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

Ivandilson Pessoa Pinto de Menezes · Fernanda Amato Gaiotto · Lucia Vieira Hoffmann · Ana Yamaguishi Ciampi · Paulo Augusto Vianna Barroso

Abstract *Gossypium mustelinum* is a wild cotton relative found only in the semiarid region of Bahia state in Brazil, and changes caused by humans in the natural habitat of this species have endangered the existence of several natural populations. Information about the occurrence and genetic composition of these populations is necessary to design effective conservation measures. The aim of this study was to characterize the in situ maintenance mode and assess the genetic diversity of *G. mustelinum* populations in the basin of the De Contas River. A sample of 205 *G. mustelinum* specimens was collected from the margins of the Jacaré, Riacho Quixaba, Riacho Serra Azul, and Riacho Riachão rivers and genotyped using 13 SSR primer pairs. In general, all *G. mustelinum* populations exhibit inadequate in situ maintenance, predominantly due to the deforestation of riparian vegetation and herbivory. The observed total genetic diversity of *G. mustelinum* was significant ($H_E = 0.489$), highly structured ($F_{ST} = 0.534$), and organized in homozygous genotypes ($F_{IS} = 0.873$). The high observed inbreeding level is consistent with the predominance of self-fertilization and geitonogamy ($t_m = 0.234$). In addition, the pattern of genetic structure tended to form groups that coincided with the collection sites, i.e., first clustering within subpopulations, then within populations, and finally within the closest populations. Thus, the observed genetic diversity is likely to be rapidly lost, and conservation measures should therefore be undertaken.

Keywords Small populations · Genetic structure · SSR markers · Wild cotton · Conservation genetics

I. P. P. de Menezes
Instituto Federal Goiano, Km 2.5, Urutaí, Goiás 75790-000, Brazil

I. P. P. de Menezes · F. A. Gaiotto
Universidade Estadual de Santa Cruz, Km 16, Ilhéus, Bahia 45662-000, Brazil

L. V. Hoffmann · P. A. V. Barroso (✉)
Embrapa Algodão, Núcleo do Cerrado, BR 153, Km 04, Goiânia, Goiás 74001-970, Brazil
e-mail: paulo.barroso@embrapa.br

A. Y. Ciampi
Embrapa Recursos Genéticos e Biotecnologia, Brasília, Distrito Federal 70770-917, Brazil

Introduction

Gossypium mustelinum Miers is the only cotton plant native to Brazil and it is endemic to the northeastern region, including the Caatinga biome, where wild populations exist in the immediate vicinity of intermittent watercourses and lakes. The Caatinga is a Brazilian biome at serious risk of desertification in many of its sectors and represents one of the least-studied ecosystems in Brazil (Castelletti et al. 2003; Prado 2003). It is an area subjected to elevated human intervention, suffering from intense environmental degradation caused by wildfires, deforestation, and agricultural activities (Castelletti et al. 2003). These actions have reduced the number of individuals and compromised the maintenance of *G. mustelinum* populations in their natural habitat (Alves et al. 2013; Barroso et al. 2010). Several populations have become extinct due to changes caused by human activities, including the clearing of riparian vegetation and the use of river beds for agriculture in the case intermittent streams, which are abundant in this region.

Brazil is the site of origin of *G. mustelinum*, one of the five allotetraploid *Gossypium* species, and an important center of distribution and diversity for two other domesticated

amphidiploid cotton species (*G. hirsutum* and *G. barbadense*) (Pickersgill et al. 1975; Wendel et al. 1994). These two domesticated species occur as crop plants, local varieties, in dooryards or as feral populations (Albrana 2013). All three species belong to the cultivated cotton primary gene pool and can be used as a source of variability for breeding programs. The allopolyploid origin created a considerable variability (Wendel et al. 1992), reduced otherwise by the history of domestication and selection (Westengen et al. 2005). The narrow genetic base of the modern cotton cultivars (*G. hirsutum*) is one if the major difficulties in breeding programs (Bertini et al. 2006). The study of identification, ecological state, and genetic relationships of new populations of *G. mustelinum* provides knowledge to enrich *Gossypium* germplasm collections, guide in situ conservation, and also improve certain characteristics of interest such as resistance to stress and fiber quality (Gardunia 2006).

The loss of genetic diversity that occurs in small populations represents a major natural bottleneck, and will limit further evolution (Ellstrand and Elam 1994). Genetic variability is essential for the adaptation of a population and is necessary for species to overcome the challenges imposed by the environment (Reed and Frankham 2003). According to the Food and Agriculture Organization (FAO), the selection pressure imposed by climatic changes may exacerbate the reduction of genetic diversity, not only in local and modern cultivars but also in wild populations of interest. In an attempt to conserve the genetic resources of cotton species in Brazil, Embrapa (the Brazilian Agricultural Research Corporation) has conducted expeditions with the goal of promoting ex situ maintenance and genetic characterization to identify in situ conservation strategies (Almeida et al. 2009; Barroso et al. 2010; de Menezes et al. 2010). Knowledge about the genetic structure, mating system, gene flow, and dispersal patterns of populations provides key information for this purpose (Vaz et al. 2009).

Considering the origin and history of the development of *G. mustelinum* in the semiarid Caatinga, which is a biome subjected to high abiotic stress, it appears that populations of this species may have developed genotypes adapted to adverse conditions but might currently be losing these adapted genotypes. To provide information for conservation of *G. mustelinum* populations in the De Contas River basin in the central southern region of Bahia state, Brazil, the sites where the genotypes occur were determined and conditions described. Reproductive system, diversity, structure, and genetic divergence were characterized using selected simple sequence repeats (SSR) markers. With the results we expected (1) to detect if *G. mustelinum* populations found in the De Contas River basin form a significant reservoir of genetic diversity, (2) to determine if nearby populations located in different streams of the same basin possess divergent structure, (3) to clarify the level of

genetic relationships among *G. mustelinum* populations and other allotetraploid *Gossypium,* and (4) to provide recommendations that help to define conservation and sampling strategies.

