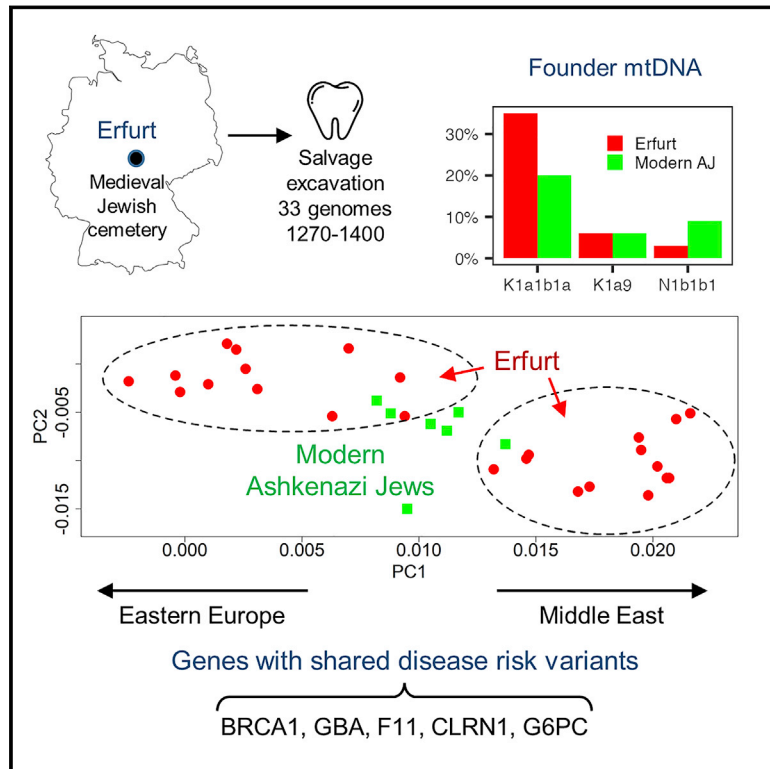


Genome-wide data from medieval German Jews show that the Ashkenazi founder event pre-dated the 14th century

Graphical abstract



Authors

Shamam Waldman, Daniel Backenroth, Éadaoin Harney, ..., Karin Sczech, Shai Carmi, David Reich

Correspondence

shai.carmi@huji.ac.il (S.C.), reich@genetics.med.harvard.edu (D.R.)

In brief

Genome-wide data was generated for 33 individuals whose skeletons were rescued from a late medieval Jewish cemetery in Erfurt, Germany. Medieval German Jews had similar genetic ancestry to modern Ashkenazi Jews but were more genetically heterogeneous, likely divided into two or more distinct groups. The medieval Jews showed evidence of an extreme founder event, including high levels of runs of homozygosity and distinctive genetic variants shared with modern Ashkenazi Jews.

Highlights

- Genome-wide data for 33 Jewish individuals from 14th-century Erfurt, Germany
- Medieval and modern Ashkenazi Jews (AJ) have similar ancestral genetic sources
- Medieval AJ were genetically heterogeneous, likely divided into two or more groups
- The individuals descend from an extreme founder event shared with modern AJ



Article

Genome-wide data from medieval German Jews show that the Ashkenazi founder event pre-dated the 14th century

Shamam Waldman,¹ Daniel Backenroth,¹ Éadaoin Harney,² Stefan Flohr,³ Nadia C. Neff,^{4,5} Gina M. Buckley,⁶ Hila Fridman,^{1,7,8} Ali Akbari,^{2,9,10} Nadin Rohland,^{9,10} Swapan Mallick,^{9,10,11} Iñigo Olalde,^{2,12,13} Leo Cooper,¹⁴ Ariel Lomes,¹ Joshua Lipson,¹⁵ Jorge Cano Nistal,¹⁶ Jin Yu,¹⁷ Nir Barzilai,¹⁸ Inga Peter,¹⁹ Gil Atzmon,^{18,20} Harry Ostrer,²¹ Todd Lencz,^{17,22,23} Yosef E. Maruvka,¹⁶ Maïke Lämmerhirt,²⁴ Alexander Beider,²⁵ Leonard V. Rutgers,²⁶ Virginie Renson,⁶ Keith M. Prufer,^{4,5} Stephan Schiffels,²⁷ Harald Ringbauer,^{2,27} Karin Sczech,²⁸ Shai Carmi,^{1,29,*} and David Reich^{2,9,10,11,*}

¹Braun School of Public Health and Community Medicine, The Hebrew University of Jerusalem, Jerusalem 9112102, Israel

²Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

³Department of Biology, University of Hildesheim, 31141 Hildesheim, Germany

⁴Department Anthropology, University of New Mexico, Albuquerque, NM 87131, USA

⁵Center for Stable Isotopes, University of New Mexico, Albuquerque, NM 87106, USA

⁶Research Reactor Center, University of Missouri, Columbia, MO 65211, USA

⁷Medical Genetics Institute, Shaare Zedek Medical Center, Jerusalem 9103102, Israel

⁸Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem 9112001, Israel

⁹Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

¹⁰Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

¹¹Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA

¹²Department of Zoology and Animal Cell Biology, University of the Basque Country UPV/EHU, 01006 Vitoria-Gasteiz, Spain

¹³Ikerbasque—Basque Foundation of Science, 48009 Bilbao, Spain

¹⁴Independent Scholar, Kalamazoo, MI 49008, USA

¹⁵Independent Scholar, New York, NY 10027, USA

¹⁶Faculty of Biotechnology and Food Engineering, Technion – Israel Institute of Technology, Haifa 3200003, Israel

¹⁷Department of Psychiatry, Division of Research, The Zucker Hillside Hospital Division of Northwell Health, Glen Oaks, NY 11004, USA

¹⁸Departments of Medicine and Genetics, Albert Einstein College of Medicine, Bronx, NY 10461, USA

¹⁹Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

²⁰Faculty of Natural Sciences, University of Haifa, Haifa 3498838, Israel

²¹Department of Pathology, Albert Einstein College of Medicine, Bronx, NY 10461, USA

²²Departments of Psychiatry and Molecular Medicine, Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY 11550, USA

²³Institute for Behavioral Science, The Feinstein Institutes for Medical Research, Manhasset, NY 11030, USA

²⁴Department of Medieval History, University of Erfurt, 99089 Erfurt, Germany

²⁵Independent Scholar, 92370 Chaville, France

²⁶Department of History and Art History, Utrecht University, Utrecht 3512 BS, the Netherlands

²⁷Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany

²⁸World Heritage Coordinator, City of Erfurt, 99084 Erfurt, Germany

²⁹Lead contact

*Correspondence: shai.carmi@huji.ac.il (S.C.), reich@genetics.med.harvard.edu (D.R.)

<https://doi.org/10.1016/j.cell.2022.11.002>

SUMMARY

We report genome-wide data from 33 Ashkenazi Jews (AJ), dated to the 14th century, obtained following a salvage excavation at the medieval Jewish cemetery of Erfurt, Germany. The Erfurt individuals are genetically similar to modern AJ, but they show more variability in Eastern European-related ancestry than modern AJ. A third of the Erfurt individuals carried a mitochondrial lineage common in modern AJ and eight carried pathogenic variants known to affect AJ today. These observations, together with high levels of runs of homozygosity, suggest that the Erfurt community had already experienced the major reduction in size that affected modern AJ. The Erfurt bottleneck was more severe, implying substructure in medieval AJ. Overall, our results suggest that the AJ founder event and the acquisition of the main sources of ancestry pre-dated the 14th century and highlight late medieval genetic heterogeneity no longer present in modern AJ.



INTRODUCTION

Ashkenazi Jews (AJ) emerged as a distinctive ethno-religious cultural group in the Rhineland in the 10th century (Frishman, 2008; Gladstein and Hammer, 2016). Since then, the AJ population expanded substantially, both geographically, first to Eastern Europe and recently beyond Europe, and in number, reaching about 10 million today (DellaPergola, 2015; Motulsky, 1995). Starting from the early days of human genetics, dozens of pathogenic recessive variants were identified in AJ (Charrow, 2004; Goodman, 1979; Ostrer, 2012), leading to the development of successful pre-conception screening programs (Gross et al., 2008; Kaback et al., 1993). A large fraction of these variants are extremely rare outside AJ and appear on the background of long shared haplotypes (e.g., Frisch et al., 2004; Hamel et al., 2011; Laitman et al., 2013; Raskin et al., 2011), implying that AJ descend from a small set of ancestral founders (Diamond, 1994; Ostrer and Skorecki, 2013; Risch et al., 2003; Slatkin, 2004). The Ashkenazi “founder event” is also evident in four mitochondrial lineages carried by 40% of AJ (Behar et al., 2006; Costa et al., 2013). More recently, studies found high rates of identical-by-descent (IBD) sharing in AJ, that is, nearly identical long haplotypes present in unrelated individuals, a hallmark of founder populations (Atzmon et al., 2010; Carmi et al., 2014a; Gusev et al., 2012; Henn et al., 2012). Quantitative modeling suggested that AJ experienced a sharp reduction in size (a “bottleneck”) in the late Middle Ages and that the (effective) number of founders was in the hundreds (Carmi et al., 2014a; Granot-Hershkovitz et al., 2018; Palamara et al., 2012; Santiago et al., 2020; Tournebize et al., 2022).

The origins of early AJ, as well as the history of admixture events that have shaped their gene pool, are subject to debate (Data S1, section 1). Genetic evidence supports a mixed Middle Eastern (ME) and European (EU) ancestry in AJ. This is based on uniparental markers with origins in either region (Behar et al., 2006, 2017; Costa et al., 2013; Hammer et al., 2000, 2009; Nebel et al., 2001), as well as autosomal studies showing that AJ have ancestry intermediate between ME and EU populations (Atzmon et al., 2010; Behar et al., 2010, 2013; Bray et al., 2010; Carmi et al., 2014a; Granot-Hershkovitz et al., 2018; Guha et al., 2012; Kopelman et al., 2020). These and other autosomal studies also showed that individuals with AJ ancestry are genetically distinguishable from those of other ancestries. Recent modeling suggested that most of the European ancestry in AJ is consistent with Southern European-related sources, and estimated the total proportion of European ancestry in AJ as 50%–70% (Carmi et al., 2014a; Xue et al., 2017; Yardumian and Schurr, 2019). The AJ population is overall highly genetically homogeneous, with no major ancestry differences based on present-day country of residence (Guha et al., 2012; Gusev et al., 2012; Kopelman et al., 2020). However, there are subtle average differences in ancestry between AJ with origins in Eastern vs. Western Europe (Behar et al., 2013; Gladstein and Hammer, 2019; Granot-Hershkovitz et al., 2018) (Data S1, section 1).

Despite the recent progress, open questions remain, including the localization of the founder event, or events, in time and space and the sources and times of the admixture events. Studying the genomes of individuals who lived closer to the time of AJ forma-

tion has the potential to shed light on these questions. We present here a DNA study of historical Jews, focusing on AJ from 14th-century Erfurt, Germany. The medieval Erfurt Jewish community existed between the late 11th century to 1454, with a short gap following a 1349 massacre (Weigelt, 2016). We report genome-wide data from 33 individuals whose skeletons were extracted in a salvage excavation. Our results demonstrate that Erfurt Ashkenazi Jews (EAJ) are genetically similar to modern Ashkenazi Jews (MAJ), implying little gene flow into the AJ gene pool since the 14th century. Further analysis demonstrates that EAJ were more genetically heterogeneous than MAJ, with multiple lines of evidence supporting the presence of two subgroups, one of which had higher Eastern European affinity compared to MAJ. The EAJ population shows strong evidence of a recent bottleneck shared with the bottleneck that affected MAJ, as alleles that are highly enriched in MAJ—including mitochondrial lineages and pathogenic variants—are also common in EAJ.

RESULTS

Historical and archaeological context, community engagement, and sample collection

The first Jewish community of Erfurt (pre-1349) was the oldest in Thuringia, and its cemetery also served nearby towns (Lämmerhirt, 2015; Toch, 1992). During the 1349 pogrom, most Jews of Erfurt and nearby communities were murdered or expelled (Weigelt, 2016). Jews returned to Erfurt in 1354 to form the second community (Lämmerhirt, 2016), which was one of the largest in Germany (Toch, 1992) (Data S1, section 2). The individuals we studied were buried in the south-western part of the medieval Jewish cemetery of Erfurt, which underwent salvage excavations in 2013 (STAR Methods and Data S1, section 3). The layout of the cemetery is shown in Figure 1A (see also Data S1, section 3).

Archaeological evidence tentatively suggests that the site was used by the second community (Data S1, section 3). Radiocarbon dating of ten individuals demonstrated that all lived between about 1270 and 1400 CE (STAR Methods and Data S2, Tables 1 and 3), but data were not informative on whether the site was used by the first or second community, due to a wiggle in the ¹⁴C calibration curve around the mid-14th century (Figure S1 and Data S1, section 3). The estimated ages at death (STAR Methods) ranged between 5 and 60 years old, with 14/33 (42%) estimated to be younger than 20 (Data S2, Table 1). The cause of death could be determined only for I14904, who was killed by several blows to the head by a sharp object.

