

INVITED REVIEW ARTICLE

The genomic prehistory of peoples speaking Khoisan languages

Brigitte Pakendorf^{1,*} and Mark Stoneking^{2,*}¹Dynamique du Langage, UMR5596, CNRS & Université de Lyon, 14 avenue Berthelot, 69007 Lyon, France and²Department of Evolutionary Genetics, MPI for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany

*To whom correspondence should be addressed. Brigitte Pakendorf, Tel: +33 472726426; Email: brigitte.pakendorf@cnrs.fr; Mark Stoneking, Tel: +49 3413550502; Email: stoneking@eva.mpg.de

Abstract

Peoples speaking so-called Khoisan languages—that is, indigenous languages of southern Africa that do not belong to the Bantu family—are culturally and linguistically diverse. They comprise herders, hunter-gatherers as well as groups of mixed modes of subsistence, and their languages are classified into three distinct language families. This cultural and linguistic variation is mirrored by extensive genetic diversity. We here review the recent genomics literature and discuss the genetic evidence for a formerly wider geographic spread of peoples with Khoisan-related ancestry, for the deep divergence among populations speaking Khoisan languages overlaid by more recent gene flow among these groups and for the impact of admixture with immigrant food-producers in their prehistory.

Introduction

In this paper, we use the term ‘Khoisan’ as a loose cover term to refer to the indigenous languages of southern Africa that do not belong to the Bantu family—and that are most saliently characterized by their heavy use of click consonants—as well as by extension to the genetic ancestry associated with the peoples who speak these languages. The term was coined by the biological anthropologist Schulze in 1928 by combining the Khoekhoe herders’ term for themselves with their term for foragers (1); variations encountered in the literature are Khoe-San and KhoeSan. Given the fact that the peoples speaking Khoisan languages are culturally and linguistically distinct and each has their own particular history, all umbrella terms are flawed; it is thus of crucial importance to keep in mind that use of a single term does not signify a unified entity.

Nowadays, peoples speaking Khoisan languages live mainly in Namibia, Botswana and Angola, with smaller numbers in Zambia, Zimbabwe, South Africa, Lesotho and Eswatini (2) (Fig. 1). In South Africa, most historically known Khoisan groups have given up their original languages and have merged with so-called Coloured populations (3). Two groups often included in recent genetic studies are the Karretjie people, itinerant sheepshearers of the Karoo who are probably partly descendants of !Xam hunter-gatherers (4), and the †Khomani. This latter is a group of people with diverse Khoisan-related ancestries tracing back to the southern Kalahari, who joined together in the 1990s in order to file a land rights claim (5).

Although an initial broad classification of the languages of Africa identified a single Khoisan phylum comprising three branches in southern Africa plus two languages—Sandawe and Hadza—spoken in East Africa (6), nowadays specialists of these

Received: August 25, 2020. Revised: September 28, 2020. Accepted: October 8, 2020

© The Author(s) 2020. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

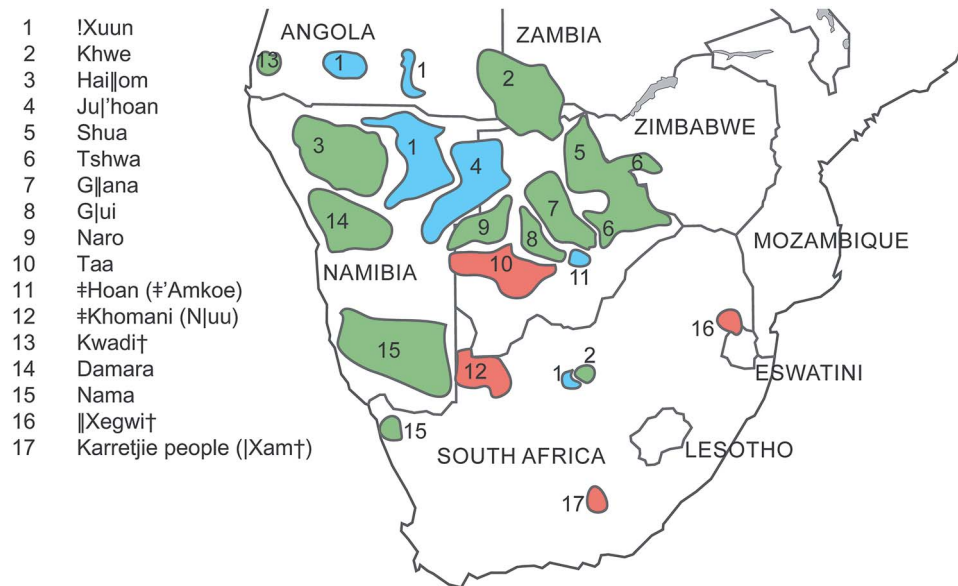


Figure 1. Map showing the approximate location of Khoisan-speaking groups, based on ethnolinguistic data from Güldemann (1) and Hitchcock (2), and information on the Karretjie from Schlebusch et al. (52). Colours indicate the language family affiliation: blue, Kx'a; red, Tuu; green, Khoe-Kwadi. Languages that are extinct are indicated by crosses. Some of the ǀKhomani still remember N|uu. ǀAmkoe is the actual name of the language spoken by the ǀHoan, but as the initial publication presenting genetic data from this group referred to them by the old language name, this has been maintained in genetic publications.

languages agree that there are three distinct language families in southern Africa, namely Kx'a, Tuu and Khoe-Kwadi (1). Of these, Kx'a and Tuu might ultimately descend from a shared ancestor, but that has not yet been conclusively demonstrated (7). The Khoe-Kwadi languages are not related to either the Kx'a or the Tuu languages (1,8). As for the East African languages, although there is no demonstrable relationship between Hadza and any of the southern African Khoisan languages, there is some indication that Sandawe might be related to the Khoe-Kwadi family; however, this, too, needs further corroboration (7).

