

Genetic polymorphism detection in semi-domesticated perennial cotton (*Gossypium* ssp.) using an ISSR marker system and its application for molecular interspecific differentiation

Fernando dos Santos Araújo (✉ fernandoaraujoagro177@gmail.com)

University of Pernambuco: Universidade de Pernambuco <https://orcid.org/0000-0002-0605-1613>

Riselane de Lucena Alcântara Bruno

Paraíba Federal University: Universidade Federal da Paraíba

Nair Helena Castro Arriel

Brazilian Agricultural Research Corporation: Empresa Brasileira de Pesquisa Agropecuária

Everaldo Paulo de Medeiros

Brazilian Agricultural Research Corporation: Empresa Brasileira de Pesquisa Agropecuária

Liziane Maria de Lima

Brazilian Agricultural Research Corporation: Empresa Brasileira de Pesquisa Agropecuária

Mayara Andrade de Souza

Centro de Estudos Superiores de Maceió: Centro Universitario CESMAC

Alberício Pereira de Andrade

Federal University of Pernambuco: Universidade Federal de Pernambuco

Richeliel Albert Rodrigues Silva

Federal University of Campina Grande: Universidade Federal de Campina Grande

Francival Cardoso Felix

Universidade Federal do Paraná: Universidade Federal do Paraná

Karialane da Silva Belarmino

Universidade Federal da Paraíba: Universidade Federal da Paraíba

Research Article

Keywords: *Gossypium*, genetic resource, molecular markers, DNA fingerprinting

Posted Date: August 9th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1902987/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Molecular Biology Reports on January 18th, 2023. See the published version at <https://doi.org/10.1007/s11033-022-08165-8>.

Abstract

Background The perennial cotton species *Gossypium hirsutum* L. r. marie-galante Hutch. and *Gossypium barbadense* L. are sources of variability for creating modern cotton varieties. However, these species are similar in shape and easily confused in young stages of their life, leading to difficulties in identification based on their morphology. Thus, in this study, an Inter-Simple Sequence Repeat (ISSR) marker system was used as a measure of genetic differentiation among 28 genotypes of perennial cotton from a Brazilian collection.

Methods and Results A set of eleven ISSR primers yielded 101 bands, of which 48 (47.5%) were polymorphic. The mean values of polymorphism information content ($PIC=0.304$) and polymorphism ($P=45.8\%$) showed that the ISSR primers are moderately informative. The ISSR markers exposed the high genetic differentiation ($G_{ST}=0.598$) and a low level of gene flow ($N_m=0.338$) between species, which suggests a pattern of reproductive isolation. The Cluster Analysis, based on Nei's genetic identity, clustered the 28 genotypes into two groups consistent with the taxonomical delimitation, occurrence mode (semi-domesticated and improved types) and partial concordance with geographic origin. Bayesian model-based structural analysis also suggests the existence of two genetic groups (Delta K = 2) and high membership likelihood ($Q>0.98$).

Conclusions The ISSR marker system offers a new molecular approach to differentiate the *G. hirsutum* L. r. marie-galante and *G. barbadense* L. species. This study can expand the molecular marker resources for the identification and improvement of our knowledge about the genetic relationships between perennial cotton genotypes from Brazilian collections.

Background

The cotton fibre (*Gossypium* L.) consumed by the Brazilian textile industry comes from modern upland cotton (*Gossypium hirsutum* L. r. latifolium). In the past, this raw material was mainly obtained from perennial cotton species: *Gossypium hirsutum* L. r. marie-galante and *Gossypium barbadense* L. These allotetraploid species are not native to Brazil but are widely distributed and adapted to conditions there, which is considered a secondary diversity centre of the genus *Gossypium* L. [1–3]. The *G. barbadense* L. (kidney-cotton and non-aggregated seeds races) and *G. hirsutum* L. r. marie-galante (mocó cotton ecotype) are found “*in situ*” in the form of subsistence crops, semi-wild populations and dooryard plants [4–5].

These perennial cottons are used to spin and weave lint, and are used for ornamental and medicinal purposes by the local populations of North and Northeast Brazil [4]. In addition, they have advantageous traits and are potentially applied to improve fibre quality (*G. barbadense* L.) as well as having a high tolerance for environmental stresses (*G. hirsutum* L. r. marie-galante) [6–7], helping to develop modern cotton cultivars.

The conservation of Brazilian cotton's genetic resources has been coordinated by the Brazilian Agricultural Research Agency (Embrapa) which holds almost 1,700 Brazilian accessions of *G. hirsutum* L. r. marie-galante and *G. barbadense* L. together in the Active Germplasm Bank (AGB), from an Embrapa unit (Embrapa Cotton, Campina Grande, Paraíba, Brazil) [8]. However, the botanical identities of several accessions of perennial cotton in this collection remain undefined because these species are similar in shape and are easily confused in the young stages of their life, leading to difficulties in their identification based on their morphology. Therefore, a new identification approach is critical to the authentication and improvement of the use of this germplasm.

