

## ORIGINAL ARTICLE

# Unexpected pattern of pearl millet genetic diversity among ethno-linguistic groups in the Lake Chad Basin

AK Naino Jika<sup>1,2</sup>, Y Dussert<sup>1,8</sup>, C Raimond<sup>3</sup>, E Garine<sup>4</sup>, A Luxereau<sup>5</sup>, N Takvorian<sup>1,6</sup>, RS Djermakoye<sup>7</sup>, T Adam<sup>2</sup> and T Robert<sup>1,6</sup>

Despite of a growing interest in considering the role of sociological factors in seed exchanges and their consequences on the evolutionary dynamics of agro-biodiversity, very few studies assessed the link between ethno-linguistic diversity and genetic diversity patterns in small-holder farming systems. This is key for optimal improvement and conservation of crop genetic resources. Here, we investigated genetic diversity at 17 SSR markers of pearl millet landraces (varieties named by farmers) in the Lake Chad Basin. 69 pearl millet populations, representing 27 landraces collected in eight ethno-linguistic farmer groups, were analyzed. We found that the farmers' local taxonomy was not a good proxy for population's genetic differentiation as previously shown at smaller scales. Our results show the existence of a genetic structure of pearl millet mainly associated with ethno-linguistic diversity in the western side of the lake Chad. It suggests there is a limit to gene flow between landraces grown by different ethno-linguistic groups. This result was rather unexpected, because of the highly outcrossing mating system of pearl millet, the high density of pearl millet fields all along the green belt of this Sahelian area and the fact that seed exchanges among ethno-linguistic groups are known to occur. In the eastern side of the Lake, the pattern of genetic diversity suggests a larger efficient circulation of pearl millet genes between ethno-linguistic groups that are less numerous, spatially intermixed and, for some of them, more prone to exogamy. Finally, other historical and environmental factors which may contribute to the observed diversity patterns are discussed.

*Heredity* (2017) **118**, 491–502; doi:10.1038/hdy.2016.128; published online 25 January 2017

## INTRODUCTION

Small-holder farming systems have long been recognized as crop genetic resource diversity hot spots, where crop evolution and varietal innovations have regularly occurred since the beginning of domestication, a few millennia ago. Genetic resources are still very efficiently maintained in these farming systems for many crop species (Jarvis *et al.*, 2008), despite the claim that they are threatened in many developing areas including crop diversity centres (Dhillon *et al.*, 2004).

Local landraces (varieties named by farmers) still contribute strongly to food security of rural populations and to the resilience of farming systems (Jarvis *et al.*, 2008; Sahri *et al.*, 2014). Species and varietal diversities in small-holder farming systems are valuable for coping with environmental variability and for specific usages (Altieri, 2004). Farmers also maintain diversity in their fields because of aesthetic or cultural preferences (Bunting, 1990). Consequently, the role of farmers in developing and maintaining genetic resources and the huge importance of on-farm conservation have been widely acknowledged and the role of farmers' social organization for *in situ* conservation strategies has already been emphasized (Altieri and Merrick, 1987). However, there are as yet no general guidelines for designing concrete strategies for on-farm conservation of genetic resources. One reason is a general lack of knowledge concerning the

biogeography of genetic resources and of the processes that underlie it at different geographical scales. Moreover, the current patterns of genetic diversity across agricultural landscapes are the product of past and more recent evolutionary processes, the effects of which are difficult to untangle.

Among these processes, seed flow has a major effect on the dynamics of genetic diversity and agro-biodiversity conservation (Pressoir and Berthaud, 2004; Samberg *et al.*, 2013; Sahri *et al.*, 2014). Seed exchanges are of major importance for the resilience of farming systems which may rely on the ability of farmers to obtain seeds during shortage periods (Almekinders *et al.*, 1994). Leclerc and Coppens d'Eeckenbrugge (2012) have suggested that crop population genetic structures could reflect farmers' social organizations. Indeed, social organizations strongly influence the transmission of knowledge and uses about landraces, and the way seeds are inherited between generations and seed exchange networks are embedded in social relationships (McGuire, 2008; Wencélius and Garine, 2014). The typology of seed exchange networks also strongly influences the structure of intra-varietal genetic variability, as it has been shown for example for rice in Thailand (Pusadee *et al.*, 2009), and may impact the distribution and maintenance of varieties in farmer communities (Barbillon *et al.*, 2015). Thus, social boundaries can

<sup>1</sup>Ecologie, Systématique et Evolution, UMR 8079, University Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Orsay, France; <sup>2</sup>Faculté d'agronomie, Université Abdou Moumouni, Niamey, Niger; <sup>3</sup>CNRS, UMR 8586 Prodig, Paris, France; <sup>4</sup>Département d'Anthropologie, Université de Paris Ouest, Nanterre, France; <sup>5</sup>UMR 7206 CNRS-Paris Diderot, MNHN, Paris, France; <sup>6</sup>Sorbonne Universités, Paris, France and <sup>7</sup>Faculté des Sciences, Université Abdou Moumouni, Niamey, Niger  
Correspondence: Dr KN Jika, Ecologie, Systématique et Evolution, Université Paris-Sud, Bâtiment 360, 91405, Orsay, France.

E-mail: kaderjika@gmail.com

<sup>8</sup>Current address: SAVE, Bordeaux Sciences Agro, INRA, 33140 Villenave d'Ornon, France.

Received 22 April 2016; revised 1 November 2016; accepted 30 November 2016; published online 25 January 2017

limit the circulation of landraces, and therefore gene flow, between farmers belonging to different social entities.

A few recent studies have focused on the effect of social factors, and in particular ethno-linguistic differentiation, on the genetic diversity in several crops, and have brought concrete arguments in favour of the proposal of Leclerc and Coppens d'Eeckenbrugge (2012). Delètre *et al.* (2011) have shown a strong coincidence between ethno-linguistic boundaries and genetic discontinuities for cassava at the regional scale in Gabon. More recently, Westengen *et al.* (2014) have shown that the population genetic structure of sorghum landraces in Africa coincided with the main languages families. Labeyrie *et al.* (2014), in a study within a small-scaled contact zone between three ethno-linguistic groups on Mount Kenya, suggested that sorghum seed exchanges are limited for local landraces but not for improved varieties. However, comparing the results of these studies is not straightforward as they deal with very different geographical scales, time span and definitions of social groups. In addition, these studies have examined mainly selfing or vegetatively propagated crops, for which the contribution of pollen flow to gene dispersal among populations is expected to play a much smaller role than seed or propagule exchanges. To our knowledge, only a very few convincing examples of correspondence between genetic structure of outcrossing crops and ethno-linguistic diversity has been reported. Perales *et al.* (2005) have shown that maize landraces grown by two adjacent ethno-linguistic groups in southern Mexico were differentiated for some morpho-phenological traits but not for isoenzyme polymorphism, emphasizing the role of farmers' local selection influenced by differences in cultural traits. However, Orozco-Ramírez *et al.* (2016) have shown in four villages of southern Mexico that maize landraces genetic structure is primarily determined by differentiation between ethno-linguistic groups rather than environmental factors. This suggests most seed exchanges between villages involved partners of the same group. More in-depth studies are

therefore needed to test for relationships between population genetic structure and social boundaries to seed flow in outcrossing crops.

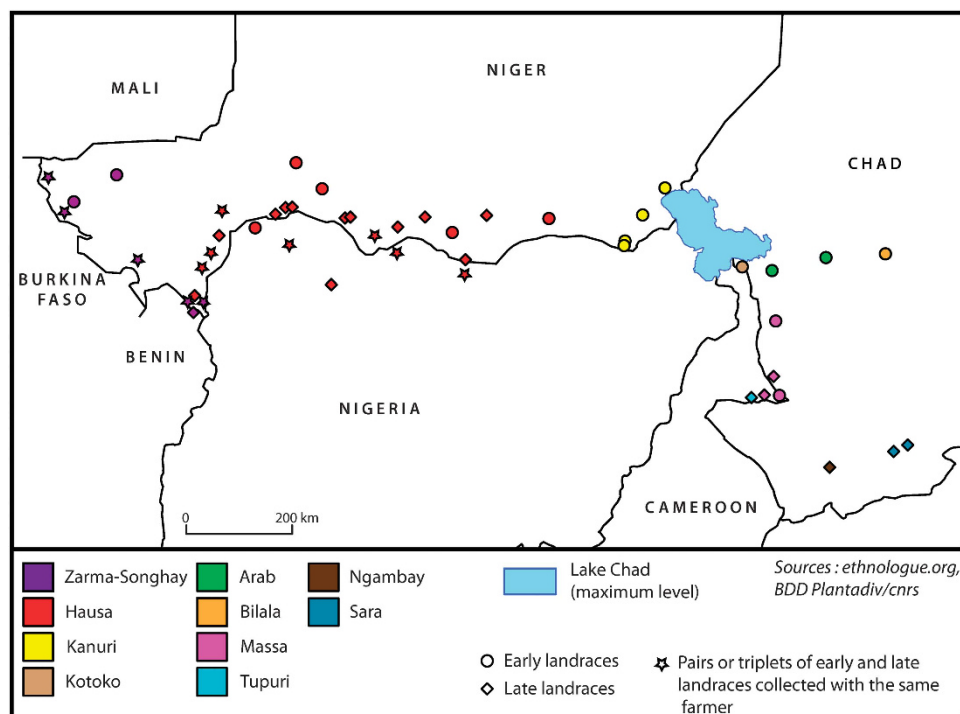
Pearl millet, a strongly outcrossing cereal mostly cultivated in Africa and India, is the staple crop for several ethno-linguistic groups, especially in the semi-arid areas of Sahel where it was probably domesticated (see Dussert *et al.* 2015, for a short review). Pearl millet is grown during the rainy seasons, often under harsh conditions. Dussert *et al.* (2015) have shown that the genetic diversity of domesticated populations is mainly structured along a west/east axis in the Sahelian region. However, their sampling did not allow them to test whether geographical limits of genetic clusters corresponded to boundaries between ethno-linguistic areas. A pioneer study on sorghum and pearl millet landraces sampled in six different and neighbouring ethno-linguistic groups in Mount Kenya has revealed no population genetic structure for pearl millet (Labeyrie *et al.*, 2016). In this study, we investigated the genetic diversity of local landrace populations of pearl millet grown by several ethno-linguistic groups using an *ad hoc* and much denser sampling in a restricted area, the Lake Chad Basin in Africa.

