The genetic landscape of Ethiopia: diversity, intermixing and the association with culture - Supplementary Material

SI Section 1. A brief history of Ethiopia with particular reference to foreign contacts

SI Section 1A. Summary

The modern state of Ethiopia, located in the Horn of Africa, encompasses approximately 1.1M km of varied climates and environments, including mountain ranges, high plateaus, forests, lakes and low, arid, extremely hot regions (Marcus, 2002). The Rift Valley, the largest geographic trench on earth which has yielded important hominid fossils, runs northeast-southwest across the country. The population of Ethiopia now exceeds 100 million with groups speaking over 80 languages classified as Semitic, Cushitic, Omotic and Nilo-Saharan. Christian, Muslim, Traditional and Jewish religions are all represented.

The country's archaeological record extends more than 3 million years with genetic and other evidence suggesting that East Africa is the likely region (Quintana-Murci et al., 1999) via which the majority of anatomically modern humans left the African continent to people the rest of the world. *Ardipithecus ramidus*, from the Afar region of northeast Ethiopia, remains the world's oldest hominid fossil find at 4.4M years ago.

It is not thought that the expansion of the Bantu speaking peoples, approximately 5,000ybp, extended into the area covered by the modern Ethiopian state (Vansina, 1995; Ashley, 2010; Clist, 1987). Contacts between the north and, respectively, Egypt and Arabia, although varying in intensity, have continued for at least the past 6,000 years (Currey, 2014; Phillipson, 2012).

While Aksum, in the northern region, was in contact with the world of antiquity (Phillipson, 2012), the south was only incorporated into what was then commonly known in the western world as Abyssinia in the late nineteenth century CE.

Before the 19th Century interaction with European powers was limited, although in the 16th Century

Portuguese emissaries had a six-year presence (Alvarez, 1961), during which plants from the Americas were introduced to local farmers. For five years from 1936 to 1941 Italy occupied the country. The independence of the modern state was recognised internationally in 1947.

SI Section 1B. The Pre-History of Ethiopia

Ethiopia can legitimately claim the longest archaeological record of any country in the world (Phillipson, 1998). The discovery in November 1974 of the first and most famous *Australopithecus* specimen named *afarensis* or Lucy (with 40% of her osteological remains) was followed by a number of hominid, and later fossil remains, including Selam, the oldest (dated 4.3 MYA) and most complete (60%) to date (Kalb, 2001; Johanson, 2009; Suwa et al., 2009; Lovejoy et al., 2009). There appears to be broad scholarly consensus on the antiquity of Ethiopian agriculture (Brandt, 1984; Phillipson, 1993). By about 7500 YBP, the Afroasiatic family of languages appeared in what is now eastern Sudan, northeast of Khartoum. Wild grass would already have been collected as food in the greater Horn of Africa region by this time. Not long afterwards, domestication of plants in the greater Horn may have begun. Nicolai Ivanovich Vavilov (1955) proposed that Ethiopia was probably one of eight centres of plant domestication in the world. At about the fourth millennium BCE, fauna and flora appear to have been introduced into the country from abroad. The fauna include cattle,[1] sheep[2] and goats and the flora wheat, barley and sorghum, all introduced from a likely West Asian source. In later times, Ethiopians maintained close ties with the Graeco-Roman and Eastern Mediterranean worlds. The earliest surviving evidence for trading ties between the greater Horn and Egypt date back to 1500 BCE, when a well-preserved wall relief from Queen Hateshepsut's Deir el-Bahari temple showed ancient Egyptian seafarers heading home from an expedition to what was known as the Land of Punt. (The Land of Punt is generally accepted by modern scholarship to be a reference to the greater Horn of Africa region (Phillips, 1997).) It is very likely that there was wider involvement in the Red Sea trade involving the Arabian Peninsula and the Eastern Mediterranean regions. Contacts with South Arabia have been better attested archaeologically than those with Nubia and Egypt.[3]

SI Section 1C. Ethiopian Christianity and contacts with Egypt, the Levant and Europe

Sometime during the second quarter of the fourth-century CE Aksum adopted Christianity (Sergew 1972). Unlike corresponding events in the Roman Empire, the first to be converted to the new creed in Ethiopia were the ruling class. Palaeographic and documentary sources show that the first Christian ruler of Aksum was King Ezana.[4] The conversion of the ruling class to the new faith seems to have resulted in important transformations that provided politico-ideological legitimacy to the monarchy and also in the sphere of the wider highland Ethiopian popular culture. The celebrated document known as the *Kibra Nagast*, historic Ethiopia's national epic, represents an epitome of this event of revolutionary proportions (Budge, 1922; Shahid, 1976). It took the new religion over a century to reach the broad masses through the evangelical activities of two groups of foreign missionaries, known as the *Tsadqan* and the "Nine Saints" (Sergew 1972). These missionaries were monks who came to Ethiopia from the Eastern Mediterranean region around the end of the fifth century CE. As sufficiently trained clergy, the missionaries appear to have completed the task of translating the Bible mainly from the Old Greek version (the Septuagint), which was translated from the older Hebrew/Aramaic versions (Polotsky 1964; Ullendorff 1968, 1980, 1987; Knibb 1988) into the local vernacular, Ge'ez. It is claimed that this literary activity led to the introduction of some foreign terms into the local language, for example, the Armenian word *adja* (adcha or adjar) for 'emmer wheat' as has long been used in Ethiopia (Harlan, 1969) and the Syriac term *haymanot* for 'religion' (Sergew, 1972). The missionaries appear to have established churches and monasteries, both built and rock-hewn, and generally helped propagate the new faith among the people.

The Alexandrine See came to have the status of spiritual suzerainty and guidance over the Ethiopian Orthodox Church (Taddesse, 1972). Invariably, all heads, known in Ge'ez as *Abun*, of the latter Church were Egyptian bishops, duly consecrated and appointed by the former down to 1959 when Ethiopian bishops begun to serve as the heads of their own Church. From the outset, a special relationship of mutual help and interdependence, appears to have been established between Church and State. The Church seems to have acted as the most prominent ideological arm of the State. In return, the latter endowed the former with massive material subsidies used to establish new churches and monasteries as well as proselytising among, first, believers of traditional religions and later Muslims. This arrangement was manifest, most markedly, during the fourteenth and fifteenth centuries (Taddesse, 1972). Throughout the medieval period, Ethiopian Orthodox Christian monks and devout lay pilgrims are known to have flocked, in comparatively large numbers, to Jerusalem where they made contact with people from other countries. An Ethiopian monk named Father Gregory provided the famous German scholar Hiob Ludolf with reliable material to write his *A New History of Ethiopia* (published in 1682). This book was the second work to provide generally correct information about the country to western Europe following the account of the Portuguese Embassy of 1520-1526 referred to below. Notwithstanding the contacts described above, Ethiopia was relatively isolated from the early-seventh century to the late nineteenth century. However, limited, more or less regular, contacts were maintained with the outside world, chiefly Egypt, the Holy Land, and, to a lesser extent, The Vatican (Taddesse, 1972; Sergew, 1972). Ideas were exchanged and knowledge filtered through in what may have arguably been a two-way traffic. Since the mid-fourteenth century, Ethiopia also maintained some communication with western Europe through visiting travellers and explorers, diplomats, Christian missionaries and scholars (Crawford, 1958; Ullendorff, 1960). In 1520 a fourteen-man Portuguese Embassy, which included a chaplain-chronicler, Father Francesco Alvares, visited the country establishing a well-documented official link between Ethiopia and a European country (Alvarez, 1961; Castanhoso, 1902). The Embassy remained until 1526 when it returned to Lisbon (Alvarez, 1961). Detailed accounts of the venture and observations were published in 1540. It is thought that the Portuguese were instrumental in introducing New World flora including pepper and perhaps also corn, cotton, and beans. Later Jesuit missionaries, led by the Spaniard Father Pedro Paez, succeeded (after years of persuasion) in converting Emperor Susneyos (r. 1607-1632) and some of his most important courtiers to Catholicism in 1622.

Jesuit involvement in the country led to religious wars. Emperor Susneyos abdicated in favour of his son, Emperor Fasil(adas) (r. 1632-1667), who founded Gondar as Imperial Capital in 1634/5 and pursued a close-door policy banning all European visitors to the country (Merid, 1971; Berry, 1976). Despite this edict, some European travellers did succeed in entering the country.

SI Section 1D. Later travellers to Ethiopia

French physician Charles Poncet (1699-1700) and the Scottish traveller James Bruce (1790, 1813), both spent time in Gondar, as did an Armenian jeweller Yohannes T'ovmacean (1764-1766) who also left an important record (Nerssian & Pankhurst, 1982). Ethio-European relations accelerated in the early nineteenth century, when many more Europeans visited the country (Ullendorff, 1960; Rubenson, 1978; Malécot, 1972) with numbers increasing in the late nineteenth and early twentieth centuries.

The Ethiopian victory at the battle of Adwa on the 2nd of March 1896 resisting Italy's colonial ambitions led to the opening of permanent diplomatic missions representing many European countries, including Italy itself (1896), France and Britain (1897), the USA (1904) and subsequently many other nations. Today the capital of Ethiopia, Addis Ababa (founded in 1889 when King Menelik of Shawa became Emperor Menelik II of Ethiopia), hosts the headquarters of the Organisation of African Unity (renamed the African Union).

Notes

1 The earliest African attestation of cattle has been dated to ca. 9000 BCE.

2 Sheep are known in the Nile Valley from at least 5000 BCE (Muigai & Hanotte, 2013).

3 Phillipson (1998: 24) has the following to say on this subject: "*It is remarkable how few are the artefacts of demonstrably Egyptian origin that have been recovered from archaeological sites in Ethiopia and Eritrea ... . Contacts with the Nile Valley, both in Nubia and further downstream, probably became stronger during Aksumite times, both through trade in raw materials (perhaps accompanied by military subjugation and through links between Christian Ethiopia and her co-religionists. This was often the route by which Ethiopian pilgrims travelled to Jerusalem, the place where most regular contact was established between Ethiopian and people from other countries*".

4 Legend has it that there were two Greek-speaking Christians from Egypt, named Frumentius and Aedesius, who had been taken captive from the Red Sea littoral and brought to the royal court in Aksum where a deceased king was survived by his minor son and the Queen Mother (Sergew 1972: pages 95-100). Frumentius and Aedesius reportedly succeeded in gradually converting the young monarch and his mother to embrace the new religion.

SI Section 2. Sampled Ethiopian groups

Today Ethiopia is one of the most populated countries in the world ("Ethiopia". The World Factbook, CIA, 2017), home to over 90 ethnic groups and 86 living languages. The largest groups are the Oromo (34.4% of the total population – source: Ethiopia People 2017, theodora.com), the Amhara (27%), the Somali (6.2%) and the Tigraway (6.1%) that make up around three-quarters of the population. The rest of the groups represent low percentages of the total population, and some of them represent minorities with less than 10,000 members. Christianity is the most practiced religion in Ethiopia (62.8%), followed by Islam (33.9%), traditional faiths (2.6%) and other religions (0.6%) (Ethiopian Central Statistical Agency, 2009). Ethiopia is administratively divided into nine regions: Afar, Amhara, Benishangul-Gumuz, Gambela, Harari, Oromia, Somali, Southern Nations, Nationalities and Peoples' and Tigray. This is subdivided into 68 zones, that in turn are subdivided into districts or woredas and two chartered cities (Addis Ababa and Dire Dawa).

Languages spoken in Ethiopia can be classified into two language families: Afroasiatic and Nilo-Saharan. Branches of the Afroasiatic languages represented in Ethiopia are Cushitic, Omotic and Semitic (www.ethnologue.com), while Nilo-Saharan languages spoken in Ethiopia have been classified as members of two groups: Chari-Nile and Koman (Bender, 1976). Previous work has suggested that linguistic affiliation is the main factor driving genetic structure in Ethiopians (Pagani et al., 2012). Table S1 and Figure 1a show the languages associated with the samples included for this work.

For the Ethiopians we genotyped those individuals whose grandparents' birthplaces were coincident, with the exception of a few ethnic groups (Meinit, Qimant, Suri, Negede-Woyto, Shinasha, Bana) where we did not find any sampled individual fulfilling this condition. In these cases, the geographical location was calculated as the average point between the birth locations of the paternal grandfather and the maternal grandmother. The latitude and longitude coordinates of birthplaces of donors were recorded as the roughly central place in the locality of their birth, which was obtained by one or other of the following means: on-site use of a GARMIN GPS unit during the data collection, information provided by a local service (Information Systems Services (ISS)) and manual searches using Google Maps, OpenStreetMap and, in a few cases, other programs. Information about elevation was obtained using the geographic coordinates of each individual in the dataset with the "Googleway" package.

Samples were collected by co-author AT in a programme that consisted of two collection expeditions a year taking place from 1998 to 2011 and which frequently involved communication through local translators in remote locations. Consistency is an extremely difficult thing to achieve in Ethiopian linguistic nomenclature. Neither the Ethnologue (www.ethnologue.com) nor many linguists consistently give primacy to the opinions of native speakers in ascribing a name to a language. Sometimes a group may declare that they speak a language that linguists may designate a dialect. Consequently the Ethnologue may use a common name for a cluster of closely related tongues. AT recorded the name of a language as declared by the speakers who were sampled. Similarly AT's practice was to use the self-adopted names of ethnic groups rather than names used to refer to them by their neighbours or other outsiders, all of which they are likely to consider derogatory. All information recorded with respect to donors was either provided by the individual donors themselves or by local informants. Information about linguistic classifications used in this study are provided in Table S1 and Extended Table 1. As examples of some of the complications, the labeled group "DawroManja" refers to a caste-like hunter-gatherer group claiming Kafa as their original language, but who presently live in the Dawro Zone and report Manja as their first language and Dawro as their second language. They claim to be descendants of a gift of slaves by the king of Kafa to the king of Dawro. The labeled group "KefaShekaManjo" refers to a caste-like hunter-gatherer group, locally called Manjo, who live amongst the Kafa and the Shekacho in the Kafa and Sheka administrative Zones (previously the Kafa-Sheka Zone). Manjo study participants living in the Sheka Zones reported their first language as Shekacho and their second language as Amharic, while Manjo living in the Kafa Zone speak Kafa as their first language.

SI Section 3. Inferring admixture in Ethiopian groups

In this section we provide some further details of the GLOBETROTTER analysis to identify and date admixture events involving the Ethiopian groups.

Dating admixture by comparing Ethiopian clusters to non-Ethiopian groups ("Ethiopia-external" analysis)

We applied GLOBETROTTER as described in the main text separately to each Ethiopian cluster, in order to identify and date admixture events using all 252 non-Ethiopian modern groups and Mota as surrogates to the putative admixing sources (i.e. "surrogates") under a model that assumes one or more pulses of admixture where two or more sources intermixed. Briefly, first CHROMOPAINTER is used to match haplotype patterns within individuals from the target (i.e. putatively admixed) group to those in a set of reference individuals (Fig 2a). In doing so, each target individual's genome is composed of a series of DNA segments, with each segment matched to (i.e. inferred to share a most recent common ancestor with) a specific reference population. GLOBETROTTER then infers admixture in the target group by modelling the decay in linkage disequilibrium among segments within each target individual that are inferred to match to different surrogate populations. For every pairing of surrogate populations, GLOBETROTTER infers the probability that two DNA segments on the same chromosome within a target individual share a most recent ancestor with those two surrogate groups, with one DNA segment matched to each surrogate, versus the centimorgan distance between the two segments. Importantly, under the pulses of admixture model assumed by GLOBETROTTER, if two surrogate groups represent the same (unknown) admixing source, the inferred probability for that pair will decrease exponentially with increasing genetic distance. In contrast, if two surrogate groups represent different admixing sources, their inferred probability will increase exponentially with increasing genetic distance (Hellenthal et al., 2014). Therefore, by studying the probability patterns among all pairs of surrogates, GLOBETROTTER can automatically infer the number of admixture events (though attempts only to characterize up to two events in practice; see below), as well as which surrogates best match genetically to each putative source involved in each event.

