality traits, 50 open

Table 1. Descriptive statistics of fiber traits among *G. barbadense* accessions grown in the Uzbekistan and USA environments.

Traits	No. of samples [†]	Mean	Min.	25 percentile	Median	75 percentile	Max.	10 percentile	90 percentile	SD	CV
Tashkent, Uzbekistan											
Micronaire[†]	247	4.25	3.0	3.9	4.3	4.6	6.3	3.6	4.9	0.51	12.22
Strength	247	38.81	26.4	37.4	38.8	40.4	49.3	35.7	42.3	2.95	7.62
Length*	247	1.30	0.95	1.26	1.32	1.36	1.5	1.19	1.4	0.08	6.51
Uniformity**	247	84.74	78.9	83.8	84.8	85.9	88.6	82.7	86.6	1.60	1.89
California, USA											
Micronaire	278	4.06	2.6	3.7	4.1	4.4	5.3	3.4	4.6	0.46	11.45
Strength	278	36.50	26.7	34.9	36.4	38.1	43.7	33.6	40.1	2.61	7.16
Length	278	1.36	1.0	1.32	1.36	1.41	1.58	1.26	1.45	0.08	6.04
Uniformity	278	87.30	81.6	86.5	87.5	88.3	90.0	85.5	89.0	1.42	1.63

[†]Samples evaluated in Uzbekistan and the USA represented a set of initially chosen 288 *G. barbadense* accessions;

[†]Characteristics and fineness of the cotton fiber maturity determined by the air permeability of the fiber sample, expressed as an index;

*Length of HVI according to the standards specified in inches (1 inch = 25.4 mm);

** Percentage of the average fiber length to the average upper length

<https://doi.org/10.1371/journal.pone.0188125.t001>

matured bolls were randomly harvested from different segments of the plant from different plants of each plot. After processing using a saw gin, samples were sent to the USDA Cotton Classing Laboratory, Visalia, CA for analyses. Fiber length (FL) in mm, fiber strength (FS) in kN m kg⁻¹, fiber micronaire (FM), and fiber uniformity (FU) in percentage were measured for 278 accessions (Table 1, S1 Data) by High Volume Instrument (HVI).

Due to use of wild-type genotypes of plant material, available in *ex situ* germplasm collection and commonly used for research, no authority permission was required to evaluate cotton germplasm resources in Uzbekistan and the USA. Plant material evaluated was not under the control of relevant regulatory bodies concerned with wildlife protection, and the field studies did not involve endangered or protected species. All field evaluations were conducted in the experimental stations, a priori assigned for research activities, which did not require a specific permission for conducting field evaluations.

Statistical analyses

Data were analyzed using analysis of variance (ANOVA) [34] for the different fiber values, and correlations analyses were performed to examine the similarity of value-responses of the different accessions and fiber traits in the two growing-environments, Uzbekistan and California.

SSR analysis

In total, 750 SSR primer pairs from different SSR collections were screened to detect polymorphisms among accessions. Genomic DNAs were isolated from leaf tissues of each germplasm or cultivar using the method of Dellaporta et al. [35]. Each accession was genotyped using 108 polymorphic SSR marker primers distributed an average of 4 SSR markers per each cotton chromosome. SSRs were chosen based on previous germplasm collection characterizations [9, 33], and based on information related to important QTLs and chromosome distribution.

Polymerase chain reaction (PCR) mixtures (10 µL) consisted of 1X reaction buffer, 1.5 mM MgCl₂, 0.2 mM dNTP, 0.3 µM primers, 25 ng template DNA, and 0.5 U Taq DNA polymerase (Applied Biosystems, Foster city, USA). Amplification was carried out in a GeneAmp 9700 (Applied Biosystems), with an initial denaturation at 95°C for 10 min, followed by 35 cycles of 1 min at 94°C, 1 min at X°C, and 1 min at 72°C, plus a 5-min final extension at 72°C. X°C

refers to the annealing temperature specified for each primer. The amplified products were separated on 3% (w/v) high resolution agarose gels (GeneMate, Radnor, USA) and visualized under UV light with ethidium bromide staining.

Genetic diversity and phylogenetic analyses

The amplified fragments of each SSR marker were scored based on fragment sizes (S2 Data). Polymorphism information content (PIC) of SSR markers was calculated using the PowerMarker software package [36]. The heterozygosity level of marker data was identified according to an average similarity frequency of alternative alleles [37]. Allele frequencies were calculated using SpaGeDi software [38]. Genetic distance and phylogenetic analyses of cotton cultivars were performed using Neighbor Joining (N-J) algorithms in PAUP* 4.0 [39]. Genetic variation within and among predefined groups and pair-wise F_{ST} genetic distances were measured by Analysis of Molecular Variance (AMOVA) [40–42] using ARLEQUIN 2.0 [43]. A Bayesian partition method of genetic differentiation among population groups was applied using HICKORY [44] software to direct estimation of F_{ST} without prior knowledge of inbreeding history [45].

Population structure and kinship analyses

The population structure of the 288 *G. barbadense* germplasm and cultivars was assessed using the model-based (Bayesian clustering) method implemented in STRUCTURE v2.3.3 [46]. The number of subgroups (K) was set from 1 to 12 based on models characterized by admixture and correlated allele frequencies. For each K, ten runs were performed separately, with 100,000 iterations carried out for each run after a burn-in period of 10,000 iterations. The true number of sub-populations was estimated using the method proposed by Evanno et al. [47]. Accessions were assigned to a subgroup if the probability of membership was greater than 70% [48]. A pairwise kinship (K-matrix) estimate for 288 *G. barbadense* accessions was calculated according to Hardy [49] using the software package SpaGeDi [38].

Pair-wise linkage disequilibrium and LD decay

The genome-wide LD between pairs of SSR marker loci in the *G. barbadense* genome was studied according to Whitt and Buckler [48] using the software package TASSEL [50]. SSR alleles with a 0.05 frequency in genotyped accessions were removed before conducting LD analyses because minor alleles are usually problematic and biased for LD estimates between pairs of loci [51, 52]. LD was estimated by a weighted average of squared allele-frequency correlations (r^2) between SSR loci. Loci were considered to be significant at p -values ≤ 0.005 among all possible SSR loci. LD was evaluated with the rapid permutation test in 10,000 shuffles. Values of LD between all pairs of SSR loci were plotted as triangle LD plots using TASSEL to estimate the general view of genome-wide LD patterns and evaluate 'block-like' LD structures. The r^2 values for pairs of SSR loci were plotted as a function of map distances (cM), and LD decay (at $r^2 > 0.1$) was estimated [48].

Association mapping of fiber quality traits

Association mapping using the mixed linear model (MLM) and general linear models (GLM) was performed for both environments and for the four major fiber quality traits data [fiber length (FL), fiber strength (FS), fiber uniformity (FU), and fiber micronaire (FM); S1 Data]. To construct marker-fiber quality trait associations using SSR and fiber data (S1 and S2 Data), the MLM test was performed according to Yu et al. [53] using the TASSEL software package

[50]. The MLM association test was simultaneously performed by accounting of multiple levels of population structure (Q-matrix) and relative kinship among the individuals (K-matrix) [50–55].

The 5% of ‘minor alleles’ filtered-SSR datasets were used for all association mapping models. Fiber trait data was imputed for missing data and normalized using algorithms implemented in TASSEL before conducting an association mapping analysis. The MLM-derived p-values were separately tested for multiple testing correction using pFDR test in QVALUE program version 1.0 [56], Sidak procedure of Bonferroni adjustment, and pACT method of Conneely and Boehnke [57]. To reliably interpret the MLM-derived significant associations, a minimum Bayes factor (BF_{min}) was calculated using the following formula: $BF_{min} = -e * p * \ln(p)$ [33, 58, 59–61]. Moreover, the MLM-derived significant associations were also subjected to comparisons with published literature information to judge obtained associations.

Results

Fiber quality properties of *G. barbadense* cultivars in the USA and Uzbekistan environments

Due to missing research plots of some accessions and/or technical errors during HVI analyses, major fiber trait measurements for 247 and 278 accessions were obtained in the Uzbekistan and USA environments, respectively (Table 1). Herein, we reference 288 accessions as a total number for the molecular set panel investigated in these two environments. A wide-range of phenotypic variation in fiber quality traits such as FL, FM, FU, FS was observed in both environments (Table 1).

The coefficient of experimental variability of traits in California (USA) and Tashkent (Uzbekistan) conditions ranged from as low as 1.63–1.89 (FU) and as high as 11.45–12.22 (FM) for the above traits. Thus, micronaire was the most variable (2.6 to 6.3) trait of all the fiber quality parameters and showed similar values of coefficient of variation in both environments. The lower variations of micronaire values (2.6 to 3.4) were observed in California’s environment, while in Tashkent, values shifted toward the high (from 4.6 to 6.3) values.

Fiber strength for all *G. barbadense* accessions in the Tashkent environment ranged from 270.0 to 503.5 kN m kg⁻¹ (27.0 g/tex⁻¹ to 50.35 g/tex⁻¹) with an average value of 396.0 kN m kg⁻¹ (39.6 g/tex⁻¹; SD = 29.6). In California, the FS had minimum and maximum values equal to 273.0 and 446.0 kN m kg⁻¹ (27.3 g/tex⁻¹ to 44.6 g/tex⁻¹) respectively, with an average of 373.0 kN m kg⁻¹ (37.3 g/tex⁻¹; SD = 26.1).

Moreover, comparisons of fiber traits among *G. barbadense* accessions from the four major geographical groups (Uzbekistan, Turkmenistan, the United States, and Africa) showed variations in Tashkent and California environments (Table 2).

Fiber trait correlations

The correlation analyses of the fiber traits of *G. barbadense* germplasm and cultivars in Uzbekistan and California environments (Table 3) showed the presence of significant positive and negative relationships between the traits studied.

Positive correlation was observed between FL and FU, FL and FS, and FU and FS. The negative correlations were observed between FM and FS (not significant in California), FM and FL, FM and FU in both environments. Significant trait correlations were observed between the same fiber traits as well as among different fiber traits in Tashkent and California-grown accessions (Table 4).

Table 2. Comparison of fiber traits among various geographical groups of *G. barbadense* cultivars evaluated in the Uzbekistan and USA environments.

Germplasm origin	Micronaire		Strength		Length		Uniformity	
	Uzb	USA	Uzb	USA	Uzb	USA	Uzb	USA
Africa	$\bar{X} = 4.26$ S = 0.59	$\bar{X} = 4.1$ S = 0.44	$\bar{X} = 37.85$ S = 3.77	$\bar{X} = 36.01$ S = 3.21	$\bar{X} = 1.29$ S = 0.10	$\bar{X} = 1.35$ S = 0.11	$\bar{X} = 84.30$ S = 1.91	$\bar{X} = 87.1$ S = 1.72
Uzbekistan	$\bar{X} = 4.31$ S = 0.49	$\bar{X} = 4.1$ S = 0.49	$\bar{X} = 39.08$ S = 2.44	$\bar{X} = 36.78$ S = 2.09	$\bar{X} = 1.31$ S = 0.08	$\bar{X} = 1.36$ S = 0.07	$\bar{X} = 84.89$ S = 1.52	$\bar{X} = 87.3$ S = 1.17
USA	$\bar{X} = 4.38$ S = 0.64	$\bar{X} = 4.0$ S = 0.44	$\bar{X} = 38.39$ S = 3.09	$\bar{X} = 36.83$ S = 2.09	$\bar{X} = 1.28$ S = 0.07	$\bar{X} = 1.35$ S = 0.07	$\bar{X} = 83.66$ S = 2.13	$\bar{X} = 87.2$ S = 0.80
Turkmenistan	$\bar{X} = 4.18$ S = 0.51	$\bar{X} = 4.0$ S = 0.47	$\bar{X} = 38.57$ S = 3.11	$\bar{X} = 36.32$ S = 3.01	$\bar{X} = 1.31$ S = 0.09	$\bar{X} = 1.36$ S = 0.09	$\bar{X} = 84.77$ S = 1.57	$\bar{X} = 87.3$ S = 1.61

Note: \bar{X} —mean; S—variance. A number of accessions for Uzbekistan (247) and the USA (278) environments were as specified in Table 1.

<https://doi.org/10.1371/journal.pone.0188125.t002>

Table 3. Correlations between fiber traits of *G. barbadense* cultivars in the Tashkent and California environments.

Fiber traits	FM	FS	FL
Tashkent, Uzbekistan			
FM	1		
FS	-0.258**	1	
FL	-0.560**	0.505**	1
FU	-0.297**	0.634**	0.610**
California, USA			
FM	1		
FS	-0.068	1	
FL	-0.427 **	0.155*	1
FU	-0.241 **	0.219**	0.774**

* Correlation is significant at a value $r \leq 0.05$;

** Correlation is significant at a value $r \leq 0.001$; FM—fiber micronaire, FS—fiber strength, FL—fiber length, FU—fiber uniformity

<https://doi.org/10.1371/journal.pone.0188125.t003>

These trait-correlations revealed the importance of the environment influencing fiber development. Similar results or similar pattern of variability for fiber traits were observed based on the analysis of variance (ANOVA), in which environmental growth condition impacted the fiber trait differences. ANOVA also revealed that the differences between these groups (Uzbekistan and California) strictly depended on fiber traits' growing conditions (Table 5).

Table 4. Comparative analysis of fiber traits correlation depending on the growing conditions.

Environment_Fiber traits	US_FM	US_FS	US_FL	US_FU
UZH_FM	0.603(**)	-0.144(*)	-0.491(**)	-0.374(**)
UZH_FS	-0.096	0.520(**)	0.278(**)	0.280(**)
UZH_FL	-0.372(**)	0.249(**)	0.689(**)	0.548(**)
UZH_FU	-0.170(**)	0.267(**)	0.324(**)	0.339(**)

* Correlation is significant at $p = 0.05$;

** Correlation is significant at $p = 0.001$

<https://doi.org/10.1371/journal.pone.0188125.t004>

Table 5. ANOVA¹ results of fiber traits depending on the growth in Uzbekistan (Tashkent) and the USA (California).

