Theor Appl Genet (2006) 113:1003–1014
DOI 10.1007/s00122-006-0360-9

ORIGINAL PAPER

Genetic diversity and gene flow among pearl millet crop/weed complex: a case study

C. Mariac · T. Robert · C. Allinne · M. S. Remigereau · A. Luxereau · M. Tidjani · O. Seyni · G. Bezancon · J. L. Pham · A. Sarr

Received: 20 September 2005 / Accepted: 3 July 2006 / Published online: 19 August 2006

Abstract Weedy plants with intermediate (domesticated × wild) phenotypes occur in most pearl millet fields in West Africa, even in the absence of wild populations. They are usually found, in high numbers, both inside and outside of drills. Questions pertaining to the evolutionary dynamics of diversity within the pearl millet complex (domesticated–weedy–wild forms) were addressed in this study. The diversity of the different components of this complex sampled in two pearl millet fields in two villages of southwestern Niger was assessed at both molecular (AFLP) and morphological levels. Results show that, in both fields, weedy plants found outside of drills are morphologically distinct from weedy plants found inside drills, despite their close similarity at AFLP markers. The data suggest some introgression from the wild to the weedy population but nevertheless that the gene flow between the parapatric wild and domesticated populations is very low. This challenges the traditional view that regular hybridization between domesticated and wild pearl millets explains the abundance of these weedy plants despite farmers' seed selection. The level of genetic differentiation between fields from the two villages was low when considering domesticated and weedy plants. This could be explained by high gene flow resulting from substantial seed exchanges between farmers. The fact that it is very difficult for farmers to keep their own selected seeds, and the consequent substantial seed exchanges between them, is probably the main factor accounting for the maintenance and dispersal of weedy pearl millets in the region, even in areas where no wild forms have been observed.

Communicated by H. C. Becker

C. Mariac and T. Robert contributed equally to the work.

T. Robert (✉) · C. Allinne · M. S. Remigereau · A. Sarr
Université Paris VI, Génétique et Evolution des Plantes Cultivées, Laboratoire Ecologie Systématique et Evolution, Bat. 360, Université Paris-Sud, 91405 Orsay, France
e-mail: thierry.robert@ese.u-psud.fr

M. Tidjani
Laboratoire de Génétique des Plantes, UMR 1097, Diversité et Génomes des Plantes Cultivées, Institut de Recherche pour le Développement (IRD), 11416 Niamey, Niger

C. Mariac · G. Bezancon · J. L. Pham
Equipe Dynadiv, UMR 1097, Diversité et Génomes des Plantes Cultivées, Institut de Recherche pour le Développement (IRD) , BP64501, 34394 Montpellier cedex 05, France

O. Seyni
Institut National de la Recherche Agronomique du Niger, 429 Niamey, Niger

A. Luxereau
Laboratoire Eco-anthropologie et ethnobiologie Département "Hommes, Natures Sociétés", UMR 5145, Muséum National d'Histoire Naturelle, Paris, France

Introduction

On-farm conservation strategies have been advocated as a sustainable approach to long-term preservation of genetic resources. This idea is supported by the fact that varietal diversity is generally very high in non-intensive farming systems (Brush 1995; Jarvis 1999) and that farmers have played a major role in the production and maintenance of this diversity since the beginning of agriculture. From an evolutionary standpoint, conservation

programmes should also aim to ensure the maintenance of evolutionary processes generating heritable variation in cultivated populations, which is the basis for future improvement and adaptation ability. Among evolutionary factors, gene flow between cultivated varieties and between domesticated and spontaneous relatives has been considered by some authors to contribute favourably to the evolution of crops in these agrosystems. Indeed, it can lead to the production of new genotypes upon which farmer selection can operate to produce enhanced phenotypes (Jarvis and Hodgkin 1999) and spread the advantageous alleles (Morjan and Rieseberg 2004). This concept is also supported by recent studies on maize (Louette et al. 1997; Pressoir and Berthaud 2004a, b) and cassava (Elias et al. 2001a, b) that have documented the role of farmer practices in promoting genetic recombination between varieties.

On the other hand, experimental and theoretical population genetics models have shown that strong gene flow can be responsible for a decrease in local adaptation through introgression of genes from non-adapted genotypes (Antonovics and Bradshaw 1970; Slatkin 1987). This has especially been modelled in the case of multi-locus selection in the framework of gene flow between domesticated and wild phenotypes (Slatkin 1995). Whether this prediction holds for non-intensive agrosystems is an important question to answer, because they often display remarkable evolutionary situations where wild, weedy and domesticated forms coexist. The actual importance of gene flow from the wild relatives for crop diversity has been questioned (Wood and Lenné 1997), although in some instances it could produce aggressive weedy forms, as clearly documented in beet (Boudry et al. 1993) and common bean (Papa and Gepts 2003). Well-supported examples of wild-cultivated gene flow, such as those in sunflower (Linder et al. 1998), alfalfa (Muller et al. 2001), rice (Suh et al. 1997) and beet (Viard et al. 2004) are still very scarce (Jarvis and Hodgkin 1999), and mainly focus on the consequences of introgression from domesticated plants into the wild relatives (Ellstrand et al. 1999). As stated by Viard et al. (2004), the extent of hybridization events between related wild and domesticated plants and the processes underlying the maintenance of weedy forms originated from this gene flow, should therefore now be assessed at both the field and the landscape levels.

Pearl millet—a highly outcrossing African crop—is an appropriate model for studying the effect of farmers' practices and gene flow in shaping the diversity within crop/weed complexes. Parapatric situations between cultivated (*Pennisetum glaucum* ssp. *glaucum*) and wild (*Pennisetum glaucum* ssp. *monodii*) forms still prevail in Africa. Allozyme analyses on wild and domesticated pearl millet samples from throughout Africa showed that diversity is patterned according to geographic distances at a large scale (Pilate-André 1992). However, at the sub-regional scale, the taxonomic status (wild versus domesticated) seemed to be a factor explaining a large part of isoenzyme polymorphism patterning. This result was confirmed by Tostain et al. (1987) in West Africa specifically. Thus, the extent of gene flow between domesticated and wild pearl millet is still questionable.

