

# Diversity of wild and cultivated pearl millet accessions (*Pennisetum glaucum* [L.] R. Br.) in Niger assessed by microsatellite markers

Cedric Mariac · Viviane Luong · Issoufou Kapran · Aïssata Mamadou ·  
Fabrice Sagnard · Monique Deu · Jacques Chantereau · Bruno Gerard ·  
Jupiter Ndjunga · Gilles Bezançon · Jean-Louis Pham · Yves Vigouroux

Received: 14 March 2006 / Accepted: 6 September 2006 / Published online: 18 October 2006  
© Springer-Verlag 2006

**Abstract** Genetic diversity of crop species in sub-Saharan Africa is still poorly documented. Among such crops, pearl millet is one of the most important staple species. In Niger, pearl millet covers more than 65% of the total cultivated area. Analyzing pearl millet genetic diversity, its origin and its dynamics is important for in situ and ex situ germplasm conservation and to increase knowledge useful for breeding programs. We developed new genetic markers and a high-throughput technique for the genetic analysis of pearl millet. Using 25 microsatellite markers, we analyzed genetic diversity in 46 wild and 421 cultivated acces-

sions of pearl millet in Niger. We showed a significantly lower number of alleles and lower gene diversity in cultivated pearl millet accessions than in wild accessions. This result contrasts with a previous study using iso-enzyme markers showing similar genetic diversity between cultivated and wild pearl millet populations. We found a strong differentiation between the cultivated and wild groups in Niger. Analyses of introgressions between cultivated and wild accessions showed modest but statistically supported evidence of introgressions. Wild accessions in the central region of Niger showed introgressions of cultivated alleles. Accessions of cultivated pearl millet showed introgressions of wild alleles in the western, central, and eastern parts of Niger.

Cedric Mariac and Viviane Luong have contributed equally to this work.

**Electronic supplementary material** Supplementary material is available in the online version of this article at <http://dx.doi.org/10.1007/s00122-006-0409-9> and is accessible for authorized users.

C. Mariac · V. Luong · J.-L. Pham · Y. Vigouroux (✉)  
Institut de Recherche pour le Développement (IRD),  
911, avenue Agropolis, BP 64501, 34394 Montpellier, France  
e-mail: yves.vigouroux@mpl.ird.fr

I. Kapran · A. Mamadou  
Institut National de la Recherche Agronomique du Niger  
(INRAN), Niamey, Niger

F. Sagnard · M. Deu · J. Chantereau  
Centre International de la Recherche Agronomique pour le  
Développement (CIRAD), Montpellier, France

B. Gerard · J. Ndjunga  
International Center of Research for the Semi-Arid tropics  
(ICRISAT), Niamey, Niger

G. Bezançon  
Institut de Recherche pour le Développement (IRD),  
Niamey, Niger

## Introduction

In Sahelian Africa, numerous traditional crops contribute to food security. Studies of the crops cultivated in these regions are relatively rare, particularly in terms of their genetic diversity. Pearl millet (*Pennisetum glaucum* [L.] R. Br.) is one of the most important crops in the whole Sahelian region from Senegal to Sudan. Pearl millet was domesticated in the Sahelian zone of western Africa (Harlan et al. 1976; Tostain 1992). In Niger, land under pearl millet represents more than 65% of a total of 7.5 millions ha of cultivated land (estimated for the 1995–1999 period, Guenguant and Banoin 2003). Studying the diversity of such important crops enables identification of landmarks for in situ germplasm conservation, the creation of core collections of accessions for genetic analysis

and the extension of knowledge useful for breeding programs. To date, the diversity of pearl millet has been studied using iso-enzyme loci (Tostain et al. 1987; Tostain and Marchais 1989; Tostain 1992, 1994), AFLP markers (Vom Brocke et al. 2003), and RFLP markers (Bhattacharjee et al. 2002). New markers such as SSCP-SNP (Bertin et al. 2005) and microsatellite loci (Allouis et al. 2000; Qi et al. 2001, 2004; Budak et al. 2003) have recently been developed. However, they have not yet been used to assess the genetic diversity of landraces of pearl millet. Microsatellite markers are a promising tool for in-depth investigations of genetic diversity of pearl millet. Consequently, we developed a new set of microsatellite markers and a high-throughput methodology for microsatellite genotyping. Using these new methods, we analyzed the genetic diversity of a large sample of wild and cultivated accessions.

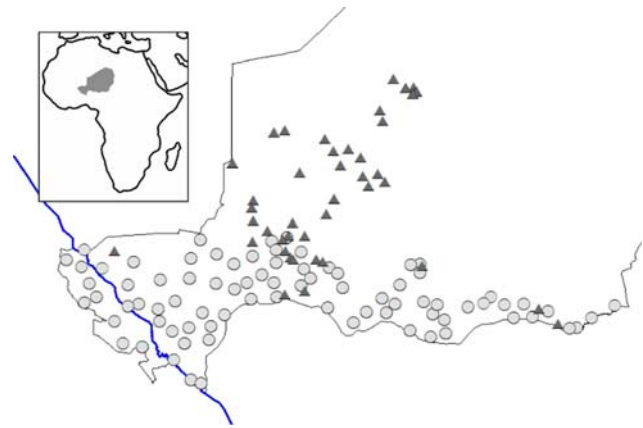
Our genetic analysis focused on Niger, the second largest pearl millet producer in Africa after Nigeria. The morphological diversity of pearl millet in Niger is the highest in West Africa (Tostain 1994). In particular, spike morphology exhibits wide variation from a very short spike in the eastern part of the country to a very long spike in the South-central part of Niger. Moreover in Niger, both wild millet (*Pennisetum glaucum* ssp. *monodii*) and cultivated pearl millet (*Pennisetum glaucum* ssp. *glaucum*) are found. Wild populations grow at latitudes between 12°N and 21°N but are found mainly in the northern part of the country (Tostain 1992) in the Air mountains. However, some wild populations have also been described in sympatry with pearl millet landraces. This situation was documented mainly near the northern limits of cultivation of pearl millet.

