Title

Population structure and infectious disease risk in southern Africa

Journal

Molecular Genetics and Genomics, 292(3)

ISSN

1617-4615

Authors

Uren, Caitlin
Möller, Marlo
van Helden, Paul D
et al.

Publication Date

2017-06-01

DOI

10.1007/s00438-017-1296-2

Peer reviewed

Mol Genet Genomics (2017) 292:499–509
DOI 10.1007/s00438-017-1296-2

REVIEW

Population structure and infectious disease risk in southern Africa

Caitlin Uren[1] · Marlo Möller[1] · Paul D. van Helden[1] · Brenna M. Henn[2] · Eileen G. Hoal[1]

Received: 12 August 2016 / Accepted: 1 February 2017 / Published online: 22 February 2017

Abstract The KhoeSan populations are the earliest known indigenous inhabitants of southern Africa. The relatively recent expansion of Bantu-speaking agropastoralists, as well as European colonial settlement along the south–west coast, dramatically changed patterns of genetic diversity in a region which had been largely isolated for thousands of years. Owing to this unique history, population structure in southern Africa reflects both the underlying KhoeSan genetic diversity as well as differential recent admixture. This population structure has a wide range of biomedical and sociocultural implications; such as changes in disease risk profiles. Here, we consolidate information from various population genetic studies that characterize admixture patterns in southern Africa with an aim to better understand differences in adverse disease phenotypes observed among groups. Our review confirms that ancestry has a direct impact on an individual's immune response to infectious diseases. In addition, we emphasize the importance of collaborative research, especially for populations in southern Africa that have a high incidence of potentially fatal infectious diseases such as HIV and tuberculosis.

Communicated by S. Hohmann.

✉ Eileen G. Hoal
egvh@sun.ac.za

[1] SA MRC Centre for TB Research, DST/NRF Centre of Excellence for Biomedical Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medical and Health Sciences, Stellenbosch University, Tygerberg, Parow 7500, South Africa

[2] Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794, USA

Keywords Population structure · Southern Africa · Disease susceptibility

Introduction

Southern Africa has a unique and complex human history reaching back at least 100,000 years (Rito et al. 2013). The region spans southern Angola, Namibia, Botswana, South Africa, Zimbabwe, and Mozambique. Many diverse ethnic groups are present in the area, including KhoeSan populations, Bantu-speaking populations, European-descent groups, and groups resulting from inter- and intra-continental admixture such as the South African "Coloured" population (de Wit et al. 2010; Daya et al. 2013; Chimusa et al. 2013). "Admixed" populations are the result of gene flow between distinct, historically divergent parental populations, such as those from different continents like Asia and Africa. The rate, extent, and timing of gene flow between genetically distinct populations have resulted in unique genetic complexity in almost all populations in southern Africa, as well as fine-scale genetic differences between populations. Patterns of allele frequency differences among populations are described as population structure and such allele frequency differences can have subtle or profound phenotypic effects, such as differential susceptibility to infectious disease.

The genetics underlying human disease phenotype variation in African populations have been under-researched. However, the NIH and Wellcome Trust-funded initiative named the Human Heredity and Health in Africa (H3Africa) (Adoga et al. 2014; Ramsay 2015) aim to improve the health of all African populations by facilitating research in the area of genomic and environmental impacts on common diseases such as trypanosomiasis, tuberculosis,

rheumatic heart disease, schizophrenia, type 2 diabetes, and other cardiometabolic diseases. We focus on recent genetic investigations of the two infectious diseases with the biggest impact on health in southern Africa, viz., tuberculosis (TB) and the human immunodeficiency virus (HIV).

Tuberculosis (TB) and the human immunodeficiency virus (HIV) have high incidence and mortality rates in southern Africa (WHO 2016). A major component of TB susceptibility is genetic, and recently, it has been established that part of this susceptibility can be attributed to a particular ancestral population, which contributed to present populations (Daya et al. 2014a, b; Chimusa et al. 2014). Understanding the role of ancestry in infectious disease risk has manifold benefits, including the identification of the most vulnerable populations, providing effective and specialized drug therapies and/or vaccines; and in the highly probable case of the identification of novel susceptibility factors, aiding in the development of new therapies.

Here, we review the population structure and prehistory of southern African populations, as inferred from recent genetic and genomic data sets, to provide a better understanding of how population structure affects disease risk. Thorough literature searches were performed using Pubmed and Google Scholar to capture a wide array of the latest studies in the field using "ancestry-related disease risk", "southern Africa population genetics", and "TB and HIV in southern Africa" as keywords.

The genetic history of the KhoeSan

The KhoeSan are indigenous inhabitants of southern Africa and their ancestors may represent the earliest divergence among extant human populations (Chen et al. 2000; Hammer et al. 2001; Knight et al. 2003; Tishkoff et al. 2007a; Brown et al. 2009, 2012; Schlebusch et al. 2009; Naidoo et al. 2010; Marean 2010; Henn et al. 2011). The genetic origin of the KhoeSan can be traced back to the emergence of modern humans in southern Africa (Gronau et al. 2011; Henn et al. 2012). Prior to 2500 years ago, all KhoeSan populations in southern Africa hunted game or fished, foraged for plants, and gathered natural products, hence the anthropological term of hunter-gatherers. Some contemporary KhoeSan populations continue to forage, while other groups have transitioned to wage labour or stock farming. The Khoekhoen are pastoralists who derive their ancestry from the original hunter-gatherer KhoeSan, but adopted sheep, goat, and cattle husbandry from east African pastoralists approximately 2000 years ago (Pleurdeau et al. 2012; Uren et al. 2016; Montinaro et al. 2017). Bantu-speaking farmers arrived in southern Africa from approximately AD 600 onwards, having migrated down both the west and east coasts of Africa, and subsequently impacted the Khoekhoe and San way of life. The expansion of Bantu-speaking agriculturalists was followed by Arab traders who sailed down the east coast at least as far as Sofala, Mozambique. The Portuguese (the first European visitors to South Africa) encountered the San and Khoekhoen in Mossel Bay, South Africa in 1487, although this and subsequent encounters were brief due to conflict with these indigenous groups. The Dutch and other Europeans began a formal settlement at the Cape of Good Hope (present-day Cape Town, South Africa) in 1652. Within a short period of time, Indian and Asian slaves were brought to the area. These historical events are depicted in Fig. 1. Over time, the Bantu-speaking, European, San, and Khoekhoen, all culturally distinct groups, intermarried with one another, a fact evident not only from their genomes but also from resulting language and cultural practices (Scheinfeldt et al. 2010). The European and Bantu expansion into San and Khoekhoe territory resulted in the decline of the indigenous populations due to conflict, disease, and resource scarcity (Fourie and van Zanden 2013).

