

Genetics and Southern African History

Cristian Capelli and Francesco Montinaro

Summary

Southern Africa past is constellated by a series of demographic events tracing back to the dawn of our species, approximately 300 thousand years ago. The intricate pattern of population movements over the millennia contributed to create an exceptional level of diversity, which is reflected into the genomic variability of Southern African groups. Although a complete characterisation of the demographic history of the subcontinent is still lacking, several decades of extensive research have contributed to shed light to the main events.

Population movements, and the introduction of new cultural features, characterise the history of southern Africa in the last five millennia and had a dramatic impact on the subcontinental genetic variability. Traces of these migrations can be identified using different genetic systems, revealing a complex history of adaptation to new selective pressures and sex biased admixture.

The historical events of the European Colonization and the Slave Trade of the last millennium, and the emergence of new cultural groups, further increased the genomic variability of human populations in this region, one of the most genetically diverse in the world.

Keywords: southern Africa genetic history, Khoe-San structure, pastoralism in southern Africa, “Bantu expansion”, colonial era genomic impact.

The genetic variability of Southern Africa

Southern Africa displays a variety of different environments distributed across several ecological and climatic backgrounds which encompass the arid Kalahari and Namib deserts, the tropical savanna of Angola and Mozambique and the humid rainforests of Madagascar. This biodiversity is mirrored by the rich cultural variation which characterises the populations living in this area. Such diversity is well exemplified by approximately one hundred different languages spoken across the whole region,¹ whose inhabitants count to more than 170 million (~17% of all sub-Saharan Africa). While anthropologists and linguists have for long time studied populations living in this region, very little was known about the genetic variability characterizing these groups until recently. In the last few years several works focused on the human groups from this area and have provided a detailed description of the distribution of variation across groups and the demographic events that have shaped such distribution. In the worldwide analysis of human population using protein polymorphisms, Cavalli-Sforza and co-workers identified Khoe-San, Somali and Mbuti (East Africa rain-forest hunter-gatherers, also known as East African Pygmies) as early diverging populations within sub-Saharan Africa. Further analysis of the genetic distances underlying the population-tree showed that a subset of Khoe-San populations differed from other sub-Saharan groups more than any sub-Saharan group from each other. Khoe-San populations varied in their affinity to other groups, with some being closer to Bantu speaking groups, a result that was interpreted as the effect of gene-flow, in agreement with the fact that these included Coloured individuals, who have West-African ancestry² (in line with other investigations focusing on the genetic variation of Southern African populations, in using the term Coloured here we note that this term is used by the South African government “to monitor progress in moving away from the apartheid-based discrimination of the past. However, membership of a population group is now based on self-perception and self-classification, not on a legal definition”. We also acknowledge that this term may have a derogatory connotation in certain cultures, but this is certainly not intended here)

Subsequent analyses based on Y chromosome and mitochondrial DNA (mtDNA), which are inherited only from one parent (the Y chromosome from the father and the mtDNA from the mother), provided insights into sex-specific population dynamics. Most of these early works focused on populations from Southern Africa which are often identified as “Khoe-San”. The term Khoe-San has been used to label a rather heterogeneous group of people with the assumption of

¹Harald Hammarström and Sebastian Nordhoff, “LangDoc: Bibliographic Infrastructure for Linguistic Typology,” *Oslo Studies in Language* 3, no. 2 (June 2011), <https://www.journals.uio.no/index.php/osla/article/download/75/199>.

²Henry Harpending, “The History and Geography of Human Genes. L. Luca Cavalli-Sforza, Paolo Menozzi, Alberto Piazza,” *The Quarterly Review of Biology* 72 (September 1997), <https://doi.org/10.1086/419930>.

shared biological, linguistic and/or cultural traits³. Linguistically, this grouping was broadly justified by the presence in the region of several languages enriched in clicks sounds, but recent investigations have suggested the existence of 3-5 different lineages⁴. Taking this in consideration, the term Khoe-San will be used here when referring to the linguistic lineages identified by Güldemann, and apply the term Khoe-San referring to the people currently living in Southern Africa who speak languages belonging to one of the three largest click-rich linguistic families (Tuu, K'xa and Khoe-Kwadi) and showing genetic affinity with groups living in the region before the arrival of Bantu speaking communities.

The early molecular investigations showed that Khoe-San-speaking populations were characterised by high within population variation and high degree of differentiation from other African and non-African populations. These evidences generated the popular view of the Khoe-San speaking Bushmen as the “earliest people” and suggested southern Africa as the place for the emergence of *H. sapiens* (the term Bushmen is used without any derogatory meaning; we note that often people in Namibia and other parts of Southern Africa refer to themselves as Bushmen). Both these views are incorrect. First, all humans descend from the same most recent common ancestor, and therefore have the same *evolutionary age*. Second, humans did not necessarily originated in a single place (Eastern Africa has been generally proposed as the possible origin for our species), but the emergence of our species probably involved gene flow between different populations across the African continent, possibly including other *Homo* populations⁵.

More recent investigations have been exploring the richness of information offered by the variation at genome level, by looking at specific positions along the DNA (SNP arrays; SNP: Single Nucleotide Polymorphisms) or by analysing the whole genome.

³Peter Mitchell, “Genetics and Southern African Prehistory: An Archaeological View,” *Journal of Anthropological Sciences = Rivista Di Antropologia: JASS / Istituto Italiano Di Antropologia* 88 (2010): 73–92.

⁴Tom Güldemann and Anne-Maria Fehn, eds., *Beyond “Khoisan”: Historical Relations in the Kalahari Basin* (Amsterdam ; Philadelphia: John Benjamins Publishing Company, 2014).

⁵M. Rosalind Harding and Gil McVean, “A Structured Ancestral Population for the Evolution of Modern Humans,” *Current Opinion in Genetics & Development* 14, no. 6 (December 2004): 667–674, <https://doi.org/10.1016/j.gde.2004.08.010>;

Joseph Lachance et al., “Evolutionary History and Adaptation from High-Coverage Whole-Genome Sequences of Diverse African Hunter-Gatherers,” *Cell* 150, no. 3 (August 3, 2012): 457–69, <https://doi.org/10.1016/j.cell.2012.07.009>; Vincent Plagnol and D. Jeffrey Wall, “Possible Ancestral Structure in Human Populations.,” *PLoS Genetics* 2 (July 2006): e105–e105, <https://doi.org/10.1371/journal.pgen.0020105>;

F. Michael Hammer et al., “Genetic Evidence for Archaic Admixture in Africa.,” *Proceedings of the National Academy of Sciences of the United States of America* 108, no. 37 (September 2011): 15123–15128, <https://doi.org/10.1073/pnas.1109300108>;

Jean-Jacques Hublin et al., “New Fossils from Jebel Irhoud, Morocco and the Pan-African Origin of Homo Sapiens.,” *Nature* 546, no. 7657 (June 2017): 289–292, <https://doi.org/10.1038/nature22336>;

PingHsun Hsieh et al., “Model-Based Analyses of Whole-Genome Data Reveal a Complex Evolutionary History Involving Archaic Introgression in Central African Pygmies.,” *Genome Research* 26 (March 2016): 291–300, <https://doi.org/10.1101/gr.196634.115>.

Placed in a continental context, Southern African populations are characterised by five different components (Figure 1), which are often shared across groups: a so-called Khoe-San component (green in Figure 1), which is modal in most Khoe-San speaking populations, and that can be present also in groups speaking other languages; a West African component (red), which is predominant in Bantu speaking populations from southern Africa and is related to the components present in the Bantu speaking groups across the continent; an eastern Africa component (orange), which is mostly present in some Khoe-San speaking groups and, finally, two non-African components, which relates to more recent European and Asian contributions (blue and purple, respectively). Interestingly, these five components appear to be associated to well-defined historical and demographic events which have characterised the history of this part of the African continent and which will be analysed in details in the following sections.

In addition, most of the ancestries described here show a high degree of heterogeneity and differentiation at a geographic local scale, highlighting the complexity of long and short-term patterns of isolation and admixture in the subcontinent.

Despite the large amount of genomic data produced in the last few years, the coverage in terms of number of individuals and populations analysed is far from homogeneous, with populations from Namibia, Botswana and South Africa representing the vast majority of the data currently available, particularly at genome-level. Nevertheless, the information based on the analysis of uniparental markers, for which more extensive datasets are available, and the recent release of data from less investigated areas have allowed scholars to provide a general overview of the regional variation and the genetic history of the subcontinent.

[Insert Figure 1] Figure 1: Southern Africa is characterised by five main genetic components.