Statistical analyses	FM	FS	FL	FU
χ^2 ; <i>p-value</i>	18.84; 0.000014	97.10; 0.000000	57.62; 0.000000	244.92; 0.000000
F-Ratio; <i>p-value</i>	23.15; 0.000002*	94.15; 0.000000*	57.88; 0.000000*	390.55; 0.000000*
Power ($\alpha = 0.05$)	0.9977	1	1	1
SS country	5.61	724.10	0.397	892.66
SS S(A)	131.51	4176.40	3.726	1241.13
MS country	5.61	724.10	0.397	892.66
MS S(A)	0.24	7.69	6.86E-03	2.28
SS Total	137.12	4900.51	4.123	2133.79
DF country	1	1	1	1
DF S(A)	574	574	574	574

¹to determine the statistical significance the F-value used for reliable level of $\alpha = 0.05$.

*values are significant at $\alpha = 0.05$

<https://doi.org/10.1371/journal.pone.0188125.t005>

Growing conditions have a direct impact on the performance of the fiber. In Uzbekistan, all accessions showed low values of FL and FU, and high values of FM and FS. Thus, in California, traits were observed to be most favorable for FL, FU, and FM in *G. barbadense* germplasm studied. Several accessions with strong stability for a single trait and/or all traits in the Tashkent and California environments were identified (Table 6).

Marker analysis

From the 750 SSR primer pairs, 108 (14%) were found to be polymorphic among *G. barbadense* germplasm and cultivars. Identified 108 SSRs primer pairs amplified 301 marker loci in our *G. barbadense* panel (S2 Data). The number of alleles ranged from 2 to 5 with an average number of 2.78 allele per SSR. Sixty SSRs (55%) amplified three or more alleles. The majority of SSRs (81%) were represented by two or three alleles on these accessions (Table 7). The

Table 6. Samples having a strong stability in the Uzbekistan and USA.

Traits	Number of cultivars	Value
Fiber strength	92	> 37g/tex
Fiber micronaire	41	$\geq 3.7 \leq 4.2$
Fiber length	9	≥ 1.5 inches
Fiber uniformity	7	> 87%
Combination of all traits	7	

<https://doi.org/10.1371/journal.pone.0188125.t006>

Table 7. A distribution of alleles among the 108 SSR markers.

Number of SSRs	Alleles	Total Alleles
48	2	96
40	3	120
15	4	60
5	5	25
Total: 108		301

<https://doi.org/10.1371/journal.pone.0188125.t007>

average polymorphic information content (PIC) among the markers was 0.29 (SD = 0.16). Mean heterozygosity (H) for all markers among the 288 accessions of *G. barbadense* was 0.33 (SD = 0.2), with the minimum and maximum values of 0.02 and 0.71, respectively. Of the 108 markers, in 49 (45.4%) markers, the heterozygosity values ranged from 0.02–0.25 and in 59 (54.6%) markers, the values ranged from 0.25–1.0.

Comparative analysis of heterozygosity and PIC values showed higher values in heterozygosity over PIC in each of the markers with an average increase of about 10.1%. Analysis of the distribution of frequencies of polymorphic alleles showed that the average frequency was 0.424 (SD = 0.33) with minimum and maximum values of 0.02 and 0.996, respectively. Out of 301 amplified/identified marker-alleles, 189 loci had a higher allele frequency than 5% and rest 113 loci turned out to be rare ($\leq 5\%$) and created minor allele frequency (MAF) in the assessed germplasm panel. Minimum, maximum and average values of minor allele frequency were 0.02, 0.05 and 0.028, respectively. Identified rare marker alleles occurred in 138 (~ 48%) accessions. The number of rare alleles on these accessions ranged from 1 to 32.

Based on our analyses, 38 SSR markers revealed high polymorphisms in this panel of long-staple cotton. These SSRs produce enough polymorphisms and can be recommended for molecular analyses of the *G. barbadense* genome (S1 Table). The chromosome locations of most SSR markers and their positions on chromosomes were determined by the consensus genetic map of tetraploid cotton reported by Blenda et al. [62].

Genetic distances and phylogeny of long-staple cotton germplasm

The average value of genetic distance (GD) among all 288 *G. barbadense* accessions was 0.19 with the smallest and largest distances of 0.01 and 0.67, respectively. The developed UPGMA dendrogram revealed two main groups «A» and «B» with the GD threshold of $> 50\%$, and five clearly distinct subgroups (Fig 1). The GD between the groups "A" and «B» was 0.65. The group "A" was genetically diverse, and samples within the group "A" were much different from each other. The average GD between samples in the group "A" was 0.31; for example, the accessions in this group were similar with an average of 69%, representing a wide genetic diversity.

Based on the fact that if GD between local populations of a single species is usually less than 0.05 [63–65], the samples belong to the same population group [66]. While if the distance is greater than 0.05 or 5% the individuals or accessions are likely to belong to different population subgroups. Thus, group «A» was separated. The group "A" was observed to be composed of two sub-groups G1 and G2. The GD between these subgroups was 0.40. The sub-group G1 predominantly consists of accessions of Africa-Egyptian genotypes (Giza, Barakat), and sub-group G2—American genotypes (Pima S1).

The group "B" included 273 cultivars or 94.8% of the analyzed accessions. The group consisted of accessions from different geographical regions of the world. The average GD in this group was much lower than in the group "A". Thus, minimum and maximum GDs in the group "B" were 0.01 and 0.65, respectively, with a mean of 0.18. Group "B" consisted of sub-groups G3, G4 and G5. Subgroup G3 includes nine samples (5 African, 3 Uzbekistan, and 1 China). The genetic distance between samples in the subgroup G3 varied from 0.22 to 0.38 with a mean of 0.31. The sub-group G4 is the largest, containing 254 accessions with average GD of 0.18. The subgroup included samples from many geographic regions. The last group G5 is represented by the Central Asian germplasm.

Molecular diversity and structure of the *G. barbadense* panel

In order to confirm the phylogenetic analysis and to support the population structure analysis, the principal component analysis (PCA) of SSR marker data was performed. PCA reduced the

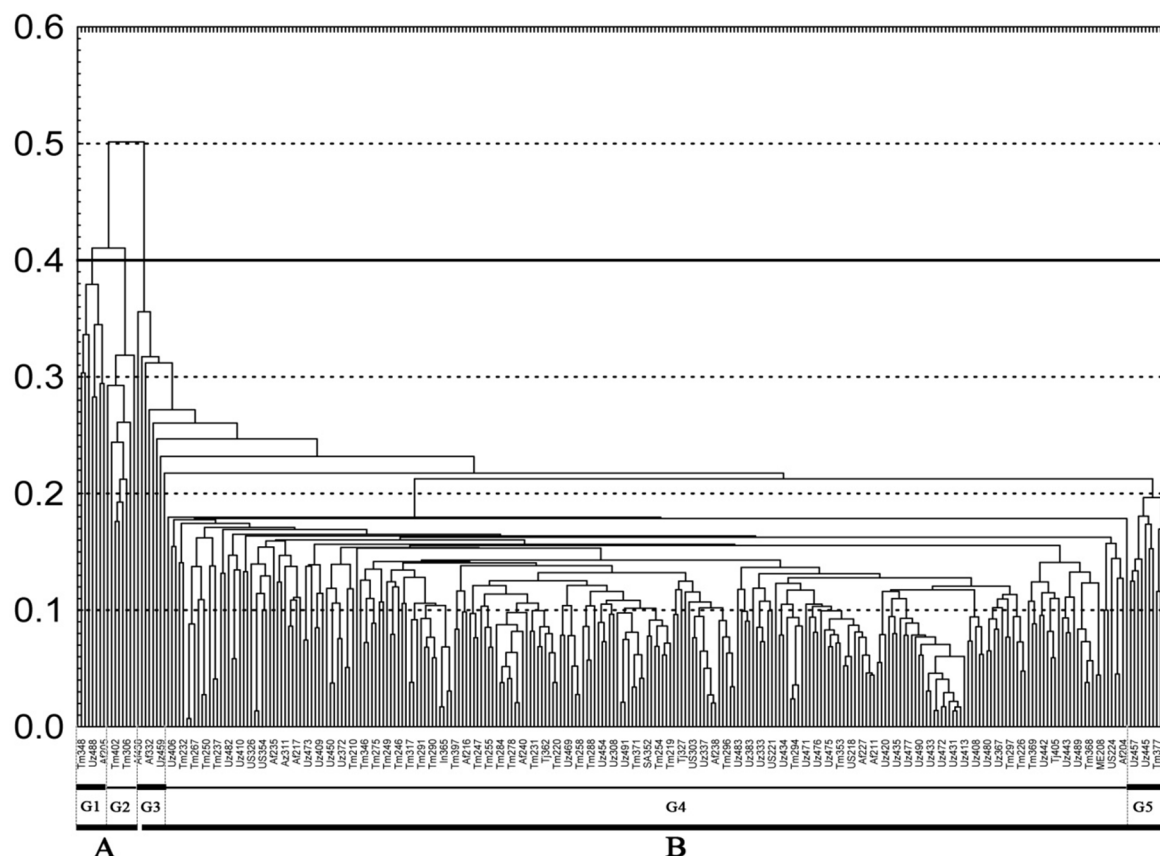


Fig 1. The UPGMA dendrogram of 288 *G. barbadense* accessions, constructed using the genotype of 301 polymorphic SSR alleles. Horizontal lines denote thresholds of genetic distances. Groups A and B are obtained on the basis of differences in > 50%, whereas subgroups G1, G2 and G3 obtained based on the upper boundary distinctions in 40%, and the subgroups G5 and G4—the upper bound of 20%.

<https://doi.org/10.1371/journal.pone.0188125.g001>

dimensionality of data and displayed all 288 *G. barbadense* accessions in a "two-dimensional" space, unlike the phylograms above. In addition, it more clearly reflects the grouping of samples and differences at the genetic level. As a result of PCA, it was determined that the first twelve components explained 51% of the variations. Of them the PC1 explains 15% of the variance, and clearly delineates the population into two subpopulations—large and small (Fig 2). The PC2 causes a 5% dispersion of 273 samples split into two main subpopulations overlapping subgroups, conventionally designated as Group A and Group B (Fig 2 and Table 8). Group A includes 108 accessions, which included most represented accessions from Uzbekistan—81 (75%), and group B comprises 165 accessions, in which the majority of genotypes are accessions from Turkmenistan—99 (60%; Fig 2 and Table 8).

Analysis of Molecular Variation (AMOVA)

To assess the genetic differentiation among and within predefined groups of a whole panel of 288 *G. barbadense* accessions, the Wright's index F_{st} (pair wise) was analyzed using statistical analysis of AMOVA. The genetic differentiation among and within groups were significant ($p \leq 0.001$), where 67.2% of total genetic variation was attributed to the difference within subpopulations, while the genetic variation between the predefined groups accounted for 32.8% of the total genetic variation (Table 9). Pairwise comparisons of the F_{st} index between the three

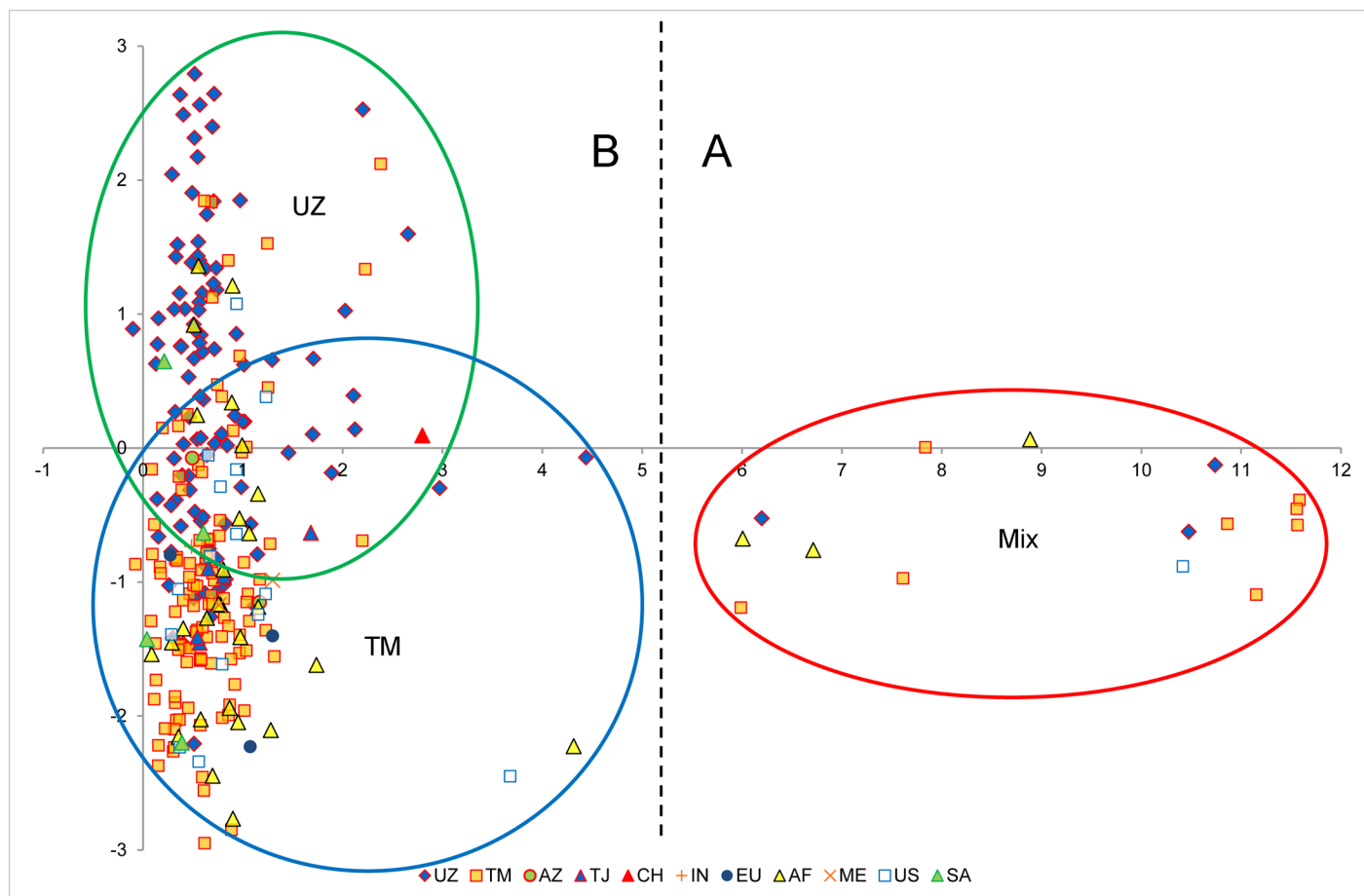


Fig 2. Principal component analysis, of 288 *G. barbadense* cultivars in the space of two main coordinate jointly by SSR genotypes. PC—the main components; (A) and (B)—subgroups represented in the majority of varieties of Uzbekistan (UZ) and Turkmenistan (TM), respectively. (Mix)—represented by the most genetically differentiated samples from several geographic regions i.e., from Turkmenistan (8), Africa (3), Uzbekistan (3), and American (1). UZ—Uzbekistan, TM—Turkmenistan, TJ—Tajikistan, AF—Africa, US—US, SA—South America AZ—Azerbaijan and ME—Middle East.