The Inter-Simple Sequence Repeat (ISSR) marker system has been used as a valuable marker for revealing inter and intraspecific variations in cotton [9]. Therefore, it is a genotyping technique which is potentially suitable for DNA fingerprinting and determining the genetic relationships of semi-domesticated perennial cotton. The ISSR marker system is a technique that involves the use of microsatellite sequences as primers to generate multilocus markers [10]. The ISSRs are highly polymorphic, relatively easy to develop, and inexpensive when compared to other methods [11].

The hypothesis in this study was that the ISSR molecular markers can provide estimates with high discrimination power between semi-domesticated perennial cotton species, with low cost and high reproducibility. In this study, an ISSR marker system was used to assess genetic differentiation and the relationships between 28 perennial cotton genotypes of *G. hirsutum* L. r. marie-galante and *G. barbadense* L. from a Brazilian cotton germplasm collection.

Materials And Methods

Cotton germplasm

This study used 28 perennial cotton genotypes (25 semi-domesticated genotypes and 3 improved genotypes) provided by the Brazilian Agricultural Research Corporation (Embrapa Cotton, Campina Grande, Brazil) (Table 1). The improved genotypes CNPA 5M, BRS 200 (*G. hirsutum* L. r. marie-galante) and Pima S7, EUA (*G. barbadense* L.) were considered as a control group (Table 1). In this investigation, improved types were defined as those which received human manipulation through breeding techniques, while semi-domesticated types were defined to include landraces, dooryard, and semi-wild cottons.

Table 1

Genotype code, botanical identity (species), occurrence mode, origin (state/country) of 28 perennial cotton genotypes (*G. hirsutum* L. r. marie-galante and *G. barbadense* L.) from a Brazilian cotton germplasm collection

ID number	Botanical identity (specie)	Occurrence mode	Origin (state/country)
1	<i>G. hirsutum</i> L. r. marie-galante	Landrace	Paraíba, Brazil
2	<i>G. hirsutum</i> L. r. marie-galante	Landrace	Rio Grande do Norte, Brazil
3	<i>G. hirsutum</i> L. r. marie-galante	Dooryard	Paraíba, Brazil
4	<i>G. hirsutum</i> L. r. marie-galante	Dooryard	Paraíba, Brazil
5	<i>G. hirsutum</i> L. r. marie-galante	Dooryard	Rio Grande do Norte, Brazil
6	<i>G. hirsutum</i> L. r. marie-galante cv. CNPA 5M	Cultivar	Paraíba, Brazil
7	<i>G. hirsutum</i> L. r. marie-galante	Semi-wild	Paraíba, Brazil
8	<i>G. hirsutum</i> L. r. marie-galante	Dooryard	Paraíba, Brazil
9	<i>G. hirsutum</i> L. r. marie-galante cv. BRS 200	Cultivar	Paraíba, Brazil
10	<i>G. hirsutum</i> L. r. marie-galante	Landrace	Paraíba, Brazil
11	<i>G. hirsutum</i> L. r. marie-galante	Landrace	Paraíba, Brazil
12	<i>G. hirsutum</i> L. r. marie-galante	Landrace	Paraíba, Brazil
13	<i>G. hirsutum</i> L. r. marie-galante	Landrace	Paraíba, Brazil
14	<i>G. hirsutum</i> L. r. marie-galante	Semi-wild	Rio Grande do Norte, Brazil
15	<i>G. hirsutum</i> L. r. marie-galante	Semi-wild	Rio Grande do Norte, Brazil
16	<i>G. hirsutum</i> L. r. marie-galante	Landrace	Pernambuco, Brazil
17	<i>G. barbadense</i> L.	Semi-wild	Minas Gerais, Brazil
18	<i>G. barbadense</i> L.	Dooryard	Minas Gerais, Brazil
19	<i>G. barbadense</i> L.	Dooryard	Minas Gerais, Brazil
20	<i>G. barbadense</i> L.	Dooryard	Minas Gerais, Brazil
21	<i>G. barbadense</i> L.	Dooryard	Minas Gerais, Brazil
22	<i>G. barbadense</i> L.	Dooryard	Pará, Brazil
23	<i>G. barbadense</i> L.	Dooryard	Pará, Brazil
24	<i>G. barbadense</i> L.	Dooryard	Amapá, Brazil
25	<i>G. barbadense</i> L.	Dooryard	Pará, Brazil
26	<i>G. barbadense</i> L.	Dooryard	Ceará, Brazil
27	<i>G. barbadense</i> L.	Dooryard	Amapá, Brazil
28	<i>G. barbadense</i> L. cv. Pima S7	Cultivar	Arizona, United States

(Insert Table 1)

Genomic DNA extraction and quantification

Genomic DNA was extracted from samples of cotyledon tissue from dry seeds (1/2 of the cotyledon) using an adapted protocol by Doyle and Doyle [12]. The concentration and quality of the genomic DNA were measured by spectrophotometry. The quality of DNA extracted was also checked by running the electrophoresis on a 0.8% agarose gel, using a lambda phage DNA standard (50, 100 and 200 ng μL^{-1}). A ratio absorbance of 260/280 nm and 260/230 nm (more than 1.8) were considered as standard.

