

## ORIGINAL ARTICLE

# Genetic diversity of *Gossypium barbadense* from the central Brazilian Amazon

Lúcia Vieira HOFFMANN<sup>1</sup>, Kálita Cristina Moreira CARDOSO<sup>2</sup>, Antônio Sabino Neto da Costa ROCHA<sup>3</sup>, Aryanny Irene Domingues de OLIVEIRA<sup>4</sup>, Aluana Gonçalves ABREU<sup>1</sup>, Caio César de Oliveira PEREIRA<sup>4</sup>, Guilherme MALAFAIA<sup>4</sup>, Ivandilson Pessoa Pinto de MENEZES<sup>4\*</sup>

<sup>1</sup> Empresa Brasileira de Pesquisa Agropecuária, Embrapa Arroz e Feijão, Rodovia GO-462, km 12, CEP 75375-000, Santo Antônio de Goiás, Brasil

<sup>2</sup> Pontifícia Universidade Católica de Goiás, PUC Goiás, Av. Universitária, 1440, CEP 74605-010, Goiânia, Goiás, Brasil

<sup>3</sup> Empresa Brasileira de Pesquisa Agropecuária, Embrapa Amazônia Ocidental, Rodovia AM-10, CEP 69010-970, Manaus, Brasil

<sup>4</sup> Instituto Federal Goiano, Rodovia Geraldo Silva Nascimento, Km. 2,5, CEP75790-000, Urutaí, Goiás, Brasil

\* Corresponding author: [ivan.menezes@ifgoiano.edu.br](mailto:ivan.menezes@ifgoiano.edu.br)

## ABSTRACT

The Amazon Basin is a center of diversity of *Gossypium barbadense* and the strategy for conservation of this genetic resource depends on the knowledge of the diversity maintained in Amazonas State. During two expeditions, in 2012 and 2014, plants were collected in ten municipalities in the state of Amazonas, in the central Brazilian Amazon region. The molecular diversity was estimated by SSR markers for 50 samples collected in 2012. The morphological diversity of 24 plants collected in 2014 was assessed *ex situ* and compared to that of 50 plants of the same and other cotton varieties from other Brazilian states. Most of plants evaluated *in situ* in Amazonas had purple petioles and veins (82%), associated to medicinal use, and kidney seeds (78%). The *ex situ* morphological analyses showed that *G. barbadense* plants from the Amazonas state: i) presented higher similarity to cotton plants from other northern Brazilian states, and ii) were grouped separately from those of other northern Brazilian states by descriptor analysis. Both the molecular ( $H=0.41$ ) and morphological ( $H=0.38\pm 0.02$ ) diversity among the collected plants was considered intermediary. Our study indicates the distinctiveness of Amazon cottons, and contributes to demonstrate the discrimination power of multicategorical traits.

**KEYWORDS:** conservation, genetic resources, cotton tree, phenotypic characterization

## Diversidade genética de *Gossypium barbadense* coletado na Amazônia central

### RESUMO

A bacia Amazônica é um centro de diversidade de *Gossypium barbadense* e a estratégia de manutenção desse recurso genético depende do conhecimento da diversidade da espécie mantida no Estado do Amazonas. Em 2012 e 2014 plantas desta espécie foram coletadas em dez municípios na região centro-leste do Estado. A diversidade molecular por marcadores microssatélites foi mensurada para 50 amostras da coleta de 2012. A diversidade morfológica de 24 plantas coletadas em 2014 foi medida *ex situ* e comparada com a de 50 amostras desta e de outras variedades de algodão de outros estados do Brasil. A maioria das plantas do Amazonas apresentou folhas arroxeadas (82%), associadas a uso medicinal, e sementes unidas, do tipo rim-de-boi (78%). A análise morfológica *ex situ* mostrou que *G. barbadense* coletado no estado do Amazonas: i) tem maior similaridade com plantas da mesma espécie de outros estados da Região Norte do Brasil e ii) se agrupam entre si de forma diferenciada das plantas de outros estados. A diversidade molecular ( $H = 0,41$ ) e morfológica ( $H = 0,38 \pm 0,02$ ) entre os acessos foi considerada intermediária. Nosso estudo indica o caráter distintivo dos algodões amazônicos, e contribui para demonstrar o poder de discriminação de variáveis multicatóricas.

**PALAVRAS-CHAVE:** conservação, recursos genéticos, algodoeiro arbóreo, caracterização fenotípica

## INTRODUCTION

Cotton produces the most important natural fiber in the world. The commercially cropped cotton is almost entirely of the species *Gossypium hirsutum*, derived from a few cultivars of the species, consequently of narrow genetic base (Moiana *et al.* 2015). *Gossypium barbadense* is the second most cultivated species, known for the best fiber quality (Liu *et al.* 2015). The Amazon basin is a center of diversity for cotton, and comprises the “kidney” cotton, classified as variety *brasiliense*, whose seeds are fused to each other in the capsule. Both species are allotetraploid ( $2n = 52$ ).

*Gossypium barbadense* occurs in all Brazilian states (Almeida *et al.* 2009). Another domesticated arboreum cotton is *G. hirsutum* var. *marie galante* (Menezes *et al.* 2010), known as Mocó cotton, localized in the Northeastern semiarid. *Gossypium barbadense* has been domesticated in Peru and distributed in Brazil in the pre-Columbian era, while Mocó cotton derived from the Antilles and Central America (Menezes *et al.* 2010; Giband *et al.* 2010). For both species, seeds have been preserved from plants maintained in domestic gardens and backyards. Remaining plants of these cotton species are planted for many purposes and occur mainly in backyards or spontaneously (Albrana 2016). During the past 15 years, about 2000 plants of non-commercial cotton (*G. barbadense*, *G. hirsutum* var. *marie galante* and the wild *G. mustelinum*) were collected. Diversity studies showed that genetic differentiation is related to the geographic origin (Almeida *et al.* 2009; Barroso *et al.* 2010; Menezes *et al.* 2010; 2014a). The *in situ* maintenance in contrasting environments for several years may have favored genetic adaptation and regional variability.

The genetic variability is an important component to the success of breeding programs, allowing the selection of suitable parents for obtaining heterosis and recombination, constituting the source of different genes and traits. Thus, diversity and discrimination studies of a wide range of genetic variability provide an opportunity for developing cultivars better adapted to the soil and climate of the producing regions, contributing to the progress of breeding selection, agriculture and production.