<https://doi.org/10.1371/journal.pone.0188125.g002>

groups revealed that the greatest genetic differentiation was present between the African group and the Turkmen group ($F_{st} = 0.58$; $p < 0.001$), and a little less variation, between the African and Uzbek group ($F_{st} = 0.57$; $p < 0.001$) (Table 10). Low-moderate genetic differentiation was found between the Uzbek and Turkmen subpopulation ($F_{st} = 0.117$).

Table 8. Differentiation of 288 *G. barbadense* accessions based on genetic and principal component analysis.

Germplasm origin	Subpopulation 1		Subpopulation 2
	Subgroup A	Subgroup B	
Uzbekistan	81	24	4
Turkmenistan	17	99	7
US	3	11	1
Africa	6	25	3
other	1	6	-
Total:	108	165	15

<https://doi.org/10.1371/journal.pone.0188125.t008>

Table 9. The AMOVA results.

Source of variation	df	Sum of squares	Variance Components	Percentage of variation	p-value
Among populations	2	1312.320	11.769	32.797	≤0.001
Within populations	285	6791.922	24.115	67.203	≤0.001
Total	287	8104.242	35.884		

<https://doi.org/10.1371/journal.pone.0188125.t009>

Table 10. Pair-wise comparisons of F_{st} values specific to each ecotype.

Origin	Africa	Uzbekistan	Turkmenistan
Africa	0.00000		
Uzbekistan	0.57568	0.00000	
Turkmenistan	0.58432	0.11723	0.00000

<https://doi.org/10.1371/journal.pone.0188125.t010>

Population structure and kinship

The model-based approach revealed the presence of at least two main subpopulations (Fig 3, K2). These two subpopulations share accessions of the total panel, 5.2% and 94.8%, respectively. Further expansion of the total population allowed us to divide it into three subpopulations, where a small mixed cluster remained unchanged (5.2%), and a large cluster was divided into two sub-populations sharing 37.5% and 57.3% of accessions, respectively (Fig 3, K3). The results of the structured population of *G. barbadense* 288 accessions were consistent with the phylogenetic analysis. Accessions, based on the genetic profile, clearly were divided into samples of mixed (accessions from all regions), Uzbek (UZ) and Turkmen (TM) cotton accessions.

The pairwise kinship analysis revealed that the majority of the pairs of cotton accessions (56%) had zero kinship values, whereas 22–23% of the pairs had a pairwise kinship value of 0.01–0.05 and 10% of the accession pairs had 10–20% relatedness. Only about 1.3% accessions had a pairwise kinship value of ≥25%.

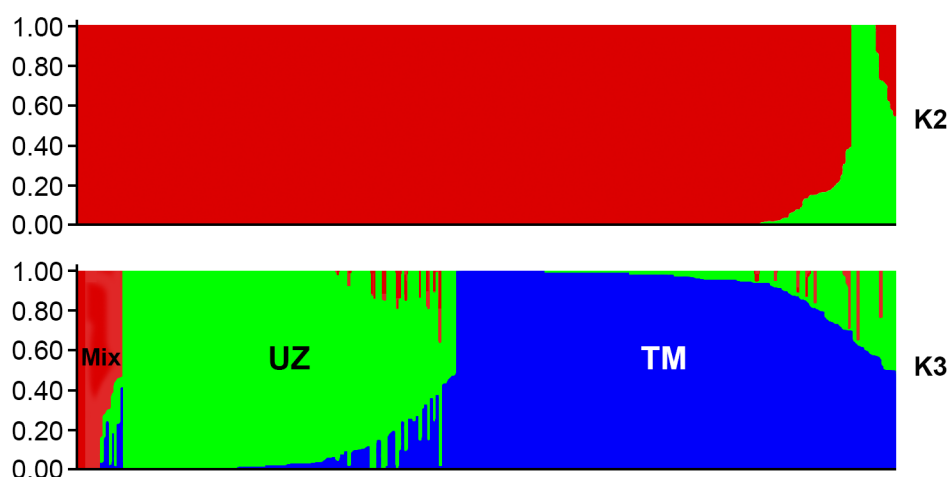


Fig 3. Summary plots of Q-matrix for the *G. barbadense* germplasm inferred from STRUCTURE analysis. K2—the division into two subpopulations: a small (green) and large (red). K3—further expansion of subpopulations on ecotypes (consistent with the results shown in Fig 2). Mix- represented by the most genetically differentiated samples from several geographic regions. UZ—Uzbek, and TK—Turkmen cotton accessions.

<https://doi.org/10.1371/journal.pone.0188125.g003>

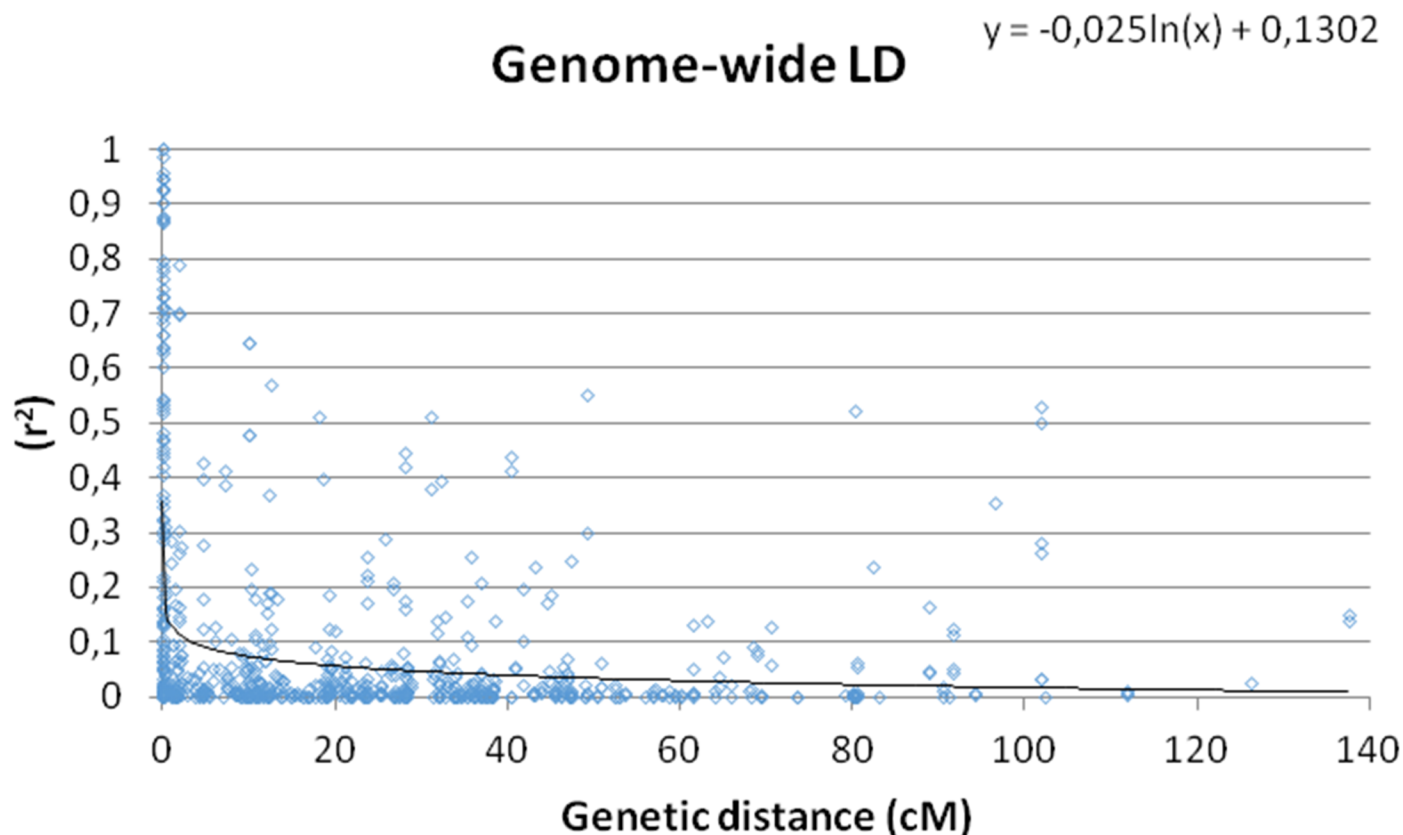


Fig 4. Scatter plot of significant r^2 values and genetic distance (cM) ($p < 0.001$) of locus pairs on whole genome of *G. barbadense* germplasm.

<https://doi.org/10.1371/journal.pone.0188125.g004>

Linkage disequilibrium (LD) and LD decay

The SSR data with 5% MAF removed set of 189 alleles were used to evaluate the extent of LD at genome level that detected pairwise LD in 17766 locus pairs in the *G. barbadense* panel. At significant threshold values ($r^2 \geq 0.05$ and $p \leq 0.005$), 16.8% (4576) of SSR marker pairs showed significant LD. By increasing the threshold to substantially higher values, $r^2 \geq 0.1$ ($p < 0.001$) and $r^2 \geq 0.2$ ($p < 0.0001$), LD was maintained in 2188 (8%) and 1187 (4.3%) of pairwise combinations of SSR markers, respectively. The linear plot of triangular graph of pairwise genome-wide LD between markers revealed significant LD blocks. This information is necessary to calculate to support the association mapping with the average distance of LD decay.

To reveal LD decay in *G. barbadense* genome, LD decay scatter plots of r^2 vs. genetic distance (cM) was generated to estimate LD decay using curvilinear regression (Fig 4). Results revealed that LD decay in *G. barbadense* genome was on average of 24.8 cM at the threshold of $r^2 \geq 0.05$. The genome-wide LD decay ($r^2 \geq 0.1$) was 3.36 cM in *G. barbadense* (Fig 4). These findings suggested the possibility of performing an efficient LD-based association mapping in the germplasm accessions of *G. barbadense* presented here.

Association mapping (AM) of fiber traits

AM analysis of SSR loci with fiber quality traits of 247 and 278 *G. barbadense* accessions, grown and evaluated in the two different ecological and geographical environmental conditions, Tashkent (Uzbekistan) and California (USA), respectively, was performed using

Table 11. Summary of SSR markers showed significant association with each of the trait studied in the Uzbekistan (UZB) and USA environments.

Traits	Number of markers with significant associations*			Markers associated with traits at $BF \leq 0.13$ ¹		
	UZB	USA	Common markers**	UZB	USA	Common markers**
Micronaire	30 (3)	28 (3)	12 (3)	9	6 (1)	3
Strength	62 (9)	59 (6)	41 (5)	6	3 (1)	2
Length	31(6)	36 (7)	22(6)	10 (1)	2	4
Uniformity	47(0)	42 (1)	25	8	5	2

Note:

* The table shows the markers that showed a significant association ($p \leq 0.05$) on the basis of MLM analysis 1.000 times permutation. In parentheses is the number of matched SSR markers associated with the described fiber traits, which are reported in other studies;

**the markers showed similar associations in both conditions.—Evaluation of these parameters in Uzbekistan and the United States have not performed.

¹BF—Bayes factor (likelihood ratio), where the value of $BF \leq 0.13$ correspond to strong ($z = 2.17$) evidence against the null hypothesis (H_0) [59, 65, 66].

<https://doi.org/10.1371/journal.pone.0188125.t011>

TASSEL software. According to the results, fiber traits varied from 1.6 to 11.4% in the USA and from 1.9 to 12.2% in the Uzbekistan environments. Therefore, not all markers associated with fiber traits in a single eco-geographic region showed association in both environments. However, a set of 100 markers retained a strong correlation and was significantly associated (MLM; $p \leq 0.05$) in both environments (Tashkent and California). For example, for fiber length, 22 markers showed significant associations: to 12—with micronaire, 41 and 25 with the strength and uniformity, respectively (Table 11 and Fig 5). Moreover, at critical values of minimum $BF \leq 0.13$ 11 SSRs retained strong associations with fiber traits (Table 11), of them 3 markers with FM (BNL3441_225, BNL3601_175, NAU2913_250), 2 with FS (BNL4003_150 and GH39_125), 4 with FL (BNL3599_200, GH75_130, NAU2913_275, NAU2913_250) and 2 with FU (BNL3902_200 and BNL3601_175).

When all fiber trait-associated SSR markers from our study were compared to reported SSR markers in previously published QTL-mapping studies, 14 SSRs [19, 67–76] revealed the same trait associations identified in our study (Tables 11 and 12). The remaining 86 SSR markers were identified for the first time in this study (Table 12).

