

INVITED REVIEW ARTICLE

Bantu-speaker migration and admixture in southern Africa

Ananyo Choudhury^{1,*,\$}, Dhriti Sengupta^{1,†,\$}, Michele Ramsay^{1,‡} and Carina Schlebusch^{2,3,4,¶}

¹Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg 2193, South Africa, ²Palaeo-Research Institute, University of Johannesburg, Auckland Park 2006, South Africa, ³Human Evolution, Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18C, SE-752 36 Uppsala 75326, Sweden and ⁴SciLifeLab, Uppsala 75237, Sweden

*To whom correspondence should be addressed at: Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, 2193, Johannesburg, South Africa. Tel: +270117176389; Email: Ananyo.Choudhury@wits.ac.za

Abstract

The presence of Early and Middle Stone Age human remains and associated archeological artifacts from various sites scattered across southern Africa, suggests this geographic region to be one of the first abodes of anatomically modern humans. Although the presence of hunter-gatherer cultures in this region dates back to deep times, the peopling of southern Africa has largely been reshaped by three major sets of migrations over the last 2000 years. These migrations have led to a confluence of four distinct ancestries (San hunter-gatherer, East-African pastoralist, Bantu-speaker farmer and Eurasian) in populations from this region. In this review, we have summarized the recent insights into the refinement of timelines and routes of the migration of Bantu-speaking populations to southern Africa and their admixture with resident southern African Khoe-San populations. We highlight two recent studies providing evidence for the emergence of fine-scale population structure within some South-Eastern Bantu-speaker groups. We also accentuate whole genome sequencing studies (current and ancient) that have both enhanced our understanding of the peopling of southern Africa and demonstrated a huge potential for novel variant discovery in populations from this region. Finally, we identify some of the major gaps and inconsistencies in our understanding and emphasize the importance of more systematic studies of southern African populations from diverse ethnolinguistic groups and geographic locations.

Introduction

Large-scale migrations and admixture have played a key role in shaping the genetic diversity of southern African people. We recognize that there are several ways of defining southern Africa as a geographical region and here our definition includes the five southernmost countries of the continent—South Africa, Namibia, Botswana, Zimbabwe and Mozambique

as well as three countries, Angola, Zambia and Malawi, that form the border between southern Africa and Central Africa. Paleoanthropological evidence establishes the presence of the genus *Homo* for the last 2 million years and anatomically modern humans from 120 thousand years ago (kya) in this geographic area (1). The southern African archeological and rock art record for the last 20 000–40 000 years, indicate an almost constant and

†Dhriti Sengupta, <http://orcid.org/0000-0001-6315-7804>

‡Michele Ramsay, <http://orcid.org/0000-0002-4156-4801>

¶Carina Schlebusch, <http://orcid.org/0000-0002-8160-9621>

§The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Received: October 9, 2020. Revised: December 15, 2020. Accepted: December 15, 2020

© The Author(s) 2020. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

wide-spread presence of a hunter-gatherer culture, with similarities to the modern-day southern African hunter-gatherer cultures (2–4). Although represented by rather small populations (often bands of less than 60 people), the present day hunter-gatherer populations from southern Africa, referred to as San communities, still live across a large geographic area that ranges from the Northern Cape Province of South Africa, large parts of Namibia, Botswana, southern Angola and southwestern Zimbabwe (5). The pronounced genetic differences, ancient population split estimates within various San groups and historical and archeological records attest to their once widespread distribution across the southern African landscape (6–10). The Khoekhoe herder communities and San hunter-gatherer communities are often referred to together with the term ‘Khoekhoe-San’ and in accordance with their immense genetic diversity have been discussed in an independent review in this issue.

Genetic evidence suggests that the migrations of three distinct groups of people in the last 2000 years had a large impact on the genetic diversity of the region (11,12). The first migration involved a small group of pastoralists from East Africa. This small immigrating population was eventually assimilated by local southern African San hunter-gatherer groups, resulting in a new population that was ancestral to the present day Khoekhoe herder populations (6,13–17). This migration was closely followed by the second set of migrations, which involved a large scale movement of Bantu-speaker farmers with a West African origin. The third and final major migration into southern Africa dates back to over the last four centuries. In addition to the arrival of several waves of colonial settlers from Europe, slave trade across the Indian Ocean also introduced ancestries from South Asia, East Asia and Madagascar (18). Subsequent admixture between these settlers, and between them and the local populations, gave rise to complex patterns of genomic admixture in some regions of southern Africa (15,18–22).