Culturally, too, there is considerable heterogeneity among the Khoisan-speaking peoples of southern Africa (9): herding groups are known historically from coastal and interior regions of the Cape, the descendants of whom are the Nama (nowadays settled mainly in southern Namibia) as well as several Coloured groups in South Africa (9,10). Furthermore, the Kwepe, small-stock pastoralists from southwestern Angola, are known to have spoken Kwadi, a language of the Khoe-Kwadi family, although this is nowadays practically extinct (9,11). Hunter-gatherers roamed the Cape interior of South Africa in historic times and are still found in the Kalahari region spanning Namibia, Botswana and parts of South Africa. But there are also groups that do not neatly fit into the herder-forager dichotomy. Foremost among these are the Damara, a peripatetic group (12) who traditionally practiced foraging, small-scale herding of goats and blacksmithing in a client relationship to the Nama and the Bantu-speaking pastoralist Herero. Along the Kavango River, the Khwe rely on fishing as well as hunting and gathering, whereas in the eastern Kalahari, the Shua and Tshwa are transitioning to food production and are in a client relationship to their Bantu neighbours in addition to their foraging subsistence.

Given this linguistic and cultural diversity, it is clear that the prehistory of the peoples speaking Khoisan languages must have been highly complex. Numerous studies over the past decade have highlighted the considerable genetic diversity found in these groups, the deep divergence among them as

well as between them and other African groups, and the impact of successive waves of migration of food-producing peoples from East and West-Central Africa as well as in historical times of European colonizers [see (13) for a recent review]. We here survey the recent literature and discuss the genetic evidence for an erstwhile wider geographic spread of peoples with Khoisan-related ancestry, for the deep divergence among populations speaking Khoisan languages overlaid by more recent gene flow among these groups and for the impact of admixture with immigrant food-producers in their prehistory. For convenience, we will refer to peoples speaking a Khoisan language as a Khoisan-speaking group and to peoples speaking a Bantu language as a Bantu-speaking group (although of course each group speaks a particular language belonging to the Kx'a, Tuu or Khoe-Kwadi families for Khoisan-speaking groups, or to the Bantu family). Throughout the paper, we follow the nomenclature of Güldemann (1), irrespective of the spelling of group names found in individual articles.

A Wider Geographic Spread in Prehistoric Times

Recent genome-wide analyses of DNA from ancient human remains in East Africa have demonstrated the presence in the past of Khoisan-related ancestry in regions as distant from the Kalahari as Tanzania and Kenya (throughout this review, for convenience, we use present-day countries to refer to the location of ancient remains that predate country formation). Thus, ~60% of the ancestry of ancient remains from Malawi dated to between 2500 and 8100 BP and ~30% of the ancestry of a 1400-year-old individual from Tanzania is related to ancestry detectable both in 2000-year-old hunter-gatherer remains from South Africa and modern Ju|'hoan (14). Similarly, an ancient individual from Kenya dated to 3500 BP shows evidence of low levels of Khoisan-related ancestry (15). In addition, there is evidence from whole genome sequences from modern

Table 1. Salient features of recent studies of whole genome sequences that included Khoisan-speaking groups

Study	Sample sizes and groups	Deepest divergence time between Khoisan-speaking and other groups	Divergence time among Khoisan-speaking groups
Fan <i>et al.</i> (19)	Four Ju'hoan ^a Two †Khomani ^a	~200 kya	~30 kya
Lorente-Galdos <i>et al.</i> (20)	Two Ju'hoan ^b One Taa ^c One †Khomani	~190 kya	n/a
Bergström <i>et al.</i> (21)	Six Ju'hoan ^d	~162 kya	n/a
Schlebusch <i>et al.</i> (16)	Five Ju'hoan Five †Khomani Five Nama Five Karretjie Five G ui/G ana ^e	~200–300 kya	~160–190 kya

^aSequences from Mallick *et al.* (53); Ju'hoan samples from the Human Genome Diversity Panel (HGDP).

^bOne sequence from Mallick *et al.* (53) and one sequence from Meyer *et al.* (54); samples from HGDP.

^cSequence from Schuster *et al.* (55), where this individual (KB1) is labelled a Tuu speaker.

^dAll samples from HGDP and include the four Ju'hoan from Mallick *et al.* (53).

^eMixed group of G|ui and G|ana individuals, see Schlebusch *et al.* (30) for details.

populations for potential long-distance migrations involving Khoisan-speaking groups, e.g. in the sharing of private alleles between the Ju'hoan and Mbuti central African rain forest foragers (16).

