

A Rare Deep-Rooting D0 African Y-Chromosomal Haplogroup and Its Implications for the Expansion of Modern Humans Out of Africa

Marc Haber,* Abigail L. Jones,[†] Bruce A. Connell,[‡] Asan,[§] Elena Arciero,* Huanming Yang,^{§,**}
Mark G. Thomas,^{††} Yali Xue,* and Chris Tyler-Smith*¹

*The Wellcome Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK, [†]Liverpool Women's Hospital, Liverpool L8 7SS, UK, [‡]Glendon College, York University, Toronto, Ontario M4N 3N6, Canada, [§]BGI-Shenzhen, Shenzhen 518083, China, **James D.

Watson Institute of Genome Science, 310008 Hangzhou, China, and ^{††}Research Department of Genetics, Evolution and Environment, University College London, WC1E 6BT, UK, and UCL Genetics Institute, University College London, WC1E 6BT, UK

ORCID ID: 0000-0003-1000-1448 (M.H.)

ABSTRACT Present-day humans outside Africa descend mainly from a single expansion out ~50,000–70,000 years ago, but many details of this expansion remain unclear, including the history of the male-specific Y chromosome at this time. Here, we reinvestigate a rare deep-rooting African Y-chromosomal lineage by sequencing the whole genomes of three Nigerian men described in 2003 as carrying haplogroup DE* Y chromosomes, and analyzing them in the context of a calibrated worldwide Y-chromosomal phylogeny. We confirm that these three chromosomes do represent a deep-rooting DE lineage, branching close to the DE bifurcation, but place them on the D branch as an outgroup to all other known D chromosomes, and designate the new lineage D0. We consider three models for the expansion of Y lineages out of Africa ~50,000–100,000 years ago, incorporating migration back to Africa where necessary to explain present-day Y-lineage distributions. Considering both the Y-chromosomal phylogenetic structure incorporating the D0 lineage, and published evidence for modern humans outside Africa, the most favored model involves an origin of the DE lineage within Africa with D0 and E remaining there, and migration out of the three lineages (C, D, and FT) that now form the vast majority of non-African Y chromosomes. The exit took place 50,300–81,000 years ago (latest date for FT lineage expansion outside Africa – earliest date for the D/D0 lineage split inside Africa), and most likely 50,300–59,400 years ago (considering Neanderthal admixture). This work resolves a long-running debate about Y-chromosomal out-of-Africa/back-to-Africa migrations, and provides insights into the out-of-Africa expansion more generally.

KEYWORDS Human Y chromosome; YAP+ Y chromosomes; phylogeography; out-of-Africa migration

HUMANS outside Africa derive most of their genetic ancestry from a single migration event 50,000–70,000 years ago, according to the current model supported by genetic data from genome-wide (Mallick *et al.* 2016; Pagani

et al. 2016), mitochondrial DNA (mtDNA) (van Oven and Kayser 2009), and Y-chromosomal (Wei *et al.* 2013; Hallast *et al.* 2015; Karmin *et al.* 2015; Poznik *et al.* 2016) analyses. The migrating population carried only a small subset of African genetic diversity, particularly strikingly for the nonrecombining mtDNA and Y chromosome where robust calibrated high-resolution phylogenies can be constructed, and in each case all non-African lineages descend from a single African lineage, L3 for mtDNA or CT-M168 for the Y chromosome. Yet there has been a long-running debate about the early spread of Y-chromosomal lineages because their current distributions do not fit a simple phylogeographical model. The CT-M168 branch diverged within a short time interval into three lineages (C-M130, DE-M145, and FT-M89), and just a few thousand years later the lineage DE-M145 further split

Copyright © Haber *et al.*

doi: <https://doi.org/10.1534/genetics.119.302368>

Manuscript received February 25, 2019; accepted for publication June 10, 2019; published Early Online June 13, 2019.

Available freely online through the author-supported open access option.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material available at FigShare: <https://doi.org/10.25386/genetics.8267861>.

¹Corresponding author: The Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK. E-mail: cts@sanger.ac.uk

into D-M174 and E-M96 (Poznik *et al.* 2016), illustrated in Supplemental Material, Figure S1. Thus, around the time of the expansion out of Africa, between one (CT-M168) and four (C-M130, D-M174, E-M96, and F-M89) of the known extant non-African lineages were in existence (plus additional African lineages). The complexity arises because three of these four early lineages (C-M130, D-M174, and FT-M89) are exclusively non-African, apart from those entering Africa through recent gene flow; while the fourth lineage (E-M96) is largely African, where it constitutes the major lineage in most African populations. The debate began in the absence of reliable calibration, and these distributions were interpreted as arising in two contrasting ways: (1) an Asian origin of DE-M145 (also known as the YAP+ lineage), implying migration of CT-M168 out of Africa followed by divergence into the four lineages outside Africa and then migration of E-M96 back to Africa (Altheide and Hammer 1997; Hammer *et al.* 1998; Bravi *et al.* 2000), or (2) an African origin of DE-M145, implying divergence of CT-M168 within Africa followed by migration of C-M130, D-M174, and FT-M89 out (Underhill and Roseman 2001; Underhill *et al.* 2001). The first scenario requires two intercontinental lineage migrations, while the second requires three and is thus slightly less parsimonious.