The objectives of this study were: (1) to develop new markers and a high-throughput method for genotyping, (2) to investigate the diversity of wild and cultivated accessions, and (3) to study introgressions between cultivated and wild pearl millet.

## Materials and methods

### Seed collection and DNA extraction

In 2003, samples were collected in 80 different villages. A total of 421 different cultivated seed accessions were collected, corresponding to about 140 different landrace names. An accession consisted of a quantity of panicles or seeds of a named variety, provided by a single farmer in one village. Sampling was conducted throughout the cultivated area of Niger (Fig. 1, S1). The area in which pearl millet is cultivated is limited by



**Fig. 1** Sampling locations of wild and cultivated pearl millet accessions. The sampling location of the 46 wild accessions (dark triangles) and 421 cultivated accessions (light gray circles). Different cultivated accessions were collected from the 80 sampling sites and are represented by a single light gray circle

rainfall, and most of northern and central Niger is not cultivated, but is used as pasture during the rainy season. In each sampling location, we collected panicles that farmers identified as their varieties. Our objective was to sample 30 panicles per variety, but this number varied depending on local availability. An average number of 21 panicles per sample was collected. We also analyzed 46 previously sampled wild accessions of pearl millet from Niger (Fig. 1, S1, Tostain 1992). For each wild and cultivated accession, one individual was studied. A total of 467 individuals was analyzed using 25 microsatellite loci.

### Microsatellite isolation

We developed a new set of microsatellite markers using public EST available in GenBank. Briefly, genomic DNA sequences were retrieved from GenBank using the name pearl millet or *Pennisetum glaucum* as query. On September 24, 2003, 2,577 sequences were retrieved. Microsatellite markers were identified using the software SSRI (<http://www.gramene.org/gramene/searches/ssrtool>). We searched for microsatellites exhibiting repeated motifs from 2 to 10 bases and at least five repeats. A total of 266 microsatellite hits were found. Some of them corresponded to a same sequence containing an interrupted microsatellite loci. Some sequences were redundant as they corresponded to the same gene. To analyze only unique ESTs, we compared each sequence with the other pearl millet ESTs using Blast software. We then selected only unique sequences and their unique microsatellite loci. The diversity of the newly developed markers was then assessed in ten wild, 12

weedy, and ten cultivated millet samples collected in Niger.

#### DNA extraction and PCR conditions

Samples of fresh leaves were harvested and ground in nitrogen. Approximately 0.2 g of powder were re-suspended with 700 µl extraction buffer (Tris 0.1 M, NaCl 1.25 M, EDTA 0.02 M, MATAB DTT 0.01 mM, PH 8) and incubated at 65°C for 4 h. Lysat was then mixed with chloroform-isoamyl alcohol (24:1), and then centrifuged. DNA was precipitated from the supernatant using isopropanol and washed with 70% ethanol. Dried pellets were re-suspended with 200 µl deionized water.

The PCR reaction mixture (11 µl final) consisted of 1× Colorless GoTaq<sup>TM</sup> Reaction Buffer (Promega M7921, Madison, WI, USA), 0.5 mM MgCl<sub>2</sub>, nucleotides dATP, dGTP, dCGT, dTTP (125 µM each), 0.1 µM primer, 1 unit of taq DNA polymerase, and 20 ng template DNA. The 5' end of the forward primer was labeled with fluorescent dye (Table 2).

The DNA and reaction mixture were dispensed using a HAMILTON starlab robot in a 384-well microtitre plate (Abgene, Epsom, UK). Plates were sealed with thermoseal Easy Peel (ABgene-0745). Silicone mats were added between plate and lid (110°C) to homogenize pressure and to limit evaporation loss. Amplifications were performed in a Biometra T1 384-well thermocycler programed for 35 cycles of 30 s at 94°C, 30 s at 55–58°C, 45 s at 72°C, and ending with 10 min at 72°C.

One microliter of 30-fold diluted amplification product and 11 µl HD formamide were mixed with 0.15 µl GS 500 Liz internal size standard and heated at 94°C for 3 min. Migration was performed using an automatic sequencer ABI Prism<sup>TM</sup> 3100 (Applied Biosystems, Foster City, CA, USA). Microsatellite alleles were scored using Genescan and Genotyper software packages (Applied Biosystems). The scoring was manually checked by two different persons. Each 384-well PCR plate included eight negative controls (no DNA).

To develop a set of markers for high-throughput genotyping, we used some of the microsatellites that we developed and also some markers developed in previous publications (Allouis et al. 2000; Qi et al. 2001, 2004; Budak et al. 2003). Our objective was to develop multiplex PCR associations of microsatellite loci and multiplex migration on ABI prism. We tested a total set of 65 microsatellite markers including the ones we actually developed. We retained only the best markers showing good yield amplification and with no surnumary bands. Surnumary bands correspond to loci

showing more than two alleles certainly link to duplicate loci or EST belonging to multigene families. Finally, we selected the markers, which could be combined for the development of multiplex PCR and migration. We ended up with a set of 25 markers including eight markers developed in this study.

