

Ethnolinguistic structuring of sorghum genetic diversity in Africa and the role of local seed systems

Ola T. Westengen^{a,b}, Mark Atam Okongo^c, Leo Onek^d, Trygve Berg^e, Hari Upadhyaya^{f,g,h}, Siri Birkeland^b, Siri Dharma Kaur Khalsa^b, Kristoffer H. Ring^a, Nils C. Stenseth^{b,1}, and Anne K. Brysting^b

^aCentre for Development and the Environment, University of Oslo, NO-0317 Oslo, Norway; ^bCentre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, NO-0316 Oslo, Norway; ^cMinistry of Agriculture, Eastern Equatoria State, Torit, South Sudan; ^dCollege of Applied and Industrial Sciences, Juba University, Juba, South Sudan; ^eNoragric, Norwegian University of Life Sciences, NO-1432 Aas, Norway; ^fInternational Crops Research Institute for the Semi-Arid Tropics, Patancheru 502 324, India; ^gDepartment of Agronomy, Kansas State University, Manhattan, KS 66506; and ^hUWA Institute of Agriculture, University of Western Australia, Crawley, WA 6009, Australia

Edited* by Hans R. Herren, Millennium Institute, Arlington, VA, and approved August 18, 2014 (received for review January 27, 2014)

Sorghum is a drought-tolerant crop with a vital role in the livelihoods of millions of people in marginal areas. We examined genetic structure in this diverse crop in Africa. On the continent-wide scale, we identified three major sorghum populations (Central, Southern, and Northern) that are associated with the distribution of ethnolinguistic groups on the continent. The codistribution of the Central sorghum population and the Nilo-Saharan language family supports a proposed hypothesis about a close and causal relationship between the distribution of sorghum and languages in the region between the Chari and the Nile rivers. The Southern sorghum population is associated with the Bantu languages of the Niger-Congo language family, in agreement with the farming-language codispersal hypothesis as it has been related to the Bantu expansion. The Northern sorghum population is distributed across early Niger-Congo and Afro-Asiatic language family areas with dry agroclimatic conditions. At a finer geographic scale, the genetic substructure within the Central sorghum population is associated with language-group expansions within the Nilo-Saharan language family. A case study of the seed system of the Pari people, a Western-Nilotic ethnolinguistic group, provides a window into the social and cultural factors involved in generating and maintaining the continent-wide diversity patterns. The age-grade system, a cultural institution important for the expansive success of this ethnolinguistic group in the past, plays a central role in the management of sorghum landraces and continues to underpin the resilience of their traditional seed system.

genetic resources | cultural selection | social-ecological adaptation

Sorghum [*Sorghum bicolor* (L.) Moench] is a drought-tolerant C₄ crop of major importance for food security in Africa (1, 2). The grain crop has played a fundamental role in adaptation to environmental change in the Sahel since the early Holocene, when the Sahara desert was a green homeland for Nilo-Saharan groups pursuing livelihoods based on hunting or herding of cattle and wild grain collecting (3, 4). The earliest archaeological evidence of human sorghum use is dated 9100–8900 B.P., and the seeds were excavated together with cattle bones, lithic artifacts, and pottery from a site close to the current border between Egypt and Sudan (5, 6). The timing of the domestication of cattle and sorghum remains contested due to limited archaeological evidence, but, at some point, the livelihoods in this region transformed from hunting and gathering into agropastoralism. Sorghum cultivation in combination with cattle herding was a successful livelihood adaptation to the dry grassland ecology, and, eventually, as the climate changed and the Sahel moved south, the agropastoral adaptation spread over large parts of the Central African steppes (7).

Recent molecular work on sorghum diversity (8–13) stands on the shoulders of J. R. Harlan and others' work from the 1960s–1980s. Diversity of sorghum types, varieties, and races has been related to movement of people, disruptive selection, geographic isolation, gene flow from wild to cultivated plants, and

recombination of these types in different environments (2, 14, 15). On the basis of morphology, Harlan and de Wet (16) classified sorghum into five basic and 10 intermediary botanical races (16). The race “bicolor” has small elongated grains, and, because of the “primitive” morphology, it is considered the progenitor of more derived races (16, 17). The race “guinea” has open panicles well adapted to high rainfall areas, and it is proposed that the “guinea margaritifera” type from West Africa represents an independent domestication (10, 12). The race “kafir” is associated with the Bantu agricultural tradition, and the race “durra” is considered well-adapted to the dryland agricultural areas along the Arabic trade routes from West Africa to India (14). The fifth race, “caudatum,” is characterized by “turtle-backed” grains, and Stemler et al. (ref. 17, p. 182) proposed that “the distribution of caudatum sorghums and Chari-Nile-speaking peoples coincide so closely that a causal relationship seems probable.” This hypothesis is considered plausible on the basis of historical linguistics, but it remains to be tested by independent evidence (3). The hypothesis is a specific version of the interdisciplinary “farming-language codispersal hypothesis,” which proposes that farming and language families have moved together through population growth and migration (18, 19).

The role of cultural selection and adaptation has been documented in many studies of domestication and translocation of crops (20, 21). The literature on the role of farmers' management in maintaining and enhancing genetic resources (22–26) is relevant to understanding how patterns of diversity visible at large spatial scales are caused by evolutionary processes

Significance

This paper shows that cultural factors play a key role in shaping the genetic structure in sorghum. We present molecular evidence of close associations between sorghum population structure and the distribution of ethnolinguistic groups in Africa. We show that traditional seed-management practices, which have played an important role for survival and expansion of agropastoral groups in the past, still are remarkably resilient to threats to human security. We argue that efforts to strengthen African sorghum seed systems are more likely to be successful when building on, rather than seeking to replace, existing traditional seed systems and landraces.

