Theor Appl Genet (2007) 114:237–248
DOI 10.1007/s00122-006-0426-8

ORIGINAL PAPER

Local genetic diversity of sorghum in a village in northern Cameroon: structure and dynamics of landraces

Adeline Barnaud · Monique Deu · Eric Garine · Doyle McKey · Hélène I. Joly

Received: 17 February 2006 / Accepted: 6 October 2006 / Published online: 7 November 2006

Abstract We present the first study of patterns of genetic diversity of sorghum landraces at the local scale. Understanding landrace diversity aids in deciphering evolutionary forces under domestication, and has applications in the conservation of genetic resources and their use in breeding programs. Duupa farmers in a village in Northern Cameroon distinguished 59 named sorghum taxa, representing 46 landraces. In each field, seeds are sown as a mixture of landraces (mean of 12 landraces per field), giving the potential for extensive gene flow. What level of genetic diversity underlies the great morphological diversity observed among landraces? Given the potential for gene flow, how well defined genetically is each landrace? To answer these questions, we recorded spatial patterns of planting and farmers' perceptions of landraces, and characterized 21 landraces using SSR markers. Analysis using distance and clustering methods grouped the 21 landraces studied into four clusters. These clusters correspond to functionally and ecologically distinct groups of landraces. Within-landrace genetic variation accounted for 30% of total variation. The average F_{is} over landraces was 0.68, suggesting high inbreeding within landraces. Differentiation among landraces was substantial and significant ($F_{st} = 0.36$). Historical factors, variation in breeding systems, and farmers' practices all affected patterns of genetic variation. Farmers' practices are key to the maintenance, despite gene flow, of landraces with different combinations of agronomically and ecologically pertinent traits. They must be taken into account in strategies of conservation and use of genetic resources.

Communicated by H. H. Geiger.

Electronic supplementary material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00122-006-0426-8 and is accessible for authorized users.

A. Barnaud (✉) · D. McKey · H. I. Joly
CIRAD, UPR 67 Gestion des Ressources Génétiques et Dynamiques Sociales, UMR 5175-CEFE (Centre d'Ecologie Fonctionnelle et Evolutive), 1919 route de Mende, Montpellier 34293, France
e-mail: adeline.barnaud@cefe.cnrs.fr

M. Deu
CIRAD, UMR 1096-PIA, TA 40/03, Avenue Agropolis, Montpellier Cedex 5, 34398, France

E. Garine
Laboratoire d'Ethnologie et de Sociologie Comparative, CNRS, Université Paris X-Nanterre, MAE,, 21 allée de l'Université, Nanterre 92023, France

Introduction

In traditional agroecosystems, farmers grow a large diversity of species and landraces. This diversity is often culturally important, and it may also lower the risk of crop failure owing to vagaries of climate, diseases, pests, and soil limitations (Teshome et al. 1997; Brush 2000). Locally adapted landraces usually produce lower yields during optimal conditions than "improved" varieties, but the relative stability of their yields provides food security to households. In addition, some local landraces are uniquely suited for particular uses. However, not all named and defined landraces have specific uses or agronomic characteristics. Landraces result from ongoing evolutionary processes of domestication, in agricultural systems where

farmers continue to select their seeds. Farmers' practices strongly influence these processes.

Understanding landrace diversity is of interest in the study of evolutionary forces under domestication, and has applications in the design of programs for the conservation, management and use in breeding programs of genetic resources. Landraces are highly pertinent for studying evolutionary forces, because they are cultivated in a dynamic situation where human and environmental selection, gene flow, and genetic drift all interact to shape genetic diversity. In conservation programs, diversity within and among landraces can confer longer-term adaptation of crop populations to fluctuating and heterogeneous environments. In breeding programs, the gene pool of landraces constitutes an important source of valuable germplasm with many specific ecological adaptations. In the particular case of sorghum, the instability of yields of improved varieties under local conditions, and the extension of cultivation to marginal lands, both point to the necessity of breeding programs to fit the needs of small-scale farmers (Haussmann et al. 2000). To understand the dynamics of diversity in such settings, it is first necessary to document the pattern of genetic diversity at a very local scale both within and among landraces. Sorghum, *Sorghum bicolor* (L.) Moench, is a main crop throughout semi-arid regions of Africa and Asia. Harlan and de Wet (1972) recognized five basic races [bicolor (B), caudatum (C), durra (D), kafir (K) and guinea (G)] and ten intermediate forms of cultivated sorghum, defined on the basis of spikelet, seed, and panicle morphology. The extent and pattern of genetic diversity within large germplasm collections of sorghum is well characterized (Djè et al. 2000; Grenier et al. 2000a, b; Casa et al. 2005; Deu et al. 2006), but it may not reflect genetic diversity of landraces at the local level. To date, sorghum landraces have been studied at the country level (Ghebru et al. 2002; Nkongolo and Nsapato 2003; Uptmoor et al. 2003) or regional level (Djè et al. 1999), but not at a local scale. Furthermore, to our knowledge, existing studies of sorghum landraces have used only few individuals and have not taken into account diversity within landraces. Ghebru et al. (2002) showed an exceptionally high level of genetic diversity among 28 Eritrean landraces compared to a sample of 32 accessions of the world sorghum collection. This is reminiscent of the great diversity of cassava landraces in a single Makushi village in Guyana compared to the world core collection (Elias et al. 2000a). Seed exchange, pollen flow, farmers' practices, and environmental pressures all affect genetic diversity in situ. It is therefore important to assess diversity and structure of landraces at the scale of a village, taking also into account farmers' knowledge associated with cultivation and uses of landraces.

Our study was conducted among Duupa farmers in the village of Wanté in subsahelian northern Cameroon. The Duupa (about 4,000 people) live in an area of 1,000 km^2 in the plain of the river Benoué and the mountainous massif of Poli. The Duupa are sedentary farmers. Sorghum cultivation is central to Duupa agriculture, which is directed towards subsistence rather than to production for markets. Forty-six sorghum landraces have been described in the village, presenting a broad spectrum of morphological diversity. Most landraces belong to the guinea race, the other races being present in intermediate forms (GC, DC, BD or KD) (Garine 1995; J. Chantereau, personal communication).

Sorghum is a wind-pollinated annual crop that is considered to be predominantly selfing (Chantereau and Nicou 1991; Doggett 1988), although outcrossing rates from 0.10 to 0.30 have been reported (Ellstrand and Foster 1983; Doggett 1988; Ollitrault et al. 1997; Djè et al. 2004). The spatial pattern in which landraces are grown in fields thus affects mating structure of landraces. Our observations show that fields are composed of several landraces, from 4 to 20 (mean 12; 19 fields surveyed) (unpublished data). Seeds from different landraces are mixed in a common bowl before sowing (Alvarez et al. 2005). Duupa farmers choose sorghum seeds mainly from their own production, with a small proportion coming from fields of other farmers in the community. Entire panicles are chosen for seeds, either in the field or during the period between harvesting and threshing.