## MATERIALS AND METHODS

### Study site and sampling strategy

The Lake Chad Basin, our study area (Figure 1), can be divided in three climatic zones: Sahelian, Sahelo-sudanian and Sudanian (Table 1). This region displays a high landrace diversity and has been proposed to be a secondary diversification centre of domesticated pearl millet, and a putative area of origin for late-flowering landraces (Tostain, 1998). Small-holder subsistence farming, using mainly local landraces, is the dominant mode of production in this region. Pearl millet is the main crop in the western side of the lake as well as in the Sahelian part of Chad, whereas sorghum is nowadays the most important cereal in northern Cameroon and in the Sudanian part of Chad.

Ethnic diversity is high in this area, being the convergence zone of three main phyla of African languages. In the western part of the Lake Chad three main



**Figure 1** Geographical location of sampled pearl millet populations. Each population is represented with a specific colour for the ethno-linguistic group to which the farmer who provided the seeds belongs. Circles and diamonds represent early and late-flowering landraces, respectively. Stars represent pairs or triplets of early and late-flowering landraces obtained from the same farmer.

**Table 1** List of sampled populations and their characteristics

Pop	Country	Village	Ethno-linguistic groups	Landraces name	Cycle length	Longitude (decimal degrees)	Latitude (decimal degrees)	Sample size	Climatic areas
1	Niger	Wanzerbé	Zarma-Songhay	Haini Kire	Early	0.36	14.74	24	Sahelian
2	Niger	Wanzerbé	Zarma-Songhay	Somno	Late	0.36	14.74	24	Sahelian
3	Niger	Taka	Zarma-Songhay	Haini Kire	Early	0.8	13.77	24	Sahelo-sudanian
4	Niger	Taka	Zarma-Songhay	Somno	Late	0.8	13.77	24	Sahelo-sudanian
5	Niger	Bandio	Zarma-Songhay	Haini Kire	Early	1.09	13.89	24	Sahelo-sudanian
6	Niger	Mangayzé	Zarma-Songhay	Haini Kire	Early	1.95	14.68	24	Sahelian
7	Niger	Tamou	Zarma-Songhay	Haini Kire	Early	2.18	12.75	24	Sudanian
8	Niger	Tamou	Zarma-Songhay	Somno	Late	2.18	12.75	24	Sudanian
9	Niger	Tanda	Zarma-Songhay	Haini Kire	Early	3.31	11.99	24	Sudanian
10	Niger	Tanda	Zarma-Songhay	Somno	Late	3.31	11.99	11	Sudanian
11 <sup>a</sup>	Benin	Mokassa	Zarma-Songhay	Somno	Late	3.4	11.78	24	Sudanian
12 <sup>a</sup>	Niger	Koira tégui	Zarma-Songhay	Haini Kire	Early	3.56	11.98	20	Sudanian
13 <sup>a</sup>	Niger	Koira tégui	Zarma-Songhay	Somno	Late	3.56	11.98	24	Sudanian
14 <sup>a</sup>	Niger	Dioundou	Hausa	Guero	Early	3.54	12.62	24	Sudanian
15 <sup>a</sup>	Niger	Dioundou	Hausa	Maiwa	Late	3.54	12.62	24	Sudanian
16 <sup>a</sup>	Niger	Lido	Hausa	Guero	Early	3.73	12.89	24	Sudanian
17 <sup>a</sup>	Niger	Lido	Hausa	Maiwa	Late	3.73	12.89	23	Sudanian
18	Niger	Bagagi	Hausa	Guero	Early	4.05	13.85	24	Sahelo-sudanian
19	Niger	Bagagi	Hausa	Maiwa	Late	4.05	13.85	24	Sahelo-sudanian
20	Niger	Bagagi	Hausa	Maiwa	Late	4.05	13.85	42	Sahelo-sudanian
21 <sup>a</sup>	Niger	Garin Mahalba	Hausa	Maiwa	Late	3.41	12.08	24	Sudanian
22 <sup>b</sup>	Niger	Sabon Gari	Hausa	Maiwa	Late	3.93	13.38	24	Sahelo-sudanian
23	Niger	Bazaga	Hausa	Maiwa	Late	5.1	13.8	24	Sahelian
24	Niger	Kakou	Hausa	Gerguera	Early	5.33	13.93	23	Sahelian
25	Niger	Kakou	Hausa	Zango	Early	5.33	13.93	31	Sahelian
26	Niger	Montere	Hausa	Maiwa	Late	5.46	13.94	24	Sahelian
27	Niger	Kalfou Dabegui	Hausa	Gerguera	Early	5.51	14.86	24	Sahelian
28	Niger	Karofane	Hausa	Gerguera	Early	6.15	14.3	23	Sahelian
29	Niger	Karofane	Hausa	Zango	Early	6.15	14.3	23	Sahelian
30	Niger	Rafin Wada	Hausa	Maiwa	Late	6.58	13.64	24	Sahelo-sudanian
31	Niger	Guidan Roumdji	Hausa	Zango	Early	6.69	13.66	22	Sahelo-sudanian
32	Niger	Bargaja	Hausa	Dam gambe	Early	7.13	13.31	23	Sahelo-sudanian
33	Niger	Bargaja	Hausa	Dam gado	Early	7.13	13.31	24	Sahelo-sudanian
34	Niger	Bargaja	Hausa	Maiwa	Late	7.13	13.31	24	Sahelo-sudanian
35	Niger	Elsinya	Hausa	Maiwa	Late	7.62	13.54	22	Sahelo-sudanian
36 <sup>b</sup>	Niger	May Jirgui	Hausa	Maiwa	Late	8.13	13.74	24	Sahelo-sudanian
37	Niger	Gomba	Hausa	Dautehama	Early	8.75	13.3	24	Sahelo-sudanian
38	Niger	Gomba	Hausa	Babarbere	Early	8.75	13.3	24	Sahelo-sudanian
39	Niger	Tinkim	Hausa	Maiwa	Late	8.97	12.88	24	Sahelo-sudanian
40	Niger	Jigawa	Hausa	Gamongi	Early	9.43	13.82	24	Sahelian