Author contributions: O.T.W., T.B., H.U., N.C.S., and A.K.B. designed research; O.T.W., M.A.O., L.O., T.B., S.B., and S.D.K.K. performed research; H.U., N.C.S., and A.K.B. contributed new reagents/analytic tools; O.T.W. and K.H.R. analyzed data; O.T.W., N.C.S., and A.K.B. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Data deposition: Data available from the Dryad Digital Repository, <http://datadryad.org> (doi: 10.5061/dryad.9fj76).

¹To whom correspondence should be addressed. Email: n.c.stenseth@ibv.uio.no.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1401646111/-DCSupplemental.

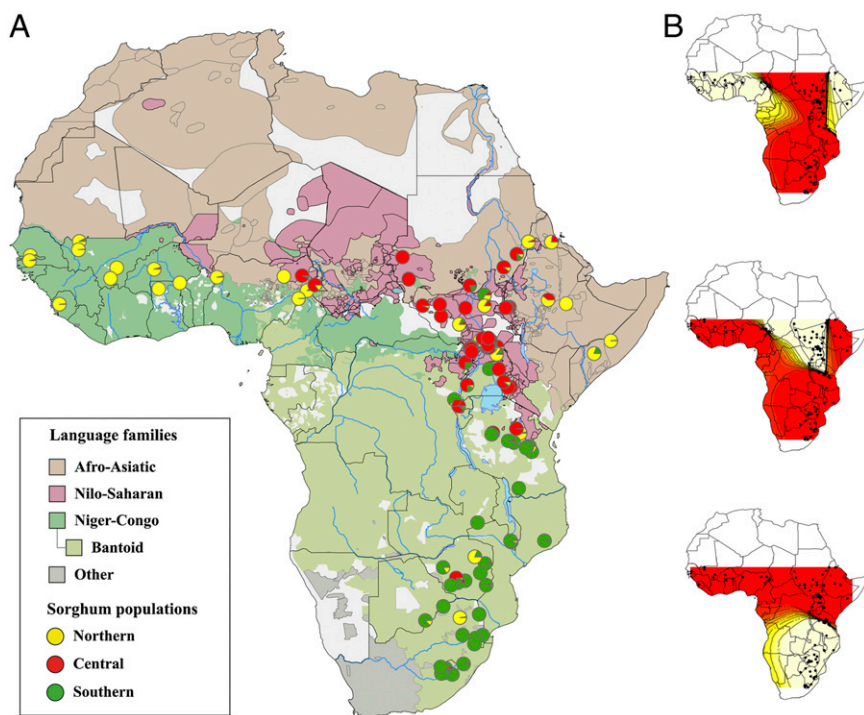


Fig. 1. Map of population structure of African sorghum. (A) Genetic population structure in sorghum based on 19 SSR markers superimposed on the distribution of language families (language distribution data from Ethnologue v. 16 compiled in ref. 62). Pie diagrams display membership of sorghum accessions to each of the three populations inferred with the program InStruct. (B) Maps of posterior probabilities of population membership in three populations inferred with GENELAND. The probability mapping area is defined by the distribution of the accessions. Lighter yellow colors denote a high probability of belonging to a population. Only accessions with collection-site coordinates in passport data are included ($n = 138$).

operating at finer scales. On-farm management of crop varieties and cultural boundaries influencing the diffusion of seeds, practices, and knowledge are important local-scale explanatory factors behind patterns of regional and continental scale associations between ethnolinguistic groups and crop genetic structure (27–30).

Knowledge on the role of social, cultural, and environmental factors in structuring crop diversity is important to assess the resilience of rural livelihoods in the face of global environmental change. Impact studies project that anthropogenic climate change will negatively affect sorghum yields in Sub-Saharan Africa (31, 32). Such projections pose questions about the availability of appropriate genetic resources and the ability of both breeding programs and local seed systems to develop the required adaptations in a timely manner (33, 34). Insight in local seed systems can contribute to more sustainable development assistance efforts aimed at building resilience in African agriculture in the face of climate change and human insecurity (25, 35).

Here, we present a study of geographic patterns in African sorghum diversity and its associations with the distribution of ethnolinguistic groups. First, we evaluate the proposed farming-language codispersal hypothesis by genotyping sorghum accessions from a continent-wide diversity panel (36). Second, to elucidate the local level mechanisms involved in generating and maintaining this diversity, we present a case study of the sorghum seed system of a group of descendants of the first Nilo-Saharan sorghum cultivators, the Pari people in South Sudan. By comparing accessions collected in 1983 with seeds sampled from the same villages in 2010 and 2013, we assess the resilience of the traditional Pari seed system during a period of civil war and climatic stress. We draw on environmental, linguistic, and anthropological evidence to understand the role of geographic, ecological, historical, and cultural factors in shaping sorghum genetic structure.

Results and Discussion

Continent-Wide Population Structure. Modeling of genetic structure in the continent-wide panel of genebank accessions consistently identifies three populations that are codistributed with language families in Africa (Fig. 1A). Bayesian model results from InStruct (37) and the complimentary algorithm in STRUCTURE (38, 39), evaluated with Evanno’s delta K method (40, 41), together indicate a likely population structure at $K = 3$ (Fig. S1A). Population assignment by the two approaches are highly correlated (Pearson’s $r = 0.93$ – 0.97) (Fig. S2A). Moreover, there is a strong correlation ($r = 0.84$ – 0.93) between membership coefficients generated on the basis of our 19 simple sequence repeats (SSRs) (Table S1) and those generated on the basis of 13,390 SNPs run on the 139 accessions in common between this study and a recent SNP study of the sorghum minicore collection (13) (Fig. S2B). These results show that our marker set reflects genome-wide diversity and that genotyping of one plant per accession is representative in this predominantly inbred species.