#### Statistical analysis

Genetic data analysis was performed using Powermarker (Liu and Muse 2005). We calculated the number of alleles, observed heterozygosity, gene diversity, polymorphic information content (PIC) and differentiation (Fst). The gene diversity was calculated as  $n/(n-1) (1 - \sum p_i^2 - H_o/2n)$ , where  $n$  is the number of individuals,  $p_i$  the frequency of the  $i$  allele and  $H_o$  the number of observed heterozygotes (Nei 1987). The PIC was calculated as  $1 - \sum p_i^2 - \sum_i \sum_{j>i} 2p_i^2 p_j^2$ , where  $p_i$  and  $p_j$  are the frequencies of the  $i$  and  $j$  alleles, respectively (Botstein et al. 1980). The differentiation between the wild and cultivated groups was tested locus by locus, and overall significance was also tested. Sample size has a large impact on the estimation of the number of alleles. We compared the number of alleles for a same sample size using the parameter allelic richness (FSTAT software, Goudet 2001). The allelic richness and gene diversity of two samples (wild and cultivated) were compared using a Wilcoxon paired test. A 1,000 bootstraps were performed to calculate the 95% confidence interval (CI) of the average allelic richness and gene diversity across loci.

We calculated the Shared Allele Distance between individuals using Powermarker. We used this matrix to statistically assess the correlation (1) between genetic distance and landrace name and (2) between genetic distance and geographical distance. We built a matrix where the geographical distance between accessions was calculated using latitude and longitude data (S1). To test whether geographical distance and genetic distance were correlated, we performed a Mantel test (Sokal and Rohlf 1995). To analyze the correlation between genetic distance and landrace name, we built a matrix as follows: if two accessions shared the same landrace name, the distance was 0, otherwise it was 1. The correlation was also tested by a Mantel test.

For each individual, we calculated the frequencies of each allele (0, 0.5, and 1) at each locus, and used this data to perform a principal component analysis (PCA) using SYSTAT.

We used a Bayesian method to determine the presence of hybrids or introgressed wild or cultivated individuals. The software Structure Version 2.1

(Pritchard et al. 2000; Falush et al. 2003) was used to perform this analysis. Parameters were set at  $K = 2$  for the number of populations, 100,000 for the burn-in time and 1,000,000 for the number of runs. Five replicates were performed. The output of this analysis is the ancestry of the two different groups: cultivated and wild groups. The ancestry value is a statistical estimation of the proportion of the genome of an individual that originated from a given population. The ancestry value varies from 0 to 1. An ancestry close to 0 or 1 in one group suggests no evidence of introgression for the individual studied. Intermediate values suggest introgression. For each individual we calculated a CI of the ancestry parameter. We also performed the same analysis using different numbers of populations ( $K$  varying from 1 to 7). Five repetitions of each assumed population number were performed. A recent simulation study proposed a methodology to assess the best  $K$ -value supported by the data (Evanno et al. 2005).

Following Evanno et al. (2005), we calculated the second order change of the likelihood function divided by the standard deviation of the likelihood ( $\Delta K$ ).

## Results

### Isolation of microsatellite loci

A total of 207 unique microsatellite loci were found in the EST sequences screened. Thus, 8% of the 2,577 sequences analyzed exhibited microsatellite loci of the minimum size specified. Di-nucleotide microsatellites are the most abundant, 169 loci were identified. We also found 32 tri-nucleotide microsatellites, five tetra-nucleotide microsatellites and only one deca-nucleotide microsatellite. The mean size of the microsatellite alleles was 9.0 repeats. We designed 58 different pairs of primers and tested them for the presence of a simple

**Table 1** List of EST derived microsatellite loci

Locus	GenBank accession	Primer sequence 5'–3'	Core motif	Number of alleles	PIC
PGIRD5	CD724362-1	CAACCCAACCCATTATACTTATCTG GCAACTCTTGCCTTTCTTGG	(GA) <sub>7</sub>	–	–
PGIRD7	CD724492-1	CGGAGACGCACTAGACTTGG CCGGATGCTCACTTCCTTAT	(GT) <sub>7</sub>	4	0.23
PGIRD12	CD724749-1	ACTCGTTCCGATGCACCTCT CGGGGAAGAGACAGGCTACT	(TA) <sub>8</sub>	8	0.44
PGIRD13	CD724750-1	CAGCAGCGAGAAGTTTAGCA GCGTAGACGGCGTAGATGAT	(AGC) <sub>8</sub>	10	0.58
PGIRD19	CD724869-1	TGAGGACCGAGAAGAAGCAT CAACACCCAACAGAACTGAA	(GT) <sub>10</sub>	6	0.44
PGIRD21	CD724961.1	GCTATTGCCACTGCTTCACA CCACCATGCAACAGCAATAA	(ACC) <sub>8</sub>	3	0.21
PGIRD25	CD725199.1	CGGAGCTCCTATCATTCCAA GCAAGCCACAAGCCTATCTC	(GA) <sub>9</sub>	9	0.53
PGIRD43	CD724428.1	GTTTCATGCAGCTTGGTTTCC AGTGACCTGGGGTACAGACG	(GAT) <sub>6</sub>	10	0.68
PGIRD44	CD725489.1	TCTCTCTCGGATCGCTGTG GCTGGTTGGTAGAGGCTGAC	(GCG) <sub>6</sub>	5	0.41
PGIRD46	CD724372.1	GAACAATTGCTTCTGTAATATTGCTT GCCGACCAAGAACTTCATACA	(CTC) <sub>6</sub>	3	0.24
PGIRD49	CD725305.1	AGCTCCTCGACGGAGAAAGT GACGGTGTGACGAAGATG	(CGG) <sub>6</sub>	5	0.48
PGIRD50	CD724894.1	CTCTCGGTTTGACGGTTTGT GGGGAAAACAAAGTTGCTCA	(TGT) <sub>6</sub>	7	0.58
PGIRD54	CD724361.1	GCCTGGGATGTGTTTCTTCT GCCTTTCATTTCCACCATGA	(GT) <sub>5</sub>	2	0.37
PGIRD55	CD725120.1	CTTTACTACGGCCACGACA GTGTGTTTGTACCGGTGTGG	(AC) <sub>6</sub>	18	0.86
PGIRD56	CD725501.1	ATCACTCCTCGATCGGTCAC ACCAGACACACGTGCCAGT	(TG) <sub>6</sub>	2	0.06
PGIRD57	CD726121.1	GGCCCCAAGTAACCTCCCTA TCAAGCTAGGGCCAATGTCT	(AG) <sub>7</sub>	11	0.67