Weedy plants with intermediate domesticated/wild phenotypes are observed in most pearl millet fields in West Africa (Brunken et al. 1977), including those where wild forms are totally absent (Busso et al. 2000), as also shown for weedy rice (Oka 1988). Their presence in drills (holes in the ground made by farmers in which they generally sow a few dozens of seeds) decreases the production of good quality seeds (large and easy to thresh). In parapatric situations, weedy plants have been considered as derived from recurrent pollen flow between domesticated and wild populations (Marchais 1994; Busso et al. 2000). Their presence in the fields results from sowing or spontaneous seedlings issued from shattered seeds. In the former case, the seeds come from candles selected by farmers to produce seeds for the next growing season. Even when farmers are able to select their seeds on plants with the best phenotypes (well domesticated), there is no way to recognize seeds issued from pollination by weedy and wild plants. The level of genetic introgression of these weedy plants by genes from wild or domesticated plants is also unknown. Obviously, the dynamics of weedy populations is dependent on farmers' practices (especially seed management and weeding). The relative impact of these practices and gene flow from wild populations on these dynamics would be a key question to address from an evolutionary standpoint. This paper presents original results on in situ phenotypic and molecular diversity in the pearl millet crop/weed complex at the field level in two localities in southwestern Niger. The goal of this study is to gain further insights into the evolutionary dynamics of this crop/weed complex in the light of farmer's practices documented in this area.

Materials and methods

Study sites and local pearl millet classification

Diversity analyses were carried out in two pearl millet fields located in two villages (Alzu and Kouré) located

about 200 km apart, in southwestern Niger (Zarma region). The Kouré site has no wild populations but a high frequency of intermediate phenotypes. In Alzu, there are parapatric situations with wild populations (*P. glaucum* ssp. *monodii*) and pearl millet fields. The field we studied in this village is in parapatry with a large wild population (5.2 ha). According to the Alzu farmers, the wild population did not exist when they arrived and settled a few decades ago. They believe that this population originated from wild millet straw that was introduced for thatching roofs.

As the genetic dynamics of both landraces and weedy populations is affected by farmers' practices, we based our approach on farmers' classification. The Alzu field (3.71 ha) was grown with two early landraces, Haini Kiré and Darancoba, which flower at the same time. The Kouré field (2.26 ha) is also planted with two landraces: Haini Kiré and a semi-early landrace (Somno), which flowers on average about 20 days later than Haini Kiré. Farmers identify landraces mainly on the basis of candle morphology, seed size and seed colour. Apart from fully non-shattering domesticated phenotypes, they identify and sometimes harvest several morphological types, mainly on the basis of their spikelet shattering ability, candle and seed sizes. They obviously identify the weedy, non-fully domesticated phenotypes (with more or less shattering spikelets, loose candles). These weedy plants are called "*soun*." *Soun* plants are always found in drills (hereafter referred to as *soun*-in) as well as outside (hereafter referred to as *soun*-out). *Soun*-in and *soun*-out are found in all fields of the region.

Soun are early flowering plants but, due to profuse tillering, the flowering of some candles overlaps the flowering period of both early and semi-early domesticated phenotypes. Consequently, like wild plants, they can pollinate domesticated components in the fields. Finally, *soun* are also harvested, especially in harsh conditions.

In the following, we will refer to these classes, based on local knowledge, as categories (Table 1). The village from which plants originate will be indicated by the suffix A (Alzu) or K (Kouré).

Sampling strategies and morphological analyses

Different categories were used to design our sampling strategy. In each field and in the wild population, 20–80 individuals per category (total of 441 individuals analysed), distributed throughout the field, were measured during the harvest period, for 12 traits: candle length, candle diameter, rachis diameter, number of spikelets per square centimeter, pedicel length, shattering score, number of abscission layer for five involucres, seed coating score, presence/absence of the longest silk on the involucre, glume length, length of the spikelet bristles and seed weight (for 1,000 seeds). These traits are representative of the domestication syndrome sensu Poncet et al. (2000), i.e. traits for which domesticated and wild forms are fully differentiated. Spikelet shattering ability is mostly determined by the presence of an abscission layer at the base of the involucre pedicel. However, variations were noted in this trait within candles in our samples. Shattering ability was therefore scored by two variables: (1) an arbitrary shattering score based on an evaluation of the easiness of making spikelets shed from the candle; (2) the number of spikelets bearing a functional abscission layer among five spikelets.

Wild phenotypes were sampled in the large population growing adjacent to the Alzu field. Plants were collected within a range of 30 m from the field. Sampling was performed along parallel transect lines spaced 5 m apart.

DNA analyses

Samples of leaves were harvested and lyophilized before grinding. Approximately 0.2 g of powder was re-suspended with 1.8 ml of buffer A (0.35 M sorbitol, 0.1 M Tris, 5 mM EDTA, 0.5% sodium bisulphite) and centrifuged for 10 min at 10,000*g*; the pellet was re-suspended with 1.8 ml of buffer A without bisulphite and centrifuged as shown next. The pellet was re-suspended with 700 µl of extraction buffer B (Tris 0.1 M, NaCl 1.25 M, EDTA 0.02 M, MATAB 4%, pH 8) and incubated at 65°C for 4 h. Chloroform:isoamyl (24:1)

Table 1 List of morphological categories used in this paper and as defined by farmers in the area of study

Morphological categories defined in the paper	Farmer classification	Phenotype
Domesticated	Varieties: *Haini Kiré* (HK), *Darancoba* (D) and *Somno* (S)	Domesticated. Virtually no shedding ability
Soun-in	*Soun* (found in drills)	Intermediate domesticated/wild. More or less pronounced shedding ability
Soun-out	*Soun* (found outside drills)	Intermediate domesticated/wild. The shedding ability is generally strong
Wild	*Sounsouna* (literally "small *soun*")	Wild plants; extreme shedding ability

extraction was conducted twice and DNA was precipitated with isopropanol and washed twice with ethanol (70%). Dried pellets were re-suspended with 200 µl of TE buffer (pH 8).

We used the AFLP protocol described in Vos et al. (1995) with the silver-staining method described in Le Thierry D'Ennequin et al. (2000). A total of five primer combinations was used after total genomic DNA digestion with Eco-R1 et MSE-1: E-AAC/M-CAA, E-ACA/M-CTT, E-AAC/M-CTC, E-AAG/M-CAC, E-AAG/M-CTT. They were selected for their polymorphism and readability in previous analyses on pearl millet.

