

Diversity and genetic structure among subpopulations of *Gossypium mustelinum* (Malvaceae)

M.F. Alves¹, P.A.V. Barroso², A.Y. Ciampi³, L.V. Hoffmann²,
V.C.R. Azevedo³ and U. Cavalcante⁴

¹Departamento de Melhoramento de Plantas, Universidade Federal de Goiás, Goiânia, GO, Brasil

²Embrapa Cotton, Brazilian Agricultural Research Corporation, Núcleo do Cerrado, Goiânia, GO, Brasil

³Embrapa Cenargen, Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, Brasil

⁴Departamento de Genética e Morfologia, Universidade de Brasília, Brasília, DF, Brasil

Corresponding author: P.A.V. Barroso
E-mail: paulo.barroso@embrapa.br

Genet. Mol. Res. 12 (1): 597-609 (2013)

Received September 25, 2012

Accepted January 15, 2012

Published February 27, 2013

DOI <http://dx.doi.org/10.4238/2013.February.27.9>

ABSTRACT. *Gossypium mustelinum* is the only cotton species native to Brazil; it is endemic to the semi-arid region of the northeast. The populations are found near perennial and semi-perennial sources of water, such as ponds or pools in intermittent streams. Problems with *in situ* conservation derive from human interference in its habitat, mainly because of excessive cattle grazing and deforestation. Establishing efficient strategies for *in situ* conservation requires knowledge of the genetic structure of the populations. We evaluated the structure and genetic variability of populations of *G. mustelinum* in the Tocó

and Capivara Rivers (State of Bahia). Two hundred and eighteen mature *G. mustelinum* plants were genotyped with SSR markers. The molecular data were used to estimate the allelic frequencies, the heterozygosity, the F statistics, and the genetic distance among the populations and among individuals. We found high genetic diversity among the populations. The F_{ST} indexes for each population were also high and strongly correlated with physical distance. The high estimated level of endogamy and the low observed heterozygosity are indicative that the populations reproduce mainly by self-fertilization and crosses between related individuals. Consequently, strategies for *in situ* preservation should include at least three occurrence sites of *G. mustelinum* from each population. For *ex situ* conservation, the collections should include as many sites as possible.

Key words: Conservation; Genetic structure; *Gossypium mustelinum*; Microsatellites; Wild cotton

INTRODUCTION

Domestication and modern plant breeding have progressively adapted cultivated species for human needs. However, the process of continuous selection associated with limited allele introduction has presumably resulted in a decrease in genetic diversity (Fu and Somers, 2009). In many cases, the narrowed diversity may jeopardize future genetic yield improvements and the quality of agricultural products (Reif et al., 2005). A main potential source of variability is found in wild-related species that are sexually compatible with cultivated species. The history of the two most important cultivated species of cotton, *Gossypium hirsutum* L. and *G. barbadense* L., is similar to that of other crops. Cotton breeding has imposed a limitation in the genetic base of cultivated forms (Wendel and Brubaker, 1993; Bertini et al., 2006; McCarty et al., 2007), and other *Gossypium* allotetraploid species are an important source of new alleles (Iqbal et al., 2001), including *G. mustelinum* Miers.

G. mustelinum is endemic in the semi-arid region of northeastern Brazil and is sexually compatible with cultivated cotton (Stephens and Phillips, 1972; Freire, 2002). It is a wild species that occurs spontaneously in small populations in the Caatinga biome (Pickersgill et al., 1975; Freire et al., 1990). The Caatinga is found only in Brazil and is characterized by small, thorny, deciduous, and xerophytic trees and permeated by annual grasses, cacti, and bromelian species. Almost all superficial water sources in this environment are intermittent, and the populations of *G. mustelinum* occur near relatively durable sources of water, such as wetlands or riparian vegetation (Barroso et al., 2010). The species has never been bred or commercially exploited (Borém et al., 2003), and there is no report of any kind of domestic use by the people who live in localities surrounding the plant populations. Probably, this species is the less known among the other allotetraploids from the same genus (Wendel et al., 1994), and it may contribute to the improvement of important agronomic traits, such as lint quality (Freire, 2002).

In the beginning of 2006, only four natural populations of *G. mustelinum* were known. One was located in the municipality of Caicó, in the State of Rio Grande do Norte, one in Jaguarari and two in Macururé, both municipalities in the State of Bahia (Neves et al., 1965; Pickersgill et al., 1975; Freire et al., 1990; Barroso et al., 2010). Intense environmental damage has been noticed and some populations have become extinct (Barroso et al., 2010). At the end of 2006, new *G. mustelinum* populations were found in Bahia. The new occurrence sites are organized in subpopulations with different individual numbers along margins of streams. The subpopulations are separated from each other by distances of less than 1 km to several kilometers.