For each Ethiopian cluster, GLOBETROTTER assigns any inferred admixture to one of four categories: (1) "one-date" involving a single purse of admixture between two sources, (2) "one-date-multiway" involving a single pulse of admixture between greater than two sources, (3) "multiple-dates" involving more than a single pulse of admixture at different times, potentially between greater than two sources, and (4) "uncertain" where the probability curves described above are challenging to categorize into (1)-(3). For (3), GLOBETROTTER attempts to only date and describe two distinct pulses of admixture, though we note these signals can be consistent with more than two pulses of admixture or continuous admixture (Hellenthal et al., 2014). Furthermore, signals concluding (1),(2),(4) may reflect a failure of GLOBETROTTER to identify genuine older admixture events and/or have inferred dates biased towards more recent admixing in the case of continuous or multiple admixture events (Hellenthal et al., 2014). In the case of (4), here we report results corresponding to (1), i.e. assuming we can only reliably characterize a single pulse of admixture between two sources, though these results should be treated with caution. For each Ethiopian cluster for which GLOBETROTTER infers admixture, we use 100 bootstrap re-samples of individuals to infer confidence intervals around inferred dates. In cases where (1), (2) or (4) are concluded by GLOBETROTTER and any of the 100 bootstraps contains an inferred date of one generation, following Hellenthal et al. (2014) we report these clusters as having no evidence of admixture, as this suggests the admixture signal is not strong (i.e. the probability curve is flat). However, in the case of (3), i.e. multiple pulses of admixture, if the earlier date bootstrap contains an estimate of one we still report these signals, since our bootstrapping procedure when assuming two dates tends to give broad CIs in practice even though the admixture signal still seems clear (e.g. see Appendix B for examples of exponential probabity curves described above).

GLOBETROTTER concludes evidence of admixture in 69 of our 78 Ethiopian clusters, with inferred dates ranging from three to 118 generations ago. These inferred dates are robust across two alternative analyses that use different surrogates, one where we simply exclude Mota as a surrogate and another where we include only five modern groups as surrogates: Bantu_SA, Han, Japanese, Mende_Sierra_Leone_MSL, Yemen (Fig S9a).

For our main analysis including all 252 surrogates, the type of admixture events inferred among our 69 clusters include 36 "one-date", 21 "one-date-multiway", six "multiple-dates" and six "uncertain" events. These inferred events typically involve a source represented by (a) Sub Saharan African groups (often including Mota) intermixing with another source represented by (b) W.Eurasian, Egyptian and/or N.African groups (Fig 3b). Importantly, visual inspection of the GLOBETROTTER probability curves shows that Somalia differs among clusters in whether it acts as a surrogate to source (a) versus source (b). In particular the evidence suggests Somalia is aligned with source (a) in 25 Ethiopian clusters primarily found more towards the north/northeast, while aligned with source (b) in 28 clusters found more towards the south/southwest (Fig 3b, Fig S9b, Appendix B). Its alignment is unclear in the 16 other clusters, two of which (ETH_CH, ETH_CW) we believe show no clear evidence of admixture given their bootstrap date estimates contain 1 (Appendix B). Note also that for other clusters the admixture type may be miscategorized by GLOBETROTTER, with e.g. ETH_AS suggesting multiple dates of admixture despite concluding one date -- in particular note how the red line is a better fit to the data than the green line in this case (Appendix B).

Consistent with their geographic locations, SOURCEFIND inferred ancestry sharing with East/West Sub-Saharan African populations is higher among the 28 clusters where Somalia is aligned to source (b) (median=37%, range=25-98%) relative to the 25 clusters where Somalia is aligned to source (a) (median=21%; range=8-27%). The inverse is true for ancestry sharing with Egyptian groups: (a)-type clusters have median=29% (range: 9-44%) sharing with Egyptian groups, while (b)-type clusters have median=3% (range=0-11%) (Fig S9c, Table S10). For 26 of these 53 clusters, GLOBETROTTER's inferred admixture conclusion is one of (2)-(4) above, hence suggesting the admixture is unclear or involves more than two sources intermixing and/or multiple episodes of (perhaps continuous) intermixing, making interpretation challenging. Among the remaining 27 of these 53 clusters where GLOBETROTTER infers a considerably simpler admixture event consisting of two sources intermixing in a single pulse, i.e. (1) above, the point-estimate dates for which Somalia represents source (a) skew older (n=16; median = 63 generations, range = 37-86) relative to those for which Somalia represents source (b) (n=11; median = 26 generations, range = 8-44) (Table S10).

A likely explanation for these type-(a) versus type-(b) patterns is that at least two separate types of intermixing occurred in the ancestors of present-day Ethiopians. The first involved older intermixing between a Sub-Saharan African source versus a West-Eurasian source (type (a) clusters), after which these admixed ancestors intermixed again with individuals from West-Central Africa (type (b) clusters).

Exploring admixture among Ethiopian clusters ("Ethiopia-internal" analysis)

We next set to determine whether there has been intermixing among Ethiopian groups. To do so, we applied GLOBETROTTER to each Ethiopian cluster while allowing all groups, including other Ethiopian clusters, to act as surrogates to putative admixing sources. This analysis inferred admixture in 70 of the 78 Ethiopian clusters, with 44 "one-date", 16 "one-date-multiway", 1 "multiple-date" and 9 "uncertain" events. The inferred dates of events were much more recent relative to the analysis that excluded Ethiopian surrogates (Fig 5a). Overall 52 (86.7%) of 60 groups that concluded "one-date" or “one-date-multiway” events had inferred dates <30 generations ago (~750-850 ya) under this analysis. This demonstrates how this analysis is capturing more recent intermixing by including Ethiopian surrogates, likely because different Ethiopian groups have been intermixing more recently.

To assess whether Ethiopian groups are intermixing with geographically nearby groups, we first arranged our clusters along a line (represented by the circle in Fig 5b) based on the geographic distance between them. To do so, we ordered groups according to their order along the first component of a principal-components-analysis (PCA) of the geographic (Haversine) distance matrix between clusters, where the latitude/longitude of each cluster is the average of that among all individuals within that cluster. For each of the 61 clusters that inferred admixture events classified into (1)-(3), i.e. excluding clusters that concluded “uncertain” admixture, we took the GLOBETROTTER inference of the most strongly signalled event (see Supp Table S5-S6). This event infers that two distinct sources intermixed, one contributing a majority of ancestry and the other contributing a minority of ancestry to the cluster’s individuals. We calculated two “geographic proximity scores” for the cluster by finding the ordinal distances, along the PCA-line, between that cluster and (i) the surrogate group that GLOBETROTTER inferred as the best genetic match to the minority-contributing source and (ii) the surrogate group that GLOBETROTTER inferred as the best genetic match to the majority-contributing source. For each of (i) and (ii), we averaged scores across all 61 clusters, giving final values of 14.26 and 12.97, respectively. To be conservative, we took the highest of these two scores, i.e. that based on the ordinal distance between the cluster and the minority-contributing source. (The higher score of the minority source makes intuitive sense, as typically the majority contributing source is more genetically similar and geographically closer to the target group than the minority contributing source. Indeed, for this reason, the majority source is often presumed to reflect the ancestors who lived in the same region as the present-day target individuals, while the minority source is presumed to have admixed with these ancestors, e.g. after migrating into the region.) We then permuted labels around the line 50,000 times, while fixing the minority-contributing sources, and recalculated our average proximity score for each permutation. The permutation-based p-value calculating the proportion of permutations whose average proximity score was less than or equal to that of the observed average proximity score was highly significant (p-val < 0.0001). Overall these results suggest that geographically nearby groups in Ethiopia have intermixed with each other more recently than the W.Eurasian and W.African-source events inferred in our admixture analysis that excludes Ethiopians as surrogates.

SI section 4. The association of genetic similarity and language classifications

This section provides further insights into the results of studying the association between genetics and linguistic classifications. Our study contained Ethiopian individuals from ethnic groups classified as belonging to the Nilo-Saharan (NS) and Afroasiatic (AA) language families. These are classified into four different within-family branches (www.ethnologue.com): the NS Satellite-Core (193 individuals), AA Cushitic (390 individuals), AA Omotic (565 individuals) and AA Semitic (95 individuals) branches. In addition our study included data from two linguistic isolates (NegedeWoyto, Shabo) not classified into these families (www.ethnologue.com). Genetic differences among individuals from these six different categories are summarized in Fig 2, Fig 3b, Fig S2 and Fig S7. In this section we describe genetic similarities between individuals of different sub-branch classifications, which are summarized in Fig S7 and Extended Tables 7-8.

We had individuals representing two distinct sub-branches within each of the AA Cushitic, AA Omotic and NS Core-Satellite branches, as well as additional classifications within most of these sub-branches (see Extended Table 1). Sub-branches within each of the above are genetically differentiable (p-val < 0.01) under both the "Ethiopia-internal" and "Ethiopia-external" analyses (Fig S7). Therefore on average people from the same language sub-branch classification (i.e. to the third tier of classification provided in www.ethnologue.com) share more recent ancestry with each other than they do with people from other language classifications, with these effects not solely resulting from recent isolation.

We also find that individuals from different linguistic classifications within each sub-branch can be significantly more genetically similar, though some genetic patterns are not consistent with linguistic classifications (Fig S7). For example, within the AA Cushitic East sub-branch, it has been suggested that individuals from Highland and Lowland linguistic categories diverged before 3,000BCE, and that Werizoid (or Dullay) languages diverged from other Lowland languages between this time and 1,000BCE (Ehret, 1976). Conflicting with this, on average Highland speakers are more genetically similar to individuals from particular Lowland groups than individuals from different Lowland groups are to each other, and Lowland speakers from the Dullay and Konso-Gidolo groups are not differentiable from each other while each being significantly different from other Lowland groups (Fig S7). However, if linguistic trees correlated with the order in which groups became isolated from one another, these genetic discrepancies could be driven by subsequent factors, such as more recent admixture events shared among Dully and Konso-Gidolo speakers that did not affect other Lowland speakers, as suggested by Black (1975).

Within the AA Omotic North sub-branch, out of six linguistic classifications only individuals from ethnicities speaking Janjer and Gonga languages are genetically distinct from all other classifications under both analyses, while individuals from ethnicities speaking the other four languages (Chara, Dizoid, Gimira, Ometo) are not always distinguishable from one another (p-val > 0.01) particularly under the "Ethiopia-external" analysis (Fig S7). All three language classifications within the AA Omotic South sub-branch are genetically distinguishable under the "Ethiopia-internal" analysis (p-val < 0.001) but not all are under the "Ethiopia-external" analysis (Fig S7), suggesting some differences may be attributable to recent isolation.

Similarly, all three language classifications (Surmic, Nilotic, Koman) within the NS Core sub-branch are genetically distinguishable under the "Ethiopia-internal" analysis (Fig S7). The Gumuz have a disputed language classification, B'aga in Ethnologue, but in Bender (1976) it is suggested that the Gumuz language may be classified as Koman. Genetically, the Gumuz are significantly most similar to Komo speakers under the "Ethiopia-external" analysis (Fig S2b, Extended Table 4), suggesting they share more recent ancestry with Koman speakers than with other NS Core sub-classifications included in this study.

SI section 5. Descriptions of cultural traits

The practices listed below are reported in (The Council of Nationalities, Southern Nations and Peoples Region, 2017). No attempt at independent verification has been attempted. Explanation of their nature is based on AT's interactions during collection seasons and knowledge of relevant publications and unpublished dissertations and theses .

Arranged marriage or marriage arranged by parents: marriage arranged by the parents of the couple with little or no involvement of the couple.

Abduction: involves a man, often assisted by members of his family and/or friends, forcibly taking a young girl or a mature woman as a wife. No parental consent is obtained.

Swift/spontaneous unions: said to occur only occasionally in southwestern parts of Wollo province to the northeast of the south-westerly bend of the Blue Nile but also elsewhere including in the SNNPR (e.g. among the Halaba). Involves a girl, well-past normal marriageable age and spur of the moment agreement.

Sororate/cousin marriages: a widower marries a sister or cousin of his deceased wife.

Wife-replacement: a widower marries a sister or close female relative of his deceased wife.

Wife-inheritance: a married man "inherits" i.e. takes as an additional wife a widow of his deceased elder brother, cousin or close kinsman with the primary objective of providing trusteeship for children and assets left behind by his deceased relative.

Belt-giving: a form of marriage that involves offering the intended bride ladies' belts as a symbolic gesture of the young man's desire to marry her. If the girl does not wish to marry the man she refuses to accept the belt. The belt may be considered a token of love and may form a small portion of the bride-price. Parents may not be able to object to the marriage.

Bead-giving: a young man offers his future bride beads as a token of his love and affection. A variant of this practice involves a young man forcibly tying beadwork round a girl's neck despite her resistance (parents may not be able to object to marriage following such an event). The beads may be considered a small pre-marriage instalment of bride-price.

Beaded necklace snatching: a form of marriage that involves an earlier snatching of a young girl's beadwork necklace. (The act of snatching the beadwork is a symbolic gesture of the man's wish to marry the woman.) If adult male members of the girl's family apprehend the snatcher, he may be beaten and dispossessed of the bead necklace, in which case he cannot lay customary claim to the girl as his future wife. If the man escapes with the girl's bead necklace the girl's family will be forced to give the girl up to the man for marriage.

Note: many groups in the SNNPR accord ideological/symbolic significance to belts and beads that revolves around 'omens and a person's fate and fortunes'.

Men moving in to marry women: the married couple move into the bride's home after marriage. This is uncommon. A women most commonly moves into a home built and owned by her husband.

Women moving in to marry men: suddenly (in circumstances in which a man and woman are in love), completely unannounced, a young woman enters the family home of a young man and, clinging to the central pole of the house, pleads with the boy's parents to allow their son to marry her.

Repeat marriages: marriages following divorce or annulment of a previous union.

Bride's butter anointment: an important part of a marriage ceremony that takes place at the groom's family home during a wedding. It involves the future mother-in-law anointing the hair of the bride with a generous amount of butter.

Male circumcision: removal of the foreskin. (May be performed on babies, young boys, teenagers and adult males, individually or in groups of similar age. May be part of initiation ceremonies.)

Female circumcision: cutting of the labia minora and/or the labia majora, with or without excision of the clitoris of young girls and women. In northern Ethiopia it is performed at an early age, while in the south-western parts of the country it takes place at a later age and is related to marriage.

Pre-marital sex: sexual intercourse with the opposite sex prior to marriage (considered unacceptable in most communities in Ethiopia but accepted as the norm by a few groups in the SNNPR).

Pre-marital pregnancy or birth: a woman becoming pregnant or having a baby prior to a marital union.

Endo/exogamy: marriage of a man or a woman to the opposite sex, respectively, within/outside their clan or lineage.

Poly/monogamy: marriage of a man to many wives or just one wife, respectively.

Marriage with bride's parental consent: marriage of a woman to a man with her parents' consent.

Marriage with brides encouraged by parents: marriage of a woman to a man whom the woman's parents prefer to other men.

Groom's parental choice marriage: marriage of a man to a woman whom the man's parents prefer to other women.

Groom's aunt arranged marriage: marriage of a man to a woman whom the man's aunt prefers to other women.

Marriage involving third-party agents: marriage between a man and a woman arranged by third-party agents (might be cousins, aunts, uncles, friends, acquaintances, any other person, related or unrelated, to the couple).

Marriage involving women intermediaries: marriage between a man and a woman arranged solely by female intermediaries (they may or may not be related to the couple).

Special unions: marriage unions between a man and a woman that do not conform to commonly accepted cultural practice followed by most members of a group (important examples of such marriages include community or religious/spiritual leaders, chiefs and kings).

Minors' marriage: marital union between underage children of the opposite sex arranged by their parents.

Mate-selection: marriage in which the couple themselves decide to marry (with little involvement of either set of parents in marriage negotiations).

Marriage with mutual spouse consent: as mate-selection.

Marriage with spouse and parental consent: marriage of a man and a woman with the consent of the couple and their parents.

Marriage by elopement/persuasive absconding: marriage of a man and woman usually without parental consent after the man persuades the woman to abscond with him.