Archaeological and genetic evidence suggests that the modern human species originated within Africa, though the precise location of origin is widely contested due to the diversity of African populations and complexity of population history (Batini and Jobling 2011). One hypothesis proposes that the earliest population divergence among humans occurred within southern Africa based on the exceptional genetic diversity present in KhoeSan groups (Henn et al. 2011). Demographic history within KhoeSan populations has widely been investigated with particular reference to their origins and thus the origins of modern humans. One such study included click-speaking Hadza and Sandawe individuals from Tanzania, ǂKhomani San from South Africa as well as 24 other African populations. Linkage disequilibrium and heterozygosity analysis from single nucleotide polymorphism (SNP) array data demonstrated that the ǂKhomani and other KhoeSan from Namibia are two of the most genetically diverse populations in the world (Henn et al. 2011). In conjunction with F_{st} patterns (a measure of genetic distance between populations), their results suggested that humans originated in southern Africa (Henn et al. 2011). This conclusion is supported by microsatellite and indel data (Tishkoff et al. 2009). It was found that the two southern African KhoeSan populations from this study clustered together when phylogenetic trees were constructed from genetic distances (F_{st}) between populations. These populations were also the most distinct populations worldwide (Tishkoff et al. 2009). This is broadly consistent with studies on mitochondrial DNA and Y chromosome, which indicated divergent genetic lineages (Behar et al. 2012; Poznik et al. 2013). Studies investigating the geographical origins of modern humans depend on contemporary populations which might not be truly representative of historical populations (Haber et al. 2016).

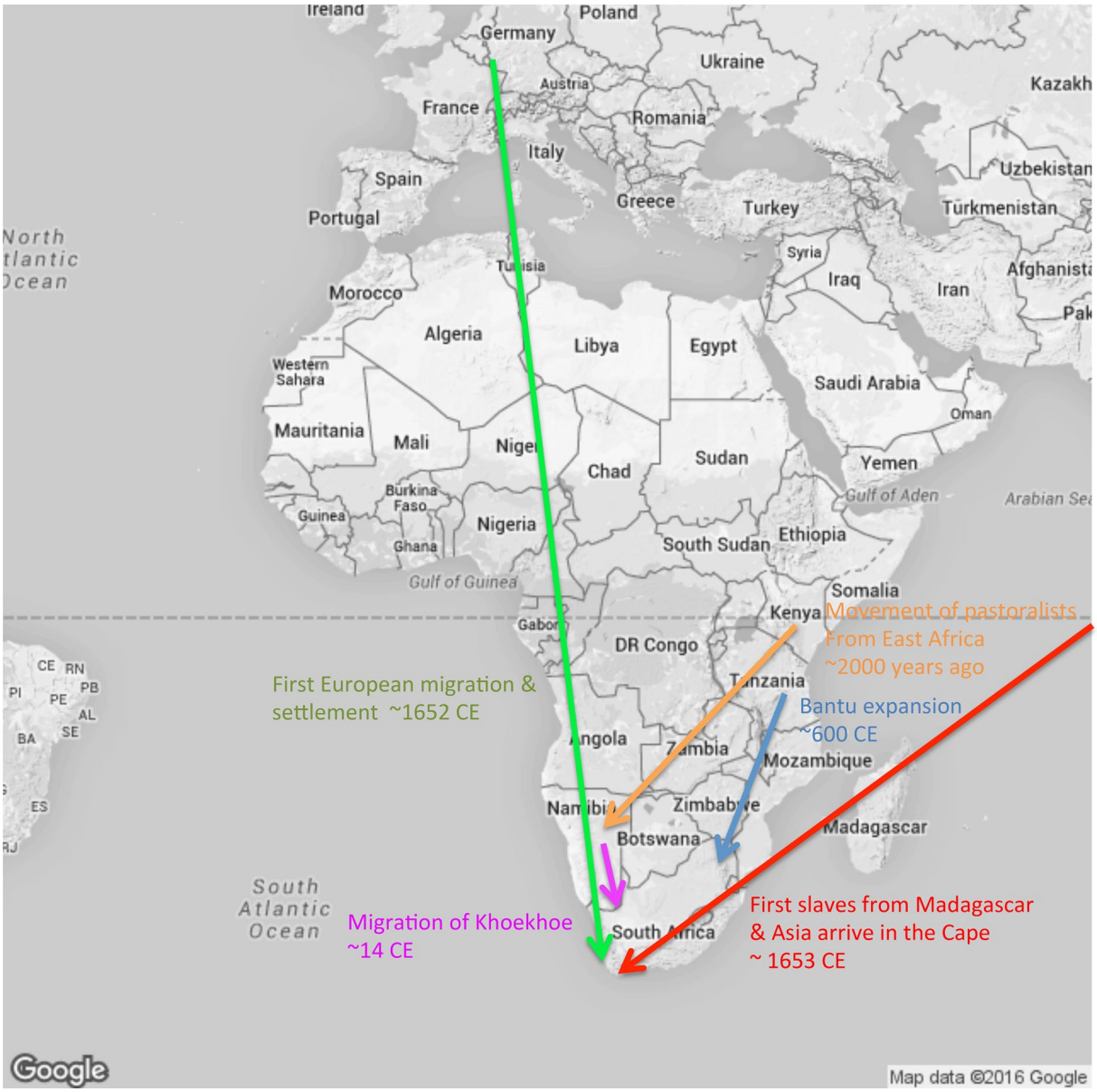

Fig. 1 Southern Africa's complex and long-standing historical migrations. Map of Africa depicting the primary population migrations into southern Africa. Geographical locations of most relevant southern African populations mentioned in this review are depicted by *black circles*

Nonetheless, under standard phylogeographic inference, and supported now by whole-genome analysis of effective population size (Mallick et al. 2016), the oldest population divergence among humans occurred in southern Africa.

In addition to determining the origins of modern humans by inferring the geographic origin of the KhoeSan, it is important to distinguish when and how the ancestors of the KhoeSan diverged from other groups. Tishkoff et al. suggested that contemporary Central African Pygmy, KhoeSan, Hadza and Sandawe hunter-gatherer populations were remnants of a larger ancient "proto-KhoeSan-Pygmy" population with the divergence into distinct populations being estimated to have occurred >35,000 years ago (Tishkoff et al. 2007a). A deeper date of divergence between the KhoeSan populations from other African populations was estimated by Veeramah et al. (2012), who re-sequenced 40 intergenic regions in individuals from the San, Eastern, and Western Pygmies as well as non-Pygmy Niger-Kordofanian populations. They concluded that the ancestors of the KhoeSan separated from a "proto-Pygmy-non-Pygmy Niger-Kordofanian group" ~100,000 years ago (Veeramah et al. 2012), which is consistent with Schlebusch et al. (2012). An even deeper date of divergence was found by analysing whole-genome sequences of six individuals

from diverse ancestral backgrounds (Gronau et al. 2011). The divergence of the San from all other human populations was postulated to occur 108,000–157,000 years ago, whereas the divergence of Eurasians from ancestral African populations occurred 38,000–64,000 years ago (Gronau et al. 2011).

After the separation between ancestors of the KhoeSan and all other populations, there has been further north–west and south–east divergence of Kalahari groups that led to deep structure within the KhoeSan (Pickrell et al. 2012; Schlebusch et al. 2012; Schlebusch and Soodyall 2012; Kidd et al. 2014). The approximate date of the divergence was estimated at ~30,000 years ago (Pickrell et al. 2012). This divergence can still be seen in extant KhoeSan populations as principal component analysis (PCA) reveals three main clusters of KhoeSan populations: namely non-KhoeSan, north-western Kalahari, and south-eastern Kalahari (Pickrell et al. 2012). According to the distribution of L0d and L0k mtDNA haplogroups, which appear virtually only within the KhoeSan groups, the more northerly southern African groups (!Xun, Ju/'hoansi, and /Gui, //Gana) cluster separately from the more southerly southern African groups ($^{\neq}$Khomani, Karretjie and some SAC populations), who have their own distinct cluster (Schlebusch and Soodyall 2012).

Recent studies investigating admixture proportions and the distribution of lactase persistence alleles in extant southern African populations have shown that the KhoeSan populations are heterogeneous and some have admixture from European, Bantu-speaking, and East African populations (Tishkoff et al. 2007b; Henn et al. 2008; Schlebusch et al. 2012; Uren et al. 2016). The gene flow can largely be attributed to three migration events. The first was by East African pastoralists 2–3 kya and the second by Bantu-speaking farmers <1 kya (Smith 2014), followed by European colonization. These migration events have contributed to the genetic diversity in a number of southern African populations, including alleles that affect phenotype. For example, lactase persistence alleles originating in East Africa were found in the Khoe-speaking Nama at a relatively high frequency, as compared to other pastoral populations such as the Himba (Ranciaro et al. 2014; Breton et al. 2014; Macholdt et al. 2014, 2015). In addition, admixture analysis in southern and eastern Africa populations shows that the Eurasian ancestry in southern Africa originated in eastern Africa (Pickrell et al. 2014). These two conclusions support the hypothesis that there was admixture between East African individuals and the native Khoe populations. East African ancestry as well as Bantu-speaking ancestry arising from the later Bantu expansion could impact numerous phenotypes in southern Africa, including TB susceptibility. This has not yet been investigated in depth, but the initial studies have shown that Bantu-speaking ancestry in the SAC population predisposes individuals to progress to active TB (Daya et al. 2014b).

Southern African Coloured populations and their link to the KhoeSan

The South Africa Coloured (SAC) population represents a highly admixed group of individuals from multiple ancestral populations (de Wit et al. 2010). Within the past ~600 years, multiple non-African and African populations have moved into southern Africa and integrated with the indigenous inhabitants (the KhoeSan). The origin of the SAC population has recently been the subject of a number of studies, concerned with quantifying the number, provenance, and proportions of the ancestral populations.

This complex gene flow pattern is crucial in the understanding of the origins of other populations in the region, such as the SAC population (predominantly found in the Western Cape of South Africa) which received over 30% of their ancestry from the KhoeSan (de Wit et al. 2010; Daya et al. 2013; Chimusa et al. 2013; Uren et al. 2016). An early study analysing genome-wide SNP data from 20 SAC, indicated four ancestral contributions to this population, namely European, South Asian, Indonesian, and Xhosa (Patterson et al. 2010). Although these results have some similarities with later ancestral determinations, there were some significant differences. First, it is not clear whether Cape Malay individuals (admixed and similar to the SAC but with higher Asian ancestry due to the slave trade) were included in the sample, which could have biased ancestry results towards Indonesia or more broadly, South Asia. Second, it is not clear which "Bushman" population was used as a proxy ancestral KhoeSan population. Since we know that each KhoeSan population differs substantially in their ancestral admixture pattern (as shown above), interpretations of KhoeSan ancestry in the SAC can be biased or even missed. As more San reference samples became available, and the within-population structure became more evident, the understanding of admixture in the SAC and the analyses thereof improved.

In 2010, a large genome-wide analysis of 959 SAC individuals was performed, where autosomal SNPs were genotyped and combined with data from distinct populations present in the Human Genome Diversity Project (Cann 2002). With the large number of putative ancestral populations available, the results fitted the historical data more accurately. It was found that ancestry proportions in the SAC were dominated by the KhoeSan ancestry which was estimated at 32–43%, followed by black African ancestry at 20–36%, European ancestry 21–28% and Asian 9–11% (de Wit et al. 2010). The accuracy of any inferences made from ancestry data can be affected by the choice of reference populations, the number of individuals used in each

reference population as well as the algorithm used. The analysis by Patterson et al. (2010) described only the continental admixture present in the SAC population, but the best proxy ancestral populations for the SAC were unknown (Patterson et al. 2010). PROXYANC was, therefore, developed by Chimusa et al. (2013) and is based on two novel algorithms, namely, population genetic differentiation and optimal quadratic programming. PROXYANC identifies the most accurate and efficient ancestral populations for a multi-way admixed population. Once the most representative populations were identified using PROXYANC, ancestry proportions were calculated using ADMIXTURE. The Xhosa (black African) contributed 33% ± 0.226, the ǂKhomani San (KhoeSan) contributed 31% ± 0.195, the Europeans contributed 16% ± 0.118, the Gujarati Indians contributed (South Asian) 13% ± 0.094, and the Chinese (East Asian) contributed 7% ± 0.0488 (Chimusa et al. 2013). The combination of Bantu-speaking ancestries (East and West African) in the SAC was estimated at 33% ± 0.04 (Uren et al. 2016). Upon the identification of a southern KhoeSan specific ancestry, both the Nama and ǂKhomani were utilized as KhoeSan reference populations for the SAC resulting in a reported ancestral contribution of 33% ± 0.03 (Uren et al. 2016). In this study, European ancestry was estimated at 12% ± 0.02, Pathan (South Asian) ancestry at 14% ± 0.02 and Chinese ancestry at 7% ± 0.01 (Uren et al. 2016). Table 1 summarises the ancestry proportions as determined by these studies. It clearly shows the increase in accuracy and correlation with historical data. In addition, as data sets and methodologies advanced, it is noteworthy that the estimation of the correct proxy ancestral population to be used, improved. The majority of studies investigating admixture proportions in southern African populations focused primarily on SAC individuals from the Western Cape. However, Petersen et al. (2013) investigated admixture proportions in other SAC populations from around South Africa. SAC individuals from the Eastern Cape had an increase in Bantu-speaking ancestry and individuals from District Six in Cape Town had an increase in Asian ancestry (slaves from India and Madagascar lived in District Six before slave trade was abolished) (Petersen et al. 2013).

mtDNA can be very informative for studying sex-biased migration and formation of complex populations. Quintana-Murci et al. (2010) investigated maternal and paternal ancestral contributions to the SAC population. Sub-Saharan Africa was the origin of the greatest proportion (79%) of the SAC maternal gene pool (Quintana-Murci et al. 2010). It is important to note that 60% of the SAC mtDNA was of the L0d lineage which, together with L0k, is specific to the KhoeSan (Quintana-Murci et al. 2010). The SAC population, therefore, contains considerable maternal input from the KhoeSan. Other mtDNA haplogroups found in the SAC population derive from the Bantu expansion (19%). The remainder of mtDNA ancestry was contributed by south and south-east Asian populations, consistent with autosomal data (Quintana-Murci et al. 2010). On the other hand, the paternal contribution to the SAC population was dominated by a contribution from sub-Saharan Africa, about twice that of the maternal contribution. The most dramatic difference observed was between the paternal (5.3%) and maternal (60%) KhoeSan ancestry (Quintana-Murci et al. 2010). These results displayed an uneven sex-specific gene flow both between and within continents and sheds light on the admixture events and social environments that brought the modern day SAC population into being.

The modern day SAC population spans much of southern Africa, with some geographical variation in terms of

Table 1 Evolution of the South African Coloured population's ancestral proportions: ancestry proportions of the SAC population from studies using different ancestry determination methods and reference populations, reported as summary means and standard error

Xhosa		European	Indonesian	South Asian
Patterson et al. (2010)				
37% ± 0.003		23% ± 0.008	18% ± 0.004	22% ± 0.009
West African	KhoeSan	European	Chinese	Indian
de Wit et al. (2010)				
24% ± 0.161	37% ± 0.148	18% ± 0.118	7% ± 0.0478	14% ± 0.093
Xhosa	KhoeSan	European	Chinese	Indian
Chimusa et al. (2013)				
33% ± 0.226	31% ± 0.195	16% ± 0.118	7% ± 0.0488	13% ± 0.094
Bantu-speaking	South African KhoeSan	European	Chinese	Pathan
Uren et al. (2016)				
33% ± 0.04	33% ± 0.03	12% ± 0.02	7% ± 0.01	14% ± 0.02

genetic and cultural characteristics. The Karretjie people (officially classified as 'Coloured') of the Great Karoo (found in the Northern Cape and Western Cape of South Africa) were analysed in a similar way to the SAC by looking at the distinction between maternal and paternal contributions. Interestingly, the KhoeSan specific clade L0d as mentioned above was present in all the Karretjie samples ($n=31$), suggesting a solely KhoeSan maternal contribution (Schlebusch and Soodyall 2012). In contrast, paternal contributions were more heterogeneous, similar to the SAC in the Western Cape (Quintana-Murci et al. 2010). This pattern is also evident in the Rehoboth Basters, a distinct group of individuals who moved from the Cape Colony to southern Namibia 150 years ago. The paternal lineage of this group is European, while the maternal lineage is KhoeSan (Petersen et al. 2013). Using a SNP array data set, ADMIXTURE analysis indicates five ancestral components, similar to the SAC, and in the PCA, the Basters cluster with the SAC (Petersen et al. 2013). Similarities in ancestry proportions were observed between the Coloured and Baster individuals but with higher European and KhoeSan ancestry in the Basters.

More recent mtDNA studies have suggested that the sex-biased admixture in southern African populations is not so straightforward as previously thought (e.g., KhoeSan maternal lineages vs. largely Bantu-speaking paternal lineages in the SAC) with analyses suggesting high levels of inter-population variance at the maternal lineage level (Barbieri et al. 2013, 2014a). This complexity evident is not only in the SAC population, but in other southern African populations who share such a complex genetic history such as Bantu-speaking populations (Barbieri et al. 2014b). These conclusions are supported by Y chromosome analyses (and thus telling of paternal lineages) where an increase in mutation rate was identified in Bantu-speaking haplogroups suggesting either a population expansion and/or an older age of paternity (Barbieri et al. 2016).

Multiple ancestries in southern Bantu-speaking populations

The majority of southern African individuals currently belong to a large variety of Bantu-speaking populations (~70% of the population). On the whole, the manner of dispersal of southern African Bantu-speakers is largely unknown and it has been previously hypothesized that there is limited population structure between individual groups. This hypothesis arose partly due to extensive population movements during the period of civil war in the early 1800's known as the "Mfecane" (Walker 1928). Later studies have attempted to identify and characterise the putative fine-scale population structure present among southern African Bantu-speaking populations by estimating F_{st} values determined from autosomal serogenetic, DNA and Y chromosome haplotypes. Southern Bantu-speaking populations tend to cluster in accordance with their linguistic grouping, with the exception of the Tsonga who clustered closer to the Venda, perhaps due to their close geographic proximity (Lane et al. 2002). In general, genetic distances as well as linguistic groupings correlated with geographical distances (Lane et al. 2002). The differences between these southern Bantu-speaking populations were, however, small.

Although there is a relative lack of identifiable structure among Bantu-speaking populations, the demographic model that governs their dispersal into southern Africa is convoluted. Utilizing mtDNA and Y chromosome data, it has been possible to distinguish between differing models so as to explain the admixture patterns we see in modern day southern African Bantu-speaking populations. There is evidence for gene flow between the migrating farmers and the indigenous foraging KhoeSan communities (represented by the Ju/'hoansi from Namibia) (Marks et al. 2014). This signal of admixture is pronounced in Bantu-speaking populations in South-East Africa, e.g., Mozambique. This finding was supported by large migration rate estimates of 1–2% per generation for ~900 years from the KhoeSan to Bantu-speakers, as calculated based on simulated data (Marks et al. 2014). High proportions of KhoeSan ancestry are observed in the Zulu and Sotho populations (i.e., ~23%) (Gurdasani et al. 2015). The Xhosa (from the Eastern Cape province in South Africa), who constitute a large proportion of South Africa's Bantu-speaking population, also derives 20% of their ancestry from the southern KhoeSan and the rest from East, West, and Central Africa (Petersen et al. 2013; Uren et al. 2016). The admixture from the KhoeSan into Bantu-speaking South African populations is similar to the European admixture proportion in African-American populations in the USA (Bryc et al. 2015); African-Americans are a canonical "admixed" population in US biomedical research. Biomedical research approaches developed for African-Americans should be considered more often when studying South African Bantu-speaking groups (e.g., admixture mapping for diseases which differ in susceptibility between the two ancestries).

Although ancestry proportions are of historical interest, they are also relevant to disease risk profiling. For example, the SAC population have arguably the highest incidence of pulmonary tuberculosis in the absence of human immunodeficiency (HIV) immunosuppression, while Bantu-speaking populations have the highest incidence of HIV/AIDS (Kenyon et al. 2013). Knowledge of the population structure and genetic ancestry can enable the mapping of possible susceptibility causing loci (i.e., via admixture mapping). This approach can be extended to other infectious diseases prevalent in southern Africa.

Genetic perspectives of infectious disease phenotypes in southern African populations

South Africa has the second highest incidence rate of TB in the world after Lesotho, a country surrounded by South Africa (WHO 2016). The disease is rife in all South African populations, including the SAC. A genome-wide association study (GWAS) indicated that excess KhoeSan ancestry predisposes the SAC population to TB and this effect was not confounded by socio-economic status (Chimusa et al. 2014). Subsequently, Daya et al. (2014b) used Ancestry Informative Markers (AIMS) in a validation study with an independent sample set of 918 cases and 507 controls. Correcting for KhoeSan ancestry affected whether a polymorphism was still significantly associated with TB or not, dependent on the frequency of the SNP in parent populations (Daya et al. 2013, 2014b). Further investigation not only indicated that KhoeSan and African non-KhoeSan ancestries are associated with an increased risk of progression to active pulmonary TB, but that European, South Asian, and East Asian ancestry are protective against TB (Daya et al. 2014b). Interestingly, intergenic class II human leukocyte antigen (*HLA*) variants were recently associated with both protection against and susceptibility to active TB disease in *M. tuberculosis*-infected individuals of European ancestry (Sveinbjornsson et al. 2016). It is not known whether these *HLA* class II variants are at appreciable frequencies in SAC populations. From admixture mapping and the resulting ancestry correlation tests, it was estimated that every 10% increase in KhoeSan ancestry in an individual correlated with a 38% increase in the odds of progressing to active pulmonary TB (Daya et al. 2014a). The regions of excess KhoeSan ancestry in TB cases include the *GADD45A* and *OSM* genes (Daya et al. 2014a), which makes them candidates for further investigation.

Although only one GWAS for TB has been performed in southern African groups, linkage and numerous candidate gene case-control association studies have identified SNP variants associated with TB susceptibility (Abel et al. 2014). As an example, the Major Histocompatibility Complex (MHC) and the Leukocyte Receptor Complex (LRC) have been implicated in altering the susceptibility to infectious diseases (Kostyu et al. 1997; Balamurugan et al. 2004; Lombard et al. 2006; Kettaneh et al. 2006). After investigating the interaction between the Human Leukocyte Antigen (HLA) type of the TB patient and the infecting *M. tuberculosis* strain (Salie et al. 2014), they hypothesized that three vaccines currently in clinical trials may not be effective in the SAC population as predicted from common HLA allele class I profiles. This is due to the significantly lower frequency in the SAC and Bantu-speaking populations compared to Europeans, of the HLA subtype that is required for bacterial epitope recognition (Salie 2014). In addition to the MHC and LRC, African genome-wide linkage studies have shown that loci in melanocortin-3-receptor (*MC3R*) and cathepsin Z (*CTSZ*) are linked to susceptibility to TB (Cooke et al. 2008), and SNPs in these genes were validated in the SAC population in a case–control association study (Adams et al. 2011).

Other variants that have been associated with TB susceptibility are involved in immune pathways in which interferon-gamma (IFN-γ) plays a crucial role (Möller et al. 2010). An intronic variant which increased IFN-γ production (therefore, affecting the host immune response) displays population specific levels of positive selection with a higher level of positive selection present in African populations (Yoruba, Mandenka, and Bantu-speaking individuals from Kenya) (Manry et al. 2011). Populations used in this study originated from sub-Saharan Africa, Europe, and East Asia, but no southern African populations were included. An association and transmission disequilibrium test (TDT) on SAC individuals noted that a promoter polymorphism in the gene for IFN-γ (*IFNG*) was significantly associated with an increased likelihood of progression to active TB (Rossouw et al. 2003). Various meta-analyses have identified other possible variants in *IFNG* that are associated with TB susceptibility once stratified based on ethnicity, but no studies have investigated this association in African populations (Tian et al. 2011). In addition to IFN-γ, tumour necrosis factor (TNF) is an important inflammatory meditator and its role in TB susceptibility is well documented, especially in sub-Saharan African populations (Mabunda et al. 2015). Meta-analyses have shown that the association between TNF-α and TB susceptibility is stratified by population (Anoosheh et al. 2011; Yi et al. 2015). Within the context of Africa, the well-characterized −308G>A polymorphism in TNF-α was associated with pulmonary TB in African populations under a dominant model but not in the Asian or Caucasian populations (Yi et al. 2015).

A linkage analysis of the quantitative tuberculin skin test (TST) reaction to injected PPD was done on 128 SAC families including 350 siblings. This study determined that a single major locus on chromosomal region 11p14 that we called *TST1*, appears to control human resistance to the bacterium, as evidenced by the lack of delayed-type hypersensitivity (Cobat et al. 2009). This work was the first report of a genetic resistance factor for TB infection as opposed to disease. In the same families, they detected a major pleiotropic locus on chromosome region 11p15, termed *TNF1*, that controlled TNF production after stimulation by both BCG alone and BCG plus IFN-γ. The close proximity of these loci suggested that there is a connection between TST negativity per se and TNF production (Cobat et al. 2013).

Toll-like receptors (TLRs) are known to play a major role in an individual's immune response to TB. The

previous association studies have led to differing results, particularly within complex populations, e.g., in southern Africa, suggesting that population substructure may be masking signals of susceptibility. For this reason, meta-analyses have looked into the relationship between TLR's and TB, considering different ethnic groups as well as increasing the power to detect any statistical associations. The major TLRs (1, 2, 4, 6, and 9) play a role in TB susceptibility in most populations (Chen et al. 2010; Sun et al. 2015; Schurz et al. 2015), but some variants were associated with TB susceptibility in one population but not in another. For example, the AG genotype of *TLR1* r4833095 and the T allele of *TLR6* rs5743810 were associated with resistance to TB across all ethnic groups studied (Asian, African, European, and Hispanic), whereas variants located in *TLR4* and *TLR2* showed associations with TB susceptibility only in the Asian and Hispanic subpopulations (Schurz et al. 2015). In contrast, variants in TLR2, TLR6, and TLR8 were found to be associated with TB susceptibility in the Chinese population (Sun et al. 2015). These variants in TLR8 (rs3764879, rs3761624, rs3788935, and rs3764880) were shown to be associated with TB susceptibility in the SAC population (Chimusa et al. 2014). Further studies investigating ethnicity as a confounding factor have been performed, but very few include African populations.

When considering susceptibility to TB, specifically the recurrence of the disease, it is relevant to discuss the effect that genetic variation has on anti-TB drug metabolism. The CYP3A4 enzyme (a human cytochrome P450) is one of the most important enzymes involved in drug metabolism. Differing allele frequencies within *CYP3A4* were found in the KhoeSan, Xhosa, and SAC populations. Of the 24 SNPs detected in *CYP3A4* in these populations, one was a functional promoter polymorphism (Drögemöller et al. 2013). Looking at drug metabolism through the lens of genetic variation leads to the prediction that although some drugs might be effective in a few populations, they might be harmful or ineffective in others (Drögemöller et al. 2013) and in the case of TB drug metabolism, differences in metabolism could lead to the development of drug resistant TB.