**Table 1 (Continued)**

Pop	Country	Village	Ethno-linguistic groups	Landraces name	Cycle length	Longitude (decimal degrees)	Latitude (decimal degrees)	Sample size	Climatic areas
41 <sup>b,c</sup>	Niger	Kilakina	Hausa	Ankoutés	Early	10.75	13.72	29	Sahelian
42	Nigeria	Sokoto	Hausa	Guerguera	Early	5.25	13.06	25	Sahelo-sudanian
43	Nigeria	Sokoto	Hausa	Maiwa	Late	5.25	13.06	24	Sahelo-sudanian
44	Nigeria	Sokoto	Hausa	Zango	Early	5.25	13.06	24	Sahelo-sudanian
45	Nigeria	Zanfara	Hausa	Damro	Late	6.24	12.18	41	Sudanian
46	Nigeria	Katsina	Hausa	Zango	Early	7.6	12.99	24	Sudanian
47	Nigeria	Katsina	Hausa	Damro	Late	7.6	12.99	24	Sudanian
48	Nigeria	Jigawa	Hausa	Zango	Early	8.94	12.57	20	Sahelo-sudanian
49	Nigeria	Jigawa	Hausa	Maiwa	Late	8.94	12.57	27	Sahelo-sudanian
50 <sup>b</sup>	Nigeria	Nigeria	Hausa	unknown <sup>d</sup>	Early	4.6	13.58	24	Sahelo-sudanian
51	Niger	Boudoum	Kanuri	Moro	Early	12.26	13.16	24	Sahelian
52	Niger	Boudoum	Kanuri	Moro	Early	12.26	13.16	24	Sahelian
53	Niger	Boudoum	Kanuri	Buduma	Early	12.26	13.16	24	Sahelian
54	Niger	Ngarwa	Kanuri	Gysré	Early	12.76	13.79	24	Sahelian
55	Niger	Ngarwa	Kanuri	Buduma	Early	12.76	13.79	21	Sahelian
56	Niger	Kabalewa	Kanuri	Buduma	Early	12.97	14.06	24	Sahelian
57	Niger	Nguigmi	Kanuri	Buduma	Early	13.11	14.25	30	Sahelian
58 <sup>b</sup>	Tchad	Nibeck	Kotoko	Fyo	Early	14.63	12.77	32	Sahelo-sudanian
59	Tchad	Farcha Ater	Arabe	Dukhum kliderie	Early	15.22	12.43	23	Sahelo-sudanian
60	Tchad	Waldalbaguimi	Arabe	Dukhum kelegue	Early	16.33	12.84	23	Sahelo-sudanian
61	Tchad	Modo	Bilala	Touigne Sara	Early	17.53	12.76	31	Sahelo-sudanian
62	Tchad	Logone Gana	Massa	Viatou	Early	15.31	11.56	23	Sudanian
63	Tchad	Teleme	Massa	Ha'na	Late	15.33	10.44	23	Sudanian
64	Tchad	Mbikou	Ngambay	Tein	Late	16.39	8.6	21	Sudanian
65	Tchad	Bedaya	Sara	Dukum	Late	17.86	8.92	23	Sudanian
66 <sup>b</sup>	Tchad	Bémouli	Sara	Tein	Late	18.12	9.04	23	Sudanian
67	Cameroon	Djondong	Massa	Muri ou guidenga	Late	15.19	10.1	24	Sudanian
68 <sup>b</sup>	Cameroon	Sirlawé	Tupuri	Tcharé Dui	Late	14.95	10.07	31	Sudanian
69 <sup>b</sup>	Cameroon	Nuldaina	Massa	Tchayda dugumba	Early	15.53	10.06	20	Sudanian

We used an average rainfall of 10 years (on the basis of daily rainfall data estimated by satellite between 2001 and 2012) of all location to delimitate Sahelian zone (between 200 and 499 mm), Sahelo-sudanian zone (500–600 mm) and Sudanian zone (above 600 mm). Farmer ethno-linguistic groups refer to human populations from which samples were obtained. Each line corresponds to one population sampled at a unique farm.

<sup>a</sup>Populations sampled in the Zarma-Songhay and Hausa mixing social area.

<sup>b</sup>Populations already included in the study of Dussert *et al.* (2015).

<sup>c</sup>Population sampled in the Hausa and Kanuri mixing social area.

<sup>d</sup>Sample obtained from ICRISAT and corresponding to a local landrace.

ethno-linguistic groups are found: the Zarma-Songhay-speaking (Nilo-Saharan language) people around the Niger River in Niger and northern Benin, the Hausa-speaking (Afro-Asiatic language) people in southern central Niger and northern Nigeria, and the Kanuri-speaking (Nilo-Saharan language) people in eastern Niger and north-eastern Nigeria. These three groups are distributed along a longitudinal axis, nearly orthogonal to the isohyets distribution. All farmers in sampled farms belonged to the ethno-linguistic group which predominates in the corresponding socio-cultural area (Figure 1 and Table 1). However, we also collected samples in villages along the contact zone between the Zarma-Songhay and the Hausa ethno-linguistic areas. In this

zone, fields of Zarma-Songhay and Hausa farmers are often located close to each other. In the remainder of the paper, we will refer to this zone as the social mixing area. We also sampled in one Hausa village (Kilakina) located in the mixing area between Hausa and Kanuri people (Table 1).

In the eastern part of the lake, our sampling included seven different ethno-linguistic groups (Figure 1) which are only a subsample of the huge ethno-linguistic diversity of this region. Our sample included: Kotoko, Massa and Arab people speaking Afro-Asiatic languages, Tupuri people speaking a Niger-Congo language, Bilala (or Boulala), Ngambaye and Sara people, whose languages belong to the Nilo-Saharan family. Ethno-linguistic groups living

in this eastern part of the lake consist of only a few hundreds of thousands people while in the other side, Zarma-Songhay, Hausa and Kanuri are each several millions (www.ethnologue.com). These ethno-linguistic groups form local states which were often made up of local people and other people that have immigrated at different times. Until now, families or lineage segments can move to join established villages, even out of their own linguistic group.

### Sampling strategy and plant material

In each village, we sampled only local landraces from farmers suggested by village heads, recognized by all villagers as good pearl millet farmers and producers for generations. Landrace identification was based on information given by farmers, who recognize varieties on the basis of inflorescence shape and seed size and colour. For each landrace, seed sampling was carried out on inflorescences identified by farmers as potential seed sources for the following year sowing. Seeds were collected by removing an apical fragment (about 15 cm long) from each of 15 inflorescences selected by farmers. However, in some cases, it was possible to collect only bulk seed samples. We will refer to each seed sample as a population (a landrace cultivated by a specific farmer in a given village) in the rest of the article.

In each village, anthropological inquiries were conducted on farmers' practices, on criteria used for landrace recognition and on landrace characteristics (origin of seeds, agronomic qualities, cycle length, culinary uses and preferences).

In most Zarma-Songhay and Hausa villages, farmers grow both early and late-flowering landraces for agronomical reasons (De Rouw, 2004), but also for culinary purposes and cultural preference as shown by our inquiries (see also Lakis *et al.*, 2012). In Niger, the current aridification process has led to the rarefaction of late-flowering landraces in the northern part of their cultivation area, but they are still present in numerous villages despite severe droughts in the last decades. Late and early-flowering landraces were therefore sampled from farmers growing both types when seeds were available (Table 1). Overall, our sample comprised 69 populations (40 early- and 29 late-flowering types), representing 27 landraces collected in 48 villages (Table 1). Here, we used landraces to refer to populations recognized and name differently by farmers. Our sample was not totally exhaustive but included the most common landraces grown by each ethno-linguistic group. Most farmers prefer to grow their own landraces, considering them to be well adapted to their soils and climate. The landraces grown by the different ethno-linguistic groups bear different names (Table 1).

### Molecular analysis

**Microsatellite genotyping.** Seedlings were grown in a greenhouse. DNA extractions were performed on young leaves using the Nucleo Spin 96 Plant II kit (Macherey–Nagel, Düren, Germany).

We used 20 microsatellite markers well distributed on the pearl millet genetic map (Supplementary Table 1), using an average of 24 individuals per population. PCR multiplexes and conditions used for genotyping are described in Dussert *et al.* (2015), and in Lakis *et al.* (2012), respectively. We also used available data for eight populations already genotyped by Dussert *et al.* (2015). Sizing of the PCR products was performed on an ABI 3130xl Genetic Analyzer. Alleles were scored using GENEMAPPER (Version 4.0). The genotype scoring was checked visually and corrected when necessary. Three markers giving ambiguous results (ctm08, PSMP2248 and PSMP2249) were removed, leaving 17 microsatellite markers for our study. Genetic diversity analyses were based on a total sample of 1689 individuals.

### Data analysis

**Genetic diversity.** Population polymorphism was assessed by estimating the allelic richness ( $A_T$ ) using FSTAT v2.9.3 (Goudet, 2001), and the observed ( $H_T$ ) and unbiased expected heterozygosity ( $H_e$ ), for each locus in each population using GENETIX v4.05.2 (Belkhir *et al.*, 2004).  $F_{IS}$  values for each population were estimated using GENETIX. The 95% confidence interval (95% CI) was computed on the basis of 1000 bootstraps, and departure from Hardy–Weinberg expectations was tested by 1000 permutations for each locus. Pairwise  $F_{ST}$  values between all pairs of populations and overall  $F_{ST}$  values were estimated

and their significance was tested by 10 000 permutations of individuals among populations using GENETIX.

Differences of genetic diversity indices among pearl millet populations were tested using Kruskal–Wallis non-parametric analysis of variance and a *post hoc* multiple comparison using Wilcoxon rank sum test adjusted by a sequential Bonferroni correction (Holm, 1979) implemented in the R software (R Development Core Team, 2010). The correlation between each diversity indice and the latitudinal coordinate of each sample location was tested by using the non-parametric Spearman's correlation coefficient.

**Genetic structure of pearl millet populations.** Population genetic distances were assessed on the basis of the Cavalli-Sforza and Edwards genetic distance (Cavalli-Sforza and Edwards, 1967) and was represented by a neighbour-joining tree built using POWERMARKER (Liu and Muse, 2005). Robustness of nodes were assessed on the basis of 10 000 bootstraps.

Analyses of molecular variance, carried out with ARLEQUIN v3.5.1.2 (Excoffier and Lischer, 2010), were used to estimate population differentiation at different hierarchical classification levels (among ethno-linguistic groups, among landraces, and between early and late-flowering landraces). Groups with only one population were excluded from these analyses. Significance of  $F$  statistics estimators was assessed using 10 000 permutations.