The Duupa practice of planting numerous landraces closely mixed in a field should lead to extensive gene flow among landraces. Three main explanations for the persistence of differentiated landraces can be proposed. First, as yet unidentified biological barriers (phenological differences, incompatibility) may minimize gene flow among landraces. Second, selection by farmers for particular combinations of morphological traits in the choice of seeds for sowing may preserve the identity of landraces. Third, it could be that landraces are not strongly differentiated. Individuals of a given landrace might, for example, share a few major genes responsible for morphological distinctiveness, while showing little differentiation at other loci (e.g., neutral SSR markers).

Molecular tools have shown their efficiency and their cost effectiveness for assessing genetic diversity. Microsatellites or simple sequence repeat (SSR) markers have been recently developed for sorghum (Brown et al. 1996; Taramino et al. 1997; Menz et al. 2002; Schloss

et al. 2002). SSR loci gave good discrimination between closely related individuals of sorghum even when only few loci were used (Djè et al. 1999; Smith et al. 2000; Ghebru et al. 2002). In this paper we assess the pattern of genetic diversity of sorghum landraces in Wanté, using a set of 14 SSR and a sample of individuals that covered most of the range of morphological and taxonomic diversity. We describe how historical factors, variation in breeding systems, and farmers' practices and perception affect patterns of genetic diversity.

Materials and methods

Study site

The village of Wanté (8°27′N, 13°18′E) is close to the unpaved road linking Poli (10 km west of Wanté) to Garoua. Wanté is thus a relatively isolated village, where approximately 20 families share lands covering 10 km^2. Duupa agriculture is characterized by a high diversity of crop species as well as by a diversity of landraces of several crops, particularly sorghum (Garine 1995; Alvarez et al. 2005). So far, 59 folk terminal taxa named by farmers have been recorded, representing 46 landraces (some synonymies exist) cultivated in the village.

Sorghum cultivation is dynamic in both space and time. The location of a farmer's field may change from year to year, and the same location may be used by different farmers in successive years. There is no perceived shortage of land. The local seed exchange networks lead to frequent seed flow among farmers. Farmers can give or receive seeds, mainly at harvest and threshing times. These exchanges follow social rules. For example, young farmers are more likely to receive seed than old farmers, and exchange of seed is never a matter of monetary transaction (Alvarez et al. 2005). Sorghum cultivation is also dynamic in time. Each year farmers select seeds for sowing their next crop, choosing the landraces they will grow in their field. From one year to the next, a farmer will always cultivate some landraces while omitting others, according to individual preferences. A farmer may also choose to use only seed from harvests of other farmers for sowing, for example, after illness or because of a change in location of his or her field (personal observation).

Farmers' practices and perceptions concerning sorghum landraces

Duupa sorghum fields are sown from April to May, corresponding to the beginning of the rainy season. Sorghum is harvested from December to January and threshed from February to March. In order to understand farmers' practices and perceptions of sorghum landraces, we conducted interviews, free listing exercises, and observations during fieldwork totaling 6 months spread over 3 years, covering one sowing time and two harvesting and threshing times. Ethnological data were obtained with the help of two Duupa interpreters who have been working with the anthropologist of the project for many years (Garine 1995). This work on management and perceptions of sorghum builds on more than 15 years of anthropological study, which has documented social structure, agricultural systems, and ethnobiological knowledge of the Duupa.

Farmers choose to grow several landraces in each field. To understand the distribution of landraces within and among fields, we studied two transects per field in 19 fields at a time when plants were mature, allowing reliable identification of panicles to landraces. Each transect was 30 m long and 1 m broad. We counted the number of panicles per landrace over the transects (mean of 92 panicles per transect) as a measure of landrace representation in fields. We calculated the average density of each landrace in fields and the percentage of fields in which a given landrace was cultivated. All transects were conducted at harvesting time in 2003–2004.

To understand farmers' perceptions of the diversity of sorghum, we used free listing methods to explore the cultural salience of the various landraces. The technique consists of asking a small set of respondents to name all items matching a given description. We asked the respondent to "name all the sorghum landraces you know". We calculated an index of item saliency, the Smith index, computed in ANTHROPAC (Borgatti 1996). The Smith index (SI) is essentially a weighted average of the inverse rank of an item across multiple free lists, where each list is weighted by the number of items in the list. A high value of SI underlines the high social value of a given landrace, whereas a low SI (or the absence of a landrace from free lists) implies a low social value. Forty-five farmers (12 women and 33 men) of Wanté were interviewed in this survey. Landrace names given here are in the Duupa language and written in italic. Duupa phonemes include the glottal stop, transcribed here as an apostrophe (‘). Our sample size, while relatively small, is exhaustive, including almost all households of the village.

Plant material

Of the 46 landraces identified in the village, only 30 were found in transects carried out in 19 fields. Therefore, in

December 2003 we extended the sampling to 27 fields spread throughout the village's lands in order to provide a better representation of the landrace diversity found in Wanté. In final, we retained for the SSR analysis 21 landraces (293 plants) for which at least two plants had been sampled (Table 1). These plants (2 to 22 per landrace) were sampled during harvest or threshing time (threshing areas are near each field). Two Duupa interpreters identified all these plants and each cultivator of the 27 fields was interviewed. From each plant, two seeds were grown in the greenhouse at the CNRS campus in Montpellier (France). DNA was extracted from leaf tissue of 3-week-old plants (one individual assayed per maternal parent) using the Qiagen DNeasy 96 plant kit.

SSR genotypes

Fourteen SSRs [listed as electronic supplementary information (EMS) S1] of known map location and distributed throughout the 10 linkage groups (Kim et al. 2005) were assayed on the 293 plants. SSRs used were chosen among those developed and mapped by Brown et al. (1996), Chittenden et al. (1994), Paterson et al. (1995), Taramino et al. (1997), Bhattramakki et al. (2000), Kong et al. (2000), Menz et al. (2002), and Schloss et al. (2002). SSRs were also chosen from a further set (gpsb) developed at CIRAD through the construction of an SSR-enriched gDNA library (Genoplante project). Among these SSRs, developed in the framework of the Generation Challenge Program, loci were chosen for their high polymorphism and for the ease with which results could be unambiguously read and scored. The M13-tails added to forward primers for each SSR were labeled with IRD700 or IRD800 fluorochromes. Plants were genotyped at the Montpellier Languedoc-Roussillon Genopole platform located on the CIRAD campus in Montpellier (France). Polymerase chain reactions (PCR) were carried out in a 10 µL reaction mix containing 25 ng (5 µL) of template DNA, 1 µl PCR buffer (10 µM Tris, 50 µM KCl and 0.01% of glycerol), 0.1 U of *Taq* DNA polymerase, 2.5 mM $MgCl_2$, 200 µM dNTPs and 0.1 µM of forward and reverse primers and M13 tail. PCR cycling conditions were as follows: 4 min initial denaturation at 94°C, 10 cycles of amplification with the shutdown method (−0.5°C per cycle) [45 s at 94°C, 1 min at TM + 5°C, 1 min 30 s at 72°C], 25