Materials and methods

Ecological characterization, tissue sampling, and DNA extraction

Expeditions were conducted to monitor *G. mustelinum* populations, and new areas of occurrence have been found in the regions of Jequié and Vitória da Conquista, Bahia, Brazil (Fig. 1). The sampling expedition covered the entire De Contas River basin, which consists of the cities of Iramaia, Jequié, Manoel Vitorino and Boa Vista, in Bahia, Brazil, and was conducted from September 3–6, 2009. During the expedition, direct observations of *G. mustelinum* populations were conducted using structured questionnaires that include geographical, morphological, phonological, and environmental issues (Albrana 2013). The leaves and seeds were used for genetic analyses, while the plants produced from the cuttings were deposited into the germplasm bank maintained in vivo by Embrapa. A total of 221 plants were analyzed, which included 205 *G. mustelinum*, seven *G. barbadense*, six *G. hirsutum,* and three *G. hirsutum* var *marie galante*. Samples of *G. hirsutum* and *G. barbadense* were sympatric with the studied populations, while *G. hirsutum* var *marie galante*, a local variety known as mocó cotton that grows in the Caatinga, had been collected previously (de Menezes et al. 2010). The collected *G. mustelinum* germplasm consisted of four populations, which were defined according to the affluent of the De Contas River where the specimens were collected (Fig. 1). The population at Riacho Riachão II (RR-II) was selected to determine the reproductive system of the plants. Twelve families, represented by the leaves of the mother plant and 16 individual seeds, were genotyped using five polymorphic SSR markers. Genomic DNA was extracted from the leaves as described previously using the DarT protocol (2009) and from the seeds as described by Schuster et al. (2004). The mother plant's leaves were compared to the DNA from the seeds to estimate the mating system.

SSR markers

A set of 13 SSR primer pairs marked with fluorochromes (6-FAM, NED, and HEX) was selected to genotype the populations of the multiplex systems (Table 2). This selection was performed on the basis of the high polymorphism observed in the closely related species

Fig. 1 Geographic location of the *G. mustelinum* populations in the Contas River basin in Bahia, Brazil
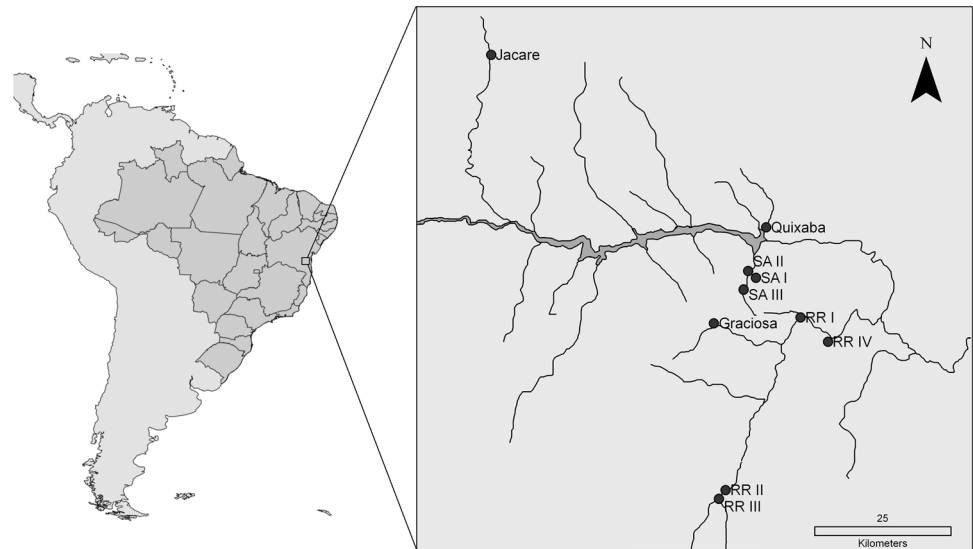
G. barbadense (Almeida et al. 2009), local *G. hirsutum* plants (de Menezes et al. 2010), and *G. mustelinum* (Barroso et al. 2007, 2010). Five out of the 13 SSR primer pairs (BNL1434, BNL203, BNL840, BNL3103, and BNL252) were selected for genotyping progenies and studying the reproductive system of subpopulation RR-II. The SSRs were amplified via PCR using a multiplex PCR Kit (Qiagen). Each amplification reaction contained 5.0 ng of DNA, 0.0025 cm^3 of 2× Qiagen multiplex PCR Master Mix (HotStar Taq DNA Polymerase, PCR amplification buffer, 3 mM MgCl$_2$), 2.5 μL of Q-solution, 0.2 μM of each primer pair (forward and reverse), and 5 μL of RNase-free water. PCR was performed in a thermocycler with an initial denaturation of 95 °C for 15 min., followed by 34 cycles, each consisting of a denaturation step at 95 °C for 1 min.; an annealing step at 55 °C (except for CIR311, for which the annealing step was 51 °C for 1.5 min.); and an extension step at 72 °C for 1 min., followed by a final extension step at 60 °C for 30 min. The amplicons of all of the multiplexes were diluted at a ratio of 1:2 with ultrapure sterile water. An aliquot of 0.5 μL of this dilution was added to a mix of 9.42 μL of DI HI-formamide and 0.08 μL of Rox GeneScan 500 (Applied Biosystems). Finally, the amplified DNA fragments were separated via capillary electrophoresis using an ABI 3100 DNA analyzer (Applied Biosystems), and allele sizes were estimated with *GeneMapper* software, version 3.5 (Applied Biosystems).