Jewish rabbinical law, which was followed by EAJ (Data S1, section 2) prohibits exhumation of Jews for most purposes, and also proscribes disturbing the dead. Rabbinical discussions on testing ancient DNA from Jewish individuals have only recently appeared (Litke, 2021; Steinberg, 2019; Vigoda, 2017) and there is no centralized authority for establishing Jewish rabbinical guidance. As part of this study, we engaged with rabbinical authorities who reviewed our proposed research plan and approved the project under the conditions that only detached teeth are used, and that the analysis is performed only on already-excavated individuals. The study was then approved by

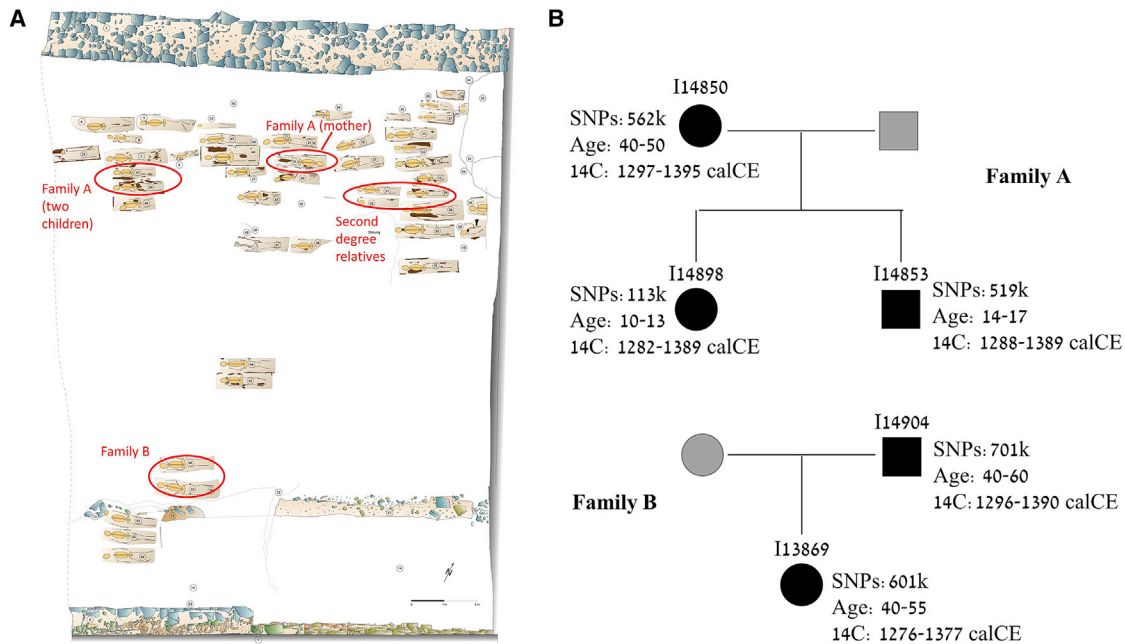


Figure 1. The medieval Jewish cemetery at Erfurt

(A) The layout of the cemetery. The inner city wall and the outer city wall are at the bottom and top of the map, respectively. Family members are marked in red ellipses (see next).

(B) The pedigrees of the two families identified based on first degree relationships. Black symbols represent individuals with DNA; gray symbols represent inferred family members. Circles: females; squares: males. For each individual, we indicate the ID, the number of genotyped SNPs, the estimated age at death, and the ¹⁴C date (95.4% probability intervals).

See also [Figure S1](#).

the Jewish community of Thuringia, Germany. Following these guidelines, we sampled teeth from 38 skeletal remains.

DNA sequencing

We followed existing protocols for DNA extraction, library preparation, and enrichment for about 1.24 million single nucleotide polymorphisms (SNPs). We then sequenced the enriched and non-enriched libraries and used multiple metrics to demonstrate very low levels of contamination ([STAR Methods](#) and [Data S2](#), Table 2). We obtained genome-wide data passing quality control for 33 individuals: 19 females and 14 males. The median proportion of human sequences was 0.03 (range 0.0003–0.67; [Data S2](#), Table 2), and the median coverage on autosomal targets was 0.45× (range 0.01–2.48×). The median number of SNPs covered by at least one sequence was 383k (range 11–842k). Children had significantly lower coverage than adults ($p = 6.7 \cdot 10^{-7}$; [Data S1](#), section 4). We found no traces of *Yersinia pestis* bacteria in the DNA sequences ([STAR Methods](#), [Data S1](#), section 4, and [Data S2](#), Table 4).

We identified two families based on first-degree relationships ([STAR Methods](#) and [Figure S1](#)): family A, with a mother, a son, and a daughter; and family B, with a father (the one killed by strokes to the head) and a daughter ([Figure 1B](#)). The two children of family A were buried next to each other, as were the two members of family B ([Figure 1A](#)). The mother of family A was buried three rows away from her children, in an orientation opposite to all other burials ([Figure 1A](#)). We also identified three distantly

related individuals, likely second-degree relatives, who were all buried next to each other ([Figure 1A](#)). Two of them were the only ones in our sample to carry the U5a1a2a mtDNA lineage ([Data S2](#), Table 1). However, the data for these three individuals were also consistent with a first-degree or no relationship ([Figure S1](#)), likely due to low coverage (13k, 15k, and 38k SNPs).

Ancestry estimation and medieval population structure

To analyze the ancestry of the EAJ individuals, we represented their genomes as “pseudo-haploids” using a single random sequence for each covered SNP. We merged EAJ with modern genomes from the Human Origins dataset (about 593k autosomal SNPs, all also enriched in EAJ), which included seven AJ and 86 other Jews. We projected the EAJ individuals on principal components learned from West Eurasian individuals of the Human Origins dataset ($n = 994$; [STAR Methods](#)). Eight EAJ had (post-merging) coverage of fewer than 50k SNPs, which did not allow reliable projection ([Data S1](#), section 5). We designated these individuals as “low-coverage” and excluded them from the principal components analysis (PCA).

In the PCA plot, EAJ individuals overlapped with MAJ, as well as with other European Mediterranean populations. Interestingly, EAJ had more variability along the European-Middle Eastern cline than MAJ ([Figure 2](#)). Higher variability in EAJ relative to MAJ was also observed when projecting a much larger MAJ sample ([Figure S2](#)) as well as in an ADMIXTURE analysis ([Alexander et al., 2009](#)) ([Data S1](#), section 5). We used *K*-means

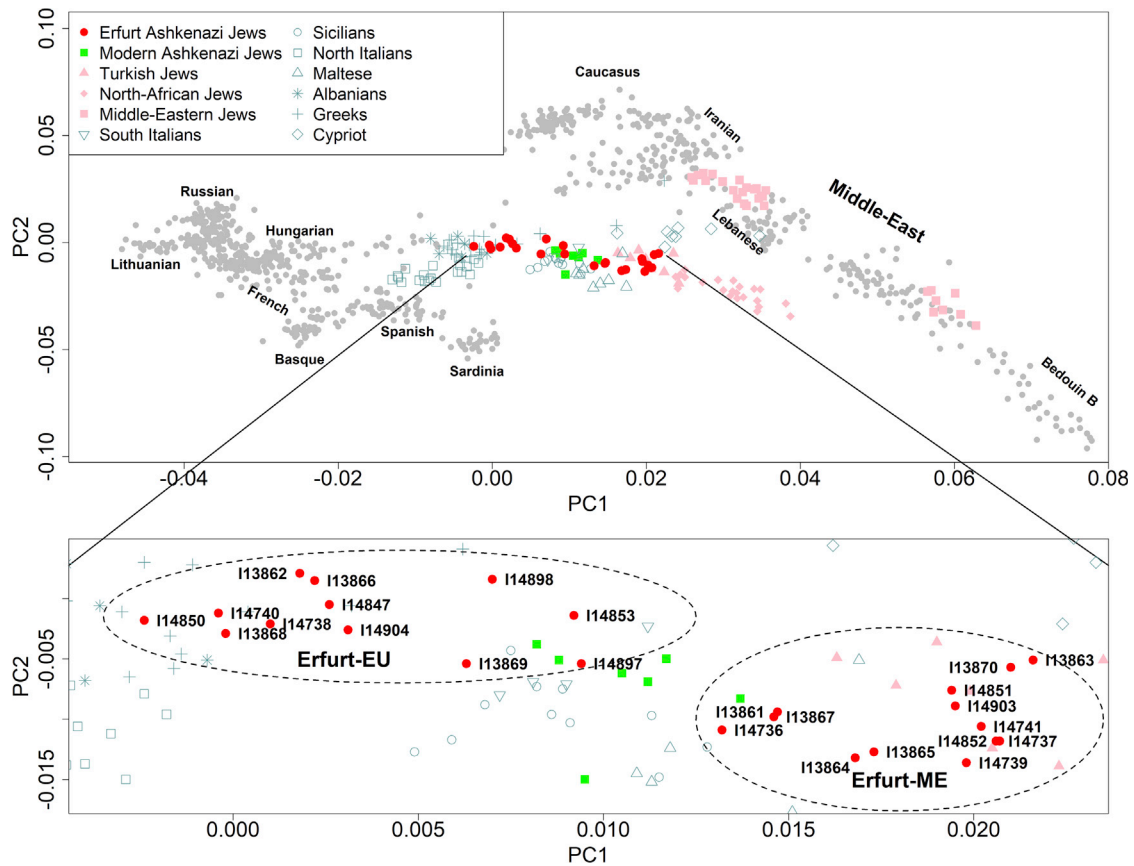


Figure 2. Principal components analysis

We learned the principal components (PCs) using West Eurasian populations (Lazaridis et al., 2014) and projected the Erfurt individuals (filled red circles) onto the inferred axes. Modern Ashkenazi Jews (green squares), Jews of non-Ashkenazi origin (pink shapes), and Mediterranean populations (teal shapes) are highlighted. The inset zooms in on the region that contains AJ individuals. See also Figure S2.

to cluster the EAJ individuals based on their PC1 and PC2 coordinates into two groups. One group, which we call “Erfurt-EU”, falls closer to individuals from European populations, while the other, which we call “Erfurt-ME”, is closer to Middle Eastern populations (Figure 2, inset).

Multiple lines of evidence provided statistical support for the dichotomization of EAJ as compared to a continuous gradient. A “gap statistic” analysis suggested that the optimal number of clusters is two, and another method showed that the clusters are significantly different (Data S1, section 6). To investigate the question from a population genetic perspective, we simulated two scenarios, either where all individuals had the same demographic history, or where one subgroup had an additional admixture event in Europe (Data S1, section 6). Summary statistics from the single group simulations (but not from the two-group simulations) were significantly different from those of EAJ, again supporting the presence of two distinct subgroups (Data S1, section 6). We also confirmed that clustering the EAJ individuals into two groups using two additional methods yielded identical clusters (Data S1, section 6). These results support the presence of at least two genetically distinct (even if possibly overlapping) groups in EAJ.

To further characterize the two Erfurt subgroups, we analyzed them separately and tested how they relate to MAJ of Eastern European or Western European origin (Data S1, section 5). Western MAJ almost entirely overlapped with Erfurt-ME in PCA. Eastern MAJ were intermediate between the two EAJ groups, and closer to Erfurt-ME (Figure S2). The Erfurt-ME group also overlapped with present-day Turkish (Sephardi) Jews (Figure 2, inset).

