"The evolutionary history of Southern Africa"

Francesco Montinaro[1,2] and Cristian Capelli[1]

1.Department of Zoology, University of Oxford, UK;

2.Estonian Biocentre, Institute of Genomics, University of Tartu, Tartu, Estonia

Contacts: francesco.montinaro@gmail.com, cristian.capelli@zoo.ox.ac.uk

Abstract

The genomic variability of Southern African groups is characterised by an exceptional degree of diversity, which is the result of long-term local evolutionary history, migrations and gene-flow. Over the last few years several investigations have characterized the signatures of these processes, revealing how ancient and more recent events have shaped the structure and ancestry composition of local populations. Here we discuss recent insights into the genetic history of the Southernmost part of the African continent provided by the analysis of modern and ancient genomes. Future work is expected to clarify the population dynamics associated with the emergence of *H. sapiens* across Africa and the details of the process of dispersion and admixture associated with the arrival of Bantu-speaking groups in the region.

Keywords: Bantu-speaking populations; Khoe-San; Southern Africa history; human origin.

Southern Africa, encompassing an area of ~6.5 million km^2 and including the 4th largest island in the world [1], is characterised by substantial ecosystemic biodiversity, mirrored by the exceptional degree of cultural and genetic diversity displayed by the human populations present in this region. Work conducted in the last few years has provided significant insights into the genetic makeup of these populations and has contributed to the understanding of their evolutionary history.

The genetic structure of Southern Africa

The genetic profiles of Southern African populations combine a few genetic components which are highly differentiated and that can be broadly identified using unsupervised clustering approaches as for example implemented in ADMIXTURE ([2], Figure 1). These different components can be linked to the different strata of genetic variation that have been arriving and merging in the region with the resident communities over the last few millennia and beyond. Asian and European ancestries can be detected at different frequencies across Southern African groups, tracing their origin back to the last 400 years of colonial presence (orange and light blue in Figure 1). These ancestries are common in groups known to have experienced historical episodes of gene-flow [3]–[7], as the Coloured people in South Africa and the Basters nation in Namibia, with substantial variation observed within and between groups (we note that several populations from Namibia and South Africa use the words Coloured and Baster when referring to themselves and such terms are used here without any derogatory connotation).

Additional ancestry components in Southern African populations relate to events extending further back in time. The most common African components detected across Southern groups occur in almost all the populations of the region (red and yellow, Figure 1). These

components mostly characterise Bantu-speaking groups and are shared with West African Niger-Congo languages speaking populations, such as (but not only) the Yoruba (Nigeria) and the Mende (Sierra Leone). The ancestry associated with these two components has been linked to the dispersal of Bantu-speaking groups from Central-West Africa 4-5,000 years ago [8], [9], whose arrival in Southern Africa has been dated as early as the 4th century CE [10]. A more localised distribution characterises the components modal in most of the Khoe-San speaking groups (Dark blue and purple in Figure 1; the term Khoe-San is used here to refer to people currently living in Southern Africa who speak languages belonging to one of the three African click-rich linguistic families: Tuu, K'xa and Khoe-Kwadi; [11]). These components reflect the long term presence of human communities in the regions and represent a deeper level of population structure dating to tens of thousands of years ago [12], [13]. Some Khoe-San speaking groups are characterised by an additional ancestry component shared with populations from Eastern Africa (violet and seagreen in Figure 1 [14]). Such ancestry has been associated with the arrival of groups from East Africa and possibly linked to the spread of pastoralism in the region approximately 2,000 years ago [3], [15]–[17], as suggested by the analysis of archaeological remains. This movement/diffusion, probably starting around 3,000 years ago from the Eastern part of the continent and brought Eurasian ancestry in Southern African populations 1.5 millennia before the colonial occupation. The genetic contribution of this component in the region is well exemplified by the distribution of the $C_{14010,}$ mutation located in the MCM6 gene (rs145946881), a variant that has been associated with lactose tolerance (see below, [18]–[20]).

This mutation is at its highest frequency in the pastoralist Nama of Namibia, who are also the group with the largest contribution of the East African ancestry. Although archaeologically this migration has been sometimes linked with the origin and/or arrival of the Khoe-Kwadi in

Southern Africa, no specific association of this ancestry with this linguistic family has been found. Furthermore, the demographic impact of this influence might be limited.

Although there is a strong correlation between linguistic affiliation and genetic ancestries, all these components are present across linguistic and cultural boundaries, emphasising the role of past and recent gene-flow in shaping local variation [21] (Figure 1). In addition, both West-African and Khoe-San ancestries can be further subdivided as the result of their evolutionary history.

Ancient population structure in Southern Africa

Nowadays, the Khoe-San components (blue and purple in Figure 1) are present in Khoe-San speaking groups in South Africa, Botswana and Namibia, but also in Bantu-speaking groups and Coloured people , as the result of the admixture events occurred in the last millennium. Extensive genetic investigation have highlighted the high heterogeneity and long term population size [22] of the Khoe-San groups, possibly summarizing different pre-Bantu elements which emerged locally over the last 50-20 ky years or more as the combination of gene-flow and isolation-by-distance [3], [12], [13], [16]. These different components are usually referred to as Northern, Central and Southern Khoe-San, which reflect their high correlation with geography but not with languages and other cultural affiliations. More generally, the relationship among genes, cultures and languages is far from being simple in Southern Africa, the current pattern of genetic and cultural variation being shaped by different combinations of cultural and demographic dynamics, including extensive gene-flow and cultural switches. Further variation has been suggested within the Southern Khoe-San, with some degree of differentiation being observed between the Khoe-San ancestry of admixed population of South Africa [12]

The analysis of genetic material extracted from archaeological remains in sub-Saharan Africa helped to clarify the origin of the early Southern Africa populations [23]–[25]. The analysis of ~2,000 years old Southern African individuals highlighted their strong affinity to extant Khoe-San populations, suggesting long-term genetic continuity in the area. Surprisingly, Southern and Eastern African Individuals from ~8100 to ~400 years ago show genetic affinity with Southern African Khoe-San and Tanzanian Hadza, to an extent that correlates with geography [25]. Ancient individuals from Malawi spanning a chronological transect of ~5,000 years (dated between 8100 and 2100BP), were characterized by a relatively low degree of heterogeneity and a substantial proportion of Khoe-San related components which appears to be absent/negligible in nowadays Bantu-speaking populations resident in the area, [25], [26] an observation suggesting population replacement.