Table 1 Names, racial classification, summary agromorphological characteristics, abundance and frequency in fields, and salience to farmers of the 21 sorghum landraces analyzed

No.	Folk taxonomy	Race	SC	GC	PASH[a]	Cycle[b]	SI	Dens[c]	%*F*[d]
1	*'angonga*	KD	Yellow	Black	3	L	0.18	<1	16
2	*baa dangkaliya*	GC	White	White	2	L	0.28	1	68
3	*beng kpankpan*	C	White	White	3	E	0.01	0	0
4	*benga*	GC	White	Black	2	E	0.30	<1	11
5	*beng naa siii*	–	Purple	Black	3	E	–	0	0
6	*bonna*	KDC	Yellow	Yellow	3	L	0.10	<1	21
7	*gbansaa*	DC	Straw	Straw	2	E	0.40	<1	16
8	*gbarat ba'a*	G	Red/white	Black	2	L	0.01	1	68
9	*goo beee*	G	White	Red	2	L	0.10	7	95
10	*goo tiii*	G	White	Black	2	L	0.07	12	100
11	no name[e]	KD	Red	Red	4	L	–	<1	11
12	*kubaze kolla*	G	Orange	Red	2	L	0.37	39	100
13	*murzumma*	G	Pink	Black	2	L	–	10	100
14	*nam vaa*	GC	White	Black	1	L	0.22	1	63
15	*see gooriya*	BD	White	Straw	1	L	0.25	<1	53
16	*se' kukka*	GC	White	Black	3	L	0.24	2	84
17	*vee nimma*	G	Red	Black	2	L	0.33	5	100
18	*yatta*	DC	Purple	Straw	2	L	0.18	<1	42
19	*yen waa*	DC	White	White	2	L	0.04	<1	37
20	*za' toota*	D	Yellow	Black	3	L	0.10	<1	63
21	*zormee*	G	Orange	Black	2	L	0.20	11	100

Race (*B* Bicolor, *C* Caudatum, *D* Durra, *G* Guinea, *K* Kafir); *SC* seed color; *GC* glume color; *SI* Smith index (see text for explanation)

[a] PASH = shape of the panicle encoded as 1: very loose; 2: loose; 3: compact; 4: very compact

[b] Cycle = early or late flowering, according to farmers

[c] Dens = percentage of panicles encountered per landrace, in transects in 19 fields (taking into account only fields in which the landrace was present)

[d] %*F* = percentage of fields in which a given landrace was cultivated (of 19 fields)

[e] Landrace for which farmers did not give a name

cycles of amplification [45 s at 94°C, 1 min at TM, 1 min 30 s at 72°C], and a final elongation of 4 min at 72°C. PCR reactions were performed on an MJ research Dyad 384 wells PCR. PCR-amplified fragments from differentially labeled SSR primers and with non-overlapping sizes were simultaneously pooled and run in the same gel (denaturing polyacrylamide gel at 6.5%) on a Li-Cor IR2. Saga GT v.2.2 (Li-Cor) was used for automated data collection and to determine allele sizes. Two internal size standards in each well, and three control samples per SSR, were also used. Each control sample was a bulk sample of three or four different individuals (Generation Challenge program).

Diversity analyses

In order to assess levels of genetic diversity, basic statistics were computed. Number of alleles, observed heterozygosity, and gene diversity (expected heterozygosity) corrected for small sample size (Nei 1978) were calculated for each SSR locus and for multi-locus genotypes of individuals (for landraces with more than 16 individuals sampled) using GENETIX 4.04 software (Belkhir et al. 2002).

Analysis of population structure

In order to assess the structure of genetic diversity within and among landraces we used four complementary approaches: *F*-statistics, a neighbor-joining analysis, a Bayesian model-based clustering method, and an analysis of molecular variance. Considering the 12 landraces for which enough individuals were sampled, F_{st}, F_{is}, and F_{it} (Weir and Cockerham 1984) were also computed as measures of the genetic diversity within and among landraces. Permutation procedures (10,000 permutations) were performed to test the significance of differences between values. Calculations were carried out using GENETIX 4.04.

The dissimilarities between all pairs of individual plants were estimated based on simple matching. Robustness was estimated by using 1,000 bootstrap resamplings. Dissimilarity matrices were computed and neighbor-joining (NJ) analyses were performed on them using Darwin software (Perrier et al. 2003).

We also used the Bayesian model-based clustering method of Pritchard et al. (2000), implemented in the software STRUCTURE 2.1. (http://www.pritch.bsd.uchicago.edu). This method assumes that each genotype in the sample may result from the admixture of an unknown number of differentiated ancestral populations, with membership coefficients totaling 1. We used the basic admixture model with unlinked loci and uncorrelated allele frequencies, with the assumed number of populations (K) varying from 1 to 21, 10 replicate runs per K value, a burning period length of 10^6, and a post-burning simulation length of 1.5×10^6. No a priori population information was used. The run showing the highest posterior probability of data was considered for each K value.

The genetic structure of sorghum landraces was further investigated by an analysis of molecular variance (AMOVA) using ARLEQUIN 3.0 (Excoffier et al. 2005). Tolerance was set to 5% of missing data per locus. The significance of the partitioning of genetic variance among groups was tested. Groups were defined according to landraces, and clusters defined by STRUCTURE analysis. As with *F*-statistics, these analyses included only the 12 landraces for which sample size was sufficient for characterizing intra-landrace genetic diversity. Furthermore, to test the variance among clusters, we only used individuals with a cluster membership probability higher than 80% (190 plants analyzed).

Results

Spatial patterns of planting and farmers' perception of landrace diversity

Analysis of spatial distribution showed great variability among landraces (Table 1). The six most frequent landraces (numbers 9, 10, 12, 13, 17, 21) together represented most of the farmers' harvest. Comparison of density in fields and percentage of fields in which a given landrace was cultivated allowed grouping of landraces into three categories: frequently present in fields and abundant when present (9, 10, 12, 13, 17, and 21), frequently present but rare in each field (2, 8, 14, 15, 16, and 20), and infrequently present and rare (1, 3, 4, 5, 6, 7, 11, 18, and 19). Analysis of free listing revealed that people cited from 2 to 20 landraces per list (mean = 9.35). The SI varied from 0.01 to 0.40, with elevated values indicating landraces reported at the beginning of the list and frequently cited. In our case, SI was highly correlated (Spearman's $r = 0.98$, $P < 0.0001$) with the frequency with which a landrace appeared in different persons' lists, and only weakly correlated with its other defining component, the order (rank) in which the landrace appeared in lists. Two groups can be defined according to the SI: the first (SI > 0.1) is composed of 11 landraces (1, 2, 4, 7, 12, 14, 15, 16, 17, 18 and 21), the second group consisting of the 9 remaining landraces (SI $\leq$ 0.1).