Interestingly, the Khoisan-related ancestry in eastern Africa is related in equal degrees to the deeply diverging lineages identified in modern-day Khoisan-speaking populations see “High Levels of Genetic Diversity in Khoisan-Speaking Peoples”, implying that the Khoisan-related groups settled in eastern Africa were genetically distinct from those living in southern Africa (14). These results mirror the results of mitochondrial DNA (mtDNA) analyses that found a complementary distribution of one of the Khoisan-specific haplogroups, L0k. Of three deeply divergent branches (L0k1a, L0k1b and L0k2), only L0k1a is found among extant Khoisan-speaking groups of Namibia and Botswana, whereas L0k1b and L0k2 are found practically exclusively in Bantu-speaking populations settled in Zambia. This implies that people genetically related to currently known Khoisan-speaking groups, yet carrying distinct lineages, were resident in regions beyond those previously attested (17). There are no historically known Khoisan-speaking groups in either Zambia or Malawi or further northeast, and modern-day populations of Malawi show no traces of Khoisan-related ancestry. It is thus clear that the incoming Bantu-speaking populations must have replaced the Khoisan-related autochthonous populations with hardly any admixture. Linguistic analyses, too, show that some Bantu languages of the Kavango–Zambezi transfrontier area borrowed words with click consonants from Khoisan languages that are nowadays extinct, in addition to borrowing words from Khwe and Ju languages (18).

High Levels of Genetic Diversity in Khoisan-Speaking Peoples

Khoisan-speaking groups are consistently found to harbour high levels of genetic diversity. Several recent studies of whole genome sequences found highest levels of genetic diversity in Khoisan-speaking individuals (16,19–21), and these individuals also have the highest frequencies, on average, of population-specific copy number variants worldwide (22). However, sample sizes and ethnolinguistic diversity of the groups analyzed remain quite limited (Table 1); there is a clear need for additional

whole genome sequence studies of further Khoisan-speaking groups.

Khoisan-speaking groups are also the first to branch off in genomic studies of African or world-wide populations (16,19,20,23), with their divergence from other populations dated to 160–300 kya (Table 1). The fact that Khoisan-related lineages are the first to diverge has sometimes been erroneously interpreted as strong evidence for an origin of modern humans in southern Africa [(24), which is based solely on mtDNA lineages; see (25,26) for substantial critiques of this paper]. However, as noted above, Khoisan-related groups were formerly more widespread, and moreover, the divergence between Khoisan-speaking groups and other African groups could, in principle, have occurred anywhere in Africa.

Khoisan-speaking peoples also show evidence of a larger effective population size over time than other African populations (16,19,20,27). All human populations show a signal of decreasing effective population size beginning around the time of the divergence of African from non-African populations, ~50–100 kya; however, Khoisan-speaking groups show less of a reduction in effective population size than do other populations (16,19,20). Some of this diversity might be due to archaic admixture from an as yet undiscovered population (28). For example, whole genome sequencing (20) suggests ~4% ancestry from an archaic ‘ghost’ population in the four Khoisan-speaking individuals analyzed.

In addition to carrying considerable amounts of genetic diversity, Khoisan-speaking populations are also quite diverged from one another, as shown by a deep split between populations residing in the northwestern Kalahari and those from the southeastern Kalahari or South Africa (29,30), respectively. Recent reanalyses of these data together with some new data (31)—providing the most complete geographical coverage of extant Khoisan-speaking groups—that focus solely on genomic segments of Khoisan-related ancestry have demonstrated a tripartite split into northern, central and southern Khoisan-speaking groups [roughly corresponding to Pickrell *et al.*'s (29) northwestern and southwestern and Schlebusch *et al.*'s (30) southern groups, respectively, and also corresponding to groups defined by ecogeographic boundaries in (32); Fig. 2]. It should be noted that these northern, central and southern genetic groupings do not correspond to a previous linguistic

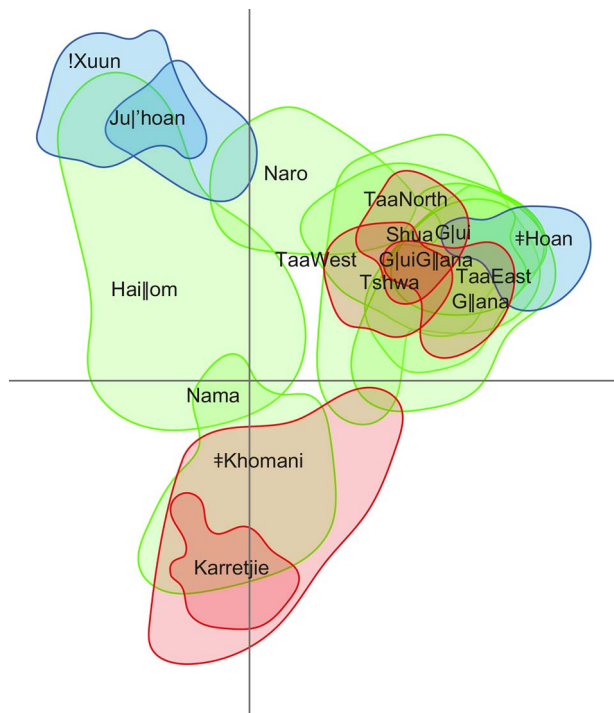


Figure 2. Plot of the first two dimensions of a multidimensional scaling analysis of Khoisan-speaking groups, illustrating the northern, central and southern genetic groupings of Khoisan-speaking groups. The plot is based on genome-wide SNP array data with non-Khoisan-related ancestry masked. The vertical axis is the first dimension and the horizontal axis is the second dimension. The contours depict 90% utilization distribution densities, i.e. the smallest area of the plot in which there is a 90% probability of locating the individuals from the same group, and are colour coded according to language family: blue, Kx'a; red, Tuu; green, Khoe. Modified from Montinaro et al. (31), which should be consulted for further details.