Khoe-San populations harbour the earliest extant diverging branches in modern humans together with substantial degree of variation, in line with the so-called "early Southern Africa divergence model" for the early history of human populations [27]

However, such a scenario has been recently challenged by the analysis of DNA retrieved from ancient remains, which demonstrated that ancient Southern Africans resulted closer to some of the ancient and modern Eastern Africans more than Western African groups. These results equally supported two alternative models: the so-called "out of East Africa" model or the "differential South East - West Admixture" model [23], [25]. The former postulates the existence of a basal west African group which contributed to the modern West African Mende and Yoruba populations, and corroborates the idea of an ancient structure in the continent (Figure 2A). On the other hand, the alternative model is compatible with a complex pattern of migration between different sub-region of the continent, which would fall within a metapopulation scenario for the origin of *H. sapiens* (Figure 2B; [28]). The complexity of the

emergence of modern humans is further highlighted by the recent comparison between Southern Africa Stone Age individuals, lacking the confounding impact of recent admixture, with a selection of modern-day and ancient individuals which pushed back the time from the most recent common ancestor of our species to 260-350k years ago and the identification of fossils compatible with early anatomically recently dated to ~315 Kya in Northern Africa [29] . These observations all provide support for a deeper and more complex evolutionary history for our species, and the characterisation of the demographic dynamics associated with this process is area of interest for future research.

The dispersal of Bantu-speaking populations within southern Africa

The West African ancestry present in the vast majority of the people present in southern Africa can be classified in several components, each having high frequencies in Central-West, East, South-West and South-East Bantu speaking groups, respectively ([30] yellow and red, Figure 1). These components show limited differentiation from each other and possibly originated relatively recently as the result of demographic processes (bottlenecks and founder effects) associated with the dispersal of Bantu speaking communities from Central-West Africa [30], [31].

Linguistic and anthropological researches indicate that the dispersal of the Niger Congo Bantu languages, started approximately 4 - 5 thousand years ago in an area close to present-day Cameroon/Nigeria border. The high affinity of Bantu languages suggests that the diffusion took place in a short period, as further supported by archeological evidence, which records agro-pastoralist communities in Southern Africa as early as 1,000-2,000 years ago.

Two main models for this dispersal have been proposed, differing for timing of split and dispersal patterns. The “early split hypothesis” speculates that the two main linguistic

branches (North West and South East) diverged at the beginning of the expansion, [32], [33]. On the contrary, the “late split hypothesis”, which is the most supported scenario according to linguistic surveys [32], [33], places the separation of the two branches later in time, approximately between 2-3 Kya. On the other hand, from the archaeological record, a higher complexity emerge, and a two phase process has been proposed, involving i) a quick dispersal in the Central-west region where the Bantu languages possibly originated and ii) and a second diffusion started from the Great Lake region and associated with the Eastern and Southern Bantu lineages.

Genetic analyses highlighted a substantial homogeneity among Bantu groups supporting a demic rather than cultural diffusion although no definitive evidence for one of the the proposed models has been found so far. The analysis of autosomal single nucleotide polymorphisms and short tandem repeats microsatellite showed a closer affinity between Eastern and Southern branches rather than Western ones [34], [35]. It should be noted here that the dynamics affecting genes and languages might be different and operate on different temporal frameworks, and these differences should be properly taken in account when attempting to reconcile the two. The recent genomic analysis of four Southern African Iron Age Individuals revealed that, among all the modern-day Western Bantu groups, the ancient samples were closer to populations from modern day Angola, as expected by the late split hypothesis, although it does not rule out alternative models.

The analysis of ancient genetic material recovered from pre- and post-expansion ages, covering most of the Bantu-speaking sub-saharan African regions, will be essential to further characterise this diffusion .

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 foragers. 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 [21], [36]–[39] of their suggested genetic homogeneity.

The dynamics of admixture between incoming Bantu speaking populations and the resident foraging communities were probably characterized by gender-biased gene-flow [36], [37].

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 Khwe-Kwadi speaker Damara, but diverge from other South West Bantu groups present in Namibia, like Owambo, Mbukushu and Kwangali. The analysis of the mitochondrial lineages of diverse Angolan populations indicates that this group may have diverged from other Bantu 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 ten of thousands today. However, additional investigations exploring autosomal and ancient DNA are needed to confirm and clarify the demographic history of these populations.