Monitoring diversity of the genetic resources should be constantly performed and used to drive variability preservation and conservation measures, especially in the Brazilian Amazon basin, where an exclusion zone for transgenic cotton cultivation has been established in 2005.

For conservation or improvement, it is necessary to conduct morphological evaluation of materials. Sets of morphological descriptors with high power of discrimination are widely used (Selvi *et al.* 2013), but, depending on the genetic basis, may not be enough, requiring the additional use of molecular markers.

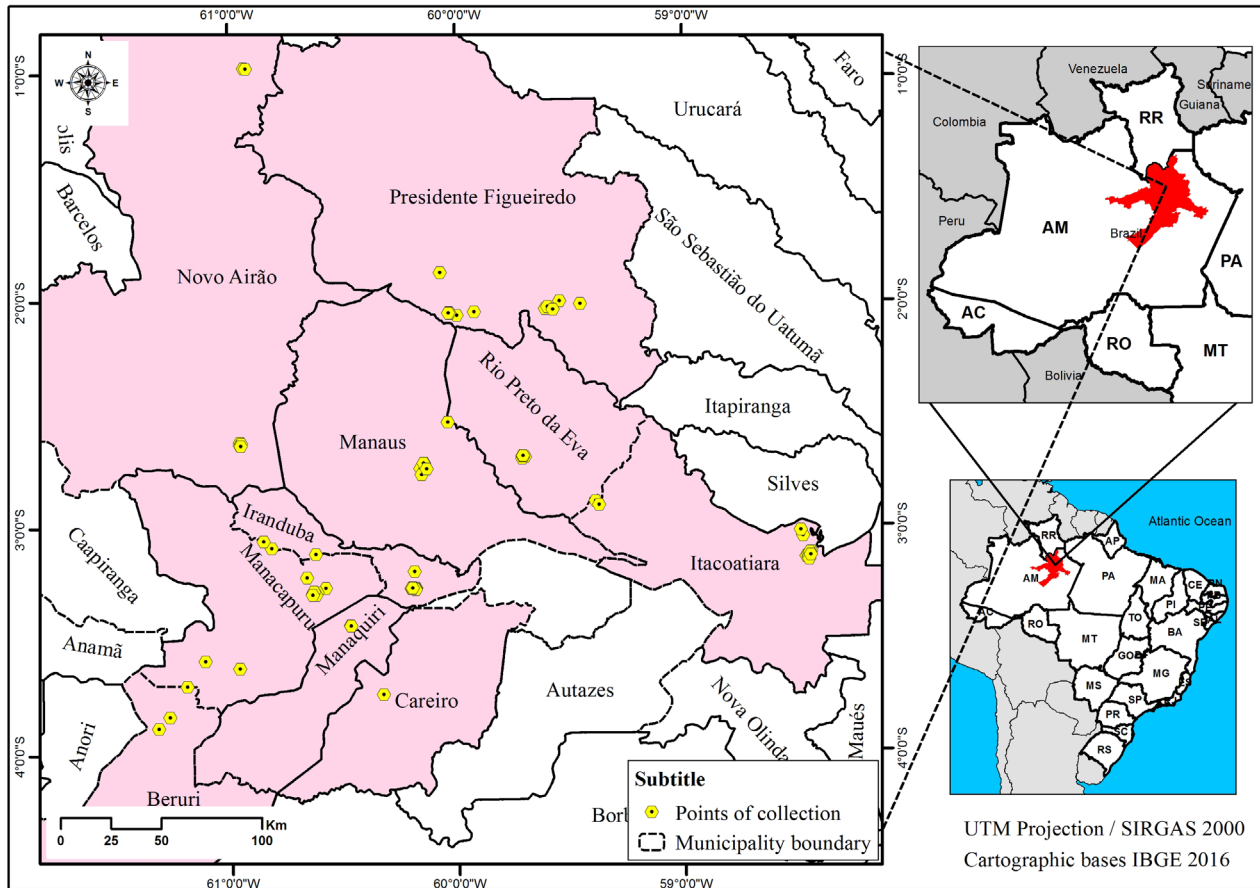
The hypothesis in this study was that the binary process of the data produced by phenotyping multicategorical traits can provide estimations with high discrimination power in the germplasm of *G. barbadense* from Amazonas, Brazil, and relate it to other cotton varieties from Brazil kept at the genebank of Embrapa, the Brazilian Agricultural Research Corporation. In this context, multicategorical traits produce different polymorphic binary markers, reducing the dissimilarity estimation error. This process, coupled with the use of morphological descriptors based on characters of easy observation and measurement, with low environmental impact and high heritability, is quite useful for diversity studies and discrimination of various species and cultivars. The specific aim of this study was to characterize and infer discrimination from cotton germplasm of different species belonging to Embrapa using multicategorical traits and SSR markers.

## MATERIAL AND METHODS

### Plant material

For simple sequence repeat (SSR) marker characterization, plants were collected, from 25 to 30 June 2012, in Presidente Figueiredo (11 plants), Manaus (five plants plus two bulks of plants of the same locality), Manacapuru (eight), Itacoatiara (nine), Rio Preto da Eva (six) and Iranduba (nine), all municipalities in Amazonas State, Brazil (Figure 1). The resulting 50 *G. barbadense* samples were analyzed for SSRs. Four *G. hirsutum* plants were included as controls for allele size: Delta Opal, LONREN, FM966 and M315.