Today 90% of residents of southern Africa are Bantu-speakers, however, genome-wide studies of populations from this region have been largely focused on Khoekhoe-San hunter-gatherers and admixed populations (15,19,22–26). Moreover, the Bantu-speaker ethnolinguistic groups that were covered in these initial studies were largely from only a few countries, such as South Africa (15,21,27,28), Namibia and Botswana (15,25,26) (Table 1) which has not allowed for a comprehensive reconstruction of the events and timelines for migration and admixture in this region. However, in the last few years there have been several investigations into the genetics of southern African Bantu-speaking ethnolinguistic groups using population-scale genomic datasets (Table 1). Although genetic contributions from early pastoralists and later European settlers form a critical component of southern African genetic diversity, in this review, we focused on the recent insights from genome wide data into the migration and admixture of Bantu-speakers in southern Africa.

Routes of Bantu-Speaker Migrations in Southern Africa

The Bantu-expansion began around ~5–4 kya in West Africa (29,30), however, the initial phases of this expansion (5–2.6 kya) were slow and confined to West-Central Africa. Most hypotheses about the Bantu-expansion routes are based on linguistics and archeology, however, archeological and linguistic inferences do not agree on several aspects (9,30–33). The linguistic ‘late-spill’ model, which proposes that climate change-induced corridors through the Central African rainforest (~2.6–2.4 kya), facilitated

rapid eastward and southward expansions, are supported by recent linguistic and genetic studies (32,34–37). After migrating south of the rainforest (probably somewhere around present-day Eastern DRC or Angola), the Bantu-speakers separated into two groups (36,38). One of these groups expanded eastward, whereas the other moved directly south giving rise to the genetically (15) and linguistically (35) distinguishable South-Eastern Bantu-speaker (SEB) and South-Western Bantu-speaker (SWB) populations, respectively. The SEB group that migrated eastward, after reaching present day Zambia, probably again split into two branches, one continued eastward while the other moved South-East (38). However, studies based on a different set of populations, suggested the southward movement of groups could have been initiated after reaching Malawi (39). The archeological record does not overlap with all aspects of the linguistic based hypotheses indicating that the expansion of the Bantu-speakers across southern Africa, instead of being a single large-scale movement, likely occurred in different phases (30,40,41). The first arrival of Bantu-speaking agro-pastoralists in southern Africa is estimated to be around 2 kya (9,31,40,42).

Admixture of Bantu-Speakers with Resident Hunter-Gatherer Populations

The study of admixture patterns in Bantu-speaker groups from southern Africa is complex and underlined by the presence of rain forest forager (RFF) gene-flow in some groups, Khoekhoe-San in others, or their absence in yet other groups (Fig. 1, Supplementary Material, Table S1) (15,21,26,28,36–39,43–45). In addition, the extent of the gene-flow also varies considerably between the SEB groups clearly differentiating them from one other. For example, the Khoekhoe-San ancestry levels vary from >20% in the South African Tswana and Sotho to only around 3% in the Chopi and Tswa from south Mozambique, whereas central and north Mozambican populations, Zambian and Malawian populations have no admixture signals with Khoekhoe-San (16,37–39,46) (Fig. 1).

The differential admixture patterns have enabled the delineation of a broad timeline for the movement of Bantu-speakers into and within southern Africa (Fig. 1, Supplementary Material, Table S1). The timelines for Khoekhoe-San admixture, irrespective of geography, date back within the last 1500 years. Moreover, while some of the studies (37,45) suggest the existence of local clines, neither overall admixture proportions nor admixture dates show any clear east to west or north to south cline across the broader geographic region (Fig. 1). Interestingly, the populations that show the least admixture, irrespective of geography, were also found to have the deepest admixture dates (Fig. 1). More careful investigations would be required to test whether these dates correspond to historical events or reflect limitations of dating techniques in scenarios of very low admixture levels. Moreover, the use of different techniques [such as GLOBETROTTER (47) in some and ALDER/MALDER (48) in others], parameters and proxy populations for dating admixture also do not allow for a direct comparison of the estimates from independent studies.

Among the southern African Bantu-speaker populations, those currently living in Angola are the only group to harbor considerable RFF admixture. The absence of clear RFF admixture in all other Bantu-speaker groups suggests that these groups passed through the rainforest without any major admixture with local groups. Several alternative scenarios, such as the RFF ancestry having been introduced to Angola from Central-Africa after the initial phase of Bantu-expansion, or that current-day southern African Bantu-speakers are from a later wave that did