ISSR analysis

The DNA amplification was performed by conventional polymerase chain reaction (PCR) using 11 ISSR primers from the Nucleic Acid-Protein Service Unit (University of British Columbia, USA). The PCR were prepared with a final volume of 25 μL , containing 1.0 mM primer, 0.25 mM dNTP's, 2.5 mM MgCl_2 , PCR buffer [100 mM Tris-HCl (pH 8.5)], 500 mM KCl], 1 unit of Taq DNA polymerase (Ludwig Biotec®), 20 ng of DNA template and ultrapure water.

DNA amplification was performed in a thermal cycler (PCR Thermal Cyclers, Amplitherm 96-Well). The protocol consisted of an initial denaturation at 94°C for 5 min, 40 amplification cycles (denaturation at 94°C for 1 min, annealing at 45°C for 1 min and extension at 72°C for 2 min) and a final extension at 72°C for 5 min. The amplicons were cooled to 4°C and stored at -20°C. All PCRs were performed in duplicate. Reactions without the DNA template were used as a negative control.

The dyes SYBR® Gold nucleic acid gel stain (Invitrogen™) and bromophenol Blue (0.01%) were added to the amplicons. The samples were applied on an agarose gel (1.5% m v^{-1}) with a TBE 0.5X running buffer (Tris, boric acid, EDTA). The applied voltage was 75 V and the migration time of the fragments was 120 min. A molecular weight standard (1 Kb Plus Express® DNA Ladder) was used. The DNA fragments (bands) were photodocumented using an Ultra Lum Electronic UV Transilluminator and Kodak GEL Logic 200 Imaging System.

Scoring ISSR band and marker efficiency analysis

The electrophoretic profiles were coded as '1' presence and '0' absence of the band at the same loci, according to the visible and repeatable bands on the gel-electrophoresis. The primer's performance was measured by calculating different parameters including the total number of bands, polymorphism (%), number of specific bands (defined as bands that occurred in only one of the species) and Polymorphism Information Content (PIC), calculated according by Botstein [13].

Genetic differentiation and relationship analysis

Based on the band profiles (presence/absence of bands), the genetic differentiation, Nei's genetic identity [14] and indirect gene flow [15] were calculated using the POPGENE software (version 1.32) [16].

A similarity dendrogram based on the UPGMA method (Unweighted Pair-Group Method using Arithmetic Averages) was conducted using a matrix of Nei's genetic identity values calculated in NTSYS-pc software (version 2.1) [17]. The structuring of the clusters was interpreted by considering the species (*G. hirsutum* L. r. *marie-galante* and *G. barbadense* L.) and the mode of occurrence (improved and semi-domesticated types).

Genetic structure and admixture analysis

STRUCTURE software (version 2.3.4.) [18] was used to infer the population structure profile and admixture detection. The program was run using the admixture model and assuming the mixed ancestry model, frequency of correlated alleles and prior information from the population (prior taxonomic classification). To estimate the most probable number of genetic groups (K-value), a burn-in of 5,000 followed by 10,000 Markov Chain Monte Carlo simulations, at 5 iterations with 10 autonomous runs, were performed with a K-value pre-set from 1 to 6.

The results of the structural analysis were analysed by STRUCTURE HARVESTER software 'on-line' (version 0.6.93) (<http://taylor0.biology.ucla.edu/structureHarvester/>) [19] to determine: $\Delta K = (|L'(K)|) / \text{sd}(L(K))$ (mean); $L'(K)$ = rate of change of the likelihood distribution (mean); $L''(K)$ = absolute value of the 2nd order rate of the likelihood distribution change

(mean); and Lnprob (K) = mean of estimated Ln probability by Evanno [20]. The genotypes were allocated to a specific cluster (K) based on the membership likelihood ($Q > 0.60$).