Supplementary Tables

Table S1 Features of the Ethiopian samples included in this work, including ethnicity, language group and self-reported first and second languages (n = sample size). "Language group" is given at the first ("Nilo-Saharan") and to the second (other languages) classification level at www.ethnologue.com; all Nilo-Saharan speakers are from the "Satellite-Core" second classification. Labels in parentheses of column 1 are used in Fig 7; labels in parenthesis of column 3 are used in Fig S6ab. Groups from the SSNPR with information about culture practices are indicated with an asterisk.

Ethnicity	Language group	Self-reported 1st language	Self-reported 2nd language	n
Afar (AFA)	Afroasiatic_Cushitic	Afar (AF)	Amharic	9
		Afar	Tigrinya	1
Agaw (AGA)	Afroasiatic_Cushitic	Agaw (AG)	Amharic	14
		Amharic (AM)	Agaw	6
		Amharic	None	10
Halaba (HAL)*	Afroasiatic_Cushitic	Halaba (AL)	Amharic	13
		Amharic	Halaba	1
Alae (ALA)*	Afroasiatic_Cushitic	Alae (GB)	Amharic	16
		Alae	Bana	1
		Alae	Affa Honso	7
		Alae	None	9
		Alae	Oromiffa	2
		Masholae (MH)	Oromiffa	5
		Oromiffa (OR)	Alae	3
		Oromiffa	Amharic	3
		Oromiffa	Masholae	1
Amhara (AMH)	Afroasiatic_Semitic	Amharic	None	25
		Amharic	Oromiffa	2
		Oromiffa	Amharic	1
		unknown	unknown	24
Anuak (ANU)	Nilo-Saharan	Anuak (AN)	Amharic	9
Arbore (ARB)*	Afroasiatic_Cushitic	Arbore (AE)	Amharic	4
		Arbore	None	3
		Arbore	Oromiffa	7
Ari_Cultivator (ARI_C)*	Afroasiatic_Omotic	Ari (AA)	Amharic	9

		Ari	None	6
Ari_Potter (ARI_P)*	Afroasiatic_Omotic	Ari	Bana	1
		Ari	None	21
		Ari	Amharic	6
Ari_Smith (ARI_S)*	Afroasiatic_Omotic	Ari	None	10
		Ari	Amharic	7
Basket (BAS)*	Afroasiatic_Omotic	Amharic	Basket	1
		Basket (BK)	Amharic	13
Bana (BAN)*	Afroasiatic_Omotic	Amharic	Bana	1
		Bana (BE)	Ari	4
		Bana	Amharic	15
		Bana	Hamar	1
		Bana	None	9
Bench (BEN)*	Afroasiatic_Omotic	Bench (BN)	Amharic	9
		Bench	Kafacho	2
		Bench	None	1
Berta (BER)	Nilo-Saharan	Berta (BT)	Amharic	10
		Berta	Oromiffa	3
Beta Israel (BET)	Afroasiatic_Semitic	Amharic	unknown	14
Bodi (BOD)*	Nilo-Saharan	Bodi (BD)	Amharic	9
		Bodi	None	8
Burji (BUR)*	Afroasiatic_Cushitic	Burji (BR)	Amharic	2
		Burji	None	3
		Burji	Oromiffa	12
		Oromiffa	Amharic	9
		Oromiffa	Burji	1
Chara (CHA)*	Afroasiatic_Omotic	Chara (CR)	Kafacho	19
Dasanech (DAS)*	Afroasiatic_Cushitic	Dasanech (DS)	Amharic	9
		Dasanech	None	8
Dawro (DAW)*	Afroasiatic_Omotic	Dawro (DW)	Amharic	13
		Dawro	None	1
DawroManja (DAM)	Afroasiatic_Omotic	Manja (MA)	Dawro	15
Dhime (DHI)*	Afroasiatic_Omotic	Dhime (DM)	Amharic	12
		Dhime	Bodi	10
		Gofa (GF)	Amharic	1

Dirasha (DIR)*	Afroasiatic_Cushitic	Amharic	Dirasha	2
		Dirasha (DR)	Amharic	11
		Dirasha	Oromiffa	1
		Oromiffa	Amharic	3
		Oromiffa	Dirasha	1
Dizi (DIZ)*	Afroasiatic_Omotic	Amharic	Dizi	3
		Dizi (DI)	Amharic	10
		Dizi	None	1
Dorze (DOR)	Afroasiatic_Omotic	Dorze (DZ)	OtherGamo	15
Gedeo (GED)*	Afroasiatic_Cushitic	Gedeo (GE)	Amharic	21
GentaGamo (GEN)*	Afroasiatic_Omotic	Amharic	GentaGamo	2
		Amharic	None	1
		Amharic	OtherGamo	1
		GentaGamo (GN)	Amharic	4
		GentaGamo	Gidicho	2
		GentaGamo	OtherGamo	3
		OtherGamo	Amharic	2
Gidicho (GID)*	Afroasiatic_Omotic	Gidicho (GG)	Amharic	7
		Gidicho	GentaGamo	7
Gofa (GOF)*	Afroasiatic_Omotic	Amharic	Gofa	3
		Amharic	None	3
		Amharic	OtherGamo	3
		Gofa	Amharic	2
		Gofa	OtherGamo	2
		OtherGamo (GM)	Amharic	2
Gumuz (GUM)	Nilo-Saharan	Gumuz (GU)	Amharic	2
		unknown	unknown	24
Gurage (GUR)*	Afroasiatic_Semitic	Amharic	Gurage	2
		Amharic	None	3
		Amharic	Oromiffa	1
		Amharic	OtherGamo	1
		Gurage (GR)	Amharic	7
		Gurage	Sidama	1
		Oromiffa	Amharic	1
Hadiya (HAD)*	Afroasiatic_Cushitic	Amharic	Hadiya	2

		Amharic	None	1
		Hadiya (HD)	Amharic	11
Hamar (HAM)*	Afroasiatic_Omotic	Amharic	Hamar	1
		Hamar (HM)	Amharic	7
		Hamar	Karo	1
		Hamar	None	6
Honsita (HON)	Afroasiatic_Cushitic	Amharic	Affa Honso	3
		Amharic	None	1
		Affa Honso (KS)	Amharic	10
		Affa Honso	Oromiffa	2
		Affa Honso	OtherGamo	1
Kafacho (KAF)	Afroasiatic_Omotic	Kafacho (KF)	Amharic	16
Kambata (KAM)*	Afroasiatic_Cushitic	Amharic	Kambata	1
		Kambata (KM)	Amharic	12
Karo (KAR)*	Afroasiatic_Omotic	Karo (KO)	Hamar	14
KefaShekaManjo (KEF)	Afroasiatic_Omotic	Kafa/Shekacho (SC)	Amharic	12
Komo (KOM)	Nilo-Saharan	Komo (KE)	Mao	1
		Komo	None	4
		Komo	Oromiffa	4
Konta (KON)*	Afroasiatic_Omotic	Konta (KN)	Amharic	16
Kore (KOR)*	Afroasiatic_Omotic	Amharic	Kore	3
		Kore (KR)	Amharic	5
		Kore	Oromiffa	2
		Oromiffa	Amharic	6
Kwegu (KWE)*	Nilo-Saharan	Karo	Kwegu	1
		Kwegu (KW)	Karo	8
		Kwegu	None	4
Maleh (MAL)*	Afroasiatic_Omotic	Maleh (ML)	Amharic	10
		Maleh	Gofa	2
Mao (MAO)	Afroasiatic_Omotic	Mao (ME)	Oromiffa	1
		Oromiffa	Amharic	5
		Oromiffa	Mao	2
		Oromiffa	None	1
Masholae (MAS)*	Afroasiatic_Cushitic	Alae	Oromiffa	1
		Masholae	Amharic	8

		Masholae	Oromiffa	5
		Oromiffa	Amharic	4
		Oromiffa	Masholae	1
Meinit (MEI)*	Nilo-Saharan	Meinit (MT)	Amharic	14
		Meinit	Bench	1
Mezhenger (MEZ)*	Nilo-Saharan	Mezhenger (MG)	Amharic	10
		Mezhenger	None	5
Mossiye (MOS)*	Afroasiatic_Cushitic	Mossiye (BU)	Amharic	2
		Mossiye	None	1
		Mossiye	OtherGamo	3
		OtherGamo	Amharic	3
		OtherGamo	Mossiye	1
		OtherGamo	Oromiffa	1
Murle (MUR)*	Nilo-Saharan	Murle (MB)	Nyangatom	1
		Nyangatom (BM)	Amharic	6
		Nyangatom	None	6
Mursi (MUS)*	Nilo-Saharan	Mursi (MU)	None	12
Nao (NAO)*	Afroasiatic_Omotic	Kafacho	Amharic	1
		Nao (NO)	Kafacho	16
NegedeWoyto (NEG)	NegedeWoyto	Amharic	None	10
Nuer (NUE)	Nilo-Saharan	Nuer (NR)	Amharic	2
		Nuer	English	1
		Nuer	None	9
Nyangatom (NYA)*	Nilo-Saharan	Nyangatom	Amharic	7
		Nyangatom	None	5
Oromo (ORO)	Afroasiatic_Cushitic	unknown	unknown	28
		Oromiffa	Amharic	7
OtherGamo (OTH)	Afroasiatic_Omotic	Amharic	OtherGamo	10
		OtherGamo	Amharic	6
Qimant (QIM)	Afroasiatic_Cushitic	Amharic	None	16
		Amharic	Qimant	1
Shabo (SHA)	Isolate?	Shabo (SH)	Shekacho	15
Shekacho (SHK)*	Afroasiatic_Omotic	Shekacho	Amharic	15
Sheko (SHE)*	Afroasiatic_Omotic	Sheko (SK)	Amharic	16
Shinasha (SHI)	Afroasiatic_Omotic	Oromiffa	Amharic	1

		Shinasha (SA) Shinasha	Amharic Oromiffa	16 1
Sidama (SID)*	Afroasiatic_Cushitic	Amharic	Sidama	2
		Sidama (SD)	Amharic	19
Somali (SOM)	Afroasiatic_Cushitic	Amharic unknown	None unknown	2 24
Suri (SUR)	Nilo-Saharan	Suri (SR)	Amharic	14
Tigraway (TIG)	Afroasiatic_Semitic	Amharic Amharic Amharic Tigrinya (TG)	None Oromiffa Tigrinya Amharic	2 1 2 8
Tsamay (TSA)*	Afroasiatic_Cushitic	Alae Bana Tsamay (TY) Tsamay Tsamay	Amharic Amharic Amharic Bana None	1 1 3 13 1
Wolayta (WOL)*	Afroasiatic_Omotic	unknown OtherGamo Wolayta (WL)	unknown Amharic Amharic	24 1 3
Wolayta_Cultivator (WOL_C)*	Afroasiatic_Omotic	Wolayta	Amharic	6
Wolayta_Potter (WOL_P)*	Afroasiatic_Omotic	Wolayta Wolayta	Amharic None	1 10
Wolayta_Smith (WOL_S)*	Afroasiatic_Omotic	Wolayta Wolayta	Amharic None	1 14
Wolayta_Tanner (WOL_T)*	Afroasiatic_Omotic	Wolayta Wolayta	Amharic None	2 7
Wolayta_Weaver (WOL_W)*	Afroasiatic_Omotic	Wolayta	None	12
Yem (YEM)*	Afroasiatic_Omotic	Yem (YM)	Amharic	13
Zayse (ZAY)*	Afroasiatic_Omotic	Amharic Amharic Zayse (ZS) Zayse	None Zayse Amharic None	1 4 12 1
Zilmamo (ZIL)*	Nilo-Saharan	Zilmamo (ZL)	Suri	13

Table S2. Non-Ethiopian samples included in this work (n = 2,812), for both the “Ethiopia-internal” and “Ethiopia-external” analyses. Reg=Region, with NA=North Africa, WA=West Africa, CA=Central Africa (Pygmies), EA=East Africa, SO=Somalia, SA=South Africa, EG=Egypt, WE=West Eurasia, SS=South Asia, ES=East Asia, CS=Central Asia/Siberia, OC=Oceania, AM=Americas

Reg	Population_label	n	Reference	Reg	Population_label	n	Reference
NA	Algerian	7	Lazaridis et al. 2014	WE	Spanish_Extremadura_IBS	5	Lazaridis et al. 2014
NA	Libyan_Jew	9	Lazaridis et al. 2014	WE	Spanish_Galicia_IBS	5	Lazaridis et al. 2014
NA	Moroccan_Jew	6	Lazaridis et al. 2014	WE	Spanish_Murcia_IBS	4	Lazaridis et al. 2014
NA	Mozabite	21	Lazaridis et al. 2014	WE	Spanish_Pais_Vasco_IBS	5	Lazaridis et al. 2014
NA	Saharawi	6	Lazaridis et al. 2014	WE	Spanish_Valencia_IBS	5	Lazaridis et al. 2014
NA	Tunisian	8	Lazaridis et al. 2014	WE	Syrian	8	Lazaridis et al. 2014
NA	Tunisian_Jew	7	Lazaridis et al. 2014	WE	Turkish	4	Lazaridis et al. 2014
NA	Libyan	5	Lazaridis et al. 2016	WE	Turkish_Adana	10	Lazaridis et al. 2014
NA	Moroccan	10	Lazaridis et al. 2016	WE	Turkish_Aydin	7	Lazaridis et al. 2014
NA	MoroccoBerber	20	this work	WE	Turkish_Balikesir	6	Lazaridis et al. 2014
WA	Esan_Nigeria_ESN	8	Lazaridis et al. 2014	WE	Turkish_Istanbul	10	Lazaridis et al. 2014
WA	Gambian_GWD	6	Lazaridis et al. 2014	WE	Turkish_Jew	8	Lazaridis et al. 2014
WA	Mandenka	17	Lazaridis et al. 2014	WE	Turkish_Kayseri	10	Lazaridis et al. 2014
WA	Mende_Sierra_Leone_MSL	8	Lazaridis et al. 2014	WE	Turkish_Trabzon	9	Lazaridis et al. 2014
WA	Yoruba	70	Lazaridis et al. 2014	WE	Ukrainian_East	6	Lazaridis et al. 2014
WA	Cameroon_Kotoko	8	this work	WE	Ukrainian_West	3	Lazaridis et al. 2014
WA	Senegal	13	this work	WE	Uzbek	10	Lazaridis et al. 2014
CA	BiakaPygmy	20	Lazaridis et al. 2014	WE	Yemen	6	Lazaridis et al. 2014
CA	MbutiPygmy	10	Lazaridis et al. 2014	WE	Yemenite_Jew	8	Lazaridis et al. 2014

EA	Buganda	100	Gurdasani et al. 2015	WE	Assyrian	11	Lazaridis et al. 2016
EA	BantuKenya	6	Lazaridis et al. 2014	WE	Iranian_Bandari	8	Lazaridis et al. 2016
EA	Datog	3	Lazaridis et al. 2014	WE	Iranian_Lor	10	Lazaridis et al. 2016
EA	Hadza	17	Lazaridis et al. 2014	WE	Iranian_Mazandarani	10	Lazaridis et al. 2016
EA	Hadza_Henn	5	Lazaridis et al. 2014	WE	Iranian_Persian	10	Lazaridis et al. 2016
EA	Kikuyu	4	Lazaridis et al. 2014	WE	Lebanese_Christian	9	Lazaridis et al. 2016
EA	Luhya_Kenya_LWK	8	Lazaridis et al. 2014	WE	Lebanese_Muslim	11	Lazaridis et al. 2016
EA	Luo	8	Lazaridis et al. 2014	WE	Romanian	10	Lazaridis et al. 2016
EA	Masai_Ayodo	2	Lazaridis et al. 2014	SS	Balochi	20	Lazaridis et al. 2014
EA	Masai_Kinyawa_MKK	10	Lazaridis et al. 2014	SS	Bengali_Bangladesh_BEB	7	Lazaridis et al. 2014
EA	Sandawe	22	Lazaridis et al. 2014	SS	Brahui	21	Lazaridis et al. 2014
EA	Tanzania_Chagga	11	this work	SS	Burusho	23	Lazaridis et al. 2014
EA	Dinka	7	Lazaridis et al. 2014	SS	Cochin_Jew	5	Lazaridis et al. 2014
SO	Somali	13	Lazaridis et al. 2014	SS	GujaratiA_GIH	5	Lazaridis et al. 2014
SA	Zulu	100	Gurdasani et al. 2015	SS	GujaratiB_GIH	5	Lazaridis et al. 2014
SA	Bantu_SA	8	Lazaridis et al. 2014	SS	GujaratiC_GIH	5	Lazaridis et al. 2014
SA	Damara	12	Lazaridis et al. 2014	SS	GujaratiD_GIH	5	Lazaridis et al. 2014
SA	Gana	8	Lazaridis et al. 2014	SS	Hazara	14	Lazaridis et al. 2014
SA	Gui	7	Lazaridis et al. 2014	SS	Kalash	18	Lazaridis et al. 2014
SA	Haiom	7	Lazaridis et al. 2014	SS	Kharia	12	Lazaridis et al. 2014
SA	Himba	4	Lazaridis et al. 2014	SS	Kusunda	10	Lazaridis et al. 2014
SA	Hoan	7	Lazaridis et al. 2014	SS	Lodhi	13	Lazaridis et al. 2014
SA	Ju_hoan_North	22	Lazaridis et al. 2014	SS	Makrani	20	Lazaridis et al. 2014
SA	Ju_hoan_South	6	Lazaridis et al. 2014	SS	Mala	13	Lazaridis et al. 2014