Very high genetic differences were observed between previously known populations of *G. mustelinum*, and the strategy of *in situ* and *ex situ* conservation had to include all the populations known at that time (Barroso et al., 2010). If the same is true for the recently discovered population, the existence of several subpopulations within each one increases the complexity of the conservation task. In view of these facts, it is almost impossible to perform *in situ* conservation of all populations and it is even more difficult to preserve all the subpopulations of each population. Therefore, knowledge of the genetic structure within populations is essential to define an achievable strategy for conservation.

Herein, we describe the conditions of two of these recently discovered *G. mustelinum* populations, with different degrees of human interference with the natural environment. We also show how these populations composed of multiple subpopulations are genetically structured.

MATERIAL AND METHODS

Population sites, collection of samples and DNA extraction

Leaves and petals were sampled from two natural populations of *G. mustelinum* found along the margins of the Capivara and Tocó Rivers, both affluents of the Paraguaçu River in the State of Bahia (Figure 1). Six subpopulations distributed from the headwaters to the mouth of each river were chosen to be sampled and named CP or TC, respectively, for the Capivara and Tocó Rivers. The sampling sites were about 3 to 12 km from each other. At least 20 plants were sampled at each site, except when fewer individuals were present. At Capivara River, about 35 km long, the samples were collected from 13 km after the headwaters (CP1 community) to 1.3 km from the formation of Juracy Magalhães Dam (CP4 community). In addition, two plant subpopulations, called RB1 and RB2, were collected at Ribeirão do Brejo, a Capivara River tributary from the west bank, which joins into the Capivara River at the location of the CP3 community. The Tocó River is about 64 km long, and plants were collected from 12 km after the river source (TC1) until approximately 1.5 km from the river mouth (TC6). The samples were dried on silica gel and stored at 15°C. The extraction of total DNA was performed according to the CTAB method described by Doyle and Doyle (1990). DNA was quantified by comparison with known standard DNA concentrations (λ DNA) on 1.0% (w/v) agarose gels using Sybr Green staining.



Figure 1. Localization of the Tocó (square) and Capivara (circle) populations of *Gossypium mustelinum*.

Microsatellite analysis

The genetic variability present in *G. mustelinum* populations was analyzed using 14 microsatellite markers (Liu et al., 2000; Nguyen et al., 2004). These markers were amplified in five multiplex PCRs (Table 1), using the Qiagen Multiplex PCR kit (Qiagen Inc., Valencia, CA, USA), according to the manufacturer protocol. For each locus, the primers were fluorescently labeled (Integrated DNA Technologies/IDT for HEX- and 6-FAM-labeled primers; Table 1). PCR was performed with an initial 15-min denaturation at 95°C, followed by 30 cycles of 60 s at 94°C, 60 s at the annealing temperature (which varied from 51° to 57°C, depending on the primer pairs), and 90 s at 72°C. A final extension at 72°C for 30 min was carried out to minimize the formation of nonspecific peaks. The PCR Multiplex products were separated on an ABI Prism model 3700 automatic sequencer (Applied Biosystems, Foster City, CA, USA). The detection and estimation of allelic size in base pairs were performed using the GeneScan software (Applied Biosystems 3.7), and the Genotyper software (Applied Biosystems 3.7) was used for peak filtering and final compilations. One plant of each of the three other cotton types present in Brazil, *G. barbadense*, *G. hirsutum* race Marie-Galante (Mocó), and upland cotton were included as controls of allele sizes.

Table 1. PCR multiplex conditions and characteristics of microsatellite loci in *Gossypium mustelinum* populations.

Multiplex	Ta (°C)	Microsatellite primer	Fluorophore	No. of alleles	Size range (bp)	Chromosome
1	57	BNL252	6-FAM	6	175-192	8
		BNL1421	HEX	8	193-235	18
		CIR311	6-FAM	4	178-187	15
2	51	CIR381*	HEX	4	246-254	2
				2	261-265	14
		CIR383	HEX	6	124-142	23
		BNL1551	6-FAM	7	175-192	16
3	55	CIR148	6-FAM	5	143-147	12
		CIR203	HEX	3	240-254	6
		BNL169*	HEX	3	194-198	20
				8	207-218	-
4	57	BNL840	HEX	3	149-157	26
		BNL3599	6-FAM	3	189-193	12
		BNL686*	6-FAM	5	141-151	9
				5	155-181	23
5	56	BNL1350	6-FAM	9	214-237	1
		BNL3649	HEX	4	191-207	A03

Ta = annealing temperature. *Primers that amplified two loci.

