Current Biology 24, 852–858, April 14, 2014 ©2014 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2014.02.041

Report

Lactase Persistence Alleles Reveal Partial East African Ancestry of Southern African Khoe Pastoralists

Gwenna Breton,[1,2,6] Carina M. Schlebusch,[1,6,*] Marlize Lombard,[3] Per Sjödin,[1] Himla Soodyall,[4] and Mattias Jakobsson[1,5,*]

[1]Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden

[2]Master Bioscience, Department of Biology, École Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon Cedex 07, France

[3]Department of Anthropology and Development Studies, University of Johannesburg, Auckland Park, Johannesburg 2006, South Africa

[4]Division of Human Genetics, School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, and National Health Laboratory Service, Braamfontein, Johannesburg 2017, South Africa

[5]Science for Life Laboratory, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden

Summary

The ability to digest milk into adulthood, lactase persistence (LP), as well as specific genetic variants associated with LP, is heterogeneously distributed in global populations [1–4]. These variants were most likely targets of selection when some populations converted from hunter-gatherer to pastoralist or farming lifestyles [5–7]. Specific LP polymorphisms are associated with particular geographic regions and populations [1–4, 8–10]; however, they have not been extensively studied in southern Africa. We investigate the LP-regulatory region in 267 individuals from 13 southern African populations (including descendants of hunter-gatherers, pastoralists, and agropastoralists), providing the first comprehensive study of the LP-regulatory region in a large group of southern Africans. The "East African" LP single-nucleotide polymorphism (SNP) (14010G>C) was found at high frequency (>20%) in a strict pastoralist Khoe population, the Nama of Namibia, suggesting a connection to East Africa, whereas the "European" LP SNP (13910C>T) was found in populations of mixed ancestry. Using genome-wide data from various African populations, we identify admixture (13%) in the Nama, from an Afro-Asiatic group dating to >1,300 years ago, with the remaining fraction of their genomes being from San hunter-gatherers. We also find evidence of selection around the *LCT* gene among Khoe-speaking groups, and the substantial frequency of the 14010C variant among the Nama is best explained by adaptation to digesting milk. These genome-local and genome-wide results support a model in which an East African group brought pastoralist practices to southern Africa and admixed with local hunter-gatherers to form the ancestors of Khoe people.

[6]Co-first author

*Correspondence: carina.schlebusch@ebc.uu.se (C.M.S.), mattias.jakobsson@ebc.uu.se (M.J.)

Results and Discussion

We sequenced 360 bp of the lactase persistence (LP)-regulatory region, encompassing all known LP-regulatory variants in 267 individuals representing 13 southern African populations (Table 1, Table S1 available online, Figure 1, and the Supplemental Experimental Procedures). The sample includes San, Khoe, and Bantu-speaking groups, which historically occupied southern Africa. The Khoe and San represent the original inhabitants of the region. Ancestors of current-day Khoe and San (hereafter referred to collectively as Khoe-San) diverged at least 100 thousand years ago from the ancestors of all other current-day humans, although subsequent admixture occurred [11]. Bantu-speaking farmers arrived in the northern parts of southern Africa 1–2 thousand years ago [12] and reached the southeastern Cape some 1.3 thousand years ago (Figure 1). Archaeological traces of domesticated animals in the southwestern Cape from about 2 thousand years ago and onward, however, point to an introduction of pastoralism by Khoe populations. Recent history indicates that the investigated groups cover diverse subsistence modes: the San were hunter-gatherers, the Khoe were pastoralists, and the Bantu-speakers practiced pastoralism and cultivated crops.

LP-Associated Polymorphisms in Southern African Groups

The single-nucleotide polymorphisms (SNPs) present in the LP-regulatory region with the greatest allele frequencies in southern African groups are 13910C>T (rs4988235) and 14010G>C (rs145946881; Table 1, Figure 1, and the Supplemental Experimental Procedures). These SNPs were shown to be directly associated with the LP phenotype, and the former was found at high frequencies in Europe and the latter at high frequencies in East Africa [4, 8, 13, 14]. Additionally, we found a 13913G>C SNP at population frequencies up to 7.5% (Table 1), which was found in Ethiopian populations, but its association with LP is unclear [15]. Three additional singleton SNPs were detected in the region (14156C>A, 14091C>T, and 13937G>A; Supplemental Discussion). The frequencies of the observed SNPs vary considerably among the study groups (Table 1). The 13910C>T allele was found in seven out of 13 groups. Its frequency is particularly high in the "Wellington Coloured" (17.5%), a group with mixed ancestry (people of mixed ancestry in South Africa often use "Coloured" as self-identification; they have, among others, European ancestors [11]). Although 13910C>T is the LP-associated SNP found in Europe, it was also identified in African groups, including the nomadic Fulani from Mali [10], Sudan [16], and Cameroon [9]. The Nama, who show low levels of European admixture [11, 17], also display a low frequency (6.8%) of the 13910C>T variant. Indeed, once individuals with recent European and Bantu-speaking admixture were removed (based on whole-genome data; see [11]), the 13910T allele was no longer present in the remaining Nama sample (Table 1).

