1

# So Close Yet So Different: Cultural Differences Among Farmers in Central Kenya Affect Their Knowledge of Sorghum (*Sorghum bicolor* [L.] Moench) Landrace Identification

V. Labeyrie[*,1,2], J. I. Kamau[3], C. Dubois[4,5], X. Perrier[4,5], and C. Leclerc[2,3]

[1]CIRAD, UPR GREEN, F-34398, Montpellier, France
[2]GREEN, University of Montpellier, CIRAD, Montpellier, France
[3]KALRO-Genetic Resources Research Institute, P.O. Box 30148-0100, Nairobi, Kenya
[4]CIRAD, UMR AGAP, F-34398, Montpellier, France
[5]AGAP, University of Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France
*Corresponding author; e-mail: vanesse.labeyrie@cirad.fr

Q2 10

**So Close Yet So Different: Cultural Differences among Farmers in Central Kenya Affect Their Knowledge of Sorghum (Sorghum bicolor [L.] Moench) Landrace Identification**. Whether knowledge of landrace identification is shared among farmers in rural societies is a matter of debate in crop diversity research, and the influence of culture on knowledge heterogeneity remains largely misunderstood. This study analyzes the heterogeneity of farmers' knowledge of crop landrace identification, and investigates factors involved in its patterns. It especially explores the effect of cultural differences by comparing how three ethnolinguistic groups identify and name sorghum diversity in the Mount Kenya region. A set of 293 panicles representing sorghum diversity in the study area was presented for identification to 96 farmers randomly selected in the three groups. A subset of 287 panicles was scored for morphological characteristics using 16 qualitative descriptors, and neutral genetic diversity of 170 of them was described using 18 SSR genetic markers. Distance-based analyses were applied to analyze knowledge patterns within and between groups and to describe the structure of sorghum morphological and genetic diversity. Results show that the degree of heterogeneity of knowledge among farmers varies strongly according to both their ethnolinguistic membership and panicle characteristics, despite their high geographic proximity. The effect of farmers' experience of landraces and of pathways for social learning on inter-individual variations of knowledge is discussed.

**Si proches et pourtant si différents: diversité culturelle et identification du sorgho (Sorghum bicolor [L.] Moench) dans la région du mont Kenya**. Un débat majeur dans le domaine de l'étude de la diversité cultivée est. de savoir si l'identification des variétés locales est. consensuelle au sein des sociétés rurales. L'influence des facteurs culturels sur l'hétérogénéité des connaissances reste en effet largement méconnue dans ce domaine. Cet article analyse l'hétérogénéité des connaissances des agriculteurs concernant l'identification des variétés de sorgho cultivées dans une localité de la région du Mont Kenya, et examine l'effet des différences culturelles en comparant comment trois groupes ethnolinguistiques identifient les différents morphotypes de sorgho. Un échantillon de 293 panicules représentant la diversité du sorgho dans la zone d'étude a été présenté pour identification à 96 agricultrices sélectionnées au hasard dans les trois groupes. Les caractéristiques morphologiques d'un sous-ensemble de 287 panicules ont été évaluées à l'aide de 16 descripteurs qualitatifs, et la diversité génétique neutre de 170 d'entre elles a été décrite au moyen de

18 marqueurs génétiques SSR. Des analyses basées sur des mesures de distance ont été utilisées pour analyser l'hétérogénéité des connaissances au sein des groupes et entre eux, et pour décrire la structure de la diversité morphologique et génétique du sorgho. Les résultats montrent que le degré d'hétérogénéité des connaissances des agriculteurs varie fortement selon qu'ils appartiennent ou non aux même groupe ethnolinguistique et ce malgré leur grande proximité géographique, et qu'il varie aussi en fonction des caractéristiques des panicules. Le rôle de l'expérience des agriculteurs concernant les différentes variétés et de l'apprentissage social sont finalement discutés.

**Key Words:**   Local ecological knowledge, identification, landrace, sorghum, Africa, ethnobotany.

## Introduction

Humans have been in close relation to crops for about 13,000 years, and a large share of the world population continues to entertain this interdependent relation in small-scale farming systems (Diamond 2002; Gepts 2004). In these systems, morphological diversity within crop species is remarkable, as stressed early by Darwin in *Origin of the Species* (1859). This diversity is generated and shaped by various evolutionary mechanisms, which are strongly influenced by human management practices (Hodgkin et al. 2007). These practices especially play on selection (Boster 1985; Louette and Smale 2000), as well as on genetic drift. They also have an impact on both seed- and pollen-mediated gene flows (Barnaud et al. 2007; McGuire 2008; vom Brocke et al. 2003).

Landraces selected and identified using vernacular names by farmers are assumed to be distinct management units, organizing the continuum of phenotypic and genetic diversity within crop species (Badstue et al. 2007; Bellon and Brush 1994; Harlan et al. 1976). In small-scale farming systems, knowledge of landrace diversity is of upmost importance for the survival of societies as part of their adaptive strategy. Humans have defined discrete categories into the crop diversity continuum just like they did for the rest of their environment, and have been using them as frames for reasoning and management (Atran and Medin 2008). These folk taxonomies involve identification, naming, and classification processes that are interrelated. According to Friedberg (1991), identification is a perceptual process through which farmers assign a plant to a class based on its perceived characteristics. Naming is the process through which these classes are labeled, mainly for communication purposes, which suppose that individuals exchanging information and planting material share a common nomenclature. Last, classification refers to the multi-level organization of classes.

A big debate exists concerning whether farmers' knowledge of landrace identification is homogeneous in farming societies, i.e., if farmers agree on the identification, naming, and classification of crop diversity. Indeed, the impressive number of landrace names inventoried on-farm in most studies questioned whether a collective consensus exists among farmers for landrace identification and naming (Jarvis et al. 2008; Sadiki et al. 2007). Furthermore, studies conducted on various species in different countries reported a poor match between farmers' landrace nomenclature and the structure of phenotypic and genetic diversity (e.g., Barnaud et al. 2007; Quiros et al. 1990; Salick et al. 1997). These studies highlighted the lack of scientific knowledge concerning farmers' landrace taxonomy. They raised debate concerning the existence of a common reference frame for landrace identification and naming shared by farmers on the local scale. They also raised questions concerning factors involved in inter-individual variations of taxonomy on the local scale.

This issue was first tackled by Boster in his work on cassava landrace identification among Aguaruna farmers (Boster 1986), based on the cultural consensus theory (Romney et al. 1986). Assuming that landraces are categories that constitute "a frame for storing and conveying experience and information" (Bulmer 1974 in Friedberg 1991), he hypothesized that farmers should exchange information and share a common experience of the landraces in order to agree on their identification and naming based on their phenotypic characteristics. He showed that differences of knowledge of landrace identification among farmers were associated with cultural differences, knowledge heterogeneity being higher between kinship groups than within. This indicates that the level of knowledge homogeneity among farmers for landraces' identification is proportional to their collective experience of crop characteristics and to the intensity of information exchange among them, and thus to

133 the strength of their social ties and cultural
134 proximity.
135    Since then, cultural consensus theory has been
136 applied to a variety of local ecological knowledge
137 domains, showing that the ways people think of and
138 classify the natural world differ across cultures.
139 Atran and Medin (2008) notably showed that dif-
140 ferent cultural groups living in the same agro-
141 ecological zones present noteworthy differences of
142 knowledge concerning their environment. These
143 studies revealed that culture influences the way
144 human societies classify the continuum of biological
145 variability surrounding them, but this issue was
146 rarely investigated with crop landraces. Most work
147 on landrace identification by farmers was conducted
148 in the frame of crop diversity studies (Nuijten and
149 Almekinders 2008; Sadiki et al. 2007), and inter-
150 individual variations of knowledge of landrace iden-
151 tification were not documented. Indeed, most stud-
152 ies document the landrace name given by one farm-
153 er to identify a plant (Soler et al. 2013), or rely on a
154 focus group to document what is the consensual
155 name for it within the community (Mucioki et al.
156 2014). Our understanding of the identification,
157 naming, and classification of crop intraspecific di-
158 versity by rural societies hence remains limited de-
159 spite its importance regarding crop genetic resources
160 conservation and property rights issues (Lapeña and
161 Halewood 2016).
162    The way farmers identify and classify landraces is
163 expected to influence their management of crop
164 infra-specific diversity, and differences of
165 management should therefore exist between
166 cultural groups. The work of Perales et al. (2005)
167 brought a key contribution to this issue, indicating
168 that two geographically close ethnolinguistic groups
169 in Mexico have divergent seed selection practices as
170 they maintain morphologically distinct maize pop-
171 ulations despite gene flows between them. Their
172 results suggest that differences in knowledge of crop
173 landrace management exist between ethnolinguistic
174 groups despite their geographical proximity. Our
175 study aims at testing this hypothesis by characteriz-
176 ing differences in knowledge concerning sorghum
177 (*Sorghum bicolor* [L.] Moench–Poaceae) landrace
178 identification between three ethnolinguistic groups
179 in a locality of the Mount Kenya region. We tested
180 three main hypotheses following Bosters' work on
181 farmers' knowledge concerning landrace identifica-
182 tion. First, farmers belonging to the same
183 ethnolinguistic group display more similar