Data analysis

For each of the 12 subpopulations, the average number of alleles per locus, the frequency of alleles exclusive of one of the populations, the observed and expected heterozygosity, and the F -statistics (Weir and Cockerham, 1984) were estimated. Bootstrap confidence intervals (95%) were obtained using 10,000 resamples. The fixation index F_{ST} (Weir and Cockerham, 1984) was estimated between populations of the two rivers, and between pairs of subpopulations at each river. The estimates were obtained using GDA (Lewis and Zaykin, 2001) and FSTAT (Goudet, 2001) programs. The genetic distance between pairs of subpopulations was estimated according to Nei (1972). The matrices of geographic distances and genetic distances were compared using the Mantel test (1967) to determine the presence of isolation-by-distance process between subpopulations. The level of associations between endogamy (f) and expected heterozygosity with the number of plants in each community was determined by the Spearman rank correlation coefficient (R).

The Structure 2.1 program (Pritchard et al., 2007) was used to perform a Bayesian clustering analysis indicating the most probable number of clusters by the K value. The software was configured for K ranging from 1 to 10, with 10 independent replicates for each K value. Each replicate had a burn-in of 50,000 iterations, followed by 500,000 MCMC repetitions. The most probable group number (K) was determined as proposed by Evanno et al. (2005). The genetic distance between the individuals was estimated according to the proportion of common alleles (Bowcock et al., 1994), using the MICROSAT program (Minch, 1997), and a dendrogram was constructed using the neighbor-joining method (Saitou and Nei, 1987).

RESULTS

At the Tocó and Capivara Rivers, *G. mustelinum* was found in riparian vegetation and on the floodplains (Table 2). All populations were typically wild and we did not find any evidence or report from the local inhabitants of use of the plants or their parts, as previously

described (Barroso et al., 2010). The populations were structured into subpopulations and the number of subpopulations as well as the number of plants per community were usually higher when the environment was best preserved. In contrast, parts of the riparian vegetation of both rivers were so depredated that the only places where *G. mustelinum* could be found were in the narrow space between the fences of the farms and the bridges and roads. In these restricted places the levels of deforestation and grazing were less intense. When there was evidence of recent and abrupt deforestation, a large number of young *G. mustelinum* plants could be found. This occurred in three places where vegetation had been cut to create pasture or to facilitate the access of the cattle to water sources. These plants grew faster and the number of individuals was higher than in intact environments. Some of the plantlets were strongly damaged by animals grazing in all three places, and in two of them part of the adult plants had been pruned to permit a better growing of the grasses. The fast and profuse growth of *G. mustelinum* in open areas is strong evidence that the species is a pioneer in its environment.

Table 2. Characteristics of two *Gossypium mustelinum* populations.

Community	Location	Latitude (N)	Longitude (E)	Altitude (m)	No.
TC1	Riachão do Jacuípe	-11°53'	-39°17'	195	120
TC2	Riachão do Jacuípe	-11°56'	-39°18'	183	150
TC3	Riachão do Jacuípe	-11°58'	-39°17'	178	30
TC4	Riachão do Jacuípe	-11°47'	-39°15'	218	40
TC5	Iehu	-11°43'	-39°12'	229	30
TC6	Conceição do Coité	-11°36'	-39°09'	294	15
CP1	Itaberaba	-12°30'	-40°19'	250	70
CP2	Itaberaba	-12°31'	-40°22'	271	550
CP3	Itaberaba	-12°21'	-40°24'	297	450
CP4	Itaberaba	-12°26'	-40°23'	281	50
RB1	Itaberaba	-12°30'	-40°27'	299	70
RB2	Itaberaba	-12°30'	-40°25'	288	40

The number of individuals of the population was estimated by the approximate number of plants at the location. Locations registered in Albrana database - Algodões Brasileiros Nativos e Naturalizados (www.cnpa.embrapa.br/albrana) - in which these sites can be consulted.

There were about 350 *G. mustelinum* plants in the 6 sampling points along the Tocó River. The plant depredation was quite intense in these sites, mainly due to the deforestation of riparian forest accompanied by extensive cattle raising. The areas where the number of plants was highest were usually protected from foraging animals by fences. There were few plants in the other places along the Tocó River, and some were protected by thorny species as described in other populations (Pickersgill et al., 1975; Barroso et al., 2010). The environmental conditions along the Capivara River were more favorable to *G. mustelinum* conservation than at the Tocó River. The less human-caused damage in the environment resulted in a better preservation of the native vegetation, and a larger number of *G. mustelinum* plants, around 1200 plants, were located in the 6 sampling points, including one community composed of more than 500 plants. The absence of livestock and agricultural practices in most of the area along the Capivara River contributed to the existence of a large number of seedlings, indicating that this population could be maintained in a sustainable way if the environmental conditions were not altered.