classification of southern African Khoisan languages into Northern (Khoisan), Central (Khoisan), Southern (Khoisan) branches (6): the northwestern/northern grouping includes not only the !Xuun and Ju|'hoan, whose languages belong to the Kx'a family, but also the Hai||om, whose language belongs to the Khoe family; the southeastern/central grouping includes populations speaking languages belonging to all three families, and the southern grouping includes both the pastoralist Nama, whose language belongs to the Khoe family, and the descendants of foragers, the Karretjie and #Khomani, whose heritage languages belonged to the Tuu family.

However, there is uncertainty as to when this deep split occurred. Initial studies based on genome-wide single-nucleotide polymorphism (SNP) array data dated the divergence to ~25–35 kya (29–31). The inclusion of genome sequence data from a 2000-year-old hunter-gatherer individual from South Africa pushed back the date of divergence between 'northern' and 'southern' Khoisan to 156–185 kya (23), which might suggest that the SNP array data underestimate divergence times due to ascertainment bias. Nevertheless, subsequent estimates based on whole genome sequences range from ~30 to ~160 kya (Table 1); different mutation rates can account for some differences in these estimates, but not all, leaving this an open question.

In any event, while this deep divergence between northern, central and southern Khoisan-speaking populations suggests

that they must have been isolated from each other for a considerable period of time, there is also evidence for gene flow among Khoisan-speaking groups taking place at a more recent time-scale. This is shown by analyses focussing on genome segments of Khoisan-specific ancestry that show a high correlation of genetic with geographic distances, and a clear signal of isolation by distance (31,33). It is therefore possible that the deep divergence times arise purely as a consequence of long-distance separation in what is actually a gradient of relatedness. However, it is also possible that the signals of isolation by distance reflect more recent processes after initial older divergence events. In particular, it has been suggested that Khoisan-speaking groups were initially split by the prehistoric lake Makgadigadi, with gene flow being reinitiated when the lake dried up around 10 kya (34). One group in particular that shows evidence for admixture are the Naro, who are both geographically (Fig. 1) and genetically (Fig. 2) intermediate between the northwestern/northern and southeastern/central groupings (29,31) and who also show evidence for gene flow from the G|ui and an ethnolinguistically undefined group from Xade in the Central Kalahari Game Reserve (33). In addition, the #Hoan, who speak a divergent language of the Kx'a family nowadays called †'Amkoe, show only 5% shared ancestry with their linguistic relatives the !Xuun and the Ju|'hoan (33), whereas they are genetically close to the neighbouring Taa (who speak a Tuu language) and the G|ui, whose language belongs to the Khoe family (31) (cf. 34,35). Distinguishing between long-term isolation by distance, versus deep divergence followed by more recent contact, may be possible when more whole genome sequence data become available.

Admixture with Immigrating Food-Producing Populations

In addition to gene flow among Khoisan-speaking groups, these have also undergone variable amounts of admixture from immigrating food-producers (23,36) (Fig. 3). Sheep and goat pastoralists are thought to have immigrated to southern Africa from East Africa a few centuries before the immigration of Iron Age agropastoralists commonly associated with the expansion of Bantu-speaking peoples into large parts of sub-Saharan Africa (37). The presence among southern African populations of the Lactase Persistence variant C-14010 (30,38,39), which is of probable East African origin (40), points towards a demic diffusion of pastoralism into southern Africa. Significantly higher frequencies of this variant in pastoralist populations than in foragers, and in groups speaking languages of the Khoe family than in Tuu- or Kx'a-speaking populations (39), support the hypothesis that the Khoe–Kwadi languages were brought to southern Africa by a migration of pastoralists from East Africa (8). Unexpectedly, however, the formerly Khoe–Kwadi-speaking pastoralist Kwepe from southwestern Angola have only low frequencies of this allele (41), in accordance with their low frequency of the East African Y-chromosome haplogroup E-M293. The spread of pastoralism and the Khoe–Kwadi languages is therefore likely to be a complex process, which might also have involved shift of Bantu-speaking groups to Khoe–Kwadi languages (42). Interestingly, two modern-day forager groups, the G|ui and the Tshwa, show evidence for ongoing positive selection for the C-14010 allele, indicating a possibly recent reversion from a herding way of life to foraging (39). Ancient DNA analyses have provided further direct evidence for admixture with East African pastoralists: a 1200-year-old specimen found in a herder context in the western Cape was shown to have ~40% ancestry related