Table 1. List of major genomic studies on southern African Bantu-speakers

Study	Country of origin	Sample size	Bantu-speaker groups ^a
Genotype array			
Schlebusch et al. (2012)	Namibia	12	Herero (South-Western Bantu-speakers)
	South Africa	20	South-Eastern Bantu-speakers
Pickrell et al. (2012)	Namibia	10	Owambo (5), Himba (5)
	Zambia	4	Mbukushu (4)
	Botswana	10	Tswana (5), Kgalagadi (5)
May et al. (2013)	South Africa	94	South-Eastern Bantu-speakers
Peterson et al. (2013)	South Africa	15	amaXhosa
Chimusa et al. (2015)	South Africa	86	Sotho (25), Xhosa (36), Zulu (25)
	Namibia	25	Herero
Gurdasani et al. (2015)	South Africa	86	Sotho
	South Africa	100	Zulu
Busby et al. (2016)	Malawi	100	Chewa
Patin et al. (2017)	Angola	42	Kimbundu (17), Kongo (10), Ovimbundu (15)
Montinaro et al. (2017) ^b	South Africa	8	Sotho
	Namibia/Angola	17	Kwangali (7), Owambo (10)
	Angola	8	Mbukushu
Semo et al. (2019)	Mozambique	200	Mwani (5), Yao (20), Makhwa (19), Nyanja (16), Sena (20), Manyika (15), Ndau (18), Tswa (6), Bitonga (17), Changana (9), Chopi (13), Ronga (3)
	Angola	39	Umbundu (20), Ganguela (6), Nyaneka (13)
Sengupta et al. (2020)	South Africa	4319	Pedi (1065), Sotho (366), Swazi (126), Tsonga (1644), Tswana (242), Venda (73), Xhosa (177), Zulu (626)
Whole genome and exome sequencing			
Gurdasani et al. (2015)	South Africa	100	Zulu
Mallick et al. (2016)	Botswana or Namibia	4	BantuHerero (2) and BantuTswana (2)
Choudhury et al. (2017)	South Africa	16	Sotho (8), Xhosa (8)
Bergström et al. (2020)	South Africa	8	Bantu SouthAfrica
Choudhury et al. (2020)	Botswana	48	Multiple Bantu speaking groups
	Zambia	41	Multiple Bantu speaking groups
Retshabile et al. (2018)	Botswana	164	18 Bantu-speaking groups
Ancient genomes^c			
Schlebusch et al. (2017)	South Africa	7	Time-period (~ 2000–300 BP)
Skoglund et al. (2017)	South Africa	3	Time period (1200–2000 BP)
	Malawi	7	Time period (2500–8100 BP)
Wang et al. (2020)	Botswana	4	Time period (900–1400 BP)

^aOnly Bantu-speaker ethnolinguistic groups from each study are shown.

^bIncludes data from Gonzalez-Santos et al. (2015).

^cIncludes all samples irrespective of their possible linguistic affiliation or lifestyle.

not mix with RFF, are also possible. The RFF admixture dates also vary quite widely from ~70 generations in Kango to 18 generations ago in Ovimbundu (36).

The Khoe-San admixture in SWB reported in two studies (36,39) seems to be much more recent (~12.94 generations ago) compared with that for all other SEB groups. In contrast, Khoe-San admixture dates for SWB in another study (10) is comparable to some of the older dates (~40 generations ago) observed for SEB groups. In-depth analysis of admixture dates from other SWB groups would be required to determine if the migration along the west was a much slower process and also if it involved two distinct waves of migration.

Population Structure within South-Eastern Bantu-Speakers

Despite a relatively recent shared history, more than 25 independent SEB languages are spoken across southern Africa. For example, just within the boundaries of present day South Africa, more than 10 different SEB languages are spoken in a somewhat

geographically stratified manner. Similarly, several different SEB languages are spoken each in Mozambique and Botswana. The possibility that some of these linguistic differences are paralleled by differences in the genetic variation of SEB speakers from South Africa was first proposed by Lane and colleagues almost two decades ago (49). Although studies conducted using population-scale genome-wide genotype data did not explicitly replicate this initial observation, a careful scrutiny of the results from some of these studies that included multiple SEB groups supported the possibility of the presence of a fine-scale population-structure (21,28,43). The Southern African Human Genome Programme pilot study (43) was the first study to reinforce the possibility of population structure within the SEB. These studies also indicated that differential Khoe-San gene-flow into various SEB ethnolinguistic groups might play a major role in the genetic differentiation of these groups.

Based on a large study of more than 5000 SEB speakers, Sengupta et al. (45) provided compelling evidence for a fine-scale population structure within the southern African SEB-speakers. The study not only demonstrated a clear correlation between genetic and linguistic variation but also showed a clear