184 knowledge than those belonging to different groups
185 because they exchange more information. Second,
186 ethnolinguistic groups display different levels of
187 knowledge heterogeneity for the identification of
188 the different landraces because they differ in their
189 experience of their characteristics. Third,
190 ethnolinguistic groups differ in the landrace names
191 they associate with the different morphotypes be-
192 cause information exchange is limited between
193 them.
194    This work builds on a previous study conducted
195 in the same area, in which samples from 14 sor-
196 ghum landraces named by famers in the three
197 ethnolinguistic groups were collected (Labeyrie
198 et al. 2014). Genetic analysis showed that the plants
199 collected belong to four main genetic clusters cor-
200 responding to differences in phenology, as well as in
201 the origins and history of the landraces. These four
202 clusters respectively correspond to (i) short-cycle
203 local landraces, (ii) long-cycle landraces, (iii) an
204 introduced variety released by research several de-
205 cades ago (*Kaguru*), and (iv) another introduced
206 variety that has just diffused in the area (*Gadam*).
207 The first two clusters included several morphotypes    Q3
208 associated with different landrace names, the occur-
209 rence frequency of which differed significantly
210 among ethnolinguistic groups. Based on these re-
211 sults, our article addresses two main questions: do
212 the three ethnolinguistic groups differ in their
213 knowledge concerning sorghum landrace identifica-
214 tion, and does their respective knowledge vary ac-
215 cording to sorghum genetic and morphological
216 characteristics?

## Material and Methods    217

### Study Site    218

219    The study site was located in the Eastern Prov-
220 ince of Kenya at the boundary between Tharaka-
221 Nithi and Embu counties (0° 24′ S, 37° 46′ E). We
222 focused on a contact zone between Chuka, Tharaka,
223 and Mbeere ethnolinguistic groups (Fig. 1). The
224 study site presents uniform agro-ecological condi-
225 tions, at an altitude of about 900 meters (m) above
226 sea level, and with mean monthly temperature rang-
227 ing between 21.7 and 23.9 degrees Celsius (°C)
228 (Jaetzold et al. 2007). The mean rainfall is about
229 700–800 millimeters (mm) per year, distributed
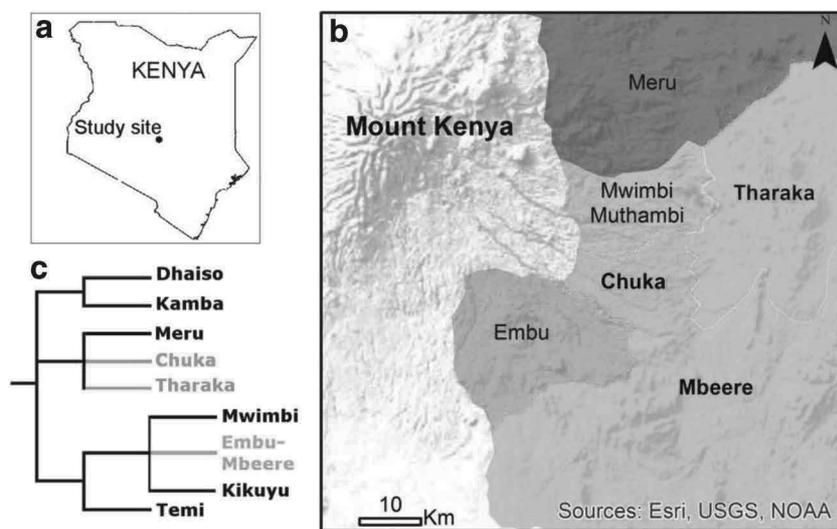230 across two rainy seasons with the long rains

**Fig. 1.** Study site location. **a** Location of the study site in Kenya. **b** Geographic distribution of ethnolinguistic groups in the Eastern Mount Kenya region. **c** Linguistic classification of Central Kenya languages according to Hammarström et al. (2015).

231 occurring from March to May and the short rains
232 from October to December.
233     According to oral history, people started to mi-
234 grate to the study area by the end of the nineteenth
235 century. The Chuka would have been the first to
236 settle in this area about one century ago, while the
237 Tharaka and the Mbeere probably settled more
238 recently, but information is lacking concerning their
239 history. The Chuka, Tharaka, and Mbeere groups
240 present cultural and linguistic differences
241 (Hammarström et al. 2015). They speak distinct
242 languages of the central Bantu cluster, which are,
243 however, largely inter-intelligible. Members of each
244 group believe in their common ascendance on
245 which is based their distinct ethnolinguistic identity
246 (Heine and Möhlig 1980; Middleton 1953).
247 Tharaka and Chuka were allied in the past and
248 consider themselves to be blood brothers, or
249 *gishiaro*, in Kimeru language (Fadiman 1993), and
250 they have limited relationship with the Mbeere
251 (Glazier 1970; Mwaniki 1973). Intermarriage is
252 frequent between the Tharaka and the Chuka,
253 while it is very uncommon between both groups
254 and the Mbeere (Labeyrie et al. 2016a). This rela-
255 tionship system is reflected by the geographical
256 organization of the three groups, the Tharaka and
257 the Chuka being spatially mixed and settled in the

258 northern part of the study site, whereas the Mbeere
259 are located separately from the two other groups in
260 the southern part of the area. The maintenance of
261 this geographical partition among ethnolinguistic
262 groups results mainly from the combination of
263 ethnolinguistic endogamy and patrilocal residence,
264 implying that most married men settle near the
265 compound of their father (Middleton 1953, pers.
266 obs.).

ETHICS STATEMENT 267

268     This work was conducted in collaboration with
269 the KALRO National Genebank of Kenya that has
270 the national mandate for the collection of plant
271 genetic resources and the documentation of accom-
272 panying information. Institutional and administra-
273 tive procedures were carefully followed prior to
274 undertaking the study, and dedicated committees
275 in KALRO granted approval for our research activ-
276 ities. We followed recommendations of the ISE
277 Code of Ethics, and the involvement of team mem-
278 ber natives from the study region contributed to
279 ensure that local procedures, rules, and customs
280 were respected, and that authorizations were
281 granted from legitimate authorities. First, govern-
282 ment administrative and local community

283 representatives were informed and kept updated of
284 the activities, and their consent was sought before
285 conducting the research. Then, the study objective
286 and the future utilization of data were explained to
287 farmers and their prior informed consent was ob-
288 tained verbally before undertaking interviews and
289 crop collection. Activities were not conducted
290 where such consent was not granted.

291                    DATA COLLECTION

292    Sorghum panicles were sampled on-farm in Jan-
293 uary and July 2011 in the three ethnolinguistic
294 groups following the strategy described previously
295 (Labeyrie et al. 2014), which aimed at representing
296 the diversity of the sorghum landraces named by
297 farmers. Seeds from the collected panicles were
298 sown in October 2011 in an experimental field
299 under controlled and uniform growing conditions.
300 A total of 293 descendants were sampled to maxi-
301 mize the range of morphological variability and
302 presented to a panel of farmers for identification.
303 Out of this set, 287 panicles were scored for mor-
304 phological descriptors as six panicles were too de-
305 graded at the end of the survey to be scored. One-
306 hundred seventy plants in this subset were scored
307 for neutral genetic microsatellite SSR markers, se-
308 lected to represent the diversity of sorghum land-
309 races named by farmers in the three ethnolinguistic
310 groups (Labeyrie et al. 2014).