Data analysis

Genetic variability was estimated based on the SSR allele frequencies, number of polymorphic alleles per SSR locus (A_P), number of private alleles (alleles counted in a single population or species), and expected (H_E) and observed (H_O) heterozygosity according to Nei (1978) using GDA software (Lewis and Zaykin 2001). Wright's *F* statistic was estimated as described by Weir and Cockerham (1984) at a range interval of 95 % with 10,000 bootstrapping replicates using *Fstat* software (Goudet 2001). The mating system was analyzed using a mixed model of reproduction on the basis of the estimates of the unilocus (t_s) and multilocus (t_m) outcrossing rates, the correlation of paternity (r_p), and the parental inbreeding rate (F_m). The SD was calculated with 1,000 bootstrapping replicates using *Mltr* software, version 3.2 (Ritland 2002). The fixation index (*F*) assuming neutral equilibrium due to the reproductive system was estimated with the following equation: $F = (1 - t_m)/(1 + t_m)$. The genetic divergence (GD) values at the population and individual levels were calculated using Nei's genetic distance (1978), and the proportion of common alleles (*ps*, then GD = 1 − *ps*) (Bowcock et al. 1994), respectively. The genetic relationship was assessed by means of the genetic distance using dendrograms, in which the neighbor joining algorithm was applied using *Mega5.05* software (Tamura et al. 2011). Mantel's test was employed to evaluate the degree of association between the genetic and geographic distances of populations and sub-populations transformed by the logarithm. An analysis of the population assignment was performed on the basis of Bayesian statistics, implemented using *Instruct* software (Gao et al. 2007). The software was configured for (v-2) and *K* ranging from 1 to 10, with ten independent replicates for each *K* value. Each replicate had a burn-in of 50,000 iterations, followed by 500,000 MCMC repetitions. The group number was determined as proposed by Evanno et al. (2005), where graphic representation was performed using *Distruct* software (Rosenberg 2004).

Table 1 Summarization of the observations made in the *G. mustelinum* occurrence areas in the basin De Contas River

Municipalities	Iramaia	Jequié						Boa Nova		Manoel Vitorino
Population	Jacaré	Quixaba	Serra Azul			Riacho Riachão				Graciosa
Subpopulation	JAC	QUI	SAI	SAII	SAIII	RRI	RRIV	RRII	RRIII	Graciosa
Altitude (m)	318	243	245	246	295	235	205	418	422	424
Size	50	30	40	50	60	27	3	60	21	1000
Degradation	+	+ −	+	+	−	−	+	+ −	+ −	−
Seed	P	P	A	P	P	P	A	P	P	P
Young plant	P	P	A	P	P	P	A	P	P	P
High (m)	<4	<5	<5	<5	<5	<3	<1	<3 e 10[1]	<7	<2.5
Grown	Sh	Sh/L	Sh/L	Sh/L	Sh/L	Sh	Sh/L	Sh/L[1]	L	Sh/L
Other cotton	Bar	NO	NO	NO	NO	Hir	NO	NO	NO	NO

P present, *A* absent, *Sh* shrub, *L* liana, *NO* no observed

Bar, *G. barbadense*; Hir, *G. hirsutum*

Results

In situ maintenance

Plants of *G. mustelinum* were found in the De Contas River basin in four municipalities, scattered in 10 different locations, with minimum and maximum distances of 2 and 91 km, respectively (Fig. 1). The plants were attributed to four populations defined according to the stream where they occurred: Jacaré, Quixaba, Serra Azul, or Riacho Riachão. Within this geographical area, individual *G. mustelinum* plants were not continuously distributed but were found in small subgroups along intermittent streams. The number of individuals at a location was generally small, except for the population found at Graciosa Farm, which was unique in that it was formed by a large number of young plants of approximately the same age, suggesting that the land had been deforested, and these *G. mustelinum* plants exhibited pioneer species behavior. All of the observed plants were distributed in riparian areas, which showed different levels of degradation. The causes of this degradation predominantly included deforestation of the riparian vegetation and forest fires, and the subsequent use of the areas for cattle and goats. The consequences of this exploitation were decreases in the number of plants as well as physical damage and reduction of seed production caused by animal feeding on the plants and their reproductive structures. The number of young plants and seeds was greater when the *G. mustelinum* grew around or were anchored to other native plants, which would not only protect the cotton plants from animals but would create a shaded environment, resulting in more vigorous, liana-like growth (Table 1). Plants of the cotton species *G. barbadense* and *G. hirsutum* were found <20 and 50 m away from the Jacaré population and RR-I subpopulation, respectively. The *G. barbadense* plants were located in

backyards and have been cultivated for decades by local residents, while the *G. hirsutum* plants were most likely derived from the use of the seeds of this species in animal feed or its dispersion during transport. There were approximately 200 small plants, with a maximum height of 30 cm, showing one to two bolls per plant. Despite the absence of signals of *G. hirsutum* introgression in the RR-I subpopulations, part of the *G. mustelinum* population from Rio Jacaré exhibited clear morphological signs of *G. barbadense* introgression.

Genetic diversity

The 13 SSR markers used in this study were polymorphic in *G. mustelinum* (Table 2). The CIR212 primer pair amplified a second locus, which was monomorphic among the *G. mustelinum* accession (153 bp) but exhibited different alleles in *G. hirsutum* (134 and 141 bp) and *G. barbadense* (126 bp). The 13 polymorphic primer pairs amplified 59 alleles. The number of alleles per locus varied from two, in CIR212b, to seven, in BNL1421, with an average of 4.77 alleles per locus observed in *G. mustelinum*. In *G. barbadense* and *G. hirsutum*, 20 and 36 alleles were detected, respectively, at the same 13 loci. Among the total amplified alleles, 39 alleles in *G. mustelinum*, 7 alleles in *G. barbadense,* and 25 alleles in *G. hirsutum* were defined as species-specific alleles (Table 2). In general, the analyzed loci were informative and presented an average value of total genetic diversity (H_E) of 0.489, with the exception of locus BNL252 (0.126), which was not very informative. The average observed heterozygosity was extremely low ($H_O = 0.031$), indicating a nearly complete absence of intralocus variation in the *G. mustelinum* cotton plants, which was consistent with the high average value of the estimated fixation index (F_{IS}) (Table 2).

