

Dealing with farmers' Ethnolinguistic differences when collecting crop diversity on-farm

Joseph Ileri Kamau¹, Vanesse Labeyrie², Grace Njeri Njoroge³, Anthony Kibira Wanjoya³, Peterson Weru Wambugu¹, Zachary Kithinji Muthamia¹ and Christian Leclerc^{2*}

¹KALRO-Genetic Resources Research Institute, P.O. Box 30148-0100, Nairobi, Kenya; ²CIRAD, UMR AGAP, F-34398 Montpellier, France and ³Jomo Kenyatta University of Agriculture & Technology, P.O. Box 62000-00200, Nairobi, Kenya

Received 15 September 2015; Accepted 19 February 2016 – First published online 28 March 2016

Abstract

Identification and characterization of the farmers' named crop varieties cultivated around the world is a major issue for conservation and sustainable use of plant genetic resources. Intraspecific diversity is strongly determined by farmers' socio-cultural environment, but this has little been documented. In this paper, we tested, on a contact zone among three ethnolinguistic groups located on the Mount Kenya region, whether farmers' socio-cultural differences have an impact on the morphological characteristics of the farmers' named sorghum varieties. Eighteen qualitative morphological traits of the panicles were measured. We first compared the morphological diversity of the named varieties among ethnolinguistic groups using multivariate analysis of homogeneity of groups' dispersion and tested their differentiation using permutational multivariate analysis of variance. Discriminant analysis of principal components was then used to categorize the morphological diversity without *a priori*, and this classification was compared with farmers' local taxonomy (vernacular names) in the three ethnolinguistic groups. Our results show that some morphotypes are peculiar to some ethnolinguistic groups and that a morphotype can bear different variety names while the same variety name can be used to identify different morphotypes. Morphological differentiation that was explained by ethnolinguistic groups was higher for local landraces than for improved varieties. Our findings imply that socio-cultural diversity of farmers and the criteria they use to identify and maintain landraces need to be considered in studying and sampling crop diversity for *in situ* as well as for *ex situ* conservation.

Keywords: cultural diversity, Farmers' variety naming system, Kenya, morphological diversity, PGR conservation, sampling strategy, sorghum

Introduction

Crop improvement for food security and adaptation to climate change directly depend on inter and intraspecific crop diversity, which is preserved *in situ* and *ex situ*. In their *ex situ* conservation effort, the national gene banks have increasingly faced the challenge of space limitation in storage cells. Aware that all crop varieties maintained by farmers

cannot be *ex situ* preserved, sampling strategies are thus needed to rationalize which and how many accessions may be preserved in the seed banks. That is particularly crucial for farmers' varieties for which the names, morphological characteristics, uses and ecological distributions remain understudied.

Collecting farmers' crop varieties diversity, however, require a thorough understanding of the dynamism of the *in situ* sources (Brush, 2000), and considerations of both biological and social components of crop diversity (Leclerc and Coppens d'Eeckenbrugge, 2012), notably

*Corresponding author. E-mail: christian.leclerc@cirad.fr

the farmers' seed selection practices that can differ among ethnolinguistic groups (Labeyrie *et al.*, 2014a, b; Mwangera *et al.*, 2014). Those components were rarely analysed jointly, but studies conducted on maize in Mexico (Pressoir and Berthaud, 2004; Perales *et al.*, 2005) showed that adjacent ethnolinguistic groups maintain morphologically different maize populations. In the absence of genetic differentiation, Perales *et al.* (2005) suggested that such morphological differentiation was related to different seed selection practices among ethnolinguistic groups.

Previous studies conducted on an ethnolinguistic contact zone in Kenya (Labeyrie *et al.*, 2014a) revealed that the overall spatial distribution patterns of sorghum – *Sorghum bicolor* (L.) Moench – genetic diversity and variety vernacular names were related to the ethnolinguistic partitioning of farmers. Indeed, the genetic clusters, which were related to the improvement status of the germplasm and its growth-cycle length, were unevenly distributed among ethnolinguistic groups, despite the fact that they live close to each other in the same ecological area. Labeyrie *et al.* (2014a) concluded that social factors shape crop genetic diversity, but no in-depth study of morphological characteristics of farmers' varieties was done.

Focusing on the same ethnolinguistic contact zone as Labeyrie *et al.* (2014a), the aim of this study was to test whether social identity of farmers, belonging to different ethnolinguistic groups, must be taken into account in a sampling strategy in order to capture today for the future the maximum farmers' crop varieties diversity. Three major questions were considered in this paper: (i) Do the different ethnolinguistic groups maintain different sorghum morphotypes? (ii) Do the different variety names correspond to distinct morphotypes? (iii) Does the same variety name correspond to the same morphotype in the different ethnolinguistic groups?