The frequency of the "East African" LP variant 14010C varies from 0% in Bantu-speaking groups to 20% in the Nama (Khoe) and 22.5% in the Askham Coloured group (Table 1). This

Table 1. Population Groups and Allele Frequencies of the Three Most Frequent Polymorphisms in the LP-Regulatory Region in Various Southern African Groups

Group Name	Main Group	Place of Origin (Country)[a]	All Individuals Sequenced for the LP-Regulatory Region (Sum: 267)				Predicted LP-Phenotype Frequency (%)[c]	Individuals with Genome-Wide Data (Recently Admixed Individuals Removed) (Sum: 117)			
			n	13910T	13913C	14010C		n	13910T	13913C	14010C
Askham Coloured	Coloured (descendants of the ≠Khomani, Nama, and Griqua)	Askham (SA)	20	0.05	0.050	0.225	45	8	–	–	–
Wellington Coloured	Coloured	Wellington (SA)	20	0.175	0.000	0.100	45	0	–	–	–
/Gui and //Gana[b]	San (Khoe speaking)	Kutse Game Reserve (BT)	20	0.000	0.000	0.075	15	7	0.000	0.000	0.071
Ju/'hoansi	San	Tsumkwe (NM)	20	0.000	0.025	0.025	5	17	0.000	0.029	0.029
Colesberg Coloured	Coloured	Colesberg (SA)	20	0.050	0.025	0.025	15	0	–	–	–
Karretjie People	San (most likely descendants of the /Xam)	Colesberg (SA)	20	0.000	0.025	0.100	20	12	0.000	0.042	0.083
≠Khomani	San (descendants of San and Khoe groups)	Askham (SA)	20	0.025	0.075	0.050	20	9	0.000	0.000	0.11
Khwe	San (Khoe speaking)	Caprivi strip and surrounding regions (AN, BT, NM)	19	0.000	0.000	0.026	5	17	0.000	0.000	0.029
Manyanga	Bantu speakers	Luozi (DRC)	11	0.000	0.000	0.000	0	0	–	–	–
Nama	Khoe	Windhoek (NM)	22	0.068	0.023	0.205	50	7	0.000	0.071	0.357
Southeastern Bantu speakers	Bantu speakers	various (SA)	41	0.024	0.000	0.073	17	19	0.000	0.000	0.079
–Sotho-Tswana[d]			(16)	0.000	0.000	0.063	(13)	(9)	0.000	0.000	0.056
–Zulu[d]			(25)	0.040	0.000	0.080	(20)	(10)	0.000	0.000	0.100
Herero (southwestern Bantu speakers)	Bantu speakers (Herero)	Windhoek (NM)	14	0.000	0.000	0.000	0	8	0.000	0.000	0.000
!Xun	San	region surrounding Menongue (AN)	20	0.000	0.050	0.050	10	13	0.000	0.077	0.038

See also Tables S1 and S2.
[a]Country abbreviations are as follows: AN, Angola; BT, Botswana; DRC, Democratic Republic of Congo; NM, Namibia; and SA, South Africa.
[b]The /Gui and //Gana group is a mixed group of San and Bantu-speaking people who had ancestries from both the /Gui and //Gana San groups and the Kgalagari Bantu-speaking group.
[c]Phenotype frequencies of LP were calculated from genotype frequencies, and no direct phenotypes were available for this study.
[d]Sotho-Tswana and Zulu are subgroups of the southeastern Bantu speakers.

variant was originally identified in East Africans from Kenya and Tanzania [4] and reported at high frequencies (58%) among the Maasai [7]. In southern Africa, it was found at lower frequencies in Herero from Angola [18] and Xhosa from South Africa [19]. We found the 14010C allele in all investigated groups, including Khoe, San, Coloured, and Bantu speakers. If we remove recently admixed individuals, the frequency of the East African allele 14010C increases to 35.7% in the Nama, who then display the largest population frequency of this allele among our sample populations (Figure 1 and Table 1). The allele frequency in the Khoe (represented by the Nama) is significantly greater than in the San (Fisher's exact test: $p = 0.0049$ [all individuals], $p = 0.0013$ [recently admixed individuals removed]), the southeastern Bantu speakers ($p = 0.043$ [all individuals], $p = 0.025$ [recently admixed individuals removed]), and the southwestern Bantu speakers ($p = 0.010$ [all individuals]). The relatively high frequencies observed in the Coloured groups might be due to their large Khoe ancestry, which is historically well known [20]. During sampling, Coloured people from Askham reported themselves to have high incidences of Khoe ancestry (Nama and Griqua), and this was confirmed genetically [11].

A few southern African populations have been investigated for LP-phenotype frequencies, including some Khoe-San groups (Ju/'hoansi and Nama) [21, 22] and southeastern Bantu speakers [23]. The Nama, a pastoralist Khoe group, were found to have high levels of LP (50%), in contrast to the Ju/'hoansi, a San hunter-gatherer group (LP = 10%), and Bantu-speaking groups (LP = 22%). On the basis of genotype frequencies of known LP mutations (13910T and 14010C), we calculated predicted phenotype frequencies (Table 1 and the Supplemental Experimental Procedures), which were similar to previous empirical phenotype data of LP [21–23] among the Nama, Ju/'hoansi, and southeastern Bantu speakers.