Statistical analyses

Morphological data were used to run univariate (ANOVA) and multivariate analyses [principal components analysis (PCA) and stepwise discriminant analyses]. Factors computed from the PCA procedure were used as synthetic variables in a two-factor nested ANOVA [categories nested under geographical (Alzu versus Kouré) localities]. A Tukey test was carried out for multiple mean comparisons. Morphological distances between groups were assessed using the stepwise discriminant analysis classification procedure with categories as a priori groups. The STATISTICA (6.0 version) software package was used.

AFLP data were used to compute dissimilarity indexes between individuals and .categories, referred to as "populations," based on the Jin and Chakraborty distance (1994). POPULATIONS (version 1.2.28, http://www.cnrs-gif/pge) software was used for this computation. An assignment procedure was implemented to evaluate the level of genetic introgression of weedy plants by genes from domesticated and wild categories. We used the "leave one out" procedure and the method proposed by Cornuet et al. (1999). This method is based on the use of distance values between individuals and reference groups. This procedure avoids hypotheses such as Hardy–Weinberg proportions and absence of linkage disequilibrium that are put forward in likelihood ratio calculations in more conventional assignment methods (Paetkau et al. 1995; Rannala and Mountain 1997). It is also appropriate for data obtained with dominant markers (e.g. AFLP). GENECLASS (version 1.0.02, http://www.ensam.inra.fr/CBGP) software was used. The reference groups used were the different categories (domesticated, *soun*-in, *soun*-out, wild) found in each locality (Alzu and Kouré).

The partition of the genetic variation (between localities and between categories) was assessed using *F*-statistics estimators from the between individuals dissimilarity matrix in a hierarchical analysis of the molecular variance (AMOVA) (Excoffier et al. 1992). ARLEQUIN version 2 (http://www.lgb.unige.ch/arlequin) software was used. The significance of the *F*-statistics results was tested by permutation procedures (1,000 permutations).

However, methods based on dissimilarity indexes estimated from dominant markers data use phenotypic (band presence versus absence) rather than genotypic information. Estimations of diversity index (Nei's genetic diversity) and population structure based on allelic frequencies (Lynch and Milligan 1994) were also obtained by using the AFLP-SURV software (http://www.ulb.ac.be/sciences/lagev/aflp-surv.html). Estimation of allelic frequencies are based on the assumption of Hardy–Weinberg proportion at each locus and have been made using the hypothesis of non-uniform prior distribution of allelic frequencies as it gives less biased estimators of allelic frequencies than alternative methods (Zhivotovsky 1999). However, concerning F_{st} values, as the two estimation methods (based on dissimilarity indexes or based on allelic frequencies) gave very similar results, only those obtained from the first one will be shown.

Results

Morphological characterization of weedy plants at the field level

An in situ morphological analysis was carried out in order to assess the level of diversity within and among *soun*-in and *soun*-out categories and to characterize them relatively to the domesticated and wild phenotypes. The overall picture of morphological diversity is given by the first two axes of the PCA (Fig. 1). The first axis (58.7% of the total phenotypic variance) is the most relevant to the differentiation between domesticated, *soun* and wild plants. It shows an opposition of domesticated phenotypes (plants having long broad candles with non-shattering spikelets and large bare seeds) to wild phenotypes (plants with loose and small candles, with shattering spikelets at maturity bearing small coated seeds). Tables 2 and 3 show the significant differences that are observed between cultivated, *soun* and wild plants on this axis. This was also true when the wild population was eliminated from the analysis. Figure 1 shows that *soun* have a large phenotypic distribution including plants with domesticated-like and wild-like phenotypes. In both categories of soun, plants showing different levels of combination of traits from domesticated and wild phenotypes are found.

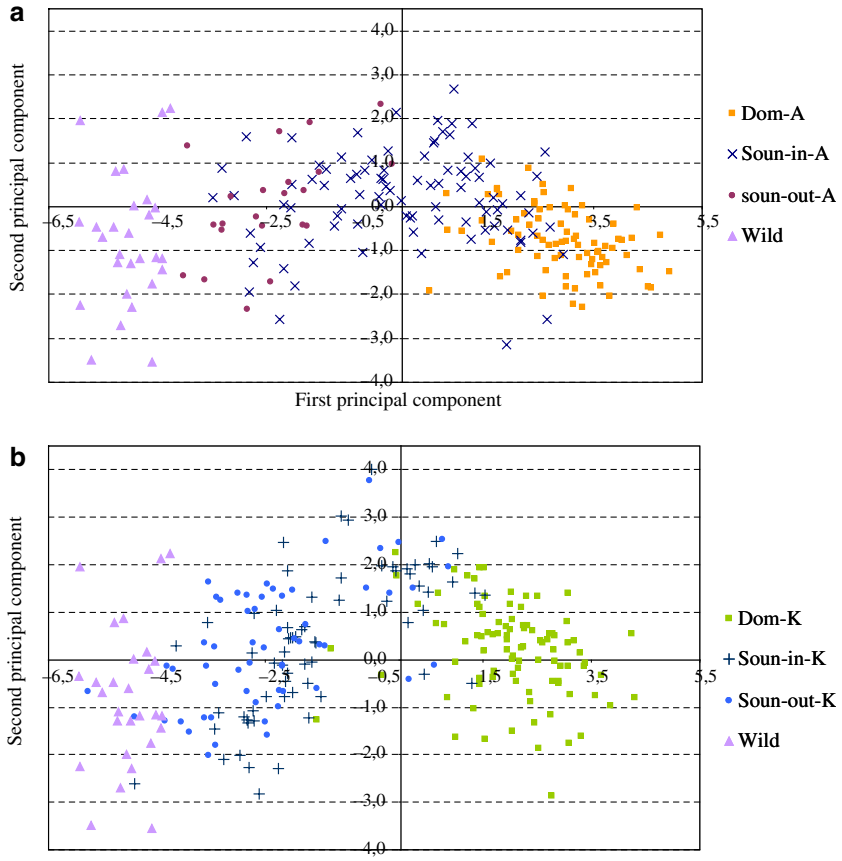

Fig. 1 Projection of domesticated, weedy and wild plants on the plane defined by the two first axes of a principal component analysis performed on domestication traits. The first and the second axes explain 58.7 and 12.4% of the total phenotypic variance, respectively. Plants from the Alzu field (**a**) and plants from the Kouré field (**b**) are presented separately for more clarity. Wild plants from Alzu are represented on each graph. The traits contributing most to the first component are: candle diameter ($r = -0.90$), rachis diameter ($r = -0.92$), candle length ($r = -0.88$) and the shattering score ($r = +0.91$). Those contributing the most to the second component are: glume length ($r = -0.87$) and the length of spikelet bristles ($r = -0.63$). *A* Alzu, *K* Kouré. The two landraces cultivated in each field were gathered together as they were in the same group as defined by using the multiple post-hoc mean comparisons Tukey test for each two first principal components (see Table 3)