*In situ*, plants were characterized for color of leaves, type of seeds, presence and color of fuzz, and presence or absence of petal spots. Seeds were collected for DNA extraction and *ex situ* preservation. A morphological *ex situ* assessment was carried out using 77 cotton plants from different species collected in different states in Brazil. The collected seeds were planted and raised in a greenhouse in Santo Antonio de Goiás (Goiás state, Brazil). Of those, 46 were of *G. barbadense* (27 from Amazonas state, six from Goiás state, two each from the states of Mato Grosso do Sul, Minas Gerais, Bahia and Pará, and one each from the states of Pernambuco, Roraima, Acre, São Paulo and Rio de Janeiro). Twenty-three accessions were from *G. hirsutum* r. *marie galante* (Mocó cotton) from 10 states (10 from Piauí, four from Paraíba, three each from Bahia and Alagoas, and one each from Sergipe, Maranhão and Rio Grande do Norte). Three accessions were of unknown origin and five were from exotic genotypes, *G. hirsutum* var. *punctatum*, a wild variety from Mexico (four plants) and *G. hirsutum latifolium* var. *australiana*, from Australia (one plant). The 27 accessions from Amazonas were collected in eight municipalities (six from Manacapuru, five from Itacoatiara, four from Novo Airão, three each from Iranduba and Rio Preto da Eva, and two each from Barcelos, Manaquiri and Presidente Figueiredo) (Figure 1). For Mocó cotton, plants with the lesser



**Figure 1.** Geographical distribution map of 84 *G. barbadense* plants evaluated *in situ* in the collection sites in central Brazilian Amazon, Brazil. This figure is in color in the electronic version.

number of seeds, possibly related to late flowering or poor productivity, were selected from the active genebank.

The morphological characterization was conducted at Embrapa, in Santo Antônio de Goiás, in greenhouse conditions (16°29'52.3"S, 49°16'53.8"W), in a completely randomized design with three replications. Plants were characterized using 23 multicategorical traits evaluated in the R4 stage (reproductive phase of the plant) (Table 1). These variables were chosen based on the standards used by the Ministry of Agriculture, Livestock and Food Supply - MAPA. Their multicategorical variations were transformed into binary data for analysis and assigned the number "1" for presence and "0" for absence of each variation in separate columns.

### DNA extraction and SSR genotyping

Genomic DNA extraction was performed from seeds obtained *in situ* using the DarT protocol (2012). Aliquots of each extracted genomic DNA were quantified using NanoDrop™. A set of 14 pairs of SSR primers marked with fluorochrome (6-FAM, NED and HEX) was selected for multiplex genotyping, with two pentaplex (CIR246, DC20027,

CIR316M, BNL3279, BNL3661 and CIR105, CIR249, JESPR153, BNL1551, BNL3482), and one tetraplex (CIR17, CIR55, CIR170, CIR373). The first pentaplex is composed of SSRs markers associated with resistance genes (Menezes *et al.* 2014b).

SSRs were amplified using Multiplex PCR Kit (Qiagen). PCR reactions, with a final volume of 5 uL, contained 1.0 ng of DNA, 2.5 uL of 2x Qiagen multiplex PCR Master mix (HotStar Taq DNA Polymerase, PCR amplification buffer, 3mM MgCl<sub>2</sub>), 0.5 of Q-solution, 0.2 μM of each pair primer (forward and reverse) and RNase free water. PCR was conducted in a thermocycler (GeneAmp Thermal Cycler 9700) with the following program: an initial denaturation at 95 °C for 15 minutes; followed by 34 cycles consisting of a denaturation step at 95 °C for 1 minute; one annealing at 55 °C or 51 °C for 1.5 minute; one extension at 72 °C for 1 minute, and 60 °C for 30-minute to final extension.

The amplified DNA fragments were separated by capillary electrophoresis in automatic fragments sequencer ABI model 3100 (Applied Biosystems). The allele size determination was carried out using GeneMapper software version 3.5 (Applied

**Table 1.** Number of multicategorical traits and their descriptors evaluated for morphological characterization of 77 cotton (*Gossypium* spp.) plants.

| N  | Multicategorical traits                | Description           |                     |                         |                        |           |
|----|--|-----------------------|---------------------|-------------------------|------------------------|-----------|
| 1  | Stem color                             | 1) Green              | 2) Purplish         | 3) Purple               |                        |           |
| 2  | Stem hairiness                         | 1) Glabrous           | 2) Little hairy     | 3) Hairy                | 4) Very hairy          |           |
| 3  | Number of lobes                        | 1) Three              | 2) Five             | 3) Seven                |                        |           |
| 4  | Leaf color                             | 1) Green              | 2) Purplish         | 3) Purple               |                        |           |
| 5  | Rib color                              | 1) Colorless          | 2) Green            | 3) Purplish             | 4) Purple              |           |
| 6  | Petiole color                          | 1) Colorless          | 2) Green            | 3) Purplish             | 4) Purple              |           |
| 7  | Nectary on leaves                      | 1) Absence            | 2) Mainrib          | 3) Main and lateral rib |                        |           |
| 8  | Leaf shape                             | 1) Palmate            | 2) Semi-digitate    | 3) Digitate             | 4) Lanceolate          |           |
| 9  | Leaf size                              | 1) Small              | 2) Medium           | 3) Big                  |                        |           |
| 10 | Leaf hairiness (upper surface)         | 1) Glabrous           | 2) Little hairy     | 3) Hairy                | 4) Very hairy          |           |
| 11 | Leaf hairiness (lower surface)         | 1) Glabrous           | 2) Little hairy     | 3) Hairy                | 4) Very hairy          |           |
| 12 | Bracteole dentation                    | 1) Less than 7        | 2) Between 7 and 12 | 3) More than 12         |                        |           |
| 13 | Bracteole length                       | 1) Small              | 2) Medium           | 3) Long                 |                        |           |
| 14 | Bracteole width                        | 1) Narrow             | 2) Normal           | 3) Large                |                        |           |
| 15 | Nectaries at the base of the bracteole | 1) Absence            | 2) Incipient        | 3) Present              |                        |           |
| 16 | Internal nectary of the bracteole      | 1) Absence            | 2) Incipient        | 3) Present              |                        |           |
| 17 | Corolla colour                         | 1) Creamy             | 2) Light yellow     | 3) Dark yellow          | 4) Yellow and red apex |           |
| 18 | Petal spot                             | 1) Absence            | 2) Incipient        | 3) Low                  | 4) Medium              | 5) Strong |
| 19 | Overlapping petals                     | 1) Little overlapping | 2) Overlapping      | 3) Very overlapping     |                        |           |
| 20 | Petal length                           | 1) Small              | 2) Medium           | 3) Long                 |                        |           |
| 21 | Stigma position in relation to anthers | 1) Below              | 2) Same height      | 3) Above                | 4) Far above           |           |
| 22 | Position of staminal column            | 1) Short              | 2) Medium           | 3) Long                 |                        |           |
| 23 | Pollen colour                          | 1) Creamy             | 2) Light yellow     | 3) Dark yellow          | 4) Orange              |           |

Biosystems). Then they were genotyped by assigning “1” for the presence and “0” for absence of the allele.