Result

Electrophoretic profiles and ISSR marker efficiency

The detailed electrophoretic profiles and marker efficiency analysis of 11 ISSR primers are given in Table 2. The PCR-ISSR amplification of genomic DNA from 28 perennial cotton genotypes yielded a total of 101 bands, of which 47.58% were polymorphic. The polymorphic band values for ISSR primers ranged from 12–73% and the average was found to be 45.8%, while the PIC ranged from 0.161 to 0.452 (an average of 0.304). The primers UBC-872, UBC-812, UBC-823 and UBC-828 were more informative, in terms of polymorphism and PIC, but the UBC-808, UBC-812, UBC-866 and UBC-828 primers revealed specific bands in two species. However, UBC-814 and UBC-823 primers only revealed specific bands of *G. hirsutum* L. r. marie-galante.

Table 2
Summary of banding profile and polymorphism revealed by 11 ISSR primers

Primer code	Sequence 5' → 3'	Total bands	Polymorphic bands	Polymorphism (%)	Private bands		PIC
					<i>G. hirsutum</i>	<i>G. barbadense</i>	
UBC-808	(AG)8 C	12	5	42	2	2	0.262
UBC-812	(GA)8 A	14	9	64	2	3	0.452
UBC-814	(CT)8 A	8	2	25	1	0	0.271
UBC-823	(TC)8 C	8	5	62	1	0	0.357
UBC-827	(AC)8 G	6	3	50	0	0	0.328
UBC-828	(TG)8 A	8	5	62	1	4	0.451
UBC-834	(AG)8 YT	9	3	33	0	0	0.250
UBC-866	(CTC)6	8	3	37	1	1	0.226
UBC-872	(GATA)4	11	8	73	0	0	0.397
UBC-884	HBH (AG)7	9	4	44	0	0	0.191
UBC-892	TAGATCTGATATCTGAATTCC	8	1	12	0	0	0.161
Total		101	48	48	8	10	-
Means		9	4	46	0.7	0.9	0.304

(Insert Table 2)

Genetic differentiation and relationship among genotypes

The interspecific genetic differentiation was high ($G_{ST}=0.598$), showing that 59.98% of the variability is distributed between the species. The indirect gene flow ($N_m = 0.338$) showed that gene exchange between species was low, which strengthens genetic differentiation between species.

The genetic relationship of 28 perennial cotton genotypes was attained from ISSR primers scoring a data set using Nei's genetic identity coefficient. The magnitude of the relatedness and disparity among the genotypes is demonstrated in Fig. 1.

(Insert Fig. 1)

The mean values of Nei's genetic identity, calculated among pairs of genotypes, ranged from 0.886 to 0.990 in *G. barbadense* L.; 0.828 to 0.971 in *G. hirsutum* L. r. marie-galante; and 0.657 to 0.714 between species. The genotypes were grouped into two clusters. One cluster included the 16 *G. hirsutum* L. r. marie-galante genotypes and the other group comprised 12 *G. barbadense* L. genotypes. The improved genotypes were grouped separately from the semi-domesticated genotypes (Fig. 1).

The interspecific genetic similarity was greater between the improved genotypes CNPA 5M and BRS 200 of the group *G. hirsutum* L. r. marie-galante and the improved genotype Pima S7 of the group *G. barbadense* L. (Fig. 1). Semi-domesticated genotypes collected from the same provinces showed close genetic relationships (Fig. 1). However, some exceptions were observed, including a genotype of *G. barbadense* L. collected in Minas Gerais State (17 genotypes) and two genotypes of *G. hirsutum* L. r. marie-galante collected in Paraíba (7 genotypes) and Rio Grande do Norte State (5 genotypes) that differed from the other genotypes collected in the same provinces (Fig. 1).

Genetic structure and admixture analysis

The genetic structure of 28 perennial cotton genotypes was estimated based on Bayesian inference. The parameters for estimating the most probable number of genetic groups (K) and membership likelihood of the genotypes to a specific cluster (K) are demonstrated in Fig. 2.

(Insert Fig. 2)

The most probable number of genetic groups formed by the resulting data set, estimated using the Evanno method, was $K = 2$ [$\text{Lnprob}(K) = -546.260$; $\text{Ln}^i(K) = 821.760$; $\text{Ln}^n(K) = 820.920$; and $\Delta K = 217.864$] (Fig. 2A-D). Regarding the membership likelihood ($Q > 0.60$), all 16 *G. hirsutum* L. r. marie-galante genotypes (green zone) and 12 *G. barbadense* L. genotypes (red zone) were recorded as highly pure (Fig. 2E). Our STRUCTURE analysis findings had a similar trend to the genotypic relatedness revealed by UPGMA clustering, resulting in all of the genotypes being in two distinct clusters.

Discussion

The ISSR primers allowed the detection of genetic polymorphism in perennial cotton. High levels of polymorphism (45.8%) represented the genetic variation in genomic ISSR regions of the combined dataset (28 genotypes of *G. hirsutum* L. r. marie-galante and *G. barbadense* L.). Therefore, this technique proved itself to be effective in differentiating species using little resources and with the possibility of high reproducibility.