Table 2 Results of the ANOVA on the first two axes of the PCA

Source of variation	*df*	First principal component			Second principal component		
		SS	*F*	*P*	SS	*F*	*P*
Locality	1	11.57	7, 26	*	47.61	36, 44	**
Morphological categories (nested in locality)	7	2,392.85	214, 55	**	53.67	5, 86	**
Error	432	688.26	–	–	564.49	–	–
Total	440	3,092.68	–	–	655.77	–	–

df degree of freedom, *SS* sum of square, *F* indicates *F*-test, *P* indicates associated probability of rejection of the null hypothesis

*$P < 1\%$

**$P < 1‰$

Soun display a higher morphological diversity than domesticated and wild plants. This is statistically confirmed by the discriminant analysis since as much as 29.7% (47/158) for *soun*-in and 36.2% (29/80) for *soun*-out were classified in other morphological categories (Table 4). Our data shows that the "locality" factor, although significant (Table 2), only explained 0.3 and 7.2% of the total phenotypic diversity for the first two principal components, respectively, despite the inclusion of wild phenotypes from Alzu in the analysis. The "category" factor explained 77.3 and 8.1% of the phenotypic variation observed for the two principal

Table 3 Repartition of each morphological category in groups as defined by using the multiple post-hoc mean comparisons Tukey test at 1% rejection level for each two first principal components

	Darancoba-A	Haini Kiré-A	Haini Kiré-K	Somno-K	*Soun*-in-A	*Soun*-in-K	*Soun*-out-A	*Soun*-out-K	Wild
First component	a	a–b	b	b	c	d	d–e	e	f
Second component	a	a	a–b–c	c	b–c	c	a–b	b–c	a–b

Morphological categories showing identical letters in the table belong to the same group as defined by the multiple post hoc mean comparisons

Table 4 Results of the discriminant analysis using the different morphological categories from the two localities as a priori groups for the classification. The table gives in each cell the results of the individual plant classification for each morphological category after the discriminant analysis (*lines*—observed morphological category, *columns*—a posteriori classification)

Observed morphological categories (farmer's classification)	Individual a posteriori classification								
	HK-A	D-A	*Soun*-in-A	*Soun*-out-A	HK-K	Somno-K	*Soun*-in-K	*Soun*-out-K	Wild
HK-A (N^a = 42; 69.0%[b])	29	6	4	0	3	0	0	0	0
D-A (N = 39; 66.7%)	7	26	2	0	3	1	0	0	0
Soun-in-A (N = 90; 58.9%)	4	4	53	9	6	0	9	5	0
Soun-out-A (N = 22; 72.7%)	0	0	2	16	0	0	2	1	1
HK-K (N = 67; 62.7%)	5	4	7	0	42	7	2	0	0
Somno-K (N = 27; 40.7%)	3	1	1	0	7	11	4	0	0
Soun-in-K (N = 68; 61.8%)	0	0	7	0	2	1	42	15	1
Soun-out-K (N = 58; 50.0%	0	0	6	5	0	0	14	29	4
Wild (N = 28; 85.7%)	0	0	0	2	0	0	0	2	24

HK-A Haini Kiré in Alzu, *D-A* Darancoba, *HK-K* Haini Kiré in Kouré

[a] Sample size

[b] Percentage of plants correctly assigned to their morphotype of origin (farmer's classification) after the discriminant analysis

components. Figure 1 and Table 3 highlight that *soun*-in and *soun*-out were significantly differentiated at the morphological level whereas s*oun* plants belonging to the same cultivation status (*soun*-in or *soun*-out) clustered together in the PCA representation even though they grew in two different villages and then in different conditions. Location of weedy plants in the field (in or outside drills) is thus associated with a morphological differentiation. Moreover, it was noteworthy that the two *soun*-out groups from the two villages were not significantly different for the first two axes (Table 3). *Soun*-out plants were phenotypically closer to wild plants than any other category, even those from Kouré that did not grow in the vicinity of wild populations (Fig. 1b; Table 4). Especially, *soun*-out have a higher seed shedding ability than *soun*-in. However, *soun*-out from the two localities were significantly different from wild phenotypes along the first axis (Table 3). On the other hand, *soun*-in plants were morphologically closer to domesticated phenotypes than *soun*-out plants, and very few of them had wild-like phenotypes (Fig. 1; Table 4). However, *soun*-in plants, irrespective of their geographic origin, were significantly different from domesticated plants along the first axis (Table 3). It is noticeable that farmers consider some *soun* plants, especially among *soun*-in, as usable plants. Nevertheless, all *soun* plants have more or less shattering spikelets.

Structure of genetic diversity and gene flow

Partitioning of molecular polymorphism among localities and categories

The five primer combinations generated a total of 133 polymorphic bands for the whole sample. No diagnostic band of the wild or domesticated groups (i.e. fixed in one group and absent in the other) was detected. Table 5 gives the levels of polymorphism found in each group. It is noticeable that the genetic diversity found in weedy plants in Alzu was a bit higher than in the cultivated varieties.

Most of the genetic variation was found within categories (Table 6). Slight but significant genetic differentiation between the genetic pools from the two localities was revealed by the AMOVA. This was mainly due to the contribution of the Alzu wild population since, as highlighted in Table 7, it was the most differentiated of all the other categories, irrespective of the sampling location (Alzu or Kouré). However, it

Table 5 Levels of molecular polymorphism within morphological categories estimated by the polymorphisme information content (PIC), the proportion of polymorphic loci (PLP) and the Nei's genetic diversity (He) estimated according to Lynch and Milligan (1994)

Morphological categories	Sample size	PIC (SE)	PPL	He (SE)
Darancoba-A	20	0.156 (0.012)	86.5	0.277 (0.014)
Haini Kiré-A	17	0.147 (0.011)	84.2	0.266 (0.014)
Soun-in-A	71	0.179 (0.011)	77.4	0.293 (0.015)
Soun-out-A	33	0.200 (0.011)	91.7	0.327 (0.014)
Wild	18	0.173 (0.013)	86.5	0.300 (0.014)
Haini Kiré-K	19	0.164 (0.012)	89.5	0.301 (0.014)
Somno-K	17	0.159 (0.012)	88.7	0.300 (0.013)
Soun-in-K	61	0.190 (0.011)	85	0.291 (0.015)
Soun-out-K	17	0.170 (0.012)	85.7	0.289 (0.015)

Standard errors (SE) for PIC were computed from 1,000 bootstrap locus resampling

also shows that domesticated and weedy plants clustered according to their geographical origin.