We used f_4 -statistics to test for evidence of gene flow between EAJ, MAJ, and other EU and ME populations (STAR Methods). We first ran tests of the form $f_4(\text{MAJ}, \text{EAJ}; X, \text{chimp})$, where X is any West Eurasian population. The results showed increasing Z scores, and hence increased affinity with MAJ as opposed to EAJ, as X changed from Eastern European to Central/Western European, Mediterranean, and Middle Eastern (Figure S3A). The same trend was observed for tests $f_4(\text{MAJ}, \text{Erfurt-EU}; X, \text{chimp})$, but the trend was opposite for tests $f_4(\text{MAJ}, \text{Erfurt-ME}; X, \text{chimp})$ (Figure S3A). These results suggest, in agreement with the PCA, that Erfurt-EU have more EU ancestry—particularly Eastern European—than Erfurt-ME or MAJ. In agreement, tests of the form $f_4(\text{Erfurt-EU}, \text{Erfurt-ME}; X, \text{chimp})$ showed increasing affinity with Erfurt-ME as X changed from Eastern

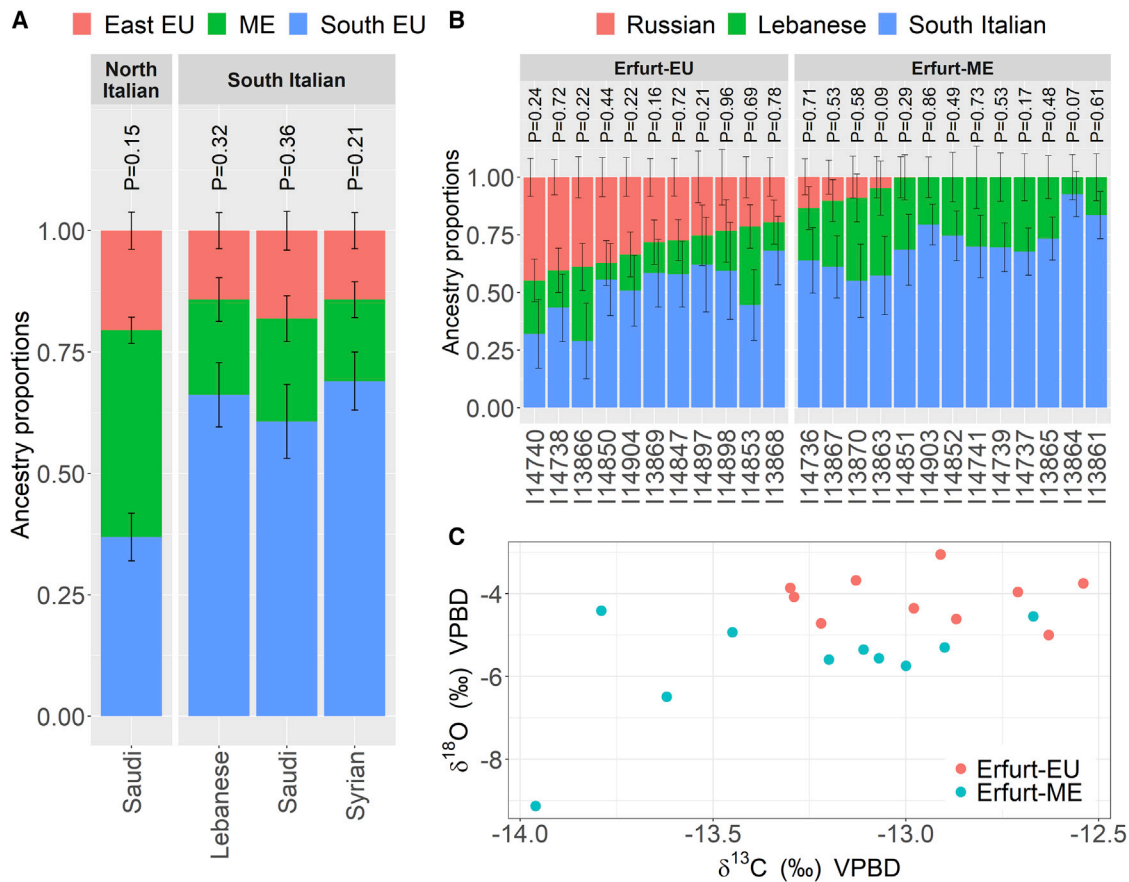


Figure 3. Models for the ancestry of Erfurt Ashkenazi Jews

(A) Each qpAdm model for the ancestry of Erfurt Jews includes a Middle Eastern, a Southern European, and an Eastern European (Russians) source. The Southern European source was either South or North Italians, as indicated at the top of each panel. The Middle Eastern source is indicated in the x axis labels. Only models with qpAdm p value >0.05 in the main analysis and in the robustness tests are shown (Table S3). Error bars represent one standard error in each direction. qpAdm p values are presented above each model.

(B) The ancestry of single Erfurt individuals, labeled by their IDs. We used qpAdm with Russian, Lebanese, and South Italian sources. The individuals are labeled by their Erfurt subgroup (EU/ME). qpAdm p values are shown for each individual. Results are not shown for low-coverage individuals (<50k SNPs), and for an additional individual who could not be modeled using these sources ($p < 0.05$).

(C) A plot of $\delta^{13}\text{C}_{\text{enamel}}$ and $\delta^{18}\text{O}_{\text{enamel}}$ stable isotope ratios for a subset of 20 Erfurt individuals with >200k SNPs. The Erfurt subgroup affiliation (EU/ME) is color-coded (legend).

See also Figure S3.

European to Central/Western European, Mediterranean, and Middle Eastern (Figure S3B). Finally, tests $f_4(\text{MAJ}, X; \text{EAJ}, \text{chimp})$, where X is any Jewish non-Ashkenazi population, were positive and very large for all X, suggesting that EAJ (including both subgroups) are closer to MAJ than to any other Jewish group (Figure S3C).

The increased affinity of Erfurt-EU with Eastern European populations may be related to the historically recorded migration of families from Bohemia, Moravia, and Silesia into the second Erfurt community (Data S1, section 2) (Lämmerhirt, 2019). To test whether some EAJ individuals had evidence of being migrants, we performed a stable isotope analysis on bioapatite derived from dental enamel (STAR Methods, Data S2, Tables 1 and 5). The $\delta^{13}\text{C}_{\text{enamel}}$ and $\delta^{18}\text{O}_{\text{enamel}}$ values are plotted for individuals with high coverage in Figure 3C, showing distinct distributions of stable isotope ratios between the two genetic

groups. The differences were significant for $\delta^{18}\text{O}$ ($p = 0.0005$; two-tailed Wilcoxon test), suggesting average differences in water sources during childhood between Erfurt-EU and Erfurt-ME. We found no correlation between the locations of the graves in the cemetery and the group affiliation (Mantel test $p = 0.46$) or PC1 and PC2 coordinates (Mantel test $p = 0.41$), showing that even though these groups were genetically distinctive, they show no evidence of being culturally or temporally segregated.

A qpWave analysis showed that EAJ and MAJ are consistent with forming a clade with respect to non-Jewish Europeans ($p = 0.15$; Table S1 and STAR Methods). This genetic similarity between EAJ and MAJ, despite living 600–700 years apart, suggests a high degree of endogamy over the period. Using simulations (STAR Methods), we inferred that any hypothetical admixture event between AJ and Eastern Europeans over the

past ≈ 20 generations must have been limited to replacing at most 2%–4% of the total AJ gene pool (corresponding to at most 0.2% replacement per generation; [Table S2](#) and [STAR Methods](#)). The same qpWave analysis with Erfurt-EU or Erfurt-ME had lower p values, particularly for Erfurt-EU, suggesting that each of these groups alone does not fully represent the entire modern AJ gene pool. When replacing EAJ or MAJ with (modern) Turkish Jews, South Italians, and Germans, all p values were under 0.05 (suggesting inconsistency with the given pair of populations being a clade), with the highest p value observed when comparing Erfurt-ME and Turkish Jews ($p = 0.01$; [Table S1](#)). We repeated the qpWave analysis to test whether pairs of populations form a clade with respect to non-Jewish Middle Eastern populations ([Table S1](#)). The p values were >0.05 for tests comparing MAJ and EAJ/Erfurt subgroups and either of these populations and Turkish Jews, suggesting that these populations have similar sources of Middle Eastern ancestry.

Quantitative ancestry modeling

We used qpAdm to test quantitative models for the ancestral sources of EAJ ([STAR Methods](#)). Based on the PCA above and previous modeling ([Xue et al., 2017](#)), we considered a model where EAJ is a mixture of the following sources: Southern European (South Italians or North Italians), Middle Eastern (Druze, Egyptians, Bedouins, Palestinians, Lebanese, Jordanians, Syrians, or Saudis), and Eastern European (Russians). We used modern populations as sources, as modeling with ancient sources was unsuccessful ([Data S1](#), section 7). Multiple models with South-Italians were plausible ($p > 0.05$; [Table S3](#)), which would be consistent with historical models pointing to the Italian peninsula as the source for the AJ population ([Data S1](#), section 16; though see below for alternatives and caveats). The mean admixture proportions (over all of our plausible models; [Table S3](#)) were 65% South Italy, 19% ME, and 16% East-EU ([Figure 3A](#)). We validated that our results did not qualitatively change when using only transversions vs. all SNPs, a different outgroup population, or fewer SNPs ([Table S3](#); [Data S1](#), section 7). Estimates of the admixture time were unreliable in our setting ([Data S1](#), section 8).

Models with other sources, in particular Mediterranean, could also fit the EAJ data ([Data S1](#), section 7). A model with North Italians as a source ([Table S3](#)) had ancestry proportions 37% North Italy, 43% ME, and 20% East-EU ([Figure 3A](#)). Models with Greek as a source had average ancestry proportions 51% Greek, 32% ME, and 17% East-EU ([Table S3](#)). Models with Spanish or North African sources (in addition to ME and East-EU) were not plausible. A model with all Levant populations merged together as the ME source fit the EAJ data ($p = 0.07$), with ancestry proportions 65% South Italy, 19% Levant, and 16% East-EU. A model with all Mediterranean populations merged as a single source (Middle Eastern, Greek, and Italian, with East-EU as the other source) fit the data ($p = 0.11$) with ancestry proportions 89% Mediterranean and 11% East-EU. Models with a Western European source (Germans) instead of Russians were not plausible. There was also no support for a minor ancestry component from East Asians.

We next used qpAdm to study the relations between EAJ, MAJ, and other Jewish groups ([Data S1](#), section 7). Erfurt-ME

could be modeled with Turkish (Sephardi) Jews (97% admixture proportion) and Germans (3%). Erfurt-EU could be modeled with Erfurt-ME (69%) and Russians (31%). Using a model with South Italian, Lebanese, and Russian sources for each EAJ individual ([Figure 3B](#)), we found striking variability in the Eastern European component, which was on average 33% in Erfurt-EU individuals but was not inferred in 9 of 13 Erfurt-ME individuals. Similar variability was observed using a North Italian source ([Data S1](#), section 7). MAJ could be modeled with Erfurt-ME (87%) and Russians (13%), or Erfurt-ME (86%) and Germans (14%). MAJ could also be modeled as having 60% ancestry from Erfurt-ME and 40% from Erfurt-EU ([Data S1](#), section 7).

Taken together, our results suggest that Erfurt-ME is a population genetically close to Sephardi Jews that has left nearly unadmixed descendants in modern AJ of Western European origin, while Erfurt-EU has experienced additional Eastern European-related admixture. The majority of AJ today likely formed as a nearly even mixture of populations represented by the two Erfurt groups. Linguistic, cultural, and onomastic studies have found differences between AJ from Western Europe, representing the early Rhineland communities, and AJ from Central and Eastern Europe ([Data S1](#), section 16). Our results suggest the hypothesis that Erfurt-ME and Erfurt-EU may be related to the Western and Central/Eastern AJ groups, respectively. The non-genetic differences nearly vanished by the 16th century ([Beider, 2015](#)), consistent with the lack of major genetic structure in modern AJ.

We caution that the specific identity of the source populations that we inferred, as well as the admixture proportions, should not be considered precise. This is due to the multiple Southern European populations that fit the EAJ data, as well as our reliance on modern populations as a proxy of the true ancestral sources. The levels of Middle Eastern ancestry in Italy were historically variable ([Aneli et al., 2021](#); [Antonio et al., 2019](#); [De Angelis et al., 2021](#); [Posth et al., 2021](#); [Raveane et al., 2019](#)), and Middle Eastern populations have also experienced demographic changes in the past two millennia, particularly African admixture ([Moorjani et al., 2011](#)) ([Data S1](#), section 16). Under the extensive set of models we studied, the ME ancestry in EAJ is estimated in the range 19%–43% and the Mediterranean European ancestry in the range 37%–65%. However, the true ancestry proportions could be higher or lower than implied by these ranges ([Data S1](#), section 16). Our results therefore should only be interpreted to suggest that AJ ancestral sources have links to populations living in Mediterranean Europe and the Middle East today.

Evidence for a founder event in the history of Erfurt Jews

Previous analyses of identical-by-descent (IBD) haplotypes ([Carmi et al., 2014a](#); [Granot-Hershkovitz et al., 2018](#); [Palamara et al., 2012](#)), mtDNA lineages ([Behar et al., 2006](#)), and pathogenic variants ([Risch et al., 2003](#); [Slatkin, 2004](#)) suggested that AJ experienced a medieval founder event (a bottleneck). However, the details of the event are yet to be fully resolved. Here, we used three sources of information — mtDNA lineages, runs of homozygosity, and MAJ-enriched variants — to test whether the EAJ population was shaped by the same founder event.