Genetic diversity

The 14 microsatellite markers generated 17 polymorphic loci (Table 1). The 17 loci

produced a total of 85 alleles, with a range of 2 to 9 alleles per locus (average \pm standard deviation: 5.0 ± 2.06). The average number of alleles per locus for the Tocó and Capivara Rivers were 3.53 (2 to 7 alleles) and 3.58 (1 to 7 alleles), respectively. All microsatellite markers showed alleles that were exclusive for at least one of the subpopulations, with frequencies ranging from 0.008 to 0.81. The only exception was one of the loci of the primer CIR381, which amplified no exclusive allele. The endogamy levels (f) of the Capivara and Tocó populations were similar and close to 0.6 (Table 3), which would classify the reproductive system as mixed in both places. In the Tocó population, the endogamy values were high and relatively similar between the subpopulations, with a tendency to preferentially perform self-pollination and crossing between relatives. The range of the variation of the endogamy values was wider in the Capivara population, and the community CPI showed the lowest value. Considering both populations, endogamy was very high and, according to Spearman's correlation coefficient, endogamy was highest when the community was smallest. This tendency was more pronounced in the Tocó ($R = -0.94$; $P < 0.001$) than in Capivara River ($R = -0.58$; $P = 0.23$). Similarly, the genetic diversity estimated by the expected heterozygosity was similar in both populations, with more constant values in subpopulations of the Tocó than in the Capivara (Table 3). The highest and the lowest values of genetic diversity within subpopulations were found in subpopulations of the affluent of the Capivara River, with estimations varying from 0.442 at RB2 to 0.019 at RB1, in which 11 loci showed only one allele. In the Tocó population, the diversity of the subpopulations was more stable, with values from 0.108 at TC6 to 0.320 at TC2. There was no association between diversity and size of subpopulations, as indicated by the low values of the Spearman coefficient of correlation ($R = 0.26$; $P = 0.41$). Considering all the populations, genetic diversity was high, as demonstrated by the expected heterozygosity of 0.482.

Table 3. Estimation for observed heterozygosity (H_o), expected heterozygosity (H_e), endogamy coefficient (f) and fixation index (F_{ST}) in populations and subpopulations.

Population	Community	Sample size	H_e	H_o	f
Capivara	CP1	21	0.091	0.033	0.642
	CP2	21	0.302	0.121	0.602
	CP3	22	0.231	0.159	0.316
	CP4	20	0.166	0.043	0.741
	RB1	22	0.019	0.014	0.274
	RB2	17	0.442	0.166	0.630
Mean	$F_{ST} = 0.634$ (0.581-0.683)		0.208	0.090	0.576 (0.491-0.637)
Tocó	TC1	12	0.274	0.076	0.730
	TC2	18	0.320	0.175	0.460
	TC3	18	0.263	0.119	0.553
	TC4	16	0.145	0.069	0.530
	TC5	20	0.217	0.081	0.632
	TC6	12	0.108	0.032	0.714
Mean	$F_{ST} = 0.547$ (0.467-0.620)		0.221	0.092	0.590 (0.510-0.665)
Among populations	$F_{ST} = 0.328$ (0.234-0.422)		0.482	0.093	0.807 (0.790-0.829)

Numbers in parentheses correspond to bootstrap 95% confidence interval.

The F_{ST} value between populations was 0.328, indicating high genetic differentiation between plants from different rivers. Considering the populations of each river separately, the differences between subpopulations were even higher: the F_{ST} values were 0.547 along the Tocó River, and 0.634 along the Capivara River. The genetic distance calculated between

pairs of subpopulations ranged from 0.02 to 0.806 around the Tocó River and from 0.13 to 0.948 around the Capivara River. The most genetically divergent subpopulations by the Tocó River were those geographically most distant: TC1, next to the headwaters, and TC6, close to the river mouth. TC1 and TC2 were the most similar and physically close. Along the Capivara River, the lowest genetic distance was also between the closest subpopulations, RB1 and RB2, and the highest between CP1 and RB2.

