

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

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

<sup>1</sup>Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA.

<sup>2</sup>Department of Genetics, Harvard Medical School, Boston, MA 02115, USA.

<sup>3</sup>Chronicle Heritage, Phoenix, AZ 85004.

<sup>4</sup>Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig.

<sup>5</sup>Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, 02115, USA.

<sup>6</sup>Broad Institute of Harvard and MIT, Cambridge, MA, 02142, USA.

<sup>7</sup>General Organisation for Antiquities and Museums (GOAM), Hadiboh, Soqotra Island, Republic of Yemen.

<sup>8</sup>Arab Regional Centre for World Heritage, Manama, Kingdom of Bahrain.

<sup>9</sup>Environmental Protection Agency on Soqotra (EPA), Hadiboh, Soqotra Island, Republic of Yemen

<sup>10</sup>Independent Researcher, UK.

### Abstract

Soqotra, an island situated at the mouth of the Gulf of Aden in the northwest Indian Ocean between African 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

---

This work is licensed under a Creative Commons Attribution 4.0 International License, which allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use.

kendra\_sirak@fas.harvard.edu, jrensburg@chronicleheritage.com.

#### AUTHOR CONTRIBUTIONS STATEMENT

K.S., J.J.V.R., and D.R. designed the research. J.J.V.R., A.S.A.A., 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.

#### CODE AVAILABILITY

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

#### COMPETING INTEREST STATEMENT

The authors declare no competing interests.

about Soqotri history derives from writings of foreign travelers who provided little detail about local people, and the geographic origins and genetic affinities of early Soqotri people has not yet been investigated directly. We report genome-wide data from 39 individuals who lived between ~650–1750 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 modeled as deriving ~86% of their ancestry from a population like 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 like 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 harbored a small population that exhibited qualitatively different marriage practices from modern Soqotri, with first-cousin unions occurring significantly less frequently than today.

---

## INTRODUCTION

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 kilometers south of the Arabian Peninsula and 250 kilometers northwest of the Horn of Africa. Soqotra, the largest island, is renowned for endemic biodiversity and rich resources which 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 Soqotra's past is now also increasing through archaeological research (Jansen van Rensburg et al., 2022).

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 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 2<sup>nd</sup> century BCE (Strauch, 2012) 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 (Naumkin, 1993; Weeks et al., 2002). The earliest account of people on Soqotra is in the *Periplus Maris Erythraei*, written in the mid-1<sup>st</sup> 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 merchants (Casson, 1989), consistent with the hypothesis that the early inhabitants of Soqatra were South Arabian tribespeople who harvested frankincense during the height of the Hadramawt Kingdom's power (Doe, 1992; Naumkin, 1993). Furthermore, the Soqotri language is closest to the ḥerēt ('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 Arabia (Morris et al., 2019).

Soqatra also played an important role in ancient long-distance trade networks and there is extensive archaeological evidence of foreign traders on Soqatra 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 Soqatra (Casson, 1989), while Cosmas Indicopleustes, a Greek traveler of the 6<sup>th</sup> century CE, reported that Christian missionaries from Persia settled on the island in the mid-1<sup>st</sup> millennium CE (Evers, 2014). 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 Soqatra also attest to the presence of Indian traders at this time (Strauch, 2012). Inscriptions from Hoq cave record traders and travelers from Roman Egypt, Palmyra, Axum, Western India, Bactria, and Gandhara (an ancient civilization centered in northwest Pakistan) (Evers, 2014) between the 2<sup>nd</sup> century BCE and 6<sup>th</sup> century CE, drawn there by the supply of aloes, incense, and Dragon's Blood (Indian cinnabar) (Strauch, 2012). 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 if 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 Soqatra for DNA (Supplementary Notes 2 and 3). A *tafone* ("mesenaa'h" in Soqotri) is a geological formation resembling a small cave which 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 (Bent, 1900; Doe, 1970, 1992; Jansen van Rensburg, 2018; Naumkin & Sedov, 1993; Naumkin, 1993; Shinnie, 1960; Weeks et al., 2002). We analyzed a tooth from each individual in most cases, and the DNA-rich petrous bone when possible (Pinhasi et al., 2015). Cortical bone from long bone shafts was analyzed 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, Supplementary Data 1, 2). The individuals we report here lived as early as 641–668 calCE (1380±15 uncalibrated years before present, BP) and mostly

during the Medieval period (7th-15th centuries CE), with one dating to the Modern Period (1660–1950 calCE (195±15 BP)) (Supplementary Data 3).

We merged the new ancient DNA data with previously published ancient data from Africans and 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 (Lazaridis et al., 2014; Mallick et al., 2016; Vyas et al., 2017) 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; 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 calCE) is shifted furthest away from the main cluster. We use pairwise *qpWave* to evaluate whether every pair of individuals can be modeled 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; 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\_Soqotri’.

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, 2C). The Soqotri additionally harbor 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\_Soqotri and present-day people from the Arabian Peninsula (‘Arabians’) or Bedouins (‘Arabian-related’; Supplementary Note 4) we carried out ancestry modeling 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 admixture (Hellenthal et al., 2014)).

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 (Fedele, 2010; McCorriston & Martin, 2010; Uerpmann et al., 2010). 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 data (Almarri et al., 2021).

To test the hypothesis that people in Arabia may harbor 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 Eurasians (Lazaridis, Alpaslan-Roodenberg, et al., 2022a), 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-modeled using either 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 (Lazaridis, Alpaslan-Roodenberg, et al., 2022a, 2022b; Lazaridis et al., 2016); 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 by either 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; Table S1 in Supplementary Note 6; 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 modeled 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 Hadramawt (Naumkin, 1993), and is also consistent with studies of uniparental markers which point to similarities between the Soqotri and non-Arabic speaking groups living in the southern Arabian Peninsula (erný et al., 2011). In 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 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 (Agranat-Tamir et al., 2020; Harney et al., 2018). 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, 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 1<sup>st</sup> millennium CE Eastern Mediterranean traders as a source (Lazaridis, Alpaslan-Roodenberg, et al., 2022c), and also found that fitting models again required additional Iranian-related ancestry (Supplementary Note 6, 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 inscriptional evidence in Hoq cave and imagery at Dahaisi cave (Jansen van Rensburg, 2018) – 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; Supplementary Note 6) and confirm that there is no genome-wide signal of African-related ancestry that cannot be completely represented by Natufians, a group which has affinity to ~15,000-year-old Iberomaurusian hunter-gatherers from Morocco (van de Loosdrecht et al., 2018).

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 Soqotra (Soares et al., 2012); 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 people (Berný et al., 2016).

### 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 (like Mbuti) to isolate the non-African part of the shared genetic drift and avoid the confounding effects of Holocene era African-related gene flow (Fig. S6 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; Supplementary Data 8). D-statistics corroborate

this strong relationship:  $D(\text{Outgroup}, \text{Yemen\_Soqotri}; X, \text{Yemeni\_Desert})$  and  $D(\text{Outgroup}, \text{Yemeni\_Desert}; \text{Yemen\_Soqotri}, X)$  show that Yemen\_Soqotri shares more alleles with Yemeni\_Desert than with any other population ( $|Z| > 8.5$  and  $7.2$ , respectively), while  $D(\text{Outgroup}, X; \text{Yemen\_Soqotri}, \text{Yemeni\_Desert})$  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 Soqotri genomes are needed to determine if and to what extent these ancestries are found on Soqotra today.

Approximately 86% of medieval Soqotri ancestry can be modeled 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 (modeled using a group with minimal Iranian-related ancestry like Pulliyar) (Supplementary Note 6). While this ancestry can be completely modeled 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 (Narasimhan et al., 2019; Shinde et al., 2019), 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 individuals (Derenko et al., 2013). The high frequency of R2c in medieval Soqotra is distinct from the present-day mtDNA landscape, where most R haplotypes are part of the R0a clade hypothesized to have arisen in Arabia (Al-Abri et al., 2012; erný et al., 2009; Gandini et al., 2016).

### 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 Naumkin, 1993), with a preference for patrilineal orthocousin marriage (Naumkin, 1993). To determine if mate preference was similar in the past, we examined runs of homozygosity (ROH), segments of the genome where the genetic material inherited from both parents is identical, in 14 medieval Soqotri individuals of sufficient coverage ( $>0.3x$  coverage of targeted autosomal SNPs) (Supplementary Data 10; Supplementary Note 8) (Ringbauer et al., 2021). Long ROH segments ( $>20$  centimorgans, cM) are indicative of parental relatedness within a few generations, while shorter ROH (4–8cM) indicate relatedness deeper in the pedigree, reflecting a limited mating pool (Ceballos et al., 2018; Ringbauer et al., 2021).