Crop diversity has been historical and continues to be shaped by farmers to satisfy their needs and farmer varieties are not only of local value (Mokuwa *et al.*, 2013). Better understanding of the social mechanisms at work in structuring crop diversity is thus crucial. Considering the farmers' variety naming systems, notably homonymy and synonymy, should help us in rationalizing both *ex situ* and *in situ* conservation effort with adequate sampling strategies and preserving farmers' varieties and valuating them in breeding programmes.

Materials and methods

Study site and sorghum variety collection

This study was carried out in the lower eastern slopes of Mount Kenya from June 2010 to August 2012 among farmers practicing low-input rain-fed agriculture. There are two

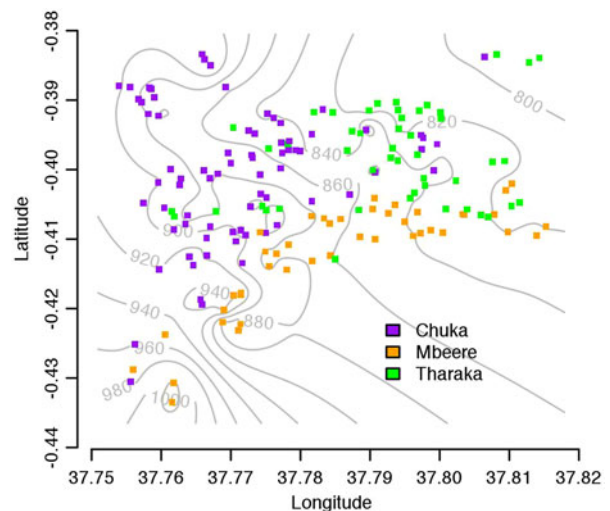


Fig. 1. Localization of farms on the ethnolinguistic contact zone of study in Mount Kenya region.

distinct rainy seasons in this region, the Long Rains (hereafter LR) occur around March to May and Short Rains (hereafter SR) around October to December (Camberlin *et al.*, 2014). The study was implemented during the two seasons on an ethnolinguistic contact zone at the boundary of Chuka, Mbeere and Tharaka ethnolinguistic groups (Fig. 1), covering about 45 km².

A total of 131 farms were randomly sampled among three ethnic groups (61 Chuka, 35 Mbeere and 35 Tharaka), representing half of the total number of households in the area, thus ensuring a good representativeness. Farms were located at similar elevation above sea level (865 ± 5.7 m – 0.95 CI) within the same agroecological zone characterized by similar soil and climate (Jaetzold *et al.*, 2007). The mean age of farmers surveyed was 45 years (SD 15.8) and ranged from 21 to 95 years, representing all generations. Mainly women were interviewed (96 women against 35 men) as they were in charge of sorghum selection, seed sowing, harvesting and trading (Pers. Obs., Middleton and Kershaw, 1965).

Farmers in the area of study grow both local varieties (landraces) and improved ones, originating from the formal varietal improvement system. Varieties can be of two types. The single-season ones complete their growth cycle during one rainy season (either LR or SR) and the ratoon ones are sown in October, cut in January and harvested in July (need two rainy seasons to complete their cycle). The origin of varieties (local or improved) and their cycle length (ratoon or single-season) were recorded from women house-heads.

A preliminary survey was carried out to estimate the frequency of farmers' sorghum varieties in the three ethnic groups, based on the vernacular variety names used by each farmer (for details, see Labeyrie *et al.*, 2014a). It was

conducted at the end of both SR (January 2011) and LR (June 2011) cropping seasons, just before harvest and provided a basis for germplasm collection.

Morphological characterization and data analysis

To undertake morphological characterization, a field experiment was set up under controlled and uniform growing conditions within the area of study. This ensured that the panicles characterized were subjected to the same environmental conditions to fully express their phenotypic traits. In doing so, possible confounding effect due to on-farm environments and farmers' practices were controlled.

We focused on the seven most frequent varieties, which were partly common among ethnolinguistic groups. Out of the total of 333 panicles collected, thus, we analysed 275 randomly selected individuals, 122 collected in the Chuka group, 63 in the Mbeere one and 90 in the Tharaka one (Table 1). On average, varieties were collected among 9.3 farms across the three ethnolinguistic groups, with a mean of 1.8 panicles per farm. The mean number of panicles originating from the three ethnolinguistic groups were similar (ANOVA; $F = 0.94$; $df = 2$; $P = 0.38$).

The morphological characterization was done at the Kenyan Genetic Resources Research Institute (GeRRI) on the panicles of the offspring of the germplasm collected on-farm. The characterization was based on 18 qualitative traits, namely, panicle shape, seed colour, grain shattering, glume adherence, glume opening, side seed shape, back seed shape, endosperm texture, glume covering, pedicellate spikelets dehiscence, awning, glume wrinkle, glume texture, grain rotation, glume colour, glume hairiness, presence of subcoat and pericarp thickness. Most traits were part of the IPGRI descriptors for sorghum (IPGRI, 1993), completed with some additional traits.