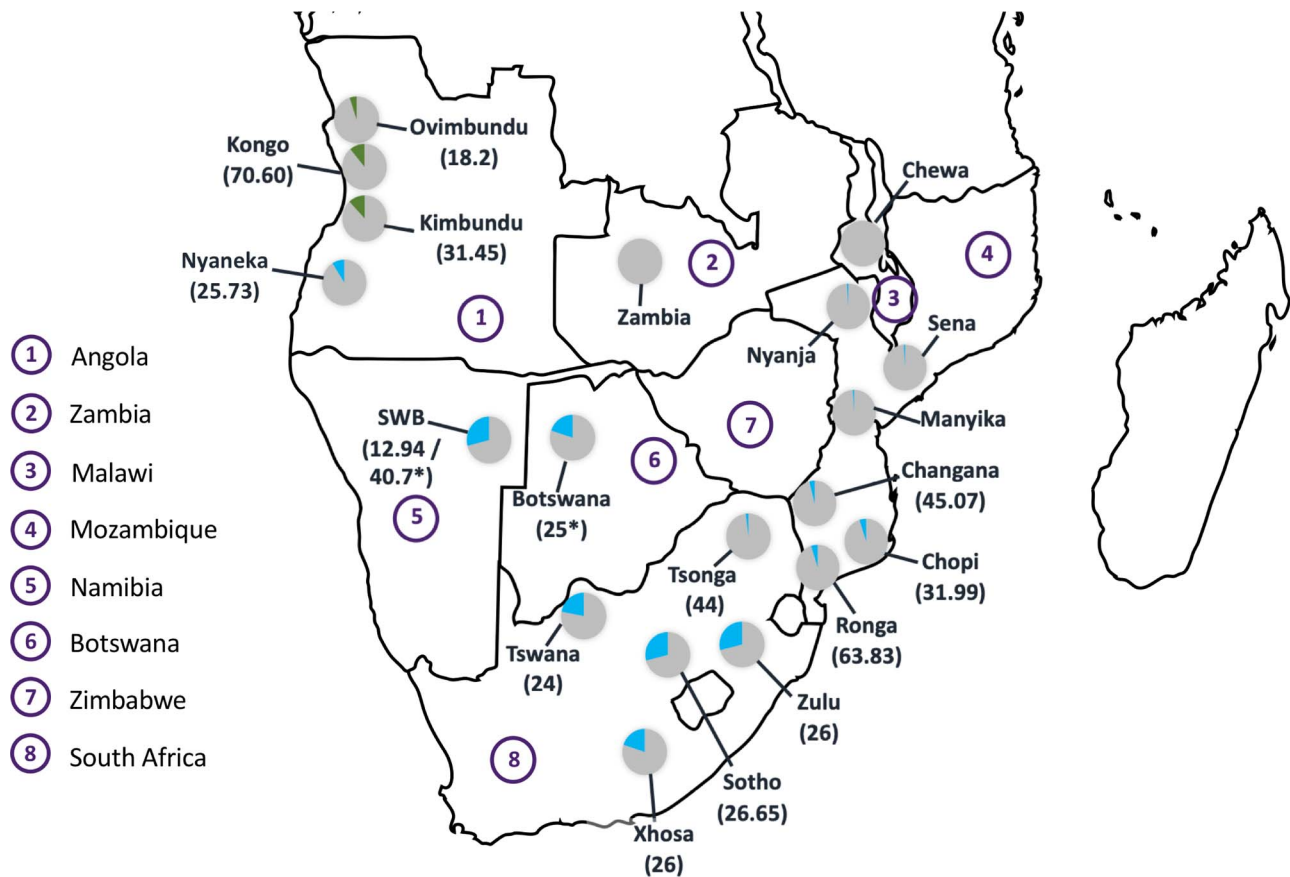


Figure 1. Hunter-gatherer (Kho-San and RFF) admixture and ancestry in Bantu-speaker groups from southern Africa. The pie charts corresponding to each population show Bantu-speaker ancestry in grey, Kho-San in blue and RFF ancestry in green. The numbers in parenthesis below population names represent the corresponding admixture dates in terms of generations using GLOBETROTTER. Admixture dates inferred using MALDER are indicated by a '*'. The geographic positioning of groups within a country is only indicative and might not correspond to the actual location of sampling. Please see [Supplementary Material, Table S1](#) for further details.

correlation between genetic variation and the geographic spread of the populations. Another interesting result from the study was the persistence of fine-scale population structure even after masking of Kho-San ancestry suggesting that the structure is not just a consequence of differential admixture but also reflects intrinsic differences in demographic history of various SEB groups. Similarly, the study of Semo *et al.* (37), also reported notable differences in levels of Kho-San admixture, which was found in southern but not in more northern SEB groups from Mozambique. They furthermore, detected a clear south bound decline of genetic diversity in SEB across Mozambique. These findings are consistent with a serial founder model of the movement of SEB from north to south across the country.

Nature of Interaction Between Kho-San and Bantu-Speakers

A clear sex-bias in ancestral contributions to a recently admixed population may provide clues to the prevailing social dynamic when the admixture between two populations occurred. For example, higher proportions of mitochondrial or X-chromosome ancestry from one of the groups suggests that more women from that particular group contributed to the genetic variation in subsequent generations compared with the male ancestors from that group (50). This could hint at scenarios where social dominance interactions are at play between admixing populations.

