bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Admixture into and within sub-Saharan Africa

George B.J. Busby[1,†], Gavin Band[1,2], Quang Si Le[1], Muminatou Jallow[3,4], Edith Bougama[5], Valentina Mangano[6], Lucas Amenga-Etego[7], Anthony Enimil[8], Tobias Apinjoh[9], Carolyne Ndila[10], Alphaxard Manjurano[11,12], Vysaul Nyirongo[13], Ogobara Doumbo[14], Kirk, A. Rockett[1,2], Dominic P. Kwiatkowski[1,2], & Chris C.A. Spencer[1,†] in association with the Malaria Genomic Epidemiology Network[15]

[1]Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN, UK
[2]Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK
[3]Medical Research Council Unit, Serrekunda, The Gambia
[4]Royal Victoria Teaching Hospital, Banjul, The Gambia
[5]Centre National de Recherche et de Formation sur le Paludisme (CNRFP), Ouagadougou, Burkina Faso
[6]Dipartimento di Sanita Publica e Malattie Infettive, University of Rome La Sapienza, Rome, Italy
[7]Navrongo Health Research Centre, Navrongo, Ghana
[8]Komfo Anokye Teaching Hospital, Kumasi, Ghana
[9]Department of Biochemistry and Molecular Biology, University of Buea, Buea, Cameroon
[10]KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya
[11]Joint Malaria Programme, Kilimanjaro Christian Medical Centre, Moshi, Tanzania
[12]Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK
[13]Malawi-Liverpool Wellcome Trust Clinical Research Programme, College of Medicine, University of Malawi, Blantyre, Malawi
[14]Malaria Research and Training Centre, Faculty of Medicine, University of Bamako, Bamako, Mali
[15]MalariaGEN consortium; http://www.malariagen.net/projects/host/consortium-members
[†]corresponding authors: GBJB george@well.ox.ac.uk; CCAS spencer@well.ox.ac.uk

February 1, 2016

Abstract

Understanding patterns of genetic diversity is a crucial component of medical research in Africa. Here we use haplotype-based population genetics inference to describe gene-flow and admixture in a collection of 48 African groups with a focus on the major populations of the sub-Sahara. Our analysis presents a framework for interpreting haplotype diversity within and between population groups and provides a demographic foundation for genetic epidemiology in Africa. We show that coastal African populations have experienced an influx of Eurasian haplotypes as a series of admixture events over the last 7,000 years, and that Niger-Congo speaking groups from East and Southern Africa share ancestry with Central West Africans as a result of recent population expansions associated with the adoption of new agricultural technologies. We demonstrate that most sub-Saharan populations share ancestry with groups from outside of their current geographic region as a result of large-scale population movements over the last 4,000 years. Our in-depth analysis of admixture provides an insight into haplotype sharing across different geographic groups and the recent movement of alleles into new climatic and pathogenic environments, both of which will aid the interpretation of genetic studies of disease in sub-Saharan Africa.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Genetic epidemiological studies aim to uncover novel relationships between genes, the environment, and disease [Malaria Genomic Epidemiology Network, 2015]. A critical component of this work is a description of patterns of genetic variation and an understanding of the genetic structure of populations. Whilst tens of thousands of genetic variants have been associated with different diseases in populations of European descent [Welter et al., 2014], despite the high burden of disease in Africa, medical genetic research on the continent has lagged behind [Need and Goldstein, 2009]. To alleviate this, several broad consortia are beginning to focus on understanding the genetic basis of infectious and non-communicable disease specifically in Africa [Malaria Genomic Epidemiology Network, 2008; H3Africa Consortium, 2014; Gurdasani et al., 2014; Malaria Genomic Epidemiology Network, 2015], and a number of recent studies have sought to describe patterns of genetic variation across the continent [Campbell and Tishkoff, 2008; Tishkoff et al., 2009; Gurdasani et al., 2014].

Genome-wide analyses of African populations are refining previous models of the continent's genetic history. One such emerging insight is the identification of clear, but complex, evidence for the movement of Eurasian ancestry back into the continent as a result of admixture over a variety of timescales [Pagani et al., 2012; Pickrell et al., 2014; Gurdasani et al., 2014; Hodgson et al., 2014a; Llorente et al., 2015]. Admixture occurs when genetically differentiated ancestral groups come together and mix, a process which is increasingly regarded as a common feature of human populations across the globe [Patterson et al., 2012; Hellenthal et al., 2014; Busby et al., 2015]. In a broad sample of 18 ethnic groups from eight countries, the African Genome Variation Project (AGVP) [Gurdasani et al., 2014] recreated previously published results to identify recent Eurasian admixture within the last 1.5 thousand years (ky) in the Fulani of West Africa [Tishkoff et al., 2009; Henn et al., 2012] and several East African groups from Kenya, older Eurasian ancestry (2-5 ky) in Ethiopian groups, consistent with previous studies of similar populations [Pagani et al., 2012; Pickrell et al., 2014], and a novel signal of ancient (>7.5 ky) Eurasian admixture in the Yoruba of Central West Africa [Gurdasani et al., 2014]. Comparisons of contemporary sub-Saharan African populations with the first ancient genome from within Africa, a 4.5 ky Ethiopian individual [Llorente et al., 2015], provide additional support for limited migration of Eurasian ancestry back into East Africa within the last 3,000 years.

Within this timescale, the major demographic change within Africa was the transition from hunting and gathering to agricultural lifestyles [Diamond and Bellwood, 2003; Smith, 2005; Barham and Mitchell, 2008; Li et al., 2014]. This shift was long and complex and occurred at different speeds, instigating contrasting interactions between the agriculturalist pioneers and the inhabitant people [Mitchell, 2002; Marks et al., 2014]. The change was initialised by the spread of pastoralism (i.e. the raising and herding of livestock) east and south from Central West Africa together with the branch of Niger-Congo languages known as Bantu [Mitchell, 2002; Barham and Mitchell, 2008]. The AGVP also found evidence of widespread hunter-gatherer ancestry in African populations, including ancient (9 ky) Khoesan ancestry in the Igbo from Nigeria, and more recent hunter-gatherer ancestry in eastern (2.5-4.5 ky) and southern (0.9-4 ky) African populations [Gurdasani et al., 2014]. The identification of hunter-gatherer ancestry in non-hunter-gatherer populations together with the timing of these latter events is consistent with the

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

known expansion of Bantu languages across African within the last 3 ky [Mitchell, 2002; Diamond and Bellwood, 2003; Smith, 2005; Barham and Mitchell, 2008; Marks et al., 2014; Li et al., 2014].

These studies have described the novel and important influence of both Eurasian and hunter-gatherer ancestry on the population genetic history of sub-Saharan Africa. Nevertheless, there is still more to understand about the nature and timing of admixture across Africa. For example, do we observe Eurasian ancestry in additional African populations? Do alternative dating methods recreate a similar timescale for these admixtures? Can we understand more about how the Eurasian ancestry entered African populations? Uncertainty about the relationships between populations within Africa also remains. For example, can we use genetics to refine the origins, interactions, and timings of the Bantu expansion? What did the ancestry of pre-Bantu populations of east and southern Africa look like? What other historical relationships can be characterised between different African ethno-linguistic groups?

Here we review these questions using the same methods as previous authors, and provide additional novel inference by using methods that utilise haplotype information. To gain a detailed understanding of the population structure and history of Africa, we analyse genome-wide data from 14 Eurasian and 46 sub-Saharan African groups. Half (23) of the African groups represent subsets of samples collected from nine countries as part of the MalariaGEN consortium. Details on the recruitment of samples in relation to studying malaria genetics are published elsewhere [Malaria Genomic Epidemiology Network, 2014, 2015]. The remaining 23 groups are from publicly available datasets from a further eight sub-Saharan African countries [Pagani et al., 2012; Schlebusch et al., 2012; Petersen et al., 2013] and the 1000 Genomes Project (1KGP), with Eurasian groups from the latter included to help understand the genetic contribution from outside of the continent (Figure 1-figure supplement 1). Although we have representative groups from all four major African linguistic macro-families (Supplementary File 1), our sample represents a significant proportion of the sub-Saharan population in terms of number, but not does not equate to a complete picture of African ethnic diversity.

To study patterns of genetic diversity we created an integrated dataset at a common set of over 328,000 high-quality SNP genotypes and use established approaches for comparing population allele frequencies across groups to provide a baseline view of historical gene flow. We apply statistical approaches to phasing genotypes to obtain haplotypes for each individual, and use previously published methods to represent the haplotypes that an individual carries as a mosaic of other haplotypes in the sample (so-called chromosome painting [Li and Stephens, 2003]). We use these data to demonstrate that haplotype-based methods have the potential to tease apart subtle relationships between closely related populations. We present a detailed picture of haplotype sharing across sub-Saharan Africa using a model-based clustering approach that groups individuals using haplotype information alone. The inferred groups reflect broad-scale geographic patterns. At finer scales, our analysis reveals smaller groups, and often differentiates closely related populations consistent with self-reported ancestry [Tishkoff et al., 2009; Bryc et al., 2010; Hodgson et al., 2014a]. We describe these patterns by measuring gene flow between populations and relate them to potential historical movements of people into and within sub-Saharan Africa. Understanding the extent to which individuals share haplotypes (which we call *coancestry*), rather than independent markers, can provide a rich description of ancestral relationships and population history [Lawson et al., 2012; Leslie et al., 2015]. For each group we use the latest analytical tools to characterise the populations

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

as mixtures of haplotypes and provide estimates for the date of admixture events [Lawson et al., 2012; Hellenthal et al., 2014; Leslie et al., 2015; Montinaro et al., 2015]. As well as providing a quantitative measure of the coancestry between groups, we identify the detailed dominant events which have shaped current genetic diversity in sub-Saharan Africa. We discuss the relevance of these observations to studying genotype-phenotype associations in Africa, particularly in the context of infectious disease.

Results

Broad-scale population structure reflects geography and language

Throughout this article we use shorthand current-day geographical and ethno-linguistic labels to describe ancestry. For example we write "Eurasian ancestry in East African Niger-Congo speakers", where the more precise definition would be "ancestry originating from groups currently living in Eurasia in groups currently living in East Africa that speak Niger-Congo languages" [Pickrell et al., 2014]. Our combined dataset included 3,350 individuals from 46 sub-Saharan African ethnic groups, and 12 non-African populations (Figure 1A and Figure 1-figure supplement 1).

As has been shown in other regions of the world [Novembre et al., 2008; Reich et al., 2009; Behar et al., 2010], analyses of population structure using principal component analysis (PCA) show that genetic relationships are broadly defined by geographical and ethno-linguistic similarity (Figure 1B,C). The first two principal components (PCs) reflect ethno-linguistic divides: PC1 splits southern Khoesan speaking populations from the rest of Africa, and PC2 splits the Afroasiatic and Nilo-Saharan speakers from East Africa from sub-Saharan African Niger-Congo speakers. The third axis of variation defines east versus west Africa, suggesting that in general, population structure in Africa largely mirrors linguistic and geographic similarity [Tishkoff et al., 2009].

Using a previously published implementation of chromosome painting (CHROMOPAINTER [Lawson et al., 2012]), we reconstructed the genomes of each study individual as mosaics of haplotype segments (or chunks) from all other individuals. We used a clustering algorithm implemented in fineSTRUCTURE [Lawson et al., 2012] to cluster individuals into groups based purely on the similarity of these paintings (Figure 1 and Figure 1-figure supplement 3). More specifically, for a given recipient individual, we can estimate the amount of their genome that is shared with (or copied from) each of a set of donor individuals based on the paintings, which are called copying vectors. These vectors are clustered hierarchically to form a tree which describes the inferred relationship between different groups (Figure 1-figure supplement 3). As such, this method uses chromosome painting to describe the coancestry between groups, which can then be visualised as a heatmap, as in Figure 1D.

Consistent with PCA, African populations tend to share more DNA with geographically proximate populations (dark colours on the diagonal; Figure 1D). Block structures on this diagonal indicate higher levels of haplotype sharing within groups, which is indicative of close genealogical relationships with other members of your group. These patterns can be seen in some of the Khoesan speaking individuals (eg. the Ju/'hoansi), several groups from the East Africa (Sudanese, Ari, and Somali groups), and the Fulani and Jola from the Gambia, each of which provides clues to the ancestral connections between the groups in

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

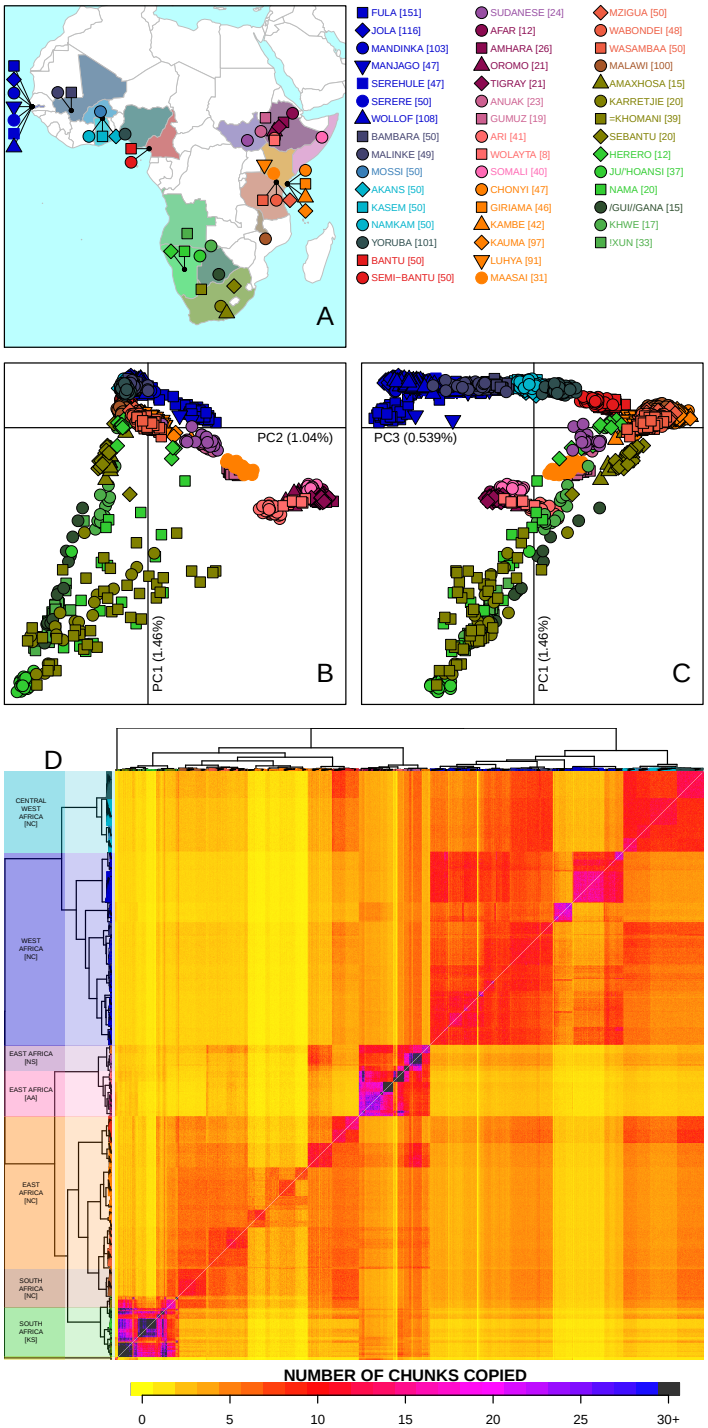

Figure 1. Sub-Saharan African genetic variation is shaped by ethno-linguistic and geographical similarity. (A) the origin of the 46 African ethnic groups used in the analysis; ethnic groups from the same country are given the same colour, but different shapes; the legend describes the identity of each point. Figure 1-figure supplement 1 and Figure 1-Source Data 1 provide further detail on the provenance of these samples. (B) PC1vPC2 shows that the first major axis of variation in Africa splits southern groups from the rest of Africa, each symbol represents an individual; PC2 reflects ethno-linguistic differences, with Niger-Congo speakers split from Afroasiatic and Nilo-Saharan speakers. (C) PC1vPC3 shows that the third principle component represents geographical separation of Niger-Congo speakers, forming a cline from west to east Africans. (D) results of the fineSTRUCTURE clustering analysis using copying vectors generated from chromosome painting; each row of the heatmap is a recipient copying vector showing the number of chunks shared between the recipient and every individual as a donor (columns);the tree clusters individuals with similar copying vectors together, such that block-like patterns are observed on the heat map; darker colours on the heatmap represent more haplotype sharing (see text for details); individual tips of the tree are coloured by country of origin, and the seven ancestry regions are identified and labelled to the left of the tree; labels in parentheses describe the major linguistic type of the ethnic groups within: AA = Afroasiatic, KS = Khoesan, NC = Niger-Congo, NS = Nilo-Saharan.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

our analysis. The heatmap also shows evidence for coancestry across regions (dark colours away from the diagonal), which can be informative of historical connections between modern-day groups. For example, east Africans from Kenya, Malawi and Tanzania tend to share more DNA with west Africans (lower right) which suggests that more haplotypes may have spread from west to east Africa (upper left) than vice versa. Using the results of the PCA and fineSTRUCTURE analyses together with ethno-linguistic classifications and geography, we defined seven groups of populations within Africa (Supplementary File 1), which we refer to as *ancestry regions* when describing gene-flow across Africa. Under this definition, there is widespread sharing of haplotypes within and between ancestry regions (Figure 1).

Haplotypes reveal subtle population structure

To quantify the extent of the genetic difference between groups we used two different metrics. First, we used the classical measure F_{ST} [Hudson et al., 1992; Bhatia et al., 2013] which measures the differentiation in allele frequencies between populations. It can be thought of as measuring the proportion of the heterozygosity at SNPs explained by the group labels. The second metric uses the similarity in copying patterns between two groups to estimate the total variation difference (TVD) at the haplotypic level. Figure 2A shows these two metrics side by side in the upper and lower diagonal. When compared to non-African populations, F_{ST} measured at our integrated set of SNPs is relatively low between many groups from West, Central, and East Africa (yellows on the upper right triangle), whereas TVD in the same populations can reveal haplotypic differences as strong as between Europe and Asia (pink and purples in lower left triangle). For example, the Chonyi from Kenya have relatively low F_{ST} but high TVD with West African groups, like the Jola (Chonyi-Jola $F_{ST} = 0.019$; Chonyi-Jola TVD $= 0.803$) suggesting that, whilst allele frequency differences between the two populations are relatively low, when we compare the populations' ancestry vectors, the haplotypic differences are some the strongest between sub-Saharan groups. In fact, whilst pairwise TVD tends to increase with pairwise F_{ST} (Pearson's correlation $R^2 = 0.79$) the relationship is not linear (Figure 2B) demonstrating that this discrepancy is not simply a function of TVD generating larger statistics. We note that high values of TVD are not sufficient to infer specific demographic events [van Dorp et al., 2015] and therefore do not interpret these differences as necessarily resulting from admixture, but as motivation to use haplotype-based methods to characterise population relationships.

Widespread evidence for admixture

We next applied similar approaches to previous authors to infer admixture between populations using correlations in allele frequencies within and between populations [Pickrell et al., 2014; Gurdasani et al., 2014]. The first approach, the three-population test (f_3 statistic [Reich et al., 2009]), uses correlations in allele frequencies between a target population and two potential source populations to identify significant departures from the null model of no admixture. Negative values are indicative of canonical admixture events where the allele frequencies in the target population are intermediate between the two source populations. Consistent with recent research [Pickrell et al., 2014; Pickrell and Reich, 2014; Gurdasani et al., 2014; Llorente et al., 2015], the majority (80%, 40/48), but not all, of the African groups surveyed

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

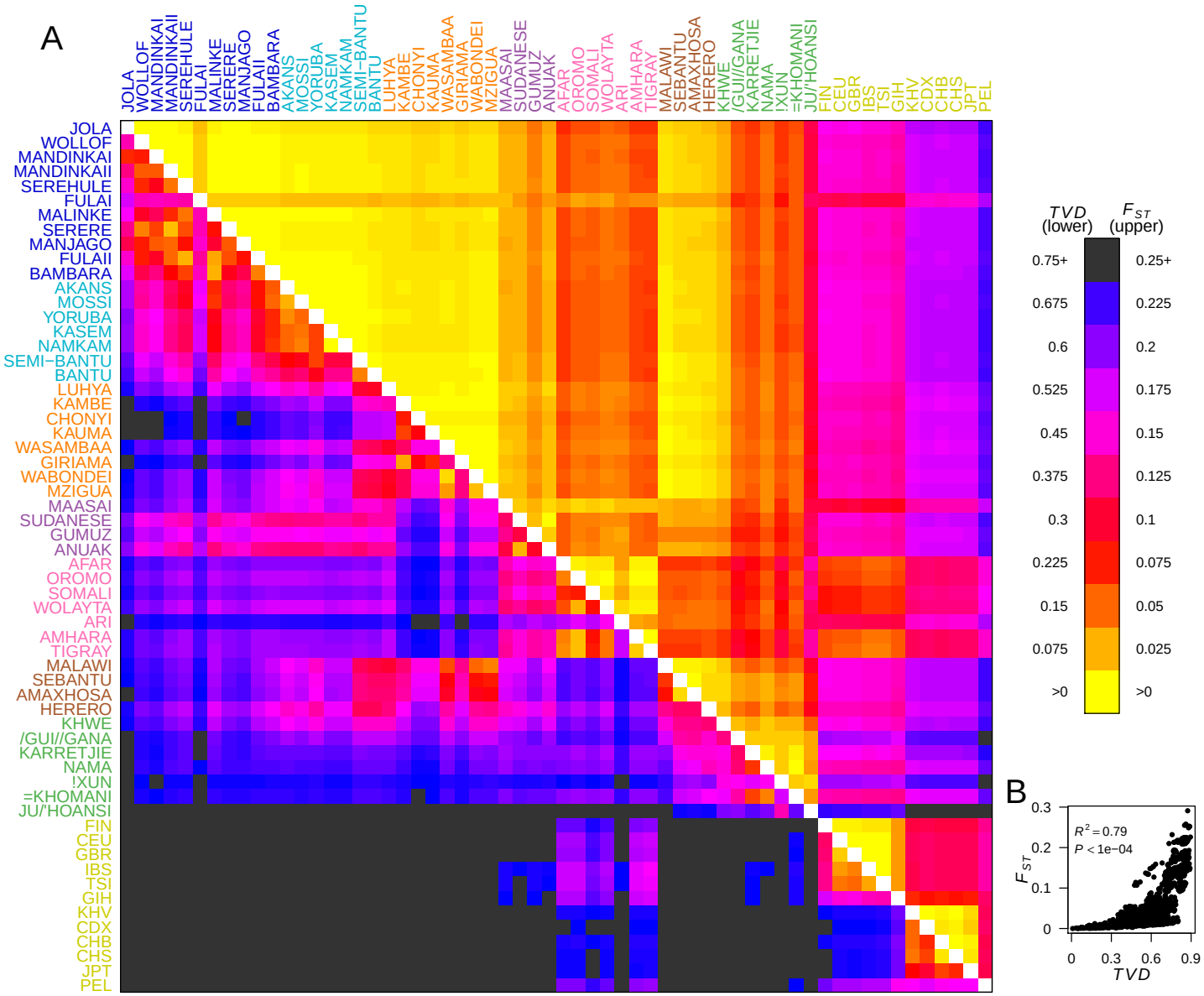

Figure 2. Haplotypes capture more population structure than independent loci. (A) For each population pair, we estimated pairwise F_{ST} (upper right triangle) using 328,000 independent SNPs, and TVD (lower left triangle) using population averaged copying vectors from CHROMOPAINTER. TVD measures the difference between two copying vectors. (B) Comparison of pairwise F_{ST} and TVD shows that they are not linearly related: some population pairs have low F_{ST} and high TVD. [Source data is detailed in Figure 2-Source Data 1 to Figure 2-Source Data 4]

showed evidence of admixture (f_3<-5), with the most negative f_3 statistic tending to involve either a Eurasian or African hunter-gatherer group [Gurdasani et al., 2014] (Supplementary Table 1).

The second approach, ALDER [Loh et al., 2013; Pickrell et al., 2014] (Supplementary Table 1) exploits

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the fact that correlations between allele frequencies along the genome decay over time as a result of recombination. Linkage disequilibrium (LD) can be generated by admixture events, and leaves detectable signals in the genome that can be used to infer historical processes from allele frequency data [Loh et al., 2013]. Following Pickrell et al. [2014] and the AGVP [Gurdasani et al., 2014], we computed weighted admixture LD curves using the ALDER [Loh et al., 2013] package to characterise the sources and timing of gene flow events . Specifically, we estimated the y-axis intercept (amplitude) of weighted LD curves for each target population using curves from an analysis where one of the sources was the target population (self reference) and the other was, separately, each of the other (non-self reference) populations. Theory predicts that the amplitude of these "one-reference" curves becomes larger the more similar the non-self reference population is to the true admixing source [Loh et al., 2013]. As with the f_3 analysis outlined above, for many of the sub-Saharan African populations, Eurasian and hunter-gatherer groups produced the largest amplitudes (Figure 3-figure supplement 1 and Figure 3-figure supplement 2), reinforcing the contribution of Eurasian admixture in our broad set of African populations.

We investigated the evidence for more complex admixture using MALDER [Pickrell et al., 2014], an implementation of ALDER which fits a mixture of exponentials to weighted LD curves to infer multiple admixture events (Figure 3 and Table Figure 3-Source Data 1). In Figure 3A, for each target population, we show the ancestry region of the two populations involved in generating the MALDER curves with the greatest amplitudes, together with the date of admixture for at most two events. Throughout, we convert time since admixture in generations to a date by assuming a generation time of 29 years [Fenner, 2005]. In general, we find that groups from similar ancestry regions tend to have inferred events at similar times and between similar groups (Figure 3), which suggests that genetic variation in groups has been shaped by shared historical events. For every event, the curves with the greatest amplitudes involved a population from a (normally non-Khoesan) African population on the one side, and either a Eurasia or Khoesan population on the other. To provide more detail on the composition of the admixture sources, we compared MALDER curve amplitudes between curves involving populations from different ancestry regions (central panel Figure 3). In general, this analysis showed that, with a few exceptions, we were unable to precisely define the ancestry of the African source of admixture, as curves involving populations from multiple different regions were not statistically different from each other ($Z < 2$; SOURCE 1). Conversely, comparisons of MALDER curves when the second source of admixture was Eurasian (yellow) or Khoesan (green), showed that these groups were ususally the single best surrogate for the second source of admixture (SOURCE 2).

We performed multiple MALDER analyses, varying the input parameters and the genetic map used. We observed a large amount of shared LD at short genetic distances between different African populations (Figure 3-figure supplement 3 and Figure 3-figure supplement 4). Such patterns may result from population genetic processes other than admixture, such as shared demographic history and population bottlenecks [Loh et al., 2013]. In the main MALDER analysis we present, we have removed the effect of this short-range LD by generating curves only after ignoring SNPs at short genetic distances where frequencies of markers at such distances are correlated between target and reference population – which can be thought of as conservative – but provide supplementary analyses where this parameter was relaxed. The main difference between the two types of analysis is that we were only able to identify ancient events

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in West Africa if we force the MALDER algorithm to start computing LD decay curves from a genetic distance of 0.5cM, irrespective of any short-range correlations in LD between populations.

Many West and Central West Africans show evidence of recent (within the last 4 ky) admixture involving African and Eurasian sources. The Mossi from Burkina Faso have the oldest inferred date of admixture, at roughly 5000BCE, but we were unable to recreate previously reported ancient admixture events in other Central West African groups [Gurdasani et al., 2014] (although see Figure 3-figure supplement 5 for results where we use different MALDER parameters). Across East Africa Niger-Congo speakers (orange) we infer admixture within the last 4 ky (and often within the last 1 ky) involving Eurasian sources on the one hand, and African sources containing ancestry from other Niger-Congo speaking African groups from the west on the other. Despite events between African and Eurasian sources appearing older in the Nilo-Saharan and Afroasiatic speakers from East Africa, we see a similar signal of very recent Central West African ancestry in a number of Khoesan groups from Southern Africa, such as the Khwe, /Gui //Gana, and !Xun, together with Malawi-like (brown) sources of ancestry in recent admixture events in East African Niger-Congo speakers.

Inference of older events relies on modelling the decay of LD over short genetic distances because recombination has had more time to break down correlations in allele frequencies between neighbouring SNPs. We investigated the effect of using European (CEU) and Central West African (YRI) specific recombination maps [Hinch et al., 2011] on the dating inference. Whilst dates inferred using the CEU map were consistent with those using the HAPMAP recombination map (Figure 3B), when using the African map dates were consistently older (Figure 3C), although still generally still within the last 7ky. There was also variability in the number of inferred admixture events for some populations between the different map analyses (Figure 3-figure supplement 6 and Figure 3-figure supplement 7).

Most events involved sources where Eurasian (dark yellow in Figure 3A) groups gave the largest amplitudes. In considering this observation, it is important to note that the amplitude of LD curves will partly be determined by the extent to which a reference population has differentiated from the target. Due to the genetic drift associated with the out-of-Africa bottleneck and subsequent expansion, Eurasian groups will tend to generate the largest curve amplitudes even if the proportion of this ancestry in the true admixing source is small [Pickrell et al., 2014] (in our dataset, the mean pairwise F_{ST} between Eurasian and African populations is 0.157; Figure 2A and Figure 2-Source Data 1). To some extent this also applies to Khoesan groups (green in Figure 3A), who are also relatively differentiated from other African groups (mean pairwise F_{ST} between Ju/'hoansi and all other African populations in our dataset is 0.095; Figure 2A and Figure 2-Source Data 1). In light of this, and the observation that curves involving groups from different ancestry regions are often no different from each other, it is therefore difficult to infer the proportion or nature of the African, Khoesan, or Eurasian admixing sources, only that the sources themselves contained African, Khoesan, or Eurasian ancestry. Moreover, given uncertainty in the dating of admixture when using different maps and MALDER parameters, these results should be taken as a guide to the general genealogical relationships between African groups, rather than a precise description of the gene-flow events that have shaped Africa.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

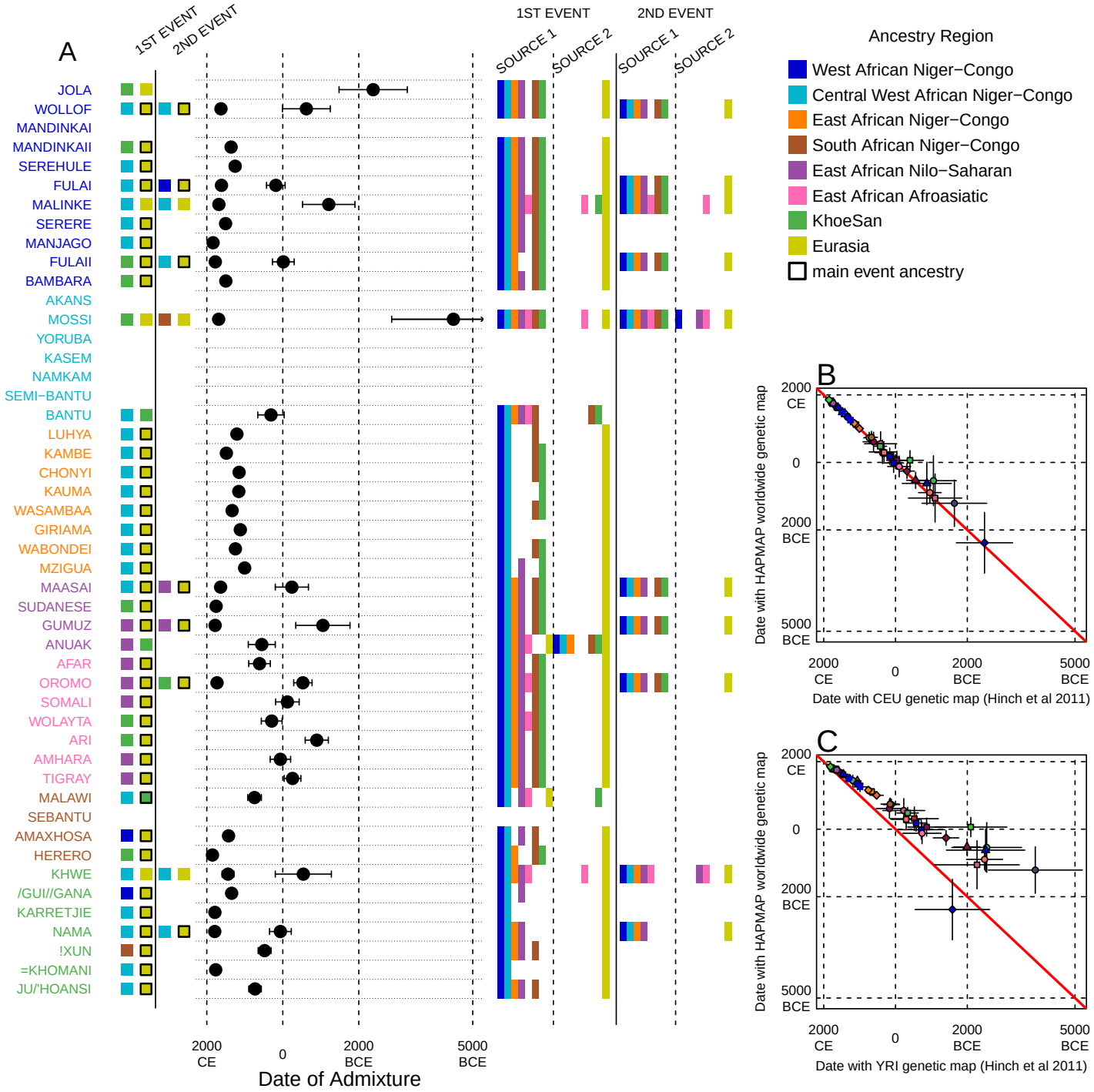

Figure 3. Inference of admixture in sub-Saharan Africa using MALDER. We used MALDER to identify the evidence for multiple waves of admixture in each population. (A) For each population, we show the ancestry region identity of the two populations involved in generating the MALDER curves with the greatest amplitudes (coloured blocks) for at most two events. The major contributing sources are highlighted with a black box. Populations are ordered by ancestry of the admixture sources and dates estimates which are shown $\pm$ 1 s.e. For each event we compared the MALDER curves with the greatest amplitude to other curves involving populations from different ancestry regions. In the central panel, for each source, we highlight the ancestry regions providing curves that are not significantly different from the best curves. In the Jola, for example, this analysis shows that, although the curve with the greatest amplitude is given by Khoesan (green) and Eurasian (yellow) populations, curves containing populations from any other African group (apart from Afroasiatic) in place of a Khoesan population are not significantly smaller than this best curve (SOURCE 1). Conversely, when comparing curves where a Eurasian population is substituted with a population from another group, all curve amplitudes are significantly smaller ($Z < 2$). (B) Comparison of dates of admixture $\pm$ 1 s.e. for MALDER dates inferred using the HAPMAP recombination map and a recombination map inferred from European (CEU) individuals from [Hinch et al., 2011]. We only show comparisons for dates where the same number of events were inferred using both methods. Point symbols refer to populations and are as in Figure 1. (C) as (B) but comparison uses an African (YRI) map. Source data can be found in Figure 3-Source Data 1

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Modelling gene flow with haplotypes

So far, our analyses have largely recapitulated recent studies of the genetic history of African populations, albeit across a broader set of sub-Saharan populations. However, we were interested in gaining a more detailed characterisation of historical gene-flow to provide an understanding and framework with which to inform genetic epidemiological studies in Africa. The chromosome painting methodology described above provides an alternative approach to inferring admixture events which directly models the similarity in haplotypes between pairs of individuals. Evidence of recent haplotype sharing suggests that the ancestors of two individuals must have been geographically proximal at some point in the past. More generally, the distance over which haplotype sharing extends along chromosomes is inverse to how far in the past the coancestry event occurred. We can use copying vectors inferred through chromosome painting to help identify those populations that share ancestry with a recipient group by fitting each vector as a mixture of all other population vectors (Figure 4A) [Leslie et al., 2015; Montinaro et al., 2015; van Dorp et al., 2015]. Figure 4A shows the contribution that each ancestry region makes to these mixtures (MIXTURE MODEL column). Almost all groups can best be described as mixtures of ancestry from different regions. For example, the copying vector of the Bantu ethnic group from Cameroon is best described as a combination of 40% Central West African Niger-Congo (sky blue), 30% Eastern Niger-Congo (orange), 25% Southern Niger-Congo (brown), and the remaining 5% coming from West African Niger-Congo (dark blue) and Khoesan-speaking (green) groups. The key insight from this analysis is that many African groups share fragments of haplotypes with groups from outside of their own ancestry region.

We used GLOBETROTTER [Hellenthal et al., 2014], a method that tests for and characterises admixture as an extension of the mixture model approach described above, to gain a more detailed understanding of recent ancestry. Admixture inference can be challenging for a number of reasons: the true admixing source population is often not well represented by a single sampled population; admixture could have occurred in several bursts, or over a sustained period of time; and multiple groups may have come together as complex convolution of admixture events. GLOBETROTTER aims to overcome some of these challenges, in part by using painted chromosomes to explicitly model the correlation structure among nearby SNPs, but also by allowing the sources of admixture themselves to be mixed [Hellenthal et al., 2014]. In addition, the approach has been shown to be relatively insensitive to the genetic map used [Hellenthal et al., 2014], and therefore potentially provides a more robust inference of admixture events, the ancestries involved, and their dates. GLOBETROTTER uses the distance between chromosomal chunks of the same ancestry to infer the time since historical admixture has occurred.

Throughout the following discussion, we refer to target populations as *recipients*, any other sampled populations used to describe the recipient population's admixture event(s) as *surrogates*, and populations used to paint both target and surrogate populations as *donors*. Including closely related individuals in chromosome painting analyses can cause the resulting painted chromosomes to be dominated by donors from these close genealogical relationships, which can mask signals of admixture in the genome [Hellenthal et al., 2014; van Dorp et al., 2015]. To help ameliorate this, we painted chromosomes for the GLOBETROTTER analysis by performing a fresh run of CHROMOPAINTER where we painted each individual from a recipient group with a set of donors which did not include individuals from within their

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

own ancestry region. We additionally painted all (59) other surrogate populations with the same set of non-local donors, and used these copying vectors, together with the non-local painted chromosomes, to infer admixture. Using this approach, we found evidence of recent admixture in all African populations (Figure 4A). To summarise these events, we show the composition of the admixing source groups as barplots for each population coloured by the contribution from each African ancestry regions and Eurasia, alongside the inferred date with confidence interval determined by bootstrapping, and the estimated proportion of admixture (Figure 4). For each event we also identify the best matching donor population to the admixture sources. We describe observations from these analyses below.

Direct and indirect gene flow from Eurasia back into Africa

In contrast to the MALDER analysis, we did not find evidence for recent Eurasian admixture in every African population (Figure 4). In particular, in several groups from South Africa and all from the Central West African ancestry region, which includes populations from Ghana, Nigeria, and Cameroon, we infer admixture between groups that are best represented by contemporary populations residing in Africa. As GLOBETROTTER is designed to identify the most recent admixture event(s) [Hellenthal et al., 2014], this observation does not rule out gene-flow from Eurasia back into these groups, but does suggest that subsequent movements between African groups were important in generating the contemporary ancestry of Central West and Southern African Niger-Congo speaking groups. With some exceptions that we describe below, we also do not observe Eurasian ancestry in all East African Niger-Congo speakers, instead finding more evidence for coancestry with Afroasiatic speaking groups. As we show later, Afroasiatic populations have a significant amount of genetic ancestry from outside of Africa, so the observation of this ancestry in several African groups identifies a route by which Eurasian ancestry may have indirectly entered the continent [Pickrell et al., 2014].

In fact, characterising admixture sources as mixtures allows us to infer whether Eurasian haplotypes are likely to have come directly into sub-Saharan Africa – in which case the admixture source will contain only Eurasian surrogates – or whether Eurasian haplotypes were brought indirectly together with sub-Saharan groups. In West African Niger-Congo speakers from The Gambia and Mali, we infer admixture involving minor admixture sources which contain mostly Eurasian (dark yellow) and Central West African (sky blue) ancestry, which most closely match the contemporary copying vectors of northern European populations (CEU and GBR) or the Fulani (FULAI, highlighted in gold in Figure 4A). The Fulani, a nomadic pastoralist group found across West Africa, were sampled in The Gambia, at the very western edge of their current range, and have previously reported genetic affinities with Niger-Congo speaking, Sudanic, Saharan, and Eurasian populations [Tishkoff et al., 2009; Henn et al., 2012], consistent with the results of our mixture model analysis (Figure 4A). Admixture in the Fulani differs from other populations from this region, with sources containing greater amounts of Eurasian and Afroasiatic ancestry, but appears to have occurred during roughly the same period (c. 0CE; Figure 5).

The Fulani represent the best-matching surrogate to the minor source of recent admixture in the Jola and Manjago, which we interpret as resulting not from specific admixture from them into these groups, but because the mix of African and Eurasian ancestries in contemporary Fulani is the best proxy for

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

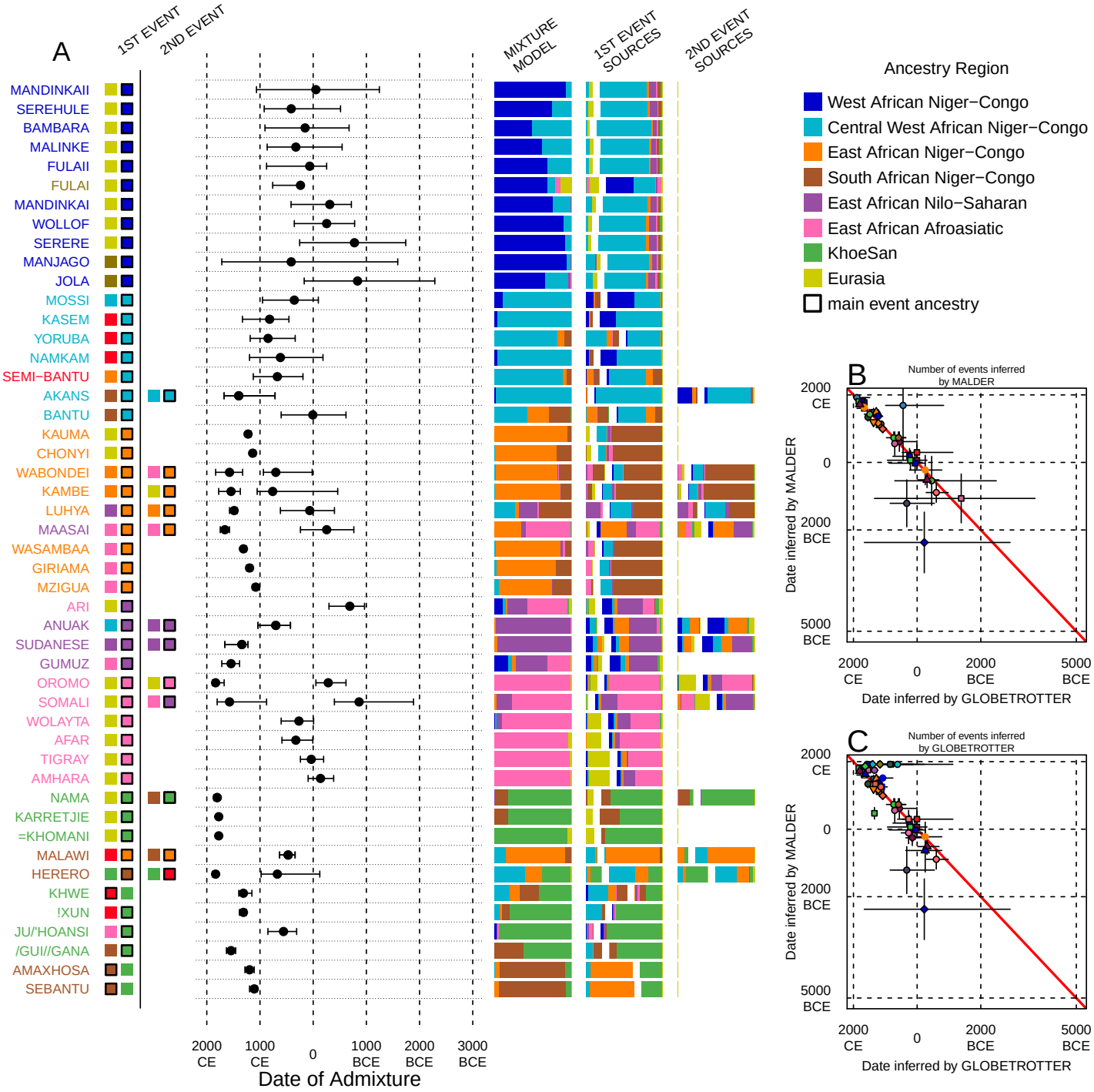

Figure 4. Inference of admixture in sub-Saharan African using GLOBETROTTER (A) For each group we show the ancestry region identity of the best matching source for the first and, if applicable, second events. Events involving sources that most closely match FULAI and SEMI-BANTU are highlighted by golden and red colours, respectively. Second events can be either multiway, in which case there is a single date estimate, or two-date in which case the 2ND EVENT refers to the earlier event. The point estimate of the admixture date is shown as a black point, with 95% CI shown with lines. MIXTURE MODEL: We infer the ancestry composition of each African group by fitting its copying vector as a mixture of all other population copying vectors. The coefficients of this regression sum to 1 and are coloured by broad ethno-linguistic origin. 1ST EVENT and 2ND EVENT SOURCES shows the ancestry breakdown of the admixture sources inferred by GLOBETROTTER, coloured by ancestry region as in the key top right. (B) and (C) Comparisons of dates inferred by MALDER and GLOBETROTTER. Because the two methods sometimes inferred different numbers of events, in (B) we show the comparison for based on the inferred number of events in the MALDER analysis, and in (C) for the number of events inferred by GLOBETROTTER. Point symbols refer to populations and are as in Figure 1 and source data can be found in Figure 4-Source Data 1.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the minor sources of admixture in this region. With the exception of the Fulani themselves, the major admixture source in groups across this region is a similar mixture of African ancestries that most closely matches contemporary Gambian and Malian surrogates (Jola, Serere, Serehule, and Malinke), suggesting ancestry from a common West African group within the last 3,000 years. The Ghana Empire flourished in West Africa between 300 and 1200CE, and is one of the earliest recorded African states [Roberts, 2007]. Whilst its origins are uncertain, it is clear that trade in gold, salt, and slaves across the Sahara, perhaps from as early as the Roman Period, as well as evolving agricultural technologies, were the driving forces behind its development [Oliver and Fagan, 1975; Roberts, 2007]. The observation of Eurasian admixture in our analysis is at least consistent with this moment in African history, and suggests that ancestry in groups from across this region of West Africa is the result of interactions through North Africa that were catalysed by trade across the Sahara.