Table 2 Genetic diversity at the individual SSR loci level across 205 *G. mustelinum* plants

Primers	Chr (genome)	Flourochrome	Size[f]	A (A_{SS})			H_E^f	H_O^f	F^f
				M	B	H			
BNL1421[a]	c13 (A)	HEX	222–251	7 (6)	2 (2)	3 (2)	0.70	0.05	0.93
BNL1551[a]	c21 (D)	6-FAM	158–188	6 (3)	1 (0)	5 (3)	0.75	0.08	0.89
BNL1434[a]	c2 (A)	NED	232–248	6 (3)	2 (1)	4 (2)	0.58	0.02	0.96
CIR203[b]	c6 (A)	HEX	240–254	4 (3)	2 (1)	3 (3)	0.36	0.01	0.97
BNL3261[b]	c12 (A)	NED	194–225	6 (5)	1 (0)	3 (3)	0.44	0.04	0.91
BNL252[c]	c8 (D)	6-FAM	169–199	4 (2)	2 (0)	2 (2)	0.13	0.00	0.95
BNL256[c]	c10	NED	213–219	4 (3)	2 (1)	1 (1)	0.36	0.02	0.96
BNL840[d]	c19 (D)	HEX	147–241	4 (3)	1 (0)	2 (2)	0.57	0.02	0.96
BNL1350[d]	c01	6-FAM	205–241	4 (3)	3 (2)	2 (2)	0.58	0.03	0.95
CIR246[d]	c14 (D)	NED	154–166	3 (1)	2 (0)	3 (1)	0.47	0.02	0.95
BNL3103[e]	c25 (D)	NED	185–204	4 (2)	1 (0)	4 (3)	0.43	0.04	0.91
CIR212b[e]	c3 (A)	NED	161–163	2 (1)	1 (0)	1 (0)	0.46	0.03	0.94
CIR311[e]	c15 (D)	6-FAM	174–184	5 (3)	1 (0)	3 (1)	0.53	0.05	0.90
Total	–	–	–	59 (39)	20 (7)	36 (25)	0.49	0.03	0.94
SE	–	–	–	–	–	–	0.04	0.01	0.01

A number of alleles, A_{SS} number of alleles specific species, H_E and H_O expected and observed heterozygosity, respectively, *F* fixation index

M, *G. mustelinum*; B, *G. barbadense*; H, *G. hirsutum*

[a–e] Set multiplex, [f] Estimative considering only *G. mustelinum*

A total of 37 % of the identified alleles were private, i.e., were present in only one of the populations. The number of private alleles was higher in the Jacaré (10 alleles) and Riacho Riachão populations (9 alleles). Nine alleles were present only in the Jacaré population and were found only in an interspecific hybrid, which were derived from *G. barbadense*. In the subpopulations at Riacho Riachão and Serra Azul, 17 and 11 private alleles were detected, respectively, which constituted approximately 40 % of the alleles in the population. Most of these private alleles among the populations were rare, being present at frequencies that were equal or below 5 %.

In Table 3, the genetic variability indices are separated as a function of the collected populations. The number of alleles ranged from 21 to 44 in the Quixaba and Riacho Riachão populations, respectively, with an average of 3.19 polymorphic alleles per locus per population. The average genetic diversity of the populations was significant ($H_E = 0.286$). It was lower in the Quixaba population (0.053) and higher in the Riacho Riachão population (0.381). The observed average heterozygosity was low ($H_O = 0.033$), as previously observed for the total sample, revealing a high homozygosity within individuals, resulting in a high inbreeding coefficient ($F_{IS} = 0.873$). In the Jacaré population, the interspecific hybrid had a large effect on the estimates of genetic diversity and inbreeding. When it was removed from the analysis, the estimates of H_E and F_{IS} changed from 0.102 and 0.044 to 0.611 and 1.00, respectively, indicating populations with more inbreeding and

similar individuals. Within the Riacho Riachão population, the RR-I and RR-II subpopulations showed the largest number of alleles (33 and 25) as well as the highest genetic diversity values (0.338 and 0.276). Although homozygous individuals predominated, a significant value of $H_O = 0.170$ was detected in the RR-I subpopulation, resulting in a reduced fixation index. Among the three Serra Azul subpopulations, SA-II presented the greatest number of alleles and the highest value of H_E (21 and 0.113), while SA-I (17 and 0.083) and SA-III (17 and 0.067) presented the lowest values. The H_O values were equally low (0.003, 0.031, and 0.015, in this order), indicating high inbreeding in all of the Serra Azul subpopulations.

Reproductive system

The estimated multilocus outcrossing rate ($t_m = 0.234 \pm 0.063$) was consistent with the estimated high inbreeding values for the populations (Tables 2, 3). These values indicated that reproduction occurred predominantly via selfing and geitonogamy. The single locus outcrossing rate ($t_s = 0.278 \pm 0.049$) was slightly higher than the multilocus crossing rate. The two estimates indicate an excess of homozygotes within individuals in the progeny. This homozygous excess was also observed by comparing the estimated parental inbreeding ($F_m = 0.999$) and that expected under equilibrium conditions ($F = 0.621$). This difference indicated that there were factors other than reproductive behavior that contributed to the increase in the

Table 3 Genetic diversity and population structure in *G. mustelinum* based on 13 SSR loci

Population	Subpopulation	S	A	A_P	H_E (SE)	H_O (SE)	F_{IS} (SE)	F_{ST}
Jacaré	–	25	38	3.4	0.10 (0.01)	0.04 (0.00)	0.61 (0.01)	–
Quixaba	–	21	21	2.7	0.05 (0.02)	0.004 (0.003)	0.93 (0.05)	–
R. Riachão	Graciosa	23	18	2.7	0.06 (0.03)	0.003 (0.003)	0.94 (0.04)	0.598*
	R. Riachão I	22	33	2.8	0.34 (0.06)	0.17 (0.04)	0.49 (0.08)	
	R. Riachão II	22	25	2.3	0.28 (0.06)	0.01 (0.004)	0.98 (0.01)	
	R. Riachão III	20	15	2.0	0.03 (0.01)	0.01 (0.004)	0.83 (0.13)	
	R. Riachão IV	3	13	0.0	0.00 (0.0)	0.00 (0.0)	1.00 (0.0)	
	Total	90	43	3.2	0.38 (0.05)	0.05 (0.01)	0.88 (0.01)	
S. Azul	S. Azul I	23	17	3.0	0.08 (0.05)	0.003 (0.003)	0.96 (0.03)	0.746*
	S. Azul II	23	21	2.3	0.11 (0.04)	0.03 (0.01)	0.72 (0.07)	
	S. Azul III	23	17	2.5	0.06 (0.04)	0.02 (0.01)	0.76 (0.18)	
	Total	69	29	3.5	0.27 (0.08)	0.02 (0.01)	0.94 (0.01)	
Mean		51.2	–	3.2	0.29 (0.03)	0.03 (0.004)	0.87 (0.02)	
Total		205	–	–	–	–	–	0.534*

A number of alleles for population, A_P number mean of polymorphic alleles, H_E and H_O expected and observed heterozygosity, respectively, F_{IS} fixation index and F_{ST} genetic differentiation, *SE* standard error

* Significant ($P < 0.05$)

inbreeding value in the subpopulation. An expressive value of the paternity correlation was estimated ($r_p = 0.422 \pm 0.17$), which indicated that the outcrosses in any of the mother progenies often share the same plant as a source for non-maternal pollen.

Genetic differentiation and structure

The genetic differentiation estimated using F_{ST} was high and differed significantly from zero when calculated among all of the populations (0.534). This value was even higher when estimated for the Riacho Riachão (0.568) and Serra Azul (0.746) subpopulations (Table 3). The difference remained significant when it was calculated pairwise in populations at the RR and SA rivers, with the exception of RR-III compared to RR-IV. Thus, both the geographic structure and the river where the subpopulation was found showed a strong effect on the genetic structure. This large difference was also verified using the average Nei's genetic distance (1978) calculated among the populations (0.348). A large average genetic distance would be consistent with the other estimated population diversity indices. The lowest genetic distance (0.107) was observed between the Quixaba and Serra Azul populations, while the greatest genetic distance (0.561) was found between the Jacaré and Quixaba populations. A high genetic divergence was also observed in populations found at Riacho Riachão (0.345) and Riacho Serra Azul (0.249).

According to the Mantel test, 76 % of the genetic divergence among the populations was due to their geographic structure, and a high and significant correlation was
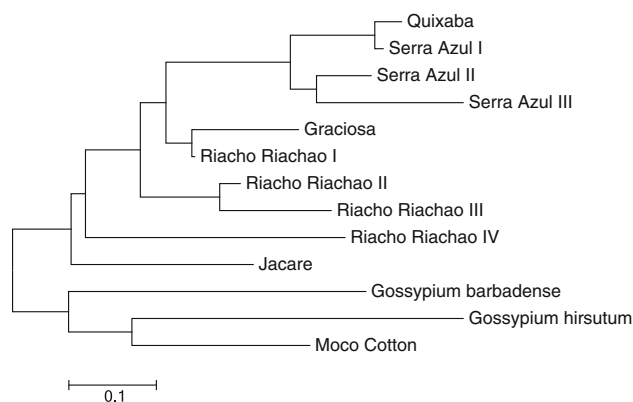

Fig. 2 Neighbor-joining tree showing the genetic relationship among the populations of the *G. mustelinum* based on Nei's genetic distance from 13 SSR loci

found between the genetic and geographic distance ($r = 0.87$, $p < 0.05$). Genetic and geographic distances did not significantly correlate within subpopulations found on Riacho Riachão and Serra Azul, but the dendrogram otherwise indicated a clustering pattern that was related to the geographic location of the collection sites on these streams (Figs. 1, 2).

The cluster analysis (Fig. 2) shows that the *G. mustelinum* subpopulations formed two well-defined groups, which separated the *G. mustelinum* population from the other evaluated *Gossypium* species.

The estimated ΔK proposed by Evanno et al. (2005) generated two peaks, $\Delta K = 2$ and $\Delta K = 5$. The level of structure detected at $\Delta K = 2$ corresponded to the higher hierarchical level populations: Riacho Riachão and Riacho