Isolation by distance was tested in the whole sample and separately for the western and eastern sides of the Lake Chad. A simple linear regression of  $F_{ST} / (1 - F_{ST})$  values by the logarithm of geographical distances was carried out, as proposed by Rousset (1997). The statistical relationship between the genetic distance matrix and the geographical distance matrix was tested using the Spearman's rank correlation coefficient (10 000 permutations) using GENEPOP v4.0.1 (Raymond and Rousset, 1995).

Population genetics structure was assessed using two different model-based Bayesian methods implemented in STRUCTURE version 2.3.4 (Pritchard *et al.*, 2000) and TESS v2.3.1 (Durand *et al.*, 2009). The algorithm of these methods assigns individual genotypes into  $K$  clusters, minimizing departure from Hardy–Weinberg proportions and linkage disequilibrium within each cluster. Additionally, TESS takes into account spatial continuity of allele frequencies (Durand *et al.* 2009). For the two software, the population number ( $K$ ) ranged from 2 to 10 (10 independent replicates of  $3.10^4$  burn-in iterations followed by  $7.10^4$  iterations for each value of  $K$ ). For STRUCTURE, we used the admixture model and determined the optimal  $K$  using the  $\Delta K$  statistic (Evanno *et al.*, 2005). For TESS, we used the conditional autoregressive (CAR) admixture model (linear trend surface, spatial parameter initially set to the default value 0.6 and subsequently updated), and results obtained with the different  $K$  values were compared using the deviation information criterion (DIC). The TESS analysis was carried out on the whole data set and on the Hausa and Zarma-Songhay areas only. For the optimal number of  $K$  found in the analysis carried out on the whole data set, 100 additional replicates were carried out, to assess if there were multiple solutions.

We used CLUMPP v1.1.2 (Jakobsson and Rosenberg, 2007) to compute a symmetrical similarity coefficient between the different replicates (*greedy* algorithm, 100 random input sequences,  $G'$  statistic), to identify the existence of distinct solutions across replicates for both TESS and STRUCTURE. Graphical displays of the individual assignment probabilities were generated using DISTRUCT v1.1 (Rosenberg, 2004).

The Q-values of TESS were used to investigate the differences among ethno-linguistic groups and the difference between early and late-flowering landraces for each cluster (C1 to C6). The analysis was performed using non-parametric analysis of variances. Moreover, we use a simple linear regression model, with the Q-values as dependent variables and latitude origin of landraces as independent variable in order to better assess the relationship between pearl millet genetic structure and the latitude of populations. These analyses were carried out using R software (R Development Core Team, 2010).

## RESULTS

### Distribution of within-population genetic diversity

Allelic richness ( $A_T$ ) and expected heterozygosity ( $H_e$ ) ranged from 2.6 to 4.0 (mean of 3.4) and from 0.4 to 0.6 (mean of 0.5), respectively (Supplementary Table 2). The mean intra-population  $H_e$  was very

similar to values found by Lakis *et al.* (2012) in four villages of south-western Niger and by Dussert *et al.* (2015) at a continental scale. Most populations displayed a mean heterozygosity deficit which has already been noticed by several authors (Lakis *et al.*, 2012; Dussert *et al.*, 2013). However, the mean deviation from Hardy–Weinberg expectations was very variable among populations, with a mean  $F_{IS}$  varying from  $-0.21$  to  $0.24$  (mean of  $0.08$ ). All the within-population genetic diversity indices were very similar among the ethno-linguistic groups (Supplementary Table 3).

Because isohyets in this region roughly follow the latitude, we tested for a correlation between population genetic diversity indices and latitude. Indeed, water availability is a strong constraint for agriculture in the northern Sahel, leading to recurrent seed shortages. This may have consequences on effective population sizes and thus on the genetic diversity of populations from the northern less rainy areas. No significant correlations were found between latitude and  $A_r$  ( $\rho = 0.1$ ,  $n = 69$ ,  $P = 0.3999$ ) or  $H_e$  ( $\rho = 0.078$ ,  $n = 69$ ,  $P = 0.52$ ).

Local landraces were highly polymorphic, with very high within-populations genetic diversity and significant genetic differentiation among populations of the same landraces. This is demonstrated by both the distribution of populations belonging to the same landraces on the genetic distance-based tree (Figure 2) and by the results of the analyses of molecular variance (Table 2). The varieties from Kanuri farmers, which were grouped together, were exceptions to this trend.

The within-population genetic diversities of early versus late-flowering landraces could be compared in the Zarma-Songhay and Hausa groups where both types of pearl millet are cultivated. In the Hausa group, early populations displayed significantly higher  $A_r$  (mean of  $3.44$  versus  $3.12$ ; Kruskal–Wallis  $\chi^2 = 9.5965$ ,  $n = 37$ ,  $P = 0.0019$ ) and  $H_e$  values (mean of  $0.54$  versus  $0.50$ ; Kruskal–Wallis  $\chi^2 = 6.3203$ ,  $n = 37$ ,  $P = 0.0119$ ) than late populations, though the differences were small. No significant differences of genetic diversity were found between the two varietal types in the Zarma-Songhay group.

#### Global pattern of genetic diversity within the lake Chad Basin

Genetic diversity in the Lake Chad Basin did not show a significant isolation by distance pattern (IBD) either for the whole sample ( $P = 0.531$ ) or for the late-flowering landraces ( $P = 0.198$ ) from the western side of the lake. The genetic distance-based tree (Figure 2) showed a clear separation between populations sampled in the eastern and western sides of the lake (bootstrap value =  $0.684$ ). This differentiation level was confirmed by the significant  $F_{ST}$  between the two groups of populations ( $F_{ST} = 0.057$ ,  $95\% \text{ CI} = (0.027–0.091)$ ).

For the TESS results, the DIC curve showed only small decreases in DIC value when increasing  $K$  (number of genetic clusters) above six (Supplementary Figure 1). In addition, when  $K$  was greater than six, solutions proposed by TESS systematically included empty clusters, a consequence of the regularization procedure implemented in the TESS algorithm. We therefore focused on solutions obtained for  $K = 6$ . Among the 100 runs carried out for  $K = 6$ , we found five solutions (named A–E, Figure 3). The solution A was the most represented (40%), the four others being in very similar proportions (between 12 and 17%). It should be noted that in solutions A and B, one of the six genetic clusters was very poorly represented (light blue). Pertaining to STRUCTURE results, the Evanno statistic gave  $K = 3$  as the best solution (Supplementary Figure 2). The difference in the optimal number of clusters inferred by the two analyses could be explained by the higher performance of TESS compared to STRUCTURE to detect the true number of genetic clusters when genetic differentiation among populations is very weak (Chen *et al.*, 2007), which is the case in our study.

The two Bayesian clustering methods confirmed the clear genetic differentiation between both sides of Lake Chad, since most individuals from each side were assigned to different genetic clusters (Figures 3a and b, Supplementary Figure 3). However, both methods also showed high admixture levels in some of the northern populations of the eastern side (Figure 3a, Supplementary Figure 3). The finer structure revealed by TESS suggested more precisely the existence of shared genome ancestries between northern individuals of the eastern side of the lake and individuals from the Kanuri area (Figure 3a).

#### Population differentiation along the north-south axis

The clustering analyses revealed a north-south axis of population differentiation reminding the rainfall gradient. Within the eastern side of the lake, populations from Cameroon and southern Chad on one side and more northern populations from near the Lake Chad on the other side showed different cluster assignment. However, this was much pronounced in the TESS results than in the STRUCTURE results. In all TESS solutions, Chadian populations from near the lake showed high admixture levels with the genetic cluster specific of Cameroon and southern Chad and with genetic clusters found in the Kanuri area (western side of the Lake Chad) except in solution E (Figure 3a).

At the western side of the Lake, in the global TESS analysis, there was a marked difference in assignment to genetic clusters between northern and southern populations within the Zarma-Songhay and the Hausa areas. Indeed, in the major solution (A), except for cluster 2 (light-blue colour), we found a highly significant relationship between individual assignment value ( $Q$ -value) and latitude for the western side of the Lake Chad populations (Supplementary Table 5). This pattern was also clearly confirmed by analyses carried out in the Zarma-Songhay and Hausa area only (Figure 4).

#### Ethno-linguistic diversity and pearl millet population structure in the western side of the Lake Chad

The genetic distance-based tree (Figure 2) showed that populations grown by Kanuri people clustered together (Figure 2), as well as most populations grown by Zarma-Songhay people, with the noteworthy exception of late-flowering populations from the southern part of the Zarma-Songhay area (Somno populations number 10, 11 and 13). Landraces grown by Hausa people were more largely distributed on the tree (Figure 2) with no obvious clustering by landrace name or flowering type (early versus late). However, it is noticeable that populations from Nigeria were grouped with nearby populations from Kanuri people.