The cluster analysis of subpopulations (Figure 2) or individuals (Figure 3) resulted in two distinct groups, each of them enclosing one population. Subgroups had a tendency to join the neighboring subpopulations. Similar results were obtained when Bayesian analysis was performed, with two different structure levels. The Bayesian analysis performed with all plants sampled in both populations resulted in a maximum ΔK when K was equal to 2, separating the two populations into different groups. As discussed by Evanno et al. (2005), these two groups correspond to the uppermost level of structuring. When each group was evaluated separately, the maximum ΔK occurred with $K = 3$ (Figure 4) for both rivers. The clusters were constituted by the geographically closest subpopulations. Generally, most individuals from each community were grouped together and the subpopulations could be distinguished from the individuals belonging to the two closest neighbor subpopulations. Putative migrants from the upstream community to the downstream community were often observed. The spatial structure of the diversity was confirmed by the Mantel test. The genetic distance between pairs of subpopulations of the Tocó River was highly correlated both with geographic distance ($r = 0.89$) and distances measured throughout the river course ($r = 0.86$). The magnitudes of correlation were slightly lower along the Capivara River: 0.75 when distances were measured in a straight line and 0.67 when measures were taken along the river bed.

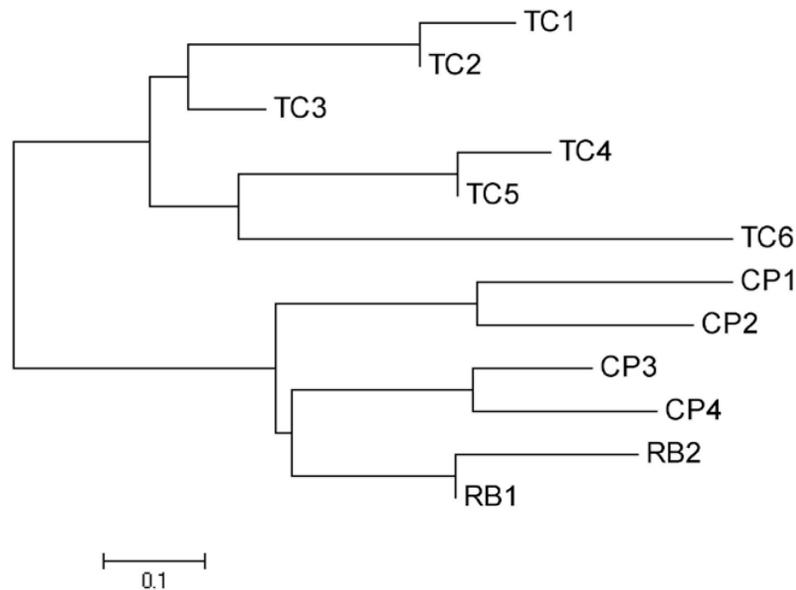


Figure 2. Pattern of genetic structuring of the communities of the populations of Tocó and Capivara Rivers, determined by the neighbor-joining method.

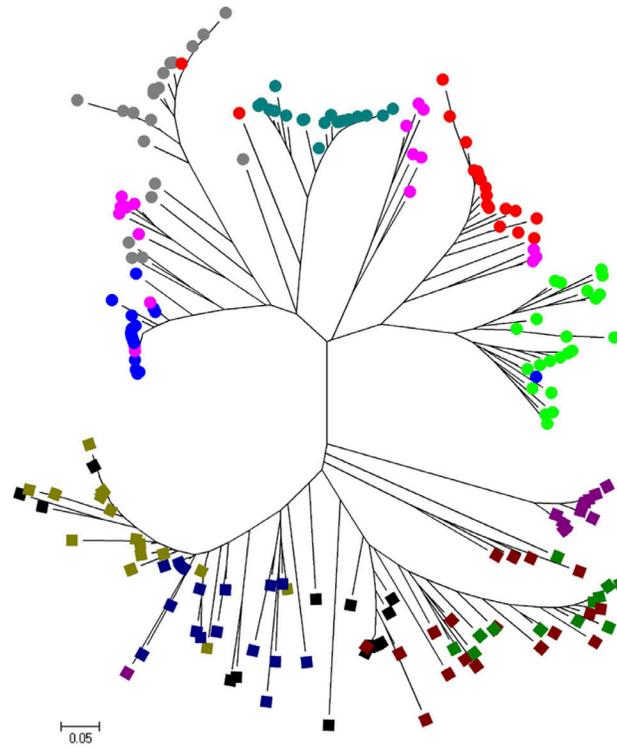


Figure 3. Pattern of genetic structuring of the individuals in Tocó and Capivara Rivers, determined by the neighbor-joining method. Tocó (squares) and Capivara (circles).