Evidence for codistribution of sorghum population structure and the distribution of language families was strengthened by spatially explicit Bayesian modeling implemented in the program GENELAND (42) for $K = 3$ (Fig. 1B). Applying a 0.90 population membership coefficient as cutoff value, we named the inferred populations according to their geographic distribution: Southern, Northern, and Central (Table S2). There are three major language families in areas where sorghum is cultivated in Africa: Niger-Congo, Afro-Asiatic, and Nilo-Saharan. We found that accessions originating from areas where only one of the three language families is represented ($n = 137$) showed a significantly higher membership coefficient to one of the populations than to the two others (Fig. S2C). The codistribution pattern was also apparent from spatially explicit inference of

population structure based on country centroid for accessions with only country-level origin data (Fig. S3).

The Central population is wedged in the middle of the distribution of the Northern population and extends southwards into the Great Lakes region in areas largely overlapping with the distribution of the Nilo-Saharan language family. Seventy-three percent of the accessions assigned to this population are from Nilo-Saharan language areas, and a further 22% are from areas where Nilo-Saharan languages are spoken together with languages from one of the two other families. There was a strong correlation between population membership and geographic origin in Nilo-Saharan language areas, both when applying a 0.90 membership coefficient cutoff and when no cutoff was applied (Spearman's $\rho = 0.76$ and $\rho = 0.68$, respectively, $P < 0.001$ for both). The association supports the notion that there is a close relationship between speakers of Nilo-Saharan languages and the Central sorghum population. About 40% of the sorghums in this population are classified as caudatums or guinea-caudatums whereas the remaining 60% have not been morphologically classified. Because the Chari-Nile language group is obsolete according to current classification of the Nilo-Saharan language family (43) and because genetic relationship in sorghum is better assessed with molecular markers than morphology (9), we consider our results to support an updated version of Stemler et al.'s hypothesis: The geographic pattern of genetic population structure in sorghum and the distribution of the Nilo-Saharan language family coincide so closely that a causal relationship seems probable.

The distribution of the Southern population corresponds with the origin and spread of Bantu sorghum agriculture. The association between the race kafir and the Bantu agricultural tradition is a well-established theory (2, 10, 14, 29, 44). In this study, we found that 90% of the sorghums in the Southern population were from the Niger-Congo language family areas where Bantu languages are spoken, and 46% of them were classified as kafir. The Bantu expansion out of a homeland near the present-day Nigeria–Cameroon border is one of the archetypal examples cited in support of the farming-language codispersal hypothesis. Beginning about 4000 B.P., Bantu speakers expanded east and south to cover most of subequatorial Africa (18, 45, 46). Linguistic, human DNA, and archaeological evidence concur that the Bantu farmers picked up sorghum cultivation in the cultural melting pot in the western parts of East Africa's Rift Valley and Great Lakes region about 2000 B.P. and subsequently spread south, stopping only when they encountered the Mediterranean climate in the Cape Province to which their crops were not adapted (46). The genetic founder event is reflected in the relatively lower genetic diversity in the Southern population in terms of total number of alleles, number of private alleles, and expected heterozygosity (Table S2).

The Northern population is distributed throughout the Sahel and the Sudanian Savanna belt from Senegal to Sudan and extends into Eritrea, Ethiopia, and Somalia. The accessions in this population are from the areas with the highest average annual temperature and with large SD on average rainfall distribution. This population is the genetically most diverse of the three inferred populations, and it probably encompasses a broad range of adaptations to past and present climates in the fluctuating Sahel ("shoreline" in Arabic). The Northern population membership coefficients were significantly higher among accessions collected in Afro-Asiatic language family areas (Fig. S2C), and accessions from the Niger-Congo language-family area clustering with this population were mainly from West-Africa, north of the region of origin and expansion of Bantu sorghum cultivation.

Some accessions deviated from the general population-language family codistribution pattern (Table S2). Most of these deviations were found close to language-family border areas and can be explained by adoption of varieties between neighboring groups whereas the long-distance outliers probably were due to

translocation events taking place at different times than the dominating dispersal events. Dispersal of cultural elements through adoption is referred to as "acculturation" (19), an explanation that is complementary rather than mutually exclusive to demic diffusion, which is dominating according to the farming-language codispersal hypothesis (47).

Testing for the impact of other spatially distributed factors augmented the explanation for the detected structure. There was no significant isolation by distance (IBD) (48) effect between genetic relatedness and geographic distance, neither with or without control for spatial dependence with a partial Mantel test ($r = -0.31$, $P = 1$) (49, 50). The use of Mantel tests beyond testing for IBD is problematic in evolutionary biology due to spatial autocorrelation (51), but, despite the bias toward identifying false positives, we did not find significant associations between genetic relatedness and distance in the environmental variables annual mean temperature and annual precipitation ($r = 0.05$ and 0.006 , respectively, $P > 0.8$). This weak relationship between population structure and ecogeographic factors alone was supported by estimates of the fixation index (F_{ST}) of genetic differentiation among a priori defined groups. The differentiation among groups defined on the basis of temperature (<20 °C, 20 – 26 °C, >26 °C) ($F_{ST} = 0.07$, $P < 0.01$) and rainfall classes (<600 mm, 600 – $1,000$ mm, $>1,000$ mm) ($F_{ST} = 0.01$, $P < 0.01$) and among Bailey's ecoregions (52) at the division level (eight divisions) (Fig. S4A) ($F_{ST} = 0.06$, $P < 0.01$) was all relatively weak compared with the differentiation among the populations inferred with InStruct ($n = 143$ with $Q > 0.90$) ($F_{ST} = 0.21$, $P < 0.01$). Moreover, as in other continent-wide assessments (9, 12) we found that the differentiation among accessions belonging to the five discrete basic races ($n = 82$) was relatively weak ($F_{ST} = 0.13$, $P < 0.01$) compared with the differentiation among the inferred populations. Thus, social and cultural factors reflected in the distribution of language families appear to be the strongest structuring factors behind the continent-wide pattern. We do not suggest, on the basis of these results, that geographic distance, ecology, and morphology are unimportant factors for explaining the population structure of sorghum, but their impact seems to be contingent on social and cultural factors.