Genomic Signatures of Adaptation

The exceptional degree of variation and gene-flow observed in Southern Africa offers the chance to explore the role played by selective pressure in shaping the genomes of Southern African populations, which may eventually shed light on the relationship between genetic and phenotypic variation, including medically relevant traits (Ronald and Akey 2005). Signatures of selection have been detected in relation to phenotypes involved with diet, physical characteristics and diseases, including responses to pathogens

Diet

One of the best example of adaptation in the region is Lactase Persistence , controlled by at least three different alleles in sub-Saharan Africa and showing strong evidence of natural selection [40], [41]. Among these, the allele known as $C_{14010,}$ located in the *MCM6* gene, is present at high frequency in different populations from Kenya and Tanzania, while reaches frequencies up to ~35% in the pastoralist Nama of Namibia, well above the estimated contribution from Eastern African (~10%), from which originated. Notably, this allele has been linked to the arrival of pastoralists in Southern Africa around ~2,000 years ago [18]–[20], [40]. The variant has not been found in four Iron-Age individuals from South Africa dated around 500 ya, and is not common across Southern african Bantu-speaking groups, suggesting different selective pressures in these populations, possibly also influenced by cultural behaviours (e.g. milk fermentation, [23]). Similarly, a polymorphism associated with lighter skin color common in European populations (rs1426654 in *SLC24A5* gene) is present in considerable proportion in some Khoe-San populations, ranging between 7% (Jul'hoan) and 40% (Nama). The average frequency of this allele in modern-day Khoe-San populations (which has been estimated to be around ~10%) is too high to be explained by recent European migration, even when positive selection is taken into account [42].

Therefore, it may be possible that this mutation arrived with first pastoralist populations from East-Africa before the 17th century colonialism, and had subsequently been the target of selective pressure [42], [43]. Although limited on the comparison of two ancient and six modern complete genomes, a scan for selection identified 5 regions with significantly high differentiation (Z-score > 6.6, [25]). Interestingly the top identified region encompasses eight taste-receptor genes on chromosome 12. Taste-receptor genes have previously been identified as potential targets of selective pressure, given their role in the detection of poisonous compounds in plants.

Physical traits

Among the former, a recent analysis based on a novel approach, in !Xun and Ju|'hoan population, identified 43 regions encompassing 60 genes, some of them previously associated with decreased body mass index (*GAD2*) and hearing (*MYO3A*). Notably, the set of genes appear to be enriched for warfarin, intercellular adhesion Molecule-1, body fat distribution and body mass index, Varicose veins, Carotid stenosis, Cholesterol (HDL) ,Alzheimer Disease and Parkinson disease, blood cells and Erythrocyte indices, and benzodiazepines (dbGaP dataset, all $p < 0.05$, analysis performed using enrichR [44], [45]).

Lastly, a score enrichment analysis revealed that, compared to ancient Southern Africa, present day Khoe-San have a significantly different allele frequency differentiation for the gene category response to radiation, possibly the results of adaptation to the life in the kalahari "refugium" for the last few thousand years [25].

Diseases and Pathogens

The role of pathogens in shaping the genomic variation via selection has been reported in several populations across the world [46]. In the last few years, several genes have been associated with resistance to Malaria and sleeping sickness (*DARC*, *HBB*, *G6PD*, *ATP2B4*,

and *APOL1 [47]–[49]*), and potentially representing targets of strong selective pressure. A recent survey based on local ancestry showed that Malagasy populations harbour a very long (61 Mb) tract in chromosome 1 putatively selected in response to malaria[50].

Notably, variants in the Duffy null allele (*DARC*, rs2814778) associated with Malaria resistance were also found in three South African Iron age individuals, but not in the Stone Age samples older than 1800 ya, which instead harboured less protective alleles in the same gene. These observations suggest that protective malaria variants reached Southern Africa mostly with Iron Age farmers , and were not present at high frequency in local Stone Age populations. Such observations are compatible with a scenario where selective pressure caused by the plasmodium infection started as a consequences of the introduction of domesticated plants and animal, and the subsequent increase of population density.

Lastly, genomic scans for selection have identified many candidate genes associated with selection among Southern Africa HG and farmers [3], [25], [27], [51]–[54]. Among the former, a recent analysis based on a novel approach, in !Xun and Jul'hoan population, identified 43 regions encompassing 60 genes, some of them previously associated with decreased body mass index (*GAD2*) and hearing (*MYO3A*). Notably, the set of genes appear to be enriched for warfarin, intercellular adhesion Molecule-1, body fat distribution and body mass index, Varicose veins, Carotid stenosis, Cholesterol (HDL), Alzheimer Disease and Parkinson disease, blood cells and Erythrocyte indices, and benzodiazepines (dbGaP dataset, all $p < 0.05$, analysis performed using enrichR [44], [45]).

A similar analysis in Khomani identified four regions with significant excess of West African ancestry (encompassing 17 genes) and three with a reduced European ancestry associated with malaria and other relevant conditions such as colorectal and lung cancer.

A scan for highly differentiated markers between Yoruba and amaXhosa identified a putatively selected SNP (rs11118642 in *HLX* gene) associated to Fryns syndrome, Hernia and acute myeloid leukemia [51].

Although limited on the comparison of two ancient and six modern complete genomes, a scan for selection identified 5 regions with significantly high differentiation (Z-score > 6.6, [25]). Interestingly the top identified region encompasses eight taste-receptor genes on chromosome 12. Taste-receptor genes have previously been identified as potential targets of selective pressure, given their role in the detection of poisonous compounds in plants. Lastly, a score enrichment analysis revealed that, compared to ancient Southern Africa, present day Khoe-San have a significantly different allele frequency differentiation for the gene category response to radiation, possibly the results of adaptation to the life in the kalahari "refugium" for the last few thousand years [25]. Lastly, a recent analysis found that selection signals enriched for genes associated with obesity and metabolism [55]

Conclusions

The steady increase in the number of available Southern African genomes [31], [52], [56] , coupled with the molecular analysis of ancient samples provided relevant insights about both recent and early population dynamics in the sub-continent, including potential adapted genes which may be relevant under a medical perspective. However, despite the recent efforts of the scientific community, genomic studies of Southern Africa populations are under-represented limiting our knowledge of the multi-layered evolutionary dynamics of this key area. In particular the issue of Archaic introgression in Africa requires more genome data and the continuous refinement of bio-statistical approaches designed to confidently identify introgressed genomic regions [57], [58]. The continuous expansion in the number of ancient

genomes, while challenging due their antiquity, state of preservation and numerosity of available material, is expected to provide more direct insights in the complex dynamics of migration and admixture that have characterized the region The analyses of both ancient and modern data is expected to overcome their respective limitations and be very fruitful in the fine characterization of the evolutionary history of the region. The biostatistical efforts that have identified several signatures of selection are now expected to be followed by the detailed characterization of the molecular basis of the suggested genotype/phenotype associations. Future work will focus on the validation of the functional aspects of these signals.