SA	Kgalagadi	5	Lazaridis et al. 2014	SS	Onge	11	Lazaridis et al. 2014
SA	Khomani	11	Lazaridis et al. 2014	SS	Pathan	19	Lazaridis et al. 2014
SA	Khwe	8	Lazaridis et al. 2014	SS	Punjabi_Lahore_PJL	8	Lazaridis et al. 2014
SA	Nama	16	Lazaridis et al. 2014	SS	Sindhi	18	Lazaridis et al. 2014
SA	Naro	8	Lazaridis et al. 2014	SS	Tiwari	15	Lazaridis et al. 2014
SA	Shua	9	Lazaridis et al. 2014	SS	Vishwabrahmin	13	Lazaridis et al. 2014
SA	Taa_East	7	Lazaridis et al. 2014	SS	India_Hindu	12	Lopez et al. 2017
SA	Taa_North	9	Lazaridis et al. 2014	SS	India_Zoroastrian	13	Lopez et al. 2017
SA	Taa_West	16	Lazaridis et al. 2014	ES	Ami_Coriell	10	Lazaridis et al. 2014
SA	Tshwa	5	Lazaridis et al. 2014	ES	Atayal_Coriell	9	Lazaridis et al. 2014
SA	Tswana	5	Lazaridis et al. 2014	ES	Cambodian	8	Lazaridis et al. 2014
SA	Wambo	5	Lazaridis et al. 2014	ES	Dai	10	Lazaridis et al. 2014
SA	Xuun	13	Lazaridis et al. 2014	ES	Daur	9	Lazaridis et al. 2014
SA	Malawi_Chewa	11	Skoglund et al. 2017	ES	Han	33	Lazaridis et al. 2014
SA	Malawi_Ngoni	4	Skoglund et al. 2017	ES	Han_NChina	10	Lazaridis et al. 2014
SA	Malawi_Tumbuka	10	Skoglund et al. 2017	ES	Hezhen	8	Lazaridis et al. 2014
SA	Malawi_Yao	9	Skoglund et al. 2017	ES	Japanese	29	Lazaridis et al. 2014
EG	Egyptian_Comas	11	Lazaridis et al. 2014	ES	Kinh_Vietnam_KHV	8	Lazaridis et al. 2014
EG	Egyptian_Metspalu	7	Lazaridis et al. 2014	ES	Korean	6	Lazaridis et al. 2014
WE	Iran_Fars	17	Broushaki et al. 2016	ES	Lahu	8	Lazaridis et al. 2014
WE	Iran_Zoroastrian	29	Broushaki et al. 2016	ES	Miao	10	Lazaridis et al. 2014
WE	Abkhasian	9	Lazaridis et al. 2014	ES	Mongola	6	Lazaridis et al. 2014
WE	Adygei	17	Lazaridis et al. 2014	ES	Naxi	9	Lazaridis et al. 2014
WE	Albanian	6	Lazaridis et al. 2014	ES	Oroqen	9	Lazaridis et al. 2014

WE	Armenian	10	Lazaridis et al. 2014	ES	She	10	Lazaridis et al. 2014
WE	Ashkenazi_Jew	7	Lazaridis et al. 2014	ES	Thai	10	Lazaridis et al. 2014
WE	Balkar	10	Lazaridis et al. 2014	ES	Tu	10	Lazaridis et al. 2014
WE	Basque_French	20	Lazaridis et al. 2014	ES	Tujia	10	Lazaridis et al. 2014
WE	Basque_Spanish	9	Lazaridis et al. 2014	ES	Uygur	10	Lazaridis et al. 2014
WE	BedouinA	25	Lazaridis et al. 2014	ES	Xibo	7	Lazaridis et al. 2014
WE	BedouinB	19	Lazaridis et al. 2014	ES	Yi	10	Lazaridis et al. 2014
WE	Belarusian	10	Lazaridis et al. 2014	CS	Aleut	7	Lazaridis et al. 2014
WE	Bulgarian	10	Lazaridis et al. 2014	CS	Altaian	7	Lazaridis et al. 2014
WE	Chechen	9	Lazaridis et al. 2014	CS	Chukchi	20	Lazaridis et al. 2014
WE	Chuvash	10	Lazaridis et al. 2014	CS	Chukchi_Sir	2	Lazaridis et al. 2014
WE	Croatian	10	Lazaridis et al. 2014	CS	Dolgan	3	Lazaridis et al. 2014
WE	Cypriot	8	Lazaridis et al. 2014	CS	Eskimo_Chaplin	4	Lazaridis et al. 2014
WE	Czech	10	Lazaridis et al. 2014	CS	Eskimo_Naukan	13	Lazaridis et al. 2014
WE	Druze	39	Lazaridis et al. 2014	CS	Eskimo_Sireniki	5	Lazaridis et al. 2014
WE	English_Cornwall_GBR	5	Lazaridis et al. 2014	CS	Even	10	Lazaridis et al. 2014
WE	English_Kent_GBR	5	Lazaridis et al. 2014	CS	Itelmen	6	Lazaridis et al. 2014
WE	Estonian	10	Lazaridis et al. 2014	CS	Kalmyk	10	Lazaridis et al. 2014
WE	Finnish_FIN	7	Lazaridis et al. 2014	CS	Koryak	9	Lazaridis et al. 2014
WE	French	54	Lazaridis et al. 2014	CS	Kyrgyz	9	Lazaridis et al. 2014
WE	French_South	7	Lazaridis et al. 2014	CS	Mansi	8	Lazaridis et al. 2014
WE	Georgian_Jew	7	Lazaridis et al. 2014	CS	Nganasan	11	Lazaridis et al. 2014
WE	Georgian_Megrels	10	Lazaridis et al. 2014	CS	Russian	22	Lazaridis et al. 2014
WE	Greek_Comas	14	Lazaridis et al. 2014	CS	Selkup	10	Lazaridis et al. 2014

WE	Greek_Coriell	6	Lazaridis et al. 2014	CS	Tajik_Pomiri	8	Lazaridis et al. 2014
WE	Hungarian_Coriell	10	Lazaridis et al. 2014	CS	Tlingit	4	Lazaridis et al. 2014
WE	Hungarian_Metspalu	10	Lazaridis et al. 2014	CS	Tubalar	22	Lazaridis et al. 2014
WE	Icelandic	12	Lazaridis et al. 2014	CS	Turkmen	7	Lazaridis et al. 2014
WE	Iranian	8	Lazaridis et al. 2014	CS	Tuvinian	10	Lazaridis et al. 2014
WE	Iranian_Jew	9	Lazaridis et al. 2014	CS	Ulchi	25	Lazaridis et al. 2014
WE	Iraqi_Jew	6	Lazaridis et al. 2014	CS	Yakut	20	Lazaridis et al. 2014
WE	Italian_Bergamo	12	Lazaridis et al. 2014	CS	Yukagir_Forest	5	Lazaridis et al. 2014
WE	Italian_EastSicilian	5	Lazaridis et al. 2014	CS	Yukagir_Tundra	14	Lazaridis et al. 2014
WE	Italian_South	5	Lazaridis et al. 2014	OC	Australian_ECCAC	3	Lazaridis et al. 2014
WE	Italian_Tuscan	8	Lazaridis et al. 2014	OC	Bougainville	10	Lazaridis et al. 2014
WE	Italian_WestSicilian	6	Lazaridis et al. 2014	OC	Papuan	14	Lazaridis et al. 2014
WE	Jordanian	9	Lazaridis et al. 2014	AM	AA_Denver	12	Lazaridis et al. 2014
WE	Kumyk	8	Lazaridis et al. 2014	AM	Algonquin	9	Lazaridis et al. 2014
WE	Lebanese	8	Lazaridis et al. 2014	AM	Aymara	5	Lazaridis et al. 2014
WE	Lezgin	9	Lazaridis et al. 2014	AM	Bolivian_LaPaz	3	Lazaridis et al. 2014
WE	Lithuanian	10	Lazaridis et al. 2014	AM	Bolivian_Pando	3	Lazaridis et al. 2014
WE	Maltese	8	Lazaridis et al. 2014	AM	Cabecar	6	Lazaridis et al. 2014
WE	Mordovian	10	Lazaridis et al. 2014	AM	Chilote	4	Lazaridis et al. 2014
WE	Nogai	9	Lazaridis et al. 2014	AM	Chipewyan	30	Lazaridis et al. 2014
WE	North_Ossetian	10	Lazaridis et al. 2014	AM	Cree	13	Lazaridis et al. 2014
WE	Norwegian	11	Lazaridis et al. 2014	AM	Guarani	5	Lazaridis et al. 2014
WE	Orcadian	13	Lazaridis et al. 2014	AM	Inga	2	Lazaridis et al. 2014
WE	Palestinian	38	Lazaridis et al. 2014	AM	Kaqchikel	5	Lazaridis et al. 2014

WE	Sardinian	27	Lazaridis et al. 2014	AM	Karitiana	12	Lazaridis et al. 2014
WE	Saudi	8	Lazaridis et al. 2014	AM	Mayan	18	Lazaridis et al. 2014
WE	Scottish_Argyll_Bute_GBR	4	Lazaridis et al. 2014	AM	Mixe	10	Lazaridis et al. 2014
WE	Spanish_Andalucia_IBS	4	Lazaridis et al. 2014	AM	Mixtec	10	Lazaridis et al. 2014
WE	Spanish_Aragon_IBS	6	Lazaridis et al. 2014	AM	Ojibwa	19	Lazaridis et al. 2014
WE	Spanish_Baleares_IBS	4	Lazaridis et al. 2014	AM	Piapoco	4	Lazaridis et al. 2014
WE	Spanish_Canarias_IBS	2	Lazaridis et al. 2014	AM	Pima	14	Lazaridis et al. 2014
WE	Spanish_Cantabria_IBS	5	Lazaridis et al. 2014	AM	Quechua_Coriell	5	Lazaridis et al. 2014
WE	Spanish_Castilla_la_Mancha_IBS	5	Lazaridis et al. 2014	AM	Quechua_RuizLinares	2	Lazaridis et al. 2014
WE	Spanish_Castilla_y_Leon_IBS	5	Lazaridis et al. 2014	AM	Surui	8	Lazaridis et al. 2014
WE	Spanish_Cataluna_IBS	5	Lazaridis et al. 2014	AM	Zapotec	10	Lazaridis et al. 2014

Table S3 Mean and inner 95% empirical quantiles of genetic similarity (1-TVD) under the "Ethiopia-internal" analysis between all pairs of individuals, or restricting to individuals that are from the same self-reported group label, whose group labels belong to the same major language group, or who speak the same first language, second language or have the same religious affiliation. Results are also shown after first conditioning on geographic distance ("Geo") or elevation difference ("Elev"), and rescaling each to span the same range as the first column.

	1-TVD	**1-TVD | Geo**	**1-TVD | Elev**
all	0.492 (0.223-0.727)	0.514 (0.254-0.739)	0.481 (0.21-0.715)
same group	0.685 (0.44-0.772)	0.618 (0.36-0.777)	0.654 (0.406-0.764)
same major language	0.541 (0.279-0.741)	0.553 (0.298-0.747)	0.525 (0.255-0.725)
same 1st language	0.617 (0.314-0.761)	0.592 (0.273-0.774)	0.593 (0.285-0.746)
same 2nd language	0.531 (0.275-0.726)	0.549 (0.302-0.734)	0.517 (0.256-0.713)
same religion	0.507 (0.233-0.722)	0.526 (0.263-0.732)	0.493 (0.212-0.709)

Table S4 Associations of genetic similarity with various factors, while accounting for (a) geographic distance under the "Ethiopia-internal" analysis, (b) elevation difference under the "Ethiopia-internal" analysis, (c) geographic distance under the "Ethiopia-external" analysis and (d) elevation difference under the "Ethiopia-external" analysis. In the second row ("Geo distance/Elevation") and second column ("All"), values to the left of the slash give the proportions of 1000 permutations that were more associated with genetic similarity than the un-permuted data, when testing for an association with (a,c) geographic distance or (b,d) elevation difference (see Methods). Values to the right of the slash give these proportions after first adjusting (a,c) genetic and geographic distance for elevation difference or (b,d) genetic and elevation distance for geographic distance. Column 2 ("All") in the subsequent rows give analogous proportions when testing an additional factor (1st column) for association with genetic similarity after accounting for spatial distance: sharing a common group label ("Group label"), having ethnicities from the same language branch ("Lang group": AA Cushitic, AA Omotic, AA Semitic, NS Satellite-Core), sharing the same first language ("1st lang"), second language ("2nd lang") or religious affiliation ("religion"). Columns 3-7 depict results when permuting in a manner to account for each other factor. Fig 4ab and Fig S5 give the maximum across all left-of-slash values in each row (*=ignored when determining this final p-value; see Methods).

Table S4a: Genetic similarity versus geographic distance, "Ethiopia-internal" analysis

	All	**group label**	**lang group**	**1st lang**	**2nd lang**	**religion**
Geo distance	0 / 0	0.011 / 0*	0 / 0	0 / 0*	0 / 0.000	0 / 0
Group label	0 / 0	NA	0 / 0	0 / 0	0 / 0	0 / 0
Lang group	0 / 0	NA	NA	0 / 0	0 / 0	0 / 0
1st lang	0 / 0	0 / 0	0 / 0	NA	0 / 0	0 / 0
2nd lang	0 / 0	1.000 / 0.998	0.371 / 0.117	1.000 / 0.999	NA	0.004 / 0.004
religion	0.974 / 0.964	0.724 / 0.529	0.845 / 0.706	0.978 / 0.910	0.398 / 0.815	NA

Table S4b: Genetic similarity versus elevation difference, "Ethiopia-internal" analysis

	All	**group label**	**lang group**	**1st lang**	**2nd lang**	**religion**
Elevation	0 / 0	0.138 / 0.659*	0 / 0.003	0.272 / 0.997*	0 / 0	0 / 0
Group label	0 / 0	NA	0 / 0	0 / 0	0 / 0	0 / 0

Lang group	0 / 0	NA	NA	0 / 0.003	0 / 0	0 / 0
1st lang	0 / 0	0.021 / 0.015	0 / 0	NA	0 / 0	0 / 0
2nd lang	0.204 / 0.190	1.0 / 1.0	0.837 / 0.772	1.0 / 1.0	NA	0.932 / 0.946
religion	1.0 / 1.0	1.0 / 1.0	1.0 / 1.0	1.0 / 1.0	0.902 / 0.947	NA

Table S4c: Genetic similarity versus geographic distance, "Ethiopia-external" analysis

	All	**group label***	**lang group**	**1st lang***	**2nd lang**	**religion**
Geo distance	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
Group label	0 / 0	NA	0 / 0	0.957 / 0	0 / 0	0 / 0
Lang group	0 / 0	NA	NA	0.485 / 0	0 / 0	0 / 0
1st lang	0 / 0	0.999 / 0.003	0.001 / 0	NA	0 / 0	0 / 0
2nd lang	0.812 / 0.666	1.0 / 1.0	1.0 / 0.999	1.0 / 1.0	NA	0.964 / 0.838
religion	1.0 / 1.0	0.989 / 0.789	1.0 / 1.0	0.996 / 0.968	1.0 / 1.0	NA

Table S4d: Genetic similarity versus elevation difference, "Ethiopia-external" analysis

	All	**group label***	**lang group**	**1st lang***	**2nd lang**	**religion**
Elevation	0 / 0	0.001 / 0.002	0 / 0	0 / 0.001	0 / 0	0 / 0
Group label	0 / 0	NA	0 / 0	0 / 0	0 / 0	0 / 0
Lang group	0 / 0	NA	NA	0 / 0	0 / 0	0 / 0
1st lang	0 / 0	0.829 / 0.006	0 / 0	NA	0 / 0	0 / 0
2nd lang	0.988 / 0.906	1.0 / 1.0	1.0 / 1.0	1.0 / 1.0	NA	0.99 / 0.971
religion	0.812 / 0.989	0.935 / 0.817	1.0 / 1.0	0.431 / 0.654	0.473 / 0.893	NA

Table S5 Median and interquartile range (IQR) of spatial distance between individuals. These values, rounded to the nearest kilometer, are shown for all pairwise combinations of individuals ("All"), or as the median/IQR across median distances of all pairwise combinations of individuals within each group label ("group label"), major language group ("lang group"), first language ("1st lang"), second language ("2nd lang") or religious affiliation ("religion").