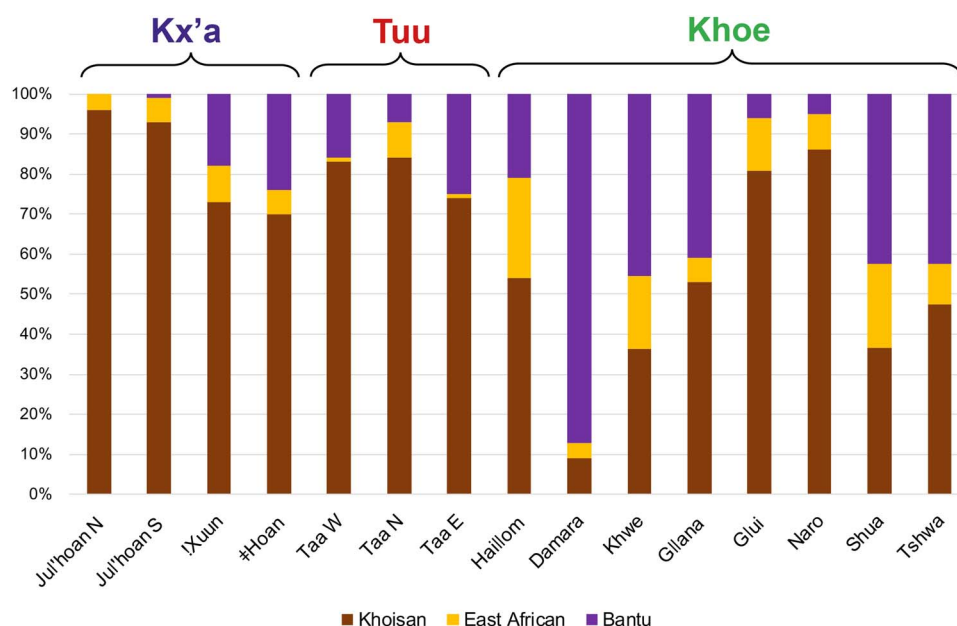


Figure 3. Variable amounts of Khoisan-related, Bantu-related and East African-related ancestries in Khoisan-speaking groups, based on Pickrell et al. (36).

to an early pastoralist from Tanzania and ~60% ancestry related to 2000-year-old South African foragers (14). Two Early Iron Age individuals from Botswana—who are likely to have spoken Bantu languages—confirm the earlier presence of East African pastoralists than Iron Age agropastoralists in the region, since they carry ancestry related to the 1200-year-old admixed herder from the western Cape (15).

The admixture with food-producing populations did not take place at the same time or to the same extent across southern Africa (29,36). Analyses of uniparental data show a strongly sex-biased signal of gene flow in southern Africa, with Khoisan-speaking populations receiving paternal lineages from food-producers, whereas Bantu-speaking groups incorporated mainly Khoisan-related maternal lineages. The intensity of this sex bias increases from North to South, possibly indicating changes in social interactions between immigrating groups and autochthonous peoples over time (35). Such changes in interactions are also implied by the varying levels of Khoisan-related ancestry detectable in modern-day Bantu-speaking populations of southern Africa: populations from Malawi do not show any evidence for Khoisan-related ancestry (14), and populations from southern Mozambique show only low levels of such ancestry [4–5% maximum (43)]. This is in contrast to populations such as the Kgalagadi and Tswana from Botswana with 33–39% and 22–24% Khoisan-related ancestry, respectively (29,36), or the Sotho, Xhosa and Zulu from South Africa with between ~10–24% Khoisan-related ancestry (43,44). Such changes in social interactions between immigrating Iron Age agropastoralists and resident Khoisan-speaking populations might also explain variable patterns of click borrowing in Bantu languages (18,45).

Ethical Considerations

Indigenous communities are playing an increasingly prominent role in genomics research, going beyond merely providing samples to being fully informed about the results and how they are presented (46,47). Even well-meaning scientists engaged

in research on indigenous peoples can fail to appreciate how their scientific statements about their results may be viewed and interpreted by the individuals and communities studied—a prominent example involved a study that sequenced the genomes of four Khoisan-speaking individuals (cf. 48). One outcome of such misunderstandings was the establishment of the San Code of Research Ethics in 2017 (<https://www.globalcodeofconduct.org/affiliated-codes/>), the first such ethics code by an indigenous African group, and a model for research involving Khoisan-speaking groups. Nonetheless, ethical difficulties continue to arise (49).

Conclusion

The stereotypical image of Khoisan-speaking peoples as Stone Age hunter-gatherers who have lived in splendid isolation since the dawn of humankind can, without any doubt, be laid to rest. These groups exhibit extensive cultural, linguistic and biological diversity. They harbour more genetic diversity, the earliest divergences and larger effective population sizes than other human populations. They used to be more widespread in former times, are likely to have engaged in long-distance migrations and they have both influenced and been influenced by at least two migrations, an earlier migration of pastoralists from eastern Africa and a later migration of agropastoralists associated with the spread of Bantu languages. Understanding the complex genomic history and structure of Khoisan-speaking populations has important implications not only for their individual histories and the history of humans in general, but also for potential variation in disease susceptibility (cf. 50,51). There is a clear need for further whole genome sequence studies of Khoisan-speaking groups, in order to achieve these goals.

Acknowledgements

M.S. acknowledges support from the Max Planck Society. The authors thank Linda Schymanski for help with the figures. B.P. is grateful to the LABEX ASLAN (ANR-10-LABX-0081)

of Université de Lyon for its financial support within the programme 'Investissements d'Avenir' (ANR-11-IDEX-0007) of the French government operated by the National Research Agency (ANR).