Polymorphism and allelic richness

All 14 loci retained were polymorphic and revealed a total of 123 alleles. The number of alleles per locus ranged from 3 (gpsb114, xcup53) to 25 (xtxp348) with a mean of 8.79 (listed as electronic supplementary information: S1). Observed heterozygosity varied widely, from 0.01 (gpsb114) to 0.39 (xcup02), with a mean of 0.11. Gene diversity, number of alleles, number of private alleles, and F_{is} are shown for each landrace in Table 2. Most landraces presented a high number of alleles. Twenty-one percent of alleles were private alleles, i.e., specific to a given landrace. The number of private alleles ranged from zero to five and was highest in intermediate forms (DK or DC). With only eight individuals sampled, *'angonga* presented five private alleles. Gene diversity varied little, from 0.30 (*se' kukka*) to 0.47 (*zormee*), except for two landraces with much lower gene diversity: *see gooriya* (0.08) and *nam vaa* (0.21). F_{is} values varied widely among landraces, from 0.53 (*gbarat ba'a*) to 0.85 (*se' kukka*).

Genetic diversity and population structure

We observed a mean F_{it} value of 0.79, which is highly significant. The average F_{is} across landraces was 0.68, suggesting high inbreeding levels within landraces. For the whole sample considered as a single population, we observed an F_{is} value of 0.78 (Jacknife SD = 0.002). There was a substantial and significant degree of differentiation among landraces (F_{st} = 0.36). The NJ analysis presented in Fig. 1 revealed that two landraces, *see gooriya* (15) and *yatta* (18), formed a distinct cluster. Among the others, guinea and guinea caudatum landraces were differentiated from non-guinea landraces. A slight differentiation was also detected within guinea/GC landraces. Four sub-groups could be identified, sub-groups G1 and G4 being mainly constituted of red-grained landraces and G2 and G3 by white-grained landraces. This analysis also shows that individual plants from the same landrace tended to cluster together, especially for the non-guinea landraces 6, 7, 11, and 20, in addition to the highly distinctive 15 and 18. However, for the guinea landraces, individual plants from a given landrace were often scattered through the four different guinea sub-groups.

Table 2 Diversity statistics for the sorghum landraces analyzed

Folk taxonomy	No.	Race	*N*	GD	F_{is}	N.AT	N.Ap
'angonga	1	KD	8	–	–	49	5
baa dangkaliya	2	GC	16	–	–	36	1
beng kpankpan	3	C	2	–	–	16	1
benga	4	GC	3	–	–	27	2
beng naa siii	5	–	2	–	–	14	–
bonna	6	KDC	8	–	–	27	1
gbansaa	7	DC	6	–	–	37	1
gbarat ba'a	8	G	18	0.37	0.53	43	2
goo beee	9	G	19	0.34	0.63	45	–
goo tiii	10	G	21	0.36	0.66	45	–
no name	11	KD	4	–	–	26	–
kubaze kolla	12	G	22	0.36	0.56	48	1
murzumma	13	G	22	0.35	0.54	43	–
nam vaa	14	GC	18	0.21	0.66	34	–
see gooriya	15	BD	18	0.08	0.78	20	–
se' kukka	16	GC	17	0.30	0.85	38	1
vee nimma	17	G	20	0.37	0.71	50	3
yatta	18	DC	18	0.31	0.59	39	2
yen waa	19	DC	15	–	–	33	3
za' toota	20	D	18	0.37	0.81	37	1
zormee	21	G	18	0.47	0.84	34	2
Mean				0.32	0.68	35	

Race (*B* Bicolor, *C* Caudatum, *D* Durra, *G* Guinea, *K* Kafir); *N* number of plants analyzed per landrace; *GD* gene diversity; *N.AT* total number of alleles observed; *N.Ap* number of private alleles; – not determined

The program STRUCTURE estimates the most likely number of clusters (*K*) by calculating the log probability of data for each value of *K*. However, detecting the 'real' number of clusters is not always straightforward, and several considerations must be taken into account. For example, the model used in the program assumes Hardy–Weinberg equilibrium within clusters, and Pritchard et al. (2000) suggested that this might lead to an overestimation of the number of clusters. For our results, the Bayesian posterior probability of data increased until $K = 2$ and to a lower extent up to $K = 6$. The largest proportion of individuals (92%) assigned to a specific cluster with a cluster membership probability higher than 80% was obtained for $K = 3$, but the percentage of individuals assigned to a specific cluster remained high (higher than 80%) until $K = 6$. At $K = 6$ the clusters obtained changed completely from those successively defined from $K = 2$ to $K = 4$. Finally, in contrast to clusters obtained with these smaller values of *K*, we could find no biological basis for the clusters defined when $K = 5$ and $K = 6$. We thus chose $K = 4$ as the most likely number of clusters.

The clusters obtained with $K = 2$ to $K = 4$ are reported in Fig. 2. The clusters identified with $K = 2$ (Fig. 2a) correspond largely to a racial structure between guinea/guinea-caudatum landraces (shown in green) and non-guinea landraces (shown in red). The groups identified when $K = 3$ (Fig. 2b) clearly revealed a subdivision of the previous non-guinea cluster and confirmed the distinction between *see gooriya* and *yatta* (shown in red) and other non-guinea landraces (shown in blue). When four clusters were considered (Fig. 2c), the previous guinea cluster was subdivided into two. One of these corresponds to the group of

