

Medieval DNA from Soqotra points to Eurasian origins of an isolated population at the crossroads of Africa and Arabia

Received: 24 July 2023

Accepted: 11 December 2023

Published online: 08 February 2024



A list of authors and their affiliations appears at the end of the paper

Soqotra, an island situated at the mouth of the Gulf of Aden in the northwest Indian Ocean between Africa and Arabia, is home to ~60,000 people subsisting through fishing and semi-nomadic pastoralism who speak a Modern South Arabian language. Most of what is known about Soqotri history derives from writings of foreign travellers who provided little detail about local people, and the geographic origins and genetic affinities of early Soqotri people has not yet been investigated directly. Here we report genome-wide data from 39 individuals who lived between ~650 and 1750 CE at six locations across the island and document strong genetic connections between Soqotra and the similarly isolated Hadramawt region of coastal South Arabia that likely reflects a source for the peopling of Soqotra. Medieval Soqotri can be modelled as deriving ~86% of their ancestry from a population such as that found in the Hadramawt today, with the remaining ~14% best proxied by an Iranian-related source with up to 2% ancestry from the Indian sub-continent, possibly reflecting genetic exchanges that occurred along with archaeologically documented trade from these regions. In contrast to all other genotyped populations of the Arabian Peninsula, genome-level analysis of the medieval Soqotri is consistent with no sub-Saharan African admixture dating to the Holocene. The deep ancestry of people from medieval Soqotra and the Hadramawt is also unique in deriving less from early Holocene Levantine farmers and more from groups such as Late Pleistocene hunter–gatherers from the Levant (Natufians) than other mainland Arabians. This attests to migrations by early farmers having less impact in southernmost Arabia and Soqotra and provides compelling evidence that there has not been complete population replacement between the Pleistocene and Holocene throughout the Arabian Peninsula. Medieval Soqotra harboured a small population that showed qualitatively different marriage practices from modern Soqotri, with first-cousin unions occurring significantly less frequently than today.

The four islands and two rocky outcrops of the Soqotra archipelago are situated in the northwest Indian Ocean at the mouth of the Gulf of Aden, approximately 380 km south of the Arabian Peninsula and 250 km northwest of the Horn of Africa. Soqotra, the largest island,

is renowned for endemic biodiversity and rich resources that have long attracted mariners and traders. Today, it is home to approximately 60,000 people with distinctive language, culture and traditions who subsist primarily through fishing, semi-nomadic pastoralism and the

cultivation of date palms. Modern Soqotri have cultural and genetic links to Arabia, southwest Asia and east and north Africa, but most of what is known about their deeper history derives from the writings of foreign traders. Importantly, our understanding of Soqatra's past is now also increasing through archaeological research¹.

The Hadramawt, a region of southern Arabia consisting of present-day eastern Yemen and a small part of southern Saudi Arabia (and in pre-modern times also including part of western Oman), has the strongest evidence of close connections to Soqotra since at least the first centuries of the Common Era (CE). Ceramics from Hajrya (located on Soqotra's north coast) are similar to those from Wadi Hadramaut (a valley east of Aden in Yemen), while inscriptions in Hoq cave on Soqotra's north coast attest to a strong southern Arabian influence beginning in the second century before the Common Era (BCE)², and similarities between dolmen-shaped structures on Soqotra and those in the Hadramawt may suggest links as far back as the second half of the first millennium BCE (refs. 3,4). The earliest account of people on Soqotra is in the *Periplus Maris Erythraei*, written in the mid-first century CE by an anonymous Greek trader from Egypt, who reported that the island ('Dioscorida') was ruled by the Hadramawt Kingdom and leased to Arabian merchants⁵, consistent with the hypothesis that the early inhabitants of Soqotra were South Arabian tribespeople who harvested frankincense during the height of the Hadramawt Kingdom's power^{3,6}. Furthermore, the Soqotri language is closest to the Šherēt ('Shahri', 'Jibbali') language of Dhofar, the region of southern Oman bordering Al Mahrah; both are Modern South Arabian languages not closely related to the Arabic spoken elsewhere in Arabia⁷.

Soqotra also played an important role in ancient long-distance trade networks, and there is extensive archaeological evidence of foreign traders on Soqotra who visited the island while traveling along maritime routes that linked Arabia with Persia, India and East Africa. The *Periplus* reported that Arab, Indian and Greek traders settled on the north coast of Soqotra⁵, while Cosmas Indicopleustes, a Greek traveller of the sixth century CE, reported that Christian missionaries from Persia settled on the island in the mid-first millennium CE (ref. 8). Although texts from this period surely reflect the biases of their authors, they are worth noting especially when corroborated by other lines of evidence. Excavations of South Asian ceramics on the western end of Soqotra also attest to the presence of Indian traders at this time². Inscriptions from Hoq cave record traders and travellers from Roman Egypt, Palmyra, Axum, Western India, Bactria and Gandhara (an ancient civilization centred in northwest Pakistan)⁸ between the second century BCE and sixth century CE, drawn there by the supply of aloes, incense and Dragon's Blood (Indian cinnabar)². The presence of foreign traders raises questions about the extent and nature of their interaction with local people, and ancient DNA provides a direct way to test whether this included genetic exchange.

An Arabic version of the Abstract is available as Supplementary Note 1.

Results and discussion

Data overview

We screened 50 skeletal elements from tafoni (sing. tafone) across six locations around Soqotra for DNA (Supplementary Notes 2 and 3). A tafone ('mesenaa'h' in Soqotri) is a geological formation resembling a small cave that has a round entrance and smooth, concave walls that forms part of the natural karst landscape. Both historically and today, Soqotri have used tafoni for storage, living quarters, animal shelters and places for burying the dead^{3,4,6,9–13}. We analysed a tooth from each individual in most cases and the DNA-rich petrous bone when possible¹⁴. Cortical bone from long bone shafts was analysed if no cranial remains were present. Fragmented skeletal material was collected to minimize disruption to the burials. We obtained genome-wide ancient DNA data passing quality control for 39 individuals across 15 tafoni but excluded from most population genetic analysis seven individuals with very low

coverage data (less than approximately 0.01 coverage of targeted positions) and an additional four individuals who are the lower-coverage first-degree relatives of another person in the dataset (Fig. 1a and Supplementary Data 1 and 2). The individuals we report here lived as early as 641–668 calibrated CE (cal CE) (1380 ± 15 uncalibrated years before present, BP) and mostly during the Medieval period (seventh to fifteenth centuries CE), with one dating to the Modern Period (1660–1950 cal CE (195 ± 15 BP)) (Supplementary Data 3).

We merged the new ancient DNA data with previously published ancient data from Africans, Eurasians and present-day people from around the world whose genomes had been shotgun sequenced or enriched for sequences covering a set of ~1.2 million single-nucleotide polymorphisms (SNPs). We also merged the new data with Affymetrix Human Origins (HO) SNP genotyping data from modern populations, including seven groups from Yemen^{15–17} and groups from other parts of Eurasia and Africa (Supplementary Note 4).

The Soqotri gene pool changed little over a millennium

We used principal component analysis (PCA) to visualize how the genomes of the medieval Soqotri relate to previously published people (Fig. 2a and Supplementary Note 4). Most Soqotri fall in a tight cluster, suggesting that the presence of foreign traders across ~1,000 years had little genetic impact or that any impact was relatively homogenous. Medieval Soqotri ancestry is qualitatively similar to that of some present-day Yemenite groups (Jew_Yemenite, Yemeni_Desert, Yemeni_Northwest, Yemeni_Highlands) but not others (Yemeni and Yemeni_Desert2) who appear to have a much higher proportion of sub-Saharan African-related ancestry. Relative to other ancient individuals, the medieval Soqotri fall nearest to ancient people from the Levant and ancient Egyptians with mostly Levantine-related ancestry but are shifted toward people with Iranian-related ancestry. Individual I21109 (dated to 641–668 cal CE) is shifted furthest away from the main cluster. We use pairwise *qpWave* to evaluate whether every pair of individuals can be modelled as descending from the same ancestral sources relative to a reference population set (they form a clade relative to that set) (Supplementary Note 5) and find that except for I21109, all Soqotri individuals are consistent with forming a single genetic group (Extended Data Fig. 1 and Supplementary Data 4). In pairwise comparisons, I21109 is rejected as forming a clade with all but five other individuals and is also rejected as forming a clade with a pool of the other Soqotri ($P = 0.003$), so we removed this individual from the pool we call 'Yemen_Seqotri'.

ADMIXTURE analysis shows that the medieval Soqotri, like most present-day groups from Yemen as well as Saudi and BedouinB, have a majority component (bright green) that is maximized in Late Pleistocene (Epipaleolithic) Natufian hunter-gatherers from the Levant (ISR_Natufian_EpiP in Fig. 2) and is inferred at high proportions in other Arabian and Levantine groups and some Africans from North and Northeast Africa who have substantial West Eurasian-related ancestry (Fig. 2b,c). The Soqotri additionally harbour an Iranian-associated component (grey; outlier I21109 has more of this component). Most distinctive about the Soqotri ADMIXTURE profile is their lack of an early European-like component (blue violet) otherwise found almost ubiquitously in people from Europe, Anatolia, the Levant and Arabia; they also have less of a sub-Saharan African-related component (purple) than modern Arabians. Across the Soqotri, we detect a minor contribution from a component maximized in some South Asian populations, consistent with records of Indian traders on the island, although we cannot reject the hypothesis that this component is wholly Central Asian in origin (see 'Medieval Soqotri have Central/South Asian-related ancestry').