The mitochondrial and Y chromosome proportions in SEB have shown the interaction among the Kho-San and Bantu-speakers to be female biased for the former and male biased for the later (43,51–56). Although a recent study based on X-autosome ancestry comparison confirmed this trend, the extent of Kho-San maternal bias was observed to be highly variable among the SEB groups (45). This suggests that despite an overarching trend of male-biased gene-flow from Bantu-speakers into the Kho-San and female biased gene-flow from Kho-San into the Bantu-speakers, these interactions were strongly influenced by locally defined factors that were unique to each of these interactions. This is in agreement with findings from uniparental marker based studies that show prevailing cultural practices and social structures (matrilineal and patrilineal) could influence such interactions (57–59).

Whole Genome Sequence Based Investigation of Southern African Populations

Kho-San populations, in accordance with their unique and pronounced diversity and ancient divergence from other populations, harbor a large number of novel (previously unidentified) variants (8,60,61). Interestingly, despite a rather recent separation of SEB from other Bantu-speaker populations, several sequencing-based studies have also shown a high rate of novel variant discovery in the SEB groups (28,38,43). Even an exome

sequencing study in a population from Botswana identified ~15% of the detected variants to be novel (44). A recent study showed a strong correlation between Khoe-San admixture proportions and novel variant detection rate, providing a possible rationale for an overall higher novel variant discovery rate in SEB populations (38).

Insights from ancient genomes

A study that sequenced the nuclear genomes of seven ancient southern African individuals, three dating to the Later Stone Age ~2 kya and four dating to the Iron Age (300–500 years ago), found that the Later Stone Age individuals were related to current-day Khoe-San hunter-gatherer individuals and the Iron Age individuals to current-day South African BS (thus containing West African ancestry) (6). This study confirmed large-scale population replacement in southern Africa, where Later Stone Age ancestors of the Khoe-San hunter-gatherers were replaced by incoming Bantu-speaking farmer groups of West African genetic ancestry, introducing the Iron Age into the region. In a follow-up investigation of the Iron Age genomes analyzed in the context of a high-resolution dataset of South African Bantu-speaking ethno-linguistic groups, it appeared that the two older Iron Age genomes were more related to Tsonga and Venda groups, whereas the two younger genomes were more related to Nguni speakers (45). This finding further supports a possible multi-stage process, with several waves of migrations, during the expansion of Bantu-speakers into Southern Africa.

South African Bantu-speakers received substantial amounts of gene-flow from local Khoe-San hunter-gatherers. The ancient Iron Age genomes showed slightly less admixture compared with most current-day populations from the region (6) and, although based on only a few samples, suggest increasing admixture over time. Interestingly, further north, the Bantu-expansion seemed to have had different demographic dynamics in terms of interaction between hunter-gatherers and incoming farmers. Current-day populations from Malawi (16) and Mozambique (37) show little to no admixture with hunter-gather groups that likely occupied the area before the Bantu-expansion. These findings indicate that the diffusion of Bantu languages and culture throughout sub-equatorial Africa was a complex process and the admixture dynamics between farmers and hunter-gatherers played an important role in creating patterns of genetic diversity. A recent ancient DNA-based study, that included samples from Botswana, further showed evidence that confirms that the movement of East-African pastoral populations into southern Africa predates the movement of Bantu-speaking farmers into the region (62). Further ancient DNA studies and better coverage of sub-equatorial Africa in terms of modern-day populations are needed to clarify the demographic dynamics, migration routes and interactions during the Bantu-expansion.

Future perspective

Against the backdrop of immense genetic diversity and complexities in the interactions of diverse ancestries that are emerging from recent studies, the currently available data provides only a fragmented view of the genomic landscape of the southern African subcontinent. As the Bantu-expansion involved differential admixture with the Khoe-San, the dating of admixture between the migrant Bantu-speaker and resident Khoe-San could be helpful in uncovering the details of this

migration. The available timeline for admixture dating is limited and in some cases not comparable across studies because of the use of different datasets, multiple dating methods and varying sample sizes. Studies based on more inclusive populations, with harmonized datasets and uniform methodological approaches could enable the generation of a congruent timeline of major migration and admixture events. The availability of population level genomic data from Zimbabwe, and large-scale datasets for more ethnolinguistic groups from Namibia, Botswana, Zambia and Angola would also be valuable in generating a more robust granular route map and timeline of Bantu-migrations across this region. The sequencing of more 'ancient genomes' from the geographic region would also enable testing and refining of the models. A better characterization of Eurasian ancestry among various southern African populations including partitioning of their source into recent Eurasian gene-flow and older gene-flow via the Indian ocean, specifically in Mozambique and South Africa, could also yield valuable information on hitherto unknown migration and admixture events. Finally, the large number of novel variants and divergent gene-flow patterns also underline the necessity of studying a wider variety of southern African populations using whole genome approaches and the importance of their inclusion in reference panels and databases.