Haplotype Background of the LP Variants among Southern African Individuals

High frequencies of the 14010C variant in some southern African populations suggest a connection to East African groups; however, it might also indicate a recurrent mutation. We merged the sequence data from the LP control region with SNP data from a genome-wide, very dense SNP data set [11], targeting 220 individuals in this study. The data were phased and further merged with phased SNP data from 50 individuals of northwestern European descent (the HapMap CEU) and 25 individuals from the East African Maasai [24] (Supplemental Experimental Procedures and Supplemental Discussion). The gene copies of southern African individuals that carry the East African LP variant (Figure 2) show the same haplotype background (stretching some 52 kb) as the East African Maasai (Figure S1A), which is also supported by local ancestry estimates of the region (Figure S2). Similarly,

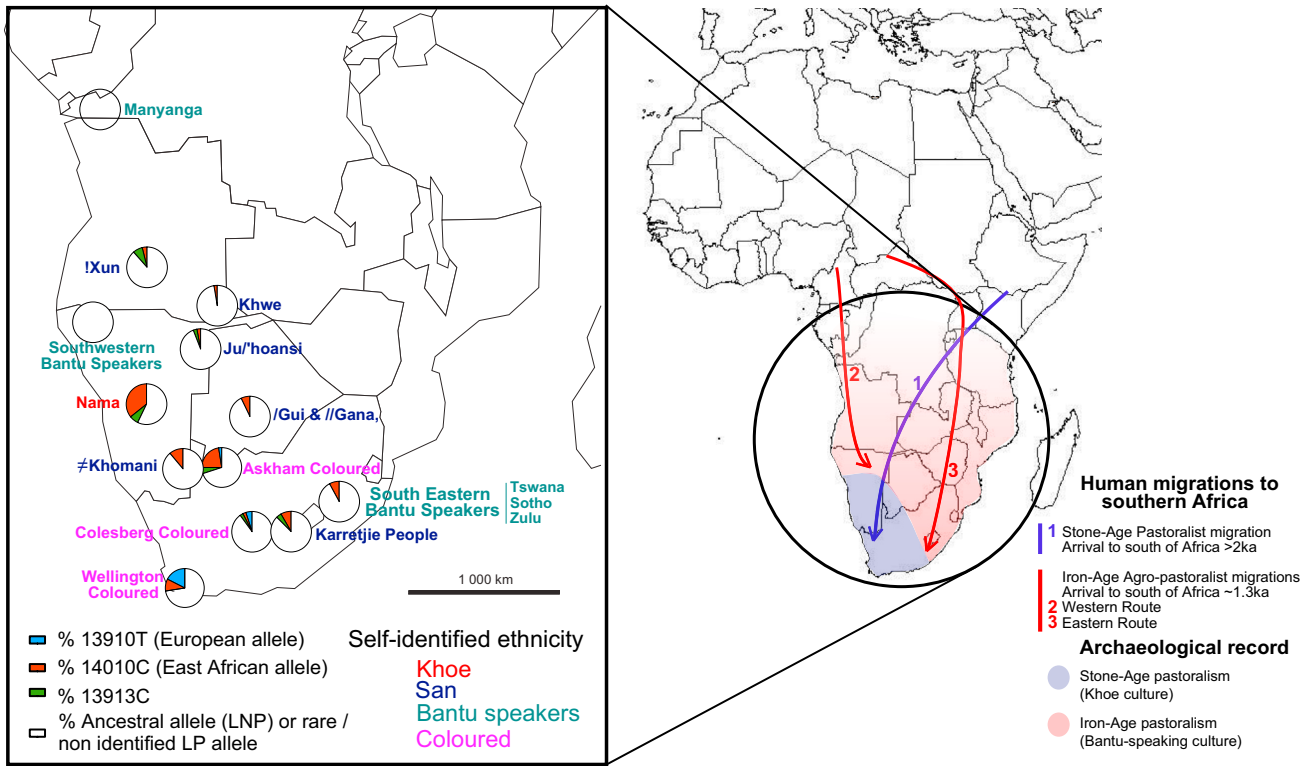

Figure 1. LP Genotype Frequencies and Human Migrations Connected to the Spread of Pastoralist Practices to Southern Africa

The pie charts display the allele frequencies of the three most frequent LP-associated variants in the LP-regulatory region in various groups of southern Africans (the pie charts display the frequencies after removal of recently admixed individuals; see Table 1 for full display of allele frequencies).

the gene copies of southern African individuals that carry the European LP variant display the same haplotype background as Europeans (Figures S1B and S1C). Therefore, the 13910T variant was most likely introduced among southern Africans through recent admixture with European colonists, whereas the 14010G variant was likely introduced by gene flow from East Africans to the Khoe.

East African Connections and Characterization of Admixture Events Based on Whole-Genome Data

To detect admixture from East Africans in various Khoe-San populations, we used a formal test of admixture (three-population f test [25]) on genome-wide SNP data (produced in [11] and [26]; Supplemental Experimental Procedures). The greatest support (low f values and highly significant Z scores) for admixture from an East African group into various Khoe-San groups was observed for the Khwe (but note the complex ancestry of this group [11]), followed by the Nama; both are Khoe-speaking groups (Table S3). The f test supported East African admixture into the Nama with various East African groups (Maasai, Afar, Amhara, and Tigray) used in conjunction with the Ju'/hoansi as parental populations. All Khoe-San groups have indications of East African admixture, except the Ju'/hoansi. Using a model-based approach of painting a population history allowing admixture (Treemix [27]), we confirm the finding of a specific East African admixture component in the Nama, distinct from potential Bantu-speaking and European contributions (Supplemental Discussion and Figure S4). This evidence of widespread East African admixture into Khoe-San groups explains the fact that the East African LP variant 14010C was recorded in all Khoe-San groups and that it was particularly high in the Nama (Table 1).