An additional very rare haplogroup, DE*, carrying variants that define DE but none of those that define D or E individually, added to this complexity. First identified in 5 out of 1247 Nigerians within a worldwide study of >8000 men (Weale *et al.* 2003), DE* chromosomes were subsequently reported in a single man among 282 from Guinea-Bissau in West Africa (Rosa *et al.* 2007) and in 2 out of 722 Tibetans within a study of 5783 East Asians (Shi *et al.* 2008). While the phylogeographic significance of these rare lineages was immediately recognized, their interpretation was hindered by the incomplete resolution of the phylogenetic branching pattern and the possibility that they might originate from back-mutations at the small numbers of variants used to define the key D and E haplogroups, or genotyping errors rather than representing deeply divergent lineages, plus the lack of a robust timescale. Large-scale sequencing of Y chromosomes has now provided both the phylogenetic resolution and the timescale needed (Wei *et al.* 2013; Hallast *et al.* 2015; Karmin *et al.* 2015; Poznik *et al.* 2016), so we have therefore reinvestigated the original Nigerian DE* chromosomes using whole-genome sequencing to clarify their phylogenetic position. We then consider the implications for the out-of-Africa/back-to-Africa debate related to Y-chromosomal lineages, and the expansion out of Africa more generally.

Materials and Methods

Samples and sequencing

We analyzed five DE* samples described previously (Weale *et al.* 2003), in the context of published worldwide

Y-chromosomal sequences including Japanese D and many E Y chromosomes (Mallick *et al.* 2016). We also included four haplogroup D samples from Tibet (Xue *et al.* 2006), which were newly sequenced for this study; the Japanese and Tibetan D chromosomes represent the deepest known split within D, since Andamanese D chromosomes lie on the same branch as the Japanese (Mondal *et al.* 2017).

Sequencing of the Nigerian samples was carried out at the Wellcome Sanger institute on the Illumina HiSeq X Ten platform (paired-end read length 150 bp) to a Y-chromosome mean coverage of $\sim 16\times$. Sequences were processed using biobambam version 2.0.79 to remove adapters, mark duplicates, and sort reads. bwa-mem version 0.7.16a was used to map the reads to the *hs37d5* reference genome. We found that two pairs of individuals were likely duplicates (Figure S2) and thus one of each pair was removed, leaving three Nigerian individuals for further analysis. The four individuals from Tibet were sequenced in the same way to a Y-chromosome mean coverage of $\sim 18\times$.

For comparative data from other haplogroups, we obtained Y-chromosome bam files for 173 males representing worldwide populations from the Simons Genome Diversity Project (Mallick *et al.* 2016).

Data analysis

Y-chromosome genotypes were called jointly from all 180 samples using FreeBayes v1.2.0 (Garrison and Marth 2012) with the arguments “-report-monomorphic” and “-ploidy 1.” Calling was restricted to 10.3 Mb of the Y chromosome previously determined to be accessible to short-read sequencing (Poznik *et al.* 2013). Then sites with depth across all samples <1900 or >11,500 (corresponding to DP/2 or DP*3), or missing in >20% of the samples, were filtered. In individuals, alleles with DP <5 or GQ <30 were excluded, and if multiple alleles were observed at a position, the fraction of reads supporting the called allele was required to be >0.8.

Genome-wide genotypes from the Nigerian samples were called using BCFtools version 1.6 (bcftools mpileup -C50 -q30 -Q30 | bcftools call -c), then merged with data from ~ 2500 people genotyped on the Affymetrix Human Origins array (Patterson *et al.* 2012; Lazaridis *et al.* 2016). Principal Component Analysis (PCA) using genome-wide SNPs was performed using EIGENSOFT v7.2.1 (Patterson *et al.* 2006) and plotted using R (R Core Team 2017).

We inferred a maximum likelihood phylogeny of Y chromosomes using RAxML v8.2.10 (Stamatakis 2014) with the arguments “-m ASC_GTRGAMMA” and “-asc-corr=stamatakis,” using only variable sites with QUAL ≥ 1 , and selecting the tree with the best likelihood from 100 runs, then replicating the tree 1000 times for bootstrap values. The tree was plotted using Interactive Tree Of Life (iTOL) v3 (Letunic and Bork 2016) and annotated with haplogroup names assigned using yHaplo (Poznik 2016) from SNPs reported by the International Society of Genetic Genealogy (ISOGG v11.01).

The ages of the internal nodes in the tree were estimated using the ρ statistic (Forster *et al.* 1996), the standard

approach for the Y chromosome. We defined the ancestral state of a site by assigning alleles as ancestral when they were monomorphic in the nine samples belonging to the A and B haplogroups in our data set. We then determined the age of a node as follows: Having an ancestral node leading to two clades, we select one sample from each clade and divide the number of derived variants found in the first sample but absent from the second, by the total number of sites having the ancestral state in both samples. We compare all possible pairs under a node and report the average value of divergence times in units of years by applying a point mutation rate of 0.76×10^{-9} mutations per site per year (Fu *et al.* 2014). We report 95% confidence intervals of the divergence times based on the 95% highest posterior density when estimating the mutation rate ($0.67\text{--}0.86 \times 10^{-9}$) (Fu *et al.* 2014). This model assumes that mutations accumulated on the chromosomes in the different lineages at similar rates, and thus expects all individuals in our data set to have comparable branch lengths from the AB root. But we found considerable differences among individuals in the number of their derived mutations from the root. This heterogeneity in the accumulation of mutations has been previously reported (Scozzari *et al.* 2014; Barbieri *et al.* 2016) and appears to be haplogroup-specific (Figure S3), and therefore in our divergence time estimates, we calibrate all lineages to have identical branch length from the root, equal to the average branch length estimated from all individuals in our data set. We first calculated the average number of mutations which accumulated on the branches of all individuals in our data set and found 768.59 derived mutations on average from the root (corresponding to $\sim 100,000$ years). We then derived a calibration coefficient α for each individual by dividing 768.59 by the normalized (in 10,000,000 bp) number of derived mutations an individual has accumulated from the root. And thus for calibrating the branches' length between any two samples when calculating the split times, we multiply α by the number of derived variants found in the first sample but absent from the second.

Data availability statement

New sequence data from the Nigerian samples are available through the European Genome-phenome Archive (EGA) under study accession number EGAS00001002674, and for Tibetan samples under study accession number EGAS00001003500.