The efficiency of individual ISSR primers was variable. Ordinarily, the efficiency of a certain ISSR primer is correlated with the level of polymorphism that could be generated among the genotypes [21]. However, our results (mean values 0.304 for PIC and 45.8% for polymorphism) were similar to those reported in several studies that assessed the genetic diversity of

Gossypium spp. using ISSR markers [22–28]. Thus, the electrophoretic profiles and ISSR marker efficiency in this study are within the expected range for *Gossypium* L.

All primers showed polymorphism, suggesting the efficacy of these ISSR markers for the assessment of genetic variation between the *Gossypium* L. species. In this way, all of the primers can be effectively used in the primary evaluation of perennial cotton germplasm. However, UBC-808, UBC-812, UBC-866, UBC-828, UBC-814, and UBC-823 primers revealed specific bands in *G. hirsutum* L. r. marie-galante (eight bands) and *G. barbadense* L. (ten bands). This finding can be useful for botanical identification, enabling authentication and qualitative identification of errors of classification in cotton germplasm collections.

The genetic polymorphism revealed by the ISSR marker system was effective in discriminating two perennial cotton species, as well as revealing intraspecific genetic relationships. Genetic analysis of population samples of *G. hirsutum* L. r. marie-galante and *G. barbadense* L. provided evidence of three levels of genetic differentiation and two levels of reproductive isolation between species.

The high interspecific genetic differentiation ($G_{ST} = 0.598$) was consistent with the other differentiation index, including genetic structuring in two groups ($K = 2$) and low genetic identity (0.657) between species. The absence of mixing ($Q > 0.98$) and the low gene flow ($Nm = 0.338$) suggest the existence of reproductive barriers between *G. hirsutum* L. r. marie-galante and *G. barbadense* L.

In this context, despite these two species being sexually compatible [29] the low gene flow can be attributed to geographic isolation. As a matter of fact, natural hybrids between *G. hirsutum* L. and *G. barbadense* L. do not occur *in situ*, or occur with low frequency due to the rare places where they appear in sympatry [30]. Another hypothesis can be attributed to the habitats where *G. hirsutum* L. r. marie-galante occurs in the semi-arid region of Northeast Brazil which, generally, does not have adequate rainfall for the establishment of *G. barbadense* L. [31].

Intraspecific genetic similarity was high in both species but the genotypes of *G. barbadense* L. (0.938) showed greater genetic identity than the genotypes of *G. hirsutum* L. r. maria galante (0.899). These results suggest that the genetic diversity in these groups is low or moderate, in agreement with the results reported by Hinze [32–33] who also reported low genetic diversity in reference groups of these species. The high similarity between the semi-domesticated genotypes of the two species indicates that there is a potential risk of decline in genetic variation, requiring the inclusion of effective conservation measures. Nonetheless, the high divergence between semi-domesticated and improved genotypes suggests that hybridization between these types can result in more heterotic combinations.

The high divergence between semi-domesticated and improved genotypes shows the usefulness of including them in breeding programs to increase the variability within germplasm collections. The smallest similarity observed between the semi-domesticated genotypes and the improved Pima S7 genotype of the *G. barbadense* L. group was also reported by [6] and attributed to the wide differentiation between the gene pools of origin. The improved genotypes CNPA 5M and BRS 200 of the group *G. hirsutum* L. r. marie-galante are different from the semi-domesticated genotypes of this group. This divergence stems from the selection of precocity and fibre colour characteristics which were applied during the creation of the improved genotypes CNPA 5M and BRS 200, respectively [34].

The divergence between semi-domesticated genotypes collected in the same province suggests that semi-domesticated populations of *G. hirsutum* L. r. marie-galante and populations of *G. barbadense* L. were originally formed by germplasm from different sources. However, the introduction of new sources of germplasm mediated by humans in these populations also seems plausible, as the exchange of seeds is a common practice among horticulturists and farmers in different Brazilian regions.

Conclusion

The ISSR marker system offers a new molecular approach to differentiate the *G. hirsutum* L. r. marie-galante and *G. barbadense* L. species. This study can expand the molecular marker resources to the identification and improvement of our knowledge of the genetic relationships between perennial cotton genotypes from germplasm collections.

Declarations

Acknowledgements

The authors express their gratitude to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Brazil) for funding scholarships (Finance Code 001), the Brazilian Agricultural Research Agency (Brazil) for providing the germplasm and institutional support, and Kyvia Pontes for her help in creating the artwork (the figures).

Funding

This study was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil (Finance Code 001) and Brazilian Agricultural Research Agency (Brazil).