We infer direct admixture from Eurasian sources in two populations from Kenya, where specifically South Asian populations (GIH, KHV) are the most closely matched surrogates to the minor sources of admixture (Figure 5). Interestingly, the Chonyi (1138CE: 1080-1182CE) and Kauma (1225CE: 1167-1254CE) are located on the so-called Swahili Coast, a region where Medieval trade across the Indian Ocean is historically documented [Allen, 1993]. In the Kambe, the third group from coastal Kenya, we infer two events, the more recent one involving local groups, and the earlier event involving a European-like source (GBR, 761CE: 461BCE-1053CE). In Tanzanian groups from the same ancestry region, we infer admixture during the same period, this time involving minor admixture sources with different, Afroasiatic ancestry: in the Giriama (1196CE: 1138-1254CE), Wasambaa (1312CE: 1254-1341CE), and Mzigua (1080: 1007-1138CE). Although the proportions of admixture from these sources differ, when we consider the major sources of admixture in East African Niger-Congo speakers, they are again similar, containing a mix of mainly local Southern Niger-Congo (Malawi), Central West African, Afroasiatic, and Nilo-Saharan ancestries.

In the Afroasiatic speaking populations of East Africa we infer admixture involving sources containing mostly Eurasian ancestry, which most closely matches the Tuscans (TSI, Figure 4). This ancestry appears to have entered the Horn of Africa in three distinct waves (Figure 5). We infer admixture involving Eurasian sources in the Afar (326CE: 7-587CE), Wolayta (268CE: 8BCE-602CE), Tigray (36CE: 196BCE-240CE), and Ari (689BCE:965-297BCE). There are no Middle Eastern groups in our analysis, and this latter group of events may represent previously observed migrations from the Arabian peninsular from the same time [Pagani et al., 2012; Hodgson et al., 2014a]. Considering Afroasiatic and Nilo-Saharan speakers separately, the ancestry of the major sources of admixture of the former are predominantly local (purple), indicative of less historical interaction with Niger-Congo speakers due to their previously reported Middle Eastern ancestry [Pagani et al., 2012]. In Nilo-Saharan speaking groups, the Sudanese (1341CE: 1225-1660), Gumuz (1544CE: 1384-1718), Anuak (703: 427-1037CE), and Maasai (1646CE: 1584-1743CE), we infer greater proportions of West (blue) and East (orange) African Niger-Congo speaking surrogates in the major sources of admixture, indicating both that the Eurasian admixture occurred into groups with mixed Niger-Congo and Nilo-Saharan / Afroasiatic ancestry, and a clear recent link with Central and West African groups.

In two Khoesan speaking groups from South Africa, the ≠Khomani and Karretjie, we infer very recent

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

direct admixture involving Eurasian groups most similar to Northern European populations, with dates aligning to European colonial period settlement in Southern Africa (c. 5 generations or 225 years ago; Figure 5) [Hellenthal et al., 2014]. Taken together, and in addition the MALDER analysis above, these observations suggest that gene flow back into Africa from Eurasia has been common around the edges of the continent, has been sustained over the last 3,000 years, and can often be attributed to specific and different historical time periods.

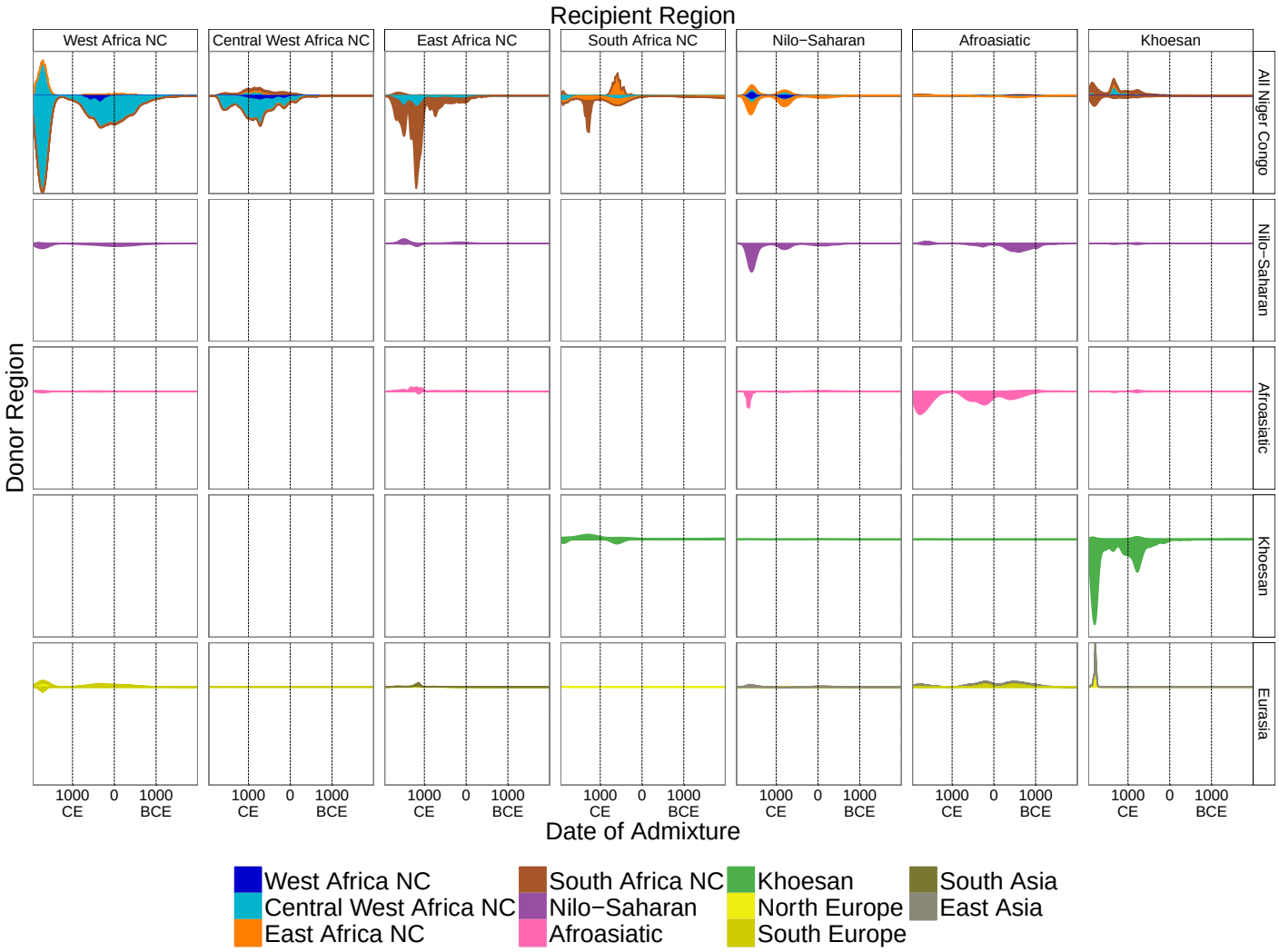

Figure 5. A timeline of recent admixture in sub-Saharan Africa. For all events involving recipient groups from each ancestry region (columns) we combine all date bootstrap estimates generated by GLOBETROTTER and show the densities of these dates separately for the minor (above line) and major (below line) sources of admixture. Dates are additionally stratified by the ancestry region of the surrogate populations (rows), with all dates involving Niger Congo speaking regions combined together (All Niger Congo). Within each panel, the densities are coloured by the geographic (country) origin of the surrogates and in proportion to the components of admixture involved in the admixture event. The integrals of the densities are proportional to the admixture proportions of the events contributing to them.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Population movements within Africa and the Bantu expansion

Admixture events involving sources that best match populations from within Africa tend to involve local groups. Even so, there are several long range admixture events of note. In the Ju/'hoansi, a San group from Namibia, we infer admixture involving a source that closely matches a local southern African Khoesan group, the Karretjie, and an East African Afroasiatic, specifically Somali, source at 558CE (311-851CE). We infer further events with minor sources most similar to present day Afroasiatic speakers in the Maasai (1660CE: 1573-1747CE and 254BCE: 764-239BCE) and, as mentioned, all four Tanzanian populations (Wabondei, Wasambaa, Giriama, and Mzigua). In contrast, the recent event inferred in the Luhya (1486: 1428-1573CE) involves a Nilo-Saharan-like minor source. The recent dates of these events imply not only that Eastern Niger-Congo speaking groups have been interacting with nearby Nilo-Saharan and Afroasiatic speakers – who themselves have ancestry from more northerly regions – after the putative arrival of Bantu-speaking groups to Eastern Africa, but also that the major sources of admixture contained both Central West and a majority of Southern Niger-Congo ancestry.

African languages can be broadly classified into four major macro-families: Afroasiatic, Nilo-Saharan, Niger-Congo, and Khoesan. Most of the sampled groups in this study, and indeed most sub-Saharan Africans, speak a language belonging to the Niger Congo lingustic phylum [Greenberg, 1972; Nurse and Philippson, 2003]. A sub-branch of this group are the so-called "Bantu" languages – a group of approximately 500 very closely related languages – that are of particular interest because they are spoken by the vast majority of Africans south of the line between Southern Nigeria/Cameroon and Somalia [Pakendorf et al., 2011]. Given the their high similarity and broad geographic range, it is likely that Bantu languages spread across Africa quickly. Whether this cultural expansion was accompanied by people is an active research question, but an increasing number of molecular studies, mostly using uni-parental genetic markers, indicate that the expanson of languages was accompanied by the diffusion of people [Beleza et al., 2005; Berniell-Lee et al., 2009; Pakendorf et al., 2011; de Filippo et al., 2012; Ansari Pour et al., 2013; Li et al., 2014; González-Santos et al., 2015]. Bantu languages can themselves be divided into three major groups: northwestern, which are spoken by groups near to the proto-Bantu heartland of Nigeria / Cameroon; western Bantu languages, spoken by groups situated down the west coast of Africa; and eastern, which are spoken across East and Central Africa [Li et al., 2014]. One particular debate concerns whether eastern languages are a primary branch that split off before the western groups began to spread south (the early-split hypothesis) or whether this occurred after the migrants had begun to spread south (the late-split hypothesis) [Pakendorf et al., 2011]. Recent linguistic [Holden, 2002; Currie et al., 2013; Grollemund et al., 2015], and genetic analyses [Li et al., 2014] support the latter.

Whilst the current dataset does not cover all of Africa, and importantly contains no hunter-gather groups outside of southern Africa, we explored whether our admixture approach could be used to gain insight into the Bantu expansion. Specifically, we wanted to see whether the dates of admixture and composition of admixture sources were consistent with either of the two major models of the Bantu expansion. The major sources of admixture in East African Niger-Congo speakers tend to have both Central West and Southern Niger-Congo ancestry, although it is predominantly the latter (Figure 4). If Eastern Niger-Congo speakers derived all of their ancestry directly from Central West Africa (i.e.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the early-split hypothesis) then we would not observe any ancestry from more southerly groups, but this is not what we see. That all East African Niger-Congo speakers that we sampled have admixture ancestry from a Southern group (Malawi), suggests that this group is more closely related to their Bantu ancestors than Central West Africans on their own. To explore this further, we performed additional GLOBETROTTER analyses where we restricted the surrogate populations used to infer admixture. For example, we re-ran the admixture inference process without allowing any groups from within the same ancestry region to contribute to the admixture event. When we disallow any East African Niger-Congo speakers from forming the admixture, Southern African Niger-Congo speakers, specifically Malawi, still contribute most to the major source of admixture. In Southern African Niger-Congo groups, restricting admixture surrogates to those outside of their region leads to East African groups contributing most to the sources of admixture. The closer affinity between East and Southern Niger-Congo Bantu components is consistent with a common origin of these groups after the split from the Western Bantu. Moreover, this observation provides additional support for the hypothesis based on linguistic phylogenetics [Grollemund et al., 2015] that the main Bantu migration moved south and then east around the Congo rainforest. We performed a further restriction, disallowing any Southern or Eastern Niger-Congo speaking groups from being involved in admixture across both the South and East African Niger-Congo regions, which led to the vast majority of the ancestry of the major sources of admixture originating from Central West Africa, and almost exclusively from Cameroon populations (Figure 4-figure supplement 1 and Figure 4-figure supplement 2).

We inferred recent admixture involving major sources most similar to East or Southern African Niger-Congo speakers in the SEBantu (1109:1051-1196CE) and AmaXhosa (1196CE: 1109-1283CE) from South Africa. We infer two admixture dates in a Bantu-speaking group from Namibia, the Herero (1834CE: 1805-1892CE and 674CE: 124BCE-979CE), and a single date in the Khoesan-speaking Khwe (1312; 1152-1399CE), involving sources that more clearly contain ancestry from the Semi-Bantu from Cameroon. In a third south-west African group, the !Xun from Angola, we infer admixture from a similar Cameroon-like source at around the same time as the Khwe (1312CE: 1254-1385CE). Assuming that Cameroon populations are the best proxy for Bantu ancestry in our dataset, these results suggest a separate more recent arrival for Niger-Congo (Bantu) ancestry in south west compared to south-east Africa, with the former coming recently directly down the west coast of Africa, specifically from Cameroon, and the latter deriving their ancestry from earlier interactions via an eastern route [de Filippo et al., 2012; Li et al., 2014].

Interestingly, in individuals from Malawi we infer a multi-way event with an older date (471: 340-631CE) involving a minor source which mostly contains ancestry from Cameroon, an event that is also seen in the Herero from Namibia. This Bantu admixture appears to have preceded that in some of the other South Africans by a few hundred years, suggesting either input from both western and eastern migrating Bantu speakers after the two branches split c.1,500 years ago, or that present day Malawi harbour ancestry from groups intermediate between the two branches [de Filippo et al., 2012]. Within this group we also see an admixture source with a significant proportion of non-Bantu (green) ancestry (2nd event, minor source), ancestry which we do not observe in the mixture model analysis, but which also evident in other south-east African Niger-Congo speakers, the AmaXhosa and SEBantu, indicating that

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

gene-flow must have occurred between the expanding Bantus and the resident hunter-gatherer groups [Marks et al., 2014]. The complex ancestry composition of the sources of these events, together with the dates of admixture, suggest large-scale interactions between groups during the Bantu expansion, and also hint that the expansion may have generated multiple waves of movement south. These events have resulted in recent haplotype sharing between groups across sub-Saharan Africa.

A haplotype-based model of gene flow in sub-Saharan Africa

Our haplotype-based analyses support a complex and dynamic picture of recent historical gene flow in Africa. We next attempted to summarise these events by producing a demographic model of African genetic history (Figure 6). Using genetics to infer historical demography will always depend somewhat on the available samples and population genetics methods used to infer population relationships. The model we present here is therefore unlikely to recapitulate the full history of the continent and will be refined in the future with additional data and methodology. We again caution that we do not have an exhaustive sample of African populations, and in particular lack significant representation from extant hunter-gatherer groups outside of southern Africa. Nevertheless, there are signals in the data that our haplotype-based approach allows us to pick up, and it is therefore possible to highlight the key gene-flow events and sources of coancestry in Africa:

(1) **Colonial Era European admixture in the Khoesan.** In two southern African Khoesan groups we see very recent admixture involving northern European ancestry which likely resulted from Colonial Era movements from the UK, Germany, and the Netherlands into South Africa [Thompson, 2001].

(2) **The recent arrival of the Western Bantu expansion in southern Africa.** Central West African, and in particular ancestry from Cameroon (red ancestry in Figure 6A), is seen in Southern African Niger-Congo and Khoesan speaking groups, the Herero, Khwe and !Xun, indicating that the gradual diffusion of Bantu ancestry reached the south of the continent only within the last 750 years. Central West African ancestry in Malawi appears to have appeared prior to this event.

(3) **Medieval contact between Asia and the East African Swahili Coast.** Specific Asian gene-flow is observed into two coastal Kenyan groups, the Kauma and Chonyi, which represents a distinct route of Eurasian, in this case Asian, ancestry into Africa, perhaps as a result of Medieval trade networks between Asia and the Swahili Coast around 1200CE.

(4) **Gene-flow across the Sahara.** Over the last 3,000 years, admixture involving sources containing northern European ancestry is seen on the Western periphery of Africa, in The Gambia and Mali. This ancestry in West Africa is likely to be the result of more gradual diffusion of DNA across the Sahara from northern Africa and across the Iberian peninsular, and not via the Middle East, as in the latter scenario we would expect to see Spanish (IBS) and Italian (TSI) in the admixture sources. We do see limited southern European ancestry in West Africa (Figs. 5 and 6D) in the Fulani, suggesting that some Eurasian ancestry may also have entered West Africa via North East Africa [Henn et al., 2012].

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

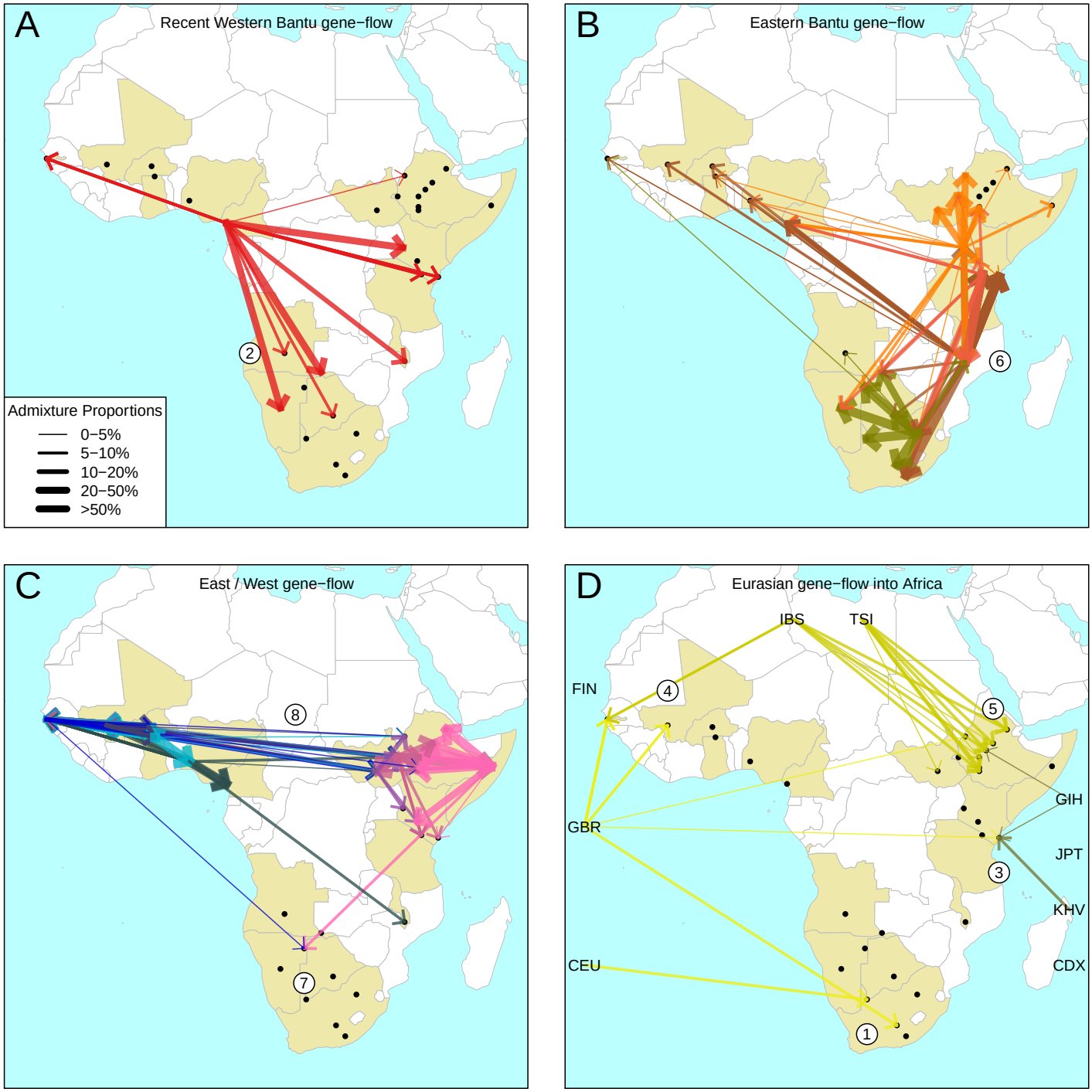

Figure 6. The geography of recent gene-flow in Africa. We summarise gene-flow events in Africa using the results of the GLOBETROTTER analysis. For each ethnic group, we inferred the composition of the admixture sources, and link recipient population to surrogates using arrows, the size of which is proportional to the amount it contributes to the admixture event. We separately plot (A) all events involving admixture source components from the Bantu and Semi-Bantu ethnic groups in Cameroon; (B) all events involving admixture sources from East and South African Niger-Congo speaking groups; (C) events involving admixture sources from West African Niger-Congo and East African Nilo-Saharan / Afroasiatic groups; (D) all events involving components from Eurasia. in (D) arrows are linked to the labelled 1KGP Eurasian groups. Arrows are coloured by country of origin, as in Figure 5. Numbers 1-8 in circles represent the events highlighted in section *A haplotype-based model of gene flow in sub-Saharan Africa*. An alternative version of this plot, stratified by date, is shown in Figure 6-figure supplement 1.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(5) **Several waves of Mediterranean / Middle Eastern ancestry into north-east Africa.** We observe southern European gene flow into East African Afroasiatic speakers over a more prolonged time period over the last 3,000 years in three major waves (Figs. 5 and 6D). We do not have Middle-Eastern groups in our analysis, so the observed Italian ancestry in the minor sources of admixture – the Tuscans are the closest Eurasian group to the Middle East – is consistent with previous results using the same samples [Pagani et al., 2012; Hodgson et al., 2014a], indicating this region as a major route for the back migration of Eurasian DNA into sub-Saharan Africa [Pagani et al., 2012; Pickrell et al., 2014].

(6) **The late split of the Eastern Bantus.** The major source of admixture in East Africa Niger-Congo speakers is consistently a mixture of Central West Africa and Southern Niger-Congo speaking groups, in particular Malawi. This result best fits a model where Bantu speakers initially spread south along the western side of the Congo rainforest before splitting off eastwards, and interacting with local groups in central south Africa – for which Malawi is our best proxy – and then moving further north-east and south (Figure 6B).

(7) **Pre-Bantu pastoralist movements from East to South Africa.** In the Ju/'hoansi we infer an admixture event involving an East African Afroasiatic source. This event precedes the arrival of Bantu-speaking groups in southern Africa, and is consistent with several recent results linking Eastern and Southern Africa and the spread of cattle pastoralism (Figs. 5 and 6C) [Pickrell et al., 2014; Ranciaro et al., 2014; Macholdt et al., 2015; Barham and Mitchell, 2008].

(8) **Ancestral connections between West Africans and the Sudan.** Concentrating on older events, we observe old "Sudanese" (Nilotic) components in very small proportions in the Gambia (Figure 4-figure supplement 1 and Figure 5) which may represent ancient expansion relationships between East and West Africa. When we infer admixture in West and Central West African groups without allowing any West Africans to contribute to the inference, we observe a clear signal of Nilo-Saharan ancestry in these groups, consistent with bidirectional movements across the Sahel [Tishkoff et al., 2009] and coancestry with (unsampled) Nilo-Saharan groups in Central West Africa. Indeed, if we look again at the PCA in Figure 1C, we observe that the Nilo-Saharan speakers are between West and East African Niger-Congo speaking individuals on PC3, an affinity which is supported by the presence of West African components in non-Niger-Congo speaking East Africans (Fig 6C).

(9) **Ancient Eurasian gene-flow back into Africa and shared hunter-gatherer ancestry.** The MALDER analysis and f_3 statistics show the general presence of ancient Eurasian and / or Khoesan ancestry across much of sub-Saharan Africa. We tentatively interpret these results as being consistent with recent research suggesting very old (>10 kya) migrations back into Africa from Eurasia [Hodgson et al., 2014a], with the ubiquitous hunter-gatherer ancestry across the continent possibly related to the inhabitant populations present across Africa prior to these more recent movements. Future research involving ancient DNA from multiple African populations will help to further characterise these observations.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

Here we present an in-depth analysis of the genetic history of sub-Saharan Africa in order to characterise its impact on present day diversity. We show that gene-flow has taken place over a variety of different time scales which suggests that, rather than being static, populations have been sharing DNA, particularly over the last 3,000 years. An unanswered question in African history is how contemporary populations relate to those present in Africa before the transition to pastoralism that began some 2,000 years ago. Whilst the f_3 and MALDER analyses show evidence for deep Eurasian and some hunter-gatherer ancestry across Africa, our GLOBETROTTER analysis provides a greater precision on the admixture sources and a timeline of events and their impact on groups in our analysis (Figure 6). The transition from foraging to pastoralism and agriculture in Africa was likely to be complex, with its impact on existing populations varying substantially. However, the similarity of language and domesticates in different parts of Africa implies that this transition spread; there are few cereals or domesticated animals that are unique to particular parts of Africa. Whilst herding has likely been going on for several thousand years in some form in Africa, 2,000 years ago much of Africa still remained the domain of hunter-gatherers [Barham and Mitchell, 2008]. Our analysis provides an attempt at timing the spread of Niger-Congo speakers east from Central West Africa. Although we do not have representative forager (hunter-gatherer) groups from all parts of sub-Saharan Africa, we observe that in addition to their local region, many of the sampled groups share ancestry with Central West African groups, which is likely the result of genes spreading with the Bantu agriculturalists as their farming technology spread across Africa.

After 0CE we begin to see admixture events shared between East and West Africa, with predominantly West to East direction, which appear to be most extensive around 1,000 years ago. During this time we see evidence of admixture from West, and West Central Africa into southern Khoesan speaking groups as well as down the Eastern side of the continent. Below, we outline some of the important technical challenges in using genetic data to interpret historical events exposed by our analyes. Nonetheless, the study presented here shows that patterns of haplotype sharing in sub-Saharan African are largely determined by historical gene-flow events involving groups with ancestry from across and outside of the continent.

Interpreting haplotype similarity as historical admixture

Our analyses that rely on correlations in allele frequencies (*Widespread evidence for admixture*) provided initial evidence that the presence of Eurasian DNA across sub-Saharan Africa is the result of gene flow back into the continent within the last 10,000 years [Gurdasani et al., 2014; Pickrell et al., 2014; Hodgson et al., 2014a]. In addition, some groups have ancient (over 5 kya) shared ancestry with hunter-gather groups (Figure 3) [Gurdasani et al., 2014]. Whilst the weighted admixture LD decay curves between pairs of populations suggests that this admixture involved particular groups, for two main reasons, the interpretation of such events is difficult. Firstly, because our dataset includes closely related groups, in many populations, multiple pairs of reference populations generate significant admixture curves. Although the MALDER analysis helps us to identify which of these groups is closest to the true admixing source, it is not always possible to identify a single best matching reference, implying that sub-Saharan African

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

groups share some ancestry with many different groups. So, as confirmed with the GLOBETROTTER analysis, it is unlikely that any single contemporary population adequately represents the true admixing source. On this basis of this analysis alone, it is not possible to characterise the composition of admixture sources.

Secondly, when we infer ancient events with ALDER, such as in the Mossi from Burkina Faso, where we estimate admixture around 5,000 years ago between a Eurasian (GBR) and a Khoesan speaking group (/Gui //Gana), we know that modern haplotypes are likely to only be an approximation of ancestral diversity [Pickrell and Reich, 2014]. Even the Ju/'hoansi, a San group from southern Africa traditionally thought to have undergone limited recent admixture, has experienced gene flow from non-Khoesan groups within this timeframe (Figure 4) [Pickrell et al., 2012, 2014]. Our result hints that the Mossi share deep ancestry with Eurasian and Khoesan groups, but any description of the historical event leading to this observation is potentially biased by the discontinuity between extant populations and those present in Africa in the past. In fact, this is one motivation for grouping populations into ancestry regions and defining admixture sources in this way. In this example we can (as we do) refer to Eurasian ancestry in general moving back into Africa, rather than British DNA in particular. Using contemporary populations as proxies for ancient groups is not the perfect approach, but it is the best that we have in parts of the world from which we do not have (for the moment) DNA from significant numbers of ancient human individuals, at sufficient quality, with which to calibrate temporal changes in population genetics.

An alternative approach is to characterise admixture as occurring between sources that have mixed ancestry themselves, to account for the fact that whilst groups in the past were not intact in the same way as they are now, DNA from such groups is nevertheless likely to be present in extant groups today. Haplotype-based methods allow for this type of analysis [Hellenthal et al., 2014; Leslie et al., 2015], and have the additional benefit of potentially identifying hidden structure and relationships (Figure 2). Whilst this approach still requires one to label groups by their present-day geographic or ethno-linguistic identity, it at least allows for the construction of admixture sources containing ancestry from multiple different contemporary groups, which is likely to be closer to the truth. It also provides a framework for describing historical admixture in terms of gene flow networks, which taps into the increasingly supported notion that the genetic variation present in the vast majority of modern human groups is the result of mixture events in the past [Pickrell and Reich, 2014; Hellenthal et al., 2014]. One drawback of such an approach is that it is not always possible to assign a specific historical population label to mixed admixture sources, but it nevertheless accommodates the need to express admixture sources as containing multiple different ancestries.

Admixture in the context of infectious disease studies

Whilst inference of historical events are interesting in there own right, they also provide an important background for conducting large-scale genetic epidemiology in sub-Saharan Africa. Two studies designs are of particular relevance: genetics association studies and inference of historical natural selection. Principally, the haplotype analysis described here suggests that while recent gene-flow events mean that absolute allele frequency differences across the sub-Sahara are small relative to between Eurasian groups,

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

there remains substantial haplotype diversity. Genetic studies often rely on assumptions about the similarity of haplotypes between two groups, either as part of testing for association, or in order to impute genetic variation that is not assayed directly. Extensive recent gene-flow can help in this regard by reducing the differences between groups. However, the clear signals of admixture suggest that novel and divergent haplotypes are likely to be pocketed with subtle variation occurring both ethno-geographically, due to differences in ancestry which pre-dates recent movements, and along the genome, either by chance or by the differential effects of natural selection.

When new haplotypes are introduced into a population their fate will partly be determined by the selective advantage they confer. These selective forces could include response to changes in infectious diseases, climate, and the cultural environmental [Coop et al., 2009; Fumagalli et al., 2015]. Additionally, important causes of morbidity and ill-health in Africa are due to highly polymorphic parasites like *Plasmodium falciparum*, where moving into new environments might lead to exposure to new strains. An implication of widespread gene-flow is that it can provide a route for potentially beneficial novel mutations to enter populations allowing them to adapt to such change. Recent examples of this phenomena include the presence of altitude adaptations in Tibetans from admixture with Sherpa [Jeong et al., 2014] and Denisovans [Huerta-Sánchez et al., 2014]; higher than expected frequencies of the Duffy-null mutation in populations from Madagascar as a result of admixture with African Bantu speaking groups [Hodgson et al., 2014b]; and the observation of shared haplotype(s) in humans and Neanderthals immunity genes belonging to the Toll-like Receptor (*TLR*) gene cluster as a result of archaic admixture [Deschamps et al., 2016; Dannemann et al., 2016].

In the context of malaria, the spread of the Duffy-null allele, an ancient mutation which arose at least 30,000 years ago [Hamblin and Di Rienzo, 2000; Hamblin et al., 2002] and which confers resistance to *P. vivax* malaria, throughout Africa is only possible through contact and gene flow between populations right across the sub-Sahara. Conversely, the same sickle-cell causing mutation appears to have recently occurred five times independently in Africa, causing multiple distinct haplotypes to be observed [Hedrick, 2011]. These mutations are young, within the order of 250-1750 years old [Currat et al., 2002; Modiano et al., 2008], so will have had limited opportunity to have been moved around by the gene-flow events that we describe.

We observed that some of this gene-flow may have coincided with the spread of Bantu languages and the concomitant introduction of new agricultural practices, which can jointly introduce both new genes and novel selective pressures into existing populations, such as specific agricultural products or unknown communicable disease. As an example, our analysis also sheds light on the spread of pastoralism into southern Africa prior to the Bantu expansion [Ehret, 1967; Guldemann, 2008; Henn et al., 2008]. Several recent resequencing studies have described the the distribution of mutations in and around the lactase gene (LCT) in African populations which are associated with the ability to digest lactase in adult life, lactase persistence (LP) [Ranciaro et al., 2014; Macholdt et al., 2015; Breton et al., 2014]. The C-14010 LP-associated variant, thought to have originated in East African Niger-Congo speaking groups [Tishkoff et al., 2007], was observed in the the San and ¡Xhosa, with the haplotype in the latter matching that of east African Bantu speaking populations [Ranciaro et al., 2014]. The same LCT variant was also observed at appreciable frequencies in the Nama [Breton et al., 2014]. We infer an admixture event in

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the Ju/'hoansi involving an Afroasiatic source, which at 732CE (616-993CE), is earlier than most of the Bantu admixture we observe in other southern African populations. In the Nama we infer a very recent admixture event (c. 1800CE) involving a source containing Niger-Congo (Malawi) ancestry. Whilst the event in the Nama is unlikely to have delivered the East African mutation to this group, the event in the Ju/'hoansi is consistent with pre-Bantu interactions between east African and southern African groups. Conversely, in the AmaXhosa the Bantu source of admixture derives most of its ancestry from East Africa, so the introduction the C-14010 mutation into this part of Africa may well have been the result of the Bantu expansion.

These examples demonstrate the utility of detailed inference exploiting the complex information that is captured by large-scale genome-wide studies of genetic diversity. Africa has an exciting opportunity to be part of bringing the genetic revolution to bear in understanding and treating both infectious and non-communicable disease. Further exploration and interpretation the the rich genetic diversity of the continent will help in this endeavour.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Materials and Methods

1 Overview of the dataset

The dataset comprises a mixture of 2,504 previously published individuals from Africa and elsewhere (see below) plus novel genotypes on 1,712 sampled by the Malaria Genomic Epidemiology Network (MalariaGEN) Figure 1-Source Data 1. The MalariaGEN samples were a subset of those collected at 8 locations in Africa as part of a consortial project on genetic resistance to severe malaria: details of the study sites and investigators involved are described elsewhere [Malaria Genomic Epidemiology Network, 2014]. Samples were genotyped on the Illumina Omni 2.5M chip in order to perform a multicentre genome-wide association study (GWAS) of severe malaria: initial GWAS findings from The Gambia, Kenya and Malawi have already been reported [Band et al., 2013; Malaria Genomic Epidemiology Network, 2015] and a manuscript describing findings at all 8 locations is in preparation

The MalariaGEN samples used in the present analysis were selected to be representative of the main ethnic groups present at each of the 8 African study sites. We screened the samples collected at each study site (typically >1000 individuals) to select individuals whose reported parental ethnicity matched their own ethnicity. This process identified 23 ethnic groups for which we had samples for approximately 50 unrelated individuals or more. For ethnic groups with more than 50 samples available, we performed a cluster analysis on cohort-wide principle components, generated as part of the GWAS, with the R statistical programming language [R Development Core Team, 2011] using the *MClust* package [Fraley et al., 2012], choosing individuals from the cluster containing the largest number of individuals, to avoid any accidental inclusion of outlying individuals and to ensure that the 50 individuals chosen were, when possible, relatively genetically homogeneous. We note that in several ethnic groups (Malawi, the Kambe from Kenya, and the Mandinka and Fula from The Gambia; Fig. 1-figure supplement 2) PCA of the genotype data showed a large amount of population structure. In these cases we chose two sub-groups of individuals from a given ethnic group, selected to represent the diversity of ancestry depicted by the PCs. In several other cases, following GWAS quality control (see below), genotype data for fewer than 50 control individuals was available, and in these cases we chose as many individuals as possible, regardless of the PC-based clustering or case/control status.

We additionally included further individuals from each of four Gambian ethnic groups: the Fula, Mandinka, Jola, Wollof. The genotype data from these individuals were included as the same individuals are also being sequenced as part of the **Gambian Genome Variation Project**[1]. These subsets included ˜30 trios from each ethnic group, information on which was used in phasing (see below). Full genome data from these individuals will be made available in the future.

1.1 Quality Control

Detailed quality control (QC) for the MalariaGEN dataset was performed for a genome-wide association study of severe malaria in Africa and is outlined in detail elsewhere [Malaria Genomic Epidemiology

[1]The Gambian Genome Variation Project will sequence a number of full genomes from four Gambian ethnic groups as a basis for improving imputation for future West African specific GWAS

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Network, 2015]. Briefly, genotype calls were formed by taking a consensus across three different calling algorithms (Illuminus, Gencall in Illumina's BeadStudio software, and GenoSNP) [Band et al., 2013] and were aligned to the forward strand. Using the data from each country separately, SNPs with a minor allele frequency of <1% and missingness <5% were excluded, and additional QC to account for batch effects and SNPs not in Hardy-Weinberg equilibrium was also performed.

1.2 Combining the MalariaGEN populations with additional populations

The post-QC MalariaGEN data was combined with published data typed on the same Illumina 2.5M Omni chip from 21 populations typed for the 1000 Genomes Project (1KGP)[2], including and accounting for duos and trios, and with publicly available data from individuals from several populations from Southern Africa (Figure 1-Source Data 1) [Consortium, 2012; Schlebusch et al., 2012]. We merged samples to the forward strand, removing any ambiguous SNPs (A to T or C to G). Merging was checked by plotting allele frequencies between populations from both datasets, which should be generally correlated (data not shown). To describe population structure across sub-Saharan Africa we combined the above dataset with further publicly available samples typed on different Illumina (Omni 1M) chips, containing individuals from southern Africa [Petersen et al., 2013] and the Horn of Africa (Somalia/Ethiopia/Sudan)[Pagani et al., 2012] to generate a final dataset containing 4,216 individuals typed on 328,176 high quality common SNPs. To obtain the final set of analysis individuals we performed additional sample QC after phasing and removed American 1KG populations (Figure 1-Source Data 1).

1.3 Phasing

We used SHAPEITv2 [Delaneau et al., 2012] to generate haplotypically phased chromosomes for each individual. SHAPEITv2 conditions the underlying hidden Markov model (HMM) from Li and Stephens [2003] on all available haplotypes to quickly estimate haplotypic phase from genotype data. We split our dataset by chromosome and phased all individuals simultaneously, and used the most likely pairs of haplotypes (using the *–output-max* option) for each individual for downstream applications. We performed 30 iterations of the MCMC and used default values for all other parameters. As mentioned, we used known pedigree relationships to improve the phasing, using family data from both the 1KG and the Gambia Sequencing Projects.

1.4 Removing non-founders and cryptically related individuals

Our dataset included individuals who were known to be closely related (1KG duos and trios; Gambia Sequencing Project trios) and, because we took multiple samples from some population groups, there was also the potential to include cryptically related individuals. After phasing we therefore performed an additional step where we first removed all non-founders from the analysis and then identified individuals with high identity by descent (IBD), which is a measure of relatedness. Using an LD pruned set of SNPs generated by recursively removing SNPs with an $R^2 > 0.2$ using a 50kb sliding window, we calculated the

[2]data downloaded on 16th October 2013 from ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20120131_omni_genotypes_and_intensities/

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

proportion of loci that are IBD for each pair of individuals in the dataset using the R package *SNPRelate* [Zheng et al., 2012] and estimated kinship using the pi-hat statistic (the proportion of loci that are identical for both alleles (IBD=2) plus 0.5* the proportion of loci where one allele matches (IBD=1); i.e. PI_HAT=P(IBD=2)+0.5*P(IBD=1)). For any pair of individuals where IBD > 0.2, we randomly removed one of the individuals. 327 individuals were removed during this step.

1.5 1KG American populations and Native American Ancestry in 1KG Peruvians

With the exception of Peru, post-phasing we dropped all 1KG American populations from the analysis (97 ASW, 102 ACB, 107 CLM, 100 MXL and 111 PUR). We used a subset of the 107 Peruvian individuals that showed a large amount of putative Native American ancestry, with little apparent admixture from non-Amerindians (data not shown). Although Amerindians are not central to this study, and it is unlikely that there has been any recurrent admixture from the New World into Africa, we nevertheless generated a subset of 16 Peruvians to represent Amerindian admixture components in downstream analyses. When this subset was used, we refer to the population as PELII. The removal of these 606 American individuals left a final analysis dataset comprising 3,283 individuals from 60 different population groups (Figure 1-Source Data 1).

2 Analysis of population structure in sub-Saharan Africa

2.1 Principal Components Analysis

We performed Principal Components Analysis (PCA) using the *SNPRelate* package in R. We removed SNPs in LD by recursively removing SNPs with an $R^2 > 0.2$, using a 50kb sliding window, resulting in a subset of 162,322 SNPs.

2.2 Painting chromosomes with CHROMOPAINTER

We used fineSTRUCTURE [Lawson et al., 2012] to identify finescale population structure and to identify high level relationships between ethnic groups. The initial step of a fineSTRUCTURE analysis involves "painting" haplotypically phased chromosomes sequentially using an updated implementation of a model initially introduced by Li and Stephens [2003] and which is exploited by the CHROMOPAINTER package [Lawson et al., 2012]. The Li and Stephens copying model explicitly relates linkage disequilibrium to the underlying recombination process and CHROMOPAINTER uses an approximate method to reconstruct each "recipient" individual's haplotypic genome as a series of recombination "chunks" from a set of sample "donor" individuals. The aim of this approach is to identify, at each SNP as we move along the genome, the closest relative genome among the members of the donor sample. Because of recombination, the identity of the closest relative will change depending on the admixture history between individual genomes. Even distantly related populations share some genetic ancestry since most human genetic variation is shared [The International HapMap Consortium, 2010; Ralph and Coop, 2013], but the amount of shared ancestry can differ widely. We use the term "painting" here to refer to the application of a different label to each of the donors, such that – conceptually – each donor is represented by a different

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

colour. Donors may be coloured individually, or in groups based on *a priori* defined labels, such as the geographic population that they come from. By recovering the changing identity of the closest ancestor along chromosomes we can understand the varying contributions of different donor groups to a given population, and by understanding the distribution of these chunks we can begin to uncover the historical relationships between groups.

2.3 Using painted chromosomes with fineSTRUCTURE

We used CHROMOPAINTER with 10 Expectation-Maximisation (E-M) steps to jointly estimate the program's parameters Ne and θ, repeating this separately for chromosomes 1, 4, 10, and 15 and weight-averaging (using centimorgan sizes) the Ne and θ from the final E-M step across the four chromosomes. We performed E-M on 5 individuals from every population in the analysis and used a weighted average of the values across all pops to arrive at final values of 190.82 for Ne and 0.00045 for θ. We ran each chromosome from each population separately and combined the output to generate a final coancestry matrix to be used for fineSTRUCTURE.

As the focus of our analysis is population structure within Africa, we used a "continental force file" to combine all non-African individuals into single populations. The processing time of the algorithm is directly related to the number of individuals included in the analysis, so reducing the number of individuals speeds the analysis up. Furthermore, fineSTRUCTURE initially uses a prior that assumes that all individuals are equally distant from each other, which in the case of worldwide populations is likely to be untrue: African populations are likely to be more closely related to each other than to non-Africa populations, for example. The result is that not all of the substructure is identified in one run.

We therefore combined all individuals from each of the non-Africa 1KG populations into "continents", which has the effect of combining all of the copying vectors from the individuals within them to look like (re-weighted) normal individuals but cannot be split and do not contribute to parameter inference, and can thus be considered as copying vectors that contain the average of the individuals within them. They are then included in the algorithm at minimal extra computational cost and exist primarily to provide chunks to (and from) the remaining groups. We combined all individuals from a labelled population (e.g. all IBS individuals were now contained in a "continent" grouping called IBS), with the exception of the three Chinese population CHB, CHD, and CHS where we combined all individuals into a single CHN continent.