For each microsatellite locus, its name, the accession number of the EST, the forward and reverse primer sequences, and the core motif of the microsatellite are presented. The number of alleles and the PIC value was estimated on a sample of 32 individual plants. For PGIRD5, polymorphism was detected but high slippage was observed at this loci and the number of alleles and PIC value was not estimated

pattern, reliable amplification and the presence of polymorphism between pearl millet accessions. A new set of 16 microsatellites was developed (Table 1). The diversity of these microsatellite loci varied from 2 to 18 for the number of alleles and 0.06–0.86 for the PIC values. One of the polymorphic locus genetic diversity parameters was not estimated because high slippage rendered diversity estimation difficult (PGIRD5).

We amplified 25 loci with 15 PCR reactions. The number of amplified microsatellite loci in a single PCR reaction varied from 4 to 1 (Table 2). The 25 microsatellite loci were combined so that an average of six microsatellite loci migrate together. PCR and migration multiplexes of microsatellite loci are presented in Table 2.

#### Diversity and differentiation between cultivated and wild pearl millet samples

To compare the number of alleles in the cultivated sample with the number of alleles in the wild sample, the effect of the sample size needs to be taken into account. We corrected the effect of the sample size by calculating allelic richness, which was calculated on the smaller size of the two samples (Goudet 2001). We found significantly lower allelic richness (Table 3) in the cultivated sample than in the wild sample (Wilcoxon test,  $Z = 3.68$ ,  $P < 0.001$ ). The average allelic richness for the cultivated sample was 6.2 compared with 8.1 for the wild sample. The 95% CI of the mean value across loci is 4.5–8.1 alleles for the cultivated sample and 6.1–10.2 alleles for the wild sample. The cultivated sample had 23% fewer alleles than the wild sample. The average  $F_{is}$  value was 0.30 for the wild sample ( $P < 0.001$ ) and 0.19 for the cultivated sample ( $P < 0.001$ ). Gene diversity was significantly lower (Wilcoxon test,  $Z = 3.29$ ,  $P < 0.001$ ) in the cultivated sample than in the wild sample. The cultivated sample showed an average gene diversity of 0.49 (CI 0.39–0.59)

compared with 0.67 (CI 0.57–0.75) for the wild sample. The cultivated sample thus showed a gene diversity that was 28% lower than in the wild sample.

Overall, the differentiation ( $F_{st}$ ) between the wild and cultivated samples was highly significant ( $P < 0.001$ ) with an average value of 0.17. This differentiation varied from 0.019 to 0.49 depending on the loci studied and was significant for each locus ( $P < 0.05$ ). PCA (Fig. 2) enabled us to explain 4.2% of the overall variation on the first axis and 1.6% on the second one. The low percentage of explained variance on the two axes is common in analyses using a high number of alleles from different microsatellite loci (Matsuoka et al. 2002). The marked differentiation between wild and cultivated samples is clear on the first axis of the PCA. The second axis differentiates different accessions of wild pearl millet.

#### Landrace names, geographical, and genetic distances

For cultivated accessions we analyzed the relationship between genetic distance and geographical distance. For this purpose, we calculated a matrix of geographical distance and a matrix of genetic distance between accessions. The relationship between these two matrices was studied using a Mantel test. There was a low correlation between genetic distance and geographical distance ( $R = 0.11$ ,  $P < 0.001$ ). We also tested the genetic proximity of accessions belonging to the same landrace. We found a very weak correlation ( $R = 0.026$ ,  $P < 0.01$ ).

#### Introgressions

We performed a Bayesian analysis to detect evidence of introgression between cultivated and wild accessions. For this purpose, we used Structure software assuming two groups ( $K = 2$ ): a cultivated and a wild

**Table 2** PCR condition and migration

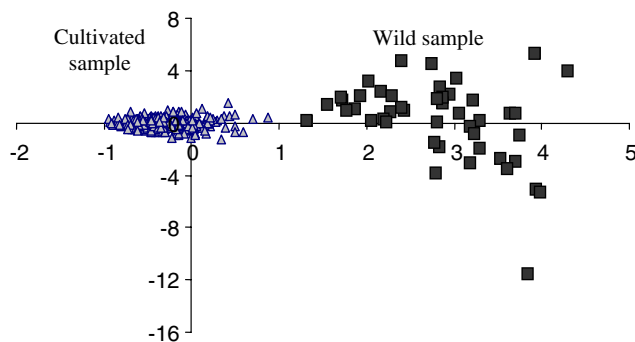
The 5' fluorescent-labels used with forward primers are NED<sup>a</sup>, VIC<sup>b</sup>, 6-FAM<sup>c</sup>, and PET<sup>d</sup>. The temperature of hybridization is given for each combination of loci. Loci from different PCR reactions were pooled and alleles were separated using a capillary ABI Prism 3100 system