Differentiation among categories within a geographical origin explains only 6.98% of the variation (Table 6). However, if the wild population is eliminated from the analysis, this value decreases to 3.37%, thus indicating that this group of plants contributes mostly to the molecular variation between categories. Moreover, the different cultivated varieties grown in a same field were genetically very similar, even for the early and late flowering varieties from Kouré (Table 7).

The matrix of pairwise F_{st} computed between categories (Table 7) provides a more detailed picture of the molecular diversity in the crop/weed/wild complex. The most differentiated categories were the wild and domesticated ones. It is noticeable that the F_{st} between the wild population and the two domesticated populations in Alzu was very high, despite their parapatric situation. The F_{st} value comparing domesticated groups from the two localities was twofold lower. Because of their very high genetic similarity, the two varieties in each field (Alzu and Kouré) were lumped into the same group (called "domesticated") for the subsequent analysis.

The assignment procedure showed that 100% of the domesticated and wild plants from Alzu were correctly assigned to their own group (Table 8), whereas it was 91.6% for the domesticated group from Kouré. Moreover, none of these domesticated plants were assigned to the wild population or the *soun*-out groups. This confirms the very significant differentiation between these domesticated and wild forms. It also shows that

Table 6 Two-factor hierarchical AMOVA analysis on the basis of AFLP data carried out on the nine morphological categories (farmer's classification) from the two localities

Sources of variation	ddl	Variance component	Per cent of total molecular variance	*F*-statistics
Among localities (among fields)-	1	0.00769	6.16	$F_{ct} = 0.061$*
Among morphological categories (within locality)	7	0.00871	6.98	$F_{sc} = 0.074$**
Within morphological categories	264	0.10840	86.86	$F_{st} = 0.131$**

*Significant at 1% level

**Significant at 1‰ level

Table 7 Matrix of pairwise F_{st} values computed between each pair of morphological categories from genetic dissimilarity indexes

	Darancoba-A	Haini Kiré-A	*Soun*-in-A	*Soun*-out-A	Wild	Haini Kiré-K	Somno-K	*Soun*-in-K	*Soun*-out-K
Darancoba-A	0								
Haini Kiré-A	−0.014 (n.s.)	0							
Soun-in-A	0.021**	0.017*	0						
Soun-out-A	0.050**	0.049**	0.014**	0					
Wild	0.259**	0.266**	0.222**	0.138**	0				
Haini Kiré-K	0.114**	0.127**	0.130**	0.141**	0.305**	0			
Somno-K	0.152**	0.171**	0.150**	0.166**	0.300**	0.042**	0		
Soun-in-K	0.096**	0.093**	0.083**	0.078**	0.230**	0.042**	0.045**	0	
Soun-out-K	0.114**	0.121**	0.099**	0.096**	0.236**	0.107**	0.098**	0.032**	0

If taking into account for multiple testing by using sequential Bonferroni correction and a 5% global error risk, conclusions were unchanged

n.s. non-significant at 5% leve

*Significant at 1% level

**Significant at 1‰ level

Table 8 Results of the direct assignment of plants from Alzu and Kouré fields based on the DAS distance (adapted from Jin and Chakraborty 1994)

Morphological categories (sample size)	Composition of the groups after the assignment procedure						
	Domestic-A	Wild-A	Soun-in-A	Soun-out-A	Domestic-K	Soun-in K	Soun-out-K
Domestic-A (37)	37	0	0	0	0	0	0
Wild-A (18)	0	18	0	0	0	0	0
Soun-in-A (71)	54	0	17	0	0	0	0
Soun-out-A (33)	13	3	16	1	0	0	0
Domestic-K (36)	2	0	0	0	33	1	0
Soun-in-K (38)	14	0	0	0	17	2	5
Soun-out-K (40)	7	0	0	0	1	7	25

our dominant AFLP markers were highly discriminant despite the fact that no diagnostic bands were found.

Structure of the genetic variability of weedy plants

The results provided by molecular analyses show a rather odd genetic diversity pattern in *soun* groups from the two fields. Differentiation between *soun*-in and *soun*-out plants from the same field was much lower than between the same type of *soun* in the two fields (Table 7). However, all of these F_{st} were statistically significant.

The genetic proximity of *soun*-in plants to domesticated plants within each field was shown by the very low F_{st} values (0.024 and 0.035 in Alzu and Kouré, respectively) and by the results from the assignment procedure (Table 8). Most *soun*-in plants sampled in the two fields were assigned to the domesticated plant group (either from the same field or from the other locality). It is noticeable that only a minority of *soun*-in plants (19 out of a total of 109 *soun*-in) were assigned to their own group. Finally, only five *soun*-in from Kouré were assigned to the *soun*-out group (also from Kouré).

Soun-out plants were more differentiated from cultivated plants than *soun*-in plants were (Table 8). However, despite their strong morphological differentiation, the F_{st} between *soun*-out and domesticated plants from Alzu was only 0.057, whereas an F_{st} value of 0.09 was obtained for the same two types of plants from Kouré.

The *soun*-out plants in Alzu had the highest genetic similarity relative to the wild population (Table 7). Three *soun*-out from Alzu were classified in the wild population (Table 8). This could be the evidence that there was gene flow from the wild population to the field (pollen and/or seed flow). Nevertheless, this result is obtained if direct assignation is carried out. No *soun* plants were classified in the wild population when a 5% probability exclusion threshold was used (data not shown). Overall, a very large majority of weedy plants were classified in other groups than the wild population and *soun*-out plants from the two villages were more similar to the domesticated group from the same village than to the wild population (Tables 7, 8).