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Author Contributions Fernando dos Santos Araújo, Riselane de Lucena Alcântara Bruno, Nair Helena Castro Arriel and Everaldo Paulo de Medeiros contributed to the study conception and design. Material preparation, data collection and analysis were performed by Liziane Maria de Lima and Mayara Andrade de Souza. The first draft of the manuscript was written by Fernando dos Santos Araújo, Riselane de Lucena Alcântara Bruno, Nair Helena Castro Arriel and Alberício Pereira de Andrade. Francival Cardoso Felix, Richeliel Albert Rodrigues Silva and Karialane da Silva Belarmino commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

1. Borém A, Freire EC, Penna JCV and Barroso PAV (2003) Considerations about cotton gene escape in Brazil: a review. *Crop Breeding and Applied Biotechnology* 3: 315–332.
2. Almeida VCD, Hoffmann LV, Yokomizo GKI, Costa JND, Giband M and Barroso PAV (2009) In situ and genetic characterization of *Gossypium barbadense* populations from the States of Pará and Amapá, Brazil. *Pesquisa Agropecuária Brasileira* 44: 719–725.
3. Menezes IPP, Hoffmann LV, Lima TH, Silva AR, Lucena VS and Barroso PAV (2017) Genetic diversity of arboreal cotton populations of the Brazilian semiarid: a remnant primary gene pool for cotton cultivars. *Genetics and Molecular Research* 16: gmr16039659.
4. Menezes IPP, Hoffmann LV, Silva JO and Barroso PAV (2018). Distribuição do modo de ocorrência no local de raças de algodoeiro Semiárido Brasileiro. *Multi-Sci Journal* 1: 39–47.
5. Hoffmann LV, Cardoso KCM, Rocha ASNC, Oliveira AID, Abreu AG, Pereira CCO, Malafaia G and Menezes IPP (2018) Genetic diversity of *Gossypium barbadense* from the central Brazilian Amazon. *Acta Amazonica* 48: 1–9.
6. Rodrigues JIS, Carvalho LP and Farias FJC (2016) Comparison of wild accessions of *Gossypium barbadense* L. from Peru and Brazil via microsatellite markers. *Bioscience Journal* 32: 1352–1363.

7. Vasconcelos UAA, Cavalcanti JJV, Farias FJC, Vasconcelos WS and Santos RC (2018) Diallel analysis in cotton (*Gossypium hirsutum* L.) for water stress tolerance. *Crop Breeding and Applied Biotechnology* 18: 24–30.
8. Albrana (2022) Algodões brasileiros nativos e naturalizados. CNPA Embrapa. <http://www.cnpa.embrapa.br/albrana>. Accessed 10 January 2022
9. Sabev P, Valkova N and Todorovska EG (2020) Molecular markers and their application in cotton breeding: progress and future perspectives. *Bulgarian Journal of Agricultural Science* 26: 816–828.
10. Marwal A and Gaur RK (2020). Molecular markers: tool for genetic analysis. In: Verma A, Singh A. (ed) *Animal Biotechnology*. Cambridge: Academic Press Elsevier, pp. 353–372.
11. Gemmill CE and Grierson ER (2021) Inter-Simple Sequence Repeats (ISSR), microsatellite-primed genomic profiling using universal primers. *Mol Plant Taxonomy* 2222: 249–262.
12. Doyle JJ and Doyle JL (1987) Isolation of plant DNA from fresh tissue. *Focus* 12: 13–15.
13. Botstein D, White RL, Skolnick M and Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32: 314–331.
14. Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* 70: 3321–3323.
15. McDermott JM and McDonald BA (1993) Gene flow in plant pathosystem. *Annual Review of Phytopathology* 31: 353–373.
16. Yeh FC, Boyle TYZ and Xian JM (1999) *POPGENE: Microsoft Window-based freeware for population genetic analysis. Version 1.31*. Edmonton: University of Alberta.
17. Rohlf FJ (2000) NTSYS-pc: numerical taxonomy and multivariate analysis system: version 2.1. Exeter Software, New York
18. Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959
19. Earl DA and Von-Holdt BM (2012) Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361.
20. Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* 14: 2611–2620.
21. Khan MMH, Rafii MY, Ramlee SI, Jusoh M, Al Mamun M and Halidu J. (2021). DNA fingerprinting, fixation-index (Fst), and admixture mapping of selected Bambara groundnut (*Vigna subterranea* [L.] Verdc.) accessions using ISSR markers system. *Scientific Reports* 1: 1–23.
22. Bilwal BB, Vadodariya KV, Rajkumar BK, Upadhyay AK and Lahane GR (2017) Genetic diversity of parents using RAPD, ISSR and SSR molecular markers in upland cotton (*Gossypium hirsutum* L.). *Bulletin of Environment, Pharmacology and Life Sciences* 6: 51–57.
23. Ghuge SB, Mehetre SS, Chimote VP, Pawar BD and Naik RM (2018) Molecular characterization of cotton genotypes using SSR, ISSR and RAPD markers in relation to fiber quality traits. *Journal of Cotton Research* 32: 1–12.
24. Nimbal S, Kuldeep J and Pawan K (2020) Molecular diversity analysis among elite genotypes of American cotton (*Gossypium hirsutum* L.). *Pantnagar Journal of Research* 18: 1–4.
25. Tajik-Khaveh Z, Iranbakhsh A, Vafaie-Tabar M and Ebadi M (2020) Study of genetic variation among some *Gossypium* spp. genotypes available in Iran using ISSR molecular marker. *Rostaniha* 21: 264–277.
26. Şahin CB, İşler N and Rustamova V (2020) Characterization of some cotton varieties using ISSR markers. *KSU Journal of Agriculture and Nature* 23: 108–116.
27. Hocaoglu-Ozyigit A, Ucar B, Altay V and Ozyigit II (2020) Genetic diversity and phylogenetic analyses of Turkish cotton (*Gossypium hirsutum* L.) lines using ISSR markers and chloroplast trnL-F regions. *Journal of Natural Fibers* 17: 1–14.