The two Bayesian clustering analyses confirmed these major trends and provided a more comprehensive view of the pattern of pearl millet genetic diversity and admixture in the Lake Chad Basin. They confirmed that landraces from the Kanuri area in Niger displayed very similar assignments. In addition, both methods converged in showing that individuals from Kanuri populations were partially assigned (with membership coefficients varying widely depending on TESS solutions) to the same cluster as the southern Hausa populations (in Nigeria). Because the population genetic structure revealed by TESS was finer than the one with STRUCTURE, we focused on the results of the former.

In the major solution (solution A), comparison of individuals assignment values ( $Q$ -values) showed significant differences between populations sampled in Hausa, Zarma-Songhay and Kanuri groups (except for the cluster 2 (light blue), between Kanuri and Zarma-Songhay; Supplementary Tables 4A and 4B). Yet, in solutions C and D, the Zarma-Songhay pearl millet populations showed partial assignment to a cluster (light-blue colour) specific of this region (Figure 3). For this reason, and because the overall genetic structure can interfere



**Figure 2** Genetic distance-based neighbour-joining tree showing the genetic similarity among sampled populations labelled by the population number given in Table 1. Coloured branches correspond to landraces belonging to different ethno-linguistic group. All populations collected in eastern side of the Lake Chad have the same colour.

with more local population structure, we carried out a Bayesian clustering analysis only on Hausa and Zarma-Songhay populations (ten runs). The best solution was obtained for  $K=4$  ( $K>4$  gave empty clusters) and very similar patterns of population classification in the different genetic clusters were obtained in nine out of the ten runs performed. We found a clear difference of population assignments between the two ethno-linguistic groups (Figure 4). This confirmed that geographical delimitation of genetic clusters coincided well with boundaries between ethno-linguistic areas in the western part of Lake Chad Basin. Additionally, high admixture levels were observed in populations sampled in the Zarma-Songhay and Hausa social mixing area. Altogether, we found significant, albeit low, genetic differentiation among pearl millet populations grown by the different ethno-linguistic groups ( $F_{CT}=0.030$ ,  $P<10^{-4}$ ). The among-population within ethno-linguistic group variance and the within-population

variance explained, respectively, the remaining 8.82% and 87.75% of the total genetic diversity.

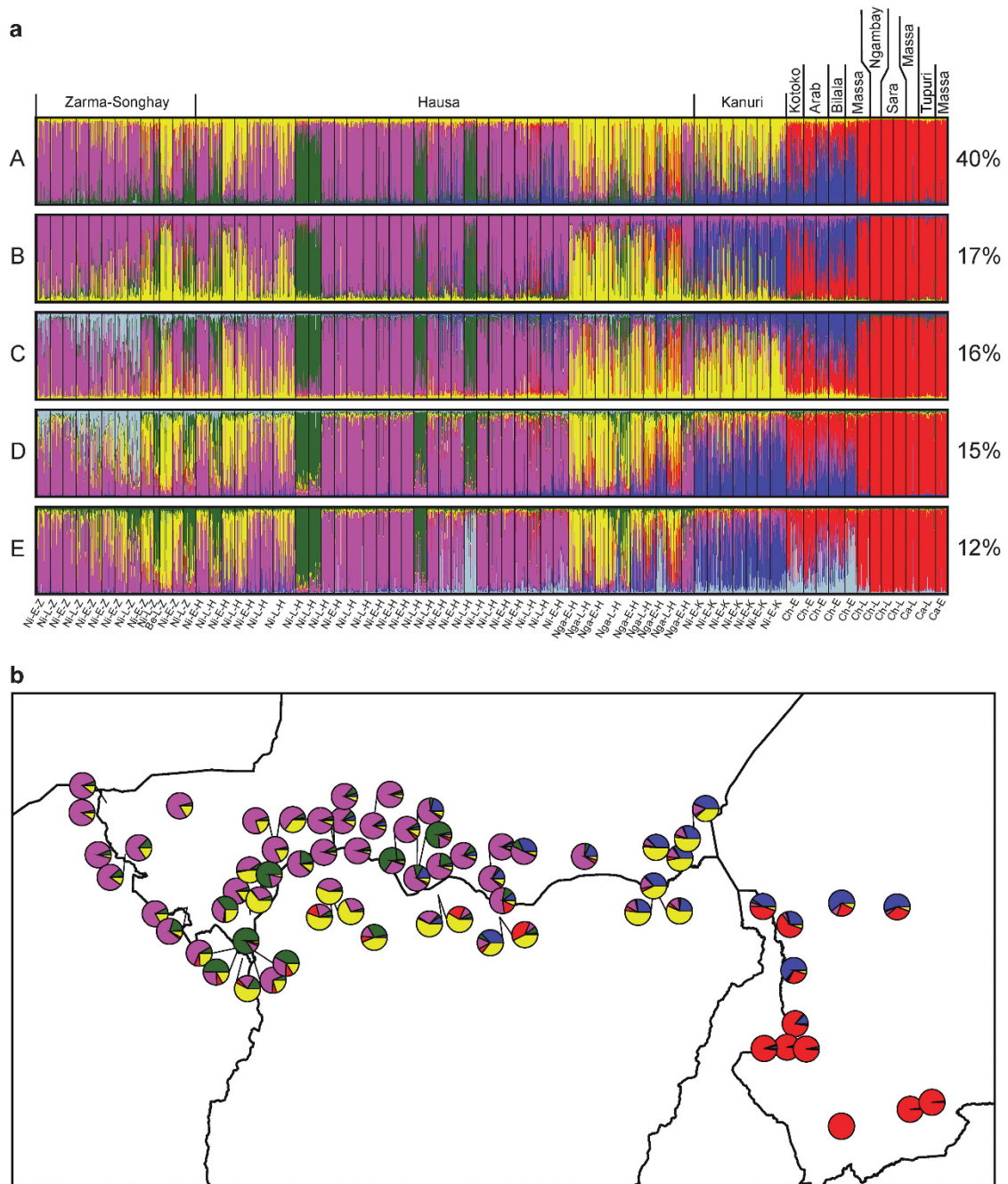
#### Relationships between local taxonomy and population genetic structure

Our sampling scheme allowed us to check if landrace classification corresponded to population genetic structure only in the Zarma-Songhay and the Hausa areas. The clustering analyses grouped some populations of different landraces from the same or neighbouring villages as being more similar than populations of the same landrace grown in distant villages (Figures 2 and 3 and Supplementary Figure 3). Landraces were not clustered according to their cycle length (early versus late-flowering), even though 4 of them (populations number 21, 22, 30 and 34 which correspond all to the late-flowering Maiwa landrace) were clearly assigned to a different group in TESS solutions (Figure 3a) as well

**Table 2 Analysis of molecular variance (AMOVA) on the whole sample**

Source of variation	Sum of Square	Variance component	Percentage of variation	F statistics
Among landraces	456.67	0.096	1.88	$F_{CT}=0.019^{***}$
Among populations within landraces	1054.26	0.44	8.70	$F_{SC}=0.089^{***}$
Within populations	10992.18	4.56	89.41	$F_{ST}=0.106^{***}$

\*\*\* $P < 10^{-4}$ .



**Figure 3** Genetic structure of pearl millet populations revealed by a clustering Bayesian analysis ( $K=6$ ) in the Lake Chad Basin. Populations are arranged in the same order as in Table 1. (a) Bar plots of all solutions obtained from the Bayesian analysis. Each thin vertical line corresponds to an individual. Coloured segments represent the proportion of each individual's genome assignment to each cluster. Numbers on the right of bar plots show how many times each solution was observed among 100 repetitions. (b) Illustration of the most probable solution (A) based on the population average proportion of genome assignment to each cluster, on the geographical map. Each pie chart represents one pearl millet population and each colour represents each inferred genetic cluster.



as in STRUCTURE (Supplementary Figure 3). These populations may contribute to the significant albeit very low differentiation we observed between early and late-flowering landraces grown in the Zarma-Songhay and Hausa areas ( $F_{CT}=0.007$ ,  $P<10^{-4}$ ). However, the analysis of the membership assignment for the most probable TESS solution (A) showed significant differences between the early and late-flowering landraces for all genetic clusters (Supplementary Table 6). Finally, couples of early and late landraces sampled in same villages of the Zarma-Songhay and the Hausa areas displayed higher genetic similarities in northern areas than in southern ones (Figure 4).

**DISCUSSION**

The main issue addressed in this paper was to investigate the relationship between pearl millet population genetic structure in the Lake Chad Basin and ethno-linguistic diversity. Two main results emerged from this study. First, we found out that differently named pearl millet landraces cultivated in the same ethno-linguistic group are not genetically differentiated. Second, genetic differentiation between pearl millet populations was better explained by ethno-linguistic differences among farmers than by spatial distance. This emphasizes the preponderance of human-driven seed flow rather than pollen flow in shaping pattern of pearl millet population genetic diversity among culturally differentiated human societies, at least at this geographical scale. Finally, a north-south differentiation was also observed at both sides of the Lake.