### Data Analysis

The intraspecific diversity was estimated using the number of polymorphic descriptors (morphological or SSR markers) (P) and the Shannon diversity index (H) calculated using the GenAlex 6.5 program (Peakall and Smouse 2012). To analyze the dissimilarity between the cotton plants, the dissimilarity matrix was calculated based on the complement of the Jaccard coefficient to the Neighbor-Joining cluster analysis by the DARwin 6.0.12 program (Perrier and Jacquemoud-Collet 2006). A Principal Coordinates Analysis (PCoA) was performed to produce the visualization of the relation between genotypes by a scale based on the dissimilarity estimated using the GenAlex 6.5 program. The dissimilarity matrix (from phenotypic and genotypic information) was also used to analysis of variance (Excoffier, et al. 1992) to verify the

partitioning of interspecific phenotypic variability between and within the groups of species.

Predictive SSRs marker analysis in *G. barbadense* was carried out by comparing the alleles of the SSR loci previously associated with resistance genes in *G. hirsutum*, using as resistance control genotypes of *G. hirsutum*: Delta Opal (for blue disease), FM 966 (for angular leaf spot), and Loren and M315 (for root knot nematodes).

## RESULTS

### *In situ* description of plants

In total, 84 plants were evaluated *in situ* in the collection sites in Amazonas (Figure 1). Most of the plants presented purplish leaves (64 out of 78 plants for which the leaf color was recorded during collection); kidney seeds (63 out of 81 plants) with no fuzz (31 out of 37 plants). The fuzz was usually brown, and a single plant presented a green fuzz, which was irregularly distributed over the seed. Flowers mostly presented



petals without spots (26 out of 31 observed plants). Most plant owners, 82 out of 84 interviewed people, declared that the plant had medicinal use. Plants were also associated with an ornamental role for 19 out of the 84 interviewees, yet only two owners declared the plant had ornamental use only. The distance between collected plants ranged from 0.2 to 323.7 km. Usually, there was a single individual of a species in the backyard, or less frequently, two individuals.

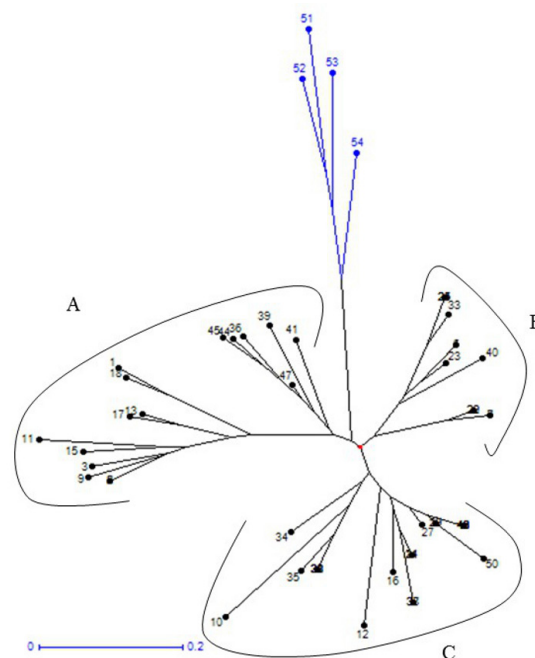
### SSR loci analysis

The evaluated markers were 14 microsatellite loci, of which four (CIR17, CIR55, CIR170, CIR 373) were considered monomorphic for all *G. barbadense* and *G. hirsutum* accessions. The others were genotyped in binary scores, “1” for presence and “0” for absence of each allele, thus forming 40 dominant genetic markers, 67.5% polymorphic in the Amazonas germplasm and 40% in *G. hirsutum* accessions. We found that among these 40 markers, 19 were exclusive to Amazonas, and ten were present only in *G. hirsutum*. The overall diversity index  $H$  was 0.43 ( $\pm 0.02$ ). The value of the average genetic dissimilarity among accessions was 0.41, ranging from 0 to 0.76, indicating intermediate average divergence among accesses. In the clustering analysis, there were three main groups, A (with 22 accessions), B (16 accessions) and C (12 accessions) (Figure 2), which were genetically different ( $\Phi_{ST} = 0.310$ ,  $p$  value = 0.001). In Figure 2, we observe the similarity of the group A to the pool of *G. hirsutum* commercial cultivars.

Among the markers, five loci are reported to be linked to disease resistance genes. DC20027 is linked to the *Rghv1* gene that confers resistance to cotton blue disease (Fang *et al.* 2010); CIR246, to the *B12* gene, which confers resistance to angular leaf spot (Xiao *et al.* 2010); BNL3279, to one locus conferring resistance to reniform nematode (Dighe *et al.* 2009); and CIR316M (Shen *et al.* 2006) and BNL3661 (Gutiérrez *et al.* 2010), which are associated with two different loci that control root-knot nematode tolerance. The allele reported in *G. hirsutum* linked to resistance to root-knot nematode, BNL3661-185 bp, is present in 40% of the *G. barbadense* Amazon accessions (Table 2). The allele DC20027-202 bp, linked in *G. hirsutum* to virus resistance, was present in only one accession, while the allele DC20027-206 bp, not yet reported for this locus, was present in 6 of the evaluated accessions.

### Ex situ morphological descriptor analysis

For *ex situ* evaluation and seed multiplication, the plants with purple leaf parts or purple bracteoles were chosen, because it is characteristic for the plants of this region, hence a priority for multiplication. Therefore there was no polymorphism for this variable, neither for leaf shape. The other 21 multicategorical traits evaluated were polymorphic. There were missing categories among polymorphic variables, which is expected since not all the variability of the genus is



**Figure 2.** Neighbor-joining tree with *G. barbadense* (black line) and *G. hirsutum* (blue line) accessions using SSRs and the Jaccard dissimilarity. This figure is in color in the electronic version.