Three supplemental figures and two supplemental tables accompany this paper:

Figure S1 Y-chromosomal phylogeny as understood before the current study.

Figure S2 PCA of worldwide populations.

Figure S3 Number of mutations from the AB root.

Table S1 SNPs defining the D0 haplogroup.

Table S2 Split-time estimates using the ρ statistic.

Supplemental material available at FigShare: <https://doi.org/10.25386/genetics.8267861>.

Results

Construction of a calibrated Y-chromosomal phylogeny

We constructed a series of phylogenetic trees based on all the Y-chromosomal sequences in our data set, or subsets of them. All showed a consistent structure, in which the Nigerian DE* chromosomes formed a clade branching from the DE lineage close to the divergence of D and E chromosomes (Figure 1A) in comparison with a set of Y chromosomes representing most of the world (Figure 1B). The Nigerian chromosomes had 489 derived SNPs exclusive to their branch in addition to a large deletion spanning $\sim 118,000$ bp (Y:28,457,736–28,576,276). All DE-M154 chromosomes shared 29 SNPs. The Nigerian chromosomes shared seven SNPs with other D chromosomes, one SNP with E chromosomes, one SNP with C1b2a chromosomes, and one SNP with an F2 chromosome (Table S1). The reads overlapping these SNPs were visually investigated using the Integrative Genomics Viewer (IGV) version 2.4.10 and seen to support the calls. We consider sharing of a single SNP as a recurrent mutation in different lineages and interpret the Nigerian chromosomes as lying on the D lineage, diverging from other D chromosomes at 71,400 years ago (Figure 1C), very soon after its divergence from E at 73,200 years ago. We name the lineage formed by the Nigerian samples D0, to reflect its position on the tree and avoid the need to rename all the other D lineages.

The three D0 chromosomes are distinguishable from one another, and have a coalescence time of ~ 2500 years (Figure 1C), consistent with their collection from different villages, languages, ethnic backgrounds, and paternal birthplaces (Weale *et al.* 2003). The autosomal genomes of these individuals confirm their genetic ancestry as West Africans (Figure S2).

Models for the expansion of Y-chromosomal lineages out of Africa

The updated phylogeny including the D0 lineage adds two key pieces of information to the debate about the phylogeography of the Y lineages $\sim 50,000\text{--}100,000$ years ago and the mode of expansion out of Africa. First, it increases the number of relevant lineages at this early time period from four to five, and second, it provides a reliable timescale for the branching times of these lineages, and thus for the lineages in existence at any particular time point.

In the phylogeny (Figure 1A), the DE lineage now contains three, rather than two, early sublineages: one exclusively African (D0), one mainly African (E), and one exclusively non-African (D). We therefore consider the implications of this revised structure for interpreting the present-day Y phylogeography as the result of male movements at different times between 28,000 and 100,000 years ago (Figure 2). To do this, we need to calibrate the phylogeny, and for this use the ancient-DNA-based mutation rate (Fu *et al.* 2014), which has been widely adopted (*e.g.*, Poznik *et al.* 2016); we consider in the *Discussion* the implications of alternative