**Local landraces and pearl millet genetic diversity**

Most genetic diversity of pearl millet in the Lake Chad Basin was found within populations rather than among landraces or ethno-linguistic groups, similar to previous observations at several geographical scales and in different studied zones in Africa (Busso *et al.*, 2000; Alline *et al.*, 2008; Dussert *et al.*, 2015; Labeyrie com. pers.). Contrarily to our study, some authors (for example, Sehgal *et al.*, 2015) found no geographic genetic structure for pearl millet. These latter results are likely due to the use of inbred lines derived from various sources (landraces, cultivars and breeding lines) instead of landraces directly obtained from farmers' fields.

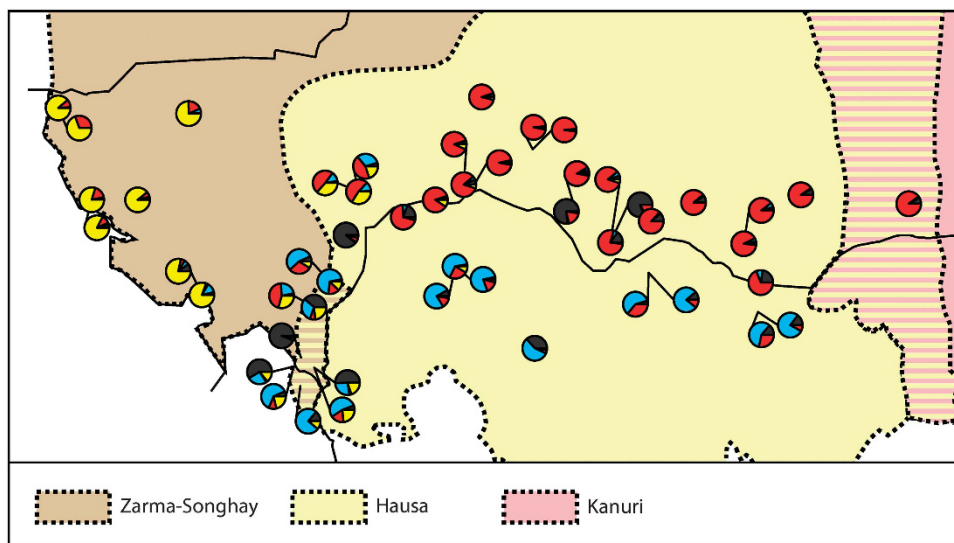
Genetic similarities were often higher between landraces sampled in the same village than between populations of the same landrace collected in different villages. This was especially observed in the western side of the lake and when comparing landrace populations sampled in the same ethno-linguistic group. This result is fully in agreement with those of Busso *et al.* (2000) in North-eastern Nigeria, and of Allinne *et al.* (2008) in the Zarma-Songhay area. Farming practices have been invoked to explain this pattern. Farmers usually grow landraces having approximately the same cycle length in the same field, promoting cross fertilization among landraces (Busso *et al.*, 2000) as is known in other outcrossing crops, for example, maize (Pressoir and Berthaud, 2004). It is however noteworthy that despite this, farmers are still able to recognize and maintain landraces with distinct phenotypes, presumably by selecting target phenotypic traits when choosing seeds for the following generation. This system associating gene flow and disruptive selection is obviously very efficient in maintaining high genetic diversity levels within pearl millet landraces, a factor that may explain their strong resilience despite severe droughts experienced in last decades in Sahel (Bezançon *et al.*, 2009).

Contrary to a previous study at the continental scale (Dussert *et al.*, 2015), we found a significant, albeit very low ( $F_{CT}=0.007$ ), genetic differentiation between early and late-flowering varieties. Lakis *et al.* (2012), in a study carried out in the Zarma-Songhay area in Niger, have shown pollen flow to occur between early and late landraces. This interpretation is supported by the higher genetic similarity between couples of early and late landraces in northern than in southern villages we observed in both the Zarma-Songhay and Hausa areas.

**Social boundaries to seed flow in pearl millet in Sahel?**

In the western side of the Lake Chad, the spatial limits of genetic clusters coincided astonishingly well with boundaries between ethno-linguistic areas. This trend was much less clear for the eastern side of the lake, but the very high number of ethno-linguistic groups and their high spatial mixing in this area as well as our sampling scheme did not allow us to test accurately for such a correspondence.

To our knowledge, our data are the first convincing demonstration for pearl millet particularly, and one of the first for an outcrossing



**Figure 4** Geographical distribution of genetic clusters inferred from the clustering Bayesian analysis carried out on pearl millet populations sampled in the Zarma-Songhay and the Hausa socio-cultural areas. Hatched zones correspond to social mixing areas. Only the major solution (found for 9 runs out of 10) is shown. Each pie chart represents one pearl millet population and each colour represents each inferred genetic cluster. The delimitation of cultural areas is approximate.

crop species in general, of a clear relationship between spatial limits of genetic groups and boundaries between cultural entities. In particular, our results contrast sharply with a recent comparative study between pearl millet and sorghum population genetic structure on Mount Kenya slopes (Labeyrie *et al.*, 2016). These authors failed to show any genetic structure for pearl millet, contrarily to sorghum, when sampling from six different ethno-linguistic neighbouring groups. The authors interpreted this result as the consequence of gene flow through both pollen dissemination and seed exchanges across the whole studied area. Recently, in maize, another strongly outcrossing species, Orozco-Ramírez *et al.* (2016) brought convincing evidence of a causal relationship between cultural differentiation of two different neighbouring communities in Mexico and genetic differentiation for microsatellite markers among maize populations they grow. They argued that a lack of seed flow between communities would explain this result. The relationship we observed between ethno-linguistic areas and delimitation of genetic clusters in the western side of the Lake Chad Basin where pearl millet is the main crop for Zarma-Songhay, Hausa and Kanuri farmers, was rather unexpected, because several factors could have promoted gene flow across large distances between populations in the whole area. Firstly, the presence of domesticated pearl millet in the Lake Chad Basin is very ancient and this region has a long history of human migration (Černý *et al.*, 2007). It was also the heart of ancient empires and has a long history of trade (Hogben and Kirk-Greene, 1966). This situation should have given opportunities for circulation of plants and among others pearl millet seeds for ages. Secondly, the density of pearl millet fields has been strongly increasing for a few decades (Ozer *et al.* 2010), such that the areas sown with the different landraces approach each other and even overlap. This could favour pollen flow among landraces, even between those grown by different ethno-linguistic groups. It seems very unlikely that environmental factors may have driven, through selection, differentiation between pearl millet populations grown in these three ethno-linguistic areas because the main environmental gradient, the rainfall level, is orthogonal to their geographical distribution. Soil fertility is also an important ecological factor, but variations are observed at very small scales, often at the field level. Rather, our results emphasize the existence of factors limiting the circulation of genes between the three large ethno-linguistic groups in the western part of the Lake Chad.

The very high symbolic value of pearl millet for Zarma-Songhay, Hausa and Kanuri rural societies, where material and symbolic life is tightly linked to this crop, likely constitutes a strong social limit to seed flow between these ethno-linguistic groups. Pearl millet seeds are generally inherited from fathers to sons, and during shortages, farmers needing to complete or even fully replenish their seed stocks before sowing time try to obtain seeds from relatives or neighbours who grow the same landraces (that is, landraces bearing the same name). Failing this, they rely on local markets (Ndjeunga, 2002) where seeds originating from local villages but also distant regions and even other countries may be sold. Nevertheless, farmers claimed they always prefer to grow their own landraces (that is, landraces of their fathers and of their immediate area), believing them to be well adapted to the local growing conditions. Seeds of non-local landraces acquired from outside sources (markets or NGOs) are mainly consumed, being sown only infrequently (Robert *et al.*, 2005). Allinne *et al.* (2008) have confirmed this preference, showing that seeds, imported for sowing by farmers in two villages of the Zarma-Songhay area in Niger, were genetically much more similar to landraces grown in the same region than to landraces from neighbouring countries. The mechanisms responsible for the relative isolation between pearl millet genetic pools

is likely effective nowadays since its fingerprint is still detectable. Nevertheless, populations sampled in the Zarma-Songhay and Hausa social mixing area displayed a high admixture pattern. This could be explained by pollen flow between landraces found in parapatric situations in this zone rather than seed exchanges. Indeed, social organization, rules of seed inheritance and landrace naming (different landraces are grown by Hausa and Zarma-Songhay farmers) are maintained in this mixing area. Further genetic studies of these pearl millet populations based on cytoplasmic genome diversity could allow confirmation or rejection of this interpretation.

Why is the picture different in the eastern part of the Lake Chad Basin? The genetic structure in this side of the Lake Chad was organized according to two geographical areas: the northern Cameroonian and southern Chadian populations on one side and other Chadian (more northern) populations on the other side. The number of sampled populations and the sampling density across ethno-linguistic groups are too low to give a definitive picture of the fine pattern of pearl millet genetic diversity in this eastern side. However, one may attempt to interpret the whole observed pattern.