We further investigated the genome-wide East African contribution in the Nama by estimating admixture proportions [28] and population stratification (principal component analysis [29, 30]; Supplemental Discussion and Figure S5) in a large collection of East African [26] and southern African populations [11] (Table S4). Several African groups with specific geographical ranges stand out as having distinct genetic ancestry (Figures 3 and S3), i.e., southern African Khoe-San (red), central African Pygmys (gray), and Niger-Congo groups with West African ancestry (light blue). There are also three more specifically East African components: green, yellow, and blue. The green component most likely corresponds to the "Ethiopian-specific" component described in [26]. The two remaining components are interlocked, but the yellow component seems associated to Afro-Asiatic groups, whereas the blue is more dominant in Nilo-Saharan groups. The three East African-specific components are present in several Khoe-San groups and are the largest in the Nama (13% in total).

Of the different East African components, the Afro-Asiatic component is largest in the Nama (11%); other components contribute less (1.1% each). The East African ancestry does not exceed 6% in the other southern African groups. To better characterize the potential admixture event between the Nama and an East African group, we compared the "Ethiopian-specific/Afro-Asiatic," "Ethiopian-specific/Nilo-Saharan," and "Afro-Asiatic/Nilo-Saharan" ratios for the Nama and the different East African groups (assuming that admixture occurred with a single East African group and that past East Africa populations are represented by contemporary

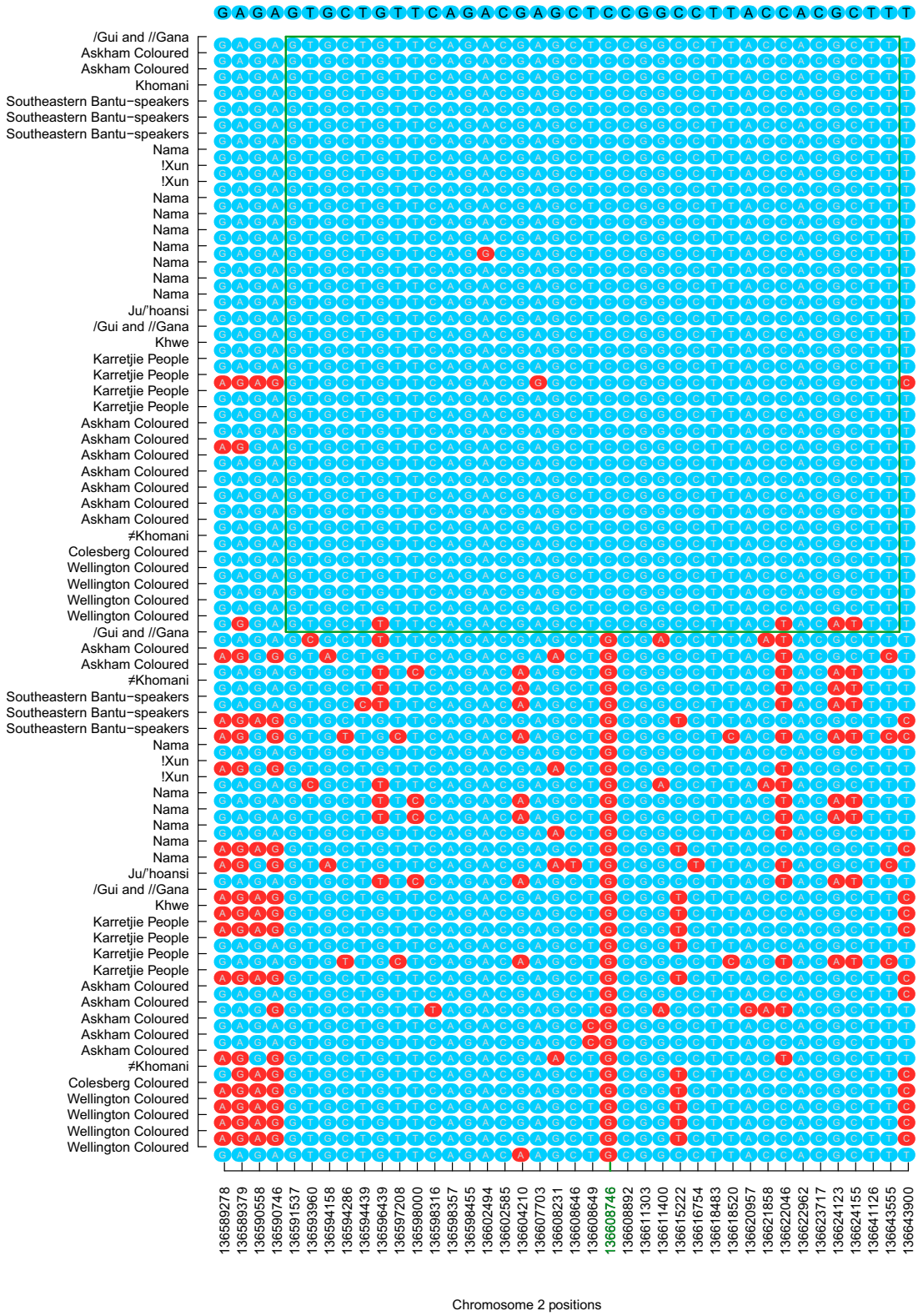

Figure 2. Display of Haplotypes over a 54.6 kb Region of Chromosome 2, Showing a Haplotype Block Surrounding the East African LP Variant