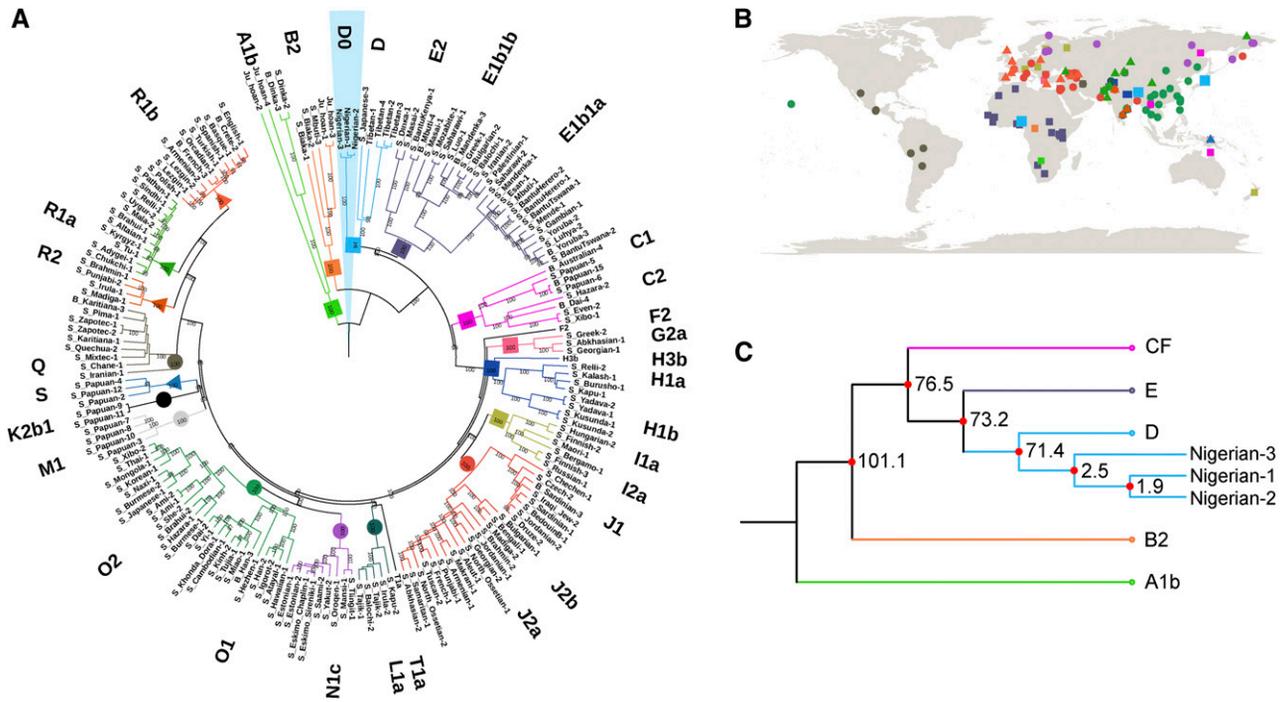


Figure 1 Y Chromosome phylogenetic tree from worldwide samples. (A) A maximum-likelihood tree of 180 Y-chromosome sequences from worldwide populations. Different branch colors and symbols represent different haplogroups assigned based on ISOGG v11.01. The Nigerian chromosomes sequenced in this study are highlighted in blue and assigned to the novel D0 haplogroup. Bootstrap values from 1000 replications are shown on the branches. (B) Map showing location of the studied individuals with colored symbols reflecting the haplogroups assigned in A. The clade consisting of the D0 and D haplogroups is represented by blue squares and is observed in Africa and East Asia. (C) Ages of the nodes leading to haplogroup D0 in the phylogenetic tree (point estimates; branch lengths are not to scale). Haplogroups D0 and D are estimated to have split 71,400 (63,100–81,000) years ago while the D0 individuals in this study coalesced 2500 (2200–2800) years ago.

mutation rates and some of the other simplifying assumptions we make here.

We consider three scenarios based on our split-time point estimates of the Y-chromosomal lineages (Table S2). First, between 101,000 years ago (divergence of the B and CT lineages) and 77,000 years ago (divergence of the DE and CF lineages) only one lineage with present-day non-African descendants is present in the phylogeny (CT; Figure 2A), so present-day Y-lineage distributions could be explained by migration of the single lineage CT out of Africa, followed by back-migration of the D0 and E lineages between 71,000 years ago (origin of D0) and 59,000 years ago (divergence of E within Africa) (Figure 2B). This and all other scenarios require migration out of E-M35 after 47,000 years ago (its origin) and before 28,500 years ago (its divergence) to explain its presence outside Africa (Figure 2, B–D). Second, between 76,000 years ago (divergence between C and FT) and 73,000 years ago (divergence between D and E), three relevant lineages are present (the C, DE, and FT lineages, Figure 2A), so migration out of these three followed by back-migration of D0 and E as above (Figure 2C) would explain the distributions. Third, between 71,000 years ago (split of D and of D0) and 57,000 years ago (divergence within FT), five relevant lineages are present, and migration out of three of these (C, D, and FT) would explain the present-day distributions without requiring back-migration (Fig-

ure 2D). For simplicity, we do not include the short intervals between these three scenarios of 500 years and 1800 years (Figure 2A and Table S2).

Discussion

The new D0 data presented in this work are based on just three Y chromosomes, but have far-ranging implications for the structure of the Y-chromosomal phylogeny and hence male movements and migration out of Africa more generally. Our phylogenetic results are consistent with three scenarios (Figure 2, B–D), and we now consider some of the complexities associated with these, and how they fit with nongenetic data.

Complexities arise because although the phylogenetic structure, including the branching order, is very robust (Wei *et al.* 2013; Hallast *et al.* 2015; Karmin *et al.* 2015; Poznik *et al.* 2016), its calibration depends entirely on the mutation rate used. The mutation rate chosen above, based on the number of mutations “missing” in a 45,000-year-old Siberian Y chromosome (Fu *et al.* 2014), has been widely adopted (Poznik *et al.* 2016; Balanovsky 2017), but a large-scale study of Icelandic pedigrees encompassing the last few centuries suggested a rate ~14% faster (Helgason *et al.* 2015). This faster mutation rate would translate directly into 14% more recent time estimates so that, for example, the Y-chromosome movements out of Africa in the three

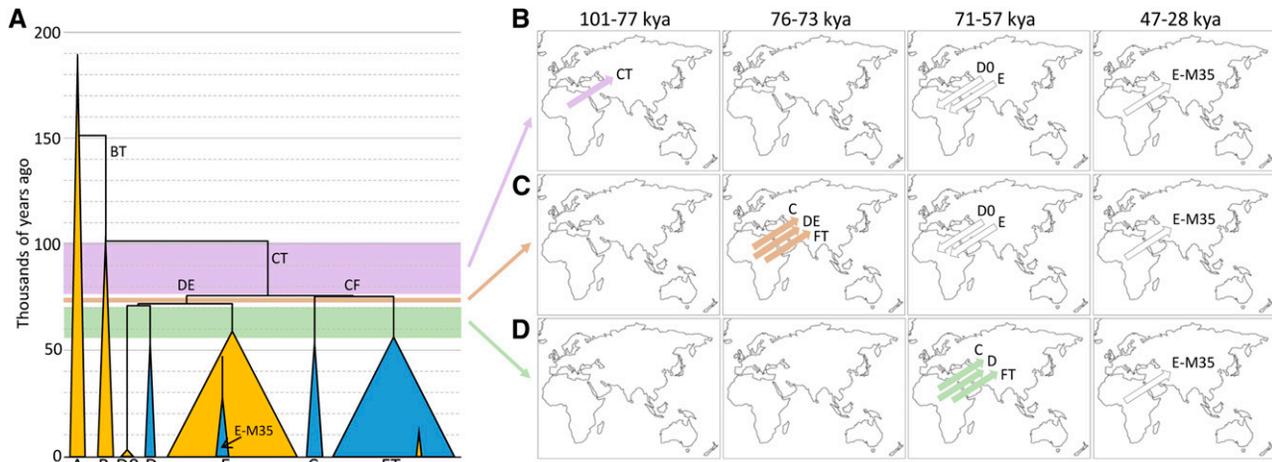


Figure 2 Models for the early movements of Y-chromosomal haplogroups out of Africa and back. (A) Simplified Y-chromosomal phylogeny showing the key lineages, including D0. Lineages currently located in Africa are colored yellow; those currently outside Africa are blue. Triangle widths are not meaningful, except that they show that E and FT are the predominant lineages inside and outside Africa, respectively. The small orange triangle in FT represents the R1b-V88 back-to-Africa migration that took place after the time period considered here. (B–D) Models for lineage movements that could lead to the present-day African or non-African distributions of lineages, using point estimates of dates derived from the phylogeny (see Table S2). The three models represent migrations out of Africa at different time intervals, indicated by the purple, brown, and green shading in A. Arrows in B–D indicate intercontinental movements and their direction, but do not represent particular locations or routes. The first colored arrow(s) represent the lineage(s) that migrated out, during the time period shown at the top of the maps. Additional uncolored arrows represent subsequent migrations and their time intervals needed to produce the present-day distributions.