	All	**group label**	**lang group**	**1st lang**	**2nd lang**	**religion**
Geo distance	245 (141-505)	33 (0-46)	233 (221-264)	39 (0-46)	55 (0-74)	310 (227-363)
Elevation	640 (289-1128)	194 (0-287)	459 (339-530)	213 (0-359)	279 (0-539)	856 (546-1024)

Table S6 Genetic similarity among individuals across religious affiliations. Average genetic similarity under the "Ethiopia-internal" analysis between two individuals with religious affiliation = Christian ("C"), Muslim ("M") or Traditional ("T"), within each of 16 Ethiopian groups with n >= 5 sampled individuals from each of at least two of these religious affiliations. Also given is the average similarity between two individuals from separate religions ("C vs M", "C vs T") within each group. Asterisks denote that, within the ethnicity, the average genetically similarity is significantly higher between two individuals from that same religion versus individuals from different religions at p-value <0.05(*) or <0.001(**), based on 1000 permutations of the two compared religious affiliations within each group. Under the "Ethiopia-external" analysis, only Christians within Murle and Muslims within each of Alae and Sidama have p-value < 0.05.

group	C (n)	M (n)	T (n)	C vs M	C vs T
Alae	0.819 (33)	0.902** (10)	0.909 (4)	0.827	--
Amhara	0.916 (11)	0.928 (17)	-- (0)	0.928	--
Ari Potter	0.79 (15)	-- (0)	0.871 (13)	--	0.821
Bana	0.751 (15)	-- (0)	0.794 (15)	--	0.746
Bodi	0.817 (9)	-- (0)	0.692 (8)	--	0.706
Chara	0.797 (10)	-- (0)	0.789 (9)	--	0.78
Dasanech	0.757 (7)	-- (0)	0.861* (10)	--	0.806
Gedeo	0.907 (10)	0.935* (11)	-- (0)	0.931	--
Gurage	0.896 (11)	0.917 (5)	-- (0)	0.901	--
Hamar	0.796 (6)	-- (0)	0.833 (9)	--	0.818
Honsita	0.85 (12)	0.927* (5)	-- (0)	0.899	--
DawroManja	0.915 (5)	-- (0)	0.898 (10)	--	0.914
Murle	0.897* (7)	-- (0)	0.848 (6)	--	0.855
Sidama	0.88 (11)	0.931* (10)	-- (0)	0.901	--
Tigraway	0.926* (8)	0.912 (5)	-- (0)	0.911	--
Tsamay	0.783 (9)	-- (1)	0.747 (9)	0.695	0.757

Table S7 Association of ancestry sharing with Mota to spatial distance. Effect sizes, standard errors and p-values for a linear regression of SOURCEFIND-inferred ancestry matching to the 4.5kya Ethiopian Mota versus geographic/elevation distance of modern individuals from Mota (Fig 4cd).

analysis	Effect size (% per km)	SE (% per km)	p-value
geographic distance	-3.13e-02	7.44e-03	0.00007
elevation difference	-11.3	2.72	0.00009

Table S8 Genetic association with cultural similarity. P-values from Mantel tests for association between genetic and cultural similarity among ethnicities ("All"), and from partial Mantel tests that accounts for one of geographic distance, elevation, or language branch (AA Cushitic, AA Omotic, AA Semitic, NS Satellite-Core) when testing for an association between genetic and cultural similarity. Within each analysis ("Ethiopia-internal", "Ethiopia-external"), the first row measures cultural similarity as the number of matching reported cultural practices across ethnicities, while the second row up-weights sharing of rare cultural practices among ethnicities (see Methods).

Analysis		**All**	**Geographic distance**	**Elevation**	**Language**
Eth-internal	equal practice weight	0.007	0.006	0.013	0.008
	up-weight rare practices	0.009	0.014	0.024	0.023
Eth-external	equal practice weight	0.206	0.161	0.433	0.269
	up-weight rare practices	0.114	0.182	0.310	0.231

Table S9 For each group *X*={*Mursi, Suri, Zilmamo*} (columns), all of which traditionally wear decorative lip plates, each row gives average genetic similarity between individuals from *X* versus those from other listed group **(a)** under the **"Ethiopia-internal"** analysis, **(b)** relative to that expected under geographic distance (fit as described in Methods) under the **"Ethiopia-internal"** analysis, **(c)** under the **"Ethiopia-external"** analysis, and **(d)** relative to that expected under geographic distance under the **"Ethiopia-external"** analysis. I.e. values reflect averages across all pairwise comparisons of individuals, with one individual from *X* and the other from the group in the row. For each of *X*={*Mursi, Suri, Zilmamo*} and the average across these three, the 10 groups with 10 highest such values are shown. Letters in parentheses give the language branch classification of that ethnicity (C=Afroasiatic Cushitic, O = Afroasiatic Omotic, N = Nilo-Saharan Satellite-Core, U = Unclassified).

Table S9a: Genetic similarity with other groups, "Ethiopia-internal" analysis

Mursi		**Suri**		**Zilmamo**		**Average**	
Mursi (N)	**0.882**	**Suri (N)**	**0.917**	**Suri (N)**	**0.89**	**Suri (N)**	**0.863**
Suri (N)	**0.783**	**Zilmamo (N)**	**0.89**	**Zilmamo (N)**	**0.876**	**Zilmamo (N)**	**0.843**
Zilmamo (N)	**0.762**	**Mursi (N)**	**0.783**	**Mursi (N)**	**0.762**	**Mursi (N)**	**0.809**
Karo (O)	0.701	Karo (O)	0.731	Karo (O)	0.72	Karo (O)	0.717
Murle (N)	0.667	Murle (N)	0.672	Murle (N)	0.662	Murle (N)	0.667
Anuak (N)	0.643	Anuak (N)	0.666	Anuak (N)	0.659	Anuak (N)	0.656
Nuer (N)	0.636	Nuer (N)	0.65	Nuer (N)	0.641	Neur (N)	0.642
Nyangatom (N)	0.632	Nyangatom (N)	0.644	Mezhenger (N)	0.635	Nyangatom (N)	0.637
Bodi (N)	0.626	Mezhenger (N)	0.633	Nyangatom (N)	0.635	Bodi (N)	0.62
Dasanech (C)	0.593	Dasanech (C)	0.63	Dasanech (C)	0.623	Mezhenger (N)	0.617

Table S9b: Genetic similarity with other groups, relative to geographic distance, "Ethiopia-internal" analysis

Mursi	**Suri**	**Zilmamo**	**Average**

Suri (N)	**0.2**	**Mursi (N)**	**0.2**	**Mursi (N)**	**0.179**	**Mursi (N)**	**0.168**
Zilmamo (N)	**0.179**	Karo (O)	0.17	Karo (O)	0.159	**Suri (N)**	**0.164**
Mursi (N)	**0.124**	**Suri (N)**	**0.159**	**Suri (N)**	**0.132**	**Zilmamo (N)**	**0.143**
Anuak (N)	0.097	**Zilmamo (N)**	**0.132**	**Zilmamo (N)**	**0.119**	Karo (O)	0.136
Murle (N)	0.092	Murle (N)	0.118	Murle (N)	0.108	Murle (N)	0.106
Nuer (N)	0.091	Anuak (N)	0.113	Anuak (N)	0.107	Anuak (N)	0.106
Karo (O)	0.078	Neur (N)	0.103	Nuer (N)	0.094	Nuer (N)	0.096
Nyangatom (N)	0.057	Nyangatom (N)	0.09	Nyangatom (N)	0.081	Nyangatom (N)	0.076
Mezhenger (N)	0.019	Dasanech (C)	0.073	Dasanech (C)	0.067	Dasanech (C)	0.047
Dasanech (C)	0.001	Mezhenger (N)	0.051	Mezhenger (N)	0.052	Mezhenger (N)	0.041

Table S9c: Genetic similarity with other groups, "Ethiopia-external" analysis

Mursi		**Suri**		**Zilmamo**		**Average**	
Mursi (N)	**0.97**	**Suri (N)**	**0.969**	**Suri (N)**	**0.962**	**Suri (N)**	**0.965**
Bodi (N)	0.966	Kwegu (N)	0.964	**Zilmamo (N)**	**0.961**	**Mursi (N)**	**0.963**
Suri (N)	**0.962**	**Zilmamo (N)**	**0.962**	Kwegu (N)	0.96	Bodi (N)	0.961
Zilmamo (N)	**0.957**	**Mursi (N)**	**0.962**	**Mursi (N)**	**0.957**	Kwegu (N)	0.96
Kwegu (N)	0.956	Bodi (N)	0.962	Bodi (N)	0.956	**Zilmamo (N)**	**0.96**
Karo (O)	0.95	Mezhenger (N)	0.957	Mezhenger (N)	0.954	Mezhenger (N)	0.953
Mezhenger (N)	0.948	Karo (O)	0.956	Karo (O)	0.953	Karo (O)	0.953
Shabo (U)	0.933	Shabo (U)	0.947	Shabo (U)	0.945	Shabo (U)	0.942
Murle (N)	0.933	Gumuz (N)	0.94	Gumuz (N)	0.939	Gumuz (N)	0.935
Nuer (N)	0.929	Komo (N)	0.938	Komo (N)	0.937	Komo (N)	0.933

Table S9d: Genetic similarity with other groups, relative to geographic distance, "Ethiopia-external" analysis

Mursi		**Suri**		**Zilmamo**		**Average**	
Gumuz (N)	0.083	Gumuz (N)	0.086	Gumuz (N)	0.085	Gumuz (N)	0.084
Komo (N)	0.062	Komo (N)	0.064	Komo (N)	0.063	Komo (N)	0.063
Nuer (N)	0.057	Kwegu (N)	0.063	Kwegu (N)	0.058	Kwegu (N)	0.055
Suri (N)	**0.055**	Karo (O)	0.056	Karo (O)	0.053	**Mursi (N)**	**0.052**
Bodi (N)	0.053	**Mursi (N)**	**0.055**	**Mursi (N)**	**0.05**	Bodi (N)	0.051
Mursi (N)	**0.051**	Bodi (N)	0.052	Mezhenger (N)	0.048	**Suri (N)**	**0.05**
Zilmamo (N)	**0.05**	Mezhenger (N)	0.051	Bodi (N)	0.046	Mezhenger (N)	0.049
Mezhenger (N)	0.048	**Suri (N)**	**0.05**	**Suri (N)**	**0.044**	Karo (O)	0.049
Kwegu (N)	0.044	**Zilmamo (N)**	**0.044**	**Zilmamo (N)**	**0.042**	**Zilmamo (N)**	**0.045**
Anuak (N)	0.042	Shabo (U)	0.043	Shabo (U)	0.041	Shabo (U)	0.041

Table S10 Summary of SOURCEFIND inferred ancestry matching to groups from each major geographic region, across **(a)** all 25 "type (a)" and **(b)** all 28 "type (b)" Ethiopian clusters for which Somalia represents the same admixing source as groups from (a) Sub-Saharan Africa or (b) N_Africa/Egypt/W_Eurasia, respectively, under GLOBETROTTER inference. Dates for which GLOBETROTTER infers a single date of admixture between two sources are given in the last row.

Table S10a: Somalia and SSAfrican groups represent the same admixing source (referred to as "type (a)" in main text and SI Section 3)

	Mean	**Median**	**Min**	**Max**
W_Africa	1.2%	1.1	0.2	2.8
E_Africa	19.3	19.3	7.1	24.5
Somalia	24.4	22.4	17.6	61.9
N_Africa	0.6	0.5	0.2	2.4
Egypt	28.6	29.1	8.6	43.7
S_Africa	0.3	0.2	0	1.1
W_Eurasia	6.9	7.0	2.7	11.4
Mota	18.4	17.6	0.1	35.2
W_Africa+E_Africa	20.5	20.5	7.6	26.6
Date (one event groups; n=16)	62 generations	63	37	86

Table S10b: Somalia and N_Africa/Egypt/W_Eurasia groups represent the same admixing source (referred to as "type (b)" in main text and SI Section 3)

	Mean	**Median**	**Min**	**Max**
W_Africa	9.9%	4.8	0.5	43.0
E_Africa	38.7	32.8	21.0	70.8
Somalia	14.8	15.9	0.4	39.5
N_Africa	0.9	1.1	0	2.1
Egypt	3.5	2.7	0	11.3
S_Africa	1.0	1.0	0	2.5
W_Eurasia	4.3	5.2	0	11
Mota	25.4	29.8	0.1	44.2
W_Africa+E_Africa	48.6	37.3	24.7	97.9
Date (one event groups; n=11)	25 generations	26	8	44

Supplementary Figures

A)

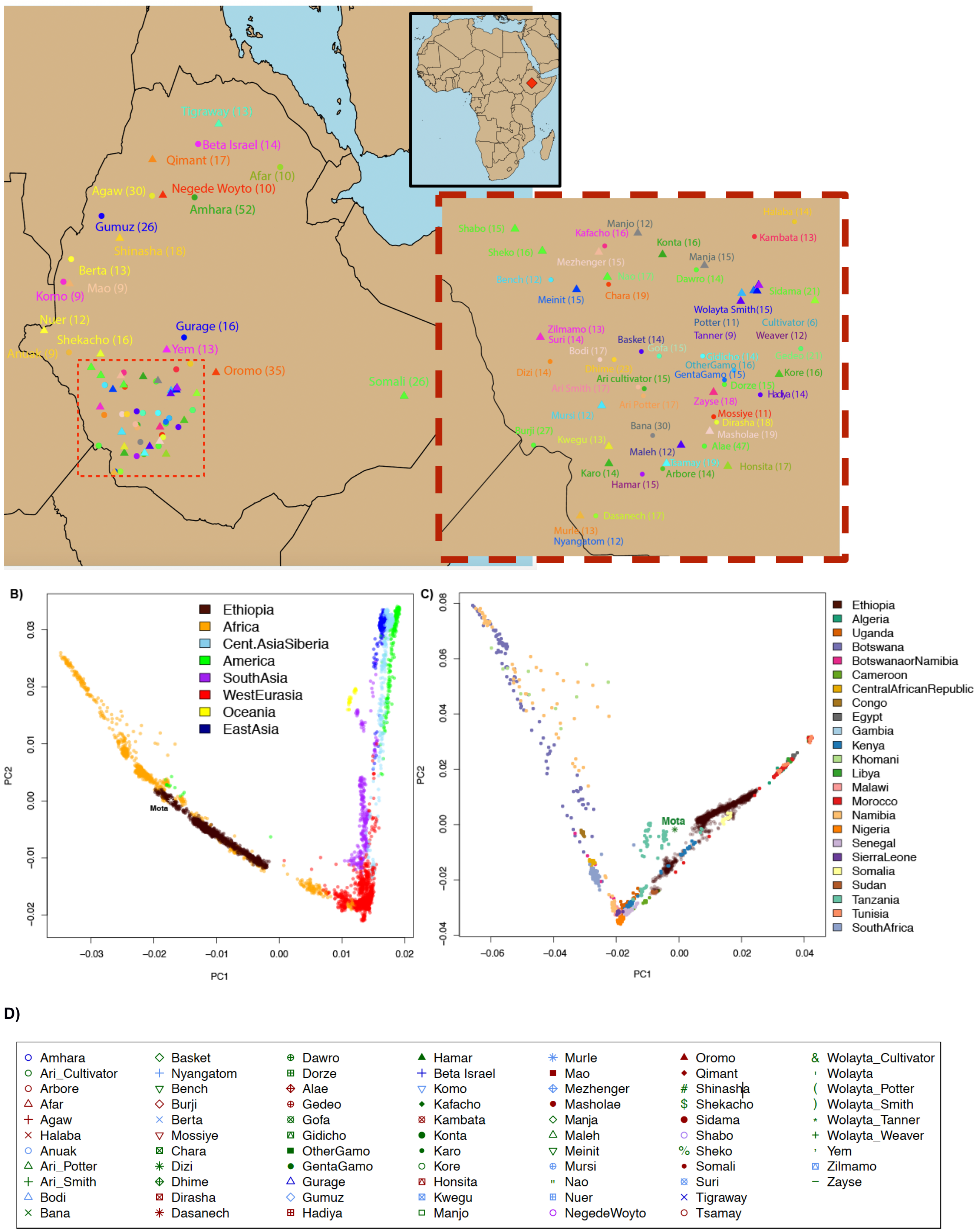


Figure S1. (A) Location of all sampled Ethiopian groups included in this study. Number of individuals from each self-reported group included in this work. Labels are placed at the average location (based on birthplace) of sampled individuals from that group. **(B)** First two principal components (PC) of a principal-components-analysis (PCA) of all the samples included in this study. **(C)** PCA of all the African samples included in this study. **(D)** Legend for Figure 1A.