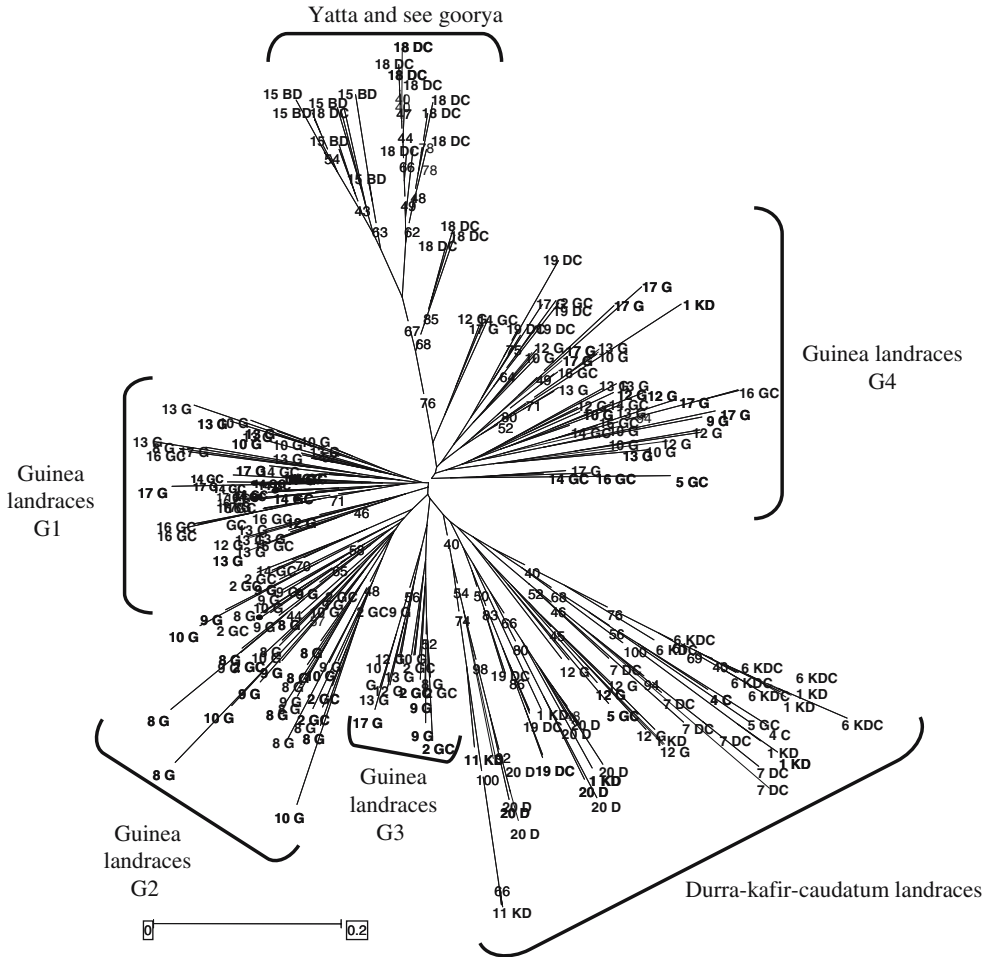

Fig. 1 Neighbor-joining tree based on 14 SSRs among sorghum plants, using the simple matching index. One thousand bootstraps were performed, and only those nodes with bootstrap values >40 are indicated in the figure. Sorghum plants are identified by the number of the landrace (see Tables 1 and 2) and by their race (*G* guinea, *C* caudatum, *D* durra, *K* kafir, *B* bicolor). *G1* and *G4* refer to two clusters of red-grained guinea landraces, and *G2* and *G3* to two clusters of white-grained guinea landraces. For some landraces, multiple points are superposed

landraces categorized by the Duupa as *toot zeyna buyya* (white-grained sorghums, shown in green in Fig. 2c), the other to *toot zeyna bee* (red-grained sorghums, shown in yellow), in agreement with the results of the NJ analysis (see previous).

Most individual plants (95%) of a given landrace were clustered together, except for four landraces (*goo tiii* [10], *nam vaa* [14], *se' kukka* [16] and *yen waa* [19]), in which individual plants were assigned to two different clusters. For *yen waa*, the placement of individuals in two different clusters might be due to an error in morphological identification at sampling time. The landrace *goo tiii* is defined only by the glume color, and its lack of strong morphological identity may help explain its weak genetic structure. Similarly, landraces 14 and 16 are morphologically defined based only on panicle shape, which may explain their weak genetic structure.

AMOVA conducted on the clusters defined by STRUCTURE analysis (for $K = 4$) and on landraces showed that all variance components were highly significant ($P < 0.001$) for both clusters and landraces (Table 3). Of the total variation, the highest percentage of SSR variability was attributed to variability among clusters (47.24%) and within landraces (30.40%).

Discussion

Our study, carried out at the scale of a single village, detected relatively low genetic diversity [average gene diversity (GD) per landrace = 0.32, average gene diversity per locus = 0.51] compared to values observed in other studies at regional scales. For example, Djè et al. (1999) reported GD = 0.83 for a sample throughout

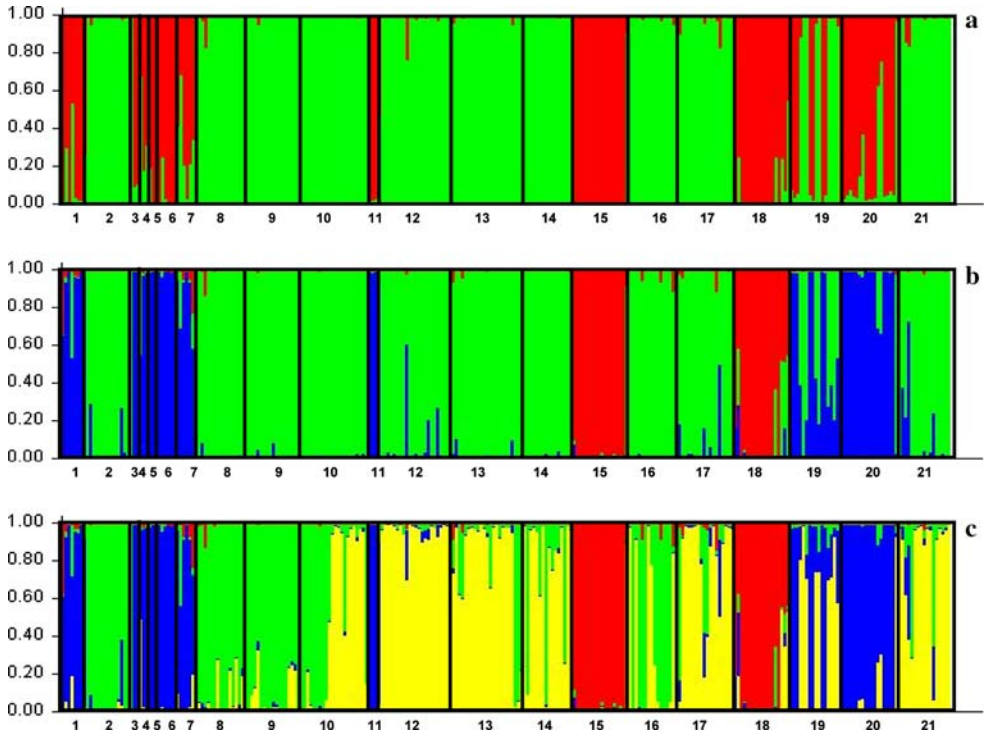

Fig. 2 Structure of the genetic diversity of the 293 sorghum plants (sorted by landrace) as estimated using the model-based Bayesian algorithm implemented in the program STRUCTURE (Pritchard et al. 2000). Cluster memberships for each plant are shown as estimated using different numbers of hypothetical clusters, $K = 2$ (**a**), $K = 3$ (**b**), and $K = 4$ (**c**). The clusters identified correspond quite well to racial and agromorphological entities. For $K = 2$, the *green* cluster corresponds to guinea/guinea-caudatum landraces, the *red* cluster to non-guinea landraces. For $K = 3$, the non-guinea landraces are further subdivided into *see gooriya* and *yatta* (in *red*) and the others (in *blue*). For $K = 4$, the *green* cluster corresponds to guinea landraces with white seeds, the *yellow* cluster to guinea landraces with red seeds, the *red* cluster to *see gooriya* and *yatta*, and the *blue* cluster to durra-kafir intermediates