The demographic impact of early Levantine farmers

To investigate the proportions of highly differentiated ancestries in the gene pools of Yemen_Seqotri and present-day people from the Arabian Peninsula ('Arabians') or Bedouins ('Arabian-related'; Supplementary

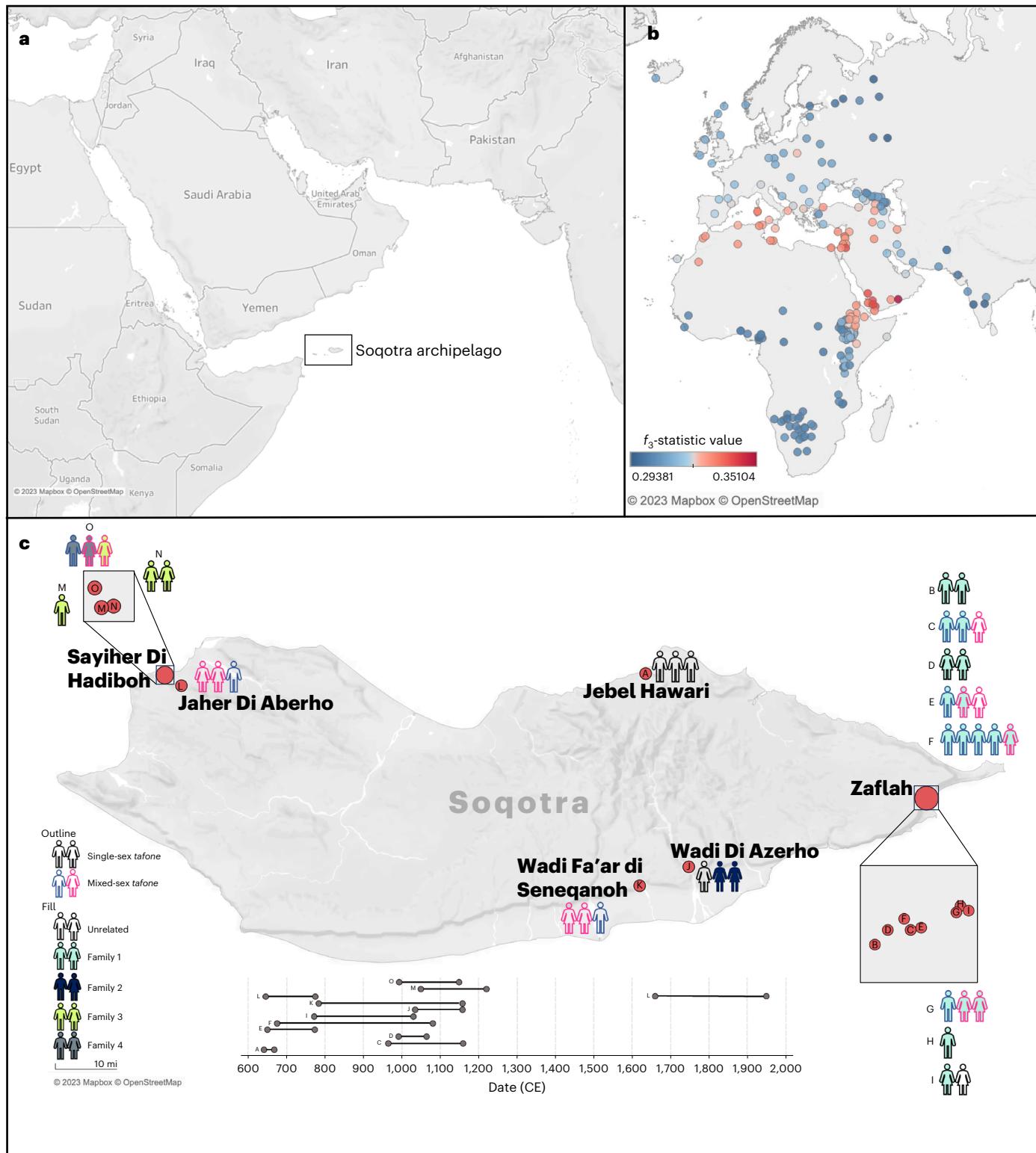


Fig. 1 | Overview of Soqotri burials with analysis of shared drift between medieval Soqotri and present-day groups. **a**, Location of Soqotra. The Soqotra archipelago, situated at the crossroads of Arabia and Africa, is identified by the black box. **b**, On the left, the 250 highest outgroup f_3 -statistics show shared drift between the medieval Soqotri and present-day populations genotyped on the Affymetrix HO array using Karitiana as an outgroup. The colour scale at the bottom left corner shows highest amount of allele sharing

with the Yemeni_Desert group. Data are in Supplementary Data 8, and additional details about outgroup f_3 -statistic analysis are in Supplementary Note 7. **c**, Burial patterns by sex and genetic relationships across 15 tafoni (A–O) on Soqotra. Legend is in the bottom left corner with additional information in Supplementary Data 14 and Supplementary Note 10. A timeline of tafone usage based on direct ^{14}C dates and dates of lineal relatives with known relationships is included (bottom centre).

Note 4), we carried out ancestry modelling using qpAdm. We used populations deeply diverged in time as proxies for ancestry sources because few if any modern populations are unadmixed descendants of populations that lived even a millennium ago in this region (for example, most groups show evidence of recent African admixture¹⁸).

Archaeological evidence has supported the hypothesis that Late Pleistocene hunter-gatherers associated with the Natufian culture inhabited the Arabian Peninsula as well as the Levant and that early farmers carrying Neolithic Levantine/Anatolian-related ancestry later spread into Arabia and had some demographic impact, although the proportion of ancestry this population contributed is a topic of debate^{19–21}. Genetic analyses of present-day groups from the Levant and Arabia show that Natufian hunter-gatherers appear to share more alleles with Arabians than with Levantine populations; however, a model in which all Levantine/Anatolian-related ancestry in Arabians derives from a farmer-related source is also consistent with the data²².

To test the hypothesis that people in Arabia may harbour variable Levantine/Anatolian-related ancestry, we applied qpAdm in a framework that models present-day Arabian or Arabian-related groups and the Soqotri as mixtures of ancient populations shown in previous work to be useful surrogates for ancestry found in Eurasians²³, plus an African-related source (Table 1). Fitting models required ancestry from ancient Levantine/Anatolian-, ancient Iranian- and African-related sources. The Levantine/Anatolian ancestry in most Arabian or Arabian-related groups can be well modelled either using a dual source of Natufian hunter-gatherers (ISR_Natufian_EpiP) plus Neolithic Anatolian farmers (TUR_Barcin_N) or, alternatively, using Neolithic Levantine farmers (Levant_PPNB) as a single source. This fit is consistent with Neolithic Levantine farmers representing a genetic intermediate between Natufians and Neolithic Anatolian farmers^{23–25}; we fit a model for Neolithic Levantine farmers as having $60.5 \pm 6.0\%$ Natufian-related and $39.5 \pm 6.0\%$ Neolithic Anatolian farmer-related ancestry ($P = 0.30$) (Supplementary Note 6).

In contrast to the majority of Arabian and Arabian-related groups, models using Neolithic Levantine farmers as a single source of the Levantine/Anatolian-related ancestry in Yemen_Soqotri are a poor fit ($P < 0.003$). However, a four-source model where Natufians are a proxy for approximately two-thirds of the total Levantine/Anatolian-related ancestry and Neolithic Anatolian farmers are a proxy for the remaining one-third fits the data ($P = 0.13$). In fact, a model without an African-related source fits the Soqotri data: in this more parsimonious model, Natufians fit as the source for over three-quarters of the total Levantine/Anatolian-related ancestry in the Soqotri gene pool ($P = 0.16$). The inferred relative ancestry proportions of Natufian-related ancestry (~66% or ~78% in a four- and three-source model, respectively) and Neolithic Anatolian farmer-related ancestry for the medieval Soqotri are higher than the inferred proportions of Natufian-related ancestry (~54–59%) and Neolithic Anatolian farmer-related ancestry in the present-day Yemeni populations not from the Hadramawt. This implies that the poor fit is driven either by slightly more Natufian-related ancestry in Yemen_Soqotri or, alternatively, by excess genetic drift shared between Natufians and Yemen_Soqotri that is not shared to the same extent between Natufians and other Arabians. In either case, we detect the effects of substructure among Late Pleistocene and early Holocene Levantine populations, with medieval Soqotri receiving relatively more ancestry from ancient groups closer to the sampled Natufians.

Two present-day Arabian groups, Yemeni_Desert and Yemeni_Desert2, both from the Hadramawt region but split into sub-groups because of variation in sub-Saharan African-related ancestry proportions, also show a signal of excess Natufian-related ancestry (Table 1, Supplementary Table 1 in Supplementary Note 6, and Supplementary Data 5). Like Yemen_Soqotri, Yemeni_Desert cannot be fit with Neolithic Levantine farmers as a single Levantine/Anatolian-related source ($P = 0.0007$) and instead can only be modelled with a dual

source of ~42% Natufian-related ancestry and ~24% Neolithic Anatolian farmer-related ancestry ($P = 0.10$). Results are qualitatively similar but less clear for Yemeni_Desert2, plausibly because of reduced power to discern distinctive ancestry due to a higher proportion of sub-Saharan African-related ancestry. The pattern of genetic similarity between these groups and the Soqotri is consistent with the archaeological evidence that Soqotra was likely peopled from, and maintained strong connections with, the Hadramawt³ and is also consistent with studies of uniparental markers that point to similarities between the Soqotri and non-Arabic speaking groups living in the southern Arabian Peninsula²⁶. By contrast, groups from parts of Yemen that were plausibly in closer contact with other parts of the Arabian Peninsula and regions further afield, as well as a group of present-day Saudi, can be fit either with Neolithic Levantine farmers or with Natufians plus Neolithic Anatolian farmers, suggesting that much of or all Levantine/Anatolian-related ancestry derives from a farmer-related source in these areas. This is consistent with a reduced displacement of hunter-gatherer-related ancestry in the ancestral gene pool of people who live in the Hadramawt and medieval Soqotra relative to people from other parts of Arabia, thus providing the strongest evidence to date that there has not been complete population replacement between the Pleistocene and Holocene throughout the Arabian Peninsula.

While the movements of Levantine farmers impacted the Soqotra archipelago relatively less than most of Arabia, the presence of both Anatolian Neolithic farmer- and Iranian-related ancestry in the Soqotri gene pool attests to additional gene flow. These ancestries could have arrived at least in part as a package from any of several different places. One possibility is that these ancestries reached Soqotra through later gene flow events from the Levant, as previous work has documented increased Anatolian Neolithic farmer- and Iranian-related ancestry in the Levant beginning in the Chalcolithic period, with a further increase in Iranian-related ancestry throughout the Bronze Age^{27,28}. We attempted to model Yemen_Soqotri using Chalcolithic or Bronze Age Levantine people as a source along with Natufians but found that additional Iranian-related ancestry was still required to fit the data (Supplementary Note 6 and Supplementary Data 6). As Soqotra was also part of key maritime trade routes that connected Arabia with India and Persia to the east and the eastern Mediterranean to the west, it is possible that either or both parts of this trade network had some genetic impact on the medieval Soqotri gene pool. We attempted to model Yemen_Soqotri using Hellenistic/Roman-era Anatolians who have a mix of Anatolian/Iranian ancestry as a proxy for first millennium CE Eastern Mediterranean traders as a source²⁹ and also found that fitting models again required additional Iranian-related ancestry (Supplementary Note 6 and Supplementary Data 6). While we cannot completely rule out that Levantine people or Eastern Mediterranean traders had some genetic impact on the medieval Soqotri, our results are consistent with the Soqotri receiving additional Zagros and/or Caucasus-related gene flow either directly from those regions or through an intermediate source not studied here.