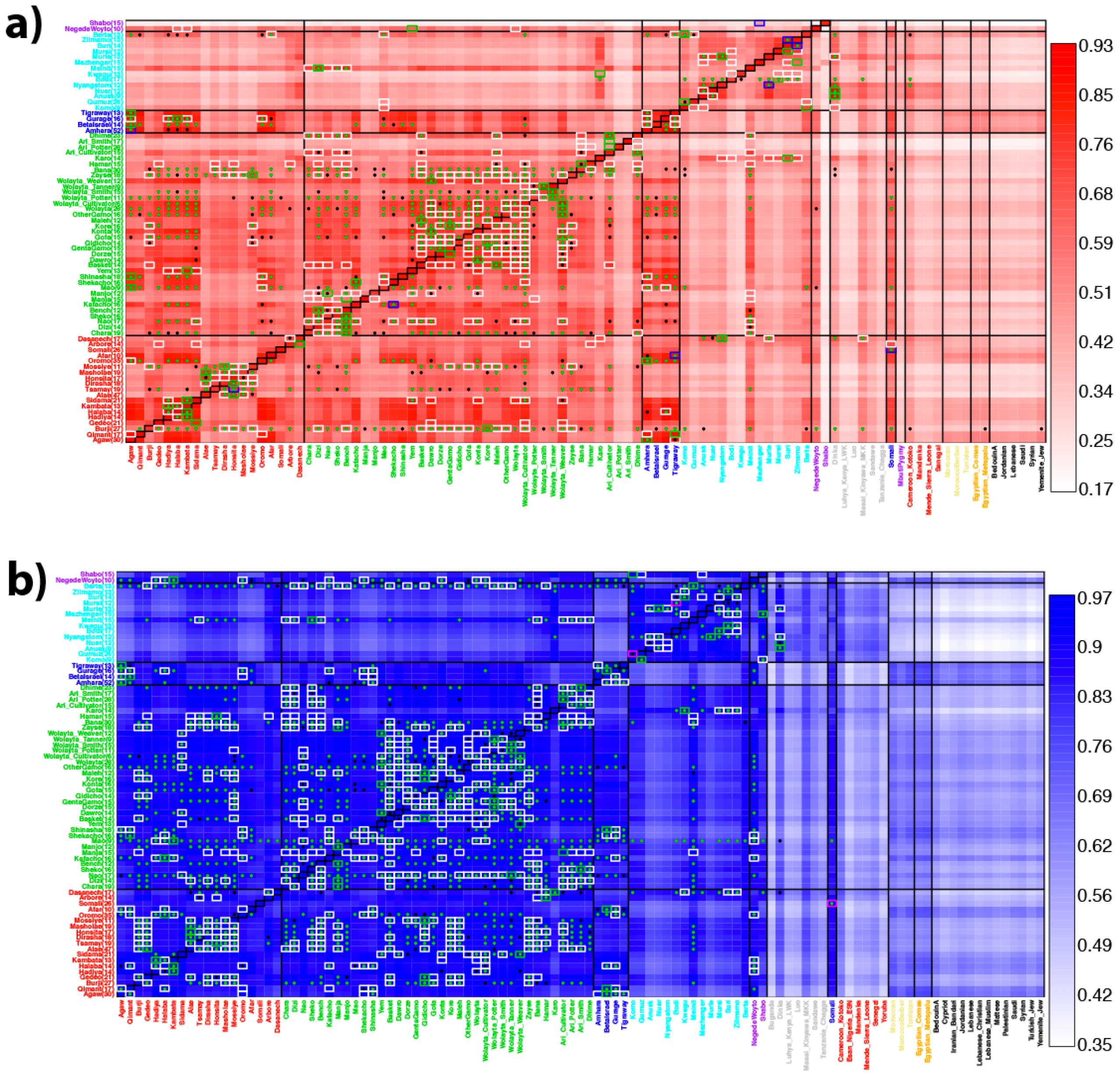


Figure S2. Average genetic similarity (1-TVD) between each Ethiopian group and other populations under the (a) "Ethiopian-internal" and (b) "Ethiopian-external" analyses. Each coloured square gives the average genetic similarity (1-TVD; legend at right) between all pairwise comparisons of individuals from the corresponding row, column (values given in Extended Tables 3-4). Columns include all Ethiopian groups, plus all non-Ethiopian groups with >= 7 sampled individuals that had relatively high genetic similarity to at least one Ethiopian group. Within each row *X*, dots denote groups (columns) for which the average genetic similarity between two individuals from group *X* is <u>not</u> significantly higher than that between an individual from *X* and an individual from the column group at Type I error rate = 0.001 (black) or 0.01 (green) (see Methods). Also within each row *X*, colored rectangles enclose: (black) the column for *X*; (green/blue/pink) the group (column) with highest average genetic similarity to *X*; (white) groups whose genetic similarity to *X* is not significantly lower than that between the group with highest genetic similarity to *X* at Type I error rate = 0.001. Blue and pink rectangles in Fig S2a and Fig S2b, respectively, signify that there are no other groups (columns) enclosed in white rectangles for the given row *X*, while green rectangles signify that there are. Ethnic group labels on axes are coloured by language classification for Ethiopian groups (legend in Fig 1a) and by major geographic region for non-Ethiopian groups (legend in Fig 1b).

[separate FigS3.pdf file]

Figure S3. FineSTRUCTURE inferred clusters of genetically homogeneous Ethiopian groups. fineSTRUCTURE's best-fitting tree relating its inferred clusters. Each leaf of the tree lists the number of individuals from each labeled group that were assigned to that cluster. Contiguous clusters of the same color were merged into one of the 78 final clusters we used in analysis; labels for these 78 clusters are provided at right. Full details in Extended Table 2.

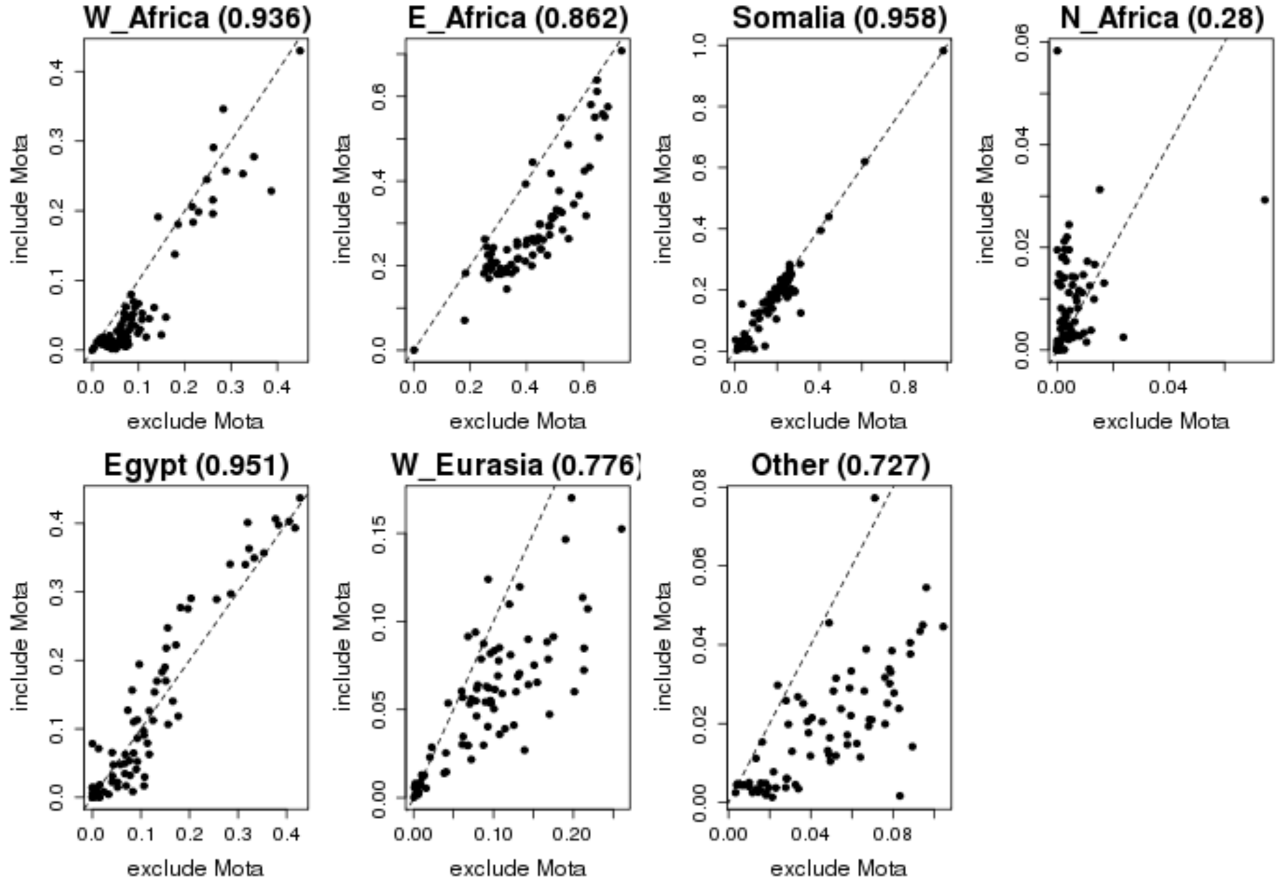


Fig S4 Proportion of haplotype matching to groups from six major demographic regions (plus "Other") between two SOURCEFIND analyses that include (y-axis) or exclude (x-axis) the 4.5kya Ethiopian Mota. Each dot represents one of the 78 Ethiopian clusters. Dashed line gives y=x; values in the title give the Pearson correlations.

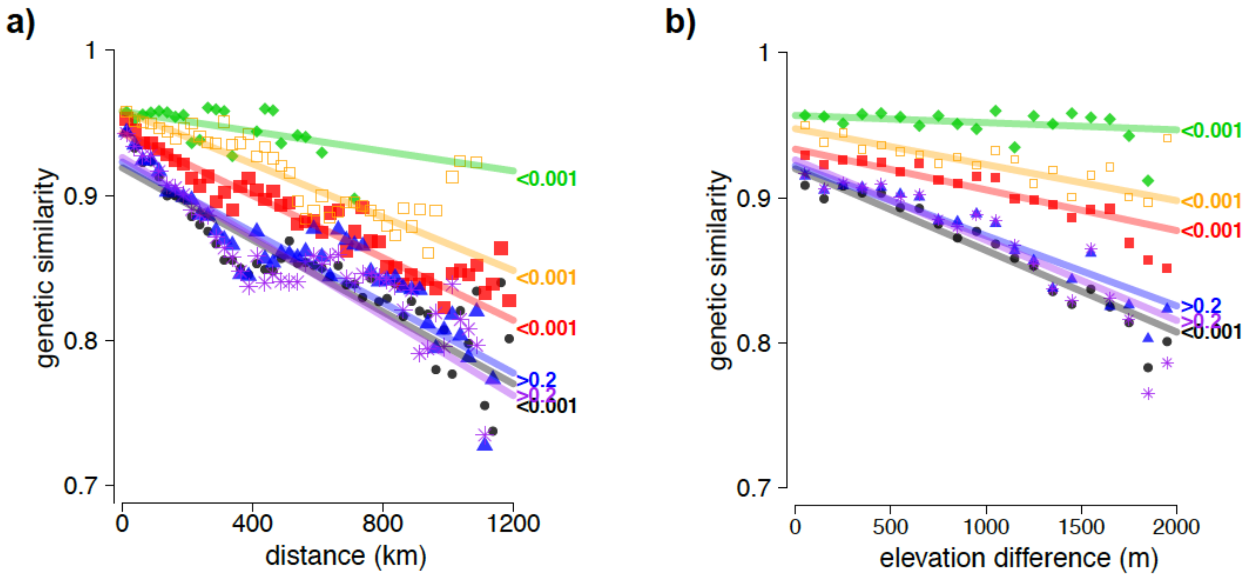


Fig S5 Genetic similarity correlates with spatial distance and shared cultural factors among Ethiopians under the "Ethiopia-external" analysis. **(a)** Fitted model for genetic similarity between pairs of individuals versus geographic distance, with points depicting the average genetic similarity within 25km bins, for all individuals (black; dots) or restricting to individuals who share group label (green; diamonds), speak the same first language (orange; open squares), speak the same second language (blue; triangles), have the same religious affiliation (purple; asterisks), or whose ethnicities are from the same language group (red; closed squares). Labels at right give permutation-based p-values when testing the null hypothesis of no increase in genetic similarity among individuals sharing the given trait (see Methods). **(b)** Analogous fitted model for genetic similarity versus elevation distance, with points depicting averages within 100km bins.