**Table 2.** Allelic composition of five microsatellite loci linked to disease resistance genes in *Gossypium barbadense*, and commercial varieties of *G. hirsutum* as a control. Alleles in bold are those linked to the resistance gene.

| Locus   | Allele1 | <i>G. barbadense</i><br>(n = 50) | <i>G. hirsutum</i> |       |       |      |
|---------|---------|----------------------------------|--------------------|-------|-------|------|
|         |         |                                  | Delta              | FM966 | Loren | M315 |
| CIR246  | 146     | 0.00                             | 1.00               | 1.00  | 0.00  | 0.00 |
|         | 156     | 1.00                             | 0.00               | 0.00  | 1.00  | 1.00 |
| DC20027 | 200     | 0.86                             | 0.00               | 1.00  | 1.00  | 1.00 |
|         | 202     | 0.02                             | 1.00               | 0.00  | 0.00  | 0.00 |
|         | 206     | 0.12                             | 0.00               | 0.00  | 0.00  | 0.00 |
| CIR316M | 192     | 0.50                             | 0.00               | 0.00  | 0.00  | 0.00 |
|         | 198     | 0.00                             | 1.00               | 1.00  | 1.00  | 0.00 |
|         | 201     | 0.00                             | 1.00               | 1.00  | 1.00  | 1.00 |
|         | 204     | 0.50                             | 0.00               | 0.00  | 0.00  | 0.00 |
| BNL3279 | 210     | 0.00                             | 0.00               | 0.00  | 0.00  | 1.00 |
|         | 106     | 0.37                             | 0.00               | 0.00  | 0.00  | 0.00 |
|         | 114     | 0.49                             | 1.00               | 1.00  | 0.00  | 1.00 |
|         | 116     | 0.00                             | 0.00               | 0.00  | 1.00  | 0.00 |
|         | 124     | 0.03                             | 1.00               | 1.00  | 1.00  | 1.00 |
| BNL3661 | 130     | 0.10                             | 0.00               | 0.00  | 0.00  | 0.00 |
|         | 144     | 0.01                             | 0.00               | 0.00  | 0.00  | 0.00 |
|         | 181     | 0.10                             | 0.00               | 0.00  | 0.00  | 0.00 |
|         | 185     | 0.40                             | 0.00               | 0.00  | 0.00  | 0.50 |
| BNL3661 | 189     | 0.50                             | 0.00               | 0.00  | 0.00  | 0.50 |
|         | 191     | 0.00                             | 0.50               | 0.50  | 0.50  | 0.00 |
|         | 193     | 0.00                             | 0.50               | 0.50  | 0.50  | 0.00 |

in this sample. The stem hairiness, number of lobes, color of the ribs, nectary on leaves, nectary at the base of the bracteole, internal nectary of the bracteole and petal spot variables, though being polymorphic, showed no genotypes classified as “very hairy”, “seven”, “purple”, “absent”, “incipient”, “incipient” and “incipient”, respectively. On binary analysis of multicategorical traits, it was possible to obtain 63 polymorphic binaries variables of 80 possibles (Table 1), which were used to establish the phenotypic diversity and dissimilarity of the studied germplasm.

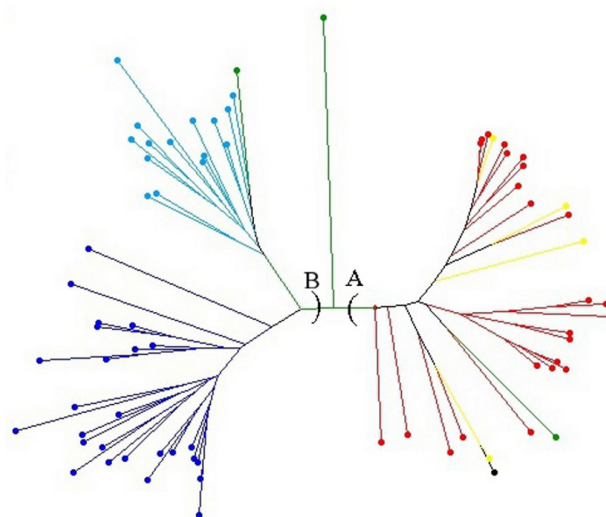
Considering all the genotypes used for morphological analysis, the total diversity index (H) was 0.40 (± 0.03). The average H of the groups was 0.26 (± 0.02) ranging from 0.15 (± 0.03) in the wild *G. hirsutum* species to 0.35 (± 0.03) for the *G. barbadense* species (Table 3).

The total diversity data calculated corroborate the expressive average dissimilarity between the 77 cotton plants, equal to 0.61 (± 0.02). The wild *G. hirsutum* and Mocó cotton showed the lowest values, 0.37 (± 0.05) and 0.40 (± 0.03), respectively. The highest values were observed for *G. barbadense* species (0.52 ± 0.03), followed by the non-classified genotypes (0.69 ± 0.07). Considering the specific groups of cotton, there was a high dissimilarity between them, with a mean value of 0.62 (± 0.06), the highest value between the unknown and *G. barbadense* groups (0.74) and the lowest between the Australian and wild groups (0.38) (Table 4).

The interspecific phenotypic variability between groups of species was high and significant (0.40, *p*-value < 0.01) when

estimated by analysis of variance. The largest portion of this variation was intragroup, with 61%, compared to intergroup, with 39% of the analyzed variability. Considering only the *G. barbadense* from Amazonas, the average dissimilarity value was intermediate (0.38 ± 0.02), ranging from 0 to 0.61. The dissimilarity between *G. barbadense* from Amazonas to the others was high, with an average value of 0.74. The greatest distance was to the Australian variety of *G. hirsutum* (0.81), followed by wild *G. hirsutum* (0.77), non-classified genotypes (0.76), Mocó cotton (0.75) and the other *G. barbadense* (0.63).

The 77 cotton plants analyzed through the 21 multicategorical traits were separated by the neighbor-joining tree into two groups (A and B) and an intermediate unit (Figure 3). The A group indicated the background of *G. hirsutum*, composed by the Mocó cotton accesses, wild *G. hirsutum*, the Australian variety, *G. hirsutum latifolium*, and one unknown access. The B group was formed almost



**Figure 3.** Neighbor-joining tree of 77 cotton plants of the genus *Gossypium* based on Jaccard dissimilarity using 21 morphological descriptors. Each colored line represents a different cotton plant indicating the studied groups: blue - *G. barbadense* (dark blue only access from Amazonas); red - Mocó cotton; green - unknown; black - *G. hirsutum* var. *australiana*; yellow - *G. hirsutum* var. *punctatum*. This figure is in color in the electronic version.