The assignment method we used is based on multilocus phenotypic information (dissimilarity indexes computed from the band presence/absence). This could lead to bias in the assignment of recombinant genotypes to hypothetical parental groups according to their introgression level if band presence frequencies are distorted in favour of one parental group. For example, a true F1 hybrid could preferentially be assigned to one parental group if this group displays an excess of locus with high frequencies of band presence relatively to the other parental group. Indeed, overall, the hybrid would also show an excess of band presence. In order to evaluate whether such biases in our assignment analysis were likely, we evaluated the eventuality of occurrence of such biases in our data by checking the distribution of band frequencies in the different morphological groups. The homogeneity χ^2 test [$\chi^2 = 39.73$ (n.s.), 54 *df* for all groups] clearly indicates that the distribution of the bands could be considered statistically as homogenous among groups. The same conclusion holds if one compares only the domesticated and wild groups in Alzu [$\chi^2 = 3.59$ (n.s.), 2 *df*]. Thence our assignment results are very likely unbiased which supports strongly the accuracy of the conclusions drawn from this analysis.

Moreover, the F_{st} between *soun*-out in Alzu and the wild population was even higher than the value obtained between the two *soun*-out groups sampled in the two localities (Table 7). Overall, this pattern of genetic diversity in *soun* groups strongly suggests that the level of genetic introgression of weedy plants by genes from the wild population in parapatry at "neutral loci" should be low.

Instead, data suggest that in Alzu at least, *soun*-out plants were largely introgressed by genes from cultivated (domesticated and *soun*-in) plants, as 39.4% of

which were classified with domesticated plants and 48.5% with *soun*-in plants, whereas only one plant was assigned to its own group. In the Kouré field, the situation was slightly different because a majority of *soun*-out plants were assigned to their own group.

Discussion

Structure of the morphological and genetical diversity in the pearl millet crop/weed complex

A significant structure pattern involving four differentiated groups (domesticated, *soun*-in, *soun*-out and wild) was observed, with respect to both morphological traits and molecular markers (AFLP), despite the broad continuous distribution of weedy phenotypes, ranging from nearly domesticated to nearly wild phenotypes. In both sites, *soun* found in drills (*soun*-in) were morphologically and genetically closer to domesticated phenotypes than *soun*-out. This was expected as *soun*-in originated from seed planted by farmers, and then mostly from selected candles derived from well-domesticated phenotypes. *Soun* can be considered as intermediate recombinant phenotypes between the domesticated and wild phenotypes. The high diversity observed within and between *soun* categories was therefore probably because of the multi-factorial inheritance of the domesticated phenotypes (Poncet et al. 2000). The marked phenotypic differences between *soun*-in and *soun*-out plants could thus reflect different introgression levels by wild genes at loci involved in "domestication" traits. Investigations carried out in several fields of the two villages have shown that *soun* phenotypes could represent up to 46% of the plants found in drills. The main question emerging from our results concerns the roles of the different factors controlling the abundance and the maintenance of these different types of weedy plants in the fields in this area.

Contribution of *soun* seeds to farmer seed stock

Soun-in and *soun*-out obviously therefore contribute, through pollination, to further introgression of traits from the wild into the selected seeds for next sowing. Thus, at least a part of *soun*-in plants probably originated from recurrent backcrosses between *soun*-in or *soun*-out and plants with domesticated phenotypes (mainly as the female parent). Contribution of the *soun*-out genetic pool to *soun*-in plants could also explain the high diversity of *soun*-in plants compared with domesticated ones. Their intermediate phenotypes could be explained by the dominance of alleles conferring the wild phenotype, as has been shown in QTL studies in pearl millet (Poncet et al. 2000). However, as total selection against dominant phenotypes should lead to a rapid decrease in the frequency of such phenotypes, quantification of the farmers' selection procedure is necessary to gain insight into the mechanisms underlying *soun* persistence in fields. A sizeable proportion (11.4%) of plants identified by farmers to be domesticated was in fact morphologically closer to *soun*-in plants. *Soun*-in seeds can therefore be unconsciously added to the seed stock through the harvest of not fully domesticated phenotypes.

The relatively low genetic differentiation between cultivated plants from Alzu and Kouré fields suggests that gene flow between the two localities has occurred, at least until a recent past. The harsh agricultural conditions in Niger lead to chronic deficits in seed production. Farmers used to get their seeds not only via their own production but also from neighbours, local sellers or regional markets where seeds harvested throughout the region and even other regions of Niger are sold (Robert et al. 2004). A similar situation has been observed in Rajasthan—India (Vom Brocke et al. 2003) where frequent importation of seed lots by farmers explains the very low differentiation between pearl millet populations cultivated in distant villages. This seems also to be the case for cultivated landrace populations in more selfing species [e.g. sorghum in Marocco (Djè et al. 1999) and tetraploid wheat in Ethiopia (Tsegaye et al. 1996)]. These data show that seed flow through seed exchanges between farmers rather than pollen flow could be therefore the main mechanism shaping molecular diversity in landraces at the regional scale. This raises the question of the consequences of seed exchanges on local adaptation and morphological evolution of cultivated populations. Our investigations have shown that seed sold to farmers on the markets is sometimes issued from non-selected plants, even *soun*. The contribution of external seed sources could therefore account for the high presence of *soun* in drills, which in turn could contribute, through their progenies, to the seed bank and therefore to *soun*-out populations.

Gene flow from the wild versus self-propagation of *soun*

The existence of a gene flow from the wild population to the field was actually suggested by the molecular analysis. However, despite mechanisms favouring potential gene flow between the wild and cultivated populations (domesticated and weedy plants), such as

outcrossing, parapatry and phenological similarity, the high genetic differentiation between them suggests that this gene flow is rather limited. The fact that this differentiation level is much higher than the differentiation between varieties from the two distant villages is consistent with this assumption. These patterns were not expected in Alzu if the *soun* plants were issued mostly from regular hybridization events between domesticated and wild populations. It could also be assumed that the wild population has experienced major demographic bottlenecks in the recent past. However, the similar level of genetic diversity observed in the wild population in comparison with the diversity observed within the other groups does not support this hypothesis.

Whether this finding could be generalized to wild-crop gene flow in pearl millet is still to be investigated. Nevertheless, it is consistent with previous data on this species. Low level of gene flow between cultivated and wild populations in one parapatry situation has already been hypothesized by Marchais (1994) on the basis of isoenzyme polymorphism analyses of progenies from wild and intermediate phenotypes. Pre-zygotic (Robert et al. 1991; Renno et al. 1997) or post-zygotic (Amoukou and Marchais 1993) isolation mechanisms already documented in pearl millet could explain this mating behaviour. This has also been shown in maize for which several components of the floral biology contribute to limit gene flow, especially from maize to teosinte (Baltazar et al. 2005). A similar high genetic differentiation between domesticated and wild neighbouring populations has also been described in the outcrossing *Beta* crop/wild complex in north of France (Viard et al. 2004). In this case, non-concordant peaks in flowering periods of wild and domesticated beets were considered as the main factor explaining this pattern.