311 *Landrace Identification Experiment*

312    The set of 293 panicles harvested in the experi-
313 mental field was presented to a panel of informants
314 from the three ethnolinguistic groups. Thirty-two
315 female informants were randomly chosen in each
316 group, and their ethnolinguistic group was record-
317 ed. Only women were interviewed because they are
318 in charge of sorghum seed selection, sowing, har-
319 vesting, and trading according to the local gendered
320 division of labor (pers. obs.). Following the proce-
321 dure used by Boster (1986), each informant was
322 independently asked to identify each of the 293
323 panicles that were successively presented to her. A
324 field assistant recorded the name used by each in-
325 formant to identify each panicle. Spelling standard-
326 ization was later done to ensure that differences were
327 not due to variation in pronunciation among
328 informants.

329 *Morphological and Genetic Characterization*

330    Out of the 293 panicles harvested, 287 were
331 scored for 16 qualitative morphological descrip-
332 tors at the Kenya Agricultural and Livestock
333 Research Organization, Genetic Resources Re-
334 search Institute, Muguga. The study was limit-
335 ed to the characteristics of the panicles because
336 the selection of seeds by farmers is done on the
337 panicle only, at home before threshing, and
338 thus without considering the characteristics of
339 the whole plant. Only qualitative descriptors
340 were scored because they are the main criteria
341 on which farmers base their perceptual distinc-
342 tiveness (Gibson 2009). The 16 descriptors were
343 selected for their polymorphism in the sorghum
344 population studied and their ease of scoring. They
345 included the main criteria that farmers reported
346 using for identifying their sorghum landraces, ac-
347 cording to information collected during semi-
348 directive interviews.
349    Traits scored (Electronic Supplementary Ma-
350 terial [ESM] Appendix 1) concerned the whole
351 panicle shape, seed characteristics (color, lateral
352 shape, shattering, endosperm texture, sub-coat pres-
353 ence, and pericarp thickness) and glume character-
354 istics (color, opening, adherence, covering, awning,
355 hairiness, texture, presence of a transversal wrinkle,
356 and pedicelate spikelet). Parts of these descriptors
357 were selected among those recommended by the
358 IPGRI (1993), and more precise descriptors of seed
359 and spikelets were added among some of those used
360 by Snowden (1936). Procedures of double charac-
361 terization of a set of panicles randomly selected
362 made it possible to ensure the consistency of oper-
363 ators in scoring morphological traits, and double
364 data entry was performed to limit typing errors.
365    DNA extraction, amplification, migration, and
366 alleles' size scoring for 18 microsatellite SSR loci
367 were done on 170 plants out of the 287 for which
368 panicles were morphologically scored. The study of
369 the genetic diversity of the sorghum population,
370 including these 170 individuals, was conducted in
371 a previous study, and full methodological details are
372 provided in Labeyrie et al. (2014).

373                  STATISTICAL ANALYSES

374    Statistical analyses were conducted to describe,
375 on the one hand, the patterns of sorghum genetic

376 and morphological diversity and, on the other hand,
377 the patterns of knowledge heterogeneity among
378 informants.

*Analysis of Sorghum Genetic and Morphological* 379
*Diversity* 380

381 First, the structure of sorghum genetic diversity
382 was described. A discriminant analysis of principal
383 components (DAPC, Jombart et al. 2010) was used
384 to identify and then describe clusters from the ge-
385 netic diversity of the 170 sorghum panicles collect-
386 ed. The *K*-means method was performed prior to
387 running the discriminant analysis using the algo-
388 rithm included in DAPC function, and the optimal
389 number of clusters to describe the diversity was
390 determined based on the Bayesian information cri-
391 terion (BIC) curve. Analyses were performed using
392 the R package *adegenet*, version 2.0.1 (R. Core
393 Team 2016; Jombart 2008).
394 Second, the morphological diversity of the 287
395 panicles was described by performing a principal
396 coordinates analysis (PCoA) on the morphological
397 dissimilarity matrix. Dissimilarity between panicle
398 pairs was computed with the simple matching in-
399 dex, i.e., dividing the number of traits for which
400 both panicles share the same modalities by the total
401 number of traits. Correspondence between the
402 structure of sorghum morphological and genetic
403 diversity was assessed by displaying genetic clusters
404 in colors on the PCoA scatterplot. Analyses were
405 performed using the R package *ade4* version 1.7–6
406 (Dray and Dufour 2007).

*Measurement of Informants' Consistency in Naming* 407
*Panicles* 408

409 We computed the number of informants in
410 each ethnolinguistic group who cited the same
411 name to identify each sorghum panicle. A
412 cross-table was built by a group, crossing the
413 list of the 293 panicles in one way and the list
414 of the names given for identification in the
415 second way, to calculate the number of infor-
416 mants at each combination of panicle x name.
417 We considered that a panicle was named con-
418 sistently in a given group, i.e., knowledge het-
419 erogeneity was low within it, when more than
420 half of informants (60%) used the same name
421 to identify this panicle. On the contrary, when
422 less than 60% of informants used the same
423 name, we considered that the panicle was

424 named inconsistently, i.e., no consensus exists
425 among farmers for its identification.

*Distance-Based Analysis of Knowledge on Panicle* 426
*Identification* 427

428 In this paper, differences of knowledge between
429 ethnolinguistic groups were measured by compar-
430 ing, on the one hand, their level of knowledge
431 heterogeneity, i.e., inter-individual variations in
432 the identification of panicles, and on the other
433 hand, their landrace identification, i.e., the name
434 they associated to each panicle. First, we analyzed
435 patterns of informants' knowledge for the whole
436 panicle set ($n = 287$), and then separately for subsets
437 of panicles corresponding to each of the genetic
438 clusters.
439 The heterogeneity of knowledge within the
440 group for the identification of the panicle set was
441 measured as the dispersion of informants' answers.
442 This heterogeneity degree was compared between
443 the Chuka, Tharaka, and Mbeere groups. For this
444 purpose, a similarity index was computed for each
445 pair of informants as the proportion of panicles they
446 named identically, and a distance index was then
447 constructed by subtracting the similarity index to
448 one. Average within-group dispersion, i.e., the av-
449 erage distance of individuals to group centroid in
450 the space of the simple matching distance index, was
451 used as a measure of knowledge heterogeneity with-
452 in groups. We conducted analyses to assess farmers'
453 knowledge heterogeneity for the identification of
454 the whole panicle set on one hand, and of each set
455 corresponding to each of the four genetic groups on
456 the other hand. Knowledge heterogeneity degree
457 was compared among ethnolinguistic groups by
458 testing if the average within-group dispersion was
459 equivalent among groups through running an anal-
460 ysis of multivariate homogeneity of groups' disper-
461 sions (PERMDISP2), which was performed on the
462 distance matrix (Anderson 2006). Further, pairwise
463 Tukey's honest significant difference (HSD) tests
464 were performed to test for the significance of
465 pairwise differences in the average dispersion be-
466 tween ethnolinguistic groups. PERMDISP2 and
467 Tukey's HSD analyses were run globally on the
468 whole panicle set.
469 Then, we tested if significant differences in land-
470 race identification exist between ethnolinguistic
471 groups, i.e., if they used different names to identify
472 panicles in each set. This was done by testing if the
473 centroids were equivalent for all groups, using a

474 non-parametric multivariate analysis of variance
475 (ADONIS, Anderson 2001). This test is a multi-
476 variate equivalent of ANOVA, based on the com-
477 parison of within- and between-group average dis-
478 persion. ADONIS results can be confidently
479 interpreted only if within-group dispersion is equiv-
480 alent among ethnolinguistic groups, which is tested
481 by PERMDISP2. A principal coordinates analysis
482 (PCoA) was conducted on the knowledge distance
483 matrix to visualize knowledge patterns within and
484 between groups. All distance-based analyses were
485 performed using the R package *vegan* version 2.4–
486 0 (Oksanen et al. 2012).

487 Both the analysis of multivariate homogeneity of
488 groups' dispersions (PERMDISP2) and the non-
489 parametric multivariate analysis of variance
490 (ADONIS) were run separately for each genetic
491 cluster after performing them on the whole panicle
492 set ($n = 287$). The correspondence between the
493 structure of sorghum morphological diversity and
494 knowledge patterns was assessed by displaying the
495 "consistent" names (i.e., used by more than 60% of
496 informants) using different colors on the PCoA
497 scatterplot of morphological distances for each
498 ethnolinguistic group.