All individuals that carry at least one copy of the East African variant, 14010C, at position 136,608,746 on chromosome 2 were extracted. The haplotypes were sorted according to the variant they contain at position 136,608,746. The top sequence shows the major allele in the Nama population. Variants that differ from the Nama consensus sequence are shown red, while variants identical to the Nama consensus sequence are shown in blue for every SNP. The y axis displays the population group of the individual haplotypes and the x axis shows the SNP position on chromosome 2. The East African LP variant is highlighted in green on the x axis. A green block outlines the common haplotype background associated with the East African LP variant. See also Figure S1 (for inclusion of the Maasai haplotypes) and Figure S2.

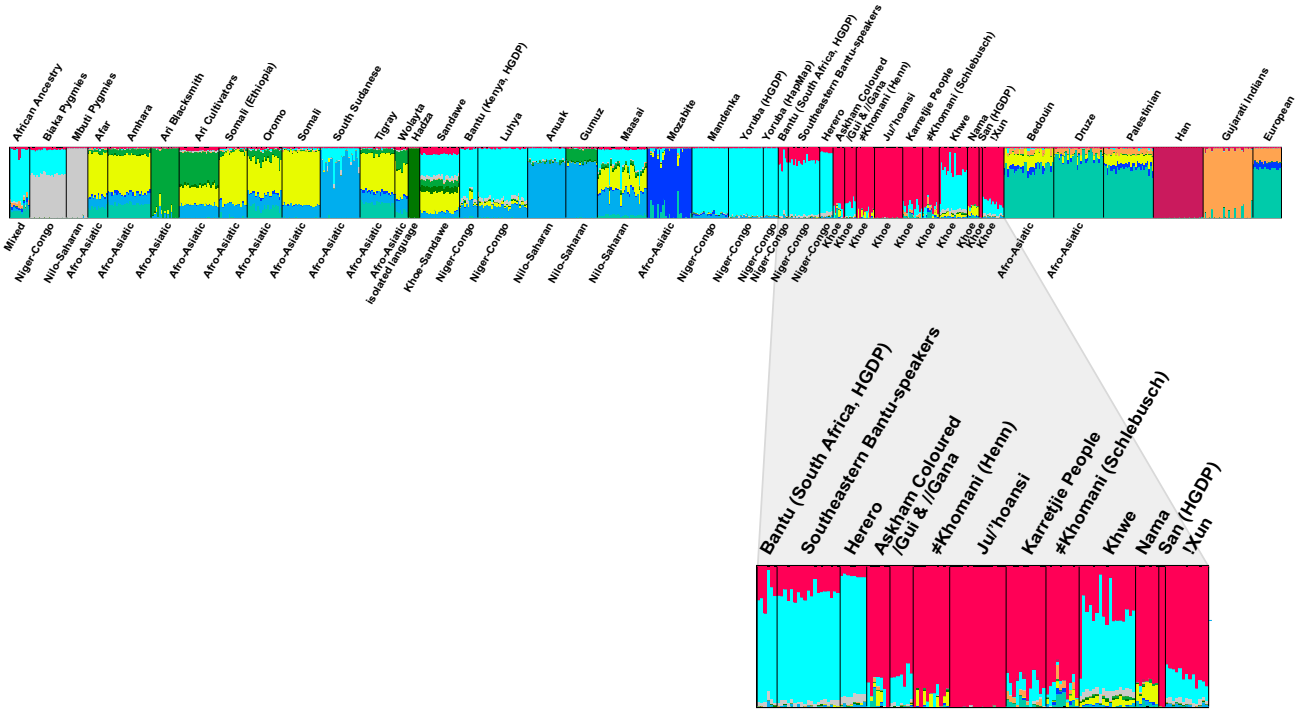

Figure 3. Genetic Clustering Analysis

(A) Clustering of 766 individuals (based on 233,363 SNPs) assuming 11 clusters (see Figure S3 for additional choices of clusters).

(B) Zoomed-in view of southern African populations that were sequenced for the LP-regulatory region.

See also Figures S3–S5 and Table S4.

groups). The best match to the ratios of the East African components (considering five to 11 allowed clusters) in the Nama was found for three Ethiopian groups, the Afar, Amhara, and Tigray. These populations originate from northern Ethiopia, speaking Afro-Asiatic languages from the Cushitic and Semetic divisions, and the Afar are traditionally pastoralists. Khoe ancestors probably admixed with a group related to and/or ancestral to these populations. The comparison of the ratios were similar among Askham Coloured and ≠Khomani, which is consistent with a first admixture event involving Khoe and an East African group, and subsequent gene flow among Khoe-San groups (accordant with low frequencies of the East African LP variant among all Khoe-San).

