Theor Appl Genet (2008) 117:489–497
DOI 10.1007/s00122-008-0793-4

ORIGINAL PAPER

Phylogeny and origin of pearl millet (*Pennisetum glaucum* [L.] R. Br) as revealed by microsatellite loci

Ibrahima Oumar · Cédric Mariac · Jean-Louis Pham · Yves Vigouroux

Received: 19 October 2007 / Accepted: 2 May 2008 / Published online: 27 May 2008
© Springer-Verlag 2008

Abstract During the last 12,000 years, different cultures around the world have domesticated cereal crops. Several studies investigated the evolutionary history and domestication of cereals such as wheat in the Middle East, rice in Asia or maize in America. The domestication process in Africa has led to the emergence of important cereal crops like pearl millet in Sahelian Africa. In this study, we used 27 microsatellite loci to analyze 84 wild accessions and 355 cultivated accessions originating from the whole pearl millet distribution area in Africa and Asia. We found significantly higher diversity in the wild pearl millet group. The cultivated pearl millet sample possessed 81% of the alleles and 83% of the genetic diversity of the wild pearl millet sample. Using Bayesian approaches, we identified intermediate genotypes between the cultivated and wild groups. We then analyzed the phylogenetic relationship among accessions not showing introgression and found that a monophyletic origin of cultivated pearl millet in West Africa is the most likely scenario supported by our data set.

Communicated by H. H. Geiger.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-008-0793-4) contains supplementary material, which is available to authorized users.

I. Oumar · C. Mariac · J.-L. Pham · Y. Vigouroux (✉)
IRD, UMR DIAPC Diversité et Adaptation des Plantes Cultivées (SupAgro, INRA, IRD, UMII) Institut de Recherche pour le Développement, 911, Avenue Agropolis, BP 64501, Montpellier 34394, France
e-mail: yves.vigouroux@mpl.ird.fr

I. Oumar
LMBA Laboratoire de Microorganisme et de Biomolécule Active Faculté des Sciences de Tunis, 2092 El-ManarII, Tunis, Tunisie

The development of different civilizations is tightly linked to the evolution of agriculture. Different centers of domestication have been described in Africa, Asia, the Middle East and America (Harlan 1975). Among the domesticated plants, cereals occupy a prominent place and contribute heavily to the human diet as an important source of starch. Studies on cereal crops evolution have shown that both multiple or single domestication events are possible (Second 1982; Nesbitt and Samuel 1996; Zohary and Hopf 2000; Badr et al. 2000; Salamini et al. 2002; Matsuoka et al. 2002; Londo et al. 2006). In Asian rice, *Oryza sativa*, at least two domestication events have taken place (Londo et al. 2006), one in the South of the Himalaya mountain region range (eastern India, Myanmar and Thailand) for *O. sativa* ssp. *indica* and the other in southern China for *O. sativa* ssp. *japonica*. A third geographical center of domestication for rice (*Oryza glaberrima*) was also found in Africa (Second 1982). The domestication of maize (*Zea mays*) in southern Mexico has been traced to a single center (Matsuoka et al. 2002). In the Middle East, wheat (*Triticum monococcum* and *Triticum dicoccoides*) and barley (*Hordeum vulgare*) have also been traced to a single domestication center (Nesbitt and Samuel 1996; Heun et al. 1997; Zohary and Hopf 2000; Badr et al. 2000; Ozkan et al. 2002). However, a more recent study on barley suggests two domestication events based on Bayesian analysis (Morrell and Clegg 2007).

Important cereals have been domesticated in Africa. Among them, pearl millet (*Pennisetum glaucum* [L.] R. Br.) is a major crop of the sub-Sahelian area from Senegal to Sudan, and it is cultivated in eastern Africa and Asia. The evolutionary history of pearl millet is not yet clearly established. The wild progenitor of pearl millet was identified as *Pennisetum glaucum* ssp. *monodii* (Harlan 1975; Brunken 1977). This species is found across the Sahel zone

in Africa (Harlan 1975; Brunken 1977). Some authors proposed that pearl millet is the product of multiple domestications (Harlan 1975; Porteres 1976), but others proposed a single domestication (Marchais and Tostain 1993). Several authors agree that domestication has taken place in Africa (Harlan 1975; Porteres 1976; Marchais and Tostain 1993), but different geographical origins for this crop have been proposed along the Sahelian zone from Mauritania to Sudan. To enhance our understanding of the evolutionary history of this important crop, we performed diversity and phylogenetic analyses that incorporated cultivated and wild pearl millet plants.

One substantial difficulty in assessing the origin and domestication process of cultivated crops is that the studies are based on current populations. Among the problems of such an analysis is the occurrence of hybridization between the crops and their wild relatives. Such hybridization creates admixed populations. In several crops, such hybridization processes led to what is called the crop/weed complex (Harlan 1975). In pearl millet, an outcrossing species, weedy populations are found throughout the area of cultivation of pearl millet in west Africa, east Africa, southern Africa and Asia (Brunken 1977). Occurrence of hybridization hampers phylogenetic analyses. However, new statistical methods are now available to identify hybrid genotypes (Pritchard et al. 2000; Falush et al. 2003). These methods permit us to identify and thus exclude hybrids from computations, which allows for finer analysis of the domestication process.

To analyze the diversity and evolutionary history of pearl millet, we used microsatellite loci (Allouis et al. 2000; Qi et al. 2001; 2004; Budak et al. 2003; Mariac et al. 2006a). Our aims were to (1) compare the diversity between wild and cultivated pearl millet samples; (2) identify the hybrid genotypes between wild and cultivated samples; and (3) document the evolutionary history and domestication of pearl millet.