Conflict of Interest statement. None declared.

References

- Güldemann, T. (2014) Khoisan' linguistic classification today. In Güldemann, T. and Fehn, A.-M. (eds), *Beyond 'Khoisan'. Historical Relations in the Kalahari Basin*. CILT, John Benjamins, Amsterdam, pp. 1–40.
- Hitchcock, R.K. (2020) The plight of the Kalahari San: hunter-gatherers in a globalized world. *J. Anthropol. Res.*, **76**, 164–184.
- Schenck, M. (2008) *Land, Water, Truth, and Love - Visions of Identity and Land Access: From Bain's Bushmen to Khomani San*, BA Hon Thesis. Mount Holyoke College, South Hadley, Massachusetts.
- Traill, A. (2007) !Khwa-ka Hhouiten Hhouiten 'The rush of the storm': the linguistic death of |Xam. In Skotnes, P. (ed), *Claim to the Country: The Archive of Lucy Lloyd and Wilhelm Bleek*. Ohio University Press, Athens, OH, pp. 130–147.
- South African Human Rights Commission (2004) *Report on the Inquiry into Human Rights Violations in the Khomani San Community, Andriesvale-Askham Area, Kalahari*. South African Human Rights Commission, Braamfontein, South Africa.
- Greenberg, J.H. (1963) *The Languages of Africa*. Indiana University Center in Anthropology, Folklore and Linguistics, and The Hague, Mouton, Bloomington.
- Güldemann, T. (2018) Historical linguistics and genealogical language classification in Africa. In Güldemann, T. (ed), *The Languages and Linguistics of Africa*. The World of Linguistics, De Gruyter Mouton, Berlin, Boston, pp. 58–444.
- Güldemann, T. (2008) A linguist's view: Khoe-Kwadi speakers as the earliest food-producers of southern Africa. *South. Afr. Humanit.*, **20**, 93–132.
- Barnard, A. (1992) *Hunters and Herders of Southern Africa. A Comparative Ethnography of the Khoisan Peoples*. Cambridge Studies in Social and Cultural Anthropology; Cambridge University Press, Cambridge.
- Robins, S. (2000) Land struggles and the politics and ethics of representing 'bushman' history and identity. *KronoScope*, **26**, 56–75.
- Oliveira, S., Fehn, A.-M., Aço, T., Lages, F., Gayà-Vidal, M., Pakendorf, B., Stoneking, M. and Rocha, J. (2018) Matriclans shape populations: insights from the Angolan Namib Desert into the maternal genetic history of southern Africa. *Am. J. Phys. Anthropol.*, **165**, 518–535.
- Bollig, M. (2005) Singing smiths and hunting ritual entrepreneurs: transitions between forager and peripatetic communities in Africa. In Berland, J.C. and Rao, A. (eds), *Customary Strangers. New Perspectives on Peripatetic Peoples in the Middle East, Africa and Asia*. Praeger, New York, pp. 195–232.
- Montinaro, F. and Capelli, C. (2018) The evolutionary history of southern Africa. *Curr. Opin. Genet. Dev.*, **53**, 157–164.
- Skoglund, P., Thompson, J.C., Prendergast, M.E., Mittnik, A., Sirak, K., Hajdinjak, M., Salie, T., Rohland, N., Mallick, S., Peltzer, A. et al. (2017) Reconstructing prehistoric African population structure. *Cell*, **171**, 59–71.e21.
- Wang, K., Goldstein, S., Bleasdale, M., Clist, B., Bostoen, K., Bakwa-Lufu, P., Buck, L.T., Crowther, A., Dème, A., McIntosh, R.J. et al. (2020) Ancient genomes reveal complex patterns of population movement, interaction, and replacement in sub-Saharan Africa. *Sci. Adv.*, **6**, eaaz0183.
- Schlebusch, C.M., Sjödin, P., Breton, G., Günther, T., Naidoo, T., Hollfelder, N., Sjöstrand, A.E., Xu, J., Gattepaille, L.M., Vicente, M. et al. (2020) Khoe-San genomes reveal unique variation and confirm the deepest population divergence in *Homo sapiens*. *Mol. Biol. Evol.*, **37**, 2944–2954.
- Barbieri, C., Vicente, M., Rocha, J., Mpoloka, S.W., Stoneking, M. and Pakendorf, B. (2013) Ancient substructure in early mtDNA lineages of southern Africa. *Am. J. Hum. Genet.*, **92**, 285–292.
- Gunnink, H., Sands, B., Pakendorf, B. and Bostoen, K. (2015) Prehistoric language contact in the Kavango-Zambezi trans-frontier area: Khoisan influence on southwestern Bantu languages. *J. Afr. Lang. Linguist.*, **36**, 193–232.
- Fan, S., Kelly, D.E., Beltrame, M.H., Hansen, M.E.B., Mallick, S., Ranciaro, A., Hirbo, J., Thompson, S., Beggs, W., Nyambo, T. et al. (2019) African evolutionary history inferred from whole genome sequence data of 44 indigenous African populations. *Genome Biol.*, **20**, 82.
- Lorente-Galdos, B., Lao, O., Serra-Vidal, G., Santpere, G., Kuderna, L.F.K., Arauna, L.R., Fadhlaoui-Zid, K., Pimenoff, V.N., Soodyall, H., Zalloua, P. et al. (2019) Whole-genome sequence analysis of a Pan African set of samples reveals archaic gene flow from an extinct basal population of modern humans into sub-Saharan populations. *Genome Biol.*, **20**, 77.
- Bergström, A., McCarthy, S.A., Hui, R., Almarri, M.A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J. et al. (2020) Insights into human genetic variation and population history from 929 diverse genomes. *Science*, **367**, eaay5012.
- Almarri, M.A., Bergström, A., Prado-Martinez, J., Yang, F., Fu, B., Dunham, A.S., Chen, Y., Hurler, M.E., Tyler-Smith, C. and Xue, Y. (2020) Population structure, stratification, and introgression of human structural variation. *Cell*, **182**, 189–199.e15.
- Schlebusch, C.M., Malmström, H., Günther, T., Sjödin, P., Coutinho, A., Edlund, H., Munters, A.R., Vicente, M., Steyn, M., Soodyall, H. et al. (2017) Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science*, **358**, 652–655.
- Chan, E.K.F., Timmermann, A., Baldi, B.F., Moore, A.E., Lyons, R.J., Lee, S.-S., Kalsbeek, A.M.F., Petersen, D.C., Rautenbach, H., Förtsch, H.E.A. et al. (2019) Human origins in a southern African palaeo-wetland and first migrations. *Nature*, **575**, 185–189.
- Ackermann, R.R., Athreya, S., Black, W., Cabana, G.S., Hare, V., Pickering, R. and Schroeder, L. (2019) Upholding 'good science' in human origins research: a response to Chan et al. *AfricArXiv*. doi: [10.31730/osf.io/qtjfp](https://doi.org/10.31730/osf.io/qtjfp).
- Schlebusch, C.M., Loog, L., Groucutt, H.S., King, T., Rutherford, A., Barbieri, C., Barbujani, G., Chikhi, L., Jakobsson, M., Eriksson, A. et al. (2019) Human origins in southern African palaeo-wetlands?: strong claims from weak evidence. doi: [10.20944/preprints201911.0193.v1](https://doi.org/10.20944/preprints201911.0193.v1).
- Kim, H.L., Ratan, A., Perry, G.H., Montenegro, A., Miller, W. and Schuster, S.C. (2014) Khoisan hunter-gatherers have been