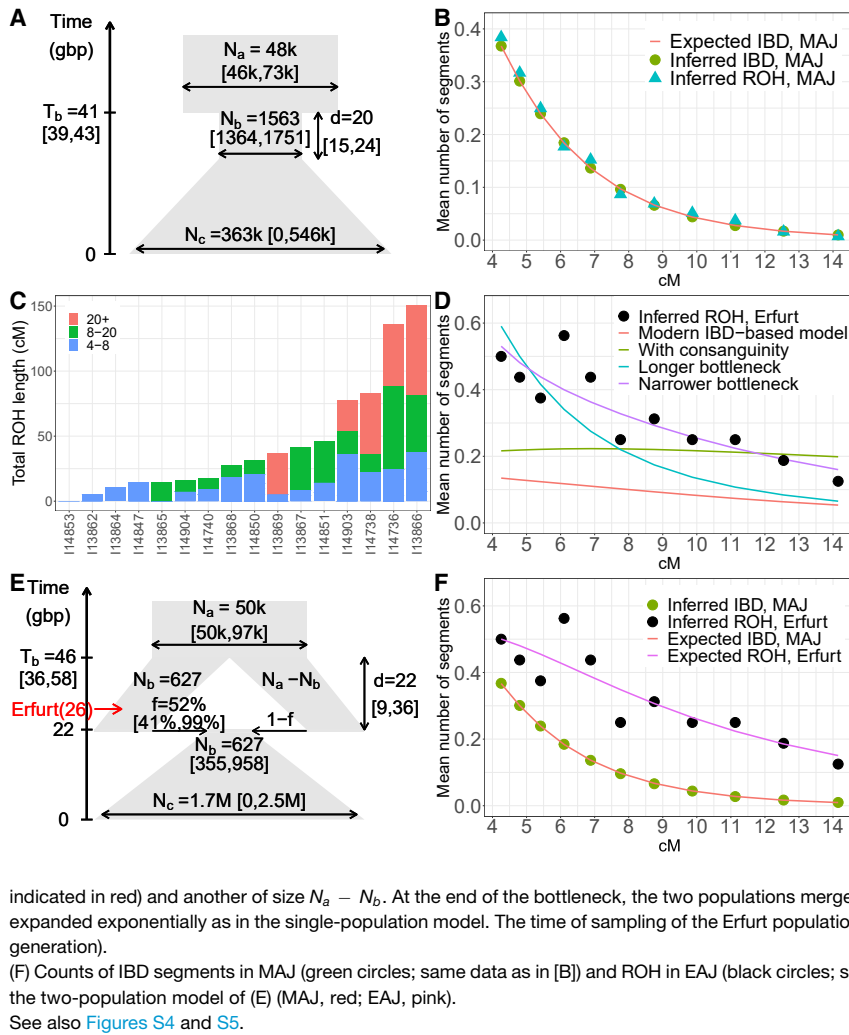


Figure 4. Models for AJ demographic history based on ancient and modern haplotypes

(A) A single-population model for the demographic history of AJ, inferred based on modern IBD sharing (Table S5, model [A]). In the diagram, the y axis represents the time in generations before present (gbp) and the width is (schematically) proportional to the effective population size. The 95% confidence intervals (CI) were computed using bootstrapping and are indicated near each parameter.

(B) The mean number of IBD and ROH segments (per pair of haploid autosomal genomes) in modern AJ vs. segment length (11 bins between 4 and 15 cM). Each symbol (circles for IBD, triangles for ROH) is placed at the middle of its corresponding bin. The red line shows the expected number of segments per bin based on the demographic model of (A).

(C) The total length of ROH segments in 16 EAJ individuals with >400k SNPs. The bars are colored proportionally to the contribution of segments of different lengths (legend).

(D) ROH counts in EAJ (circles) and the expected number based on various models (lines). The model inferred using modern IBD (panel [A]) is in red. The same model but allowing consanguinity in EAJ (Table S5, model [D]) is in green. Both models fit poorly to short ROH segments. A model similar to (A) but with either a narrower or a longer bottleneck (Table S5, models [E] and [F], respectively) are in purple and teal, respectively.

(E) A two-population model inferred jointly using IBD in MAJ and ROH in EAJ (Table S5, model [H]). According to the model, the ancestral population split $T_b = 46$ generations ago into one population of effective size $N_b = 627$ (representing Erfurt,

indicated in red) and another of size $N_a - N_b$. At the end of the bottleneck, the two populations merged with proportions $f = 52\%$ and $1 - f$, respectively, and expanded exponentially as in the single-population model. The time of sampling of the Erfurt population is shown at 26 generations ago (assuming 25 years per generation).

(F) Counts of IBD segments in MAJ (green circles; same data as in [B]) and ROH in EAJ (black circles; same data as in [D]), and the expectations (lines) based on the two-population model of (E) (MAJ, red; EAJ, pink). See also Figures S4 and S5.

We list the mtDNA lineages of EAJ in Data S2, Table 1 and report the number of carriers of the four most common Ashkenazi lineages (Behar et al., 2006) in Table S4. Remarkably, among 31 unrelated EAJ, 11 (35%) carried the K1a1b1a lineage, which is nearly absent in individuals of non-Jewish ancestry (Behar et al., 2006). This is greater than the 20% frequency in MAJ ($p = 0.04$; two-tailed binomial test; Table S4). All 11 carriers had identical sequences except for a single C/T polymorphism at position 16223 (C count: 3/11; Data S1, section 9). A joint Bayesian analysis (Bouckaert et al., 2019) of MAJ and EAJ K1a1b1a carriers (accounting for the known date of the EAJ individuals; STAR Methods) suggested a median posterior time to the most recent common ancestor about 1500 years ago, slightly earlier than previous point estimates (Behar et al., 2006; Costa et al., 2013), but with very high uncertainty (95% highest posterior density interval: 650–6700; Data S1, section 9).

The frequency of K1a1b1a in Erfurt-ME (7/13 = 54%) was higher than the frequency in Erfurt-EU (2/10 = 20%). The difference was not statistically significant, given the small sample size ($p = 0.20$; two-tailed Fisher's Exact test); however, a higher prevalence of K1a1b1a in Erfurt-ME is concordant with its higher

frequency in AJ of Western European origin (Behar et al., 2006; Costa et al., 2013). Among the other AJ founder lineages, two EAJ individuals carried the K1a9 lineage, one carried N1b1b1 (N1b2), and none carried K2a2a1 (Table S4). Overall, the mtDNA results provide evidence that EAJ were affected by the same bottleneck that affected MAJ.

Runs of homozygosity (ROH) appear in individuals whose parents are genetically related and are also expected in founder populations (Severson et al., 2019). We detected ROH in EAJ using hapROH (Ringbauer et al., 2021) (STAR Methods and Figure S4). We focused on 16 individuals covered in at least 400k SNPs (Data S1, section 10). The EAJ individuals had substantially higher levels of ROH compared to most other individuals from the historical period (Figure S4), with an average of 44 cM per individual in segments greater than 4 cM (30 cM in segments >8 cM and 14 cM in segments >20 cM; Figure 4C). These results provide additional support to the hypothesis that EAJ were affected by a strong bottleneck.

We next considered variants that are specific (or nearly specific) to MAJ and are also present in EAJ. We define AJ founder alleles as minor alleles (not necessarily disease causing) in SNPs

targeted by our in-solution enrichment that have frequency $>0.5\%$ in MAJ, $<0.01\%$ in Europeans (both using gnomAD [Karczewski et al., 2020]), and $\leq 1\%$ in the Middle East (STAR Methods). Overall, we identified 216 AJ founder alleles. Among these, 15 were present in at least one EAJ individual (Data S2, Table 1). To determine whether this number is expected under the scenario where EAJ post-dated the bottleneck, we used MAJ allele frequencies and ran binomial simulations to estimate the number of observed alleles given the EAJ sample size and per-individual coverage (STAR Methods). The [2.5,97.5]-percentiles of the simulated allele counts in the EAJ-like individuals were [14,32] (Figure S5C). (These expected allele counts are likely somewhat overestimated, due to the conditioning on alleles whose MAJ frequency was above a cutoff [Data S1, section 11]). The presence of 15 founder alleles in EAJ is thus within the range expected if EAJ have already experienced the AJ bottleneck (Data S1, section 11).

Inferring a demographic model of the founder event

To quantitatively estimate the AJ bottleneck parameters, we inferred demographic models based on IBD sharing (Palamara et al., 2012). We started with a model of an ancestral population of a constant (effective) size that experienced a bottleneck of effective size N_b starting T_b generations ago and lasting d generations, followed by an exponential expansion (Figure 4A). We used whole-genome sequencing data for $n = 637$ MAJ individuals (Carmi et al., 2014a; Lencz et al., 2018) and inferred IBD segments using IBDseq (STAR Methods). We then estimated the bottleneck parameters using a maximum likelihood approach (STAR Methods): $N_b = 1563$ (diploid individuals; 95% confidence interval (CI): [1364–1751]), $T_b = 41$ (generations; 95% CI: [39–43]), and $d = 20$ (generations; 95% CI: [15–24]; see Figure 4A and Table S5). The predicted counts of IBD segments (based on the model of Figure 4A) provide good fit to the counts from the MAJ data (Figure 4B). Assuming 25 years per generation, our model places the onset of the bottleneck about 1000 years ago, at the formation of the early AJ communities (Data S1, section 16). We inferred a very similar model using ROH in the same MAJ individuals (Data S1, section 12) with good fit to the empirical data (Figure 4B).

We next sought to determine whether our inferred demographic model provides a good fit to the ancient EAJ data (after accounting for the ≈ 650 -year difference; STAR Methods), which would be expected given the genetic similarity between MAJ and EAJ. However, ROH counts in EAJ exceeded the expectation based on the model (Figure 4A), in particular for short and intermediate segments (Figure 4D), indicating that the Erfurt bottleneck must have been more severe. Given previous simulation results (Ringbauer et al., 2021) and further exploration here (Figure S4), the gap is unlikely to be due to false positive ROH calls. Modeling consanguinity in EAJ also did not explain the excess of short ROH (Figure 4D and Data S1, section 12). To fit the ancient data, the Erfurt bottleneck must have been either 3.0-fold narrower or 2.4-fold longer compared to the expectation based on modern IBD (Figure 4D, Table S5, and Data S1, section 12). However, these bottleneck parameters did not fit the modern data (Figure S5A). The best fit model to both modern IBD and ancient ROH data also did not provide good overall fit (Fig-

ure S5B, Table S5, and Data S1, section 12). Taken together, these results suggest that the single-population (and single-bottleneck) model of Figure 4A does not fully describe the AJ demographic history.

Given that the Erfurt population only represents a single AJ medieval site, we hypothesized that a missing component in our model is medieval substructure. We therefore considered a model in which the AJ population split at the onset of the bottleneck into two groups experiencing different bottleneck intensities, one of them represented by EAJ (Figure 4E) and the other unsampled. We showed by simulations that our method can infer the model parameters even in the presence of extreme imbalance between the amount of modern and ancient data (Data S1, section 13). In the best fit model (Figure 4E and Table S5), the bottleneck started $T_b = 46$ generations ago (95% CI: [36,58]) and lasted $d = 22$ generations (95% CI: [9,36]). The group represented by EAJ experienced a narrow bottleneck of size $N_b = 627$ (95% CI: [355, 958]) and contributed 52% (95% CI: [41,99]%) to the MAJ gene pool. The remaining contribution came from a second group that did not experience the initial bottleneck but contracted only at the end of the bottleneck, when merging with the other AJ group (Figure 4E). This model fits well both modern and ancient data (Figure 4F). We used parametric bootstrapping to demonstrate that the favorable fit is not due to over-fitting (STAR Methods and Data S1, section 13).

Our model makes several assumptions: that all demographic events happened in discrete epochs; that the non-Erfurt subgroup experienced no initial bottleneck; that there was no gene flow from non-AJ populations or between the AJ subgroups; that population splitting and merging coincided with the start and end of the bottleneck, respectively; and that the bottleneck remained of a constant size throughout its duration. Given that these assumptions are very specific, our results should not be interpreted as a statement on the correct form of the demographic model: we cannot rule out alternative models, particularly more complex ones. Our results serve as a proof-by-example illustrating how a model of substructure, with different groups undergoing different bottleneck intensities, can reproduce the modern and ancient haplotype sharing data.

Pathogenic variants

If EAJ have experienced the AJ founder event, we expect them to carry some of the pathogenic variants enriched in present-day AJ. We compiled a list of 62 pathogenic variants (Carmi et al., 2014a), after excluding variants with high frequency in Europeans and East-Asians (STAR Methods and Data S2, Table 6). As only six of these were included in our 1240k enriched SNPs, we used imputation to search for additional variants. We used a reference panel of 702 MAJ whole-genomes (Carmi et al., 2014a; Lencz et al., 2018) and two imputation methods. The first, which we call PHCP (Pseudo-Haploid ChromoPainter), takes pseudo-haploid SNP data as input. It is based on the diploid Li and Stephens hidden Markov model (Lawson et al., 2012; Li and Stephens, 2003), as we describe in the STAR Methods. The second is GLIMPSE, based on genome-wide sequence data (Rubinacci et al., 2021). We validated the high accuracy of PHCP using Mendelian consistency and masking

Table 1. High-confidence AJ-enriched pathogenic variants detected in Erfurt

Disease	Gene	Variant	Carrier ID and PHCP probability of carrying at least one alternate allele	Frequency in modern AJ (gnomAD)	Frequency in non-Finnish Europeans (gnomAD)	Gene included in Ashkenazi PCS panels
Retinitis pigmentosa 59	<i>DHDDS</i>	c.124A>G	I14903, 0.985	0.52%	0.004%	4/4
Gaucher disease, type 1	<i>GBA</i>	c.84dupG	I13861, 1.000	0.08%	0.002%	4/4
Usher syndrome, type 3	<i>CLRN1</i>	p.N48K	I14897, 0.980	0.58%	0.008%	4/4
Factor XI deficiency	<i>F11</i>	p.E135X	I13870, genotyped	1.71%	0.03%	3/4
Factor XI deficiency	<i>F11</i>	p.F301L	I13865, genotyped	2.38%	0.03%	3/4
Cystic fibrosis	<i>CFTR</i>	p.G542X	I14736, 1.000	0.15%	0.04%	4/4
Parkinson's disease	<i>LRRK2</i>	p.G2019S	I14739, genotyped	0.84%	0.03%	0/4
Acyl CoA dehydrogenase deficiency	<i>ACADS</i>	c.319C>T	I14737, 0.999	1.88%	0.02%	0/4
Familial Mediterranean fever	<i>MEFV</i>	p.V726A	I14739, 0.993	3.93%	0.09%	2/4
Glycogen storage disease, type 1A	<i>G6PC</i>	p.R83C	I13870, 0.970	0.66%	0.05%	4/4
Breast/ovarian cancer predisposition	<i>BRCA1</i>	c.68_69delAG	I13861, 1.000	0.41%	0.009%	0/4

For each of the 11 variants, we indicate the disease, the gene, and the variant in HGVS (Human Genome Variation Society) nomenclature. The c.68_69delAG BRCA1 variant is also known as 185delAG. Variants that were detected in sequences covering enriched SNPs are designated as "genotyped". For imputed variants, we provide the marginal posterior probability in PHCP for having at least one alternate allele. We further provide the carrier IDs, the allele frequency in MAJ and non-Finnish Europeans (gnomAD), and the number (out of four) of Ashkenazi-specific pre-conception screening (PCS) panels where the gene is included (STAR Methods).

of genotyped founder alleles, as well as confirmed high concordance between the two methods (Data S1, section 14 and Data S2, Table 6).