Medieval Soqotri have little sub-Saharan African ancestry

While there is archaeological evidence of African influence on Soqotra—for example, the Aksumite conquest is directly attested by inscriptive evidence in Hoq cave and imagery at Dahaisi cave¹¹—our data argue against substantial sub-Saharan African admixture into the medieval Soqotri gene pool. Specifically, ancestry models fit without an African-related source at all (African-related ancestry proportions overlap zero at two standard errors in a four-source model, while a three-source model excluding an African-related source also fits; Table 1).

We tested the robustness of a three-source ancestry model by rotating individuals or groups with sub-Saharan African-related ancestry into the reference set (Supplementary Data 7 and Supplementary Note 6) and confirm that there is no genome-wide signal of

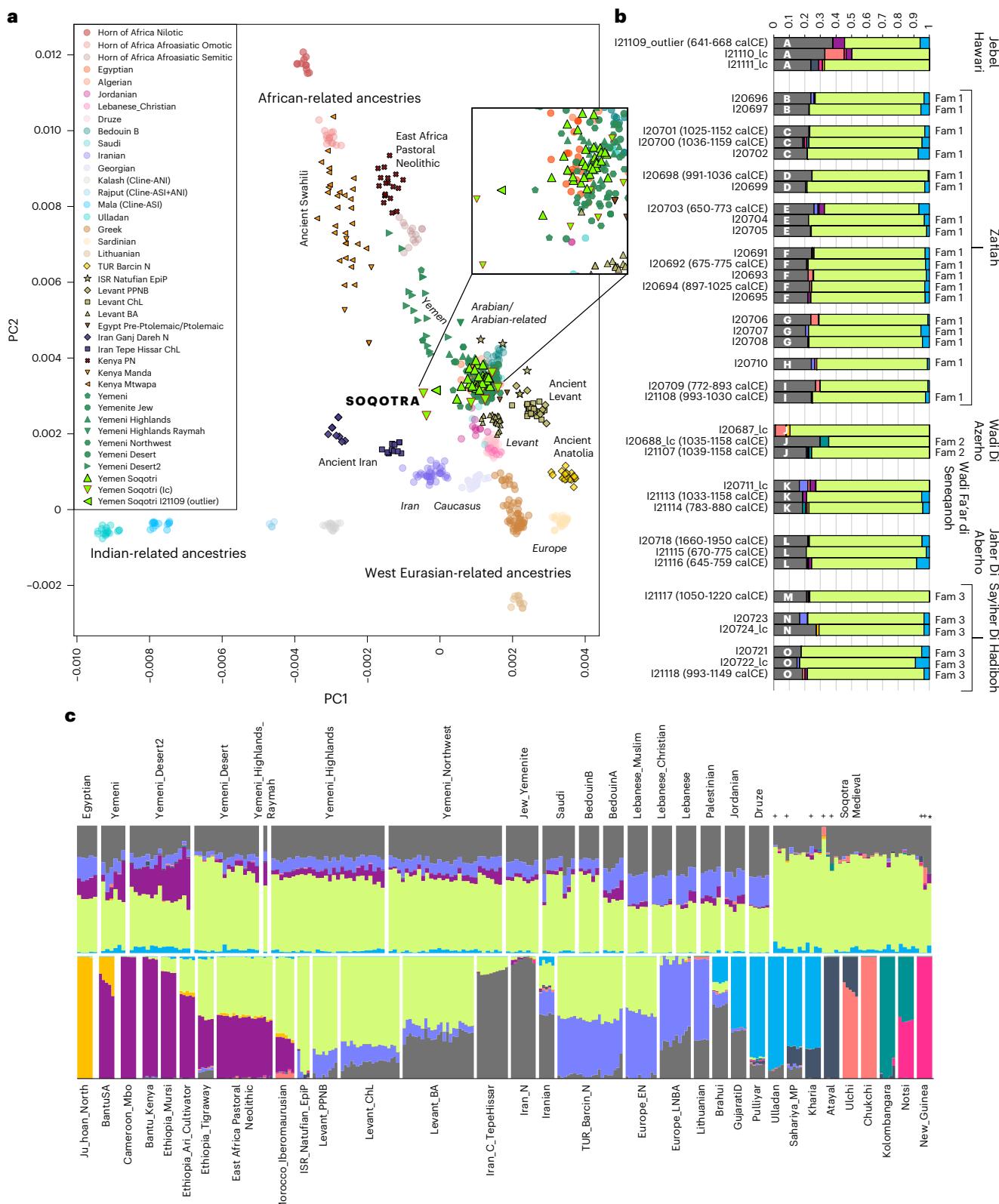


Fig. 2 | Qualitative analyses of medieval Soqotri DNA. a, PCA. We computed axes with present-day individuals (Supplementary Note 4) and projected ancient individuals onto PC1 and PC2. Inset provides a zoomed view of the Soqotra cluster. **b,** Unsupervised ADMIXTURE at $K = 10$ components for 39 Soqotri individuals. Individuals with $<15,000$ SNPs are marked '_lc' denoting low coverage; we include calibrated dates when available. Individuals are grouped by site starting on the northeast of the island and moving clockwise (site name written next to the plot), then by burial tafone (labelled A–O); see

Supplementary Data 14). Members of four families are identified (Fam 1 to Fam 4; see Supplementary Data 13). **c,** Additional results of unsupervised ADMIXTURE at $K = 10$ components. Data for the Soqotri and relevant present-day groups from Egypt and the Near East are on top. The Soqotri are in the same order as in **b** (top to bottom in **b** corresponding to right to left in **c**), with low coverage individuals denoted with a plus symbol and the outlier by an asterisk. A reference panel for the 10 components is on the bottom.

Table 1 | qpAdm models of the distal ancestry of present-day Arabian or Arabian-related groups and Yemen_Soqotri

Target	Model 1					Model 2					Proportion of Levant/Anatolian ancestry modelled with ISR_Natufian_EpiP
	Levant_PPNB	Iran_N	Eth_4500 BP	P value	ISR_Natufian_EpiP	Iran_N	TUR_Barcin_N	Eth_4500 BP	P value		
BedouinA	59.3±2.4	33.1±2.3	7.6±0.5	0.04174	31.7±5.5	35.5±2.7	25.1±4.9	7.8±1.0	0.00740	55.81%	
BedouinB	65.5±2.6	30.4±2.6	4.1±0.6	0.21748	34.1±5.6	32.7±2.9	28.7±5.1	4.5±1.1	0.10627	54.30%	
Saudi	61.9±2.6	36.2±2.6	2.0±0.6	0.36931	32.0±5.1	38.8±2.9	26.9±4.8	2.4±1.0	0.54233	54.33%	
Jew_Yemenite	63.2±2.7	34.1±2.7	2.7±0.6	0.21456	33.6±5.3	36.4±2.9	27.0±4.9	3.0±1.0	0.50612	55.45%	
Yemeni	48.8±2.5	36.8±2.5	14.4±0.7	0.14294	26.4±4.8	38.6±2.7	20.5±4.6	14.6±1.0	0.65806	56.29%	
Yemeni_Desert	67.5±2.8	28.8±2.8	3.7±0.6	0.00072	41.9±6.7	31.3±3.5	23.4±6.2	3.3±1.2	0.10329	64.17%	
Yemeni_Desert2	43.4±2.4	28.5±2.4	281±0.6	0.00342	27.7±5.9	31.4±2.7	13.4±5.2	27.5±1.2	0.02966	67.40%	
Yemeni_HIGHLANDS	62.9±2.5	31.1±2.5	6.1±0.5	0.01850	35.4±5.2	33.5±2.8	25.1±4.8	6.1±1.0	0.32483	58.51%	
Yemeni_HIGHLANDS_Raymah	61.4±4.4	19.0±4.4	19.6±1.3	0.27552	34.2±7.5	21.8±4.5	24.4±7.3	19.6±1.7	0.39685	58.36%	
Yemeni_Northwest	64.2±2.5	31.9±2.5	3.9±0.6	0.03131	33.2±4.9	33.9±2.7	28.5±4.5	4.3±0.9	0.45075	53.81%	
Yemen_Soqotri	65.6±2.9	32.3±2.8	2.1±0.6	0.00264	42.0±6.7	35.1±3.5	21.3±6.1	1.6±1.2	0.13450	66.35%	
Yemen_Soqotri (no African source)	67.6±2.9	32.4±2.9	-	0.00004	49.3±5.2	36.6±3.9	14.2±5.1	-	0.15674	77.64%	

Data shown as percentage ancestry (\pm s.e.) unless indicated otherwise. Model 1 uses Neolithic Levantine farmers (Levant_PPNB) as a single source of Levantine/Anatolian ancestry, while Model 2 uses a dual-source of Natufian hunter-gatherers (ISR_Natufian_EpiP) and Neolithic Anatolian farmers (TUR_Barcin_N) for Levantine/Anatolian ancestry. We include ancestry proportions and one standard error. Results significant at $P=0.01$ – 0.05 are in italics; results significant at $P>0.05$ are in bold and italics (we interpret a P value above 0.05 to be consistent with a plausible admixture model, while a P value between 0.01 and 0.05 indicates a marginal fit of the model). The last column reports the percentage of Levantine/Anatolian-related ancestry contributed by a Natufian-related (ISR_Natufian_EpiP-related) source for each group.

African-related ancestry that cannot be completely represented by Natufians, a group which has affinity to ~15,000-year-old Iberomauritanian hunter-gatherers from Morocco³⁰.

With mitochondrial DNA (mtDNA), we find the only evidence for African ancestry in medieval Soqotra. Specifically, we identify one individual belonging to African-associated haplogroup L3h2, today distributed primarily within northern parts of East Africa and Soqotra³¹; all other medieval Soqotri belong to the Eurasian-associated N and R branches of the mtDNA phylogeny (Supplementary Data 2), attesting to the strong Eurasian genetic connection that is also detected in mtDNA data from present-day people³².

Medieval Soqotri have Central/South Asian-related ancestry

To test which populations share the most genetic drift with Yemen_Soqotri, we computed outgroup f_3 -statistics using an indigenous population from Brazil (Karitiana) as Outgroup rather than an African population (such as Mbuti) to isolate the non-African part of the shared genetic drift and avoid the confounding effects of Holocene era African-related gene flow (Supplementary Fig. 6 shows results with Mbuti as Outgroup). This analysis confirms that the medieval Soqotri share more drift with Yemeni_Desert than any other present-day genotyped population (Fig. 1b, Supplementary Note 7 and Supplementary Data 8). D -statistics corroborate this strong relationship: D (Outgroup, Yemen_Soqotri; X, Yemeni_Desert) and D (Outgroup, Yemeni_Desert; Yemen_Soqotri, X) show that Yemen_Soqotri shares more alleles with Yemeni_Desert than with any other population ($|Z| > 8.5$ and 7.2, respectively), while D (Outgroup, X; Yemen_Soqotri, Yemeni_Desert) passes the threshold for significance (max $Z = 4.3$) only if a population with sub-Saharan African-related ancestry is X (Supplementary Data 9). We show with qpAdm that Yemeni_Desert has ~3% additional West African-related and ~18% Anatolian-related ancestry when Yemen_Soqotri is used as a proxy ancestry source ($P = 0.12$) (Supplementary Note 6). Present-day Soqotrigenesomes are needed to determine whether and to what extent these ancestries are found on Soqotra today.