TB is also the leading cause of death in HIV-infected patients (WHO 2016). A recent study hypothesized that HIV positive individuals who live in a TB endemic area and do not develop active disease could provide insight into TB resistance (Sobota et al. 2016). A GWAS was done in HIV positive individuals from Tanzania and Uganda and identified a locus at chromosome 5q33.3 which may offer protection against TB (Sobota et al. 2016).

Sub-Saharan Africa has the highest HIV incidence rate in the world (WHO 2016). As the effectiveness of antiretrovirals has reached the stage where normal life expectancy is possible, research is moving towards focusing on the cause of death for infected individuals, and identifying host factors contributing to HIV infection. Viral co-receptors CCR5 and CXCR4 are the most crucial and polymorphisms in either can confer protection or susceptibility to the virus (Feng et al. 1996; Alkhatib et al. 1996; Choe et al. 1996). Specifically, *CCR5*Δ32 has been shown to confer protection to HIV infection in Europeans (Samson et al. 1996). Given the high HIV incidence in southern Africa, it is noteworthy that there is a statistically significant difference in the activation and expression levels of CCR5 between two South African populations, namely "South African Africans" and "South African Caucasians" (Picton et al. 2012). This could result in altered susceptibility to HIV-1 infection as well as affect the progression of the infection itself. This study did not include ancestral information to determine the origin of the phenotypes under study, which may play a pivotal role in the identification of the variant associated with the phenotype.

Conclusion

Southern African populations are unique in their culture, history, and languages, and may be the cradle of humankind; this uniqueness is supported by population genetic analyses. Understanding the genetic structure of these populations is not only important to reconstruct human evolutionary history, but also has implications for the study of disease risk. Significant KhoeSan ancestry in many present-day southern African populations reflects how recent migration into this region resulted in the absorption of indigenous KhoeSan groups into many ethnicities. For example, population structure analyses of the SAC population in the Western Cape indicate substantial (>30%) ancestry from the KhoeSan who were present at the Cape during the initial European colonization. It is also clear that there are ancestry-linked genetic factors contributing to infectious disease susceptibility in southern Africa, particularly with regard to TB. This may also be true for other infectious and chronic diseases not mentioned in this review, but due to lack of available studies, these warrant further investigation. Research into the link between ancestry and disease risk is sparse and the lack of publically available data is one reason for this. Collaborative networks, especially with respect to sample collection and analysis, would greatly facilitate these investigations. Although there has been some progress in encouraging genetic studies of populations displaying adverse phenotypes in Africa, more research needs to be based on southern African populations, especially in the fields of HIV and TB.

Acknowledgements Authors thank the National Research Foundation of South Africa for funding.

Author contributions BMH and MM conceived the review topic. CU wrote the manuscript with review from MM, PvH, BMH, and EGH.

Compliance with ethical standards

Conflict of interest Authors declare that no conflict of interest exists.

Ethical approval This article does not contain any studies, performed by the authors, involving human or animal participants.

References

Abel L, El-Baghdadi J, Bousfiha AA, et al (2014) Human genetics of tuberculosis: a long and winding road. Phil Trans R Soc B 369:20130428. doi:10.1098/rstb.2013.0428

Adams LA, Möller M, Nebel A, et al (2011) Polymorphisms in MC3R promoter and CTSZ 3′UTR are associated with tuberculosis susceptibility. Eur J Hum Genet 19:676–681. doi:10.1038/ejhg.2011.1

Adoga MP, Fatumo SA, Agwale SM (2014) H3Africa: a tipping point for a revolution in bioinformatics, genomics and health research in Africa. Source Code Biol Med 9:10. doi:10.1186/1751-0473-9-10

Alkhatib G, Combadiere C, Broder CC et al (1996) CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. Science 272:1955–1958

Anoosheh S, Farnia P, Kargar M (2011) Association between TNF-alpha (-857) gene polymorphism and susceptibility to tuberculosis. Iran Red Crescent Med J 13:243–248.

Balamurugan A, Sharma SK, Mehra NK (2004) Human leukocyte antigen class I supertypes influence susceptibility and severity of tuberculosis. J Infect Dis 189:805–811. doi:10.1086/381689

Barbieri C, Vicente M, Rocha J et al (2013) Ancient substructure in early mtDNA lineages of Southern Africa. Am J Hum Genet 92:285–292. doi:10.1016/j.ajhg.2012.12.010

Barbieri C, Güldemann T, Naumann C et al (2014a) Unraveling the complex maternal history of Southern African Khoisan populations. Am J Phys Anthropol 153:435–448. doi:10.1002/ajpa.22441

Barbieri C, Vicente M, Oliveira S, et al (2014b) Migration and interaction in a contact zone: mtDNA variation among Bantu-speakers in Southern Africa. PLoS One 9:e99117. doi:10.1371/journal.pone.0099117

Barbieri C, Hübner A, Macholdt E et al (2016) Refining the Y chromosome phylogeny with southern African sequences. Hum Genet 135:541–553. doi:10.1007/s00439-016-1651-0

Batini C, Jobling MA (2011) The jigsaw puzzle of our African ancestry: unsolved, or unsolvable? Genome Biol 12:118. doi:10.1186/gb-2011-12-6-118

Behar DM, van Oven M, Rosset S et al (2012) A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet 90:675–684. doi:10.1016/j.ajhg.2012.03.002

Breton G, Schlebusch CM, Lombard M, et al (2014) Lactase persistence alleles reveal partial East african ancestry of southern african Khoe pastoralists. Curr Biol 24:852–858. doi:10.1016/j.cub.2014.02.041

Brown KS, Marean CW, Herries AIR et al (2009) Fire as an engineering tool of early modern humans. Science 325:859–862. doi:10.1126/science.1175028

Brown KS, Marean CW, Jacobs Z et al (2012) An early and enduring advanced technology originating 71,000 years ago in South Africa. Nature 491:590–593. doi:10.1038/nature11660

Bryc K, Durand EY, Macpherson JM et al (2015) The genetic ancestry of African Americans, Latinos, and European Americans across the US. Am J Hum Genet 96:37–53. doi:10.1016/j.ajhg.2014.11.010

Cann HM (2002) A human genome diversity cell line panel (letter). Science 296:261. doi:10.1126/science.296.5566.261b

Chen YS, Olckers A, Schurr TG et al (2000) mtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations. Am J Hum Genet 66:1362–1383. doi:10.1086/302848

Chen Y-C, Hsiao C-C, Chen C-J et al (2010) Toll-like receptor 2 gene polymorphisms, pulmonary tuberculosis, and natural killer cell counts. BMC Med Genet 11:17. doi:10.1186/1471-2350-11-17

Chimusa ER, Daya M, Möller M et al (2013) Determining ancestry proportions in complex admixture scenarios in South Africa using a novel proxy ancestry selection method. PLoS One. doi:10.1371/journal.pone.0073971

Chimusa ER, Zaitlen N, Daya M et al (2014) Genome-wide association study of ancestry-specific TB risk in the South African Coloured population. Hum Mol Genet 23:796–809. doi:10.1093/hmg/ddt462

Choe H, Farzan M, Sun Y et al (1996) The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. Cell 85:1135–1148

Cobat A, Gallant CJ, Simkin L et al (2009) Two loci control tuberculin skin test reactivity in an area hyperendemic for tuberculosis. J Exp Med 206:2583–2591

Cobat A, Hoal EG, Gallant CJ, et al (2013) Identification of a major locus, TNF1, that controls BCG-triggered tumor necrosis factor production by leukocytes in an area hyperendemic for tuberculosis. Clin Infect Dis Off Publ Infect Dis Soc Am 57:963–970.