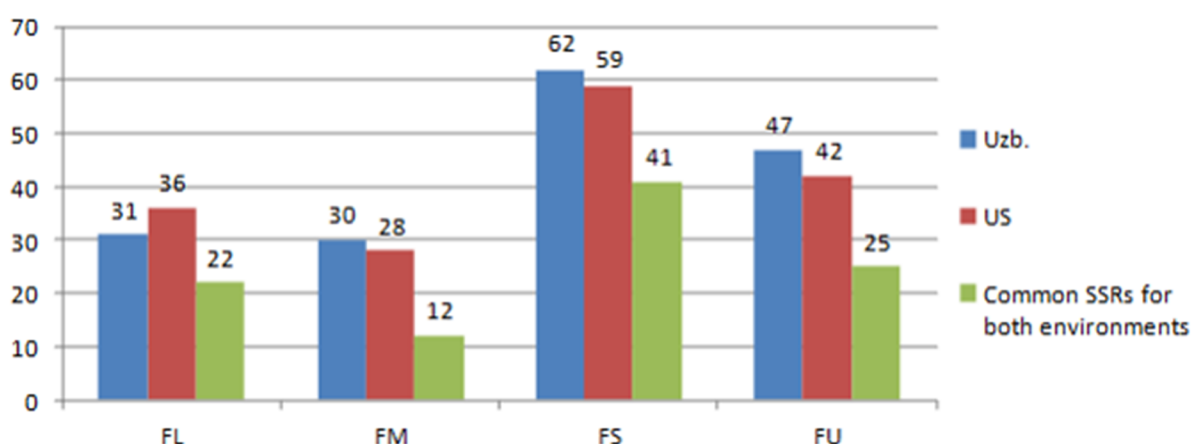


Fig 5. Result of association mapping of fiber quality traits in a particular region. Markers showed significant association (MLM; $p \leq 0.05$) both in Uzbekistan (Uzb.), and the United States (US) environments.: FL-fiber length, FM- micronaire, FS- fiber strength, FU- uniformity.

<https://doi.org/10.1371/journal.pone.0188125.g005>

Table 12. SSR markers showed significant fiber trait-association (MLM; $p \leq 0, 05$) in both environments (USA and Uzbekistan).

#	SSR marker*	USA		Uzbekistan		Match **	Chromosome***	Association with other traits (reference)
		F (MLM)	p (MLM)	F (MLM)	p (MLM)			
Fiber length								
1	BNL1421_220	5,3046	0,0221	4,5953	0,0331	[71]	13; 18	FL, FS, FE [77]
2	BNL1421_250	8,7632	0,0034	10,3148	0,0015	[71]	13; 18	FL, FS, FE [77]
3	BNL1495_150	7,4088	0,0070	11,1068	0,0010	[75]	13; 18	FE [78]; FL, FS, FE [77]; Lint% [72]
4	BNL1521_170	20,2535	0,0000	17,2683	0,0000	[19]	24; 15	FS [19,79]; FM, FE [70]; FM [21]
5	BNL1611_190	11,1161	0,0010	7,0728	0,0084	-	19	Lint index
6	BNL1705_200	13,6747	0,0003	8,1739	0,0046	[76]	21	FF; FE [76]; Lint% [80]
7	BNL3171_250	11,2913	0,0009	10,7198	0,0012	[81]	14; 21	FS, FL [81]; NSPB [82]
8	BNL3398_190	13,7112	0,0003	7,8360	0,0055	-	3; 13	
9	BNL3442_130	7,3967	0,0070	5,1340	0,0244	-	11	FS [81]
10	BNL3599_200 ¹	14,5256	0,0002	16,0080	0,0001	-	12; 25; 26	Near (2cM) the CF44ss gene, differential expressed during fiber development [83]
11	BNL3599_225	14,0670	0,0002	12,6243	0,0005	-	12; 25; 26	
12	BNL3902_200	10,5747	0,0013	4,0639	0,0450	-	15	FM, FF [70,19]; FE, FS [78]
13	BNL625_260	7,5997	0,0063	9,1459	0,0028	-	11	
14	GH118_150	14,4985	0,0002	13,6357	0,0003		9	
15	GH39_100	8,4011	0,0041	5,9578	0,0154	[75]	8	
16	GH58_150	11,2451	0,0009	7,3450	0,0072	-	9	
17	GH75_130 ²	3,9857	0,0470	7,2289	0,0000	-	1	
18	GH77_130	15,6853	0,0001	12,1496	0,0006	-	25	
19	GH83_150	14,4985	0,0002	13,6357	0,0003	-	5	
20	NAU2913_275 ²	28,0884	0,0000	33,1437	0,0000	-	26	
21	NAU2913_250 ¹	10,9679	0,0011	7,3916	0,0072	-	26	
22	TMB1660_200	4,7863	0,03	4,0461	0,0459	-	13	
Micronaire								
23	BNL1317_200	5,3110	0,0221	4,4295	0,0364	[19,2,70]	9; 23	FL; FS; FF; FM; FE [70,19,84–87]; [21,86,88,89]; [19,84,85,88]; [70]; Lint% [72]; Phe, Val, Ile [83]
24	BNL1611_250	4,6319	0,0324	4,3260	0,0386	-	19	Lint index
25	BNL2655_170	4,5173	0,0346	7,1258	0,0081	[89]	24	-
26	BNL2655_180	6,5472	0,0111	7,2491	0,0076	[89]	24	-
27	BNL3171_270	4,7169	0,0309	5,5791	0,0190	-	21	FS, FL [81]; NSPB [82]
28	BNL3441_225 ¹	4,4178	0,0366	5,5588	0,0192	[78]	3	Seed Yield, Lint Yield [76]
29	BNL3599_200	4,569	0,0338	4,6125	0,0331	-	12; 25; 26	-
30	BNL3601_150	7,1592	0,0081	4,3647	0,0382	-	22	-
31	BNL3601_175 ²	14,7358	1,69E-04	7,7107	0,0001	-	22	Maturity, Wall thickness [72]
32	BNL3792_275	3,7064	0,0558	3,9364	0,0489		8	-
33	BNL3977_200	7,3660	0,0071	6,4973	0,0114	-	19	-
34	NAU2913_250 ¹	6,8974	0,0092	22,7003	0,0000	-	26	-
Fiber strength								
35	BNL1521_170	4,4852	0,0352	11,4486	0,0008	[19]	24	FS [79]; FM [21]; FM, FE [70]
36	BNL1667_275	6,3563	0,0124	10,2118	0,0016	-	15	-
37	BNL2590_300	9,5001	0,0023	12,4752	0,0005	-	9; 23	-
38	BNL2632_250	5,9613	0,0154	9,6192	0,0022	-	11; 21	-
39	BNL2960_170	5,0139	0,0261	5,6441	0,0183	-	10	FL, FE [75]
40	BNL2967_180	4,5632	0,034	10,5314	0,0014	-	12	-
41	BNL3071_150	3,9221	0,0488	6,6112	0,0108		20; 10	-
42	BNL3259_210	6,9678	0,0089	10,0291	0,0017	[81]	3; 14	-
43	BNL3408_150	7,5481	0,0065	6,9507	0,0089	-	3	-
44	BNL3442_130	9,8947	0,0019	9,0888	0,0029	-	11	-

(Continued)

Table 12. (Continued)

#	SSR marker*	USA		Uzbekistan		Match **	Chromosome***	Association with other traits (reference)
		F (MLM)	p (MLM)	F (MLM)	p (MLM)			
45	BNL3452_180	7,1697	0,0079	8,7972	0,0033	[81]	3; 11	-
46	BNL3479_240	11,6504	0,0008	7,4832	0,0067	-	13; 18	-
47	BNL3482_130	7,7441	0,0058	4,4109	0,0368	[72]	26; 20	ELO [33]; MIC, BW, seed index [72]
48	BNL3590_225	6,3393	0,0125	10,0275	0,0017	-	2; 17	Lint%, BN [80]
49	BNL3599_200	4,9390	0,0272	11,2787	0,0009	-	12; 26	-
50	BNL3955_175	7,1704	0,0079	15,7115	0,0001	-	6; 17; 22	-
51	BNL3955_200	5,3898	0,0211	4,7147	0,0309	-	6; 17; 22	-
52	BNL3992_170	11,6470	0,0008	14,6615	0,0002	-	5	-
53	BNL3994_120	14,2212	0,0002	4,7149	0,0309	-	4; 22; 26	-
54	BNL3994_130	5,5912	0,0189	4,4562	0,0359	-	4; 22; 26	-
55	BNL4003_150 ¹	5,3124	0,0221	10,2666	0,0015	-	17	-
56	BNL4061_200	12,8342	0,0004	7,2074	0,0078	-	13	-
57	BNL625_260	18,7211	0,0000	18,2298	0,0000	-	11	-
58	CM209_120	9,1795	0,0027	6,4197	0,0119	-	9	Node number, resistance to <i>V. dahlia</i> [90]
59	CM209_150	4,9767	0,0266	4,5798	0,0334	-	-	-
60	GH107_250	4,8334	0,0289	5,2264	0,0231	-	4	-
61	GH110_140	9,1454	0,0028	7,4042	0,0070	-	20	-
62	GH117_160	11,2309	0,0009	18,7498	0,0000	-	4	-
63	GH117_170	12,2549	0,0006	15,2407	0,0001	-	4	-
64	GH118_150	20,3746	0,0000	24,2525	0,0000	-	9	-
65	GH171_250	9,2628	0,0026	9,7270	0,0020	-	24	-
66	GH200_150	7,6600	0,0061	19,2938	0,0000	-	22	-
67	GH39_100	17,4070	0,0000	17,7736	0,0000	-	6	-
68	GH39_125 ¹	14,8501	0,0002	5,5328	0,0195	-	6	-
69	GH58_150	18,7026	0,0000	24,3992	0,0000	-	10	-
70	GH75_150	7,9448	0,0052	11,8654	0,0007	-	1	-
71	GH77_130	14,0671	0,0002	15,6514	0,0001	-	-	-
72	GH82_190	14,8755	0,0001	20,6424	0,0000	-	12; 6	-
73	GH83_150	20,3746	0,0000	24,2525	0,0000	-	5; 21	-
74	NAU2709_150	19,9144	0,0000	23,9373	0,0000	-	23	-
75	NAU3620_150	11,3272	0,0009	10,2668	0,0015	-	-	-
Fiber uniformity								
76	BNL1495_135	8,0936	0,0048	4,6559	0,0320	-	13	FE [78]; FL, FS, FE [77]; FY [72]; FL [75]
77	BNL1521_170	16,9956	0,0001	17,4674	0,0000	-	24	-
78	BNL1611_190	12,8852	0,0004	17,3193	0,0000	-	19	-
79	BNL2590_300	3,9382	0,0484	6,2594	0,0131	-	9	-
80	BNL2960_170	6,4625	0,0117	5,3490	0,0216	-	10	FL, FE [75]
81	BNL2967_200	7,8402	0,0055	5,8171	0,0166	-	12	-
82	BNL3398_190	10,0170	0,0018	17,7981	0,0000	-	3	-
83	BNL3599_200	18,9071	0,0000	15,2568	0,0001	-	12	-
84	BNL3599_225	16,0116	0,0001	9,1933	0,0027	-	12	-
85	BNL3601_175 ¹	12,2255	0,0006	4,6528	0,0320	-	22	-
86	BNL3902_200 ¹	7,7856	0,0057	13,9393	0,0002	-	15	-
87	BNL3955_175	9,0779	0,0029	16,1893	0,0001	-	6; 17; 22	-
88	BNL4061_200	6,9722	0,0088	3,9938	0,0468	-	13	-
89	BNL625_260	6,4257	0,0119	8,5918	0,0037	-	11	-
90	GH107_250	4,1634	0,0424	7,9804	0,0051	-	4	-
91	GH118_150	14,3306	0,0002	16,9257	0,0001	-	9	-

(Continued)

Table 12. (Continued)

#	SSR marker*	USA		Uzbekistan		Match **	Chromosome***	Association with other traits (reference)
		F (MLM)	p (MLM)	F (MLM)	p (MLM)			
92	GH39_125	7,6613	0,0061	9,8348	0,0019	-	6	-
93	GH58_150	14,4281	0,0002	14,8799	0,0001	-	10; 9	-
94	GH77_130	14,1736	0,0002	11,3857	0,0009	-	-	-
95	GH77_175	5,0152	0,0261	8,9700	0,0030	-	-	-
96	GH82_190	14,4588	0,0002	21,1300	0,0000	-	6	-
97	GH83_150	14,3306	0,0002	16,9257	0,0001	-	4; 21	-
98	NAU3620_150	11,5049	0,0008	8,8581	0,0032	-	5	-
99	NAU3860_250	4,2711	0,0399	3,9286	0,0487	-	12; 26	-
100	NAU5015_180	10,4259	0,0014	4,7976	0,0295	-	5	-

*—SSRs with significant ($\leq 0,05$) fiber trait associations;

**—The same marker-trait association matched with other studies (reference),

***—Chromosome location of SSR marker;

¹—Marker showed significant association at Bayes factor $BF \leq 0,15$ [61],

²—Marker with very strong association at $BF \leq 0,02$; FL—fiber length; FS—fiber strength; FE—fiber elongation; FM—fiber micronaire; Lint%—lint yield, Phe—phenylalanine, Val—valine, Ile—leucine, NSPB—number seeds per boll.

<https://doi.org/10.1371/journal.pone.0188125.t012>

Discussion

Gathering information about genetic diversity and population structure is essential for providing insights into the breeding history and genetic relationship of crop germplasm. In this research, the first SSR marker-based molecular genetic study was conducted of *G. barbadense* cotton germplasm from the Uzbekistan cotton collection. It is believed that 10–30% of accessions may be enough to represent 70–90% of the genetic diversity of a whole germplasm collection [91]. The 288 accessions studied here represent almost 29% of the entire long-staple cotton genetic collection preserved in IG&PEB, Tashkent, Uzbekistan.