To characterize the morphological diversity of sorghum and compare it among ethnolinguistic groups, we conducted analyses first at the population level, considering all the panicles collected in each ethnolinguistic group, and second at the variety level, implementing analyses separately for each named variety.

Morphological diversity of the sorghum populations and of each named varieties were compared among ethnolinguistic groups using a procedure for the analysis of multivariate homogeneity of group dispersion (Anderson, 2006; PERMDISP2, Anderson *et al.*, 2006). This analysis was conducted on a dissimilarity matrix computed using Bray–Curtis distance index from the *vegdist* procedure in the R package *vegan* (Oksanen *et al.*, 2015).

Principal coordinates analysis (PCoA) was used to visualize the patterns of the morphological diversity. To test the morphological differentiation of the sorghum

Table 1. Sampling scheme of characterized panicles

	No. pan	No. farms	Pan/farm	Variety status
CHUKA				
Gadam	19	10	1.9	Improved
Kaguru	20	18	1.1	Improved
Mbura imwe	14	9	1.6	Landraces
Mugeta	20	10	2	Landraces
Muruge mbura ciiri	12	7	1.7	Landraces
Muruge mbura imwe	19	7	2.7	Landraces
Ngirigacha	18	9	2	Improved
Means	17	10	1.9	
Total	122			
MBEERE				
Gadam	18	8	2.2	Improved
Kaguru	19	15	1.3	Improved
Muruge mbura ciiri	6	5	1.2	Landraces
Ngirigacha	20	13	1.5	Improved
Means	16	10.2	1.6	
Total	63			
THARAKA				
Gadam	14	6	2.3	Improved
Kaguru	20	13	1.5	Improved
Mbura imwe	18	7	2.6	Landraces
Mugeta	13	7	1.9	Landraces
Muruge mbura imwe	8	5	1.6	Landraces
Ngirigacha	17	9	1.9	Improved
Means	15	7.8	2	
Total	90			

populations and of each named variety among ethnolinguistic groups, a permutational multivariate analysis of variance using distance matrices (perMANOVA, Anderson, 2001) was performed using the *adonis* procedure of the R package *vegan*. Both ethnolinguistic identity of farmers and variety names were included in the model in order to control for the potential interaction between these two factors.

A discriminant analysis of principal components (DAPC, Jombart *et al.*, 2010) was implemented on the whole dataset, including all varieties cultivated in the ethnolinguistic groups. The DAPC procedure in the R package *adegenet* (Jombart, 2008) was used to identify, without *a priori*, morphological categories in our data.

According to our in depth ethnographic observation of seed selection practices, farmers in our study site select seeds for planting only on the basis of morphological characteristics of the panicles. They do not take into account

vegetative or phenological characteristics of the plants, as seed selection takes place at home, after harvesting. We thus considered only morphological characteristics of the panicles in our analysis. Morphological categories were used as a reference in order to identify which variety names were synonyms or homonyms, while considering different ethnolinguistic groups. The distribution of these categories across groups was described, as well as their correspondence with farmers' sorghum variety names.

All analyses were performed using R (ver. 3.1.1, Team, 2014), with the packages *vegan* (ver. 2.0-10) and *Gmisc* (ver. 0.6.8) for plot and table output, as well as *knitr* (ver. 1.6) for reproducible research (Xie, 2013).

Results

Morphological differences of sorghum populations among ethnolinguistic groups

No difference was observed among ethnolinguistic groups concerning the extent of sorghum population morphological diversity according to the analysis of multivariate homogeneity of group dispersion (PERMDISP2; $F = 1.1$; $df = 2$; $P = 0.33$), and this was observed for both local ($F = 2.2$; $df = 2$; $P = 0.11$) and improved varieties ($F = 0.4$; $df = 2$; $P = 0.67$).

However, there were significant morphological differences of sorghum populations among ethnolinguistic groups according to the outputs of the perMANOVA. This was observed globally (perMANOVA; $F = 4.7$; $df = 2$; $P < 0.001$, see online Supplementary Table S1), for local landraces ($F = 5.9$; $df = 2$; $P < 0.001$) as well as for improved varieties ($F = 2.5$; $df = 2$; $P = 0.04$). Indeed, part of the accessions collected in the Chuka (Fig. 2, plan 1–2) and Tharaka (plan 2–3, not showed) groups presented morphological differences according to the PCoA display.

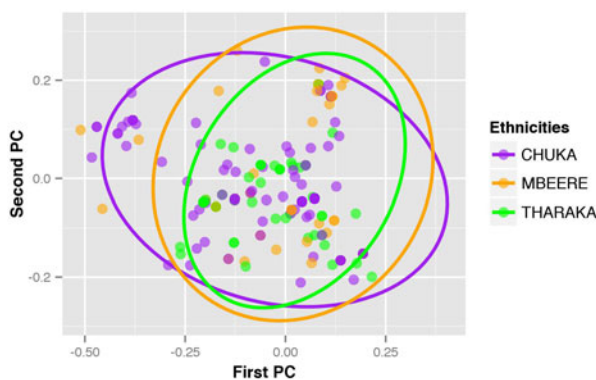


Fig. 2. Morphological diversity of sorghum in the three ethnolinguistic groups: PCoA graphical display (purple: Chuka, green: Tharaka, orange: Mbeere). Axis 1–2 (32 and 17%).