Approximately 86% of medieval Soqotri ancestry can be modelled by Yemeni_Desert, with an additional ~14% proxied by a group with Iranian-related ancestry that has up to 2% contribution from the Indian

sub-continent (modelled using a group with minimal Iranian-related ancestry such as Pulliyar) (Supplementary Note 6). While this ancestry can be completely modelled with present-day Iranian people as a surrogate, an artificially produced population mixed with 80% or more Iranian ancestry and the remainder Indian ancestry also provides working models (Supplementary Note 6). While our ability to discern specifically Indian-related ancestry is constrained by the Iranian-related ancestry shared between Central Asian and many South Asian populations, our data are consistent with a proportion of Indian-related ancestry in the Soqotri gene pool that is close to our limit of definitive detection. mtDNA haplogroup U2b2, which is nearly exclusive to South Asia^{33,34}, further suggests a Soqotri matrilineal connection to this region. Possible Iranian matrilineal ancestry is documented by the presence of haplogroup R2c in five ancient Soqotri individuals³⁵. The high frequency of R2c in medieval Soqotra is distinct from the present-day mtDNA landscape, where most R haplotypes are part of the ROa clade hypothesized to have arisen in Arabia^{36–38}.

First-cousin unions were infrequent in medieval Soqotra

Like many other ethnically Muslim groups, modern Soqotri have a cultural preference for marriage between cousins. At least ~35% and up to ~40% of married respondents in an ethnographic study of Soqotri today were part of a first-cousin union (this range reflects ~5% of marriages classified as occurring with a ‘father’s/mother’s relative’ in ref. 3), with a preference for patrilineal orthocousin marriage³. To determine whether mate preference was similar in the past, we examined runs of homozygosity (ROHs), segments of the genome where the genetic material inherited from both parents is identical, in 14 medieval Soqotri individuals of sufficient coverage (>0.3 times coverage of targeted autosomal SNPs)³⁹ (Supplementary Data 10 and Supplementary Note 8). Long ROH segments (>20 cM (centiMorgans)) are indicative of parental relatedness within a few generations, while shorter ROH (4–8 cM) indicate relatedness deeper in the pedigree, reflecting a limited mating pool^{39,40}.

Over half of the medieval Soqotri have at least one long ROH, which has an appreciable chance of arising when parents are related at the level of fourth cousins or closer, indicating that many unions

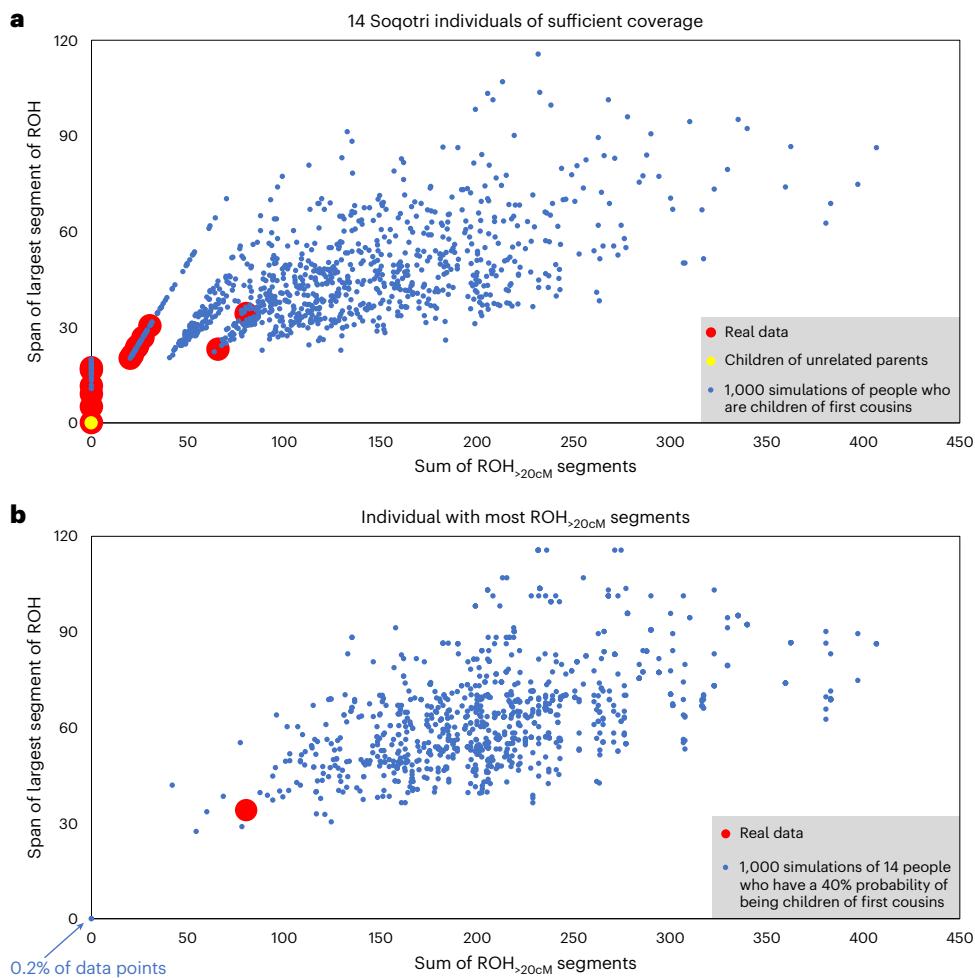


Fig. 3 | Medieval Soqotri had fewer long ROH than expected from modern ethnographic studies. We focused on long ROH (>20 cM) as it is unlikely to arise except for unions of people in the same extended family, and hence inferences based on it are relatively unaffected by artefacts due to small island population sizes. **a**, One thousand simulations of individual ROH profiles. We compare the real data to 1,000 simulations of people whose parents are first cousins, under the simplifying assumption that all other genealogical relationships between peoples' parents are very deep in the past. Treated as a single hypothesis test, I20705 (who has the most ROH out of the 14 ancient individuals for which we had ROH results) is in the $P = 0.26$ percentile (one-sided test) of simulations with

respect to the total summed ROH > 20 cM and the $P = 0.29$ percentile (one-sided test) of simulations with respect to the largest ROH segment. While this is in the range expected for the product of a first-cousin union when treated as a single hypothesis test, it is surprisingly low given that modern ethnographic rates would predict about five first-cousin unions in the data. **b**, One thousand simulated sets of 14 individuals. We compare the maximum ROH observed in 1,000 simulations of 14 individuals assuming the modern ethnographic rate of 40% first-cousin unions and find that the real data are at the $P = 0.007$ percentile (one-sided test) of simulations with respect to the total summed ROH > 20 cM and $P = 0.009$ percentile (one-sided test) of simulations with respect to largest ROH segment.

among medieval Soqotri were among people who belonged to the same extended family (see supplementary figure S16 of ref. 39). ROH data also provide evidence for an absence of close kin unions, defined here as first cousins or closer. Individual I20705 from Zaflah, who has the largest amount of ROH_{>20cM} in our dataset (80.3 cM ROH_{>20cM} with a max length of 34.3 cM), is at the lower end of the range expected for the offspring of first cousins (26th percentile from the perspective of summed ROH_{>20cM}; 29th percentile from the perspective of the longest segment; Supplementary Data 15). That the individual with the most ROH_{>20cM} is at the low end of the range expected for first cousins is surprising under the null hypothesis that the rate of first-cousin unions in medieval times is the same as today. Specifically, in a sample of 14 individuals, we would expect about five to be the offspring of first-cousin unions using the modern rate and would expect that the individual with the most ROH would be at a higher percentile of the first-cousin distribution.

We carried out simulations to assess the proportion of first-cousin unions in a sample of 14 individuals where all individuals have less ROH_{>20cM} than the maximum in our data (Fig. 3). Simulations show

that the data are incompatible with the modern value of 40% cousin unions ($P = 0.007$), and instead the $P = 0.05$ threshold corresponds to a maximum of 26% first-cousin unions (Supplementary Note 8). The data are also compatible with no first-cousin unions among the 14 medieval Soqotri: the total ROH_{>20cM} in I20705 is in the 95th percentile expected for second-cousin unions (81st percentile from the perspective of the largest single segment), which is unsurprising in a scenario that assumes that our sample of 14 medieval individuals includes multiple second-cousin unions (for caveats, see Supplementary Note 8).

Consistent with a founder effect and consanguineous unions, there is low diversity in Y chromosome haplogroups among medieval Soqotri males: all except one belong to J2-M172 and specifically to the same branch of J2b-Z534 (outlier I21109 belongs to J1a2a1a2d2b-Z2324). The near ubiquity of J2 haplogroups and paucity of J1 lineages is distinct from most of present-day mainland Yemen, where J1-M267 haplogroups dominate⁴¹. This is consistent with the differences in the population history between Soqotra and most of the Arabian Peninsula that we identify through deep ancestry modelling (Supplementary Note 9).

We used the distribution of ROH among medieval Soqotri individuals with sufficient coverage to estimate an effective community size (N_e) of 668 individuals (95% confidence interval (CI) 550–828 individuals). An estimate of N_e only in the hundreds of individuals provides further evidence of the relative isolation of the Soqotri population: despite a location along a key trade route, mating took place in a limited group resulting in a signal of genomic insularity, such as has been observed in other island communities (for example, ref. 42) (estimates for some present-day groups in Supplementary Data 11).

No evidence of sex segregation in burial tafone on Soqotra

A long-standing subject of debate in the archaeology of Soqotra is whether tafone that had multiple interments were segregated by sex^{3,4,6,9,10,13,43}. While collective burials in caves have been recorded in Saudi Arabia as well as in mainland Yemen, little is known about whether they represent families or tribe members⁴⁴.

Ancient DNA provide an opportunity to resolve these questions. We determined genetic sex for the 39 Soqotri individuals in our dataset, 21 who were consistent with being genetically female (XX) and 18 who were consistent with being genetically male (XY) (Supplementary Data 2). Leveraging a method that compares the mean mismatch rate of all autosomal SNPs covered for both individuals⁴⁵ and identity-by-descent (IBD)-based methods (Supplementary Data 12), we identify four genetic families (Fig. 2c) where people share relationships as close as first degree and as distant as fifth to seventh degrees (Supplementary Data 13). Seven out of 13 tafoni with the remains of multiple individuals contained members of both sexes, providing no support for a preference of placing individuals of the same sex together in death (Supplementary Data 14 and Supplementary Note 10). Closely related individuals were sometimes, but not always, interred in the same tafone. The absence of relatives buried across sites is consistent with a preference for returning people to their home burial area for interment if they died elsewhere.