Assessment of *G. barbadense* accessions revealed a wide range of diversity in fiber quality traits within specific environments and between environments indicating the existence of useful genetic variation for these traits within the collection. Correlations of fiber quality traits between the USA (California) and Uzbekistan (Tashkent) environments demonstrated different performance of the same long-staple cotton accessions, which reflects the effect of the environment on the development of fiber quality traits. This should be taken into account when breeding for these traits. Nevertheless, several accessions were identified with stable fiber trait performance in both environments (Table 6). By definition, stability is the ability of an accession or genotype to show minimum variability in the interaction with the environment [92]. Thus, identified stable *G. barbadense* accessions that demonstrated the best values for single or all fiber traits in both (USA and Uzbekistan) environments should be primarily considered for breeding programs.

Genetic diversity analysis revealed a narrower genetic base of long-staple cotton germplasm based on SSR markers compared to Upland (*G. hirsutum*) cotton. This result is consistent with earlier studies [9, 33, 19, 93–95]. In addition, the genetic diversity was observed to be lower than previous reports from other studies of *G. barbadense* germplasm [95–99]. An explanation for this finding could be that previous studies used small sample size and/or low numbers of markers. Another explanation of this phenomenon is that in this study the majority of *G. barbadense* accessions from the IG&PEB Uzbek germplasm collection belong to Uzbekistan and Turkmenistan cotton germplasm that are closely related genetically and historically [5].

In this context, comparison of accessions for all clusters showed that the Turkmen cultivars have wider introgression/selection compared to the Uzbek accessions. The presence of groups (clusters) is a reflection of the genetic differentiation of populations as a result of the introduction of genes of genetically distinct forms. Thus, according to the cluster analysis, it can be hypothesized that studied *G. barbadense* collection was formed by the introduction/introgression of African (including Egyptian), African-American and American genotypes. In addition, the genetic relationship, identified between each of the studied accessions, is important for selecting breeding material and in the creation of improved germplasm and cultivars. It is also important to notice that as a result of many years of breeding, the population of *G. barbadense* cultivars formed genotypes specific to agro-ecological conditions of the Central Asian region, and clearly was traced by the genetic isolation of the Uzbek and Turkmen cultivars.

The average number of alleles per SSR marker (2.78) was higher than reported elsewhere (1.72—[97]; 1.66—[98]; 1.60—[95]), and the average PIC value (0.29) was lower than the previously reported from the Chinese's *G. barbadense* germplasm (0.32—[98], but close to values reported for *G. hirsutum* germplasm (0.28 [18] and 0.30 [80]). On the other hand, the same SSR marker set showed different fragment sizes and polymorphism compared to the *G. hirsutum* accessions from the Uzbek germplasm collection, in which the average allele number was higher (5.5) per SSR, whereas PIC value was much lower—0.082 [33]. As a result, the selected 108 markers, used in this study, are highly suitable to detect allelic variation in *G. barbadense* germplasm.

Population structure and differentiation of *G. barbadense* germplasm

To avoid spurious associations in LD-based AM, a detailed knowledge about population structure in a germplasm panel is of great importance. A model-based MLM approach using population structure information [46] is the most reliable method to correct spurious associations. However, under certain scenarios it is difficult to obtain accurate estimates of the actual number of subpopulations (K) [47, 93]. Generally, K is assumed to be the value with the highest estimated $\text{LnP}(D)$ generated by STRUCTURE [46]. The $\text{LnP}(D)$ value in real data tends to increase with increasing K and might not show a mode for the true K . Therefore, to avoid false associations an *ad hoc* measure ΔK proposed by the Evanno et al. [38] approach was used to detect the true K present in the SSR marker data.

The *ad hoc* measure ΔK [47] values proposed in this study indicated two groups as the most biologically meaningful population structure of the 288 *G. barbadense* germplasm panel. Similar clustering results have been reported for population structure of long-staple germplasm from other studies of *G. barbadense* [95, 97], including a recent comparative study of genome-wide divergence and population demographic histories for *G. hirsutum* and *G. barbadense* using genome-anchored SNPs [100]. In this study, several different methods (UPGMA clustering, PCoA, and Bayesian-based approach) were used to determine the level and pattern of genetic diversity and population structure present in the *G. barbadense* germplasm accessions based on SSR markers. Thus, grouping based on clustering analysis was an agreement with available background information of these accessions. As a result, the methods adopted here roughly reveal a similar level of population structure.

The AMOVA revealed a clear genetic structure of the germplasm accessions. The high variability of genetic loci within a population might be due to several factors. For example, widespread species pollinated by insects have high intrapopulation variability [101]. The high degree of cross-pollination patterns within the population serve as indicator of intensive breeding events such as hybridization. On the other hand, from an evolutionary point of view, domestication of long-staple cotton has been relatively recent (~ 2500 years BC) [102, 103]. It

is known that a population that has passed through a bottleneck has a temporarily disrupted mutation balance among the loci with an excess of heterozygosity [104]. The results of this study re-highlight the presence of the bottleneck in the recent past of cotton domestication.

The introduction of cotton germplasm and cultivars to new environments leads to formation of novel allele combinations in different loci allowing their adaptability to local stresses. This gives rise to several breeds within a gene pool. One example of this 'cultivar-introduction' is the "Acala" cultivar from the USA, which was introduced in Central Asia [105]. Analysis of 56 *G. barbadense* cultivars from China revealed 8% of genetic differentiation among populations (probably because a common cultivar-introduction) and 92% within populations [98]. In our study, the level of inter-population differentiation was much higher and accounted for 32.8%, while within population variation was 67.2%. This was similar to results of Upland cotton germplasm collection analysis where the genetic differentiation within and among populations of *G. hirsutum* accessions was 31.4% and 65.84%, respectively [106].

The AMOVA results demonstrate a significant correlation between genetic differentiation of accessions and their geographic origin. In different populations of the same species, there is always present historical evidence of interbreeding, even if an admixture does not exist at present. According to Wright, an $F_{st} > 0.25$ corresponds to a high level of genetic differentiation [107,108]. In this study, the F_{st} value in the *G. barbadense* diverse germplasm was equal to 0.328 ($p \leq 0.001$), indicating distinct population structure. Pairwise F_{st} analysis revealed strong differentiation between African and Turkmenistan ($F_{st} = 0.584$), and African and Uzbekistan ($F_{st} = 0.575$) germplasm. The differentiation between Uzbekistan and Turkmenistan gene pools was much lower ($F_{st} = 0.117$). This could be due to interbreeding events not only within a geographic niche, but also from similar introductions. This result also indicates that both gene pools (Uzbekistan and Turkmenistan) arose from common ancestors with a tendency of slight isolation, according to the ratio of allele frequencies identified in these groups or populations.

According to genetic relationship-patterns (Fig 1), it can be assumed that during initial development of the "group B" germplasm, ancestors from the Egyptian and Egyptian-American or/and American gene pools might be involved. It was also interesting to notice that group "A" was observed to be composed of two sub-groups G1 and G2, consisting of accessions of Africa-Egyptian (Giza, Barakat) and American genotypes (Pima S1) (Fig 1). Even though *G. barbadense* is indigenous to the northern part of South America and extends into Mesoamerica and the Caribbean [24], the Egyptian-Giza and America Pima-S series have been reported to have an interconnected breeding history. The Sea Island lineage also known as long-staple cotton contributed to the development of the Egyptian cotton [7, 24, 25]. This was later reintroduced in the 1940s into the United States as a part of the genetic base of the Pima gene pool of Pima-S series germplasm releases by USDA-ARS [22, 25]. This is the first report in which molecular data and historic breeding records provide similar evidence for the *G. barbadense* history. It could also be concluded that ancestors of the Egyptian and Egyptian-American or/and American germplasm were used in the development of the Turkmen and Uzbek *G. barbadense* germplasm. Thus, our results provide important insights into the evolutionary and breeding processes that influenced the structure of genetic variation within and among populations, which is the key point in association genetics studies.

Application of LD-based association mapping approaches

Association mapping (AM) is a very effective method of combining information on the genotype, phenotype, population structure and the LD in plants [28, 54]. The estimation of LD decay during AM is of great importance. In this study, the first attempt to apply the LD-based AM of fiber quality traits of *G. barbadense* germplasm from the Uzbek cotton collection was

made. The most appropriate measure of the LD for AM studies in plants is the squared correlation coefficient r^2 [26], which also points to marker-trait correlation [26,109–111]. In this study, 16.8% of SSR marker pairs showed significant pairwise linkage disequilibrium at $r^2 \geq 0.05$. At the higher values of $r^2 \geq 0.1$ and $r^2 \geq 0.2$, 8% and 4.3% of SSR marker pairs showed significant LD, respectively. The value of $r^2 \geq 0.1$ was a threshold for significant LD [112]. The results differ from those obtained from studies of different *G. hirsutum* germplasm collections [33, 80, 113, 114]. This indicates some differences in the formation of the LD pattern between pairs of loci in the genomes of *G. hirsutum* and *G. barbadense* species, which requires detailed comparative studies in the future.

The observed percentage of SSR loci in LD for *G. barbadense* genome in our study, as mentioned above (4.3–16.8%), is significantly lower than that of other crops such as corn, barley and wheat, where the percentage of markers in LD has been reported at 49–57%, 52–86% and 45–100%, respectively [115–119]. The low level of pairwise LD between SSR loci might be due to high levels of recombination rate in the genomes of allopolyploid cottons [120], as well as, mutations and experimental hybridization in the recent history of cultivated cotton germplasm [33]. In this study, an average genome-wide LD decay for *G. barbadense* accessions was 3.36cM at $r^2 \geq 0.1$ and 0.6cM at $r^2 \geq 0.2$. A recent study of 219 *G. barbadense* cultivars and land-race accessions of widespread origin using the genome-wide SNPs suggested a genome-wide LD decay was longer with an average of 128Kb compared to *G. hirsutum* with an average decay of 117Kb [100]. The fast LD decay of *G. barbadense* germplasm illustrates the significant potential for LD-based association mapping for agronomic traits. Taking into account that the average length of recombination block in the genome of tetraploid cotton is around 5,200 cM, with an average of 400kb per 1cM [121], the block size of ~ 5 cM is sufficient for reliable association mapping [33]. Therefore, our findings suggest a great possibility for association mapping of *G. barbadense* genome.

Several studies of the LD decay in a whole genome scale in diverse *G. hirsutum* germplasm collections found that the LD decay varied from 25 to 5 cM at r^2 threshold of 0.1 and from 6 to 1 cM at $r^2 \geq 0.2$ [9, 18, 19, 33, 113, 122]. This indicates that the size of the LD blocks may vary depending on the sample size and the population studied although the structure of LD haplotype blocks found to be considerably similar between *G. hirsutum* and *G. barbadense* [100]. Moreover, the composition of germplasm plays a key role in the LD variations, in other words the genetic distance over which LD decays depends on the genetic diversity present in the population [123]. Therefore, further characterization of the population structure and LD levels in *G. barbadense* germplasm collected from all over the world will be a benefit for association mapping of complex traits in long-staple cotton. In our study, the average size of the LD blocks in the genome of *G. barbadense* is less than that of *G. hirsutum*, which suggests a greater genetic variability. A large part of the genetic variability observed in modern *G. barbadense* germplasm may be due to introgression with *G. hirsutum* [124]. This also may be due to intensive breeding programs of Upland cotton that are ten times more than those dedicated to *G. barbadense* accessions [125,126].

Fiber quality trait associated markers

Linkage mapping is a powerful tool for identifying the genetic basis of quantitative traits in plants. However, association mapping is another effective approach for connecting phenotypes and genotypes in plants when information on population structure and LD is available [54]. The LD-based AM has recently gained popularity among plant geneticists and become a powerful approach to dissecting complex traits in many crops [28, 29]. In the current study, a number of major fiber trait-associated SSR markers were identified in the two diverse

environments of Uzbekistan and USA. Only markers that showed significant associations in both GLM and MLM were considered for further analysis. Among them, 100 SSR markers were associated with fiber quality traits in both environments. Furthermore, 14 SSRs associated with main fiber quality traits in our study coincided with reported fiber quality trait-associated SSRs from QTL-mapping studies in various experimental populations (Tables 11 and 12). At the same time, an additional 86 yet-unreported in literature SSR markers, associated with fiber quality traits in *G. barbadense* cotton germplasm, were detected (Table 12).

In a previous study of *G. hirsutum* germplasm, 25 fiber quality traits were significantly associated with SSR markers in Uzbekistan and Mexican environments [33]. In analyses of 56 cultivars of *G. arboreum* germplasm, 30 fiber trait-associated SSRs were identified [127]. Two independent association-mapping studies of 99 and 241 cultivars from the Chinese *G. hirsutum* germplasm collection, revealed 70 and 48 fiber quality trait-associated SSR markers, respectively [19, 80]. Another AM study using 220 cultivars from the US Upland germplasm collection identified 129 fiber trait-associated SSR markers [128]. Notably, several of the identified SSR markers were also reported by previous studies. For example, BNL1521 associated with FL and FS in this study showed the same trait associations in Upland cotton [19]. In previous reports, this marker was also found to be associated with FM and FE [70], FS [79] and FM [21]. Thus, BNL1521 is the high-priority candidate DNA marker for MAS in cotton breeding to improve fiber quality traits.

Association of markers with two or more fiber quality traits indicates the close location of some genes controlling these traits that have been repeatedly observed in many studies [21, 23, 75, 81, 129–131]. Analysis of chromosomal location of identified markers revealed clustering of positively correlated fiber traits on the same chromosome segments (S1 Table). However, two markers were negatively associated with correlated traits (FM-FL and FM-FU). Similar findings were reported by Cai et al. [19] where two markers were associated with FM-FL and FM-FS and were negatively correlated. This suggests the possibility of a joint transfer and inheritance of these traits, thereby bypassing the obstacles in the form of negative correlations.