Thus, our AFLP data strongly suggested that direct hybridization (pollen flow) between cultivated and wild populations was probably not the main process underlying the abundance of *soun* (in and outside the drills) in the Alzu field. Our data showed that the genetic differentiation between *soun*-out and cultivated plants from the same field (domesticated and *soun*-in) was low although significant. This suggests that the seed bank from which *soun*-out seeds originated mainly included seeds derived from crosses between *soun* (shattering) plants or between *soun* and domesticated plants. However, founder effects, and the fact that plants from different generations and even from different geographical origins (because of seed flows due to farmer exchanges) probably contribute to the seed bank, could be causal factors for the maintenance of a low, but significant differentiation. A minor contribution of wild pollinators to the diversity of *soun*-out populations could be another explanation for this phenomenon. Furthermore, spatial variation in the level of gene flow between adjacent wild and weedy populations could occur (Viard et al. 2004). Thus, additional studies including several parapatric situations between domesticated and wild populations and a more direct assessment of pollen flow through progeny analyses are needed to confirm our hypotheses.

Weeding and thinning efficiency

Weeding usually aims at eliminating plants found outside drills, including *soun*-out that are issued from spontaneous seed dispersal. However, the efficiency of this practice varies according to farmers and even years. We observed high *soun*-out plant densities after weeding (up to 30 plants/100 m^2). In such cases, these weedy plants probably contributed significantly, through pollination, to the production of *soun* plants in harvested seed stocks. Farmers hardly differentiate *soun*-in and domesticated seedlings, i.e. before flowering (A. Luxereau, personal communication). Thus, thinning seedlings has also a very low efficiency in the elimination of *soun*-in. Furthermore, *soun* plants are sometimes deliberately kept for two reasons. First, *soun* that are not eliminated before flowering are used as a food harvest complement, especially in harsh conditions. Second, farmers consider *soun* to be stressed pearl millets and not as originating from hybridization events. Elimination of these plants does not therefore appear to farmers as being related to maintenance of the landrace identity.

Fitness of wild-like phenotypes and soil seed banks

It is still unclear how intermediate phenotypes can be maintained in fields in the absence of recurrent gene flow from wild populations in situations where pollen from domesticated plants is abundant. Theoretically, because of farmer's mode of selection, recurrent backcrosses with domesticated phenotypes should lead to a *soun* phenotypic evolution towards a more domesticated phenotype through generations. However, our results clearly demonstrate that *soun*-out, *soun*-in and domesticated phenotypes are well differentiated at morphological level and farmers claims that *soun* have always been observed in pearl millet fields. Moreover, phenotypic similarity of *soun*-out plants found in different fields was observed. Several factors might explain the persistence of the phenotypic identity of *soun* populations. Quantitative genetics models have shown that population differentiation can be high at the trait

(phenotypic) level and low at the allelic frequency level for the majority of QTLs underlying the trait, especially when both gene flow and diversifying selection are strong (Le Corre and Kremer 2003). Phenotypes such as *soun* found outside the drills could be favoured by natural selection. This probably occurs since *soun* outside the drills generally have a higher shattering ability than plants found in drills. They should therefore contribute more than others to the seed bank. Furthermore, progenies issued from these plants will also tend to have a high shattering ability as this phenotype is dominant in pearl millet (Poncet et al. 1998). Consequently, if this selective advantage of shattering phenotypes is effective, a hitch-hiking process may lead to an increase in other wild characters such as candle morphology. This would be reinforced by linkage of domesticated genes in pearl millet (Poncet et al. 2002).

Conclusion

Our genetic results and investigations on farmers' practices suggest that the abundance of intermediate phenotypes in fields in southwestern Niger is mainly because of farmers' practices rather than the direct pollen flow and hybridization between domesticated and wild pearl millets. Seed flow through exchanges between farmers could be the main mechanism for *soun* maintenance and dispersal in the Zarma region, and even at a larger scale. Moreover, farmers have no means or even willingness to specifically eliminate *soun* in drills. This provides an opportunity for reinforcing seed banks via shattering plants and could lead to long-term persistence of *soun*-out plant populations. Thus, *soun*-out populations, rather than wild ones, provide a regular source of traits from the wild to be further introgressed in domesticated phenotypes. Hence, it is crucial to analyse the role of changes in farmers' practices and their economical and ecological constraints in order to better understand the evolution of crop/weed complexes.

Acknowledgements The authors wish to thank the technicians of the IRD team at Niamey for their technical assistance. The authors also thank A. Kaïrou, D. Bani, G. Idrissa and H. Abdoulaye, all students from CRESA-Niamey, for their precious contribution to this work. This work was supported by the Bureau des Ressources Génétiques (France).

References

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

Antonovics J, Bradshaw AD (1970) Evolution in closely adjacent plant populations. VIII. Clinal patterns at a mine boundary. Heredity 25:349–362

Baltazar MB, Sanchez-Gonzalez JJ, de la Cruz-Larios L, Schoper JB (2005) Pollination between maize and teosinte: an important determinant of gene flow in Mexico. Theor Appl Genet 110:519–526

Boudry P, Mörchen M, Saumitou-Laprade P, Vernet P, Van Dijk H (1993) The origin and evolution of weed beets: consequences for the breeding and release of herbicide resistant transgenic sugar beets. Theor Appl Genet 87:471–478

Brunken J, de Wet JMJ, Harlan JR (1977) The morphology and domestication of pearl millet. Econ Bot 31:163–174

Brush SB (1995) In-situ conservation of landraces in centers of crop diversity. Crop Sci 35:346–354

Busso CS, Devos KM, Ross G, Mortimore M, Adams WM, Ambrose MJ, Alldrick S, Gale MD (2000) Genetic diversity within and among landraces of pearl millet (*Pennisetum glaucum*) under farmer management in West-Africa. Genet Resour Crop Evol 47:561–568

Cornuet JM, Piry S, Luikart G, Estoup A, Solignac M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics 153(4):1989–2000

Djè Y, Forcioli D, Ater M, Lefèbvre C, Vekemans X (1999) Assessing population genetic structure of sorghum landraces from North-Western Morocco using allozyme and microsatellite markers. Theor Appl Genet 99:157–163