**Table 3.** Description of germplasm diversity of 77 cotton plants of different autotetraploid species using 21 multicategorical traits. P (%) - polymorphism percentage; N - sample size; H - Shannon diversity index, SE - standard errors.

| Species   | N  | P (%) | H (SE)        |
|---|----|-------|---------------|
| <i>G. barbadense</i>                                  | 46 | 93.7  | 0.35 (± 0,03) |
| <i>G. hirsutum</i> r. <i>marie galante</i>            | 23 | 57.1  | 0.23 (± 0,03) |
| <i>G. hirsutum</i> var. <i>punctatum</i>              | 4  | 28.6  | 0.15 (± 0,03) |
| Unknown cotton  | 3  | 54    | 0.29 (± 0,04) |
| <i>G. hirsutum latifolium</i> var. <i>australiana</i> | 1  | -     | -             |
| Total   | 77 | 100   | 0.40 (± 0,03) |

**Table 4.** Jaccard dissimilarity matrix obtained from pair-wise comparison of 21 multicategorical traits in a group of cotton species and varieties (*Gossypium* spp).

|   | <i>G. barbadense</i> | <i>G. hirsutum</i> r. <i>marie galante</i> | Unknown cotton | <i>G. hirsutum latifolium</i> var. <i>australiana</i> | <i>G. hirsutum</i> var. <i>punctatum</i> |
|---|----------------------|--|----------------|---|--|
| <i>G. barbadense</i>                                  | 0.00                 |  |                |   |  |
| <i>G. hirsutum</i> r. <i>marie galante</i>            | 0.71                 | 0.00                                       |                |   |  |
| Unknown cotton  | 0.74                 | 0.64                                       | 0.00           |   |  |
| <i>G. hirsutum latifolium</i> var. <i>australiana</i> | 0.72                 | 0.50                                       | 0.69           | 0.00  |  |
| <i>G. hirsutum</i> var. <i>punctatum</i>              | 0.73                 | 0.43                                       | 0.66           | 0.38  | 0.00                                     |

exclusively by *G. barbadense* plants and included another non-previously classified cotton. Furthermore, in B group there was a clear differentiation by region, in which the northern access of Brazil formed a single subgroup (27 accesses from Amazonas, in dark blue, one from Acre, one from Pará and another from Roraima); and another subgroup formed by the other *G. barbadense* cottons studied. Finally, one of the unknown accesses grouped between the A and B groups.

The principal coordinates analysis (PCoA) (Figure 4), corroborated the neighbor-joining cluster analysis, confirming the separation of *G. hirsutum* and *G. barbadense*, and demonstrated the distinction between the cotton plants from the state of Amazonas and the other states of Brazil. Furthermore, PCoA indicated the proximity of two of the non-classified accesses to *G. hirsutum* and another one to *G. barbadense*.

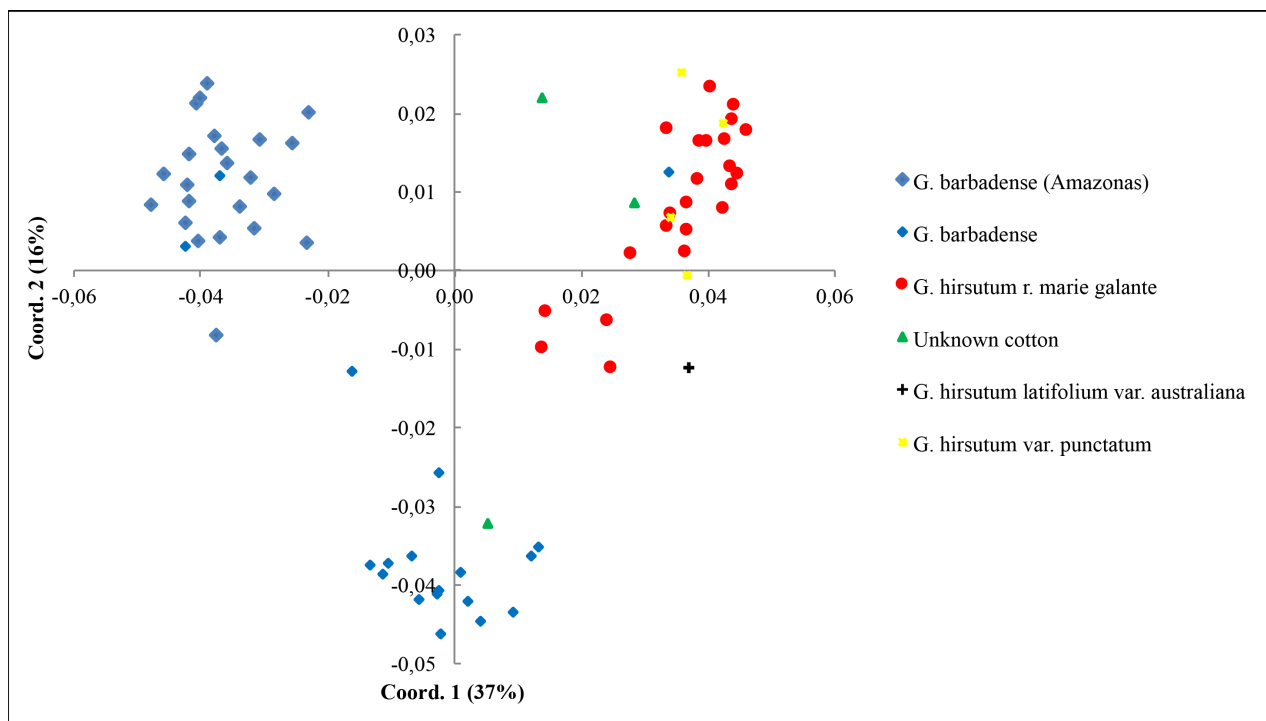
## DISCUSSION

*Gossypium barbadense* from Amazonas is a very different gene pool when compared to commercial cultivars, demonstrated both by DNA markers and morphological traits, by the estimates of genetic differentiation and cluster analysis. The set of markers used in *G. barbadense* germplasm revealed a slightly higher genetic diversity than that reported in the states of Amapá and Pará (Almeida *et al.* 2009). When compared to Moco cotton, also found as arborescent individuals in

backyards, the diversity was lower (Menezes *et al.* 2010), but higher than the values estimated for cultivars (Moiana *et al.* 2015; Pereira *et al.* 2015).