Discussion

Genome-level data from medieval Soqotri people shine new light on an isolated population with very little sub-Saharan African-related ancestry and strong genetic links to the Hadramawt region. The medieval Soqotri had relatively more ancestry from ancient groups closer to Late Pleistocene Natufian hunter-gatherers, a signal also documented in present-day groups in the Hadramawt; by contrast, other present-day Arabian groups can be modelled as deriving all of their Levantine/Anatolian-related ancestry from a Neolithic farmer-related source. The reduced displacement of Natufian-related ancestry in Soqotra and the Hadramawt provides strong evidence that there has not been complete population replacement between the Pleistocene and Holocene throughout the entirety of the Arabian Peninsula.

The Soqotri gene pool remained stable over a millennium despite extensive archaeological evidence of foreign presence on the island, consistent with Soqotri oral history referring to the plateaus of the island interior that served as a refuge for local people to escape the attentions of foreign traders. Although traders and mariners were drawn to Soqotra by its rich resources and location along major Indian Ocean trade routes, they could arrive only by sea and rarely moved from coastal areas. Despite the remarkable archaeological evidence of foreign presence such as the Indian Brāhmī, Ethiopian Ge'ez, Roman Egyptian, and Palmyrene inscriptions within Hoq cave dating between the second century BCE and the sixth century CE, the Soqotri gene pool can be ~86% modelled with Yemeni_Desert as a proxy source, suggesting migration from the Hadramawt had a dominant influence shaping the medieval Soqotri gene pool. How the remaining ancestry that is best proxied by an Iranian-related source with up to 2% ancestry from the Indian sub-continent was introduced to the Soqotri gene pool is unclear but is plausibly linked to the extensive maritime trade network that connected Arabia with Persia and India as well as with the eastern Mediterranean. Although the Indian-related ancestry in the Soqotri gene pool

is close to or below our limit of definitive detection, archaeological and historical evidence support extensive Indian activity on Soqotra. For example, Indian inscriptions in Hoq cave are nearly all written in informal varieties of Brāhmī script and in Sanskrit or vernacularized Sanskrit dated to the second to fourth centuries CE (ref. 2), while text from the *Periplus* records India's provision of female slaves to Soqotra^{46–48}. Both the *Periplus* and epigraphic evidence specifically provide a record of traders coming from Barygaza (present-day Broach) and Hāthab in Gujarat on the northwest coast of India; epigraphic evidence also suggests Indians traveling to Soqotra from the Malabar coast of southwest India^{2,49}, while a single text is written in the Kharosthi script exclusively used in the northwest of India⁵⁰. The presence of Indians on Soqotra is continuously mentioned from the 10th to 13th centuries, although they were increasingly portrayed as pirates who at least had some children with the coastal population⁵¹.

We also find evidence of a shifting practice toward an increased frequency of cousin marriage from medieval to present-day Soqotra. This reduction in the rate of first-cousin unions would be consistent with a similar pattern observed in medieval northwest Pakistan where none of four individuals buried in a Muslim context from 900–500 BP had evidence of close parental relatedness, in sharp contrast to modern patterns of high rates³⁹. Considered together, these results not only suggest a relatively recent shift in cultural practice in Soqotra toward a preference for first-cousin marriages but also raise the possibility that this has been part of a broader cultural change toward preferences for such marriages in the Muslim world. In particular, it is possible that the spread of the practice of 'Bint'amm' marriages between paternal cousins—which is common in the Muslim world today and has been hypothesized to have spread from the Arabian Peninsula with the movement of Arab tribes in the seventh to eighth centuries^{52–54}—became so widespread only during the medieval-to-modern transition. These results are tentative, and an important topic for future work is to test this hypothesis in other parts of the Islamic world where first-cousin unions are common today.

Finally, a long-standing subject of debate in the archaeology of Soqotra regards burial patterns of people in natural tafone on Soqotra, first documented in the seventeenth century by the Carmelite P. Vincenzo who mentioned that each family on Soqotra had a cave in which they placed their deceased⁵⁵. This practice of burying relatives in caves was reported to reflect family members having no strength left to dig graves due to epidemics or drought. However, until now, it has been unclear whether tafone were segregated by sex. We show that 7 out of 13 tafoni containing the remains of multiple individuals included individuals of both sexes and provide evidence of familial as well as matrilineal and patrilineal relationships among people buried in a tafone, which is consistent with a strong preference for interment in one's local burial area in a family tafone, if possible, or with other tribespeople in other cases. An important opportunity for obtaining further insights will come from collecting genome-wide data from modern Soqotri people, which would make it possible to understand how they relate to people who lived on the island a millennium ago.

Methods

Community engagement

Ancient DNA can be used to investigate geographic origins, genetic affinities and population structure of people who lived long ago. In Soqotra, an important part of our project was to directly involve the local community in telling the story of the people who lived in the same region over the past two millennia. Long before ancient DNA analysis began, we established a relationship with multiple local people, specifically the Soqotra Governorate, the Director of the General Organization of Antiquities and Museums (GOAM) Soqotra, local heritage practitioners from the Soqotra Heritage Project and villagers living in the areas from which samples were taken. Collection of human skeletal material for this study was undertaken by a team of women and men from the

Soqotra Heritage Project and GOAM who received education and training in the identification, recording and assessment of the burials from which samples were taken. This included information about how and why samples of skeletal material are collected for ancient DNA studies. This knowledge, coupled with direct involvement in recording and collecting samples, provided local people with a deeper understanding of the fieldwork and helped them better communicate and educate local governmental officials, communities, students and school children during the various outreach activities that were carried out during and after fieldwork was completed. Permissions for the collection of samples were provided by GOAM and village leaders (sheiks), who either accompanied the team or provided guides to help in the location and identification of areas where sampling was permitted. Permits for export and analysis of all skeletal samples were obtained from GOAM (Supplementary Note 2).

Laboratory work and bioinformatics analysis

In ancient DNA-dedicated clean room facilities at Harvard Medical School, we generated ~40 mg of powder from bone and tooth samples, specifically targeting skeletal elements known to be especially DNA-rich^{14,56,57}. Details about sample selection can be found in Supplementary Note 2, and the skeletal element used for each individual can be found in Supplementary Data 1 and 2. We performed DNA extraction using silica-coated magnetic beads to support robotic clean-ups⁵⁸. From the extracts, we prepared dual-barcoded libraries using double-strand ligation, and we treated these libraries with uracil-DNA glycosylase (UDG) in a modified partial UDG preparation ('partial'), leaving a reduced damage signal at both ends (5' C-to-T, 3' G-to-A) (ref. 59; additional details in Supplementary Data 1). To generate SNP capture data, we used in-solution target hybridization to simultaneously enrich for sequences that overlap the mitochondrial genome and about 1.2 million genome-wide SNPs^{60–63}. We then added two seven-base pair indexing barcodes to the adapters of each double-stranded library and carried out sequencing using an Illumina HiSeqX10 instrument with 2 × 101 cycles and reading the indices with 2 × 7 cycles.

We merged paired-end sequences, retaining reads that showed no more than one mismatch between the forward and reverse base if base quality was ≥20 or three mismatches if base quality was <20 before alignment. A custom toolkit (available at <https://github.com/DReichLab/ADNA-Tools>) was used for merging and trimming adapters and barcodes. Merged sequences were mapped to the Reconstructed Sapiens Reference Sequence (RSRS)⁶⁴ and the human reference genome version hg19 using the same command in BWA 0.7.15-r1140⁶⁵ with the parameters -n 0.01, -o 2 and -l 16500. Duplicate molecules were removed following alignment using the Picard MarkDuplicates tool of the Broad Institute (v.2.17.10; <http://broadinstitute.github.io/picard/>). We trimmed two terminal bases from the DNA fragments in all partial-UDG libraries to reduce damage-induced errors.

We evaluated the authenticity of the isolated DNA by retaining individuals with a minimum of 3% of cytosine-to-thymine substitutions at the end of the sequenced fragments⁵⁹, point estimates of mtDNA contamination below 5% (estimates made with contamMix v.1.0–12)⁶⁶; and point estimates of X chromosome contamination in males, with sufficient coverage below 3% (estimates made with ANGSD v. 0.921–3-g40ac3d6 (htslib: 1.7) build(26 January 2018, 13:16:35))⁶⁷ (Supplementary Data 1).

We determined SNPs by randomly sampling an overlapping read with minimum mapping quality of ≥10 and base quality of ≥20, generating 'pseudo-haploid' calls. Individuals with fewer than 20,000 covered SNPs out of ~1.24 million were excluded from quantitative analyses. One individual from each of four pairs of first-degree relatives in the dataset was excluded from population genetics analysis when the data were grouped; in all cases, we retained the higher-coverage individual (Supplementary Data 1 and 2). Library information for all individuals studied as part of this work is in Supplementary Data 1, and we also

report negative results (samples that failed to yield authentic ancient DNA) in this table to prevent duplicative efforts in the future.

Radiocarbon dates and isotope data

We obtained 18 accelerator mass spectrometry radiocarbon dates (¹⁴C) from the Pennsylvania State University Radiocarbon Laboratory. We sonicated all of the bone samples in successive washes of American Chemical Society (ACS)-grade methanol, acetone and dichloromethane for 30 min each at room temperature, followed by three washes in Nanopure water to rinse to remove possible contaminants (conservants and adhesives). We extracted bone collagen and purified this using a modified Longin method with ultrafiltration (>30 kDa gelatin)⁶⁸. If collagen yields were low and amino acids were poorly preserved, we used a modified XAD process (XAD amino acids)⁶⁹. The method used for each sample can be found in Supplementary Data 3.

For quality assurance, we measured carbon and nitrogen concentrations and C/N ratios of all of the extracted and purified collagen/amino acid samples using a Costech elemental analyser (ECS 4010). We evaluated sample quality by percentage of crude gelatin yield, percentage of C, percentage of N and C/N ratios before accelerator mass spectrometry ¹⁴C dating. C/N ratios for all direct radiocarbon samples except one (I20701) fell between 3.1 and 3.3, indicating excellent preservation⁷⁰. We combusted collagen/amino acid samples (~2.1 mg) for 3 h at 900 °C in vacuum-sealed quartz tubes with CuO and Ag wires. Sample CO₂ was reduced to graphite at 550 °C using H₂ and a Fe catalyst, and we drew off reaction water with Mg(ClO₄)₂⁷¹. We pressed graphite samples into targets in aluminium boats, loaded them onto a target wheel and performed all ¹⁴C measurements using a modified National Electronics Corporation compact spectrometer with a 0.5 MV accelerator (NEC 1.5SDH-1). We corrected the ¹⁴C ages for mass-dependent fractionation with measured δ¹³C values⁷² and compared these with samples of whale bone from the Pleistocene (>48,000 ¹⁴C BP), bison bone from the late Holocene (~1,850 ¹⁴C BP), cow bone from the late 1800s CE and OX-2 oxalic acid standards. We calibrated ¹⁴C ages using OxCal v.4.4⁷³ and the IntCal20 Northern Hemisphere curve⁷⁴.