scenarios presented above would be 87,000–66,000, 65,000–63,000, and 61,000–49,000 years ago, respectively. These differences between mutation rates inferred in different ways should be seen within the context of a wider debate about human mutation rates, previously based largely on autosomal data (Scally and Durbin 2012). Each mutation rate is also accompanied by its own uncertainty, leading to the 95% confidence intervals in Table S2, which include the mutation rate uncertainty. We also assume that the mutation rate is constant over time and does not differ between lineages. The first assumption is very reasonable for the time period of most interest here, 50,000–60,000 years, when the mutation rate averaged over 45,000 years (Fu *et al.* 2014) is used. A flexible mutation rate that assumed a real increase in recent times would have little influence on these estimates since the Fu *et al.* rate already includes the last few centuries. Differences in mutation rate between lineages need further investigation, but would not be sufficient to affect the scenarios presented in Figure 2. For these reasons, we believe that the Fu *et al.* rate, averaged over 45,000 years, is the appropriate one to use for the times of interest here.

These genetic times can be compared with dates from nongenetic sources for modern humans outside Africa. The 45,000-year-old Siberian fossil (Fu *et al.* 2014) was reliably dated using carbon-14, while a ~43,000-year-old fragment of human maxilla from the Kent's Cavern site in the UK was dated using Bayesian modeling of stratigraphic, chronological, and archaeological data (Higham *et al.* 2011). Archaeological deposits at Boodie Cave in Australia were dated to ~50,000 years ago using optically stimulated luminescence

(Veth *et al.* 2017). Thus, there is strong support for the widespread presence of modern humans outside Africa 45,000–50,000 years ago. Earlier dates have also been reported, for example the Madjedbebe rock shelter in northern Australia dated by optically stimulated luminescence to at least 65,000 years ago (Clarkson *et al.* 2017), a modern human cranium from Tam Pa Ling, Laos was dated by Uranium-Thorium to ~63,000 years ago (Demeter *et al.* 2012), and 80 teeth from Fuyan Cave in southern China dated using the same method to 80,000–120,000 years ago (Liu *et al.* 2015), raising the possibility of a substantially earlier exit (Bae *et al.* 2017). Such early archaeological dates also, however, raise the question of whether or not the humans associated with them contributed genetically to present-day populations (Mallick *et al.* 2016; Pagani *et al.* 2016). Archaeological data alone therefore do not provide an unequivocal date for the migration of the ancestors of present-day humans out of Africa.

All non-Africans carry ~2% Neanderthal DNA in their genomes (Green *et al.* 2010), and Neanderthal fossils have only been reported outside Africa. The geographical distribution of Neanderthals thus suggests that mixing probably occurred outside Africa, and the ubiquitous presence of Neanderthal DNA in present-day non-Africans is most easily explained if the mixing took place once, soon after the migration out. This mixing has been dated with some precision using the length of the introgressed segments in the 45,000-year-old (43,210–46,880 years) Siberian male (Ust'-Ishim) to 232–430 generations before he lived, *i.e.*, 49,900–59,400 years ago assuming a generation time of 29 years (Fu *et al.* 2014). If this date represented the time of the migration out of Africa, it would

exclude the first two scenarios (Figure 2, B and C). Thus, the combination of Y phylogenetic structure and dating of the out-of-Africa migration based on the 45,000-year-old Siberian fossil (Fu *et al.* 2014) favors the third scenario (Figure 2D) involving the migration out of C, D, and FT between 50,300 years ago (lower bound of the FT diversification, Table S2) and 59,400 years ago (upper bound of the introgression; see Figure 3), which is in accordance with suggested models incorporating an African origin of the DE lineages (Underhill and Roseman 2001; Underhill *et al.* 2001). According to this interpretation, the reported Tibetan DE* chromosomes (Shi *et al.* 2008) would most likely represent back-mutations or genotyping errors at the one SNP used to define haplogroup D, but require further investigation.

mtDNA sequences also provide a robust phylogeny which demonstrates that non-African mtDNAs descend from a single African branch with rapid diversification outside Africa into the M and N lineages and many subsequent branches (Ingman *et al.* 2000; Devèze *et al.* 2019). Dating using ancient mtDNA suggests a separation of non-African from African lineages after 62,000–95,000 years ago (Fu *et al.* 2013), while an analysis of present-day mtDNAs suggested divergence outside Africa 57,000–65,000 years ago (Fernandes *et al.* 2012). These estimates are based on <1% of the sequence length used from the Y chromosome but are nonetheless very consistent.