Locus combination multiplexed in a single PCR	Temperature (°C)	Migration pool number
PGIRD46 <sup>a</sup> , CTM8 <sup>b</sup> , PSMP2249 <sup>c</sup> , PSMP2247 <sup>c</sup>	55	1
PSMP2201 <sup>d</sup> , PSMP2237 <sup>d</sup>	55	
PSMP2227 <sup>a</sup> , PGIRD12 <sup>b</sup>	55	
PSMP2266 <sup>c</sup> , PGIRD21 <sup>c</sup>	55	2
PSMP2208 <sup>a</sup> , PSMP2085 <sup>a</sup>	58	
PGIRD44 <sup>b</sup> , PGIRD 13 <sup>b</sup>	58	
PSMP73 <sup>d</sup>	58	3
PSMP32 <sup>b</sup>	55	
CTM10 <sup>b</sup> , PGIRD43 <sup>c</sup>	55	
PSMP2214 <sup>d</sup>	55	4
PGIRD7 <sup>d</sup>	55	
PGIRD25 <sup>d</sup> , PSMP2219 <sup>a</sup> , PSMP2220 <sup>a</sup>	55	
PSMP2206 <sup>c</sup> , PSMP2202 <sup>b</sup>	55	



**Table 3** Diversity of the wild and cultivated samples

Marker	Allelic richness		Gene diversity	
	Cultivated sample	Wild sample	Cultivated sample	Wild sample
PSMP2085	5.2	7.0	0.52	0.83
PSMP2201	3.5	4.0	0.30	0.71
PSMP2202	2.0	2.0	0.49	0.18
PSMP2206	6.5	6.9	0.71	0.74
PSMP2208	9.0	15.4	0.53	0.88
PSMP2214	4.2	4.9	0.72	0.63
PSMP2219	5.2	5.0	0.41	0.72
PSMP2220	8.7	12.8	0.75	0.88
PSMP2227	4.1	3.9	0.17	0.27
PSMP2237	5.2	8.0	0.67	0.79
PSMP2247	4.8	7.5	0.18	0.72
PSMP2249	3.3	4.0	0.50	0.18
PSMP2266	7.8	15.7	0.77	0.92
CTM10	15.9	17.5	0.91	0.92
CTM8	4.3	5.7	0.31	0.64
PGIRD12	6.3	8.0	0.50	0.75
PGIRD13	4.7	6.8	0.49	0.67
PGIRD21	2.2	3.0	0.06	0.36
PGIRD25	4.6	8.6	0.49	0.75
PGIRD43	3.7	5.9	0.48	0.65
PGIRD44	3.8	3.0	0.22	0.65
PGIRD46	3.6	4.0	0.19	0.61
PGIRD7	3.1	5.9	0.08	0.46
PSMP32	15.0	18.7	0.89	0.87
PSMP73	19.4	18.3	0.88	0.87
Mean	6.2 (4.5–8.1)	8.1 (6.1–10.2)	0.49 (0.39–0.59)	0.67 (0.57–0.75)

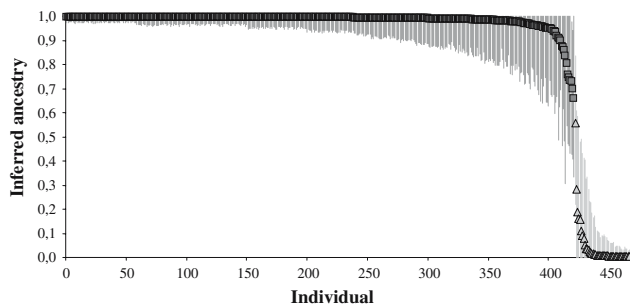
**Fig. 2** Principal component analysis of the wild and cultivated accessions. The first axis explains 4.2% of the variance and the second axis 1.6%. A clear distinction is apparent between the cultivated accession and the wild accessions on the first axis of the PCA. The second axis differentiates different wild accessions

group. Even if it is perfectly biologically sound, the choice of a fixed number of populations is arbitrary. Consequently, we also performed this analysis where  $K$  could vary from 1 to 7. This analysis showed a large increase of likelihood from  $K = 1$  to 2, a smaller increase from  $K = 2$  to 3 after which likelihood leveled off from  $K = 4$  to 7. The methodology of Evanno et al. (2005) strongly supported  $K = 2$  as the best number of groups. However, the analysis of introgression between

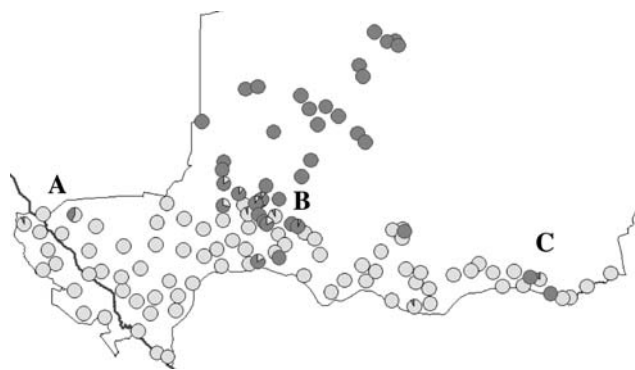
the cultivated and wild group was also performed for  $K = 3$  and results were similar to  $K = 2$ .