2.4 Using fineSTRUCTURE to inform population groupings

The fineSTRUCTURE analysis identified 154 clusters of individuals, grouped on the basis of copying-vector similarity (Fig. 1-figure supplement 3). Some ethnic groups, such as Yoruba, Mossi, Jola and Ju/'hoansi form clusters containing only individuals from their own ethnic groups. In other cases, individuals from several different ethnic groups are shared across different clusters. We see this particularly in groups from The Gambia and Kenya. These are the two countries where the most ethnic groups were sampled, seven and four, respectively, and this putative mixing could be a result of sampling several

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

different ethnic groups from a limited geographic area. The reason that we do not see the same extent of haplotype sharing in Cameroon, for example, may be because we only have two ethnic groups sampled from that area. The mixed nature of the clusters in this region could also be the results of real ancestral relationships: some of the ethnic groups in The Gambia, for example, may have history of choosing partners from outside their own ethnic group.

As mentioned above, we combined data from different sources, and in some groups (e.g. the Fula) we specifically chose different groups of individuals in an attempt to cover the broad spectrum of ancestry present in that group. We used the fineSTRUCTURE tree to visually group individuals based on their ancestry. Our aim here is to try to maximise the number of individuals that we can include within an ethnic group, without merging together individuals that are distant on the tree. We also decided *not* to use the fineSTRUCTURE clusters themselves as analytical groups because of difficulties with the interpretation of the history of such clusters. We were interested in identifying the major admixture events that have occurred in the history of different populations, and it is not clear what an analytical group that is defined as, for example, a mixture of Manjago, Mandinka and Serere individuals, would mean in our admixture analyses. In practice, this meant that we used the original geographic population labelled groups for all populations except in the Fula and Mandinka from The Gambia, where individuals fell into two distinct groups of clusters. Here we defined two clusters for each group, with the the two groups suffixed with an "I" or "II' (Figure 1-figure supplement 3).

2.5 Defining ancestry regions

We used a combination of genetic and ethno-linguistic information (see Supplementary File 1 below) to define seven ancestry regions in sub-Saharan Africa. The ancestry regions are reported in Figure 1-Source Data 1 and closely match the high level groupings we observed in the fineSTRUCTURE tree, with the following exceptions:

1. East African Niger-Congo speakers
   (a) The two ethnic groups from Cameroon – Bantu and Semi-Bantu – were included in the Central West African Niger-Congo ancestry region despite clustering more closely with East African groups from Kenya and Tanzania in Figure 1-figure supplement 3. In a preliminary fineSTRUCTURE analysis based on the MalariaGEN and 1KG populations only, using c. 1 million SNPs, the Cameroon populations clustered with other Central West African groups, and not East Africans.
   (b) Malawi was included in the South African Niger-Congo ancestry region, despite being an outlying cluster in a clade with East African Niger-Congo speaking groups. A preliminary fineSTRUCTURE analysis based on the MalariaGEN, 1KG and Schlebusch populations, clustered Malawi with the Herero and SEBantu speakers.
2. Southern Africa
   (a) We treated Southern African individuals slightly differently: even though the fineSTRUCTURE analysis did not split them into two separate clades of Khoesan and Niger-Congo

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

speaking individuals, we nevertheless did. Schlebusch et al. [2012] showed that these populations were inter-related and admixed, two properties in the data we were hoping to uncover. The final ancestry region assignments are outlined in Figure 1-Source Data 1.

2.6 Estimating pairwise F_{ST}

We used *smartpca* in the EIGENSOFT [Patterson et al., 2006] package to estimate pairwise F_{ST} between all populations. This implementation uses the Hudson estimator recently recommended by Bhatia et al. [2013]. Results are shown in Figure 2, Figure 2-Source Data 1 and Figure 2-Source Data 2.

2.7 Comparing sets of copying vectors

We used Total Variation Distance (TVD) to compare copying vectors [Leslie et al., 2015; van Dorp et al., 2015]. As the copying vectors are discrete probability distributions over the same set of donors, TVD is a natural metric for quantifying the difference between them. For a given pair of groups A and B with copying vectors describing the copying from n donors, a and b, we can compute TVD with the following equation:

$$TVD = 0.5 \times \sum_{i=1}^{n} (|a_i - b_i|)$$

3 Analysis of admixture in sub-Saharan African populations

We used a combination of approaches to explore admixture across Africa. Initially, we employed commonly used methods that utilise correlations in allele frequencies to infer historical relationships between populations. To understand ancient relationships between African groups we used the f_3 statistic [Reich et al., 2009] to look for shared drift components between a test population and two reference groups. We next used ALDER [Loh et al., 2013] and MALDER [Pickrell et al., 2014] – an updated implementation of ALDER that attempts to identify multiple admixture events – to identify admixture events through explicit modelling of admixture LD by generating weighted LD curves. The weightings of these curves are based on allele frequency differences, at varying genetic distances, between a test population and two putative admixing groups.

To identify more recent events we used two methods which aim to more fully model the mixed ancestry in a population by utilising the distribution and length of shared tracts of ancestry as identified with the CHROMOPAINTER algorithm [Lawson et al., 2012; Hellenthal et al., 2014]. We outline the details of this analysis below, but note here that, because this approach is based on the comparison and analysis of painted chromosomes, it offers a different perspective from approaches based on comparisons of allele frequencies.

3.1 Inferring admixture with the f_3 statistic and ALDER

We computed the f_3 statistic, introduced by Reich et al. [2009], as implemented in the TREEMIX package [Pickrell and Pritchard, 2012]. These tests are a 3-population generalization of F_{ST}, equal to the inner

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

product of the frequency differences between a group X and two other groups, A and B. The statistic, commonly denoted f_3(X:A,B) is proportional to the correlated genetic drift between A and X and A and B. If X is related in a simple way to the common ancestor with A and B, we expect this quantity to be positive. Significantly negative values of f_3 suggest that X has arisen as a mixture of A and B, which is thus an unambiguous signal of mixture. Standard errors are computed using a block jackknife procedure in blocks of 500 SNPs (Supplementary Table 1).

This analysis shows that sub-Saharan African populations tend to have either a hunter-gatherer or Eurasian source contributing to their most significant f_3 statistics (Supplementary Table 1). That is, there is clear evidence of admixture involving both Eurasian and hunter-gatherer groups across sub-Saharan Africa. Whilst we do not infer admixture using this statistic in the Jola, Kasem, Sudanese, Gumuz, and Ju/'hoansi, in most other groups the most significant f_3 statistic includes either the Ju/'hoansi or a 1KGP European source (GBR, CEU, FIN, or TSI). Niger-Congo speaking groups from Central West and Southern Africa tend to show most significant statistics involving the Ju/'hoansi, where as West and East African and Southern Khoesan speaking groups tended to show most significant statistics involving European sources, consistent with an recent analysis on a similar (albeit smaller) set of African populations [Gurdasani et al., 2014].

We used ALDER [Patterson et al., 2012; Loh et al., 2013] to test for the presence of admixture LD in different populations. This approach works by generating weighted admixture curves for pairs of populations and tests for admixture. As noted in Loh et al. [2013] the use of f_3 statistics and weighted LD curves are somewhat complementary, and there are several reasons why f_3 statistics might pick up signals of admixture when ALDER does not. In particular, admixture identified using f_3 statistics but not by ALDER is potentially related to more ancient events because whilst shared drift signals will still be present, admixture LD will have been broken over (potentially) millennia of recombination.

As previously shown by Loh et al. [2013] and Pickrell et al. [2014], weighted LD curves can be used to identify the source of the gene flow by comparing curves computed using different reference populations. This is possible because theory predicts that the amplitude (i.e. the y-axis intercept) of these curves becomes larger as one uses reference populations that are closer to the true mixing populations. Loh et al. [2013] demonstrated that this theory holds even when using the admixed population itself as one of the reference populations. Pickrell et al. [2014] used this concept to identify west Eurasian ancestry in a number of East African and Khoesan speaking groups from southern Africa.

We thus initially ran ALDER in "one-reference" mode, where for each focal population, we generated curves involving itself with every other reference population in turn. We used the average amplitude of the curves generated in this way to identify the groups important in describing admixture in the history of the focal group. Figure 3-figure supplement 1 shows comparative plots to those by Pickrell et al. [2014] for a selection of African populations, including the Ju/'hoansi, who we also infer to have largest curve amplitudes with Eurasian groups, consistent with that previous analysis. We summarise the results of the analysis of curve amplitudes across all populations in Figure 3-figure supplement 2, which shows the rank of the curve amplitudes for each reference population across different sub-Saharan African groups. Across much of Africa, amplitudes from non-African (European) groups are amongst the highest, indicative of admixture from Eurasian-like sources.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Next, for each focal ethnic group in turn, we used ALDER to characterise admixture using all other ethnic groups as potential reference groups (i.e. in two-population mode). In effect, this approach compares every pair of reference groups, identifying those pairs that show evidence of shared admixture LD (P <0.05 after multiple-hypothesis testing). As many of the groups are closely related, we often observed more than one pair of ethnic groups as displaying admixture in a given focal population, the results of which are highly correlated. In Supplementary Table 1, we show the evidence for admixture only for the pair of groups with the lowest P-value for each focal group. Dates for admixture events were generated using a generation time of 29 years [Fenner, 2005] and the following equation:

$$D = 1950 - (n + 1) * g$$

where D is the inferred date of admixture, n is the inferred number of generations since admixture, and g is the generation time in years.

3.2 Inferring multiple waves of admixture in African populations using weighted LD curves

We used MALDER [Pickrell et al., 2014], an implementation of ALDER designed to fit multiple exponentials to LD decay curves and therefore characterise multiple admixture events to allele frequency data. For each event we recorded (a) the curve, C, with the largest overall amplitude $C^{max}_{Pop1;Pop2}$, and (b) the curves which gave the largest amplitude where each of the two reference populations came from a different ancestry region, and for which a significant signal of admixture was inferred. To identify the source of an admixture event we compared curves involving populations from the same ancestry region as the two populations involved in generating $C^{max}_{Pop1;Pop2}$. For example, in the Jola, the population pair that gave C^{max} were the Ju/'hoansi and GBR. Substituting these populations for their ancestry regions we get $C^{max}_{Khoesan;Eurasia}$. To understand whether this event represents a specific admixture involving the Khoesan in the history of the Jola, we identified the amplitude of the curves from (b) of the form $C^{max}_{M;Eurasia}$, where M represents a population from any ancestry region other than Eurasia that gave a significant MALDER curve. We generated a Z-score for this curve comparison using the following formula [Pickrell et al., 2014]:

$$Z = \frac{C^{max}_{Khoesan;Eurasia} - C^{max}_{Khoesan;M}}{\sqrt{se(C^{max}_{Khoesan;Eurasia})^2 + se(C^{max}_{Khoesan;M})^2}}$$

The purpose of this was to determine, for a given event, whether the sources of admixture could be represented by a single ancestry region, in which case the overall $C^{max}_{ancestry1;ancestry2}$ will be significantly greater than curves involving other regions, or whether populations from multiple ancestry regions can generate admixture curves with similar amplitudes, in which case there will be a number of ancestry regions that best represent the admixing source. We combined all values of M where the Z-score computed from the above test gave a value of < 2, and define the sources of admixture in this way.

To identify the major source of admixture, we performed a similar test. We determined the regional identity of the two populations used to generate C^{max}. In the example above, these are Khoesan and Eurasia. Separately for each region, we identify the curve, C, with the maximum amplitude where either of the two reference populations was from the Khoesan region, $C^{max}_{Khoesan}$ as well as the curve where neither of the reference populations was Khoesan, $C^{max}_{notKhoesan}$. We compute a Z-score as follows:

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$Z = \frac{C^{max}_{Khoesan} - C^{max}_{notKhoesan}}{\sqrt{se(C^{max}_{Khoesan})^2 + se(C^{max}_{notKhoesan})^2}}$$

This test generates two Z-scores, in this example, one for the Khoesan/not-Khoesan comparison, and one for the Eurasia/not-Eurasia comparison. We assign the main ancestry of an event to be the region(s) that generate Z > 2. If neither region generates a Z-score > 2, then we do not assign a major ancestry to the event.

3.3 Comparisons of MALDER dating using the HAPMAP worldwide and African-specific recombination maps

Recombination maps inferred from different populations are correlated on a broad scale, but differ in the fine-scale characterisation of recombination rates [Hinch et al., 2011]. Having observed significantly older dates in many West and Central West African groups, we investigated the effect of recombination map choice by recomputing MALDER results with an African specific genetic map which was inferred through patterns of LD from the HAPMAP Yoruba (YRI) sample (Fig. 3C).

We next re-inferred admixture parameters with MALDER using all populations with the African (YRI) and additionally with a European (CEU) map [Hinch et al., 2011]. We show comparison of the dates inferred with these different maps in the main paper, and here we shows the equivalent figures to Figure 3 for events inferred using the African (Fig. 3-figure supplement 6) and European (Fig. 3-figure supplement 7) maps.

3.4 Analysis of the minimum genetic distance over which to start curve fitting when using ALDER/MALDER

A key consideration when using weighted LD to infer admixture parameters is the minimum genetic distance over which to begin computing admixture curves. Short-range LD correlations between two reference populations and a target may not only be the result of admixture, but may also be due to demography unrelated to admixture, such as shared recent bottlenecks between the target population and one of the references, or from an extended period of low population size [Loh et al., 2013]. Indeed, the authors of the ALDER algorithm specifically incorporate checks into the default ALDER analysis pipeline that define the threshold at which a test population shares short-range LD with with either of the two reference populations. Subsequent curve analyses then ignore data from pairs of SNPs at smaller distances than this correlation threshold [Loh et al., 2013].

The authors nevertheless provide the option of over-riding this LD correlation threshold, allowing the user to define the minimum genetic distance over which the algorithm will begin to compute curves and therefore infer admixture. So there are (at least) two different approaches that can be used to infer admixture using weighted LD. The first is to infer the minimum distance to start building admixture curves from the data (the default), and the second is to assume that any short-range correlations that we observe in the data result from true admixture, and prescribe a minimum distance over which to infer admixture.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3.5 All African populations share correlated LD at short genetic distances

We tested these two approaches by inferring admixture using MALDER/ALDER using a minimum distance defined by the data on the one hand, and a prescribed minimum distance of 0.5cM on the other. This value is commonly used in MALDER analyses, for example by the African Genome Variation Project [Gurdasani et al., 2014]. For each of the 48 African populations as a target, we used ALDER to infer the genetic distance over which LD correlations are shared with every other population as a reference. In Figure 3-figure supplement 3 we show the distribution of these values across all targets for each reference population.

Across all African populations we observe LD correlations with other African populations at genetic distances > 0.5cM, with median values ranging between 0.7cM when GUMUZ is used as a reference to 1.4cM when FULAII is used as a reference. In fact, when we further explore the range of these values across each region separately (Fig. 3-figure supplement 3), we note that, as expected, these distances are greater between more closely related groups.

When we plot the inverse of these distributions, that is, the range of these values for each target population separately, we see that the median minimum distance that all sub-Saharan African populations is always greater than 0.5cM (Fig. 3-figure supplement 4). Taken together, these results suggest that all African populations share some LD over short genetic distances, that may be the result of shared demography or admixture. In order to reduce the confounding effect of demography, with the exception of the next section, all ALDER/MALDER analyses presented in the paper were performed after accounting for this short range shared LD.

3.6 Results of MALDER analysis using a fixed minimum genetic distance of 0.5cM

In order to compare our MALDER analysis to previously published studies, we show the results of the MALDER analysis where we fixed the minimum genetic distance to 0.5cM (Figure 3-figure supplement 5). The main differences between this analysis and that presented in the main part of the paper are:

1. Ancient (>5ky) admixture in Central West African populations where the main analysis found no signal of admixture
2. A second ancient admixture in Malawi c.10ky
3. More of the events appear to involve Eurasian and Khoesan groups mixing.

3.7 Chromosome painting for mixture model and GLOBETROTTER

For the mixture model and GLOBETROTTER analyses, we generated painted samples where we disallowed closely related groups from being painting donors. In practice, this meant removing all populations from the same ancestry region as a given population from the painting analysis. The exception to this are populations from the Nilo-Saharan and Afroasiatic ancestry regions. In these groups, no population from either ancestry region was used as painting donors. We refer to this as the "non-local" painting analysis.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3.8 Modelling populations as mixtures of each other using linear regression

Copying vector summaries generated from painted chromosomes describe how populations relate to one another in terms of the relative time to a common shared ancestor, subsequent recent admixture, and population-specific drift [Hellenthal et al., 2014; Leslie et al., 2015]. For the following analysis, we used the GLOBETROTTER package to generate the mixing coefficients used in Figure 4.

Given a number of potential admixing donor populations, a key step in assessing the extent of admixture in a given population k is to identify which of these donors is relevant; that is, we want to identify the set T^* containing all populations $l \neq k \in [1, ..., K]$ believed to be involved in any admixture generative to population k. Using copying vectors from the non-local painting analysis, we generate an initial estimate of the mixing coefficients that describe the copying vector of population k by fitting f^k as a mixture of f^l where $l \neq k \in [1, ..., K]$. The purpose of this step is to assess the evidence for putative admixture in our populations, as described by Hellenthal et al. [2014] and Leslie et al. [2015]. In practice, we remove the self-copying (drift) element from these vectors, i.e. we set $f^k_l = 0$, and rescale each population's copying vector such that $\sum_{i=1}^{K} f^l_i = 1.0$ for all $l = k \in [1, ..., K]$.

We assume a standard linear model form for the relationship between f^k and terms f^l for $l \neq k \in [1, ..., K]$:

$$f^k = \sum_{l \neq k}^{K} \beta^k_l f^l + \epsilon$$

where ϵ is a vector of errors which we seek to choose the β terms to minimise using non negative least squares regression with the R "nnls" package. Here, β^k_l is the coefficient for f^l under the mixture model, and we estimate the β^k_ls under the constraints that all $\beta^k_l \geq 0$ and $\sum_{l \neq k}^{K} \beta^k_l = 1.0$. We refer to the estimated coefficient for the l^{th}population as $\hat{\beta}^k_l$; to avoid over-fitting we exclude all populations for which $\hat{\beta}^k_l < 0.001$ and rescale so that $\sum_{l \neq k}^{K} \hat{\beta}^k_l = 1.0$. T^* is the set of all populations whose $\hat{\beta}^k_l > 0.001$.

The $\hat{\beta}^k_l$s represent the mixing coefficients that describe a recipient population's DNA as a linear combination of the set T^* donor populations. This process identifies donor populations whose copying vectors match the copying vector of the recipient, as inferred by the painting algorithm.

3.9 Overview of GLOBETROTTER analysis pipeline

In the current setting we are interesting in identifying the general historical relationships between the different African and non-African groups in our dataset. We used GLOBETROTTER [Hellenthal et al., 2014] to characterise patterns of ancestral gene flow and admixture. Individuals tend to share longer stretches of DNA with more closely related individuals, so we used a focused approach where we disallowed copying from local populations.

GLOBETROTTER was originally described by Hellenthal et al. [2014] and a detailed description of the algorithm and the extensive validation of the method is presented in that paper. Here we run over the general framework as used in the current study, with the key difference between our approach and the default use of the algorithm being that we do not allow any groups from within the same ancestry region as a target group to be donors in the painting analysis. Throughout we use GLOBETROTTERv2.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

For a given test population k:

1. We define the set of populations present in the same broad ancestry region as k as m, with the caveats outlined below. Using CHROMOPAINTER, we generate painting samples using a reduced set of donors T^* , that included only those populations not present in the same region of Africa as k, i.e. $T^* = l \neq m \in [1, ..., K]$. The effect of this is to generate mosaic painted chromosomes whose ancestral chunks do not come from closely related individuals or groups, which can mask more subtle signals of admixture.

2. For each population, l in T^* , we generate a copying-vector, f^l, allowing all individuals from l to copy from every individual in T^*; i.e. we paint every population l with the same set of restricted donors as k in (1). For each recipient and surrogate group in turn, we sum the chunklengths donated by all individuals within all of our final donor groups (i.e. all 59 groups: including the recipient's own group and average across all recipients to generate a single 59 element copying vector for each recipient group.

3. To account for noise due to haplotype sharing among groups, we perform a non-negative-least-squares regression (mixture model; outlined above) that takes the copying vector of the recipient group as the response and the copying vectors for each donor group as the predictors. We take the coefficients of this regression, which are restricted to be ≥ 0 and to sum to 1 across donors, as our initial estimates of mixing coefficients describing the genetic make-up of the recipient population as a mixture of other sampled groups.

4. Within and between every pairing of 10 painting samples generated for each haploid of a recipient individual, we consider every pair of chunks (i.e. contiguous segements of DNA copied from a single donor haploid) separated by genetic distance g. For every two donor populations, we tabulate the number of chunk pairs where the two chunks come from the two populations. This is done in a manner to account for phasing switch errors, a common source of error when inferring haplotypes.

5. An appropriate weighting and rescaling of the curves calculated in step 4 gives us the observed coancestry curves illustrating the decay in ancestry linkage disequilibrium versus genetic distance. There is one such curve for each pair of donor populations.

6. We find the maximum likelihood estimate (MLE) of rate parameter λ of an exponential distribution fit to all coancestry curves simultaneously. Specifically, we perform a set of linear regressions that takes each curve in turn as a response and the exponential distribution with parameter λ as a predictor, finding the λ that minimizes the mean-squared residuals of these regressions. This value of λ is our estimated date of admixture. We take the coefficients from each regression. (In the case of 2 dates, we fit two independent exponential distributions with separate rate parameters to all curves simultaneously and take the MLEs of these two rate parameters as our estimates of the two respective admixture dates. We hence get two sets of coefficients, with each set representing the coefficients for one of the two exponential distributions.)

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

7. We perform an eigen decomposition of a matrix of values formed using the coefficients inferred in step 6. (In the case of 2 dates, we perform an eigen decomposition of each of the two matrices of coefficients, one for each inferred date.)

8. We use the eigen decomposition from step 7 and the copying vectors to infer both the proportion of admixture α and the mixing coefficients that describe each of the admixing source groups as a linear combination of the donor populations. (In the case of 2 dates, we perform separate fits on each of the two eigen decompositions described in step 7 to describe each admixture event separately.)

9. We re-estimate the mixing coefficients of step 3 to be $\hat{\alpha}$ times the inferred mixing coefficients of the first source plus $1 - \hat{\alpha}$ times the inferred mixing coefficients of the second source.

10. We repeat steps 5-9 for five iterations.

11. We repeat steps 4-5 using a new set of coancestry curves that should eliminate any putative signal of admixture (by taking into account the background distribution of chunks, the so-called **null procedure**), normalize our previous curves using these new ones, and repeat steps 6-10 to re-estimate dates using these normalized curves. We generate 101 date estimates via bootstrapping and assess the proportion of inferred dates that are $= 1$ or ≥ 400, setting this proportion as our empirical p-value for showing *any* evidence of admixture.

12. Using values calculated in the final iteration of step 10, we classify the admixture event into one of five categories as: (A) 'no admixture', (B) 'uncertain', (C) 'one date', (D) 'multiple dates' and (E) 'one date, multiway'.

3.10 Inferring admixture with GLOBETROTTER

We use the painting samples from (1) and the copying-vectors from (2) detailed in Section 3.9 to implement GLOBETROTTER, characterising admixture in group k; the intuition being that any admixture observed is likely to be representative of gene-flow from across larger geographic scales.

We report the results of this analysis in Figure 4-Source Data 1 as well as Figures 4, 5, and 6. We generate date estimates by simultaneously fitting an exponential curve to the coancestry curves output by GLOBETROTTER and generate confidence intervals based on 100 bootstrap replicates of the GLOBETROTTER procedure, each time bootstrapping across chromosomes. Because it is unlikely that the true admixing group is present in our set of donor groups, GLOBETROTTER infers the sources of admixture as mixtures of donor groups, which are in some sense equivalent to the β coefficients described above, but are inferred using the additional information present in the coancestry curves. We infer the composition of the admixing sources by using the βs output by GLOBETROTTER from the two (or more) sources of admixture to arrive at an understanding of the genetic basis of the the admixing source groups. These contrasts show us the contribution of each population – which we sum together into regions – to the admixture event and thus provide further intuition into historical gene flow.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3.11 Defining GLOBETROTTER admixture events

The GLOBETROTTER algorithm provides multiple metrics as evidence that admixture has taken place which are combined to arrive at an understanding of the nature of the observed admixture event. In particular, as the authors suggest, to generate an admixture P value, we ran GLOBETROTTER's "NULL" procedure, which estimates admixture parameters accounting for unusual patterns of LD, and then inferred 100 date bootstraps using this inference, identifying the proportion of inferred dates(s) that are ≤ 1 or ≥ 400.

Although the algorithm provides a "best-guess" for observed admixture event, we performed the following post-GLOBETROTTER filtering to arrive at our final characterisation of events. We outline the full GLOBETROTTER output in Figure 4-Source Data 2.

1. **Southern African populations** In all Southern African groups we present the results of the GLOBETROTTER runs where results are standardised by using the "NULL" individual see Hellenthal et al. [2014] for further details. We also note that in both the AmaXhosa and SEBantu GLOBETROTTER found evidence for two admixture events but on running the date bootstrap inference process, in both populations the most recent date confidence interval contained 1 generation, suggesting that the dating is not reliable. Inspection of the coancestry curves in this case showed that evidence for a single date of admixture.

2. **East Africa Afroasiatic speaking populations** In all Afroasiatic groups we present the results of the GLOBETROTTER runs where results are standardised by using a "NULL" individual see Hellenthal et al. [2014] for further details.

3. **West African Niger-Congo speaking populations** In all West African Niger-Congo speaking groups, with the exception of the Jola, GLOBETROTTER found evidence for two dates of admixture. In all cases the most recent event was young (1-10 generations) and the date bootstrap confidence interval often contained very small values. Inspection of the coancestry curves showed a sharp decrease at short genetic distances – consistent with the old inferred event – but there was little evidence of a more recent event based on these curves. In groups from this region we therefore show inference of a single date, which we take to be the older of the two dates inferred by GLOBETROTTER.

In all other cases we used the result output by GLOBETROTTER using the default approach.

3.12 Comparison of weighted LD curve dates with GLOBETROTTER dates

Noting that there were differences between the dates inferred by the two dating methods we employed, we compared the dates generated by ALDER/MALDER with those inferred from GLOBETROTTER. Figure 3B shows a comparison of dates using the MALDER event inference; that is, for each population, we used the MALDER inference (either one or two dates) and used the corresponding GLOBETROTTER date inference (either one or two dates) irrespective of whether GLOBETROTTER's inferred event was different to that of MALDER. Figure 3C is the opposite: we use GLOBETROTTER's event inference to

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

define whether we select one or two dates, and then use MALDER's two date inferences if two dates are inferred, or ALDER's inference if MALDER infers two dates and GLOBETROTTER infers one. Each point represents a comparison of dates for a single ethnic group, with the symbol and colour defining the ethnic group as in previous plots. In general there appears to be an inflation in dates inferred from West African groups (blues) with MALDER/ALDER compared to GLOBETROTTER.

3.13 Analysis of admixture using sets of restricted surrogates

Recall that for GLOBETROTTER analyses two painting steps are required. One needs to (a) paint target individuals with a set of "donor" individuals to generate mosaic painted chromosomes, and (b) paint all potential "surrogate" groups with the same set of painting donors, such that we then describe admixture in the target individuals with this particular set of surrogate groups. One major benefit of GLOBETROTTER is its ability to represent admixing source groups as mixtures of surrogates.

To infer admixture in sub-Saharan African groups, we painted individuals with a set of painting donors which did not include other individuals from the same ancestry region. We refer to these as "non-local" donors. We then painted individuals from *all* ethnic groups with the same reduced set of painting donors, allowing us to include all ethnic groups as surrogates in the inference process. This allows us to infer sources of admixture for which the components can potentially come from any ethnic group. For example, when inferring admixture in the Jola from the West African Niger-Congo ancestry region, we painted all Jola individuals without including individuals from any other ethnic group from the West African Niger-Congo region. We then painted all ethnic groups with the same set of non-West African Niger-Congo donors, allowing us to model admixture using information from all ethnic groups which have been painted with the same set of painting donors.

The purpose of using this approach was to "mask" the effects of including many closely-related individuals in the painting process: including such individuals in a painting analysis causes the resultant mosaic chromosomes to be dominated by chunklengths from these local groups. As mentioned, this allows us to concentrate on identifying the ancestral processes that have occurred at broader spatio-temporal scales. In effect, removing closely-related individuals from the painting step means that we identify the changing identity of the non-local donor along chromosomes and use these identities to infer historical admixture processes.

3.14 Removing non-local surrogates

In the main analysis we inferred admixture in each of the 48 target sub-Saharan African ethnic groups using all other 47 sub-Saharan African and 12 Eurasian groups as surrogates. We were interested in seeing how the admixture inference changed as we removed surrogate groups from the analysis. Masking surrogates like this provides further insight into the historical relationships between groups. By removing non-local surrogates, we can infer admixture parameters and characterize admixture sources as mixtures of this reduced set of surrogates. Given that, by definition, local groups are more closely related to the target of interest, this approach effectively asks who, outside of the targets region is next best at describing the sources of admixture.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We performed several "restricted surrogate" analyses, for different sets of targets, where we infer admixture using sub-sets of surrogates. One aim of the this analysis was to track the spread of Niger-Congo ancestry in the four Niger-Congo ancestry regions. For example, in the full analysis, the major sources of admixture in East African groups tended to be dominated by Southern African Niger-Congo (specifically Malawi) components. If we remove South African Niger-Congo groups from the admixture inference, how is the admixture source now composed?

We performed the following restricted surrogate analyses:

1. **No local region**: for all 48 African groups, we re-ran GLOBETROTTER without allowing any surrogates from the same ancestry region.

2. **No local, east or south**: for groups from the East African and South African Niger-Congo ancestry regions, we disallowed groups from both East and South African Niger-Congo regions from being surrogates. In effect, this asks where in West/Central African is their Niger-Congo ancestry likely to come from.

3. **No local or west**: For West and Central West African groups, we disallowed both West and Central West African Niger-Congo groups from being admixture surrogates. In effect, this asks where in East/South African their ancestry comes from.

4. **No local or Malawi**: As previously noted (Materials and Methods Section 2.5), Malawi was included in the South Africa Niger-Congo ancestry region. There is some evidence, for example from the fineSTRUCTURE analysis, that Malawi is closely related to the East African groups. We therefore wanted to assess whether the inference of a large amount of South African Niger-Congo ancestry in the major sources of admixture in East African Niger-Congo groups was a function of the genetic proximity of Malawi to East Africa. We removed East African Niger-Congo and Malawi as surrogates, and re-inferred admixture parameters.

3.15 Summary of inferred gene-flow from West to East and South Africa

In Figure 4-figure supplement 1 we show the changing composition of GLOBETROTTER's inferred sources of admixture. Figure 4-figure supplement 2 shows these same results with Niger-Congo surrogates only coloured, and with Malawi and Cameroon groups additionally emphasized. This analysis highlights several key insights:

1. **West and Central West Africa Niger-Congo.** Removing local surrogates from the West African Niger-Congo ancestry region does not substantially change the admixture source inference as admixture in groups from this region already involved Central West African donors. Sources of admixture in Central West African are replaced by West African Niger-Congo components in the non-local analysis. In both regions, admixture source components contain limited ancestry from East and South Africa regions. Interestingly, when we remove all West and Central West African surrogates, we observe similar major sources of admixture across groups from this region, which contains a significant amount of Nilo-Saharan / Afroasiatic ancestry (specifically Sudanese). We

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

do not have Nilo-Saharan populations from Central and West Africa, although these exist. This observation may represent gene-flow between Nilo-Saharans populations across sub-Saharan Africa.

2. **East African Niger-Congo.** The major sources of admixture inferred in groups from this region tend to contain a majority of Southern African (Malawi) ancestry (Fig. 4-figure supplement 2) whether or not local surrogates are removed from the analysis. When we remove East African Niger-Congo groups and Malawi from the analysis, other South African Niger-Congo speaking groups (SEBantu, Herero, Amaxhosa) are involved in the admixture source. Subsequent removal of these groups leads to an admixture source that is almost exclusively made up of groups from Cameroon, suggesting that Niger-Congo ancestry in East and Southern Africa is most closely related to populations from this part of Central West Africa. Intriguingly, the remainder of major admixture source is made up of a small amount of Khoesan ancestry. Whilst we do not have autochtonous groups from East Africa, this component likely represents ancestry from groups that were present in the area before the major West to East population migrations.

3. **East African Nilo-Saharan and Afroasiatic.** In the full analysis the major source of admixture tends to contain local (purple) groups (Fig. 4-figure supplement 1). When we remove local surrogates from the analysis we see an increase in West and East African donors in the major sources of admixture.

4. **South African Niger-Congo.** We similarly observe mostly East African Niger-Congo ancestry in South African Niger-Congo speakers whether or not we remove non-local surrogates, although there is also a significant Malawi component in the SEBantu and Amaxhosa in the full analysis, which is replaced by East African Niger-Congo groups in the non-local analysis. As in East African Niger-Congo speakers, when we disallow any East or South African Niger-Congo surrogates, the major sources of admixture are almost exclusively from Cameroon donors.

5. **South African Khoesan.** Several Khoesan groups show evidence of specific admixture events involving Central West African donors, in the Khwe, /Gui//Gana, and Xun. When we remove Khoesan groups from being surrogates, interestingly we see that the major sources of admixture now tend to contain ancestry from Southern African Niger-Congo populations, which specifically are not Malawi-like (Fig. 4-figure supplement 2). We also observe a small amount of West African (dark blue) ancestry in these admixture sources, and very little Central West African ancestry (and none from Cameroon). This suggests that the non-Khoesan ancestry in groups from this region is unlikely to be associated with the migrations that spread the Cameroon (Bantu) ancestry into East and South Africa. We also observe a small amount of East African Nilo-Saharan / Afroasiatic ancestry in the major sources of admixture in the non-local analysis.

A key insight from these analyses is that a large amount of the ancestry in East and Southern Africa Niger-Congo speaking groups comes from populations currently present in Central West Africa. The opposite is not true for Western and Central West African groups, who share DNA amongst themselves, but when local surrogates are removed, they instead choose mainly non-Niger-Congo speaking groups as admixture sources.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4 Plotting date densities

For each admixture event we split the admixture sources into their constituent components (i.e. we used the β coefficients inferred by GLOBETROTTER) at the appropriate admixture proportions. For a given event, these components sum to 1. We multiplied these components by 100 to estimate the percentage of ancestry from a given event that originates from each donor group. We then assigned each of the components the set of date bootstraps associated with the event. For example, in the Kauma we infer an admixture event with an admixture proportion α of 6% involving a minor source containing the following coefficients: Massai 0.02 Afar 0.15 GBR 0.26 GIH 0.57. We multiply each of these coefficients by α to obtain a final proportion that each group gives to the admixture event: Massai 0.5 Afar 1 GBR 1.5 GIH 3. We assign all the inferred date bootstraps for the Kauma to each of the populations in these proportions. In this example, GIH has twice the density of GBR. We then additionally sum components across the same country to finally arrive at the density plots in Figures 5.

5 Gene-flow maps

We generated maps with the *rworldmaps* package in R. To generate arrows, we combined the inferred ancestral components (i.e. 1ST and 2ND EVENT SOURCES in Figure 3) for each population and estimated the proportion of a group's ancestry coming from each component, summed across all surrogates from a particular country. For example, if an admixture contained source contains components from both the Jola and Wollof (both from The Gambia), then these components were added together. As such, the arrows point from the country of component origin to the country of the recipient. We then plot only those arrows which relate to events pertaining to the different broad gene-flow events. For each map, we used plot arrows for any event involving the following:

(A) **Recent Western Bantu gene-flow**: any admixture source which has a component from either of the two Cameroon ethnic groups, Bantu and Semi-Bantu.

(B) **Eastern Bantu gene-flow**: any admixture source which has a component from Kenya, Tanzania, Malawi, or South Africa (Niger-Congo speakers).

(C) **East / West gene-flow**: any admixture event which has a component from Gambia, Burkina Faso, Ghana, Mali, Nigeria, Ethiopia, Sudan or Somalia.

(D) **Eurasian gene-flow into Africa**: any admixture event which has a component from any Eurasian population.

An alternative map stratified by time window, rather than admixture component is shown in Figure 6-figure supplement 1.

6 Analysis and plotting code

Code used for analyses and plotting will be made available at `https://github.com/georgebusby/popgen`

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

We thank all the MalariaGEN study sites that contributed samples to this analysis: a list of researchers involved at each study site can be found at `https://www.malariagen.net/projects/host/consortium-members`.

MalariaGEN is funded by the Wellcome Trust (WT077383/Z/05/Z, 090770/Z/09/Z) and the Bill and Melinda Gates Foundation through the Foundation for the National Institutes of Health (566). Genotyping was performed at the Wellcome Trust Sanger Institute, partly funded by its core award from the Wellcome Trust (098051/Z/05/Z). This research was also supported by Centre grants from the Wellcome Trust (090532/Z/09/Z) and the Medical Research Council (G0600718). C.C.A.S. was supported by a Wellcome Trust Career Development Fellowship (097364/Z/11/Z).

The Malaria Research and Training Center–Bandiagara Malaria Project (MRTC-BMP) in Mali group is supported by an Interagency Committee on Disability Research (ICDR) grant from the National Institute of Allergy and Infectious Diseases/US National Institutes of Health (NIAID/NIH) to the University of Maryland and the University of Bamako (USTTB) and by the Mali-NIAID/NIH International Centers for Excellence in Research (ICER) at USTTB. The Kenya Medical Research Institute (KEMRI)–Wellcome Trust Programme is funded through core support from the Wellcome Trust. This paper is published with the permission of the director of KEMRI. C.M.N. is supported through a strategic award to the KEMRI–Wellcome Trust Programme from the Wellcome Trust (084538). The Joint Malaria Programme, Kilimanjaro Christian Medical Centre in Tanzania received funding from a UK MRC grant (G9901439).

We thank Clare Bycroft and Lucy van Dorp for critically evaluating the manuscript and Francesco Montinaro for insightful discussions on interpretation of MALDER analyses. Genotype data for the MalariaGen samples included in this paper will be made available at the European Nucleotide Archive (accession number TBC).

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Allen, J. D. V. (1993). *Swahili Origins: Swahili Culture & the Shungwaya Phenomenon*. James Currey Publishers.

Ansari Pour, N., Plaster, C. A., and Bradman, N. (2013). Evidence from Y-chromosome analysis for a late exclusively eastern expansion of the Bantu-speaking people. *European Journal of Human Genetics*, 21(4):423–429.

Band, G., Le, Q. S., Jostins, L., Pirinen, M., Kivinen, K., Jallow, M., Sisay-Joof, F., Bojang, K., Pinder, M., Sirugo, G., Conway, D. J., Nyirongo, V., Kachala, D., Molyneux, M., Taylor, T., Ndila, C., Peshu, N., Marsh, K., Williams, T. N., Alcock, D., Andrews, R., Edkins, S., Gray, E., Hubbart, C., Jeffreys, A., Rowlands, K., Schuldt, K., Clark, T. G., Small, K. S., Teo, Y. Y., Kwiatkowski, D. P., Rockett, K. A., Barrett, J. C., Spencer, C. C. A., and Malaria Genomic Epidemiological Network ¶ (2013). Imputation-Based Meta-Analysis of Severe Malaria in Three African Populations. *PLoS Genet*, 9(5):e1003509.

Barham, L. and Mitchell, P. (2008). *The First Africans: African Archaeology from the Earliest Toolmakers to Most Recent Foragers*. Cambridge University Press, Cambridge ; New York, 1 edition edition.

Behar, D. M., Yunusbayev, B., Metspalu, M., Metspalu, E., Rosset, S., Parik, J., Rootsi, S., Chaubey, G., Kutuev, I., Yudkovsky, G., and others (2010). The genome-wide structure of the Jewish people. *Nature*, 466(7303):238–242.

Beleza, S., Gusmão, L., Amorim, A., Carracedo, A., and Salas, A. (2005). The genetic legacy of western Bantu migrations. *Human Genetics*, 117(4):366–375.

Berniell-Lee, G., Calafell, F., Bosch, E., Heyer, E., Sica, L., Mouguiama-Daouda, P., van der Veen, L., Hombert, J.-M., Quintana-Murci, L., and Comas, D. (2009). Genetic and Demographic Implications of the Bantu Expansion: Insights from Human Paternal Lineages. *Molecular Biology and Evolution*, 26(7):1581–1589.

Bhatia, G., Patterson, N., Sankararaman, S., and Price, A. L. (2013). Estimating and interpreting FST: The impact of rare variants. *Genome Research*, 23(9):1514–1521.

Breton, G., Schlebusch, C. M., Lombard, M., Sjödin, P., Soodyall, H., and Jakobsson, M. (2014). Lactase Persistence Alleles Reveal Partial East African Ancestry of Southern African Khoe Pastoralists. *Current Biology*, 24(8):852–858.

Bryc, K., Auton, A., Nelson, M. R., Oksenberg, J. R., Hauser, S. L., Williams, S., Froment, A., Bodo, J.-M., Wambebe, C., Tishkoff, S. A., and Bustamante, C. D. (2010). Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proceedings of the National Academy of Sciences*, 107(2):786–791.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Busby, G. B. J., Hellenthal, G., Montinaro, F., Tofanelli, S., Bulayeva, K., Rudan, I., Zemunik, T., Hayward, C., Toncheva, D., Karachanak-Yankova, S., Nesheva, D., Anagnostou, P., Cali, F., Brisighelli, F., Romano, V., Lefranc, G., Buresi, C., Ben Chibani, J., Haj-Khelil, A., Denden, S., Ploski, R., Krajewski, P., Hervig, T., Moen, T., Herrera, R. J., Wilson, J. F., Myers, S., and Capelli, C. (2015). The Role of Recent Admixture in Forming the Contemporary West Eurasian Genomic Landscape. *Current Biology*, 25(19):2518–2526.

Campbell, M. C. and Tishkoff, S. A. (2008). African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annual Review of Genomics and Human Genetics*, 9(1):403–433.

Consortium, T. . G. P. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491(7422):56–65.

Coop, G., Pickrell, J., Novembre, J., Kudaravalli, S., Li, J., Absher, D., Myers, R., Cavalli-Sforza, L., Feldman, M., and Pritchard, J. (2009). The role of geography in human adaptation. *PLoS Genetics*, 5(6).

Currat, M., Trabuchet, G., Rees, D., Perrin, P., Harding, R. M., Clegg, J. B., Langaney, A., and Excoffier, L. (2002). Molecular Analysis of the β-Globin Gene Cluster in the Niokholo Mandenka Population Reveals a Recent Origin of the βS Senegal Mutation. *The American Journal of Human Genetics*, 70(1):207–223.

Currie, T. E., Meade, A., Guillon, M., and Mace, R. (2013). Cultural phylogeography of the Bantu Languages of sub-Saharan Africa. *Proceedings of the Royal Society of London B: Biological Sciences*, 280(1762):20130695.

Dannemann, M., Andrés, A. M., and Kelso, J. (2016). Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. *The American Journal of Human Genetics*, 98(1):22–33.

de Filippo, C., Bostoen, K., Stoneking, M., and Pakendorf, B. (2012). Bringing together linguistic and genetic evidence to test the Bantu expansion. *Proceedings of the Royal Society B: Biological Sciences*, 279(1741):3256–3263.

Delaneau, O., Marchini, J., and Zagury, J.-F. (2012). A linear complexity phasing method for thousands of genomes. *Nature Methods*, 9(2):179–181.

Deschamps, M., Laval, G., Fagny, M., Itan, Y., Abel, L., Casanova, J.-L., Patin, E., and Quintana-Murci, L. (2016). Genomic Signatures of Selective Pressures and Introgression from Archaic Hominins at Human Innate Immunity Genes. *The American Journal of Human Genetics*, 98(1):5–21.

Diamond, J. and Bellwood, P. (2003). Farmers and their languages: The first expansions. *Science*, 300(5619):597–603.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ehret, C. (1967). Cattle-Keeping and Milking in Eastern and Southern African History: The Linguistic Evidence. *The Journal of African History*, 8(1):1–17.

Fenner, J. (2005). Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *American Journal of Physical Anthropology*, 128(2):415–423.

Fraley, C., Raftery, A. E., Murphy, T. B., and Scrucca, L. (2012). *mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation*. Number 597. Department of Statistics, University of Washington.

Fumagalli, M., Moltke, I., Grarup, N., Racimo, F., Bjerregaard, P., Jørgensen, M. E., Korneliussen, T. S., Gerbault, P., Skotte, L., Linneberg, A., Christensen, C., Brandslund, I., Jørgensen, T., Huerta-Sánchez, E., Schmidt, E. B., Pedersen, O., Hansen, T., Albrechtsen, A., and Nielsen, R. (2015). Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science*, 349(6254):1343–1347.

González-Santos, M., Montinaro, F., Oosthuizen, O., Oosthuizen, E., Busby, G. B. J., Anagnostou, P., Destro-Bisol, G., Pascali, V., and Capelli, C. (2015). Genome-Wide SNP Analysis of Southern African Populations Provides New Insights into the Dispersal of Bantu-Speaking Groups. *Genome Biology and Evolution*, 7(9):2560–2568.

Greenberg, J. H. (1972). Linguistic evidence regarding Bantu origins. *The Journal of African History*, 13(02):189–216.