Conclusions

The spread of Bantu-speaking groups from their West African homeland, across sub-equatorial Africa was a complex process. It likely involved both rapid and gradual expansions and multiple distinct waves. A significant feature of the dispersal of Bantu-speaking populations across southern Africa has been its accompaniment of a bidirectional gene-flow with local populations in certain regions, whereas in other regions there are limited to no admixture and a complete population replacement seems to be the most plausible explanation. Especially more towards the south, substantial Khoe-San ancestry was introduced into the SEB and SWB, and there is Bantu-speaker ancestry present in many (but not all) present-day Khoe-San populations. The variability of the Khoe-San admixture in the SEB and SWB groups, both in terms of ancestry proportions and estimated time for admixture are starting to contribute to hypotheses on the dynamics of the Bantu-expansion in southern Africa. Ancient genomes from the Southern African Iron age, on the other hand, have the potential to provide additional temporal evidence for benchmarking the migration, settlement and admixture timelines. Accumulating evidence is suggesting that differential Khoe-San admixture along with distinctive demographic histories have resulted in a fine-scale population-structure among SEB groups. Cautious sampling and appropriate statistical approaches are therefore required for studies on the genetic architecture of complex diseases and traits in southern African populations. Future genetic studies on ancient and modern southern African genomes, together with inferences from the archeological and linguistic fields will continue to clarify the complexities of the expansion of Bantu-speaking farmers into the southern region of the continent.

Supplementary Material

Supplementary Material is available at HMG online.

Conflict of Interest statement. None declared.

Funding

A.C. and D.S. are funded by the National Institutes of Health (National Human Genome Research Institute, NHGRI) AWI-Gen Collaborative Centre under award number U54HG006938, as part of the H3Africa Consortium. M.R. is a South African Research Chair in Genomics and Bioinformatics of African populations hosted by the University of the Witwatersrand, funded by the Department of Science and Technology and administered by National Research Foundation of South Africa (NRF). C.S. was funded by the European Research Council (ERC—no. 759933) and the Knut and Alice Wallenberg foundation. This paper describes the views of the authors and does not necessarily represent the official views of the funders.

References

- Dusseldorp, G., Lombard, M. and Wurz, S. (2013) Pleistocene homo and the updated stone age sequence of South Africa. *S. Afr. J. Sci.*, **109**, 46–52.
- Deacon, H.J. and Deacon, J. (1999) *Human Beginnings in South Africa: Uncovering the Secrets of the Stone Age*. In *David Philip*. Cape Town, South Africa.
- Wadley, L., Hodgskiss, T. and Grant, M. (2009) Implications for complex cognition from the hafting of tools with compound adhesives in the middle stone age, South Africa. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 9590–9594.
- d’Errico, F., Backwell, L., Villa, P., Degano, I., Lucejko, J.J., Bamford, M.K., Higham, T.F.G., Colombini, M.P. and Beaumont, P.B. (2012) Early evidence of San material culture represented by organic artifacts from border cave, South Africa. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 13214–13219.
- Güldemann, T. and Vossen, R. (2000) Khoisan. In Heine, B. and Nurse, D. (eds), *2000. African Languages: An Introduction*. Cambridge University Press, Cambridge, UK, pp. 99–122.
- Schlebusch, C.M., Malmström, H., Günther, T., Sjödin, P., Coutinho, A., Edlund, H., Munters, A.R., Vicente, M., Steyn, M., Soodyall, H., Lombard, M. and Jakobsson, M. (2017) Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science*, **358**, 652–655.
- Choudhury, A., Aron, S., Sengupta, D., Hazelhurst, S. and Ramsay, M. (2018) African genetic diversity provides novel insights into evolutionary history and local adaptations. *Hum. Mol. Genet.*, **27**, R209–R218.
- 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.
- Schlebusch, C.M. and Jakobsson, M. (2018) Tales of human migration, admixture, and selection in Africa. *Annu. Rev. Genomics Hum. Genet.*, **19**, 405–428.
- Schlebusch, C.M., Prins, F., Lombard, M., Jakobsson, M. and Soodyall, H. (2016) The disappearing San of southeastern Africa and their genetic affinities. *Hum. Genet.*, **135**, 1365–1373.
- Montinaro, F. and Capelli, C. (2018) The evolutionary history of southern Africa. *Curr. Opin. Genet. Dev.*, **53**, 157–164.
- Uren, C., Möller, M., van Helden, P.D., Henn, B.M. and Hoal, E.G. (2017) Population structure and infectious disease risk in southern Africa. *Mol. Gen. Genomics.*, **292**, 499–509.
- 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.
- 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.
- 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., Soodyall, H. and Jakobsson, M. (2012) Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science*, **338**, 374–379.
- 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.
- 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. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 2632–2637.
- Hollfelder, N., Erasmus, J.C., Hammaren, R., Vicente, M., Jakobsson, M., Greeff, J.M. and Schlebusch, C.M. (2020) Patterns of African and Asian admixture in the Afrikaner population of South Africa. *BMC Biol.*, **18**, 16.
- de Wit, E., Delport, W., Rugamika, C.E., Meintjes, A., Möller, M., van Helden, P.D., Seoighe, C. and Hoal, E.G. (2010) Genome-wide analysis of the structure of the South African coloured population in the western cape. *Hum. Genet.*, **128**, 145–153.
- Petersen, D.C., Libiger, O., Tindall, E.A., Hardie, R.A., Hanick, L.I., Glashoff, R.H., Mukerji, M., Indian Genome Variation Consortium, Fernandez, P., Haacke, W., Schork, N.J. and Hayes, V.M. (2013) Complex patterns of genomic admixture within southern Africa. *PLoS Genet.*, **9**, 10–13.
- Chimusa, E.R., Meintjes, A., Tchanga, M., Mulder, N., Seoighe, C., Soodyall, H. and Ramesar, R. (2015) A genomic portrait of haplotype diversity and signatures of selection in indigenous southern African populations. *PLoS Genet.*, **11**, e1005052.
- Daya, M., van der Merwe, L., Gignoux, C.R., van Helden, P.D., Möller, M. and Hoal, E.G. (2014) Using multi-way admixture mapping to elucidate TB susceptibility in the South African coloured population. *BMC Genomics*, **15**, 1021.
- Henn, B.M., Gignoux, C.R., Jobin, M., Granka, J.M., Macpherson, J.M., Kidd, J.M., Rodríguez-Botigué, L., Ramachandran, S., Hon, L., Brisbin, A. et al. (2011) Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 5154–5162.
- 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.
- 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.
- 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.
- May, A., Hazelhurst, S., Li, Y., Norris, S.A., Govind, N., Tikly, M., Hon, C., Johnson, K.J., Hartmann, N., Staedtler, F. and