Materials and methods

Plant material

We sampled 439 accessions (one plant each) of pearl millet from world collections assembled from 1975 to 1989 by IRD, ICRISAT and IPBGR (Fig. 1, Supplementary material S1). This sample comprised 355 cultivated accessions grown in the entire traditional cultivation area in Africa and Asia. We analyzed 84 wild plants originating from West African savannah zones from Senegal to Sudan. They represent most of the known populations of wild pearl millet. The complete data including the accession name, the country of origin and geographical coordinates of the collecting points are provided as a supplementary data file (S1).

Fig. 1 Geographic distribution of the pearl millet accessions used in this study. The sample included 355 cultivated (*gray dots*) and 84 wild accessions (*black triangles*) collected from North Africa (Algeria, Morocco, Tunisia), West Africa (Benin, Burkina Faso, Ghana, Guinea, Mali, Mauritania, Niger, Nigeria, Togo, Sierra Leone, Senegal), Central Africa (Cameroon, Chad, Sudan, RCA), East Africa (Malawi, Zimbabwe, Zambia, Tanzania, Kenya, Namibia) and Asia (India, Pakistan, Yemen)

DNA extraction and microsatellite amplification

DNA from individual plants was extracted from fresh leaves ground in liquid nitrogen. The powder was resuspended with 750 µl extraction buffer (Tris 0.1 M PH 8, NaCl 1.25 M, EDTA 0.02 M, MATAB 4%). After incubation for 4 h at 65°C, DNA extraction was performed using chloroform:isoamyl alcohol (24: 1). After centrifugation (9000g, 10 min, 4°C), the DNA was precipitated from the supernatant using 550 µl of isopropanol (centrifugation $9000 \times g$, 10 min, 4°C), washed with 70% ethanol, dried and resuspended in 200 µl sterile water. The positions of the different individual DNA extractions were then randomized on 96 well-plates regardless of their nature (wild/cultivated) or their geographical origin.

Out of the 27 microsatellite loci used in this study, 25 had been previously used in assessments of pearl millet diversity (Mariac et al. 2006a). Two additional loci were used in this study and correspond to PSMP2231 and PSMP2261 (Allouis et al. 2001). Eleven of the loci used in this study have already been placed on a pearl millet genetic map (Qi et al. 2004). They are spread over linkage group 2 (PSMP2237, PSMP2201, PSMP2206, PSMP2231), linkage group 3 (PSMP2216, PSMP2214), linkage group 4 (PSMP205), linkage group 5 (PSMP2208, PSMP2202, PSMP2202) and linkage group 7 (PSMP2266).

The PCR reaction was carried out as previously described (Mariac et al. 2006a). Pooled PCR products were size separated by capillary electrophoreses using an ABI Prism 3130 DNA analyzer (Applied Biosystems). Each of the 384 well PCR plates included eight negative PCR controls. Sizes were determined and scored using GeneMapper V.3.7. The scoring was manually checked by two different persons.

Diversity and differentiation between wild and cultivated samples

Analysis of genetic data was performed using Power Marker v.2.5 (Liu and Muse 2005). The number of shared alleles between groups, or "common alleles", and the number of alleles present in one population only, or "specific alleles", were calculated. The number of alleles, gene diversity and differentiation (F_{ST}) per locus between groups were also calculated. The gene diversity was calculated as $n/(n-1)(1-\rho_i^2-H_o/2n)$ where n is the number of individuals, ρ_i the frequency of the i allele and H_o the number of observed heterozygotes. Differentiation (F_{ST}) between the wild and cultivated groups was tested locus by locus, and its overall significance was tested (G-test, 10,000 randomizations) with FSTAT (Goudet 2001). F_{IS} were calculated and tested (Hardy-Weinberg test, 10,000 randomization) with FSTAT. The number of alleles is dependent on the sample size. To compare the number of alleles between our two samples, we calculated a parameter called "allelic richness" using FSTAT. The allelic richness (R) was calculated using the formula (Petit et al. 1998):

$$R=\sum_{i=1}^{k}\left(1-\frac{C_{2N-N_i}^{2n}}{C_{2N}^{2n}}\right),$$

where N_i was the number of the i allele among the population of the largest size N (2 N chromosomes), and n was the number of individuals analyzed for the smallest population (2n chromosmes), and k was the total number of alleles for the locus studied.

The allelic richness and gene diversity were compared between the wild and the cultivated samples using a Wilcoxon paired test. Using the allelic frequencies of each individual, we performed a principal component analysis with Statistica v.6 (Statsoft France 2001. STATISTICA v.6, http://www.statsoft.com). Patterson et al. (2006) proposed a statistical test to assess if an eigenvector of a PCA analysis reflects real population structure or merely noise.

To assess significance of the first eigenvalues, we considered the matrix (m, n) where m was the number of plants and n was the number of markers. We first calculated n', a theoritical statistical parameter modeling an "effective number of markers" to take into account non-independence between markers (Patterson et al. 2006).

$$n'=\frac{(m+1)\left(\sum_{i=1}^{m-1}\lambda_i\right)^2}{(m-1)\sum_{i=1}^{m-1}(\lambda_i)^2+\left(\sum_{i=1}^{m-1}\lambda_i\right)^2},$$

where m was the number of plants, λ_i was the ith eigenvalue sorted in decreasing order.