- the largest population throughout most of modern-human demographic history. *Nat. Commun.*, **5**, 5692.
28. Hammer, M.F., Woerner, A.E., Mendez, F.L., Watkins, J.C. and Wall, J.D. (2011) Genetic evidence for archaic admixture in Africa. *P. Natl. Acad. Sci. U S A*, **108**, 15123–15128.
 29. Pickrell, J.K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Güldemann, T., Kure, B., Mpoloka, S.W., Nakagawa, H., Naumann, C. et al. (2012) The genetic prehistory of southern Africa. *Nat. Commun.*, **3**, 1143.
 30. Schlebusch, C.M., Skoglund, P., Sjödin, P., Gattepaille, L.M., Hernandez, D., Jay, F., Li, S., De Jongh, M., Singleton, A., Blum, M.G.B. et al. (2012) Genomic variation in seven Khoesan groups reveals adaptation and complex African history. *Science*, **338**, 374–379.
 31. Montinaro, F., Busby, G.B.J., Gonzalez-Santos, M., Oosthuizen, O., Oosthuizen, E., Anagnostou, P., Destro-Bisol, G., Pascali, V.L. and Capelli, C. (2017) Complex ancient genetic structure and cultural transitions in southern African populations. *Genetics*, **205**, 303–316.
 32. Uren, C., Kim, M., Martin, A.R., Bobo, D., Gignoux, C.R., van Helden, P.D., Möller, M., Hoal, E.G. and Henn, B.M. (2016) Fine-scale human population structure in southern Africa reflects ecogeographic boundaries. *Genetics*, **204**, 303–314.
 33. Vicente, M., Jakobsson, M., Ebbesen, P. and Schlebusch, C.M. (2019) Genetic affinities among southern Africa hunter-gatherers and the impact of admixing farmer and herder populations. *Mol. Biol. Evol.*, **36**, 1849–1861.
 34. Barbieri, C., Güldemann, T., Naumann, C., Gerlach, L., Berthold, F., Nakagawa, H., Mpoloka, S.W., Stoneking, M. and Pakendorf, B. (2014) Unraveling the complex maternal history of southern African Khoisan populations. *Am. J. Phys. Anthropol.*, **153**, 435–448.
 35. Bajić, V., Barbieri, C., Hübner, A., Güldemann, T., Naumann, C., Gerlach, L., Berthold, F., Nakagawa, H., Mpoloka, S.W., Roewer, L. et al. (2018) Genetic structure and sex-biased gene flow in the history of southern African populations. *Am. J. Phys. Anthropol.*, **167**, 656–671.
 36. Pickrell, J.K., Patterson, N., Loh, P.-R., Lipson, M., Berger, B., Stoneking, M., Pakendorf, B. and Reich, D. (2014) Ancient west Eurasian ancestry in southern and eastern Africa. *P. Natl. Acad. Sci. U S A*, **111**, 2632–2637.
 37. Lander, F. and Russell, T. (2018) The archaeological evidence for the appearance of pastoralism and farming in southern Africa. *PLoS One*, **13**, e0198941.
 38. Breton, G., Schlebusch, C.M., Lombard, M., Sjödin, P., Soodyall, H. and Jakobsson, M. (2014) Lactase persistence alleles reveal partial east African ancestry of southern African Khoe pastoralists. *Curr. Biol.*, **24**, 852–858.
 39. Macholdt, E., Lede, V., Barbieri, C., Mpoloka, S.W., Chen, H., Slatkin, M., Pakendorf, B. and Stoneking, M. (2014) Tracing pastoralist migrations to southern Africa with lactase persistence alleles. *Curr. Biol.*, **24**, 875–879.
 40. Tishkoff, S.A., Reed, F.A., Ranciaro, A., Voight, B.F., Babbitt, C.C., Silverman, J.S., Powell, K., Mortensen, H.M., Hirbo, J.B., Osman, M. et al. (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.*, **39**, 31–40.