Dataset assembly

We merged the newly published ancient DNA data with previously published data from ancient Africans and Eurasians and present-day people from around the world whose genomes had been shotgun sequenced or enriched for sequences covering a canonical set of ~1,233,013 million SNPs (1240k data)^{60–63}, as well as whole-genome sequence data from ancient and present-day African and Near Eastern groups^{6,23–25,27–30,33,75–92}. We also merged the new data with Affymetrix HO SNP genotyping data (597,573 SNPs), which allows us to leverage data from a more diverse range of present-day populations from Eurasia and Africa^{15,17,25,75,82,92–106}. Whenever we needed to co-analyse 1240k data and HO array data, we restricted to SNPs also covered on the HO SNP array.

Uniparental haplogroups

We called mtDNA haplogroups using Haplotype2¹⁰⁷, reported in Supplementary Data 2. We determined Y chromosome haplogroups using both targeted SNPs and off-target sequences that aligned to the Y chromosome based on comparisons to the Y chromosome phylogenetic tree with Yfull version 8.09 (<https://www.yfull.com/>). We provide two notations for Y chromosome haplogroups in Supplementary Data 2. The first uses a label based on the terminal mutation, while the second describes the associated branch of the Y chromosome tree based on the nomenclature of the International Society of Genetic Genealogy (<http://www.isogg.org>) version 15.73 (July 2020).

PCA

We performed PCA with smartpca v.18200¹⁰⁸ and the parameters lsqproject: YES, 'newshrink': YES, 'numoutlieriter': 2, and 'hiprec': YES.

We computed eigenvectors of the covariance matrix of SNPs from 1,198 present-day individuals from 63 modern populations genotyped on the Affymetrix HO SNP array. We projected all other individuals (including ancient individuals) onto the first two principal components (PC1 and PC2). Full details of PCA set-up are in Supplementary Note 4.

ADMIXTURE (v.1.3.0)

We ran unsupervised ADMIXTURE¹⁰⁹ using 2,934 world-wide individuals either genotyped on the HO array or pulled down on the HO SNP set. We used convertf to transform the dataset from PACKEDANCESTRYMAP format to the PACKEDPED format for pruning in PLINK2¹¹⁰ and pruned SNPs based on linkage disequilibrium, using the indep-pairwise option with a window size of 200 variants, a step size of 25 variants, and a pairwise threshold of 0.4. We ran three replicates of ADMIXTURE with random seeds with $K = 3$ to $K = 12$ ancestral reference populations and present $K = 10$ based on low cross-validation errors with the goal of not overfitting.

Genetic relatedness

We tested genetic relatedness between every pair of newly reported individuals following the method described in ref. 45 and also applied IBD analysis on the X chromosome of males with sufficient coverage (>400,000 SNPs) to further specify the nature of genetic relationships (for example, avuncular) and identify more distant relatives (for example, fifth- to seventh-degree relatives). Results are in Supplementary Data 13, with discussion of genetic relationships in terms of burial groupings in 15 tafone in Supplementary Data 14. We use IBD analysis to show some additional more distant relationships for eight individuals from Zaflah with sufficient coverage in Supplementary Data 12.

Analysis of consanguineous genetic segments

We identified ROHs using the Python package hapROH (<https://test.pypi.org/project/hapROH/>), which uses 5,008 genomes from the 1000 Genomes project haplotype panel (1000 Genomes Project¹¹¹) as the reference panel. We restricted our analyses to 14 medieval Soqotri individuals who had a minimum coverage of 300,000 SNPs and called ROH > 4 cM. We used the default parameters of hapROH which are optimized for ancient DNA data. In Supplementary Data 10 we report the total sum amount and number of ROH in length bins 4–8 cM, 8–12 cM, 12–20 cM and >20 cM as well as the longest ROH block for each individual. We include comparative plots for Arabian and Arabian-related modern populations in Supplementary Note 8.

qpWave (v.1540)

We used qpWave from ADMIXTOOLS¹⁰² to test clinality between every pair of medieval Soqotri individuals in our analysis dataset. We used 'allsnps: YES' which calls qpfstats to calculate the relevant f_4 -statistics and set Han.SDG as the base population ('basepop: Han.SDG') for all qpfstats calculations (<https://github.com/DReichLab/AdmixTools/blob/master/qpfstats.pdf>). We set 'inbreed: NO' but also added the 'inbreedlistname' parameter so that any population with more than one individual would instead be analysed as if we set 'inbreed: YES'. Results are in Supplementary Data 4, with additional details in Supplementary Note 5.

qpAdm (v.1770)

We used the qpAdm framework⁶² to model ancestry and estimate ancestry proportions. qpAdm uses f_4 -statistics to detect share drift between the population of interest (the 'target' population) and possible admixing source proxies relative to a set of differentially related outgroup populations (referred to as the 'reference set'). It evaluates whether a population of interest can be plausibly modelled as descending from a common ancestor of one or more source populations and produces unbiased estimates of ancestry proportions from these sources⁶². For models consistent with the data ($P > 0.05$), qpAdm

estimates proportions of admixture for the target population from the specified source populations without requiring an explicit model for how the reference populations are related. We used 'allsnps: YES' which calls qpfstats to compute f -statistics and set Han.SDG as the base population ('basepop: Han.SDG') for all qpfstats calculations (<https://github.com/DReichLab/AdmixTools/blob/master/qpfstats.pdf>). We set 'inbreed: NO', but also added the 'inbreedlistname' parameter so any population with more than one individual would be analysed as if we set 'inbreed: YES'. Additional details are in Supplementary Note 6.

f -statistics

We computed f_4 -statistics using qpDstat (v.994) from ADMIXTOOLS¹⁰² with 'f4 mode: YES' and 'inbreed: YES.' We computed 'outgroup' f_3 -statistics using qp3Pop (v.670) from ADMIXTOOLS¹⁰² with 'inbreed: YES.' Additional details about f_3 -statistics as used in this work are in Supplementary Note 7.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The aligned sequences are available through the European Nucleotide Archive under accession number PRJEB66485. Genotype data are available at <https://reich.hms.harvard.edu/datasets>. Any other relevant data are available from the corresponding authors upon reasonable request.

Code availability

The custom code used in this study is available at <https://github.com/DReichLab/ADNA-Tools>.

References

1. Jansen van Rensburg, J. et al. The Socotra Heritage Project (2017–2020). *Bull. Soc. Arab. Stud.* **22**, 26–27 (2022).
2. Strauch, I. *Foreign Sailors on Socotra: The Inscriptions and Drawings from the Cave Hoq* (Hempen, 2012).
3. Naumkin, V. V. A. *Island of the Phoenix: An Ethnographic Study of the People of Socotra* (Paul & Company Pub Consortium, 1993).
4. Weeks, L., Morris, M., McCall, B. & Al-Zubairy, K. A recent archaeological survey on Soqotra. Report on the preliminary expedition season, January 5th–February 2nd 2001. *Arab. Archaeol. Epigr.* **13**, 95–125 (2002).
5. Casson, L. *The Periplus Maris Erythraei: Text with Introduction, Translation, and Commentary* (Princeton Univ. Press, 1989).
6. Doe, B. *Socotra: Island of Tranquillity* (Immel, 1992).
7. Morris, M., Watson, J. C. E. & Eades, D. *A Comparative Cultural Glossary across the Modern South Arabian Language Family* (Oxford Univ. Press, 2019).
8. Evers, K. Cave of revelations: Indian Ocean trade in light of the Socotran graffiti. *J. Indian Ocean Archaeol.* **10**, 19–37 (2014).
9. Bent, J. T. *Southern Arabia* (Smith, Elder and Co., 1900).
10. Doe, B. *Socotra; an Archaeological Reconnaissance in 1967* (Field Research Projects, 1970).
11. Jansen van Rensburg, J. Rock art of Soqotra, Yemen: a forgotten heritage revisited. *Arts* **7**, 99 (2018).
12. Naumkin, V. V. & Sedov, A. V. Monuments of Socotra. *Topoi Orient-Occident* **3**, 569–623 (1993).
13. Shinnie, P. L. Socotra. *Antiquity* **34**, 100–110 (1960).
14. Pinhasi, R. et al. Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS ONE* **10**, e0129102 (2015).
15. Lazaridis, I. et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
16. Mallick, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).