Fig. 3 Attribution of the populations of 205 *G. mustelinum* plants on the basis of Bayesian analysis using the *Instruct* program. **a** and **b** are graphic representations of the different genetic groups for $\Delta K = 2$ (**a**) and $\Delta K = 5$ (**b**)
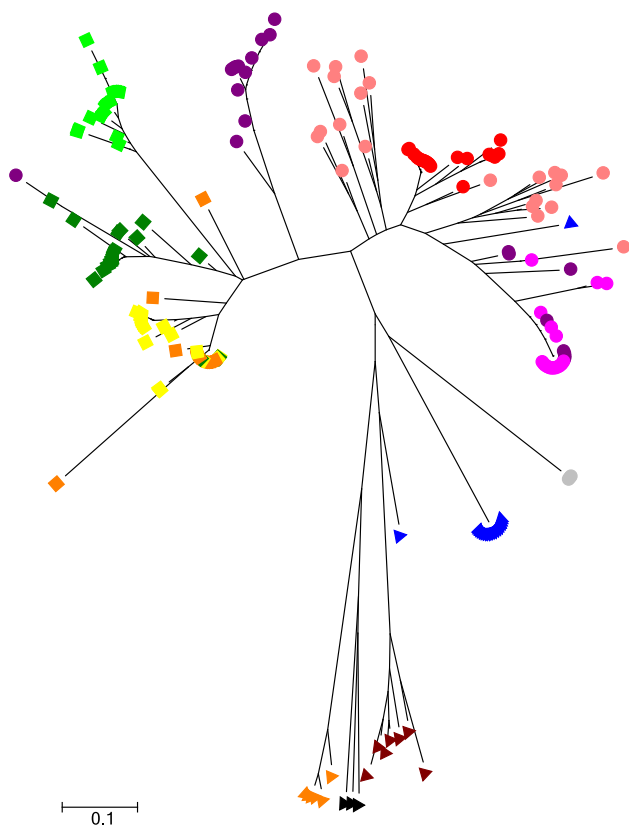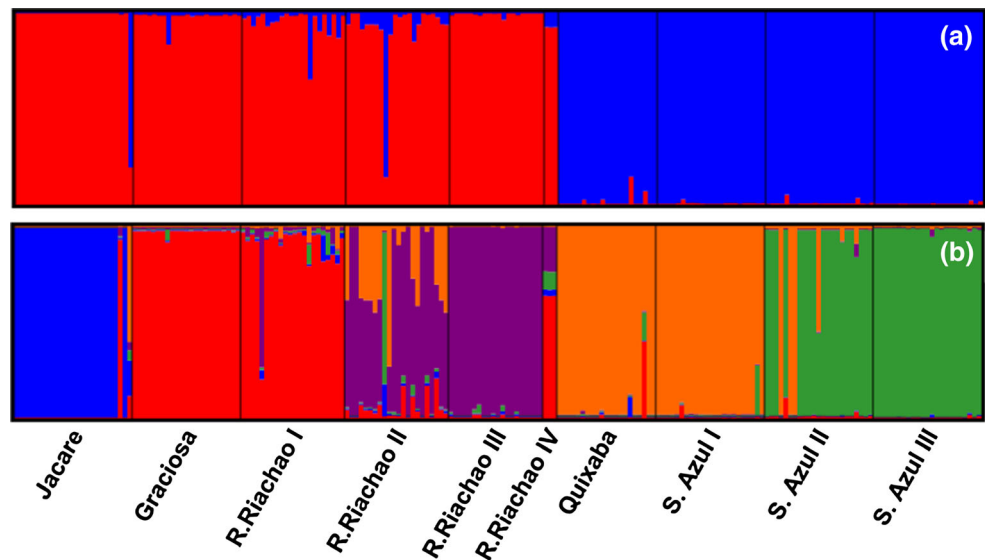
Fig. 4 Neighbor-joining tree of the 205 *G. mustelinum* genotypes including some individuals of *G. hirsutum*, *G. barbadense* and *mocó* cotton based on proportion of common alleles from 13 SSR loci. Each population's individuals are indicated using the following *symbols*: ▲, Jacaré; ■, Quixaba; ■, Serra Azul I; ■, Serra Azul II; ■, Serra Azul III; ●, Graciosa; ●, Riacho Riachão I; ●, Riacho Riachão II; ●, Riacho Riachão III; ●, Riacho Riachão IV; ▲, *G. barbadense*; ▲, *G. hirsutum* L.; ▲, mocó cotton. (Color figure online)

Serra Azul, with the Jacaré and Quixaba populations being anchored in the Riacho Riachão and Serra Azul populations, respectively (Fig. 3a).

At $\Delta K = 5$ (Fig. 3b), the inferred groups were congruent with the groups indicated in the dendrogram by population and subpopulation (Figs. 2, 4). Furthermore, they corresponded to most of the populations defined according to the creek of occurrence (Fig. 1). This clustering makes clear that the connection among subgroups is strongest within the same river, which confirmed that a large number of plants shared the same mixed ancestry and/or formed the same genetic group. An exception to this pattern was observed for Quixaba and SA-I, which formed a single genetic group and furthermore, the significant ancestry between this group and RR-II (Fig. 3b), although they are geographically isolated rivers.

The results of individual clustering on the basis of genetic distance were similar to those obtained using *Instruct* software, which separated the plants into groups according to the creek or river of occurrence (Fig. 4). The RR-II plants formed an intermediate branch between the group including plants from the Sierra Azul/Quixaba populations and another group formed by plants from Riacho Riachão. The Jacaré population formed a branch that was separated from the other populations and was located near a branch that included other species, most likely influenced by introgressions from *G. barbadense*.

Discussion

Negative interference by humans has been observed at every site of *G. mustelinum* occurrence (Alves et al. 2013;

Barroso et al. 2010). In Ceará state, where the species was first reported, it is no longer found (Barroso et al. 2010). Prior to 2009, six out of the eight native populations were reported to be small (Alves et al. 2013; Neves et al. 1965; Pickersgill et al. 1975), and two of them were highly endangered, similar to the RR-IV subpopulation studied in this work. In the De Contas River basin. The situation of the species in De Contas River was not as critical as described previously; however, it is also in a poor state of conservancy. If the current conditions are maintained, the evolutionary capacity of these remaining populations and subpopulations will be compromised.