The Local Origin of Genetic Structure: The Case of the Pari and Nyithin Sorghum. The Pari community, whose sorghum we explored in-depth, lives in the Lafon villages located around a solitary rocky hilltop on the river Nile's flood plain in the southeastern part of South Sudan. The Pari language belongs to the Luo group of the Western-Nilotic branch of the Nilotic language group, together with larger language groups in South Sudan like the Dinka and Nuer languages. The Nilotic languages originated somewhere in the area between the White and the Blue Nile 6,000–7,000 y B.P., and Nilotic speakers have spread southwards in repeated expansions that have continued well into historic times (7). The majority of the family lineages in Lafon claim origin in an Anuak homeland to the north, but the Pari society harbors a mixture of different Western-Nilotic cultural elements (53). The Pari livelihood is characterized as a "multiple subsistence economy," and sorghum cultivation is an essential pillar, supplemented by husbandry, hunting, fishing, and collection of wild food (54, 55). The Pari society is organized in an age-grade system; young men are enrolled in groups based on age, and these age sets pass over different age grades over the course of their lives (56). Ehret (7) considers the age-grade system a particularly important factor for explaining the success of the historic Nilotic expansions. The age-grade system remains a fundamentally important institution with political, legal, military, ritual, and economic functions for the Pari (56), and it also plays an important role in the traditional seed system in Lafon (54). The ruling age grade, the *mojomiji*, decides when sowing shall commence, and all Pari lineages ritually mix a gourd bowl of seeds from a central granary with their

own seeds before sowing. This seed-management practice connects sorghum fields and granaries in Lafon in a metapopulation, and different morphological types, known as landraces, are managed as part of a landrace complex. The lowest level of classification is given to sorghum plants belonging to a visually distinct or otherwise characteristic landrace, and a higher-level classification distinguishes the Pari sorghum from other sorghums. The Pari sorghum is called “nyithin,” which means “the small one,” and informants from the Pari as well as from neighboring groups referred to nyithin as the sorghum that “came with the Pari.” The folk taxonomy of the nyithin landrace complex recognizes 12 named landraces, including one said to appear from time to time in farmers’ fields as a bad omen (i.e., *bendi-kirikik*). To assess the local genetic structure and temporal dynamics of the nyithin, we genotyped 20 seed lots collected from granaries and fields in 2010 and 2013, including 13 seed lots of nyithin collected in Lafon and nearby villages and 7 seed lots of other landraces collected among neighboring communities. Due to lack of sufficient plant descriptors and partly incompatibility of folk taxonomic classifications with scientific classification criteria, classification of the in situ seed lots below the nyithin category is limited to three folk taxonomic landrace names (i.e., “acar,” “adel,” and “lwalo”) from the 2010/2013 collection. The data on the ex situ accessions collected in 1983 that we used for comparison classify three accessions as nyithin and one as lwalo (Dataset S1).

Neighbor-joining (NJ) analysis supported the differentiation of three continent-wide clusters also when the in situ sampled seed lots were included (Fig. S5). The 1983 samples of the landraces from Lafon, conserved as genebank accessions at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), cluster firmly within the Central population of the continent-wide panel. The NJ analysis revealed one outlier

among the nyithin seed lots, which clusters within the Southern sorghum population together with two genebank accessions from Bari language areas and the “kabir” variety of the Acholi community (one accession collected in 1983 and one seed lot collected in 2013). The Acholi speak a Southern Luo dialect and have the main part of their homeland in Northern Uganda. The outlier nyithin seed lot probably represents a recent introduction and was not included in the following analyses.

Substructure analysis of the accessions clustering to the Central population, including the in situ sampled seed lots, identified five subpopulations (Fig. 2 and Table S3). Assigning accessions and seed lots to populations based on a >0.80 membership coefficient cutoff, we identified a North-Eastern subpopulation with some Afro-Asiatic affinity, a North-Western subpopulation including the accession from Cameroon, and three subpopulations with origin mainly in Nilotic language areas: Central-North, Central-South, and a latitude-wise more broadly distributed North-South subpopulation.