However, the part of morphological variation explained by the ethnolinguistic groups was low (2%) as compared with that explained by variety names (49%). Model outputs also indicated significant interaction between ethnolinguistic groups and varieties, suggesting that named varieties were unevenly distributed among ethnolinguistic groups.

The part of differentiation that was explained by ethnolinguistic groups was higher for local landraces (6%) than for improved varieties (2%). This suggests that morphological characteristics of varieties differ among ethnolinguistic groups more for landraces than for improved varieties.

Results of DAPC explained the morphological differentiation of sorghum populations among ethnolinguistic groups. According to the Bayesian information criterion (BIC), the optimal number of morphological categories to describe our data was seven (online Supplementary Fig. S1). In general, there were several morphological traits that contributed to DAPC (see online Supplementary Fig. S2 for contribution of morphological traits to the principal components one to four). For example, pericarp thickness and seed colour contributed in distinguishing the morphological categories in the first DAPC (Fig. 3), while seed colour, seed shape and panicle shape contributed to the second PC. Glume characteristics and presence of subcoat were discriminant traits for the third and the fourth PC, respectively.

Categories A and B were characterized by their grey seed colour (online Supplementary Fig. S3). Category C was characterized by its red seed and black glume (online Supplementary Fig. S4), and category D was the only one with cream seed. Categories E, F and G were characterized by white seed and red glume, but F and G were morphologically similar in being characterized by a compact elliptic panicle (see online Supplementary Fig. S5 for panicle shape distribution).

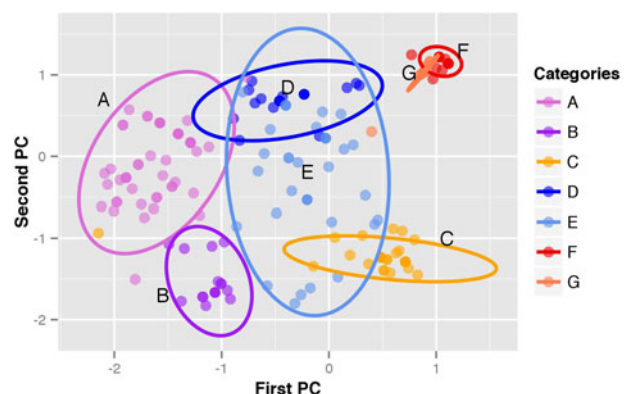


Fig. 3. Scatter plot of the DAPC based on 18 traits, 275 panicles. First and second PCs.

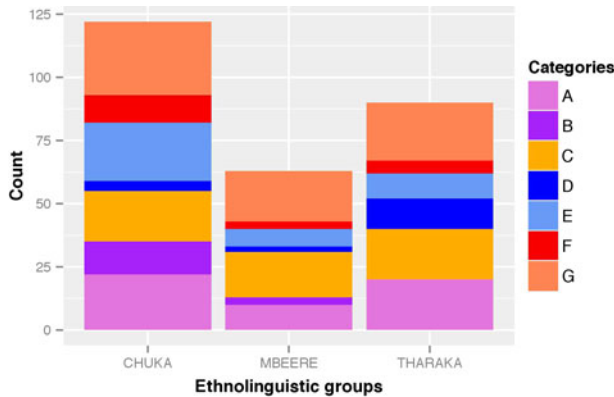


Fig. 4. Distribution of the seven morphological categories (A–G) across the three ethnolinguistic groups.

Figure 4 shows that the morphological categories were unevenly distributed among ethnolinguistic groups ($\chi^2 = 28.8$; $df = 12$; $P = 0.004$). The category B was mainly present in the Chuka group, rare in the Mbeere one and absent in the Tharaka group. The category D was found mainly in the Tharaka group and was rare in the two others. The categories

E and F were more frequent in the Chuka group than in the two others.

Correspondence between farmers' varietal nomenclature and sorghum morphological categories

When testing the correspondence between the morphological categories and farmers' variety names, we first observed that some morphological categories corresponded to one variety name while others included panicles bearing many variety names (Table 2 and online Supplementary Fig. S6). For instance, 95% of the panicles assigned to the category C were named *Kaguru*, and 88% of those assigned to the category B were named *Muruge mbura ciiri*. On the contrary, the rest of the categories included panicles bearing different variety names.