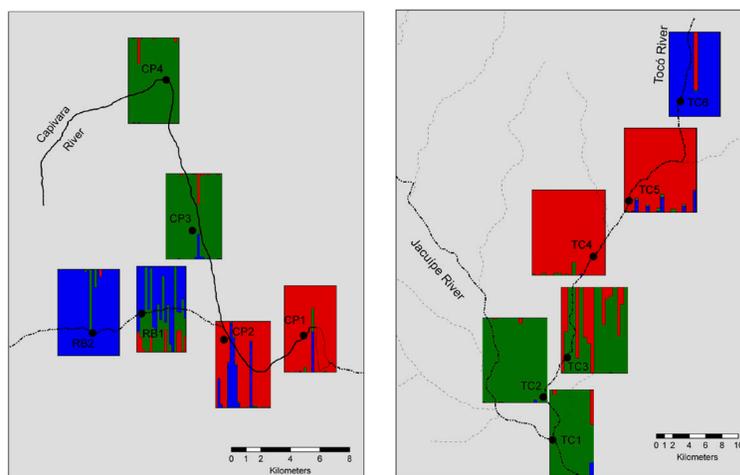


Figure 4. Pattern of genetic structuring of individuals inside communities in Tocó and Capivara populations determined by the Bayesian analysis.

DISCUSSION

Part of the subpopulations included in this study showed strong signs of degradation while others were adequately preserved, showing indications that they were able to remain preserved *in situ*, unlike those previously known. The main problem of the maintenance of both populations was environmental damage to their habitats, wood exploitation, and agricultural practices, notably ruminant livestock. The same impacts of cattle and goat raising observed on other populations of *G. mustelinum* (Pickersgill et al., 1975; Barroso et al., 2010) were also present in the Capivara and Tocó River populations. Animals eat leaves, young shoots and seeds, mainly during the long dry season in the semi-arid regions of northeastern Brazil. Although feeding on adult plants rarely causes the plants' death, damage caused by the animals reduces the amount of seeds the plants can produce. The young plants are killed by feeding. As a result, the substitution of the dead plants due to anthropic or natural causes is slower than is needed to avoid the decline in subpopulations situated in places where livestock has a greater impact.

The allele numbers observed in the Capivara and Tocó populations were larger than the ones obtained in studies of genetic diversity with other cotton species in Brazil, including upland cotton (Bertini et al., 2006; Menezes et al., 2008), *G. barbadense* (Almeida et al., 2009) and Mocó cotton (Menezes et al., 2010). The differences observed between the estimations may be due to the fact that *G. mustelinum* is a wild species that has never been submitted to breeding (Borém et al., 2003). The number of alleles found was more than twice the sum observed in the three populations known until 2006 (Barroso et al., 2010). The elevated number of alleles probably was due to the markedly better preservation of the populations along the margins of the Tocó and Capivara Rivers when compared to previously known populations, which were found distributed over a wider geographic area and composed of a greater number of individuals. The number of alleles that were present exclusively along the Tocó or Capivara Rivers was equivalent: 25 alleles were exclusive to individuals from the Capivara River and 24 were present only in the Tocó River area. These represented 57.64% of the total number of alleles, a high value that is indicative of huge allelic differences. The high quantity of exclusive alleles was probably due to genetic isolation among populations. The absence of gene flow between the populations, associated with genetic drift, resulted in fixation and loss of different alleles. Some exclusive alleles showed a frequency below 5%, and these were restricted to one of the plant subpopulations along the river, which may result from introgressions or genetic drift. It is unlikely that these rare alleles have been introgressed from other cotton species, since the majority of them were present in other populations of *G. mustelinum* (Cavalcante, 2011). The number of alleles that was exclusive to the populations was also greater than that found among the populations of Macururé, Jaguarari, and Caicó (Barroso et al., 2010), which together showed 26 private alleles.

G. mustelinum has characteristics that favor crosses mediated by pollinating insects. It has large pollen grains that are viscous and covered with spikes; the stigmas are longer than the anthers, the petals have strong red spots, and the floral nectaries produce large amounts of nectar. Nevertheless, high endogamy coefficients were observed. The endogamy coefficient in each community was independent of the diversity of the community and proportional to the number of individuals present in the field. Since the smallest populations were usually in the worst conditions of conservation, the highest levels of endogamy were due to the bottleneck caused by reduction of individuals resulting from human activities.