Grollemund, R., Branford, S., Bostoen, K., Meade, A., Venditti, C., and Pagel, M. (2015). Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences*, 112(43):13296–13301.

Guldemann, T. (2008). A linguist's view : Khoe-Kwadi speakers as the earliest food-producers of southern Africa.

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M. O., Choudhury, A., Ritchie, G. R. S., Xue, Y., Asimit, J., Nsubuga, R. N., Young, E. H., Pomilla, C., Kivinen, K., Rockett, K., Kamali, A., Doumatey, A. P., Asiki, G., Seeley, J., Sisay-Joof, F., Jallow, M., Tollman, S., Mekonnen, E., Ekong, R., Oljira, T., Bradman, N., Bojang, K., Ramsay, M., Adeyemo, A., Bekele, E., Motala, A., Norris, S. A., Pirie, F., Kaleebu, P., Kwiatkowski, D., Tyler-Smith, C., Rotimi, C., Zeggini, E., and Sandhu, M. S. (2014). The African Genome Variation Project shapes medical genetics in Africa. *Nature*, advance online publication.

H3Africa Consortium (2014). Research capacity. Enabling the genomic revolution in Africa. *Science (New York, N.Y.)*, 344(6190):1346–1348.

Hamblin, M. T. and Di Rienzo, A. (2000). Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *American Journal of Human Genetics*, 66(5):1669–1679.

Hamblin, M. T., Thompson, E. E., and Di Rienzo, A. (2002). Complex signatures of natural selection at the duffy blood group locus. *American Journal of Human Genetics*, 70(2):369–383.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hedrick, P. W. (2011). Population genetics of malaria resistance in humans. *Heredity*, 107(4):283–304.

Hellenthal, G., Busby, G. B. J., Band, G., Wilson, J. F., Capelli, C., Falush, D., and Myers, S. (2014). A Genetic Atlas of Human Admixture History. *Science*, 343(6172):747–751.

Henn, B. M., Botigué, L. R., Gravel, S., Wang, W., Brisbin, A., Byrnes, J. K., Fadhlaoui-Zid, K., Zalloua, P. A., Moreno-Estrada, A., Bertranpetit, J., Bustamante, C. D., and Comas, D. (2012). Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. *PLoS Genet*, 8(1):e1002397.

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. *Proceedings of the National Academy of Sciences*, 105(31):10693.

Hinch, A. G., Tandon, A., Patterson, N., Song, Y., Rohland, N., Palmer, C. D., Chen, G. K., Wang, K., Buxbaum, S. G., Akylbekova, E. L., Aldrich, M. C., Ambrosone, C. B., Amos, C., Bandera, E. V., Berndt, S. I., Bernstein, L., Blot, W. J., Bock, C. H., Boerwinkle, E., Cai, Q., Caporaso, N., Casey, G., Adrienne Cupples, L., Deming, S. L., Ryan Diver, W., Divers, J., Fornage, M., Gillanders, E. M., Glessner, J., Harris, C. C., Hu, J. J., Ingles, S. A., Isaacs, W., John, E. M., Linda Kao, W. H., Keating, B., Kittles, R. A., Kolonel, L. N., Larkin, E., Le Marchand, L., McNeill, L. H., Millikan, R. C., Murphy, Musani, S., Neslund-Dudas, C., Nyante, S., Papanicolaou, G. J., Press, M. F., Psaty, B. M., Reiner, A. P., Rich, S. S., Rodriguez-Gil, J. L., Rotter, J. I., Rybicki, B. A., Schwartz, A. G., Signorello, L. B., Spitz, M., Strom, S. S., Thun, M. J., Tucker, M. A., Wang, Z., Wiencke, J. K., Witte, J. S., Wrensch, M., Wu, X., Yamamura, Y., Zanetti, K. A., Zheng, W., Ziegler, R. G., Zhu, X., Redline, S., Hirschhorn, J. N., Henderson, B. E., Taylor Jr, H. A., Price, A. L., Hakonarson, H., Chanock, S. J., Haiman, C. A., Wilson, J. G., Reich, D., and Myers, S. R. (2011). The landscape of recombination in African Americans. *Nature*, 476(7359):170–175.

Hodgson, J. A., Mulligan, C. J., Al-Meeri, A., and Raaum, R. L. (2014a). Early Back-to-Africa Migration into the Horn of Africa. *PLoS Genet*, 10(6):e1004393.

Hodgson, J. A., Pickrell, J. K., Pearson, L. N., Quillen, E. E., Prista, A., Rocha, J., Soodyall, H., Shriver, M. D., and Perry, G. H. (2014b). Natural selection for the Duffy-null allele in the recently admixed people of Madagascar. *Proceedings of the Royal Society B: Biological Sciences*, 281(1789):20140930.

Holden, C. J. (2002). Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. *Proceedings of the Royal Society of London B: Biological Sciences*, 269(1493):793–799.

Hudson, R., Slatkin, M., and Maddison, W. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics*, 132(2):583–589.

Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., Ni, P., Wang, B., Ou, X., Huasang, Luosang, J., Cuo, Z. X. P., Li, K., Gao, G., Yin, Y., Wang, W., Zhang, X., Xu, X., Yang, H., Li, Y., Wang, J., Wang, J., and Nielsen, R. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, 512(7513):194–197.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Jeong, C., Alkorta-Aranburu, G., Basnyat, B., Neupane, M., Witonsky, D. B., Pritchard, J. K., Beall, C. M., and Di Rienzo, A. (2014). Admixture facilitates genetic adaptations to high altitude in Tibet. *Nature Communications*, 5.

Lawson, D. J., Hellenthal, G., Myers, S., and Falush, D. (2012). Inference of Population Structure using Dense Haplotype Data. *PLoS Genetics*, 8(1):e1002453.

Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T., Hutnik, K., Royrvik, E. C., Cunliffe, B., Wellcome Trust Case Control Consortium 2, International Multiple Sclerosis Genetics Consortium, Lawson, D. J., Falush, D., Freeman, C., Pirinen, M., Myers, S., Robinson, M., Donnelly, P., and Bodmer, W. (2015). The fine-scale genetic structure of the British population. *Nature*, 519(7543):309–314.

Li, N. and Stephens, M. (2003). Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics*, 165(4):2213–2233.

Li, S., Schlebusch, C., and Jakobsson, M. (2014). Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples. *Proceedings of the Royal Society B: Biological Sciences*, 281(1793):20141448.

Llorente, M. G., Jones, E. R., Eriksson, A., Siska, V., Arthur, K. W., Arthur, J. W., Curtis, M. C., Stock, J. T., Coltorti, M., Pieruccini, P., Stretton, S., Brock, F., Higham, T., Park, Y., Hofreiter, M., Bradley, D. G., Bhak, J., Pinhasi, R., and Manica, A. (2015). Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science*, page aad2879.

Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J. K., Reich, D., and Berger, B. (2013). Inferring Admixture Histories of Human Populations Using Linkage Disequilibrium. *Genetics*, 193(4):1233–1254.

Macholdt, E., Slatkin, M., Pakendorf, B., and Stoneking, M. (2015). New insights into the history of the C-14010 lactase persistence variant in Eastern and Southern Africa. *American Journal of Physical Anthropology*, 156(4):661–664.

Malaria Genomic Epidemiology Network (2008). A global network for investigating the genomic epidemiology of malaria. *Nature*, 456(7223):732–737.

Malaria Genomic Epidemiology Network (2014). Reappraisal of known malaria resistance loci in a large multicenter study. *Nature Genetics*, 46(11):1197–1204.

Malaria Genomic Epidemiology Network (2015). A novel locus of resistance to severe malaria in a region of ancient balancing selection. *Nature*, 526(7572):253–257.

Marks, S. J., Montinaro, F., Levy, H., Brisighelli, F., Ferri, G., Bertoncini, S., Batini, C., Busby, G. B., Arthur, C., Mitchell, P., Stewart, B. A., Oosthuizen, O., Oosthuizen, E., D'Amato, M. E., Davison, S., Pascali, V., and Capelli, C. (2014). Static and moving frontiers: the genetic landscape of Southern African Bantu-speaking populations. *Molecular Biology and Evolution*, page msu263.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mitchell, P. (2002). *The archaeology of Southern Africa*. Cambridge University Press, Cambridge, UK.

Modiano, D., Bancone, G., Ciminelli, B. M., Pompei, F., Blot, I., Simporé, J., and Modiano, G. (2008). Haemoglobin S and haemoglobin C: 'quick but costly' versus 'slow but gratis' genetic adaptations to Plasmodium falciparum malaria. *Human Molecular Genetics*, 17(6):789–799.

Montinaro, F., Busby, G. B. J., Pascali, V. L., Myers, S., Hellenthal, G., and Capelli, C. (2015). Unravelling the hidden ancestry of American admixed populations. *Nature Communications*, 6.

Need, A. C. and Goldstein, D. B. (2009). Next generation disparities in human genomics: concerns and remedies. *Trends in Genetics*, 25(11):489–494.

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A., Auton, A., Indap, A., King, K., Bergmann, S., Nelson, M., Stephens, M., and Bustamante, C. (2008). Genes mirror geography within Europe. *Nature*, 456(7218):98–101.

Nurse, D. and Philippson, G. (2003). *The Bantu Languages*. Number 4 in Routledge Language Family Series. Routledge, London, UK.

Oliver, R. A. and Fagan, B. M. (1975). *Africa in the Iron Age: C.500 BC-1400 AD*. Cambridge University Press.

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., and 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.

Pakendorf, B., Bostoen, K., and de Filippo, C. (2011). Molecular Perspectives on the Bantu Expansion: A Synthesis. *Language Dynamics and Change*, 1(1):50–88.

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(3):1065–1093.

Patterson, N., Price, A. L., and Reich, D. (2006). Population Structure and Eigenanalysis. *PLoS Genetics*, 2(12):e190.

Petersen, D. C., Libiger, O., Tindall, E. A., Hardie, R.-A., Hannick, L. I., Glashoff, R. H., Mukerji, M., Fernandez, P., Haacke, W., Schork, N. J., Hayes, V. M., and Indian Genome Variation Consortium (2013). Complex Patterns of Genomic Admixture within Southern Africa. *PLoS Genet*, 9(3):e1003309.

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., Reich, D., and Pakendorf, B. (2012). The genetic prehistory of southern Africa. *Nature Communications*, 3:1143.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Pickrell, J. K., Patterson, N., Loh, P.-R., Lipson, M., Berger, B., Stoneking, M., Pakendorf, B., and 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.

Pickrell, J. K. and Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics*, 8(11):e1002967.

Pickrell, J. K. and Reich, D. (2014). Toward a new history and geography of human genes informed by ancient DNA. *Trends in Genetics*, 30(9):377–389.

R Development Core Team (2011). R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.

Ralph, P. and Coop, G. (2013). The Geography of Recent Genetic Ancestry across Europe. *PLoS Biol*, 11(5):e1001555.

Ranciaro, A., Campbell, M. C., Hirbo, J. B., Ko, W.-Y., Froment, A., Anagnostou, P., Kotze, M. J., Ibrahim, M., Nyambo, T., Omar, S. A., and Tishkoff, S. A. (2014). Genetic Origins of Lactase Persistence and the Spread of Pastoralism in Africa. *The American Journal of Human Genetics*, 94(4):496–510.

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., and Singh, L. (2009). Reconstructing Indian population history. *Nature*, 461(7263):489–494.

Roberts, J. (2007). *The New Penguin History of the World*. Penguin Books, London, UK, 5th edition.

Schlebusch, C. M., Skoglund, P., Sjödin, P., Gattepaille, L. M., Hernandez, D., Jay, F., Li, S., Jongh, M. D., Singleton, A., Blum, M. G. B., Soodyall, H., and Jakobsson, M. (2012). Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History. *Science*, 338(6105):374–379.

Smith, A. B. (2005). *African Herders: Emergence of Pastoral Traditions*. Rowman Altamira.

The International HapMap Consortium (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, 467(7311):52–58.

Thompson, L. (2001). *A history of South Africa*. Yale University Press, USA, 3rd edition.

Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J. B., Awomoyi, A. A., Bodo, J.-M., Doumbo, O., Ibrahim, M., Juma, A. T., Kotze, M. J., Lema, G., Moore, J. H., Mortensen, H., Nyambo, T. B., Omar, S. A., Powell, K., Pretorius, G. S., Smith, M. W., Thera, M. A., Wambebe, C., Weber, J. L., and Williams, S. M. (2009). The Genetic Structure and History of Africans and African Americans. *Science*, 324(5930):1035–1044.

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., Ghori,

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

J., Bumpstead, S., Pritchard, J. K., Wray, G. A., and Deloukas, P. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics*, 39(1):31–40.

van Dorp, L., Balding, D., Myers, S., Pagani, L., Tyler-Smith, C., Bekele, E., Tarekegn, A., Thomas, M. G., Bradman, N., and Hellenthal, G. (2015). Evidence for a Common Origin of Blacksmiths and Cultivators in the Ethiopian Ari within the Last 4500 Years: Lessons for Clustering-Based Inference. *PLoS Genet*, 11(8):e1005397.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., and Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research*, 42(Database issue):D1001–D1006.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28(24):3326–3328.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figures and Figure Supplements

Figure 1: Sub-Saharan African genetic variation is shaped by ethno-linguistic and geographical similarity. . . . 5
Figure 2: Haplotypes capture more population structure than independent loci 7
Figure 3: Inference of admixture in sub-Saharan Africa using MALDER . . . 10
Figure 4: Inference of admixture in sub-Saharan African using GLOBETROTTER . . . 13
Figure 5: A timeline of recent admixture in sub-Saharan Africa. . . . 15
Figure 6: The geography of recent gene-flow in Africa . . . 19
Figure 1-figure supplement 1: Map of populations used in the analysis. . . . 54
Figure 1-figure supplement 2: An example of hierarchical clustering to chose two groups of similar individuals from the Fula based on a PCA of the Gambia. . . . 55
Figure 1-figure supplement 3: fineSTRUCTURE analysis of the full dataset . . . 56
Figure 3-figure supplement 1: Weighted LD amplitudes for a selection of 9 ethnic groups . . . 57
Figure 3-figure supplement 2: Comparison of weighted LD amplitude scores across all African ethnic groups . . . 58
Figure 3-figure supplement 3: Comparison of the minimum distance to begin computing admixture LD . . . 59
Figure 3-figure supplement 4: Comparison of the minimum distance to begin computing admixture LD split by region . . . 60
Figure 3-figure supplement 5: Results of the MALDER analysis computing weighted admixture decay curves from 0.5cM . . . 61
Figure 3-figure supplement 6: Results of MALDER for all populations using an African specific recombination map . . . 62
Figure 3-figure supplement 7: Results of MALDER for all populations using a European specific recombination map . . . 63
Figure 4-figure supplement 1: Admixture source inference by GLOBETROTTER after sequentially removing local surrogates from the analysis . . . 64
Figure 4-figure supplement 2: Admixture source inference by GLOBETROTTER after sequentially removing local surrogates from the analysis . . . 65
Figure 6-figure supplement 1: Gene-flow in Africa over the last 2,000 years. . . . 66

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Files

Supplementary File 1 A note on ethnolinguistic groupings 67

Source Data Tables

Figure 1-Source Data 1 Overview of sampled populations 69
Figure 2-Source Data 1 Pairwise F_{ST} for African populations 70
Figure 2-Source Data 2 Pairwise F_{ST} for Eurasian populations 71
Figure 2-Source Data 3 Pairwise TVD for African populations 72
Figure 2-Source Data 4 Pairwise TVD for Eurasian populations 73
Figure 3-Source Data 1 The evidence for multiple waves of admixture in African populations using MALDER and the HAPMAP recombination map. 74
Figure 3-Source Data 2 The evidence for multiple waves of admixture in African populations using MALDER and the African recombination map. 75
Figure 3-Source Data 3 The evidence for multiple waves of admixture in African populations using MALDER and the European recombination map. 76
Figure 3-Source Data 4 The evidence for multiple waves of admixture in African populations using MALDER and the HAPMAP recombination map and a mindis of 0.5cM. 77
Figure 4-Source Data 1 Results of the main GLOBETROTTER analysis. 79
Figure 4-Source Data 2 Results of the main GLOBETROTTER analysis. 81
Supplementary Table 1 Evidence for admixture across the ethnic groups included in the study using f_3 tests and ALDER. 85

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Supplements

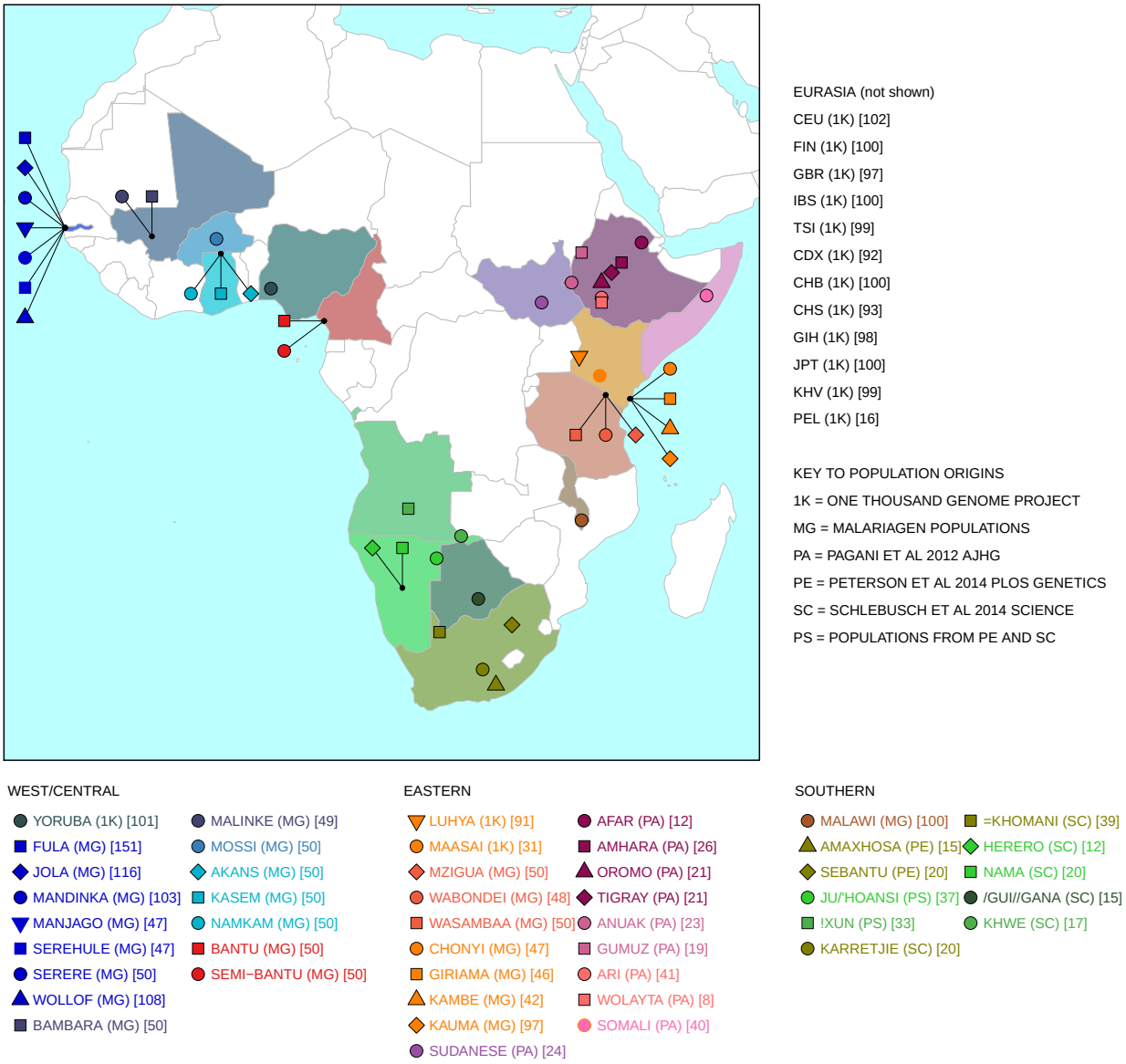

Figure 1-figure supplement 1. Map of populations used in the analysis. Population names are coloured by the country of origin; positions of the countries are shown on the map. Individual point labels, which are used throughout this paper, are shown for each population in the legend. Sample provenance is shown immediately after the population name in circular parentheses and final number of individuals is shown in square parentheses.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

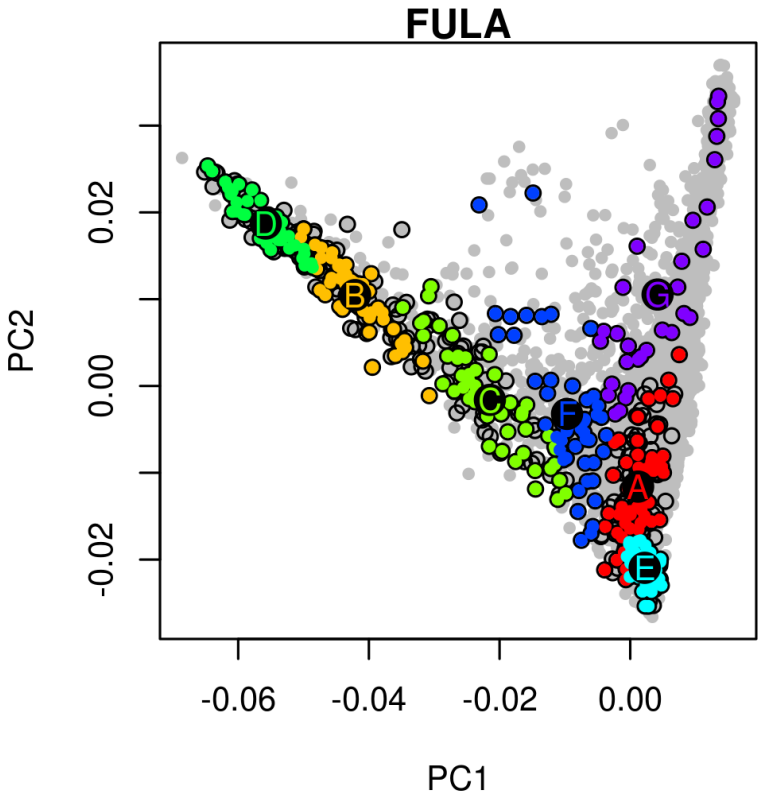

Figure 1-figure supplement 2. An example of hierarchical clustering to chose two groups of similar individuals from the Fula based on a PCA of the Gambia. Projected onto a PCA of Gambian genetic variation where each point represents an individual, all Fula individuals are coloured, with the colour depicting their cluster assignment, based on the *MClust* clustering algorithm. We chose individuals from the green (D) and light blue (E) clusters to maximise the representation of Fula genetic variation. Note that the majority of the individuals from the other 6 Gambian ethnic groups occur in the right arm of the PCA. An analogous process was preformed for all ethnic groups from the MalariaGEN dataset where more than 50 individuals were available.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

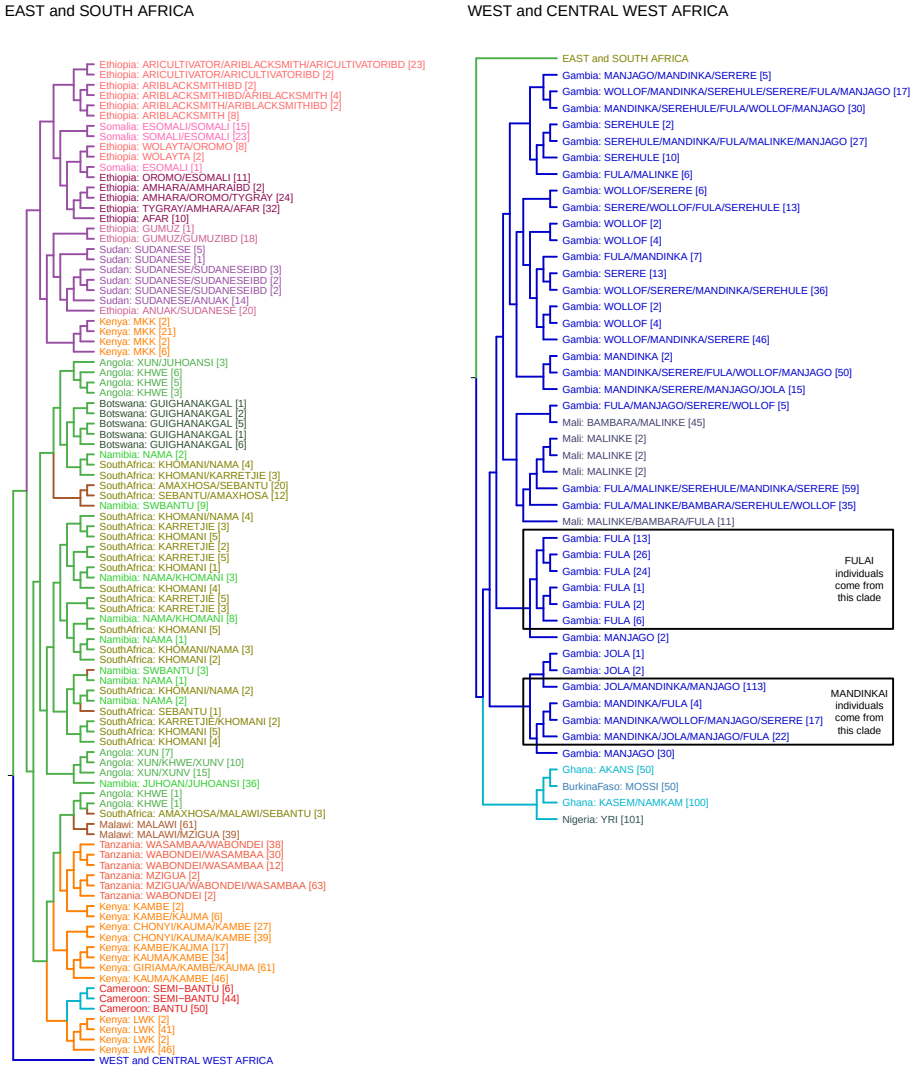

Figure 1-figure supplement 3. fineSTRUCTURE analysis of the full dataset. We show the tree output from a single run of the fineSTRUCTURE algorithm. To aid reading, the tree has been split in two, East and Southern African groups are on the left, West and Central West African groups are on the right. Leaves are labelled by the identity of the individuals within them, with the total number of individuals in the clusters shown in parentheses. Leaves are coloured by the country of origin (as in Fig. 1-figure supplement 1) and branches are coloured by the final ancestry region that the clusters were assigned to. Note that although Malawi and Cameroon individuals were located in a clade with mostly East African individuals, they were assigned to Southern and Central West African ancestry regions, respectively. Clades containing outlying individuals from the Fula and Mandinka are also shown.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

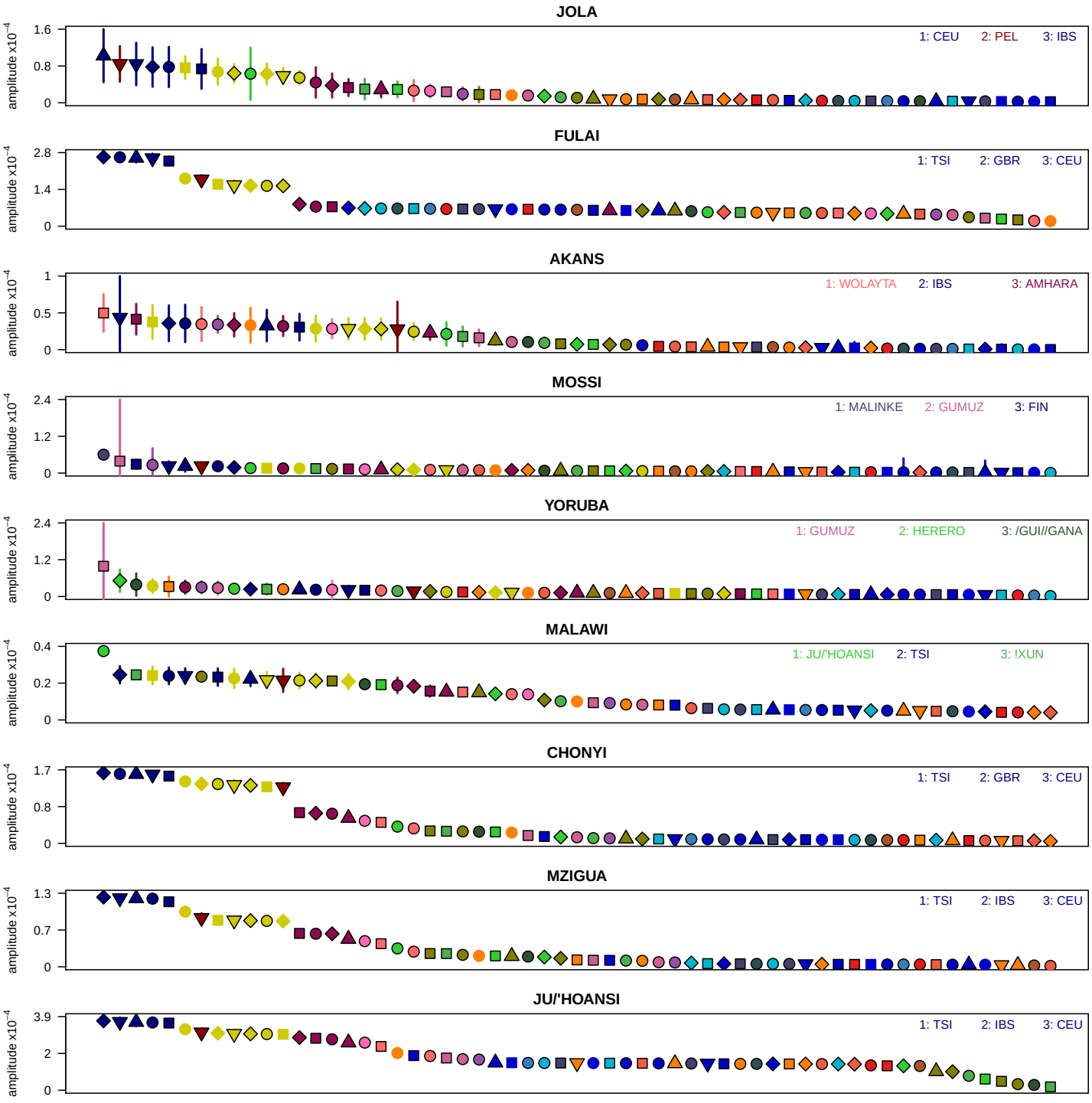

Figure 3-figure supplement 1. Weighted LD amplitudes for a selection of 9 ethnic groups. For a given test population we show the amplitude ($\pm$ 1 s.e.) computed using a test population and every other population as the second reference. Plotted are the fitted amplitudes for each set of curves with the population used labelled beneath, with populations ordered by amplitude. A large number of population showed a similar profile to (A), that is with Eurasian populations showing the highest amplitudes. Other populations, e.g. Malawi, obtained the largest amplitudes from an African population.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

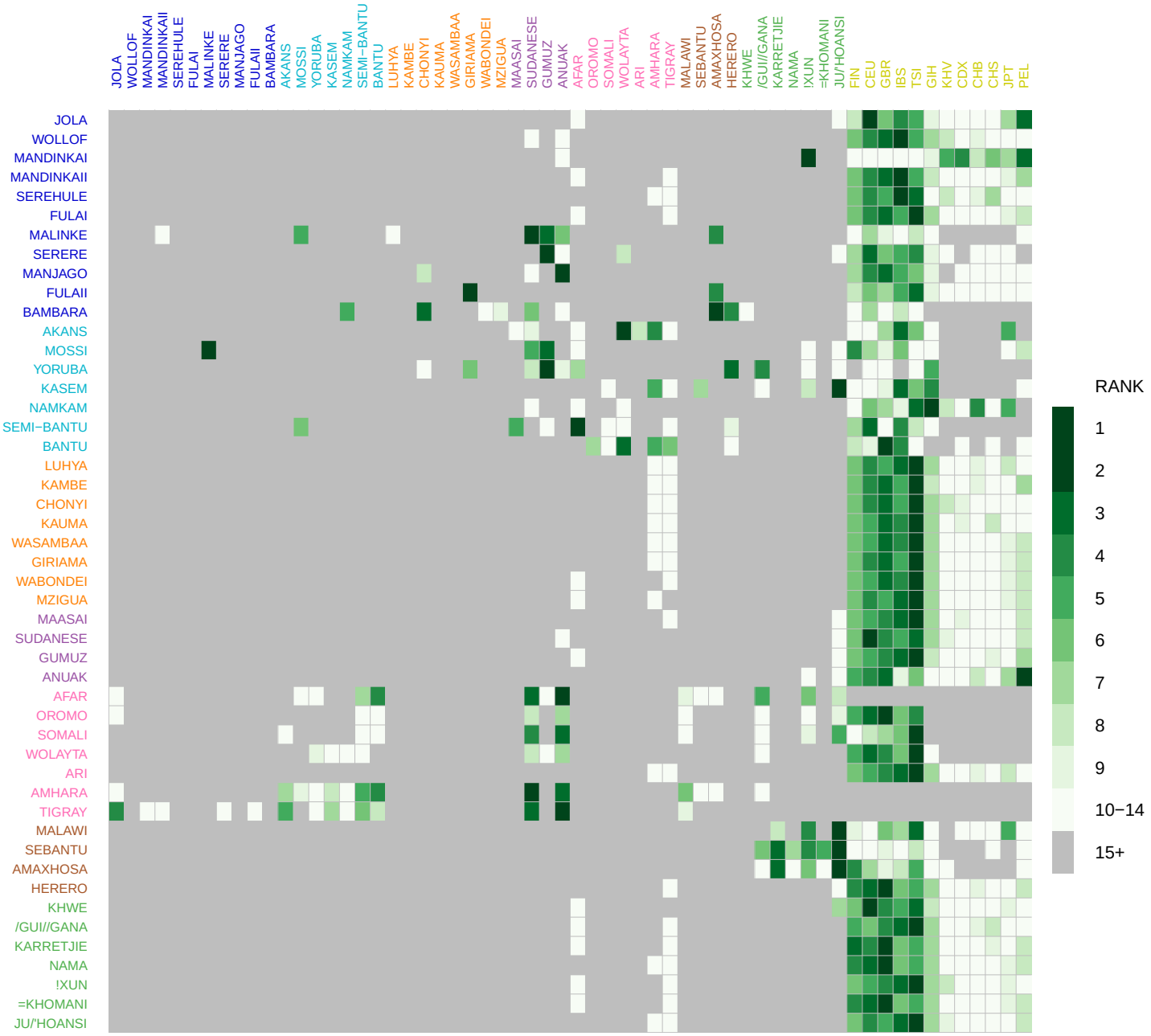

Figure 3-figure supplement 2. Comparison of weighted LD amplitude scores across all African ethnic groups. For a given test population we computed the ALDER amplitude (y-axis intercept) using the test population and every other population as the second reference. We then ranked the amplitudes across a given test population: populations who gave the top-ranked (i.e. largest) amplitude are in green, with those beneath a rank 15 shown in grey. This analysis shows that for many populations the reference populations giving the largest amplitudes (i.e. have the highest rank) are often non-African groups.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

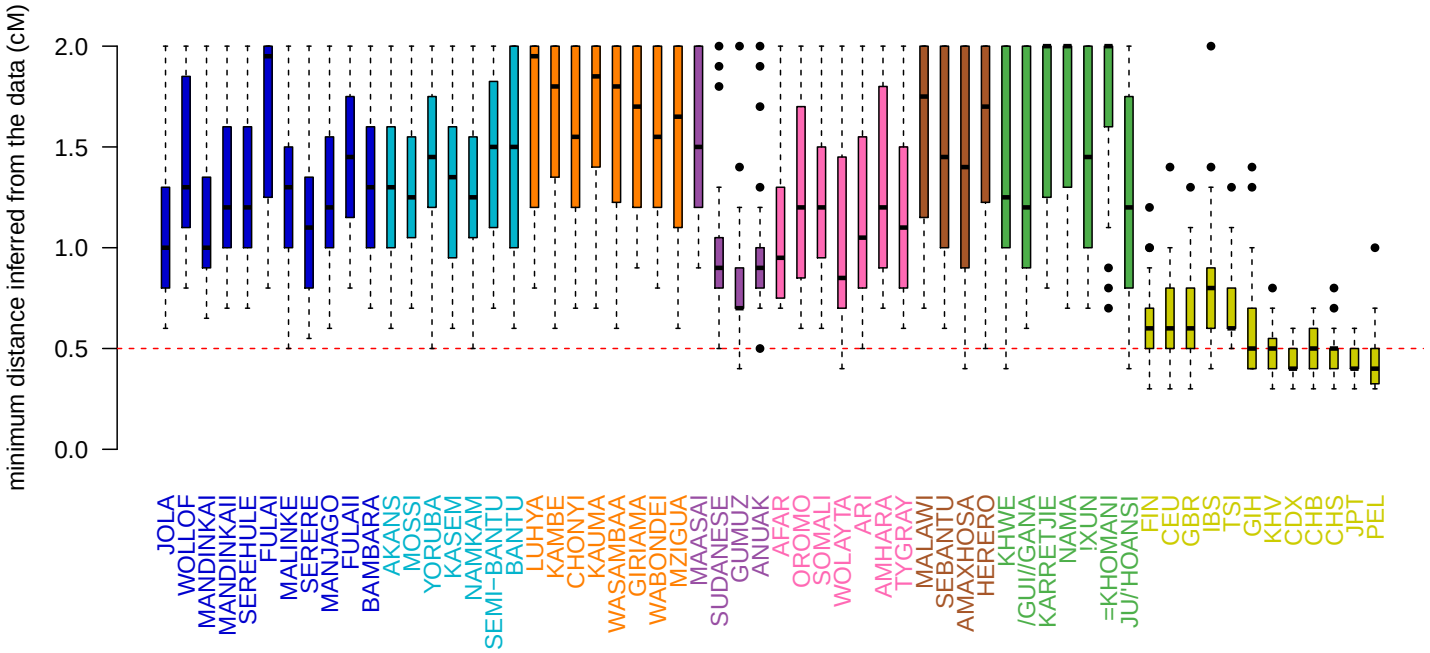

Figure 3-figure supplement 3. Comparison of the minimum distance to begin computing admixture LD. For each of the 48 African populations as a target, we used ALDER to compute the minimum distance over which short-range LD is shared with each of the 47 other African and 12 Eurasian reference populations. Here we show boxplots showing the distribution of minimum inferred genetic distances (y-axis) over which LD is shared for each of the reference populations separately (x-axis). We performed two analyses using weighted LD, one using these values of the minimum distance inferred from the data, and another where this distance was forced to be 0.5cM (dotted red line). The comparisons show that all African populations share LD correlations at distances > 0.5cM with all other African populations. Note that ALDER computes LD correlations at distances < 2cM.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

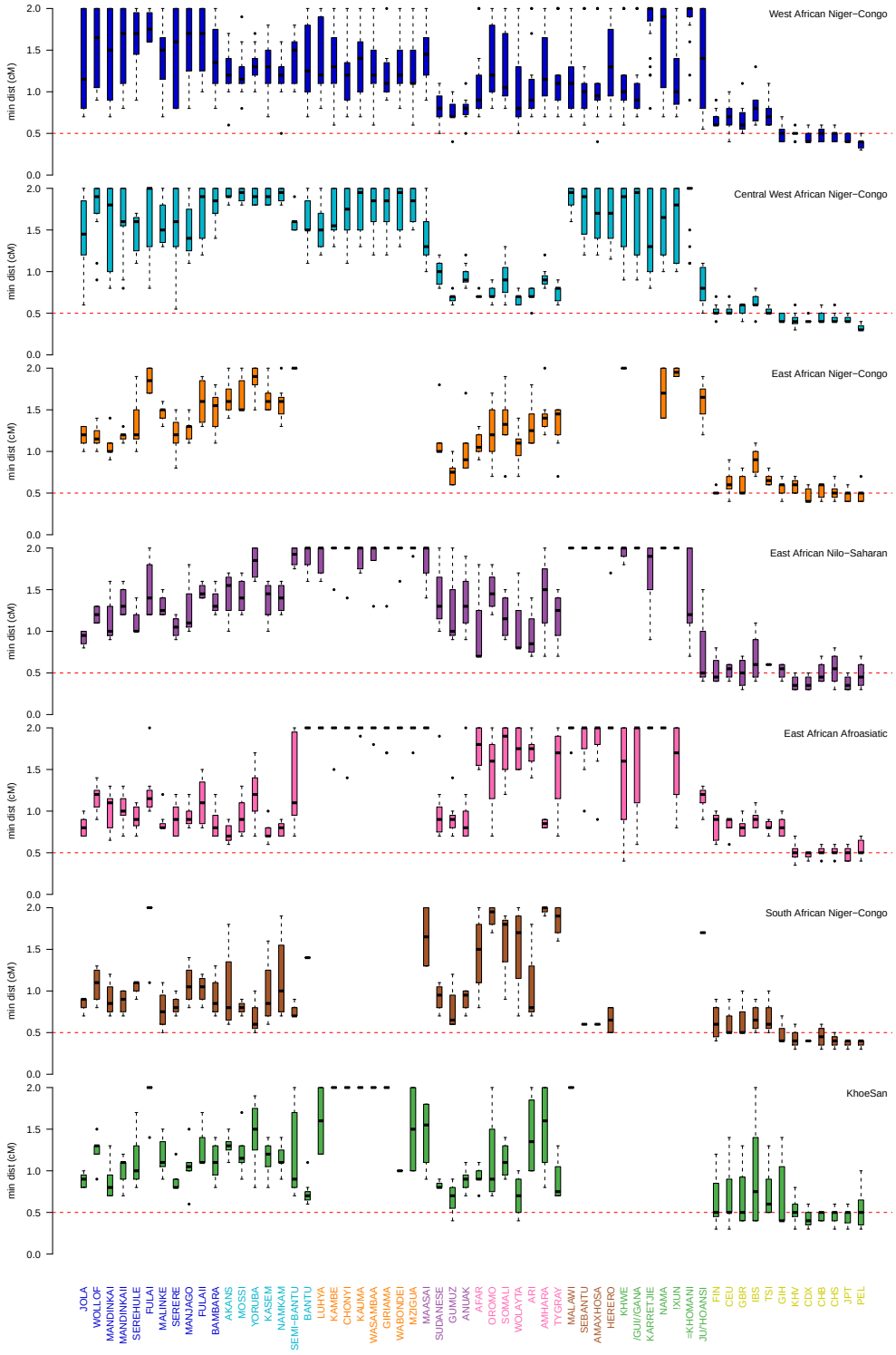

Figure 3-figure supplement 4. Comparison of the minimum distance to begin computing admixture LD split by region. As in Figure 3-figure supplement 3 except distances are stratified by region. The comparisons show that all African populations share LD correlations at distances > 0.5cM with all other African populations. Note that ALDER computes LD correlations at distances < 2cM.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

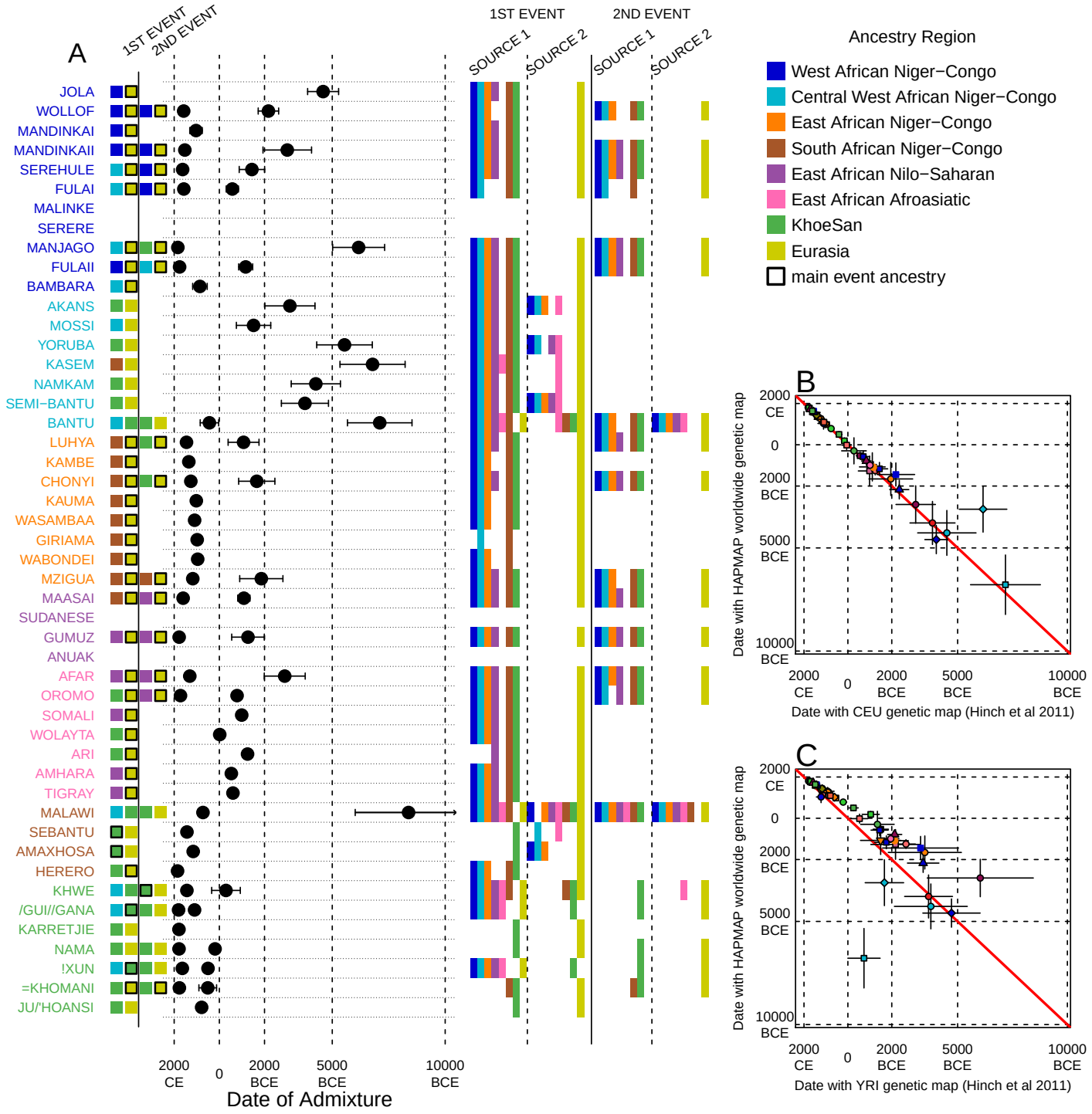

Figure 3-figure supplement 5. Results of the MALDER analysis computing weighted admixture decay curves from 0.5cM. As in the main analyses, the algorithm was run independently three times with the HAPMAP, YRI, and CEU genetic maps. The main results shown here are from the HAPMAP analysis. For each population, we show the ancestry region identity of the two populations involved in generating the MALDER curves with the greatest amplitudes (which are the closest to the true admixing sources amongst the reference populations) for at most two events. The sources generating the greatest amplitude are highlighted with a black box. Populations are ordered by ancestry of the admixture sources and dates estimates which are shown ± 1 s.e. (B) Comparison of dates of admixture ± 1 s.e. for MALDER dates inferred using the HAPMAP recombination map and a recombination map inferred from European (CEU) individuals from Hinch et al. [2011]. We only show comparisons for dates where the same number of events were inferred using both methods. Point symbols refer to populations and are as in Figure 1. (C) as (B) but comparing with an African (YRI) map.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