The exploitation of *G. barbadense* as genetic resource is indicated by the presence of the allele of the microsatellite marker CIR 3661, reported to be linked to the *Meloidogyne incognita* resistance gene, which was present in 30 out of the 38 analyzed plants. Furthermore, one plant exhibited the allele reported to be linked to the blue disease resistance gene in loco DC20027. None presented alleles associated with resistance to bacterial blight or any other nematode resistance linked loci, although sources of resistance different from those described in the literature may be present.

The purple color of the leaf parts and bracteoles is expressive in North Brazil. Plants with purple leaves comprise in average half of the observed *G. barbadense* found in Acre, Pará and Roraima (Albrana 2016). In Amapá it was present, but not frequent (8 out of 115 plants) and in Maranhão, in Northeast Brazil, it was more frequent (five out of twelve observed plants). Only one plant with purple leaves was collected in Bahia, Mato Grosso do Sul, Tocantins and Piauí. In Amazonas it has been associated with medicinal properties such as treating uterine hemorrhaging (Ososki *et al.* 2002) and wounds, corroborating the antimicrobial properties of *G. arboretum* (Annan and Houghton 2008) and *G. hirsutum* (Omojasola and Awe 2004). Kidney seeds are typical of east



**Figure 4.** Principal Coordinates Analysis (PCoA) using 77 cotton plants of different autotetraploid species based on Jaccard dissimilarity using 21 morphological descriptors. This figure is in color in the electronic version.

Andes (Percy and Wendel 1990) and are more frequent than separated seeds in the North and Central Brazilian regions (Albrana 2016).

The multicategoric variables showed expressive variability, with 91.3% polymorphism. Various multicategorical traits presented discrepant distribution, with some morphological traits much more frequent than others, contributing to reduce diversity estimates. Nevertheless, we observed total diversity indices close to those calculated by DNA markers for different species of cotton plants in Brazil (Almeida *et al.* 2009; Menezes *et al.* 2010; Moiana *et al.* 2015).

Vidal Neto *et al.* (2008) also found a significant variability for morphological variables in permanent *Gossypium* accesses from the Active Germplasm Bank of Embrapa, highlighting the potential use as a complementary tool in breeding programs. Some studies have shown that utility, demonstrating associations of this type of single heritage (discrete variables) with disease and pest resistance behavior and fiber characteristics (Juhász *et al.* 2013; Neto Vidal *et al.* 2005).

The significant dissimilarity shown by both morphological and molecular markers demonstrates the importance of maintaining both *in situ* and *ex situ* conservation of *Gossypium barbadense*, to preserve Amazon plants, which are distinguished from other plants, even from other states of the North Region (Roraima, Acre and Pará). However, this organization of the cotton germplasm studied was not detected in the joint analysis of the original data, considering the quantitative, multicatameric and molecular variables (codominant) using Gower distance (unpublished results).

Our analysis allowed to classify individuals into well-defined groups, separating the cotton according to species and regional variation, such as the *G. barbadense* cotton from the North Region and other Brazilian states. In another analysis of this germplasm, based on the Euclidean distance, no regional structure was observed (Cardoso *et al.* 2015). Therefore, the binary approach to multicategoric variables was important for the discrimination of genotypes, indicating the greater power of distinction for conservation studies, improvement, and protection of cotton varieties.

## CONCLUSIONS

*Gossypium barbadense* plants currently found in the central region of Amazonas state (Brazil) were clearly differentiated from others in Brazil, as they formed a well-defined group when analyzed with the use of multicategorical traits and SSR markers. Morphologically, a high proportion of plants presented purple color of the leaf veins, petiole and bracteoles, and kidney seeds with no fuzz. The proportion of effective use as a medicinal plant is also remarkable. Although collections covered only a small proportion of the Amazonas state, a relatively high diversity was reported. Monitoring

and exploring an extended area is recommended. *Ex situ* maintenance has been initiated.

## ACKNOWLEDGEMENTS

We thank Empresa Brasileira de Pesquisa Agropecuária (Embrapa) for financial support, Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG) for providing a scholarship, and Paulo Henrique G. Marques for his assistance in the production of the geographical distribution map. The study was conducted under authorization 13973 of the Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA).

## REFERENCES

- Albrana, 2016. Brazilian Native and Naturalized Cotton [online]. (<https://www.cnpa.embrapa.br/albrana/english/index.html>). Accessed on 10/04/2016.
- Almeida, V.C.; Hoffmann, L.V.; Yokomizo, G.K.; Costa, J.N.; Giband, M.; Barroso, P.A.V. 2009. *In situ* and genetic characterization of *Gossypium barbadense* L. from the States of Pará and Amapá, Brazil. *Pesquisa Agropecuária Brasileira*, 44: 719-725.
- Annan, K.; Houghton, P.J. 2008. Antibacterial, antioxidant and fibroblast growth stimulation of aqueous extracts of *Ficus asperifolia* Miq. and *Gossypium arboreum* L., wound-healing plants of Ghana. *Journal of Ethnopharmacology*, 119: 141-144.
- Barroso, P.A.V.; Hoffmann, L.V.; Batista, C.E.; Freitas, R.B.; Alves, M.F.; Silva, U.C.; Andrade, F.P. 2010. *In situ* conservation and genetic diversity of three populations of *Gossypium mustelinum* Miers (ex Watt). *Genetic Resources and Crop Evolution*, 57: 343-349.
- Cardoso, K.C.M.; Abreu, A.G.; Cardoso, R.M.; Rocha, A.S.N.C.; Hoffmann, L.V. 2016. Morfologia de algodoeiros de diferentes estados do Brasil (<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/132356/1/p74.pdf>). Accessed on 18/07/2017.
- Giband, M.; Dessauw, D.; Barroso, P.A.V. 2010. Cotton: Taxonomy, origin and domestication. In: Wakelyn, P.J.; Chaudhry, M.R. (Ed.). *Cotton: Technology for the 21<sup>st</sup> century*. International cotton Advisory committee, Washington, DC, p.3-15.
- DArT. 2012. Protocolo de extração de DNA de plantas. ([http://www.DiversityArrays.com/pub/DArT\\_DNA\\_isolation.pdf](http://www.DiversityArrays.com/pub/DArT_DNA_isolation.pdf)). Accessed on 10/04/2012.
- Dighe, N.D.; Robinson, A.F.; Bell, A.A.; Menz, M.A.; Cantrell, R.G.; Stelly, D.M. 2009. Linkage Mapping of Resistance to Reniform Nematode in Cotton following Introgression from *Gossypium longicalyx* (Hutch. & Lee). *Crop Science*, 49: 1151-1164.
- Excoffier, L.; Smouse, P.E.; Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131: 479-491.
- Fang, D.D.; Xiao, J.; Canci, P.C.; Cantrell, R.G. 2010. A new SNP haplotype associated with blue disease resistance gene in cotton (*Gossypium hirsutum* L.). *Theoretical and Applied Genetics*, 120: 943-953.