Using  $K = 2$ , ancestry and its CI were estimated for the 421 cultivated and the 46 wild accessions. The ancestry estimated for an individual in the two groups sums to 1. Consequently, we report ancestry in the cultivated group (Fig. 3). The vast majority of cultivated plants showed strong ancestry higher than 95% in the cultivated cluster. Only six cultivated plants (1.4%) showed ancestry lower than 80%: individual M395 (Mean ancestry 0.76 and CI: 0.47–1.0), individual M043 (Mean ancestry 0.75 and CI: 0.46–1.0), individual M413 (Mean ancestry 0.74 and CI: 0.43–1.0), individual M005 (Mean ancestry 0.66 and CI: 0.36–1.0), individuals M357 (Mean ancestry 0.73 and CI: 0.43–0.98), and individual M234 (Mean ancestry 0.70 and CI: 0.41–0.94). Most of the wild accessions showed very low ancestry ( $< 5\%$ ) in the cultivated cluster (Fig. 3), and consequently high ancestry in the wild cluster. Two wild plants (4.3%) showed ancestry higher than 20% in the cultivated cluster: individual PE08082 (Mean ancestry 0.56 and CI: 0.22–0.84) and individual PE08112 (Mean ancestry 0.28 and CI: 0.0–0.61).

To analyze the relationship between ancestry estimates and geography, we plotted the ancestry values on a map (Fig. 4). The ancestry at a given sampling site



**Fig. 3** Individual ancestry in the cultivated group for each of the wild and cultivated accessions. A Bayesian analysis was performed to detect ancestry of each individual in the cultivated and wild sample (see text for details). The ancestry estimated in the cultivated sample is presented for the 424 different cultivated accessions (*square*) and the 46 wild accessions (*triangle*). The 95% CI was estimated for the ancestry coefficient and plotted. Intermediate ancestries are found for some cultivated and wild accessions suggesting hybridization between the two groups



**Fig. 4** Geographical projection of ancestry estimates. The ancestry estimated using a Bayesian method was calculated for each individual. The average ancestry for each sampling location was also calculated (see text for details). Ancestries were then plotted on a map to analyze regional patterns

was calculated as the mean of the ancestry of the accessions sampled at this site. Average ancestries higher than 95% in the cultivated pool were set at 100%. Each sampling location is represented by the ancestry in the cultivated group (light gray) and in the wild group (dark gray). Both ancestry values sum to 1 and are represented by a pie chart.

The geographical locations of cultivated introgressed individuals are in the region of Ayrourou (A, Fig. 4) and Tahoua (B, Fig. 4). Slight traces of introgression were also detected in the eastern part of Niger (C, Fig. 4).

## Discussion

Our study aims to compare diversity between wild and cultivated accessions and study introgressions between

cultivated and wild pearl millet. We choose to maximize the number of different accessions by sampling one individual per accession. This approach permits to maximize diversity inside each group and have been previously used to study crops and their wild relatives (Matsuoka et al. 2002; Fukunaga et al. 2005). This approach is complementary to frequency-based approach using allele frequency of accessions.

### A high genetic diversity in wild populations

Our results with microsatellite loci show a highly significant difference in diversity between wild and cultivated samples. Cultivated samples presented a much lower level of diversity than the wild sample. A previous study in West Africa using iso-enzyme analysis (Tostain 1992) found no difference in diversity between wild and cultivated samples. Using iso-enzymes, Tostain (1992) found a higher level of diversity in cultivated samples for some loci compared with those of wild accessions (Pgi A, Pgm A), a lower level of diversity for other loci (Got A, Pgd A, Cat A) and sometimes no difference at all (Adh A, Est A). In our study, the vast majority of microsatellite loci (20/25) displayed higher allelic richness in the wild sample than in the cultivated sample. Microsatellite loci have the particularity to display high mutation rate (Vigouroux et al. 2002). One consequence could be, as it is observed in maize (Vigouroux et al. 2005), that microsatellite loci in the cultivated group would recover diversity quicker after a domestication bottleneck than other types of markers displaying lower mutation rates. In which case, one would expect a smaller difference in diversity between wild and cultivated group when observed on microsatellite loci, than when observed on iso-enzymes. As we observe the contrary, we conclude that the high mutation rate of microsatellite loci does not explain the difference we observed. However, the very low variability of iso-enzyme markers may induce a bias in the estimation of the levels of diversity between wild and cultivated samples. Indeed, the levels of diversity in wild pearl millet have been studied using iso-enzymes previously identified as polymorphic in cultivated millet (Tostain 1992). This bias may level off the genetic difference between samples. Markers with a higher-genetic diversity should not present a similar bias.

Finally, another previous study found a significant correlation between the frequency of Pgm and Adh alleles and an environmental gradient (Leblanc and Pernes 1983) and suggested that selection may occur at these iso-enzyme loci in pearl millet. Selection at iso-enzyme loci has been detected in numerous

organisms (Endler 1986). It is not unlikely that this occurs in pearl millet and also that this phenomenon is responsible for leveling off the differences between wild and cultivated samples. However, we cannot exclude that some form of selection was underway for some of microsatellite loci we used. Another difference between our study and the previous ones may be the limited geographical origin of the cultivated and wild samples used. Our study was restricted to samples from Niger while the previous one used a wider sample (Tostain 1992). To explain the difference observed, the diversity of landraces from Niger should be the lowest among the cultivated accessions in western Africa and the diversity of wild accessions among the highest. Previous studies do not support these two facts (Tostain and Marchais 1989; Tostain 1992). Therefore, it would be surprising if the marked difference we found in terms of genetic diversity between wild and cultivated accessions was limited to pearl millet from Niger.

The average differentiation ( $G_{ST}$ ) estimated by Tostain (1992) between wild and cultivated samples in West Africa was 0.13. This value was calculated globally using wild and cultivated accessions from West Africa and is similar to the value we found in Niger. In a previous study (Tostain 1992) wild and cultivated accessions exhibited a similar distribution in the PCA analysis. In our PCA analysis, there was a marked variation between accessions of wild pearl millet: different accessions presented a unique allele combination. Compared with the variation in wild accessions, cultivated pearl millet is more homogenous. Microsatellites may be a more powerful tool to differentiate accessions within the wild and cultivated groups than iso-enzyme markers. Significant  $F_{IS}$  values were found. This result suggests the existence of a genetic structure within wild and cultivated samples. We actually observed a low correlation between genetic distance and geographical distance in the cultivated sample, suggesting a slight structuration. However, homogamy may also partly explain this finding. Homogamy has been observed in outbreeding crops in traditional agricultural settings (Pressoir and Berthaud 2004a) and has been proven to create positive  $F_{IS}$  values within fields. Further studies are needed to fully assess the role of homogamy and genetic structure in pearl millet.