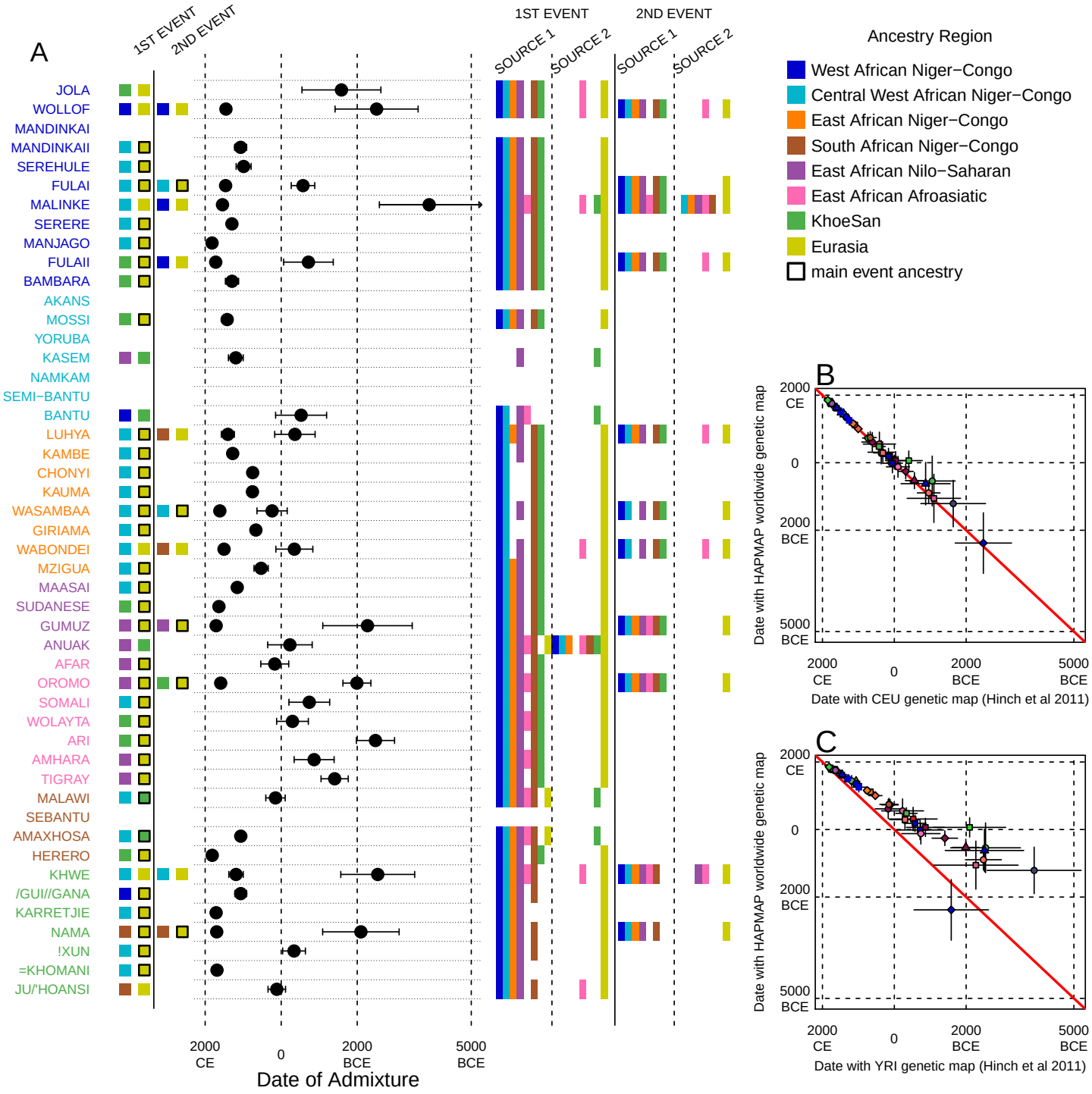

Figure 3-figure supplement 6. Results of MALDER for all populations using an African specific recombination map. We used MALDER to identify the evidence for multiple waves of admixture in each population. (A) For each population, we show the ancestry region identity of the two populations involved in generating the MALDER curves with the greatest amplitudes (which are the closest to the true admixing sources amongst the reference populations) for at most two events. The sources generating the greatest amplitude are highlighted with a black box. Populations are ordered by ancestry of the admixture sources and dates estimates which are shown ± 1 s.e. (B) Comparison of dates of admixture ± 1 s.e. for MALDER dates inferred using the HAPMAP recombination map and a recombination map inferred from European (CEU) individuals from Hinch et al. [2011]. We only show comparisons for dates where the same number of events were inferred using both methods. Point symbols refer to populations and are as in Figure 1. (C) as (B) but comparing with an African (YRI) map.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

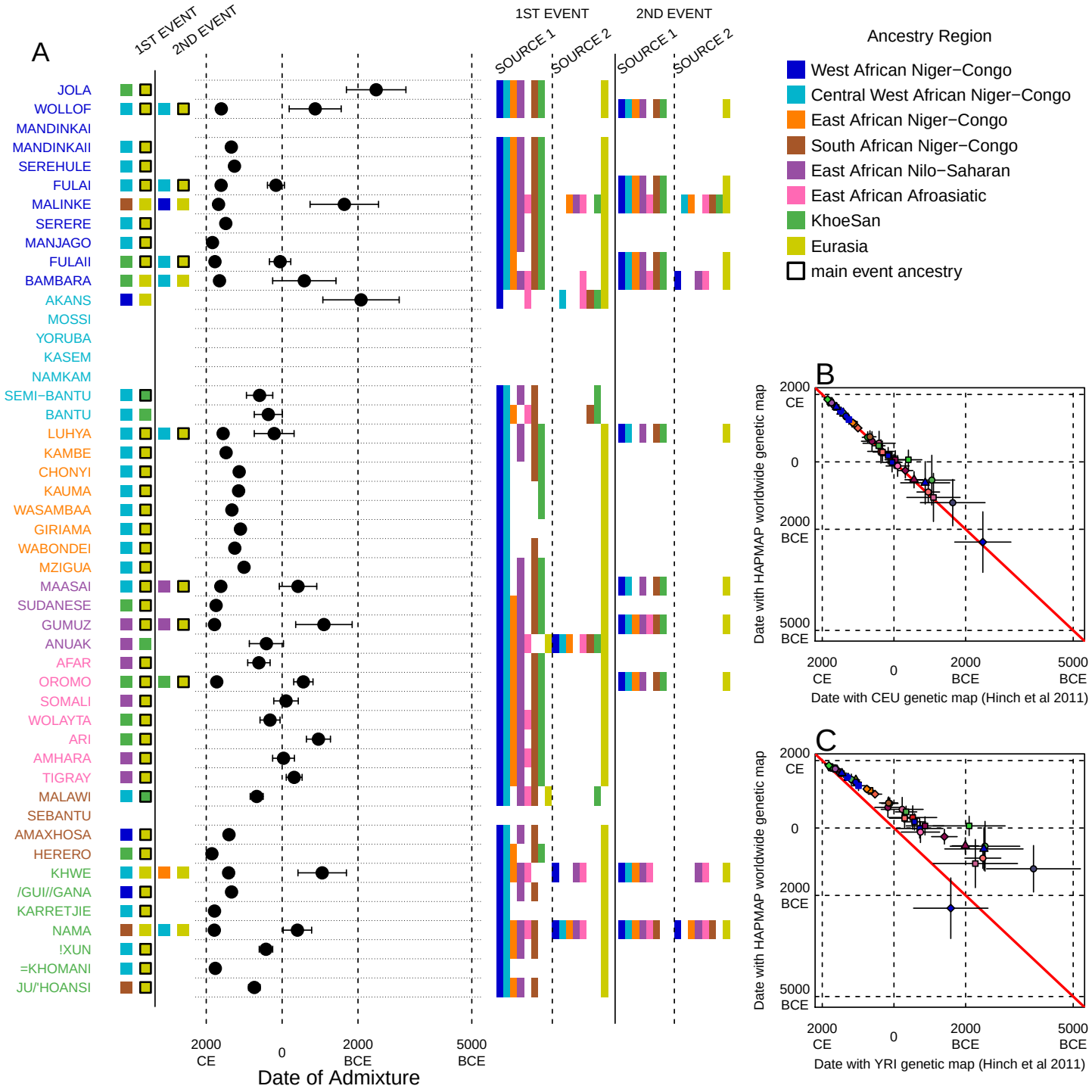

Figure 3-figure supplement 7. Results of MALDER for all populations using a European specific recombination map. We used MALDER to identify the evidence for multiple waves of admixture in each population. (A) For each population, we show the ancestry region identity of the two populations involved in generating the MALDER curves with the greatest amplitudes (which are the closest to the true admixing sources amongst the reference populations) for at most two events. The sources generating the greatest amplitude are highlighted with a black box. Populations are ordered by ancestry of the admixture sources and dates estimates which are shown ± 1 s.e. (B) Comparison of dates of admixture ± 1 s.e. for MALDER dates inferred using the HAPMAP recombination map and a recombination map inferred from European (CEU) individuals from Hinch et al. [2011]. We only show comparisons for dates where the same number of events were inferred using both methods. Point symbols refer to populations and are as in Figure 1. (C) as (B) but comparing with an African (YRI) map.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4-figure supplement 1. Admixture source inference by GLOBETROTTER after sequentially removing local surrogates from the analysis. In addition to the Full analysis, we show the inferred composition of admixture sources for different, restricted surrogate analyses. Components and y-axis labels are coloured by ancestry region. In each case we show admixture sources inferred by GLOBETROTTER for a single date of admixture.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

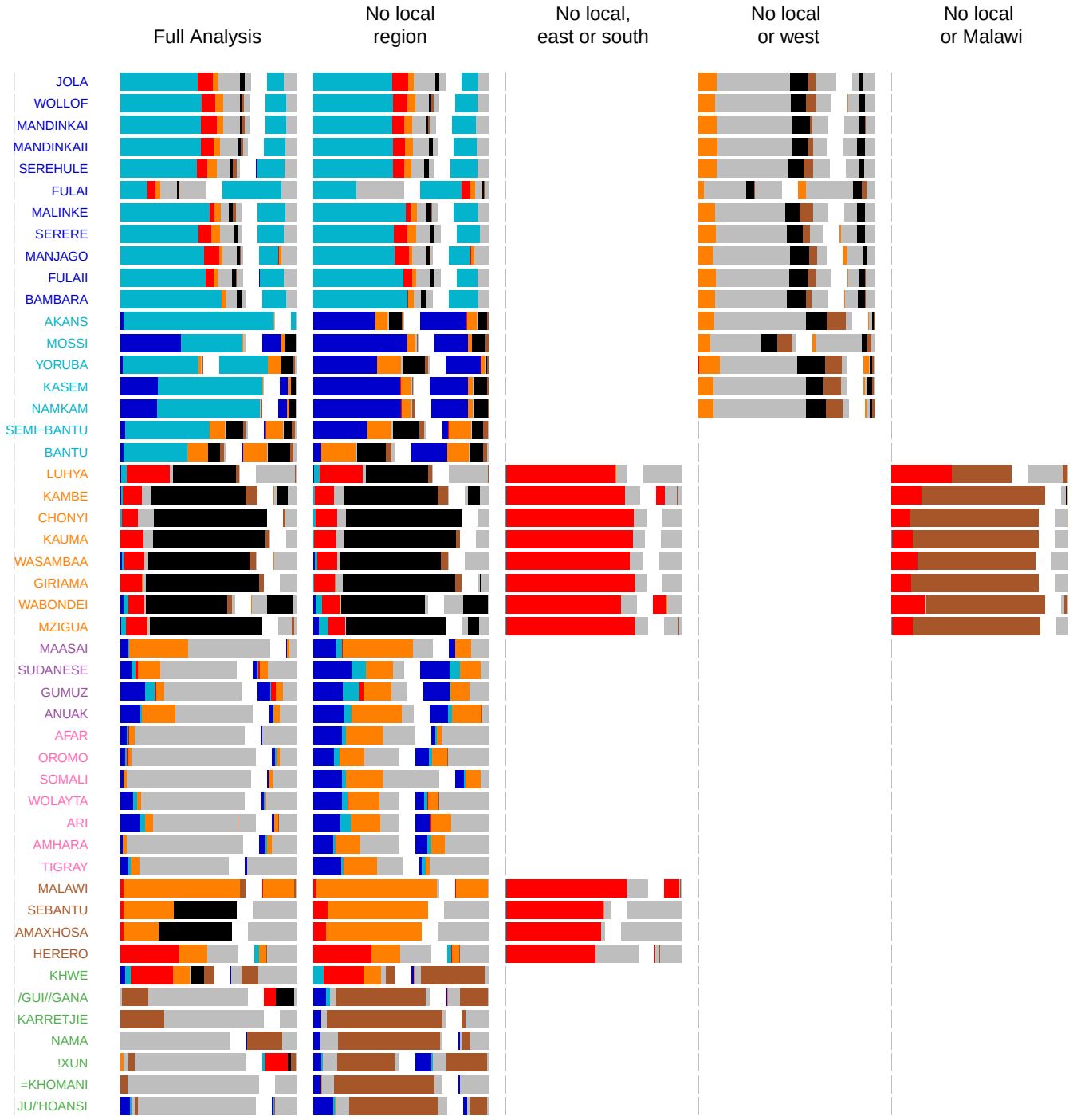

Figure 4-figure supplement 2. Admixture source inference by GLOBETROTTER after sequentially removing local surrogates from the analysis. The results are the same as Figure 4-figure supplement 1, but only Niger-Congo speaking groups are coloured. We highlight Malawi components in black, and Cameroon (Bantu and Semi-Bantu) in red.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

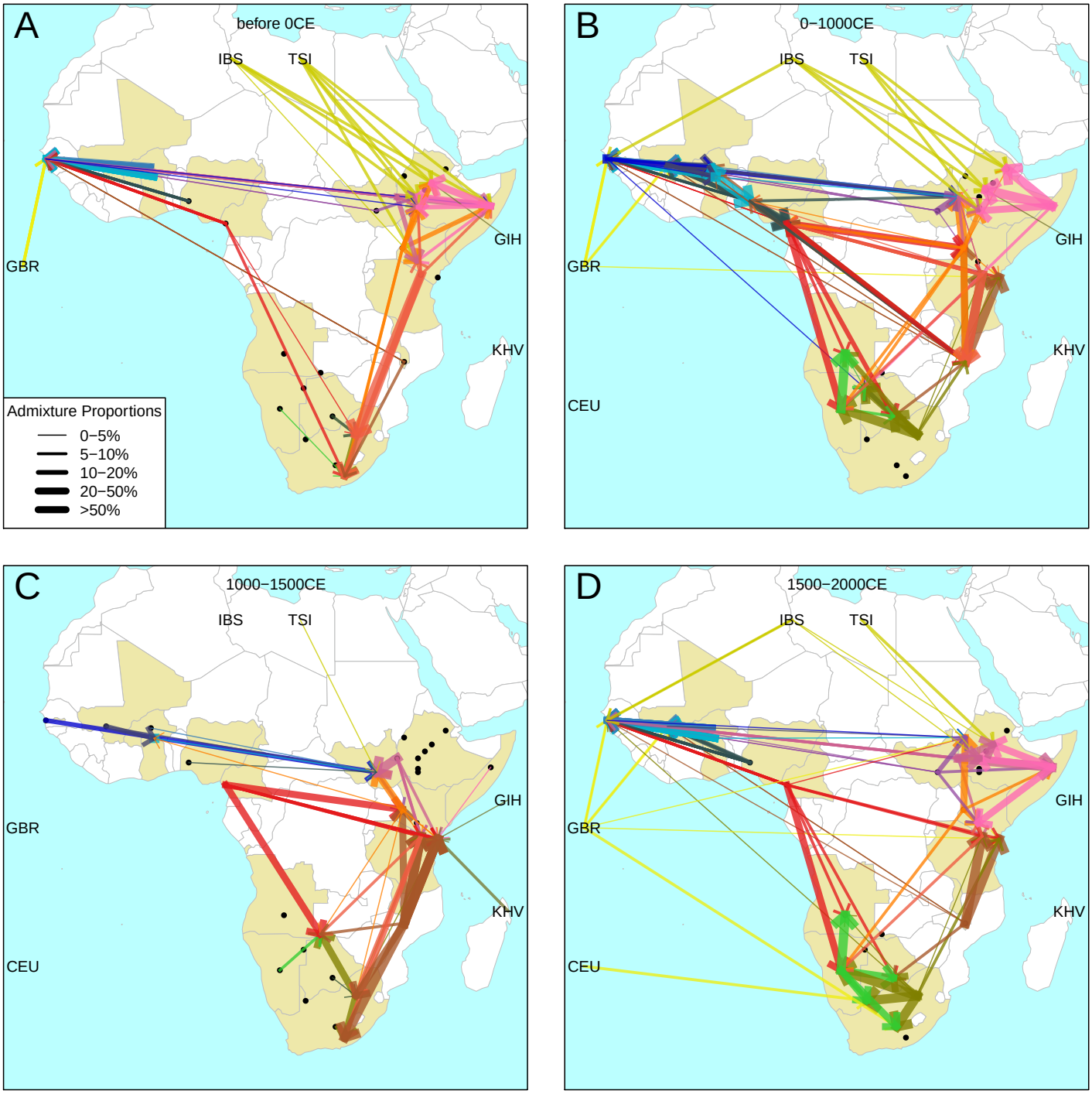

Figure 6-figure supplement 1. Gene-flow in Africa over the last 2,000 years. Using the results of the GLOBETROTTER analysis we show the connections between different groups in sub-Saharan Africa over time. For each population, we inferred the date of admixture and the composition of the admixing sources. We link each recipient population to its donor components using arrows, the size of which is proportional to the amount it contributes to the admixture event. Arrows are coloured by country of origin, as in Figure 4 in the main text.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary File 1 A note on ethnolinguistic groupings

The results of the population genetic analysis shows that population structure is largely the result of ethno-linguistic similarity, which itself is largely but not completely correlated with geographical proximity. These divisions are shown below and referred to in the text, together with the latest Ethnologue classification[‡] of the languages spoken, where possible.

1. 1st major Niger-Congo speaking group from West Africa: Gambian and Malian ethnic groups
   - Niger-Congo, **Mande** {Mandinka, Malinke, Bambara}
   - Niger-Congo, Atlantic-Congo, Atlantic, **Northern, Senegambian, Fula-Wolof** {Fula, Wollof}
   - Niger-Congo, Atlantic-Congo, Atlantic, **Northern, Senegambian, Serer** {Serere, Serehule}
   - Niger-Congo, Atlantic-Congo, Atlantic, **Northern, Bak** {Jola}
2. 2nd major Niger-Congo speaking group from West Africa: Ghana/BF/Nigerian ethnic groups:
   - Niger-Congo, Atlantic-Congo, Atlantic, **Volta-Congo**, **Kwa**{Akan, Yoruba}
   - Niger-Congo, Atlantic-Congo, Atlantic, **Volta-Congo**, **North**, **Gur** {Mossi, Kasem, Namkam?}
3. A Central and Eastern African Niger-Congo / "Bantoid" speaking group, split into two subdivisions:
   (a) North Western
   - Niger-Congo, Atlantic-Congo, Atlantic, Volta-Congo, **Benue-Congo**, **exNarrow-Bantu** {Cameroon: Bantu, Semi-Bantu?}

   (b) Eastern
   - Niger-Congo, Atlantic-Congo, Atlantic, Volta-Congo, **Benue-Congo**, **Bantoid**, **Southern, Narrow-Bantu**, **Central**, **I** {Masaba-Luhya?}
   - Niger-Congo, Atlantic-Congo, Atlantic, Volta-Congo, **Benue-Congo**, **Bantoid**, **Southern, Narrow-Bantu**, **Central**, **E** {Mijikenda (Kenya)}
   - Niger-Congo, Atlantic-Congo, Atlantic, Volta-Congo, **Benue-Congo**, **Bantoid**, **Southern, Narrow-Bantu**, **Central**, **F-G** {Tanzania}
   - Niger-Congo, Atlantic-Congo, Atlantic, Volta-Congo, **Benue-Congo**, **Bantoid**, **Southern, Narrow-Bantu**, **Central**, **N** {Malawi (Chewa)}
4. Afroasiatic and Nilo-Saharan speakers from the Horn of Africa:
   - Afroasiatic, **Cushitic** {Afar, Somali, Oromo}
   - Afroasiatic, **Semitic** {Amhara, Tigrayan}

[‡]information accessed on 28th July 2014 at https://www.ethnologue.com/

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

- Afroasiatic, **Omotic** {Ari, Wolayta}
- Nilo-Saharan, **Komuz** {Gumuz}
- Nilo-Saharan, **Eastern Sudanic**, **Nilotic** {Maasai}
- Nilo-Saharan, **Eastern Sudanic**, **Nilotic** {Anuak}
- Nilo-Saharan {Sudanese}

5. Khoesan and Bantu speaking groups from Southern Africa

   (a) Khoesan

   - Khoesan, Southern Africa, **Northern** {Ju/'hoansi, !Xun}
   - Khoesan, Southern Africa, **Central** {Nama}
   - Khoesan, Southern Africa, **Central**, **Tshu-Khwe**, **Northwest** {/Gui//Gana, Khwe}
   - Khoesan, Southern Africa, **Southern** {≠Khomani}
   - uncertain {Karretijie}

   (b) Southern Bantu speakers

   - Niger-Congo, Atlantic-Congo, Atlantic, Volta-Congo, **Benue-Congo**, **Bantoid**, **Southern, Narrow-Bantu**, **Central**, **R** {Herero}
   - Niger-Congo, Atlantic-Congo, Atlantic, Volta-Congo, **Benue-Congo**, **Bantoid**, **Southern, Narrow-Bantu**, **Central**, **S** {Amaxhosa, SEBantu}

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Source Data

Figure 1-Source Data 1. Overview of sampled populations describing the continent, region, numbers of individuals used, and the source of any previously published datasets.

Ancestry Region	Country	Ethnic Group	Origin	N(phased)	IBD	REL	N(final)
Western Africa Niger-Congo	Gambia	JOLA	this study	147	0	31	116
Western Africa Niger-Congo	Gambia	MANJAGO	this study	47	0	0	47
Western Africa Niger-Congo	Gambia	SEREHULE	this study	47	0	0	47
Western Africa Niger-Congo	Gambia	SERERE	this study	50	0	0	50
Western Africa Niger-Congo	Gambia	WOLLOF	this study	135	0	27	108
Western Africa Niger-Congo	Gambia	MANDINKAI	this study	60	0	28	32
Western Africa Niger-Congo	Gambia	MANDINKAII	this study	71	0	0	71
Western Africa Niger-Congo	Gambia	FULAI	this study	97	0	25	72
Western Africa Niger-Congo	Gambia	FULAII	this study	79	0	0	79
Western Africa Niger-Congo	Mali	BAMBARA	this study	50	0	0	50
Western Africa Niger-Congo	Mali	MALINKE	this study	49	0	0	49
Central West Africa Niger-Congo	BurkinaFaso	MOSSI	this study	50	0	0	50
Central West Africa Niger-Congo	Ghana	AKANS	this study	50	0	0	50
Central West Africa Niger-Congo	Ghana	KASEM	this study	50	0	0	50
Central West Africa Niger-Congo	Ghana	NAMKAM	this study	50	0	0	50
Central West Africa Niger-Congo	Nigeria	YORUBA	1KG	161	0	60	101
East Africa Niger-Congo	Cameroon	BANTU	this study	50	0	0	50
East Africa Niger-Congo	Cameroon	SEMI-BANTU	this study	50	0	0	50
East Africa Niger-Congo	Kenya	LUHYA	1KG	100	9	0	91
East Africa Niger-Congo	Kenya	CHONYI	this study	47	0	0	47
East Africa Niger-Congo	Kenya	KAUMA	this study	97	0	0	97
East Africa Niger-Congo	Kenya	KAMBE	this study	42	0	0	42
East Africa Niger-Congo	Kenya	GIRIAMA	this study	46	0	0	46
East Africa Niger-Congo	Tanzania	MZIGUA	this study	50	0	0	50
East Africa Niger-Congo	Tanzania	WABONDEI	this study	48	0	0	48
East Africa Niger-Congo	Tanzania	WASAMBAA	this study	50	0	0	50
East Africa Nilo-Saharan	Kenya	MAASAI	1KG	31	0	0	31
East Africa Nilo-Saharan	Sudan	SUDANESE	Pagani 2012	24	0	0	24
East Africa Nilo-Saharan	Ethiopia	GUMUZ	Pagani 2012	19	0	0	19
East Africa Nilo-Saharan	Ethiopia	ANUAK	Pagani 2012	23	0	0	23
East Africa Nilo-Saharan	Somalia	SOMALI	Pagani 2012	40	0	0	40
East Africa Afroasiatic	Ethiopia	ARI	Pagani 2012	41	0	0	41
East Africa Afroasiatic	Ethiopia	AFAR	Pagani 2012	12	0	0	12
East Africa Afroasiatic	Ethiopia	TIGRAY	Pagani 2012	21	0	0	21
East Africa Afroasiatic	Ethiopia	AMHARA	Pagani 2012	26	0	0	26
East Africa Afroasiatic	Ethiopia	WOLAYTA	Pagani 2012	8	0	0	8
East Africa Afroasiatic	Ethiopia	OROMO	Pagani 2012	21	0	0	21
South Africa Niger-Congo	Malawi	MALAWI	this study	100	0	0	100
South Africa Niger-Congo	Namibia	HERERO	Schlebusch 2012	12	0	0	12
South Africa Niger-Congo	SouthAfrica	SEBANTU	Schlebusch 2012	20	0	0	20
South Africa Niger-Congo	SouthAfrica	AMAXHOSA	Peterson 2013	15	0	0	15
South Africa KhoeSan	SouthAfrica	KARRETJIE	Schlebusch 2012	20	0	0	20
South Africa KhoeSan	SouthAfrica	≠KHOMANI	Schlebusch 2012	39	0	0	39
South Africa KhoeSan	Namibia	NAMA	Schlebusch 2012	20	0	0	20
South Africa KhoeSan	Namibia	KHWE	Schlebusch 2012	17	0	0	17
South Africa KhoeSan	Angola	!XUN	Schlebusch 2012 Peterson 2013	33	0	0	33
South Africa KhoeSan	Botswana	/GUI //GANA	Schlebusch 2012	15	0	0	15
South Africa KhoeSan	Namibia	JU/'HOANSI	Schlebusch 2012 Peterson 2013	37	0	0	37
Europe	Finland	FIN	1KG	100	0	0	100
Europe	NorthEurope	CEU	1KG	104	0	2	102
Europe	Britain	GBR	1KG	101	1	3	97
Europe	Spain	IBS	1KG	150	0	50	100
Europe	Italy	TSI	1KG	100	1	0	99
Asia	India	GIH	1KG	100	2	0	98
Asia	Vietnam	KHV	1KG	121	1	21	99
Asia	China	CDX	1KG	100	8	0	92
Asia	China	CHB	1KG	101	0	1	100
Asia	China	CHS	1KG	150	7	50	93
Asia	Japan	JPT	1KG	100	0	0	100
Americas	Peru	PELII	1KG	105[§]	0	0	16
				3699	29	298	3283

[§]89 admixed Peruvians and 517 individuals from five other 1KG American populations were removed from further analysis after phasing

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Figure 2-Source Data 1. Pairwise F_{ST} for African populations. We used *smartpca* to compute F_{ST} for each pair of populations, upper left diagonal, together with standard errors computed using a block jacknife. F_{ST} has been multiplied by 1000

	JOLA	WOLLOF	MANDINKAI	MANDINKAII	SEREHULE	FULAI	MALINKE	SERERE	MANJAGO	FULAII	BAMBARA	AKANS	MOSSI	YORUBA	KASEM	NAMKAM	SEMI-BANTU	BANTU
JOLA		4	1	3	5	23	6	3	3	5	7	10	9	10	10	9	11	13
WOLLOF	0.10		2	1	2	17	3	1	3	2	4	7	6	7	6	6	8	10
MANDINKAI	0.10	0.10		1	2	18	3	1	2	2	4	7	6	7	7	6	8	10
MANDINKAII	0	0	0.10		1	18	2	1	1	1	3	5	4	6	5	5	7	8
SEREHULE	0.10	0	0.10	0.10		17	1	2	3	1	2	5	4	6	5	5	7	8
FULAI	0.20	0.20	0.20	0.20	0.20		19	18	20	17	20	23	22	24	23	23	24	25
MALINKE	0.10	0.10	0.10	0.10	0.10	0.20		2	4	1	1	3	2	4	3	3	5	7
SERERE	0.10	0	0.10	0	0.10	0.20	0.10		2	1	3	6	5	6	6	5	8	9
MANJAGO	0.10	0.10	0.10	0.10	0.10	0.20	0.10	0.10		2	4	7	6	7	7	7	8	10
FULAII	0.10	0	0.10	0	0	0.20	0	0	0.10		1	3	3	4	3	3	5	6
BAMBARA	0.10	0.10	0.10	0.10	0.10	0.20	0.10	0.10	0.10	0		2	2	3	2	2	4	6
AKANS	0.10	0.10	0.10	0.10	0.10	0.30	0.10	0.10	0.10	0.10	0.10		2	2	2	2	4	5
MOSSI	0.10	0.10	0.10	0.10	0.10	0.20	0.10	0.10	0.10	0.10	0.10	0.10		2	1	1	4	5
YORUBA	0.10	0.10	0.10	0.10	0.10	0.20	0.10	0.10	0.10	0.10	0.10	0	0.10		2	2	2	4
KASEM	0.10	0.10	0.10	0.10	0.10	0.30	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10		0	4	6
NAMKAM	0.10	0.10	0.10	0.10	0.10	0.30	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10		4	5
SEMI-BANTU	0.10	0.10	0.10	0.10	0.10	0.30	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10		2
BANTU	0.10	0.10	0.10	0.10	0.10	0.30	0.10	0.10	0.20	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	
LUHYA	0.10	0.10	0.10	0.10	0.10	0.20	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
KAMBE	0.10	0.10	0.10	0.10	0.10	0.20	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
CHONYI	0.20	0.20	0.20	0.10	0.20	0.20	0.10	0.20	0.20	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
KAUMA	0.10	0.10	0.10	0.10	0.10	0.20	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
WASAMBAA	0.10	0.10	0.10	0.10	0.10	0.20	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
GIRIAMA	0.10	0.10	0.20	0.10	0.10	0.30	0.10	0.10	0.20	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
WABONDEI	0.10	0.10	0.10	0.10	0.10	0.20	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
MZIGUA	0.10	0.10	0.10	0.10	0.10	0.20	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
MAASAI	0.30	0.30	0.30	0.30	0.30	0.20	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30
WOLAYTA	0.50	0.50	0.50	0.50	0.50	0.40	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
SOMALI	0.50	0.40	0.50	0.40	0.40	0.30	0.40	0.40	0.50	0.40	0.40	0.50	0.50	0.50	0.50	0.50	0.40	0.50
ARI	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30
SUDANESE	0.30	0.20	0.30	0.20	0.20	0.30	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
GUMUZ	0.40	0.30	0.40	0.30	0.30	0.40	0.30	0.40	0.40	0.30	0.30	0.30	0.30	0.30	0.40	0.30	0.30	0.30
OROMO	0.50	0.40	0.50	0.40	0.40	0.30	0.50	0.50	0.50	0.40	0.40	0.50	0.50	0.50	0.50	0.50	0.50	0.50
AMHARA	0.50	0.50	0.50	0.50	0.50	0.40	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
AFAR	0.60	0.60	0.60	0.60	0.60	0.40	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60
ANUAK	0.20	0.20	0.20	0.20	0.20	0.30	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
TIGRAY	0.60	0.50	0.60	0.50	0.50	0.40	0.60	0.50	0.50	0.50	0.50	0.60	0.60	0.50	0.50	0.60	0.60	0.50
MALAWI	0.10	0.10	0.10	0.10	0.10	0.20	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
SEBANTU	0.20	0.20	0.20	0.20	0.20	0.30	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.10	0.20	0.10
AMAXHOSA	0.20	0.20	0.20	0.20	0.20	0.30	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
HERERO	0.30	0.30	0.30	0.30	0.30	0.40	0.30	0.30	0.30	0.30	0.30	0.30	0.20	0.20	0.30	0.30	0.20	0.20
KHWE	0.30	0.30	0.30	0.30	0.30	0.40	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30
/GUI//GANA	0.50	0.50	0.50	0.50	0.50	0.60	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
KARRETJIE	0.50	0.50	0.50	0.50	0.50	0.60	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
NAMA	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.30
!XUN	0.50	0.50	0.50	0.50	0.50	0.60	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
≠KHOMANI	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
JU/'HOANSI	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.60	0.70	0.70	0.60	0.60
FIN	1.10	1	1.10	1.10	1	0.90	1.10	1.10	1.10	1	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10
CEU	1.10	1	1.10	1.10	1	0.80	1.10	1.10	1.10	1	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10
GBR	1.10	1	1.10	1.10	1	0.90	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10	1.10
IBS	1	1	1	1	1	0.80	1	1	1	1	1	1	1	1	1	1	1	1
TSI	1	1	1	1	1	0.80	1	1	1	1	1	1.10	1	1	1	1	1.10	1.10
GIH	1	0.90	1	1	0.90	0.70	1	1	0.90	0.90	1	1	1	1	1	1	1	1
KHV	1.20	1.20	1.20	1.20	1.20	1.10	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20
CDX	1.20	1.20	1.20	1.20	1.20	1.10	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20
CHB	1.20	1.20	1.20	1.20	1.20	1.10	1.20	1.20	1.20	1.20	1.20	1.30	1.20	1.20	1.20	1.20	1.30	1.20
CHS	1.20	1.20	1.20	1.20	1.20	1.10	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20
JPT	1.20	1.20	1.20	1.20	1.20	1.10	1.20	1.20	1.20	1.20	1.20	1.30	1.20	1.20	1.20	1.20	1.20	1.20
PEL	1.40	1.40	1.40	1.40	1.40	1.30	1.40	1.40	1.40	1.40	1.40	1.40	1.40	1.40	1.40	1.40	1.40	1.40

	LUHYA	KAMBE	CHONYI	KAUMA	WASAMBAA	GIRIAMA	WABONDEI	MZIGUA	MAASAI	WOLAYTA	SOMALI	ARI	SUDANESE	GUMUZ	OROMO	AMHARA	AFAR	ANUAK	TIGRAY
JOLA	15	15	19	17	15	17	14	14	32	52	58	54	30	46	57	66	69	30	68
WOLLOF	12	12	15	13	12	14	11	11	27	46	51	48	26	41	50	59	61	26	60
MANDINKAI	12	12	15	13	12	14	11	11	28	48	53	50	27	42	52	61	64	26	63
MANDINKAII	11	11	14	12	11	13	10	10	27	47	52	48	26	41	51	60	63	25	62
SEREHULE	11	10	14	12	10	12	10	10	26	45	51	48	25	40	50	59	61	25	60
FULAI	24	23	27	25	22	26	23	24	23	32	35	42	36	45	32	38	41	35	38
MALINKE	10	9	13	11	9	11	9	8	27	47	53	49	25	41	52	61	64	25	63
SERERE	12	11	15	13	11	13	11	10	28	47	53	49	26	41	52	61	63	26	62
MANJAGO	12	12	16	14	12	14	11	11	29	48	54	50	27	42	53	62	64	27	64
FULAII	9	9	12	10	9	11	8	8	26	45	52	47	25	40	50	59	62	24	61
BAMBARA	9	9	12	10	9	10	8	8	27	48	53	49	25	41	52	62	64	24	63
AKANS	9	8	12	10	8	10	8	7	29	50	56	50	26	42	55	64	67	25	66
MOSSI	9	9	12	10	8	10	8	7	28	49	54	50	25	41	53	62	65	25	64
YORUBA	8	7	10	9	7	9	6	6	28	49	55	50	25	41	54	64	66	25	65
KASEM	9	9	12	10	9	10	8	8	28	49	55	51	26	42	54	63	66	25	65
NAMKAM	9	9	12	10	9	10	8	8	28	49	55	50	25	41	54	63	66	25	65
SEMI-BANTU	6	5	8	6	5	7	4	4	26	48	54	48	24	39	53	62	65	23	64
BANTU	6	5	8	7	6	7	5	4	27	48	55	48	24	40	53	63	66	24	65
LUHYA		5	8	6	4	7	4	4	18	38	45	40	17	31	43	52	55	16	54
KAMBE	0.10		3	1	2	1	2	2	20	39	45	41	23	36	43	52	55	22	54
CHONYI	0.10	0.10		5	5	6	5	5	25	44	51	46	27	40	49	58	61	26	60
KAUMA	0.10	0.10	0.10		4	3	3	3	23	41	48	43	24	38	46	55	58	24	57
WASAMBAA	0.10	0.10	0.10	0.10		4	1	1	19	37	43	39	22	34	41	50	53	21	52
GIRIAMA	0.10	0.10	0.10	0.10	0.10		4	4	23	43	49	45	25	38	47	56	59	24	58
WABONDEI	0.10	0.10	0.10	0.10	0.10	0.10		1	20	39	46	41	22	35	44	53	56	21	55
MZIGUA	0.10	0.10	0.10	0.10	0.10	0.10	0.10		22	42	48	43	23	37	46	56	58	22	57
MAASAI	0.20	0.20	0.30	0.20	0.20	0.20	0.20	0.20		13	17	23	18	25	14	21	23	16	22
WOLAYTA	0.50	0.40	0.50	0.40	0.40	0.50	0.40	0.50	0.30		11	15	38	34	3	7	12	35	8
SOMALI	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.20	0.30		31	42	45	7	9	10	41	9
ARI	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.20	0.30	0.30		38	32	22	31	36	35	33
SUDANESE	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.50	0.40	0.30		22	42	52	54	2	54
GUMUZ	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.50	0.40	0.30	0.30		41	51	55	19	53
OROMO	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.20	0.30	0.10	0.30	0.50	0.50		2	6	40	2
AMHARA	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.30	0.30	0.20	0.30	0.50	0.50	0.10		6	50	1
AFAR	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.40	0.40	0.20	0.40	0.60	0.60	0.20	0.20		53	5
ANUAK	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.50	0.50	0.30	0.10	0.30	0.50	0.50	0.60		52
TIGRAY	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.30	0.30	0.20	0.30	0.60	0.50	0.20	0.10	0.20	0.60	
MALAWI	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.30	0.50	0.40	0.30	0.20	0.30	0.50	0.50	0.60	0.20	0.60
SEBANTU	0.10	0.10	0.20	0.10	0.10	0.20	0.10	0.10	0.30	0.50	0.50	0.40	0.30	0.40	0.50	0.60	0.70	0.30	0.60
AMAXHOSA	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.30	0.50	0.50	0.40	0.30	0.40	0.50	0.60	0.70	0.30	0.60
HERERO	0.20	0.20	0.30	0.20	0.20	0.20	0.20	0.20	0.40	0.60	0.60	0.40	0.40	0.40	0.60	0.60	0.70	0.40	0.60
KHWE	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.60	0.50	0.40	0.40	0.40	0.50	0.60	0.70	0.40	0.60
/GUI//GANA	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.60	0.80	0.70	0.60	0.60	0.60	0.70	0.80	0.80	0.60	0.80
KARRETJIE	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.60	0.70	0.70	0.60	0.60	0.60	0.70	0.80	0.80	0.60	0.80
NAMA	0.30	0.30	0.40	0.30	0.30	0.30	0.30	0.40	0.40	0.60	0.50	0.40	0.40	0.50	0.50	0.60	0.70	0.40	0.60
!XUN	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.70	0.60	0.50	0.50	0.60	0.60	0.70	0.80	0.60	0.70
≠KHOMANI	0.40	0.40	0.40	0.30	0.40	0.40	0.40	0.40	0.40	0.60	0.50	0.40	0.40	0.50	0.50	0.60	0.70	0.40	0.60
JU/'HOANSI	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.70	0.80	0.80	0.70	0.70	0.80	0.80	0.90	0.90	0.70	0.90
FIN	1	1	1.10	1	1	1	1	1	0.80	0.80	0.60	0.80	1.10	1.10	0.60	0.50	0.60	1.10	0.50
CEU	1	1	1.10	1	1	1	1	1	0.80	0.80	0.60	0.80	1.10	1.10	0.60	0.50	0.60	1.10	0.50
GBR	1	1	1.10	1	1	1	1	1	0.80	0.80	0.60	0.80	1.10	1.10	0.60	0.50	0.60	1.10	0.50
IBS	1	1	1	1	1	1	1	1	0.80	0.70	0.60	0.80	1	1	0.50	0.50	0.60	1.10	0.50
TSI	1	1	1	1	1	1	1	1	0.80	0.70	0.60	0.80	1.10	1	0.60	0.50	0.60	1.10	0.50
GIH	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.70	0.70	0.50	0.70	1	0.90	0.50	0.50	0.60	1	0.50
KHV	1.20	1.10	1.20	1.20	1.10	1.20	1.20	1.20	1	1	1	1	1.20	1.20	0.90	0.90	1	1.20	0.90
CDX	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.10	1.10	1	1.10	1.20	1.20	1	1	1	1.20	0.90
CHB	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.10	1	1	1.10	1.20	1.20	1	0.90	1	1.20	0.90
CHS	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.10	1	1	1.10	1.20	1.20	1	0.90	1	1.20	0.90
JPT	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.10	1	1	1.10	1.20	1.20	1	0.90	1	1.20	0.90
PEL	1.40	1.40	1.40	1.40	1.40	1.40	1.40	1.40	1.30	1.30	1.20	1.30	1.40	1.40	1.20	1.20	1.30	1.40	1.20