In northern Cameroon and southern Chad, sorghum is widely preferred in these more humid zones. Pearl millet is grown only in years of late onset of the rainy season (Violon, submitted). It is also not the main crop in other sampled Chadian villages where irrigation is possible (for example, Nibeck village) or sorghum is grown in natural basins (for example, Farcha Ater). Farmers of these regions give less emphasis to pearl millet seed conservation than farmers in the western side of the Lake Chad, therefore promoting seed exchanges between neighbouring ethno-linguistic groups. This region is also composed of many relatively small ethno-linguistic groups with closely intertwined spatial distributions, a situation favoring exogamy in some cases. For example, it is common for Massa and Tupuri, for which sorghum seed movements are strongly correlated with matrimonial exchanges (Wencélius and Garine, 2014). Similarly, seed exchanges within and between ethno-linguistic groups are strongly determined by marriage rules for sorghum on Mount Kenya (Labeyrie *et al.*, 2016). In this example, seeds are preferentially exchanged between relatives. Further anthropological inquiries and seed sampling across the whole Lake Chad Basin should allow us to test if this explanation can be generalized.

#### Ancient historical processes and population genetic structure

Leclerc and Coppens d'Eeckenbrugge (2012) emphasized the major role of differences in farmers' culture and social organization on crop diversity. Our data clearly supported a major role of farmers' culture and of the social status of pearl millet on seed exchanges across the Lake Chad Basin, at least in the western side. However, ancient historical factors may have influenced the circulation of genetic resources and in particular pearl millet seeds across the Lake Chad Basin. The complex and still poorly known history of human migration and settlements in this region makes the interpretation of crop population genetic patterns difficult. However, it is tempting to associate some of our results with major historical facts which occurred in the Lake Chad Basin. For example, the region around the Lake Chad belonged previously to the Kanem and then Bornu Empire (probably before the eighth century according to Hogben and Kirk-Greene, 1966). The high admixture levels observed between populations around the Lake (northern Chad, Kanuri villages, northern Nigeria) may be a consequence of important seed movements across the Kanem-Bornu Empire for centuries, and probably still today. More samples, especially from the Nigerian part and the northern part of the ancient Kanem-Bornu region would be useful to test further this hypothesis.

Our results differ widely from those of Westengen *et al.* (2014), who found for sorghum local landraces three well differentiated genetic clusters co-distributed with the three main language families in Africa (Afro-Asiatic, Niger-Congo and Nilo-Saharan). They suggested that their data supported an ancient co-dispersion event of sorghum landraces and human populations (and therefore languages) from agriculture and domestication centres of origin, a few thousands of years ago. For pearl millet, we found no close correspondence between genetic clusters and language families. For example, languages of Zarma-Songhay and Kanuri people belong to the Nilo-Saharan family, and they are spatially separated by the Hausa group speaking an Afro-Asiatic language belonging to the Chadic family (a sub-group of the Afro-Asiatic family, as the Massa language, in Cameroon). However our data showed that the landraces grown by these different human groups belong to different genetic clusters. We rather believe that the complex history of human migration and admixture in the Lake Chad Basin and the fragmentation of the spatial distribution of language families in this region have erased the possible fingerprint of an ancient co-diffusion event, if it occurred for pearl millet. However, the clear differentiation between most pearl millet populations from the eastern side of the lake and those from the western side could trace back to the very ancient history of people from both sides who were not connected within the same exchange networks for centuries (Lange and Barkindo, 1990).

#### Rainfall gradient and pearl millet genetic diversity

Environmental factors may have also contributed to the observed genetic pattern (Vigouroux *et al.*, 2011). We observed a differentiation between populations sampled in the northern and southern parts of the investigated area, particularly for the Hausa region, but also to some extent in the Zarma-Songhay region and in the eastern side of the Lake Chad. This differentiation may be partially linked to the north-south rainfall gradient. Local adaptation to rainfall conditions may have played a role in limiting gene flow between populations along this gradient, through post-migration selection against introgression of non-local alleles and associated genomic regions (Lenormand, 2002). For example, in barley, the pattern of landrace genetic diversity is strongly linked to an altitudinal gradient, as a likely result of both natural and anthropic selection (Hadado *et al.*, 2010). This hypothesis bears further exploration, especially because global climatic change is expected to modify rainfall and even more temperature in the Sahel region, with drastic changes on pearl millet and sorghum yields (Sultan *et al.*, 2013). To distinguish between the effect of limits to seed exchanges and of post-migration selection on population genetic structure could be difficult. However, consequences on genomic diversity of these two processes are expected to be different because barriers to gene flow impact the whole genome while selection for local adaptation is supposed to target more specific genomic regions. Whole genome scan analyses of population differentiation along environmental gradients, in particular rainfall gradients, should provide a more detailed picture of pearl millet diversity patterns and may help to a better understanding of the role of local adaptation versus other factors in the circulation of pearl millet genes across socio-cultural landscapes.

#### Implications for pearl millet genetic resources management

Genetic clusters put in evidence could contribute to define sampling schemes for the conservation of pearl millet genetic resources in the Lake Chad Basin. However, this study highlights the complex and polymorphic relationship between pearl millet population genetic structure and anthropological and geographical factors. Our results

do not allow us to draw a full comprehensive picture of the evolutionary mechanisms that have shaped the pattern of genetic diversity in the Lake Chad Basin. Indeed, dissociating the effects of past demographic events, local adaptation and current dynamics of gene flow, the latter two being essential to define appropriate gene pools to be conserved, is highly challenging.

This study stresses the difficulty in determining at what spatial scale the notion of landrace is really relevant as a unit of sampling and conservation of genetic resources on the basis of farmers naming systems and of genetic diversity assessed by molecular markers. We are convinced for example that local adaptation contributes in some extent to the north-south pattern of genetic diversity we observed. Sanon *et al.* (2014) have demonstrated how photoperiodism variation in pearl millet landraces from Burkina Faso, which has been a neglected trait for pearl millet improvement for long, is finely tuned according to the latitude of origin. This could probably be a result of adaptation to variability of the rainy season, one of the main constraints in the Sahelian region.

Our results advocate for the claim of previous authors to take into account the social organization of farmers to target crop gene pools representing valuable genetic resources for local populations. In that sense, the major effect of rules of seed inheritance on the spatial coverage of gene flow could be a common denominator to, but sharply contrasted among, all the societies inhabiting this region. In particular, our inquiries have revealed, in the western side of the Lake, the strong attachment of Sahelian farmers to their own pearl millet landraces. Agronomic traits, such as photoperiodism, tolerance to drought or pathogens, although essential for farmers, are not sufficient to explain this attachment. Purely symbolic and aesthetic values, but also the way these varieties answer to the different uses farmers have for pearl millet seeds or other parts of the plant (Adjanohoun, 1980), should be investigated more in depth to understand all the elements which add up to make some landraces probably more than others unique and original genetic resources for farmers at local scales.

#### DATA ARCHIVING

Data available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.hg61r>

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### ACKNOWLEDGEMENTS

We thank all the farmers who agreed to give us their grains and their kind participation to anthropological inquiries. We also thank for their help traditional and administrative authorities of localities we visited. We thank the Gentyane platform at the INRA Research Unit Génétique, Diversité et Ecophysiologie des Céréales (UMR 1095), Clermont-Ferrand, for genotyping on ABI sequencer. We would also like to thank Jacqui Shykoff for her very helpful comments on the manuscript, Benjamin Sultan and Antoine Martin for providing us climatic data.

- 
- Adjanohoun EJ (Ed) (1980). *Contribution aux études ethnobotaniques et floristiques au Niger*. Agence de Coopération Culturelle et Technique: Paris.
- Almekinders CJM, Louwaars NP, De Bruijn GH (1994). Local seed systems and their importance for an improved seed supply in developing countries. *Euphytica* **78**: 207–216.
- Allinne C, Mariac C, Vigouroux Y, Bezançon G, Couturon E, Moussa D *et al.* (2008). Role of seed flow on the pattern and dynamics of pearl millet (*Pennisetum glaucum* [L.] R. Br.) genetic diversity assessed by AFLP markers: a study in south-western Niger. *Genetica* **133**: 167–178.
- Altieri MA, Merrick L (1987). In situ conservation of crop genetic resources through maintenance of traditional farming systems. *Economic Botany* **41**: 86–96.