We then calculated two parameters $\mu(m,n')$ and $\sigma(m,n')$:

$$\mu(m,n')=\left(\frac{\sqrt{n'-1}+\sqrt{m}}{n'}\right)^2$$

$$\sigma(m,n')=\frac{\sqrt{n'-1}+\sqrt{m}}{n'}\left(\frac{1}{\sqrt{n'-1}}+\frac{1}{\sqrt{m}}\right)^{1/3}$$

The first eigenvalue λ_1 was then normalized so that the sum of all eigenvalues was $m-1$.

$$l_1=\frac{(m-1)\lambda_1}{\sum_{i=1}^{m-1}\lambda_i}$$

The test statistics was :

$$x=\frac{l_1-\mu(m,n')}{\sigma(m,n')}$$

P value was obtained from a Tracy-Widom density law table (Patterson et al. 2006). To test the significance of λ_2, the procedure was reiterated with m reset to $m-1$ as suggested by Patterson et al. (2006).

The phylogenetic relationships between accessions and groups of accessions were assessed using the neighbor-joining method implemented in Power Marker v.2.5 (Liu and Muse 2005). We used the shared allele distance, a distance frequently used for microsatellite loci (Bowcock et al. 1994; Vilà et al. 2001; Rosenberg et al. 2001). To assess the degree of statistical support for the different branches in the phylogenies, we performed a 1,000 bootstrap analysis on the data set. The consensus tree was then obtained using the program CONSENSE v.3.6 (Felsenstein 2005) on 1,000 phylogenic trees. The trees were edited and colored using the computer program TreeDyn v.1.9.3 (Chevenet 2004).

Population structure analyses

To assess population structure and the presence of hybrids or introgressed wild or cultivated individuals, we used a Bayesian method. This analysis was performed using the model-based program STRUCTURE (Pritchard et al. 2000) that infers the number of groups K (populations). The

Bayesian analysis allowed inference of the number K of populations and allele frequencies in these populations based on the empirical genotypic data. We used the admixture model with a burn-in period of 100,000 and 500,000 replicates. Five independent runs were performed without prior information on groups. The output of this analysis is an ancestry value (q) calculated for each individual that corresponds to the proportion of its genome derived from the different inferred groups. We performed the analysis for different numbers of assumed populations ($K = 1$ to $K = 10$) in order to determine the number of populations (K) supported by our data. The "accepted" number of populations (K) is often identified using the maximal value of likelihood (Pritchard et al. 2000; Hampton et al. 2004) in respect to K, $L(K)$. Recently, Evanno et al. (2005), using simulated data, found that the distribution of $L(K)$ did not show a clear mode for the true K value. They proposed an ad hoc quantity based on the second order change of the likelihood with respect to K (ΔK). The parameter $\Delta K = m\ (|L''K|)/s[L(K)]$, was calculated as the mean of absolute values of the second order change of the likelihood distribution divided by the standard deviation of likelihood. The modal value of this distribution is interpreted as the true number of populations K (Evanno et al. 2005). To detect the presence of hybrid or introgressed individuals between the cultivated and the wild groups we also performed the analysis with $K = 2$ recording the 95% confidence interval (CI) of the ancestry parameters (q). Introgressed individuals were detected according to their ancestry value (q). An ancestry close to 0 or 1 suggests no evidence of introgression for this individual. Intermediate values of ancestry suggest introgression. An arbitrary threshold of the ancestry (q) has to be chosen to assign an individual to a group (cultivated/wild). A plant was assigned to a group if its ancestry is higher than 0.90 and if the lower bound of the 95% CI of this value is higher than 0.60. We therefore based this threshold not only on the mean ancestry value but also on the 95% CI of this value.

Results

Diversity and differentiation between wild and cultivated samples

Using 27 *loc*i, 404 different alleles were observed on 439 individuals. The overall genetic diversity was 0.58. We found that 18.6% of the alleles are specific to the cultivated sample and 13.4% are specific to the wild. The average number of alleles per *locus* was 12.2 in the cultivated sample and 13.0 in the wild sample. The number of alleles can be largely influenced by the sample size; therefore, to compare this parameter between cultivated and wild samples, we calculated the allelic richness (Goudet 2001). The average allelic richness per locus was 9.6 in the cultivated sample and 11.9 in the wild sample (Table 1). The allelic richness was significantly lower in the cultivated sample than in the wild sample (Wilcoxon paired test, $Z = 4.1$, $P < 0.001$). The cultivated sample possessed 81% of the allelic richness of the wild sample. Gene diversity per locus varied from 0.11 to 0.95. The average value of gene diversity was 0.60 in the cultivated sample and 0.72 in the wild sample (Table 1). Gene diversity (Table 1) of the cultivated sample was significantly lower compared to the wild sample (Wilcoxon paired test, $Z = 3.5$, $P < 0.001$). The cultivated sample possessed 83% of the gene diversity found in the wild sample.

The differentiation (F_{ST}) by locus varied from 0.005 to 0.32, with an average value of 0.10. We found a highly significant differentiation between the wild and cultivated sample (G-test, $P < 0.001$). F_{IS} was positive and statistically significant (H.W.-test, $P < 0.001$) for all loci. Their average value was 0.26, showing a heterozygote deficiency in both cultivated and wild samples.