We furthermore estimated the date of East African admixture into the Nama using patterns of linkage disequilibrium decay [25] (Table S3). With the Maasai and Ju/'hoansi as potential parental populations to the Nama, we obtain an admixture date of 39.4 generations ago (1,143 years ago, SE 74 years, assuming 29 years per generation). If we use the Afar, Amhara, and Tigray (albeit for a data set containing fewer SNPs), admixture dates were somewhat older (around 43 generations and 1,255 years ago). The date of 39–43 generations is distinctly older than the estimated 7–24 generations (Table S3) of the inferred admixture events between Bantu speakers and Khoe-San. Furthermore, small amounts of Bantu-speaking admixture into the Nama and/or the East African parental population(s) will reduce the estimated time of admixture for the East African component in the Nama such that the true admixture time could be somewhat older than ~1.3 thousand years ago. In addition, continuous East African gene flow (after the initial event) into the Nama would also yield an underestimation of the first admixture time.

These admixture-time estimates are more recent than archeological evidence of livestock in southern Africa. The earliest goat bones in central Namibia date to 2190 ± 40 before present (BP) and 2270 ± 40 BP [31]. There were sheep in northern Namaqualand (South Africa) by 1625 ± 25 BP [32], and possibly by 2105 ± 65 BP [33]. In the southern Cape, sheep were present at Blombos Cave by 1960 ± 50 BP, the earliest evidence of domesticated animals in the southernmost regions of Africa [34].

Our estimated admixture dates of humans associated with pastoralism fit well with an archaeological hypothesis [35] suggesting that early evidence of domesticated animals in southwestern Africa represents hunter-gatherers with sheep and/or goats, rather than full-scale pastoralism. According to this scenario, hunter-gatherers obtained livestock through cultural diffusion and a gift-giving system, but were not pastoralists in the strict sense. The pastoralist lifeway of the Khoe probably only developed later, possibly after infiltration by an East African group around 1.3 thousand years ago [35]. This infiltration changed pottery styles, and, subsequently, more robust and frequent evidence of pastoralism appeared in the southern African archeological record [35]. The later date agrees well with our estimated time of admixture (but the genetically estimated admixture time could be downwardly biased; see above).

Selection on LP Variants in Southern African Populations

A selective advantage of being LP in societies whose members consume fresh milk from domestic animals is evident [3, 4], and LP-associated variants were under strong positive selection in some groups, displaying one of the strongest selection signals in the human genome [5–7]. In the southern African groups, we found evidence of selection at the *LCT* region

in the Nama (enrichment of extreme integrated haplotype score values [iHS] [36]: p = 0.035), Khwe (p = 0.013), and /Gui and //Gana (p = 0.019) (Supplemental Experimental Procedures and Figure S6); all three groups are Khoe-speaking groups (Tables 1 and S1). The signal, however, is distinctly weaker than in, e.g., the Maasai [5], which is possibly related to the shorter time of pastoralism in southern Africa compared to East Africa [12].

Given a frequency of 57.8% of the 14010C allele in the Maasai [4] and the 13% East African admixture fraction in the Nama, the expected frequency of the 14010C allele in the Nama is 7.5%. The observed frequency of the 14010C allele in the Nama is much greater: 35%, with recently admixed individuals removed. We investigated whether the observed high frequency of 14010C in the Nama sample could be explained by genetic drift alone, but found this to be highly unlikely (Supplemental Discussion). We also estimated the selection coefficient necessary to explain the fast increase in the frequency of 14010C since the admixture event, finding it comparable to estimates for particular East African populations [4] (Supplemental Discussion and Figure S7). Cumulatively, the data are best explained by a scenario in which adaptation to LP occurred both among East African groups (prior to southward migration) and in the Nama (after admixture with incoming East Africans).

An Afro-Asiatic East African Group Brought Pastoralism to Southern Africa

The introduction of pastoralism to southern Africa is vigorously debated, and our study assists in deciphering the processes. It is widely accepted that pastoralism spread from North to East Africa, and later to southern Africa [12, 35]. Europeans, who reached the Cape of Good Hope in the 1600s, encountered hunter-gatherers (San) and various pastoralist groups (Khoe) [12, 20, 35]. Several hypotheses aim to explain how the Khoe developed pastoralism. These range from the demic-diffusion model (in which an East African or another external group with livestock migrated south and were the ancestors of the Khoe) to the cultural-diffusion model (in which indigenous southern African hunter-gatherers adopted a pastoralist lifestyle, e.g., through a gift-giving system). Arguments for both scenarios (and several intermediary variations) are drawn from archaeological and linguistic records [12, 31, 32, 35, 37, 38], as well as opinions that Bantu speakers brought cattle herding to the ancestors of the Khoe [39]. However, recent discovery of a cow horn core at an archaeological site in the western Cape, South Africa, dated to >1.6 thousand years ago [32], makes it too old to have been introduced by Bantu speakers.

In Africa, genetic variants associated to LP are mainly found in pastoralist populations; therefore, study of LP-associated variants in different groups helps to generate knowledge about past relationships [18, 19]. We demonstrate that the East African LP variant is present at greater frequencies in traditionally pastoralist Khoe groups, here represented by the Nama, compared to San hunter-gatherers and Bantu-speaking farmers. The haplotype background for the East African LP variant in southern Africans was the same as the East African haplotype background. The presence of the East African LP variant in the Khoe suggests an admixture event, probably associated with the spread of pastoralism into southern Africa. This hypothesis is supported by analyses of genome-wide data that detect a 13% admixture fraction (most similar to East African Afro-Asiatic groups) in the Nama. Given the small but significant East African genetic component in the Nama, an exclusive cultural- or demic-diffusion model seems unlikely. Our results support an interpretation of a small group of Afro-Asiatic pastoralists from East Africa who migrated to southern Africa where they became assimilated by a local San hunter-gatherer group. This group adopted a pastoralist lifestyle and was the ancestors of the Khoe. Thus—consistent with some linguistic hypotheses on the origin of the Khoe languages [37] and with Y chromosome haplogroup similarities [40]—our analyses of the *LCT* region and genome-wide data among southern Africans show that the pastoralist Khoe originate from a San group that adopted pastoralism, with introgression from an East African Afro-Asiatic group that migrated south prior to ~1.3 thousand years ago.