We discovered 11 high-confidence pathogenic variants in the EAJ genomes (STAR Methods and Table 1). Five additional variants were detected with low confidence (STAR Methods and Data S2, Table 6). The high confidence variants were carried by eight EAJ individuals, with each variant appearing once (Data S2, Table 6). Six of the discovered variants were previously dated using modern genomic data, with their estimated times of origin consistent with presence in the 14th century (Data S1, section 14). Most of variants are in genes included in pre-conception screening panels (Table 1). A caveat is that imputation demonstrates the presence of *haplotypes* carrying the variants, not the variants themselves. However, the low false positive rate we observed with genotyped founder alleles (Data S1, section 14) argues against this scenario being common.

Finally, we assessed variants associated with selected non-medical or polygenic phenotypes. Variants for lactase persistence, eye and hair pigmentation, and (putatively) plague risk had similar allele frequency between EAJ and MAJ (Data S1, section 15). Osteological stature estimates for 13 adults were correlated with polygenic scores for height, although not significantly (Data S1, section 15).

DISCUSSION

By analyzing genome-wide data from historical AJ individuals, we refine our understanding of early AJ origins. The ancestry of EAJ was closely related to that of MAJ, as evidenced by the PCA, ADMIXTURE, and qpWave analyses, suggesting a high degree of continuity of AJ ancestry over the past \approx 700 years. However, EAJ individuals had more variable ancestry than MAJ and

were stratified by the presence of a minor Eastern European-related ancestry component. Multiple lines of evidence suggest that the EAJ population had already experienced a "bottleneck" shared with MAJ: the high frequency of Ashkenazi founder mtDNA lineages, the presence of AJ-enriched pathogenic variants and other alleles, and long runs of homozygosity. In agreement with previous studies (Carmi et al., 2014a; Granot-Hershkovitz et al., 2018; Santiago et al., 2020), we date the onset of AJ expansion to about 20–25 generations ago (Data S1, section 16).

Our ancient DNA data allowed us to identify patterns in the history of AJ that would not have been detectable from modern genetic variation. Specifically, our results suggest that the AJ population was more structured during the Middle Ages than it is today. Within Erfurt, one group had an enrichment of Eastern European-related ancestry (Figures 2 and 3B), while the other had ancestry close to that of MAJ of Western European origin and modern Sephardi Jews (Figures 2 and S2). The two groups also had distinct levels of enamel $\delta^{18}\text{O}$ (Figure 3C). Our results cannot exclude even finer divisions within Erfurt, and medieval AJ may have been additionally structured beyond Erfurt (Figure 4E). In contrast, present-day AJ are remarkably homogeneous (Guha et al., 2012; Gusev et al., 2012; Kopelman et al., 2020). This suggests that even though the overall sources of ancestry remained similar between medieval and modern AJ, endogamy and within-AJ mixture since medieval times have contributed to the homogenization of the AJ gene pool.

We can speculate on the identity of the two Erfurt subgroups based on non-genetic data suggesting that in the late Middle Ages, the AJ population was divided linguistically and culturally along a west/east axis (Beider, 2015; Katz, 1993) (Data S1, section 16). The Western AJ group, which may correspond to Erfurt-ME, likely represented descendants of the early AJ settlers in the Rhineland. One source explicitly mentions Erfurt as lying on the

boundary between the two groups (Weinreich, 2008), and in the 14th century, given names in Erfurt were typical of both Western and Eastern AJ (Beider, 2001). These points, together with evidence on migration into the second Erfurt community from the East (Data S1, section 2), may explain why individuals belonging to both genetic groups were present in Erfurt. The increased genetic diversity in Erfurt does not contradict our inference that EAJ experienced a more severe bottleneck than MAJ: while EAJ descended from fewer founders compared to MAJ, these founders were likely more variable in terms of their European genetic ancestry.

Limitations of this study

We caution against interpreting the qpAdm models for the ancestral sources of EAJ as quantifying direct contributions of specific populations to the early AJ gene pool. This is because (1) the wide range of inferred ancestry proportions across models; (2) the historical fluctuations in ME ancestry in Italy (Antonio et al., 2019); and (3) the large space of models not explored here. Instead, the results reflect genetic links between the ancestral sources of AJ and modern populations, who may or may not have genetic continuity with their ancient antecedents. Further caveats are our inability to identify a satisfactory model for modern AJ and the technical difficulties when modeling an ancient population using modern sources with qpAdm (Harney et al., 2021) (although see our robustness tests; Table S3 and Data S1, section 7).

As with other ancient DNA studies, our historical inferences are based on a single site in time and space. This implies that our data may not be representative of the full genetic diversity of early AJ, as we have indeed inferred (Figure 4E). However, even for a single site, our sample size was relatively large (>30), and we were able to capture substructure not present in MAJ. Further, we were able to show genetic continuity with MAJ, as well as sharing of the same founder event. Another limitation is the reliance of our demographic models on a relatively small number of short runs of homozygosity (Figures 4D and S5), which are difficult to infer from pseudo-haploid data. In general, our inferred demographic model (Figure 4E) should not be interpreted as a full and precise demographic reconstruction; rather, it should be viewed as a simplified model (perhaps one among many) that captures the main features of the observed genetic data.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Archaeology
- METHOD DETAILS
 - Age at death estimation

- Radiocarbon dating
- Isotope analysis
- DNA extraction and sequencing
- Bioinformatics and quality control
- Pathogen DNA scan

● QUANTIFICATION AND STATISTICAL ANALYSES

- Detecting genetic relatives
- Investigating genetic ancestry using PCA
- Admixture modeling with *ADMIXTOOLS*
- Detecting admixture with *qpWave*
- Inferring admixture models with *qpAdm*
- Inferring admixture time with *DATES*
- Maximal post-medieval gene flow into AJ
- Correlation between EAJ grave location and group affiliation
- Aligning the mtDNA sequences of modern and ancient K1a1b1a carriers
- Estimating the time to the most recent common ancestor (TMRCA) of the K1a1b1 carriers
- Detecting runs of homozygosity in the ancient genomes
- Ashkenazi founder alleles in Erfurt Jews
- Binomial simulations of founder allele counts
- Detecting IBD and ROH segments in modern genomes
- Demographic inference using IBD and ROH segments
- Defining pathogenic founder variants
- Imputation using pseudo-haploid *ChromoPainter (PHCP)*
- Imputation using *GLIMPSE*
- Analyzing the pathogenic founder variants detected in Erfurt Jews

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2022.11.002>.

ACKNOWLEDGMENTS

We are thankful to Ephraim Shoham-Steiner for consultations when planning this study and to Rabbi Ze'ev Litke, who undertook a ruling on contexts in which ancient DNA studies could be permissible. We thank the Jewish community of Thuringia for approving the study. We thank Douglas Kennett, Nick Patterson, Lily Agranat-Tamir, and Maria Stürzebecher for discussions; Shaul Stampfer, Sergio DellaPergola, and Itsik Pe'er for commenting on the manuscript; Kim Callan, Fatma Zalzal, Kristin Stewardson, Nicole Adamski, and Ann Marie Lawson for wet laboratory work; Rebecca Bernardos for sample management; and Iosif Lazaridis, Matthew Mah, Adam Micco, and Zhao Zhang for bioinformatics support. The study was funded by the Israel Science Foundation grant 407/17 and the United States–Israel Binational Science Foundation grant 2017024 to S.C., by the National Science Foundation (USA) grants 1912776 and 0922374 to V.R., by the MCIN/AEI/ 10.13039/501100011033 and by “ESF Investing in your future” grant “Ayudas para contratos Ramón y Cajal” to I.O., and by the following grants to D.R.: NIH grants GM100233 and HG012287; the Allen Discovery Center program, a Paul G. Allen Frontiers Group advised program of the Paul G. Allen Family Foundation; John Templeton Foundation grant 61220; a private gift from Jean-François Clin; and the Howard Hughes Medical Institute.

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.

AUTHOR CONTRIBUTIONS

S.C. and D.R. conceived, designed, and supervised the study. K.S. performed the archaeological excavation. S.F. performed the physical anthropology analyses. D.B. and J.C.N. performed the mitochondrial DNA analysis. N.C.N., G.M.B., K.M.P., and V.R. performed the isotope analyses. M.L. provided details on the history of the Erfurt community. L.V.R., A.B., L.C., A.L., and J.L. provided historical context on Jewish origins. N.R. supervised the laboratory experiments. S.M. supervised the bioinformatics analyses. I.O. performed the relatedness analysis. A.A. performed imputation using GLIMPSE. É.H. performed the pathogen DNA analysis. H.R. detected runs of homozygosity. H.F. provided information on carrier screening. L.C., A.L., and J.L. inferred refined uniparental lineages. S.W. performed all other data analyses. H.O., N.B., G.A., I.P., and T.L. provided the Ashkenazi genomic reference data. J.Y. performed quality control on the reference data. Y.E.M. provided input on the mtDNA analysis. S.S. provided input on the population genetic analyses. S.W. and S.C. drafted the manuscript with input from other authors and revised it together with D.R.; H.R., H.O., T.L., L.V.R., and S.S. provided feedback on the manuscript.

DECLARATION OF INTERESTS

S.C. is a paid consultant and holds stock options at MyHeritage. D.B. is an employee and shareholder at The Janssen Pharmaceutical Companies of Johnson & Johnson. É.H. is an employee of 23andMe.