This discussion has thus far assumed that present-day distributions of Y haplogroups are relevant to events 50,000–100,000 years ago and thus that Y phylogeography carries information about the major migration out of Africa. Ancient population structure within Africa that separated C, D, and FT from other Y haplogroups beginning after 76,000 years ago with migration out only 50,000–59,000 years ago would also fit the evidence presented above. Present-day Y-chromosomal structure in Africa has been massively shaped by events in the last 10,000 years, including the Bantu-speaker expansion in central and southern Africa (Poznik *et al.* 2016; Patin *et al.* 2017) and entry of Eurasian lineages into northern and central Africa (Haber *et al.* 2016; D’Atanasio *et al.* 2018), and is thus a poor guide to structure before 10,000 years ago. Despite this, it is striking that western central Africa is the location of the deepest-rooting A00 lineage in Cameroon (Mendez *et al.* 2013), a major location of the A0 lineage in Cameroon, The Gambia, and Ghana (Scozzari *et al.* 2014; Poznik *et al.* 2016) and the D0 lineage in Nigeria and Guinea-Bissau (Weale *et al.* 2003; Rosa *et al.* 2007). This retention of the deepest Y-chromosomal diversity in western central Africa contrasts with the autosomal genetic structure, where the deepest roots have been reported in southern African hunter-gatherers (Gronau *et al.* 2011; Schlebusch *et al.* 2012, 2017; Veeramah *et al.* 2012; Mallick *et al.* 2016; Skoglund *et al.* 2017), perhaps supporting the hypothesis of deep population structure (Henn *et al.* 2018; Scerri *et al.* 2018). Analysis of ancient African DNA from 50,000 to 100,000 years ago would provide consid-

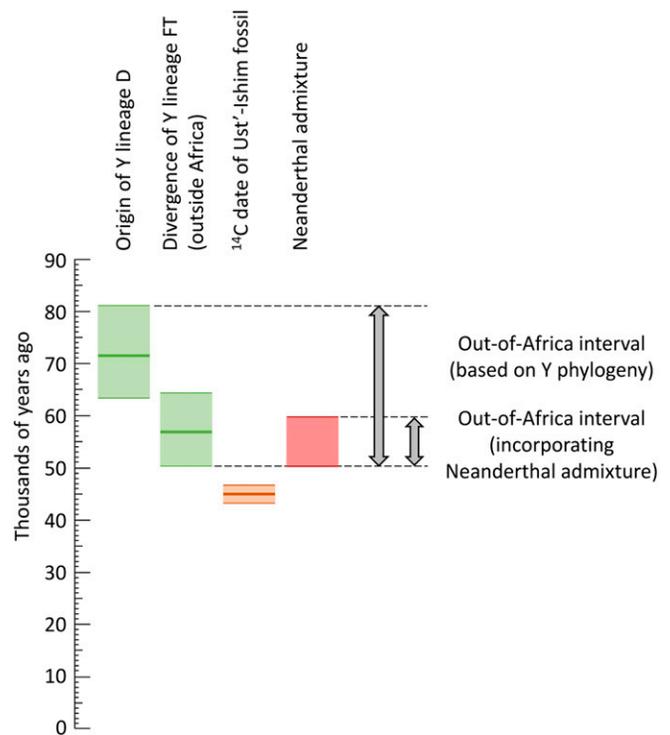


Figure 3 Estimation of the time of the out-of-Africa migration incorporating information from Y-chromosomal lineages (green, this work), archaeological dates (brown, Fu *et al.* 2014), and ancient DNA (red, Fu *et al.* 2014).

erably more information on Y-haplogroup distributions at this time, but is not currently available. In the meantime, further focus on present-day Y-chromosomal lineages in central and western Africa to understand more about deep African lineages seems warranted, and this current study illustrates the broad insights that can sometimes be revealed by very rare lineages.

In conclusion, sequencing of the D0 Y chromosomes and placement of them on a calibrated Y-chromosomal phylogeny identify the most likely model of Y-chromosomal exit from Africa: an origin of the DE lineage inside Africa and expansion out of the C, D, and FT lineages. It suggests an exit time interval that overlaps with the time of Neanderthal admixture estimated from autosomal analyses, and slightly refines it. These findings are consistent with a shared history of Y chromosomes and autosomes, and illustrate how study of Y lineages may lead to general new insights.

Acknowledgments

We thank the sample donors for making this work possible, and the Sanger DNA Pipelines Bespoke Sequencing Team for especial efforts in generating sequences from small amounts of DNA. Our work was supported by Wellcome (098051).