Admixture plot illustrating the genetic composition of southern African populations. Each bar shows the individual composition in ancestry (each colour represent an ancestral component). 138,000 genome-wide markers from different Illumina arrays were used. For clarity, we show only a representative subset of southern Africa groups⁶, and also included Western⁷ (Yoruba from

⁶Carina M Schlebusch et al., “Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History,” *Science (New York, N.Y.)* 338, no. 6105 (October 19, 2012): 374–79, <https://doi.org/10.1126/science.1227721>; Desiree C Petersen et al., “Complex Patterns of Genomic Admixture within Southern Africa,” *PLoS Genetics* 9, no. 3 (2013): e1003309, <https://doi.org/10.1371/journal.pgen.1003309>; Francesco Montinaro et al., “Unravelling the Hidden Ancestry of American Admixed Populations,” *Nature Communications* 6 (March 24, 2015), <https://doi.org/10.1038/ncomms7596>; Caitlin Uren et al., “Fine-Scale Human Population Structure in Southern Africa Reflects Ecogeographic Boundaries.,” *Genetics* 204, no. 1 (September 2016): 303–314, <https://doi.org/10.1534/genetics.116.187369>.

⁷1000 Genomes Project Consortium et al., “A Global Reference for Human Genetic Variation.,” *Nature* 526, no. 7571 (October 2015): 68–74, <https://doi.org/10.1038/nature15393>.

Nigeria) and Eastern Africa⁸ (Anuak and Amhara from Ethiopia), European (British), South and Southeastern Asian (Gujarati Indians and Han chinese) groups⁹. Credits: Francesco Montinaro.

Early genetic structure in Southern Africa

One of the most differentiated component emerging from the analysis of African populations is commonly represented by groups who speak Khoe-San languages (Figure 2A), and represents the ancestral genetic component present in the southernmost area of the continent prior to the arrival of pastoralists and agriculturalists.

Nowadays, this ancestry is present in Khoe-San speaking groups in South Africa, Botswana and Namibia, but also in Bantu-speaking groups and Coloured people, as a result of the admixture events occurred in the last 2000 years.

Genetic analyses of modern populations from Southern Africa have demonstrated the existence of at least three different ancestral components (“Khoe-San components”, Figure 2B) in the area, which diverged approximately at least 20,000 years ago or possibly even much earlier¹⁰.

In details, the first Khoe-San ancestry component is common in extant populations from the area at the North of the Kalahari desert, the Ju|’hoan and the !Xun, which trace their roots to the region of modern day Namibia and Angola. Interestingly, some of the Bantu speaking populations present in this area show traces of this ancestry, suggesting the occurrence of recent episodes of gene-flow.

[Insert Figure 2] Figure 2: Khoe-San populations are characterised by three main ancestries.

A. Geographic distribution of the three Khoe-San linguistic lineages described in the text. Modified from <https://www2.hu-berlin.de/kba/languages.html>. **B.** Multi-dimensional Scaling of Khoesan-specific genetic fragments. To remove the confounding effect of recent admixture events, we considered only genetic fragments classified as Khoe-San by means of a “Local ancestry approach”,

⁸Luca Pagani et al., “Ethiopian Genetic Diversity Reveals Linguistic Stratification and Complex Influences on the Ethiopian Gene Pool,” *The American Journal of Human Genetics* 91, no. 1 (July 13, 2012): 83–96, <https://doi.org/10.1016/j.ajhg.2012.05.015>.

⁹Consortium et al., “A Global Reference for Human Genetic Variation.”

¹⁰M. Carina Schlebusch et al., “Southern African Ancient Genomes Estimate Modern Human Divergence to 350,000 to 260,000 Years Ago.,” *Science (New York, N.Y.)* 358, no. 6363 (November 2017): 652–655, <https://doi.org/10.1126/science.aao6266>; Pontus Skoglund et al., “Reconstructing Prehistoric African Population Structure.,” *Cell* 171, no. 1 (September 2017): 59–71.e21, <https://doi.org/10.1016/j.cell.2017.08.049>; Joseph K Pickrell et al., “The Genetic Prehistory of Southern Africa,” *Nature Communications* 3 (2012): 1143, <https://doi.org/10.1038/ncomms2140>; Francesco Montinaro et al., “Complex Ancient Genetic Structure and Cultural Transitions in Southern African Populations.,” *Genetics*, November 2016, <http://www.ncbi.nlm.nih.gov/pubmed/27838627>.

and estimated the genetic distance between populations¹¹. The distance matrix is shown in a Multi-dimensional scaling scatter-plot. Redrawn from¹². Credits: Alessandro Corliano' and Francesco Montinaro.

The second Khoe-San ancestry component is common to all the populations from the area in the Central Kalahari. Notably, these populations belong to all the three Khoe-San linguistic lineages mentioned above, and also include, some of the Bantu-speaking individuals previously analysed, as the result of contacts with populations characterised by these ancestral component¹³

The last Khoe-San ancestry component is mainly present in populations living on the southernmost tip of the continent, such as the Khomani, the Nama, and the Karretjie¹⁴. The Khomani are a Hunter-gatherer population speaking a language belonging to the Tuu linguistic lineage (N|uu language of the !Ui subgroup), whereas the Nama, a Khoe-Kwadi speaking group from Namibia, are characterised by a pastoralist subsistence. The Bantu-speaking and Coloured populations from South Africa and Lesotho draw most of the Khoe-San ancestry assimilated in their genomes from this group. However, the two groups show some non-negligible difference in their Khoe-San specific ancestry, possibly reflecting the overlap of differential ancient and recent admixture history.

Interestingly, multiple regression and allele frequency analyses showed a high and significant correlation between the genetic and geographic distance and the Khoe-San ancestry components in the area, suggesting the existence of Isolation by Distance dynamics, in which individuals are more likely to mate with individuals from neighbouring populations, generating a cline of allele frequencies.

In the last few years, the analysis of genetic material from archaeological remains in the continent shed new light into the origin of the ancestry related to the Khoe-San populations¹⁵. The analysis of 15 ancient African individuals from South Africa, Malawi, Tanzania and Kenya, dated between ~500 and 8,000 years before present, showed that Southern African individuals from ~2,000 years ago are genetically close to extant Khoe-San populations, pointing towards a long-

¹¹Francesco Montinaro et al., "Complex Ancient Genetic Structure and Cultural Transitions in Southern African Populations.," *Genetics*, November 2016, <http://www.ncbi.nlm.nih.gov/pubmed/27838627>.

¹²Montinaro et al.

¹³Montinaro et al.; Uren et al., "Fine-Scale Human Population Structure in Southern Africa Reflects Ecogeographic Boundaries."

¹⁴Carina M Schlebusch et al., "Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History," *Science (New York, N.Y.)* 338, no. 6105 (October 19, 2012): 374–79, <https://doi.org/10.1126/science.1227721>.

¹⁵M. Gallego Llorente et al., "Ancient Ethiopian Genome Reveals Extensive Eurasian Admixture throughout the African Continent.," *Science (New York, N.Y.)* 350, no. 6262 (November 2015): 820–822, <https://doi.org/10.1126/science.aad2879>; Schlebusch et al., "Southern African Ancient Genomes Estimate Modern Human Divergence to 350,000 to 260,000 Years Ago.," Skoglund et al., "Reconstructing Prehistoric African Population Structure."

standing genetic continuity in the area. Surprisingly, the same investigation highlighted that Southern and Eastern African Individuals from a time interval between ~8100 and ~400BP are related with Southern African Khoe-San and Tanzanian Hadza, to an extent which correlates with geography¹⁶. As an example, seven ancient individuals from Malawi, dating between 8100 and 2100BP, were characterized by a relatively low degree of heterogeneity, pointing towards demographic continuity for at least 5 millennia. However, this component appears to be currently absent in the region, which is inhabited by Bantu-speaking populations with no detectable signature of gene flow from Khoe-San groups¹⁷.

Lineages present in Khoe-San populations have been previously described as the earliest extant diverging branches in modern humans. However, it has been recently shown that ancient Southern Africans populations are more related to some Eastern (both ancient and modern) than Western African groups, which in turns show differential relationships with Southern and Eastern Africa. These results challenge the “early Southern Africa divergence model”, while supporting two equally likely models involving an “out of East Africa” or a “differential Southeast-West Admixture” scenario (Figure 3). The former postulates the existence of a basal west African group which contributed to the modern West African Mende and Yoruba populations, and corroborating the idea of an ancient structure in the continent (Figure 3A). On the other hand, the latter is even compatible with a complex pattern of migration between different sub-regions of the continent (Figure 3B).

A similar analysis based on three ancient individuals has shown that the genomic features of Stone Age individuals from South Africa closely resemble the one of modern day Southern Khoe-San populations¹⁸.

The comparison between Southern Africa Stone Age individuals and a selection of modern-day and ancient individuals pushed back the Time from most recent common ancestor (TMRCA) of our species to 260k-350k years ago. Such a deep TMRCA fits well with fossils displaying facial features compatible with early anatomically humans which have been recently dated to ~315Kya, and the presence in the same time period of modern looking (or transitional) humans present in southern Africa, as attested by the Florisbad skull¹⁹. This observations, together with the new date

¹⁶Skoglund et al., “Reconstructing Prehistoric African Population Structure.”

¹⁷Pontus Skoglund et al., “Reconstructing Prehistoric African Population Structure.,” *Cell* 171, no. 1 (September 2017): 59–71.e21, <https://doi.org/10.1016/j.cell.2017.08.049>; Bj George Busby et al., “Admixture into and within Sub-Saharan Africa.,” *ELife* 5 (June 2016), <https://doi.org/10.7554/eLife.15266>.