Acknowledgments

The authors would like to thank St Hugh's College, the Leverhulme Trust ("The genetic landscape of Southern Africa human populations"; RPG-2013-298), the Wenner-Gren foundation, the Boise Trust Fund and The John Fell Fund in Oxford for supporting their work in Southern Africa; The ministries of Health of Namibia, Lesotho and South Africa, and the Lesotho Ministries of Local Government, Tourism and Environment; Alessandro Corliano' for his invaluable help in figures editing; Davide Marnetto for his help in the interpretation of the enrichment analysis; Luca Pagani and Ludovica Molinaro for providing comments on early version of the manuscript; all the students, researchers and collaborators that over the last ten years contributed to the projects focusing on the genetic variation of Southern African populations and finally all the people who contributed to the realisation of these projects by donating their DNA.

The authors declare no conflict of interest.

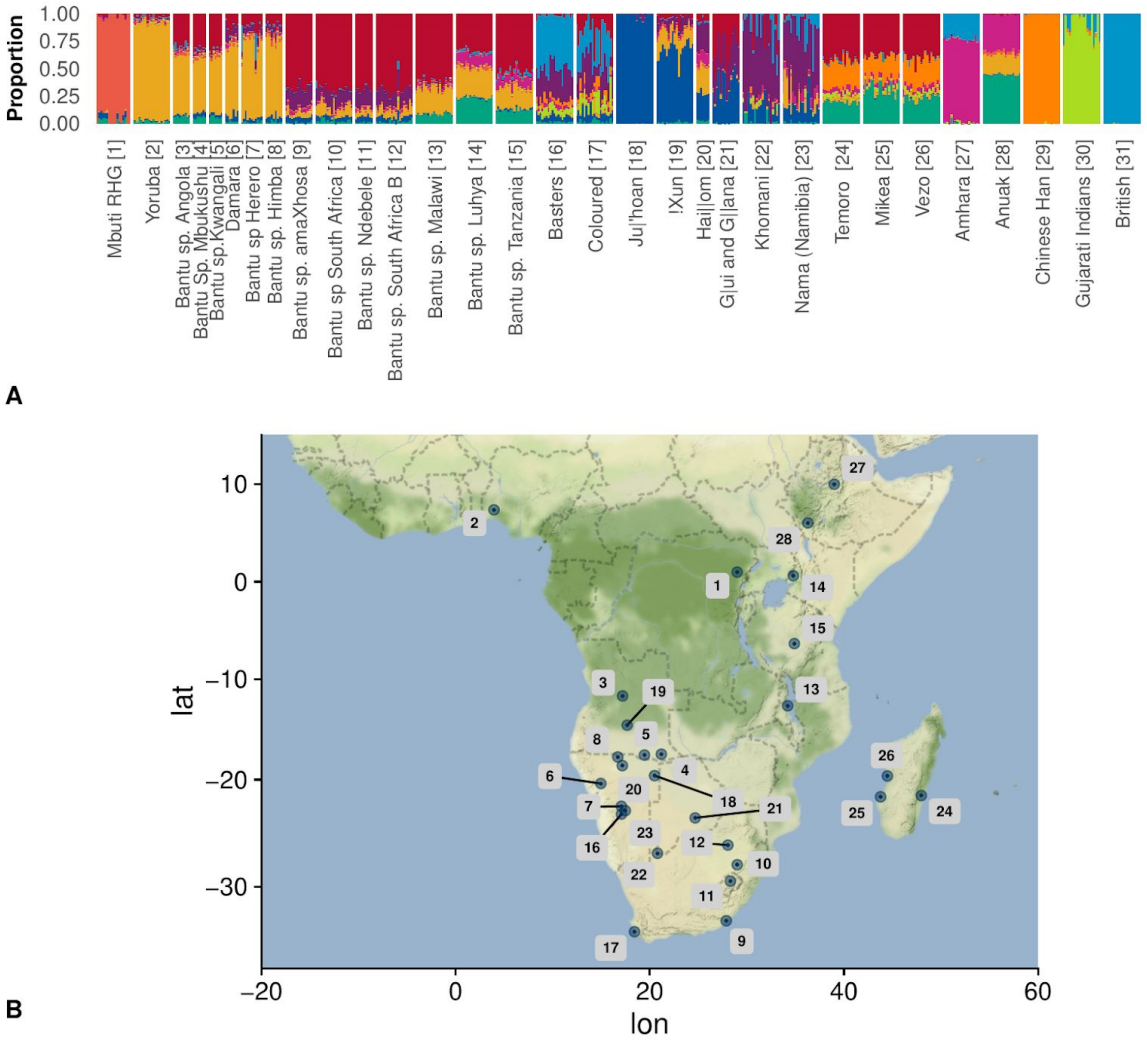

Figure 1: The genetic structure of southern Africa: A. Admixture plot of a set of 31 Southern African and other genetically relevant sub Saharan and Eurasian populations[5], [12], [26], [59]–[63]. We analysed ~138, 000 LD-pruned autosomal markers, performed 10 ADMIXTURE runs for each K and selected K=10, characterised by the lowest cross-validation error. Different runs were combined using CLUMMP and *"distruct by many Ks"[64], [65]*. Numbers in square brackets refer to geographic location in the map. **B.** Geographic location of the analysed African populations.