The substructure analysis assigned nearly all nyithins to the Central-North subpopulation with an average population membership coefficient of 0.89 (Dataset S1). The only nyithin not assigned to this subpopulation was a 1983 accession, which showed more affinity with the North-Eastern population, which also includes the accessions collected in the purported Pari homeland along the Sobat River. Population membership in the Central-North subpopulation was positively associated with origin in Western-Nilotic language areas ($\rho = 0.42$, $P < 0.001$), and the population membership was significantly different in Western-Nilotic areas compared with those in Eastern Nilotic and other language areas ($\chi^2 = 10.20$, $P < 0.01$). Two accessions from Dinka and Gbaya areas and two seed lots from Otuho (also known as Lotuko) areas also clustered in the Central-North subpopulation, suggesting a relationship with the nyithins. In the case of the two seed lots from Otuho areas, the relationship tallied with the information given by the farmers who donated them; the seeds were said to be nyithin originally acquired from Lafon in 1972. The farmers were able to precisely date this event because the seeds were acquired when the Addis Ababa accord ended the first Sudanese civil war and allowed them to return from exile. There was insignificant differentiation of these seed lots from those of Lafon ($F_{ST} = 0.10$, $P > 0.05$), showing that the nyithin was genetically characteristic after 40 y. This example of intercommunal adoption of landraces provides a local level example of the spread of crop varieties by acculturation found in language-family border zones in the continental scale analysis.

We explored temporal patterns of diversity to assess the resilience of the Pari seed system. The 30-y period between the collecting of landraces in Lafon has been characterized by environmental stress and human insecurity in the area. The average annual temperature in Lafon has shown an increasing trend through the period, and there has been large interannual variation both in rainfall and temperature (Fig. S4 B–E). In 1993, during the civil war in Sudan, all six Lafon villages were burned down, and the entire population (ca. 30,000) was displaced and resettled in several new settlements further away from the hill (57). This dramatic incidence came on top of occasional interethnic hostilities between the Pari and neighboring groups. With this environmental and historical backdrop, we assessed changes in genetic diversity over the period. Samples from the two temporal groups (1983 and 2010/2013 collecting) were intermixed in the NJ analysis and the genetic differentiation among the 1983 accessions ($F_{ST} = 0.59–0.85$, $P < 0.05$) and among the 2010/2013 seed lots ($F_{ST} = 0.07–0.60$, not all significant) were in most cases higher than the differentiation between the two groups ($F_{ST} = 0.17$, $P < 0.05$). In the case of the variety lwalo, for which passport data allowed us to compare a named variety of nyithin from 1983 and 2010, we found that it was insignificantly differentiated over the period as measured by F_{ST} at $P < 0.01$ level. Overall, the 2010/2013 group was more diverse than

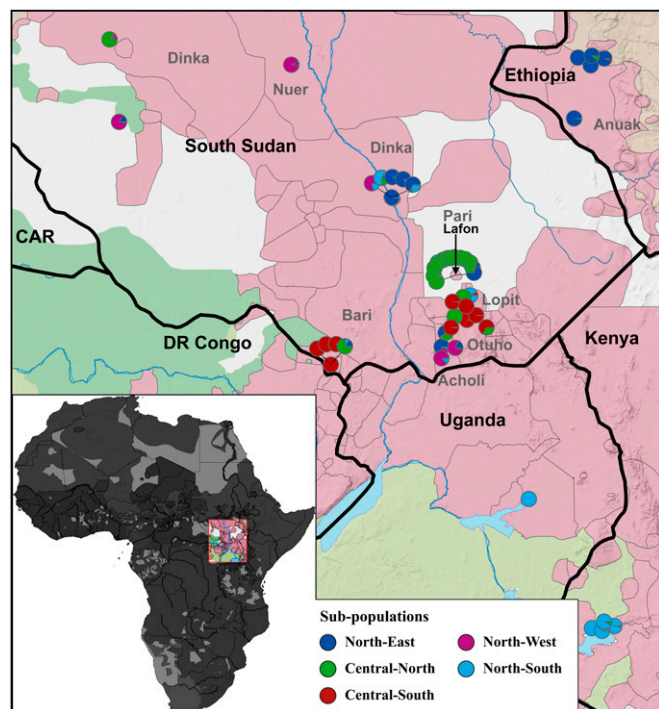


Fig. 2. Subpopulation structure from InStruct analysis of accessions belonging to the Central population (Fig. 1) and in situ sampled seed lots from South Sudan. Pie diagrams display inferred membership in the five subpopulations, and samples with origin in the same locations are displayed by fans. The location of the Lafon villages, the home of the Pari people, is indicated with an arrow (language distribution data from Ethnologue v. 16 compiled in ref. 62).

the 1983 group, with seven times more private alleles after using rarefaction to control for different sample sizes. Caution is important when inferring diachronic patterns of diversity based on comparison of ex situ conserved accessions and in situ sampled seed lots. An obvious caveat in our case is that, during both the 1983 collecting and the 2010/2013 collecting of nyithin, we were unable to sample the full range of landraces included in the nyithin landrace complex in the Pari folk taxonomy. Furthermore, ex situ conserved accessions have undergone purification and forced selfing during regeneration (9), resulting in reduced heterozygosity compared with the original landrace. Thus, we limit our inference of the detected patterns to say that there is no evidence for loss of genetic diversity within nyithin. Similar conclusions were drawn in studies of pearl millet and sorghum landrace diversity in Niger (11, 58). A molecular comparison of sorghum landraces, sampled in the same villages separated by 26 y, marked by major social and environmental change, found no evidence of genetic erosion (11). Whether or not nyithin and other landraces and affiliated seed systems harbor the necessary diversity to adapt to projected climate change (32) (Fig. S4F) remains an open question, but our findings suggest that the traditional seed system of the Pari is remarkably resilient.

Conclusion

The farming-language codispersal hypothesis is commonly framed as a proposed correlation between the spread of human genes and languages out of an agricultural homeland (18). The association between languages and genetic structure in sorghum presented in this study provides crop evolutionary evidence for such codispersal. Our findings suggest that languages and sorghum seeds have moved together both in the case of Bantu and Nilo-Saharan expansions. We propose that this codispersal ultimately is adaptive; drought-hardy sorghum varieties have probably played a key role in allowing their cultivators to expand into new areas during periods of climate change in the past. Thus, the relationship between ethnolinguistic groups and sorghum population structure reflects the dispersal of successful social-ecological adaptations.