¹⁸Schlebusch et al., “Southern African Ancient Genomes Estimate Modern Human Divergence to 350,000 to 260,000 Years Ago.”

¹⁹Rainer Grün et al., “Direct Dating of Florisbad Hominid,” *Nature* 382, no. 6591 (August 1996): 500–501, <https://doi.org/10.1038/382500a0>.

proposed for *Homo naledi* (236,000 to 335,000 Years Old),²⁰ suggests the coexistence of multiple hominin populations at the dawn of our species.

Figure 3: Two possible models may explain the different relationship between Southern

African and Western or Eastern groups. A. The out of East Africa model postulates that after the expansion of populations from the East, Western groups (WA2) admixed with a basal African lineage (WA1). **B.** The same pattern can be explained by the presence of complex ancient genetic structure throughout the continent, which limited the contact between Southern and Western populations. Notably, different Western groups show asymmetric relationship with southern Africa (Based on Skoglund et al 2017; Credits: Alessandro Corliano', Francesco Montinaro).

Genetic evidence for a migration related to pastoralism

The arrival of pastoralism in Southern Africa around 2000 years ago has been the subject of an extensive debate across the archaeologists for long time²¹. From a genetic perspective, the presence of a signature possibly related with the arrival of sheep and goat herders was for long time elusive, suggesting a cultural rather than demic diffusion. The first evidence for such a genetic signature was initially found by looking at the variation of the Y chromosome. The Y chromosome haplotype (a haplotype is a DNA sequence defined by a specific combination of DNA variants and inherited from the same ancestor) defined by the E-M293 mutation was identified in populations from Southern (Namibia and South Africa) and Eastern Africa (Tanzania). Y chromosomes haplotypes carrying the M293 mutation were originally found at the highest frequency in the Khwe from the Caprivi strip in Namibia and the Datog from Tanzania²². The analysis of microsatellite (Short Tandem Repeats [STRs]: short DNA motifs which are tandemly repeated and that vary in number between different alleles) variation associated with this mutation was interpreted as supporting an expansion through Tanzania to southern-central Africa, independent of the dispersal of Bantu-speaking people and dating to 2,000 years ago²³.

²⁰R. Lee Berger et al., "And Pleistocene Hominin Evolution in Subequatorial Africa.," *ELife* 6 (May 2017), <https://doi.org/10.7554/eLife.24234>; Hgm Paul Dirks et al., "The Age of and Associated Sediments in the Rising Star Cave, South Africa.," *ELife* 6 (May 2017), <https://doi.org/10.7554/eLife.24231>; John Hawks et al., "New Fossil Remains of from the Lesedi Chamber, South Africa.," *ELife* 6 (May 2017), <https://doi.org/10.7554/eLife.24232>.

²¹Karim Sadr, "The Neolithic of Southern Africa," *The Journal of African History* 44, no. 2 (2003): 195–209.

²²M. Brenna Henn et al., "Y-Chromosomal Evidence of a Pastoralist Migration through Tanzania to Southern Africa.," *Proceedings of the National Academy of Sciences of the United States of America* 105, no. 31 (August 2008): 10693–10698, <https://doi.org/10.1073/pnas.0801184105>.

²³Henn et al.

More recent work conducted at genome level have confirmed this recent link between Southern and Eastern Africa and offered the opportunity to quantify this contribution in populations across Southern Africa as well as provide further indication of when such contribution have entered populations in Southern Africa. In details, harnessing the decay of admixture Linkage-Disequilibrium (LD) patterns (a measure of how the association between alleles changes in relation to their distance along chromosomes), several studies identified Western Eurasian ancestry in Southern African groups dating between 900 and 1,800 years ago²⁴. Similar signatures were also detected in Eastern African populations, dating to around 3,000 years ago. By extensive populations comparisons authors suggested that the best proxies for the Eurasian Ancestry in the South are represented by Eastern African groups. The amount of the ancestry putatively related to these ancient migration varies across populations, but is present in all the Khoe-San populations tested so far. Notably, strong signatures were detected in the Nama, a Khoe-speaking pastoralist group in Namibia, which showed the strongest contribution at 14%, while it was lower than 6% in all the other Southern African populations tested²⁵.

Although the demographic impact of this contribution from East Africa has not probably been as high from the genetic point of view as other migration events (as for example the arrival of Bantu-speaking communities), it certainly had important effects on the genetics of Southern Africa. An example is the distribution of the genetic variant linked to the ability of digesting lactose after weaning (rs145946881, also known as C₁₄₀₁₀, located in the *MCM6* gene), the disaccharide sugar present in milk. Adults are not able to digest lactose and drinking milk results particularly unpleasant, with effects ranging from flatulence to severe diarrhoea. The variant conferring the ability to digest lactose common in Eastern Africa was found also in Southern Africa (C₁₄₀₁₀) at frequencies between 20-35% in the pastoralist Nama²⁶ (Figure 4). The haplotypes associated with this variant on chromosome 2 were shared across populations in Southern Africa, and were characterized by low degree of variation and relatively extended linkage disequilibrium, genomic features usually associated with signatures of selection. In fact the Nama showed a higher than expected frequency of the C₁₄₀₁₀ allele (~35%) when considering the estimated degree of Eastern African ancestry (~10%). The genetic background on which the mutation is present in Southern

²⁴Joseph K Pickrell et al., "Ancient West Eurasian Ancestry in Southern and Eastern Africa," *Proceedings of the National Academy of Sciences of the United States of America* 111, no. 7 (February 18, 2014): 2632–37, <https://doi.org/10.1073/pnas.1313787111>; Schlebusch et al., "Southern African Ancient Genomes Estimate Modern Human Divergence to 350,000 to 260,000 Years Ago."; Skoglund et al., "Reconstructing Prehistoric African Population Structure."

²⁵Pickrell et al., "Ancient West Eurasian Ancestry in Southern and Eastern Africa."

²⁶Enrico Macholdt et al., "Tracing Pastoralist Migrations to Southern Africa with Lactase Persistence Alleles," *Current Biology: CB* 24, no. 8 (April 2014): 875–879, <https://doi.org/10.1016/j.cub.2014.03.027>; Gwenna Breton et al., "Lactase Persistence Alleles Reveal Partial East African Ancestry of Southern African Khoe Pastoralists," *Current Biology: CB* 24, no. 8 (April 14, 2014): 852–58, <https://doi.org/10.1016/j.cub.2014.02.041>.

African populations is the same as the one in Eastern Africa. Interestingly, the signal of selection in this region appears weaker in the Nama than in other African pastoralist populations, as for example the Maasai, which might be related to the relatively recent introduction of this allele in Southern African groups.

Another signature of selection related with this ancestry component refers to a polymorphism (rs1426654 in *SLC24A5* gene) associated with lighter skin colour common in Europe that is present in considerable proportion in some Khoe-San populations, ranging between 7% (Jul'hoan) and 49% (Nama)²⁷. The average allele frequency of this polymorphism in modern-day Khoe-San populations (around ~10%²⁸), is too high to be explained simply by a recent introgression following European migration, even when positive selection is taken into account. Therefore, it may be possible that this mutation, potentially originating in Eurasia, arrived with the first pastoralist populations from East-Africa before the 17th century colonialism, and has subsequently been the target of selective pressure.

[Insert Figure 4] **Figure 4:** Early contacts between Eastern African pastoralist and Southern African foragers revealed by the distribution of the Lactase persistence associated allele C₁₄₀₁₀. Allele frequency data for the C-14010 allele are from²⁹. Credits: Francesco Montinaro.

The “Bantu expansion”

The spread of the Bantu languages, belonging to the Niger-Congo family, across most of Sub-Saharan Africa, has been often referred to as one of the most dramatic events occurred in the African continent³⁰. By the end of the 20th century approximately five hundred Bantu languages (from “ba-ntu”, meaning “human being”) were spoken by ~300 million people in sub Saharan Africa, in an area larger than Europe (11 million km²)³¹.

²⁷R. Alicia Martin et al., “An Unexpectedly Complex Architecture for Skin Pigmentation in Africans,” *Cell* 171, no. 6 (November 2017): 1340–1353.e14, <https://doi.org/10.1016/j.cell.2017.11.015>.

²⁸Martin et al.

²⁹Alessia Ranciaro et al., “Genetic Origins of Lactase Persistence and the Spread of Pastoralism in Africa,” *The American Journal of Human Genetics* 94, no. 4 (April 3, 2014): 496–510, <https://doi.org/10.1016/j.ajhg.2014.02.009>; Gwenna Breton et al., “Lactase Persistence Alleles Reveal Partial East African Ancestry of Southern African Khoe Pastoralists,” *Current Biology: CB* 24, no. 8 (April 14, 2014): 852–58, <https://doi.org/10.1016/j.cub.2014.02.041>; Enrico Macholdt et al., “Tracing Pastoralist Migrations to Southern Africa with Lactase Persistence Alleles,” *Current Biology: CB* 24, no. 8 (April 2014): 875–879, <https://doi.org/10.1016/j.cub.2014.03.027>.