Received: May 8, 2022

Revised: August 26, 2022

Accepted: November 1, 2022

Published: November 30, 2022

REFERENCES

- Agranat-Tamir, L., Waldman, S., Martin, M.A.S., Gokhman, D., Mishol, N., Eshel, T., Cheronet, O., Rohland, N., Mallick, S., Adamski, N., et al. (2020). The genomic history of the bronze age southern levant. *Cell* **181**, 1146–1157.e11.
- Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664.
- Amorim, C.E.G., Vai, S., Posth, C., Modi, A., Koncz, I., Hakenbeck, S., La Rocca, M.C., Mende, B., Bobo, D., Pohl, W., et al. (2018). Understanding 6th-century barbarian social organization and migration through paleogenomics. *Nat. Commun.* **9**, 3547.
- Aneli, S., Caldon, M., Saube, T., Montinaro, F., and Pagani, L. (2021). Through 40,000 years of human presence in Southern Europe: the Italian case study. *Hum. Genet.* **140**, 1417–1431.
- Antonio, M.L., Gao, Z., Moots, H.M., Lucci, M., Candilio, F., Sawyer, S., Oberreiter, V., Calderon, D., Devitofranceschi, K., Aikens, R.C., et al. (2019). Ancient Rome: A genetic crossroads of Europe and the Mediterranean. *Science* **366**, 708–714.
- Atzmon, G., Hao, L., Pe'er, I., Velez, C., Pearlman, A., Palamara, P.F., Morrow, B., Friedman, E., Oddoux, C., Burns, E., and Ostrer, H. (2010). Abraham's children in the genome era: major Jewish diaspora populations comprise distinct genetic clusters with shared Middle Eastern Ancestry. *Am. J. Hum. Genet.* **86**, 850–859.
- Behar, D.M., Metspalu, E., Kivisild, T., Achilli, A., Hadid, Y., Tzur, S., Pereira, L., Amorim, A., Quintana-Murci, L., Majamaa, K., et al. (2006). The matrilineal ancestry of Ashkenazi Jewry: portrait of a recent founder event. *Am. J. Hum. Genet.* **78**, 487–497.
- Behar, D.M., Metspalu, M., Baran, Y., Kopelman, N.M., Yunusbayev, B., Gladstein, A., Tzur, S., Sahakyan, H., Bahmanimehr, A., Yepiskoposyan, L., et al. (2013). No evidence from genome-wide data of a Khazar origin for the Ashkenazi Jews. *Hum. Biol.* **85**, 859–900.
- Behar, D.M., Saag, L., Karmin, M., Gover, M.G., Wexler, J.D., Sanchez, L.F., Greenspan, E., Kushniarevich, A., Davydenko, O., Sahakyan, H., et al. (2017). The genetic variation in the R1a clade among the Ashkenazi Levites' Y chromosome. *Sci. Rep.* **7**, 14969.
- Behar, D.M., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.L., Silva, N.M., Kivisild, T., Torroni, A., and Villems, R. (2012). A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684.
- Behar, D.M., Yunusbayev, B., Metspalu, M., Metspalu, E., Rosset, S., Parik, J., Rootsi, S., Chaubey, G., Kutuev, I., Yudkovsky, G., et al. (2010). The genome-wide structure of the Jewish people. *Nature* **466**, 238–242.
- Beider, A. (2001). *A Dictionary of Ashkenazic Given Names: Their Origins, Structure, Pronunciation, and Migrations* (Avotaynu).
- Beider, A. (2015). *Origins of Yiddish Dialects* (Oxford: Oxford University Press).
- Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., et al. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650.
- Bray, S.M., Mulle, J.G., Dodd, A.F., Pulver, A.E., Wooding, S., and Warren, S.T. (2010). Signatures of founder effects, admixture, and selection in the Ashkenazi Jewish population. *Proc. Natl. Acad. Sci. USA* **107**, 16222–16227.
- Briggs, A.W., and Heyn, P. (2012). Preparation of Next-Generation Sequencing Libraries from Damaged DNA. In *Ancient DNA*, S. B. and H. M., eds. (Humana Press), p. 143.
- Bronk Ramsey, C. (2009). Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**, 337–360.
- Browning, B.L., and Browning, S.R. (2013). Detecting identity by descent and estimating genotype error rates in sequence data. *Am. J. Hum. Genet.* **93**, 840–851.
- Buikstra, J.E., and Ubelaker, D.H. (1994). *Standards for Data Collection from Human Skeletal Remains* (Fayetteville: Arkansas Archeological Survey).
- Carmi, S., Hui, K.Y., Kochav, E., Liu, X., Xue, J., Grady, F., Guha, S., Upadhyay, K., Ben-Avraham, D., Mukherjee, S., et al. (2014a). Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins. *Nat. Commun.* **5**, 4835.
- Carmi, S., Wilton, P.R., Wakeley, J., and Pe'er, I. (2014b). A renewal theory approach to IBD sharing. *Theor. Popul. Biol.* **97**, 35–48.
- Charrow, J. (2004). Ashkenazi Jewish genetic disorders. *Fam. Cancer* **3**, 201–206.
- Chintalapati, M., Patterson, N., and Moorjani, P. (2022). Reconstructing the spatiotemporal patterns of admixture during the European Holocene using a novel genomic dating method. *eLife* **11**, e77625. <https://doi.org/10.7554/eLife.77625>.
- Costa, M.D., Pereira, J.B., Pala, M., Fernandes, V., Olivieri, A., Achilli, A., Perigo, U.A., Rychkov, S., Naumova, O., Hatina, J., et al. (2013). A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. *Nat. Commun.* **4**, 2543.
- Dabney, J., Knapp, M., Glocke, I., Gansauge, M.T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Pääbo, S., Arsuaga, J.L., and Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. USA* **110**, 15758–15763.
- De Angelis, F., Veltre, V., Romboni, M., Di Corcia, T., Scano, G., Martínez-Labarga, C., Catalano, P., and Rickards, O. (2021). Ancient genomes from a rural site in Imperial Rome (1st–3rd cent. CE): a genetic junction in the Roman Empire. *Ann. Hum. Biol.* **48**, 234–246.
- DellaPergola, S. (2015). *World Jewish Population, 2015* (Berman Jewish DataBank in Cooperation with Association for the Social Scientific Study of Jewry).
- Diamond, J.M. (1994). Human genetics. Jewish lysosomes. *Nature* **368**, 291–292.
- Fridman, H., Behar, D.M., Carmi, S., and Levy-Lahad, E. (2020). Preconception carrier screening yield: effect of variants of unknown significance in partners of carriers with clinically significant variants. *Genet. Med.* **22**, 646–653.

- Frisch, A., Colombo, R., Michaelovsky, E., Karpati, M., Goldman, B., and Pleg, L. (2004). Origin and spread of the 1278insTATC mutation causing Tay-Sachs disease in Ashkenazi Jews: genetic drift as a robust and parsimonious hypothesis. *Hum. Genet.* *114*, 366–376.
- Frishman, A. (2008). Early Ashkenazi Jews: Since Their Settlement in North-West Europe to the First Crusade (Tel Aviv: Hakibbutz Hameuchad).
- Fu, Q., Hajdinjak, M., Moldovan, O.T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., et al. (2015). An early modern human from Romania with a recent Neanderthal ancestor. *Nature* *524*, 216–219.
- Fu, Q., Mittnik, A., Johnson, P.L.F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., et al. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* *23*, 553–559.
- Gansauge, M.T., Aximu-Petri, A., Nagel, S., and Meyer, M. (2020). Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nat. Protoc.* *15*, 2279–2300.
- Gansauge, M.T., Gerber, T., Glocke, I., Korlevic, P., Lippik, L., Nagel, S., Riehl, L.M., Schmidt, A., and Meyer, M. (2017). Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* *45*, e79.
- Giffin, K., Lankapalli, A.K., Sabin, S., Spyrou, M.A., Posth, C., Kozakaitė, J., Friedrich, R., Miliuskienė, Ž., Jankauskas, R., Herbig, A., and Bos, K.I. (2020). A treponemal genome from an historic plague victim supports a recent emergence of yaws and its presence in 15th century Europe. *Sci. Rep.* *10*, 9499.
- Gladstein, A., and Hammer, M.F. (2016). Population Genetics of the Ashkenazim. In *eLS* (John Wiley).
- Gladstein, A.L., and Hammer, M.F. (2019). Substructured Population Growth in the Ashkenazi Jews Inferred with Approximate Bayesian Computation. *Mol. Biol. Evol.* *36*, 1162–1171.
- Goodman, R.M. (1979). Genetic Disorders Among the Jewish People (The Johns Hopkins University Press).
- Granot-Hershkovitz, E., Karasik, D., Friedlander, Y., Rodriguez-Murillo, L., Dorajoo, R., Liu, J., Sewda, A., Peter, I., Carmi, S., and Hochner, H. (2018). A study of Kibbutzim in Israel reveals risk factors for cardiometabolic traits and subtle population structure. *Eur. J. Hum. Genet.* *26*, 1848–1858.
- Gross, S.J., Pletcher, B.A., Monaghan, K.G., and Professional Practice and Guidelines Committee, and Guidelines, C. (2008). Carrier screening in individuals of Ashkenazi Jewish descent. *Genet. Med.* *10*, 54–56.
- Guha, S., Rosenfeld, J.A., Malhotra, A.K., Lee, A.T., Gregersen, P.K., Kane, J.M., Pe'er, I., Darvasi, A., and Lencz, T. (2012). Implications for health and disease in the genetic signature of the Ashkenazi Jewish population. *Genome Biol.* *13*, R2.
- Gusev, A., Palamara, P.F., Aponte, G., Zhuang, Z., Darvasi, A., Gregersen, P., and Pe'er, I. (2012). The architecture of long-range haplotypes shared within and across populations. *Mol. Biol. Evol.* *29*, 473–486.
- Hamel, N., Feng, B.J., Foretova, L., Stoppa-Lyonnet, D., Narod, S.A., Imyanitov, E., Sinilnikova, O., Tihomirova, L., Lubinski, J., Gronwald, J., et al. (2011). On the origin and diffusion of BRCA1 c.5266dupC (5382insC) in European populations. *Eur. J. Hum. Genet.* *19*, 300–306.
- Hammer, M.F., Behar, D.M., Karafet, T.M., Mendez, F.L., Hallmark, B., Erez, T., Zhivotovsky, L.A., Rosset, S., and Skorecki, K. (2009). Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish priesthood. *Hum. Genet.* *126*, 707–717.
- Hammer, M.F., Redd, A.J., Wood, E.T., Bonner, M.R., Jarjanazi, H., Karafet, T., Santachiara-Benerecetti, S., Oppenheim, A., Jobling, M.A., Jenkins, T., et al. (2000). Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. *Proc. Natl. Acad. Sci. USA* *97*, 6769–6774.
- Harney, E., Patterson, N., Reich, D., and Wakeley, J. (2021). Assessing the performance of qpAdm: a statistical tool for studying population admixture. *Genetics* *217*, iyaa045.
- Henn, B.M., Hon, L., Macpherson, J.M., Eriksson, N., Saxonov, S., Pe'er, I., and Mountain, J.L. (2012). Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. *PLoS One* *7*, e34267.
- Herbig, A., Maixner, F., Bos, K.I., Zink, A., Krause, J., and Huson, D.H. (2016). MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman. Preprint at bioRxiv. <https://doi.org/10.1101/050559>.
- Hübner, R., Key, F.M., Warinner, C., Bos, K.I., Krause, J., and Herbig, A. (2019). HOPS: automated detection and authentication of pathogen DNA in archaeological remains. *Genome Biol.* *20*, 280.
- Kaback, M., Lim-Steele, J., Dabholkar, D., Brown, D., Levy, N., and Zeiger, K. (1993). Tay-Sachs disease-carrier screening, prenatal diagnosis, and the molecular era. An international perspective, 1970 to 1993. The International TSD Data Collection Network. *JAMA* *270*, 2307–2315.
- Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* *581*, 434–443.
- Katz, D. (1993). East and West, khes and shin and the origins of Yiddish. In *Studies in Jewish culture in honour of Chone Shmeruk, I. Bartal, E. Mendelsohn, and C. Turniansky, eds.* (Jerusalem: The Zalman Shazar Center for Jewish History), pp. 9–37.
- Kennett, D.J., Plog, S., George, R.J., Culleton, B.J., Watson, A.S., Skoglund, P., Rohland, N., Mallick, S., Stewardson, K., Kistler, L., et al. (2017). Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nat. Commun.* *8*, 14115.
- Kopelman, N.M., Stone, L., Hernandez, D.G., Gefel, D., Singleton, A.B., Heyer, E., Feldman, M.W., Hillel, J., and Rosenberg, N.A. (2020). High-resolution inference of genetic relationships among Jewish populations. *Eur. J. Hum. Genet.* *28*, 804–814.
- Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinf.* *15*, 356.
- Laitman, Y., Feng, B.J., Zamir, I.M., Weitzel, J.N., Duncan, P., Port, D., Thirthagiri, E., Teo, S.H., Evans, G., Latif, A., et al. (2013). Haplotype analysis of the 185delAG BRCA1 mutation in ethnically diverse populations. *Eur. J. Hum. Genet.* *21*, 212–216.
- Lämmerhirt, M. (2015). Die Anfänge der jüdischen Besiedlung in Thüringen. *Zeitschrift für Thüringische Geschichte* *69*, 57–91.
- Lämmerhirt, M. (2016). Die Schutzauflösung 1453 und das Ende der zweiten jüdischen Gemeinde. In *Die Erfurter jüdische Gemeinde im Spannungsfeld zwischen Stadt, Erzbischof und Kaiser (Jena/Quedlinburg: Busert & Stadel)*, pp. 124–137.
- Lämmerhirt, M. (2019). Neuedition des Liber Judeorum der Stadt Erfurt. In *Wohnen, beten, handeln Das hochmittelalterliche jüdische Quartier antepontem in Erfurt Mit einer Neuedition des Liber Judeorum der Stadt Erfurt*, B. Perlich, ed. (Petersberg: Michael Imhof), pp. 297–399.
- Lawson, D.J., Hellenthal, G., Myers, S., and Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS Genet.* *8*, e1002453.
- Lazaridis, I., Alpaslan-Roodenberg, S., Acar, A., Açıkkol, A., Agelarakis, A., Aghikyan, L., Akyüz, U., Andreeva, D., Andrijašević, G., Antonović, D., et al. (2022). The genetic history of the Southern Arc: A bridge between West Asia and Europe. *Science* *377*, eabm4247.
- Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* *513*, 409–413.
- Lencz, T., Yu, J., Palmer, C., Carmi, S., Ben-Avraham, D., Barzilai, N., Bressman, S., Darvasi, A., Cho, J.H., Clark, L.N., et al. (2018). High-depth whole genome sequencing of an Ashkenazi Jewish reference panel: enhancing sensitivity, accuracy, and imputation. *Hum. Genet.* *137*, 343–355.

- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Li, N., and Stephens, M. (2003). Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* 165, 2213–2233.
- Litke, Z. (2021). DNA Sampling from Skeletal Remains - Halakhic Considerations. (translated). <https://din.org.il/2021/09/11/%d7%93%d7%92%d7%99%d7%9e%d7%95%d7%aa-%d7%93%d7%a0%d7%90-%d7%9e%d7%a9%d7%9c%d7%93%d7%99%d7%9d-%d7%a2%d7%aa%d7%99%d7%a7%d7%99%d7%9d-%d7%a1%d7%95%d7%92%d7%99%d7%95%d7%aa-%d7%94%d7%9c%d7%9b/>.
- Lohse, J.C., Culleton, B.J., Black, S.L., and Kennett, D.J. (2014). A precise chronology of Middle to Late Holocene bison exploitation in the far Southern Great Plains. *J. Tex. Archeol. Hist.* 1, 94.
- Margaryan, A., Lawson, D.J., Sikora, M., Racimo, F., Rasmussen, S., Moltke, I., Cassidy, L.M., Jørsboe, E., Ingason, A., Pedersen, M.W., et al. (2020). Population genomics of the Viking world. *Nature* 585, 390–396.
- Maricic, T., Whitten, M., and Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* 5, e14004.
- Monroy Kuhn, J.M., Jakobsson, M., and Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. *PLoS One* 13, e0195491.
- Moorjani, P., Patterson, N., Hirschhorn, J.N., Keinan, A., Hao, L., Atzmun, G., Burns, E., Ostrer, H., Price, A.L., and Reich, D. (2011). The history of African gene flow into Southern Europeans, Levantines, and Jews. *PLoS Genet.* 7, e1001373.
- Motulsky, A.G. (1995). Jewish diseases and origins. *Nat. Genet.* 9, 99–101.
- Mullen, K., Ardia, D., Gil, D., Windover, D., and Cline, J. (2011). DEoptim: An R Package for Global Optimization by Differential Evolution. *J. Stat. Software* 40, 1.
- Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., and Durbin, R. (2016). BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* 32, 1749–1751.
- Nebel, A., Filon, D., Brinkmann, B., Majumder, P.P., Faerman, M., and Oppenheim, A. (2001). The Y chromosome pool of Jews as part of the genetic landscape of the Middle East. *Am. J. Hum. Genet.* 69, 1095–1112.
- Ostrer, H. (2012). *Legacy: A Genetic History of the Jewish People* (New York: Oxford University Press).
- Ostrer, H., and Skorecki, K. (2013). The population genetics of the Jewish people. *Hum. Genet.* 132, 119–127.
- Palamara, P.F., Lencz, T., Darvasi, A., and Pe'er, I. (2012). Length distributions of identity by descent reveal fine-scale demographic history. *Am. J. Hum. Genet.* 91, 809–822.
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. *Genetics* 192, 1065–1093.
- Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* 2, e190.
- Posth, C., Zaro, V., Spyrou, M.A., Vai, S., Gnecci-Ruscione, G.A., Modi, A., Peltzer, A., Mötsch, A., Nägele, K., Vågø, Å.J., et al. (2021). The origin and legacy of the Etruscans through a 2000-year archeogenomic time transect. *Sci. Adv.* 7, eabi7673.
- Raskin, L., Schwenker, F., Freytsis, M., Tischkowitz, M., Wong, N., Chong, G., Narod, S.A., Levine, D.A., Bogomolny, F., Aronson, M., et al. (2011). Characterization of two Ashkenazi Jewish founder mutations in MSH6 gene causing Lynch syndrome. *Clin. Genet.* 79, 512–522.
- Raveane, A., Aneli, S., Montinaro, F., Athanasiadis, G., Barlera, S., Birolo, G., Boncoraglio, G., Di Blasio, A.M., Di Gaetano, C., Pagani, L., et al. (2019). Population structure of modern-day Italians reveals patterns of ancient and archaic ancestries in Southern Europe. *Sci. Adv.* 5, eaaw3492.
- Reimer, P.J., Austin, W.E.N., Bard, E., Bayliss, A., Blackwell, P.G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R.L., Friedrich, M., et al. (2020). The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal BP). *Radiocarbon* 62, 725–757.
- Ringbauer, H., Coop, G., and Barton, N.H. (2017). Inferring Recent Demography from Isolation by Distance of Long Shared Sequence Blocks. *Genetics* 205, 1335–1351.
- Ringbauer, H., Novembre, J., and Steinrücken, M. (2021). Parental relatedness through time revealed by runs of homozygosity in ancient DNA. *Nat. Commun.* 12, 5425.
- Risch, N., Tang, H., Katzenstein, H., and Ekstein, J. (2003). Geographic distribution of disease mutations in the Ashkenazi Jewish population supports genetic drift over selection. *Am. J. Hum. Genet.* 72, 812–822.
- Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20130624.
- Rubinacci, S., Ribeiro, D.M., Hofmeister, R.J., and Delaneau, O. (2021). Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat. Genet.* 53, 120–126.
- Santiago, E., Novo, I., Pardiñas, A.F., Saura, M., Wang, J., and Caballero, A. (2020). Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium. *Mol. Biol. Evol.* 37, 3642–3653.
- Severson, A.L., Carmi, S., and Rosenberg, N.A. (2019). The Effect of Consanguinity on Between-Individual Identity-by-Descent Sharing. *Genetics* 212, 305–316.
- Slatkin, M. (2004). A population-genetic test of founder effects and implications for Ashkenazi Jewish diseases. *Am. J. Hum. Genet.* 75, 282–293.
- Steinberg, A. (2019). Halakhic considerations on exhuming the Yemenite Jewish children (translated). *Med Ethics.* 29, 111–116. <https://www.medethics.org.il/article/%D7%A4%D7%AA%D7%99%D7%97%D7%AA-%D7%A7%D7%91%D7%A8%D7%99-%D7%99%D7%9C%D7%93%D7%99-%D7%AA%D7%99%D7%9E%D7%9F-%D7%A9%D7%99%D7%A7%D7%95%D7%9C%D7%99%D7%9D-%D7%94%D7%9C%D7%9B%D7%AA%D7%99%D7%99/>.
- Stoukal, M., and Hanáková, H. (1978). Die Länge der Längsknochen altslawischer Bevölkerungen - Unter besonderer Berücksichtigung von Wachstumsfragen. *Homo* 29, 53.
- Toch, M. (1992). Siedlungsstruktur der Juden Mitteleuropas im Wandel von Mittelalter zur Neuzeit. In *Juden in der christlichen Umwelt während des späten Mittelalters*, A. Haverkamp and F.-J. Ziwes, eds. (Berlin: Duncker & Humblot), pp. 29–39.
- Tournebise, R., Chu, G., and Moorjani, P. (2022). Reconstructing the history of founder events using genome-wide patterns of allele sharing across individuals. *PLoS Genet.* 18, e1010243.
- Ubelaker, D.H. (1978). *Human Skeletal Remains: Excavation, Analysis, Interpretation* (New Brunswick, NJ: Taraxacum).
- Vågø, Å.J., Herbig, A., Campana, M.G., Robles García, N.M., Warinner, C., Sabin, S., Spyrou, M.A., Andrades Valtueña, A., Huson, D., Tuross, N., et al. (2018). *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat. Ecol. Evol.* 2, 520–528.
- van Klinken, G.J. (1999). Bone collagen quality indicators for palaeodietary and radiocarbon measurements. *J. Archaeol. Sci.* 26, 687–695.
- Veeramah, K.R., Rott, A., Groß, M., van Dorp, L., López, S., Kirsanow, K., Sell, C., Blöcher, J., Wegmann, D., Link, V., et al. (2018). Population genomic analysis of elongated skulls reveals extensive female-biased immigration in Early Medieval Bavaria. *Proc. Natl. Acad. Sci. USA* 115, 3494–3499.
- Vigoda, M. 2017 Exhumation and Genetic Testing for Identification - a Review of Halakhic Questions (translated). https://www.gov.il/BlobFolder/dynamiccollecresultitem/ptihat-kevarim-yaldei-teiman/he/legislative-opinion_126-ptihat%20kevarim%20yaldei%20teiman.doc

- Wasik, K., Berisa, T., Pickrell, J.K., Li, J.H., Fraser, D.J., King, K., and Cox, C. (2021). Comparing low-pass sequencing and genotyping for trait mapping in pharmacogenetics. *BMC Genom.* *22*, 197.
- Weigelt, C.M. (2016). Das Erfurter Pestpogrom 1349. Eine kritische Rekonstruktion. In *Die Erfurter jüdische Gemeinde im Spannungsfeld zwischen Stadt, Erzbischof und Kaiser (Jena/Quedlinburg: Bussert & Stadeler)*, pp. 30–122.
- Weinreich, M. (2008). *History of the Yiddish Language* 1st ed., 1 (United States: Yale University Press).
- Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.J., Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* *44*, W58–W63.
- Xue, J., Lencz, T., Darvasi, A., Pe'er, I., and Carmi, S. (2017). The time and place of European admixture in Ashkenazi Jewish history. *PLoS Genet.* *13*, e1006644.
- Yardumian, A., and Schurr, T.G. (2019). The Geography of Jewish Ethnogenesis. *J. Anthropol. Res.* *75*, 206–234.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
2× HI-RPM hybridization buffer	Agilent Technologies	5190-0403
Herculase II Fusion DNA Polymerase	Agilent Technologies	600679
Pfu Turbo Cx Hotstart DNA Polymerase	Agilent Technologies	600412
50% PEG 8000	Anatrace	OPTIMIZE-82 100 ML
0.5 M EDTA pH 8.0	BioExpress	E177
Sera-Mag™ SpeedBead Carboxylate-Modified [E3] Magnetic Particles	Cytiva Life Sciences	65152105050250
silica magnetic beads	G-Biosciences	786-916
10 × T4 RNA Ligase Buffer	New England Biolabs	B0216L
Bst DNA Polymerase2.0, large frag.	New England Biolabs	M0537
UGI	New England Biolabs	M0281
USER enzyme	New England Biolabs	M5505
Buffer PB	QIAGEN	19066
Buffer PE concentrate	QIAGEN	19065
1 M Tris-HCl pH 8.0	Sigma Aldrich	AM9856
1 M NaOH	Sigma Aldrich	71463
20% SDS	Sigma Aldrich	5030
3 M Sodium Acetate (pH 5.2)	Sigma Aldrich	S7899
5 M NaCl	Sigma Aldrich	S5150
Ethanol	Sigma Aldrich	E7023
Guanidine hydrochloride	Sigma Aldrich	G3272
Isopropanol	Sigma Aldrich	650447
PEG-8000	Sigma Aldrich	89510
Proteinase K	Sigma Aldrich	P6556
Tween-20	Sigma Aldrich	P9416
Water	Sigma Aldrich	W4502
10× Buffer Tango	Thermo Fisher Scientific	BY5
50× Denhardt's solution	Thermo Fisher Scientific	750018
AccuPrime Pfx Polymerase (2.5 U/ul)	Thermo Fisher Scientific	12344032
ATP	Thermo Fisher Scientific	R0441
dNTP Mix	Thermo Fisher Scientific	R1121
Dyna MyOne Streptavidin C1 beads	Thermo Fisher Scientific	65002
FastAP (1 U/μL)	Thermo Fisher Scientific	EF0651
GeneAmp 10× PCR Gold Buffer	Thermo Fisher Scientific	4379874
Human Cot-I DNA	Thermo Fisher Scientific	15279011
Klenow Fragment (10 U/μL)	Thermo Fisher Scientific	EP0052
Maxima Probe qPCR 2xMM	Thermo Fisher Scientific	K0233
Maxima SYBR Green kit	Thermo Fisher Scientific	K0251
Maxima SYBR Green kit	Thermo Fisher Scientific	K0253
Salmon sperm DNA	Thermo Fisher Scientific	15632-011
SSC Buffer (20×)	Thermo Fisher Scientific	AM9770
T4 DNA Ligase	Thermo Fisher Scientific	EL0012
T4 DNA Ligase, HC (30U/μL)	Thermo Fisher Scientific	EL0013