- Altieri MA (2004). Linking ecologists and traditional farmers in the search for sustainable agriculture. *Front Ecol Environ* 2: 35–42.
- Barbillon P, Thomas M, Goldringer I, Hospital F, Robin S (2015). Network impact on persistence in a finite population dynamic diffusion model: Application to an emergent seed exchange network. *J Theor Biol* 365: 365–376.
- Belkhir K, Borsari P, Chiki L, Raufaste N, Bonhomme FG (2004). 4.05. *Logiciel Sous Windows TM Pour la Genétique des Populations*. 4. Montpellier, France: Laboratoire Genome, Populations, Interactions, CNRS UMR 5000. Université de Montpellier II.
- Bezançon G, Pham JL, Deu M, Vigouroux Y, Sagnard F, Mariac C *et al.* (2009). Changes in the diversity and geographic distribution of cultivated millet (*Pennisetum glaucum* (L.) R. Br.) and sorghum (*Sorghum bicolor* (L.) Moench) varieties in Niger between 1976 and 2003. *Genet Resour Crop Evol* 56: 223–236.
- Bunting AH (1990). The pleasures of diversity. *Biol J Linn Soc* 39: 79–87.
- Busso CS, Devos KM, Ross G, Mortimore M, Adams WM, Ambrose MJ *et al.* (2000). Genetic diversity within and among landraces of pearl millet (*Pennisetum glaucum*) under farmer management in West Africa. *Genet Resour Crop Evol* 47: 561–568.
- Cavalli-Sforza LL, Edwards AW (1967). Phylogenetic analysis. Models and estimation procedures. *Am J Hum Genet* 19: 233.
- Černý V, Salas A, Hajek M, Zaloudková M, Brdička R (2007). A bidirectional corridor in the Sahel-Sudan belt and the distinctive features of the Chad Basin populations: a history revealed by the mitochondrial DNA genome. *Ann Hum Genet* 71: 433–452.
- Chen C, Durand E, Forbes F, François O (2007). Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Mol Ecol Notes* 7: 747–756.
- De Rouw A (2004). Improving yields and reducing risks in pearl millet farming in the African Sahel. *Agr Syst* 81: 73–93.
- Delêtre M, McKey DB, Hodgkinson TR (2011). Marriage exchanges, seed exchanges, and the dynamics of manioc diversity. *Proc Natl Acad Sci USA* 108: 18249–18254.
- Dhillon BS, Dua RP, Brahmi P, Bisht IS (2004). On-farm conservation of plant genetic resources for food and agriculture. *Curr Sci* 87: 557–559.
- Durand E, Jay F, Gaggiotti OE, François O (2009). Spatial inference of admixture proportions and secondary contact zones. *Mol Biol Evol* 26: 1963–1973.
- Dussert Y, Remigereau MS, Fontaine MC, Snirc A, Lakis G, Stoeckel S *et al.* (2013). Polymorphism pattern at a miniature inverted-repeat transposable element locus downstream of the domestication gene *Teosinte-branched1* in wild and domesticated pearl millet. *Mol Ecol* 22: 327–340.
- Dussert Y, Snirc A, Robert T (2015). Inference of domestication history and differentiation between early- and late-flowering varieties in pearl millet. *Mol Ecol* 24: 1387–1402.
- Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14: 2611–2620.
- Excoffier L, Lischer HE (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10: 564–567.
- Goudet J (2001). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from: <http://www2.unil.ch/popgen/softwares/fstat.htm>.
- Hadado TT, Rau D, Bitocchi E, Papa R (2010). Adaptation and diversity along an altitudinal gradient in Ethiopian barley (*Hordeum vulgare* L.) landraces revealed by molecular analysis. *BMC Plant Biol* 10: 1.
- Hogben SJ, Kirk-Greene AH (1966). *The Emirates of Northern Nigeria: a Preliminary Survey of their Historical Traditions*. Oxford University Press: London, UK.
- Holm S (1979). A simple sequentially rejective multiple test procedure. *Scand J Stat* 6: 65–70.
- Jakobsson M, Rosenberg NA (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801–1806.
- Jarvis DI, Brown AH, Cuong PH, Collado-Panduro L, Latournerie-Moreno L, Gyawali S *et al.* (2008). A global perspective of the richness and evenness of traditional crop-variety diversity maintained by farming communities. *Proc Natl Acad Sci USA* 105: 5326–5331.
- Labeyrie V, Deu M, Barnaud A, Calatayud C, Buiron M, Wambugu P *et al.* (2014). Influence of ethno-linguistic diversity on the sorghum genetic patterns in subsistence farming systems in Eastern Kenya. *PLoS One* 9: e92178.
- Labeyrie V, Thomas M, Muthamia ZK, Leclerc C (2016). Seed exchange networks, ethnicity, and sorghum diversity. *Proc Natl Acad Sci USA* 113: 98–103.
- Labeyrie V, Deu M, Dussert Y, Rono B, Lamy F, Marangu C *et al.* (2016). Past and present dynamics of sorghum and pearl millet diversity in Mount Kenya region. *Evol Appl* 9: 1241–1257.
- Lakis G, Ousmane AM, Sanoussi D, Habibou A, Badamassi M, Lamy F *et al.* (2012). Evolutionary dynamics of cycle length in pearl millet: the role of farmer's practices and gene flow. *Genetica* 139: 1367–1380.
- Lange D, Barkindo BW (1990). La région du Tchad en tant que carrefour. Histoire générale de l'Afrique, 3. In: El Fasi M (ed.). *Histoire générale de l'Afrique*, Vol. 3. UNESCO: Paris, pp 465–488.
- Leclerc C, Coppens d'Eeckenbrugge G (2012). Social organization of crop genetic diversity. The G × E × S interaction model. *Diversity* 4: 1–32.
- Lenormand T (2002). Gene flow and the limits to natural selection. *Trends Ecol Evol* 17: 183–189.
- Liu K, Muse SV (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21: 2128–2129.
- McGuire SJ (2008). Securing access to seed: Social relations and sorghum seed exchange in eastern Ethiopia. *Hum Ecol* 36: 217–229.
- Ndjeunga J (2002). Local village seed systems and pearl millet seed quality in Niger. *Exp Agr* 38: 149–162.
- Orozco-Ramirez Q, Ross-Ibarra J, Santacruz-Varela A, Brush S (2016). Maize diversity associated with social origin and environmental variation in Southern Mexico. *Heredity* 116: 477–484.
- Ozer P, Hountondji Y, Niang AJ, Karimoune S, Laminou Manzo O, Salmon M (2010). Désertification au Sahel: historique et perspectives. *Bullet Soc Géograph Liège* 54: 69–84.
- Perales HR, Benz BF, Brush SB (2005). Maize diversity and ethnolinguistic diversity in Chiapas, Mexico. *Proc Natl Acad Sci USA* 102: 949–954.
- 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.
- Pusadee T, Jamjod S, Chiang YC, Rerkasem B, Schaal BA (2009). Genetic structure and isolation by distance in a landrace of Thai rice. *Proc Natl Acad Sci USA* 106: 13880–13885.
- Raymond M, Rousset F (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Heredity* 86: 248–249.
- R Development C. O. R. E (2010). *TEAM (2010): R, a language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria.
- Robert T, Mariac C, Allinne C, Ali K, Beidari Y, Bezançon G *et al.* (2005). *Gestion des semences et dynamiques des introgressions entre variétés cultivées et entre formes domestiques et spontanées des mils (Pennisetum glaucum spp. glaucum) au Sud-Niger*. In Colloque National du BRG: Un Dialogue pour la Diversité Génétique. Lyon, France.
- Rosenberg NA (2004). DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes* 4: 137–138.
- Rousset F (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145: 1219–1228.
- Sahri A, Chentoufi L, Arbaoui M, Ardisson M, Belqadi L, Birouk A *et al.* (2014). Towards a comprehensive characterization of durum wheat landraces in Moroccan traditional agrosystems: analysing genetic diversity in the light of geography, farmers' taxonomy and tetraploid wheat domestication history. *BMC Evol Biol* 14: 1.
- Samberg LH, Fishman L, Allendorf FW (2013). Population genetic structure in a social landscape: barley in a traditional Ethiopian agricultural system. *Evol Appl* 6: 1133–1145.
- Sanon M, Hoogenboom G, Traore SB, Sarr B, Garcia AG, Somé L *et al.* (2014). Photoperiod sensitivity of local millet and sorghum varieties in West Africa. *NJAS-Wagen J Life Sci* 68: 29–39.
- Sehgal D, Skot L, Singh R, Srivastava RK, Das SP, Taunk J *et al.* (2015). Exploring potential of pearl millet germplasm association panel for association mapping of drought tolerance traits. *PLoS One* 10: e0122165.
- Sultan B, Roudier P, Quirion P, Alhassane A, Muller B, Dingkuhn M *et al.* (2013). Assessing climate change impacts on sorghum and millet yields in the Sudanian and Sahelian savannas of West Africa. *Environ Res Lett* 8: 014040.
- Tostain S (1998). *Le mil, une longue histoire: hypothèses sur sa domestication et ses migrations*. In *Plantes et Paysages D'Afrique* (Chastenet.). Paris, France.
- Vigouroux Y, Mariac C, De Mita S, Pham JL, Gérard B, Kapran I *et al.* (2011). Selection for earlier flowering crop associated with climatic variations in the Sahel. *PLoS One* 6: e19563.
- Wencélius J, Garine É (2014). Dans les sillons de l'alliance. Ethnographie de la circulation des semences de sorgho dans l'Extrême-Nord du Cameroun. *Les Cahiers d'Outre Mer* 265: 93–116.
- Westengen OT, Okongo MA, Onk L, Berg T, Upadhyaya H, Birkeland S *et al.* (2014). Ethnolinguistic structuring of sorghum genetic diversity in Africa and the role of local seed systems. *Proc Natl Acad Sci USA* 111: 14100–14105.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

© The Author(s) 2017

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)