³⁰Jared M Diamond, *Guns, Germs, and Steel: The Fates of Human Societies* (New York: W.W. Norton & Co., 1997).

³¹Hammarström and Nordhoff, “LangDoc: Bibliographic Infrastructure for Linguistic Typology.”

Linguistic and anthropological researches indicate that this “expansion” started between 4 and 5 thousand years ago, in an area close to present-day Cameroon/Nigeria border, and involved both the movements of people (demes) and technology³² (ideas). The high similarities between Bantu languages suggest that the diffusion occurred in a relatively short period, a hypothesis supported by archaeological evidence for agro-pastoralist communities settlements in the southernmost part of the continent dated as early as 1,000-2,000 years ago.

Geneticists have joined the debate by exploring the variation present in Bantu speaking groups and their neighbouring populations³³. One interesting characteristic emerging when comparing the genetics of Bantu speaking groups is the overall similarity showed by these populations, despite their occurrence across thousands of kilometres (Figure 6). This long distance homogeneity has been interpreted as supporting the demic model for the dispersal of Bantu languages and the associated Iron and agro pastoralist technologies³⁴. Nevertheless, differences can be found among Bantu speaking groups. Some of these differences are related to local events of gene flow with neighbour populations, whereas others are caused by demographic processes associated with the dispersal.

Although the broad dynamics of the dispersal process have been uncovered, little is still known about the path of this expansion and the local dynamics of the phenomenon.

Two main linguistic hypotheses have been promoted over the years, which propose different timing of split and dispersal patterns. The “early split hypothesis” postulates that the two main linguistic branches (North West and South East) diverged at the beginning of the expansion³⁵, the Bantu speakers east of the great lakes being closely related to the one from Northern Congo. On the contrary, the “late split hypothesis” places the separation of the two branches later in time, approximately between 2-3 Kya, after a first migration towards south-east, through the Congo rainforest. The latter is the most supported scenario, according to the latest linguistic investigations, which evaluated similarities between more than 400 Bantu languages.

The dispersal from the Cameroon homeland has been described as a two-step process in the archaeological context: a first step moving from Cameroon to the surrounding areas and a second step

³²J. Vansina, “New Linguistic Evidence and ‘the Bantu Expansion,’” *The Journal of African History* 36, no. 2 (1995): 173–95; Immanuel Ness, ed., *The Encyclopedia of Global Human Migration* (John Wiley & Sons, Inc., 2013), <https://doi.org/10.1002/9781444351071>.

³³Sarah A Tishkoff et al., “The Genetic Structure and History of Africans and African Americans,” *Science (New York, N.Y.)* 324, no. 5930 (May 22, 2009): 1035–44, <https://doi.org/10.1126/science.1172257>.

³⁴Peter Mitchell, “Early Farming Communities of Southern and South-Central Africa,” *The Oxford Handbook of African Archaeology*, July 4, 2013, <https://doi.org/10.1093/oxfordhb/9780199569885.013.0045>.

³⁵Rebecca Grollemund et al., “Bantu Expansion Shows That Habitat Alters the Route and Pace of Human Dispersals,” *Proceedings of the National Academy of Sciences* 112, no. 43 (October 27, 2015): 13296–301, <https://doi.org/10.1073/pnas.1503793112>; Thomas E Currie et al., “Cultural Phylogeography of the Bantu Languages of Sub-Saharan Africa,” *Proceedings. Biological Sciences / The Royal Society* 280, no. 1762 (2013): 20130695, <https://doi.org/10.1098/rspb.2013.0695>.

from the Great Lakes region. The archaeological record points to a steady dispersal of related pottery traditions that started from the area where and at the time when Bantu languages possibly originated, ~4-5kya. Comb stamped pottery appears just south of the Bantu languages homeland around 2,600-1,500 kya, from 2,600 years ago in Gabon, 2,400-2,100 years ago at the convergence of main rivers of the Congo Basin (Imbonga tradition), and 2,300-2,000 years ago in the Republic of Congo (Ngovo group). Iron technology follows temporally the pottery, but might not have played major roles in shaping subsistence strategies³⁶. Interestingly, evidences of pearl millet (a savannah crop) appear around 2,300-2,100 years ago, and the word indicating this crop has been reconstructed in Narrow West Bantu languages³⁷. Overall the order of appearance of the various aspects of Bantu speaking societies (subsistence, metallurgy, pottery) is not clear and might be occurring at different times, but a North to South dispersal pattern has been suggested. The second dispersal step is centred on the Great Lake region and has been associated with the spread of the Eastern Bantu languages. Data from Uganda, Rwanda, Burundi and Tanzania associate the beginning of the Iron Age with the Urewe pottery tradition around 1,000 years ago. The latter dispersed rapidly, in the form of related styles in Eastern and Southern Africa. Linguistically, the Proto-Great Lake Bantu languages have been shown to have references for yams and goats, and later for cattle, miller and sorghum, which might have been received from Sudanic speakers in the region³⁸. Most of the pottery in south-central and southern Africa (*thick wares*) is related to the Urewe ceramics, which, together with the Kalundu tradition, has been suggested to be related to the Chifumbaze Complex (Figure 5). Both Urewe and Kalundu pottery makers were probably Eastern-Bantu speakers. Urewe-related pottery travelled south along two routes, one, represented by the Mwabulambo and Nkope traditions, travelled more internally through Malawi (4th century) and later Zimbabwe; the other, represented by the Kwale pottery, moved at similar times along the coast. In the west, the presence of the Naviundu tradition (a very different pottery style from the Urewe and Kalundu) signal the arrival of Western-Bantu speakers in northern Namibia, Botswana, and western Zambia (Figure 5). The second branch of the Chifumbaze Complex represented by the Kalundu pottery, originating in central Angola in the early first-millennium CE, moved from West to South across Zambia and Botswana, reaching South Africa,

³⁶ Pierre de Maret, "Archaeologies of the Bantu Expansion," July 4, 2013, <http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199569885.001.0001/oxfordhb-9780199569885-e-43>.

³⁷Stefanie Kahlheber, Koen Bostoen, and Katharina Neumann, "Early Plant Cultivation in the Central African Rain Forest: First Millennium BC Pearl Millet from South Cameroon," *Journal of African Archaeology* 7, no. 2 (2009): 253–72. Kahlheber, Bostoen, and Neumann.

³⁸Christopher Ehret, *An African Classical Age: Eastern and Southern Africa in World History 1000 BC to AD 400*, 1st Edition edition (Charlottesville; Oxford: University of Virginia Press, 2001). Ehret.

with the earliest ceramic evidences dating to the 6th-/8th-century in Limpopo and southeastern Botswana³⁹.

[Insert Figure 5 here] Figure 5: Possible dispersal routes of Early Farming cultures in sub-Equatorial Africa. Redrawn from Mitchell and Lane 2013. Credits: Alessssandro Corliano'

DNA based investigations have suggested some degree of genetic homogeneity among Bantu speakers across subcontinent, which is compatible with the hypothesis of a demic rather than only cultural diffusion. However, none of the genetic investigations conducted so far found a strong support for any of the two models based on linguistics. The comparison of the genetic (unilinear and autosomal) and geographic distances drawn on the basis of different expansion models provided contrasting results, although overall it showed a marginal support for the late split hypothesis⁴⁰.

More recently, stronger support for the late split hypothesis has been found⁴¹, and the analysis of single nucleotide polymorphisms (SNPs) and autosomal microsatellites showed that the Eastern and Southern branches are more closely related than Western ones⁴².

The recent genomic analysis of four Iron Age Individuals from South Africa (Champagne Castle, Eland Cave, Mfongosi, Newcastle) dated between 300 to 500 years ago has revealed a strong relatedness with populations from South Africa speaking Southeaster Bantu languages (Zulu, Sotho, and Xhosa). When the early vs late split model was tested, among all the Western Bantu groups, the Iron Age sample was closer to populations from modern day Angola, as expected by the late split hypothesis⁴³

³⁹ Mitchell.Peter Mitchell, "Early Farming Communities of Southern and South-Central Africa," *The Oxford Handbook of African Archaeology*, July 4, 2013, <https://doi.org/10.1093/oxfordhb/9780199569885.013.0045>.Mitchell.

⁴⁰ de Filippo et al.Cesare de Filippo et al., "Bringing Together Linguistic and Genetic Evidence to Test the Bantu Expansion," *Proceedings. Biological Sciences / The Royal Society* 279, no. 1741 (August 22, 2012): 3256–63, <https://doi.org/10.1098/rspb.2012.0318>.de Filippo et al.

⁴¹George Busby et al., "Admixture into and within Sub-Saharan Africa," *BioRxiv*, February 1, 2016, 038406, <https://doi.org/10.1101/038406>; Etienne Patin et al., "The Impact of Agricultural Emergence on the Genetic History of African Rainforest Hunter-Gatherers and Agriculturalists," *Nature Communications* 5 (February 4, 2014), <https://doi.org/10.1038/ncomms4163>.