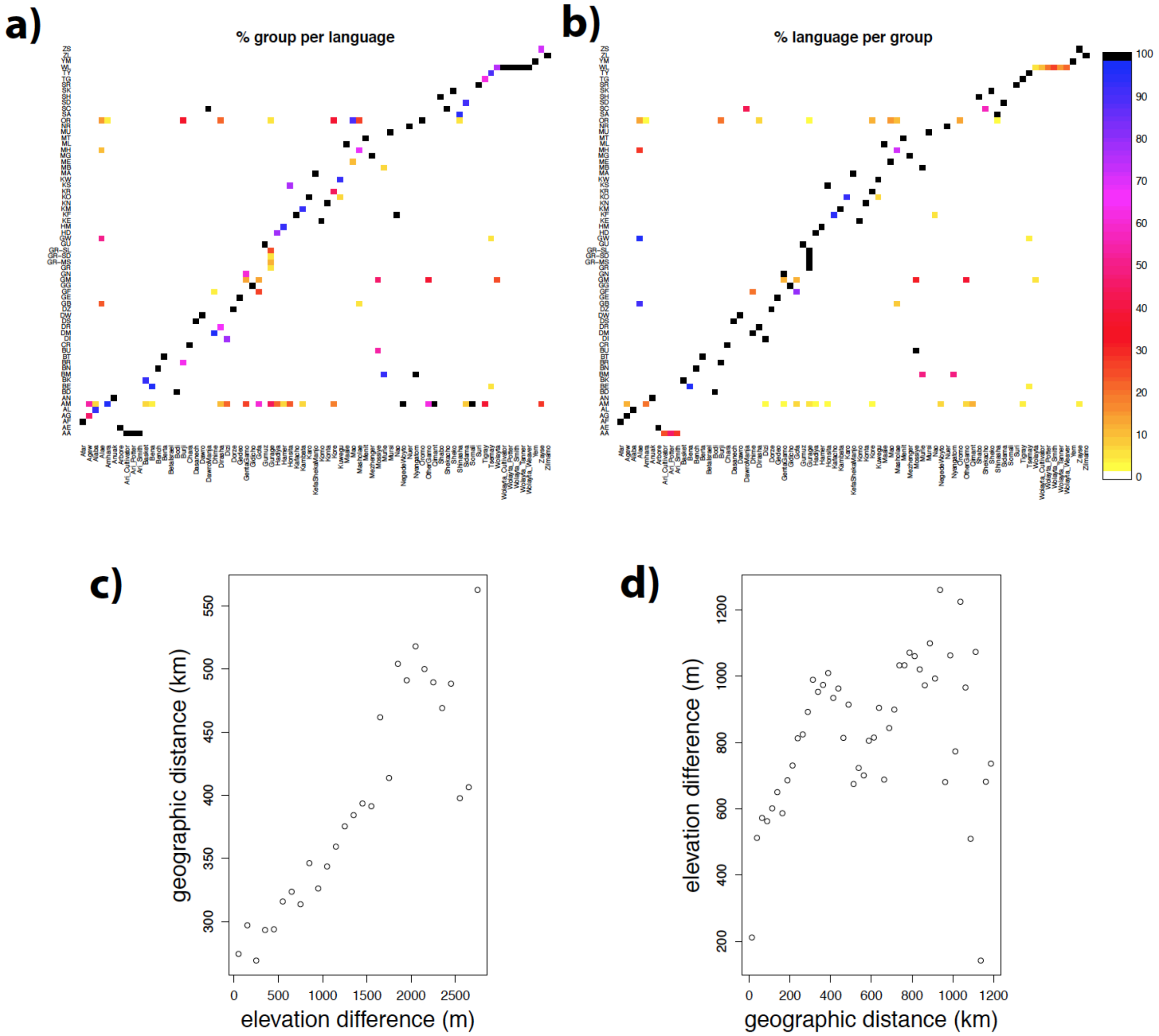


Fig S6 Correlation between first language and ethnicity, and between geographic and elevation distance. **(a)** Percentage of individuals from each group label (column) that speak the given first language (row). **(b)** Percentage of individuals from each first language (row) that fall into each group label (column). I.e. in the left plot, columns sum to 1, while in the right plot, rows sum to 1. The key for each linguistic label is given in Table S1. **(c)** Average geographic distance among individuals within 100-meter bins of elevation difference among individuals. **(d)** Average elevation difference among individuals within 25-kilometer bins of geographic distance.

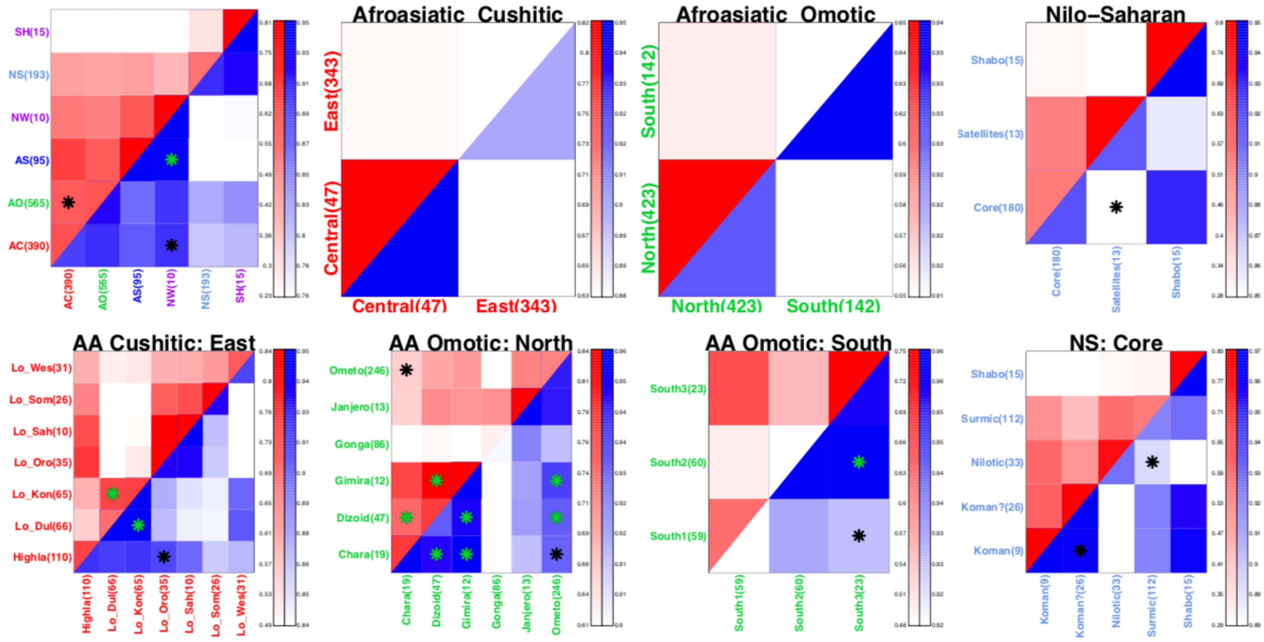


Fig S7 Average genetic similarity (1-TVD) between Ethiopians whose ethnicities are classified into different language groups. Row 1 gives results between within-family branches (column 1; second tier at www.ethnologue.com) and between sub-branches (columns 2-4; third tier at www.ethnologue.com), while row 2 gives results for further sub-classifications within sub-branches. Each plot gives results under the "Ethiopia-internal" (upper left triangle, red-scale) and "Ethiopia-external" (lower right triangle, blue-scale) analyses. Each square (row=A, column=B) within a plot gives the average genetic similarity (1-TVD; legend at right of each plot) between all pairwise comparisons of individuals where one individual is from A and the other is from B. Asterisks within each square indicate a lack of significant genetic differentiation between the two classifications at a Type I error level of 0.001 (black) or 0.01 (green), based on 100K permutations individuals' language classifications (see Methods). Numbers in parentheses on the axes give the number of individuals in each language classification, with AC: Afroasiatic Cushitic; AO: Afroasiatic Omotic; AS: Afroasiatic Semitic; NW: Negede-Woyto; NS: Nilo-Saharan; SH: Shabo; Lo_(Dul/Kon/Oro/Sah/Som/Wes): Lowland Dullay/Konso-Gidole/Oromo/Saho-Afar/Somali/Western. "Koman?" refers to the Gumuz. Though the Shabo have no official language family classification, we compared them to Nilo-Saharan sub-family classifications.

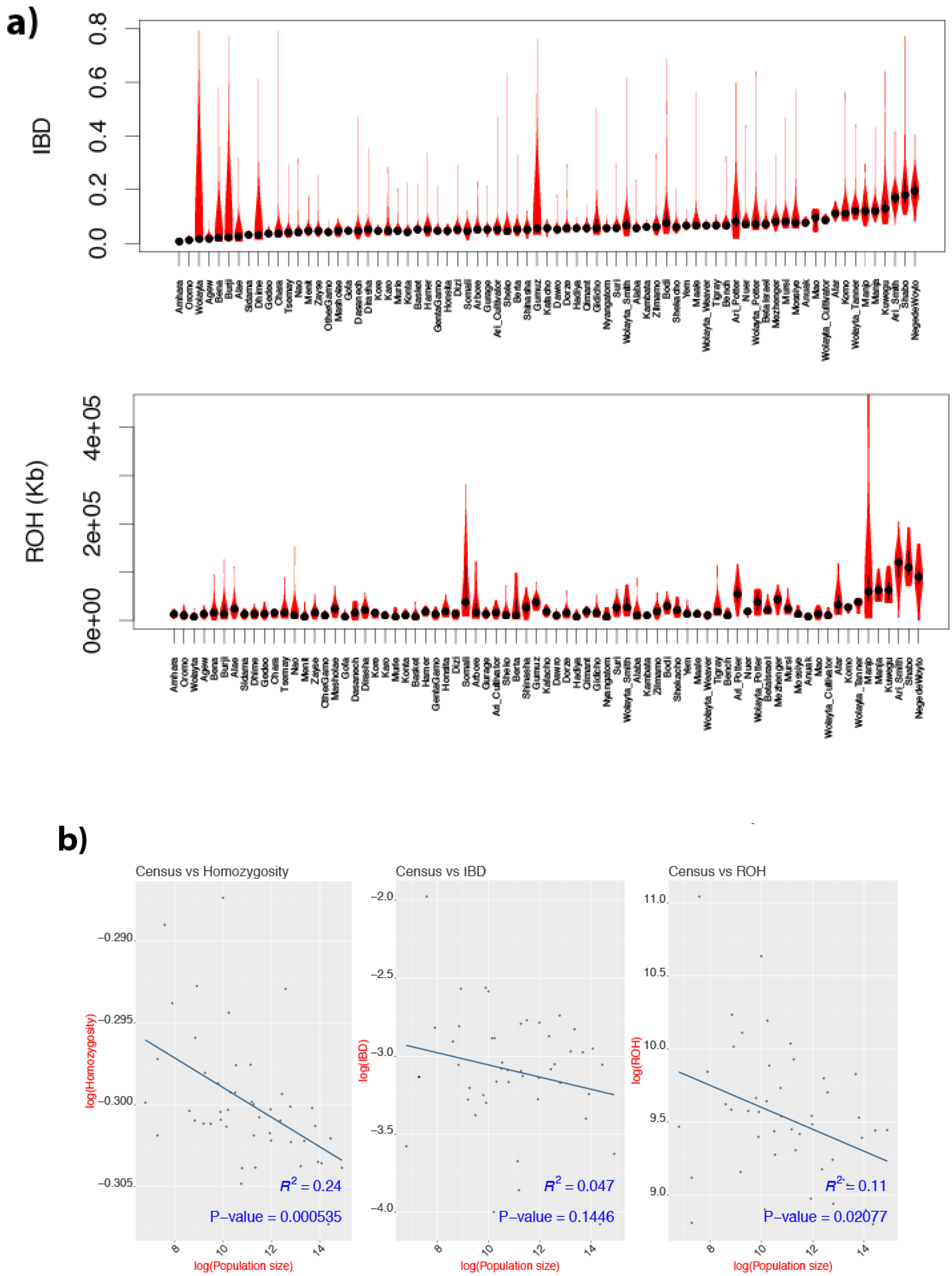
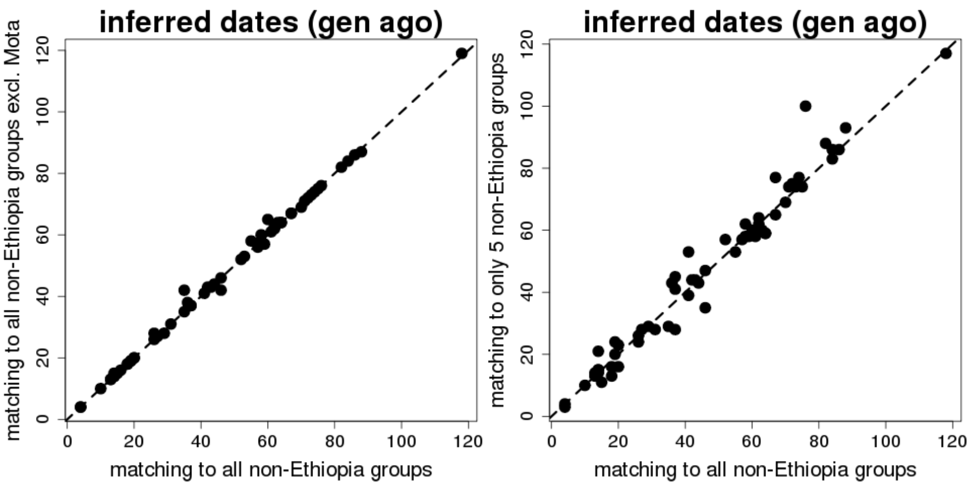

a)
IBD
ROH (Kb)
b)
Census vs Homozygosity
Census vs IBD
Census vs ROH
log(Homozygosity)
log(IBD)
log(ROH)
log(Population size)
R^2 = 0.24
P–value = 0.000535
R^2 = 0.047
P–value = 0.1446
R^2 = 0.11
P–value = 0.02077

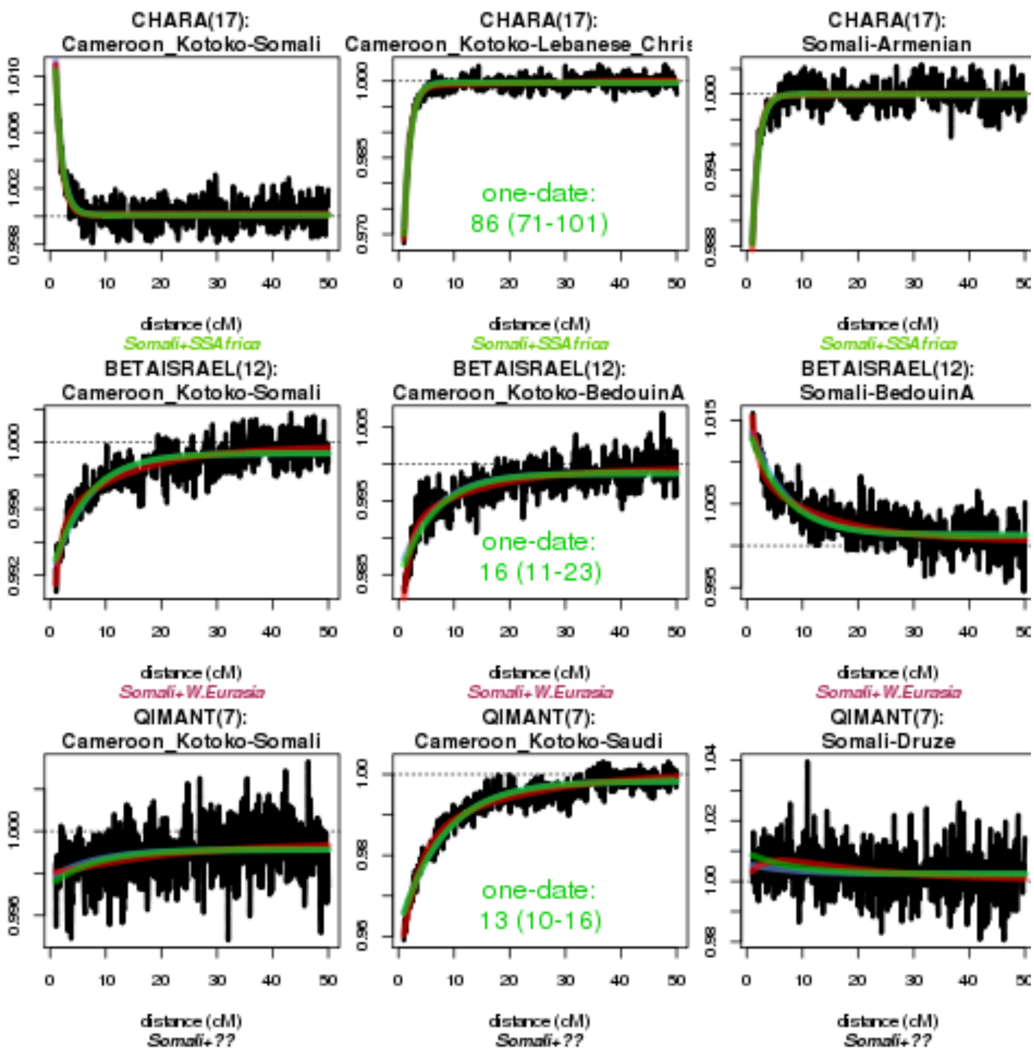
Figure S8 Genetic homogeneity estimates for the Ethiopian groups, and correlation with population census size. **(a)** Mean estimates (dots) and distributions (red) of Identity-by-Descent (IBD) sharing among individuals and Runs-of-homozygosity (ROH) within individuals within each ethnic group. The Negede Woyto, the Shabo and the Ari Blacksmiths share relatively longer haplotype segments with members of their own group, consistent with being genetically isolated from other populations. On the contrary, the Amhara and the Oromo show the lowest levels of within-group genetic homogeneity, suggesting that these individuals are not isolated and instead intermarry with (or have absorbed people from) other Ethiopian populations. **(b)** Linear regression between homozygosity, IBD and ROH against log population census size. In general, we note a significant increase in the genetic diversity (i.e. reduced homogeneity) of ethnic groups with larger population sizes. As expected, the highest levels of homozygosity were detected in populations with low census sizes, likely reflecting an elevated degree of endogamy and marriage between close kin.