A fiber traits gene-cluster was identified near markers BNL1421 and BNL1495. The BNL1421, associated with FL in this study, as well as, in a study of *G. arboreum* germplasm [71], was associated with FE [78] and FY [72] in *G. hirsutum* and located within a chromosome segment that is rich to fiber quality traits [77] (Table 12). The BNL1495, associated with FL in this study, as well as, in a study of *G. hirsutum* germplasm [75], was also located within a group of markers associated with FE [78]. Furthermore, an estimated distance between BNL1421 and BNL1495 is ~1.8cM, implying the clustering of fiber quality genes within the selected chromosome segment. BNL1521, located on Ch24 and associated with FL in the current study and in a study of *G. hirsutum* germplasm [19], were reported to be associated with FS [19, 79], FM [70, 74] and FE [70]. BNL1705 associated with FL in this study were also reported to be associated with FL [76] and FY [80]. BNL1317 associated with FM herein was also associated with the same fiber trait in other studies [8, 19, 21], FE [19], and FL [70].

Furthermore, BNL1317 was associated with a QTL for phenylalanine content [132], which plays a key role in phenylpropanoid pathway during cotton fiber cell wall formation [133–135]. Thus, BNL1317 is another high-priority candidate marker for MAS. Marker BNL3601, significantly associated with FM ($BF \leq 0.02$) in our study, was also reported to be associated with fiber maturity and fiber cell wall thickness [62], that is directly related to micronaire [136, 137]. Therefore, these SSRs should be very useful for fiber quality improvement of cotton cultivars by means of marker-assisted selection (MAS).

In this context, it should be noted that, so far, cotton lags behind on MAS application and success compared to other crops [9, 30, 33]. Many molecular markers tagged through numerous traditional QTL-mapping studies, except those monogenically inherited resistance traits

(e.g. [138, 139]), have had a limited success in cotton breeding programs [139]. This may be primarily connected with (1) complexity and polyploidy of the cotton genome, (2) polygenic and epigenetic nature of inheritance of many important QTLs including fiber traits, which are greatly impacted from G by E interactions and (3) specificity of tagged molecular markers to a bi-parentally-derived mapping population, making markers meaningless when other populations or genotypes are used [30, 139].

Differing from QTL-mapping approach, LD-based association mapping using germplasm resources helps to associate more biologically meaningful markers in a large number of germplasm accessions, shaped under many historic meiotic events [9, 30, 33, 123]. Therefore, molecular markers associated with important traits using LD-based association mapping should be efficient to be used in MAS programs. For instance, previous efforts on association mapping in a large set of Upland cultivars and exotic landrace stock germplasm [9, 33] have helped us to design a successful molecular breeding program in Uzbekistan. In a short time, using SSR markers associated with fiber length, strength and micronaire traits, novel cotton cultivars series “Ravnaq” (translates from Uzbek as “Advance”) with improved fiber quality traits have been developed, which are currently under evaluation of State Variety Testing Stations of Uzbekistan [139, 140]. This exemplifies the usefulness of genome-wide association mapping studies of cotton that should be highly efficient with application of recently developed genome-anchored SNPs [100] because of genome wide scale and considering many alleles and genetic interactions.

Conclusions

Thus, in a set of 288 *G. barbadense* germplasm resources from the Uzbekistan cotton collection, for the first time, a genetic diversity, population structure, and the extent of genome-wide linkage disequilibrium were assessed for Pima or extra-long staple cotton genome. Efforts have helped to perform LD-based association mapping of fiber quality traits evaluated in two diverse environments (Uzbekistan and USA) using a highly polymorphic set of simple sequence repeat (SSR) markers. Results have provided important insights into the evolutionary and breeding processes that influence the structure of genetic variation within a population and among populations. Also, there is a lower level of LD compared to the Upland cotton genome or other agricultural crops. Model based-association mapping efforts have further identified strongly associated novel and previously reported SSR markers with major fiber quality traits. Results should help to expand our knowledge of the breeding history and germplasm peculiarities of Pima cotton. Identified SSR markers and candidate gene sequence associated fiber quality traits in *G. barbadense* foster cotton improvement programs using marker-assisted selection.

Supporting information

S1 Data. Fiber quality trait data measured for four major fiber traits from Uzbekistan (UZB) and the USA environments. This data set was used for trait analyses and association mapping studies.
(XLSX)

S2 Data. TASSEL formatted SSR marker data set for 288 *G. barbadense* panel used in this study. Note that 5% MAF is not filtered and can be done using TASSEL.
(XLSX)

S1 Table. A set of SSR markers with high polymorphism for studies of the *G. barbadense* genome.
(DOC)

Acknowledgments

We thank Academy of Sciences of Uzbekistan, and Committee for Coordination Science and Technology Development of Uzbekistan for basic science (FA-F5-T030). We greatly acknowledge the Office of International Research Programs (OIRP) of the United States Department of Agriculture (USDA)–Agricultural Research Service (ARS) and U.S. Civilian Research and Development Foundation (CRDF) for international cooperative grants UZB-TA-31017, which were devoted to study *G. barbadense* germplasm resources. We thank Uzbekistan Government, Ministry of Foreign Economic Relations, Trade and Investments of Uzbekistan for capacity funding of the Center of Genomics and Bioinformatics, Uzbekistan. We greatly acknowledge the Uzbek and US partner laboratories for the assistance with molecular mapping experiments described in this work. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. The US Department of Agriculture is an equal opportunity provider and employer.

Author Contributions

Conceptualization: Mauricio Ulloa, Ibrokhim Y. Abdurakhmonov.

Data curation: Alisher A. Abdullaev, Sofiya M. Rizaeva, Ibrokhim Y. Abdurakhmonov.

Formal analysis: Alisher A. Abdullaev, Ilkhom B. Salakhutdinov, Sharof S. Egamberdiev, Ernest E. Khurshut, Ibrokhim Y. Abdurakhmonov.

Funding acquisition: Mauricio Ulloa, Ibrokhim Y. Abdurakhmonov.

Investigation: Sofiya M. Rizaeva.

Methodology: Alisher A. Abdullaev, Ilkhom B. Salakhutdinov, Sharof S. Egamberdiev, Mauricio Ulloa, Ibrokhim Y. Abdurakhmonov.

Project administration: Ibrokhim Y. Abdurakhmonov.

Resources: Alisher A. Abdullaev, Sofiya M. Rizaeva, Mauricio Ulloa, Ibrokhim Y. Abdurakhmonov.

Software: Ilkhom B. Salakhutdinov.

Supervision: Ibrokhim Y. Abdurakhmonov.

Validation: Alisher A. Abdullaev, Ilkhom B. Salakhutdinov, Ernest E. Khurshut, Mauricio Ulloa.

Visualization: Alisher A. Abdullaev, Ilkhom B. Salakhutdinov, Sharof S. Egamberdiev, Ernest E. Khurshut.

Writing – original draft: Alisher A. Abdullaev, Ibrokhim Y. Abdurakhmonov.

Writing – review & editing: Mauricio Ulloa, Ibrokhim Y. Abdurakhmonov.

References

1. Campbell BT, Jones MA. Assessment of genotype × environment interactions for yield and fiber quality in cotton performance trials. *Euphytica*. 2005; 144:69–78. <https://doi.org/10.1007/s10681-005-4336-7>
2. Rungis D, Llewellyn D, Dennis ES, Lyon BR. Simple sequence repeat (SSR) markers reveal low levels of polymorphism between cotton (*Gossypium hirsutum* L.) cultivars. *Aust J Agric Res*. 2005; 56:301–307. <https://doi.org/10.1071/AR04190>

3. Culp TW, Harrell DC. Breeding methods for improving yield and fiber quality of Upland cotton (*Gossypium hirsutum* L.). Crop Sci. Madison, WI: Crop Science Society of America; 1973; 13:686–689. <https://doi.org/10.2135/cropsci1973.0011183X001300060030x>
4. Campbell BT, Saha S, Percy R, Frelichowski J, Jenkins JN, Park W, et al. Status of the global cotton germplasm resources. Crop Sci. 2010; 50:1161–1179. <https://doi.org/10.2135/cropsci2009.09.0551>
5. Abdurakhmonov IY. World Cotton Germplasm Resources. 1st ed. Rijeka: InTech Press; 2014. 320 p. <https://doi.org/10.5772/56978> <https://www.intechopen.com/books/world-cotton-germplasm-resources>
6. Wendel JF, Brubaker CL, Percival AE. Genetic diversity in *Gossypium hirsutum* and the origin of Upland cotton. Am J Bot. 1992; 79:1291–1310. <https://doi.org/10.2307/2445058>
7. Niles GA, Feaster CV. Breeding. In: Kohel RJ, Lewis CF, editors. Cotton. Madison, WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America; 1984. pp. 201–231. <https://doi.org/10.2134/agronmonogr24.c7>
8. Wang F, Gong Y, Zhang C, Liu G, Wang L, Xu Z, et al. Genetic effects of introgression genomic components from Sea Island cotton (*Gossypium barbadense* L.) on fiber related traits in upland cotton (*G. hirsutum* L.). Euphytica. 2011; 181:41–53. <https://doi.org/10.1007/s10681-011-0378-1>
9. Abdurakhmonov IY, Kohel RJ, Yu JZ, Pepper AE, Abdullaev AA, Kushanov FN, et al. Molecular diversity and association mapping of fiber quality traits in exotic *G. hirsutum* L. germplasm. Genomics. 2008; 92:478–487. <https://doi.org/10.1016/j.ygeno.2008.07.013> PMID: 18801424
10. Kim HJ, Triplett BA. Cotton fiber growth *in planta* and *in vitro* models for plant cell elongation and cell wall biogenesis. Plant Physiol. 2001; 127:1361–1366. PMID: 11743074
11. Paterson AH. Genetics and genomics of cotton. Paterson AH, editor. Plant Genetics and Genomics: Crops and Models. New York, NY: Springer Science + Business Media; 2009. 518 p. <https://doi.org/10.1007/978-0-387-70810-2>
12. Zhao Y, Wang H, Chen W, Zhao P, Gong H, Sang X, et al. Regional association analysis-based fine mapping of three clustered QTL for verticillium wilt resistance in cotton (*G. hirsutum* L.). BMC Genomics. 2017; 18:661. <https://doi.org/10.1186/s12864-017-4074-y> PMID: 28841857
13. Ademe MS, He S, Pan Z, Sun J, Wang Q, Qin H, et al. Association mapping analysis of fiber yield and quality traits in Upland cotton (*Gossypium hirsutum* L.). Mol Genet Genomics. 2017; in press. <https://doi.org/10.1007/s00438-017-1346-9> PMID: 28748394
14. Iqbal MA, Rahman MU. Identification of Marker-trait associations for lint traits in cotton. Front Plant Sci. 2017; 8:86. <https://doi.org/10.3389/fpls.2017.00086> PMID: 28220132
15. Islam MS, Thyssen GN, Jenkins JN, Zeng L, Delhom CD, McCarty JC, et al. A MAGIC population-based genome-wide association study reveals functional association of GhRBB1_A07 gene with superior fiber quality in cotton. BMC Genomics. 2016; 17(1):903. <https://doi.org/10.1186/s12864-016-3249-2> PMID: 27829353
16. Zhao YL, Wang HM, Shao BX, Chen W, Guo ZJ, Gong HY et al. SSR-based association mapping of salt tolerance in cotton (*Gossypium hirsutum* L.). Genet Mol Res. 2016; 15(2): gmr.15027370. PMID: 27323090
17. Jia Y, Sun X, Sun J, Pan Z, Wang X, He S, et al. Association mapping for epistasis and environmental interaction of yield traits in 323 cotton cultivars under 9 different environments. PLoS ONE. 2014; 9(5):e95882. <https://doi.org/10.1371/journal.pone.0095882> PMID: 24810754
18. Zhao Y, Wang H, Chen W, Li Y. Genetic structure, linkage disequilibrium and association mapping of Verticillium wilt resistance in elite cotton (*Gossypium hirsutum* L.) germplasm population. PLoS ONE. 2014; 9(1):e86308. <https://doi.org/10.1371/journal.pone.0086308> PMID: 24466016
19. Cai C, Ye W, Zhang T, Guo W. Association analysis of fiber quality traits and exploration of elite alleles in Upland cotton cultivars/accessions (*Gossypium hirsutum* L.). J Integr Plant Biol. 2014; 56(1):51–62. <https://doi.org/10.1111/jipb.12124> PMID: 24428209
20. Mei H, Zhu X, Zhang T. Favorable QTL alleles for yield and its components identified by association mapping in Chinese Upland cotton cultivars. PLoS One. 2013; 8(12):e82193. <https://doi.org/10.1371/journal.pone.0082193> PMID: 24386089
21. Wang XQ, Yu Y, Li W, Guo HL, Lin ZX, Zhang XL. Association analysis of yield and fiber quality traits in *Gossypium barbadense* with SSRs and SRAPs. Genet Mol Res. 2013; 12(3):3353–62. <https://doi.org/10.4238/2013.September.4.1> PMID: 24065676
22. Ulloa M, Percy R, Huttmacher RB, Zhang J. The future of cotton breeding in the Western United States. J Cotton Sci. 2009; 13:246–255.
23. Ulloa M, Brubaker C, Chee P. Cotton. In: Kole C, editor. Technical crops. Springer-Verlag Berlin Heidelberg; 2007. pp. 1–49. https://doi.org/10.1007/978-3-540-34538-1_1
24. Hutchinson J, Manning H. The sea island cotton. Mem Cotton Res Station Trinidad Ser Genet. 1945; 25:80–92.