Reversely, most varieties names included mainly panicles assigned to only one morphological category. This was the case for *Kaguru* (93% of the panicles bearing this name were assigned to category C), for *Gadam* (88% of the panicles assigned to the close categories F and G),

Table 2. Correspondence between the seven morphological categories (A–G, columns) and the variety names in the three ethnolinguistic groups (rows)

Varieties	Groups	A	B	C	D	E	F	G	Total
Gadam	CHUKA	1	0	0	0	1	4	13	19
	MBEERE	0	0	0	0	2	3	13	18
	THARAKA	0	0	0	2	0	1	11	14
Kaguru	CHUKA	0	0	18	0	1	0	1	20
	MBEERE	0	0	17	0	2	0	0	19
	THARAKA	0	0	20	0	0	0	0	20
Mbura imwe	CHUKA	1	0	1	1	5	2	4	14
	MBEERE	0	0	0	0	0	0	0	0
	THARAKA	10	0	0	0	3	1	4	18
Mugeta	CHUKA	0	2	0	2	12	0	4	20
	MBEERE	0	0	0	0	0	0	0	0
	THARAKA	1	0	0	10	1	0	1	13
Muruge mbura ciiri	CHUKA	0	11	0	1	0	0	0	12
	MBEERE	3	3	0	0	0	0	0	6
	THARAKA	0	0	0	0	0	0	0	0
Muruge mbura imwe	CHUKA	17	0	0	0	2	0	0	19
	MBEERE	0	0	0	0	0	0	0	0
	THARAKA	7	0	0	0	1	0	0	8
Ngrigacha	CHUKA	3	0	1	0	2	5	7	18
	MBEERE	7	0	1	2	3	0	7	20
	THARAKA	2	0	0	0	5	3	7	17
Total		52	16	58	18	40	19	72	275

Muruge mbura imwe (89% of the panicles assigned to category A) and *Muruge mbura ciiri* (78% of the panicles assigned to category B). On the contrary, some varieties names were used for different morphological categories. This was notably the case for *Mugeta*, *Mbura imwe* and *Ngirigacha*.

Farmers' variety names, however, were fairly consistent with morphological classification, even if some morphological categories included panicles bearing different variety names. Mismatches between farmers' variety names and morphological classification could be due to the fact that farmers refer to different classification criteria according to their ethnolinguistic membership. Indeed, some morphological categories included panicles bearing different names among ethnolinguistic groups. For instance, the category A was mainly named *Muruge mbura imwe* in the Chuka group, while part of the Tharaka named it *Mbura imwe* and part of the Mbeere named it *Ngirigacha*.

Our analysis also revealed the presence of varieties of homonyms, that is, different morphotypes were identified by the same variety name in different ethnolinguistic groups. For instance, *Mugeta* cultivated by Tharaka farmers was mainly assigned to the category D, while *Mugeta* cultivated by Chuka farmers was assigned to the category E. Individuals named *Mbura imwe* were mainly assigned to the category A in the Tharaka group, and to the categories E, F, G in the Chuka group. The named varieties *Ngirigacha* and *Mbura imwe* covered an especially large range of morphological categories as compared with the other named varieties.

Significant morphological differentiation was observed for the local varieties *Mugeta* (perMANOVA; $F = 6.5$; $df = 1$; $P = < 0.001$; $R^2 = 17.3\%$) and *Mbura imwe* ($F = 3.6$; $df = 1$; $P = 0.022$; $R^2 = 10.8\%$) between the Chuka and Tharaka groups. It was also observed for the improved variety *Ngirigacha* ($F = 2.46$; $df = 1$; $P = 0.044$; $R^2 = 8.7\%$).

The extent of panicles' morphological diversity differed also among ethnolinguistic groups for *Kaguru*, *Muruge mbura imwe* and *Muruge mbura ciiri* varieties according to the outputs of the analysis of multivariate homogeneity of group dispersion (online Supplementary Table S2). These differences were nonetheless marginally significant, except for *Kaguru*.

Discussion

Focusing on an ethnolinguistic contact zone, the aim of this study was to test whether farmers belonging to three different ethnolinguistic groups (Chuka, Tharaka and Mbeere) maintained different sorghum morphotypes and whether different sorghum morphotypes corresponded to different farmers' variety names. A major scientific interest in focusing on an ethnolinguistic contact zone was to limit the

environmental variations while maximizing social and cultural diversity of farmers.

Different ethnolinguistic groups maintain morphologically different varieties

This study showed that the three ethnolinguistic groups maintained morphologically different varieties despite their high geographic proximity. Two categories of morphotypes were unevenly distributed. The category D was mainly grown by the Tharaka and corresponded to the single-season variety called *Mugeta* by farmers in this group. The second, category B, was mainly grown by the Chuka and corresponded to the ratoon variety called *Muruge mbura ciiri*. These results coincide with previous studies conducted on maize in Mexico, notably that of Perales *et al.*, (2005) who showed that adjacent ethnolinguistic groups maintained morphologically different maize populations despite the occurrence of gene flows between them. They suggested that such differences were probably maintained by divergent seed selection practices.