## Results

### KNOWLEDGE HETEROGENEITY WITHIN AND BETWEEN ETHNOLINGUISTIC GROUPS

502 A large number of landrace names were cited
503 during the identification experiment (Chuka, 30
504 names; Tharaka, 36; Mbeere, 39), but only a few
505 names were used consistently at least one time, i.e.,
506 by more than 60% of the informants, to identify the
507 same panicle (Chuka, 7 names; Tharaka, 6; Mbeere,
508 5). Most landrace names used consistently were
509 common to the three ethnolinguistic groups, but
510 some were peculiar to one or two groups. In most
511 cases, each consistent name was used to identify
512 several panicles, but some were used to identify only
513 one or two panicles.

514 Significant differences of knowledge heterogene-
515 ity level were observed between ethnolinguistic
516 groups for the identification of the whole panicle
517 set ($n = 293$). Indeed, within-group dispersion (dis-
518 tance to a group's centroid) differed significantly
519 between groups, as indicated by PERMDISP2 re-
520 sults ($F = 8.55$, $p$ value < 0.001). A Tukey HSD test
521 further showed that the mean within-group disper-
522 sion was significantly lower in the Chuka group

523 (0.27) than in the Tharaka (0.39) and Mbeere
524 (0.38) groups, indicating a higher knowledge ho-
525 mogeneity in the former group. Differences of
526 knowledge between the Mbeere and the two other
527 groups are displayed along the second PCoA axis
528 (Fig. 2), but the significance of these differences
529 cannot be assessed based on ADONIS results be-
530 cause of differences in within-group dispersion.
531 Furthermore, two different knowledge subgroups
532 are distinguished within the Mbeere group along
533 the first PCoA axis, and part of the Tharaka appears
534 close to the Chuka, while the rest presents differ-
535 ences in knowledge.

536 Strong differences in consistency were observed
537 among panicles. The proportion of informants who
538 used the same landrace name varied among pani-
539 cles, ranging from a minimum of 19% in the Chuka
540 and Mbeere groups and 25% in the Tharaka group,
541 to a maximum of 100% in the three groups (ESM
542 Fig. 1). Overall, the proportion of panicles identi-
543 fied consistently varied strongly among
544 ethnolinguistic groups, with 80% in the Chuka
545 group, 40% in Tharaka, and 32% in Mbeere ($n =$
546 293 panicles).

547 Farmers' knowledge was homogeneous within
548 and between ethnolinguistic groups for some pani-
549 cles, which were identified highly consistently in all
550 groups. This was especially noticed for a set of
551 panicles named *Kaguru* by a large majority of infor-
552 mants (Fig. 3). In other cases, the level of knowledge
553 heterogeneity differed between groups as some pan-
554 icles were identified highly consistently in one
555 group and not in the others. This was especially
556 striking for a set of panicles named *Gadam* by most
557 Chuka informants, but identified inconsistently in
558 the two other groups. A similar situation was ob-
559 served for panicles consistently named *Murugue* and
560 *Mugeta* by the Chuka and Tharaka, and for panicles
561 consistently named *Ngirigacha* by the Mbeere,
562 which were identified inconsistently in the other
563 groups. Last, a high degree of knowledge heteroge-
564 neity was observed both within and between
565 ethnolinguistic groups for some panicles that were
566 named inconsistently in all groups.

### DIFFERENCES IN KNOWLEDGE ACCORDING TO PANICLES' CHARACTERISTICS

569 An adequate number of classes to describe the
570 genetic diversity in our dataset was $K = 4$ according
571 to the BIC criterion in $K$-means algorithm. Genetic
572 clusters matched partially with the structure of mor-
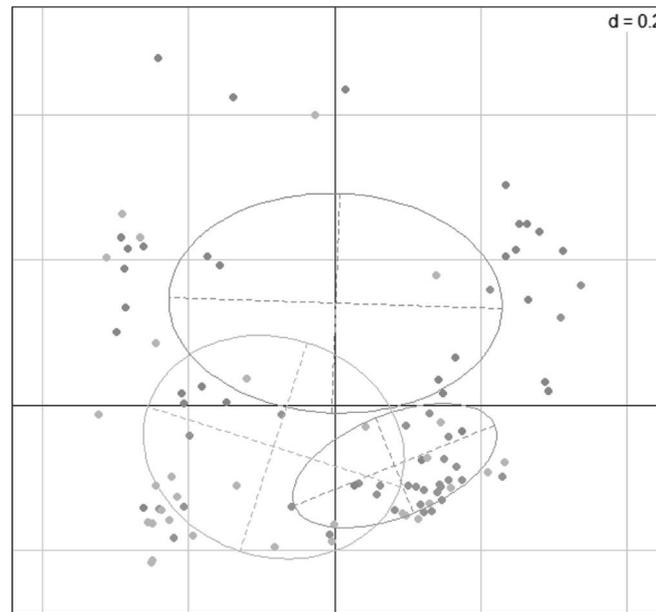573 phological diversity, some genetic clusters being

**Fig. 2.** Knowledge similarity among individuals according to their ethnolinguistic membership. Plot of the two first axis of the PCoA based on knowledge distance matrix between informants ($n = 96$ informants; the first component expresses 23% of the total variation, and the second one expresses 17%). Ethnolinguistic groups are displayed in colors (red, Chuka; green, Tharaka; blue, Mbeere).

morphologically distinct while others overlapped (Fig. 4). G2 was highly homogeneous and distinct morphologically, while G1, G3, and G4 were heterogeneous and overlapped. Overlap was especially high between G3 and G4.

Knowledge heterogeneity was similar among ethnolinguistic groups for G1, G3, and G4 clusters, as no significant differences of answer dispersion were detected by PERMDISP2 ($1.41 < F < 1.93$, $p$ value $> 0.1$). Differences of knowledge heterogeneity between ethnolinguistic groups were significant only for the genetic cluster G2 ($F = 11.04$, $p$ value $< 0.001$; ESM Appendix 2). Indeed, the large majority of Chuka informants agree on the identification of panicles assigned to the G2 genetic group that they named *Gadam*, while Mbeere and Tharaka informants presented a high level of knowledge heterogeneity.

Differences of knowledge on panicle identification were observed between ethnolinguistic groups for some genetic clusters (G3 and G4) while not for others (G1). For G1, no significant differences of identification were observed (ADONIS: $F = 1.71$, $p$ value $= 0.044$; ESM Fig. 2) as knowledge homogeneity was very high both within and between groups. Indeed, this cluster

was mainly composed of panicles consistently named *Kaguru* by informants in all groups (Fig. 5). An ADONIS test on centroid difference between groups was significant for three clusters (G2 to G4). Significant differences of groups' centroid for G3 and G4 traduce differences of identification among groups because within-group dispersion was similar for these clusters. However, such interpretation cannot be applied to G2 because significant differences of within-group dispersion were detected by PERMDISP2.