Table 3 Summary of results of AMOVA

Source of variation[a]			
	Among clusters or landraces	Among landraces within clusters	Within landraces
$K = 4$	47.24	22.36	30.40
Landraces	59.34	–	40.66

Values given are percentage of variation

[a] All sources of variation were significant ($P < 0.001$)

five regions in Morocco, and Uptmoor et al. (2003) reported GD = 0.59 for 23 landraces from southern Africa. We detected a considerable number of private alleles. This number is probably over-estimated due to the small sample size of some landraces analyzed, and to the fact that all landraces analyzed represented only about one-half of all landraces present in the village. We found great within-landrace genetic variability (30.4%), suggesting that increased sample sizes per landrace might well have detected further variation. These considerations suggest that our results underestimate the genetic diversity existing in the village.

The 21 landraces studied were grouped in four clusters, which corresponded to morphologically and ecologically distinct groups. Several factors could contribute to the pattern found.

First, history has an important influence. The observed genetic structure is in part a legacy of structure in ancestral sorghum populations, in which geographic differentiation, leading to the five races, has been followed by complex patterns of diffusion allowing secondary contact. Many studies have indicated that the organisation of sorghum diversity is linked to geographic and/or racial classifications (Aldrich et al. 1992; Tao et al. 1993; Deu et al. 1994, 2006; Cui et al. 1995; de Oliveira et al. 1996; Menkir et al. 1997, Djè et al. 2000). Our results also reflect such racial structure. Diffusion is an ongoing process, and landraces cultivated in Wanté vary in the time since their introduction into the village. A reasonable hypothesis is that more recently introduced landraces are genetically more distinct. In our sample, a landrace that has not yet been given a name by the Duupa (no. 11 in Table 1) is the most recently introduced, and all its individuals are close to each other in the NJ tree (Fig. 1).

Second, the predominantly autogamous breeding system of sorghum also contributes to explaining the patterns of genetic diversity and structure observed. Djè et al. (1999) reported a value of $F_{is} = 0.63$ (comparing different fields). In our study, average F_{is} (comparing different landraces, not fields) was 0.68. These F_{is} values are both consistent with a predominantly autogamous mating system. A selfing rate of $s = 0.81$ would be expected under the mixed mating system model with $s = 2F_{is}/(1+F_{is})$. Similar values have been found in experimental fields (Ellstrand and Foster 1983: mean $s = 0.70$) and for guinea races (Ollitrault et al. 1997: mean $s = 0.81$). We found a high and significant degree of differentiation among landraces ($F_{st} = 0.36$), although our F_{st} values are lower than those observed in previous studies. Djè et al. (2000) reported $F_{st} = 0.68$ in a world collection (25 accessions) and Ghebru et al. (2002) reported $F_{st} = 0.50$ among 28 Eritrean accessions. The lower F_{st} values we found could be explained by the fact that our study took place in situ at a very local scale, with the potential for high gene flow among landraces, which are planted closely mixed in fields. Nevertheless, the significant differentiation we found among landraces suggests the existence of biological barriers to gene flow. The first such barrier might be differences in flowering time. Panicle morphology may also impose a partial barrier (Djè et al. 2004). Based on variation in F_{is} values among landraces, we suspect variation in mating system among landraces. Analysis using STRUCTURE showed great admixture among guinea landraces. In contrast, *see gooriya* (15) and *yatta* (18) appeared in a single highly distinct cluster, with less than 1% admixture from other clusters (except for a few individuals of *yatta* which showed 21–75% admixture), suggesting that little gene flow occurs between these and other landraces. For *see gooriya*, this result is not surprising because its extremely long glumes likely impose cleistogamy (E. Garine and A. Barnaud, personal observation). Studies on outcrossing rates and phenology of different landraces are currently in progress.

Third, farmers' practices and perceptions also influence genetic diversity and structure. Farmers cultivate several landraces in the same field, so the size of the seed lots sown per landrace is small. Furthermore, variable social valuation of landraces leads to variation in abundance of landraces in the fields. The effect of drift is thus variable, and in some cases may be quite high. Farmers' practices also reflect selection. Farmers select panicles of each landrace for the next sowing; their selection may preserve landrace identity. The accuracy with which farmers discriminate the diversity of their crop population has important evolutionary implications, because it is closely related to the level of conscious selection that farmers can apply. This accuracy, and the level of conscious selection, may both vary among landraces. Among guinea landraces, farmers distinguish landraces with red seeds (*toot zeyna bee*) and landraces with white seeds (*toot zeyna buyya*). Landraces of the first group are sown in abundance because they are considered to better resist bird predation and to be more adapted to prepare beer than white-seeded landraces (results from interviews). However, because white flour is preferred for cooking, white-seeded landraces are nevertheless sown, but in lower abundance in the fields. This was confirmed by data from our transects. Together, white-seeded landraces accounted for only 24% of all plants encountered in the transects, whereas red-seeded landraces accounted for 65% (Table 1). These differences in farmers' practices may explain the genetic differentiation observed between the two clusters of landraces. Farmers may also pay particular attention to maintaining rare landraces, as suggested by several observations. For example, some landraces are typically only grown at low density, but are planted in numerous fields. Also, some landraces grown only at low density have high salience for farmers, as shown by analysis of results of free-listing exercises. The landrace *gbansaa* (landrace 7 in Table 2), for instance, is always planted at low density but has a very high saliency (SI = 0.40). The attention focused by farmers on rare landraces may imply selection towards a more defined morphotype. This, associated with genetic drift, could lead to a more strongly defined genetic identity. Indeed, all individuals of *gbansaa* are grouped on the NJ tree. Similarly, some landraces with particularly high social value, such as *yatta* (18), which in ceremonies is linked to Duupa relationships with their ancestors (E. Garine and A. Barnaud, personal observation) and which is held by the Duupa to be the most ancient of their landraces, can be easily distinguished from other landraces by SSR markers. Selection on guinea landraces allows maintenance of the panicle characteristics of this race, along with any genetically linked loci, while the rest of the genome evolves. In Wanté, farmers select against *genkya* (folk name for off-type panicles), which can help to maintain morphological differences among landraces (Dickinson and Antovonics 1973). This finding could explain the maintenance of guinea landraces despite gene flow among them. Louette and Smale (2000) show similar results on maize in Mexico. The importance of potential gene flow, through both pollen dispersal and seed exchange, can reduce the effect of drift and selection, which both tend to decrease diversity and to structure landraces. Pressoir and Berthaud

(2004) show that seed exchange between villages leads to the absence of isolation by distance in maize populations. Thus, whereas recent introduction may account for the genetic distinctiveness of some landraces (as suggested earlier), farmers' practices may be more important for maintaining distinctiveness in other landraces. The pattern in which landraces are planted (e.g., at low density but in many fields), their different cultural roles, and their specific uses, may all have an impact on the structure of genetic diversity within and among landraces. There is little evidence for spatial genetic structure (e.g., variation among fields for a given landrace) in Duupa sorghum populations. As in other situations (Pressoir and Berthaud 2004), this appears to be due to widespread seed exchanges between farmers.