	MALAWI	SEBANTU	AMAXHOSA	HERERO	KHWE	/GUI//GANA	KARRETJIE	NAMA	!XUN	≠KHOMANI	JU/'HOANSI
JOLA	15	19	20	21	29	58	66	51	68	59	106
WOLLOF	12	16	17	18	27	56	63	48	66	56	104
MANDINKAI	12	16	17	18	26	56	62	48	65	56	104
MANDINKAII	11	15	16	17	25	55	61	47	64	55	102
SEREHULE	11	15	16	17	25	55	61	47	65	55	103
FULAI	27	31	31	29	38	69	68	50	77	56	114
MALINKE	9	13	14	16	24	54	61	47	64	55	102
SERERE	11	15	16	17	26	56	62	47	65	56	104
MANJAGO	12	16	17	18	26	55	62	48	65	56	103
FULAII	9	13	14	15	23	52	59	44	62	53	
BAMBARA	8	13	13	15	24	54	61	47	64	55	102
AKANS	7	12	13	14	23	52	60	46	62	55	101
MOSSI	8	13	13	14	24	55	62	48	65	56	104
YORUBA	6	11	12	13	22	53	61	46	63	55	102
KASEM	8	13	14	15	24	55	63	48	65	57	104
NAMKAM	8	13	14	15	24	55	63	48	65	57	104
SEMI-BANTU	4	8	9	11	19	49	57	43	59	51	97
BANTU	4	8	9	11	19	48	56	42	58	51	96
LUHYA	6	10	10	13	20	48	54	40	58	48	96
KAMBE	3	7	8	12	19	47	53	39	57	46	95
CHONYI	6	10	11	15	22	50	56	43	60	50	99
KAUMA	4	8	9	13	20	48	55	41	59	48	95
WASAMBAA	3	7	8	12	18	45	51	37	55	44	91
GIRIAMA	5	9	10	14	21	49	55	42	59	49	99
WABONDEI	2	6	7	11	18	45	52	38	55	45	92
MZIGUA	2	6	7	11	18	46	53	39	57	47	94
MAASAI	28	29	29	28	34	62	59	40	69	46	110
WOLAYTA	49	49	48	45	50	77	69	48	82	52	114
SOMALI	56	57	56	52	60	89	80	59	95	62	119
ARI	49	48	48	48	49	73	68	50	78	55	117
SUDANESE	26	29	29	30	36	65	69	52	73	60	114
GUMUZ	41	42	42	43	47	73	75	57	80	64	117
OROMO	54	54	54	49	56	85	75	53	90	56	120
AMHARA	64	64	64	58	66	96	84	61	101	64	118
AFAR	67	67	67	61	70		88	65	105	68	120
ANUAK	25	28	28	30	35	63	67	50	71	59	115
TIGRAY	66	66	66	60	68	98	85	63	103	66	118
MALAWI		5	6	12	19	46	55	42	57	50	99
SEBANTU	0.10		1	14	13	30	36	28	41	34	77
AMAXHOSA	0.20	0.20		15	13	30	35	26	40	33	74
HERERO	0.20	0.30	0.30		21	48	53	36	56	46	91
KHWE	0.30	0.20	0.30	0.40		24	29	20	26	25	54
/GUI//GANA	0.50	0.50	0.50	0.60	0.40		17	18	18	18	32
KARRETJIE	0.50	0.50	0.50	0.60	0.40	0.30		8	18	6	30
NAMA	0.30	0.30	0.40	0.40	0.30	0.30	0.30		19	3	36
!XUN	0.50	0.50	0.50	0.60	0.40	0.30	0.20	0.30		19	15
≠KHOMANI	0.40	0.30	0.40	0.50	0.30	0.30	0.20	0.10	0.20		33
JU/'HOANSI	0.60	0.60	0.60	0.70	0.50	0.40	0.30	0.40	0.20	0.30	
FIN	1.10	1.10	1.10	1.10	1.10	1.20	1.20	1.10	1.20	1	1.30
CEU	1.10	1.10	1.10	1.10	1.10	1.20	1.20	1.10	1.20	1	1.30
GBR	1.10	1.10	1.10	1.10	1.10	1.20	1.20	1.10	1.20	1

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2-Source Data 2. Pairwise F_{ST} for Eurasian populations. We used *smartpca* to compute F_{ST} for each pair of populations, upper left diagonal, together with standard errors computed using a block jacknife. F_{ST} has been multiplied by 1000

	FIN	CEU	GBR	IBS	TSI	GIH	KHV	CDX	CHB	CHS	JPT	PEL
JOLA	160	157	158	151	153	143	182	186	184	186	186	225
WOLLOF	151	148	149	142	144	135	175	178	177	178	178	217
MANDINKAI	154	151	152	145	147	138	177	180	179	180	180	219
MANDINKAII	153	150	151	144	146	137	176	179	178	180	180	218
SEREHULE	150	148	148	142	143	135	174	178	176	178	178	216
FULAI	112	108	109	102	103	102	147	151	149	151	151	188
MALINKE	154	152	152	146	148	138	177	181	180	181	181	220
SERERE	154	151	152	145	147	137	177	180	179	180	180	219
MANJAGO	154	152	152	146	147	138	178	181	180	181	181	220
FULAII	152	149	150	143	145	136	175	179	177	179	179	217
BAMBARA	155	152	153	147	148	139	178	181	180	181	182	220
AKANS	158	156	157	150	152	142	181	184	183	184	184	223
MOSSI	156	154	154	148	149	140	179	182	181	182	182	221
YORUBA	158	155	156	150	151	141	180	183	182	184	184	222
KASEM	157	155	155	149	150	141	180	183	182	183	183	222
NAMKAM	157	155	155	149	150	141	180	183	182	183	183	222
SEMI-BANTU	157	154	155	149	150	140	179	183	181	183	183	221
BANTU	158	155	156	150	151	141	180	183	182	183	183	222
LUHYA	147	144	145	138	140	130	170	174	172	174	174	213
KAMBE	142	140	140	134	135	127	167	170	169	171	171	209
CHONYI	149	146	147	141	142	133	173	176	175	176	176	215
KAUMA	146	143	144	138	139	130	170	174	172	174	174	213
WASAMBAA	142	139	140	133	135	126	167	170	169	170	170	209
GIRIAMA	148	145	146	140	141	132	172	175	174	175	176	214
WABONDEI	145	142	143	137	138	129	169	173	171	173	173	212
MZIGUA	148	145	146	140	141	132	172	175	174	175	175	214
MAASAI	104		101	94	95	92	137	141	139	141	141	179
WOLAYTA	77	73	73	67	67	70	120	124	122	123	123	159
SOMALI	78	73	74	67	67	71	122	126	124	126	126	162
ARI	115	112	112	106	106	103	147	150	149	150	150	189
SUDANESE	152	150	150	144	146	135	174	178	176	178	178	218
GUMUZ	147	144	145	139	140	130	170	173	172	173	174	213
OROMO	66	61	62	55	55	61	113	117	115	117	117	153
AMHARA	59	53	54	47	46	55	111	115	112	114	114	149
AFAR	64	58	59	52	52	60	114	118	116	117	117	153
ANUAK	150	148	149	143	144	134	173	176	175	176	177	216
TIGRAY	57	51	51	44	44	54	109	113	111	113	113	148
MALAWI	158	156	157	150	152	142	181	184	183	184	184	223
SEBANTU	158	156	157	150	152	142	181	184	183	184	184	223
AMAXHOSA	157	155	156	150	151	141	180	183	182	184	184	222
HERERO	145	143	143	138	139	131	172	175	174	175	176	213
KHWE	160	157	158	152	153	144	183	186	185	187	187	225
/GUI//GANA	190	187	188	182	183	173	212	215	214	215	216	250
KARRETJIE	165	162	163	157	159	150	190	193	192	193	194	232
NAMA	142	139	139	134	135	130	172	176	174	176	176	213
!XUN	193	191	192	186	187	178	216	220	218	220	220	250
≠KHOMANI	138	135	136	131	132	126	168	172	170	172	172	210
JU/'HOANSI	226	224	224	219	220	210	249	250	250	250	250	250
FIN		6	7	10	12	36	101	106	102	104	104	127
CEU	0.10		0	2	4	35	109	113	110	112	112	134
GBR	0.10	0		3	4	35	109	113	110	112	112	134
IBS	0.10	0.10	0.10		2	36	108	113	110	112	111	136
TSI	0.10	0.10	0.10	0		34	108	113	110	112	112	137
GIH	0.30	0.30	0.30	0.30	0.30		72	77	75	77	76	114
KHV	0.90	1	1	1	1	0.60		2	7	4	14	111
CDX	0.90	1	1	1	1	0.70	0.10		9	6	17	115
CHB	0.90	1	1	1	1	0.70	0.10	0.10		1	7	107
CHS	0.90	1	1	1	1	0.70	0.10	0.10	0		9	110
JPT	0.90	1	1	1	1	0.70	0.20	0.20	0.10	0.10		107
PEL	1.20	1.30	1.30	1.30	1.30	1.10	1.20	1.20	1.20	1.20	1.20	

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2-Source Data 3. Pairwise TVD for African populations. TVD has been multiplied by 1000

	JOLA	WOLLOF	MANDINKAI	MANDINKAII	SEREHULE	FULAI	MALINKE	SERERE	MANJAGO	FULAII	BAMBARA	AKANS	MOSSI	YORUBA	KASEM	NAMKAM	SEMI-BANTU	BANTU	LUHYA	KAMBE	CHONYI	KAUMA	WASAMBAA	GIRIAMA	WABONDEI	MZIGUA	MAASAI	WOLAYTA	SOMALI	ARI	SUDANESE	GUMUZ	OROMO	AMHARA	AFAR	ANUAK	TIGRAY	MALAWI	SEBANTU	AMAXHOSA	HERERO	KHWE	/GUI//GANA	KARRETJIE	NAMA	IXUN	≠KHOMANI	JU/'HOANSI
JOLA		421	210	364	443	524	487	376	320	447	514	565	558	595	584	583	633	660	702	770	803	800	730	778	727	723	700	670	709	760	612	643	681	695	706	603	691	724	734	739	711	722	757	775	766	783	784	852
WOLLOF	421		230	166	217	415	284	132	223	226	321	402	393	450	439	437	498	538	596	696	749	745	638	715	634	631	613	587	637	701	481	539	603	621	634	469	617	647	665	672	606	641	692	700	686	731	713	843
MANDINKAI	210	230		165	272	429	337	184	165	281	373	445	436	490	477	474	535	571	627	718	765	761	664	734	659	655	640	612	660	717	518	568	626	645	657	508	640	670	683	690	636	662	710	724	711	744	736	846
MANDINKAII	364	166	165		144	412	209	68	146	155	249	341	333	395	384	381	448	491	564	670	727	723	607	689	604	601	582	561	612	691	442	509	577	596	610	428	592	618	637	644	572	613	668	675	664	722	695	839
SEREHULE	443	217	272	144		404	138	171	236	84	182	285	278	344	332	329	401	459	539	649	709	705	583	669	580	577	556	538	594	681	412	487	556	578	593	397	572	595	614	621	545	590	649	653	653	719	684	836
FULAI	524	415	429	412	404		415	420	436	397	434	498	490	543	530	528	590	626	668	751	796	792	700	767	698	695	666	635	677	729	562	598	647	662	674	552	658	708	720	726	668	702	743	742	733	776	752	851
MALINKE	487	284	337	209	138	415		234	294	86	74	207	199	267	258	254	346	409	502	622	683	679	549	642	544	540	531	525	589	682	388	477	547	573	590	372	567	559	578	586	512	554	623	638	652	715	683	831
SERERE	376	132	184	68	171	420	234		159	180	269	353	345	403	393	390	453	494	566	670	727	723	609	690	605	602	585	563	615	693	445	512	580	599	612	431	594	619	637	645	574	614	668	676	666	723	697	839
MANJAGO	320	223	165	146	236	436	294	159		241	324	392	383	439	429	427	484	524	585	687	740	736	627	705	624	620	599	575	626	695	464	526	591	610	623	452	606	636	654	661	590	631	683	686	673	726	702	841
FULAII	447	226	281	155	84	397	86	180	241		127	238	230	298	287	284	367	430	518	634	695	691	563	655	559	556	546	533	595	685	401	483	553	578	594	386	572	574	593	600	528	569	634	645	655	716	686	833
BAMBARA	514	321	373	249	182	434	74	269	324	127		157	150	217	209	205	318	388	484	608	670	666	534	628	529	524	525	525	592	686	382	479	548	577	594	366	570	540	561	570	496	538	611	636	654	713	685	830
AKANS	565	402	445	341	285	498	207	353	392	238	157		77	116	157	149	250	334	441	571	637	633	497	591	488	480	501	526	597	691	374	478	552	581	598	352	575	491	518	527	446	494	579	632	653	706	684	826
MOSSI	558	393	436	333	278	490	199	345	383	230	150	77		150	140	132	276	359	464	590	652	647	517	609	509	502	518	531	600	695	381	484	556	585	602	362	579	514	539	547	471	515	594	638	659	711	690	830
YRI	595	450	490	395	344	543	267	403	439	298	217	116	150		190	184	190	273	404	536	607	604	461	556	452	442	482	532	602	697	376	479	557	587	604	358	581	447	477	486	401	462	567	634	654	701	684	824
KASEM	584	439	477	384	332	530	258	393	429	287	209	157	140	190		9	310	387	498	615	670	666	546	630	539	531	545	546	613	706	403	500	569	597	614	389	590	538	562	570	502	540	613	650	672	715	701	832
NAMKAM	583	437	474	381	329	528	254	390	427	284	205	149	132	184	9		304	383	494	612	667	663	543	627	535	527	542	544	611	706	400	498	568	597	613	386	590	534	558	567	497	536	611	649	671	715	700	832
SEMI-BANTU	633	498	535	448	401	590	346	453	484	367	318	250	276	190	310	304		126	293	449	552	551	345	474	335	325	453	527	597	693	373	475	554	584	601	357	578	327	364	374	338	439	549	615	632	688	663	813
BANTU	660	538	571	491	459	626	409	494	524	430	388	334	359	273	387	383	126		287	443	547	545	331	466	320	310	456	536	606	700	401	499	563	593	610	388	587	307	347	357	338	436	545	617	634	685	665	810
LUHYA	702	596	627	564	539	668	502	566	585	518	484	441	464	404	498	494	293	287		413	535	534	288	449	282	274	405	510	560	660	404	483	528	554	565	391	549	301	341	351	364	448	555	617	633	694	666	818
KAMBE	770	696	718	670	649	751	622	670	687	634	608	571	590	536	615	612	449	443	413		222	185	339	93	336	332	555	641	671	726	589	630	654	669	676	583	666	354	390	406	494	540	585	642	667	708	691	829
CHONYI	803	749	765	727	709	796	683	727	740	695	670	637	652	607	670	667	552	547	535	222		258	464	265	461	457	630	702	728	777	669	701	714	728	733	662	725	460	500	514	583	612	638	693	725	738	740	837
KAUMA	800	745	761	723	705	792	679	723	736	691	666	633	647	604	666	663	551	545	534	185	258		463	215	460	456	625	696	723	772	664	695	709	723	728	657	720	461	500	514	581	608	634	689	721	735	736	837
WASAMBAA	730	638	664	607	583	700	549	609	627	563	534	497	517	461	546	543	345	331	288	339	464	463		378	48	98	456	558	601	671	479	543	576	596	606	470	593	176	256	286	402	457	545	607	623	692	657	815
GIRIAMA	778	715	734	689	669	767	642	690	705	655	628	591	609	556	630	627	474	466	449	93	265	215	378		375	371	582	663	692	746	616	652	676	691	697	610	688	381	419	433	519	556	598	660	686	713	709	831
WABONDEI	727	634	659	604	580	698	544	605	624	559	529	488	509	452	539	535	335	320	282	336	461	460	48	375		64	463	566	610	680	480	549	584	605	615	471	602	147	237	269	391	445	540	604	620	690	653	814
MZIGUA	723	631	655	601	577	695	540	602	620	556	524	480	502	442	531	527	325	310	274	332	457	456	98	371	64		467	570	615	687	478	551	588	609	619	467	606	125	226	258	384	441	539	606	622	689	655	813
MAASAI	700	613	640	582	556	666	531	585	599	546	525	501	518	482	545	542	453	456	405	555	630	625	456	582	463	467		359	409	561	340	376	373	400	415	310	396	497	514	521	447	514	582	569	593	699	610	816
WOLAYTA	670	587	612	561	538	635	525	563	575	533	525	526	531	532	546	544	527	536	510	641	702	696	558	663	566	570	359		239	479	392	350	75	137	176	367	134	599	601	608	527	580	628	565	587	705	581	816
SOMALI	709	637	660	612	594	677	589	615	626	595	592	597	600	602	613	611	597	606	560	671	728	723	601	692	610	615	409	239		549	472	433	224	235	215	449	229	646	651	657	581	631	665	597	610	727	600	831
ARI	760	701	717	691	681	729	682	693	695	685	686	691	695	697	706	706	693	700	660	726	777	772	671	746	680	687	561	479	549		589	545	517	538	548	572	542	720	719	723	673	705	721	675	669	751	684	837
SUDANESE	612	481	518	442	412	562	388	445	464	401	382	374	381	376	403	400	373	401	404	589	669	664	479	616	480	478	340	392	472	589		313	419	458	477	84	452	506	537	547	465	527	601	605	623	704	648	820
GUMUZ	643	539	568	509	487	598	477	512	526	483	479	478	484	479	500	498	475	499	483	630	701	695	543	652	549	551	376	350	433	545	313		374	409	436	287	407	579	590	597	516	569	632	596	612	705	633	822
OROMO	681	603	626	577	556	647	547	580	591	553	548	552	556	557	569	568	554	563	528	654	714	709	576	676	584	588	373	75	224	517	419	374		70	124	395	72	618	622	628	549	601	644	578	596	713	589	822
AMHARA	695	621	645	596	578	662	573	599	610	578	577	581	585	587	597	597	584	593	554	669	728	723	596	691	605	609	400	137	235	538	458	409	70		124	435	50	640	644	650	573	624	659	591	604	722	595	827
AFAR	706	634	657	610	593	674	590	612	623	594	594	598	602	604	614	613	601	610	565	676	733	728	606	697	615	619	415	176	215	548	477	436	124	124		454	117	651	655	661	585	635	668	599	610	727	600	830
ANUAK	603	469	508	428	397	552	372	431	452	386	366	352	362	358	389	386	357	388	391	583	662	657	470	610	471	467	310	367	449	572	84	287	395	435	454		427	496	527	537	452	517	592	594	617	701	639	817
TIGRAY	691	617	640	592	572	658	567	594	606	572	570	575	579	581	590	590	578	587	549	666	725	720	593	688	602	606	396	134	229	542	452	407	72	50	117	427		636	641	647	569	620	657	589	604	721	595	827
MALAWI	724	647	670	618	595	708	559	619	636	574	540	491	514	447	538	534	327	307	301	354	460	461	176	381	147	125	497	599	646	720	506	579	618	640	651	496	636		194	233	378	448	529	604	633	681	660	805
SEBANTU	734	665	683	637	614	720	578	637	654	593	561	518	539	477	562	558	364	347	341	390	500	500	256	419	237	226	514	601	651	719	537	590	622	644	655	527	641	194		44	343	377	409	468	514	586	541	714
AMAXHOSA	739	672	690	644	621	726	586	645	661	600	570	527	547	486	570	567	374	357	351	406	514	514	286	433	269	258	521	608	657	723	547	597	628	650	661	537	647	233	44		342	379	395	453	499	577	525	709
SWBANTU	711	606	636	572	545	668	512	574	590	528	496	446	471	401	502	497	338	338	364	494	583	581	402	519	391	384	447	527	581	673	465	516	549	573	585	452	569	378	343	342		355	447	449	430	593	492	723
KHWE	722	641	662	613	590	702	554	614	631	569	538	494	515	462	540	536	439	436	448	540	612	608	457	556	445	441	514	580	631	705	527	569	601	624	635	517	620	448	377	379	355		367	465	451	469	507	607
/GUI//GANA	757	692	710	668	649	743	623	668	683	634	611	579	594	567	613	611	549	545	555	585	638	634	545	598	540	539	582	628	665	721	601	632	644	659	668	592	657	529	409	395	447	367		350	357	473	401	603
KARRETJIE	775	700	724	675	653	742	638	676	686	645	636	632	638	634	650	649	615	617	617	642	693	689	607	660	604	606	569	565	597	675	605	596	578	591	599	594	589	604	468	453	449	465	350		202	519	186	643
NAMA	766	686	711	664	653	733	652	666	673	655	654	653	659	654	672	671	632	634	633	667	725	721	623	686	620	622	593	587	610	669	623	612	596	604	610	617	604	633	514	499	430	451	357	202		483	129	615
XUN	783	731	744	722	719	776	715	723	726	716	713	706	711	701	715	715	688	685	694	708	738	735	692	713	690	689	699	705	727	751	704	705	713	722	727	701	721	681	586	577	593	469	473	519	483		521	414
≠KHOMANI	784	713	736	695	684	752	683	697	702	686	685	684	690	684	701	700	663	665	666	691	740	736	657	709	653	655	610	581	600	684	648	633	589	595	600	639	595	660	541	525	492	507	401	186	129	521		640
JU/'HOANSI	852	843	846	839	836	851	831	839	841	833	830	826	830	824	832	832	813	810	818	829	837	837	815	831	814	813	816	816	831	837	820	822	822	827	830	817	827	805	714	709	723	607	603	643	615	414	640	
FIN	886	859	869	854	847	860	851	856	854	852	854	859	860	864	868	868	863	869	855	867	891	889	855	880	857	861	790	664	706	813	823	798	635	610	630	816	595	882	877	879	825	870	874	781	794	884	757	894
CEU	857	827	838	821	814	829	817	823	822	819	820	825	827	830	836	836	828	833	814	837	865	862	815	851	820	825	741	608	659	769	776	750	578	552	577	768	535	849	845	847	790	835	841	747	760	851	724	866
GBR	861	832	842	826	819	834	822	828	827	823	825	830	832	836	841	840	833	839	820	842	869	866	820	856	825	830	748	616	666	776	783	757	587	562	584	775	546	854	850	852	795	841	846	753	765	856	729	870
IBS	827	797	808	790	783	800	786	792	792	788	789	794	797	800	806	805	798	807	785	812	848	845	786	828	793	799	702	567	616	731	733	708	536	511	534	726	493	825	821	822	765	810	816	723	734	829	700	856
TSI	835	805	816	798	791	808	793	800	800	795	797	802	804	808	813	813	806	813	791	817	850	846	791	832	799	804	705	572	621	734	739	713	541	515	540	731	497	831	827	828	772	816	822	729	740	834	705	856
GIH	841	812	822	805	798	814	801	807	807	803	805	810	812	816	821	821	814	820	799	820	850	847	799	834	804	810	729	597	646	754	763	738	566	540	564	756	522	835	831	832	776	820	826	733	745	837	709	857
KHV	871	846	854	842	836	844	840	843	841	841	843	848	848	853	855	854	852	857	847	855	876	873	847	867	848	852	782	716	736	805	815	790	711	707	713	808	704	870	865	867	813	859	862	769	785	871	757	881
CDX	878	856	863	852	846	853	850	853	851	851	853	858	859	863	864	864	862	867	858	866	884	882	857	876	859	863	793	738	759	815	825	800	736	735	740	818	732	879	874	876	822	868	871	783	800	879	779	890
CHB	869	844	852	840	834	842	838	841	839	839	841	846	847	851	853	853	850	856	845	854	875	872	845	865	846	850	780	705	728	803	813	788	701	697	703	806	695	869	864	865	811	857	860	767	783	869	751	880
CHS	871	846	854	842	837	844	840	844	842	842	843	848	849	853	855	855	853	858	848	856	877	874	847	867	849	853	782	720	740	805	816	790	715	714	719	808	711	871	866	867	813	859	863	770	787	871	761	882
JPT	873	850	858	846	840	847	844	847	845	845	847	852	852	857	859	858	856	861	852	860	880	877	851	870	853	856	786	720	741	809	819	794	716	714	719	812	712	874	869	871	817	863	866	773	790	874	763	885
PELII	861	833	843	828	822	834	826	830	829	827	829	835	836	840	843	843	840	846	831	843	867	864	830	855	833	837	765	633	682	789	799	774	615	600	615	792	591	859	854	856	801	846	851	757	771	860	734	874

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2-Source Data 4. Pairwise TVD for Eurasian populations. TVD has been multiplied by 1000

	FIN	CEU	GBR	IBS	TSI	GIH	KHV	CDX	CHB	CHS	JPT	PEL
JOLA	886	857	861	827	835	841	871	878	869	871	873	861
WOLLOF	859	827	832	797	805	812	846	856	844	846	850	833
MANDINKAI	869	838	842	808	816	822	854	863	852	854	858	843
MANDINKAII	854	821	826	790	798	805	842	852	840	842	846	828
SEREHULE	847	814	819	783	791	798	836	846	834	837	840	822
FULAI	860	829	834	800	808	814	844	853	842	844	847	834
MALINKE	851	817	822	786	793	801	840	850	838	840	844	826
SERERE	856	823	828	792	800	807	843	853	841	844	847	830
MANJAGO	854	822	827	792	800	807	841	851	839	842	845	829
FULAII	852	819	823	788	795	803	841	851	839	842	845	827
BAMBARA	854	820	825	789	797	805	843	853	841	843	847	829
AKANS	859	825	830	794	802	810	848	858	846	848	852	835
MOSSI	860	827	832	797	804	812	848	859	847	849	852	836
YORUBA	864	830	836	800	808	816	853	863	851	853	857	840
KASEM	868	836	841	806	813	821	855	864	853	855	859	843
NAMKAM	868	836	840	805	813	821	854	864	853	855	858	843
SEMI-BANTU	863	828	833	798	806	814	852	862	850	853	856	840
BANTU	869	833	839	807	813	820	857	867	856	858	861	846
LUHYA	855	814	820	785	791	799	847	858	845	848	852	831
KAMBE	867	837	842	812	817	820	855	866	854	856	860	843
CHONYI	891	865	869	848	850	850	876	884	875	877	880	867
KAUMA	889	862	866	845	846	847	873	882	872	874	877	864
WASAMBAA	855	815	820	786	791	799	847	857	845	847	851	830
GIRIAMA	880	851	856	828	832	834	867	876	865	867	870	855
WABONDEI	857	820	825	793	799	804	848	859	846	849	853	833
MZIGUA	861	825	830	799	804	810	852	863	850	853	856	837
MAASAI	790	741	748	702	705	729	782	793	780	782	786	765
WOLAYTA	664	608	616	567	572	597	716	738	705	720	720	633
SOMALI	706	659	666	616	621	646	736	759	728	740	741	682
ARI	813	769	776	731	734	754	805	815	803	805	809	789
SUDANESE	823	776	783	733	739	763	815	825	813	816	819	799
GUMUZ	798	750	757	708	713	738	790	800	788	790	794	774
OROMO	635	578	587	536	541	566	711	736	701	715	716	615
AMHARA	610	552	562	511	515	540	707	735	697	714	714	600
AFAR	630	577	584	534	540	564	713	740	703	719	719	615
ANUAK	816	768	775	726	731	756	808	818	806	808	812	792
TIGRAY	595	535	546	493	497	522	704	732	695	711	712	591
MALAWI	882	849	854	825	831	835	870	879	869	871	874	859
SEBANTU	877	845	850	821	827	831	865	874	864	866	869	854
AMAXHOSA	879	847	852	822	828	832	867	876	865	867	871	856
SWBANTU	825	790	795	765	772	776	813	822	811	813	817	801
KHWE	870	835	841	810	816	820	859	868	857	859	863	846
/GUI//GANA	874	841	846	816	822	826	862	871	860	863	866	851
KARRETJIE	781	747	753	723	729	733	769	783	767	770	773	757
NAMA	794	760	765	734	740	745	785	800	783	787	790	771
!XUN	884	851	856	829	834	837	871	879	869	871	874	860
≠KHOMANI	757	724	729	700	705	709	757	779	751	761	763	734
JU/'HOANSI	894	866	870	856	856	857	881	890	880	882	885	874
FIN		341	354	378	378	523	733	754	725	740	738	620
CEU	341		50	107	136	452	704	732	695	711	711	573
GBR	354	50		126	153	463	709	735	700	716	716	579
IBS	378	107	126		87	431	702	730	693	709	710	568
TSI	378	136	153	87		416	692	720	682	698	699	558
GIH	523	452	463	431	416		617	645	608	624	625	483
KHV	733	704	709	702	692	617		146	157	137	253	488
CDX	754	732	735	730	720	645	146		217	191	300	516
CHB	725	695	700	693	682	608	157	217		34	193	474
CHS	740	711	716	709	698	624	137	191	34		214	495
JPT	738	711	716	710	699	625	253	300	193	214		491
PELII	620	573	579	568	558	483	488	516	474	495	491	

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3-Source Data 1. The evidence for multiple waves of admixture in African populations using MALDER and the HAPMAP recombination map. For each event in each ethnic group we show the largest inferred amplitude and date of an admixture event involving two reference populations (Pop1 and Pop2). We additionally provide the ancestry region identity of the two main reference populations, together with Z scores for curve comparisons between this best curve and those containing populations from different ancestry regions. We use a cut-off of $Z < 2$ to decide whether sources from multiple ancestries best describe the admixture source.

Ethnic Group	amp	Date	Date (CI)	Pop1	Pop1Anc	Pop1Anc-West Africa NC(Z)	Pop1Anc-Central West Africa NC(Z)	Pop1Anc-East Africa NC(Z)	Pop1Anc-Nilo-Saharan(Z)	Pop1Anc-Afroasiatic(Z)	Pop1Anc-South Africa NC(Z)	Pop1Anc-Khoesan(Z)	Pop1Anc-Eurasia(Z)	Pop1Anc-(Z)	Pop2	Pop2Anc	Pop2Anc-West Africa NC(Z)	Pop2Anc-Central West Africa NC(Z)	Pop2Anc-East Africa NC(Z)	Pop2Anc-Nilo-Saharan(Z)	Pop2Anc-Afroasiatic(Z)	Pop2Anc-South Africa NC(Z)	Pop2Anc-Khoesan(Z)	Pop2Anc-Eurasia(Z)	Pop2Anc-(Z)
JOLA	7.3e-05	148	31	JU/'HOANSI	Khoesan	0.57	0.43	0.72	0.69	2.25	0.22	0.00		0.22	GBR	Eurasia				3.27	2.39	3.33		0.00	1.89
WOLLOF	1e-04	88	22	AKANS	Central West Africa NC	0.00	0.00	0.63	1.11	3.42	0.35	0.17		0.00	GBR	Eurasia	6.51		6.43	5.74	3.57		6.16	0.00	3.21
WOLLOF	2.3e-05	10	3	NAMKAM	Central West Africa NC	-0.22	0.00	0.79	1.35	3.39	0.23	0.18		-0.22	TSI	Eurasia	6.66		6.66	6.01	3.81		6.45	0.00	3.81
MANDINKAI																									
MANDINKAII	3.7e-05	19	3	JU/'HOANSI	Khoesan	0.39	0.06	0.78	1.04	4.21	0.33	0.00		0.06	IBS	Eurasia	7.56	7.66	7.95	7.53	4.68	7.80		0.00	4.67
SEREHULE	6.3e-05	23	5	AKANS	Central West Africa NC	0.43	0.00	0.52	1.01	3.47	0.23	0.20		0.20	IBS	Eurasia	6.17		6.22	5.52	3.43	6.26	5.88	0.00	3.43
FULAI	0.00044	60	8	JOLA	West Africa NC	0.00	0.10	1.02	1.15	4.59	0.40	0.79		0.10	CEU	Eurasia			9.03	8.17	5.28		8.99	0.00	5.28
FULAI	0.00016	11	2	AKANS	Central West Africa NC	-0.19	0.00	0.87	1.17	4.72	0.26	0.72		-0.19	IBS	Eurasia			9.82	8.91	5.78		9.81	0.00	5.78
MALINKE	8.9e-05	108	24	YORUBA	Central West Africa NC	0.03	0.00	0.27	0.90	1.55	0.24	0.11		0.03	PEL	Eurasia	2.47		2.55	2.27	1.43		2.03	0.00	1.39
MALINKE	1.9e-05	8	2	BANTU	Central West Africa NC	-0.27	0.00	-0.18	0.30	1.24	0.10	-0.44		0.10	TSI	Eurasia	2.63		2.68	2.32	1.52		1.79	0.00	1.74
SERERE	2.5e-05	14	3	BANTU	Central West Africa NC	0.80	0.00	1.00	1.19	4.83	0.53	0.69		0.53	GBR	Eurasia	8.39		8.46	7.48	4.64	8.42	7.96	0.00	4.64
MANJAGO	3.4e-05	3	1	AKANS	Central West Africa NC	0.24	0.00	0.92	1.73	6.49	0.25	0.86		0.24	GBR	Eurasia	12.37		13.06	12.08	8.21	12.96	12.40	0.00	8.21
FULAII	2e-05	5	1	JU/'HOANSI	Khoesan	0.09	0.20	1.16	2.17	5.69	0.36	0.00		0.09	GBR	Eurasia	10.33	10.21	10.51	9.27	5.86	10.37		0.00	5.86
FULAII	7.6e-05	67	10	SEMI-BANTU	Central West Africa NC	0.18	0.00	0.78	1.74	5.30	-0.02	-0.01		0.18	TSI	Eurasia	10.01		9.96	8.85	5.56		9.96	0.00	5.56
BAMBARA	2.2e-05	15	4	/GUI//GANA	Khoesan	0.62	0.31	0.69	0.67	2.94	0.25	0.00		0.25	IBS	Eurasia	5.62		5.63	4.97	3.35	5.63		0.00	3.35
AKANS																									
MOSSI	0.00056	221	56	MALAWI	South Africa NC	0.05	0.15	0.06	0.72	0.88	0.00	0.65		0.05	GBR	Eurasia	0.83			0.92	0.58			0.00	0.35
MOSSI	5.2e-06	8	2	/GUI//GANA	Khoesan	0.16	-0.62	-0.50	0.85	1.34	-0.43	0.00		-0.43	CEU	Eurasia					1.13			0.00	1.13
YORUBA																									
KASEM																									
NAMKAM																									
SEMI-BANTU																									
BANTU	2.3e-05	56	12	MOSSI	Central West Africa NC	0.00	0.00	0.20	0.77	0.25	1.06			0.00	JU/'HOANSI	Khoesan						0.93	0.00		0.81
LUHYA	7e-05	25	2	AKANS	Central West Africa NC	0.97	0.00		7.52	11.82	1.74	2.63		0.97	TSI	Eurasia	14.05			9.63	4.45		12.19	0.00	4.45
KAMBE	0.00014	15	2	YORUBA	Central West Africa NC	0.23	0.00		2.10	6.27	1.21	1.21		0.23	TSI	Eurasia	10.74			9.31	5.67	10.67	9.83	0.00	5.67
CHONYI	0.00014	27	2	YORUBA	Central West Africa NC	0.17	0.00		3.28	9.02	1.89	1.65		0.17	TSI	Eurasia	14.04			13.89	7.89		13.04	0.00	7.89
KAUMA	0.00015	26	2	SEMI-BANTU	Central West Africa NC	0.21	0.00		3.09	8.19		1.39		0.21	TSI	Eurasia	13.20			12.05	6.52		11.74	0.00	6.52
WASAMBAA	0.00012	20	3	NAMKAM	Central West Africa NC	0.14	0.00		2.40	6.34	1.14	1.44		0.14	TSI	Eurasia	9.03			7.79	3.78	8.96	7.50	0.00	3.78
GIRIAMA	0.00014	28	2	YORUBA	Central West Africa NC	0.63	0.00		4.08	9.30		2.09		0.63	TSI	Eurasia	13.07			12.71	6.51		12.43	0.00	6.51
WABONDEI	1e-04	23	2	MOSSI	Central West Africa NC	0.16	0.00		2.94	8.04	1.41	1.97		0.16	TSI	Eurasia	11.24			10.52	5.30	12.03	9.59	0.00	5.30
MZIGUA	0.00011	32	4	YORUBA	Central West Africa NC	0.18	0.00		1.98	5.11		1.12		0.18	TSI	Eurasia	7.41			7.11	3.66		6.68	0.00	3.66
MAASAI	0.00037	74	15	SUDANESE	Nilo-Saharan	0.23	0.08	0.50	0.00	3.35	0.09	0.54		0.08	TSI	Eurasia	5.09	5.66	5.58		2.94	5.62	5.30	0.00	2.79
MAASAI	8e-05	10	2	SEMI-BANTU	Central West Africa NC	0.44	0.00	0.47	-0.08	4.06	0.27	0.65		0.27	TSI	Eurasia	6.94			7.36	3.72	7.35	7.41	0.00	3.72
SUDANESE	1.7e-05	6	2	JU/'HOANSI	Khoesan	0.42	0.22	0.83	1.39	4.11	0.55	0.00		0.22	GBR	Eurasia	5.28	5.55		4.38	2.62	5.81		0.00	2.62
GUMUZ	1.9e-05	5	1	SUDANESE	Nilo-Saharan	0.53	0.27	0.62	0.00	2.23	0.59	0.27		0.27	GBR	Eurasia	5.62	5.88	5.87		3.68	5.51	5.33	0.00	3.58
GUMUZ	7.9e-05	103	25	SUDANESE	Nilo-Saharan	0.66	0.35	1.09	0.00	2.14	0.54	0.65		0.65	TSI	Eurasia	5.46	5.73	5.70		3.63	5.36	5.37	0.00	3.77
ANUAK	3.3e-05	47	12	SUDANESE	Nilo-Saharan	1.62	1.59	1.71	0.00	1.47	2.40		0.81	0.21	JU/'HOANSI	Khoesan	1.28	1.27	1.23		2.23	0.55	0.00	1.40	0.21
AFAR	0.00039	45	10	SUDANESE	Nilo-Saharan	0.39	0.25	0.56	0.00	2.08	0.33	0.39		0.25	TSI	Eurasia	5.03	4.99			3.78	5.03	5.09	0.00	3.68
OROMO	0.00066	84	8	JU/'HOANSI	Khoesan	0.79	0.58	0.75	0.33	2.07	0.61	0.00		0.33	TSI	Eurasia	6.72	7.50	7.46	7.05	4.62	7.52		0.00	4.62
OROMO	4.9e-05	7	2	GUMUZ	Nilo-Saharan	-0.07	-0.20	0.17	0.00	1.33	-0.15	-0.74		-0.74	TSI	Eurasia	6.57		7.13		5.73		6.47	0.00	3.91
SOMALI	0.00045	70	11	SUDANESE	Nilo-Saharan	0.30	0.09	0.79	0.00	2.74	0.29	0.66		0.09	TSI	Eurasia	5.50	5.80	5.74		3.96	5.78		0.00	3.87
WOLAYTA	0.00056	56	9	JU/'HOANSI	Khoesan	0.72	0.48	0.82	0.39	1.20	0.37	0.00		0.37	TSI	Eurasia	4.53	5.16	5.09	4.58	2.87	5.11		0.00	2.87
ARI	0.00045	97	10	JU/'HOANSI	Khoesan	0.95	0.72	1.09	0.70	3.38	0.61	0.00		0.61	TSI	Eurasia				5.11	2.82	6.01		0.00	2.82
AMHARA	0.00052	64	9	SUDANESE	Nilo-Saharan	0.47	0.25	0.87	0.00	2.06	0.40	0.28		0.25	GBR	Eurasia	6.68	6.77	6.85		4.93	6.83		0.00	4.37
TIGRAY	0.00067	75	8	SUDANESE	Nilo-Saharan	0.61	0.42	0.78	0.00	2.47	0.46	0.43		0.42	TSI	Eurasia	7.97				5.92	8.29		0.00	5.87
MALAWI	3.9e-05	41	6	YORUBA	Central West Africa NC	0.34	0.00		1.33	1.17			0.25	0.25	JU/'HOANSI	Khoesan				4.68	4.19		0.00	2.74	2.74
SEBANTU																									
AMAXHOSA	5e-05	17	3	MANDINKAI	West Africa NC	0.00	0.15		1.78	5.18	1.79			0.15	GBR	Eurasia		6.93		6.36	3.88			0.00	3.82
HERERO	0.00025	2	0	JU/'HOANSI	Khoesan	1.50	1.06	1.88	2.81	6.67	1.00	0.00		1.00	GBR	Eurasia	12.04	12.06	12.16	11.40	8.11	12.58		0.00	8.11
KHWE	0.00017	85	25	SEMI-BANTU	Central West Africa NC	0.64	0.00	0.12	0.89	1.60				0.12	PEL	Eurasia	2.35		2.53	1.95	0.98			0.00	0.98
KHWE	3.9e-05	17	5	BANTU	Central West Africa NC	0.00	0.00	0.11	0.46	1.15				0.00	TSI	Eurasia	2.01		2.19	2.04	1.53			0.00	1.53
/GUI//GANA	3.8e-05	20	4	JOLA	West Africa NC	0.00	0.22		1.60	3.27				0.22	TSI	Eurasia		7.15		6.02	3.45			0.00	3.01
KARRETJIE	0.00047	5	0	BANTU	Central West Africa NC	0.74	0.00		2.33	7.00				0.74	GBR	Eurasia	19.76			19.43	13.76			0.00	13.76
NAMA	0.00027	4	0	BANTU	Central West Africa NC	0.53	0.00	0.73	1.53	5.33				0.53	GBR	Eurasia	14.83		14.81	14.59	10.23			0.00	10.23
NAMA	0.00024	64	10	BANTU	Central West Africa NC	0.82	0.00	0.85	1.65	8.99				0.82	GBR	Eurasia	14.83		14.83	14.59	10.57			0.00	10.57
!XUN	0.00011	50	6	MALAWI	South Africa NC	0.30	0.05	0.66	1.53	2.99	0.00			0.05	TSI	Eurasia	5.36		5.99	4.90	2.65			0.00	2.65
=KHOMANI	5e-04	5	0	BANTU	Central West Africa NC	0.73	0.00		2.06	14.81				0.73	GBR	Eurasia	19.98			19.70	14.19			0.00	14.19
JU/'HOANSI	9e-05	41	6	AKANS	Central West Africa NC	0.21	0.00	0.59	1.76	4.73	0.06			0.06	TSI	Eurasia	8.70		9.51	7.87	4.17	9.44		0.00	3.68

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3-Source Data 2. The evidence for multiple waves of admixture in African populations using MALDER and the African recombination map. Columns as in Figure 3-Source Data 1