The differentiation observed between the two populations was high and represented 32.8% of total variability. This value reveals that migration among populations, if it exists, is very rare. The low migration frequency can be explained by the geographic isolation between the two populations, separated by 127 km. This distance cannot be overcome by mechanisms of seed dispersal of the species or by insect pollinators. The contact between the two populations must be necessarily mediated by bridge populations or humans. If bridge populations were present, the efficiency to mediate gene flow between the two populations studied was low. The genetic differentiation between these populations was lower than that found for the *G. mustelinum* populations in Macururé, Jaguarari, and Caicó (Barroso et al., 2010), and this is evidence that founder effects or bottlenecks that happened during the history of the populations were more pronounced in the Capivara and Tocó River populations.

The genetic differentiation between subpopulations of each population was still more intense than between populations and represented 54.7% of the variability present in the population from the Tocó River and 63% for the Capivara River. These values are extremely high and indicate the existence of a very strong structure and great differences between the subpopulations. This was confirmed by the association of geographical and genetic distances present within each population and showed that gene flow between subpopulations was not uniform, occurring preferentially among closer populations. Bayesian analysis made it clear that gene flow occurred almost exclusively between neighboring subpopulations (Figure 4). Therefore, the migration structure of this species follows a stepping-stone model. In this model, it is assumed that gene flow occurs between adjacent populations. Consequently, the genetic distance between the subpopulations increases when the geographic distance increases (Kimura and Weiss, 1964). As a result, a gradient of diversity is formed through the rivers from their headwaters to their mouths. Bayesian analysis also permits identifying that migration occurs almost in a unidirectional way, from upstream to downstream subpopulations. Migration mediated by pollen or seeds transportation by animals could not explain the pattern observed. If they were prevalent, geographically close subpopulations would be influenced in the same intensity, and migration rates downstream and upstream would be similar. The transport of seeds by the river's water is the only possible path to explain the pattern of migration between subpopulations observed, as had been previously suggested by Freire et al. (1990). The dispersal of *G. mustelinum* by water can also be evidenced by observation of plant characteristics in the natural habitat, which was also noticed by Freire et al. (1990). Open bolls do not retain the seeds in it, which are dispersed. This is the reason why in some places local inhabitants call *G. mustelinum* "chain cotton" (in Portuguese, *algodão de corrente*), because seeds are arranged "hanging" onto each other forming a structure similar to a chain. Additionally, seeds were observed floating on the surface of ponds near the *G. mustelinum* plants. Ponds are frequently observed along the river bed during the drought season, when there is not enough water to flow. Thus, during the rainy season it is probable that the current distributes the seeds along the river.

The high genetic differences between the populations and between subpopulations, revealed by F_{ST} values, show that genetic conservation requires maintenance of both populations. However, it would be difficult to implement conservation strategies that include all the subpopulations, but considering the results obtained by Bayesian diversity analysis, at least three sites per population should be conserved. Measures to preserve the riparian vegetation of the rivers could be combined with the *in situ* preservation of the species. Considering the fact that most *G. mustelinum* individuals occur within the range that must be preserved as deter-

mined by Brazilian law, compliance with the law would be sufficient to allow the conservation of these populations. A second strategy would be the adoption of the species for the restoration of riparian forests in Brazilian states with semi-arid climates. Projects that involve the restoration of many riparian forests are being developed, and *G. mustelinum* could be included on the list of species used in reforestation, especially in the State of Bahia, where most *G. mustelinum* plants are located, assisting in the recovery of these forests and simultaneously increasing the occurrence of this rare and endemic species in Brazil.

The actual distribution of the species is not known, and other populations may occur in other affluents of the Paraguaçu River basin, so new collections should be made. Diversity studies may lead to the conclusion that the populations are adequately represented in *ex situ* collections. This is the best guarantee of long-term preservation of this important genetic resource for cotton breeding.

ACKNOWLEDGMENTS

Research supported by Embrapa, FINEP, and CAPES. J.N. Dantas assisted in explorations. This study was conducted with IBAMA authorization (#13973).

REFERENCES

- Almeida VC, Hoffmann LV, Yokomizo GK, Costa JN, et al. (2009). *In situ* and genetic characterization of *Gossypium barbadense* L. from the States of Pará and Amapá, Brazil. *Pesq. Agropec. Bras.* 44: 719-725.
- Barroso PAV, Hoffmann LV, Freitas RB, Batista CEA, et al. (2010). *In situ* conservation and genetic diversity of three populations of *Gossypium mustelinum* Miers ex Watt. *Genet. Resour. Crop Evol.* 57: 343-349.
- Bertini CHCM, Schuster I, Sedyama T, Barros EG, et al. (2006). Characterization and genetic diversity analysis of cotton using microsatellites. *Genet. Mol. Biol.* 29: 321-329.
- Borém A, Freire EC, Penna JCV and Barroso PAV (2003). Considerations about gene escape of cotton in Brazil. *Crop Breed. Appl. Biotechnol.* 3: 315-332.
- Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, et al. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368: 455-457.
- Cavalcante U (2011). Caracterização Genética de Populações Naturais de *Gossypium mustelinum* Miers ex Watt. Master's thesis, UFRPE, Recife.
- Doyle JJ and Doyle JL (1990). Isolation of plant DNA from fresh tissue. *Focus* 12: 13-15.
- Evanno G, Regnaut S and Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-2620.
- Freire EC (2002). Fluxo gênico entre algodoeiros convencionais e transgênicos. *Rev. Bras. Ol. Fibras.* 6: 471-482.
- Freire EC, Moreira JAN, Miranda AR, Percival AE, et al. (1990). Identificação de Novos Sítios de Ocorrência de *Gossypium mustelinum* no Brasil. Volume 10. Pesquisa em Andamento. Embrapa Algodão, Campina Grande.
- Fu Y and Somers D (2009). Genome-wide reduction of genetic diversity in wheat breeding. *Crop Sci.* 49: 161-168.
- Goudet J (2001). FSTAT: Program to Estimate and Test Gene Diversities and Fixation Indices (Software). Version 2.9.3. Available at [<http://www2.unil.ch/popgen/softwares/fstat.htm>]. Accessed July 27, 2011.
- Iqbal MJ, Reddy OUK, El-Zik KM and Pepper AE (2001). A genetic bottleneck in the 'evolution under domestication' of upland cotton *Gossypium hirsutum* L. examined using DNA fingerprinting. *Theor. Appl. Genet.* 103: 547-554.
- Kimura M and Weiss GH (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49: 561-576.
- Lewis PO and Zaykin D (2001). Genetic Data Analysis: Computer Program for the Analysis of Allelic Data, Version 1.0 (d16c). Available at [<http://hydrodictyon.eeb.uconn.edu/people/plewis/>]. Accessed July 9, 2011.
- Liu S, Saha S, Stelly D, Burr B, et al. (2000). Chromosomal assignment of microsatellite loci in cotton. *J. Hered.* 91: 326-332.
- Mantel N (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27: 209-220.
- McCarty JC, Wu J and Johnie NJ (2007). Use of primitive derived cotton accessions for agronomic and fiber traits

- improvement: variance components and genetic effects. *Crop Sci.* 47: 100-110.
- Menezes IPP, Hoffmann LV, Alves MF, Morello CL, et al. (2008). Genetic distance among advanced lineages of cotton germplasm using RAPD and microsatellite markers. *Pesq. Agropec. Bras.* 10: 1339-1347.
- Menezes IPP, Barroso PAV, Hoffmann LV, Lucena VS, et al. (2010). Genetic diversity of mocó cotton (*Gossypium hirsutum* race *marie-galante*) from the northeast of Brazil: implications for conservation. *Botany* 88: 765-773.
- Minch E (1997). MICROSAT. Version 1.5b. Stanford University Medical Center, Stanford. Available at [<http://genetics.stanford.edu/hppl/projects/microsat/>]. Accessed July 18, 2011.
- Nei M (1972). Genetic distance between populations. *Am. Nat.* 106: 283-292.
- Neves OS, Cavaleri PA, Grid-Papp IL and Fuzzato MG (1965). Algodoeiro selvagem do Nordeste do Brasil. *Bragantia* 24: 19-25.
- Nguyen TB, Giband M, Brottier P, Risterucci AM, et al. (2004). Wide coverage of the tetraploid cotton genome using newly developed microsatellite markers. *Theor. Appl. Genet.* 109: 167-175.
- Pickersgill B, Barrett SCH and Andrade D (1975). Wild Cotton in Northeast Brazil. *Biotropica* 7: 42-54.
- Pritchard JK, Stephens M and Donnelly P (2007). Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
- Reif JC, Melchinger AE and Frisch M (2005). Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. *Crop Sci.* 45: 1-7.
- Saitou N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406-425.
- Stephens SG and Phillips LL (1972). The history and geographical distribution of a polymorphic system in New World cottons. *Biotropica* 4: 49-60.
- Weir BS and Cockerham CC (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.
- Wendel JFE and Brubaker CL (1993). RFLP diversity in *Gossypium hirsutum* L. and new insights into the domestication of cotton. *Am. J. Bot.* 80: 71.
- Wendel JFE, Rowley R and Stewart MCD (1994). Genetic diversity in phylogenetic relationships of the Brazilian endemic cotton, *Gossypium mustelinum* (Malvaceae). *Plant Syst. Evol.* 192: 49-59.