The study of the Pari and their nyithin sorghums provides a local-level window into the kinds of mechanisms that have generated and maintained the genetic patterns we see at the level of language families. The seed system of the Pari is governed by the ruling age grade, a social institution with deep Nilotic cultural roots. The Pari sorghums are sufficiently genetically distinct to be recognized as a separate population from the sorghum of neighboring groups and at the same time sufficiently heterogeneous to harbor a number of different genetically characteristic landraces. Maintenance of this two-tiered diversity pattern is possible due to the combination of a low, but pertinent, outcrossing rate in sorghum and the seed-mixing practice in the traditional seed system. Despite relocation of the entire Pari population and dramatic disturbance of normal livelihood activities during the civil war, the traditional seed system has prevailed, and we do not find evidence of significant genetic erosion.

The low adoption of improved sorghum varieties in Africa is explained by institutional shortcomings on the supply side, as well as by demand-side factors related to large differences in agroecological constraints and local end-product preferences (59). Thus, reluctance to adopt modern varieties is the flip side of the kind of social-ecological adaptation described in the case of the Pari seed system. Despite lower achievable yields, communities relying on subsistence agriculture often choose to continue to cultivate local varieties because yields are stable and predictable and consumption characteristics are well known (25, 27). Insights from the literature on seed systems elucidate this two-sided reason for the low adoption. On the one hand, informal seed systems are important local safety nets (22, 26, 33); on the other hand, introduction of modern varieties in an ad hoc

manner can increase vulnerability in the affected communities (35). These insights provide important background for considering the sustainability of development assistance initiatives aimed at modernizing the seed sector and introducing new sorghum varieties in Africa. Efforts to build resilience to current and future environmental change require understanding of the social and cultural context and are more likely to be successful when building on, rather than seeking to replace, existing traditional seed systems and landraces.

Materials and Methods

Collecting. In 1983, a germplasm-collecting mission led by the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) visited Lafon and collected sorghum seeds (60). Equipped with the collecting report from 1983 and the contribution in this study by one of the participants in the 1983 mission (T.B.), we revisited Lafon and neighboring villages in November 2010 and January 2013 and resampled seed lots of nyithin. We obtained permits to conduct the study as well as export permits for seeds and silica-dried material from the Ministry of Agriculture and Forestry in South Sudan. In Lafon, we were permitted to carry out the collecting by the ruling age grade, and seed lots (seeds of a named landrace sampled from one field or one granary) were sampled under the supervision of the owners. We acquired information about the sampled landraces and the local seed system in interviews with individual farmers and members of the ruling age grade.

Plant Material. The panel used to assess continent-wide genetic structure consisted of 200 accessions sourced from the major global sorghum collection at ICRISAT, including 138 African and 4 Indian accessions from the minicore collection (36). The panel used to assess local genetic structure and temporal dynamics partially overlaps with the continent-wide dataset in addition to 20 seed lots sampled in situ. All together, 358 plants from 220 accessions and seed lots were genotyped (Dataset S1). As in other large-scale assessments of sorghum genetic diversity (9, 11, 12), we included one individual per accession to infer population structure, and, as in ref. 8, we included more than one individual per seed lot for local-scale calculation of diversity indices, F_{ST} among landraces, and NJ analysis.

Genotyping. DNA was extracted from 15- to 30-mg dried-leaf samples using an E.Z.N.A. plant DNA Mini Kit (Omega Bio-tek), according to the manufacturer's protocol. Genotyping was done with 19 simple sequence repeat (SSR) markers (Table S1). The markers were amplified using a M13 tailing approach (61) and were separated with capillary electrophoresis on an ABI 3730 sequencer (Applied Biosystems). For further details, see *SI Materials and Methods*.

Data Analysis. The language distributions in Fig. 1A are based on spatial data from Ethnologue v. 16 (www.ethnologue.com/) compiled in ref. 62. Climate data (average data for ~1950–2000 in 2.5 arc minutes resolution) were obtained from ref. 63, and ecoregion division was obtained from ref. 52. Scoring of genotypes was done in GeneMapper v. 3.7 (Applied Biosystems). We used two complimentary Bayesian model-based programs to estimate population structure: STRUCTURE v. 2.3.3 (38, 39) and InStruct (37). Whereas STRUCTURE assumes random mating within subpopulations, InStruct calculates expected genotype frequencies on the basis of selfing rates and is particularly suitable for predominantly inbreeding organisms such as sorghum. We ran InStruct with 1×10^5 burn-in, 2×10^5 iteration steps, and a thinning interval of 10 steps, assuming different starting points (37). We ran STRUCTURE with 10 independent runs for each value of K from 1 to 9, with a burn-in period of 5×10^5 followed by 10^6 iterations. The most probable number of groups, K , was determined with STRUCTURE HARVESTER (40), calculating the ad hoc measure in change in likelihood between successive K values, ΔK (41). We correlated structure membership coefficients obtained for the same accessions generated by InStruct and STRUCTURE and between STRUCTURE generated memberships inferred on the basis of the 13,390 SNPs from ref. 13 and our 19 SSRs for 139 common accessions using R (www.r-project.org/). Spatially explicit Bayesian modeling was done on the 138 genbank accessions for which latitude and longitude data were available using the program GENELAND (42). We ran 10^5 Markov chain Monte Carlo iterations with a thinning interval of 100 using the uncorrelated frequency model as a prior for the allele frequencies. The posterior probabilities of population membership for each pixel were computed using a burn-in of 200, and the spatial domain was set to 500 times 500. To test for IBD, we tested the association between a matrix of individual kinship coefficients, estimated according to ref. 64 with the program SPAGED1 (65), with geographic distance, using the R package

Vegan to conduct both simple Mantel test and partial Mantel test controlling for spatial dependence according to ref. 50. Genetic differentiation between a priori defined ecogeographic and morphological groups was calculated by using Weir and Cockerham's F_{ST} estimator (66) in SPAGEDI.