Ethnolinguistic groups differed significantly in their identification of panicles in G3 (ADONIS: $n = 97$, $F = 7.89$, $p$ value $< 0.001$). This genetic cluster comprised three morphological sub-clusters corresponding to different landrace names that were identified with very different consistency levels by the different groups (Fig. 6). A first sub-cluster was composed of a large number of panicles consistently named *Muruge* by the Chuka, while the Tharaka and Mbeere identified consistently only a small part of them. A second sub-cluster included panicles consistently named *Mugeta* by the Tharaka and to a lesser extent by the Chuka, while the Mbeere were inconsistent in their identification. Last, a third sub-cluster included a few panicles consistently named
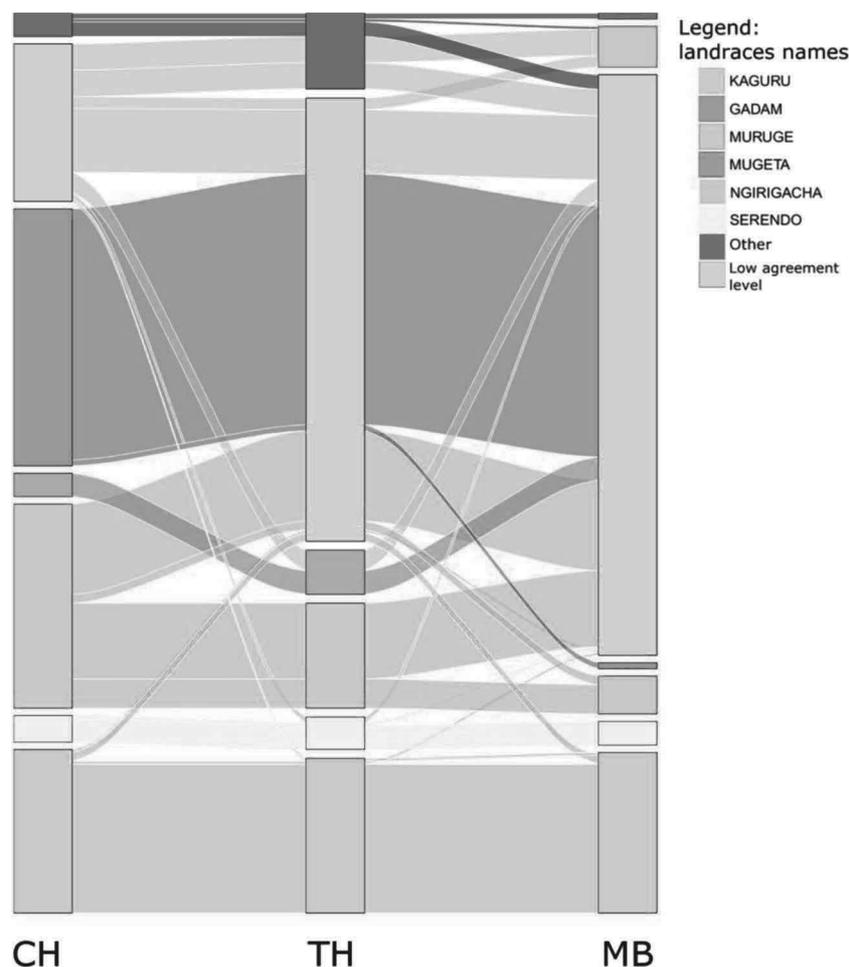
**Fig. 3.**  Comparison of panicle identification between ethnolinguistic groups. Boxes represent the proportion of panicles (*y*-axis) identified consistently (in colors) or not (in gray) in each ethnolinguistic group (*x*-axis). Flows among boxes in the different ethnolinguistic groups represent the share of panicles that were identified similarly or differently between groups. For instance, the width of the flow between box A in group X and box B in group Y represents the proportion of panicles identified as landrace A in group X that was identified as landrace B in group Y.

*Ngirigacha* by the Mbeere only. Ethnolinguistic groups also differed significantly in their identification of panicles in G4 (ADONIS: $n = 17$, $F = 3.62$, $p$ value < 0.001). This cluster mainly included panicles identified consistently as *Muruge*, and, in a lesser extent, *Serendo* and *Kaguru* in proportions differing among groups. In addition, G3 and G4 also included a large share of panicles identified inconsistently in proportions that varied strongly between groups. These panicles were displayed in between the morphological groups corresponding to panicles identified consistently, for which

knowledge homogeneity was high, on the PCoA of morphological distance.

G3 and G4 genetic clusters displayed a large morphological heterogeneity and overlapped. Part of the plants in these clusters hence presented morphological similarities despite their genetic differences. Interestingly, some panicles presenting genetic differences were named similarly because of their morphological similarity. In particular, panicles named *Muruge* in all ethnolinguistic groups presented morphological similarity although they belong to both G3 and G4 genetic clusters. Similarly, some
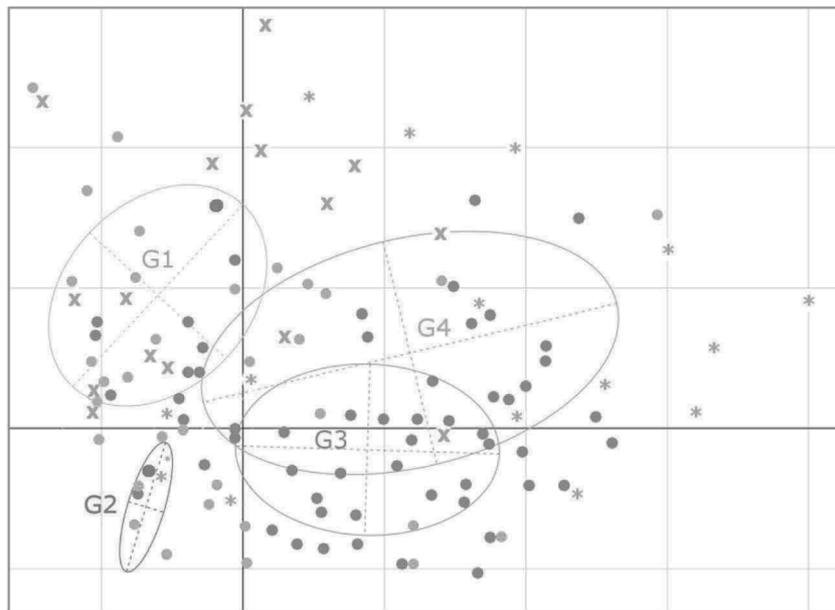
# AUTHOR'S PROOF

**Fig. 4.** Match between genetic and morphological diversity. Plot of the first two axes of the PCoA based on panicles' morphological traits ($n = 287$ panicles, 16 traits; the first component expresses 26% of the total variation, and the second one expresses 14%). Genetic clusters are displayed in colors ($n = 170$, panicles morphotyped but not genotyped are displayed in gray).
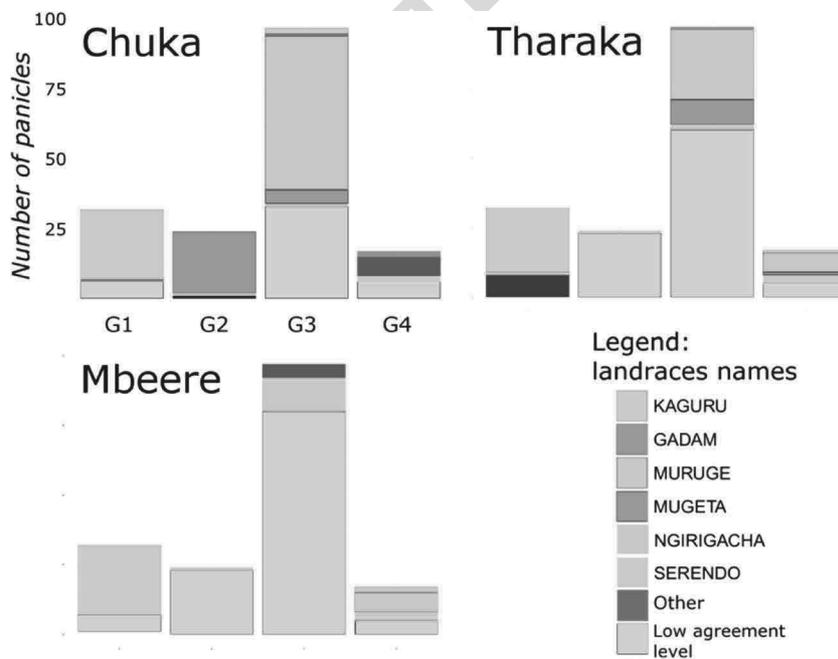


**Fig. 5.** Match between genetic clusters and landrace identification. Bars represent the number of panicles identified consistently (> 60% of informants, in colors) and inconsistently (< 60%, in black) in each genetic cluster.
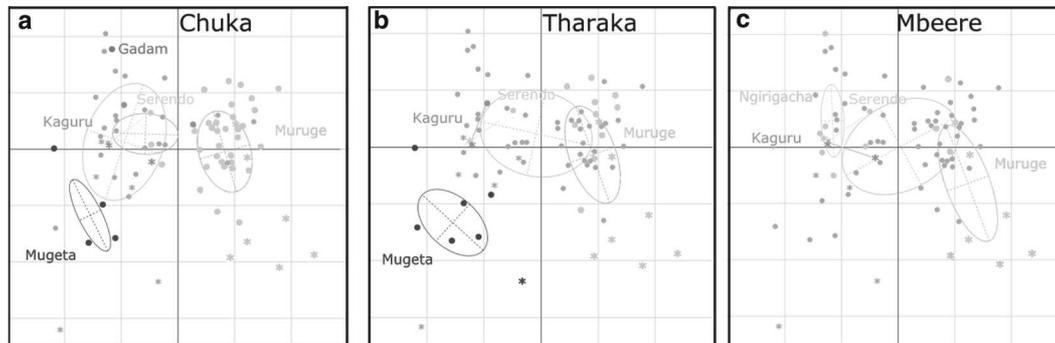
**Fig. 6.**    Correspondence between morphological diversity and landrace identification for panicles assigned to G3 (dots) and G4 (stars) genetic clusters (*n* = 114 panicles, 16 traits). Plot of the two first axes of the PCoA based on panicle morphological traits (variability expressed: 1st Co = 23%, 2nd Co = 15%). Colors correspond to landrace names identified consistently in each ethnolinguistic group, and panicles identified inconsistently are displayed in gray.