Elias M, McKey D, Panaud O, Anstett MC, Robert T (2001a) Traditional management of cassava morphological and genetic diversity by the Makushi Amerindians (Guyana, South America): perspectives for on-farm conservation of crop genetic resources. Euphytica 120:143–157

Elias M, Penet L, Vindry P, McKey D, Panaud O, Robert T (2001b) Unmanaged sexual reproduction and the dynamics of genetic diversity of a vegetatively propagated crop plant, cassava (*Manihot esculenta* Crantz) in a traditional farming system. Mol Ecol 10(8):1895–1908

Ellstrand N, Prentice HC, Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. Annu Rev Ecol Syst 30:539–563

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Jarvis D (1999) Strengthening the scientific basis of in situ conservation of agricultural biodiversity on farm. Bot Lithuania (Suppl 2):79–90

Jarvis D, Hodgkin T (1999) Wild relatives and crop cultivars: detecting natural introgression and farmer selection of new genetic combinations in agroecosystems. Mol Ecol 8:(s1)S159–S173

Jin L, Chakraborty R (1994) Population dynamics of DNA fingerprint patterns within and between populations. Genet Res 63:1–9

Le Corre V, Kremer A (2003) Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. Genetics 164:1205–1219

Le Thierry d'Ennequin M, Panaud O, Toupance B, Sarr A (2000) Assessment of genetic relationships between *Setaria italica* and its wild relative *S. viridis* using AFLP markers. Theor Appl Genet 100(7):1061–1066

Linder CR, Taha I, Seiler GJ, Snow AA, Rieseberg LH (1998) Long term introgression of crop genes into wild sunflower populations. Theor Appl Genet 96:339–347

Louette D, Charrier A, Berthaud J (1997) In situ conservation of maize in Mexico: genetic diversity and maize seed management in a traditional community. Econ Bot 51:20–38

Lynch M, Milligan BG (1994) Analysis of population structure with RAPD markers. Mol Ecol 3:91–99

Marchais L (1994) Wild pearl millet population (*Pennisetum glaucum*, Poaceae) integrity in agricultural Sahelian areas. An example from Keita (Niger). Plant Syst Evol 189:233–245

Morjan CL, Rieseberg LH (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. Mol Ecol 13:1341–1356

Muller M-H, Prosperi JM, Santoni S, Ronfort J (2001) How mitochondrial DNA diversity can help to understand the dynamics of wild-cultivated complexes. The case of Medicago sativa in Spain. Mol Ecol 10(12):2753–2764

Oka HI (1988) Weedy forms of rice. In: Oka (eds) Origin of cultivated rice. Elsevier and Japanese Science Society Press, Amsterdam, pp. 107–114

Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. Mol Ecol 4:347–354

Papa R, Gepts P (2003) Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. Theor Appl Genet 106:239–250

Pilate-André S (1992) Etude de l'organisation de la diversité génétique du complexe des mils pénicillaires (Pennisetum sp.) par les marqueurs enzymatiques et par l'analyse moléculaire de la région Adh. Thèse de 3ème cycle, Université Paris-XI

Poncet V, Lamy F, Enjalbert J, Joly H, Sarr A, Robert T (1998) Genetic analysis of the domestication syndrome in Pearl Millet. Heredity 81(6):648–659

Poncet V, Lamy F, Devos K, Gale M, Sarr A, Robert T (2000) Genetic control of domestication traits in pearl millet (*Pennisetum glaucum* L., Poaceae). Theor Appl Genet 100:147–159

Poncet V, Allouis S, Devos KM, Lamy F, Sarr A, Robert T (2002) Comparative analysis of QTLs affecting domestication traits among two domesticated × wild pearl millet (*Pennisetum glaucum* L., Poaceae) crosses and cereals. Theor Appl Genet 410:965–975

Pressoir G, Berthaud J (2004a) Patterns of population structure in maize landraces from the central valley 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

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. Proc Natl Acad Sci USA 94:9197–9221

Renno JF, Winkel T, Bonnefous F, Bezançon G (1997) Experimental study of gene flow between wild and cultivated *Pennisetum glaucum*. Can J Bot 75:925–931

Robert T, Lespinasse R, Pernès J, Sarr A (1991) Gametophytic competition as influencing gene flow between wild and cultivated forms of pearl millet (*Pennisetum typhoïdes*). Genome 34:195–200

Robert T, Luxereau A, Mariac C, Kairou A, Allinne A, Bani J, Beidari Y, Bezançon G, Cayeux S, Couturon E, Dedieu V, Moussa D, Sadou MS, Seydou M, Seyni O, Tidjani M, Sarr A (2004) 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* ssp. *glaucum* et ssp. *monodii*) dans deux localités du Niger. Actes du 4ème colloque national. Le patrimoine génétique: la diversité et la ressource. Bureau des Ressources Génétiques, Paris, pp. 223–245

Slatkin M (1987) Gene flow and the geographic structure of natural populations. Science 236:787–792

Slatkin M (1995) Epistatic selection opposed by immigration in multiple locus genetic systems. J Evol Biol 8:623–633

Suh HS, Sato YI, Morishima H (1997) Genetic characterization of weedy rice (*Oryza sativa* L.) based on morpho-physiology, isozymes and RAPD markers. Theor Appl Genet 94:316–321

Tostain S, Riandey MF, Marchais L (1987) Enzyme diversity in pearl millet (*Pennisetum glaucum*) 1. West Africa. Theor Appl Genet 74:188–193

Tsegaye S, Tesemma T, Belay G (1996) Relationships among tetraploid wheat (*Triticum turgidum* L.) landrace populations revealed by isozyme markers and agronomic traits. Theor Appl Genet 93:600–605

Viard F, Arnaud JF, Delescluse M, Cuguen J (2004) Tracing back seed and pollen flow within the crop-wild *Beta vulgaris* complex: genetic distinctiveness vs. hot spots of hybridization over a regional scale. Mol Ecol 13:1357–1364

Vom Brocke K, Christinck A, Weltzien ER, Presterl T, Geiger HH (2003) Farmer's seed systems and management practices determine pearl millet genetic diversity patterns in semiarid regions of India. Crop Sci 43:1680–1689

Vos P, Hogers R, Bleeker M, Reljans M, Van delee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 21:4407–4414

Wood D, Lenné JM (1997) The conservation of agrobiodiversity on-farm: questioning the emerging paradigm. Biodivers Conserv 6:109–129

Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. Mol Ecol 8:907–913