25. Percy RG, Wendel JF. Allozyme evidence for the origin and diversification of *Gossypium barbadense* L. Theor Appl Genet. 1990; 79:529–542. <https://doi.org/10.1007/BF00226164> PMID: 24226459
26. Flint-Garcia SA, Thornsberry JM, Buckler ES. Structure of linkage disequilibrium in plants. Annu Rev Plant Biol. 2003; 54:357–74. <https://doi.org/10.1146/annurev.arplant.54.031902.134907> PMID: 14502995
27. Würschum T, Liu W, Gowda M, Maurer HP, Fischer S, Schechert A, et al. Comparison of biometrical models for joint linkage association mapping. Heredity 2012; 108:332–340. <https://doi.org/10.1038/hdy.2011.78> PMID: 21878984
28. Gupta PK, Rustgi S, Kulwal PL. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. Plant Mol Biol. 2005; 57:461–485. <https://doi.org/10.1007/s11103-005-0257-z> PMID: 15821975
29. Gupta PK, Kulwal PL, Jaiswal V. Association mapping in crop plants. Advances in Genetics. Elsevier; 2014. pp. 109–147. <https://doi.org/10.1016/B978-0-12-800271-1.00002-0> PMID: 24880734
30. Abdurakhmonov IY, Abdurakimov A. Application of association mapping to understanding the genetic diversity of plant germplasm resources. Int J Plant Genomics. 2008; 2008:574927. <https://doi.org/10.1155/2008/574927> PMID: 18551188
31. Saeed M, Wangzhen G, Tianzhen Z. Association mapping for salinity tolerance in cotton (*Gossypium hirsutum* L.) germplasm from US and diverse regions of China. Aust J Crop Sci. 2014; 8:338–346.
32. Liu G, Mei H, Wang S, Li X, Zhu X, Zhang T. Association mapping of seed oil and protein contents in upland cotton. Euphytica. Springer Netherlands; 2015; 205:637–645. <https://doi.org/10.1007/s10681-015-1450-z>
33. Abdurakhmonov IY, Saha S, Jenkins JN, Buriev ZT, Shermatov SE, Scheffler BE, et al. Linkage disequilibrium based association mapping of fiber quality traits in *G. hirsutum* L. variety germplasm. Genetica. 2009; 136:401–417. <https://doi.org/10.1007/s10709-008-9337-8> PMID: 19067183
34. Steel RGD, Torrie JH. Principles and procedures of statistics: a biometrical approach. New York: McGraw-Hill; 1980. 633 p.
35. Dellaporta SL, Wood J, Hicks JB. A plant DNA miniprep: Version II. Plant Mol Biol Report. 1983; 1:19–21. <https://doi.org/10.1007/BF02712670>
36. Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 2005; 21:2128–2129. <https://doi.org/10.1093/bioinformatics/bti282> PMID: 15705655
37. Li Y, Li Y, Wu S, Han K, Wang Z, Hou W, et al. Estimation of multilocus linkage disequilibrium in diploid populations with dominant markers. Genetics. 2007; 176:1811–1821. <https://doi.org/10.1534/genetics.106.068890> PMID: 17565957
38. Hardy OJ, Vekemans X. SPAGeDi: a versatile computer program to analyze spatial genetic structure at the individual or population levels. Mol Ecol Notes. 2002; 2:618–620. <https://doi.org/10.1046/j.1471-8286.2002.00305.x>
39. Swofford DL. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4.0b10. Sinauer Assoc Sunderland MA. 2001
40. Reynolds J, Weir BS, Cockerham CC. Estimation of the co-ancestry coefficient: basis for a short-term genetic distance. Genetics. 1983; 105:767–779 PMID: 17246175
41. Weir BS, Cockerham CC. Estimating F-Statistics for the analysis of population structure. Evolution. 1984; 38:1358–1370. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x> PMID: 28563791
42. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 1992; 131:479–491. PMID: 1644282
43. Schneider S, Roessli D, Excoffier L. Arlequin: a software for population genetics data analysis User manual ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva; Geneva: 2000.
44. Holsinger KE, Lewis PO. Hickory: a package for analysis of population genetic data v1.1. University of Connecticut: Department of Ecology & Evolutionary Biology; 2007. pp. 1–42.
45. Holsinger KE, Lewis PO, Dey D. A bayesian approach to inferring population structure from dominant markers. EEB Artic. 2002; 1:1–14.
46. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155:945–959. PMID: 10835412
47. 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. <https://doi.org/10.1111/j.1365-294X.2005.02553.x> PMID: 15969739

48. Whitt SR, Buckler ES. Using natural allelic diversity to evaluate gene function. In: Grotewold E, editor. *Plant Functional Genomics*. New Jersey: Humana Press; 2003. pp. 123–140. <https://doi.org/10.1385/1-59259-413-1:123>
49. Hardy OJ. Estimation of pairwise relatedness between individuals and characterization of isolation by distance processes using dominant genetic markers. *Mol Ecol*. 2003; 12:1577–1588. <https://doi.org/10.1046/j.1365-294X.2003.01835.x> PMID: 12755885
50. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007; 23:2633–2635. <https://doi.org/10.1093/bioinformatics/btm308> PMID: 17586829
51. Mohlke KL, Lange EM, Valle TT, Ghosh S, Magnuson VL, Silander K, et al. Linkage disequilibrium between microsatellite markers extends beyond 1 cM on chromosome 20 in Finns. *Genome Res*. 2001; 11:1221–1226. <https://doi.org/10.1101/gr.173201> PMID: 11435404
52. McRae AF, McEwan JC, Dodds KG, Wilson T, Crawford AM, Slate J. Linkage disequilibrium in domestic sheep. *Genetics*. 2002; 160:1113–1122. PMID: 11901127
53. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet*. 2006; 38:203–208. <https://doi.org/10.1038/ng1702> PMID: 16380716
54. Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES. Dwarf8 polymorphisms associate with variation in flowering time. *Nat Genet*. 2001; 28:286–289. <https://doi.org/10.1038/90135> PMID: 11431702
55. Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, et al. An Arabidopsis example of association mapping in structured samples. *PLoS Genet*. 2007; 3:e4. <https://doi.org/10.1371/journal.pgen.0030004> PMID: 17238287
56. Storey JD, Tibshirani R. SAM thresholding and false discovery rates for detecting differential gene expression in DNA microarrays. In: Parmigiani G, Garrett ES, Irizarry RA, Zeger SL, editors. *The analysis of gene expression data: methods and software*. Springer New York; 2003. pp. 272–290. https://doi.org/10.1007/0-387-21679-0_12
57. Conneely KN, Boehnke M. So many correlated tests, so little time! Rapid adjustment of p values for multiple correlated tests. *Am J Hum Genet*. 2007; 81:1158–1168. <https://doi.org/10.1086/522036> PMID: 17966093
58. Goodman SN. Of p-values and Bayes: a modest proposal. *Epidemiology*. 2001; 12:295–297. <https://doi.org/10.1097/00001648-200105000-00006> PMID: 11337600
59. Katki HA. Invited commentary: evidence-based evaluation of p values and Bayes factors. *Am J Epidemiol*. 2008; 168:384–388. <https://doi.org/10.1093/aje/kwn148>
60. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann Intern Med*. 1999; 130:1005–1013. <https://doi.org/10.7326/0003-4819-130-12-199906150-00019> PMID: 10383350
61. Goodman SN. Introduction to Bayesian methods I: measuring the strength of evidence. *Clin Trials*. 2005; 2:282–290. <https://doi.org/10.1191/1740774505cn0980a> PMID: 16281426
62. Blenda A, Fang DD, Rami J-F, Garsmeur O, Luo F, Lacape J-M. A high density consensus genetic map of tetraploid cotton that integrates multiple component maps through molecular marker redundancy check. *PLoS ONE*. 2012; 7:e45739. <https://doi.org/10.1371/journal.pone.0045739> PMID: 23029214
63. Nei M. Genetic distance between populations. *Am Nat*. The University of Chicago Press; 1972; 106:283–292. <https://doi.org/10.1086/282771>
64. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*. 1978; 89:583–590. PMID: 17248844
65. Nei M. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA*. 1973; 70:3321–3323. <https://doi.org/10.1073/pnas.70.12.3321> PMID: 4519626
66. Chakraborty R. Gene diversity analysis in nested subdivided populations. *Genetics*. 1980; 96:721–726.
67. Song X, Wang K, Guo W, Zhang J, Zhang T. A comparison of genetic maps constructed from haploid and BC 1 mapping populations from the same crossing between *Gossypium hirsutum* L. and *Gossypium barbadense* L. *Genome*. 2005; 48:378–390. <https://doi.org/10.1139/g04-126> PMID: 16121235
68. Zhang Z-S, Xiao Y-H, Luo M, Li X-B, Luo X-Y, Hou L, et al. Construction of a genetic linkage map and QTL analysis of fiber-related traits in upland cotton (*Gossypium hirsutum* L.). *Euphytica*. 2005; 144:91–99. <https://doi.org/10.1007/s10681-005-4629-x>
69. Lin Z, He D, Zhang X, Nie Y, Guo X, Feng C, et al. Linkage map construction and mapping QTL for cotton fibre quality using SRAP, SSR and RAPD. *Plant Breed*. 2005; 124:180–187. <https://doi.org/10.1111/j.1439-0523.2004.01039.x>

70. Shen X, Guo W, Zhu X, Yuan Y, Yu JZ, Kohel RJ, et al. Molecular mapping of QTLs for fiber qualities in three diverse lines in Upland cotton using SSR markers. *Mol Breed*. 2005; 15:169–181. <https://doi.org/10.1007/s11032-004-4731-0>
71. Kantartzi SK, Stewart JM. Association analysis of fibre traits in *Gossypium arboreum* accessions. *Plant Breed*. 2008; 127:173–179. <https://doi.org/10.1111/j.1439-0523.2008.01490.x>
72. Wu J, Gutierrez OA, Jenkins JN, McCarty JC, Zhu J. Quantitative analysis and QTL mapping for agronomic and fiber traits in an RI population of upland cotton. *Euphytica*. 2009; 165:231–245. <https://doi.org/10.1007/s10681-008-9748-8>
73. Thiyagu K, Nadarajan N, Boopathi NM, Gunasekaran M. Molecular genetic diversity and marker phenotype association of locally adopted germplasm for genetic improvement of cotton. *World Cotton Research Conference on Technologies for Prosperity*. Mumbai 7–11 November, 2011: India Publishers, New Delhi; 2011. p. 104.
74. Wang X, Yu Y, Sang J, Wu Q, Zhang X, Lin Z. Intrasppecific linkage map construction and QTL mapping of yield and fiber quality of *Gossypium barbadense*. *Aust J Crop Sci*. 2013; 7:1252–1261.
75. Liang Q, Hu C, Hua H, Li Z, Hua J. Construction of a linkage map and QTL mapping for fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *Chin Sci Bull*. 2013; 58:3233–3243. <https://doi.org/10.1007/s11434-013-5807-1>
76. Yu J, Zhang K, Li S, Yu S, Zhai H, Wu M, et al. Mapping quantitative trait loci for lint yield and fiber quality across environments in a *Gossypium hirsutum* × *Gossypium barbadense* backcross inbred line population. *Theor Appl Genet*. 2013; 126:275–287. <https://doi.org/10.1007/s00122-012-1980-x> PMID: 23064252
77. Frelichowski JE, Palmer MB, Main D, Tomkins JP, Cantrell RG, Stelly DM, et al. Cotton genome mapping with new microsatellites from Acala “Maxxa” BAC-ends. *Mol Genet Genomics*. 2006; 275:479–491. <https://doi.org/10.1007/s00438-006-0106-z> PMID: 16501995
78. Park Y-H, Alabady MS, Ulloa M, Sickler B, Wilkins TA, Yu J, et al. Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. *Mol Genet Genomics*. 2005; 274:428–441. <https://doi.org/10.1007/s00438-005-0037-0> PMID: 16187061
79. Zhang T, Yuan Y, Yu J, Guo W, Kohel RJ. Molecular tagging of a major QTL for fiber strength in Upland cotton and its marker-assisted selection. *Theor Appl Genet*. 2003; 106:262–268. <https://doi.org/10.1007/s00122-002-1101-3> PMID: 12582851
80. Qin H, Chen M, Yi X, Bie S, Zhang C, Zhang Y, et al. Identification of associated SSR markers for yield component and fiber quality traits based on frame map and Upland cotton collections. *PLoS One*. 2015; 10:e0118073. <https://doi.org/10.1371/journal.pone.0118073> PMID: 25635680
81. Lacape J-M, Nguyen TB. Mapping quantitative trait loci associated with leaf and stem pubescence in cotton. *J Hered*. 2005; 96:441–444. <https://doi.org/10.1093/jhered/esi052> PMID: 15829730
82. He D-H, Lin Z-X, Zhang X-L, Nie Y-C, Guo X-P, Zhang Y-X, et al. QTL mapping for economic traits based on a dense genetic map of cotton with PCR-based markers using the interspecific cross of *Gossypium hirsutum* × *Gossypium barbadense*. *Euphytica*. 2006; 153:181–197. <https://doi.org/10.1007/s10681-006-9254-9>
83. Liu C, Yuan D, Zhang X, Lin Z. Isolation, characterization and mapping of genes differentially expressed during fibre development between *Gossypium hirsutum* and *G. barbadense* by cDNA-SRAP. *J Genet*. 2013; 92:175–181. <https://doi.org/10.1007/s12041-013-0238-y> PMID: 23970073
84. Shen X, Van Becelaere G, Kumar P, Davis RF, May OL, Chee P. QTL mapping for resistance to root-knot nematodes in the M-120 RNR Upland cotton line (*Gossypium hirsutum* L.) of the Auburn 623 RNR source. *Theor Appl Genet*. 2006; 113:1539–1549. <https://doi.org/10.1007/s00122-006-0401-4> PMID: 16960714
85. Shen X, Guo W, Lu Q, Zhu X, Yuan Y, Zhang T. Genetic mapping of quantitative trait loci for fiber quality and yield trait by RIL approach in Upland cotton. *Euphytica*. 2007; 155:371–380. <https://doi.org/10.1007/s10681-006-9338-6>
86. Wang J, Guo W, Zhang T. QTL mapping for fiber quality properties in cotton cultivar Yumian 1. *Acta Agron Sin*. 2007; 33:1915–1921.
87. Qin H, Guo W, Zhang Y-M, Zhang T. QTL mapping of yield and fiber traits based on a four-way cross population in *Gossypium hirsutum* L. *Theor Appl Genet*. 2008; 117:883–894. <https://doi.org/10.1007/s00122-008-0828-x> PMID: 18604518
88. Kalivas A, Xanthopoulos F, Kehagia O, Tsiftaris AS. Agronomic characterization, genetic diversity and association analysis of cotton cultivars using simple sequence repeat molecular markers. *Genet Mol Res*. 2011; 10:208–217. <https://doi.org/10.4238/vol10-1gmr998> PMID: 21341213