28. Sevindik E, Çayir EM, Emrebaş T and Filiz E (2021) Assessment of phylogenetic relationships of some worldwide-cultivated cotton genotypes (*Gossypium hirsutum* L.) by using ISSR and RAPD markers. *Biologica Nyssana* 12: 131–139.
29. Stephens SG (1946) The genetics of "Corky". The New World alleles and their possible role as an interspecific isolating mechanism. *Journal of Genetics* 47: 150–161.
30. Pereira GS, Sousa RL, Araújo RL, Hoffmann LV, Silva EF and Barroso PAV (2012). Selective fertilization in interspecific crosses of allotetraploid species. *Botany* 90: 159–166.
31. Freire EC (2000) Distribuição, coleta, uso e preservação das espécies silvestres de algodão no Brasil. Embrapa. <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/272012/distribuicao-coleta-uso-e-preservacao-das-especies-silvestres-de-algodao-no-brasil>. Accessed 20 Jan 2022.
32. Hinze LL, Fang DD, Gore MA, Scheffler BE, Yu JZ, Frelichowski J and Percy RG (2015) Molecular characterization of the *Gossypium* diversity reference set of the US national cotton germplasm collection. *Theoretical and Applied Genetics* 128: 313–327.
33. Hinze LL, Gazave E, Gore MA, Fang DD, Scheffler BE, Yu JZ, Jones DC, Frelichowski J, Percy RG (2016) Genetic diversity of the two commercial tetraploid cotton species in the *Gossypium* diversity reference set. *Journal of Heredity* 107: 274–286.
34. Embrapa (2001) BRS 200 marron: cultivar de algodão de fibra colorida. Embrapa Algodão. <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/276144/brs-200-marrom-cultivar-de-algodao-de-fibra-colorida> . Accessed 20 Jan 2022.