Supplemental Information

Supplemental Information includes Supplemental Introduction, Supplemental Discussion, Supplemental Experimental Procedures, seven figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2014.02.041.

Author Contributions

C.M.S. and M.J. conceived of the study. C.M.S. and H.S. collected and prepared the samples. G.B. carried out the molecular laboratory work and sequence processing with assistance from C.M.S. G.B. and P.S. conducted the population genetic analyses with assistance and guidance from C.M.S. and M.J. M.L. provided archeological contexts. G.B., C.M.S., and M.J. wrote the manuscript with contributions from all authors. All authors read and approved the final manuscript.

Acknowledgments

We are grateful to all subjects who participated in our research and would like to thank members of the Working Group of Indigenous Minorities in Southern Africa (WIMSA) and the South African San Council for facilitating research among the San people. We would also like to acknowledge Michael De Jongh from the University of South Africa (UNISA), who participated in and facilitated the sample collection among the Karretjie People, and Professor Trefor Jenkins and colleagues in the Division of Human Genetics for assistance with other fieldwork and processing of samples. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Projects p2013083 and p2012202. G.B. was financially supported by the European Union (Erasmus) and the Région Rhône-Alpes, France (Explora'sup). C.M.S. would like to thank the Wenner Gren Foundation (Sweden) for financial support. M.L. is funded by the African Origins Platform of the South African National Research Foundation. This work was funded by a grant to M.J. from the Swedish Research Council.

Received: November 26, 2013
Revised: December 20, 2013
Accepted: February 15, 2014
Published: April 3, 2014

References

1. Ingram, C.J., Mulcare, C.A., Itan, Y., Thomas, M.G., and Swallow, D.M. (2009). Lactose digestion and the evolutionary genetics of lactase persistence. Hum. Genet. *124*, 579–591.
2. Itan, Y., Jones, B.L., Ingram, C.J., Swallow, D.M., and Thomas, M.G. (2010). A worldwide correlation of lactase persistence phenotype and genotypes. BMC Evol. Biol. *10*, 36.
3. Gerbault, P., Liebert, A., Itan, Y., Powell, A., Currat, M., Burger, J., Swallow, D.M., and Thomas, M.G. (2011). Evolution of lactase persistence: an example of human niche construction. Philos. Trans. R. Soc. Lond. B Biol. Sci. *366*, 863–877.
4. 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., et al. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. Nat. Genet. *39*, 31–40.