In our case study, the morphological category D was not previously found to be genetically different from the rest of the single-season local varieties (Labeyrie *et al.*, 2014a), while it presents clear morphological differences notably of its seed and glume colours, endosperm texture and sub-coat characteristics. The absence of genetic differentiation allows us to suggest that this morphotype is consciously maintained by the Tharaka farmers through their seed selection practices, which thwart the homogenization that is induced by gene flows. Thus, it is likely that Tharaka farmers have specific seed selection criterion to maintain this particular morphotype. A comparison of the seed selection practices and classification criterion among ethnolinguistic groups would enable to precisely determine the mechanisms involved in these patterns.

The case of the morphological category B corresponding to the ratoon variety *Muruge mbura ciiri* is different as it was previously found to be also genetically different from the rest of the varieties (Labeyrie *et al.*, 2014a). This differentiation is potentially maintained by the phenological time-lag, ratoon varieties being late flowering. Hence, the uneven distribution of this variety is most likely due to its limited diffusion through seed exchange or commercialization among ethnolinguistic groups. Further characterization of the seed diffusion pathways for this variety would be necessary to further understand its distribution patterns.

A second observation suggesting differences in farmers' management practices among ethnolinguistic groups is the differences in the extent of the morphological diversity of *Kaguru* variety. *Kaguru* is of improved origin and was introduced in the area about 15 years ago. This variety is

morphologically more diverse in the Chuka group than in the two others, and a similar observation was done for the neutral genetic diversity (Labeyrie *et al.*, 2014a). This suggests that seed selection practices for this variety may be less stringent by Chuka farmers, who appear to select a wider range of phenotypes to reproduce this variety as compared with the Tharaka and Mbeere farmers.

Cultural differences of varietal nomenclature and identification criteria

Our results showed that varietal nomenclature and identification criteria partly differ among ethnolinguistic groups. Indeed, we observed that on an ethnolinguistic contact zone different variety names can be used to identify the same morphotype (synonyms), and reversely, the same variety name (e.g. *Mugeta*) can be used to identify different morphotypes. Such differences in varieties identification and naming were observed among villages in Gambia (Nuijten and Almekinders, 2008), and were found to be related to kinship units in Amazonia (Boster, 1986). According to this latter author, varietal nomenclature constitutes a frame for storing and exchanging information concerning the planting material. He stated that the variety name used by farmers to identify a given morphotype depends on the experience they have from its characteristics and also on the cultural context where their learning took place.

The fact that different variety names were used by members of the same ethnolinguistic group to identify the same morphological category could be of different origins. First, farmers' classification could differ from ours. Second, farmers potentially attributed different weights to the various morphological criteria, while we attributed equal weights to all of them. Another reason is that variety identification skills probably differ among farmers, some farmers being less experimented than others and using laxly some variety names.

Identification criterion and nomenclature are hence largely shaped by the learning and knowledge diffusion networks. The differences of nomenclature and identification criteria we observed among ethnolinguistic groups probably traduce the limited diffusion of seeds as well as knowledge concerning varieties identification and naming among them, and this calls for further study of the learning and information diffusion pathways. This would notably require an in-depth comparison of the varietal identification criterion and nomenclature among the three groups.

Our analysis reaches the one implemented on rice varieties in West Africa (Mokuwa *et al.*, 2014). Indeed, farmers' varieties are adapted to their conditions as a result of not only genotype and environment interactions, but also as

a result of social interactions (for a review, see Leclerc and Coppens d'Eeckenbrugge, 2012).

Implications for the collection and conservation of crop genetic resources

This study shows that the diversity of farmers' socio-cultural organization, which has been largely neglected in the collection and conservation programmes of crop plant genetic resources, should be taken into consideration. Indeed, farmers' are the major stakeholders in seed diffusion and selection, the two major drivers of crop evolution *in situ* (Brush, 1995). Ethnolinguistic differences reflect the organization of societies (Barth, 1969) and cultural differences frequently correspond to limited diffusion of knowledge and material, notably planting material. In Kenya, about 67 language groups have been censused (Lewis *et al.*, 2009), corresponding to distinct cultural units. As the germplasm collection in Kenya was mainly designed to capture the diversity between and within plant species according to different agroecological zones, it is likely that it neglected part of the diversity which is maintained by different ethnolinguistic groups living in the same agroecological zone.

Completing the collection of germplasm for its characterization and conservation to better represent the ethnolinguistic diversity of Kenya would help to preserve the large phenotypic and genetic diversity which is maintained in smallholder farming systems, but highly threatened. Indeed, farmers living in the arid and semi-arid zones of Eastern Africa are already facing crop variety conservation issues. In these areas, farmers have been turning to the cultivation of maize in place of sorghum and millet (Smale and Jayne, 2003; Smale *et al.*, 2011; Leclerc *et al.*, 2014) which are considered as more adapted to drought than maize. Since the year 1980, the increasing popularity of maize – *Zea mays* (L.) – at the expense of sorghum – *Sorghum bicolor* (L.) Moench – and millet – *Eleusine coracana* (L.) – has reduced the cropping systems capacity to cope with drought (Leclerc *et al.*, 2014; Mwongera *et al.*, 2014). It is reasonable to think that farmers would want to come back to the growing of sorghum after experimenting with maize. But traditional sorghum varieties may not be available at the national gene bank to give them back to farmers. For this reason, combining the completion of crop genetic resources collection for *ex situ* conservation with participatory programmes for *in situ* conservation in the different ethnolinguistic groups is urgent to preserve this heritage.