ACKNOWLEDGMENTS. We thank the Pari for their willingness to share information about their seed system and for allowing us to sample their sorghums. We thank the officials in the Ministry of Agriculture and Forestry in South Sudan for permits to carry out the current study and for assistance

during fieldwork. We thank the anthropologist Eisei Kurimoto for sharing insights and literature about the Pari. We thank Sandeep Sukumaran, Julian Ramirez-Villegas, and Dag Terje Endresen for assistance with access and visualization of climate and ecogeographic data. We thank the two anonymous referees for very valuable comments. This work was supported by a grant from the University of Oslo, and fieldwork was supported by the Norwegian University Cooperation Program for Capacity Development on Postwar Livelihood and Environment Studies, managed by the Norwegian University of Life Sciences and the University of Juba.

1. National Research Council (1996) *Grains, Lost Crops of Africa* (National Academies Press, Washington, DC), Vol 1, p 383.
2. Doggett H (1988) *Sorghum* (Longman Scientific and Technical, New York).
3. Ehret C (2002) *The Civilizations of Africa: A History to 1800* (James Currey, Oxford).
4. Bellwood P (2013) *First Migrants. Ancient Migration in Global Perspective* (Wiley Blackwell, West Sussex, UK).
5. Wendorf F, Schild R (1998) Nabta Playa and its role in northeastern African prehistory. *J Anthropol Archaeol* 17(2):97–123.
6. Wendorf F, et al. (1992) Saharan exploitation of plants 8,000 years BP. *Nature* 359(6397):721–724.
7. Ehret C (2002) Language family expansion: Broadening our understandings of cause from an African perspective. *Examining the Farming/Language Dispersal Hypothesis*, eds Bellwood P, Renfrew AC (McDonald Institute for Archaeological Research, University of Cambridge, Cambridge, UK), pp 163–176.
8. Barnaud A, Deu M, Garine E, McKey D, Joly HI (2007) Local genetic diversity of sorghum in a village in northern Cameroon: Structure and dynamics of landraces. *Theor Appl Genet* 114(2):237–248.
9. Billot C, et al. (2013) Massive sorghum collection genotyped with SSR markers to enhance use of global genetic resources. *PLoS ONE* 8(4):e59714.
10. Deu M, Rattunde F, Chanterreau J (2006) A global view of genetic diversity in cultivated sorghums using a core collection. *Genome* 49(2):168–180.
11. Deu M, et al. (2010) Spatio-temporal dynamics of genetic diversity in *Sorghum bicolor* in Niger. *Theor Appl Genet* 120(7):1301–1313.
12. Morris GP, et al. (2013) Population genomic and genome-wide association studies of agronomic traits in sorghum. *Proc Natl Acad Sci USA* 110(2):453–458.
13. Wang Y-H, et al. (2013) Genetic structure and linkage disequilibrium in a diverse, representative collection of the C4 model plant, *Sorghum bicolor*. *G3* 3(5):783–793.
14. De Wet J, Huckabay J (1967) The origin of *Sorghum bicolor*. II. Distribution and domestication. *Evolution* 21(4):787–802.
15. Harlan J, Stemler A (1976) The races of sorghum in Africa. *Origins of African Plant Domestication*, eds Harlan J, de Wet JM, Stemler AB (Mouton Publishers, The Hague).
16. Harlan JR, de Wet MJM (1972) A simplified classification of cultivated sorghum. *Crop Sci* 12(2):172–176.
17. Stemler AB, Harlan JR, de Wet JM (1975) Caudatum sorghums and speakers of Char-Nile languages in Africa. *J Afr Hist* 16(2):161–183.
18. Diamond J, Bellwood P (2003) Farmers and their languages: The first expansions. *Science* 300(5619):597–603.
19. Jobling MA, Hollar E, Hurler M, Kivisild T, Tyler-Smith C (2013) *Human Evolutionary Genetics* (Garland Science, New York), 2nd Ed.
20. Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. *Nature* 457(7231):843–848.
21. Larson G, et al. (2014) Current perspectives and the future of domestication studies. *Proc Natl Acad Sci USA* 111(17):6139–6146.
22. Bellon M, Brush S (1994) Keepers of maize in Chiapas, Mexico. *Econ Bot* 48(2):196–209.
23. Louette D, Charrier A, Berthaud J (1997) In situ conservation of maize in Mexico: Genetic diversity and maize seed management in a traditional community. *Econ Bot* 51(1):20–38.
24. Mekbib F (2008) Farmers' breeding of sorghum in the center of diversity, Ethiopia. I. Socioecotype differentiation, varietal mixture and selection efficiency. *J New Seeds* 9(1):43–67.
25. Almekinders CJM, Louwaars NP, Debruijn GH (1994) Local seed systems and their importance for an improved seed supply in developing countries. *Euphytica* 78(3):207–216.
26. Jarvis DI, 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(14):5326–5331.
27. Perales HR, Benz BF, Brush SB (2005) Maize diversity and ethnolinguistic diversity in Chiapas, Mexico. *Proc Natl Acad Sci USA* 102(3):949–954.
28. Labeyrie V, et al. (2014) Influence of ethnolinguistic diversity on the sorghum genetic patterns in subsistence farming systems in eastern Kenya. *PLoS ONE* 9(3):e92178.