This study addressed the pattern of genetic diversity of sorghum landraces at the local scale. We have shown great diversity at the scale of a single village, and demonstrated the dynamism of this diversity in both space and time. Our results suggest that despite sorghum's predominantly autogamous mating system there is substantial gene flow among landraces, and thus underline the role of farmers' practices in the maintenance of landrace identity and the favoring of genetic diversity.

Two features are striking in the management of diversity by farmers in Wanté: many landraces are grown at the village level and landraces are sown mixed in each field.

Although many landraces are grown, only a few predominate. Of the 46 landraces grown in the village, six were grown by all farmers and produced most of the harvest. Only 30 landraces were encountered in transects, and therefore the 16 other landraces must be cultivated by a very small number of farmers or at a very low density. Studies on other crops have also shown that even though many landraces are cultivated at the village level, only a few of them collectively represent the bulk of the harvest. Caillon and Lanouguère-Bruneau (2005) showed that 83% of taro [*Colocasia esculenta* L. (Schott)] planted by 12 farmers in a village in Vanuatu belonged to 6 cultivars of a total of 46. Similar results were found for cassava (*Manihot esculenta* Crantz) in an Amerindian village in Guyana (Elias et al. 2000b). Kshirsagar and Pandey (1995) reported that farmers planted 33 landraces of rice in a village in India. Of these, six landraces covered more than 50% of the land cultivated with landraces. Louette (2000) showed that among 26 landraces of maize cultivated in the Cuzalapa valley in Mexico, only four were cultivated by a large percentage of farmers.

Farmers in Wanté grew landraces mixed in fields, the consequence of mixing seeds of all landraces each farmer possesses in a common bowl for sowing. The resulting spatial distribution of landraces is thus very different from that in certain root crops, for example, where several landraces are planted in each field, but in contiguous monovarietal patches (Elias et al. 2000b). About 25% of Duupa farmers cultivated fewer than 9 landraces, whereas 40% of farmers cultivated more than 15, but in all cases landraces were thoroughly mixed in each field. For sorghum, such mixing of landraces seems common, but most studies do not provide clear data regarding spatial patterns of cultivation. Nevertheless, Teshome et al. (1999) showed an average of 9.75 landraces per field (from 1 to 24 landraces) in transects in 260 sorghum fields in Ethiopia.

No farmer in Wanté cultivated all landraces. Conservation of the diversity of landraces is thus a problem that must be considered at the village level. The landraces cultivated by few farmers belong to the group of forms intermediate between races (durra-caudatum; kafir-durra; bicolor-caudatum) and this part of the genetic diversity may therefore be more likely to be lost.

Farmers' practices may be ultimately influenced by ecological considerations, such as risk avoidance, but in a proximate sense they are shaped by cultural perceptions. Understanding these cultural perceptions is important, because culture—and along with it patterns of crop diversity—may be subject to rapid change. Among the possible consequences are the loss of diversity and thus of long-term adaptive potential of crop populations.

Acknowledgments This work is part of the PhD thesis of the first author. This research was funded by a grant from the "Institut Français de la Biodiversité". We thank Claire Billot and the Montpellier Languedoc-Roussillon Genopole platform for technical assistance. We acknowledge the help of J. Chantereau of CIRAD in the racial characterization of landraces. We thank the government of Cameroon for allowing us to carry out our fieldwork in Wanté, and Aboubacar Moussa and the MEADEN for their assistance in Cameroon. Research assistance by Simon Monné and Celestin Khasah is gratefully acknowledged. Our greatest debt of gratitude is to the Duupa farmers of Wanté for providing the research material, for their hospitality, and for their interest in participating in our study. We thank two anonymous reviewers for valuable suggestions.

References

Aldrich PR, Doebley J, Schertz KF, Stec A (1992) Patterns of allozyme variation in cultivated and wild *Sorghum bicolor*. Theor Appl Genet 85:451–460

Alvarez N, Garine E, Khasah C, Dounias E, Hossaert-McKey M, McKey D (2005) Farmers' practices, metapopulation dynamics, and conservation of agricultural biodiversity on-farm: a case study of sorghum among the Duupa in subsahelian Cameroon. Biol Conserv 121:533–543

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2002) genetix 4.04, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France (available from: http://www.univ-montp2.fr/~genetix/genetix/genetix.htm)

Bhattramakki D, Dong J, Chhabra AK, Hart GE (2000) An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. Genome 43:988–1002

Borgatti SP (1996) ANTHROPAC 4.0. Analytic Technologies, Natick, MA, USA

Brown SM, Hopkins MS, Mitchell SE, Senior ML, Wang TY, Duncan RR, Gonzalez Candelas F, Kresovich S (1996) Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. Theor Appl Genet 93:190–198

Brush SB (2000) The issue of in situ conservation of crop genetic resources. In: Brush SB (ed) Genes in the field: on-farm conservation of crop diversity. IDRC/IPGRI/Lewis Publishers, Boca Raton, pp 3–26

Caillon S, Lanouguère-Bruneau V (2005) Gestion de l'agrobiodiversité dans un village de Vanua Lava (Vanuatu): Stratégies de sélection et enjeux sociaux. Journal de la Société des Océanistes 49:129–148

Casa AM, Mitchell SE, Hamblin MT, Sun H, Bowers JE, Paterson AH, Aquadro CF Kresovich S (2005) Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. Theor Appl Genet 111:23–30

Chantereau J, Nicou R (1991) Le sorgho. Maisonneuve & Larose, Paris

Chittenden LM, Schertz KF, Lin YR, Wing RA, Paterson AH (1994) A detailed RFLP map of *Sorghum bicolor* × *Sorghum propinquum* suitable for high density mapping suggests ancestral duplication of *Sorghum* chromosomes or chromosomal segments. Theor Appl Genet 87:925–933

Cui YX, Xu GW, Magill CW, Schertz KF, Hart GE (1995) RFLP-based assay of *Sorghum-bicolor* (L.) Moench genetic diversity. Theor Appl Genet 90:787–796

de Oliveira AC, Richter T, Bennetzen JL (1996) Regional and racial specificities in sorghum germplasm assessed with DNA markers. Genome 39:579–587