Over half of the medieval Soqotri have at least one long ROH, which have an appreciable chance of arising when parents are related at the level of fourth cousins or closer, indicating that many unions among medieval Soqotri were among people who belonged to the same



extended family (see Figure S16 of Ringbauer et al., 2021). ROH data also provide evidence for an absence of close kin unions, defined here as first cousins or closer. I20705 from Zaflah, who has the largest amount of ROH<sub>>20cM</sub> in our dataset (80.3cM ROH<sub>>20cM</sub>, with a max length of 34.3cM) is at the lower end of the range expected for the offspring of first cousins (26<sup>th</sup> percentile from the perspective of summed ROH<sub>>20cM</sub>; 29<sup>th</sup> percentile from the perspective of the longest segment; Supplementary Data 15). That the individual with the most ROH<sub>>20cM</sub> 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<sub>>20cM</sub> 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<sub>>20cM</sub> in I20705 is in the 95<sup>th</sup> percentile expected for second cousin unions (81<sup>st</sup> percentile from the perspective of the largest single segment), which is unsurprising in a scenario which 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 dominate (Cadenas et al., 2008). This is consistent with the differences in the population history between Soqotra and most of the Arabian Peninsula that we identify through deep ancestry modeling (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 (e.g., Ariano et al., 2022) (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 (Bent, 1900; Doe, 1970, 1992; Naumkin, 1993; Shinnie, 1960; Weeks et al., 2002; Wellsted, 1840). 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 members (Luciani et al. 2018).

Ancient DNA provides 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 individuals (Olalde et al., 2019) as well as 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-degree (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; 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 shines 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 huntergatherers, a signal also documented in present-day groups in the Hadramawt; in contrast, other present-day Arabian groups can be modeled 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 *Br hm*, Ethiopian Ge'ez, Roman Egyptian, and Palmyrene inscriptions within Hoq cave dating between the 2<sup>nd</sup> century BCE and the 6<sup>th</sup> century CE, the Soqotri gene pool can be ~86% modeled 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 subcontinent 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 *Br hm* script and in Sanskrit or vernacularized Sanskrit dated to the 2<sup>nd</sup>-4<sup>th</sup> centuries CE (Strauch, 2012), while text from the *Periplus* records India's provision

of female slaves to Soqotra (Dridi, 2002; Dridi & Gorea, 2003; Robin & Gorea, 2002). Both the *Periplus* and epigraphic evidence specifically provide a record of traders coming from Barygaza (present-day Broach) and H thab in Gujarat on the northwest coast of India; epigraphic evidence also suggests Indians traveling to Soqotra from the Malabar coast of southwest India (Strauch, 2012; Strauch & Bukharin, 2004), while a single text is written in the Kharosth script exclusively used in the northwest of India (Strauch, 2016). The presence of Indians on Soqotra is continuously mentioned from the 10<sup>th</sup>-13<sup>th</sup> centuries, although they were increasingly portrayed as pirates who at least had some children with the coastal population (Al-Mujwir & ibn Ya'qub, 2008).

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 rates (Ringbauer et al., 2021). 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 7<sup>th</sup>-8<sup>th</sup> centuries (Bittles & Hamamy, 2010; Korotayev, 2000; Stern, 1939) 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 17<sup>th</sup> century by the Carmelite P. Vincenzo who mentioned that each family on Soqotra had a cave in which they placed their deceased (Yule & Cordier, 1993). 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 seven 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 last two millennia. Long before ancient DNA analysis began, we established a relationship with multiple local people, specifically the Soqatra Governorate, the Director of the General Organization of Antiquities and Museums (GOAM) Soqatra, local heritage practitioners from the Soqatra Heritage Project (SHP), 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 SHP 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 ~40mg of powder from bone and tooth samples, specifically targeting skeletal elements known to be especially DNA-rich (Hansen et al., 2017; Pinhasi et al., 2015; Pinhasi et al., 2019). 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-ups (Rohland et al., 2018). 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) (Rohland et al., 2015, 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 (Fu et al., 2015; Fu, Mittnik, et al., 2013; Haak et al., 2015; Mathieson et al., 2015). 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 exhibited no more than one mismatch between the forward and reverse base if base quality was ≥ 20, or three mismatches if base quality was <20 prior to 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) (Behar et al., 2012) and the human reference genome version hg19 using the *samse* command in BWA 0.7.15-r1140 (Li & Durbin, 2009) 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 fragments (Rohland et al., 2015), point estimates of mtDNA contamination below 5% (estimates made with contamMix v.1.0–12; Fu, Meyer, et al., 2013), 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(Jan 26 2018 13:16:35)) (Korneliusson et al., 2014) (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 (AMS) radiocarbon dates ( $^{14}\text{C}$ ) from the Pennsylvania State University (PSU) Radiocarbon Laboratory. We sonicated all of the bone samples in successive washes of American Chemical Society (ACS) grade methanol, acetone, and dichloromethane for 30 minutes each at room temperature, followed by three washes in Nanopure water to rinse in order to remove possible contaminants (conservants and adhesives). We extracted bone collagen and purified using a modified Longin method with ultrafiltration (>30 kDa gelatin) (Kennett et al., 2017). If collagen yields were low and amino acids were poorly preserved, we used a modified XAD process (XAD Amino Acids) (Lohse et al., 2014). 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 analyzer (ECS 4010). We evaluated sample quality by percentage of crude gelatin yield, percentage of C, percentage of N and C/N ratios before AMS  $^{14}\text{C}$  dating. C/N ratios for all directly radiocarbon samples except one (I20701) fell between 3.1 and 3.3, indicating excellent preservation (Van Klinken, 1999). 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  $\text{CO}_2$  was reduced to graphite at 550 °C using  $\text{H}_2$  and a Fe catalyst, and we drew off reaction water with  $\text{Mg}(\text{ClO}_4)_2$  (Santos et al., 2004). We pressed graphite samples into targets in aluminum boats and loaded them onto a target wheel and performed all  $^{14}\text{C}$  measurements using a modified National Electronics Corporation compact spectrometer with a 0.5 MV accelerator (NEC 1.5SDH-1). We corrected the  $^{14}\text{C}$  ages for mass-dependent

fractionation with measured  $\delta^{13}\text{C}$  values (Stuiver & Polach, 1977) and compared with samples of Pleistocene whale bone (>48,000  $^{14}\text{C}$  BP), late Holocene bison bone (~1,850  $^{14}\text{C}$  BP), late AD 1800s cow bone and OX-2 oxalic acid standards. We calibrated  $^{14}\text{C}$  ages using OxCal v.4.4 (Bronk Ramsey & Lee, 2013) and the IntCal20 northern hemisphere curve (Reimer et al., 2020).

### 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) (Fu et al., 2015; Fu, Mittnik, et al., 2013; Haak et al., 2015; Mathieson et al., 2015), as well as whole-genome sequence (WGS) data from ancient and present-day African and Near Eastern groups (Agranat-Tamir et al., 2020; Brielle et al., 2023; Fan et al., 2019; Fregel et al., 2018; Fu et al., 2014; Gallego-Llorente et al., 2015; Harney et al., 2018; Jones et al., 2015; Kılınç et al., 2016; Lazaridis, Alpaslan-Roodenberg, et al., 2022a, 2022b, 2022c; Lazaridis et al., 2016; Lipson et al., 2020; Lipson et al., 2022; Mallick et al., 2016; Meyer et al., 2012; Narasimhan et al., 2019; Prendergast et al., 2019; Prüfer et al., 2014; Raghavan et al., 2014; Schlebusch et al., 2017; Schuenemann et al., 2017; Sirak et al., 2021; Skoglund et al., 2015; Skoglund et al., 2017; van de Loosdrecht et al., 2018). We also merged the new data with Affymetrix Human Origins (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 (Biagini et al., 2019; Brielle et al., 2023; Broushaki et al., 2016; Changmai et al., 2022; Flegontov et al., 2019; Jeong et al., 2019; Lazaridis et al., 2016; Lazaridis et al., 2014; Lipson et al., 2018; Lipson et al., 2020; Liu et al., 2020; López et al., 2021; Nakatsuka et al., 2017; Patterson et al., 2012; Pickrell et al., 2012; Qin & Stoneking, 2015; Skoglund et al., 2016; Skoglund et al., 2017; Vyas et al., 2017; Wang et al., 2021). Whenever we needed to co-analyze 1240k data and HO array data, we restricted to SNPs also covered on the HO SNP array.

### Uniparental haplogroups.

We called mitochondrial DNA (mtDNA) haplogroups using HaploGrep2 (Weissensteiner et al., 2016), reported in Supplementary Data 2. We determined Y chromosome haplogroups using both targeted SNPs as well as 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.18200 (Patterson et al., 2006) and the parameters `lsqproject: YES`, `newshrink: YES`, `numoutlieriter: 2`, and `hiprec: YES`. We computed eigenvectors of the covariance matrix of SNPs from 1198 present-day individuals from 63 modern populations genotyped on the Affymetrix Human Origins (HO) SNP array. We

projected all other individuals (including ancient individuals) onto the first two principal components (PC1 and PC2). Full details of PCA setup are in Supplementary Note 4.