Population structure

To detect the population structure of our total sample, we used a Bayesian approach implemented in STRUCTURE (Pritchard et al. 2000; Falush et al. 2003). This program allowed us to assign individuals to K populations and to determine the proportion of genome of each individual in each assigned population. We performed this analysis where K could vary from 1 to 10. The modal value of this distribution was proposed to be the number of populations K supported by the data. Our analysis strongly supported $K = 2$ as the number of populations supported by our data set. The two groups were formed of cultivated and wild

Table 1 Average number of alleles, allelic richness and gene diversity in a global collection of wild and cultivated pearl millets

	Cultivated sample	Wild sample	*P* value
Sample size	355	84	
Allelic richness	9.6	11.9	<0.001
Gene diversity	0.60	0.72	<0.001
Sample size*	289	64	
Allelic richness	8.4	11.3	<0.001
Gene diversity	0.58	0.73	<0.001

The number of alleles, allelic richness and gene diversity are presented for the cultivated and wild samples. The differences in gene diversity and allelic richness between samples are tested using a Wilcoxon paired test. The *P* value is reported. Allelic richness and the gene diversity were estimated using a sample of smaller size (*asterisk*) where hybrid genotypes were excluded (see text for details). The same Wilcoxon paired test is performed on this data set and the *P* value is reported

accessions. The ancestry and its confidence interval (CI) using $K = 2$ were estimated for the 355 cultivated and 84 wild accessions. The ancestries estimated for an individual in the two groups sum to 1. We analyzed the ancestry in the wild clusters, q_W (Fig. 2). Most of the wild accessions showed a strong ancestry in the wild cluster with ancestry values higher than 95%. A total of 20 plants (24%) showed an ancestry lower than 90% or a lower bound of the 95% CI lower than 0.60. The cultivated accessions had a very low ancestry in this wild cluster. A total of 64 cultivated accessions (18%) had an ancestry higher than 10% or a higher bound of the 95% CI higher than 0.40 in the wild cluster.

If we exclude from the overall sample genotypes showing admixed ancestry based on the average ancestry value and its confidence interval, we obtained 355 individual plants with 291 cultivated accessions and 64 wild accessions. Genetic parameters were computed for these reduced samples. Compared to results presented above for samples including introgressed individuals, we found a higher differentiation between the cultivated and wild samples ($F_{ST} = 0.14$). The allelic richness in the cultivated group was 8.4 in the cultivated group and 11.3 in the wild group (Table 1). The average gene diversity was 0.58 and 0.73 in the cultivated and wild groups, respectively (Table 1). Thus the cultivated sample possessed 74% of the allelic richness and 79% of the gene diversity found in the wild sample. All the differences between cultivated and wild samples were highly significant.

Principal component and phylogeny analyses

The principal component analysis (Fig. 3) explained 2.9% of the first component and 1.6% of the second. The statistical significance for these first two axes of the PCA using Patterson et al. (2006) procedure was strong ($x = 107.9$, $P < <0.001$ and $x = 22.4$, $P < <0.001$, respectively). Despite the low value of the variance explained, the PCA revealed the real population structure. The differentiation between wild and cultivated samples was clear on the first axis of PCA. The second axis further differentiated the wild accessions. This principal component analysis showed that

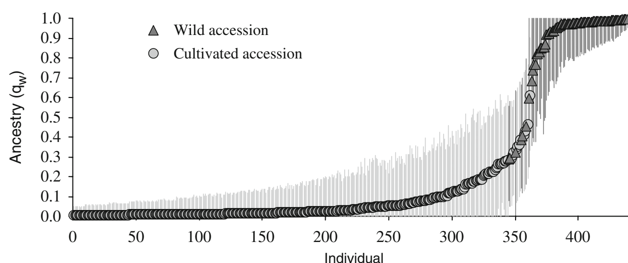

Fig. 2 Individual ancestry in the pearl millet wild and cultivated groups. For each individual, ancestry in the wild cluster (q_w) values and its confidence intervals at 95% were calculated (*gray vertical lines*). Ancestries were calculated assuming two populations using the STRUCTURE software

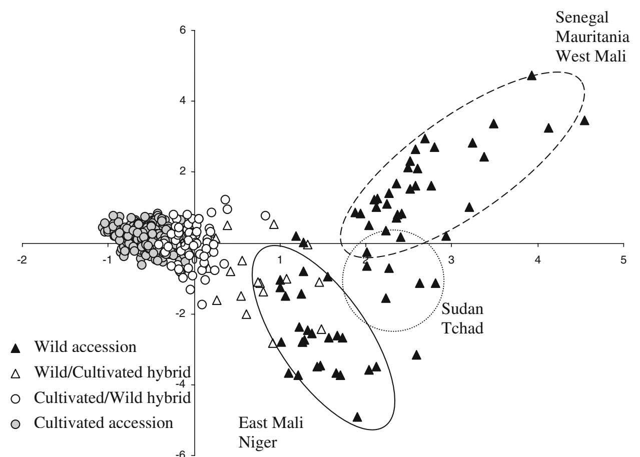

Fig. 3 Principal component analysis of pearl millet microsatellite diversity. The principal component analysis were performed on the allele frequencies of each individual. The analysis was done on 355 cultivated plants and 84 wild plants. The first component explains 2.89% and the second 1.55 % of the total variation. The first two PCA axes are supported by a high level of significance (see text for details)

all cultivated accessions were regrouping into a single cloud. Hybrid genotypes identified using the Bayesian analysis tended to have, as expected, an intermediate position of the first axis of the PCA.