The high estimated value of F_{ST} indicates a high level of reproductive isolation among populations. The subpopulations at Riacho Riachão and Serra Azul presented the two greatest F_{ST} values and hence are the most isolated populations. Even in the populations separated by the smallest distances there was strong differentiation, suggesting low levels of gene flow among and within the populations at each river. It was consistent with the genetic structure observed in situ, where G. mustelinum plants occurred in small clusters separated from one another by several kilometers. The significant Mantel's correlation between physical and genetic distances of the populations allows the conclusion of the existence of isolation by distance mechanism among the populations. Hence the gene flow was low, but enough to provide a geographical structure of the diversity among populations. The same was not true when we evaluated the subpopulations inside the same watercourse, where genetic distances are not explained by geographical isolation, so it is likely that genetic drift, resulting from bottleneck or founder effect, is the main evolutionary effect and cancels out the reduced levels of gene flow.

The neighborhood gene flow detected using individual cluster analysis revealed a hierarchical genetic organization, where individuals tended to first cluster within subpopulations, then within populations, and finally, with other close populations. Although this effect may have occurred via pollen or seed dispersal, it is unlikely to be attributed to pollen because it did not show a mutual genetic influence either within or between populations, which was also indicated by the large differences between the populations (F_{ST}; Nei's distance; private alleles). This effect was uni-directional, following the route of seed dispersal by water along the course of the river, as suggested by Freire (2002) and in accordance with the stepping-stone model observed in other G. mustelinum populations by Alves et al. (2013).

Sixty-one percent of the SSR allele compositions amplified for G. mustelinum were shown to be specific to this species when compared to G. barbadense or G. hirsutum (Almeida et al. 2009; de Menezes et al. 2010,

CMD 2013). Although little is known regarding the uniqueness of the value of G. mustelinum, it can certainly be used as an allelic source to enlarge the genetic base of G. hirsutum cultivars. Thus, the wild species must be further studied to identify its value as genetic resource (Freire 2002; Gardunia 2006, Hague et al. 2007).

Although the basin of the De Contas River was found to be less degraded compared to the basin of the Paraguaçu River (Alves et al. 2013), the observed genetic diversities were similar, with H_E values of 0.49 and 0.48, respectively. Both presented high genetic diversity compared to the populations studied by Barroso et al. (2010) ($H_E = 0.25$). The lowest estimate, 0.08, had been obtained by Wendel et al. (1994), probably due to the small number of individuals and reduced amount of information provided by isozyme in relation to SSR markers. Therefore, the genetic diversity of G. mustelinum is greater than previously thought and also when compared to the allotetraploid cultivated cotton relatives from Brazil, such as G. barbadense, with an H_E of 0.39 (Almeida et al. 2009), and G. hirsutum L. cultivars, with an H_E of 0.440 (Bertini et al. 2006), while it was slightly lower than the H_E reported for G. hirsutum r. marie galante (0.52) (de Menezes et al. 2010).

The G. mustelinum plants collected at the Jacaré River form the only group showing phenotypic evidence of the introgression of G. barbadense. Compared to other visited populations with the typical G. mustelinum morphological pattern, the petals presented a darker yellow color, the fruits were longer, more fusiform, and with gossypol glands more prominent. Besides the morphological evidences, the alleles present in individuals were typical from G. mustelinum. There was only one exception, the F_1 hybrid heterozygous in all SSR loci, with one allele from G. mustelinum and another from G. barbadense. Probably the morphological traits were the resemblances were identified are controlled by loci not linked to SSR markers. To form the hybrid, the pollen of G. barbadense was provided to dooryard plants of G. barbadense present in houses situated <20 m from G. mustelinum plants. Although there are sexual barriers between the tetraploid Gossypium species (McGregor 1976; Pereira et al. 2012), they are only partial and allow spontaneous interspecific crossings, as observed in this and other studies (Neves et al. 1968). The presence of interspecific Sympatry between G. mustelinum and other allotetraploid Gossypium species must be avoided whenever in situ conservation is desired. The in situ conservation can be achieved with the implementation of the Brazilian forest Code, which designates the permanent preservation of up to 30 m of riparian vegetation along both margins of waterways that are less than 10 m wide. Because the genetic diversity is highly structured, it is necessary to conserve most of the populations. Ex situ conservation must be performed as a

complementary measure to ensure that diversity is preserved whether environmental impact or human mediated disturbance occurs. The strong geographic structure of genetic diversity indicated that sampling must be performed in a stratified manner, and the strategy should be defined according to genetic groups determined by K populations, as indicated by Bayesian analysis. Finally, the number of individuals collected in each place must be defined considering the high level of inbreeding.

References

Albrana (2013) Brazilian native and naturalized cotton. http://www.cnpa.embrapa.br/albrana. Accessed 10 Feb 2013

Almeida VC, Hoffmann LV, Yokomizo GK, Costa JN, Giband M, Barroso PAV (2009) In situ and genetic characterization of *Gossypium barbadense* L. from the States of Pará and Amapá, Brazil. Pesq Agropec Bras 44:719–725. doi:10.1590/S0100-204X2009000700011

Alves MF, Barroso PAV, Ciampi AY, Hoffmann LV, Azevedo VCR, Cavalcante U (2013) Diversity and genetic structure among subpopulations of *Gossypium mustelinum* (Malvaceae). Genet Mol Res 12:597–609. doi:10.4238/2013.February.27.9

Barroso PAV, Hoffmann LV, Alves MF, Freitas RB (2007) Aplicabilidade de Marcadores SSR em Estudos Genéticos em *Gossypium mustelinum* Miers. Embrapa CNPA, Campina Grande. Available from Centro Nacional de Pesquisa do Algodão, Campina Grande, PB, Brazil. Publ. comunicado técnico no. 345