### **ADMIXTURE (v.1.3.0).**

We ran unsupervised ADMIXTURE (Alexander et al. 2009) using 2934 world-wide individuals either genotyped on the HO array or pulled down on the HO SNP set. We used *convert* to transform the dataset from PACKEDANCESTRYMAP format to the PACKEDPED format for pruning in PLINK2 (Chang et al. 2015) and pruned SNPs based on linkage disequilibrium (LD), 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 (Olalde et al., 2019) 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 (e.g., avuncular) and identify more distant relatives (e.g., 5<sup>th</sup>-7<sup>th</sup>-degree relatives). Results are in Supplementary Data 13, with discussion of genetic relationships in terms of burial groupings in 16 tafone in Supplementary Data 14. We use IBD analysis to show some additional more distant relationships amount eight individuals from Zaflah with sufficient coverage in Supplementary Data 12.

### **Analysis of consanguineous genetic segments.**

We identified runs of homozygosity (ROH) using the Python package hapROH (<https://test.pypi.org/project/hapROH/>) which uses 5008 genomes from the 1000 Genomes project haplotype panel (1000 Genomes Project Consortium, 2015) 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 centiMorgan (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 4–8cM, 8–12cM, 12–20cM, and >20cM 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 ADMIXTOOLS (Patterson et al., 2012) to test cladality between every pair of medieval Soqotri individuals in our analysis dataset. We used ‘allsnps: YES’ which calls *qpfstats* to calculate the relevant f-statistics and set Han.SDG as the base population (‘basepop: Han.SDG’) for all *qpfstats* calculations (<https://github.com/DReichLab/AdmixTools/blob/master/qpfs.pdf>). We set ‘inbreed: NO’ but also added the ‘inbreedlistname’ parameter so that any population with more than one individual would instead be analyzed as if we set ‘inbreed: YES’. Results are in Supplementary Data 4, with additional details in Supplementary Note 5.

**qpAdm (v.1770).**

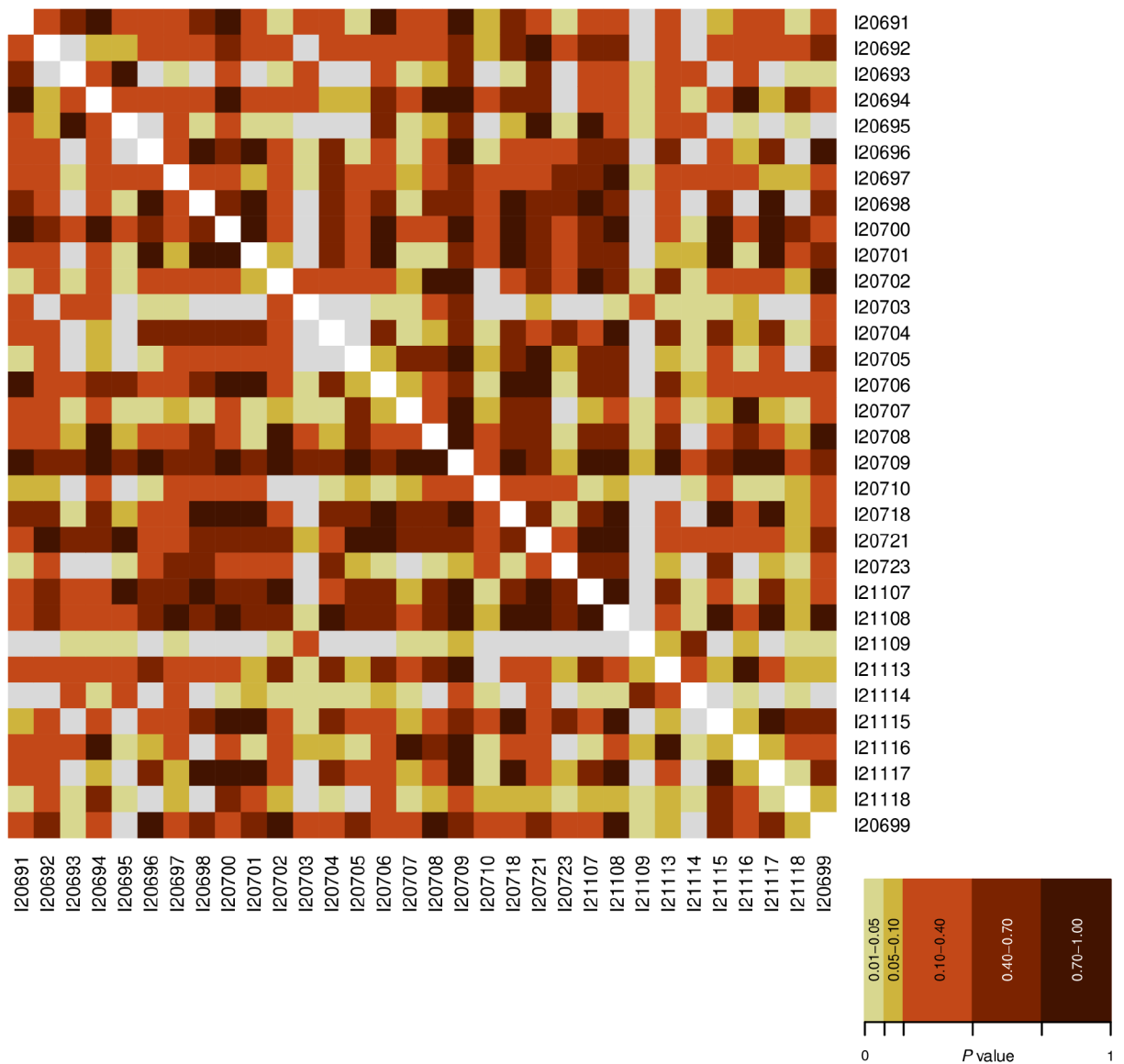
We used the *qpAdm* framework (Haak et al., 2015) to model ancestry and estimate ancestry proportions. *qpAdm* uses  $f_4$ -statistics to detect share drift between the population of interest (the ‘target’ population) and possible admixing source proxies relative to a set of differentially related outgroup populations (referred to as the ‘reference set’). It evaluates whether a population of interest can be plausibly modeled as descending from a common ancestor of one or more source populations and produces unbiased estimates of ancestry proportions from these sources (Haak et al., 2015). 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/qpfs.pdf>). We set ‘inbreed: NO’, but also added the ‘inbreedlistname’ parameter so any population with more than one individual would be analyzed 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 ADMIXTOOLS (Patterson et al., 2012) with ‘f4 mode: YES’ and ‘inbreed: YES.’ We computed ‘outgroup’  $f_3$ -statistics using *qp3Pop* (v.670) from ADMIXTOOLS (Patterson et al., 2012) with ‘inbreed: YES.’ Additional details about  $f_3$ -statistics as used in this work are in Supplementary Note 7.



## Extended Data



### Extended Data Figure 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 5. 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.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## 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 Ahmed Eissa Ahmed Saeed Al-Rumali, Mohammed Talee Hasan Ali, Ahdab Salem Saeed Awadh Al-Ameri, and Shaikhah Mubarak Omar Sulaiman. We thank Nada 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, 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 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

## 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.

## REFERENCES

- Agranat-Tamir L, Waldman S, Martin MAS, Gokhman D, Mishol N, Eshel T, Cheronet O, Rohland N, Mallick S, Adamski N, Lawson AM, Mah M, Michel M, Oppenheimer J, Stewardson K, Candilio F, Keating D, Gamarra B, Tzur S, ... Reich D (2020). The Genomic History of the Bronze Age Southern Levant. *Cell*, 181(5), 1146–1157. [PubMed: 32470400]
- Al-Abri A, Podgorná E, Rose JI, Pereira L, Mulligan CJ, Silva NM, Bayoumi R, Soares P, & erný V (2012). Pleistocene-Holocene boundary in Southern Aabia from the perspective of human mtDNA variation. *American Journal of Physical Anthropology*, 149, 21–28.
- Al-Mujwir I, & ibn Ya'qub Y (2008). A traveller in thirteenth-century Arabia: Ibn al-Mujwir's *Trikh almustabshir*. Ashgate.
- Alexander DH, Novembre J, & Lange K (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. [PubMed: 19648217]
- Almarri MA, Haber M, Lootah RA, Hallast P, Al Turki S, Martin HC, Xue Y, & Tyler-Smith C (2021). The genomic history of the Middle East. *Cell*, 184(18), 4612–4625. e4614. [PubMed: 34352227]
- Ariano B, Mattiangeli V, Breslin EM, Parkinson EW, McLaughlin TR, Thompson JE, Power RK, Stock JT, Mercieca-Spiteri B, Stoddart S, Malone C, Gopalakrishnan S, Cassidy LM, & Bradley DG (2022). Ancient Maltese genomes and the genetic geography of Neolithic Europe. *Current Biology*, 32, 2668–2680.e6. [PubMed: 35588742]
- Behar Doron M., van Oven M, Rosset S, Metspalu M, Loogväli E-L, Silva Nuno M., Kivisild T, Torroni A, & Villems R (2012). A “Copernican” Reassessment of the Human Mitochondrial DNA Tree from its Root. *The American Journal of Human Genetics*, 90(4), 675–684. [PubMed: 22482806]
- Bent JT (1900). *Southern Arabia*. Smith, Elder and Co.
- Biagini SA, Solé-Morata N, Matisoo-Smith E, Zalloua P, Comas D, & Calafell F (2019). People from Ibiza: an unexpected isolate in the Western Mediterranean. *European Journal of Human Genetics*, 27(6), 941–951. [PubMed: 30765884]
- Bittles AH, & Hamamy HA (2010). Endogamy and consanguineous marriage in Arab populations. In *Genetic Disorders among Arab Populations*, 85–108.
- Brielle ES, Fleisher J, Wynne-Jones S, Sirak K, Broomandkhoshbacht N, Callan K, Curtis E, Iliev L, Lawson AM, Oppenheimer J, Qiu L, Stewardson K, Workman JN, Zalzal F, Ayodo G, Gidna AO, Kabiru A, Kwekason A, Mabulla AZP, ... Kusimba CM (2023). Entwined African and Asian

- genetic roots of medieval peoples of the Swahili coast. *Nature*, 615(7954), 866–873. [PubMed: 36991187]
- Bronk Ramsey C, & Lee S (2013). Recent and planned developments of the program OxCal. *Radiocarbon*, 55(2), 720–730.
- Broushaki F, Thomas MG, Link V, López S, van Dorp L, Kirsanow K, Hofmanová Z, Diekmann Y, Cassidy LM, & Díez-del-Molino D (2016). Early Neolithic genomes from the eastern Fertile Crescent. *Science*, 353, 99–503.
- Cadenas AM, Zhitovovsky LA, Cavalli-Sforza LL, Underhill PA, & Herrera RJ (2008). Y-chromosome diversity characterizes the Gulf of Oman. *European Journal of Human Genetics*, 16(3), 374–386. [PubMed: 17928816]
- Casson L (1989). *The Periplus Maris Erythraei: Text with introduction, translation, and commentary*. Princeton University Press.
- Ceballos FC, Joshi PK, Clark DW, Ramsay M, & Wilson JF (2018). Runs of homozygosity: windows into population history and trait architecture. *Nature Review Genetics*, 19(4), 220–234.
- erný V, Pereira L, Kujanová M, Vašíková A, Hájek M, Morris M, & Mulligan CJ (2009). Out of Arabia—The settlement of Island Soqatra as revealed by mitochondrial and Y chromosome genetic diversity. *American Journal of Physical Anthropology*, 138(4), 439–447. [PubMed: 19012329]
- erný V, Mulligan CJ, Fernandes V, Silva NM, Alshamali F, Non A, Harich N, Cherni L, El Gaaied ABA, Al-Meerri A, & Pereira L (2011). Internal Diversification of Mitochondrial Haplogroup R0a Reveals Post-Last Glacial Maximum Demographic Expansions in South Arabia. *Molecular Biology and Evolution*, 28(1), 1–78. [PubMed: 20720154]
- erný V, Vašíková M, Poloni ES, Al-Meerri A, & Mulligan C (2016). Comprehensive View of the Population History of Arabia as Inferred by mtDNA Variation. *American Journal of Physical Anthropology*, 159, 607–616. [PubMed: 26691891]
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, & Lee JJ (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4(1), s13742–015.
- Changmai P, Jaisamut K, Kampuansai J, Kutanan W, Altunbaş NE, Flegontova O, Inta A, Yüncü E, Boonthai W, & Pamjav H (2022). Indian genetic heritage in Southeast Asian populations. *PLoS genetics*, 18(2), e1010036. [PubMed: 35176016]
- Consortium TGP (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74. [PubMed: 26432245]
- Derenko M, Malyarchuk B, Bahmanimehr A, Denisova G, Perkova M, Farjadian S, & Yepiskoposyan L (2013). Complete Mitochondrial DNA Diversity in Iranians. *Plos One*, 8(11), e80673. [PubMed: 24244704]
- Doe B (1992). *Socotra: Island of tranquillity*. Immel Publishing.
- Doe B (1970). *Socotra; an Archaeological Reconnaissance in 1967*. Coconut Grove, Fla.: Field Research Projects.
- Dridi H (2002). Indiens et proche-orientaux dans une grotte de Suqutır (Yemen). *Journal Asiatique*, 290(2), 565–610.
- Dridi H, & Gorea M (2003). Au IIIe siècle de notre ère: le voyage d'Agbar à Suqutra. *Archeologia*, 396, 48–57.
- Evers K (2014). Cave of revelations: Indian Ocean trade in light of the Socotran graffiti. *Journal of Indian Ocean Archaeology*, 10(11), 19–37.
- Fan S, Kelly DE, Beltrame MH, Hansen MEB, Mallick S, Ranciaro A, Hirbo J, Thompson S, Beggs W, Nyambo T, Omar SA, Meskel DW, Belay G, Froment A, Patterson N, Reich D, & Tishkoff SA (2019). African evolutionary history inferred from whole genome sequence data of 44 indigenous African populations. *Genome Biology*, 20(1), 1–14. [PubMed: 30606230]
- Fedele FG (2010). Early Holocene in the Highlands: Data on the Peopling of the Eastern Yemen Plateau, with a Note on the Pleistocene Evidence. In Petraglia MD & Rose JI (Eds.), *The evolution of human populations in Arabia: Paleoenvironments, prehistory and genetics* (pp. 215–236). Springer Netherlands.

- Flegontov P, Altınışık NE, Changmai P, Rohland N, Mallick S, Adamski N, Bolnick DA, Broomandkoshbacht N, Candilio F, ... & Schiffels S (2019). Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature*, 570(7760), 236–240. [PubMed: 31168094]
- Fregel R, Méndez FL, Bokbot Y, Martín-Socas D, Camalich-Massieu MD, Santana J, Morales J, Ávila-Arcos MC, Underhill PA, Shapiro B, Wojcik G, Rasmussen M, Soares AER, Kapp J, Sockell A, Rodríguez-Santos FJ, Mikdad A, Trujillo-Mederos A, & Bustamante CD (2018). Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. *Proceedings of the National Academy of Sciences*, 115(26), 6774–6779.
- Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, Patterson N, Rohland N, Lazaridis I, ... & Pääbo S (2015). An early modern human from Romania with a recent Neanderthal ancestor. *Nature*, 524(7564), 216–219. [PubMed: 26098372]
- Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, Johnson PLF, Aximu-Petri A, Prüfer K, de Filippo C, Meyer M, Zwyns N, Salazar-Garcia DC, Kuzmin YV, Keates SG, Kosintsev PA, Razhev DI, Richards MP, Peristov NV, ... Pääbo S (2014). Genome sequence of a 45,000-year-old modern human from western Siberia [Article]. *Nature*, 514(7523), 445–449. [PubMed: 25341783]
- Fu Q, Meyer M, Gao X, Stenzel U, Burbano HA, Kelso J, & Pääbo S (2013). DNA analysis of an early modern human from Tianyuan Cave, China. *Proceedings of the National Academy of Sciences*, 110(6), 2223–2227.
- Fu Q, Mittnik A, Johnson Philip L. F., Bos K, Lari M, Bollongino R, Sun C, Giemisch L, Schmitz R, Burger J, Ronchitelli Anna M., Martini F, Cremonesi Renata G., Svoboda J, Bauer P, Caramelli D, Castellano S, Reich D, Pääbo S, & Krause J (2013). A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. *Current Biology*, 23(7), 553–559. [PubMed: 23523248]
- Gallego-Llorente M, Jones ER, Eriksson A, Siska V, Arthur K, Arthur J, Curtis M, Stock JT, Coltorti M, & Pieruccini P (2015). Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science*, 350(6262), 820–822. [PubMed: 26449472]
- Gandini F, Achilli A, Pala M, Bodner M, Brandini S, Huber G, Egyed B, Ferretti L, Gómez-Carballa A, Salas A, Scozzari R, Cruciani F, Coppa A, Parson W, Semino O, Soares P, Torroni A, Richards MB, & Olivieri A (2016). Mapping human dispersals into the Horn of Africa from Arabian Ice Age refugia using mitogenomes. *Scientific Reports*, 6, 25472–25472. [PubMed: 27146119]
- Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K, Fu Q, Mittnik A, Bánffy E, Economou C, Francken M, Friederich S, Pena RG, Hallgren F, Khartanovich V, ... & Reich D (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, 522, 207–211. [PubMed: 25731166]
- Hansen HB, Damgaard PB, Margaryan A, Stenderup J, Lynnerup N, Willerslev E, & Allentoft ME (2017). Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum. *Plos One*, 12(1), e0170940. [PubMed: 28129388]
- Harney É, May H, Shalem D, Rohland N, Mallick S, Lazaridis I, Sarig R, Stewardson K, Nordenfelt S, Patterson N, Hershkovitz I, & Reich D (2018). Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation. *Nature Communications*, 9(3336), 1–11.
- Hellenthal G, Busby GB, Band G, Wilson JF, Capelli C, Falush D, & Myers S (2014). A genetic atlas of human admixture history. *Science*, 343(6172), 747–751. [PubMed: 24531965]
- Jansen van Rensburg J (2018). Rock Art of Soqatra, Yemen: A Forgotten Heritage Revisited. *Arts*, 7(4), 99.
- Jansen van Rensburg J, Forrest A, Ali MTH, Al-Ameri ASSA, Sulaiman SMO, Al-Orqabi ASA, Salem ISM, I-Rumeili AEAS, Hasan AMS, Esmail AAM, Wilkins O, & Carey B (2022). The Soqatra Heritage Project (2017 – 2020). *Bulletin of the Society of Arabian Studies*, 22, 26–27.
- Jeong C, Balanovsky O, Lukianova E, Kahbatkyzy N, Flegontov P, Zaporozhchenko V, Immel A, Wang C-C, Ixan O, ... & Krause J (2019). The genetic history of admixture across inner Eurasia. *Nature Ecology & Evolution*, 3(6), 966–976. [PubMed: 31036896]
- Jones ER, Gonzalez-Fortes G, Connell S, Siska V, Eriksson A, Martiniano R, McLaughlin RL, Llorente MG, Cassidy LM, ... & Bradley D (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nature Communications*, 6.

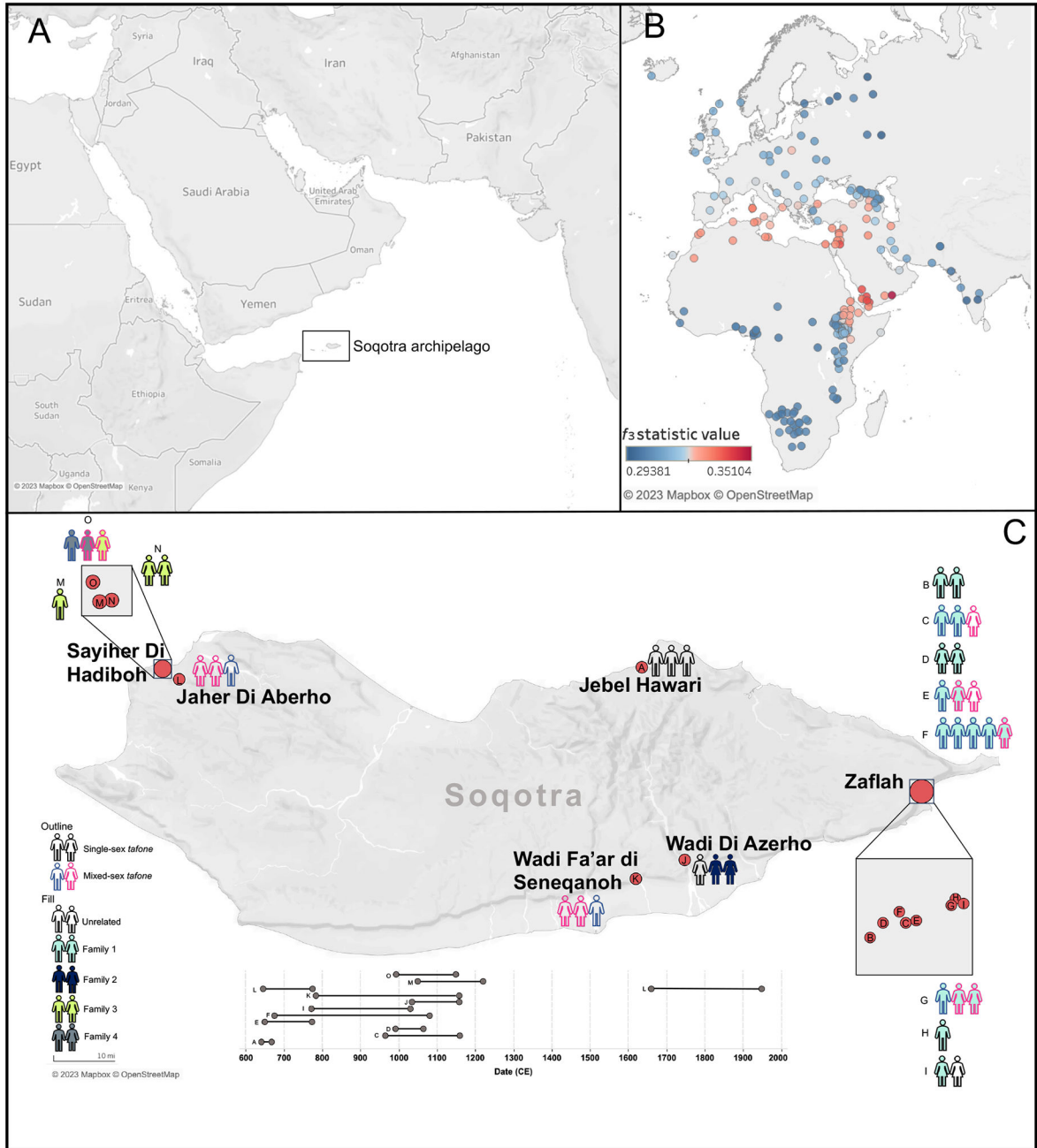
- Kennett DJ, Plog S, George RJ, Culleton BJ, Watson AS, Skoglund P, Rohland N, Mallick S, Stewardson K,... & Perry GH (2017). Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nature Communications*, 8.
- Kılınç GM, Omrak A, Özer F, Günther T, Büyükkarakaya AM, Bıçakçı E, Baird D, Dönerta HM, Ghalichi A,... & Götherström A (2016). The demographic development of the first farmers in Anatolia. *Current Biology*, 26(19), 2659–2666. [PubMed: 27498567]
- Korneliussen TS, Albrechtsen A, & Nielsen R (2014). ANGSD: analysis of next generation sequencing data. *BMC bioinformatics*, 15(1), 356. [PubMed: 25420514]
- Korotayev A (2000). Parallel-cousin (FBD) marriage, Islamization, and Arabization. *Ethnology* (39), 395–407.
- Lazaridis I, Alpaslan-Roodenberg S, Acar A, Açikkol A, Agelarakis A, Aghikyan L, Akyüz U, Andreeva D, Andrijašević G, Antonovi D, Armit I, Atmaca A, Avetisyan P, Aytek A, Bacvarov K, Badalyan R, Bakardzhiev S, Balen J, Bejko L,... & Reich D (2022a). The genetic history of the Southern Arc: A bridge between West Asia and Europe. *Science*, 377(6609), eabm4247.
- Lazaridis I, Alpaslan-Roodenberg S, Acar A, Açikkol A, Agelarakis A, Aghikyan L, Akyüz U, Andreeva D, Andrijašević G,... & Reich D (2022b). Ancient DNA from Mesopotamia suggests distinct pre-pottery and pottery Neolithic migrations into Anatolia. *Science*, 377(6609), 982–987. [PubMed: 36007054]
- Lazaridis I, Alpaslan-Roodenberg S, Acar A, Açikkol A, Agelarakis A, Aghikyan L, Akyüz U, Andreeva D, Andrijašević G, Antonovi D, Armit I, Atmaca A, Avetisyan P, Aytek A, Bacvarov K, Badalyan R, Bakardzhiev S, Balen J, Bejko L,... & Reich D (2022c). A genetic probe into the ancient and medieval history of Southern Europe and West Asia. *Science*, 377(6609), 940–951. [PubMed: 36007020]
- Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, Fernandes D, Novak M, Gamarra B, Sirak K, Connell S, Stewardson K, Harney E, Fu Q, Gonzalez-Fortes G, Jones ER, Roodenberg SA, Lengyel G, Bocquentin F,... & Reich D (2016). Genomic insights into the origin of farming in the ancient Near East. *Nature*, 536(7617), 419–424. [PubMed: 27459054]
- Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S,... & Krause J (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, 513(7518), 409–413. [PubMed: 25230663]
- Li H, & Durbin R (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. [PubMed: 19451168]
- Lipson M, Cheronet O, Mallick S, Rohland N, Oxenham M, Pietruszewski M, Pryce TO, Willis A, Matsumura H, Buckley H, Domett K, Hai NG, Hiep TH, Kyaw AA, Win TT, Pradier B, Broomandkshobacht N, Candilio F, Changmai P,... & Reich D (2018). Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science*, 361, 92–95. [PubMed: 29773666]
- Lipson M, Ribot I, Mallick S, Rohland N, Olalde I, Adamski N, Broomandkshobacht N, Lawson AM, López S, Oppenheimer J, Stewardson K, Asombang R. N. a., Bocherens H, Bradman N, Culleton BJ, Cornelissen E, Crevecoeur I, de Maret P, Fomine FLM,... & Reich D (2020). Ancient West African foragers in the context of African population history. *Nature*, 577, 665–670. [PubMed: 31969706]
- Lipson M, Sawchuk EA, Thompson JC, Oppenheimer J, Tryon CA, Ranhorn KL, de Luna KM, Sirak KA, Olalde I, Ambrose SH, Arthur JW, Arthur KJW, Ayodo G, Bertacchi A, Cerezo-Román JI, Culleton BJ, Curtis MC, Davis J, Gidna AO,... & Prendergast ME (2022). Ancient DNA and deep population structure in sub-Saharan African foragers. *Nature*, 603, 290–296. [PubMed: 35197631]
- Liu D, Duong NT, Ton ND, Van Phong N, Pakendorf B, Van Hai N, & Stoneking M (2020). Extensive Ethnolinguistic Diversity in Vietnam Reflects Multiple Sources of Genetic Diversity. *Molecular biology and evolution*, 37(9), 2503–2519. [PubMed: 32344428]
- Lohse JC, Madsen DB, Culleton BJ, & Kennett DJ (2014). Isotope paleoecology of episodic mid-to-late Holocene bison population expansions in the Southern Plains, USA. *Quaternary Science Reviews*, 102, 14–26.
- López S, Tarekegn A, Band G, van Dorp L, Bird N, Morris S, Oljira T, Mekonnen E, Bekele E, Blench R, Thomas MG, Bradman N, & Hellenthal G (2021). Evidence of the interplay of genetics and culture in Ethiopia. *Nature Communications*, 12(1), 3581.

- Luciani M, Binder M, & Alsaud AS (2018). Life and living conditions in north-west Arabia during the Bronze Age: first results from the bioarchaeological work at Qurayyah. *Proceedings of the Seminar for Arabian Studies*, 48, 185–201
- Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S,... & Reich D (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*, 538(7624), 201–206. [PubMed: 27654912]
- Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, Sirak K, Gamba C, Jones ER, Llamas B, Dryomov S, Pickrell J, Arsuaga JL, de Castro JMB, Carbonell E,... & Reich D (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, 528(7583), 499–503. [PubMed: 26595274]
- McCorrison J, & Martin L (2010). Southern Arabia's Early Pastoral Population History: Some Recent Evidence. In Petraglia MD & Rose JI (Eds.), *The evolution of human populations in Arabia: Paleoenvironments, prehistory and genetics* (pp. 237–250). Springer Netherlands.
- Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K,... & Pääbo S (2012). A high-coverage genome sequence from an archaic Denisovan individual. *Science*, 338(6104), 222–226. [PubMed: 22936568]
- Morris M, Watson JCE, & Eades D (2019). A comparative cultural glossary across the Modern South Arabian language family (Vol. *Journal of Semitic Studies supplement* 43). Oxford University Press.
- Nakatsuka N, Moorjani P, Rai N, Sarkar B, Tandon A, Patterson N, Bhavani GS, Girisha KM, Mustak MS, Srinivasan S, Kaushik A, Vahab SA, Jagadeesh SM, Satyamoorthy K, Singh L, Reich D, & Thangaraj K (2017). The promise of discovering population-specific disease-associated genes in South Asia. *Nature genetics*, 49(9), 1403–1407. [PubMed: 28714977]
- Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, Mallick S, Lazaridis I, Nakatsuka N, Olalde I, Lipson M, Kim AM, Olivieri LM, Coppa A, Vidale M, Mallory J, Moiseyev V, Kitov E, Monge J, Adamski N,... & Reich D (2019). The formation of human populations in South and Central Asia. *Science*, 365(6457), eaat7487.
- Naumkin VV, & Sedov AV (1993). Monuments of socotra. *Topoi. Orient-Occident*, 3(2), 569–623.
- Naumkin VVA (1993). *Island of the Phoenix: An ethnographic study of the people of Socotra*. Paul & Company Pub Consortium.
- Olalde I, Mallick S, Patterson N, Rohland N, Villalba-Mouco V, Silva M, Dulias K, Edwards CJ, Gandini F,... & Reich D (2019). The genomic history of the Iberian Peninsula over the past 8000 years. *Science*, 363(6432), 1230–1234. [PubMed: 30872528]
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, & Reich D (2012). Ancient admixture in human history. *Genetics*, 192(3), 1065–1093. [PubMed: 22960212]
- Patterson N, Price AL, & Reich D (2006). Population structure and eigenanalysis. *PLoS genetics*, 2(12), e190. [PubMed: 17194218]
- Pickrell JK, Patterson N, Barbieri C, Berthold F, Gerlach L, Güldemann T, Kure B, Mpoloka SW, Nakagawa H, & Naumann C (2012). The genetic prehistory of southern Africa. *Nature Communications*, 3, 1143.
- Pinhasi R, Fernandes D, Sirak K, Novak M, Connell S, Alpaslan-Roodenberg S, Gerritsen F, Moiseyev V, Gromov A, Raczky P, Anders A, Pietruszewsky M, Rollefson G, Jovanovic M, Trinhhoang H, Bar-Oz G, Oxenham M, Matsumura H, & Hofreiter M (2015). Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. *Plos One*, 10(6), e0129102. [PubMed: 26086078]
- Pinhasi R, Fernandes DM, Sirak K, & Cheronet O (2019). Isolating the human cochlea to generate bone powder for ancient DNA analysis. *Nature Protocols*, 14(4), 1194–1205. [PubMed: 30842617]
- Prendergast ME, Lipson M, Sawchuk EA, Olalde I, Ogola CA, Rohland N, Sirak KA, Adamski N, Bernardos R, Broomandkhoshbacht N, Callan K, Culleton BJ, Eccles L, Harper TK, Lawson AM, Mah M, Oppenheimer J, Stewardson K, Zalzal F,... & Reich D (2019). Ancient DNA reveals a multistep spread of the first herders into sub-Saharan Africa. *Science*, 365(6448), eaaw6275.
- Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH,... & Pääbo S (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, 505(7481), 43–49. [PubMed: 24352235]

- Qin P, & Stoneking M (2015). Denisovan ancestry in East Eurasian and native American populations. *Molecular biology and evolution*, 32(10), 2665–2674. [PubMed: 26104010]
- Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW Jr, Orlando L,... & Willerslev E (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*, 505, 87–91. [PubMed: 24256729]
- Reimer PJ, Austin WE, Bard E, Bayliss A, Blackwell PG, Ramsey CB, Butzin M, Cheng H, Edwards RL,... & Talamo S (2020). The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon*, 62(4), 725–757.
- Ringbauer H, Novembre J, & Steinrücken M (2021). Parental relatedness through time revealed by runs of homozygosity in ancient DNA. *Nature Communications*, 12(1), 5425.
- Robin CJ, & Gorea M (2002). Les vestiges antiques de la grotte de Hôq (Suqutra, Yémen)(note d'information). *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres*, 146(2), 409–445.
- Rohland N, Glocke I, Aximu-Petri A, & Meyer M (2018). Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nature Protocols*, 13(11), 2447–2461. [PubMed: 30323185]
- Rohland N, Harney E, Mallick S, Nordenfelt S, & Reich D (2015). Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 370(1660), 20130624. [PubMed: 25487342]
- Santos GM, Southon JR, Druffel-Rodriguez KC, Griffin S, & Mazon M (2004). 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(1), 165–173.
- Schlebusch CM, Malmström H, Günther T, Sjödin P, Coutinho A, Edlund H, Munters AR, Vicente M, Steyn M, Soodyall H, Lombard M, & Jakobsson M (2017). Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science*, 358(6363), 652–655. [PubMed: 28971970]
- Schuenemann VJ, Peltzer A, Welte B, van Pelt WP, Molak M, Wang C-C, Furtwängler A, Urban C, Reiter E, Nieselt K, Teßmann B, Francken M, Harvati K, Haak W, Schiffels S, & Krause J (2017). Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. *Nature Communications*, 8(1), 1–11.
- Shinde V, Narasimhan VM, Rohland N, Mallick S, Mah M, Lipson M, Nakatsuka N, Adamski N, Broomandkhoshbacht N,... & Reich D (2019). An Ancient Harappan Genome Lacks Ancestry from Steppe Pastoralists or Iranian Farmers. *Cell*, 179, 729–735. [PubMed: 31495572]
- Shinnie PL (1960). Socotra. *Antiquity*, 34(134), 100–110.
- Sirak KA, Fernandes DM, Lipson M, Mallick S, Mah M, Olalde I, Ringbauer H, Rohland N, Hadden CS, Harney É, Adamski N, Bernardos R, Broomandkhoshbacht N, Callan K, Ferry M, Lawson AM, Michel M, Oppenheimer J, Stewardson K,... & Reich D (2021). Social stratification without genetic differentiation at the site of Kulubnarti in Christian Period Nubia. *Nature Communications*, 12(1), 7283.
- Skoglund P, Mallick S, Bortolini MC, Chennagiri N, Hünemeier T, Petzl-Erler ML, Salzano FM, Patterson N, & Reich D (2015). Genetic evidence for two founding populations of the Americas. *Nature*, 525(7567), 104–108. [PubMed: 26196601]
- Skoglund P, Posth C, Sirak K, Spriggs M, Valentin F, Bedford S, Clark GR, Reepmeyer C, Petchey F,... & Reich D (2016). Genomic insights into the peopling of the Southwest Pacific. *Nature*, 538(7626), 510–513. [PubMed: 27698418]
- Skoglund P, Thompson JC, Prendergast ME, Mittnik A, Sirak K, Hajdinjak M, Salie T, Rohland N, Mallick S, Peltzer A, Heinze A, Olalde I, Ferry M, Harney E, Michel M, Stewardson K, Cerezo-Román JI, Chiumia C, Crowther A,... & Reich D (2017). Reconstructing Prehistoric African Population Structure. *Cell*, 171(1), 59–71. [PubMed: 28938123]
- Soares P, Alshamali F, Pereira JB, Fernandes V, Silva NM, Afonso C, Costa MD, Musilová E, Macaulay V, Richards MB, erný V, & Pereira L (2012). The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Molecular Biology and Evolution*, 29 (3): 915–927. [PubMed: 22096215]
- Stern GH (1939). *Marriage in Early Islam*. Royal Asiatic Society.

- Strauch I (2012). *Foreign Sailors on Socotra: The inscriptions and drawings from the cave Hoq*. Hempen Verlag.
- Strauch I (2016). Indian Inscriptions from Cave Hoq at Socotra. *Ports of the Ancient Indian Ocean*, 79–97.
- Strauch I, & Bukharin M (2004). Indian Inscriptions from the Cave Hoq on Suqatr . *Annali dell'Istituto Universitario Orientale*, 64, 121–138.
- Stuiver M, & Polach HA (1977). Discussion reporting of 14C data. *Radiocarbon*, 19(3), 355–363.
- Uerpmann H-P, Potts D, & Uerpmann M (2010). Holocene (re-)occupation of eastern Arabia. In *The Evolution of Human Populations in Arabia: Paleoenvironments, Prehistory and Genetics* (eds Petraglia MD & Rose JI) 205–214.
- van de Loosdrecht M, Bouzouggar A, Humphrey L, Posth C, Barton N, Aximu-Petri A, Nickel B, Nagel S, Talbi EH, ... & Krause J (2018). Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. *Science*, 360(6388), 548–552. [PubMed: 29545507]
- Van Klinken GJ (1999). Bone collagen quality indicators for palaeodietary and radiocarbon measurements. *Journal of Archaeological Science*, 26(6), 687–695.
- Van Rensburg JJ (2018). The rock art of Dahaisi cave, Soqatra, Yemen. *Rock Art Research: The Journal of the Australian Rock Art Research Association (AURA)*, 35(2), 180–187.
- Vyas DN, Al-Meerri A, & Mulligan CJ (2017). Testing support for the northern and southern dispersal routes out of Africa: an analysis of Levantine and southern Arabian populations. *American Journal of Physical Anthropology*, 164(4), 736–749. [PubMed: 28913852]
- Wang C-C, Yeh H-Y, Popov AN, Zhang H-Q, Matsumura H, Sirak K, Cheronet O, Kovalev A, Rohland N, ... & Reich D (2021). Genomic insights into the formation of human populations in East Asia. *Nature*, 591(7850), 413–419. [PubMed: 33618348]
- Weeks L, Morris M, McCall B, AL-ZUBAIRY, K. J. A. a., & epigraphy. (2002). A recent archaeological survey on Soqatra. Report on the preliminary expedition season, January 5th–February 2nd 2001. *Arabian Archaeology and Epigraphy* 13(1), 95–125.
- Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt H-J, Kronenberg F, Salas A, & Schönherr S (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic acids research*, 44(W1), W58–W63. [PubMed: 27084951]
- Wellsted JR (1840). *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 vols.). Henry Colburn.
- Yule H, & Cordier H (1993). *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.

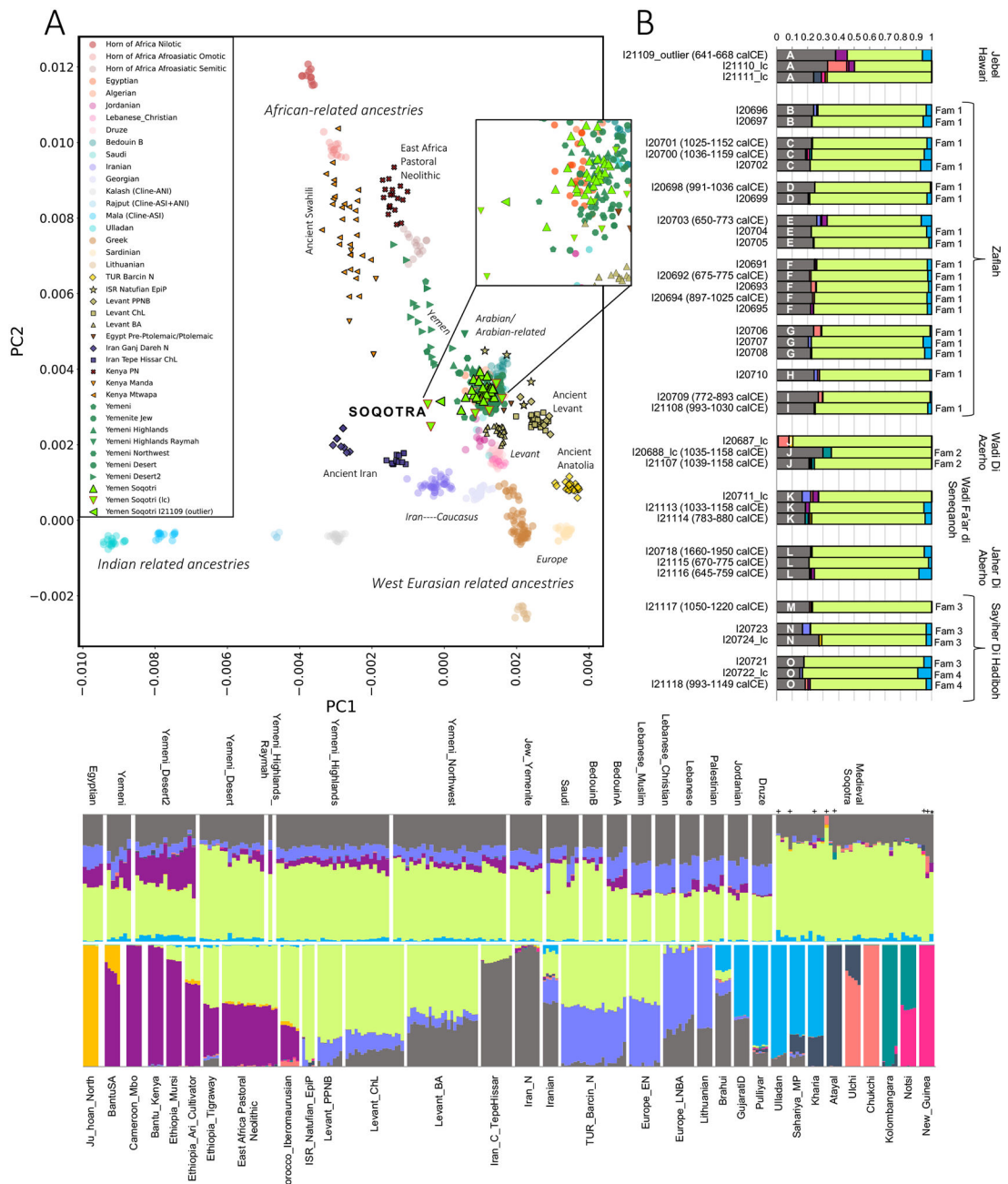




**Figure 1. Overview of Soqotri burials with analysis of shared drift between medieval Soqotri and present-day groups.**

(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 Human Origins (HO) array using Karitiana as an outgroup. The color scale is in the bottom left corner, showing 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 Soqatra. Legend is in the bottom left corner with additional

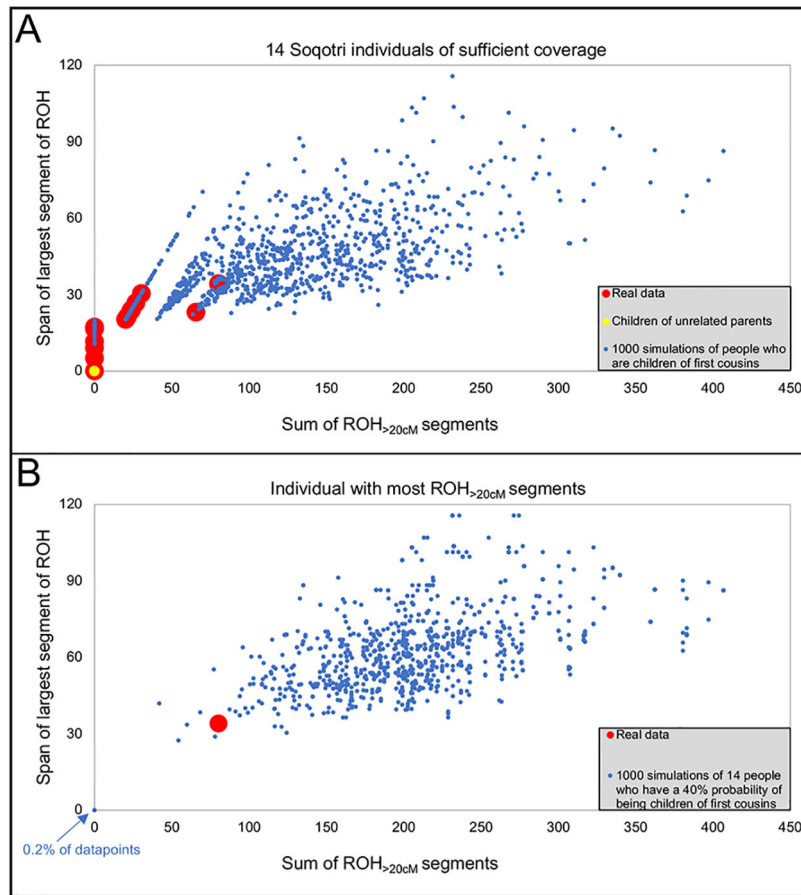
information in Supplementary Data 14 and Supplementary Note 10. A timeline of *tafone* usage based on direct  $^{14}\text{C}$  dates and dates of lineal relatives with known relationships is included (bottom center).



**Figure 2. Qualitative analyses of medieval Soqotri DNA.**

(A) PCA. We computed axes with present-day individuals (Supplementary Note 4) and projected ancient individuals onto PCs 1 and 2. Inset provides a zoomed view of the Soqotri 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* (labeled A – O), see Supplementary Data 14). Members of four families are identified (Fam 1 – 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 Panel B (top to bottom in Panel B corresponding to right to left in Panel C), with low coverage individuals denoted with a cross symbol and the outlier by an asterisk. A reference panel for the 10 components is on the bottom.



**Figure 3. Medieval Soqotri had fewer long ROH than expected from modern ethnographic studies.**

We focused on long ROH ( $>20\text{cM}$ ) 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 artifacts due to small island population sizes. (A) 1000 simulations of individual ROH profiles. We compare the real data to 1000 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 $>20\text{cM}$ , 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) 1000 simulated sets of 14 individuals. We compare the maximum ROH observed in 1000 simulations of 14 individuals assuming the modern ethnographic rate of 40% first cousin unions and find that the real data is at the  $p=0.007$  percentile (one-sided test) of simulations with respect to the total summed ROH $>20\text{cM}$ , and  $p=0.009$  percentile (one-sided test) of simulations with respect to largest ROH segment.

**Table 1.**  
***qpAdm* models of the distal ancestry of present-day Arabian or Arabian-related groups and Yemen\_Soqotri.**

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.

Target	% ancestry ( $\pm$ % s.e.)				% ancestry ( $\pm$ % s.e.)					Proportion of Levantine/Anatolian-related ancestry in the model (ISR_Natufian_EpiP-related)
	Levant_PPNB	Iran_N	Eth_4500 BP	p-value	ISR_Natuf_EpiP	Iran_N	TUR_Barcin_N	Eth_4500 BP	p-value	
BedouinA	59.3 $\pm$ 2.4	33.1 $\pm$ 2.3	7.6 $\pm$ 0.5	<i>0.04174</i>	31.7 $\pm$ 5.5	35.5 $\pm$ 2.7	25.1 $\pm$ 4.9	7.8 $\pm$ 1.0	0.00740	55.8
BedouinB	65.5 $\pm$ 2.6	30.4 $\pm$ 2.6	4.1 $\pm$ 0.6	<b><i>0.21748</i></b>	34.1 $\pm$ 5.6	32.7 $\pm$ 2.9	28.7 $\pm$ 5.1	4.5 $\pm$ 1.1	<b><i>0.10627</i></b>	54.3
Saudi	61.9 $\pm$ 2.6	36.2 $\pm$ 2.6	2.0 $\pm$ 0.6	<b><i>0.36931</i></b>	32.0 $\pm$ 5.1	38.8 $\pm$ 2.9	26.9 $\pm$ 4.8	2.4 $\pm$ 1.0	<b><i>0.54233</i></b>	54.3
Jew_Yemenite	63.2 $\pm$ 2.7	34.1 $\pm$ 2.7	2.7 $\pm$ 0.6	<b><i>0.21456</i></b>	33.6 $\pm$ 5.3	36.4 $\pm$ 2.9	27.0 $\pm$ 4.9	3.0 $\pm$ 1.0	<b><i>0.50612</i></b>	55.4
Yemeni	48.8 $\pm$ 2.5	36.8 $\pm$ 2.5	14.4 $\pm$ 0.7	<b><i>0.14294</i></b>	26.4 $\pm$ 4.8	38.6 $\pm$ 2.7	20.5 $\pm$ 4.6	14.6 $\pm$ 1.0	<b><i>0.65806</i></b>	56.2
Yemeni_Desert	67.5 $\pm$ 2.8	28.8 $\pm$ 2.8	3.7 $\pm$ 0.6	0.00072	41.9 $\pm$ 6.7	31.3 $\pm$ 3.5	23.4 $\pm$ 6.2	3.3 $\pm$ 1.2	<b><i>0.10329</i></b>	<b>64.1</b>
Yemeni_Desert2	43.4 $\pm$ 2.4	28.5 $\pm$ 2.4	28.1 $\pm$ 0.6	0.00342	27.7 $\pm$ 5.9	31.4 $\pm$ 2.7	13.4 $\pm$ 5.2	27.5 $\pm$ 1.2	<i>0.02966</i>	<b>67.4</b>
Yemeni_Highlands	62.9 $\pm$ 2.5	31.1 $\pm$ 2.5	6.1 $\pm$ 0.5	<i>0.01850</i>	35.4 $\pm$ 5.2	33.5 $\pm$ 2.8	25.1 $\pm$ 4.8	6.1 $\pm$ 1.0	<b><i>0.32483</i></b>	58.5
Yemeni_Highlands_Raymah	61.4 $\pm$ 4.4	19.0 $\pm$ 4.4	19.6 $\pm$ 1.3	<b><i>0.27552</i></b>	34.2 $\pm$ 7.5	21.8 $\pm$ 4.5	24.4 $\pm$ 7.3	19.6 $\pm$ 1.7	<b><i>0.39685</i></b>	58.3
Yemeni_Northwest	64.2 $\pm$ 2.5	31.9 $\pm$ 2.5	3.9 $\pm$ 0.6	<i>0.03131</i>	33.2 $\pm$ 4.9	33.9 $\pm$ 2.7	28.5 $\pm$ 4.5	4.3 $\pm$ 0.9	<b><i>0.45075</i></b>	53.8
Yemen_Soqotri	65.6 $\pm$ 2.9	32.3 $\pm$ 2.8	2.1 $\pm$ 0.6	0.00264	42.0 $\pm$ 6.7	35.1 $\pm$ 3.5	21.3 $\pm$ 6.1	1.6 $\pm$ 1.2	<b><i>0.13450</i></b>	<b>66.3</b>
Yemen_Soqotri (no African source)	67.6 $\pm$ 2.9	32.4 $\pm$ 2.9	-	0.00004	49.3 $\pm$ 5.2	36.6 $\pm$ 3.9	14.2 $\pm$ 5.1	-	<b><i>0.15674</i></b>	<b>77.6</b>