⁴²Etienne Patin et al., "Dispersals and Genetic Adaptation of Bantu-Speaking Populations in Africa and North America.," *Science (New York, N.Y.)* 356, no. 6337 (May 2017): 543–546, <https://doi.org/10.1126/science.aal1988>; Sen Li, Carina Schlebusch, and Mattias Jakobsson, "Genetic Variation Reveals Large-Scale Population Expansion and Migration during the Expansion of Bantu-Speaking Peoples," *Proc. R. Soc. B* 281, no. 1793 (October 22, 2014): 20141448, <https://doi.org/10.1098/rspb.2014.1448>.

⁴³

However, this observation is still compatible with other models and/or the coexistence of multiple migration waves. Furthermore, it is not clear how the admixture dynamics between different Bantu speaking groups after their settlement affected the present-day pattern of variation.

The analysis of additional genetic material recovered from pre- and post-expansion ages, covering most of the Bantu-speaking sub-Saharan Africa, will be crucial to shed light on the global and local dynamics of this tremendous expansion.

Nevertheless, the genomic analysis from modern day populations provided several insights both on the genetic structure related to these Bantu speaking groups or their recent admixture dynamics.

It is well known that a series of admixture events took place during the expansion when Bantu speaking newcomers met the resident foraging groups, as shown for Rainforest and Southern African foragers⁴⁴. The degree of their interaction and the impact of these episodes of gene-flow had in shaping the current distribution of genetic variation has been only marginally explored. Similarly, multiple waves chronologically separated but occurring along similar paths have been also proposed on the basis of archaeological data. In this context, the emerging picture related to the dispersal of the Bantu-speaking communities across Africa appears as characterized by a high degree of complexity, which included overlapping diffusion waves and the admixture with local populations along the main migration routes, suggesting a genomic variation extending beyond the simplistic consensus of their suggested genetic homogeneity.

Gene-flow between Bantu speaking farmers and Rainforest Hunter-gatherers (also referred to as Pygmies), east African Nilotic and Afro-Asiatic populations⁴⁵ has been reported. In Southern Africa similar admixture events involving farmers and resident hunter-gatherers have occurred; however they did not result in a homogenous assimilation. Gene-flow occurred differently in different regions, in line with what proposed by the “frontier model” of interaction⁴⁶. In areas where the dispersal was rapid as facilitated by environmental conditions compatible with the agro-pastoralist

Schlebusch et al. M. Carina Schlebusch et al., “Southern African Ancient Genomes Estimate Modern Human Divergence to 350,000 to 260,000 Years Ago,” *Science (New York, N. Y.)* 358, no. 6363 (November 2017): 652–655, <https://doi.org/10.1126/science.aao6266>.

⁴⁴Giovanni Destro-Bisol et al., “The Analysis of Variation of MtDNA Hypervariable Region 1 Suggests That Eastern and Western Pygmies Diverged before the Bantu Expansion,” *The American Naturalist* 163, no. 2 (February 2004): 212–26, <https://doi.org/10.1086/381405>; Patin et al., “The Impact of Agricultural Emergence on the Genetic History of African Rainforest Hunter-Gatherers and Agriculturalists”; Sarah J. Marks et al., “Migration Distance Rather than Migration Rate Explains Genetic Diversity in Human Patrilocal Groups,” *Molecular Ecology* 21, no. 20 (2012): 4958–4969, <https://doi.org/10.1111/j.1365-294X.2012.05689.x>.

⁴⁵Luca Pagani et al., “Ethiopian Genetic Diversity Reveals Linguistic Stratification and Complex Influences on the Ethiopian Gene Pool,” *The American Journal of Human Genetics* 91, no. 1 (July 13, 2012): 83–96, <https://doi.org/10.1016/j.ajhg.2012.05.015>; Patin et al., “Dispersals and Genetic Adaptation of Bantu-Speaking Populations in Africa and North America.”; Skoglund et al., “Reconstructing Prehistoric African Population Structure.”

⁴⁶T. Humphreys et al., “Frontiers: Southern African Archaeology Today,” 1984, <http://usir.salford.ac.uk/7483/>.

package, limited gene-flow between foragers and farmers occurred (as for example along the Eastern part of Southern Africa, e.g. Mozambique and Malawi); on the contrary, in regions where climatic and ecological boundaries were instead present, the dispersal process stalled for prolonged periods of time, a situation that facilitated cultural and genetic exchange, as in the central part of South Africa⁴⁷. The dynamics of admixture between incoming Bantu speaking populations and the resident foraging communities were probably characterized by gender-biased gene-flow. Comparison between the estimated amounts of Khoe-San Y chromosome and mtDNA haplotypes present in Bantu speaking groups showed consistently higher maternal than paternal Khoe-San contributions, in line with ethnographic and historical evidences reporting females being more commonly accepted in Bantu speaking communities⁴⁸. Overall these episodes of gene-flow contribute to genetic variants present in resident populations and shaped at least in part the differences present across Bantu speakers. In addition to variation due to admixture from neighbouring groups, an additional source of variation was generated by the process of dispersal itself, when this was associated with bottlenecks, founder events and isolation. The main components common across all Bantu-speaking groups are in fact not similarly distributed, showing clines and differential distributions, as the result of genetic drift⁴⁹.

Beside drift and gene-flow, local demographic histories and multiple dispersal events might have additionally contributed to the current pattern of relationships among Bantu speaking groups. The Bantu speaking pastoralist groups Himba and Herero in Namibia form a genetically related group with the Khoe-Kwadi speaker Damara, but diverge from other South West Bantu speaking groups present in Namibia, like Owambo, Mbukushu and Kwangali. Linguistically all these groups speak very similar languages, all part of the so-called South Western Bantu (the languages spoken by Himba, Herero and Owambo are all part of the R subgroup; Mbukushu and Kwangali speak languages belonging to the less related K branch⁵⁰) but genetically they clearly cluster separately (Figure 6), even when admixture is controlled for. The differentiation of these two groups of populations may be caused by cultural processes, as for example a linguistic shift following different waves of Bantu speaking groups, or demographic events, the result of a distinct and possibly early dispersal of Bantu speaking populations, or the result of recent severe isolation.

⁴⁷Miguel Gonzalez-Santos et al., "Genome-Wide SNP Analysis of Southern African Populations Provides New Insights into the Dispersal of Bantu Speaking Groups," n.d.; Marks et al., "Migration Distance Rather than Migration Rate Explains Genetic Diversity in Human Patrilocal Groups"; Montinaro et al., "Complex Ancient Genetic Structure and Cultural Transitions in Southern African Populations."

⁴⁸Marks et al., "Migration Distance Rather than Migration Rate Explains Genetic Diversity in Human Patrilocal Groups."

⁴⁹Gonzalez-Santos et al., "Genome-Wide SNP Analysis of Southern African Populations Provides New Insights into the Dispersal of Bantu Speaking Groups."

⁵⁰Grollemund et al., "Bantu Expansion Shows That Habitat Alters the Route and Pace of Human Dispersals."

The analysis of the mitochondrial lineages of diverse Angolan populations suggests that the latter hypothesis is the most supported, indicating that the Herero and the Himba may have diverged from other Bantu-speaking groups in recent times, after a severe bottleneck. If such a scenario is indeed proven to be supported, this would imply a demographic event resulting in a very strong reduction of diversity affecting groups which number in the tens of thousands. However, additional investigations exploring autosomal and ancient DNA are needed to confirm and clarify the demographic history of these populations.

[Insert Figure 6] Figure 6: Bantu-speaking groups show different genetic ancestries

composition. Admixture plot illustrating the heterogeneity of Bantu speaking populations. Each bar shows the individual composition in ancestry (each colour represent an ancestral component). For clarity, we show only Bantu populations accompanied by Southern Hunter-gatherers (Ju|'hoan and !Xun)⁵¹, Western Africa⁵² (Yoruba from Nigeria) , Rain Forest Hunter Gatherers⁵³ (Mbuti from the Democratic Republic of Congo) and Eastern Africa (Anuak and Amhara from Ethiopia), European (British) and South Asian (Gujarati Indians) groups. Credits: Francesco Montinaro.

The modern era

The genetic investigation of several groups in southern Africa has provided evidence for the presence of genetic components related to extra continental influences possibly associated with the Age of Exploration and the subsequent colonial period. Some of the signatures associated with non-African sources have been explained by the introgression of Western Eurasia ancestry and by the contribution of Eastern African groups in association with the dispersal of pastoralism, but signatures of more recent contribution have been also highlighted.

The colonisation of different Southern Africa areas mostly by Portuguese, Dutch, German and English from the 16th century, and the Indian Ocean slave trade, with the deportation of slaves from South and South East Asia as well as Madagascar, generated a melting pot whose signatures can be detected in populations nowadays.

European-related ancestries have been reported in Namibia and South Africa, particularly, but not exclusively, in populations with a known history of admixture with European colonists, as for example the Basters in Namibia and several Coloured groups in South Africa.

The Basters in particular have a very unique history which traces back to the earliest colonists who settled at the Cape of Good Hope starting in 1652. These early settlers comprised Europeans and

⁵¹Schlebusch et al., "Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History"; Petersen et al., "Complex Patterns of Genomic Admixture within Southern Africa."

⁵²Consortium et al., "A Global Reference for Human Genetic Variation."