Ethnic Group	amp	Date	Date (CI)	Pop1	Pop1Anc	Pop1Anc-West Africa NC(Z)	Pop1Anc-Central West Africa NC(Z)	Pop1Anc-East Africa NC(Z)	Pop1Anc-Nilo-Saharan(Z)	Pop1Anc-Afroasiatic(Z)	Pop1Anc-South Africa NC(Z)	Pop1Anc-Khoesan(Z)	Pop1Anc-Eurasia(Z)	Pop1Anc-(Z)	Pop2	Pop2Anc	Pop2Anc-West Africa NC(Z)	Pop2Anc-Central West Africa NC(Z)	Pop2Anc-East Africa NC(Z)	Pop2Anc-Nilo-Saharan(Z)	Pop2Anc-Afroasiatic(Z)	Pop2Anc-South Africa NC(Z)	Pop2Anc-Khoesan(Z)	Pop2Anc-Eurasia(Z)	Pop2Anc-(Z)
JOLA	2.4e-05	121	36	/GUI//GANA	Khoesan	0.39	0.32	0.44	0.34		0.32	0.00		0.32	CDX	Eurasia	2.53			2.36	1.54	2.61		0.00	1.54
WOLLOF	0.00017	153	38	MANDINKAII	West Africa NC	0.00	0.80	1.00	1.25	2.00	0.83	1.03		0.80	TSI	Eurasia		2.82	2.81	2.68	1.57	2.79	2.68	0.00	1.57
WOLLOF	2.4e-05	16	5	JOLA	West Africa NC	0.00	0.87	1.10	1.41	2.23	1.01	1.15		0.87	TSI	Eurasia		3.18	3.18	3.17	1.67	3.12	3.00	0.00	1.67
MANDINKAI																									
MANDINKAII	3.7e-05	29	5	SEMI-BANTU	Central West Africa NC	0.53	0.00	0.93	1.12	4.07	0.39	0.12		0.12	IBS	Eurasia	6.33		7.70	7.02	4.51	7.79	7.22	0.00	4.51
SEREHULE	5.7e-05	32	7	BANTU	Central West Africa NC	0.60	0.00	0.61	0.97	3.71	0.40	0.22		0.22	IBS	Eurasia	6.60		6.61	5.85	3.80		6.26	0.00	3.80
FULAI	0.00041	86	11	SEMI-BANTU	Central West Africa NC	0.08	0.00	0.89	1.28	5.26	0.21	0.81		0.08	IBS	Eurasia	11.60		11.46	10.28	6.63		11.57	0.00	6.45
FULAI	0.00015	16	2	AKANS	Central West Africa NC	-0.15	0.00	0.75	0.87	4.09	0.04	0.43		-0.15	TSI	Eurasia	8.62		8.53	7.68	5.13		8.59	0.00	5.13
MALINKE	0.00021	200	45	BAMBARA	West Africa NC	0.00	0.58	0.63	0.80	0.99	0.39	0.57		0.39	JPT	Eurasia		1.21	1.21	1.15	0.78	1.15		0.00	0.78
MALINKE	1.9e-05	13	2	SEMI-BANTU	Central West Africa NC	-0.24	0.00	0.03	0.56	1.26	-0.15	-0.22		-0.15	TSI	Eurasia	2.06		2.17	2.04	0.99		0.88	0.00	1.25
SERERE	2.5e-05	22	4	BANTU	Central West Africa NC	0.60	0.00	1.08	1.38	4.94	0.83	0.92		0.60	GBR	Eurasia	6.65		8.32	7.37	4.63	8.29	7.90	0.00	4.63
MANJAGO	3e-05	4	1	AKANS	Central West Africa NC	0.13	0.00	0.47	0.92	3.36	0.11	0.44		0.11	GBR	Eurasia	6.41		6.76	6.23	4.21	6.72	6.41	0.00	4.21
FULAII	7.1e-05	91	22	JOLA	West Africa NC	0.00	0.11	0.52	0.68	2.07	0.27	0.53		0.11	TSI	Eurasia		3.16	3.08	2.88	1.97	3.12	2.89	0.00	1.94
FULAII	1.8e-05	7	2	JU/'HOANSI	Khoesan	-0.93	-0.96	-0.41	-0.38	2.79	-0.84	0.00		-0.93	GBR	Eurasia	6.08	5.71		5.84	4.11	5.23		0.00	4.11
BAMBARA	2.1e-05	22	6	/GUI//GANA	Khoesan	0.45	0.15	0.56	0.60	2.78	0.01	0.00		0.01	IBS	Eurasia	5.36				3.17			0.00	3.17
AKANS																									
MOSSI	7.3e-06	17	5	/GUI//GANA	Khoesan	0.17	0.24	0.29	0.73	4.92	0.15	0.00		0.15	CEU	Eurasia	5.47			5.27	3.13			0.00	3.13
YORUBA																									
KASEM	1.9e-06	25	7	GUMUZ	Nilo-Saharan				0.00					4.91	JU/'HOANSI	Khoesan	3.26				2.83		0.00		2.10
NAMKAM																									
SEMI-BANTU																									
BANTU	2.3e-05	84	23	FULAI	West Africa NC	0.00	0.13		0.96	0.42				0.13	JU/'HOANSI	Khoesan		3.04	2.93			2.31	0.00		0.63
LUHYA	8.7e-05	79	18	MALAWI	South Africa NC	0.89	0.18	0.61	1.41	4.02	0.00	0.65		0.18	TSI	Eurasia	4.30	3.87		3.29	1.57		3.09	0.00	1.57
LUHYA	2.2e-05	18	6	YORUBA	Central West Africa NC	-0.18	0.00	-0.35	0.98	4.25	-1.00	0.85		-1.00	TSI	Eurasia	5.93			3.98	2.02	4.23	4.83	0.00	2.02
KAMBE	0.00013	22	3	YORUBA	Central West Africa NC	0.08	0.00		1.77	5.34	0.24	0.69		0.08	TSI	Eurasia	8.56			7.99	4.93		8.52	0.00	4.93
CHONYI	0.00013	40	3	YORUBA	Central West Africa NC	0.02	0.00		3.09	9.17	0.31	1.66		0.02	TSI	Eurasia	14.27			13.25	7.94		13.21	0.00	7.94
KAUMA	0.00014	40	3	SEMI-BANTU	Central West Africa NC	0.33	0.00		3.02	7.44	0.21	0.75		0.21	TSI	Eurasia	11.04			10.01	5.92	11.85	10.72	0.00	5.92
WASAMBAA	0.00013	58	14	BANTU	Central West Africa NC	0.12	0.00		1.26	3.93	0.03	0.46		0.03	TSI	Eurasia	5.25			4.39	2.47	4.90	4.54	0.00	2.29
WASAMBAA	2.5e-05	11	3	YORUBA	Central West Africa NC	0.30	0.00		1.25	3.96	0.21	0.91		0.21	GBR	Eurasia	5.22			4.36	2.37	4.88	4.78	0.00	2.37
GIRIAMA	0.00014	43	3	YORUBA	Central West Africa NC	0.49	0.00		3.88	8.92	0.35	1.66		0.35	TSI	Eurasia	12.40			10.92	6.16	12.29	11.83	0.00	6.16
WABONDEI	0.00013	78	17	AMAXHOSA	South Africa NC	0.22	0.02		1.00	2.43	0.00	0.11		0.02	GBR	Eurasia	2.97	2.87		2.85	1.58		3.41	0.00	1.59
WABONDEI	3e-05	14	4	BANTU	Central West Africa NC	0.25	0.00		1.21	2.50	-0.01	0.37		0.25	TSI	Eurasia	3.55			3.02	1.98	3.11	3.15	0.00	1.98
MZIGUA	0.00011	48	6	YORUBA	Central West Africa NC	0.16	0.00	0.63	1.86	4.41	0.11	0.63		0.11	TSI	Eurasia	6.34		6.78	5.54	3.15	6.02	5.67	0.00	3.15
MAASAI	0.00018	26	4	YORUBA	Central West Africa NC	0.05	0.00	1.37	1.24	5.63	0.76	1.37		0.05	TSI	Eurasia	8.66		9.16	8.42	4.89	8.92	8.73	0.00	4.89
SUDANESE	1.6e-05	10	3	/GUI//GANA	Khoesan	0.42	0.10	0.69	0.53	3.42	0.35	0.00		0.10	GBR	Eurasia	5.33	5.53	5.66	4.46	2.56			0.00	2.56
GUMUZ	1.8e-05	7	1	SUDANESE	Nilo-Saharan	0.65	0.31	0.80	0.00	3.03	0.78	0.32		0.31	GBR	Eurasia	7.71	8.03			5.03	7.56	7.33	0.00	4.99
GUMUZ	7e-05	145	41	SUDANESE	Nilo-Saharan	-0.52	-0.91	0.21	0.00	1.99	1.08	-0.46		-0.91	PEL	Eurasia	9.39	9.92			5.41	9.06	9.04	0.00	8.86
ANUAK	3.4e-05	74	20	SUDANESE	Nilo-Saharan	1.33	1.30	1.56	0.00	1.26	1.72		0.71	0.35	JU/'HOANSI	Khoesan	1.06	1.05	1.12		0.78	1.37	0.00	1.27	0.35
AFAR	0.00032	60	13	SUDANESE	Nilo-Saharan	0.43	0.25	0.64	0.00	2.16	0.35	0.39		0.25	FIN	Eurasia	5.50	5.46			4.22	5.50	5.55	0.00	4.20
OROMO	0.00064	135	13	JU/'HOANSI	Khoesan	0.72	0.51	0.34	0.20	1.10	0.56	0.00		0.20	TSI	Eurasia	6.85	6.82	6.68	6.38	3.33	6.86		0.00	3.33
OROMO	5.4e-05	11	2	GUMUZ	Nilo-Saharan	-0.16	-0.29	0.05	0.00	0.55	-0.27	-0.85		-0.85	TSI	Eurasia	8.91	8.97	9.39		5.86	9.40	8.36	0.00	6.21
SOMALI	0.00036	92	18	AKANS	Central West Africa NC	0.35	0.00	0.75	0.36	2.40	0.24	0.73		0.24	TSI	Eurasia	4.74		4.93	4.91	3.55		4.65	0.00	3.39
WOLAYTA	0.00048	76	14	JU/'HOANSI	Khoesan	0.66	0.50	0.73	0.29	0.91	0.41	0.00		0.29	CEU	Eurasia	3.93		4.39	4.09	2.53	4.39		0.00	2.53
ARI	0.00045	152	17	JU/'HOANSI	Khoesan	0.72	0.59	0.83	0.56	3.33	0.54	0.00		0.54	TSI	Eurasia				3.33	2.07			0.00	2.07
AMHARA	0.00048	96	18	ANUAK	Nilo-Saharan	0.34	0.17	0.63	0.00	1.27	0.22	0.18		0.17	TSI	Eurasia	4.77	4.75	4.76		3.32	4.75		0.00	2.60
TIGRAY	0.00064	115	12	SUDANESE	Nilo-Saharan	0.55	0.38	0.75	0.00	2.55	0.41	0.32		0.32	TSI	Eurasia	7.49				5.15	7.76		0.00	5.01
MALAWI	3.7e-05	61	9	YORUBA	Central West Africa NC	0.36	0.00	0.70	1.36	1.22	1.10		0.50	0.36	JU/'HOANSI	Khoesan	4.92			4.37	3.87		0.00	2.75	2.75
SEBANTU																									
AMAXHOSA	0.00018	30	4	KASEM	Central West Africa NC	0.20	0.00	0.17	0.93	1.39	1.34		0.43	0.17	JU/'HOANSI	Khoesan	7.76		7.90	7.56	6.89		0.00	5.54	5.43
HERERO	0.00025	4	1	JU/'HOANSI	Khoesan	0.97	0.70	1.09	1.78	4.15	0.63	0.00		0.63	GBR	Eurasia	7.51	7.52	7.57	7.08	5.03	7.60		0.00	5.03
KHWE	0.00026	154	34	SEMI-BANTU	Central West Africa NC	0.82	0.00	0.15	1.06	1.55	0.46			0.15	PEL	Eurasia	2.08		2.14	1.68	1.06	2.22		0.00	0.72
KHWE	3.9e-05	25	7	YORUBA	Central West Africa NC	0.18	0.00	0.06	0.60	1.45	-0.30			-0.30	TSI	Eurasia	2.66		2.71	2.26	1.81	2.92		0.00	1.36
/GUI//GANA	3.7e-05	30	5	JOLA	West Africa NC	0.00	0.10	0.71	1.85	3.56	1.72			0.10	TSI	Eurasia		7.54	7.65	6.45	3.74	7.36		0.00	3.12
KARRETJIE	0.00046	7	1	BANTU	Central West Africa NC	0.59	0.00	0.70	1.92	5.71				0.59	GBR	Eurasia	16.31		16.32	15.31	11.00			0.00	11.00
NAMA	0.00029	8	1	MALAWI	South Africa NC	0.59	0.08	0.77	1.57	5.15	0.00			0.08	GBR	Eurasia	13.97		13.96	13.16	9.63			0.00	9.63
NAMA	0.00045	139	35	MALAWI	South Africa NC	1.00	0.02	1.20	-1.95	5.99	0.00			1.20	PEL	Eurasia	13.80		13.79	12.62	9.88			0.00	9.88
!XUN	1e-04	78	10	SEMI-BANTU	Central West Africa NC	0.13	0.00	0.32	1.33	2.59	0.02			0.02	TSI	Eurasia	5.37		5.90	4.81	2.25			0.00	2.17
=KHOMANI	0.00048	8	1	BANTU	Central West Africa NC	0.55	0.00	0.93	1.56	5.30				0.55	GBR	Eurasia	15.25		15.26	15.03	10.47			0.00	10.47
JU/'HOANSI	1e-04	62	8	AMAXHOSA	South Africa NC	0.93	0.80	1.10	1.84	3.13	0.00			0.80	GBR	Eurasia	4.04	5.01	5.00	3.80	1.91			0.00	1.91

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3-Source Data 3. The evidence for multiple waves of admixture in African populations using MALDER and the European recombination map. Columns as in Figure 3-Source Data 1

Ethnic Group	amp	Date	Date (CI)	Pop1	Pop1Anc	Pop1Anc-West Africa NC(Z)	Pop1Anc-Central West Africa NC(Z)	Pop1Anc-East Africa NC(Z)	Pop1Anc-Nilo-Saharan(Z)	Pop1Anc-Afroasiatic(Z)	Pop1Anc-South Africa NC(Z)	Pop1Anc-Khoesan(Z)	Pop1Anc-Eurasia(Z)	Pop1Anc-(Z)	Pop2	Pop2Anc	Pop2Anc-West Africa NC(Z)	Pop2Anc-Central West Africa NC(Z)	Pop2Anc-East Africa NC(Z)	Pop2Anc-Nilo-Saharan(Z)	Pop2Anc-Afroasiatic(Z)	Pop2Anc-South Africa NC(Z)	Pop2Anc-Khoesan(Z)	Pop2Anc-Eurasia(Z)	Pop2Anc-(Z)
JOLA	7.1e-05	152	27	JU/'HOANSI	Khoesan	0.43	0.39	0.92	1.09	2.97	0.21	0.00		0.21	GBR	Eurasia	3.10			3.75	2.42			0.00	2.32
WOLLOF	0.00011	96	24	SEMI-BANTU	Central West Africa NC	0.03	0.00	0.62	0.93	3.18	0.29	0.41		0.03	GBR	Eurasia	5.84		5.77	5.13	3.16	5.72	5.59	0.00	3.16
WOLLOF	2.4e-05	11	3	YORUBA	Central West Africa NC	0.45	0.00	0.67	1.01	3.10	0.18	0.31		0.45	TSI	Eurasia	5.73		5.74	5.37	3.26	5.61	5.56	0.00	3.53
MANDINKAI																									
MANDINKAII	3.9e-05	20	3	SEMI-BANTU	Central West Africa NC	0.51	0.00	1.04	0.99	4.40	0.23	0.19		0.19	IBS	Eurasia	8.38		8.39	7.59	4.80	8.43	7.91	0.00	4.80
SEREHULE	6e-05	23	4	BANTU	Central West Africa NC	0.43	0.00	0.60	1.05	3.94	0.07	0.07		0.07	IBS	Eurasia	7.41		7.42	6.59	4.09	7.16	7.09	0.00	4.09
FULAI	0.00043	61	8	SEMI-BANTU	Central West Africa NC	0.09	0.00	1.27	1.27	5.42	0.40	0.63		0.09	GBR	Eurasia	11.76		11.65	10.62	6.64		11.74	0.00	6.43
FULAI	0.00016	11	2	AKANS	Central West Africa NC	-0.20	0.00	0.80	0.94	4.22	0.03	0.50		-0.20	IBS	Eurasia	8.82		8.76	8.02	5.19		8.80	0.00	5.19
MALINKE	0.00012	123	31	SEREHULE	West Africa NC	0.00	0.30	0.23	0.64	0.87	0.22	0.27		0.22	PEL	Eurasia		1.38	1.39		0.82	1.38	1.15	0.00	0.82
MALINKE	1.9e-05	8	2	MALAWI	South Africa NC	-0.52	-0.13	-0.40	0.13	0.84	0.00	-0.67		-0.13	TSI	Eurasia	2.03		1.97	1.91	1.25		1.29	0.00	1.25
SERERE	2.5e-05	15	3	BANTU	Central West Africa NC	0.68	0.00	0.75	1.02	4.58	0.41	0.25		0.25	GBR	Eurasia	7.98		8.10	7.10	4.32	8.07	7.58	0.00	4.32
MANJAGO	3.4e-05	3	1	YORUBA	Central West Africa NC	0.23	0.00	0.95	1.74	6.45	0.22	0.88		0.22	GBR	Eurasia	12.37		13.07	12.09	8.20	12.95	12.39	0.00	8.20
FULAII	2e-05	5	1	JU/'HOANSI	Khoesan	0.11	0.21	1.13	2.11	5.37	0.38	0.00		0.11	GBR	Eurasia	9.78	9.90	9.93	8.76	5.51			0.00	5.51
FULAII	7e-05	64	10	SEMI-BANTU	Central West Africa NC	0.10	0.00	0.73	1.68	5.01	0.13	-0.08		0.13	TSI	Eurasia	9.51		9.46	8.52	5.30	9.37	9.56	0.00	5.44
BAMBARA	5.2e-05	86	29	YORUBA	Central West Africa NC	0.06	0.00	0.18	0.35	0.97	0.09	0.33		0.06	IBS	Eurasia	1.70			1.58	0.99			0.00	0.99
BAMBARA	1.3e-05	9	2	/GUI//GANA	Khoesan	-0.20	-0.56	-0.33	-0.01	0.57	-0.36	0.00		-0.56	FIN	Eurasia					1.48	2.49		0.00	1.48
AKANS	2.3e-05	138	35	MANDINKAII	West Africa NC	0.00				0.26				0.26	GBR	Eurasia		1.98			0.58	1.76	0.48	0.00	0.48
MOSSI																									
YORUBA																									
KASEM																									
NAMKAM																									
SEMI-BANTU	1e-05	46	12	AKANS	Central West Africa NC	0.23	0.00		1.10		1.00			0.23	JU/'HOANSI	Khoesan	2.94			2.56			0.00		2.56
BANTU	2.1e-05	54	13	MOSSI	Central West Africa NC	0.01	0.00	0.41		0.54	1.79		2.07	0.01	JU/'HOANSI	Khoesan			2.94			0.84	0.00		0.84
LUHYA	7.8e-05	59	18	NAMKAM	Central West Africa NC	0.12	0.00		0.61	3.90	0.09	0.51		0.09	TSI	Eurasia	6.09			4.88	2.47		4.26	0.00	2.47
LUHYA	2.6e-05	13	3	YORUBA	Central West Africa NC	0.06	0.00		0.56	3.58	-0.14	0.77		0.06	TSI	Eurasia	5.84			4.42	2.27		4.50	0.00	2.27
KAMBE	0.00014	15	2	BANTU	Central West Africa NC	0.20	0.00		1.95	5.82	1.10	0.94		0.20	TSI	Eurasia	9.29			9.26	5.37	9.95	9.17	0.00	5.37
CHONYI	0.00014	27	2	YORUBA	Central West Africa NC	0.05	0.00		3.21	8.89	1.91	1.50		0.05	TSI	Eurasia	13.89			13.70	7.74		12.78	0.00	7.74
KAUMA	0.00015	27	2	SEMI-BANTU	Central West Africa NC	0.41	0.00		3.18	7.98		1.30		0.41	TSI	Eurasia	12.84			11.74	6.37		11.38	0.00	6.37
WASAMBAA	0.00012	21	3	NAMKAM	Central West Africa NC	0.21	0.00		2.46	6.63		1.56		0.21	TSI	Eurasia	8.67			8.16	3.97		7.87	0.00	3.97
GIRIAMA	0.00014	28	2	YORUBA	Central West Africa NC	0.58	0.00		3.89	8.81		2.05		0.58	TSI	Eurasia	12.38			11.97	6.15		11.80	0.00	6.15
WABONDEI	1e-04	23	2	NAMKAM	Central West Africa NC	0.36	0.00		3.78	9.03	1.59	2.31		0.36	TSI	Eurasia	12.37			11.60	5.91	13.21	10.54	0.00	5.91
MZIGUA	0.00011	32	4	YORUBA	Central West Africa NC	0.23	0.00		1.97	5.14	0.90	1.11		0.23	TSI	Eurasia	7.28			6.99	3.60	7.89	6.54	0.00	3.60
MAASAI	0.00039	81	17	SUDANESE	Nilo-Saharan	0.26	0.17		0.00	2.77	0.22	0.59		0.17	TSI	Eurasia	4.02	4.53			2.35	4.47	4.25	0.00	2.16
MAASAI	8.6e-05	11	2	SEMI-BANTU	Central West Africa NC	0.17	0.00		-0.23	2.85	0.12	0.43		0.12	TSI	Eurasia	4.24			4.52	2.38	4.52	4.28	0.00	2.38
SUDANESE	1.7e-05	6	2	/GUI//GANA	Khoesan	0.34	0.11	0.79	0.47	3.31	0.43	0.00		0.11	GBR	Eurasia	5.21	5.65	5.53	4.34	2.53	5.70		0.00	2.53
GUMUZ	1.8e-05	5	1	SUDANESE	Nilo-Saharan	0.50	0.25	0.64	0.00	2.23	0.59	0.24		0.24	GBR	Eurasia	5.58	5.84	5.82		3.68	5.45	5.31	0.00	3.55
GUMUZ	8.6e-05	104	26	SUDANESE	Nilo-Saharan	-0.55	-0.90	-0.07	0.00	1.22	-0.67	-0.99		-0.07	PEL	Eurasia	6.14	6.54	6.51		3.49	5.92	6.20	0.00	5.82
ANUAK	3.3e-05	52	16	SUDANESE	Nilo-Saharan	1.56	1.53	1.82	0.00	1.45	2.49		0.53	0.05	JU/'HOANSI	Khoesan	1.39	1.27	1.23		0.83	0.44	0.00	0.73	0.05
AFAR	0.00038	45	10	SUDANESE	Nilo-Saharan	0.23	0.09	0.54	0.00	2.01	0.17	0.37		0.09	TSI	Eurasia	4.84	4.80			3.63			0.00	3.57
OROMO	0.00063	86	9	JU/'HOANSI	Khoesan	0.75	0.52	0.76	0.22	2.19	0.60	0.00		0.22	TSI	Eurasia	7.99	7.99	8.01	7.48	4.39	8.01		0.00	4.39
OROMO	5e-05	7	2	JU/'HOANSI	Khoesan	0.75	0.64	0.76	0.86	2.19	0.65	0.00		0.86	TSI	Eurasia	8.02	8.14	8.03	7.48	4.39	8.01		0.00	6.19
SOMALI	0.00043	70	11	SUDANESE	Nilo-Saharan	0.30	0.12	0.93	0.00	2.60	0.15	0.31		0.12	TSI	Eurasia	5.26	5.54	5.48		3.83	5.52	5.28	0.00	3.75
WOLAYTA	0.00054	55	9	JU/'HOANSI	Khoesan	0.88	0.69	1.00	0.46	1.32	0.59	0.00		0.46	CEU	Eurasia	4.79	5.47	5.38	4.90	3.03	5.41		0.00	3.03
ARI	0.00045	99	11	JU/'HOANSI	Khoesan	0.83	0.67	1.03	0.63	3.20	0.59	0.00		0.59	TSI	Eurasia		5.17		4.34	2.73	5.15		0.00	2.73
AMHARA	0.00052	68	10	ANUAK	Nilo-Saharan	0.41	0.21	0.58	0.00	1.70	0.27	0.06		0.06	TSI	Eurasia	5.83	6.18	6.17		4.35	5.98		0.00	3.70
TIGRAY	0.00066	77	7	SUDANESE	Nilo-Saharan	0.65	0.44	0.86	0.00	2.83	0.47	0.44		0.44	TSI	Eurasia	9.02		8.98		6.29	8.86	8.73	0.00	6.27
MALAWI	4.1e-05	43	6	YORUBA	Central West Africa NC	0.42	0.00		1.53	1.34	1.12		0.51	0.42	JU/'HOANSI	Khoesan				5.50	4.79		0.00	3.15	3.15
SEBANTU																									
AMAXHOSA	5.2e-05	18	4	BAMBARA	West Africa NC	0.00	0.25		1.74	5.17	1.83			0.25	GBR	Eurasia		6.78		6.24	3.80			0.00	3.80
HERERO	0.00024	2	0	/GUI//GANA	Khoesan	1.06	0.64	1.29	2.36	6.25	0.58	0.00		0.58	GBR	Eurasia	12.56	12.59	12.65	11.80	8.44	12.70		0.00	8.44
KHWE	0.00019	103	22	WASAMBAA	East Africa NC	0.48	0.06	0.00	0.73	1.52				0.06	PEL	Eurasia	2.40	2.37		1.98	1.12			0.00	0.99
KHWE	4.4e-05	18	5	BANTU	Central West Africa NC	-0.04	0.00	0.09	0.48	1.04				-0.04	TSI	Eurasia	1.83		2.05	1.49	0.77			0.00	0.77
/GUI//GANA	3.8e-05	20	4	JOLA	West Africa NC	0.00	0.16		1.69	3.46	1.33			0.16	TSI	Eurasia		7.65		6.45	3.64			0.00	3.16
KARRETJIE	0.00047	5	0	BANTU	Central West Africa NC	0.76	0.00		2.33	6.97				0.76	GBR	Eurasia	19.78			19.45	13.76			0.00	13.76
NAMA	0.00027	5	0	MALAWI	South Africa NC	0.07	0.03	0.08	0.17	0.56	0.00			0.03	GBR	Eurasia	1.46	1.46	1.46	1.45	1.03			0.00	1.03
NAMA	0.00031	80	13	BANTU	Central West Africa NC	0.07	0.00	0.01	0.13	0.46	-0.01			0.07	FIN	Eurasia	1.24		1.24	1.22	0.95	1.24		0.00	1.03
!XUN	0.00011	52	6	BANTU	Central West Africa NC	0.37	0.00	0.58	1.89	3.84	0.08			0.08	TSI	Eurasia	7.06		7.61	6.26	3.51			0.00	3.16
=KHOMANI	0.00049	6	0	BANTU	Central West Africa NC	0.73	0.00		2.10					0.73	GBR	Eurasia	20.41			20.13				0.00	20.13
JU/'HOANSI	8.6e-05	41	5	MALAWI	South Africa NC	0.15	0.15	0.50	1.51	3.60	0.00			0.15	TSI	Eurasia	6.13	6.68	6.68	5.56	2.96			0.00	2.96

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3-Source Data 4. The evidence for multiple waves of admixture in African populations using MALDER and the HAPMAP recombination map and a mindis of 0.5cM. Columns are as in Figure 3-Source Data 1. Here we show the results for the MALDER analysis where we over-ride any short-range LD and define a minimum distance of 0.5cM from which to start computing admixture LD curves

Ethnic Group	amp	Date	Date (CI)	Pop1	Pop1Anc	Pop1Anc-West Africa NC(Z)	Pop1Anc-Central West Africa NC(Z)	Pop1Anc-East Africa NC(Z)	Pop1Anc-Nilo-Saharan(Z)	Pop1Anc-Afroasiatic(Z)	Pop1Anc-South Africa NC(Z)	Pop1Anc-Khoesan(Z)	Pop1Anc-Eurasia(Z)	Pop1Anc-(Z)	Pop2	Pop2Anc	Pop2Anc-West Africa NC(Z)	Pop2Anc-Central West Africa NC(Z)	Pop2Anc-East Africa NC(Z)	Pop2Anc-Nilo-Saharan(Z)	Pop2Anc-Afroasiatic(Z)	Pop2Anc-South Africa NC(Z)	Pop2Anc-Khoesan(Z)	Pop2Anc-Eurasia(Z)	Pop2Anc-(Z)
JOLA	0.00016	225	24	MANDINKAI	West Africa NC	0.00	0.24	0.78	1.07	3.65	0.45	0.56		0.24	GBR	Eurasia		5.34	5.19	4.32	2.64	5.12	4.66	0.00	2.64
WOLLOF	0.00019	141	16	JOLA	West Africa NC	0.00	0.36	1.83	2.89	8.97	0.90	0.78		0.36	TSI	Eurasia		13.75	14.09	12.53	7.70	13.63	12.55	0.00	7.70
WOLLOF	2.9e-05	12	3	JOLA	West Africa NC	0.00	0.55	1.83	2.90	8.99	0.90	1.29		0.55	TSI	Eurasia		13.89	14.13	12.53	7.70	13.94	13.50	0.00	7.70
MANDINKAI	7.5e-05	31	10	JOLA	West Africa NC	0.00	0.25	0.82	1.09	3.11	0.51	0.62		0.25	TSI	Eurasia		4.36	4.59	4.62	2.71	4.40	4.39	0.00	2.71
MANDINKAII	0.00016	170	37	JOLA	West Africa NC	0.00	0.15	0.59	1.03	3.61	0.25	0.37		0.15	IBS	Eurasia		5.25	5.42	4.62	2.65	4.76	4.58	0.00	2.65
MANDINKAII	2.6e-05	14	5	JOLA	West Africa NC	0.00	0.32	0.92	1.14	3.75	0.30	0.45		0.32	CEU	Eurasia		5.37	5.41	4.59	2.59	5.29	4.55	0.00	2.59
SEREHULE	0.00018	116	19	JOLA	West Africa NC	0.00	0.28	1.03	1.96	6.12	0.50	0.83		0.28	TSI	Eurasia		9.70	9.95	8.98	5.73	9.64	9.42	0.00	5.73
SEREHULE	2.6e-05	10	3	AKANS	Central West Africa NC	0.02	0.00	0.91	1.73	6.58	0.11	0.55		0.11	IBS	Eurasia	10.78		12.00	10.65	6.76	10.87	10.86	0.00	6.76
FULAI	6e-04	86	9	JOLA	West Africa NC	0.00	0.47	2.57	3.61	13.21	0.98	2.25		0.47	TSI	Eurasia		25.91	25.75	23.18	14.98	24.99	24.12	0.00	14.98
FULAI	0.00019	12	2	AKANS	Central West Africa NC	-0.50	0.00	2.08	3.08	12.59	0.51	1.76		-0.50	IBS	Eurasia	25.07		25.02	22.44	14.50	24.28	23.53	0.00	14.50
MALINKE																									
SERERE																									
MANJAGO	3.3e-05	3	0	AKANS	Central West Africa NC	0.08	0.00	0.97	1.80	6.48	0.23	0.92		0.08	GBR	Eurasia	12.26		12.97	12.00	8.15	12.87	12.31	0.00	8.15
MANJAGO	0.00024	279	40	JU/'HOANSI	Khoesan	-1.13	-1.23	-0.25	0.73	5.76	-0.61	0.00		-1.13	TSI	Eurasia	12.86	13.37	13.41	12.78	9.42	12.92		0.00	9.42
FULAII	0.00013	106	11	AKANS	Central West Africa NC	0.04	0.00	1.03	1.60	6.88	0.40	0.92		0.04	TSI	Eurasia	10.92		12.61	10.98	6.87	11.57	11.18	0.00	6.71
FULAII	2.2e-05	6	1	JOLA	West Africa NC	0.00	-0.03	0.96	1.57	6.92	0.35	1.13		-0.03	GBR	Eurasia		10.70	11.25	11.07	6.48	10.75	10.04	0.00	6.48
BAMBARA	4.9e-05	37	11	AKANS	Central West Africa NC	0.33	0.00	0.50	0.81	2.70	0.13	0.21		0.13	TSI	Eurasia	3.76		4.81	4.12	2.52	4.45	4.41	0.00	2.52
AKANS	7.7e-05	174	38	JU/'HOANSI	Khoesan	0.64	0.41	0.76	1.05	2.01	0.35	0.00		0.35	IBS	Eurasia	1.59	1.43	1.96	2.08	0.95	2.13		0.00	0.95
MOSSI	7.3e-05	118	26	BANTU	Central West Africa NC	0.09	0.00	0.40	0.80	2.53	0.12	0.13		0.09	GBR	Eurasia	3.50		3.94	3.22	2.01	3.45	3.23	0.00	1.78
YORUBA	0.00013	257	42	/GUI//GANA	Khoesan	0.44	0.21	0.49	1.01	2.48	0.10	0.00		0.10	CEU	Eurasia	1.55	1.24	2.43	1.96	1.11	2.59		0.00	1.11
KASEM	0.00021	300	50	AMAXHOSA	South Africa NC	0.18	0.02	0.32	0.85	1.85	0.00	0.33		0.02	IBS	Eurasia	2.37	2.49	2.97	2.28	1.36		2.49	0.00	1.15
NAMKAM	0.00013	214	38	!XUN	Khoesan	0.22	0.11	0.48	0.62	2.14	0.00	0.00		0.00	TSI	Eurasia	3.11	3.05	3.22	2.77	1.47	3.33		0.00	1.47
SEMI-BANTU	1e-04	197	36	JU/'HOANSI	Khoesan	1.60	1.30	1.55	1.50		0.89	0.00		0.89	IBS	Eurasia	1.34	1.05	1.38	1.74	0.87	2.71		0.00	0.87
BANTU	0.00017	311	49	JU/'HOANSI	Khoesan	0.36	0.30	0.43			0.07	0.00		0.07	GBR	Eurasia	1.85	1.45	1.92	1.16	1.06	2.66		0.00	1.06
BANTU	2.1e-05	51	14	YORUBA	Central West Africa NC	0.57	0.00	0.27	0.15	0.14	1.39		-1.81	0.57	JU/'HOANSI	Khoesan	3.01		2.62	2.05	1.25	1.02	0.00	-1.42	1.88
LUHYA	0.00013	103	24	JU/'HOANSI	Khoesan	0.52	0.49	0.78	0.46	4.66	0.37	0.00		0.37	TSI	Eurasia	5.35	5.32	5.58	5.45	3.13	7.19		0.00	3.13
LUHYA	4.1e-05	16	3	MALAWI	South Africa NC	0.39	0.21	0.65	0.16	5.61	0.00	-0.06		0.21	TSI	Eurasia	10.24	9.26	11.20	9.63	5.17		9.85	0.00	5.33
KAMBE	0.00018	20	2	MALAWI	South Africa NC	0.98	0.32	0.88	3.69	9.41	0.00	1.23		0.32	TSI	Eurasia	14.36	14.30	15.34	13.51	8.35		14.33	0.00	8.00
CHONYI	0.00018	123	28	JU/'HOANSI	Khoesan	0.94	0.75	1.06	1.90	4.24	0.21	0.00		0.21	IBS	Eurasia	4.59	4.60	4.48	4.96	3.02	5.73		0.00	3.02
CHONYI	0.00012	23	2	MALAWI	South Africa NC	0.46	0.31	0.56	1.25	3.88	0.00	-0.58		0.31	TSI	Eurasia	5.64	5.49	6.72	5.52	3.30		5.63	0.00	3.51
KAUMA	0.00019	31	2	MALAWI	South Africa NC	1.52	0.67	1.28	5.87	14.76	0.00	1.75		0.67	TSI	Eurasia	21.44	21.18	23.06	19.97	12.10		21.42	0.00	11.46
WASAMBAA	0.00018	28	4	MALAWI	South Africa NC	1.08	0.50	0.95	3.70	8.45	0.00	1.54		0.50	TSI	Eurasia	11.02	10.81	12.27	10.10	5.83		10.99	0.00	5.33
GIRIAMA	0.00017	32	1	MALAWI	South Africa NC	2.33	0.78	2.11	9.54	22.80	0.00	3.41		0.78	TSI	Eurasia	31.59	31.09	34.24	29.08	17.33		31.40	0.00	16.28
WABONDEI	0.00016	33	3	MALAWI	South Africa NC	1.35	0.70	1.48	5.50	13.06	0.00	2.46		0.70	TSI	Eurasia	16.92	16.61	19.24	16.10	9.40		16.92	0.00	8.67
MZIGUA	0.00015	130	33	AMAXHOSA	South Africa NC	1.26	0.75	0.98	2.75	6.66	0.00	0.20		0.20	GBR	Eurasia	8.19	7.88	9.60	8.48	4.81		8.67	0.00	4.11
MZIGUA	8.8e-05	26	5	MALAWI	South Africa NC	0.50	0.24	0.71	1.54	4.32	0.00	-0.25		0.24	TSI	Eurasia	5.66	5.67	6.72	5.37	3.18		5.95	0.00	3.23
MAASAI	0.00053	104	9	SUDANESE	Nilo-Saharan	1.17	0.65	1.85	0.00	10.69	0.60	1.05		0.60	TSI	Eurasia	19.62	21.14	21.11		15.32	21.17	19.21	0.00	11.84
MAASAI	9.9e-05	11	2	MALAWI	South Africa NC	0.82	0.10	1.26	-0.60	9.60	0.00	0.42		0.10	TSI	Eurasia	18.61	18.96	20.17	19.72	11.30		18.97	0.00	10.88
SUDANESE																									
GUMUZ	1.9e-05	5	1	SUDANESE	Nilo-Saharan	0.61	0.33	0.68	0.00	2.65	0.68	0.30		0.30	GBR	Eurasia	6.71	6.96	6.95		5.24	7.21	6.25	0.00	4.16
GUMUZ	8.2e-05	110	25	SUDANESE	Nilo-Saharan	0.81	0.43	1.25	0.00	2.58	0.65	0.29		0.29	TSI	Eurasia	6.52	6.76	6.90		5.10	7.00	6.27	0.00	4.41
ANUAK																									
AFAR	0.001	166	31	ANUAK	Nilo-Saharan	0.59	0.37	0.64	0.00	2.55	0.44	0.56		0.37	TSI	Eurasia	6.03	6.28	6.40		5.12	6.33	6.20	0.00	4.15
AFAR	0.00013	21	6	SUDANESE	Nilo-Saharan	0.72	0.60	0.73	0.00	2.69	0.57	0.74		0.60	TSI	Eurasia	6.15	6.48	6.53		5.20	6.45	6.33	0.00	4.38
OROMO	0.00073	93	7	SUDANESE	Nilo-Saharan	1.58	1.05	2.14	0.00	7.99	1.10	0.42		0.42	TSI	Eurasia	22.22	22.84	23.25		18.51	22.96	22.13	0.00	15.00
OROMO	5.4e-05	7	2	JU/'HOANSI	Khoesan	1.44	0.96	1.77	1.51	7.87	1.05	0.00		1.51	TSI	Eurasia	22.98	24.38	24.89	23.06	15.63	24.84		0.00	15.63
SOMALI	0.00075	100	5	SUDANESE	Nilo-Saharan	1.15	0.87	1.60	0.00	7.53	0.90	1.25		0.87	TSI	Eurasia	17.28	17.76	18.06		18.06	17.82	17.45	0.00	11.67
WOLAYTA	0.00069	67	8	JU/'HOANSI	Khoesan	1.43	1.19	1.58	0.75	2.93	1.00	0.00		0.75	TSI	Eurasia	9.97	11.00	10.94	10.12	6.52	11.10		0.00	6.52
ARI	0.00056	109	7	JU/'HOANSI	Khoesan	2.39	2.07	2.77	1.91	9.13	1.87	0.00		1.87	TSI	Eurasia	12.67	14.16	14.05	12.60	7.83	14.34		0.00	7.83
AMHARA	0.00074	85	6	SUDANESE	Nilo-Saharan	0.89	0.59	1.23	0.00	5.02	0.69	0.29		0.29	TSI	Eurasia	14.33	14.91	15.11		12.28	14.96	14.57	0.00	9.77
TIGRAY	0.00082	87	7	SUDANESE	Nilo-Saharan	0.86	0.64	1.18	0.00	4.62	0.68	0.77		0.64	TSI	Eurasia	12.52	12.90	13.07		10.72	12.96	12.67	0.00	9.02
MALAWI	0.00021	355	81	/GUI//GANA	Khoesan	0.42	0.36	0.42	0.65	1.73	0.06	0.00		0.06	IBS	Eurasia	0.99	0.84	1.00	0.81	0.58	1.36		0.00	0.58
MALAWI	4e-05	41	6	YORUBA	Central West Africa NC	-0.08	0.00	0.06	-0.30	-0.37	0.53		-0.83	0.06	JU/'HOANSI	Khoesan	1.66		1.59	1.06	-0.01	0.49	0.00	-0.47	0.49
SEBANTU	0.00024	17	3	JU/'HOANSI	Khoesan	8.68	8.48	8.66	9.27	10.18	2.89	0.00		2.89	TSI	Eurasia	2.08	1.99	2.34	2.91	1.83	9.12		0.00	1.83
AMAXHOSA	0.00026	26	2	JU/'HOANSI	Khoesan	14.45	14.23	14.54	16.19	18.29	5.37	0.00		5.37	IBS	Eurasia	1.07	0.67	1.26	2.74	2.37	14.21		0.00	0.67
HERERO	0.00025	2	0	JU/'HOANSI	Khoesan	1.70	1.24	1.96	3.05	7.17	1.17	0.00		1.17	GBR	Eurasia	11.83	12.56	12.67	11.90	8.43	13.60		0.00	8.43
KHWE	0.00039	76	22	JU/'HOANSI	Khoesan	6.38	6.21	6.55	6.72	7.75	3.15	0.00		3.15	TSI	Eurasia	2.61	2.75	3.13	2.93	1.68	8.29		0.00	1.68
KHWE	2e-04	17	4	NAMKAM	Central West Africa NC	0.07	0.00	0.14	0.06	-1.06	1.97		-2.70	0.07	JU/'HOANSI	Khoesan	2.29		2.39	2.34	2.03	1.13	0.00	0.73	1.13
/GUI//GANA	4e-04	28	3	JU/'HOANSI	Khoesan	23.55	23.00	23.20	24.05	24.99	8.24	0.00		8.24	IBS	Eurasia	4.03	4.22	4.64	5.19	3.03	21.66		0.00	3.03
/GUI//GANA	0.00014	4	1	NAMKAM	Central West Africa NC	0.11	0.00	0.17	0.51	-1.21	8.82		-3.35	0.54	JU/'HOANSI	Khoesan	11.43		11.39	11.10	10.90	4.07	0.00	10.44	4.07
KARRETJIE	0.0011	5	0	JU/'HOANSI	Khoesan	14.37	13.73	14.43	15.56	18.63	11.44	0.00		11.44	GBR	Eurasia	15.47	18.14	18.14	15.87	10.22	24.23		0.00	10.22
NAMA	5e-04	5	0	JU/'HOANSI	Khoesan	6.00	5.60	6.16	6.74	9.17	2.97	0.00		2.97	GBR	Eurasia	9.48	10.56	10.71	9.78	6.79	10.40		0.00	6.79
NAMA	0.00079	60	5	JU/'HOANSI	Khoesan	6.13	5.72	6.48	6.98	9.34	3.04	0.00		3.04	TSI	Eurasia	9.25	10.38	10.64	9.57	6.47	10.21		0.00	6.47
!XUN	0.00047	49	5	JU/'HOANSI	Khoesan	15.44	14.86	15.60	16.79	18.79	7.67	0.00		7.67	TSI	Eurasia	4.92	5.33	6.25	5.66	2.90	19.07		0.00	2.90

Continued on next page

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3-Source Data 4 Continued from previous page

Ethnic Group	amp	Date	Date (CI)	Pop1	Pop1Anc	Pop1Anc-West Africa NC(Z)	Pop1Anc-Central West Africa NC(Z)	Pop1Anc-East Africa NC(Z)	Pop1Anc-Nilo-Saharan(Z)	Pop1Anc-Afroasiatic(Z)	Pop1Anc-South Africa NC(Z)	Pop1Anc-Khoesan(Z)	Pop1Anc-Eurasia(Z)	Pop1Anc-(Z)	Pop2	Pop2Anc	Pop2Anc-West Africa NC(Z)	Pop2Anc-Central West Africa NC(Z)	Pop2Anc-East Africa NC(Z)	Pop2Anc-Nilo-Saharan(Z)	Pop2Anc-Afroasiatic(Z)	Pop2Anc-South Africa NC(Z)	Pop2Anc-Khoesan(Z)	Pop2Anc-Eurasia(Z)	Pop2Anc-(Z)
!XUN	0.00013	10	1	MOSSI	Central West Africa NC	0.23	0.00	0.82	-0.12	-1.61	8.78		-2.87	0.23	JU/'HOANSI	Khoesan	10.14		10.37	10.14	9.62	4.91	0.00	7.78	4.91
=KHOMANI	0.00094	5	1	JU/'HOANSI	Khoesan	4.80	4.52	4.84	5.27	6.81	1.90	0.00		1.90	GBR	Eurasia	6.96	8.06	8.03	7.10	4.74	9.79		0.00	4.74
=KHOMANI	0.00051	49	13	JU/'HOANSI	Khoesan	6.12	5.75	6.14	6.45	7.77	1.93	0.00		1.93	TSI	Eurasia	6.81	7.94	7.96	6.94	4.53	9.75		0.00	4.53
JU/'HOANSI	0.00029	39	3	!XUN	Khoesan	11.56	11.12	11.64	12.76	14.32	2.67	0.00		2.67	TSI	Eurasia	7.98	9.92	10.17	7.52	3.67	15.73		0.00	3.67

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4-Source Data 1. Results of the main GLOBETROTTER analysis. *Analysis* refers to whether the main or masked analysis was used to produce the final result. Admixture P-values are based on 100 bootstrap replicates of the NULL procedure. Our resulting inference, *res* can be: 1D (two admixing sources at a single date); 1MW (multiple admixing sources at a single date); 2D (admixture at multiple dates); NA (no-admixture); U (uncertain). max(R_1) refers to the R^2 goodness-of-fit for a single date of admixture, taking the maximum value across all inferred coancestry curves. FQ_1 is the fit of a single admixture event (i.e. the first principal component, reflecting admixture involving two sources) and FQ_2 is the fit of the first two principal components capturing the admixture event(s) (the second component might be thought of as capturing a second, less strongly-signalled event. M is the additional R^2 explained by adding a second date versus assuming only a single date of admixture; we use values above 0.35 to infer multiple dates (although see Supplementary Text for details). As well as the final result, for each event we show the inferred dates, αs and best matching sources for 1D, 1MW, and 2D inferences. Inferred dates are in years(+ 95% CI; B=BCE, otherwise CE); the proportion of admixture from the minority source (source 1) is represented by α. Date confidence intervals are based on 100 bootstrap replicates of the date inference

Cluster	Analysis	P	res	max(R_1)	FQ_1	FQ_2	M	1D	1Dα1	1D src1	1D src2	1MW α	1MW src3	1MW src4	2D1	2D1α	2D1 src1	2D1 src2	2D2	2D2α	2D2 src1	2D2 src2
SEREHULE	main	<0.01	1D	0.943	1	1	0.36	413 (514B-921)	0.11	GBR	JOLA											
MANDINKAII	main	<0.01	1D	0.925	1	1	0.39	51B (1244B-1066)	0.1	GBR	JOLA											
WOLLOF	main	<0.01	1D	0.938	1	1	0.63	254B (779B-355)	0.09	GBR	JOLA											
MANDINKAI	main	<0.01	1D	0.806	1	1	0.47	312B (718B-416)	0.15	GBR	JOLA											
SERERE	main	<0.01	1D	0.885	1	1	0.41	776B (1740B-254)	0.08	GBR	JOLA											
FULAI	main	<0.01	1D	0.969	1	1	0.82	239 (224-761)	0.19	IBS	WOLLOF											
BAMBARA	main	<0.01	1D	0.901	1	1	0.36	152 (675B-906)	0.06	CEU	MALINKE											
FULAII	main	<0.01	1D	0.939	1	1	0.57	65 (253B-876)	0.1	GBR	MALINKE											
MANJAGO	main	<0.01	1D	0.779	0.985	1	0.50	413 (1590B-1718)	0.21	FULAI	JOLA											
JOLA	main	<0.01	1D	0.642	1	1	0.08	834B (2287B-169)	0.17	FULAI	SERERE											
MALINKE	main	<0.01	1D	0.91	1	1	0.50	326 (544B-865)	0.11	GBR	BAMBARA											
YORUBA	main	<0.01	1D	0.465	0.999	1	0.03	848 (338-1184)	0.48	SEMI-BANTU	AKANS											
KASEM	main	<0.01	1D	0.305	0.997	1	0.03	819 (456-1329)	0.1	SEMI-BANTU	MOSSI											
NAMKAM	main	<0.01	1D	0.197	0.996	0.999	0.02	616 (184B-1197)	0.11	SEMI-BANTU	MOSSI											
SEMI-BANTU	main	<0.01	1D	0.43	0.999	1	0.03	674 (192-1126)	0.2	MZIGUA	YORUBA											
BANTU	main	<0.01	1D	0.447	0.995	0.999	0.05	7 (617B-602)	0.34	MALAWI	YORUBA											
MOSSI	main	<0.01	1D	0.41	0.997	1	0.05	355 (97B-952)	0.21	YORUBA	KASEM											
AKANS	main	<0.01	1MW	0.623	0.791	0.999	0.14	1399 (717-1675)	0.03	MALAWI	KASEM	0.29	KASEM	NAMKAM								
MAASAI	main	<0.01	2D	0.918	0.992	1	0.61								1660 (1573-1747)	0.06	TIGRAY	LUHYA	254B (764B-239)	0.35	AFAR	LUHYA
CHONYI	main	<0.01	1D	0.989	0.981	1	0.18	1138 (1080-1182)	0.08	KHV	WASAMBAA											
MZIGUA	main	<0.01	1D	0.988	0.999	1	0.19	1080 (1007-1138)	0.11	AFAR	WABONDEI											
KAMBE	main	<0.01	2D	0.989	0.987	1	0.46								1544 (1370-1776)	0.14	KAUMA	MZIGUA	761 (461B-1053)	0.06	GBR	MZIGUA
WABONDEI	main	<0.01	2D	0.985	0.998	0.999	0.48								1573 (1326-1834)	0.28	WASAMBAA	MZIGUA	703 (19-936)	0.1	TIGRAY	MZIGUA
LUHYA	main	<0.01	2D	0.993	0.998	0.999	0.59								1486 (1428-1573)	0.25	SUDANESE	MZIGUA	65 (400B-616)	0.29	WASAMBAA	MZIGUA
GIRIAMA	main	<0.01	1D	0.989	0.999	1	0.21	1196 (1138-1254)	0.1	OROMO	MZIGUA											
WASAMBAA	main	<0.01	1D	0.988	1	1	0.34	1312 (1254-1341)	0.14	TIGRAY	MZIGUA											
KAUMA	main	<0.01	1D	0.992	0.982	1	0.29	1225 (1167-1254)	0.06	GIH	MZIGUA											
MALAWI	main.null	<0.01	1MW	0.894	0.967	0.997	0.12	471 (340-631)	0.17	SEMI-BANTU	MZIGUA	0.16	AMAXHOSA	MZIGUA								
ANUAK	main	<0.01	1MW	0.575	0.961	0.997	0.04	703 (427-1037)	0.17	YORUBA	SUDANESE	0.33	SUDANESE	SUDANESE								
ARI	main.null	<0.01	1D	0.866	0.995	0.999	0.06	689B (965B-297B)	0.14	TSI	GUMUZ											
SUDANESE	main	<0.01	1MW	0.789	0.782	0.999	0.21	1341 (1225-1660)	0.27	GUMUZ	ANUAK	0.25	ANUAK	ANUAK								
GUMUZ	main	<0.01	1D	0.785	0.987	0.999	0.26	1544 (1384-1718)	0.24	ARI	ANUAK											
AFAR	main.null	<0.01	1D	0.895	1	1	0.14	326 (7-587)	0.23	TSI	SOMALI											
OROMO	main.null	<0.01	2D	0.922	1	1	0.50								1834 (1674-1892)	0.17	IBS	WOLAYTA	283B (617B-51B)	0.27	TSI	ARI
SOMALI	main.null	<0.01	2D	0.939	1	1	0.40								1573 (876-1805)	0.04	TSI	WOLAYTA	863B (1887B-399B)	0.47	TIGRAY	GUMUZ
WOLAYTA	main.null	<0.01	1D	0.825	1	1	0.18	268 (8B-602)	0.22	TSI	ARI											
TIGRAY	main.null	<0.01	1D	0.947	1	1	0.20	36 (196B-240)	0.35	TSI	ARI											

Continued on next page

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4-Source Data 1 Continued from previous page

Cluster	Analysis	P	res	$\max(R_1)$	FQ_1	FQ_2	M	$1D$	$1D\alpha 1$	$1D$ src1	$1D$ src2	$1MW$ α	$1MW$ src3	$1MW$ src4	$2D1$	$2D1\alpha$	$2D1$ src1	$2D1$ src2	$2D2$	$2D2\alpha$	$2D2$ src1	$2D2$ src2
AMHARA	main.null	<0.01	1D	0.947	1	1	0.26	138B (385B-95)	0.35	TSI	ARI											
AMAXHOSA	main.null	<0.01	1D	0.982	1	1	0.49	1196 (1109-1283)	0.32	KARRETJIE	MALAWI											
SEBANTU	main.null	<0.01	1D	0.982	1	1	0.57	1109 (1051-1196)	0.3	KARRETJIE	MALAWI											
HERERO	main.null	<0.01	2D	0.816	0.998	1	0.34								1834 (1805-1892)	0.24	NAMA	AMAXHOSA	674 (124B-979)	0.44	NAMA	SEMI-BANTU
KARRETJIE	main.null	<0.01	1D	0.997	0.999	1	0.33	1776 (1776-1805)	0.1	GBR	/GUI//GANA											
JU/'HOANSI	main.null	<0.01	1D	0.832	0.986	1	0.08	558 (311-851)	0.11	SOMALI	KARRETJIE											
=KHOMANI	main.null	<0.01	1D	0.998	0.999	1	0.09	1776 (1747-1805)	0.13	CEU	KARRETJIE											
/GUI//GANA	main.null	<0.01	1D	0.865	0.995	1	0.33	1544 (1457-1631)	0.24	MALAWI	JU/'HOANSI											
NAMA	main.null	<0.01	1MW	0.991	0.93	0.999	0.52	1805 (1805-1863)	0.12	CEU	JU/'HOANSI	0.23	HERERO	=KHOMANI								
!XUN	main.null	<0.01	1D	0.951	0.999	1	0.21	1312 (1254-1385)	0.28	SEMI-BANTU	JU/'HOANSI											
KHWE	main.null	<0.01	1D	0.911	0.985	0.999	0.16	1312 (1152-1399)	0.4	JU/'HOANSI	SEMI-BANTU											

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND4.0 International license.