- Ramsay, M. (2013) Genetic diversity in black south Africans from Soweto. *BMC Genomics*, **14**, 644.
28. Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A. et al. (2015) The African genome variation project shapes medical genetics in Africa. *Nature*, **517**, 327–332.
 29. Greenberg, J.H. (1972) Linguistic evidence regarding Bantu origins. *J. Afr. Hist.*, **13**, 189–216.
 30. Mitchell, P. and Lane, P. (2013) *The Oxford Handbook of African Archaeology*. Oxford University Press, Oxford, UK.
 31. Li, S., Schlebusch, C. and Jakobsson, M. (2014) Genetic variation reveals large-scale population expansion and migration during the expansion of bantu-speaking peoples. *Proc. Biol. Sci.*, **281**, 20141448.
 32. de Filippo, C., Bostoen, K., Stoneking, M. and Pakendorf, B. (2012) Bringing together linguistic and genetic evidence to test the Bantu expansion. *Proc. R. Soc. B Biol. Sci.*, **279**, 3256–3263.
 33. Pakendorf, B., de Filippo, C. and Bostoen, K. (2011) Molecular perspectives on the Bantu expansion: a synthesis. *Lang. Dyn. Change*, **1**, 50–88.
 34. Bostoen, K. (2018) *The Bantu Expansion*. Oxford Research Encyclopedia of African History. Oxford University Press, Oxford, UK.
 35. Grollemund, R., Branford, S., Bostoen, K., Meade, A., Venditti, C. and Pagel, M. (2015) Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proc. Natl. Acad. Sci. U. S. A.*, **112**, 13296–13301.
 36. Patin, E., Lopez, M., Grollemund, R., Verdu, P., Harmant, C., Quach, H., Laval, G., Perry, G.H., Barreiro, L.B., Froment, A. et al. (2017) Dispersals and genetic adaptation of bantu-speaking populations in Africa and North America. *Science*, **356**, 543–546.
 37. Semo, A., Gayà-Vidal, M., Fortes-Lima, C., Alard, B., Oliveira, S., Almeida, J., Prista, A., Damasceno, A., Fehn, A.M., Schlebusch, C. and Rocha, J. (2020) Along the Indian Ocean coast: genomic variation in Mozambique provides new insights into the Bantu expansion. *Mol. Biol. Evol.*, **37**, 406–416.
 38. Choudhury, A., Aron, S., Botigué, L.R., Sengupta, D., Botha, G., Bensellak, T., Wells, G., Kumuthini, J., Shriner, D., Fakim, Y.J. et al. (2020) High-depth genome sequencing in diverse African populations informs migration history and human health. *Nature*, **586**, 741–748.
 39. Busby, G.B., Band, G., Si Le, Q., Jallow, M., Bougama, E., Mangano, V.D., Amenga-Etego, L.N., Enimil, A., Apinjoh, T., Ndila, C.M. et al. (2016) Admixture into and within sub-Saharan Africa. *elife*, **5**, e15266.
 40. Huffman, T.N. (2007) *Handbook to the Iron Age: The Archaeology of Pre-colonial Farming Societies in Southern Africa*. University of KwaZulu-Natal Press, Durban, South Africa.
 41. Huffman, T.N. (2006) Bantu migrations in southern Africa. In Soodyall, H. (ed), *The prehistory of Africa: Tracing the lineage of modern man*. Jonathon Ball, Johannesburg, South Africa, pp. 97–108.
 42. Lander, F. and Russell, T. (2018) The archaeological evidence for the appearance of pastoralism and farming in southern Africa. *PLoS One*, **13**, e0198941.
 43. Choudhury, A., Ramsay, M., Hazelhurst, S., Aron, S., Bardien, S., Botha, G., Chimusa, 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.