Cooke GS, Campbell SJ, Bennett S et al (2008) Mapping of a novel susceptibility locus suggests a role for MC3R and CTSZ in human tuberculosis. Am J Respir Crit Care Med 178:203–207. doi:10.1164/rccm.200710-1554OC

Daya M, van der Merwe L, Galal U, et al (2013) A panel of ancestry informative markers for the complex five-way admixed South African coloured population. PLoS One 8:e82224. doi:10.1371/journal.pone.0082224

Daya M, van der Merwe L, Gignoux CR et al (2014a) Using multiway admixture mapping to elucidate TB susceptibility in the South African Coloured population. BMC Genom 15:1021. doi:10.1186/1471-2164-15-1021

Daya M, van der Merwe L, van Helden PD, et al (2014b) The role of ancestry in TB susceptibility of an admixed South African population. Tuberc Edinb Scotl 94:413–420. doi:10.1016/j.tube.2014.03.012

de Wit E, Delport W, Rugamika CE et al (2010) Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. Hum Genet 128:145–153. doi:10.1007/s00439-010-0836-1

Drögemöller B, Plummer M, Korkie L, et al (2013) Characterization of the genetic variation present in CYP3A4 in three South African populations. Front Genet 4:17. doi:10.3389/fgene.2013.00017

Feng Y, Broder CC, Kennedy PE, Berger EA (1996) HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. Science 272:872–877

Fourie J, van Zanden JL (2013) GDP in the Dutch Cape colony: the national accounts of a slave-Based society. South Afr J Econ 81:467–490. doi:10.1111/saje.12010

Gronau I, Hubisz MJ, Gulko B et al (2011) Bayesian inference of ancient human demography from individual genome sequences. Nat Genet 43:1031–1034. doi:10.1038/ng.937

Gurdasani D, Carstensen T, Tekola-Ayele F et al (2015) The African genome variation project shapes medical genetics in Africa. Nature 517:327–332. doi:10.1038/nature13997

Haber M, Mezzavilla M, Xue Y, Tyler-Smith C (2016) Ancient DNA and the rewriting of human history: be sparing with Occam's razor. Genome Biol. doi:10.1186/s13059-015-0866-z

Hammer MF, Karafet TM, Redd AJ et al (2001) Hierarchical patterns of global human Y-chromosome diversity. Mol Biol Evol 18:1189–1203

Henn BM, Gignoux C, Lin AA, et al (2008) Y-chromosomal evidence of a pastoralist migration through Tanzania to southern Africa. Proc Natl Acad Sci 105:10693–10698. doi:10.1073/pnas.0801184105

Henn BM, Gignoux CR, Jobin M et al (2011) Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. Proc Natl Acad Sci USA 108:5154–5162. doi:10.1073/pnas.1017511108

Henn BM, Cavalli-Sforza LL, Feldman MW (2012) The great human expansion. Proc Natl Acad Sci USA 109:17758–17764. doi:10.1073/pnas.1212380109

Kenyon C, Buyze J, Colebunders R (2013) HIV prevalence by race co-varies closely with concurrency and number of sex partners in South Africa. PLoS One 8:e64080. doi:10.1371/journal.pone.0064080

Kettaneh A, Seng L, Tiev KP et al (2006) Human leukocyte antigens and susceptibility to tuberculosis: a meta-analysis of case–control studies. Int J Tuberc Lung Dis 10:717–725

Kidd JM, Sharpton TJ, Bobo D et al (2014) Exome capture from saliva produces high quality genomic and metagenomic data. BMC Genom 15:262. doi:10.1186/1471-2164-15-262

Knight A, Underhill PA, Mortensen HM et al (2003) African Y chromosome and mtDNA divergence provides insight into the history of click languages. Curr Biol 13:464–473. doi:10.1016/S0960-9822(03)00130-1

Kostyu DD, Hannick LI, Traweek JL et al (1997) HLA class I polymorphism: structure and function and still questions. Hum Immunol 57:1–18

Lane A b., Soodyall H, Arndt S et al (2002) Genetic substructure in South African Bantu-speakers: Evidence from autosomal DNA and Y-chromosome studies. Am J Phys Anthropol 119:175–185. doi:10.1002/ajpa.10097

Lombard Z, Brune AE, Hoal EG et al (2006) HLA class II disease associations in southern Africa. Tissue Antigens 67:97–110. doi:10.1111/j.1399-0039.2006.00530.x

Mabunda N, Alvarado-Arnez LE, Vubil A et al (2015) Gene polymorphisms in patients with pulmonary tuberculosis from Mozambique. Mol Biol Rep 42:71–76. doi:10.1007/s11033-014-3741-1

Macholdt E, Lede V, Barbieri C, et al (2014) Tracing pastoralist migrations to southern Africa with lactase persistence alleles. Curr Biol CB 24:875–879. doi:10.1016/j.cub.2014.03.027

Macholdt E, Slatkin M, Pakendorf B, Stoneking M (2015) New insights into the history of the C-14010 lactase persistence variant in Eastern and Southern Africa. Am J Phys Anthropol 156:661–664. doi:10.1002/ajpa.22675

Mallick S, Li H, Lipson M et al (2016) The Simons genome diversity project: 300 genomes from 142 diverse populations. Nature 538:201–206. doi:10.1038/nature18964

Manry J, Laval G, Patin E et al (2011) Evolutionary genetics evidence of an essential, nonredundant role of the IFN-γ pathway in protective immunity. Hum Mutat 32:633–642. doi:10.1002/humu.21484

Marean CW (2010) Pinnacle point cave 13B (Western Cape Province, South Africa) in context: the Cape Floral kingdom, shellfish, and modern human origins. J Hum Evol 59:425–443. doi:10.1016/j.jhevol.2010.07.011

Marks SJ, Montinaro F, Levy H, et al (2014) Static and moving frontiers: the genetic landscape of Southern African Bantu-speaking populations. Mol Biol Evol: msu263. doi:10.1093/molbev/msu263

Möller M, Nebel A, van Helden PD et al (2010) Analysis of eight genes modulating interferon gamma and human genetic susceptibility to tuberculosis: a case-control association study. BMC Infect Dis 10:154. doi:10.1186/1471-2334-10-154

Montinaro F, Busby GBJ, Gonzalez-Santos M et al (2017) Complex ancient genetic structure and cultural transitions in Southern African populations. Genetics 205:303–316. doi:10.1534/genetics.116.189209

Naidoo T, Schlebusch CM, Makkan H, et al (2010) Development of a single base extension method to resolve Y chromosome haplogroups in sub-Saharan African populations. Investig Genet 1:6. doi:10.1186/2041-2223-1-6

Patterson N, Petersen DC, van der Ross RE et al (2010) Genetic structure of a unique admixed population: implications for medical research. Hum Mol Genet 19:411–419. doi:10.1093/hmg/ddp505

Petersen DC, Libiger O, Tindall EA et al (2013) Complex patterns of genomic admixture within southern Africa. PLoS Genet 9:e1003309. doi:10.1371/journal.pgen.1003309

Pickrell JK, Patterson N, Barbieri C et al (2012) The genetic prehistory of southern Africa. Nat Commun 3:1143. doi:10.1038/ncomms2140

Pickrell JK, Patterson N, Loh P-R, et al (2014) Ancient west Eurasian ancestry in southern and eastern Africa. Proc Natl Acad Sci 111:2632–2637. doi:10.1073/pnas.1313787111

Picton ACP, Shalekoff S, Paximadis M, Tiemessen CT (2012) Marked differences in CCR5 expression and activation levels in two South African populations. Immunology 136:397–407. doi:10.1111/j.1365-2567.2012.03592.x

Pleurdeau D, Imalwa E, Détroit F et al (2012) "Of Sheep and Men": earliest direct evidence of Caprine domestication in Southern Africa at Leopard Cave (Erongo, Namibia). PLoS One 7:e40340. doi:10.1371/journal.pone.0040340

Poznik GD, Henn BM, Yee M-C et al (2013) Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. Science 341:562–565. doi:10.1126/science.1237619

Quintana-Murci L, Harmant C, Quach H et al (2010) Strong maternal Khoisan contribution to the South African coloured population: a case of gender-biased admixture. Am J Hum Genet 86:611–620. doi:10.1016/j.ajhg.2010.02.014

Ramsay M (2015) Growing genomic research on the African continent: the H3Africa Consortium. South Afr Med J Suid-Afr Tydskr Vir Geneeskd 105:1016–1017.

Ranciaro A, Campbell MC, Hirbo JB et al (2014) Genetic origins of lactase persistence and the spread of pastoralism in Africa. Am J Hum Genet 94:496–510. doi:10.1016/j.ajhg.2014.02.009

Rito T, Richards MB, Fernandes V, et al (2013) The first modern human dispersals across Africa. PLoS One 8:e80031. doi:10.1371/journal.pone.0080031

Rossouw M, Nel HJ, Cooke GS et al (2003) Association between tuberculosis and a polymorphic NFkappaB binding site in the interferon gamma gene. Lancet 361:1871–1872. doi:10.1016/S0140-6736(03)13491-5

Salie M (2014) The role of the major histocompatibility complex and the leukocyte receptor complex genes in susceptibility to tuberculosis in a South African population. University of Stellenbosch, Stellenbosch

Salie M, van der Merwe L, Möller M et al (2014) Associations between human leukocyte antigen class I variants and the

Mycobacterium tuberculosis subtypes causing disease. J Infect Dis 209:216–223. doi:10.1093/infdis/jit443

Samson M, Libert F, Doranz BJ et al (1996) Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 382:722–725. doi:10.1038/382722a0

Scheinfeldt LB, Soi S, Tishkoff SA (2010) In the light of evolution. National Academy of Sciences, Washington, DC

Schlebusch CM, Soodyall H (2012) Extensive population structure in San, Khoe, and mixed ancestry populations from southern Africa revealed by 44 short 5-SNP haplotypes. Hum Biol 84:695–724

Schlebusch CM, Naidoo T, Soodyall H (2009) SNaPshot minisequencing to resolve mitochondrial macro-haplogroups found in Africa. Electrophoresis 30:3657–3664. doi:10.1002/elps.200900197

Schlebusch CM, Skoglund P, Sjödin P et al (2012) Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. Science 338:374–379. doi:10.1126/science.1227721

Schurz H, Daya M, Möller M, et al (2015) TLR1, 2, 4, 6 and 9 variants associated with tuberculosis susceptibility: a systematic review and meta-analysis. PLoS One 10:e0139711. doi:10.1371/journal.pone.0139711

Smith A (2014) The origins of Herding in Southern Africa: debating the "Neolithic" model. LAP LAMBERT Academic Publishing, Saarbrücken

Sobota RS, Stein CM, Kodaman N et al (2016) A Locus at 5q33.3 confers resistance to tuberculosis in highly susceptible individuals. Am J Hum Genet 98:514–524. doi:10.1016/j.ajhg.2016.01.015

Sun Q, Zhang Q, Xiao H, Bai C (2015) Toll-like receptor polymorphisms and tuberculosis susceptibility: a comprehensive meta-analysis. J Huazhong Univ Sci Technol Med Sci Hua Zhong Ke Ji Xue Xue Bao Yi Xue Ying Wen Ban Huazhong Keji Daxue Xuebao Yixue Yingdewen Ban 35:157–168. doi:10.1007/s11596-015-1405-6

Sveinbjornsson G, Gudbjartsson DF, Halldorsson BV et al (2016) HLA class II sequence variants influence tuberculosis risk in populations of European ancestry. Nat Genet 48:318–322. doi:10.1038/ng.3498

Tian C, Zhang Y, Zhang J et al (2011) The +874T/A polymorphism in the interferon-γ gene and tuberculosis risk: an update by meta-analysis. Hum Immunol 72:1137–1142. doi:10.1016/j.humimm.2011.07.310

Tishkoff SA, Gonder MK, Henn BM et al (2007a) History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. Mol Biol Evol 24:2180–2195. doi:10.1093/molbev/msm155

Tishkoff SA, Reed FA, Ranciaro A et al (2007b) Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet 39:31–40. doi:10.1038/ng1946

Tishkoff SA, Reed FA, Friedlaender FR et al (2009) The genetic structure and history of Africans and African Americans. Science 324:1035–1044. doi:10.1126/science.1172257

Uren C, Kim M, Martin AR et al (2016) Fine-scale human population structure in Southern Africa reflects ecogeographic boundaries. Genetics. doi:10.1534/genetics.116.187369

Veeramah KR, Wegmann D, Woerner A et al (2012) An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. Mol Biol Evol 29:617–630. doi:10.1093/molbev/msr212

Walker EA (1928) A history of Southern Africa. Longmans, Green, Harlow

WHO (2016) Global tuberculosis report 2016. http://www.who.int/tb/publications/global_report/en/. Accessed 20 Dec 2016

Yi Y-X, Han J-B, Zhao L et al (2015) Tumor necrosis factor alpha gene polymorphism contributes to pulmonary tuberculosis susceptibility: evidence from a meta-analysis. Int J Clin Exp Med 8:20690–20700