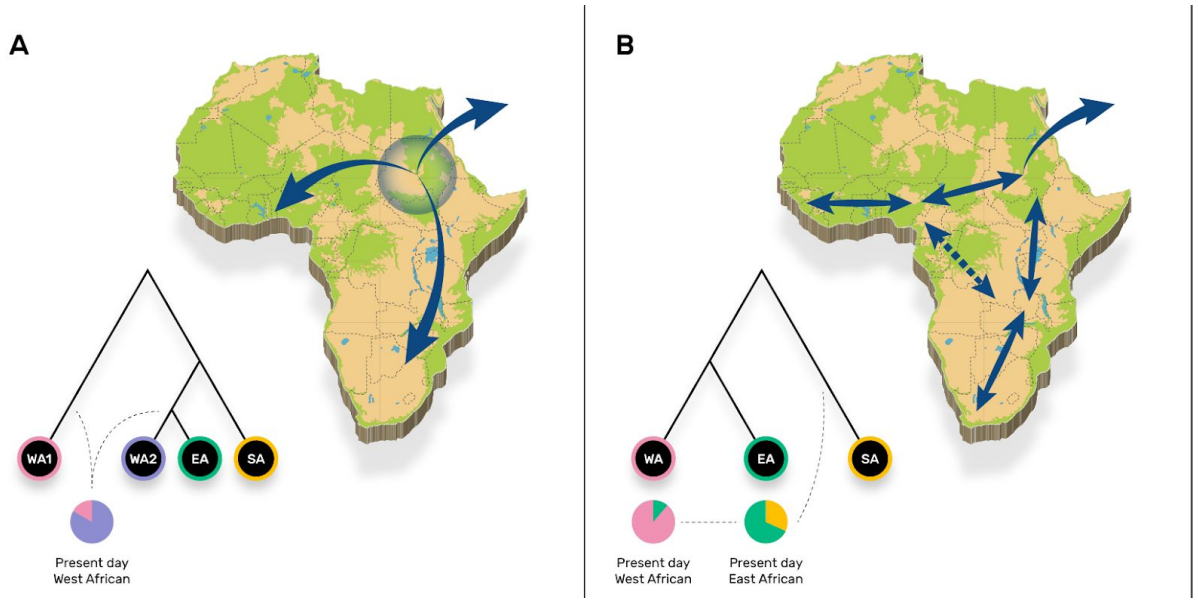

Figure 2: Proposed evolutionary models for African genetic structure. ***A.)*** Western Africa groups have ancestry from a basal western African lineage (WA1). The major source of western African ancestry (WA2) is more related to eastern Africans (EA) and non-Africans than Southern African Khoe-San (SA).
B.) West Africa populations have gene flow from a population related to both southern and eastern Africa, supporting a more complex pattern of isolation-by-distance. (Redrawn from Skoglund et al. 2017).