Ethnolinguistic contact zones are especially interesting places for the collection of crop genetic resources and the implementation of *in situ* conservation programmes, as a large diversity of farmers' varieties is maintained within a limited area despite the increasing flows of varieties

through market and introduction of improved planting material. There was a clear correspondence between the farmers' nomenclature and the genetic structure of individuals assigned to genetic cluster (Labeyrie *et al.*, 2014a). However, the analysis of the morphological characteristics of the panicles suggests that the landraces presented morphological differences that were not detected with neutral genetic markers, as shown for sorghum in East Africa (Labeyrie *et al.*, 2014a) and for rice in West Africa (Mokuwa *et al.*, 2014).

Furthermore, our results call for a better consideration of the characteristics of varieties relevant for farmers in conservation programmes. The classification criteria used by farmers and the one used by gene banks managers usually differ and this is a potential gap for an efficient conservation strategy directed towards the benefit of farmers. Indeed, germplasm collectors in the national gene banks usually refer to the botanical classification. In the case of the sorghum botanical classification (Harlan and de Wet, 1972; de Wet, 1978), panicle shape and spikelet morphology allow classifying the worldwide sorghum diversity in five botanical races. But our results show that those characteristics do not correspond to those that are relevant for farmers who distinguish different varieties where collectors see only one botanical race. Barnaud *et al.* (2007) observed that the sorghum botanical classification rarely corresponds to the one of farmers, who can manage up to 40 varieties in the same village. It is thus unlikely that the usual gene bank sampling methods adequately captures all the diversity of farmer named crop varieties.

Last, our study shows that basing sampling strategies on local variety names can be misleading as some names correspond to different morphotypes and some morphotypes bear different names. This problem has been reported in a number of studies for different species (Quiros *et al.*, 1990; Barnaud *et al.*, 2007; Rabbi *et al.*, 2010). To cope with this problem, 'focus groups' of farmers could help in identifying synonyms and homonyms. However, the organization of such focus groups is not adapted to areas where habitat is largely scattered, such as our area of study. Indeed, it is difficult, time consuming and costly to contact farmers to organize a focus group which is representative of the different gender, ages, economic classes and also of the different ethnolinguistic groups. Our experience in Mount Kenya region shows that in such areas presenting a scattered habitat, more time and deeper knowledge of the local social organization is required to map out and collect crop diversity.

Conclusion

This study highlighted the major importance of considering farmers' ethnolinguistic and more generally farmers' social organization for the collection and conservation of crop

genetic resources. It also raises the major question of how to efficiently conserve the diversity still maintained in small-scale farming systems, for their future benefit.

Supplementary material

To view supplementary material for this article, please visit <https://doi.org/10.1017/S147926211600006X>

Acknowledgements

This study was supported by Agropolis Foundation in the framework of the Agropolis Resource Centre for Crop Conservation, Adaptation and Diversity project (ARCAD) No. 0900-001. V. Labeyrie was supported by Agropolis Foundation through a Ph.D. grant and received financial support from Agricultural Research for Development (CIRAD) for this study. The authors acknowledge the support of Kabururu farmers (for availing the sorghum panicles), the Kenya Agricultural and Livestock Research Organization (KALRO, formerly KARI), the Regional Universities Forum for Capacity Building in Agriculture (RUFORUM) and Jomo Kenyatta University of Agriculture and Technology (JKUAT).

References

- Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26: 32–46.
- Anderson MJ (2006) Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* 62: 245–253.
- Anderson MJ, Ellingsen KE and McArdle BH (2006) Multivariate dispersion as a measure of beta diversity. *Ecology Letters* 9: 683–693.
- Barnaud A, Deu M, Garine E, McKey D and Joly HI (2007) Local genetic diversity of sorghum in a village in Northern Cameroon: structure and dynamics of landraces. *Theoretical and Applied Genetics* 114: 237–248.
- Barth F (1969) *Ethnic Groups and Boundaries. The Social Organization of Culture Difference*. Long Grove, Illinois: Little, Brown & Co.
- Boster JS (1986) Exchange of varieties and information between Aguaruna Manioc cultivators. *American Anthropologist* 88: 428–436.
- Brush SB (1995) In situ conservation of landraces in centers of crop diversity. *Crop Science* 35: 346–354.
- Brush SB (2000) *Genes in the Field: on-Farm Conservation of Crop Diversity*. Ottawa, Canada: IDRC.
- Camberlin P, Boyard-Michaud J, Philippon N, Baron C, Leclerc C and Mwongera C (2014) Climatic gradients along the windward slopes of Mount Kenya and their implication for crop risks. Part 1: climate variability. *International Journal of Climatology* 34: 2136–2152. doi: 10.1002/joc.3427
- de Wet MJM (1978) Systematics and evolution of *Sorghum* (Gramineae). *American Journal of Botany* 65: 477–484.
- Harlan J and de Wet MJM (1972) A simplified classification of cultivated sorghum. *Crop Science* 12: 172–176.