Deu M, Gonzalez-de-Leon D, Glaszmann JC, Degremont I, Chantereau J, Lanaud C, Hamon P (1994) RFLP diversity in cultivated sorghums in relation to racial differentiation. Theor Appl Genet 88:838–844

Deu M, Rattunde F, Chantereau J (2006) A global view of genetic diversity in cultivated sorghum using a core collection. Genome 49:168–180

Dickinson H, Antovonics J (1973) Theoretical considerations of sympatric divergence. Am Nat 107:256–274

Djè Y, Forcioli D, Ater M, Lefèbre 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

Djè Y, Heuertz M, Lefèbre C, Vekemans X (2000) Assessment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatellite markers. Theor Appl Genet 100:918–925

Djè Y, Heuertz M, Ater M, Lefèbre C, Vekemans X (2004) *In situ* estimation of outcrossing rate in sorghum landraces using microsatellite markers. Euphytica 138:205–212

Doggett H (1988) Sorghum, 2nd edn. Wiley, New York

Elias M, Panaud O, Robert T (2000a) Assessment of genetic variability in a traditional cassava (*Manihot esculenta* Crantz) farming system, using AFLP markers. Heredity 85:219–230

Elias M, Rival L, McKey D (2000b) Perception and management of cassava (*Manihot esculenta* Crantz) diversity among Makushi Amerindians of Guyana (South America). J Ethnobiol 20:239–265

Ellstrand NC, Foster KW (1983) Impact of population structure on the apparent outcrossing rate of grain sorghum (*Sorghum bicolor*). Theor Appl Genet 66:323–327

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50

Garine E (1995) Le système agraire des Duupa du Massif de Poli (Nord Cameroun). Thèse de Doctorat, Université Paris X (Nanterre), France

Ghebru B, Schmidt RJ, Bennetzen JL (2002) Genetic diversity of Eritrean sorghum landraces assessed with simple sequence repeat (SSR) markers. Theor Appl Genet 105:229–236

Grenier C, Bramel-Cox PJ, Noirot M, Rao KEP, Hamon P (2000a) Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs non-random sampling procedures A. Using morpho-agronomical and passport data. Theor Appl Genet 101:190–196

Grenier C, Deu M, Kresovich S, Bramel-Cox PJ, Hamon P (2000b) Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs non-random sampling procedures B. Using molecular markers. Theor Appl Genet 101:197–202

Harlan JR, de Wet JMJ (1972) Simplified classification of cultivated sorghum. Crop Sci 12:172–176

Haussmann BIG, Obilana AB, Ayiecho PO, Blum A, Schipprack W, Geiger HH (2000) Yield and yield stability of four population types of grain sorghum in a semi-arid area of Kenya. Crop Sci 40:319–329

Kim J-S, Klein PE, Klein RR, Price HJ, Mullet JE, Stelly DM (2005) Chromosome identification and nomenclature of *Sorghum bicolor*. Genetics 169:955–965

Kong L, Dong J, Hart GE, (2000) Characteristics, linkage-map positions, and allelic differentiation of *Sorghum bicolor* (L.) Moench DNA simple-sequence repeats (SSRs). Theor Appl Genet 101:438–448

Kshirsagar KG, Pandey S (1995) Diversity of rice cultivars in a rainfed village in the Orissa state of India. In: Proceedings of the seminar 'Using diversity: enhancing and maintaining genetic resources on farm', June 19–21 1995. International Development Research Center, New Delhi

Louette D (2000) Traditional management of seed and genetic diversity: what is a landrace? In: Brush SB (ed) Genes in the field: on-farm conservation of crop diversity. IDRC/IPGRI/Lewis Publishers, Boca Raton, pp 109–142

Louette D, Smale M (2000) Farmers' seed selection practices and traditional maize varieties in Cuzalapa, Mexico. Euphytica 113:25–41

Menkir A, Goldsbrough P, Ejeta G (1997) RAPD based assessment of genetic diversity in cultivated races of sorghum. Crop Sci 37:564–569

Menz MA, Klein RR, Mullet JE, Obert JA, Unruh NC, Klein PE (2002) A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP (R), RFLP and SSR markers. Plant Mol Biol 48:483–499

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

Nkongolo KK, Nsapato L (2003) Genetic diversity in *Sorghum bicolor* (L.) Moench accessions from different ecogeographical regions in Malawi assessed with RAPDs. Genet Resour Crop Evol 50:149–156

Ollitrault O, Noyer JL, Chantereau J, Glaszmann JC (1997) Structure génétique et dynamique des variétés traditionnelles de sorgho au Burkina Faso. In: Begic A (Ed), Gestion des ressources génétiques de plantes en Afrique des savanes. IER-BRG Solagral, Bamako, Mali, pp 231–240

Paterson AH, Lin Y, Li Z, Schertz KF, Doebley JF, Pinson SRM, Liu SC, Stansel JW, Irvine JE (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714–1718

Perrier X, Flori A, Bonnot F (2003) Methods for data analysis. In: Hamon P, Seguin M, Perrier X, Glazmann JC (eds) Genetic diversity of cultivated tropical plants. Science Publishers, Inc and Cirad, Montpellier, pp 31–63

Pressoir G, Berthaud J (2004) Patterns of population structure in maize landraces from the Central Valleys of Oaxaca in Mexico. Heredity 92:88–94

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Schloss SJ, Mitchell SE, White GM, Kukatla R, Bowers JE, Paterson AH, Kresovich S (2002) Characterization of RFLP probe sequences for gene discovery and SSR development in *Sorghum bicolor* (L.) Moench. Theor Appl Genet 105:912–920

Smith JSC, Hopkins MS, Mitchell SE, Dean RE, Woodman WL, Lee M, Porter K (2000) Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. Crop Sci 40:226–232

Tao Y, Manners JM, Ludlow MM, Henzell RG (1993) DNA polymorphisms in grain-sorghum (*Sorghum bicolor* (L.) Moench). Theor Appl Genet 86:679–688

Taramino G, Tarchini R, Ferrario S, Lee M, Pe ME (1997) Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor*. Theor Appl Genet 95:66–72

Teshome A, Baum BR, Fahrig L, Torrance JK, Arnason TJ, Lambert JD (1997) Sorghum [*Sorghum bicolor* (L.) Moench] landrace variation and classification in north Shewa and south Welo, Ethiopia. Euphytica 97:255–263

Teshome A, Fahrig L, Torrance JK, Lambert JD, Arnason TJ, Baum BR (1999) Maintenance of Sorghum (*Sorghum bicolor*, Poaceae) landrace diversity by farmers' selection in Ethiopia. Econ Bot 53:79–88

Uptmoor R, Wenzel W, Friedt W, Donaldson G, Ayisi K (2003) Comparative analysis on the genetic relatedness of *Sorghum bicolor* accessions from Southern Africa by RAPDs, AFLPs and SSRs. Theor Appl Genet 106:1316–1325

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