⁵³Jun Z. Li et al., "Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation," *Science* 319, no. 5866 (February 22, 2008): 1100–1104, <https://doi.org/10.1126/science.1153717>.

people from Asia and Africa, imported as slaves. The descendants of these mixed group were the founders of the Coloured people, some of which moved in 1870 to the region later to become Namibia, where they identified themselves as the Baster nation (1872). Signatures of gene-flow related to South and Southeast Asia have been found in the Coloured communities in the Western Cape, whose ancestry contribution is estimated around 10%. European, while West-African and Khoe-San ancestries made up the remaining 90% almost in equal parts. However, substantial inter-individual variation in ancestry composition was reported across self-identified Coloured people, underlying the complex demographic and cultural processes shaping Coloured identity⁵⁴. Coloured communities from other areas are characterised by similar variability, where differences in the average ancestry composition were also observed. Coloured from Wellington are characterized by lower Khoe-San ancestry (14%; ⁵⁵) than Coloured from Colesberg (33%; ⁵⁶) Cape Town (31%; ⁵⁷) or Upington (61% ; ⁵⁸), and variation was also reported for other ancestries. Wellington displayed the largest European (28%), and Asian (South and South East, both at 17%) contributions across the sampled coloured communities, while the West African ancestry was estimated as 26-22% except in Upington, where it was below 10%⁵⁹. Notably, the pattern of admixture in Coloured populations was strongly sex-biased. The Coloured from Cape Town show greatly reduced European ancestry in the X chromosome markers, which instead is enriched for Indonesian ancestry⁶⁰. Y chromosome and mtDNA analysis on the same dataset confirmed male/female imbalance contributions for Europeans (higher male contribution), and Khoe-San and India (higher female contribution⁶¹).

The Baster population has been reported to have largest Khoe-San maternal and the largest European paternal contributions across Coloured communities (both more than 90%). At genome level, the average European ancestry in the Basters is estimated at almost 50%, while the Khoe-San

⁵⁴Erika de Wit et al., "Genome-Wide Analysis of the Structure of the South African Coloured Population in the Western Cape.," *Human Genetics* 128, no. 2 (August 2010): 145–153, <https://doi.org/10.1007/s00439-010-0836-1>.

⁵⁵Desiree C Petersen et al., "Complex Patterns of Genomic Admixture within Southern Africa," *PLoS Genetics* 9, no. 3 (2013): e1003309, <https://doi.org/10.1371/journal.pgen.1003309>.

⁵⁶Schlebusch et al., "Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History."

⁵⁷Petersen et al., "Complex Patterns of Genomic Admixture within Southern Africa."

⁵⁸Michelle Daya et al., "A Panel of Ancestry Informative Markers for the Complex Five-Way Admixed South African Coloured Population.," *PloS One* 8, no. 12 (December 2013), <https://doi.org/10.1371/journal.pone.0082224>.

⁵⁹Daya et al.

⁶⁰Nick Patterson et al., "Genetic Structure of a Unique Admixed Population: Implications for Medical Research.," *Human Molecular Genetics* 19, no. 3 (February 2010): 411–419, <https://doi.org/10.1093/hmg/ddp505>.

⁶¹Luis Quintana-Murci et al., "Strong Maternal Khoisan Contribution to the South African Coloured Population: A Case of Gender-Biased Admixture," *American Journal of Human Genetics* 86, no. 4 (April 9, 2010): 611–20, <https://doi.org/10.1016/j.ajhg.2010.02.014>.

ancestry reaches more than 28%. West-African and Asian ancestries are estimated at less than 6% and more than 17% respectively⁶².

One particular source of variation in the Coloured communities which has not been addressed so far in genetic investigations has been Madagascar and the Malagasy people. The historical records for immigration in South Africa are incomplete and possibly biased towards the overall composition of the European settlers more than slaves, which makes difficult the task of estimating the demographic impact of the slave trade in general, and the contributions from Madagascar in particular.

Slaves were initially imported from Angola and Dahomey (West Africa) and later from Mozambique⁶³. The majority of the Indian Ocean slaves were traded through private slave traders⁶⁴, from outposts such as the Sunda Islands, Moluccas, Ceylon, India and Bengal⁶⁵.

Records from 1672 to 1682 give an account of five batches of slaves totalling 569 Malagasy slaves⁶⁶ while between 1680 and 1731 half of the imported slaves were from Madagascar while the Indian and Indonesian slaves comprised a third of the total⁶⁷. The contribution from Madagascar was certainly substantial though it was not possible to estimate the contribution from isotope analysis⁶⁸.

As an indication, in 1753, 26% of slaves were from the east coast of Africa, 25% from Madagascar, 26% from South Asia 26% and 23% from the Indonesian archipelago⁶⁹. The records indicate a decline in the number of resident Malagasy and Mozambican slaves at the Cape over the years, despite the importation of a substantial number of people. Malagasy/Indonesian and South Asian ancestry in the Khoe-San populations is suggested by the presence of mitochondrial haplogroups M36 (present in Dravidian from India) and M37c3c (present in Austronesian-speaking from South East Asia) in the Nama and Khomani people and likewise in the Cape Coloured⁷⁰.

⁶²Petersen et al., "Complex Patterns of Genomic Admixture within Southern Africa."

⁶³The Afrikaners: Biography of a People - Hermann Giliomee, Hermann Buhr Giliomee (Tafelberg Publishers Limited, Cape Town, South Africa., 2018).

⁶⁴Linda Mbeki and Matthias van Rossum, "Private Slave Trade in the Dutch Indian Ocean World: A Study into the Networks and Backgrounds of the Slavers and the Enslaved in South Asia and South Africa," *Slavery & Abolition* 38, no. 1 (January 2017): 95–116, <https://doi.org/10.1080/0144039X.2016.1159004>.

⁶⁵T. Jenkins and T. G. Nurse, "Blood Group Gene Frequencies.," *South African Medical Journal = Suid-Afrikaanse Tydskrif Vir Geneeskunde* 46, no. 18 (April 1972), <http://www.ncbi.nlm.nih.gov/pubmed/5034965>.

⁶⁶Jenkins and Nurse.

⁶⁷The Afrikaners: Biography of a People - Hermann Giliomee, Hermann Buhr Giliomee.

⁶⁸M. Lisette Kootker et al., "Dynamics of Indian Ocean Slavery Revealed through Isotopic Data from the Colonial Era Cobern Street Burial Site, Cape Town, South Africa (1750-1827).," *PloS One* 11, no. 6 (June 2016), <https://doi.org/10.1371/journal.pone.0157750>.

⁶⁹Kootker et al.

⁷⁰Erika de Wit et al., "Genome-Wide Analysis of the Structure of the South African Coloured Population in the Western Cape.," *Human Genetics* 128, no. 2 (August 2010): 145–153, <https://doi.org/10.1007/s00439-010-0836-1>; Patterson et al., "Genetic Structure of a Unique Admixed Population: Implications for Medical Research."

Genome-wide SNP data have too showed the presence of Indonesian ancestry in the Cape Coloured community⁷¹, but there is no discussion on whether this is through Malagasy ancestry or through direct contributions from Indonesian exiles⁷². Overall, the role of slaves imported from Madagascar in shaping the variation present in the coloured communities is still unknown. The genetic profile of the Malagasy population is a combination of West-African and Asian ancestries, which came together 1,000-2,000 years ago⁷³.

The fact that such ancestries also came together more recently when the European colonists imported slaves directly from coastal Africa, India and Southeast Asia, makes the identification and dissection of the specific Malagasy contribution more complicated. Future work is expected to focus also on the characterisation of such additional source of variation in the populations of Southern Africa

Future perspectives

The recent attention that geneticists have given to the population of Southern Africa provided new insights into the demographic events that have shaped the human genetic variation of the subcontinent. The collection of data from areas only partially investigated so far and the increase in the amount of genomic data available for these populations is expected to provide additional details on these processes. One particular aspect of interest is the possibility that gene flow between *Homo sapiens* and other *Homo* populations might have occurred in Africa, as it did outside Africa with Neanderthals and Denisovans (the latter indicates a previously unknown hominin from the Altai mountains in Siberia and represented so far only by a handful of remains but a full genome⁷⁴). Together with more samples from extant populations, a very promising aspect is the increase in the number of human remains from which DNA has been extracted and characterised⁷⁵.

⁷¹Patterson et al., "Genetic Structure of a Unique Admixed Population: Implications for Medical Research."; de Wit et al., "Genome-Wide Analysis of the Structure of the South African Coloured Population in the Western Cape."; R. Emile Chimusa et al., "A Genomic Portrait of Haplotype Diversity and Signatures of Selection in Indigenous Southern African Populations.," PLoS Genetics 11, no. 3 (March 2015), <https://doi.org/10.1371/journal.pgen.1005052>.

⁷²Petersen et al., "Complex Patterns of Genomic Admixture within Southern Africa"; Schlebusch et al., "Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History."