Figures

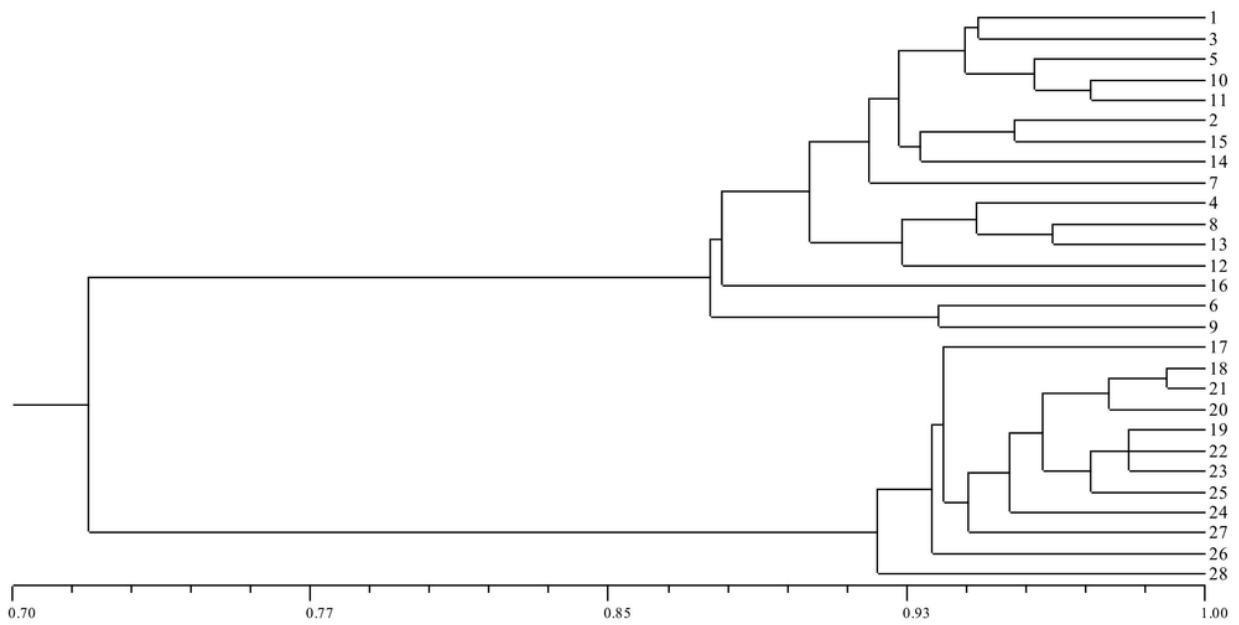


Figure 1

Dendrogram UPGMA illustrating the relationship between 28 perennial cotton genotypes based on Nei's genetic identity. The numeric value beneath the branches displayed the Nei's genetic identity value, while the numeric value on the right indicates genotype ID numbers according to Table 1 (1 to 16 = *G. hirsutum* L. r. *marie-galante*; 17 to 28 = *G. barbadense* L.)

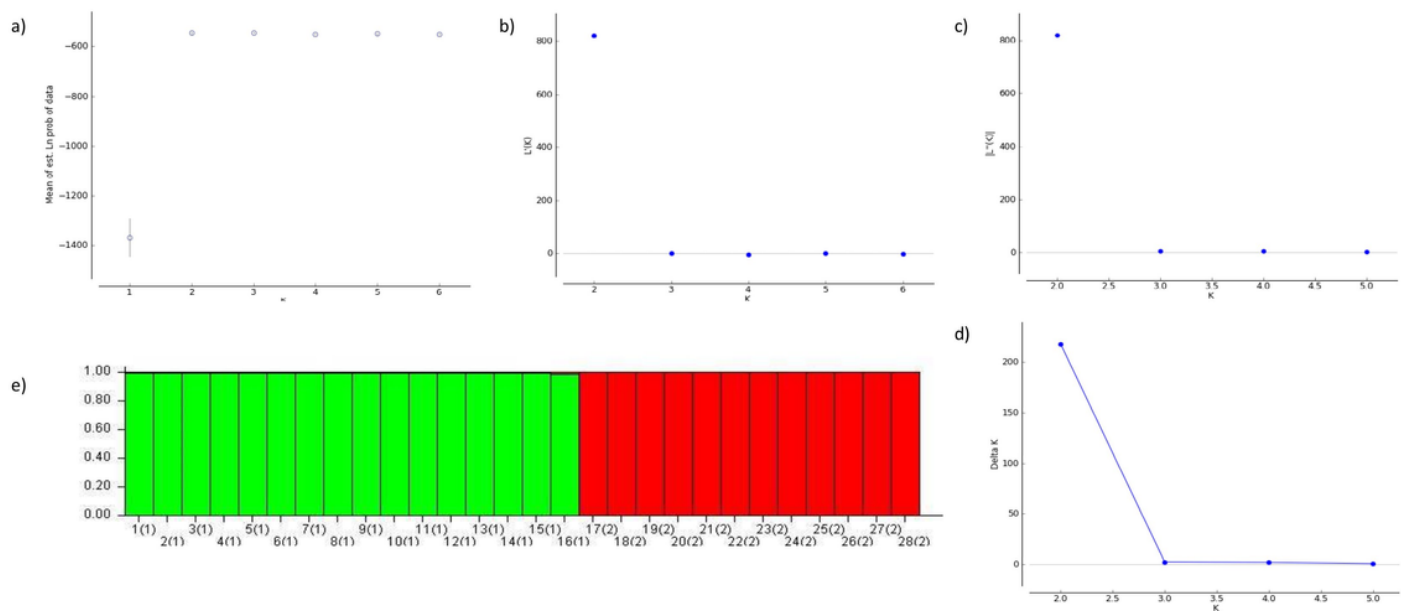


Figure 2

Genetic structure of 28 perennial cotton genotypes by Bayesian inference. (A) Mean of estimated Ln probability change rate of the likelihood distribution; (B) change rate of the likelihood distribution (mean); (C) absolute value of the 2nd order change rate of the likelihood distribution (mean); (D) Delta K (mean); (E) Bar plot based on estimated membership coefficient values (Q) for maximum K value (Delta K = 2). Each vertical line represents an accession and different colours represent the estimated membership coefficients. The numbers at the base of each vertical line indicate the genotypes' ID numbers according to Table 1 (1 to 16 = *G. hirsutum* L. r. marie-galante; 17 to 28 = *G. barbadense* L.) while the values inside the parentheses indicate the two groups based on prior taxonomic classification (1 = *G. hirsutum* L. r. marie-galante; 2 = *G. barbadense* L.)