Overall, these results show that wild populations may be an interesting source of new alleles and new allele combinations, which could be useful to broaden the genetic basis of cultivated accessions. Our study is however limited to Niger. A larger study is necessary to address this question on a broader geographical scale.

## A weak relationship between landrace name and genetic distance

We analyzed the relationship between genetic distance and landrace name for cultivated accessions. We found that accessions bearing the same landrace name were genetically closer than those having different names, but this correlation was very weak. The correlation was also low but somewhat higher if we considered the geographical distance between accessions. These results are in good agreement with a previous study on diversity between and within accessions (Busso et al. 2000). Landraces are identified by their morphological and phenological differences. However such morphological differences do not appear to result in large genetic differences (Busso et al. 2000). The landrace name seems to be a poor reflection of genetic distance between accessions. Even if the correlation is relatively low, geographical distance provides more information than landrace names in terms of genetic distance between accessions. In this context, a sampling strategy aimed at maximizing diversity should include the geographical distance between accessions as an important component of sampling. With this aim in mind, sampling different landraces from different geographical locations would certainly be the best strategy. High-morphological differentiation between landraces obtained from the same farmer and low-genetic differentiation between these landraces have also been observed in maize, another open-pollinated cereal (Pressoir and Berthaud 2004b). This study established that this pattern is explained by recurrent farmer selection for a given morphology in a context of high cross-pollination between landraces (Pressoir and Berthaud 2004b). A similar situation may also exist in pearl millet, which would explain the low-genetic differentiation between morphologically different landraces (Robert et al. 2002).

## Introgression between wild and cultivated plants

In the central zone of Tahoua or the western part of Niger, sympatric populations of wild and cultivated pearl millet are found (Tostain 1992). In this region, evidence of wild and cultivated introgressed individuals are found (A, Ayourou and B, Tahoua, Fig. 4), suggesting significant wild to cultivated and cultivated to wild gene flows. Indeed, some of the ancestry CI statistically support the occurrence of introgression in the cultivated group (individual M357, M234) and in the wild group (individual PE08082). Northern wild accessions in the Air mountains did not show any signs of significant introgression.



We found a higher percentage of wild plants introgressed by cultivated alleles (4.2%) than cultivated plants introgressed by wild alleles (1.4%). This result suggests an asymmetric gene flow from cultivated to wild populations. We should however, be cautious in interpreting such data since it is based on relatively few plants. A finer regional study would certainly enable us to evaluate introgression in these regions more precisely. A previous study on the diversity of wild and cultivated accessions (Tostain 1992) showed that sympatric wild populations are genetically closer to cultivated populations. Our data explained this finding by the occurrence of introgression between wild and cultivated populations.

A previous experimental study carried out in the Keita area, East of Tahoua, Niger (Marchais 1994), showed that wild/cultivated crosses are common in sympatric situations but also that wild populations maintain their genetic distinctiveness. In our dataset, some wild populations in the central and eastern parts of Niger seem to have maintained their genetic distinctness although they are in sympatry with large cultivated pearl millet populations. Phenology (Renno and Winkel 1996), pollen competition (Sarr et al. 1988; Robert et al. 1991), and reproductive barriers (Amoukou and Marchais 1993) could explain this result. However, we cannot exclude the possibility that some cultivated alleles were introgressed by numerous backcrosses in the wild populations but these backcrosses could not be detected with the number of microsatellites we used.

## Conclusions

We assessed the genetic diversity of wild and cultivated samples of pearl millet in Niger. To do so, we developed new microsatellite loci and high-throughput methods. We found significantly lower genetic diversity in the cultivated sample than in the wild sample in terms of the number of alleles and gene diversity. Differentiation is strong between the wild and cultivated compartments. We also found significant evidence of introgressions between cultivated and wild accessions in sympatric areas.

**Acknowledgments** We thank the different people who participated in the 2003 sampling and particularly Djibo Moussa, Moussa Tidjani, and H. Yahaya Bissala. We thank Ousmane Seyni, Sylvain Santoni for useful help during genotyping. We thank Serge Tostain for helpful discussions. The article was significantly improved by thorough comments of three reviewers. This work was supported by Institut de la Recherche pour le Développement (IRD) and partly by the Institut Français de la Biodiversité (IFB).