650  panicles assigned to G4 but morphologically similar
651  to G1 were named *Kaguru.*

652              **Discussion**

653      In this paper, we analyzed farmers' knowledge of
654  sorghum landrace identification in an
655  ethnolinguistic contact zone. Our results show that
656  knowledge varies according to both individuals'
657  cultural identity and panicle characteristics. We first
658  assessed whether farmers within ethnolinguistic
659  groups shared knowledge concerning landrace iden-
660  tification by measuring within-group knowledge
661  heterogeneity, and then tested whether the names
662  used to identify panicles differed between groups.
663  Results showed that groups differed significantly in
664  their level of knowledge heterogeneity, the Chuka
665  displaying the highest homogeneity for the identifi-
666  cation of the whole panicle set. Furthermore,
667  within-group knowledge heterogeneity varied
668  strongly among panicles on one hand, and differed
669  between groups for some panicles on the other
670  hand. This indicates that knowledge heterogeneity
671  within and between ethnolinguistic groups is related
672  to panicle characteristics.
673      We further conducted analyses to test for the
674  effect of panicle genetic and morphological charac-
675  teristics on knowledge heterogeneity within and
676  between ethnolinguistic groups. Results first
677  showed that within-group knowledge heterogeneity
678  differed between groups for only one genetic cluster
679  (G2), indicating that Chuka shared common
680  knowledge on the identification of panicles in this
681  cluster, while Tharaka and Mbeere do not. Second-
682  ly, we found that knowledge was highly similar and

683  homogeneous among ethnolinguistic groups for the
684  identification of one cluster (G1), while significant
685  differences were observed among them for the iden-
686  tification of two clusters (G3 and G4). Further-
687  more, ethnolinguistic groups differ in their level of
688  knowledge heterogeneity for the identification of
689  the different morphotypes within G3 and G4.
690      The knowledge patterns we described reveal in-
691  sights on farmers' experience of the different land-
692  races, and on the diffusion of knowledge within and
693  between ethnolinguistic groups. Indeed, according
694  to Boster (1986), landrace identification, naming,
695  and classification regarding their morphological
696  characteristics are socially learned and further con-
697  structed by individuals through their direct experi-
698  ence with the plants. He identified three major
699  processes involved in inter-individual differences of
700  knowledge concerning landrace identification: (i)
701  differences among individuals in their learning
702  sources and pathways; (ii) differences in their expe-
703  rience of the landrace and its morphological charac-
704  teristics; and (iii) differences in the time or the
705  willingness individuals have for acquiring experi-
706  ence in this domain. As there is little support for
707  significant differences of inter-individual variations
708  of time and willingness to learn between
709  ethnolinguistic groups, the knowledge patterns we
710  observed reflect, on the one hand, modalities of
711  knowledge transmission within and between groups
712  and, on the other hand, differences between groups
713  in their level of experience concerning landrace
714  identification. This leads us to discuss knowledge
715  patterns regarding the history and characteristics of
716  panicles and knowledge transmission modalities
717  within and between groups (Cavalli-Sforza and

718 Feldman 1981; Leclerc and Coppens
719 d'Eeckenbrugge 2012; Reyes-García et al. 2009).

720 PANICLE CHARACTERISTICS AND HISTORY IN
721 RELATION TO ETHNOLINGUISTIC GROUPS

722 Panicles presented to informants display different
723 morphological characteristics, and have different
724 histories and origins. These differences help explain
725 the strong variations of within-group knowledge
726 heterogeneity between panicles and its variations
727 between ethnolinguistic groups for the same pani-
728 cle. Indeed, farmers share knowledge on panicles
729 presenting morphological characteristics with which
730 they are collectively familiar. Several characteristics
731 of landraces can contribute to building collective
732 knowledge, such as the time since they are cultivat-
733 ed, their popularity, and whether they can be easily
734 identified and distinguished based on their morpho-
735 logical characteristics.

736 A previous genetic study, which included the 170
737 individuals we analyzed here, showed that sorghum
738 landraces in our study area present different histo-
739 ries and agronomic characteristics (Labeyrie et al.
740 2014). It identified four genetic clusters matching
741 with those we identified, showing that clusters G1
742 and G2 were introduced varieties released by the
743 formal breeding system, whereas G3 and G4 were
744 local landraces with different agro-morphological
745 characteristics.

746 Knowledge was highly homogeneous within
747 and between groups for the identification of
748 G1 panicles that were named *Kaguru* by most
749 informants. By contrast, differences were observed
750 among groups for G2 as knowledge was highly
751 homogeneous in the Chuka group where it was
752 named *Gadam*, while it was heterogeneous in the
753 Mbeere and Tharaka groups. Differences in the
754 dates of dissemination of these two improved vari-
755 eties likely explain why G1 was identified consis-
756 tently in all groups whereas G2 was only in the
757 Chuka group. Indeed, *Kaguru* was introduced sev-
758 eral decades ago and has been widely cultivated and
759 sold in markets in the area, while *Gadam* was re-
760 leased in the area only two years before our study.
761 Our results further suggest that G2 introduction
762 started in the Chuka group, who display a strong
763 collective experience of its identification. Such high
764 knowledge uniformity may result from massive dis-
765 semination of G2 under the well-defined "*Gadam*"
766 name by Kenyan agricultural extension services in
767 the Chuka group. By contrast, knowledge hetero-
768 geneity for G2 identification in the Tharaka and

769 Mbeere groups indicates that they are not yet famil-
770 iar with its characteristics and name.

771 Clusters G3 and G4 included panicles identified
772 consistently with various landrace names as well as
773 panicles identified inconsistently, in proportions
774 that vary strongly between ethnolinguistic groups.
775 This variety of landrace names was associated with
776 differences in panicles' morphological characteristics
777 within these clusters, the different landrace names
778 corresponding to different morphotypes. Knowl-
779 edge heterogeneity was similar among groups for
780 these genetic clusters, but it varied between
781 morphotypes within genetic clusters, especially
782 G3. Ethnic groups appeared to be respectively fa-
783 miliar with the identification of different
784 morphotypes, some being identified highly consis-
785 tently in one ethnolinguistic group while not in
786 others and conversely. The Chuka in particular
787 consistently identified a set of morphologically sim-
788 ilar panicles as *Muruge*, while knowledge within the
789 two other groups was more heterogeneous. A sim-
790 ilar situation was observed for a set of panicles
791 identified as *Mugeta* by most Tharaka informants.
792 This probably results from differences in experience
793 that ethnolinguistic groups developed over time
794 concerning these different landraces. This hypothe-
795 sis is supported by previous results showing that
796 *Muruge* and *Mugeta* landraces were respectively in-
797 troduced by the Chuka and the Tharaka in the
798 study area, which explains their respective higher
799 knowledge uniformity for the identification of cor-
800 responding morphotypes (Labeyrie et al. 2014,
801 2016b).

802 Last, our study showed that knowledge hetero-
803 geneity was very high in all ethnolinguistic groups
804 for some panicles. Our hypothesis is that these
805 panicles present a combination of traits that do
806 not correspond to the traits on which the landrace
807 classification system used by most informants is
808 based. Either these panicles may result from crosses
809 and combine morphological characteristics of sever-
810 al landraces, which is confusing for informants, or
811 they may belong to recently introduced or rare
812 landraces, whose characteristics are not yet familiar
813 to most informants.