- IPGRI (1993) Descriptors for Sorghum [Sorghum bicolor (L.) Moench]. Rome, Italy: IBP-GR/ICRISAT.
- Jaetzold R, Schmidt H, Hornetz B and Shisanya C (2007) *Farm Management Handbook of Kenya (2nd edition), vol.II, part C East Kenya. Natural Conditions and Farm Management Information*. Nairobi: Ministry of Agriculture.
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403–1405.
- Jombart T, Devillard S and Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11: 94.
- Labeyrie V, Deu M, Barnaud A, Calatayud C, Buiron M, Wambugu P, ... Leclerc C (2014a) Influence of ethnolinguistic diversity on the sorghum genetic patterns in subsistence farming systems in Eastern Kenya. *PLoS ONE* 9: e92178. doi: 10.1371/journal.pone.0092178
- Labeyrie V, Rono B and Leclerc C (2014b) How social organization shapes crop diversity: an ecological anthropology approach among Tharaka farmers of Mount Kenya. *Agriculture and Human Values* 31: 97–107.
- Leclerc C and Coppens d'Eeckenbrugge G (2012) Social organization of crop genetic diversity. The GxExS interaction model. *Diversity* 4: 1–32. doi: 10.3390/d401001
- Leclerc C, Mwangera C, Camberlin P and Moron V (2014) Cropping system dynamics, climate variability, and seed losses among East African smallholder farmers: a retrospective survey. *Weather, Climate and Society* 6: 354–370. doi: 10.1175/WCAS-D-13-00035.1
- Lewis MP, Simons GF and Fennig CD (2009) *Ethnologue: Languages of the World*. (Vol. 9). Dallas, Texas: SIL international.
- Middleton J and Kershaw G (1965) *The Central Tribes of the North-eastern Bantu. The Kikuyu, Including Embu, Meru, Mbere, Chuka, Mwimbi, Tharaka, and the Kamba of Kenya*. London: International African Institute.
- Mokuwa A, Nuijten E, Okry F, Teeken B, Maat H, Richards P and Struik PC (2013) Robustness and strategies of adaptation among farmer varieties of African rice (*Oryza glaberrima*) and Asian rice (*Oryza sativa*) across West Africa. *PLoS ONE* 8: e34801.
- Mokuwa A, Nuijten E, Okry F, Teeken B, Maat H, Richards P and Struik PC (2014) Processes underpinning development and maintenance of diversity in rice in West Africa: evidence from combining morphological and molecular markers.
- Mwongera C, Boyard-Michaud J, Baron C and Leclerc C (2014) Social process of adaptation to environmental changes: how Eastern African societies intervene between crops and climate. *Weather, Climate and Society* 6: 341–353. doi: 10.1175/WCAS-D-13-00034.1
- Nuijten E and Almekinders CJ (2008) Mechanisms explaining variety naming by farmers and name consistency of rice varieties in the Gambia. *Economic Botany* 62: 148–160.
- Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ and Suggests M (2015) The vegan package. *Community Ecology Package version 2.3-0*. <http://CRAN.R-project.org/package=vegan>
- Perales HR, Benz BF and Brush SB (2005) Maize diversity and ethnolinguistic diversity in Chiapas, Mexico. *Proceedings of the National Academy of Sciences of the United States of America* 102: 949–954.
- Pressoir G and Berthaud J (2004) Population structure and strong divergent selection shape phenotypic diversification in maize landraces. *Heredity* 92: 95–101.
- Quiros C, Brush SB, Douches DS, Zimmerer KS and Huestis G (1990) Biochemical and folk assessment of variability of Andean cultivated potatoes. *Economic Botany* 44: 254–266.
- Rabbi IY, Geiger HH, Haussmann BI, Kiambi D, Folkertsma R and Parzies HK (2010) Impact of farmers' practices and seed systems on the genetic structure of common sorghum varieties in Kenya and Sudan. *Plant Genetic Resources* 8: 116–126.
- Smale M and Jayne TS (2003) Maize in Eastern and Southern Africa: "seeds" of success in retrospect. *EPTD Discussion* 97: 1–87.
- Smale M, Byerlee D and Jayne T (2011) *Maize Revolutions in Sub-Saharan Africa*. World Bank Policy Research Working Paper Series 5659: 1–32.
- Team RC (2014) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. 2012. ISBN 3-900051-07-0.
- Xie Y (2013) knitr: a general-purpose package for dynamic report generation in R. *R Package Version* 1: 1–11.