To determine the relationship between accessions, we performed a phylogenetic analysis of our sample of 439 (355 cultivated and 84 wild) pearl millet accessions using 27 loci microsatellites (Fig. 4a). Wild accessions, which were found in the cultivated group in the phylogenetic tree were detected as admixed individuals using STRUCTURE: PE02592 ($q_w = 0.29$), PE08105 ($q_w = 0.29$), PE08197 ($q_w = 0.32$), PE05708 ($q_w = 0.40$), PE11359 ($q_w = 0.60$), PE05714 ($q_w = 0.83$), PE08146 ($q_w = 0.86$) and PE06652 ($q_w = 0.87$). Only one cultivated accession (PE00150) showed clustering in the wild cluster.

To analyze the origin of cultivated pearl millet, we first excluded the hybrid genotypes identified using the Bayesian method. We pooled the individual plants into 85 eco-geographically defined groups to allow a bootstrap analysis. Each eco-geographic group consisted of individuals of similar longitude and latitude (S1). We then performed a bootstrap analysis with this smaller number of taxonomic units. The phylogeny for ecogeographic group sample shows that cultivated pearl millet was monophyletic in 987 of 1,000 phylogenetic trees (Fig. 4b). Wild accessions from eastern of Mali and northwestern Niger were genetically close to the cultivated group.

Discussion

In the present study, gene diversity in a world collection of cultivated pearl millet was estimated to be 0.58, a value

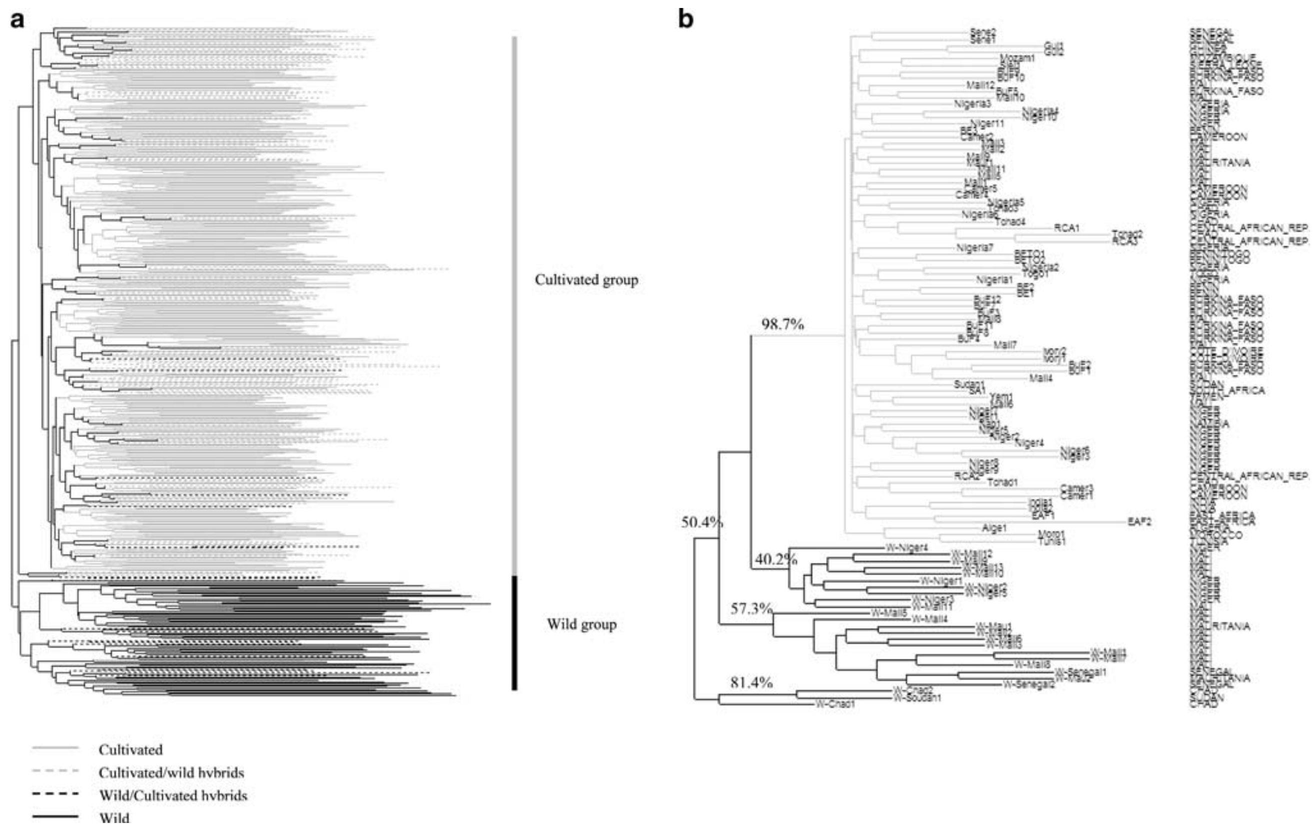

Fig. 4 Phylogenetic relationship between each individual plant. A first phylogenetic analysis has been performed on all of the 355 cultivated samples and 84 wild samples (**a**). The phylogeny is based on the shared allele genetic distance. The distance was calculated using the genotype obtained using 27 microsatellite loci. Using the geographical origin of each accession, we created 85 ecogeographically defined groups to perform a bootstrap analysis (**b**). The number reported on the tree corresponds to the percentage of the value of bootstrap among a 1,000 bootstrap replicates

similar to those observed in studies using microsatellite markers for world collections of cereals like *indica* rice, *japonica* rice and wheat with 0.55, 0.47 (Garris et al. 2005) and 0.56 (Kuleung et al. 2006), respectively. In maize, gene diversity in a world cultivated sample was estimated at 0.74 (Vigouroux et al. 2005). These differences in gene diversity are the product of the effective size of wild populations, the strength of the domestication bottleneck, the mutation rate of the SSR loci used and the evolutionary history of cultivated populations (Thuillet et al. 2005; Vigouroux et al. 2005). Diversity in cultivated plants is generally lower than that found in their wild relatives. In pearl millet, the gene diversity in cultivated pearl millet represented 74% of the diversity found in wild populations. Such a figure was lower than that found in maize where gene diversity in the cultivated sample was 88% of the gene diversity of the wild sample (Vigouroux et al. 2005).