29. Leclerc C, Coppens d'Eeckenbrugge G (2011) Social organization of crop genetic diversity: The $G \times E \times S$ interaction model. *Diversity* 4(1):1–32.
30. Delêtre M, McKey DB, Hodkinson TR (2011) Marriage exchanges, seed exchanges, and the dynamics of manioc diversity. *Proc Natl Acad Sci USA* 108(45):18249–18254.
31. Schlenker W, Lobell DB (2010) Robust negative impacts of climate change on African agriculture. *Environ Res Lett* 5(1):014010.
32. Ramirez-Villegas J, Jarvis A, Läderach P (2013) Empirical approaches for assessing impacts of climate change on agriculture: The EcoCrop model and a case study with grain sorghum. *Agric For Meteorol* 170:67–78.
33. Bellon MR, Hodson D, Hellin J (2011) Assessing the vulnerability of traditional maize seed systems in Mexico to climate change. *Proc Natl Acad Sci USA* 108(33):13432–13437.
34. Burke MB, Lobell DB, Guarino L (2009) Shifts in African crop climates by 2050, and the implications for crop improvement and genetic resources conservation. *Glob Environ Chang Hum Policy Dimens* 19(3):317–325.
35. McGuire S, Sperling L (2013) Making seed systems more resilient to stress. *Glob Environ Chang Hum Policy Dimens* 23(3):644–653.
36. Upadhyaya HD, et al. (2009) Developing a mini core collection of sorghum for diversified utilization of germplasm. *Crop Sci* 49(5):1769–1780.
37. Gao H, Williamson S, Bustamante CD (2007) A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics* 176(3):1635–1651.
38. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959.
39. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164(4):1567–1587.
40. Earl DA (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4(2):359–361.
41. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol Ecol* 14(8):2611–2620.
42. Guillot G, Mortier F, Estoup A (2005) GENELAND: A computer package for landscape genetics. *Mol Ecol Notes* 5(3):712–715.
43. Dimmendaal GJ, Goodman MJ (2013) Entry for “Nilo-Saharan languages.” *Encyclopaedia Britannica*. Available at Britannica.com.
44. Harlan JR (1976) *Origins of African Plant Domestication* (Walter de Gruyter, Berlin).
45. Currie TE, Meade A, Guillon M, Mace R (2013) Cultural phylogeography of the Bantu Languages of sub-Saharan Africa. *Proc R Soc B: Biol Sci* 280(1762).
46. Ehret C (1998) *An African Classical Age: Eastern and Southern Africa in World History, 1000 BC to AD 400* (Univ of Virginia Press, Charlottesville).
47. Gepts P, Famula TR, Bettinger RL, Brush SB, Damania AB (2012) *Biodiversity in Agriculture: Domestication, Evolution, and Sustainability* (Cambridge Univ Press, Cambridge, UK).
48. Wright S (1943) Isolation by distance. *Genetics* 28(2):114–138.
49. Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* 27(2):209–220.
50. Meirmans PG (2012) The trouble with isolation by distance. *Mol Ecol* 21(12):2839–2846.
51. Guillot G, Rousset F (2013) Dismantling the Mantel tests. *Methods Ecol Evol* 4(4):336–344.
52. Bailey RG, Hogg HC (1986) A world ecoregions map for resource reporting. *Environ Conserv* 13(03):195–202.
53. Kurimoto E (1996) People of the river: Subsistence economy of the Anywaa (Anuak) of Western Ethiopia. *Essays in Northeast African Studies*, Senri Ethnological Studies, eds Sato S, Kurimoto E (National Museum of Ethnology, Osaka, Japan), No. 43, pp 29–58.
54. Kurimoto E (1984) Agriculture in the multiple subsistence economy of the Pari. *Agriculture and Land Utilization in the Eastern Zaire and the Southern Sudan*, ed Sakamoto K (Kyoto University, Kyoto, Japan), pp 23–52.
55. Takei E (1984) Variation and geographical distribution of cultivated plants in the Southern Sudan. *Agriculture and Land Utilization in the Eastern Zaire and the Southern Sudan*, ed Sakamoto K (Kyoto University, Kyoto, Japan), pp 53–76.
56. Kurimoto E (1995) *Coping with Enemies: Graded Age System Among the Pari of Southeastern Sudan* (Osaka University, Osaka, Japan), pp 261–311.
57. Kurimoto E (2007) *Report of the Field Research in Lafon, Eastern Equatoria State: Assessment of the General Conditions and Livelihood of the Pari People* (Osaka University, Osaka, Japan).
58. Bezañon G, 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(2):223–236.
59. Ejeta G (2010) African Green Revolution needn't be a mirage. *Science* 327(5967):831–832.
60. Mengesha MH (1983) *Southern Sudan Trip Report* (ICRISAT, Patancheru, India).
61. Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nat Biotechnol* 18(2):233–234.
62. GMI (2013) *World Language Mapping System: Ethnologue* (Global Mapping International, Colorado Springs, CO), Version 16.0.
63. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25(15):1965–1978.
64. Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understorey shrub, *Psychotria officinalis* (Rubiaceae). *Am J Bot* 82(11):1420–1425.
65. Hardy OJ, Vekemans X (2002) SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* 2(4):618–620.
66. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38(6):1358–1370.