## References

- Allouis S, Qi X, Lindup S, Gale MD, Devos KM (2000) Construction of a BAC library of pearl millet, *Pennisetum glaucum*. Theor Appl Genet 102:1200–1205
- Amoukou AI, Marchais L (1993) Evidence of partial reproductive barrier between wild and cultivated pearl millets (*Pennisetum glaucum*). Euphytica 67:19–26
- Bertin I, Zhu JH, Gale MD (2005) SSCP-SNP in pearl millet—a new marker system for comparative genetics. Theor Appl Genet 110:1467–1472
- Bhattacharjee R, Bramel PJ, Hash CT, Kolesnikova-Allen MA, Khairwal IS (2002) Assessment of genetic diversity within and between pearl millet landraces. Theor Appl Genet 105:666–673
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331
- Budak H, Pedraza F, Cregan PB, Baenzinger PS, Dweikat I (2003) Development and utilization of SSRs to estimate the degree of genetic relationships in a collection of pearl millet germplasm. Crop Sci 43:2284–2290
- Busso CS, Devos KM, Ross G, Mortimore M, Adams WM, Ambrose MJ, Alldrick S, Gale M-D (2000) Genetic diversity within and among landraces of pearl millet (*Pennisetum glaucum*) under farmer management in West Africa. Genet Res Crop Evol 47:561–568
- Endler JA (1986) Natural selection in the wild. Princeton University Press, Princeton, NJ
- 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
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587
- Fukunaga K, Hill J, Vigouroux Y, Matsuoka Y, Sanchez GJ, Liu K, Buckler ES, Doebley J (2005) Genetic diversity and population structure of teosinte. Genetics 169:2241–2254
- Goudet J (2001) FSTAT, a program to estimate and test gene diversity and fixation indices (version 2.9.3). Available from <http://www.unil.ch/izea/software/fstat.html>. Updated from Goudet (1995)
- Guenguant J-P, Banoïn M (2003) Dynamique des populations, disponibilités des terres et adaptation des régimes fonciers: le cas du Niger. CIRCRED-FAO, Paris, 157p
- Harlan J, De Wet J, Stemler A (1976) Origins of African plant domestication. Mouton Publishers, La Hague
- Leblanc JM, Pernes J (1983) Enzyme polymorphism of *Pennisetum americanum* in the Ivory Coast. Jpn J Genet 58:121–131
- Liu K, Muse S (2005) PowerMarker: new genetic data analysis software. Version 3.23. Free program distributed by the author over the internet from <http://www.powermarker.net>
- Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York
- Marchais L (1994) Wild pearl millet (*Pennisetum glaucum*, Poaceae) integrity in agricultural Sahelian areas. An example from Keita (Niger). Pl Syst Evol 189:233–245
- Matsuoka Y, Vigouroux Y, Goodman M-M, Sanchez J, Buckler E, Doebley J (2002) A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99:6080–6084
- Pressoir G, Berthaud J (2004a) Patterns of population structure in maize landraces from the Central Valleys of Oaxaca in Mexico. Heredity 92:88–94

- Pressoir G, Berthaud J (2004b) Population structure and strong divergent selection shape phenotypic diversification in maize landraces. *Heredity* 92:95–101
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Qi X, Lindup S, Pittaway TS, Allouis S, Gale MD, Devos KM (2001) Development of simple sequence repeat markers from bacterial artificial chromosome without sub-cloning. *Biotechniques* 31:355–362
- Qi X, Pittaway TS, Lindup S, Liu H, Wateran E, Padi FK, Hash CT, Zhu J, Gale MD, Devos KM (2004) An integrated genetic map and new set of simple sequence repeat markers for pearl millet, *Pennisetum glaucum*. *Theor Appl Genet* 109:1485–1493
- Renno JF, Winkel T (1996) Phenology and reproductive effort of cultivated and wild forms of *Pennisetum glaucum* under experimental conditions in the Sahel: implications for the maintenance of polymorphism in the species. *Can J Bot* 74:959–964
- Robert T, Lespinasse R, Pernès J, Sarr A (1991) Gametophytic competition as influencing gene flow between wild and cultivated forms of pearl millet (*Pennisetum typhoides*). *Genome* 34:195–200
- Robert T, Luxereau A, Mariac C, Ali K, Allinne C, Bani J, Beidari Y, Bezançon G, Cayeux S, Couturon E, Dedieu V, Moussa D, Sadou M, Seydou M, Seyni O, Tidjani M, Sarr A (2002) Gestion de la diversité en milieu paysan: influence de facteurs anthropiques et des flux de gènes sur la variabilité génétique des formes cultivées et spontanées du mil (*Pennisetum glaucum* (L.) R. Br.) dans deux localités du Niger. BRG, Actes du 4ème Colloque National «Le patrimoine génétique: la diversité et la ressource». La Châtre, France
- Sarr A, Sandmeier M, Pernès J (1988) Gametophytic competition in pearl millet *Pennisetum typhoides* (Stapf. and Hubb.). *Genome* 30:924–929
- Sokal R-R, Rohlf F-J (1995) *Biometry: the principles and practice of statistics in biological research*, 3rd edn. W.H. Freeman and Co., New York
- Tostain S (1992) Enzyme diversity in pearl millet (*Pennisetum glaucum* L.) 3. Wild millet. *Theor Appl Genet* 83:733–742
- Tostain S (1994) Isoenzymatic classification of pearl millet (*Pennisetum glaucum*, *Poaceae*) landraces from Niger (West Africa). *Pl Syst Evol* 193:81–93
- Tostain S, Marchais L (1989) Enzyme diversity in pearl millet (*Pennisetum glaucum*). 2. Africa and India. *Theor Appl Genet* 77:634–640
- Tostain S, Riandley MF, Marchais L (1987) Enzyme diversity in pearl millet (*Pennisetum glaucum*). 1. West Africa. *Theor Appl Genet* 74:188–193
- Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith JS, Doebley J (2002) Rate and pattern of mutation at microsatellite loci in maize. *Mol Biol Evol* 19:1251–1260
- Vigouroux Y, Mitchell S, Matsuoka Y, Hamblin M, Kresovich S, Smith JS, Jaqueth J, Smith OS, Doebley J (2005) An analysis of genetic diversity across the maize genome using microsatellites. *Genetics* 169:1617–1630
- Vom Brocke K, Christinck A, Weltzien ER, Prestlerl T, Geiger HH (2003) Farmers' seed systems and management practices determine pearl millet genetic diversity patterns in semi-arid regions of India. *Crop Sci* 43:1680–1689