- Gutiérrez, O.A.; Jenkins, J.N.; McCarty, J.C.; Wubben, M.J.; Hayesn, R.W.; Callahan, F.E. 2010. SSR markers closely associated with genes for resistance to root-knot nematode on chromosomes 11 and 14 of Upland cotton. *Theoretical and Applied Genetics*, 121: 1323-37.
- Juhasz, A.C.P.; Condé, A.B.T.; Rabelo, H. O.; Pimenta, S.; Soares, B.O. 2013. Variabilidade genética de populações oriundas do cruzamento de algodão branco e colorido para melhoria de qualidade de fibra no norte de Minas Gerais. *Revista Agrarian*, 6: 368-375.
- Khaleequr, R.; Arhiya, S.; Shafeequr, R. 2012. *Gossypium herbaceum* Linn: An Ethnopharmacological Review. *Journal of Pharmaceutical and Scientific Innovation*, 1: 1-5.
- Liu, X.; Zhao, B.; Zheng, Hua-J.; Hu, Y.; Lu, G.; Yang, Chang-Q.; *et al.* 2015. *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. *Scientific Reports*, 5: 1-14.
- Menezes, I.P.P.; Barroso, P.A.V.; Hoffmann, L.V.; Lucena, V.S.; Giband, M. 2010. Genetic diversity of mocó cotton (*Gossypium hirsutum* race *marie-galante*) from the northeast of Brazil: implications for conservation. *Botany*, 88: 1-9.
- Menezes, I.P.P.; Gaiotto, F.A.; Hoffmann, L.V.; Ciampi, A.Y.; Barroso, P.A.V. 2014. Genetic diversity and structure of natural populations of *Gossypium mustelinum*, a wild relative of cotton, in the basin of the De Contas River in Bahia, Brazil. *Genetica*, 142: 99-108.
- Menezes, I.P.P.; Gaiotto, F.A.; Suassuna, N.D.; Hoffmann, L.V.; Barroso, P.A.V. 2014. Susceptibility of *Gossypium mustelinum* populations to the main cotton diseases in Brazil. *Journal of Agricultural Science*, 6: 39-47.
- Moiana, L.D.; Vidigal Filho, P.S.; Gonçalves-Vidigal, M.C.; Carvalho, L.P. 2015. Genetic diversity and population structure of upland cotton Brazilian cultivars (*Gossypium hirsutum* L. race *latifolium* H.) using SSR markers. *Australian Journal of Crop Science*, 9: 143-152.
- Osocki, A.L.; Lohr, P.; Reiff, M.; Balick, M.J.; Kronenberg, F.; Fugh-Berman, A.; O'Connor, B. 2002. Ethnobotanical literature survey of medicinal plants in the Dominican Republic used for women's health conditions. *Journal of Ethnopharmacology*, 79: 285-298.
- Omojasola, P.F.; Awe, S. 2004. The antibacterial activity of the leaf extracts of *Anacardium occidentale* and *Gossypium hirsutum* against some selected microorganisms. *Bioscience Research Communications*, 16: 25-28.
- Percy, R.G.; Wendel, J.F. 1990. Allozyme evidence for origin and diversification of *Gossypium barbadense* L. *Theoretical and Applied Genetics*, 79: 529-42.
- Peakall, R.; Smouse, P.E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, 28: 2537-2539.
- Perrier, X.; Jacquemoud-Collet, J.P. 2006. DARwin software. (<http://darwin.cirad.fr/>). Accessed on 21/01/2016.
- Selvi, D.T.; Srimathi, P.; Senthil, N.; Ganesan, K.N. 2013. Distinctness, uniformity and stability (DUS) characterization on phenological traits and assessing the diversity of inbreds in maize (*Zea mays* L.). *African Journal of Agricultural Research*, 8: 6086-6092.
- Shen, X.; Becelaere, G.V.; Kumar, P.; Davis, R.F.; May, O.L.; Chee, P. 2006. 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. *Theoretical and Applied Genetics*, 113: 1539-1549.
- Vidal Neto, F.C.; Silva, F.P.; Isidro, R.; Araújo, G.P.; Costa, J.N.; Barroso, P.A.V.; Andrade, F.P. 2008. *Caracterização de Acessos de Gossypium Pertencentes ao Banco de Germoplasma da Embrapa Algodão*. Comunicado Técnico 362, Embrapa Algodão, Campina Grande, 9p.
- Vidal Neto, F.C.; Silva, F.P.; Bleicher, E.; Melo, F.I.O. 2005. Mutantes morfológicos de algodoeiro herbáceo como fonte de resistência ao bicudo. *Pesquisa Agropecuária Brasileira*, 40: 123-128.
- Xiao, J.; Fang, D.D.; Bhatti, M.; Hendrix, B.; Cantrell, R. 2010. A SNP haplotype associated with a gene resistant to *Xanthomonas axonopodis* pv. *malvacearum* in upland cotton (*Gossypium hirsutum* L.). *Molecular Breeding*, 25: 593-602.

RECEIVED: 22/12/2016

ACCEPTED: 30/09/2017

ASSOCIATE EDITOR: Alexandre Pio Viana



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.