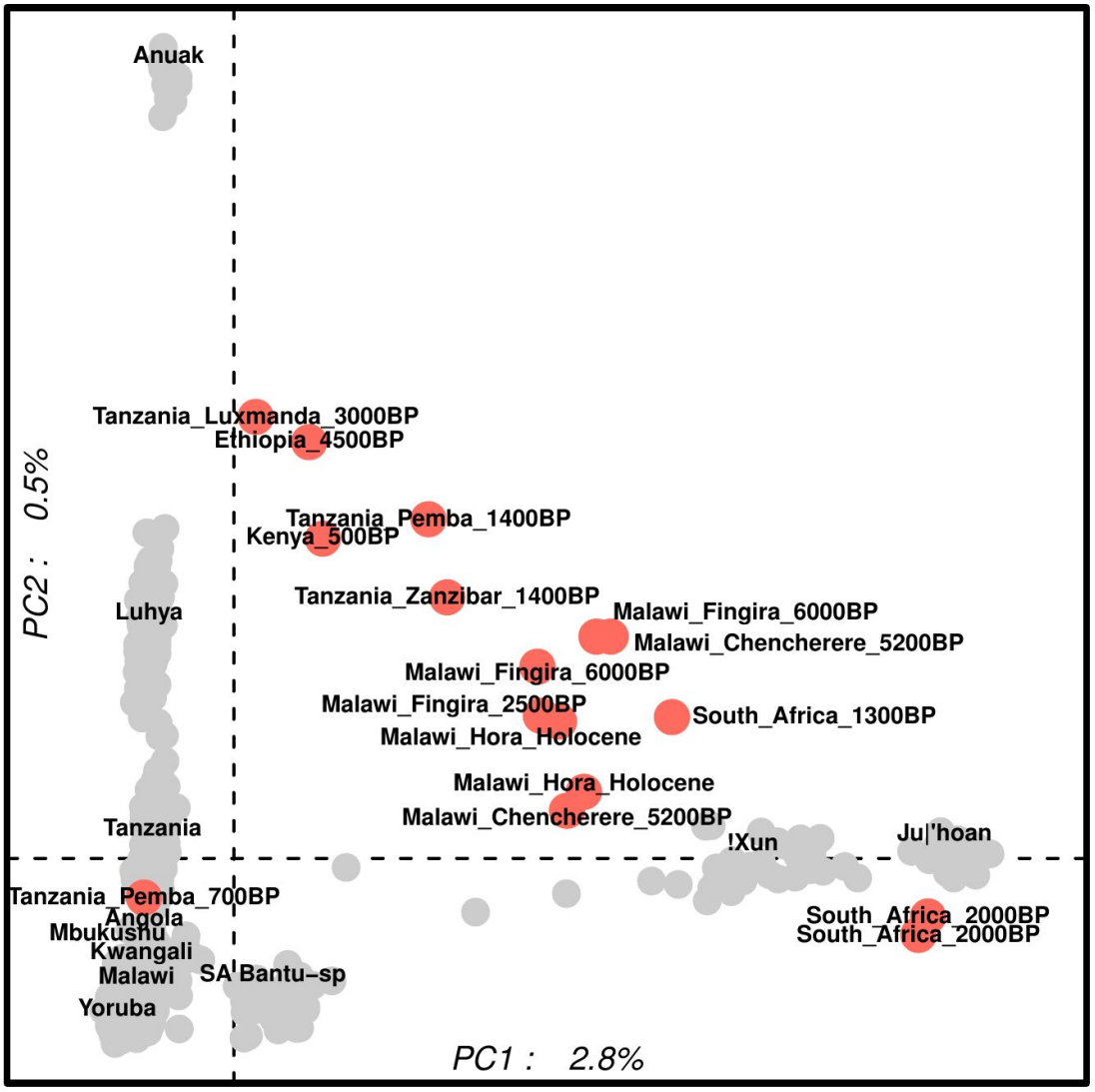

Figure 3: The analysis of ancient individuals from Southern Africa reveals the existence of an early East-South genetic cline. Principal Component Analysis of 16 individuals[24], [25] projected onto the PCs defined by 615 individuals [3], [12], [13], [62], [66] from 10 Sub-saharan populations. We analysed ~138,000 LD-pruned autosomal markers and performed the analysis using smartpca tool implemented in EIGENSTRAT software [67].

Annotated references

Choudhury2017

Pilot study for The Southern African Human Genome Programme, presenting high quality genomes for 24 South African individuals and identifying potential targets of natural selection.

Schlebusch 2017

Genomic analysis of seven ancient South African individuals revealing complex admixture dynamics in the area and providing an estimated date of 260-350 kya for the deepest split among human populations

Skoglund 2017

This genomic analysis of 16 ancient individuals from Southern and Eastern informs on the early population structure and history of Africa

Busby 2016

Genome wide analysis of 48 sub-saharan groups that uncovers the full admixture history of the continent over the last 4,000 years.

Dirks 2017

Indirect and direct dating of the remains of Homo naledi, uncovering the relatively young age of the fossils of the newly discovered species (236-335 kya)

Martin 2017

Evolutionary and genetic study of skin pigmentation in ~500 african Khoe-San individuals, revealing adaptation and uncovering new variants associated with this trait.

Crawford 2017

Genome wide association analysis of 1530 African individuals, which identified variants significantly associated with skin pigmentation and provided hints of their physiological basis through functional analyses.

Mallick 2016

This study of 300 genomes from 142 populations provides a global overview of worldwide genetic structure and early human evolution.

References:

[1] "World Bank Open Data | Data," 29-Apr-2018. [Online]. Available: https://data.worldbank.org/. [Accessed: 29-Apr-2018].

[2] D. H. Alexander, J. Novembre, and K. Lange, "Fast model-based estimation of ancestry in unrelated individuals.," *Genome Res.*, vol. 19, no. 9, pp. 1655–1664, Oct. 2009.

[3] C. M. Schlebusch *et al.*, "Genomic variation in seven Khoe-San groups reveals adaptation and complex African history.," *Sci.*, vol. 338, no. 6105, pp. 374–379, Oct. 2012.

[4] E. de Wit *et al.*, "Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape.," *Hum. Genet.*, vol. 128, no. 2, pp. 145–153, Aug. 2010.

[5] D. C. Petersen *et al.*, "Complex patterns of genomic admixture within southern Africa.," *PLoS Genet.*, vol. 9, no. 3, p. e1003309, Mar. 2013.

[6] N. Patterson *et al.*, "Genetic structure of a unique admixed population: implications for