Barroso PAV, Hoffmann LV, Batista CE, Freitas RB, Alvez MF, Silva UC, de Andrade FP (2010) In situ conservation and genetic diversity of three populations of *Gossypium mustelinum* Miers (ex Watt). Genet Resour Crop Evol 57:343–349. doi:10.1007/s10722-009-9472-9

Bertini MCH, Schuster I, Sediyama T, Barros EG, Moreira MA (2006) Characterization and diversity analysis of cotton cultivars using microsatellites. Genet Mol Biol 29:321–329

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolution with polymorphic microsatellites. Nature 368:455–457. doi:10.1038/368455a0

Castelletti CHM, Santos AMM, Tbarelli M, Siva JMC (2003) Quanto ainda resta da Caatinga? Uma estimativa preliminar. In: Leal IR, Tabarelli M, Silva JMC (eds) Ecologia e conservação da caatinga. Universitária da UFPE, Recife, pp 719–734

CMD (2013) Cotton data marker base. http://www.cottonmarker.org. Accessed 1 Jan 2013

DArT (Diversity Arrays Technology) (2009) Protocolo de extração de DNA de plantas. http://www.DiversityArrays.com/pub/DArT_DNA_isolation.pdf. Accessed 10 Jan 2009

de Menezes IPP, Barroso PAV, Hoffmann LV, Lucena VS, 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. doi:10.1139/B10-045

Ellstrand NC, Elam DR (1994) Population genetic consequences of small population size: implications for plant conservation. Annu Rev Ecol Syst 24:217–242. doi:10.1146/annurev.es.24.110193.001245

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620. doi:10.1111/j.1365-294X.2005.02553

Freire EC (2002) Viabilidade de cruzamentos entre algodoeiros transgênicos e comerciais e silvestres do Brasil. *Rev bras ol fibros* 1:465–470. http://www.cnpa.embrapa.br/ojs/index.php/RBOF/article/view/132

Gao H, Williamson S, Bustamante CD (2007) An MCMC approach for joint inference population structure and inbreeding rates from multi-locus genotype data. Genetics 176:1635–1651. doi:10.1534/genetics.107.072371

Gardunia BW (2006) Introgression from *Gossypium mustelinum* and *Gossypium tomentosum* into upland cotton, *Gossypium hirsutum*. Thesis, Texas A&M University

Goudet J (2001) FSTAT: program to estimate and test gene diversities and fixation indices (software). Version 2.9.3. http://www2.unil.ch/popgen/softwares/fstat.htm. Accessed 20 Dec 2009

Hague SS, Smith CW, Stelly DM, Souder C, Thaxton P (2007) Combining ability and utility of interspecific *Gossypium* matings. International Cotton Advisory Committee. http://wcrc.confex.com/wcrc/2007/techprogram/MEETING.HTM. Accessed 1 June 2009

Lewis P, Zaykin D (2001) Genetic data analysis: computer program for the analyses of allelic data. Version 1.1 (win32). http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php. Accessed 20 Dec 2009

McGregor SE (1976) Insect pollination of cultivated crop plants. U.S.D.A. Agriculture Handbook No. 496

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590

Neves OS, Cavaleri PA, Gridi-Papp IL, Fuzzato MG (1965) Algodoeiro selvagem nordeste do Brasil. Bragantia 24:19–25

Neves OS, Cavaleri PA, Ferraz CAM, Fuzzato MG, Silva NM, Schmidt W, Corrêa DM (1968) Distribuição geográfica atual dos algodoeiros perenes no Brasil. Bragantia 35:437–475

Pereira GS, Sousa RL, Araújo RL, Silva EF, Hoffmann LV, Barroso PAV (2012) Selective fertilization interspecific crosses of allotetraploid species *Gossypium*. Botany 90:159–166. doi:10.1139/b11-094

Pickersgill B, Barrett SCH, Lima AD (1975) Wild cotton in northeast Brazil. Biotropica 7:42–54

Prado DE (2003) As caatingas da America do Sul. In: Leal IR, Tabarelli M, Silva JMC (eds) Ecologia e conservação da caatinga. Universitária da UFPE, Recife, pp 3–74

Reed HD, Frankham R (2003) Correlation between fitness and genetic diversity. Conserv Biol 1:230–237. doi:10.1046/j.1523-1739.2003.01236

Ritland K (2002) Extensions of models for the estimation of mating systems using an independent loci. Heredity 88:221–228. doi:10.1038/sj.hdy.6800029

Schuster I, Queiroz VT, Teixeira AI, Barros EG, Moreira MA (2004) Determinação da pureza varietal de sementes de soja com o auxílio de marcadores moleculares microssatélites. Pesquisa Agropecuaria Brasileira 39:247–253

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138. doi:10.1046/j.1471-8286.2003.00566

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) Mega5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739. doi:10.1093/molbev/msr121

Vaz ARC, Borba TCO, Brondani C, Rangel PHN, Camargo GSO, Telles MPC, Diniz Filho JAF, Brondani RPV (2009) Genetic analysis of a local population of *Oryza glumaepatula* using SSR markers: implications for management and conservation programs. Genetica 137:221–231. doi:10.1007/s10709-009-9393-8

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370. doi:10.2307/2408641

Wendel JF, Brubaker CL, Perciva IAE (1992) Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. Am J Bot 79:1291–1310

Wendel JF, Rowley R, Stewart JMD (1994) Genetic diversity in and phylogenetic relationships of the Brazilian endemic cotton, *Gossypium mustelinum* (Malvaceae). Plant Syst Evol 192:49–59. doi:10.1007/bf00985907

Westengen OT, Huamán Z, Heun M (2005) Genetic diversity and geographic pattern in early South American cotton domestication. Theor Appl Genet 110:392–402. doi:10.1007/s00122-004-1850-2