 44. Retshabile, G., Mlotshwa, B.C., Williams, L., Mwesigwa, S., Mboowa, G., Huang, Z., Rustagi, N., Swaminathan, S., Katagirya, E., Kyobe, S. et al. (2018) Whole-exome sequencing reveals uncaptured variation and distinct ancestry in the southern African population of Botswana. *Am. J. Hum. Genet.*, **102**, 731–743.
 45. Sengupta, D., Choudhury, A., Fortes-Lima, C., Aron, S., Whitelaw, G., Bostoen, K., Gunnink, H., Chousou-Polydouri, N., Delius, P., Tollman, S. et al. (2020) Genetic-substructure and complex demographic history of south African bantu speakers. *bioRxiv*, 2020.08.11.243840.
 46. González-Santos, M.G., Montinaro, F., Oosthuizen, O., Oosthuizen, E., Busby, G.B.J., Anagnostou, P., Destro-Bisol, G., Pascali, V. and Capelli, C. (2015) Genome-wide snp analysis of southern african populations provides new insights into the dispersal of bantu-speaking groups. *Genome Biol. Evol.*, **7**, 2560–2568.
 47. Hellenthal, G., Busby, G.B.J., Band, G., Wilson, J.F., Capelli, C., Falush, D. and Myers, S. (2014) A genetic atlas of human admixture history. *Science*, **343**, 747–751.
 48. Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D. and Berger, B. (2013) Inferring admixture histories of human populations using linkage disequilibrium. *Genetics*, **193**, 1233–1254.
 49. Lane, A.B., Soodyall, H., Arndt, S., Ratshikhopha, M.E., Jonker, E., Freeman, C., Young, L., Morar, B. and Toffie, L. (2002) Genetic substructure in south African bantu-speakers: evidence from autosomal DNA and Y-chromosome studies. *Am. J. Phys. Anthropol.*, **119**, 175–185.
 50. Goldberg, A. and Rosenberg, N.A. (2015) Beyond 2/3 and 1/3: the complex signatures of sex-biased admixture on the X chromosome. *Genetics*, **201**, 263–279.
 51. Schlebusch, C.M. (2010) *Genetic Variation in Khoisan-Speaking Populations from southern Africa*. University of the Witwatersrand Johannesburg, South Africa.
 52. 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.
 53. 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.
 54. Barbieri, C., Hübner, A., Macholdt, E., Ni, S., Lippold, S., Schröder, R., Mpoloka, S.W., Purps, J., Roewer, L., Stoneking, M. and Pakendorf, B. (2016) Refining the Y chromosome phylogeny with southern African sequences. *Hum. Genet.*, **135**, 541–553.
 55. 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.
 56. Marks, S.J., Montinaro, F., Levy, H., Brisighelli, F., Ferri, G., Bertoni, S., Batini, C., Busby, G.B., Arthur, C., Mitchell, P. et al. (2015) Static and moving frontiers: the genetic landscape of southern African bantu-speaking populations. *Mol. Biol. Evol.*, **32**, 29–43.
 57. 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.

58. 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. The role of matrilineality in shaping patterns of Y chromosome and mtDNA sequence variation in southwestern Angola. *Eur. J. Hum. Genet.*, **27**, 475–483.
59. Montinaro, F., Davies, J. and Capelli, C. (2016) Group membership, geography and shared ancestry: genetic variation in the Basotho of Lesotho. *Am. J. Phys. Anthropol.*, **160**, 156–161.
60. 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.
61. 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**, 6484.
62. 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.