Literature Cited

- Altheide, T. K., and M. F. Hammer, 1997 Evidence for a possible Asian origin of YAP+ Y chromosomes. *Am. J. Hum. Genet.* 61: 462–466. [https://doi.org/10.1016/S0002-9297\(07\)64077-4](https://doi.org/10.1016/S0002-9297(07)64077-4)
- Bae, C. J., K. Douka, and M. D. Petraglia, 2017 On the origin of modern humans: Asian perspectives. *Science* 358: 1269. <https://doi.org/10.1126/science.aai9067>
- Balanovsky, O., 2017 Toward a consensus on SNP and STR mutation rates on the human Y-chromosome. *Hum. Genet.* 136: 575–590. <https://doi.org/10.1007/s00439-017-1805-8>
- Barbieri, C., A. Hübner, E. Macholdt, S. Ni, S. Lippold *et al.*, 2016 Refining the Y chromosome phylogeny with southern African sequences. *Hum. Genet.* 135: 541–553. <https://doi.org/10.1007/s00439-016-1651-0>
- Bravi, C. M., G. Bailliet, V. L. Martinez-Marignac, and N. O. Bianchi, 2000 Origin of YAP+ lineages of the human Y-chromosome. *Am. J. Phys. Anthropol.* 112: 149–158. [https://doi.org/10.1002/\(SICI\)1096-8644\(2000\)112:2<149::AID-AJPA2>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1096-8644(2000)112:2<149::AID-AJPA2>3.0.CO;2-M)
- Clarkson, C., Z. Jacobs, B. Marwick, R. Fullagar, L. Wallis *et al.*, 2017 Human occupation of northern Australia by 65,000 years ago. *Nature* 547: 306–310. <https://doi.org/10.1038/nature22968>
- D'Atanasio, E., B. Trombetta, M. Bonito, A. Finocchio, G. Di Vito *et al.*, 2018 The peopling of the last Green Sahara revealed by high-coverage resequencing of trans-Saharan patrilineages. *Genome Biol.* 19: 20. <https://doi.org/10.1186/s13059-018-1393-5>
- Demeter, F., L. L. Shackelford, A. M. Bacon, P. Düringer, K. Westaway *et al.*, 2012 Anatomically modern human in Southeast Asia (Laos) by 46 ka. *Proc. Natl. Acad. Sci. USA* 109: 14375–14380. <https://doi.org/10.1073/pnas.1208104109>
- Devièse, T., D. Massilani, S. Yi, D. Comeskey, S. Nagel *et al.*, 2019 Compound-specific radiocarbon dating and mitochondrial DNA analysis of the Pleistocene hominin from Salkhit Mongolia. *Nat. Commun.* 10: 274. <https://doi.org/10.1038/s41467-018-08018-8>
- Fernandes, V., F. Alshamali, M. Alves, M. D. Costa, J. B. Pereira *et al.*, 2012 The Arabian cradle: mitochondrial relicts of the first steps along the southern route out of Africa. *Am. J. Hum. Genet.* 90: 347–355. <https://doi.org/10.1016/j.ajhg.2011.12.010>
- Forster, P., R. Harding, A. Torroni, and H. J. Bandelt, 1996 Origin and evolution of Native American mtDNA variation: a reappraisal. *Am. J. Hum. Genet.* 59: 935–945.
- Fu, Q., A. Mittnik, P. L. F. Johnson, K. Bos, M. Lari *et al.*, 2013 A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* 23: 553–559. <https://doi.org/10.1016/j.cub.2013.02.044>
- Fu, Q., H. Li, P. Moorjani, F. Jay, S. M. Slepchenko *et al.*, 2014 Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 514: 445–449. <https://doi.org/10.1038/nature13810>
- Garrison, E., and G. Marth, 2012 Haplotype-based variant detection from short-read sequencing. [arXiv:1207.3907](https://arxiv.org/abs/1207.3907).
- Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel *et al.*, 2010 A draft sequence of the Neandertal genome. *Science* 328: 710–722. <https://doi.org/10.1126/science.1188021>
- Gronau, I., M. J. Hubisz, B. Gulko, C. G. Danko, and A. Siepel, 2011 Bayesian inference of ancient human demography from individual genome sequences. *Nat. Genet.* 43: 1031–1034. <https://doi.org/10.1038/ng.937>
- Haber, M., M. Mezzavilla, A. Bergstrom, J. Prado-Martinez, P. Hallast *et al.*, 2016 Chad genetic diversity reveals an African history marked by multiple Holocene Eurasian migrations. *Am. J. Hum. Genet.* 99: 1316–1324. <https://doi.org/10.1016/j.ajhg.2016.10.012>
- Hallast, P., C. Batini, D. Zadik, P. Maisano Delser, J. H. Wetton *et al.*, 2015 The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades. *Mol. Biol. Evol.* 32: 661–673. <https://doi.org/10.1093/molbev/msu327>
- Hammer, M. F., T. Karafet, A. Rasanayagam, E. T. Wood, T. K. Altheide *et al.*, 1998 Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol. Biol. Evol.* 15: 427–441. <https://doi.org/10.1093/oxfordjournals.molbev.a025939>
- Helgason, A., A. W. Einarsson, V. B. Guethmundsdottir, A. Sigurethsson, E. D. Gunnarsdottir *et al.*, 2015 The Y-chromosome point mutation rate in humans. *Nat. Genet.* 47: 453–457. <https://doi.org/10.1038/ng.3171>
- Henn, B. M., T. E. Steele, and T. D. Weaver, 2018 Clarifying distinct models of modern human origins in Africa. *Curr. Opin. Genet. Dev.* 53: 148–156. <https://doi.org/10.1016/j.gde.2018.10.003>
- Higham, T., T. Compton, C. Stringer, R. Jacobi, B. Shapiro *et al.*, 2011 The earliest evidence for anatomically modern humans in northwestern Europe. *Nature* 479: 521–524. <https://doi.org/10.1038/nature10484>
- Ingman, M., H. Kaessmann, S. Paabo, and U. Gyllensten, 2000 Mitochondrial genome variation and the origin of modern humans. *Nature* 408: 708–713. <https://doi.org/10.1038/35047064>
- Karmin, M., L. Saag, M. Vicente, M. A. Wilson Sayres, M. Jarve *et al.*, 2015 A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* 25: 459–466. <https://doi.org/10.1101/gr.186684.114>
- Lazaridis, I., D. Nadel, G. Rollefson, D. C. Merrett, N. Rohland *et al.*, 2016 Genomic insights into the origin of farming in the ancient Near East. *Nature* 536: 419–424. <https://doi.org/10.1038/nature19310>
- Letunic, I., and P. Bork, 2016 Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44: W242–W245. <https://doi.org/10.1093/nar/gkw290>
- Liu, W., M. Martinon-Torres, Y. J. Cai, S. Xing, H. W. Tong *et al.*, 2015 The earliest unequivocally modern humans in southern China. *Nature* 526: 696–699. <https://doi.org/10.1038/nature15696>
- Mallick, S., H. Li, M. Lipson, I. Mathieson, M. Gymrek *et al.*, 2016 The simons genome diversity project: 300 genomes from 142 diverse populations. *Nature* 538: 201–206. <https://doi.org/10.1038/nature18964>
- Mendez, F. L., T. Krahn, B. Schrack, A. M. Krahn, K. R. Veeramah *et al.*, 2013 An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree. *Am. J. Hum. Genet.* 92: 454–459 (erratum: *Am. J. Hum. Genet.* 92: 637). <https://doi.org/10.1016/j.ajhg.2013.02.002>
- Mondal, M., A. Bergström, Y. Xue, F. Calafell, H. Laayouni *et al.*, 2017 Y-chromosomal sequences of diverse Indian populations and the ancestry of the Andamanese. *Hum. Genet.* 136: 499–510. <https://doi.org/10.1007/s00439-017-1800-0>
- Pagani, L., D. J. Lawson, E. Jagoda, A. Morseburg, A. Eriksson *et al.*, 2016 Genomic analyses inform on migration events during the peopling of Eurasia. *Nature* 538: 238–242. <https://doi.org/10.1038/nature19792>
- Patin, E., M. Lopez, R. Grollemund, P. Verdu, C. Harmant *et al.*, 2017 Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. *Science* 356: 543–546. <https://doi.org/10.1126/science.aal1988>
- Patterson, N., A. L. Price, and D. Reich, 2006 Population structure and eigenanalysis. *PLoS Genet.* 2: e190. <https://doi.org/10.1371/journal.pgen.0020190>

- Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland *et al.*, 2012 Ancient admixture in human history. *Genetics* 192: 1065–1093. <https://doi.org/10.1534/genetics.112.145037>
- Poznik, G. D., 2016 Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. *bioRxiv*: 088716.
- Poznik, G. D., B. M. Henn, M. C. Yee, E. Sliwerska, G. M. Euskirchen *et al.*, 2013 Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males vs. females. *Science* 341: 562–565. <https://doi.org/10.1126/science.1237619>
- Poznik, G. D., Y. Xue, F. L. Mendez, T. F. Willems, A. Massaia *et al.*, 2016 Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat. Genet.* 48: 593–599. <https://doi.org/10.1038/ng.3559>
- R Core Team, 2017 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Rosa, A., C. Ornelas, M. A. Jobling, A. Brehm, and R. Villems, 2007 Y-chromosomal diversity in the population of Guinea-Bissau: a multiethnic perspective. *BMC Evol. Biol.* 7: 124. <https://doi.org/10.1186/1471-2148-7-124>
- Scally, A., and R. Durbin, 2012 Revising the human mutation rate: implications for understanding human evolution. *Nat. Rev. Genet.* 13: 745–753 (erratum: *Nat. Rev. Genet.* 13: 824). <https://doi.org/10.1038/nrg3295>
- Scerri, E. M. L., M. G. Thomas, A. Manica, P. Gunz, J. T. Stock *et al.*, 2018 Did our species evolve in subdivided populations across Africa, and why does it matter? *Trends Ecol. Evol.* 33: 582–594. <https://doi.org/10.1016/j.tree.2018.05.005>
- Schlebusch, C. M., P. Skoglund, P. Sjödin, L. M. Gattepaille, D. Hernandez *et al.*, 2012 Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. *Science* 338: 374–379. <https://doi.org/10.1126/science.1227721>
- Schlebusch, C. M., H. Malmstrom, T. Gunther, P. Sjödin, A. Coutinho *et al.*, 2017 Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science* 358: 652–655. <https://doi.org/10.1126/science.aao6266>
- Scozzari, R., A. Massaia, B. Trombetta, G. Bellusci, N. M. Myres *et al.*, 2014 An unbiased resource of novel SNP markers provides a new chronology for the human Y chromosome and reveals a deep phylogenetic structure in Africa. *Genome Res.* 24: 535–544. <https://doi.org/10.1101/gr.160788.113>
- Shi, H., H. Zhong, Y. Peng, Y. L. Dong, X. B. Qi *et al.*, 2008 Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol.* 6: 45. <https://doi.org/10.1186/1741-7007-6-45>
- Skoglund, P., J. C. Thompson, M. E. Prendergast, A. Mittnik, K. Sirak *et al.*, 2017 Reconstructing prehistoric African population structure. *Cell* 171: 59–71. <https://doi.org/10.1016/j.cell.2017.08.049>
- Stamatakis, A., 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Underhill, P. A., and C. C. Roseman, 2001 The case for an African rather than an Asian origin of the human Y-chromosome YAP insertion, pp. 43–56 in *Genetic, Linguistic and Archaeological Perspectives on Human Diversity in Southeast Asia: Recent Advances in Human Biology*, edited by L. Jin, M. Seielstad, and C. Xiao. World Scientific Publishing, Singapore. https://doi.org/10.1142/9789812810847_0004
- Underhill, P. A., G. Passarino, A. A. Lin, P. Shen, M. Mirazon Lahr *et al.*, 2001 The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* 65: 43–62. <https://doi.org/10.1046/j.1469-1809.2001.6510043.x>
- van Oven, M., and M. Kayser, 2009 Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* 30: E386–E394. <https://doi.org/10.1002/humu.20921>
- Veeramah, K. R., D. Wegmann, A. Woerner, F. L. Mendez, J. C. Watkins *et al.*, 2012 An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. *Mol. Biol. Evol.* 29: 617–630. <https://doi.org/10.1093/molbev/msr212>
- Veth, P., I. Ward, T. Manne, S. Ulm, K. Ditchfield *et al.*, 2017 Early human occupation of a maritime desert, Barrow Island, North-West Australia. *Quat. Sci. Rev.* 168: 19–29. <https://doi.org/10.1016/j.quascirev.2017.05.002>
- Weale, M. E., T. Shah, A. L. Jones, J. Greenhalgh, J. F. Wilson *et al.*, 2003 Rare deep-rooting Y chromosome lineages in humans: lessons for phylogeography. *Genetics* 165: 229–234.
- Wei, W., Q. Ayub, Y. Chen, S. McCarthy, Y. Hou *et al.*, 2013 A calibrated human Y-chromosomal phylogeny based on resequencing. *Genome Res.* 23: 388–395. <https://doi.org/10.1101/gr.143198.112>
- Xue, Y., T. Zerjal, W. Bao, S. Zhu, Q. Shu *et al.*, 2006 Male demography in East Asia: a north-south contrast in human population expansion times. *Genetics* 172: 2431–2439. <https://doi.org/10.1534/genetics.105.054270>

Communicating editor: S. Ramachandran