 41. Pinto, J.C., Oliveira, S., Teixeira, S., Martins, D., Fehn, A.-M., Aço, T., Gayà-Vidal, M. and Rocha, J. (2016) Food and pathogen adaptations in the Angolan Namib desert: tracing the spread of lactase persistence and human African trypanosomiasis resistance into southwestern Africa. *Am. J. Phys. Anthropol.*, **161**, 436–447.
 42. Oliveira, S., Hübner, A., Fehn, A.-M., Aço, T., Lages, F., Pakendorf, B., Stoneking, M. and Rocha, J. (2019) The role of matrilineality in shaping patterns of Y chromosome and mtDNA sequence variation in southwestern Angola. *Eur. J. Hum. Genet.*, **27**, 475–483.
 43. Semo, A., Gayà-Vidal, M., Fortes-Lima, C., Alard, B., Oliveira, S., Almeida, J., Prista, A., Damasceno, A., Fehn, A.-M., Schlebusch, C. et al. (2020) Along the Indian Ocean coast: genomic variation in Mozambique provides new insights into the Bantu expansion. *Mol. Biol. Evol.*, **37**, 406–416.
 44. Choudhury, A., Ramsay, M., Hudson, E.R., Christoffels, A., Gamielien, J., Sefid-Dashti, M.J. et al. (2017) Whole-genome sequencing for an enhanced understanding of genetic variation among south Africans. *Nat. Commun.*, **8**, 2062.
 45. Pakendorf, B., Gunnink, H., Sands, B. and Bostoen, K. (2017) Prehistoric Bantu-Khoisan language contact. *Lang. Dyn. Chang.*, **7**, 1–46.
 46. Garrison, N.A., Hudson, M., Ballantyne, L.L., Garba, I., Martinez, A., Tualii, M., Arbour, L., Caron, N.R. and Rainie, S.C. (2019) Genomic research through an indigenous lens: understanding the expectations. *Annu. Rev. Genom. Hum. G.*, **20**, 495–517.
 47. Hudson, M., Garrison, N.A., Sterling, R., Caron, N.R., Fox, K., Yracheta, J., Anderson, J., Wilcox, P., Arbour, L., Brown, A. et al. (2020) Rights, interests and expectations: indigenous perspectives on unrestricted access to genomic data. *Nat. Rev. Genet.*, **21**, 377–384.
 48. Chennells, R. and Steenkamp, A. (2018) International genomics research involving the San people. In Schroeder, D., Cook, J., Hirsch, F., Fenet, S. and Muthuswamy, V. (eds), *Ethics Dumping. Case Studies from North-South Research Collaborations*. Springer Briefs in Research and Innovation Governance, Springer, Cham, pp. 15–22.
 49. Stokstad, E. (2019) Genetics lab accused of misusing African DNA. *Science*, **2019**, 555–556.
 50. Thami, P.K. and Chimusa, E.R. (2019) Population structure and implications on the genetic architecture of HIV-1 phenotypes within southern Africa. *Front. Genet.*, **10**, 905.
 51. Swart, Y., van Eeden, G., Sparks, A., Uren, C. and Möller, M. (2020) Prospective avenues for human population genomics and disease mapping in southern Africa. *Mol. Genet. Genomics*, **295**, 1079–1089.
 52. Schlebusch, C.M., de Jongh, M. and Soodyall, H. (2011) Different contributions of ancient mitochondrial and Y-chromosomal lineages in 'Karretjie people' of the great Karoo in South Africa. *J. Hum. Genet.*, **56**, 623–630.
 53. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A. et al. (2016) The Simons genome diversity project: 300 genomes from 142 diverse populations. *Nature*, **538**, 201–206.
 54. Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C. et al. (2012) A high-coverage genome sequence from an archaic Denisovan individual. *Science*, **338**, 222–226.
 55. Schuster, S.C., Miller, W., Ratan, A., Tomsho, L.P., Giardine, B., Kasson, L.R., Harris, R.S., Petersen, D.C., Zhao, F., Qi, J. et al. (2010) Complete Khoisan and Bantu genomes from southern Africa. *Nature*, **463**, 943–947.