This study and a previous study restricted to Niger (Mariac et al. 2006a) found a higher diversity for the wild pearl millet sample compared to cultivated sample. These results contrast with a previous study based on isozyme data where no differences between wild and cultivated samples were observed (Tostain 1992). Low diversity observed at isoenzyme data or selection occurring at isoenzyme markers might explain the difference between diversity assessments done with microsatellite and isoenzyme loci (Mariac et al. 2006a).

The differentiation between wild and cultivated groups (not including introgressed individuals) was 0.14. This differentiation was twice as high as the one found, $F_{ST} = 0.07$, between maize and its wild ancestor, teosinte (Vigouroux et al. 2005). Microsatellites in the present study were mainly di-nucleotide loci and differentiation was even lower in maize when we considered di-nucleotide loci only, $F_{ST} = 0.044$ (Vigouroux et al. 2005). The higher differentiation observed between cultivated and wild pearl millet samples might certainly be linked to difference of strength of the bottleneck or selection effect during domestication. Compared to maize, domestication in pearl millet may have been performed on a smaller sample of wild plants and/or during a longer period of time.

Population structure of our sample revealed two major groups wild and cultivated based on a Bayesian approach. Study of the effectiveness of this Bayesian method has

shown that clustering was dependent on the number of markers, the number of individuals and the differentiation between populations (Rosenberg et al. 2005). In the present analysis, we might have only detected the major differences between two major groups. To identify more subtle differences within cultivated and wild groups, a larger number of markers should certainly be used. Our sampling strategy was to maximize the number of accessions to obtain a representative sample of both wild and cultivated accessions. A sampling strategy maximizing the number of individuals per accession may also be more effective to uncover these subtle differences inside wild or cultivated groups.

We found significant evidence of introgression between wild and cultivated pearl millet samples. Introgressed wild and cultivated individuals are found mainly at latitudes where wild populations are in sympatry with cultivated pearl millet (Clement et al. 1993). The existence of such hybridization through the Sahelian zone may contribute to the occurrence of weedy populations commonly found in pearl millet. The frequency of this wild/cultivated hybridization is however variable and several factors limit or enhance gene flow locally, such as flowering phenology, pollen competition and reproductive barriers (Sarr et al. 1988; Robert et al. 1991; Amoukou and Marchais 1993; Renno and Winkel 1996). However, weedy populations have also been found to occur in regions largely isolated from wild populations (Brunken 1977; Clément et al. 1993; Miura and Terauchi 2005; Mariac et al. 2006b). We also found in our study hybrid genotypes in eastern Africa, southern Africa and India. These results were in agreement with that previously observed based on morphological data (Brunken 1977). Thus hybridization and proximity to wild populations may not be the only factor explaining the occurrence and the maintenance of weedy populations. A recent study in Mali suggested that weedy morphology was governed by a major locus (Miura and Terauchi 2005). Weedy plants generally flowered earlier and were also harvested before maturity when food was scarce (Miura and Terauchi 2005; Mariac et al. 2006b). So, weedy populations may also be maintained by direct or indirect human actions (Mariac et al. 2006b).

The existence of hybrids made the analysis of the origin of pearl millet rather difficult. Hybrids found using Bayesian analysis created continuity in principal component analysis between wild and cultivated accessions. If we excluded hybrids based on restrictive rules, we found a clearer monophyletic origin of the cultivated population and a high bootstrap value for a monophyletic origin. Recently, some authors criticized the dominant markers and neighbor joining phylogeny framework to infer a monophyletic origin (Allaby and Brown 2003). Several authors have used such an approach to infer a single origin of a cultivated crop (Badr et al. 2000; Ozkan et al. 2002). However, Salamini et al. (2004) stated that PCA and phylogenic analysis together were powerful tools for investigating monophyletic origin. Pearl millet cultivated samples formed a monophyletic group in the eco-geographical phylogenetic tree and presented only one homogeneous group in the principal component analysis. These two results suggested that a single domestication of pearl millet was the most likely scenario.

Our results were based on microsatellite markers and these markers might not be the most effective markers for phylogenetic studies because of homoplasy or size constraints (Estoup et al. 2002). However, homoplasy and size constraints tended to minimize differentiation between populations and therefore the result of a monophyletic origin of cultivated plants was to these regards a conservative result. However, genetic distances between wild individuals separated by longer time scale might be underestimated using these markers.