Figure 4-Source Data 2. Results of the main GLOBETROTTER analysis. *Analysis* refers to whether the main or masked analysis was used to produce the final result. Admixture P-values are based on 100 bootstrap replicates of the NULL procedure. Our resulting inference, *res* can be: 1D (two admixing sources at a single date); 1MW (multiple admixing sources at a single date); 2D (admixture at multiple dates); NA (no-admixture); U (uncertain). max(R_1) refers to the R^2 goodness-of-fit for a single date of admixture, taking the maximum value across all inferred coancestry curves. FQ_1 is the fit of a single admixture event (i.e. the first principal component, reflecting admixture involving two sources) and FQ_2 is the fit of the first two principal components capturing the admixture event(s) (the second component might be thought of as capturing a second, less strongly-signalled event. M is the additional R^2 explained by adding a second date versus assuming only a single date of admixture; we use values above 0.35 to infer multiple dates (although see Supplementary Text for details). As well as the final result, for each event we show the inferred dates, αs and best matching sources for 1D, 1MW, and 2D inferences. Inferred dates are in years(+ 95% CI; B=BCE, otherwise CE); the proportion of admixture from the minority source (source 1) is represented by α. Date confidence intervals are based on 100 bootstrap replicates of the date inference

Cluster	Analysis	P	res	max(R_1)	FQ_1	FQ_2	M	$1D$	$1D\alpha 1$	$1D$ src1	$1D$ src2	$1MW$ α	$1MW$ src3	$1MW$ src4	$2D1$	$2D1\alpha$	$2D1$ src1	$2D1$ src2	$2D2$	$2D2\alpha$	$2D2$ src1	$2D2$ src2
AFAR	main	<0.01	1D	0.935	1	1	0.23	558 (253-820)	0.22	TSI	SOMALI	0.43	AMHARA	OROMO	1196 (993-1892)	0.16	TSI	WOLAYTA	1443B (2284B-428)	0.29	TSI	ARI
AFAR	main.null	<0.01	1D	0.895	1	1	0.14	326 (7-587)	0.23	TSI	SOMALI	0.37	OROMO	AMHARA	1167 (1007-1892)	0.14	TSI	WOLAYTA	1530B (2183B-413)	0.31	TSI	ARI
AKANS	main	<0.01	1MW	0.623	0.791	0.999	0.14	1399 (717-1675)	0.03	MALAWI	KASEM	0.29	KASEM	NAMKAM	1805 (1645-1892)	0.26	MOSSI	NAMKAM	384 (1791B-1255)	0.04	MALAWI	KASEM
AKANS	main.null	<0.01	1D	0.253	0.996	0.998	0.06	935 (66B-1545)	0.04	MALAWI	KASEM	0.1	MOSSI	NAMKAM	1805 (1572-1892)	0.06	SEMI-BANTU	NAMKAM	283B (3744B-1040)	0.06	MZIGUA	MOSSI
AMAXHOSA	main	<0.01	1D(2D)	0.986	1	1	0.43	1225 (1167-1283)	0.31	KARRETJIE	MALAWI	0.34	SEBANTU	SEBANTU	1312 (1239-1892)	0.3	KARRETJIE	MALAWI	4111B (6092B-1080)	0.32	KARRETJIE	MALAWI
AMAXHOSA	main.null	<0.01	1D(2D)	0.982	1	1	0.49	1196 (1109-1283)	0.32	KARRETJIE	MALAWI	0.32	SEBANTU	SEBANTU	1312 (1225-1617)	0.31	KARRETJIE	MALAWI	2458B (4187B-646)	0.32	KARRETJIE	MALAWI
AMHARA	main	<0.01	2D	0.954	1	1	0.39	7 (196B-167)	0.35	TSI	ARI	0.25	TIGRAY	AFAR	1573 (1121-1791)	0.23	TSI	OROMO	631B (1299B-370B)	0.36	TSI	ARI
AMHARA	main.null	<0.01	1D	0.947	1	1	0.26	138B (385B-95)	0.35	TSI	ARI	0.3	TIGRAY	TIGRAY	1631 (1079-1892)	0.21	TSI	OROMO	602B (1081B-312B)	0.37	TSI	ARI
ANUAK	main	<0.01	1MW	0.575	0.961	0.997	0.04	703 (427-1037)	0.17	YORUBA	SUDANESE	0.33	SUDANESE	SUDANESE	1892 (1137-1892)	0.18	GUMUZ	SUDANESE	471 (271B-964)	0.16	YORUBA	SUDANESE
ANUAK	main.null	<0.01	1MW	0.449	0.962	0.996	0.05	587 (311-1109)	0.14	YORUBA	SUDANESE	0.25	GUMUZ	SUDANESE	1225 (1109-1892)	0.33	SUDANESE	SUDANESE	297 (516B-1037)	0.14	YORUBA	SUDANESE
ARI	main	<0.01	1D	0.885	0.994	0.999	0.08	602B (965B-283B)	0.15	TSI	GUMUZ	0.24	WOLAYTA	SOMALI	1399 (1108-1892)	0.23	WOLAYTA	MAASAI	1124B (1501B-485B)	0.1	TSI	GUMUZ
ARI	main.null	<0.01	1D	0.866	0.995	0.999	0.06	689B (965B-297B)	0.14	TSI	GUMUZ	0.21	WOLAYTA	SOMALI	1312 (1138-1892)	0.46	SOMALI	GUMUZ	1327B (1866B-440B)	0.11	TSI	GUMUZ
BAMBARA	main	<0.01	1D(2D)	0.901	1	1	0.36	1370 (1181-1501)	0.11	GBR	MALINKE	0.3	SERERE	MALINKE	1747 (1631-1892)	0.21	FULAI	MALINKE	152 (675B-906)	0.06	CEU	MALINKE
BAMBARA	main.null	<0.01	1D	0.897	1	1	0.31	1341 (1167-1457)	0.1	GBR	MALINKE	0.22	SERERE	MALINKE	1718 (1529-1892)	0.2	FULAI	MALINKE	51B (1473B-819)	0.06	CEU	MALINKE
BANTU	main	<0.01	1D	0.447	0.995	0.999	0.05	7 (617B-602)	0.34	MALAWI	YORUBA	0.31	MALAWI	MALAWI	1776 (1295-1892)	0.43	MZIGUA	MALAWI	109B (718B-503)	0.36	MALAWI	YORUBA
BANTU	main.null	<0.01	1D	0.588	0.998	1	0.02	747B (1257B-311B)	0.5	MALAWI	YORUBA	0.29	MALAWI	MALAWI	1283 (1239-1892)	0.36	MZIGUA	MALAWI	312B (1098B-66)	0.48	MALAWI	YORUBA
CHONYI	main	<0.01	1D	0.989	0.981	1	0.18	1138 (1080-1182)	0.08	KHV	WASAMBAA	0.24	LUHYA	MALAWI	1254 (1167-1733)	0.2	MALAWI	WASAMBAA	1356B (3259B-950)	0.08	KHV	MZIGUA
CHONYI	main.null	<0.01	1D	0.988	0.975	1	0.24	1109 (1022-1138)	0.07	CDX	GIRIAMA	0.3	LUHYA	MALAWI	1283 (1167-1632)	0.13	CDX	WASAMBAA	254B (1827B-848)	0.06	KHV	MZIGUA
FULAI	main	<0.01	1D(2D)	0.969	1	1	0.82	1225 (1109-1254)	0.23	IBS	SEREHULE	0.44	WOLLOF	WOLLOF	1660 (1631-1878)	0.46	BAMBARA	IBS	239 (224-761)	0.19	IBS	WOLLOF
FULAI	main.null	<0.01	1D(2D)	0.966	1	1	0.82	1196 (1138-1254)	0.24	IBS	SEREHULE	0.4	WOLLOF	WOLLOF	1689 (1602-1791)	0.42	BAMBARA	WOLLOF	297 (7-558)	0.2	IBS	WOLLOF
FULAII	main	<0.01	1D(2D)	0.939	1	1	0.57	1399 (1312-1472)	0.13	GBR	MALINKE	0.39	SERERE	BAMBARA	1660 (1602-1834)	0.23	FULAI	MALINKE	65 (253B-876)	0.1	GBR	MALINKE
FULAII	main.null	<0.01	1D(2D)	0.939	1	1	0.56	1399 (1283-1457)	0.13	GBR	MALINKE	0.14	SERERE	MALINKE	1660 (1602-1834)	0.23	FULAI	MALINKE	51B (472B-906)	0.11	GBR	MALINKE
GIRIAMA	main	<0.01	1D	0.989	0.999	1	0.21	1196 (1138-1254)	0.1	OROMO	MZIGUA	0.18	SEMI-BANTU	MALAWI	1370 (1225-1685)	0.21	WASAMBAA	MZIGUA	210 (2679B-964)	0.1	OROMO	MZIGUA
GIRIAMA	main.null	<0.01	1D	0.989	1	1	0.28	1196 (1109-1240)	0.11	OROMO	MZIGUA	0.2	SEMI-BANTU	MALAWI	1399 (1268-1660)	0.19	WASAMBAA	MZIGUA	22B (1667B-819)	0.09	OROMO	MZIGUA
/GUI//GANA	main	<0.01	1D(2D)	0.904	0.995	1	0.40	1544 (1428-1631)	0.25	MALAWI	KARRETJIE	0.15	KHWE	AMAXHOSA	1747 (1544-1892)	0.2	MALAWI	JU/'HOANSI	877 (286B-1196)	0.28	KHWE	KARRETJIE
/GUI//GANA	main.null	<0.01	1D	0.865	0.995	1	0.33	1544 (1457-1631)	0.24	MALAWI	JU/'HOANSI	0.12	KHWE	AMAXHOSA	1834 (1660-1892)	0.19	MALAWI	JU/'HOANSI	935 (383-1196)	0.27	MALAWI	KARRETJIE
GUMUZ	main	<0.01	1D	0.785	0.987	0.999	0.26	1544 (1384-1718)	0.24	ARI	ANUAK	0.42	ANUAK	ANUAK	1747 (1631-1892)	0.27	ARI	ANUAK	1385B (3694B-1358)	0.21	WOLAYTA	ANUAK
GUMUZ	main.null	<0.01	1MW	0.748	0.974	0.996	0.14	1573 (1428-1747)	0.19	ARI	ANUAK	0.46	ANUAK	ANUAK	1718 (1631-1892)	0.22	ARI	ANUAK	1124B (3500B-1544)	0.29	WOLAYTA	ANUAK
JOLA	main	<0.01	1D	0.642	1	1	0.08	225B (1067B-1791)	0.18	FULAI	SERERE	0.49	MANDINKAI	MALINKE	1892 (1790-1892)	0.44	MANJAGO	WOLLOF	834B (2287B-169)	0.17	FULAI	SERERE
JOLA	main.null	<0.01	1D	0.652	1	1	0.04	1907B (3204B-674B)	0.11	GBR	SERERE	0.47	MANDINKAI	FULAII	1834 (1469-1892)	0.37	SERERE	WOLLOF	2487B (3044B-34B)	0.09	GBR	SERERE
JU/'HOANSI	main	<0.01	1MW	0.866	0.974	0.998	0.10	732 (616-993)	0.15	SOMALI	KARRETJIE	0.33	NAMA	KARRETJIE	1892 (1616-1892)	0.26	NAMA	KARRETJIE	587 (311-805)	0.16	SOMALI	KARRETJIE
JU/'HOANSI	main.null	<0.01	1D	0.832	0.986	1	0.08	558 (311-851)	0.11	SOMALI	KARRETJIE	0.48	!XUN	KARRETJIE	1805 (1108-1892)	0.15	KARRETJIE	!XUN	413 (586B-761)	0.13	SOMALI	KARRETJIE
KAMBE	main	<0.01	2D	0.989	0.987	1	0.46	1283 (1254-1399)	0.07	GBR	MZIGUA	0.34	LUHYA	MALAWI	1544 (1370-1776)	0.14	KAUMA	MZIGUA	761 (461B-1053)	0.06	GBR	MZIGUA
KAMBE	main.null	<0.01	2D	0.988	0.986	1	0.49	1225 (1210-1370)	0.07	GBR	MZIGUA	0.32	LUHYA	MALAWI	1602 (1457-1776)	0.16	KAUMA	MZIGUA	790 (296-1022)	0.06	GBR	MZIGUA

Continued on next page

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Figure 4-Source Data 2 Continued from previous page

Cluster	Analysis	P	res	max(R_1)	FQ_1	FQ_2	M	1D	1$D\alpha$1	1D src1	1D src2	1MW α	1MW src3	1MW src4	2D1	2$D1\alpha$	2D1 src1	2D1 src2	2D2	2$D2\alpha$	2D2 src1	2D2 src2
KARRETJIE	main	<0.01	1D(2D)	0.998	0.999	1	0.38	1776 (1747-1805)	0.1	GBR	/GUI//GANA	0.19	CDX	=KHOMANI	1805 (1747-1878)	0.1	GBR	/GUI//GANA	1602 (1123-1776)	0.15	AMAXHOSA	=KHOMANI
KARRETJIE	main.null	<0.01	1D	0.997	0.999	1	0.33	1776 (1776-1805)	0.1	GBR	/GUI//GANA	0.21	CDX	=KHOMANI	1805 (1747-1892)	0.09	GBR	/GUI//GANA	1660 (1092-1776)	0.37	AMAXHOSA	=KHOMANI
KASEM	main	<0.01	1D	0.305	0.997	1	0.03	819 (456-1329)	0.1	SEMI-BANTU	MOSSI	0.37	MOSSI	NAMKAM	1892 (1282-1892)	0.45	MOSSI	MOSSI	790 (10518B-1414)	0.07	SEMI-BANTU	MOSSI
KASEM	main.null	<0.01	1D	0.278	0.997	1	0.02	645 (87B-1138)	0.16	YORUBA	MOSSI	0.37	MOSSI	NAMKAM	1399 (1138-1892)	0.45	AKANS	MOSSI	616 (6103B-1312)	0.06	MALAWI	MOSSI
KAUMA	main	<0.01	1D	0.992	0.982	1	0.29	1225 (1167-1254)	0.06	GIH	MZIGUA	0.42	LUHYA	MALAWI	1515 (1341-1878)	0.2	KAMBE	MZIGUA	674 (138B-1080)	0.07	GIH	MZIGUA
KAUMA	main.null	<0.01	1D	0.991	0.986	1	0.31	1196 (1167-1254)	0.08	GIH	MZIGUA	0.44	WASAMBAA	MALAWI	1573 (1370-1834)	0.17	KAMBE	MZIGUA	761 (166-1022)	0.08	GIH	MZIGUA
=KHOMANI	main	<0.01	1D	0.998	0.999	1	0.18	1776 (1747-1805)	0.13	CEU	KARRETJIE	0.27	HERERO	KARRETJIE	1776 (1747-1820)	0.13	CEU	KARRETJIE	312B (356B-1762)	0.19	WOLAYTA	KARRETJIE
=KHOMANI	main.null	<0.01	1D	0.998	0.999	1	0.09	1776 (1747-1805)	0.13	CEU	KARRETJIE	0.27	HERERO	KARRETJIE	1776 (1747-1892)	0.13	CEU	KARRETJIE	1472B (1142B-1776)	0.21	IBS	KARRETJIE
KHWE	main	<0.01	1D	0.928	0.981	0.998	0.21	1341 (1225-1428)	0.41	/GUI//GANA	SEMI-BANTU	0.28	MZIGUA	SEBANTU	1486 (1370-1675)	0.43	/GUI//GANA	SEMI-BANTU	457B (2490B-703)	0.45	/GUI//GANA	SEMI-BANTU
KHWE	main.null	<0.01	1D	0.911	0.985	0.999	0.16	1312 (1152-1399)	0.4	JU/'HOANSI	SEMI-BANTU	0.27	MZIGUA	SEBANTU	1573 (1370-1820)	0.43	!XUN	MALAWI	268 (2009B-979)	0.42	/GUI//GANA	SEMI-BANTU
LUHYA	main	<0.01	2D	0.993	0.998	0.999	0.59	1370 (1341-1428)	0.28	SUDANESE	MZIGUA	0.46	WASAMBAA	MZIGUA	1486 (1428-1573)	0.25	SUDANESE	MZIGUA	65 (400B-616)	0.29	WASAMBAA	MZIGUA
LUHYA	main.null	<0.01	2D	0.991	0.998	0.999	0.59	1341 (1341-1399)	0.29	SUDANESE	MZIGUA	0.44	WASAMBAA	MZIGUA	1486 (1442-1544)	0.25	SUDANESE	MZIGUA	123 (313B-486)	0.26	ANUAK	MZIGUA
MALAWI	main	<0.01	1MW	0.895	0.929	0.992	0.08	587 (413-761)	0.21	SEMI-BANTU	MZIGUA	0.16	SEBANTU	MZIGUA	1225 (703-1892)	0.41	MZIGUA	MZIGUA	167B (2506B-500)	0.14	YORUBA	MZIGUA
MALAWI	main.null	<0.01	1MW	0.894	0.967	0.997	0.12	471 (340-631)	0.17	SEMI-BANTU	MZIGUA	0.16	AMAXHOSA	MZIGUA	1863 (1297-1892)	0.38	MZIGUA	MZIGUA	355 (21-457)	0.18	SEMI-BANTU	MZIGUA
MALINKE	main	<0.01	1D(2D)	0.91	1	1	0.50	1457 (1384-1602)	0.14	GBR	BAMBARA	0.31	SERERE	BAMBARA	1718 (1674-1834)	0.24	FULAI	FULAII	326 (544B-865)	0.11	GBR	BAMBARA
MALINKE	main.null	<0.01	1D(2D)	0.916	1	1	0.43	1370 (1355-1602)	0.12	GBR	BAMBARA	0.39	JOLA	BAMBARA	1718 (1645-1834)	0.23	FULAI	BAMBARA	355 (591B-747)	0.08	GBR	BAMBARA
MANDINKAI	main	<0.01	1D(2D)	0.806	1	1	0.47	1573 (1399-1689)	0.16	FULAI	JOLA	0.44	JOLA	SEREHULE	1805 (1747-1878)	0.19	FULAI	JOLA	312B (718B-416)	0.15	GBR	JOLA
MANDINKAI	main.null	<0.01	1D(2D)	0.796	1	1	0.42	1544 (1355-1675)	0.16	FULAI	JOLA	0.49	SEREHULE	JOLA	1805 (1747-1863)	0.19	FULAI	JOLA	428B (690B-256)	0.15	GBR	JOLA
MANDINKAII	main	<0.01	1D(2D)	0.925	1	1	0.39	1225 (1123-1341)	0.14	GBR	JOLA	0.3	JOLA	MALINKE	1631 (1486-1892)	0.2	FULAI	JOLA	51B (1244B-1066)	0.1	GBR	JOLA
MANDINKAII	main.null	<0.01	1D(2D)	0.923	1	1	0.38	1196 (1051-1327)	0.19	FULAI	JOLA	0.43	JOLA	FULAII	1573 (1500-1892)	0.24	FULAI	JOLA	109B (843B-1138)	0.17	GBR	JOLA
MANJAGO	main	<0.01	1D(2D)	0.779	0.985	1	0.50	1747 (1718-1834)	0.18	FULAI	JOLA	0.39	SERERE	SEREHULE	1892 (1790-1892)	0.23	FULAI	JOLA	413 (1590B-1718)	0.21	FULAI	JOLA
MANJAGO	main.null	<0.01	1D(2D)	0.766	0.989	1	0.48	1776 (1718-1863)	0.19	FULAI	JOLA	0.38	SERERE	WOLLOF	1892 (1805-1892)	0.17	FULAI	JOLA	326 (880B-1660)	0.2	FULAI	SERERE
MAASAI	main	<0.01	2D	0.918	0.992	1	0.61	1312 (1225-1486)	0.49	LUHYA	SOMALI	0.27	WABONDEI	LUHYA	1660 (1573-1747)	0.06	TIGRAY	LUHYA	254B (764B-239)	0.35	AFAR	LUHYA
MAASAI	main.null	<0.01	2D	0.911	0.992	1	0.56	1254 (1123-1443)	0.49	SOMALI	LUHYA	0.27	WABONDEI	LUHYA	1631 (1515-1747)	0.06	TIGRAY	LUHYA	109B (603B-297)	0.37	SOMALI	LUHYA
MOSSI	main	<0.01	1D	0.41	0.997	1	0.05	355 (97B-952)	0.21	YORUBA	KASEM	0.2	AKANS	KASEM	1892 (1485-1892)	0.27	KASEM	KASEM	442 (840B-979)	0.15	SEMI-BANTU	KASEM
MOSSI	main.null	<0.01	1D	0.417	0.998	1	0.04	65 (443B-573)	0.28	YORUBA	BAMBARA	0.18	NAMKAM	KASEM	1892 (1399-1892)	0.24	KASEM	KASEM	181 (501B-776)	0.2	SEMI-BANTU	KASEM
MZIGUA	main	<0.01	1D	0.988	0.999	1	0.19	1080 (1007-1138)	0.11	AFAR	WABONDEI	0.32	LUHYA	MALAWI	1196 (1109-1733)	0.13	WASAMBAA	WABONDEI	370B (1156B-993)	0.12	WASAMBAA	WABONDEI
MZIGUA	main.null	<0.01	1D	0.986	0.999	1	0.18	1022 (935-1095)	0.09	AFAR	WABONDEI	0.23	SEMI-BANTU	MALAWI	1341 (1108-1791)	0.17	WASAMBAA	WABONDEI	268 (385B-863)	0.1	AFAR	WABONDEI
NAMA	main	<0.01	1D(2D)	0.992	0.982	1	0.70	1805 (1805-1834)	0.3	HERERO	=KHOMANI	0.21	CEU	JU/'HOANSI	1834 (1834-1863)	0.31	HERERO	=KHOMANI	210 (152-935)	0.15	HERERO	=KHOMANI
NAMA	main.null	<0.01	1MW(2D)	0.991	0.93	0.999	0.52	1805 (1805-1863)	0.12	CEU	JU/'HOANSI	0.23	HERERO	=KHOMANI	1834 (1819-1892)	0.11	CEU	JU/'HOANSI	573B (428B-1153)	0.29	GBR	KARRETJIE
NAMKAM	main	<0.01	1D	0.197	0.996	0.999	0.02	616 (184B-1197)	0.11	SEMI-BANTU	MOSSI	0.15	MOSSI	KASEM	1892 (973-1892)	0.4	AKANS	MOSSI	268 (1458B-1196)	0.12	SEMI-BANTU	MOSSI
NAMKAM	main.null	<0.01	1D	0.18	0.999	1	0.03	123 (1243B-1037)	0.17	SEMI-BANTU	MOSSI	0.2	MOSSI	KASEM	1892 (920-1892)	0.17	YORUBA	MOSSI	515B (2417B-820)	0.28	YORUBA	MOSSI
OROMO	main	<0.01	2D	0.931	1	1	0.61	355 (152-558)	0.28	TSI	ARI	0.2	AMHARA	AFAR	1805 (1660-1892)	0.15	IBS	WOLAYTA	312B (690B-36B)	0.28	TSI	ARI
OROMO	main.null	<0.01	2D	0.922	1	1	0.50	239 (79-515)	0.27	TSI	ARI	0.21	AMHARA	AFAR	1834 (1674-1892)	0.17	IBS	WOLAYTA	283B (617B-51B)	0.27	TSI	ARI
SEBANTU	main	<0.01	1D(2D)	0.986	1	1	0.55	1167 (1051-1225)	0.29	KARRETJIE	MALAWI	0.28	AMAXHOSA	AMAXHOSA	1254 (1196-1399)	0.27	KARRETJIE	MALAWI	2951B (3604B-848B)	0.31	KARRETJIE	MALAWI
SEBANTU	main.null	<0.01	1D(2D)	0.982	1	1	0.57	1109 (1051-1196)	0.3	KARRETJIE	MALAWI	0.34	AMAXHOSA	AMAXHOSA	1254 (1196-1641)	0.28	KARRETJIE	MALAWI	2777B (4025B-390)	0.31	KARRETJIE	MALAWI
SEMI-BANTU	main	<0.01	1D	0.43	0.999	1	0.03	674 (192-1126)	0.2	MZIGUA	YORUBA	0.36	YORUBA	YORUBA	1167 NA	0.19	MZIGUA	YORUBA	7 NA	0.42	MALAWI	YORUBA
SEMI-BANTU	main.null	<0.01	1D	0.535	1	1	0.02	167B (965B-474)	0.27	MZIGUA	YORUBA	0.48	YORUBA	YORUBA	1080 NA	0.11	MZIGUA	YORUBA	1153B NA	0.27	MZIGUA	YORUBA
SEREHULE	main	<0.01	1D(2D)	0.943	1	1	0.36	1109 (935-1225)	0.12	GBR	JOLA	0.46	MANJAGO	BAMBARA	1689 (1515-1863)	0.25	FULAI	SERERE	413 (514B-921)	0.11	GBR	JOLA
SEREHULE	main.null	<0.01	1D	0.94	1	1	0.35	1080 (964-1225)	0.09	GBR	JOLA	0.41	JOLA	MALINKE	1544 (1326-1834)	0.22	FULAI	JOLA	22B (1204B-822)	0.08	GBR	JOLA
SERERE	main	<0.01	1D(2D)	0.885	1	1	0.41	1080 (761-1356)	0.14	GBR	JOLA	0.47	MANJAGO	FULAII	1602 (1500-1791)	0.24	FULAI	JOLA	776B (1740B-254)	0.08	GBR	JOLA
SERERE	main.null	<0.01	1D(2D)	0.878	1	1	0.38	1051 (761-1356)	0.14	GBR	JOLA	0.34	FULAII	MANJAGO	1602 (1515-1805)	0.24	FULAI	JOLA	950B (1986B-211)	0.08	GBR	JOLA

Continued on next page

Figure 4-Source Data 2 Continued from previous page

Cluster	Analysis	P	res	max(R_1)	FQ_1	FQ_2	M	$1D$	$1D\alpha 1$	$1D$ src1	$1D$ src2	$1MW$ α	$1MW$ src3	$1MW$ src4	$2D1$	$2D1\alpha$	$2D1$ src1	$2D1$ src2	$2D2$	$2D2\alpha$	$2D2$ src1	$2D2$ src2
SOMALI	main	<0.01	2D	0.933	1	1	0.60	268 (36-529)	0.38	ANUAK	TIGRAY	0.12	WOLAYTA	WOLAYTA	1573 (1370-1791)	0.18	MAASAI	WOLAYTA	921B (1458B-413B)	0.46	TIGRAY	GUMUZ
SOMALI	main.null	<0.01	2D	0.939	1	1	0.40	36 (196B-268)	0.39	ANUAK	TIGRAY	0.07	WASAMBAA	WOLAYTA	1573 (876-1805)	0.04	TSI	WOLAYTA	863B (1887B-399B)	0.47	TIGRAY	GUMUZ
SUDANESE	main	<0.01	1MW	0.789	0.782	0.999	0.21	1341 (1225-1660)	0.27	GUMUZ	ANUAK	0.25	ANUAK	ANUAK	1660 (1469-1892)	0.36	ANUAK	ANUAK	254B (979B-1284)	0.28	GUMUZ	ANUAK
SUDANESE	main.null	<0.01	1MW	0.658	0.856	0.999	0.13	1138 (1196-1689)	0.31	GUMUZ	ANUAK	0.19	ANUAK	ANUAK	1892 (1674-1892)	0.15	ANUAK	ANUAK	790 (8B-1515)	0.23	GUMUZ	ANUAK
HERERO	main	<0.01	1D(2D)	0.851	0.998	1	0.46	1631 (1747-1863)	0.41	SEMI-BANTU	NAMA	0.19	KAMBE	AMAXHOSA	1834 (1834-1892)	0.26	NAMA	AMAXHOSA	558 (298B-935)	0.43	NAMA	MALAWI
HERERO	main.null	<0.01	1D	0.816	0.998	1	0.34	1631 (1718-1863)	0.41	SEMI-BANTU	NAMA	0.19	KAMBE	AMAXHOSA	1834 (1805-1892)	0.24	NAMA	AMAXHOSA	674 (124B-979)	0.44	NAMA	SEMI-BANTU
TIGRAY	main	<0.01	1D	0.956	1	1	0.26	152 (51B-370)	0.32	TSI	ARI	0.31	AMHARA	AMHARA	1341 (819-1776)	0.21	IBS	OROMO	602B (1968B-138B)	0.32	TSI	ARI
TIGRAY	main.null	<0.01	1D	0.947	1	1	0.20	36 (196B-240)	0.35	TSI	ARI	0.42	AMHARA	AMHARA	1573 (862-1892)	0.16	IBS	AFAR	399B (1562B-94B)	0.35	TSI	ARI
WABONDEI	main	<0.01	2D	0.985	0.998	0.999	0.48	1138 (1080-1240)	0.11	TIGRAY	MZIGUA	0.5	MALAWI	WASAMBAA	1573 (1326-1834)	0.28	WASAMBAA	MZIGUA	703 (19-936)	0.1	TIGRAY	MZIGUA
WABONDEI	main.null	<0.01	2D	0.983	0.996	0.999	0.53	1109 (1051-1196)	0.1	AFAR	MZIGUA	0.49	WASAMBAA	MALAWI	1573 (1266-1820)	0.28	WASAMBAA	MZIGUA	587 (52B-863)	0.1	AFAR	MZIGUA
WASAMBAA	main	<0.01	1D	0.988	1	1	0.34	1312 (1254-1341)	0.14	TIGRAY	MZIGUA	0.3	LUHYA	MALAWI	1370 (1341-1834)	0.12	TIGRAY	MZIGUA	631B (1226B-1138)	0.16	WOLAYTA	MZIGUA
WASAMBAA	main.null	<0.01	1D	0.988	1	1	0.34	1254 (1225-1341)	0.15	AMHARA	MZIGUA	0.29	LUHYA	MALAWI	1486 (1384-1863)	0.15	AMHARA	MZIGUA	210 (283B-1066)	0.13	OROMO	MZIGUA
WOLAYTA	main	<0.01	1D	0.891	1	1	0.29	268 (138B-602)	0.22	TSI	ARI	0.26	OROMO	SOMALI	1312 (1138-1834)	0.14	TSI	SOMALI	1182B (1893B-144)	0.27	TSI	ARI
WOLAYTA	main.null	<0.01	1D	0.825	1	1	0.18	268 (8B-602)	0.22	TSI	ARI	0.26	OROMO	SOMALI	1457 (1137-1892)	0.15	TIGRAY	SOMALI	573B (1330B-388)	0.24	TSI	ARI
WOLLOF	main	<0.01	1D(2D)	0.938	1	1	0.63	1225 (1138-1341)	0.13	GBR	JOLA	0.4	MALINKE	JOLA	1631 (1544-1733)	0.19	FULAI	JOLA	254B (779B-355)	0.09	GBR	JOLA
WOLLOF	main.null	<0.01	1D(2D)	0.942	1	1	0.58	1167 (1080-1283)	0.12	GBR	JOLA	0.36	FULAII	MANDINKAI	1573 (1515-1762)	0.19	FULAI	JOLA	399B (951B-370)	0.08	GBR	JOLA
!XUN	main	<0.01	1D(2D)	0.957	0.999	1	0.39	1341 (1254-1399)	0.27	SEMI-BANTU	JU/'HOANSI	0.08	SOMALI	/GUI//GANA	1602 (1312-1892)	0.21	SEMI-BANTU	JU/'HOANSI	819 (1008B-1196)	0.17	SEMI-BANTU	JU/'HOANSI
!XUN	main.null	<0.01	1D	0.951	0.999	1	0.21	1312 (1254-1385)	0.28	SEMI-BANTU	JU/'HOANSI	0.07	SOMALI	/GUI//GANA	1805 (1341-1892)	0.15	SEMI-BANTU	JU/'HOANSI	1080 (329B-1167)	0.27	SEMI-BANTU	JU/'HOANSI
YORUBA	main	<0.01	1D	0.465	0.999	1	0.03	848 (338-1184)	0.48	SEMI-BANTU	AKANS	0.29	AKANS	AKANS	1167 (1036-1892)	0.49	AKANS	SEMI-BANTU	1501B (2896B-1182)	0.25	MOSSI	SEMI-BANTU
YORUBA	main.null	<0.01	1D	0.498	1	1	0.04	355 (429B-761)	0.37	SEMI-BANTU	AKANS	0.26	AKANS	AKANS	1109 (905-1892)	0.5	SEMI-BANTU	AKANS	2342B (4107B-778)	0.16	BANTU	AKANS

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Tables

bioRxiv preprint doi: https://doi.org/10.1101/038406; this version posted February 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Table 1. Evidence for admixture across the ethnic groups included in the study using f_3 tests and ALDER. For each group, we report the f_3 test with the most negative value. Source.1.f_3 and Source.2.f_3 refer to the two groups between which gene flow must have occured in the ancestors of the ethnic group under consideration. Groups without results are those where no negative f_3 statistic was found are shown with no results. For the ALDER analysis, we show the two sources of admixture, Source.2.LD and Source.1.LD, with the lowest reported admixture P-values. Dates for the ALDER events involving these groups are shown in generations, Date.gens, and in years, Date.

Ethnic.Group	Source.1.f_3	Source.2.f_3	f_3(Z-score)	Source.1.LD	Source.2.LD	P.value.LD	Date.gens	Date
JOLA								
MANJAGO	JOLA	GBR	-11.091	MAASAI	GIH	5.3e-08	3.27 +/- 0.52	1826CE (1811-1841CE)
SEREHULE	JOLA	TSI	-17.778	WOLAYTA	IBS	1.6e-05	30.98 +/- 5.81	1023CE (854-1191CE)
SERERE	JOLA	TSI	-12.360	SOMALI	TSI	0.013	21.61 +/- 5.46	1294CE (1136-1453CE)
WOLLOF	JOLA	TSI	-23.384	ARI	IBS	6.3e-15	23.00 +/- 2.73	1254CE (1175-1333CE)
MANDINKAI	JOLA	FULAI	-20.677					
MANDINKAII	JOLA	FULAI	-18.910	MZIGUA	AFAR	1.1e-09	24.15 +/- 3.51	1221CE (1119-1322CE)
FULAI	JOLA	TSI	-55.429	ARI	AMAXHOSA	2e-21	16.06 +/- 1.21	1455CE (1420-1490CE)
FULAII	FULAI	YORUBA	-25.483	MAASAI	AFAR	1.8e-10	8.69 +/- 1.22	1669CE (1634-1704CE)
BAMBARA	FULAI	AKANS	-10.291	MANDINKAII	IBS	0.049	15.27 +/- 4.21	1478CE (1356-1600CE)
MALINKE	FULAI	AKANS	-8.553					
MOSSI	AKANS	TSI	-3.905					
AKANS	NAMKAM	JU/'HOANSI	-9.038					
KASEM								
NAMKAM	KASEM	CHONYI	-1.890					
YORUBA	NAMKAM	JU/'HOANSI	-2.862					
BANTU	MOSSI	JU/'HOANSI	-13.894					
SEMI-BANTU	MOSSI	JU/'HOANSI	-14.844					
MALAWI	NAMKAM	JU/'HOANSI	-7.927	YORUBA	JU/'HOANSI	5.9e-07	38.47 +/- 5.71	805CE (640-971CE)
LUHYA	MALAWI	TIGRAY	-32.492	MANJAGO	JU/'HOANSI	1.2e-12	42.03 +/- 5.26	702CE (550-855CE)
CHONYI	MALAWI	TSI	-14.987	OROMO	TSI	4.6e-25	30.42 +/- 2.81	1039CE (957-1120CE)
KAUMA	MALAWI	TSI	-34.562	GUMUZ	CEU	2.4e-27	29.55 +/- 2.43	1064CE (994-1135CE)
KAMBE	MALAWI	TSI	-44.547	WOLAYTA	IBS	6.7e-17	15.07 +/- 1.69	1484CE (1435-1533CE)
GIRIAMA	MALAWI	TSI	-21.929	ARI	AMHARA	1.4e-08	23.28 +/- 3.16	1246CE (1154-1338CE)
MZIGUA	MALAWI	TSI	-34.835	MANDINKAI	IBS	6.9e-21	32.96 +/- 3.33	965CE (869-1062CE)
WABONDEI	MALAWI	TSI	-42.835	ANUAK	CEU	9.8e-20	30.71 +/- 3.19	1030CE (938-1123CE)
WASAMBAA	MALAWI	TIGRAY	-50.513	KASEM	ARI	1.9e-17	15.16 +/- 1.67	1481CE (1433-1530CE)
MAASAI	SUDANESE	TSI	-66.343	GUMUZ	GBR	9.8e-21	31.33 +/- 3.17	1012CE (920-1104CE)
SUDANESE				MANDINKAI	TIGRAY	0.036	7.76 +/- 2.10	1696CE (1635-1757CE)
SOMALI	SUDANESE	TSI	-67.795	GUMUZ	FIN	3.7e-11	68.34 +/- 9.14	61BCE (326B-204CE)
ARI	JU/'HOANSI	TSI	-23.214					
ANUAK	SUDANESE	ARI	-5.653					
GUMUZ				SOMALI	TSI	0.00019	5.78 +/- 1.19	1753CE (1719-1788CE)
AFAR	SUDANESE	TSI	-64.721					
TIGRAY	SUDANESE	TSI	-86.578	SEMI-BANTU	GBR	7e-31	79.21 +/- 6.59	376BCE (567B-185BCE)
AMHARA	SUDANESE	TSI	-87.571	BAMBARA	GBR	5.6e-18	76.91 +/- 7.81	309BCE (536B-83BCE)
WOLAYTA	ANUAK	TSI	-66.705	/GUI //GANA	CEU	6e-08	50.92 +/- 8.12	444CE (209-680CE)
OROMO	SUDANESE	TSI	-90.232	BANTU	GBR	1.5e-45	59.19 +/- 4.08	204CE (86-323CE)
HERERO	MALAWI	FIN	-5.953	AMHARA	GBR	9.1e-08	2.29 +/- 0.37	1855CE (1844-1865CE)
NAMA	JU/'HOANSI	CEU	-81.913	SUDANESE	JPT	8e-32	5.43 +/- 0.45	1764CE (1750-1777CE)
SEBANTU	MALAWI	KARRETJIE	-46.315					
AMAXHOSA	MOSSI	JU/'HOANSI	-39.259	ARI	WOLAYTA	0.019	27.01 +/- 6.35	1138CE (954-1322CE)
KARRETJIE	JU/'HOANSI	GBR	-46.547	SUDANESE	CDX	2.3e-28	4.88 +/- 0.42	1779CE (1767-1792CE)
≠KHOMANI	JU/'HOANSI	FIN	-86.201	NAMKAM	CHS	6.8e-47	5.54 +/- 0.38	1760CE (1749-1771CE)
KHWE	MOSSI	JU/'HOANSI	-55.836					
!XUN	YORUBA	JU/'HOANSI	-55.053	SEREHULE	SUDANESE	0.002	21.99 +/- 5.02	1283CE (1138-1429CE)
/GUI //GANA	MALAWI	JU/'HOANSI	-36.212	SUDANESE	ARI	0.021	7.92 +/- 2.06	1691CE (1632-1751CE)
JU/'HOANSI				ARI	GIH	2.5e-07	34.11 +/- 5.63	932CE (769-1095CE)