17. Vyas, D. N., Al-Meeri, A. & Mulligan, C. J. Testing support for the northern and southern dispersal routes out of Africa: an analysis of Levantine and southern Arabian populations. *Am. J. Phys. Anthropol.* **164**, 736–749 (2017).
18. Hellenthal, G. et al. A genetic atlas of human admixture history. *Science* **343**, 747–751 (2014).
19. Fedele, F. G. in *The Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics* (eds Petraglia, M. D. & Rose, J. I.) 215–236 (Springer, 2010).
20. McCroriston, J., & Martin, L. in *The Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics* (eds Petraglia, M. D. & Rose, J. I.) 237–250 (Springer, 2010).
21. Uerpmann, H.-P., Potts, D., & Uerpmann, M. in *The Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics* (eds Petraglia, M. D. & Rose, J. I.) 205–214 (Springer, 2010).
22. Almarri, M. A. et al. The genomic history of the Middle East. *Cell* **184**, 4612–4625. e4614 (2021).
23. Lazaridis, I. et al. The genetic history of the Southern Arc: a bridge between West Asia and Europe. *Science* **377**, eabm4247 (2022).
24. Lazaridis, I. et al. Ancient DNA from Mesopotamia suggests distinct pre-pottery and pottery Neolithic migrations into Anatolia. *Science* **377**, 982–987 (2022).
25. Lazaridis, I. et al. Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).
26. Černý, V. et al. Internal diversification of mitochondrial haplogroup rOa reveals post-last glacial maximum demographic expansions in South Arabia. *Mol. Biol. Evol.* **28**, 71–78 (2011).
27. Agranat-Tamir, L. et al. The genomic history of the bronze age southern levant. *Cell* **181**, 1146–1157 (2020).
28. Harney, É. et al. Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation. *Nat. Commun.* **9**, 1–11 (2018).
29. Lazaridis, I. et al. A genetic probe into the ancient and medieval history of Southern Europe and West Asia. *Science* **377**, 940–951 (2022).
30. van de Loosdrecht, M. et al. Pleistocene north African genomes link near eastern and sub-Saharan African human populations. *Science* **360**, 548–552 (2018).
31. Soares, P. The expansion of mtDNA haplogroup L3 within and out of Africa. *Mol. Biol. Evol.* **29**, 915–927 (2012).
32. Černý, V., Čížková, M., Poloni, E. S., Al-Meeri, A. & Mulligan, C. Comprehensive view of the population history of Arabia as inferred by mtDNA variation. *Am. J. Phys. Anthropol.* **159**, 607–616 (2016).
33. Narasimhan, V. M. et al. The formation of human populations in South and Central Asia. *Science* **365**, eaat7487 (2019).
34. Shinde, V. et al. An ancient Harappan genome lacks ancestry from steppe pastoralists or Iranian farmers. *Cell* **179**, 729–735 (2019).
35. Derenko, M. et al. Complete mitochondrial DNA diversity in Iranians. *PLoS ONE* **8**, e80673 (2013).
36. Al-Abri, A. et al. Pleistocene–Holocene boundary in southern Arabia from the perspective of human mtDNA variation. *Am. J. Phys. Anthropol.* **149**, 21–28 (2012).
37. Černý, V. et al. Out of Arabia—the settlement of Island Socotra as revealed by mitochondrial and Y chromosome genetic diversity. *Am. J. Phys. Anthropol.* **138**, 439–447 (2009).
38. Gandini, F. et al. Mapping human dispersals into the Horn of Africa from Arabian Ice Age refugia using mitogenomes. *Sci. Rep.* **6**, 25472–25472 (2016).
39. Ringbauer, H., Novembre, J. & Steinrücken, M. Parental relatedness through time revealed by runs of homozygosity in ancient DNA. *Nat. Commun.* **12**, 5425 (2021).
40. Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M. & Wilson, J. F. Runs of homozygosity: windows into population history and trait architecture. *Nat. Rev. Genet.* **19**, 220–234 (2018).
41. Cadena, A. M., Zhivotovsky, L. A., Cavalli-Sforza, L. L., Underhill, P. A. & Herrera, R. J. Y-chromosome diversity characterizes the Gulf of Oman. *Eur. J. Hum. Genet.* **16**, 374–386 (2008).
42. Ariano, B. et al. Ancient maltese genomes and the genetic geography of neolithic Europe. *Curr. Biol.* **32**, 2668–2680.e6 (2022).
43. Wellsted, J. R. *Travels to the City of the Caliphs, Along the Shores of the Persian Gulf and the Mediterranean: Including a Voyage to the Coast of Arabia, and a Tour on the Island of Socotra* (2 volumes) (Henry Colburn, 1840).
44. Luciani, M., Binder, M. & Alsaud, A. S. Life and living conditions in north-west Arabia during the Bronze Age: first results from the bioarchaeological work at Qurayyah. *Proceedings of the Seminar Arabian Studies* **48**, 185–20 (2018).
45. Olalde, I. et al. The genomic history of the Iberian Peninsula over the past 8000 years. *Science* **363**, 1230–1234 (2019).
46. Dridi, H. Indiens et proche-orientaux dans une grotte de Suqutrā (Yemen). *J. Asiatique* **290**, 565–610 (2002).
47. Dridi, H. & Gorea, M. Au II^e siècle de notre ère: le voyage d'Agbar à Suqutra. *Archeologia* **396**, 48–57 (2003).
48. Robin, C. J. & Gorea, M. Les vestiges antiques de la grotte de Höq (Suqutra, Yémen)(note d'information). *C. R. Séances Acad. Inscr. B.-Lett.* **146**, 409–445 (2002).
49. Strauch, I. & Bukharin, M. Indian inscriptions from the Cave Höq on Suqutrā. *Ann. dell'Istituto Universitario Orient. Napoli* **64**, 121–138 (2004).
50. Strauch, I. in *Ports of the Ancient Indian Ocean* (eds. Boussac, M.-F. et al.) 79–97 (Primus Books, 2016).
51. Al-Mujāwir, I., & ibn Ya'qūb, Y. *A Traveller in Thirteenth-Century Arabia: Ibn al-Mujāwir's Tārikh al-Mustabṣir* (Ashgate, 2008).
52. Bittles, A. H., & Hamamy, H. A. in *Genetic Disorders among Arab Populations* (ed. Teebi, A. S.) 85–108 (Springer, 2010).
53. Korotayev, A. Parallel-cousin (FBD) marriage, islamization, and arabization. *Ethnology* **39**, 395–407 (2000).
54. Stern, G. H. *Marriage in Early Islam* (Royal Asiatic Society, 1939).
55. Yule, H., & Cordier, H. *The Travels of Marco Polo: The Complete Yule-Cordier Edition: Including the Unabridged Third Edition (1903) of Henry Yule's Annotated Translation, as Revised by Henri Cordier, Together with Cordier's Later Volume of Notes and Addenda (1920)* (Dover Publications/Constable, 1993).
56. Hansen, H. B. et al. Comparing ancient DNA preservation in petrous bone and tooth cementum. *PLoS ONE* **12**, e0170940 (2017).
57. Pinhasi, R., Fernandes, D. M., Sirak, K. & Cheronet, O. Isolating the human cochlea to generate bone powder for ancient DNA analysis. *Nat. Protoc.* **14**, 1194–1205 (2019).
58. Rohland, N., Glocke, I., Aximu-Petri, A. & Meyer, M. Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nat. Protoc.* **13**, 2447–2461 (2018).
59. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S. & Reich, D. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B* **370**, 20130624 (2015).
60. Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
61. Fu, Q. et al. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl. Acad. Sci. USA* **110**, 2223–2227 (2013).
62. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
63. Mathieson, I. et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
64. Behar, D. M. et al. A 'Copernican' reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012).

65. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
66. Fu, Q. et al. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
67. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinform.* **15**, 356 (2014).
68. Kennett, D. J. et al. Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun.* **8**, 14115 (2017).
69. Lohse, J. C., Madsen, D. B., Culleton, B. J. & Kennett, D. J. Isotope paleoecology of episodic mid-to-late Holocene bison population expansions in the Southern Plains, USA. *Quat. Sci. Rev.* **102**, 14–26 (2014).
70. Van Klinken, G. J. Bone collagen quality indicators for palaeodietary and radiocarbon measurements. *J. Archaeol. Sci.* **26**, 687–695 (1999).
71. Santos, G. M., Southon, J. R., Druffel-Rodriguez, K. C., Griffin, S. & Mazon, M. Magnesium perchlorate as an alternative water trap in AMS graphite sample preparation: a report on sample preparation at KCCAMS at the University of California, Irvine. *Radiocarbon* **46**, 165–173 (2004).
72. Stuiver, M. & Polach, H. A. Discussion reporting of ^{14}C data. *Radiocarbon* **19**, 355–363 (1977).
73. Bronk Ramsey, C. & Lee, S. Recent and planned developments of the program OxCal. *Radiocarbon* **55**, 720–730 (2013).
74. Reimer, P. J. et al. The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* **62**, 725–757 (2020).
75. Brielle, E. S. et al. Entwined African and Asian genetic roots of medieval peoples of the Swahili coast. *Nature* **615**, 866–873 (2023).
76. Fan, S. et al. African evolutionary history inferred from whole genome sequence data of 44 indigenous African populations. *Genome Biol.* **20**, 1–14 (2019).
77. Fregel, R. et al. Ancient genomes from north Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. *Proc. Natl Acad. Sci. USA* **115**, 6774–6779 (2018).
78. Fu, Q. et al. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014).
79. Gallego-Llorente, M. et al. Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science* **350**, 820–822 (2015).
80. Jones, E. R. et al. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).
81. Kılınç, G. M. et al. The demographic development of the first farmers in Anatolia. *Curr. Biol.* **26**, 2659–2666 (2016).
82. Lipson, M. et al. Ancient West African foragers in the context of African population history. *Nature* **577**, 665–670 (2020).
83. Lipson, M. et al. Ancient DNA and deep population structure in sub-Saharan African foragers. *Nature* **603**, 290–296 (2022).
84. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
85. Prendergast, M. E. et al. Ancient DNA reveals a multistep spread of the first herders into sub-Saharan Africa. *Science* **365**, eaaw6275 (2019).
86. Prüfer, K. et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
87. Raghavan, M. et al. Upper palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
88. Schlebusch, C. M. et al. Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science* **358**, 652–655 (2017).
89. Schuenemann, V. J. et al. Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. *Nat. Commun.* **8**, 1–11 (2017).
90. Sirak, K. A. et al. Social stratification without genetic differentiation at the site of Kulubnarti in Christian Period Nubia. *Nat. Commun.* **12**, 7283 (2021).
91. Skoglund, P. et al. Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104–108 (2015).
92. Skoglund, P. et al. Reconstructing prehistoric African population structure. *Cell* **171**, 59–71 (2017).
93. Biagini, S. A. et al. People from Ibiza: an unexpected isolate in the Western Mediterranean. *Eur. J. Hum. Genet.* **27**, 941–951 (2019).
94. Broushaki, F. et al. Early Neolithic genomes from the eastern Fertile Crescent. *Science* **353**, 99–503 (2016).
95. Changmai, P. et al. Indian genetic heritage in Southeast Asian populations. *PLoS Genet.* **18**, e1010036 (2022).
96. Flegontov, P. et al. Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature* **570**, 236–240 (2019).
97. Jeong, C. et al. The genetic history of admixture across inner Eurasia. *Nat. Ecol. Evol.* **3**, 966–976 (2019).
98. Lipson, M. et al. Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science* **361**, 92–95 (2018).
99. Liu, D. et al. Extensive ethnolinguistic diversity in Vietnam reflects multiple sources of genetic diversity. *Mol. Biol. Evol.* **37**, 2503–2519 (2020).
100. López, S. et al. Evidence of the interplay of genetics and culture in Ethiopia. *Nat. Commun.* **12**, 3581 (2021).
101. Nakatsuka, N. et al. The promise of discovering population-specific disease-associated genes in South Asia. *Nat. Genet.* **49**, 1403–1407 (2017).
102. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
103. Pickrell, J. K. et al. The genetic prehistory of southern Africa. *Nat. Commun.* **3**, 1143 (2012).
104. Qin, P. & Stoneking, M. Denisovan ancestry in east Eurasian and Native American populations. *Mol. Biol. Evol.* **32**, 2665–2674 (2015).
105. Skoglund, P. et al. Genomic insights into the peopling of the Southwest Pacific. *Nature* **538**, 510–513 (2016).
106. Wang, C.-C. et al. Genomic insights into the formation of human populations in East Asia. *Nature* **591**, 413–419 (2021).
107. Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**, W58–W63 (2016).
108. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
109. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
110. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
111. Consortium, T. G. P. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

Acknowledgements

We thank N. Adamski, R. Bernardos and Z. Zhang for contributions to data generation, processing and curation. We acknowledge the help and support we received from the Soqatra Heritage Project team, including A. E. A. S. Al-Rumali, M. T. H. Ali, A. S. S. A. Al-Ameri and S. M. O. Sulaiman. We thank N. Salem for translation of the manuscript title and abstract to Arabic. The work of J. Jansen van Rensburg in Soqatra was supported by the National Geographic Society (NGS-61123R-19). D. Reich is an Investigator of the Howard Hughes Medical Institute (HHMI), and the ancient DNA laboratory work and analyses were also supported by National Institutes of Health grant HG012287, by John Templeton Foundation grant 61220, by the Allen Discovery

Center program, which is a Paul G. Allen Frontiers Group advised program of the Paul G. Allen Family Foundation, and by a gift from J.-F. Clin. This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 licence to the public and a sublicensable licence to HHMI in their research articles. Pursuant to those licences, the author-accepted manuscript of this article can be made freely available under a CC BY 4.0 licence immediately upon publication.