medical research.," *Hum. Mol. Genet.*, vol. 19, no. 3, pp. 411–419, Feb. 2010.
[7] M. Daya *et al.*, "A panel of ancestry informative markers for the complex five-way admixed South African coloured population.," *PloS one*, vol. 8, no. 12, p. e82224, Dec. 2013.
[8] S. A. Tishkoff *et al.*, "The genetic structure and history of Africans and African Americans.," *Sci.*, vol. 324, no. 5930, pp. 1035–1044, May 2009.
[9] P. de Maret, "Archaeologies of the Bantu Expansion," in *The Oxford Handbook of African Archaeology*, Oxford University Press, 2013.
[10] P. Mitchell, "Early Farming Communities of Southern and South-Central Africa," in *The Oxford Handbook of African Archaeology*, Oxford University Press, 2013.
[11] T. Güldemann and A.-M. Fehn, Eds., *Beyond "Khoisan": Historical relations in the Kalahari Basin*. John Benjamins Publishing Company, 2014.
[12] F. Montinaro *et al.*, "Complex Ancient Genetic Structure and Cultural Transitions in Southern African Populations.," *Genetics*, vol. 205, no. 1, pp. 303–316, Jan. 2017.
[13] C. Uren *et al.*, "Fine-Scale Human Population Structure in Southern Africa Reflects Ecogeographic Boundaries.," *Genetics*, vol. 204, no. 1, pp. 303–314, Sep. 2016.
[14] L. Molinaro and L. Pagani, "The evolutionary history of Eastern Africa," *Curr. Opinion*. .
[15] B. M. Henn *et al.*, "Y-chromosomal evidence of a pastoralist migration through Tanzania to southern Africa.," *Proc. Natl. Acad. Sci. United States Am.*, vol. 105, no. 31, pp. 10693–10698, Aug. 2008.
[16] J. K. Pickrell *et al.*, "The genetic prehistory of southern Africa.," *Nat. Commun.*, vol. 3, p. 1143, 2012.
[17] J. K. Pickrell *et al.*, "Ancient west Eurasian ancestry in southern and eastern Africa.," *Proc. Natl. Acad. Sci. United States Am.*, vol. 111, no. 7, pp. 2632–2637, Feb. 2014.
[18] G. Breton, C. M. Schlebusch, M. Lombard, P. Sjödin, H. Soodyall, and M. Jakobsson, "Lactase persistence alleles reveal partial East African ancestry of southern African Khoe pastoralists.," *Curr. Biol. : CB*, vol. 24, no. 8, pp. 852–858, Apr. 2014.
[19] E. Macholdt, M. Slatkin, B. Pakendorf, and M. Stoneking, "New insights into the history of the C-14010 lactase persistence variant in Eastern and Southern Africa.," *Am. J. Phys. Anthropol.*, vol. 156, no. 4, pp. 661–664, Apr. 2015.
[20] E. Macholdt *et al.*, "Tracing pastoralist migrations to southern Africa with lactase persistence alleles.," *Curr. Biol. : CB*, vol. 24, no. 8, pp. 875–879, Apr. 2014.
[21] S. J. Marks *et al.*, "Static and moving frontiers: the genetic landscape of Southern African Bantu-speaking populations.," *Mol. Biol. Evol.*, vol. 32, no. 1, pp. 29–43, Jan. 2015.
[22] H. L. Kim, A. Ratan, G. H. Perry, A. Montenegro, W. Miller, and S. C. Schuster, "Khoisan hunter-gatherers have been the largest population throughout most of modern-human demographic history.," *Nat. Commun.*, vol. 5, p. 5692, Dec. 2014.
[23] C. M. Schlebusch *et al.*, "Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago.," *Sci.*, vol. 358, no. 6363, pp. 652–655, Nov. 2017.
[24] M. Gallego Llorente *et al.*, "Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent.," *Sci.*, vol. 350, no. 6262, pp. 820–822, Nov. 2015.
[25] P. Skoglund *et al.*, "Reconstructing Prehistoric African Population Structure.," *Cell*, vol. 171, no. 1, p. 59–71.e21, Sep. 2017.
[26] G. B. Busby *et al.*, "Admixture into and within sub-Saharan Africa.," *eLife*, vol. 5, Jun.

2016.
[27] B. M. Henn *et al.*, “Hunter-gatherer genomic diversity suggests a southern African origin for modern humans.,” *Proc. Natl. Acad. Sci. United States Am.*, vol. 108, no. 13, pp. 5154–5162, Mar. 2011.
[28] R. M. Harding and G. McVean, “A structured ancestral population for the evolution of modern humans.,” *Curr. Opin. Genet. & Dev.*, vol. 14, no. 6, pp. 667–674, Dec. 2004.
[29] J.-J. Hublin *et al.*, “New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens.,” *Nature*, vol. 546, no. 7657, pp. 289–292, Jun. 2017.
[30] M. González-Santos *et al.*, “Genome-Wide SNP Analysis of Southern African Populations Provides New Insights into the Dispersal of Bantu-Speaking Groups.,” *Genome Biol. Evol.*, vol. 7, no. 9, pp. 2560–2568, Sep. 2015.
[31] A. Choudhury *et al.*, “Whole-genome sequencing for an enhanced understanding of genetic variation among South Africans.,” *Nat. Commun.*, vol. 8, no. 1, p. 2062, Dec. 2017.
[32] R. Grollemund, S. Branford, K. Bostoen, A. Meade, C. Venditti, and M. Pagel, “Bantu expansion shows that habitat alters the route and pace of human dispersals.,” *Proc. Natl. Acad. Sci. United States Am.*, vol. 112, no. 43, pp. 13296–13301, Oct. 2015.
[33] T. E. Currie, A. Meade, M. Guillon, and R. Mace, “Cultural phylogeography of the Bantu Languages of sub-Saharan Africa.,” *Proceedings. Biol. Sci.*, vol. 280, no. 1762, p. 20130695, Jul. 2013.
[34] E. Patin *et al.*, “Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America.,” *Sci.*, vol. 356, no. 6337, pp. 543–546, May 2017.
[35] S. Li, C. Schlebusch, and M. Jakobsson, “Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples.,” *Proceedings. Biol. Sci.*, vol. 281, no. 1793, Oct. 2014.
[36] S. J. Marks, H. Levy, C. Martinez-Cadenas, F. Montinaro, and C. Capelli, “Migration distance rather than migration rate explains genetic diversity in human patrilocal groups.,” *Mol. Ecol.*, vol. 21, no. 20, pp. 4958–4969, Oct. 2012.
[37] C. de Filippo, K. Bostoen, M. Stoneking, and B. Pakendorf, “Bringing together linguistic and genetic evidence to test the Bantu expansion.,” *Proceedings. Biol. Sci.*, vol. 279, no. 1741, pp. 3256–3263, Aug. 2012.
[38] C. Barbieri, A. Butthof, K. Bostoen, and B. Pakendorf, “Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia.,” *Eur. J. Hum. Genet. : EJHG*, vol. 21, no. 4, pp. 430–436, Apr. 2013.
[39] C. Barbieri *et al.*, “Migration and interaction in a contact zone: mtDNA variation among Bantu-speakers in Southern Africa.,” *PloS one*, vol. 9, no. 6, p. e99117, Jun. 2014.
[40] A. Ranciaro *et al.*, “Genetic origins of lactase persistence and the spread of pastoralism in Africa.,” *Am. J. Hum. Genet.*, vol. 94, no. 4, pp. 496–510, Apr. 2014.
[41] S. A. Tishkoff *et al.*, “Convergent adaptation of human lactase persistence in Africa and Europe.,” *Nat. Genet.*, vol. 39, no. 1, pp. 31–40, Jan. 2007.
[42] A. R. Martin *et al.*, “An Unexpectedly Complex Architecture for Skin Pigmentation in Africans.,” *Cell*, vol. 171, no. 6, p. 1340–1353.e14, Nov. 2017.
[43] N. G. Crawford *et al.*, “Loci associated with skin pigmentation identified in African populations.,” *Sci.*, vol. 358, no. 6365, Nov. 2017.
[44] E. Y. Chen *et al.*, “Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.,” *BMC Bioinforma.*, vol. 14, p. 128, Apr. 2013.
[45] M. V. Kuleshov *et al.*, “Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.,” *Nucleic acids Res.*, vol. 44, no. W1, pp. W90–W97, Jul. 2016.