89. Pillay M, Myers GO. Genetic diversity in cotton assessed by variation in ribosomal RNA genes and AFLP markers. *Crop Sci.*; 1999; 39:1881. <https://doi.org/10.2135/cropsci1999.3961881x>
90. Bolek Y, El-Zik KM, Pepper AE, Bell AA, Magill CW, Thaxton PM, et al. Mapping of Verticillium wilt resistance genes in cotton. *Plant Sci.* 2005; 168:1581–1590. <https://doi.org/10.1016/j.plantsci.2005.02.008>
91. Xu H, Mei Y, Hu J, Zhu J, Gong P. Sampling a core collection of island cotton (*Gossypium barbadense* L.) based on the genotypic values of fiber traits. *Genet Resour Crop Evol.* 2006; 53:515–521. <https://doi.org/10.1007/s10722-004-2032-4>
92. Eberhart SA, Russell WA. Stability parameters for comparing varieties. *Crop Sci.* 1966; 6:36. <https://doi.org/10.2135/cropsci1966.0011183X000600010011x>
93. Liu S, Cantrell RG, McCarty JC, Stewart JM. Simple sequence repeat–based assessment of genetic diversity in cotton race stock accessions. *Crop Sci.* 2000; 40:1459–1469. <https://doi.org/10.2135/cropsci2000.4051459x>
94. Lacape J-M, Dessauw D, Rajab M, Noyer J-L, Hau B. Microsatellite diversity in tetraploid *Gossypium* germplasm: assembling a highly informative genotyping set of cotton SSRs. *Mol Breed.* 2006; 19:45–58. <https://doi.org/10.1007/s11032-006-9042-1>
95. Hinze LL, Gazave E, Gore MA, Fang DD, Scheffler BE, Yu JZ, et al. Genetic diversity of the two commercial tetraploid cotton species in the *Gossypium* diversity reference set. *J Hered.* 2016; 107:274–286. <https://doi.org/10.1093/jhered/esw004> PMID: 26774060
96. Adawy SS. An evaluation of the utility of simple sequence repeat loci (SSR), expressed sequence tags (ESTs) and expressed sequence tag microsatellites (EST-SSR) as molecular markers in cotton. *J Appl Sci Res.* 2007; 3:1581–1588.
97. Boopathi NM, Gopikrishnan A, Selvam NJ, Iyanar K, Muthuraman S, Saravanan N, et al. Genetic diversity assessment of *G. barbadense* accessions to widen cotton (*Gossypium* spp.) gene pool for improved fibre quality. *J Cotton Res Dev.* 2008; 22:135–138.
98. Wang XQ, Feng CH, Lin ZX, Zhang XL. Genetic diversity of Sea-Island cotton (*Gossypium barbadense*) revealed by mapped SSRs. *Genet Mol Res.* 2011; 10:3620–3631. PMID: 22183945
99. Abdellatif KF, Khidr YA, El-Mansy YM, El-Lawendey MM, Soliman YA. Molecular diversity of Egyptian cotton (*Gossypium barbadense* L.) and its relation to varietal development. *J Crop Sci Biotechnol.* 2012; 15:93–99. <https://doi.org/10.1007/s12892-011-0120-5>
100. Reddy UK, Nimmakayala P, Abburi VL, Reddy CVC, Saminathan T, Percy RG, et al. Genome-wide divergence, haplotype distribution and population demographic histories for *Gossypium hirsutum* and *Gossypium barbadense* as revealed by genome-anchored SNPs. *Sci Rep.* 2017; 7:41285. <https://doi.org/10.1038/srep41285> PMID: 28128280
101. Hamrick JL, Godt MJW. Allozyme diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL, Weir BS, editors. *Plant population genetics, breeding, and genetic resources.* Sunderland, Massachusetts: Sinauer Associates Inc.; 1990. pp. 43–63.
102. McGowan JC. History of extra-long staple cottons. A Thesis Submitted to the Faculty of Department of History of the University of Arizona. 1960. http://arizona.openrepository.com/arizona/bitstream/10150/553949/1/AZU_TD_BOX256_E9791_1960_82.pdf [verified on October2, 2017]
103. Stephens SG. A reexamination of the cotton remains from Huaca Prieta, North Coastal Peru. *Am Antiq. Society for American Archaeology*; 1975; 40:406–419. <https://doi.org/10.2307/279327>
104. Luikart G, Cornuet J-M. Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conserv Biol.*; 1998; 12:228–237.
105. Simongulyan NG., Mukhamedkhanov SR., Shafrin AN. Genetics, breeding and seed production of cotton. Tashkent: “Mehnat”; 1987. p. 18 (In Russian).
106. Tyagi P, Gore MA, Bowman DT, Campbell BT, Udall JA, Kuraparthi V. Genetic diversity and population structure in the US Upland cotton (*Gossypium hirsutum* L.). *Theor Appl Genet.* 2014; 127:283–295. <https://doi.org/10.1007/s00122-013-2217-3> PMID: 24170350
107. Wright S. *Evolution and the genetics of populations, Volume 4: Variability within and among natural populations.* University of Chicago Press; 1978. 590 p.
108. Hartl DL, Clark AG. *Principles of population genetics.* 4th edition. Sinauer AD, editor. Sunderland, Massachusetts: Sinauer Associates; 2007.
109. Jorde LB. Linkage disequilibrium and the search for complex disease genes. *Genome Res.* 2000; 10:1435–1444. <https://doi.org/10.1101/gr.144500> PMID: 11042143
110. Abdallah JM, Goffinet B, Cierco-Ayrolles C, Pérez-Enciso M. Linkage disequilibrium fine mapping of quantitative trait loci: a simulation study. *Genet Sel Evol.* 2003; 35: 513–532. <https://doi.org/10.1186/1297-9686-35-6-513> PMID: 12939203

111. Oraguzie NC, Wilcox PL, Rikkerink EHA, de Silva HN. Linkage disequilibrium. In: Oraguzie NC, Rikkerink EHA, Gardiner SE, de Silva HN, editors. Association mapping in plants. New York: Springer; 2007. pp. 11–39.
112. D'hoop BB, Paulo MJ, Kowitwanich K, Sengers M, Visser RGF, van Eck HJ, et al. Population structure and linkage disequilibrium unravelled in tetraploid potato. *Theor Appl Genet*. 2010; 121:1151–1170. <https://doi.org/10.1007/s00122-010-1379-5> PMID: 20563789
113. Fang DD, Hinze LL, Percy RG, Li P, Deng D, Thyssen G. A microsatellite-based genome-wide analysis of genetic diversity and linkage disequilibrium in Upland cotton (*Gossypium hirsutum* L.) cultivars from major cotton-growing countries. *Euphytica*. 2013; 191:391–401. <https://doi.org/10.1007/s10681-013-0886-2>
114. Zhang Z, Li J, Muhammad J, Cai J, Jia F, Shi Y, et al. High resolution consensus mapping of quantitative trait loci for fiber strength, length and micronaire on chromosome 25 of the Upland cotton (*Gossypium hirsutum* L.). *PLoS One*. 2015; 10:e0135430. <https://doi.org/10.1371/journal.pone.0135430> PMID: 26262992
115. Kraakman ATW. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics*. 2004; 168:435–446. <https://doi.org/10.1534/genetics.104.026831> PMID: 15454555
116. Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC. Linkage disequilibrium in European elite maize germplasm investigated with SSRs. *Theor Appl Genet*. 2005; 111:723–730. <https://doi.org/10.1007/s00122-005-2057-x> PMID: 15997389
117. Maccaferri M, Sanguineti MC, Noli E, Tuberosa R. Population structure and long-range linkage disequilibrium in a durum wheat elite collection. *Mol Breed*. 2005; 15:271–290. <https://doi.org/10.1007/s11032-004-7012-z>
118. Malysheva-Otto L, Ganai M, Röder M. Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). *BMC Genet*. 2006; 7:1–14.
119. Stich B, Maurer HP, Melchinger AE, Frisch M, Heckenberger M, van der Voort JR, et al. Comparison of linkage disequilibrium in elite European maize inbred lines using AFLP and SSR markers. *Mol Breed*. 2006; 17:217–226. <https://doi.org/10.1007/s11032-005-5296-2>
120. Brubaker CL, Paterson AH, Wendel JF. Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. *Genome*. 1999; 42:184–203. <https://doi.org/10.1139/g98-118>
121. Paterson AH, Smith RH. Future horizons: biotechnology of cotton improvement. *Cotton: origin, history, technology, and production*. New York, NY, USA: John Wiley & Sons; 1999. pp. 415–432
122. Mei H, Ai N, Zhang X, Ning Z, Zhang T. QTLs conferring FOV 7 resistance detected by linkage and association mapping in Upland cotton. *Euphytica*. 2014; 197:237–249. <https://doi.org/10.1007/s10681-014-1063-y>
123. Soto-Cerda BJ, Cloutier S. Association mapping in plant genomes. In: Caliskan M, editor. Genetic diversity in plants. InTech; 2012. pp. 29–54.
124. Percy RG. The worldwide gene pool of *Gossypium barbadense* L. and its improvement. In: Paterson AH, editor. "Genetics and genomics of cotton" *Plant Genetics and Genomics: Crops and Models* Volume 3. New York, NY: Springer US; 2009. pp. 69–97. https://doi.org/10.1007/978-0-387-70810-2_3
125. Wu Z, Soliman KM, Zipf A, Saha S, Sharma GC, Jenkins JN. Isolation and characterization of genes differentially expressed in fiber of *Gossypium barbadense* L. *J Cotton Sci*. 2005; 9:166–174.
126. Jenkins JN. Cotton. In: OECD, editor. *Traditional crop breeding practices: an historical review to serve as a baseline for assessing the role of modern biotechnology*. Paris: Organization for Economic Co-Operation and Development (OECD); 1993. pp. 61–70.
127. Zhang T, Qian N, Zhu X, Chen H, Wang S, Mei H, et al. Variations and transmission of QTL alleles for yield and fiber qualities in Upland cotton cultivars developed in China. *PLoS One*. 2013; 8:e57220. <https://doi.org/10.1371/journal.pone.0057220> PMID: 23468939
128. Badigannavar A. Characterization of quantitative traits using association genetics in tetraploid and genetic linkage mapping in diploid cotton (*Gossypium* spp). A PhD Dissertation Submitted to the Graduate Faculty of the Louisiana State University. 2010.
129. Shappley ZW, Jenkins JN, Zhu J, McCarty JC Jr., Zhu J. Quantitative trait loci associated with agronomic and fiber traits of upland cotton. *J Cotton Sci*. 1998; 2:153–163.
130. Karademir E, Karademir C, Ekininci R, Gencer O. Relationship between yield, fiber length and other fiber-related traits in advanced cotton strains. *Not Bot Horti Agrobot Cluj-Napoca*. 2010; 38:111–116.
131. Asif M, Mirza JI, Zafar Y. Genetic analysis for fiber quality traits of some cotton genotypes. *Pak J Bot*. 2008; 40:1209–1215.

132. Liu H, Quampah A, Chen J, Li J, Huang Z, He Q, et al. QTL Mapping based on different genetic systems for essential amino acid contents in cottonseeds in different environments. PLoS One. 2013; 8: e57531. <https://doi.org/10.1371/journal.pone.0057531> PMID: 23555562
133. Fang L, Tian R, Li X, Chen J, Wang S, Wang P, et al. Cotton fiber elongation network revealed by expression profiling of longer fiber lines introgressed with different *Gossypium barbadense* chromosome segments. BMC Genomics. 2014; 15:838. <https://doi.org/10.1186/1471-2164-15-838> PMID: 25273845
134. Fan L, Hu W-R, Yang Y, Li B. Plant special cell—cotton fiber. In: Mworia JK, editor. Botany. Shanghai, China: InTech; 2012. pp. 211–226.
135. Tuttle JR, Nah G, Duke MV, Alexander DC, Guan X, Song Q, et al. Metabolomic and transcriptomic insights into how cotton fiber transitions to secondary wall synthesis, represses lignification, and prolongs elongation. BMC Genomics; 2015; 16:477. <https://doi.org/10.1186/s12864-015-1708-9> PMID: 26116072
136. Chee PW, Campbell BT. Bridging classical and molecular genetics of cotton fiber quality and development. In: Paterson AH, editor. Genetics and Genomics of Cotton. New York, NY: Springer US; 2009. pp. 283–311. https://doi.org/10.1007/978-0-387-70810-2_12
137. Meredith WR. Genetics and management factors influencing textile fiber quality. In: Chewing C, editor. 7th Ann Cotton Incorporated Engineered Fiber Selection System Res Forum. Raleigh, N.C.: Cotton Incorporated; 1994. pp. 256–261.
138. Jenkins JN, McCarty JC, Wubben MJ, Hayes R, Gutierrez OA, Callahan F, et al. SSR markers for marker assisted selection of root-knot nematode (*Meloidogyne incognita*) resistant plants in cotton (*Gossypium hirsutum* L.). Euphytica. 2012; 183:49–54
139. Bolek Y, Hayat K, Bardak A, Azhar MT. Molecular breeding of cotton. In: Abdurakhmonov IY, editor. Cotton Research. Rijeka: InTech Press; 2016. Pp. 123–166.
140. Abdurakhmonov IY. Modern high-biotechnologies for improvement of superior fibre, productive and early maturing Upland cotton cultivars. In: Cotton India 2015–16: Weaving the world of cotton together; 2016 Feb 22–24; Goa, India: Cotton Association of India; 2016. p.36–44.