⁷³Denis Pierron et al., "Genome-Wide Evidence of Austronesian–Bantu Admixture and Cultural Reversion in a Hunter-Gatherer Group of Madagascar," *Proceedings of the National Academy of Sciences* 111, no. 3 (January 21, 2014): 936–41, <https://doi.org/10.1073/pnas.1321860111>.

⁷⁴

Skoglund et al., "Reconstructing Prehistoric African Population Structure."

⁷⁵Alan G. Morris et al., "First Ancient Mitochondrial Human Genome from a Prepastoralist Southern African," *Genome Biology and Evolution* 6, no. 10 (September 10, 2014): 2647–53, <https://doi.org/10.1093/gbe/evu202>; Schlebusch et al., "Southern African Ancient Genomes Estimate

The molecular investigation of ancient human remains is certainly a promising direction of research expected to contribute to our understanding the history of Southern Africa. Similar work done in Europe so far has revolutionised the consensus picture of the ancient history of the continent. A similar impact can be envisioned for the African continent once more and more samples will be analysed. Current limitations in our ability of retrieving genetic material due to DNA degradations will be possibly bypassed by additional technical breakthrough, which will make possible to push the boundaries of what is currently considered feasible, as it has been the case over the last decades.

While more extensive molecular analysis of ancient African samples is surely to come, an exhaustive survey of continental variation via ancient samples might be difficult to reach given the smaller density of human archaeological remains available for DNA analysis across the continent, when compared to Europe; as such, an approach combining data from ancient and modern samples is expected to be very beneficial to the process of understanding the history of Southern Africa and overall of the African continent. Africa is still a part of the world for which whole genomes are still a few, but international efforts are in place to correct this situation⁷⁶.

As more genome data is accumulated and novel bio-statistical approaches are developed, our ability to recover the fine details of the demographic and evolutionary process that have shaped the variation of human populations is expected to improve, and in doing so, complement and enrich the work by fellow archaeologists, linguists and anthropologists. Over the last few years DNA analysis has contributed to sketch the ancient and recent history behind the variation of extant populations in Southern Africa. Despite this, there are still uncertainties in several aspects of our history of the African continent. For example, is not clear if humans admixed with other *Homo* populations and the exact geographical and temporal extension of population structure among the early modern humans in Africa is unknown. Furthermore the regional pattern of admixture events from the rest of the continent is far to be complete, with only a fraction of regions and historical periods investigated.

Further Reading

Modern Human Divergence to 350,000 to 260,000 Years Ago.”; Skoglund et al., “Reconstructing Prehistoric African Population Structure.”

⁷⁶Deepti Gurdasani et al., “The African Genome Variation Project Shapes Medical Genetics in Africa,” *Nature* 517, no. 7534 (January 15, 2015): 327–32, <https://doi.org/10.1038/nature13997>; Ananyo Choudhury et al., “Whole-Genome Sequencing for an Enhanced Understanding of Genetic Variation among South Africans,” *Nature Communications* 8, no. 1 (December 2017), <https://doi.org/10.1038/s41467-017-00663-9>.

[Busby, George Bj et al. “Admixture into and within Sub-Saharan Africa.” *eLife* 5 \(2016\): n. pag. *eLife*.](#)

[Gomez, Felicia, Jibril Hirbo, and Sarah A Tishkoff. “Genetic Variation and Adaptation in Africa: Implications for Human Evolution and Disease.” *Cold Spring Harbor perspectives in biology* 6.7 \(2014\): a008524. *Cold Spring Harbor perspectives in biology*.](#)

[“Hunters and Herders Southern Africa Comparative Ethnography Khoisan Peoples | Social and Cultural Anthropology | Cambridge University Press.” N.p., 16 Feb. 2018.](#)

[Gurdasani, Deepti et al. “The African Genome Variation Project Shapes Medical Genetics in Africa.” *Nature* 517.7534 \(2015\): 327–332. *Nature*.](#)

[Patin, Etienne et al. “Dispersals and Genetic Adaptation of Bantu-Speaking Populations in Africa and North America.” *Science \(New York, N.Y.\)* 356.6337 \(2017\): 543–546. *Science \(New York, N.Y.\)*.](#)

[Pickrell, Joseph K, Nick Patterson, Po-Ru Loh, et al. “Ancient West Eurasian Ancestry in Southern and Eastern Africa.” *Proceedings of the National Academy of Sciences of the United States of America* 111.7 \(2014\): 2632–2637. *Proceedings of the National Academy of Sciences of the United States of America*.](#)

[Pickrell, Joseph K, Nick Patterson, Chiara Barbieri, et al. “The Genetic Prehistory of Southern Africa.” *Nature communications* 3 \(2012\): 1143. *Nature communications*.](#)

[Schlebusch, Carina M et al. “Southern African Ancient Genomes Estimate Modern Human Divergence to 350,000 to 260,000 Years Ago.” *Science \(New York, N.Y.\)* 358.6363 \(2017\): 652–655. *Science \(New York, N.Y.\)*.](#)

[Skoglund, Pontus et al. “Reconstructing Prehistoric African Population Structure.” *Cell* 171.1 \(2017\): 59–71.e21. *Cell*.](#)

[“The Encyclopedia of Global Human Migration - Wiley Online Library.” N.p., 16 Feb. 2018.](#)

[Vansina, J. “New Linguistic Evidence and ‘The Bantu Expansion.’” *The Journal of African History*](#)

Acknowledgments

The authors would like to thank the funding bodies which supported their work on African populations and the several people who contributed to the realisation of these projects over the years⁷⁷

Notes

Berger, Lee R. et al. “And Pleistocene Hominin Evolution in Subequatorial Africa.” *eLife* 6 (2017): n. pag. *eLife*.

Breton, Gwenna et al. “Lactase Persistence Alleles Reveal Partial East African Ancestry of Southern African Khoe Pastoralists.” *Current biology: CB* 24.8 (2014): 852–858. *Current biology: CB*.

Chimusa, Emile R et al. “A Genomic Portrait of Haplotype Diversity and Signatures of Selection in Indigenous Southern African Populations.” *PLoS genetics* 11.3 (2015): e1005052. *PLoS genetics*.

Choudhury, Ananyo et al. “Whole-Genome Sequencing for an Enhanced Understanding of Genetic Variation among South Africans.” *Nature communications* 8.1 (2017): 2062. *Nature*

⁷⁷The authors would like to thank in particular St Hugh’s College and the Leverhulme Trust (“The genetic landscape of Southern Africa human populations”; RPG-2013-298) for the support. Further support has been received from the Boise Trust Fund and The John Fell Fund in Oxford, and The Wenner-Gren foundation. Authors would also like to additionally acknowledge the contributions of Miguel Gonzales-Santos, who provided relevant references for the Bantu dispersal section; Ryan Daniels, who gave suggestions for the contribution of Malagasy to South Africa genetic variation, and Alessandro Corliano for the help provided in the editing of the figures. CC would like to thank all the students, post-docs and collaborators that over the last ten years contributed to the research into the genetic variation of Southern African populations and particularly the people who contributed to the realisation of these projects by donating their DNA. The ministries of Health of Namibia, Lesotho and South Africa, and the Lesotho Ministries of Local Government, Tourism and Environment are also acknowledged for their support.

communications.

Crawford, Nicholas G et al. “Loci Associated with Skin Pigmentation Identified in African Populations.” *Science (New York, N.Y.)* 358.6365 (2017): n. pag. *Science (New York, N.Y.)*.

Currie, Thomas E et al. “Cultural Phylogeography of the Bantu Languages of Sub-Saharan Africa.” *Proceedings. Biological sciences* 280.1762 (2013): 20130695. *Proceedings. Biological sciences*.

Daya, Michelle et al. “A Panel of Ancestry Informative Markers for the Complex Five-Way Admixed South African Coloured Population.” *PloS one* 8.12 (2013): e82224. *PloS one*.

de Filippo, Cesare et al. “Bringing Together Linguistic and Genetic Evidence to Test the Bantu Expansion.” *Proceedings. Biological sciences* 279.1741 (2012): 3256–3263. *Proceedings. Biological sciences*.

de Wit, Erika et al. “Genome-Wide Analysis of the Structure of the South African Coloured Population in the Western Cape.” *Human genetics* 128.2 (2010): 145–153. *Human genetics*.

Diamond, Jared. *Guns, Germs and Steel: A Short History of Everybody for the Last 13,000 Years*. Vintage, 1997.

Dirks, Paul Hgm et al. “The Age of and Associated Sediments in the Rising Star Cave, South Africa.” *eLife* 6 (2017): n. pag. *eLife*.

Gallego Llorente, M. et al. “Ancient Ethiopian Genome Reveals Extensive Eurasian Admixture throughout the African Continent.” *Science (New York, N.Y.)* 350.6262 (2015): 820–822. *Science (New York, N.Y.)*.

González-Santos, Miguel et al. “Genome-Wide SNP Analysis of Southern African Populations Provides New Insights into the Dispersal of Bantu-Speaking Groups.” *Genome biology and evolution* 7.9 (2015): 2560–2568. *Genome biology and evolution*.