a)



inferred dates (gen ago)
matching to all non-Ethiopia groups excl. Mota
matching to all non-Ethiopia groups
inferred dates (gen ago)
matching to only 5 non-Ethiopia groups
matching to all non-Ethiopia groups

b)



CHARA(17): Cameroon_Kotoko-Somali
CHARA(17): Cameroon_Kotoko-Lebanese_Chris
one-date: 86 (71-101)
CHARA(17): Somali-Armenian
distance (cM)
Somali+SSAfrica
BETAISRAEL(12): Cameroon_Kotoko-Somali
BETAISRAEL(12): Cameroon_Kotoko-BedouinA
one-date: 16 (11-23)
BETAISRAEL(12): Somali-BedouinA
Somali+W.Eurasia
QIMANT(7): Cameroon_Kotoko-Somali
QIMANT(7): Cameroon_Kotoko-Saudi
one-date: 13 (10-16)
QIMANT(7): Somali-Druze
Somali+??

c)

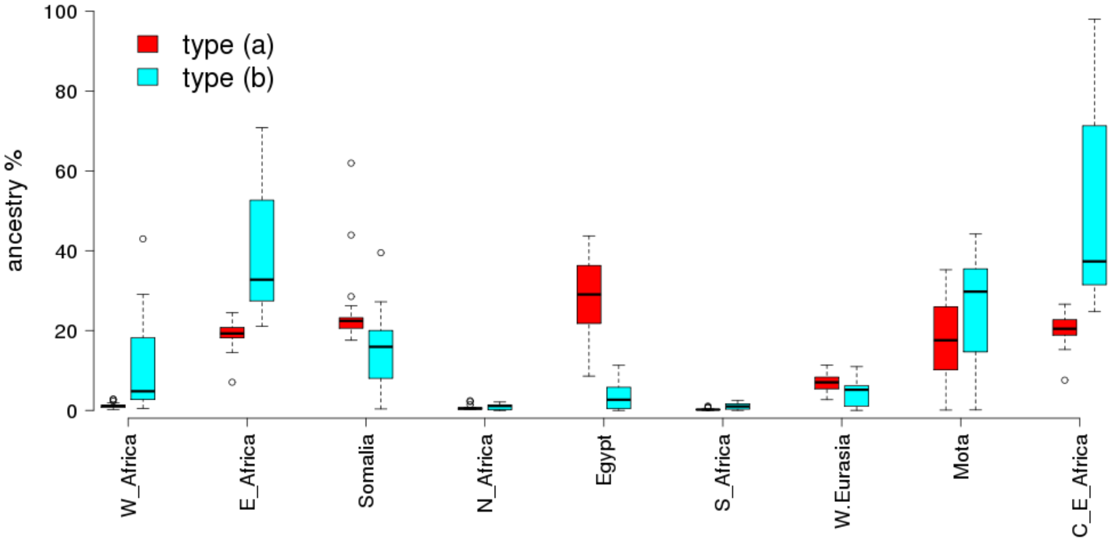


Fig S9 GLOBETROTTER/SOURCEFIND inference identifying and dating admixture events. **(a)** Robustness of GLOBETROTTER inferred dates of admixture (in generations ago) across Ethiopian clusters when using all non-Ethiopian populations as surrogates to admixing sources (x-axis; "Ethiopia-external" analysis) versus all non-Ethiopian surrogates excluding Mota (y-axis, left) or only the five non-Ethiopian surrogates {Bantu_SA, Han, Japanese, Mende_Sierra_Leone_MSL, Yemen} (y-axis, right). Clusters where either of the two compared analyses concluded "multiple-dates" of admixture are not depicted, as this can strongly affect date estimates. **(b)** GLOBETROTTER probability curves testing for admixture in three Ethiopian clusters, each inferring a single admixture event between two sources (i.e. "one-date"). Each row gives results for one Ethiopian cluster (given in the title), with the black lines in each plot depicting how the (scaled) probability of matching two DNA segments within that cluster's individuals to the surrogates listed just below the cluster name changes according to the centimorgan (cM) distance between the two DNA segments. The first column surrogates are Somalia and a West African population, the second column surrogates are West African and West Eurasian populations, and the third column surrogates are Somalia and a West Eurasian population. Green lines depict the model fit when assuming a single pulse of admixture, cyan lines when assuming a single pulse of admixture between two sources, red lines when assuming two pulses of admixture with distinct dates. In the middle plot, we provide the inferred date (in generations ago) and 95% CI. The first row depicts a cluster where we concluded that Somalia is aligned with Sub-Saharan African populations (i.e. "type (a) clusters"), as demonstrated clearly by the curve patterns (in that the Cameroon-Kotoko versus Somali curve decreases with increasing cM distance). In contrast, the second row depicts a cluster where we concluded that Somalia is aligned with West Eurasian populations ("type (b) clusters"). For the last row, the lack of pattern in the column 1 and 3 plots involving Somalia precluded a conclusion on which group Somali was aligned with. For analogous plots for all clusters, see Appendix B. **(c)** SOURCEFIND inferred ancestry proportions matching to groups from each major geographic region, across all 25 "type (a)" Ethiopian clusters and 28 "type (b)" Ethiopian clusters (see main text) for which Somalia represents the same admixing source as groups from Sub-Saharan Africa or from N_Africa/Egypt/W_Eurasia, respectively, under GLOBETROTTER inference.

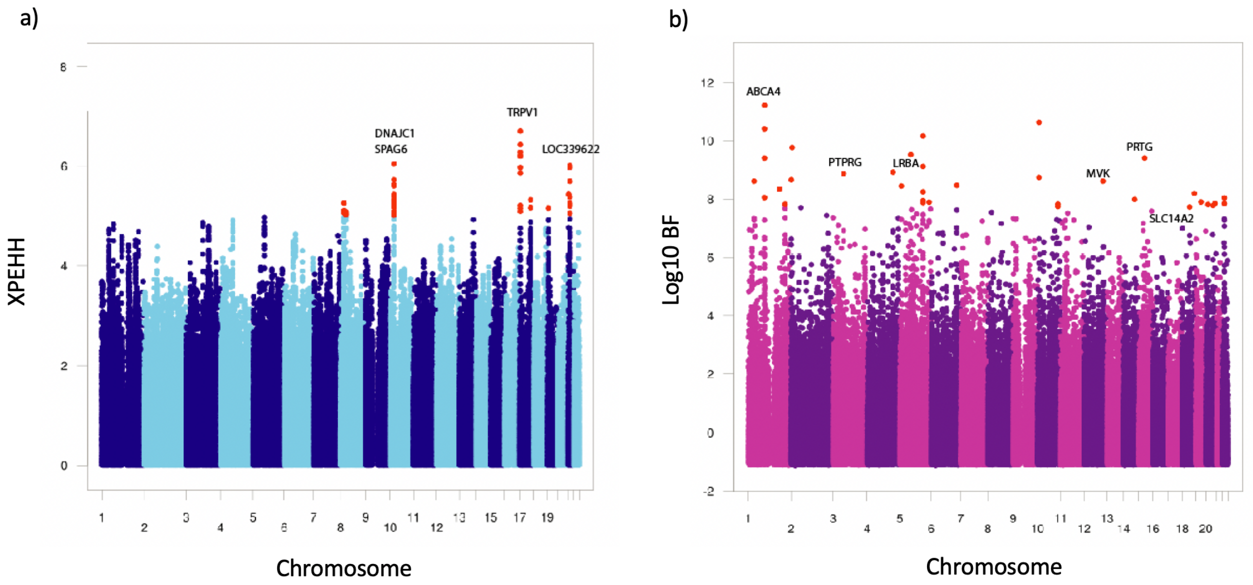


Fig S10. Selection tests to identify SNPs associated with elevation. (a) XPEHH scores when comparing 126 individuals living at the 10th percentile of elevation (0-563 meters) vs 119 individuals living at the 90th percentile of elevation (2302-3362 meters). Red dots highlight SNPs with scores >=5, with text giving the nearest genes to these SNPs. **(b)** Bayenv scores testing for an association between all 75 Ethiopian groups' allele frequencies and elevation. Red dots highlight SNPs in the 99.99th percentile, with text giving the nearest genes to these SNPs.

References

Alvarez, F. *The Prester John of the Indies: A True Relation of the Lands of the Prester John, being the narrative of the Portuguese Embassy to Ethiopia in 1520* (Beckingham, C. F. and Huntingford, G. W. B. (eds.) Cambridge, 1961).

Ashley, C. Z. Towards a socialised archaeology of ceramics in Great Lakes Africa. *Afr. Archaeol. Rev*. **27**, 135–163 (2010).

Bender, M. L. *Nilo-Saharan Overview In: The Non-Semitic Languages of Ethiopia* (Bender, M. L. (ed) African Studies Center, Michigan State University, Michigan, 1976)

Berry, L. B. *The Solomonic Monarchy at Gonder, 1630-1755: An Institutional Analysis of Kingship in the Christian Kingdom of Ethiopia* (PhD thesis, Boston: Boston University, 1976).

Black, P. *Linguistic Evidence on the Origins of the Konsoid Peoples* (Marcus and Hinnant (eds), Michigan State University, 1975)

Brandt, S. A. New perspectives on the origins of food production in Ethiopia. In: *From Hunters to Farmers. The Causes and Consequences of Food Production in Africa*. (J. D. Clark and S. A. Brandt (eds) Berkeley: University of California Press, 1984).

Budge, E. A. W. A Guide to the Babylonian and Assyrian Antiquities. (London: Printed Trustees, 1922).

Castanhoso, D. *Historia das Cousas que o muy Esforçado Capitao Dom Cristovao da Gama fez nos Reinos de Preste Joao com Questrocentos Portugueses que Consigo Levou*, translated in: Whiteway R. S., *The Portuguese Expedition to Abyssinia in 1541-1543 as narrated by Castanhoso* (London 1902, repr. Wiesbaden 1967).

Clist, B. A critical reappraisal of the chronological framework of the early Urewe Iron Age industry. *Muntu* **6**, 35–62 (1987)

Crawford, O. G. S. *Ethiopian Itineraries, circa 1400–1524* (Cambridge University Press, 1958).

Currey, J. *Foundations of an African Civilisation: Aksum and the northern Horn* (Oxford, 2014).

Ehret, C. *Cushitic Prehistory, In: The Non-Semitic Languages of Ethiopia* (Benrder, L.M. (ed)

African Studies Center, Michigan State University, Michigan, 1976).

Ethiopian Central Statistical Agency. *2007 Ethiopian census, first draft. (*2009)

Harlan, J. R. Ethiopia: A Center of Diversity. *Economic Botany* **23**, 309-314 (1969).

Hellenthal, G. et al. A genetic atlas of human admixture history. *Science* **343**, 747–751 (2014).

Johanson, D. C. *Lucy (Australopithecus afarensis). In: Evolution: The First Four Billion Years*.(Michael Ruse; Joseph Travis. (eds) Cambridge, Massachusetts: The Belknap Press of Harvard University Press, 2009).

Kalb, J. E. *Adventures in the Bone Trade: The Race to Discover Human Ancestors in Ethiopia's Afar Depression*(New York: Copernicus Books. 2001)

Knibb, M. A. Hebrew and Syriac elements in the Ethiopic version of Ezekiel? *Journal of Semitic Studies* **33**, 11-35 (1988).

Lovejoy, C. O. Re-examining human origins in light of Ardipithecus ramidus. *Science***326**, 74e1:8 (2009).

Malecot, G. Les Voyageurs francajs et les relations entre la France et L'Abyssinle de 1835 à 1870. Outre-Mers. *Revue d'histoire* **212**, 279-353 (1971).

Marcus, H. G. *A History of Ethiopia, Updated Edition* (Berkley, Los Angeles, London: The Regents of the University of California, 2002).

Merid, W. A. *Southern Ethiopia and the Christian Kingdom, 1508-1708, with Special Reference to the Galla Migrations and their Consequences* (Ph.D. Thesis, School of Oriental and African Studies (S.O.A.S.), University of London, 1971).

Muigai, A. W. T. & Hanotte, O. The Origin of African Sheep: Archaeological and Genetic Perspectives. *African Archaeological Review* **30**, 39-50 (2013)

Nersessian, V. & Pankhurst, R. The visit to Ethiopia of Yohannes T'Ovmacean, an Armenian jeweller, in 1764-66. *Journal of Ethiopian Studies* **15**, 79-104 (1982).

Pagani, L. et al. Ethiopian Genetic Diversity Reveals Linguistic Stratification and Complex Influences on the Ethiopian Gene Pool. *Am. J. Hum. Genet.* **91**, 83–96 (2012).

Phillips, J. Punt and Aksum: Egypt and the Horn of Africa. *J. African Hist.* **38**, 423-457 (1997).

Phillipson, D. W. *Ancient Ethiopia: Aksum, Its Predecessors and Successors* (London: The British

Museum Press, 1998).

Phillipson, D. W. *Foundations of an African Civilisation: Aksum & The Northern Horn 1000 BC – AD 1300* (Suffolk & Rochester, James Currey, 2012).

Phillipson, D. W. *The antiquity of cultivation and herding in Ethiopia. In The Archaeology of Africa: Food, metals and towns* (Thurstan Shaw, Paul Sinclair, Bassey Andah, and Alex Okpoko (eds) London, Routledge, 1993).

Polotsky, H. J. 1964. Aramaic, Syriac, and Ge'ez. *Journal of Semitic Studies* **9**, 1-10 (1964).

Quintana-Murci, L. et al. Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nat. Genet.* **23**, 437-441 (1999).

Sergew, H. S. *Ancient and Medieval Ethiopian History to 1270* (Addis Ababa: Artistic Printing Press. 1972).

Shahid, I. The Kebra Nagast in the Light of Recent Research, *Le Museón*, **89** (1976).

Suwa, G, et al. The Ardipithecus ramidus skull and its implications for hominid origins. *Science*. **326**, 68e1:7 (2009)

Sven, R. *The Survival of Ethiopian Independence* (Heinemann, London, 1978).

Taddesse, T. *Church and State in Ethiopia, 1270-1527* (London: The Clarendon Press, 1972).

The Council of Nationalities (CoN), Southern Nations and Peoples Region (SNNPR). *A Profile of the Nations, Nationalities and Peoples of the Southern Region* (Awassa, Ethiopia, 2017).

Ullendorf, E. 1980. Hebrew, Aramaic and Greek: the versions underlying Ethiopic translations of Bible and intertestamental Literature, In: *The Bible World. Essays in Honor of Cyrus H. Gordon* (G. Rendsburg, R. Adler, M. Arfa and N. H. Winter (eds.), New York, 1980).

Ullendorf, E. *Ethiopia and the Bible—The Schweich Lectures of the British Academy* (London, 1968).

Ullendorf, E. Hebrew elements in the Ethiopic Old Testament. *Jerusalem Studies in Arabic and Islam* **9,** 42–50 (1987).

Ullendorf, E. *The Ethiopians. An Introduction to Country and People* (London and New York, 1960).

Vansina, J. New Linguistic Evidence and 'the Bantu Expansion'. *J. Afr. Hist.* **36**, 173–195 (1995).