Our phylogeny and principal component analysis showed that the wild plants from eastern Mali through northwestern Niger are the closest to the cultivated group. These wild populations span the area from the interior delta of Niger to the Aïr Mountains. This result suggests that the cradle of pearl millet domestication might be located somewhere in this rather large region. A previous study based on iso-enzyme suggested a most western origin of pearl millet domestication in Mauritania and Western Mali (Tostain 1992). This previous hypothesis was based on lower genetic distance between cultivated and wild populations in these regions. Such a result could be easily explained by the existence of wild/cultivated hybridization (Mariac et al. 2006b). With such hybrid populations, it is difficult to conclude on pearl millet origin without the confounding effect of recent ancestry. Our approach to this problem was to identify and then discard hybrid genotypes to assess wild and cultivated genetic relationship. Using this methodology, we identified eastern Mali and western Niger as the possible regions of domestication of pearl millet, but this result should be regarded with caution. Climate and human settlement have changed in Africa in the last 8,000 years (Salzmann and Hoelzmann 2005; Kuper and Kröpelin 2006) and some wild populations may have moved or disappeared since pearl millet domestication. Further studies will be needed to confirm and precisely map the location of pearl millet domestication. One of the oldest archeobotanical evidences of the cultivation of pearl millet has been found in Mauritania dating around 3500 BP at DharTichitt (Amblard and Pernes 1989). However, proof of cultivation of pearl millet was also discovered in Ghana around 3460 BP at Birmi (D'Andrea et al. 2001; D'andrea and Casey 2002), and near the Lake Chad in Nigeria between 3500–3300 BP (Klee et al. 2004).

Evidence of pearl millet cultivation was found also in India around the same time frame (Fuller et al. 2004). So, it seems that pearl millet cultivation was widespread around 3500 years BP in West Africa. Domestication of pearl millet is certainly older than this date, and the oldest proof of cultivation does not really prove that domestication process took place in Mauritania. Tostain (1998) proposed a date of domestication around 8000 years BP, and a spread of pearl millet cultivation in Asia around 4500–5000 years BP. These dates are compatible with archeological data even though the archeological oldest evidence found in Mauritania dates back to only 3500 BP (Amblard and Pernes 1989). In the future, archeological remains will hopefully permit a more precise estimation of the timing of pearl millet domestication and of the spread of the cultivation of this important Sahelian crop.

Acknowledgments I.O. was funded by the Agence Universitaire de la Francophonie (AUF). This work was supported by the Institut de Recherche pour le Développement - IRD. We thank S. Tostain and G. Bezançon for their discussion and help in the choice of pearl millet samples, and R. Oldham for help during redaction of the article. The two anonymous reviewers significantly improved the previous version of the manuscript.

References

Allaby RG, Brown TA (2003) AFLP data and the origins of domesticated crops. Genome 46:448–453

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

Amblard S, Pernes J (1989) The identification of the cultivated pearl millet (*Pennisetum*) amongst plant impressions on pottery from Oued Chebbi (Dhar Oualata, Mauritania). Afr Archaeol Rev 7:117–126

Amoukou AI, Marchais L (1993) Evidence of partial reproductive barrier between wild and cultivated pearl millet (*Pennisetum glaucum*). Euphytica 67:19–26

Badr A, Müller K, Schäfer-Pregl R, El Rabey H, Effgen S et al (2000) On the origin and domestication history of barley (*Hordeum vulgare*). Mol Biol Evol 17:499–510

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

Brunken JN (1977) A systematic study of *Pennisetum* sect. *pennisetum* (*Gramineae*). Am J Bot 64:161–176

Budak H, Pedraza F, Cregan PB, Baenziger 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

Chevenet F (2004) Treedyn—dynamic graphic editor for exploring phylogenetic or classification trees. Institut de recherche pour le développement, Montpellier, France (http://www/viradium.mpl.ird.fr/treedyn/)

Clément JC, Bezançon G, Billard G (1993) Prospections des mils cultivés et sauvages d'Afrique de l'Ouest. In: Hamon S (eds) Le mil en Afrique diversité génétique et agrophysiologique: Potentialités et contraintes pour l'amélioration génétique et l'agriculture. Actes de la réunion thématique sur le mil (*Pennisetum glaucum,* L.), Montpellier du 24 au 26 novembre 1992

D'Andrea AC, Casey J (2002) Pearl millet and Kintampo subsistence. Afr Archaeol Rev 19:147–173

D'Andrea AC, Klee M, Casey J (2001) Archaeological evidence for pearl millet (*Pennisetum glaucum*) in sub-saharan West Africa. Antiquity 75:341–348

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

Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetic analysis. Mol Ecol 11:1591–1604

Falush D, Stephens M, Pritchard JK (2003) Inferences of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6 distributed by the author. Department of Genome Sciences, University of Washington, Seattle

Fuller D, Korisettar R, Venkatasubbaiah PC, Jones MK (2004) Early plant domestications in southern India: some preliminary archaeobotanical results. Veget Hist Archaeobot 13:115–129

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631–1638

Goudet J (2001) FSTAT, a program to estimate and test gene diversity and fixation indices (version 2.3.9.2). Accessed on http://www.unil.ch./izea/softwares/fstat.html

Harlan JR (1975) Crops and man. American Society of Agronomy and Crop Science Society of America, Madison, Wisconsin, p 295

Hampton JO, Spencer PBS, Alpers DL (2004) Molecular techniques, wildlife management and importance of genetic population structure and dispersal: a case study with feral pigs. J Appl Ecol 41:735–743

Heun M, Schäfer-Pregl R, Klawan D, Castagna R, Accerbi M, Borghi B, Salamini F (1997) Site of einkorn wheat domestication identified by DNA fingerprinting. Science 278:1312–1314

Klee M, Zach B, Stika H-P (2004) Four thousand years of plant exploitation in the lake Chad basin (Nigeria), part III: plant impressions in postherds from the final stone age Gajiganna culture. Veget Hist Archaeobot 13:131–142

Kuleung C, Baenziger PS, Kachmanand SD, Dweikat I (2006) Evaluating the genetic diversity of triticale with wheat and rye SSR Markers. Crop Sci 46:1692–1700