Grollemund, Rebecca et al. “Bantu Expansion Shows That Habitat Alters the Route and Pace of Human Dispersals.” *Proceedings of the National Academy of Sciences of the United States of America* 112.43 (2015): 13296–13301. *Proceedings of the National Academy of Sciences of the United States of America*.

Güldemann, Tom, and Anne-Maria Fehn, eds. *Beyond “Khoisan”: Historical Relations in the Kalahari Basin*. John Benjamins Publishing Company, 2014. John Benjamins Publishing Company.

Gurdasani, Deepti et al. “The African Genome Variation Project Shapes Medical Genetics in Africa.” *Nature* 517.7534 (2015): 327–332. *Nature*.

Hammarström, Harald, and Sebastian Nordhoff. “LangDoc: Bibliographic Infrastructure for Linguistic Typology.” *Oslo Studies in Language* 3.2 (2011): n. pag. *Oslo Studies in Language*.

Hammer, Michael F. et al. “Genetic Evidence for Archaic Admixture in Africa.” *Proceedings of the National Academy of Sciences of the United States of America* 108.37 (2011): 15123–15128. *Proceedings of the National Academy of Sciences of the United States of America*.

Harding, Rosalind M, and Gil McVean. “A Structured Ancestral Population for the Evolution of Modern Humans.” *Current opinion in genetics & development* 14.6 (2004): 667–674. *Current opinion in genetics & development*.

Hawks, John et al. “New Fossil Remains of from the Lesedi Chamber, South Africa.” *eLife* 6 (2017): n. pag. *eLife*.

Henn, Brenna M et al. “Y-Chromosomal Evidence of a Pastoralist Migration through Tanzania to Southern Africa.” *Proceedings of the National Academy of Sciences of the United States of America* 105.31 (2008): 10693–10698. *Proceedings of the National Academy of Sciences of the United States of America*.

Hsieh, PingHsun et al. “Model-Based Analyses of Whole-Genome Data Reveal a Complex Evolutionary History Involving Archaic Introgression in Central African Pygmies.” *Genome research* 26 (2016): 291–300. *Genome research*.

Hublin, Jean-Jacques et al. “New Fossils from Jebel Irhoud, Morocco and the Pan-African Origin of Homo Sapiens.” *Nature* 546.7657 (2017): 289–292. *Nature*. Web. 7 June 2017.

Humphreys, T. et al. “Frontiers: Southern African Archaeology Today.” (1984): n. pag. *British Archaeological Reports*.

Jenkins, T., and G. T Nurse. “Blood Group Gene Frequencies.” *South African medical journal* = *Suid-Afrikaanse tydskrif vir geneeskunde* 46.18 (1972): 560. *South African medical journal* = *Suid-Afrikaanse tydskrif vir geneeskunde*.

Kootker, Lisette M et al. “Dynamics of Indian Ocean Slavery Revealed through Isotopic Data from the Colonial Era Cobern Street Burial Site, Cape Town, South Africa (1750-1827).” *PloS one* 11.6 (2016): e0157750. *PloS one*.

Lachance, Joseph et al. “Evolutionary History and Adaptation from High-Coverage Whole-Genome Sequences of Diverse African Hunter-Gatherers.” *Cell* 150.3 (2012): 457–469. *Cell*.

Li, Sen, Carina Schlebusch, and Mattias Jakobsson. “Genetic Variation Reveals Large-Scale Population Expansion and Migration during the Expansion of Bantu-Speaking Peoples.” *Proceedings. Biological sciences* 281.1793 (2014): n. pag. *Proceedings. Biological sciences*.

Macholdt, Enrico, Montgomery Slatkin, et al. “New Insights into the History of the C-14010 Lactase Persistence Variant in Eastern and Southern Africa.” *American journal of physical anthropology* 156.4 (2015): 661–664. *American journal of physical anthropology*.

Macholdt, Enrico, Vera Lede, et al. “Tracing Pastoralist Migrations to Southern Africa with Lactase Persistence Alleles.” *Current biology: CB* 24.8 (2014): 875–879. *Current biology: CB*.

Marks, Sarah J. et al. “Static and Moving Frontiers: The Genetic Landscape of Southern African Bantu-Speaking Populations.” *Molecular biology and evolution* 32.1 (2015): 29–43. *Molecular biology and evolution*.

Martin, Alicia R. et al. “An Unexpectedly Complex Architecture for Skin Pigmentation in Africans.” *Cell* 171.6 (2017): 1340–1353.e14. *Cell*.

Mbeki, Linda, and Matthias van Rossum. “Private Slave Trade in the Dutch Indian Ocean World: A Study into the Networks and Backgrounds of the Slavers and the Enslaved in South Asia and South Africa.” *Slavery & Abolition* 38.1 (2017): 95–116. *Slavery & Abolition*.

Mitchell, Peter. “Early Farming Communities of Southern and South-Central Africa.” *The Oxford Handbook of African Archaeology*. Oxford University Press, 2013. *The Oxford Handbook of African Archaeology*.

---. "Genetics and Southern African Prehistory: An Archaeological View." *Journal of anthropological sciences = Rivista di antropologia: JASS* 88 (2010): 73–92.

Montinaro, Francesco et al. "Complex Ancient Genetic Structure and Cultural Transitions in Southern African Populations." *Genetics* (2016): n. pag. *Genetics*.

Morris, Alan G et al. "First Ancient Mitochondrial Human Genome from a Prepastoralist Southern African." *Genome biology and evolution* 6.10 (2014): 2647–2653. *Genome biology and evolution*.

Ness, Immanuel, ed. *The Encyclopedia of Global Human Migration*. John Wiley & Sons, Inc., 2013. *John Wiley & Sons, Inc.*

Pagani, Luca et al. "Ethiopian Genetic Diversity Reveals Linguistic Stratification and Complex Influences on the Ethiopian Gene Pool." *American journal of human genetics* 91.1 (2012): 83–96. *American journal of human genetics*.

Patin, Etienne et al. "Dispersals and Genetic Adaptation of Bantu-Speaking Populations in Africa and North America." *Science (New York, N.Y.)* 356.6337 (2017): 543–546. *Science (New York, N.Y.)*.

Patterson, Nick et al. "Genetic Structure of a Unique Admixed Population: Implications for Medical Research." *Human molecular genetics* 19.3 (2010): 411–419. *Human molecular genetics*.

Petersen, Desiree C et al. "Complex Patterns of Genomic Admixture within Southern Africa." *PLoS genetics* 9.3 (2013): e1003309. *PLoS genetics*.

"Piazza, A., Menozzi, P. and Cavalli-Sforza, L.L.: The History and Geography of Human Genes (Paperback) | Princeton University Press." N.p., 13 Feb. 2018. Web. 13 Feb. 2018.

Pickrell, Joseph K, Nick Patterson, Po-Ru Loh, et al. "Ancient West Eurasian Ancestry in Southern and Eastern Africa." *Proceedings of the National Academy of Sciences of the United States of America* 111.7 (2014): 2632–2637. *Proceedings of the National Academy of Sciences of the United States of America*.

Pickrell, Joseph K, Nick Patterson, Chiara Barbieri, et al. "The Genetic Prehistory of Southern

[Africa.” *Nature communications* 3 \(2012\): 1143. *Nature communications*.](#)

[Pierron, Denis et al. “Genomic Landscape of Human Diversity across Madagascar.” *Proceedings of the National Academy of Sciences of the United States of America* 114.32 \(2017\): E6498–E6506. *Proceedings of the National Academy of Sciences of the United States of America*.](#)

[Plagnol, Vincent, and Jeffrey D Wall. “Possible Ancestral Structure in Human Populations.” *PLoS genetics* 2 \(2006\): e105–e105. *PLoS genetics*.](#)

[Quintana-Murci, Lluís et al. “Strong Maternal Khoisan Contribution to the South African Coloured Population: A Case of Gender-Biased Admixture.” *American journal of human genetics* 86.4 \(2010\): 611–620. *American journal of human genetics*.](#)

[Schlebusch, Carina M, Pontus Skoglund, et al. “Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History.” *Science \(New York, N.Y.\)* 338.6105 \(2012\): 374–379. *Science \(New York, N.Y.\)*.](#)

[Schlebusch, Carina M, Helena Malmström, et al. “Southern African Ancient Genomes Estimate Modern Human Divergence to 350,000 to 260,000 Years Ago.” *Science \(New York, N.Y.\)* 358.6363 \(2017\): 652–655. *Science \(New York, N.Y.\)*.](#)

[Skoglund, Pontus et al. “Reconstructing Prehistoric African Population Structure.” *Cell* 171.1 \(2017\): 59–71.e21. *Cell*.](#)

[The Afrikaners: Biography of a People - Hermann Giliomee, Hermann Buhr Giliomee. Tafelberg Publishers Limited, Cape Town, South Africa., 2018.](#)

[Uren, Caitlin et al. “Fine-Scale Human Population Structure in Southern Africa Reflects Ecogeographic Boundaries.” *Genetics* 204.1 \(2016\): 303–314. *Genetics*.](#)

[Vansina, J. “New Linguistic Evidence and ‘The Bantu Expansion.’” *The Journal of African History* 36.02 \(1995\): n. pag. *The Journal of African History*.](#)