Author contributions

K.S., J.J.V.R. and D.R. designed the research. J.J.V.R., A.S.A.A.-O., E.M.A.S., A.M.S.H. and D.C.B. excavated and/or contributed to the acquisition and description of the osteological and archaeological data. K.S., E.B., B.C., I.L., D.R. and H.R. carried out population genetics analyses. M.M., S.M., A.M. and A.K. contributed to data generation, processing and curation. K.C., E.C., A.M.L., J.N.W. and F.Z. performed laboratory work under the supervision of N.R. K.S. and J.J.V.R wrote and revised the manuscript with input from all co-authors.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at

<https://doi.org/10.1038/s41559-024-02322-x>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41559-024-02322-x>.

Correspondence and requests for materials should be addressed to Kendra Sirak or Julian Jansen Van Rensburg.

Peer review information *Nature Ecology & Evolution* thanks Diyendo Massilani, Miranda Morris and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

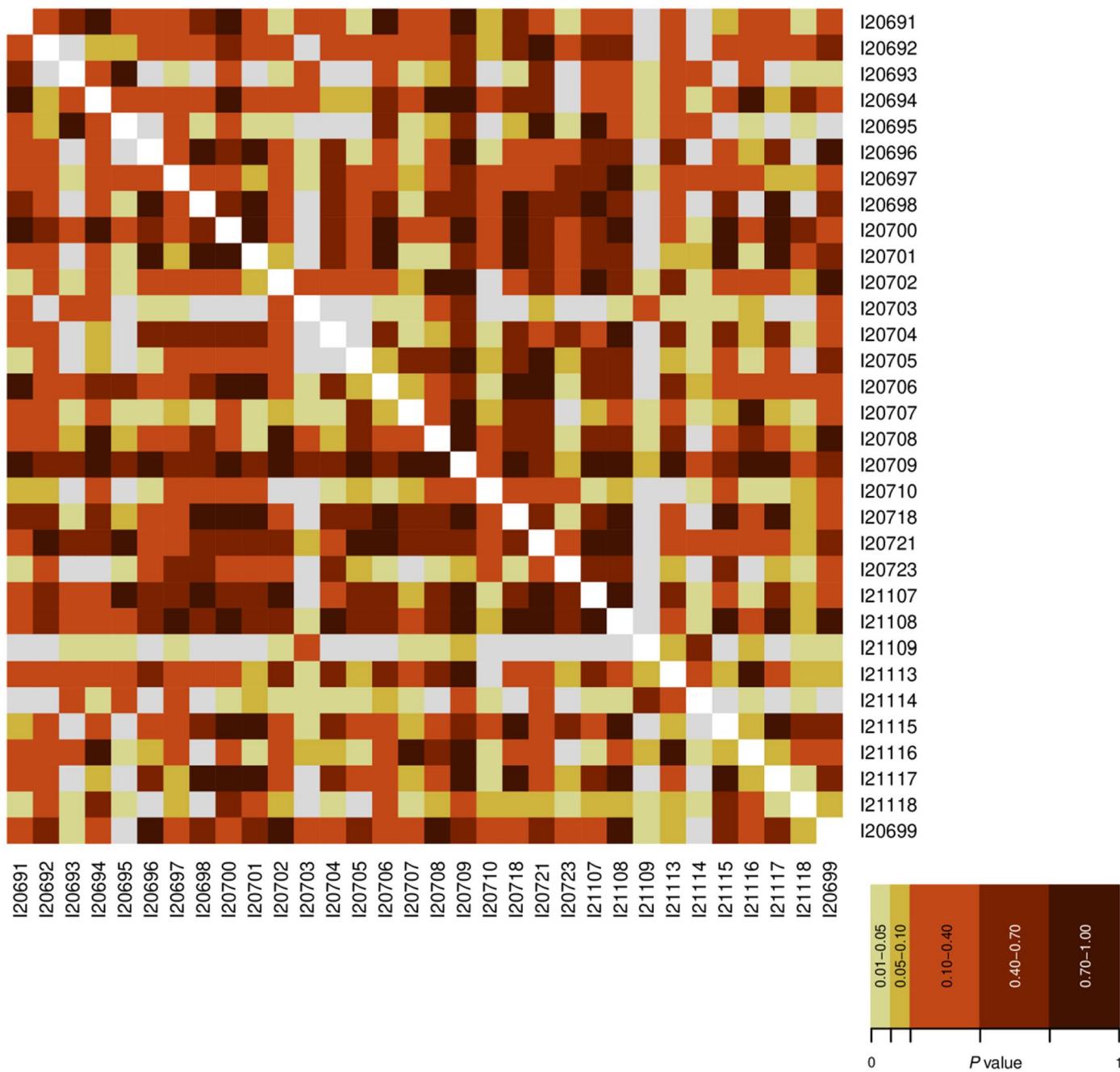
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2024

Kendra Sirak  ^{1,2}✉, **Julian Jansen Van Rensburg**  ³✉, **Esther Brielle**  ^{1,2}, **Bowen Chen**  ¹, **Iosif Lazaridis**  ^{1,2}, **Harald Ringbauer**  ^{1,4}, **Matthew Mah**  ^{1,2,5,6}, **Swapan Mallick** ^{1,2,5,6}, **Adam Micco** ^{2,5}, **Nadin Rohland** ^{2,6}, **Kimberly Callan** ^{2,5}, **Elizabeth Curtis** ^{2,5}, **Aisling Kearns** ^{2,5}, **Ann Marie Lawson** ^{2,5}, **J. Noah Workman**  ^{2,5}, **Fatma Zalzala**  ^{2,5}, **Ahmed Saeed Ahmed Al-Orqbi** ⁷, **Esmail Mohammed Ahmed Salem** ⁸, **Ali Mohammed Salem Hasan** ⁹, **Daniel Charles Britton** ¹⁰ & **David Reich**  ^{1,2,5,6}

¹Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA. ²Department of Genetics, Harvard Medical School, Boston, MA, USA. ³Chronicle Heritage, Phoenix, AZ, USA. ⁴Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. ⁵Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA. ⁶Broad Institute of Harvard and MIT, Cambridge, MA, USA. ⁷General Organisation for Antiquities and Museums (GOAM), Hadiboh, Republic of Yemen. ⁸Arab Regional Centre for World Heritage, Manama, Kingdom of Bahrain. ⁹Environmental Protection Agency on Socotra (EPA), Hadiboh, Republic of Yemen. ¹⁰Independent Researcher, London, UK. ✉e-mail: kendra_sirak@fas.harvard.edu; jrensburg@chronicleheritage.com



Extended Data Fig. 1 | Heatmap of pairwise *qpWave* tests for 32 medieval Soqotri individuals with sufficient SNP coverage. Data are in Supplementary Data 4 and details are in Supplementary Note 4. We interpret pairwise model

p-values > 0.01 (colored cells) to be consistent with the two individuals forming a clade relative to the reference population set, and *p*-values < 0.01 (grey cells) to be inconsistent.

Corresponding author(s): Kendra Sirak

Last updated by author(s): December 4, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection BWA v0.7.15-r1140, Picard MarkDuplicates, contamMix v.1.0-12, OxCal v.4.4

Data analysis ANGSD v. 0.921-3-g40ac3d6, HaploGrep2, hapROH, ADMIXTURE v1.3.0, PLINK1.9, smartpca v.18200, ADMIXTOOLS (qpWave v.1540, qpAdm v.1770, qpDstat v.994, qp3Pop v.670)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The aligned sequences will be made available through the European Nucleotide Archive upon manuscript acceptance under accession number PRJEB66485 (accession number provided in the manuscript). Genotype data will be made available at <https://reich.hms.harvard.edu/datasets>. Any other relevant data will be made available by the corresponding author upon reasonable request.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	We report molecular sex for the ancient Soqotri individuals (i.e., we report if an individual was XX, female, or XY, male). This allowed us to examine sex-specific patterns of burial on Soqatra.
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	Permission for this study was provided by the Soqotri Governorate and General Organization of Antiquities and Museums (GOAM) with additional support from the support from the Environmental Protection Agency on Soqatra (EPA) as well as the Soqatra Heritage Project (SHP) team. All archaeological sites studied as part of this project were selected in partnership with the local GOAM representative and village leaders.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Population genetics analyses were performed on DNA data generated from ancient human skeletal material. Population genetic statistics, primarily testing genetic relationships by measuring allele-sharing patterns across populations were computed using genome-wide SNP genotypes.
Research sample	We report genome-wide information for 39 individuals who lived during the Medieval and Modern Periods on the island of Soqatra.
Sampling strategy	We collected 50 skeletal samples from cave burials at six locations across the island of Soqatra. We obtained working genome-level data from 39 individuals. We targeted approximately 1.2M SNPs which effectively covers almost all independent loci (due to linkage disequilibrium) and provides good power in analyses of population history.
Data collection	DNA from the ancient remains was collected, sequenced, and processed to provide SNP genotype calls.
Timing and spatial scale	Ancient individuals were sampled from six locations across the island of Soqatra. Ancient individuals lived between the 7th-15th centuries CE with one individual dating to the Modern Period (1650-1950CE).
Data exclusions	9 skeletal samples did not yield working data as assessed by pre-established ancient DNA quality criteria. 2 additional samples were found to be duplicates of other samples in our dataset.
Reproducibility	All attempts to reproduce were successful.
Randomization	Samples were group based on statistical analysis using qpWave..
Blinding	Analyses were performed either for all individuals separately or as a single group.

Did the study involve field work? Yes No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology and archaeology
<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	Dual use research of concern
<input checked="" type="checkbox"/>	Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging