

RESEARCH ARTICLE



WILEY

Later Stone Age human hair from Vaalkrans Shelter, Cape Floristic Region of South Africa, reveals genetic affinity to Khoe groups

Alexandra Coutinho¹ | Helena Malmström^{1,2} | Hanna Edlund¹ |
 Christopher S. Henshilwood^{3,4} | Karen L. van Niekerk³ | Marlize Lombard² |
 Carina M. Schlebusch^{1,2,5} | Mattias Jakobsson^{1,2,5} 

¹Human Evolution, Department of Organismal Biology, Uppsala University, Uppsala, Sweden

²Palaeo-Research Institute, University of Johannesburg, Johannesburg, South Africa

³SFF Centre for Early Sapiens Behaviour (SapienCE), University of Bergen, Bergen, Norway

⁴Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, South Africa

⁵Science for Life Laboratory, Uppsala, Sweden

Correspondence

Mattias Jakobsson, Human Evolution, Department of Organismal Biology, Uppsala University, Norbyvägen 18C, 75236 Uppsala, Sweden.

Email: mattias.jakobsson@ebc.uu.se

Funding information

Göran Gustafssons Stiftelse för Naturvetenskaplig och Medicinsk Forskning; Knut och Alice Wallenbergs Stiftelse; Norges Forskningsråd, Grant/Award Number: 262618; South African National Research Foundation; Vetenskapsrådet, Grant/Award Numbers: 621-2014-5211, 642-2013-8019

Abstract

Previous studies show that the indigenous people of the southern Cape of South Africa were dramatically impacted by the arrival of European colonists starting ~400 years ago and their descendants are today mixed with Europeans and Asians. To gain insight on the occupants of the Vaalkrans Shelter located at the southernmost tip of Africa, we investigated the genetic make-up of an individual who lived there about 200 years ago. We further contextualize the genetic ancestry of this individual among prehistoric and current groups. From a hair sample excavated at the shelter, which was indirectly dated to about 200 years old, we sequenced the genome (1.01 times coverage) of a Later Stone Age individual. We analyzed the Vaalkrans genome together with genetic data from 10 ancient (pre-colonial) individuals from southern Africa spanning the last 2000 years. We show that the individual from Vaalkrans was a man who traced ~80% of his ancestry to local southern San hunter-gatherers and ~20% to a mixed East African-Eurasian source. This genetic make-up is similar to modern-day Khoekhoe individuals from the Northern Cape Province (South Africa) and Namibia, but in the southern Cape, the Vaalkrans man's descendants have likely been assimilated into mixed-ancestry “Coloured” groups. The Vaalkrans man's genome reveals that Khoekhoe pastoralist groups/individuals lived in the southern Cape as late as 200 years ago, without mixing with non-African colonists or Bantu-speaking farmers. Our findings are also consistent with the model of a Holocene pastoralist migration, originating in Eastern Africa, shaping the genomic landscape of historic and current southern African populations.

KEYWORDS

admixture, genomics, human history

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *American Journal of Physical Anthropology* published by Wiley Periodicals LLC.

1 | INTRODUCTION

The Cape Floristic Region is located on the southwestern tip of Africa between latitudes 31° and 34.5° South and longitudes 18° and 26° East. Towards the south and west the region is surrounded by the Indian and Atlantic Oceans, whereas the interior margin is formed by the more arid Karoo biomes. For its human populations, this region has been an interesting refuge and/or *cul-de-sac* through the ages as attested by its rich Stone Age archeological record with sites occupied from more than a million years ago to as recent as the last century (Lombard et al., 2012). From the fossil record we know that *Homo heidelbergensis* lived here since about 600,000 years ago (Dusseldorp et al., 2013), and at Klasies River Mouth one of the largest (~40 specimens) early *Homo sapiens* collections have been found dating from ~110,000 years ago (Grine et al., 2017). This area's role in the origins of modern human behavior and cognition is also well known from artifacts such as engraved ochre, shell beads, and sophisticated hunting technologies dating to between about 80,000 and 60,000 years ago (Henshilwood et al., 2018; Lombard, 2020; Tylén et al., 2020; Vanhaeren et al., 2019). Furthermore, one of the earliest known records of domestication in South Africa (Henshilwood, 1996) has been attributed to the presence of sheep at Blombos Cave, dated to about 2,000 years ago. Taken together, this long and likely continuous occupation by *Homo* and ancestral *Homo sapiens* in the Cape Floristic Region points to its importance for understanding human history and development.

Today the Khoe-San live in scattered groups across southern Africa, and can be grouped into San (historically foragers/hunter-gatherers) and Khoekhoe (historically herders/pastoralists), a grouping that is reflected in genetic stratification (Montinaro et al., 2017; Schlebusch et al., 2012; Uren et al., 2016; Vicente et al., 2019). They are the descendants of southern African Later Stone Age populations, mixed with East African Later Stone Age pastoralist groups (Breton et al., 2014; Schlebusch et al., 2017). Current-day San descendants often display genetic stratification that can be grouped into Northern San (Kx'a speakers from southern Angola, northern Namibia, and Botswana), Southern San (Tuu speakers from southern Botswana and South Africa) and Central San (Kalahari-Khoe speakers from central Botswana). The Khoekhoe herders likely emerged from San groups mixing between 2,000 and 1,200 years ago with incoming migrant herders from East African (Breton et al., 2014; Macholdt et al., 2014; Schlebusch et al., 2017). Historically, several Khoekhoe-speaking groups persisted throughout the Cape in South Africa (De Jongh, 2016). Their socio-cultural identity, however, became diluted as a result of interactions with incoming European and Asian groups over the last 400 years so that today their descendants form part of the so-called Cape Coloured population (De Jongh, 2016; de Wit et al., 2010; Montinaro et al., 2017; Schlebusch et al., 2012). Thus, teasing apart the factors that contributed to the introduction of domestic animals and the arrival of herding populations in southern Africa remains a challenge (Bollig et al., 2013; Morris et al., 2014; Sadr, 2015; Smith, 2005).

From the DNA of living Khoe-San descendants it is known that the deepest divergence among modern humans was found between them and other Africans and non-Africans (Gronau et al., 2011;

Montinaro et al., 2017; Schlebusch et al., 2012, 2020; Veeramah et al., 2012). The oldest modern human population divergence is currently calculated at 350,000–260,000 years ago based on the genome of a southern African Stone Age hunter-gatherer boy (closely related to Southern San) who lived ~2,000 years ago (before the arrival of herders in that area) compared to genomes of non-Khoe-San individuals (Lombard et al., 2018; Schlebusch et al., 2017). We also know that among Khoe and San groups, the Northern and Southern San are the most genetically distinct from one another, with the Central San being genetically intermediate (Schlebusch et al., 2012; Vicente et al., 2019). The population divergence between the Northern and the Southern San is estimated to be between 150,000 and 190,000 years ago (Schlebusch et al., 2017, 2020).

Recently, it was demonstrated that all living Khoe-San groups, including the Ju|'hoansi, who were long thought to be unaffected by recent migrations and admixture, harbor significant amounts of genetic material coming from mixing with incoming East African pastoralists, Bantu-speaking farmers and/or non-African colonists (Pickrell et al., 2014; Schlebusch et al., 2017; Schlebusch & Jakobsson, 2018). Sometime between 2,000 and 1,200 years ago the genetic make-up of pre-historic southern African individuals change and starts displaying (modest) genetic admixture with incoming East African-Eurasian herding group/s (Breton et al., 2014; Henn et al., 2008; Macholdt et al., 2014; Pickrell & Pritchard, 2012; Schlebusch et al., 2012, 2017; Skoglund et al., 2017). The Nama, a pastoralist Khoe-Kwadi speaking group, further displays a distinct East African-Eurasian genetic component (Breton et al., 2014; Pickrell et al., 2012, 2014; Schlebusch et al., 2012, 2017; Skoglund et al., 2017). Interestingly, at about 1,200 years ago, a pastoralist individual from Kasteelberg in the Western Cape in South Africa was found to have distinctly more East African/Eurasian ancestry compared to Stone-Age groups and to modern-day Khoe-San groups (Skoglund et al., 2017). Furthermore, the lactase persistence allele found in Khoe-Kwadi groups is the same as in East African populations, including the Maasai, and the variant is located on a large unique haplotype block shared between East Africans and Khoe-Kwadi groups (Breton et al., 2014; Macholdt et al., 2014). This lactase persistence variant has been found in high frequencies in pastoralist populations in Africa (Breton et al., 2014; Macholdt et al., 2014; Tishkoff et al., 2007), and the gene is linked to one of the most striking examples of gene-culture co-evolution, where the persistence variants have become important for groups drinking unprocessed milk.

From about 400 years ago, Khoe-San groups of the Cape Floral Region came into contact with a range of European travelers, colonists, settlers and their slaves, brought from Malaysia and other East, South and Southeast Asian regions—all contributing to the complexity of the genetic landscape (Hollfelder et al., 2020; Montinaro et al., 2017; Schlebusch & Jakobsson, 2018). Bantu-speaking farmers from West Africa reached north-eastern South Africa about 1,800 years ago, but never settled in the arid west or the southern winter-rainfall areas including the Cape Floral Region (Lombard et al., 2020). In recent decades, however, descendants of these groups, together with migrants from across Africa became part of the workforce and entrepreneurial

scenery of urban South Africa throughout the region. However, by controlling for recent admixture from European colonists, Bantu-speaking farmers and East African pastoralists, the genetic diversity among Khoe-San groups can be shown to be geographically structured over the greater area of southern Africa, indicating modest levels of gene-flow among Khoe-San groups in pre-historic times (Vicente et al., 2019).

It is still unclear which East African group/s introduced pastoralism to southern Africa and with whom they mixed to form the early southern African Khoekhoe populations (De Jongh, 2016; Morris, 2014). Whereas the effects of mixing with East African pastoralist groups can be seen to a certain extent in the genomes of modern-day Nama (some of whom are still pastoralists), most living Khoe-San groups have been impacted by subsequent admixture events with other groups, including non-Africans (Schlebusch & Jakobsson, 2018). As a result, the genetic relationship between pre-historic Khoekhoe herding groups and autochthonous San hunter gatherer populations became masked (De Jongh, 2016; Montinaro et al., 2017). Genetic data from human remains from archeologically well-characterized sites can improve our understanding of the interactions between groups and the processes related to the introduction of pastoralism to southern Africa. Here we report on the genome obtained from the hair of a person who lived approximately 200 years ago at the Later Stone Age Vaalkrans Shelter (Henshilwood & van Niekerk, 2009), in the Cape Floral Region (Figure 1).

We explored the ancestry of this individual and searched for signs of any mixing with colonists. We further compared the genomic data obtained from the Vaalkrans Shelter hair sample to that of previously published ancient southern African individuals (Figure 2), spanning different times and different contexts to reveal genetic affinities among individuals and groups. Finally, we examined the impact of the migrations that introduced pastoralism to the southern tip of Africa.

2 | RESULTS

2.1 | DNA extraction and sequence results

Following DNA extraction of the hair sample from Vaalkrans Shelter, 8 DNA libraries were built for different dilutions using two different library protocols (see Section 5, Tables 1 and S1). We sequenced the libraries to completion and reached 1.01× genome coverage for the Vaalkrans man. As is expected for hair samples (Gilbert et al., 2006), mitochondrial coverage was very high (4,846×). Contamination estimates were low for each library (Table 1) and the expected deamination patterns for ancient DNA (Sawyer et al., 2012) were distinct (Figure S1). The average read length of merged reads (restricted to read >35 bps) were between 63 and 82 bps, similar to previous sequence length estimates from southern African ancient DNA sequences obtained from bone and teeth (Schlebusch et al., 2017). The Vaalkrans hair sample had more than twice the proportion of human DNA than those extracted from other Stone and Iron Age African bone and teeth samples, for which the highest proportion is 20% (Schlebusch et al., 2017). The molecular sex estimation and the presence of Y-chromosome DNA in the Vaalkrans individual show that the individual was a man (Mittnik et al., 2016; Tables S3 and S4).

2.2 | Uniparental markers

The Vaalkrans man's mitochondrial haplogroup was inferred to be L0d3b1 (Table S2). Haplogroup L0d forms part of the deepest diverging mitochondrial haplogroup clade in humans, namely L0 (L0d is suggested to have emerged ~100,000 years ago), and is found in high frequencies in Khoe-San and Coloured groups (Barbieri et al., 2014; Schlebusch



FIGURE 1 Vaalkrans Shelter before excavation; see Henshilwood and van Niekerk (2009) for detailed archeological context

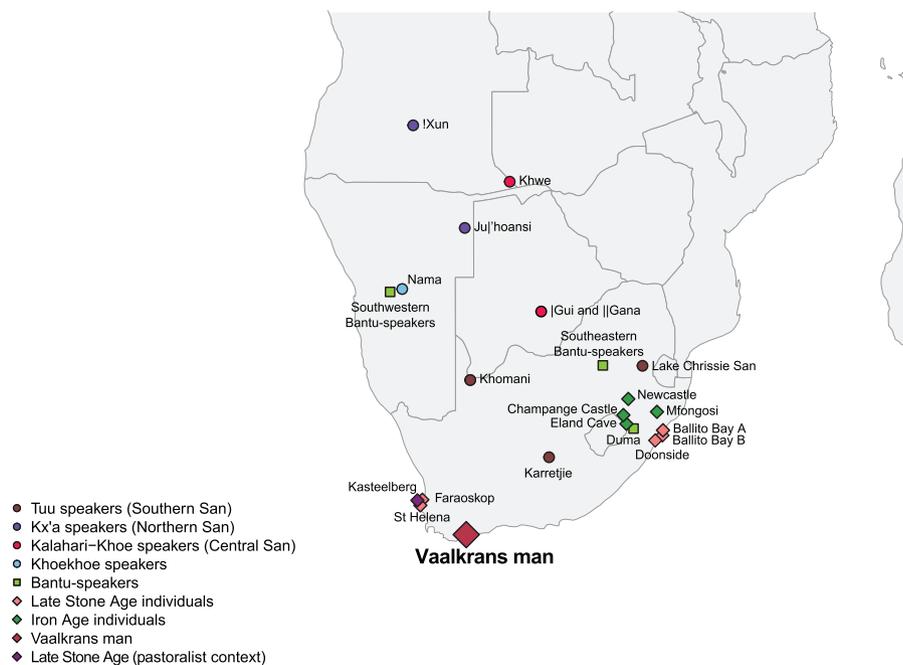


FIGURE 2 Map of southern Africa showing where the remains of prehistoric individuals were excavated and where modern-day groups are known to have resided in historic times. Modern-day Khoe-San populations are displayed by filled circles, Bantu speakers are displayed by squares, and pre-historic/historic individuals are displayed by diamonds

TABLE 1 Summary of sequence results with a focus on the average read length, average proportion human DNA, genome coverage, mitochondrial coverage, and contamination estimates for the libraries sequenced for the Vaalkrans individual

Library	Dilution	Number of libraries	Mean read length	Mean prop. human DNA (%)	Genome coverage	Mitochondrial coverage	Mitochondrial contamination estimate (ContamMix)	Autosomal contamination estimate (%)
Blunt	1:1, 1:2, 1:4, 1:10	4 ^a	63	42	0.25	622.71	0.01	0.075
UDG-treated	1:4	4	82	40	0.76	4223.14	0.01	0.274
Merged data		8	82		1.01	4845.85	0.01	0.485

^aOne library was made for each of four dilutions of this extract.

et al., 2013). Although rare compared to haplogroups L0d1 and L0d2 among Khoe-San and Coloured groups in southern Africa, haplogroup L0d3 has its highest frequency in the current Eastern Cape and Western Cape provinces of South Africa. It is found in highest frequencies in descendants of Southern Khoe-San populations such as the Karretjie people (Schlebusch et al., 2011; Schlebusch et al., 2013). The haplogroup is however wide-spread along the African east coast and is found in low frequencies in individuals from Mozambique and Kenya as well as one occurrence in Kuwait (Schlebusch et al., 2013). A Neighbour Joining tree analyses of the Vaalkrans individuals' mtDNA together with comparative L0d3 sequences from southern Africa, Tanzania and Kuwait (Figure S2) showed that the Vaalkrans individual grouped with other southern African L0d3 sequences and were closely associated with Karretjie and Coloured (from Colesberg) individuals. Today, these groups still live close to the Vaalkrans Shelter. Intriguingly, this haplogroup appears to be absent or at very low frequencies in Khoe-San populations from Namibia, Botswana and Angola. Another individual from a southern African pastoralist context dated to 1200 years ago displayed the L0d1 haplogroup (Skoglund et al., 2017), common among

modern-day Khoe-San groups. Thus far, none of the other DNA-typed southern African ancient individuals belong to the L0d3 lineage found in the Vaalkrans man. Instead Later Stone Age individuals and one Iron Age individual display other L0d haplogroups while other Iron Age individuals belonged to the L3e haplogroup (Morris et al., 2014; Schlebusch et al., 2017; Skoglund et al., 2017).

Analysis of the Y-chromosome of the Vaalkrans man shows that he likely belonged to the E1b1b1 haplogroup with the M35 mutation (E-M35*; Tables S3 and S4). Although the E1b1b haplogroup is widely distributed throughout Africa and southern Europe, the haplogroup E1b1b represented by the M35 mutation includes a paraphyletic subclade defined by an extra mutation, M293, which is more restricted in its distribution to eastern Africa and southern Africa (Henn et al., 2008). The sequence data for the Y-chromosome of the Vaalkrans man did not, however, cover that specific position. The eastern African individual from Mota who lived about 4,500 years ago, excavated from a foraging context, also carried an E1b1 haplotype (Gallego Llorente et al., 2015), although more ancestral than that of the Vaalkrans man, while other individuals from southern Africa dated to

the Later Stone Age display A1b haplotypes (Schlebusch et al., 2017; Skoglund et al., 2017).

2.3 | The genomic ancestry of the Vaalkrans man

We analyzed the genome sequence data of the Vaalkrans man together with Later Stone Age and Iron Age individuals from the KwaZulu-Natal province of South Africa (Schlebusch et al., 2017), Later Stone Age individuals from the Cape province (Skoglund et al., 2017), as well as modern-day southern African Khoe-San and Bantu-speaker groups typed using the Illumina 2.5 Omni array (Schlebusch et al., 2012, 2016), and other modern-day African and non-African groups (Pagani et al., 2012; The 1000 Genomes Project Consortium, 2015), denoted in the “SA-KGP-EA” dataset (see Section 5 and Figure S5 for an overview of the included populations). The locations of the southern African individuals/groups are displayed in Figure 2.

Principal components analysis (PCA) showed the first component separating (ancient and modern) southern Africans from other Africans (Figure 3). The second component separates West (Yoruba) and East (Maasai) Africans. The Later Stone Age individuals excavated at St Helena and Faraoskop Rock Shelter in the Western Cape Province of South Africa group with the Later Stone Age Ballito Bay A and B individuals from KwaZulu-Natal, whereas the Vaalkrans man and the Kasteelberg individual group together near the Nama (the only Khoekhoe-speaking herder group that remain today). The Northern

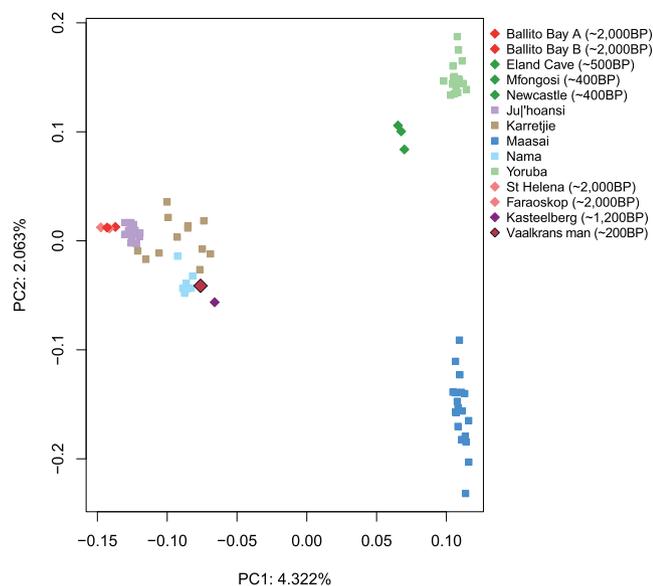


FIGURE 3 Principal components analysis plot of the Vaalkrans man (maroon diamond) and four southern African Later Stone Age individuals and three southern African Iron Age individuals (Schlebusch et al., 2012; Skoglund et al., 2017). The Kasteelberg individual was found in a pastoralist context (purple diamond). Both the Vaalkrans man and the Kasteelberg individual group with the (modern-day) Nama, a pastoralist Khoe-San group, whereas the other Later Stone Age individuals (Ballito Bay A and B, St Helena and Faraoskop) group together, at the extreme end of PC1

San (Ju|'hoansi) form a tight group between the Stone Age southern African individuals and the Southern San (Karretjie People). The PCA results are reflected in f_3 -statistic analyses where the Vaalkrans man shows affinity to the Nama (Figure S3B).

Admixture proportions were also summarized for the SA-KGP-EA dataset (Figures 4 and S5), and the Vaalkrans man's ancestry consists of a major Southern Khoe-San component and an East African component (15–32%). The Later Stone Age individuals from southern Africa who lived 2,000 years ago are genetically similar to modern-day southern Khoe-San groups and the Vaalkrans man, but without the East African admixture component. The genetic makeup of the 1,200-year-old remains of the pastoralist from Kasteelberg is similar to that of the Vaalkrans man, but carries a slightly greater fraction East African ancestry (24–29%, Figures 4, S4F, S5). Both these individuals grouped close to the Nama Khoekhoe population on the PCA, which is consistent with the observation that the Nama has the highest proportion of East African ancestry among all the modern-day Khoe-San groups (Figure 4, Schlebusch et al., 2017).

2.4 | The origin of the East African ancestry component

We used f -statistics to investigate the potential source of East African admixture in the Vaalkrans individual. The model comprised of chimpanzee as an outgroup, the Vaalkrans individual, the Ju|'hoansi representing the Khoe-San component, and another African population/ancient individual (Amhara, Somali, Oromo, Maasai, the Mota individual, Yoruba, Luhya, Southeast Bantu speakers or Southwest Bantu speakers) as the fourth group.

The Amhara as the source East African population gives the highest D -value (Figure 5). The negative D -values for the model when West African (Yoruba, or groups of West African descent; Luhya, Southeast Bantu speakers or Southwest Bantu speakers) populations are included as the fourth population suggest that the Vaalkrans man does not share ancestry, nor was he admixed with the ancestors of these populations (Figures 5, S4A and Table S5). Estimates of admixture fractions, f_4 -ratios (Table S6), result in qualitatively similar observations when we use Amhara as the donor population for the East African migrants to southern Africa.

We further examined the relationships of the Vaalkrans man's ancestry components using an explicit model approach (qpGraph). We first built a model to verify the source of admixture into the Vaalkrans man (Figure S4D) based on a southern African ancestral source population, related to the Ballito Bay A boy from KwaZulu-Natal and an early East African source population related to the Mota individual. This model resulted in an East African admixture fraction of 32% (z -score -0.062) for the Vaalkrans man. The same model was used for estimating the East African admixture fraction for the Kasteelberg individual, resulting in a similar outcome (Figure S4F, East African ancestry fraction 29%, z -score 0.092). An alternative (similar) model that tests West African (i.e., Bantu-speaker) admixture instead of East African as the admixture source into the Vaalkrans man could be rejected (Figure S4A, z -score 66.5). A Model with a direct contribution from a European source

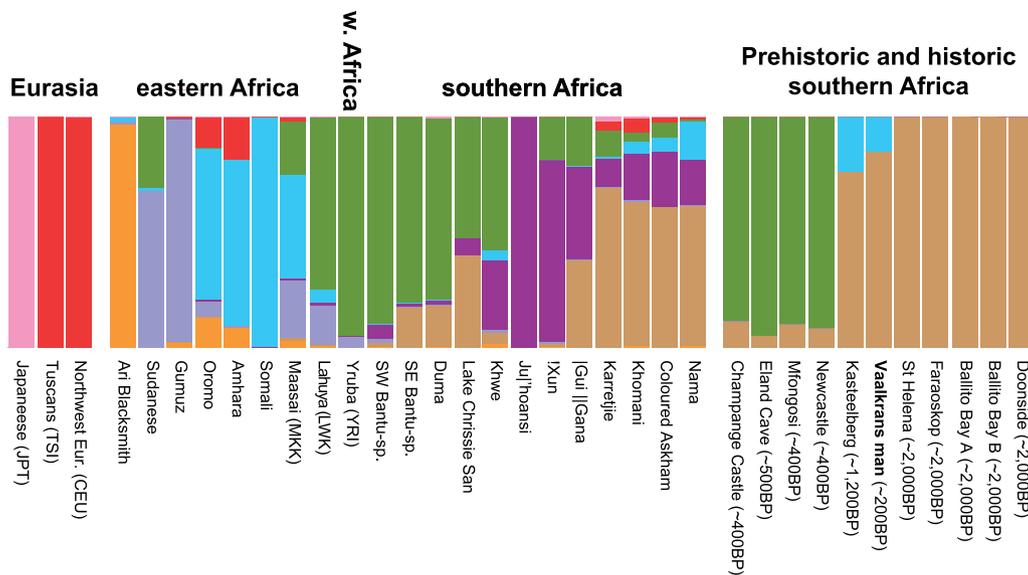


FIGURE 4 Estimated ancestry components, assuming eight ancestral clusters, for a subset of relevant groups and individuals, extracted from the SA-KGP-EA dataset. The ancestry components in the figure for modern-day populations display population averages. The results for the complete set of individuals, and other assumptions on the number of ancestral clusters, can be seen in Figure S5. The “non-Khoe-San” component (blue) and the “Khoe-San” component (brown) is very similar assuming 7 to 11 ancestral clusters, see figure S5

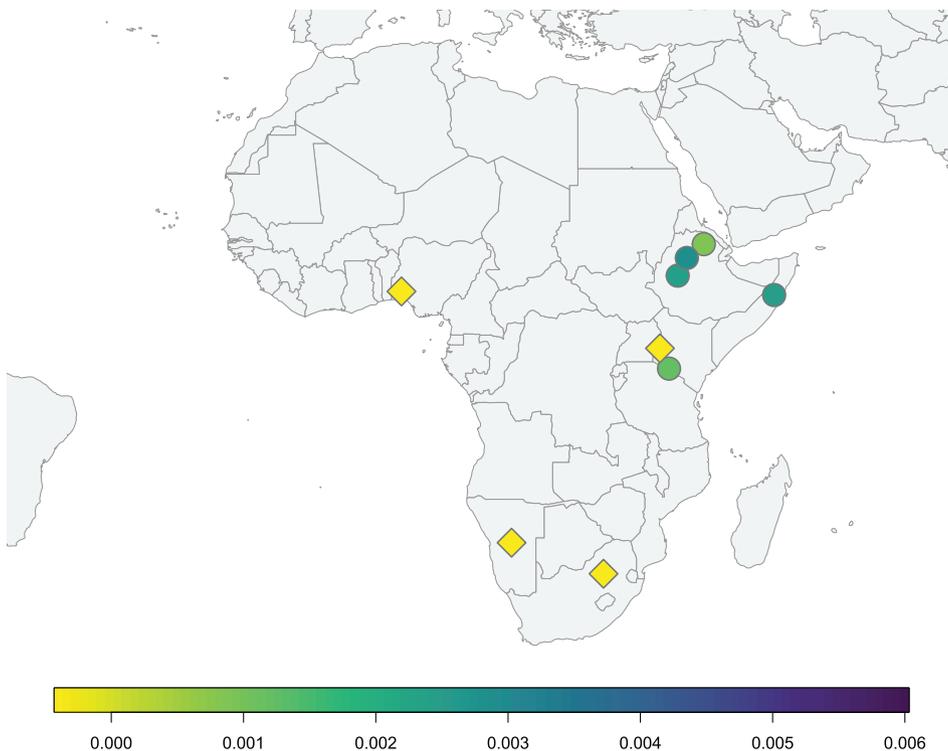


FIGURE 5 Genetic similarity of the “non-Khoe-San” component of the Vaalkrans man’s genetic make-up among East and West African groups. The D -values for the computation D (chimpanzee, Vaalkrans man; Jul’hoansi, test group) are indicated with a color on the map. The test populations (from the SA-KGP-EA dataset) include East African populations, the Mota individual, West African Yoruba, and two Bantu-speaking groups from South Africa. The best match occurs with the Amhara from Ethiopia. Bantu-speaking populations are indicated by diamond symbols

into the Vaalkrans man was also rejected (Figure S4E). We further investigated alternative sources to the East African component, including some fraction of Neolithic Europeans, and these models (Figures S4B and S4C) give slightly worse fits to the data, but are still not statistically rejected. Hence, we conclude that the source of the East African component in the Vaalkrans and modern-day Khoe-San groups is related to, but somewhat different from the Amhara of today (Schlebusch et al., 2017; Skoglund et al., 2017).

2.5 | The East African ancestry over time

The inferred East African admixture fractions dramatically increase between 2,000 and 1,200 years ago in southern African indigenous groups (Figure 6). The Ballito Bay, Faraoskop, and St Helena individuals would have lived before the arrival of East African pastoralists to southern Africa, and therefore they would have no East African ancestry (Schlebusch et al., 2017). The Kasteelberg individual, dated to

1,200 years ago, and excavated from a pastoralist context, lived some time after the East African migration to southern Africa. Interestingly, the greatest level of the East African ancestry component among southern African individuals was not observed for the present day pastoralists (e.g., Nama), instead it was greater some 1,200–2000 years ago. This may be a reflection of an initial migration and admixture of East African pastoralists followed by subsequent and continuous admixture with surrounding Khoe-San groups. It has been shown that all modern-day Khoe-San groups have admixture with various groups, including non-Africans (Schlebusch et al., 2012, 2017), making fine-scale interpretations of the past migration and admixture events based on present-day populations difficult. The Vaalkrans man does not show recent admixture from non-African or West African groups, verifying the East African ancestry component's existence in the southern Cape until historic times (Figure 6).

2.6 | Lactase persistence

The SNPs rs4988235 (13910*T, with the derived variant common among Europeans), rs41525747 (13907*G, common in Ethiopia), rs869051967 (14009*G, a variant common in Arab groups from Africa) and rs41380347 (13915*G, a variant common among Middle Easterners), and rs145946881 (14010*C, with the derived variant common among Kenyans and Tanzanians) are all associated with lactase persistence (in the *MCM6/LCT* gene complex) (Schlebusch & Jakobsson, 2018; Tishkoff et al., 2007). The genome data for the Vaalkrans man showed one derived allele covering the SNP rs145946881 (14010*C) that is associated with lactase

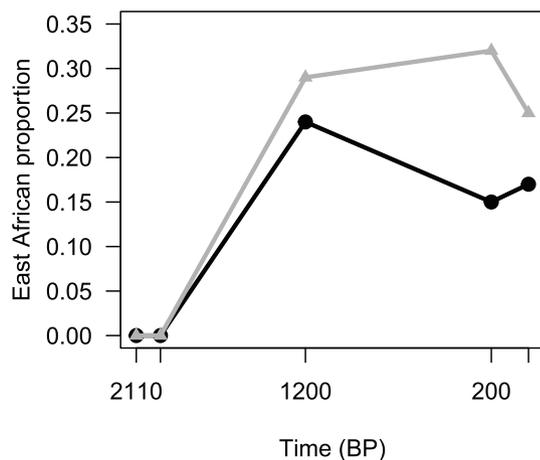


FIGURE 6 Proportion of East African ancestry in southern African individuals/groups as a function of time. The fraction of East African ancestry dramatically increases for southern African individuals between 2,000 and 1,200 years ago. Gray represents values estimated using qpGraph and black represents values from admixture (the blue “non-Khoe-San” component in Figure 4). The point-estimates of East African ancestry in the graph were taken from the following individuals (from oldest to youngest): Ballito Bay B, Ballito Bay A, Kasteelberg, Vaalkrans, and (present-day) Nama as a group

persistence (Table S7). His genome also covered two other lactase associated SNPs, and for these the non-persistence variants were found, Table S7). The 14010*C variant at rs145946881 has been found in relative high frequencies in the Nama, between 20% and 36%, and at low frequencies (2.5–10%) in San hunter gatherer populations (Breton et al., 2014; Macholdt et al., 2014). The presence of the East African lactase persistence variant is a further indication of the partial East African pastoralist ancestry of the Vaalkrans man.

3 | DISCUSSION

Genetic investigations of historic and prehistoric individuals from southern Africa reveals an interesting and complex human migration history for the region. The Vaalkrans man displays a dual ancestry, with one part related to southern San groups and one part related to East Africans. Similar ancestries are seen in modern-day Nama Khoekhoe, and are also reflected in the Kasteelberg pastoralist individual who lived 1,200 years ago. f_4 -Statistics verify that the ancestry of the Vaalkrans man as well as all modern-day Khoe-San groups is the result of admixture between Later Stone Age southern African foragers and an East African (with some input from a Eurasian group) group similar to the Amhara (Table S6). Ancestry results are also reflected in the uniparental markers for the Vaalkrans man, who exhibited a mitochondrial haplogroup common among southern Khoe-San and a Y-chromosome variant that is common among East African pastoralist groups and the Nama (Tables S3 and S4, Henn et al., 2008, Schlebusch, 2010). The Vaalkrans man also carried a lactase persistence variant, which is common among (some) East African pastoralist populations and the Nama (Breton et al., 2014; Macholdt et al., 2014). Thus, the Vaalkrans man was likely part of the Cape Khoekhoe pastoralist populations whose genetic makeup of East African intogression dates to before 1200 years ago (Skoglund et al., 2017). The Nama (the only remaining Khoekhoe pastoralist population) has a small fraction of additional Eurasian ancestry, which is likely a result of recent admixture (Pickrell et al., 2012; Schlebusch et al., 2012), but this ancestry component is not present in the two ancient pastoralist individuals from southern Africa (Vaalkrans and Kasteelberg, Figure 4).

Despite having a younger age than the Iron Age individuals from KwaZulu-Natal (Schlebusch et al., 2017), who displayed a predominant West African ancestry (Figure 1 in Schlebusch et al., 2017), the Vaalkrans man shows no West African ancestry. Current-day Nama individuals display small amounts of West African ancestry, likely from recent admixture (Pickrell et al., 2012; Schlebusch et al., 2012). Bantu speakers started to arrive in the northern parts of today's South Africa ~1800 years ago and gradually migrated southwards (Huffman, 1989; Steyn et al., 2019), admixing with Khoe-San populations along the way (Schlebusch et al., 2017). By the 1600s, Bantu speakers occupied much of the eastern part of southern Africa up to the Great Fish River in the Eastern Cape. The Western parts of South Africa and the Western Cape was occupied by Khoekhoe herders (Schlebusch et al., 2017;

Schlebusch & Jakobsson, 2018). Today, their descendants are likely assimilated into the mixed Coloured population of the region. While some groups such as the southern Cape Hessequa retain their Khoekhoe identity (De Jongh, 2016; Henshilwood & van Niekerk, 2009), they appear to have mixed with other groups (De Jongh, 2016; C. Henshilwood & van Niekerk, 2009). These groups are related to the Vaalkrans man, and are possibly among his descendants.

The distinct Southern San affinity for the Vaalkrans and Kasteelberg individuals (in contrast to Northern San) implies that gene flow was hindered/restricted between incoming East African herder and San groups during earlier times. In addition, most modern-day Khoe-San groups display admixture with West Africans and/or non-Africans (Figures 3 and 4), which is absent in the Vaalkrans and Kasteelberg individuals. Even though only a handful of historic/prehistoric individuals from southern Africa have been genetically investigated at this date (as opposed to the DNA data of modern-day groups represented by many individuals), the genetic data from the Vaalkrans man is clear; he traces his genetic ancestry back to East Africans and the Southern San and he was likely a Khoekhoe herder who lived 200 years ago.

4 | CONCLUSION

The genome of the Vaalkrans man could be sequenced from a hair sample to 1.01x coverage, resulting in rich genetic information about the ancestry of this individual. Our results further support the finding that East African herders arrived in the southernmost tip of Africa between 2,000 and 1,200 years ago, which led to the formation of the southern African Khoekhoe herding populations by mixing with the ancestors of the Southern San. Analysis of the pastoralist DNA from Kasteelberg, Vaalkrans and the only surviving pastoralist population today, the Nama, reveals a distinct genetic signature for southern African herder groups comprising East African and Southern San ancestry that has prevailed through time to the present day. The East African component is genetically similar to the current-day Amhara groups.

While the information from the genetics of historic and prehistoric individuals is beneficial for archeological investigations, as it adds genetic data along the temporal dimension, these data are also crucial for understanding modern-day population genetic investigations. For instance, the variable East African ancestry among current day Khoe-San groups will confound and potentially cause misleading results for many demographic investigations of the genetic landscape and human history earlier than 2,000 years ago if the East African component is not accounted for (e.g., Vicente et al., 2019). It can cause underestimates of population divergences or overestimates of genetic affinities among groups. Nonetheless, the time horizon on this East African component among all Khoe-San groups, and its magnitude, was first revealed by studying 2000-year-old remains of Later Stone Age southern Africans (Schlebusch et al., 2017; Skoglund et al., 2017). With the sequencing of additional, and perhaps older, ancient

individuals from Africa, the complex genetic structure and history of prehistoric populations of Africa may be better understood. In particular, our study exemplifies how state-of-the-art paleo-genomic approaches can help in understanding the context of archeological sites that do not present skeletal human remains, by examining remains such as human hair.

5 | METHODS

5.1 | The Vaalkrans Shelter

Vaalkrans Shelter is in the De Hoop Nature Reserve located along a rocky shoreline on the southern Cape coast of South Africa, approximately 20 m from the Indian Ocean and 11 m above sea level. The hair was recovered in quadrat N5c from the topmost layer AA that consists of medium brown to tan colored compacted eolian sand containing a dense concentration of shellfish. Similar clumps of hair were recovered from other layers at the site. Layer AA was not radiocarbon dated. Two charcoal samples from Square M5b, Layer ACA and Square N5a, Level ACB, provided radiocarbon ages of 220 ± 45 (Pta-9192) years BP and 140 ± 35 (Pta-9187) years BP respectively (C. Henshilwood & van Niekerk, 2009). The hair sampled for this study is probably slightly younger than these dates. Henshilwood (C. S. Henshilwood, 2008) showed, based on historical records of European travelers in the southern Cape region from the late 1600s, that indigenous Khoe-San people were living as herders and coastal hunter-gatherers in the De Hoop/Riversdale Plain area into the 1800s. The presence of copper and an iron chunk in the upper layers, close to where the hair was found, suggest the final occupants of the site may have had contact with European material culture or people, perhaps also through salvage from Portuguese or Dutch shipwrecks that litter this coast.

Organic preservation at Vaalkrans is excellent. Plant remains, particularly geophytes, are well preserved, as are tortoise scutes. The most abundant remains at the site are shellfish, particularly *Perna perna* (brown mussel), *Turbo sarmaticus* (giant periwinkle), *Diloma* species (periwinkles), *Haliotis midae* (abalone), and various limpet species. Other faunal remains are far less numerous compared to the shellfish remains. Most bones are from small animals such as hyrax (*Procavia capensis*), dune mole rats (*Bathyergus suillus*) and tortoise (*Chersina angulata*). Fish and snake bones were also present. No human bone remains were found at the site.

Only a few stone artifacts were found, mostly informal quartz/quartzite pieces (C. Henshilwood & van Niekerk, 2009), sometimes associated with herder sites (Beaumont et al., 1995; Orton & Parsons, 2018), which may be further implied by the presence of a thin-walled pottery sherd found at Vaalkrans (C. Henshilwood & van Niekerk, 2009). The close-by sites of Klipdrift Shelter, was occupied between about 65,000 and 59,000 years ago during the Middle Stone Age (C. S. Henshilwood et al., 2014), and Klipdrift Cave occupied between 13,700 and 10,700 years ago during the Later Stone Age (Ryano et al., 2017), serve as evidence for the time depth of human occupation and the presence of ancient hunter-gatherer populations

on this landscape. Today, the descendants of the Hessequa Cape Khoekhoe still live in the region (De Jongh, 2016).

5.2 | DNA-extraction

It has been shown previously that hair (both from modern-day individuals and ancient material) is resistant to contamination and/or can be easily decontaminated, making hair a good source for endogenous DNA (Gilbert et al., 2006). As hair grows, hair shaft cells become keratinized, providing a physical barrier between the endogenous DNA in the keratinized cells and the outside environment. Hair is also hydrophobic, which prevents damage of the endogenous DNA (Gilbert et al., 2006).

One small tuft of circa 1 cm hair shafts was washed in 0.5% sodiumhypochlorite (NaOCl) and washed three times in UV irradiated DNA-free water. A total of 1 mL of digestion buffer containing 10 mM Tris-HCl (pH 8.0), 10 mM NaCl₂, 2% (w/v) sodium dodecyl sulfate (SDS), 5 mM CaCl₂, 2.5 mM ethylene-diamine-tetra-acetic acid (EDTA) (pH 8.0), 40 mM dithiothreitol (DTT) and 10 mg/mL proteinase K solution (Life Technologies) was added to the hair shaft sample followed by incubation in 14 h at 37°C (Gilbert et al., 2004; Vilstrup et al., 2013). After 14 h of incubation, an additional 10 mg/mL Proteinase K was added followed by incubation at 55°C until the hair shaft was completely digested. The DNA-containing solution was concentrated to 200 µL using Amicon Ultra-4 PLTK Ultracel-PL membrane 30 kDa (Millipore) and purified using Minelute PCR purification kit according to manufacturer's protocol and eluted in 110 µL Eb-buffer (Qiagen).

5.3 | Library preparation

Blunt-end Illumina multiplex sequence libraries were prepared from 20 µL of extracted DNA excluding the fragmentation step (Günther et al., 2015; Meyer & Kircher, 2010; Schlebusch et al., 2017). To remove inhibitors, the DNA extract was diluted 1:2, 1:4, and 1:10 in distilled and UV-irradiated H₂O. A total of four sequencing libraries were prepared including two library negatives. To determine the optimal number of PCR cycles for the amplification of the libraries, quantitative-PCR (qPCR) was performed. The sequencing libraries were amplified in duplicate reactions in a total volume of 50 µL including negative library- and PCR controls (one for each sample). The cycling conditions were 12 min at 94°C, followed by 20 cycles of 30 s at 94°C, 30 s at 60°C and 45 s at 72°C and a final extension step for 10 min at 72°C.

In addition, four damage-repair sequencing libraries were prepared from 1:4 dilution and treated with USER enzyme in order to repair post-mortem deaminated sites (Briggs et al., 2010). The libraries were amplified in duplicate reactions using Accuprime DNA polymerase (Thermo Fisher Scientific) in a total volume of 50 µL. The cycling conditions were 2 min at 95°C, followed by 20 cycles of 15 s at 95°C, 30 s at 60°C and 60 s at 68°C and a final extension step for 5 min at 68°C (Gansauge & Meyer, 2013). For both Blunt-end and USER-

treated sequencing libraries, duplicate samples were pooled and purified using Agencourt AMPure XP beads (Beckman Coulter) according to manufacturer's protocol and eluted in 40 µL TET buffer. The quality of the purified libraries was analyzed on a TapeStation 2200 using the High Sensitivity D1000 Kit (Agilent Technologies) and the quantity was measured using Qubit 3.0 (Life Technologies). Samples were pooled at equimolar concentrations and sequenced on the HiSeqX system (Illumina) with 150 bp paired-end v2.5 chemistry.

Libraries were sequenced on 1/5th or 1/6th of an Illumina HiSeq XTen lane at SciLifeLab SNP&Seq Platform in Uppsala using paired-end 150 bp chemistry. As the DNA libraries contained substantial levels of human DNA and the fraction of very short (<35 bp) fragments was modest for the initial sequencing results, all libraries were re-sequenced on 1/8th of an Illumina HiSeq XTen lane to increase genome coverage.

5.4 | Sequence analysis and preparation for PCA and admixture

Paired-end sequence reads were merged and trimmed, and mapped against the human reference genome (build 37) (Günther et al., 2018; Kircher, 2012). PCR duplicates were collapsed to make consensus sequences for each set of fragments with identical start and stop position. Sequences were further merged to library level depending on the type of library method used to generate sample libraries (blunt-end or damage-repair libraries) to be quality assessed. The blunt-end and damage-repair libraries were finally merged to create one final sequence-file for subsequent population genetic analyses.

Deamination patterns were generated and checked for each library (Sawyer et al., 2012). Contamination estimates were calculated using ContamMix, where sample mitochondrial sequences are compared to 311 reference mitochondrial genomes in order to determine whether sample mitochondrial genomes map better to one of the reference genomes or not (Fu et al., 2013). Estimates are given as an estimate of authenticity as opposed to an estimate of contamination seen in methods such as that of Green et al. (2008). Autosomal contamination was also estimated using VerifyBamID (Jun et al., 2012). Genetic sex determination followed the method developed in Mitnik et al. (Mitnik et al., 2016), where ratios of sequence fragments mapping to the X and autosomal chromosomes is used to determine whether the sample is female or male.

A mitochondrial consensus sequence was constructed for each library-level and individual-level merged sample sequence using ANGSD (minimum base quality score of 30, and a requirement of at least 3 reads to verify the DNA sequence) (Korneliusson et al., 2014). The online haplotyping tool "Haplofind" (Vianello et al., 2013), and PhyloTree mtDNA Build 17 (February 18, 2016) (van Oven, 2015), were used to call haplotypes for all three libraries to ensure a reliable mitochondrial haplotype call. Haplotype/s are reported against the Reconstructed Sapiens Reference Sequence (RSRS; Behar et al., 2012).

A mtDNA phylogenetic tree was constructed by collating the Vaalkrans mtDNA sequence with; all LO haplotypes identified in

Schlebusch et al. (2013), all L0d3 sequences published in Schlebusch et al., 2013, all Tanzanian and Kuwait L0d3 sequences published in Behar et al. (2008) and Gonder et al. (2007). Since most L0d3 comparative sequences were only sequenced for HVRI and II, only these regions were analyzed. Sequences were aligned to the Reconstructed Sapiens Reference Sequence and a Neighbour Joining tree was constructed in Mega v.4. A Neanderthal sequence were used to root the tree.

Y-chromosome variants were called using the samtools mpileup function (Samtools v 1.3) (Li & Durbin, 2009). Y-chromosome haplotype was called using PhylotreeY (van Oven, 2015) as well as ISOGG (International Society of Genetic Genealogy) (v 11.224, <http://isogg.org>). Haplotype analysis excluded transition sites from sample bam files, and only used sequence data with a minimum base and mapping quality score of 30 for each site. Nomenclature from the ISOGG database was used to present the haplotype for the Vaalkrans sample.

To be able to perform population genetics analysis on the Vaalkrans man, as the data contained both blunt-end and damage-repair data, the sequence data was merged together with the comparative dataset of Schlebusch et al. (2017), as well as additional ancient southern African individuals, from St Helena, Faraoskop and Kasteelberg in South Africa (Skoglund et al., 2017). In brief, the comparative dataset (referred to as SA-KGP-EA dataset) was composed of the ancient southern African Later Stone Age and Iron Age individuals from the KwaZulu-Natal province (SA) (Schlebusch et al., 2017), southern African Khoe-San and Bantu-speaker groups typed using the Illumina 2,5 Omni array (Schlebusch et al., 2012, 2016), 6 populations from the 1000 genomes database (LWK, MKK, YRI, TSI, JPT, CEU) (The 1000 Genomes Project Consortium, 2015), and 6 East African populations typed on the Illumina 1 M Omni array (SOMALI, AMHARA, OROMO, GUMUZ, ARI-BLACKSMITH, SUDANESE) (Pagani et al., 2012). Further details on this comparative dataset can be found in Schlebusch et al. (2017).

5.5 | Population genetic analyses

Principal components analysis was carried out on a haploidized version of the merged dataset of the Vaalkrans man along with the SA-KGP-EA dataset, as well as the 3 ancient southern African shotgun-sequenced individuals from Skoglund et al. (2017). To create a haploid dataset, one allele is chosen randomly at each locus. The program package EIGENSOFT and smartpca was used to generate a PCA using only transversion sites for each of the ancient samples (Patterson et al., 2006; Price et al., 2006). The Vaalkrans man and the ancient individuals from Schlebusch et al. (2017) and Skoglund et al. (2017) were projected onto the data of modern-day individuals. While the above comparative dataset compares the Vaalkrans man in a broader context, we wanted to investigate his ancestry in a more specific African context, therefore a PCA was generated for the merged dataset of the Vaalkrans man with a selection of different modern Khoe-San populations, as well as populations from West Africa and other

southern African ancient individuals. Genetic affinities among combinations of the Vaalkrans man and other individuals from the SA-KGP-EA dataset were investigated using D-statistics.

The program ADMIXTURE was used to infer the ancestry components for the Vaalkrans individual, previously published prehistoric individuals and modern-day individuals. Before analysis, the dataset was LD pruned in PLINK 1.9 (<https://www.cog-genomics.org/plink2>) with the parameters “-indep-pairwise 200 25 0.4.” Fifty iterations were run for $K = 2$ to $K = 15$. Results from ADMIXTURE were summarized for each K using PONG (Alexander et al., 2009; Behr et al., 2016). Results were visualized using R (R Core Team, 2017). A summary admixture graph was generated using R and the Q-values of the best iteration run for the k -values which exhibited enough resolution to show Northern and Southern Khoe-San groups as separate ancestry components.

Genetic affinities (focusing on the East African component in modern-day Khoe-San individuals and prehistoric individuals from southern and eastern Africa [Gallego Llorente et al., 2015; Schlebusch et al., 2012, 2017; Skoglund et al., 2017]) were investigated further using D-tests (Figure 5) and f_4 -statistics (Patterson et al., 2012). One of the modern-day East African populations, as well as the Stone Age East African Mota individual was used to represent the East African component in order to determine whether the East African cluster for these samples was a mixed Eurasian/East African population, or an East African population without Eurasian admixture, and which population was the best proxy as the source of East African admixture into the above samples. f_4 -Ratio statistics was then calculated to establish the proportion of East African ancestry in each modern and ancient Khoe-San population (see supplemental material). QpGraph (Patterson et al., 2012) was also utilized to model a population history connecting the different populations that form the admixture proportions of the Vaalkrans individual (Figure 6 and see supplemental material).

The Vaalkrans individual's genome was analyzed for the presence of lactase persistence variants using the samtools mpileup function where a minimum base and mapping quality of 30 was used for sample reads (Samtools v 1.3) (Li & Durbin, 2009). Lactase persistence variants affiliated with lactase persistence in European, Middle Eastern, and African populations were investigated in the sample for the specific positions of these SNPs (Schlebusch & Jakobsson, 2018; Tishkoff et al., 2007).

ACKNOWLEDGEMENTS

We thank P. Keene and W. Black at the IZIKO South African Museum for help with sampling the archeological hair and T. Günther for technical assistance. This project was supported by grants from the Knut and Alice Wallenberg Foundation (to M.J.), the Swedish Research Council (no. 642-2013-8019 to M.J. and no. 621-2014-5211 to C.S.), the Göran Gustafsson Foundation (to M.J.), the European Research Council to C.S. (no. 759933) and an African Origins Platform grant from the South African National Research Foundation (to M.L.). Excavations at Vaalkrans Shelter were supported by grants from the University of Bergen, Norway. This work was partly supported by the

Research Council of Norway through its Centres of Excellence funding scheme, Centre for Early Sapiens Behaviour (SapienCE), project number 262618. Sequencing was performed at SciLifeLab, National Genomics Infrastructure, Uppsala, and computations were performed at Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX). Research permits for sampling remains of hair from Vaalkrans were obtained from Heritage Western Cape (14070301GT0704E) and IZIKO Museums.

AUTHOR CONTRIBUTIONS

Alexandra Coutinho: Data curation; formal analysis; investigation; methodology; project administration; validation; visualization; writing-original draft. **Helena Malmström:** Conceptualization; data curation; formal analysis; investigation; methodology; supervision; validation; visualization; writing-review and editing. **Hanna Edlund:** Data curation; investigation; methodology; project administration; supervision; writing-review and editing. **Christopher Henshilwood:** Conceptualization; investigation; project administration; resources; writing-review and editing. **Karen van Niekerk:** Investigation; resources; writing-review and editing. **Marlize Lombard:** Conceptualization; investigation; project administration; supervision; writing-original draft; writing-review and editing. **Carina Schlebusch:** Conceptualization; data curation; formal analysis; investigation; methodology; software; supervision; visualization; writing-original draft; writing-review and editing. **Mattias Jakobsson:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; software; supervision; validation; visualization; writing-original draft; writing-review and editing.

DATA AVAILABILITY STATEMENT

Data are available from the European Nucleotide Archive (EBI-ENA) under accession no. PRJEB40005.

ORCID

Mattias Jakobsson  <https://orcid.org/0000-0001-7840-7853>

REFERENCES

- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Barbieri, C., Güldemann, T., Naumann, C., Gerlach, L., Berthold, F., Nakagawa, H., Mpoloka, S. W., Stoneking, M., & Pakendorf, B. (2014). Unraveling the complex maternal history of Southern African Khoisan populations. *American Journal of Physical Anthropology*, 153(3), 435–448. <https://doi.org/10.1002/ajpa.22441>
- Beaumont, P. B., Smith, A. B., & Vogel, J. C. (1995). Before the Einiqua: The archaeology of the frontier zone. *Einiqualand: Studies of the Orange River Frontier*, 236, 264.
- Behar, D. M., Van Oven, M., Rosset, S., Metspalu, M., Loogväli, E. L., Silva, N. M., ... Vilems, R. (2012). A “copernican” reassessment of the human mitochondrial DNA tree from its root. *American Journal of Human Genetics*, 90(4), 675–684. <https://doi.org/10.1016/j.ajhg.2012.03.002>
- Behr, A. A., Liu, K. Z., Liu-Fang, G., Nakka, P., & Ramachandran, S. (2016). Pong: Fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics*, 32(18), 2817–2823. <https://doi.org/10.1093/bioinformatics/btw327>
- Bollig, M., Schnegg, M., & Wotzka, H.-P. (2013). *Pastoralism in Africa: Past, present, and futures*. Berghahn Books.
- Breton, G., Schlebusch, C. M., Lombard, M., Sjödin, P., Soodyall, H., & Jakobsson, M. (2014). Lactase persistence alleles reveal partial east African ancestry of southern African Khoe pastoralists. *Current Biology*, 24(8), 852–858. <https://doi.org/10.1016/j.cub.2014.02.041>
- Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., & Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Research*, 38(6), e87–e87. <https://doi.org/10.1093/nar/gkp1163>
- De Jongh, M. (2016). *A Forgotten First People: The Southern Cape Hessequa*. Watermark Press. Retrieved from https://books.google.se/books?id=g_RrnQAACAAJ
- de Wit, E., Delport, W., Rugamika, C. E., Meintjes, A., Möller, M., van Helden, P. D., Seoighe, C., & Hoal, E. G. (2010). Genome-wide analysis of the structure of the South African Coloured population in the Western Cape. *Human Genetics*, 128(2), 145–153. <https://doi.org/10.1007/s00439-010-0836-1>
- Doron, M. B., Richard, V., Himla S., Jason, B.S, Luisa, P., Ene, M., Rosaria, S., Heeran, M., Shay, T., David, C., Jaume, B., Lluis, Q.M., Chris Tyler-Smith, R. Spencer Wells, Saharon Rosset. (2008). The Genographic Consortium. The Dawn of Human Matrilineal Diversity. *American Journal of Human Genetics*, 82(5), 1130–1134.
- Dusseldorp, G., Lombard, M., & Wurz, S. (2013). Pleistocene Homo and the updated Stone Age sequence of South Africa. *South African Journal of Science*, 109(5/6), #0042.
- Fu, Q., Mittnik, A., Johnson, P. L. F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., Ronchitelli, A. M., Martini, F., Cremonesi, R. G., Svoboda, J., Bauer, P., Caramelli, D., Castellano, S., Reich, D., Pääbo, S., & Krause, J. (2013). A revised time-scale for human evolution based on ancient mitochondrial genomes. *Current Biology*, 23(7), 553–559. <https://doi.org/10.1016/J.CUB.2013.02.044>
- Gallego Llorente, M., Jones, E. R., Eriksson, A., Siska, V., Arthur, K. W., Arthur, J. W., ... Manica, A. (2015). Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science (New York, N.Y.)*, 350(6262), 820–822. <https://doi.org/10.1126/science.aad2879>
- Gansauge, M.-T., & Meyer, M. (2013). Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols*, 8(4), 737–748. <https://doi.org/10.1038/nprot.2013.038>
- Gilbert, M. T. P., Menez, L., Janaway, R. C., Tobin, D. J., Cooper, A., & Wilson, A. S. (2006). Resistance of degraded hair shafts to contaminant DNA. *Forensic Science International*, 156(2–3), 208–212. <https://doi.org/10.1016/j.forsciint.2005.02.021>
- Gilbert, M. T. P., Wilson, A. S., Bunce, M., Hansen, A. J., Willerslev, E., Shapiro, B., Higham, T. F. G., Richards, M. P., O'Connell, T. C., Tobin, D. J., Janaway, R. C., & Cooper, A. (2004). Ancient mitochondrial DNA from hair. *Current Biology*, 14(12), R463–R464. <https://doi.org/10.1016/j.cub.2004.06.008>
- Gonder, M. K., Mortensen, H. M., Reed, F. A., de Sousa, A., & Tishkoff, S. A. (2007). Whole-mtDNA genome sequence analysis of ancient African lineages. *Molecular Biology and Evolution*, 24, 757–768.
- Green, R. E., Malaspinas, A. S., Krause, J., Briggs, A. W., Johnson, P. L. F., Uhler, C., Meyer, M., Good, J. M., Maricic, T., Stenzel, U., Prüfer, K., Siebauer, M., Burbano, H. A., Ronan, M., Rothberg, J. M., Egholm, M., Rudan, P., Brajković, D., Kučan, Ž., ... Pääbo, S. (2008). A complete neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell*, 134(3), 416–426. <https://doi.org/10.1016/j.cell.2008.06.021>
- Grine, F. E., Wurz, S., & Marean, C. W. (2017). The middle stone age human fossil record from Klasies River Main site. *Journal of Human Evolution*, 103, 53–78. <https://doi.org/10.1016/J.JHEVOL.2016.12.001>

- Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., & Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics*, 43(10), 1031–1034. <https://doi.org/10.1038/ng.937>
- Günther, T., Malmström, H., Svensson, E. M., Omrak, A., Sánchez-Quinto, F., Kılınc, G. M., Krzewińska, M., Eriksson, G., Fraser, M., Edlund, H., Munters, A. R., Coutinho, A., Simões, L. G., Vicente, M., Sjölander, A., Jansen Sellevold, B., Jørgensen, R., Claes, P., Shriver, M. D., ... Jakobsson, M. (2018). Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biology*, 16(1), e2003703. <https://doi.org/10.1371/journal.pbio.2003703>
- Günther, T., Valdiosera, C., Malmström, H., Ureña, I., Rodríguez-Varela, R., Sverrisdóttir, Ó. O., ... Jakobsson, M. (2015). Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proceedings of the National Academy of Sciences of the United States of America*, 112(38), 11917–11922. <https://doi.org/10.1073/pnas.1509851112>
- Henn, B. M., Gignoux, C., Lin, A. A., Oefner, P. J., Shen, P., Scozzari, R., ... Underhill, P. A. (2008). Y-chromosomal evidence of a pastoralist migration through Tanzania to southern Africa. *Proceedings of the National Academy of Sciences of the United States of America*, 105(31), 10693–10698. <https://doi.org/10.1073/pnas.0801184105>
- Henshilwood, C. (1996). A revised chronology for pastoralism in southernmost Africa: New evidence of sheep at c. 2000 b.p. from Blombos Cave, South Africa. *Antiquity*, 70(270), 945–949. <https://doi.org/10.1017/S0003598X00084210>
- Henshilwood, C., & van Niekerk, K. (2009). *Report on the 2006, 2007, & 2009 excavations at Vaalkrans Shelter*. Cape Town: Cape Nature.
- Henshilwood, C. S., d'Errico, F., van Niekerk, K. L., Dayet, L., Queffelec, A., & Pollarolo, L. (2018). An abstract drawing from the 73,000-year-old levels at Blombos Cave, South Africa. *Nature*, 562(7725), 115–118.
- Henshilwood, C. S., van Niekerk, K. L., Wurz, S., Delagnes, A., Armitage, S. J., Rifkin, R. F., Douze, K., Keene, P., Haaland, M. M., Reynard, J., Discamps, E., & Mienies, S. S. (2014). Klipdrift Shelter, southern Cape, South Africa: Preliminary report on the Howiesons Poort layers. *Journal of Archaeological Science*, 45, 284–303. <https://doi.org/10.1016/J.JAS.2014.01.033>
- Henshilwood, C. S. (2008). *Holocene prehistory of the southern cape, South Africa: Excavations at Blombos Cave and Blombosfontein Nature Reserve* (75th ed.). Cambridge, MA: Cambridge Monographs in African Archaeology.
- Hollfelder, N., Erasmus, J. C., Hammaren, R., Vicente, M., Jakobsson, M., Greeff, J. M., & Schlebusch, C. M. (2020). Patterns of African and Asian admixture in the Afrikaner population of South Africa. *BMC Biology*, 18(1), 1–13.
- Huffman, T. N. (1989). Ceramics, settlements and late iron age migrations. *The African Archaeological Review*, 7, 155–182 Retrieved from <http://www.jstor.org/stable/25130514>
- Jun, G., Flickinger, M., Hetrick, K. N., Romm, J. M., Doheny, K. F., Abecasis, G. R., Boehnke, M., & Kang, H. M. (2012). Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *The American Journal of Human Genetics*, 91(5), 839–848. <https://doi.org/10.1016/J.AJHG.2012.09.004>
- Kircher, M. (2012). Analysis of high-throughput ancient DNA sequencing data. *Methods in Molecular Biology*, 840, 197–228. https://doi.org/10.1007/978-1-61779-516-9_23
- Korneliusson, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, 15(1), 356.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Lombard, M. (2020). The tip cross-sectional areas of poisoned bone arrowheads from southern Africa. *Journal of Archaeological Science: Reports*, 33, 102477.
- Lombard, M., Caruana, M., van der Walt, J., & Högberg, A. (2020). The Keimoes 3 “desert kite” site, South Africa: An aerial LiDAR and microtopographic exploration. *Antiquity*, 94(373), 197–211.
- Lombard, M., Jakobsson, M., & Schlebusch, C. (2018). Ancient human DNA: How sequencing the genome of a boy from Ballito Bay changed human history. *South African Journal of Science* scieloza.114(1/2), Art. #a0253,-3p. <http://dx.doi.org/10.17159/sajs.2018/a0253>.
- Lombard, M., Wadley, L. Y. N., Deacon, J., Wurz, S., Parsons, I., Mohapi, M., ... Mitchell, P. (2012). South African and Lesotho Stone Age Sequence updated. *The South African Archaeological Bulletin*, 67(195), 123–144 Retrieved from <http://www.jstor.org/stable/23631399>
- Macholdt, E., Lede, V., Barbieri, C., Mpoloka, S. W., Chen, H., Slatkin, M., Pakendorf, B., & Stoneking, M. (2014). Tracing Pastoralist Migrations to Southern Africa with Lactase Persistence Alleles. *Current Biology*, 24(8), 875–879. <https://doi.org/10.1016/J.CUB.2014.03.027>
- Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 2010(6), pdb-prot5448. <https://doi.org/10.1101/pdb.prot5448>
- Mittnik, A., Wang, C.-C., Svoboda, J., & Krause, J. (2016). A molecular approach to the sexing of the triple burial at the Upper Paleolithic Site of Dolní Věstonice. *PLoS One*, 11(10), e0163019.
- Montinaro, F., Busby, G. B. J., Gonzalez-Santos, M., Oosthuizen, O., Oosthuizen, E., Anagnostou, P., Destro-Bisol, G., Pascali, V. L., & Capelli, C. (2017). Complex ancient genetic structure and cultural transitions in Southern African populations. *Genetics*, 205(1), 303–316. <https://doi.org/10.1534/genetics.116.189209>
- Morris, A. G. (2014). Going full circle on Khoekhoe origins. *The Digging Stick*, 31(1), 1–4 Retrieved from https://www.archaeologysa.co.za/sites/default/files/attachments/publications/2015/12/27/ds_2014_april.pdf
- Morris, A. G., Heinze, A., Chan, E. K. F., Smith, A. B., & Hayes, V. M. (2014). First ancient mitochondrial human genome from a pre-pastoralist Southern African. *Genome Biology and Evolution*, 6(10), 1–16. <https://doi.org/10.1093/gbe/evu202>
- Orton, J., & Parsons, I. (2018). Looking beneath the surface: Later Stone Age remains at Klipgats Pan, Bushmanland, South Africa. *Southern African Humanities*, 31(1), 181–204.
- Pagani, L., Kivisild, T., Tarekegn, A., Ekong, R., Plaster, C., Gallego Romero, I., Ayub, Q., Mehdi, S. Q., Thomas, M. G., Luiselli, D., Bekele, E., Bradman, N., Balding, D. J., & Tyler-Smith, C. (2012). Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian Gene Pool. *The American Journal of Human Genetics*, 91(1), 83–96. <https://doi.org/10.1016/J.AJHG.2012.05.015>
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient admixture in human history. *Genetics*, 192(3), 1065–1093. <https://doi.org/10.1534/genetics.112.145037>
- Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, 2(12), e190.
- Pickrell, J. K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Güldemann, T., Kure, B., Mpoloka, S. W., Nakagawa, H., Naumann, C., Lipson, M., Loh, P. R., Lachance, J., Mountain, J., Bustamante, C. D., Berger, B., Tishkoff, S. A., Henn, B. M., Stoneking, M., ... Pakendorf, B. (2012). The genetic prehistory of southern Africa. *Nature Communications*, 3, 1143. <https://doi.org/10.1038/ncomms2140>
- Pickrell, J. K., Patterson, N., Loh, P.-R., Lipson, M., Berger, B., Stoneking, M., ... Reich, D. (2014). Ancient west Eurasian ancestry in southern and eastern Africa. *Proceedings of the National Academy of Sciences of the United States of America*, 111(7), 2632–2637. <https://doi.org/10.1073/pnas.1313787111>
- Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8(11), e1002967. <https://doi.org/10.1371/journal.pgen.1002967>

- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8), 904–909.
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: Author. Retrieved from <https://www.r-project.org>
- Ryano, K. P., Wurz, S., van Niekerk, K. L., & Henshilwood, C. S. (2017). The Technology of the Early Oakhurst Lithic Techno-Complex from Klipdrift Cave, Southern Cape, South Africa. *African Archaeological Review*, 34(1), 93–119. <https://doi.org/10.1007/s10437-017-9246-2>
- Sadr, K. (2015). Livestock first reached southern Africa in two separate events. *PLoS One*, 10(8), e0134215. <https://doi.org/10.1371/journal.pone.0134215>
- Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., & Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One*, 7(3), e34131.
- Schlebusch C. M. 2010. *Genetic variation in Khoisan-speaking populations from southern Africa (Doctoral dissertation)*. University of the Witwatersrand, Johannesburg: University of the Witwatersrand.
- Schlebusch, C. M., De Jongh, M., & Soodyall, H. (2011). Different contributions of ancient mitochondrial and Y-chromosomal lineages in Karretjie people of the Great Karoo in South Africa. *Journal of Human Genetics*, 56(9), 623–630. <https://doi.org/10.1038/jhg.2011.71>
- Schlebusch, C. M., F Prins, F., Lombard, M., Jakobsson, M., & Soodyall, H. (2016). The disappearing San of southeastern Africa and their genetic affinities. *Human Genetics*, 135, 1365–1373.
- Schlebusch, C. M., & Jakobsson, M. (2018). Tales of human migration, admixture, and selection in Africa. *Annual Review of Genomics and Human Genetics*, 19(1), 405–428. <https://doi.org/10.1146/annurev-genom-083117-021759>
- Schlebusch, C. M., Lombard, M., & Soodyall, H. (2013). MtDNA control region variation affirms diversity and deep sub-structure in populations from southern Africa. *BMC Evolutionary Biology*, 13(1), 56. <https://doi.org/10.1186/1471-2148-13-56>
- Schlebusch, C. M., Malmström, H., Günther, T., Sjödin, P., Coutinho, A., Edlund, H., Munters, A. R., Vicente, M., Steyn, M., Soodyall, H., Lombard, M., & Jakobsson, M. (2017). Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science (New York, NY)*, 358(6363), 652–655. <https://doi.org/10.1126/science.aao6266>
- Schlebusch, C. M., Sjödin, P., Breton, G., Günther, T., Naidoo, T., Hoffelder, N., ... Vicente, M. (2020). Khoe-San genomes reveal unique variation and confirm the deepest population divergence in *Homo sapiens*. *Molecular Biology and Evolution*, 37, 2944–2954.
- Schlebusch, C. M., Skoglund, P., Sjödin, P., Gattepaille, L. M., Hernandez, D., Jay, F., ... Jakobsson, M. (2012). Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science (New York, NY)*, 338(6105), 374–379. <https://doi.org/10.1126/science.1227721>
- Skoglund, P., Thompson, J. C., Prendergast, M. E., Mitnik, A., Sirak, K., Hajdinjak, M., Salie, T., Rohland, N., Mallick, S., Peltzer, A., Heinze, A., Olalde, I., Ferry, M., Harney, E., Michel, M., Stewardson, K., Cerezoromán, J. I., Chiumia, C., Crowther, A., ... Reich, D. (2017). Reconstructing prehistoric African population structure. *Cell*, 171(1), 59–71. <https://doi.org/10.1016/j.cell.2017.08.049>
- Smith, A. B. (2005). *African herders: Emergence of pastoral traditions*. Alta Mira Press.
- Steyn, M., Whitelaw, G., Botha, D., Vicente, M., Schlebusch, C. M., & Lombard, M. (2019). Four Iron Age women from KwaZulu-Natal: Biological anthropology, genetics and archaeological context. *Southern African Humanities*, 32(1), 23–56 Retrieved from <https://journals.co.za/content/journal/10520/EJC-15f8f9fcd4>
- The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74. <https://doi.org/10.1038/nature15393>
- Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., Powell, K., Mortensen, H. M., Hirbo, J. B., Osman, M., Ibrahim, M., Omar, S. A., Lema, G., Nyambo, T. B., Ghorji, J., Bumpstead, S., Pritchard, J. K., Wray, G. A., & Deloukas, P. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics*, 39(1), 31–40. <https://doi.org/10.1038/ng1946>
- Tylén, K., Fusaroli, R., Rojo, S., Heimann, K., Fay, N., Johannsen, N. N., Riede, F., & Lombard, M. (2020). The evolution of early symbolic behavior in *Homo sapiens*. *Proceedings of the National Academy of Sciences of the United States of America*, 117(9), 4578–4584.
- Uren, C., Kim, M., Martin, A. R., Bobo, D., Gignoux, C. R., van Helden, P. D., Möller, M., Hoal, E. G., & Henn, B. M. (2016). Fine-scale human population structure in southern Africa reflects Ecogeographic boundaries. *Genetics*, 204(1), 303–314.
- van Oven, M. (2015). PhyloTree build 17: Growing the human mitochondrial DNA tree. *Forensic Science International: Genetics Supplement Series*, 5, e392–e394.
- Vanhaeren, M., Wadley, L., & d'Errico, F. (2019). Variability in Middle Stone Age symbolic traditions: The marine shell beads from Sibudu Cave, South Africa. *Journal of Archaeological Science: Reports*, 27, 101893.
- Veeramah, K. R., Wegmann, D., Woerner, A., Mendez, F. L., Watkins, J. C., Destro-Bisol, G., Soodyall, H., Louie, L., & Hammer, M. F. (2012). An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. *Molecular Biology and Evolution*, 29(2), 617–630. <https://doi.org/10.1093/molbev/msr212>
- Vianello, D., Sevinci, F., Castellani, G., Lomartire, L., Capri, M., & Franceschi, C. (2013). HAPLOFIND: A new method for high-throughput mtDNA Haplogroup assignment. *Human Mutation*, 34(9), 1189–1194.
- Vicente, M., Jakobsson, M., Ebbesen, P., & Schlebusch, C. M. (2019). Genetic affinities among Southern Africa hunter-gatherers and the impact of admixing farmer and herder populations. *Molecular Biology and Evolution*, 36(9), 1849–1861. <https://doi.org/10.1093/molbev/msz089>
- Vilstrup, J. T., Seguin-Orlando, A., Stiller, M., Ginolhac, A., Raghavan, M., Nielsen, S. C. A., Weinstock, J., Froese, D., Vasiliev, S. K., Ovodov, N. D., Clary, J., Helgen, K. M., Fleischer, R. C., Cooper, A., Shapiro, B., & Orlando, L. (2013). Mitochondrial phylogenomics of modern and ancient equids. *PLoS One*, 8(2), e55950. <https://doi.org/10.1371/journal.pone.0055950>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Coutinho A, Malmström H, Edlund H, et al. Later Stone Age human hair from Vaalkrans Shelter, Cape Floristic Region of South Africa, reveals genetic affinity to Khoe groups. *Am J Phys Anthropol*. 2021;174:701–713. <https://doi.org/10.1002/ajpa.24236>