814 The existence of a collective coherence among
815 farmers for landrace identification has been debated
816 in several studies (Sadiki et al. 2007). Some studies
817 reported high knowledge heterogeneity among
818 farmers, such as Salick et al. (1997), on cassava
819 nomenclature among Amuesha in the Peruvian
820 Amazon. Others found higher knowledge homoge-
821 neity within villages than between villages, and

postulated that geographic distance that limits seed circulation also limits knowledge transmission (Nuijten and Almekinders 2008). However, knowledge of different farmers for the identification of the same plant was not measured since Bosters' work was based on the cultural consensus framework (1986). Our study shows that farmers' consistency in identifying crops varies strongly depending on plants' characteristics and history, but varies also according to the cultural background of farmers. Hence, farmers living in the same location can present strong differences of knowledge because of their cultural differences. Our results are thus in line with those of Perales et al. (2005) on maize in Mexico, and contribute to explaining the differences in selection practices they observed between adjacent ethnolinguistic groups.

KNOWLEDGE TRANSMISSION MODALITIES

Knowledge heterogeneity within and between ethnolinguistic groups indicates that communication and landrace circulation is limited between them. Indeed, we would expect high knowledge homogeneity both within and between groups if circulation was not limited (Romney et al. 1986). Sharing a common nomenclature depends on the path for learning and knowledge transmission. As landrace names are used for communicating, learning what is the culturally appropriate name to label a given morphotype is essential for farmers. Socially related farmers exchange more information concerning landraces than farmers that are not related, and they are hence more likely to share a common taxonomy (Boster 1986). First, the overall higher knowledge homogeneity in the Chuka group indicates more intense knowledge transmission between individuals within this group than in the two others. Interestingly, this can be put in relation to results from a previous study in the same area (Labeyrie et al. 2016a), which showed that the seed circulation network was more dense and cohesive in the Chuka group than in the two others, possibly because it was the first group to settle in the study area. It is likely that these properties of seed circulation networks enhance knowledge homogenization between individuals concerning landraces.

Second, our results indicate that different modalities of knowledge diffusion exist for the different genetic clusters and morphotypes. Knowledge for G1 identification was highly homogeneous, indicating intense knowledge diffusion within and between ethnolinguistic groups. These results are in line with those of previous studies showing that G1 corresponds to *Kaguru* improved variety, which is cultivated by most farmers and sold on the market (Labeyrie et al. 2014). Then, major differences of knowledge heterogeneity for G2 between the Chuka and the other groups suggest first that sources of knowledge differed between them, as diffusion of the *Gadam* improved variety appears to have been bound to the former group. Furthermore, knowledge diffusion appeared limited between groups for this variety, but this is not surprising as time since its introduction was very short when the study was conducted. Knowledge patterns concerning identification of the different morphotypes within G3 indicate a higher knowledge similarity between Chuka and Tharaka than between these groups and the Mbeere. Higher knowledge similarity was especially observed between the two former groups for the identification *Muruge* and *Mugeta* landraces. This is likely related to the intensity of interpersonal seed exchanges between Chuka and Tharaka, which is the major seed circulation modality for these varieties, while exchanges with the Mbeere were very rare (Labeyrie et al. 2016a). These seed circulation patterns were linked to a strong alliance relationship between Chuka and Tharaka groups.

Several studies reported a coincidence between seed circulation networks and knowledge patterns. For instance, cassava circulation was found to be more intense among kin, who also display more similar knowledge on landrace identification among the Aguaruna in the Amazon (Boster 1986), and similar results were observed among members of the same village for rice in Gambia (Nuijten and Almekinders 2008). Other studies showed that people exchanging more seeds were also the most knowledgeable, for instance among home garden keepers in the Catalan Pyrenees (Calvet-Mir et al. 2012), or among *caboclo* cassava farmers in Brazilian Amazonia (Kawa et al. 2013). Our results are partly in line with these studies, as we found a correspondence between the structure of seed circulation networks and knowledge patterns for some landraces, but interestingly, this was not the case for others. For instance, knowledge for *Gadam* identification was strongly shared by the Chuka but not with the Tharaka, despite our previous study which had not detected any limitation to seed circulation between both groups. In addition, knowledge for *Kaguru* identification was shared by the three groups, despite this same study showing that seed circulation was limited between the Mbeere and the

two other groups. An explanation for such a discrepancy between the structure of seed circulation networks among farmers and their knowledge patterns could be that they obtain a large share of their seed on the local market as well as through extension services and NGOs. That is especially the case for introduced varieties released by research (unpublished results). Furthermore, these results suggest that farmers learn about landrace identification through different pathways depending on its nature, i.e., local or introduced, and that seed and knowledge of landraces do not necessarily circulate through the same channels.

## Conclusion

This study shows that knowledge concerning sorghum landrace identification differs between adjacent ethnolinguistic groups and that it varies according to landraces' characteristics. First, our results indicate that farmers in each ethnolinguistic group present a high level of collective consistency in naming what they consider their own sorghum landraces and anciently introduced varieties. This suggests that the consensus level of landrace identification reflects how long it has been cultivated by the human group. Second, our results suggest that social pathways for learning could play a major role in shaping knowledge, as geographically close cultural groups present major differences in knowledge on landraces.

These results open perspectives to understand farmers' seed selection practices, a major driver of crop evolution and adaptation in situ. Indeed, the way individuals perceive, represent, and classify their environment affects their management practices (Atran and Medin 2008). The effect of intercultural differences was especially observed in maize seed selection practices in Mexico, where populations of this crop present divergent morphological characteristics between adjacent villages (Pressoir and Berthaud 2004) and ethnolinguistic groups (Perales et al. 2005). Our study suggests that such divergence in crop selection practices could result from differences in the identification and classification of landraces by the different human groups. It advocates for further integration of anthropology in crop diversity research as crops not only are biological objects but also bear the imprint of the societies in which they are grown, exchanged, and selected (Harlan 1975).

## Literature Cited

Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecology 26(1):32–46.

———. 2006. Distance-based tests for homogeneity of multivariate dispersions. Biometrics 62(1): 245–253.

Atran, S. and D. Medin. 2008. The native mind and the cultural construction of nature. Cambridge, Massachusetts: MIT Press.

Badstue, L. B., M. R. Bellon, J. Berthaud, A. Ramirez, D. Flores, and X. Juárez. 2007. The dynamics of farmers' maize seed supply practices in the central valleys of Oaxaca, Mexico. World Development 35(9):1579–1593.

Barnaud, A., M. Deu, E. Garine, D. McKey, and H. Joly. 2007. Local genetic diversity of sorghum in a village in northern Cameroon: Structure and dynamics of landraces. Theoretical and Applied Genetics 114(2):237–248.

Bellon, M. R. and S. B. Brush. 1994. Keepers of maize in Chiapas, Mexico. Economic Botany 48(2):196–209.

Boster, J. S.1985. Selection for perceptual distinctiveness: Evidence from Aguaruna cultivars of Manihot esculenta. Economic Botany 39(3): 310–325.

———.1986. Exchange of varieties and information between Aguaruna manioc cultivators. American Anthropologist 88(2):428–436.

Cavalli-Sforza, L. L. and M. W. Feldman. 1981. Cultural transmission and evolution: A quantitative approach. Princeton, New Jersey: Princeton University Press.

Darwin, C. 1859. On the origins of species by means of natural selection. London, England: Murray.

Diamond, J. 2002. Evolution, consequences and future of plant and animal domestication. Nature 418(6898):700–707.

Dray, S. and A. B. Dufour. 2007. The ade4 package: Implementing the duality diagram for ecologists. Journal of Statistical Software 22(4):1–20.

Fadiman, J. A. 1993. When we began there were witchmen: An oral history from Mount Kenya. Berkeley: University of California Press.

Friedberg, C. 1991. Operative aspects of folk classification. In: Man and a half: Essays in Pacific anthropology and ethnobiology in honour of Ralph Bulmer. Ed. A. Pawley, 102–109. Auckland, New Zealand: The Polynesian Society

Gepts, P. 2004. Crop domestication as a long-term selection experiment. Plant Breeding Review 24(2):1–44.