Kuper R, Kröpelin S (2006) Climate-controlled Holocene occupation in the Sahara: motor of Africa's evolution. Science 313:803–807

Liu K, Muse S (2005) PowerMarker: new genetic data analyis software. Version 3.25. free program distributed by author over the internet from http://www.powerMarker.net

Londo JP, Chiang Y-C, Hung K-H, Chiang T-Y, Schall BA (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. Proc Natl Acad Sci USA 103:9578–9583

Marchais L, Tostain S (1993) Evaluation de la diversité génétique des Mils (*Pennisetum glaucum,* (L.) R. BR.) au moyen de marqueurs enzymatiques et relation entre formes sauvages et cultivées. In: Hamon S (eds) *Le mil en Afrique diversité génétique et agrophysiologique: Potentialités et contraintes pour l'amélioration génétique et l'agriculture*. Actes de la réunion thématique sur le mil (*Pennisetum glaucum,* L.), Montpellier du 24 au 26 novembre 1992

Mariac C, Luong V, Kapran I, Mamadou A, Sagnard F, Deu M, Chantereau J, Gerard B, Ndjeunga J, Bezancon G, Pham JL, Vigouroux Y (2006a) Diversity of wild and cultivated pearl

millet accessions (*Pennisetum glaucum* [L.] R. Br.) in Niger assessed by microsatellite markers. Theor Appl Genet 114:49–58

Mariac C, Robert T, Allinne C, Remigereau MS, Luxereau A, Tidjani M, Seyni O, Bezancon G, Pham JL, Sarr A (2006b) Genetic diversity and gene flow among pearl millet crop/weed complex: a case study. Theor Appl Genet 113:1003–1014

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley J (2002) A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci U S A 99:6080–6084

Miura R, Terauchi R (2005) Genetic control of weediness traits and the maintenance of sympatric crop-weed polymorphism in pearl millet (*Pennisetum glaucum*). Mol Ecol 14:1251–1261

Morrell PL, Clegg MT (2007) Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. Proc Natl Acad Sci U S A 104:3289–3294

Nesbitt M, Samuel D (1996) From staple crop to extinction ? The archeology and history of the hulled wheats. In: Padulosi S, Hammer K, Heller J (eds) *hulled wheats*. Proceeding of the 1st international workshop on hulled wheats, pp 41–100

Ozkan H, Brandolini A, Schafer-Pregl R, Salamini F (2002) AFLP analysis of a collection of tetraploid wheats indicates the origin of emmer and hard wheat domestication in southeast Turkey. Mol Biol Evol 19:1797–1801

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. Plos Genetics 2(12):e190

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. Cons Biol 12:844–855

Porteres R (1976) African cereals: eleusine, fonio, black fonio, teff, Brachiaria, Paspalum, Pennisetum and African rice. In: Harlan J, De Wet J, Stemler A (eds) *Origins of African plant domestication*. Mouton Publishers, La Hague

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, Waterman 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 J-F, 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, Pernes J (1991) Gametophytic competition as influencing gene flow between wild and cultivated forms of pearl millet (*Pennisetum typhoides)*. Genome 34:195–200

Rosenberg NA, Woolf E, Pritchard JK, Schaap T, Gefel D, Shpirer I, Lavi U, Bonne-Tamir B, Hillel J, Feldman MW (2001) Distinctive genetic signature in the Libyan Jews. Proc Natl Acad Sci USA 98:858–863

Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, Feldman MW (2005) Clines, clusters, and the effect of study design on the inference of human population structure. Plos Genetics 1:e70

Salamini F, Heun M, Brandolini A, Özkan H, Wunder J (2004) Comment on « AFLP data and the origins of domesticated crop ». Genome 47:615–620

Salzmann U, Hoelzmann P (2005) The Dahomey gap: an abrupt climatically induced rain forest fragmentation in West Africa during the late Holocene. Holocene 15:190–199

Sarr A, Sandmeier M, Pernès J (1988) Gametophytic competition in pearl millet *Pennisetum typhoides* (Sapf. And Hubb). Genome 30:924–929

Second G (1982) Origin of the genetic diversity of cultivated rice (*Oryza* ssp.): study of the polymorphim scored at 40 isoenzyme loci. Jpn J Genet 46:448–453

Thuillet AC, Bataillon T, Poirier S, Santoni S, David JL (2005) Estimation of long-term effective population sizes through the history of durum wheat using microsatellite data. Genetics 169:1589–1599

Tostain S (1992) Enzyme diversity in pearl millet (*Pennisetum glaucum)* 3 Wild millet. Theor Appl Genet 83:736–742

Tostain S (1998) Le mil, une longue histoire, hypothèses sur sa domestication et ses migrations. In: Chastenet N (ed) Plantes et paysages d'Afrique. Karthala, Paris, pp 461–490

Vigouroux Y, Mitchell S, Matsuoka Y, Hamblin M, Kresovich S, Smith S, Jaqueth J, Smith O, Doebley J (2005) An analysis of genetic diversity across the maize genome using microsatellites. Genetics 169:617–1630

Vilà C, Leonard JA, Göthersthröm A, Marklund S, Sandberg K, Lidén K, Wayne RK, Ellegren H (2001) Widespread origins of domestic horse lineage. Science 291:474–477

Zohary D, Hopf M (2000) Domestication of plants in the old world: the origin and spread of cultivated plants in West Asia, Europe and the Nile Valley. Oxford University Press, New York