[46] M. Fumagalli, M. Sironi, U. Pozzoli, A. Ferrer-Admetlla, L. Pattini, and R. Nielsen, "Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution," *PLoS Genet.*, vol. 7, no. 11, p. 1002355, 2011.
[47] R. E. Howes *et al.*, "The global distribution of the Duffy blood group.," *Nat. Commun.*, vol. 2, p. 266, 2011.
[48] G. Bedu-Addo, S. Meese, and F. P. Mockenhaupt, "An ATP2B4 polymorphism protects against malaria in pregnancy.," *J. Infect. Dis.*, vol. 207, no. 10, pp. 1600–1603, May 2013.
[49] K. F. McManus, A. M. Taravella, B. M. Henn, C. D. Bustamante, M. Sikora, and O. E. Cornejo, "Population genetic analysis of the DARC locus (Duffy) reveals adaptation from standing variation associated with malaria resistance in humans.," *PLoS Genet.*, vol. 13, no. 3, p. e1006560, Mar. 2017.
[50] D. Pierron *et al.*, "Strong selection during the last millennium for African ancestry in the admixed population of Madagascar.," *Nat. Commun.*, vol. 9, no. 1, p. 932, Mar. 2018.
[51] E. R. Chimusa *et al.*, "A genomic portrait of haplotype diversity and signatures of selection in indigenous southern African populations.," *PLoS Genet.*, vol. 11, no. 3, p. e1005052, Mar. 2015.
[52] D. Gurdasani *et al.*, "The African Genome Variation Project shapes medical genetics in Africa.," *Nature*, vol. 517, no. 7534, pp. 327–332, Jan. 2015.
[53] K. Rockett *et al.*, "Inferring adaptive gene-flow in recent African history," *bioRxiv*, Jan. 2017.
[54] K. A. Owers, P. Sjödin, C. M. Schlebusch, P. Skoglund, H. Soodyall, and M. Jakobsson, "Adaptation to infectious disease exposure in indigenous Southern African populations.," *Proceedings. Biol. Sci.*, vol. 284, no. 1852, Apr. 2017.
[55] L. A. Sugden, E. G. Atkinson, A. P. Fischer, S. Rong, B. M. Henn, and S. Ramachandran, "Localization of adaptive variants in human genomes using averaged one-dependence estimation.," *Nat. Commun.*, vol. 9, no. 1, p. 703, Feb. 2018.
[56] S. Mallick *et al.*, "The Simons Genome Diversity Project: 300 genomes from 142 diverse populations.," *Nature*, vol. 538, no. 7624, pp. 201–206, Oct. 2016.
[57] S. R. Browning, B. L. Browning, Y. Zhou, S. Tucci, and J. M. Akey, "Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture.," *Cell*, vol. 173, no. 1, p. 53–61.e9, Mar. 2018.
[58] S. Sankararaman and A. Durvasula, "Recovering signals of ghost archaic admixture in the genomes of present-day Africans," *bioRxiv*, Jan. 2018.
[59] J. Z. Li *et al.*, "Worldwide human relationships inferred from genome-wide patterns of variation.," *Sci.*, vol. 319, no. 5866, pp. 1100–1104, Feb. 2008.
[60] 1000 Genomes Project Consortium *et al.*, "A global reference for human genetic variation.," *Nature*, vol. 526, no. 7571, pp. 68–74, Oct. 2015.
[61] D. Pierron *et al.*, "Genomic landscape of human diversity across Madagascar.," *Proc. Natl. Acad. Sci. United States Am.*, vol. 114, no. 32, pp. E6498–E6506, Aug. 2017.
[62] L. Pagani *et al.*, "Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool.," *Am. J. Hum. Genet.*, vol. 91, no. 1, pp. 83–96, Jul. 2012.
[63] D. Pierron *et al.*, "Genome-wide evidence of Austronesian-Bantu admixture and cultural reversion in a hunter-gatherer group of Madagascar.," *Proc. Natl. Acad. Sci. United States Am.*, vol. 111, no. 3, pp. 936–941, Jan. 2014.
[64] M. Jakobsson and N. A. Rosenberg, "CLUMPP: a cluster matching and permutation

program for dealing with label switching and multimodality in analysis of population structure.," *Bioinforma.*, vol. 23, no. 14, pp. 1801–1806, Jul. 2007.
[65] N. M. Kopelman, J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose, "Clumpak: a program for identifying clustering modes and packaging population structure inferences across K.," *Mol. Ecol. Resour.*, vol. 15, no. 5, pp. 1179–1191, Sep. 2015.
[66] G. B. J. Busby *et al.*, "The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269.," *Proceedings. Biol. Sci.*, vol. 279, no. 1730, pp. 884–892, Mar. 2012.
[67] A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich, "Principal components analysis corrects for stratification in genome-wide association studies," *Nat. Genet.*, vol. 38, no. 8, pp. 904–909, 2006.