Gibson, R. W. 2009. A review of perceptual distinctiveness in landraces including an analysis of how its roles have been overlooked in plant breeding for low-input farming systems. Economic Botany 63(3):242–255.

Glazier, J. 1970. Ritual and social conflict: Circumcision and oath-taking in Mbeere. Nairobi, Kenya: Institute for Development Studies, University of Nairobi.

Hammarström, H., R. Forkel, M. Haspelmath, and S. Banak. 2015. Glottolog 2.4. Leipzig, Germany: Max Planck Institute for Evolutionary Anthropology.

Harlan, J. R. 1975. Crop and man. Foundations for modern crop science series. Madison, Wisconsin: American Society of Agronomy.

Harlan, J. R., J. M. J. De Wet, and A. B. L. Stemler. 1976. Origins of African plant domestication. The Hague, Netherlands: De Gruyter-Mouton.

Heine, B. and W. J. G. Möhlig. 1980. Language and dialect atlas of Kenya. Vol. I. Berlin, Germany: Dietrich Reimer Verlag.

Hodgkin, T., R. Rana, J. Tuxill, D. Balma, A. Subedi, I. Mar, R. Karamura, R. Valdivia, L. Collado, L. Latournerie, et al. 2007. Seed systems and crop genetic diversity in agroecosystems. In: Managing biodiversity in agricultural ecosystems, eds. D. Jarvis, C. Padoch, and D. Cooper, 77–116. New York: Colombia University Press.

IPGRI (International Plant Genetic Resources Institute). 1993. Descriptors for sorghum (Sorghum bicolor [L.] Moench). Rome, Italy: IBPGR / ICRISAT.

Jaetzold, R., H. Schmidt, B. Hornetz, and C. Shisanya. 2007. Farm management handbook. Vol. II, Part C, East Kenya. Subpart C1, Eastern Province. Nairobi, Kenya: Ministry of Agriculture / GTZ.

Jarvis, D., A. Brown, P. Cuong, L. Collado-Panduro, L. Latournerie-Moreno, S. Gyawali, T. Tanto, M. Sawadogo, I. Mar, M. Sadiki, et al. 2008. A global perspective of the richness and evenness of traditional crop-variety diversity maintained by farming communities. Proceeding of the National Academy of Sciences of the USA 105(14):5326–5331.

Jombart, T. 2008. adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics 24(11):1403–1405.

Jombart, T.., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genetics 11:94. https://doi.org/10.1186/1471-2156-11-94.

Kawa, N. C., C. McCarty, and C. R. Clement. 2013. Manioc varietal diversity, social networks, and distribution constraints in rural Amazonia. Current Anthropology 54(6):764–770.

Labeyrie, V., M. Deu, A. Barnaud, C. Calatayud, M. Buiron, P. Wambugu, S. Manel, J.-C. Glaszmann, and C. Leclerc. 2014. Influence of ethnolinguistic diversity on the sorghum genetic patterns in subsistence farming systems in Eastern Kenya. PloS One 9:3. https://doi.org/10.1371/journal.pone.0092178.

Labeyrie, V., M. Thomas, Z. K. Muthamia, and C. Leclerc. 2016a. Seed exchange networks, ethnicity, and sorghum diversity. Proceeding of the National Academy of Sciences of the USA 113(1):98–103.

Labeyrie, V., M. Deu, Y. Dussert, B. Rono, F. Lamy, C. Marangu, D. Kiambi, C. Calatayud, G. Coppens d'Eeckenbrugge, and T. Robert. 2016b. Past and present dynamics of sorghum and pearl millet diversity in Mount Kenya region. Evolutionary Applications 9(10):1241–125.

Lapeña, I. and M. Halewood. 2016. Farmers' varieties and farmers' rights: Challenges at the crossroads of agriculture, taxonomy and law. In: Farmers' crop varieties and farmers' rights, ed. M. Halewood, 15–38. New York: Routledge.

Leclerc, C., and G. Coppens d'Eeckenbrugge. 2012. Social organization of crop genetic

diversity. The G x E x S interaction model. Diversity 4(1):1–32.

Louette, D. and M. Smale. 2000. Farmers' seed selection practices and traditional maize varieties in Cuzalapa, Mexico. Euphytica 113(1):25–41.

McGuire, S. 2008. Securing access to seed: Social relations and sorghum seed exchange in eastern Ethiopia. Human Ecology 36(3):217–229.

Middleton, J. 1953. The central tribes of the northeastern Bantu. London, U.K.: International African Institute.

Mucioki, M., S. K. Mucioki, and T. Johns. 2014. Intraspecific diversity and seed management of pearl millet (*Pennisetum glaucum*) in Tharaka, Kenya: A persistent and valued traditional food crop. Economic Botany 68(4):397–409.

Mwaniki, H. S. K. 1973. The living history of Embu and Mbeere to 1906. Nairobi, Kenya: East African Literature Bureau.

Nuijten, E. and C. J. Almekinders. 2008. Mechanisms explaining variety naming by farmers and name consistency of rice varieties in the Gambia. Economic Botany 62(2):148–160.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. O'Hara, G. L. Simpson, P. Solymos, M. Stevens, and H. Wagner. 2012. Vegan: Community ecology package. Version 2.4–0.

Perales, H., B. Benz, and S. Brush. 2005. Maize diversity and ethnolinguistic diversity in Chiapas, Mexico. Proceedings of the National Academy of Sciences of the USA 102(3):949–954.

Pressoir, G. and J. Berthaud. 2004. Population structure and strong divergent selection shape phenotypic diversification in maize landraces. Heredity 92(2):95–101.

Quiros, C. F., S. Brush, D. S. Douches, K. S. Zimmerer, and G. Huestis. 1990. Biochemical and folk assessment of variability of Andean cultivated potatoes. Economic Botany 44(2): 254–266.

R. Core Team. 2016. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Reyes-García, V., J. Broesch, L. Calvet-Mir, N. Fuentes-Peláez, T. W. McDade, S. Parsa, S. Tanner, T. Huanca, W. R. Leonard, and M. R. Martínez-Rodríguez. 2009. Cultural transmission of ethnobotanical knowledge and skills: An empirical analysis from an Amerindian society. Evolution and Human Behavior 30(4):274–285.

Romney, A. K., S. C. Weller, and W. H. Batchelder. 1986. Culture as consensus: A theory of culture and informant accuracy. American Anthropologist 88(2):313–338.

Sadiki, M., D. Jarvis, D. Rijal, J. Bajracharya, N. Hue, T. Camacho-Villa, L. Burgos-May, M. Sawadogo, D. Balma, D. Lope, et al. 2007. Variety names: An entry point to crop genetic diversity and distribution in agroecosystems? In: Managing biodiversity in agricultural ecosystems, eds. D. Jarvis, C. Padoch, and D. Cooper, 34–76. New York: Colombia University Press.

Salick, J., N. Cellinese, and S. Knapp. 1997. Indigenous diversity of cassava: Generation, maintenance, use and loss among the Amuesha, Peruvian Upper Amazon. Economic Botany 51(1):6–19.

Snowden, J. D. 1936. The cultivated races of sorghum. Kew, U.K.: Royal Botanic Gardens.

Soler, C., A. Saidou, T. Hamadou, M. Pautasso, J. Wencelius, and H. Joly. 2013. Correspondence between genetic structure and farmers' taxonomy: A case study from dry-season sorghum landraces in northern Cameroon. Plant Genetic Resources: Characterization and Utilization 11(1): 36–49.

vom Brocke, K., A. Christinck, R. E. Weltzien, T. Presterl, and H. H. Geiger. 2003. Farmers' seed systems and management practices determine pearl millet genetic diversity patterns in semiarid regions of India. Crop Science 43(5):1680–1689.

# AUTHOR'S PROOF

## AUTHOR QUERIES

**AUTHOR PLEASE ANSWER ALL QUERIES.**

Q1. Please check the suggested running page title if appropriate. Otherwise, please provide a short running title with a maximum of 65 characters including spaces.

Q2. Please check if the affiliations are presented correctly.

Q3. Please check if the modification is appropriate in the sentence "The first two clusters included several morphotypes...."

Q4. Ref. "Calvet-Mir et al. 2012" is cited in the body but its bibliographic information is missing. Kindly provide its bibliographic information in the list.

UNCORRECTED PROOF