5. Schlebusch, C.M., Sjödin, P., Skoglund, P., and Jakobsson, M. (2013). Stronger signal of recent selection for lactase persistence in Maasai than in Europeans. Eur. J. Hum. Genet. *21*, 550–553.
6. Kelley, J.L., and Swanson, W.J. (2008). Positive selection in the human genome: from genome scans to biological significance. Annu. Rev. Genomics Hum. Genet. *9*, 143–160.
7. Wagh, K., Bhatia, A., Alexe, G., Reddy, A., Ravikumar, V., Seiler, M., Boemo, M., Yao, M., Cronk, L., Naqvi, A., et al. (2012). Lactase persistence and lipid pathway selection in the Maasai. PLoS ONE *7*, e44751.
8. Enattah, N.S., Sahi, T., Savilahti, E., Terwilliger, J.D., Peltonen, L., and Järvelä, I. (2002). Identification of a variant associated with adult-type hypolactasia. Nat. Genet. *30*, 233–237.
9. Ingram, C.J., Elamin, M.F., Mulcare, C.A., Weale, M.E., Tarekegn, A., Raga, T.O., Bekele, E., Elamin, F.M., Thomas, M.G., Bradman, N., and Swallow, D.M. (2007). A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? Hum. Genet. *120*, 779–788.
10. Lokki, A.I., Järvelä, I., Israelsson, E., Maiga, B., Troye-Blomberg, M., Dolo, A., Doumbo, O.K., Meri, S., and Holmberg, V. (2011). Lactase persistence genotypes and malaria susceptibility in Fulani of Mali. Malar. J. *10*, 9.
11. Schlebusch, C.M., Skoglund, P., Sjödin, P., Gattepaille, L.M., Hernandez, D., Jay, F., Li, S., De Jongh, M., Singleton, A., Blum, M.G., et al. (2012). Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. Science *338*, 374–379.
12. Smith, A.B. (2005). African Herders. Emergence of Pastoral Traditions (Oxford: AltaMira Press).
13. Jensen, T.G., Liebert, A., Lewinsky, R., Swallow, D.M., Olsen, J., and Troelsen, J.T. (2011). The -14010*C variant associated with lactase persistence is located between an Oct-1 and HNF1α binding site and increases lactase promoter activity. Hum. Genet. *130*, 483–493.
14. Lewinsky, R.H., Jensen, T.G., Møller, J., Stensballe, A., Olsen, J., and Troelsen, J.T. (2005). T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity in vitro. Hum. Mol. Genet. *14*, 3945–3953.
15. Jones, B.L., Raga, T.O., Liebert, A., Zmarz, P., Bekele, E., Danielsen, E.T., Olsen, A.K., Bradman, N., Troelsen, J.T., and Swallow, D.M. (2013). Diversity of lactase persistence alleles in Ethiopia: signature of a soft selective sweep. Am. J. Hum. Genet. *93*, 538–544.
16. Enattah, N.S., Trudeau, A., Pimenoff, V., Maiuri, L., Auricchio, S., Greco, L., Rossi, M., Lentze, M., Seo, J.K., Rahgozar, S., et al. (2007). Evidence of still-ongoing convergence evolution of the lactase persistence T-13910 alleles in humans. Am. J. Hum. Genet. *81*, 615–625.
17. Pickrell, J.K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Güldemann, T., Kure, B., Mpoloka, S.W., Nakagawa, H., Naumann, C., et al. (2012). The genetic prehistory of southern Africa. Nat. Commun. *3*, 1143.
18. Coelho, M., Sequeira, F., Luiselli, D., Beleza, S., and Rocha, J. (2009). On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. BMC Evol. Biol. *9*, 80.
19. Torniainen, S., Parker, M.I., Holmberg, V., Lahtela, E., Dandara, C., and Jarvela, I. (2009). Screening of variants for lactase persistence/non-persistence in populations from South Africa and Ghana. BMC Genet. *10*, 31.
20. Giliomee, H., and Mbenga, B. (2010). New History of South Africa (Cape Town: Tafelberg).
21. Nurse, G.T., and Jenkins, T. (1974). Letter: Lactose intolerance in San populations. BMJ *3*, 809.
22. Jenkins, T., Lehmann, H., and Nurse, G.T. (1974). Public health and genetic constitution of the San ("Bushmen"): carbohydrate metabolism and acetylator status of the Kung of Tsumkwe in the North-western Kalahari. BMJ *2*, 23–26.
23. Segal, I., Gagjee, P.P., Essop, A.R., and Noormohamed, A.M. (1983). Lactase deficiency in the South African black population. Am. J. Clin. Nutr. *38*, 901–905.
24. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A.; 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.
25. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.
26. Pagani, L., Kivisild, T., Tarekegn, A., Ekong, R., Plaster, C., Gallego Romero, I., Ayub, Q., Mehdi, S.Q., Thomas, M.G., Luiselli, D., et al. (2012). Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. Am. J. Hum. Genet. *91*, 83–96.
27. Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. *8*, e1002967.
28. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.
29. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.
30. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909.
31. Pleurdeau, D., Imalwa, E., Détroit, F., Lesur, J., Veldman, A., Bahain, J.J., and Marais, E. (2012). "Of sheep and men": earliest direct evidence of caprine domestication in southern Africa at Leopard Cave (Erongo, Namibia). PLoS ONE *7*, e40340.
32. Orton, J., Mitchell, P., Klein, R.G., Steele, T., and Horsburgh, K.A. (2013). An early date for cattle from Namaqualand, South Africa: implications for the origins of herding in southern Africa. Antiquity *87*, 108–120.
33. Sealy, J., and Yates, R. (1994). The chronology of the introduction of pastoralism to the Cape, South Africa. Antiquity *68*, 58–67.
34. Henshilwood, C.S. (1996). A revised chronology for pastoralism in southernmost Africa: new evidence of sheep at ca. 2000 B.P. from Blombos Cave, South Africa. Antiquity *70*, 945–949.
35. Sadr, K. (2013). A Short History of Early Herding in Southern Africa. In Pastoralism in Africa: Past, Present and Future, *Volume 171*, M. Bollig, M. Schnegg, and H.-P. Wotzka, eds. (New York: Berghahn Books).
36. Pickrell, J.K., Coop, G., Novembre, J., Kudaravalli, S., Li, J.Z., Absher, D., Srinivasan, B.S., Barsh, G.S., Myers, R.M., Feldman, M.W., and Pritchard, J.K. (2009). Signals of recent positive selection in a worldwide sample of human populations. Genome Res. *19*, 826–837.
37. Güldemann, T. (2008). A linguist's view: Khoe-Kwadi speakers as the earliest food-producers of southern Africa. In Khoekhoe and the origins of herding in southern Africa, *Volume 20*, K. Sadr and F.-X. Fauvelle-Aymar, eds. (Pietermaritzburg: Southern African Humanities), pp. 93–132.
38. Smith, A. (2005). The concepts of 'Neolithic' and 'Neolithisation' for Africa? Before Farming *1*, 1–6.
39. Smith, A.B. (2006). Excavations at Kasteelberg and the origins of the Khoekhoen in the Western Cape, South Africa. British Archaeological Reports International series: 1537 (Oxford: Archaeopress).
40. Henn, B.M., Gignoux, C., Lin, A.A., Oefner, P.J., Shen, P., Scozzari, R., Cruciani, F., Tishkoff, S.A., Mountain, J.L., and Underhill, P.A. (2008). Y-chromosomal evidence of a pastoralist migration through Tanzania to southern Africa. Proc. Natl. Acad. Sci. USA *105*, 10693–10698.