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
T4 DNA Polymerase	Thermo Fisher Scientific	EP0062
T4 Polynucleotide Kinase	Thermo Fisher Scientific	EK0032
2× HI-RPM hybridization buffer	Agilent Technologies	5190-0403
2% Sodium Hypochlorite Solution	Millipore Sigma	Cat# XX0637-76
Acetic Acid, Glacial (TraceMetal Grade)	Fisher Chemical	Cat# A507-P212
7 M HNO ₃ (Optima)	Fisher Chemical	Cat# A467-2
6 M HCL (TraceMetal Grade)	Fisher Chemical	Cat# A508-4
0.05 M HNO ₃ (Optima)	Fisher Chemical	Cat# A467-2
30% H ₂ O ₂ (GR ACS Grade)	Millipore Sigma	Cat# HX0635-2
0.1 M CH ₃ COOH (GR ACS Grade)	Millipore Sigma	Cat# AX0073-6
Critical commercial assays		
MinElute PCR Purification Kit	QIAGEN	28006
NextSeq® 500/550 High Output Kit v2 (150 cycles)	Illumina	FC-404-2002
HiSeq X Reagent Kits	Illumina	FC-501-2501
Deposited data		
Raw and analyzed data from 33 newly reported ancient humans	This paper	ENA: PRJEB53475
A stable isotope analysis of 20 ancient samples reported in this study	This paper	IsoBank. ID: 686
Human Origins genotype data	Reich Lab website	https://reich.hms.harvard.edu/allen-ancient-dna-resource-aad-downloadable-genotypes-present-day-and-ancient-dna-data
Western and Eastern Ashkenazi Jews	Behar et al. (2013)	http://www.evolutioon.ut.ee/MAIT/jew_data/
Ashkenazi Genome Consortium sequencing data	Carmi et al. (2014a); Lencz et al. (2018)	https://ega-archive.org/studies/EGAS00001000664
Italy Imperial, Late Antiquity and Medieval sequencing data	Antonio et al. (2019)	https://reich.hms.harvard.edu/allen-ancient-dna-resource-aad-downloadable-genotypes-present-day-and-ancient-dna-data
Canaanites SNP enrichment	Agranat-Tamir et al. (2020)	https://reich.hms.harvard.edu/allen-ancient-dna-resource-aad-downloadable-genotypes-present-day-and-ancient-dna-data
Germany Early Medieval sequencing data	Veeramah et al. (2018)	https://reich.hms.harvard.edu/allen-ancient-dna-resource-aad-downloadable-genotypes-present-day-and-ancient-dna-data
Hungary Langobard SNP enrichment and sequencing data	Amorim et al. (2018)	https://reich.hms.harvard.edu/allen-ancient-dna-resource-aad-downloadable-genotypes-present-day-and-ancient-dna-data
Denmark Viking sequencing data	Margaryan et al. (2020)	https://reich.hms.harvard.edu/allen-ancient-dna-resource-aad-downloadable-genotypes-present-day-and-ancient-dna-data
Software and algorithms		
OxCal 4.4	Bronk Ramsey (2009)	https://c14.arch.ox.ac.uk/oxcal.html
BWA version 0.7.15	Li and Durbin (2009)	https://bio-bwa.sourceforge.net/
contamMix version 1.0–12	Fu et al. (2013)	https://github.com/DReichLab/ADNA-Tools
ANGSD	Korneliussen et al. (2014)	http://www.popgen.dk/angsd/index.php/ANGSD
HaploGrep2	Weissensteiner et al. (2016)	https://haplogrep.i-med.ac.at/
MALT	Herbig et al. (2016); Vagene et al. (2018)	https://uni-tuebingen.de/fakultaeten/mathematisch-naturwissenschaftliche-fakultaet/fachbereiche/informatik/lehrstuehle/algorithms-in-bioinformatics/software/malt/
READ	Monroy Kuhn et al. (2018)	https://bitbucket.org/tguenther/read
smartPCA	Patterson et al. (2006)	https://reich.hms.harvard.edu/software
ADMIXTOOLS version 5.1	Patterson et al. (2012)	https://reich.hms.harvard.edu/software
Samtools	Li (2011)	http://www.htslib.org/download/
BEAST 2 v2.6.6	Bouckaert et al. (2019)	https://www.beast2.org/
hapROH version 0.1a8	Ringbauer et al. (2021)	https://pypi.org/project/hapROH/

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
bcftools/ROH	Narasimhan et al. (2016)	https://samtools.github.io/bcftools/howtos/roh-calling.html
IBDseq	Browning and Browning (2013)	https://faculty.washington.edu/browning/ibdseq.html
PHCPImpute	This paper	https://github.com/ShamamW/PHCPImpute (DOI: https://doi.org/10.5281/zenodo.7233296)
GLIMPSE v1.0.0	Rubinacci et al. (2021)	https://odelaneau.github.io/GLIMPSE/
ADMIXTURE version 1.3.0	Alexander et al. (2009)	http://dalexander.github.io/admixture/download.html
DATES	Chintalapati et al. (2022)	https://github.com/MoorjaniLab/DATES_v3600
HOPS	Hubler et al. (2019)	https://github.com/rhuebler/HOPS
Other		
Sr-Spec Resin (50–100 µm)	Eichrom Technologies	Cat# Sr-B25-S
gnomad	Karczewski et al. (2020)	https://gnomad.broadinstitute.org/

RESOURCE AVAILABILITY**Lead contact**

Further information and requests for resources or reagents should be directed to and will be fulfilled by the lead contact, Shai Carmi (shai.carmi@huji.ac.il).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The genotypes for the ancient Erfurt individuals have been deposited at <https://reich.hms.harvard.edu/datasets>. The sequencing reads have been deposited at the European Nucleotide Archive under accession number [PRJEB53475](https://doi.org/10.5555/PRJEB53475). All other ancient genomes and the Human Origins array data for modern populations are publicly available at <https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>. The modern Ashkenazi Jewish genomes are available at <https://ega-archive.org/datasets/EGAD00001000781>. The isotope data have been deposited at IsoBank under ID number 686.
- Original code of PHCP imputation has been deposited at <https://github.com/ShamamW/PHCPImpute> (DOI: <https://doi.org/10.5281/zenodo.7233296>) and is publicly available. All other software packages used in this study were previously published.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS**Archaeology**

The medieval Jewish cemetery of Erfurt served the Jews of Erfurt and the surrounding area ([Data S1](#), section 2). After the expulsion of the Jews from Erfurt in 1454, a barn and a granary were built on top of the cemetery. An archaeological rescue investigation was performed in 2013 during the conversion of the granary into a multi-story car garage ([Data S1](#), section 3).

The excavated region, of about 16 × 12 m, is located between the inner and outer city walls. This section was probably used for burial only after the construction of the outer wall, as before that time it was likely used for fortification. Given archaeological evidence that the outer wall was built in the second half of the 14th century ([Data S1](#), section 3), the excavated region was probably used by the second community (the one that returned to Erfurt five years after the 1349 pogrom).

The excavation identified 47 burials, which is likely a small fraction of the total number of graves in the cemetery ([Data S1](#), section 3). Our genetic study was performed in 2018, using only detached teeth, after receiving the approval of the Jewish community of Thuringia. We collected 38 teeth, one per individual, mostly molars ([Data S2](#), Table 1). In 2021, all skeletons were reburied in the Jewish cemetery of the 19th-century community ([Data S1](#), section 3).

METHOD DETAILS**Age at death estimation**

We estimated the age at death for non-adult individuals by assessing the developmental stage of the dentition ([Ubelaker, 1978](#)) and the length of the long bones ([Stloukal and Hanáková, 1978](#)). In adult individuals, we estimated the age at death by established

methods based on the stage of degeneration of the pubic symphysis face, the auricular joint face, the sternal rib ends, and additionally the cranial suture closure, as summarized e.g., in (Buikstra and Ubelaker, 1994). Adult age was determined in an individual when the epiphyses were fused.

Radiocarbon dating

We performed accelerator mass spectrometry radiocarbon dating on purified collagen extracted from tooth dentin at the Pennsylvania State University Accelerator Mass Spectrometry (AMS) laboratory using previously described methods (Kennett et al., 2017; Lohse et al., 2014). We calibrated dates using OxCal 4.4 (Bronk Ramsey, 2009) and the IntCal20⁴⁸ calibration curve (Reimer et al., 2020). We assessed sample quality using stable isotope analysis. We found that carbon-to-nitrogen ratios for all collagen samples fell between 3.19 and 3.24, within the range of 2.9–3.6 expected for good collagen preservation (van Klinken, 1999).

The estimated dates are reported in Data S2, Table 1 (95.4% probability intervals) and the full measurements are reported in Data S2, Table 3. Our first attempt at dating the father and daughter from Family B (I14904 and I13869, respectively) yielded results inconsistent with their relationship (95.4% CI 1266–1298 calCE for the daughter and 1288–1398 calCE for the father). We radiocarbon dated these samples again, obtaining the same result for the father (1297–1395 calCE). For the daughter, the second run suggested a range of dates consistent with the relationship (1278–1380 calCE), due to the appearance of a small post-1350 peak (Figure S1 C).

We note that we dated teeth, which form in childhood and early adulthood. Thus, all dates should not be interpreted as representing date of death, but instead a date during an earlier time of life.

Isotope analysis

To test the hypothesis that some Erfurt individuals were migrants, we selected enamel samples for $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ isotope analysis at the University of New Mexico Center for Stable Isotopes. Among samples with sufficient material, we considered 20 samples with coverage >200k SNPs, which were all confidently assigned to an Erfurt subgroup (EU/ME).

Enamel surfaces were cleaned with a rotary tool and a 200 μm endmill used to remove about 30 mg of enamel, avoiding any dentin. Samples were crushed in an agate mortar with UltraPure water (resistivity = 18.2 M Ω cm) to keep chips from shattering. About 15 mg of sample was transferred to a 2 mL centrifuge tube with about 0.1 M Acetic Acid and agitated for about 30 s using a VortexGenie to sufficiently mix the sample and solution. Samples were reacted for about 4 h, then rinsed to neutrality with UltraPure water and micro-centrifuged for 1 min at 3000 rpm between each rinse. After the final rinse samples were freeze dried overnight (about 12 h). Next, 7–8 mg of sample was weighed into 12 mL glass exetainers for analysis. An additional 1 mg aliquot of sample was separated and placed on the ThermoFisher Nicolet Summit FTIR diamond ATR for QC/QA analysis. Samples were analyzed on a Thermo Scientific Delta V IRMS with a GasBench carbonate device for $\delta^{13}\text{C}_{\text{enamel}}$ and $\delta^{18}\text{O}_{\text{enamel}}$.

The full results of the isotope analysis are reported in Data S2, Table 5, and the final values of $\delta^{13}\text{C}_{\text{enamel}}$ and $\delta^{18}\text{O}_{\text{enamel}}$ appear in Data S2, Table 1.

DNA extraction and sequencing

Of the 38 teeth, four did not have root material and were not further processed. For the remaining 34, we performed the following steps. In dedicated clean rooms at Harvard Medical School, we drilled into the roots of the teeth to obtain powder from cementum and dentin. We extracted DNA using a protocol meant to retain short molecules (Dabney et al., 2013); converted the DNA into individual barcoded and/or indexed double-stranded and single-stranded libraries treated to remove characteristic ancient DNA damage (Briggs and Heyn, 2012; Gansauge et al., 2017, 2020; Rohland et al., 2015); enriched for approximately 1.24 million SNPs (Fu et al., 2015) and mitochondrial DNA (Maricic et al., 2010); and sequenced the enriched and non-enriched libraries on Illumina instruments (Data S2, Table 2).

Bioinformatics and quality control

We merged read pairs, requiring at least 15 overlapping base pairs, and allowing no more than one mismatch if base quality was at least 20, and as many as three mismatches if base quality was less than 20, and chose the nucleotide of higher quality when we observed mismatches. We mapped the sequences to the human genome reference sequence hg19 (GRCh37, https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/) and the inferred mitochondrial ancestral sequence RSRS (Behar et al., 2012) using the *samse* command of *BWA* version 0.7.15 using parameters *-n* 0.01, *-o* 2, and *-l* 16500 (Li and Durbin, 2009). We removed duplicate molecules that mapped to the same start and stop positions and (for double-stranded libraries) that had the same molecular barcodes.

We analyzed the data to assess ancient DNA authenticity based on the following metrics (Data S2, Table 2). First, the ratio of Y to X + Y chromosome sequences: uncontaminated females should have a very low (<0.030) and males should have a high (>0.35) ratio in the type of data we produced. Second, the estimated rate of variation in the mitochondrial genome and the X chromosome in males at known polymorphisms: uncontaminated individuals should be consistent with very little variation. We estimated the degree of contamination with *contamMix* version 1.0–12 (Fu et al., 2013) for the mitochondrial DNA and *ANGSD* for the X chromosome (Korneliussen et al., 2014). Third, we required an appreciable rate of cytosine to thymine damage in the final nucleotide, as expected for genuine ancient DNA. We initially obtained DNA data for 34 individuals. However, one individual was represented at only 52 SNPs, and was omitted from all analyses.

We determined mitochondrial haplogroups using *HaploGrep2* (Weissensteiner et al., 2016). The procedure for Y-chromosome haplogroup determination is described in Supplementary Text S5 of (Lazaridis et al., 2022), using the YFull YTree v. 8.09 phylogeny (https://github.com/YFullTeam/YTree/blob/master/ytrees/tree_8.09.0.json), obtaining information about SNPs from ISOGG YBrowse (https://ybrowse.org/gbrowse2/gff/snps_hg38.csv; accessed Oct 18, 2020), lifting coordinates from hg38 to hg19 using liftOver and intersecting with the SNPs present in the v. 8.09 tree. The haplogroup calls were converted into letter-number haplogroup designations (e.g., J2a1) using the phylogeny of the International Society of Genetic Genealogy v. 15.73. We could not infer the haplogroup of four males due to insufficient coverage over the informative Y-SNP targets. The method we used to determine the terminal mitochondrial and Y chromosome lineages is described in [Data S1](#), section 4.

Pathogen DNA scan

We screened the Erfurt individuals for the presence of pathogens using *MALT* (Herbig et al., 2016; Vagene et al., 2018). We used a custom *RefSeq* genomic dataset containing bacteria, viruses, eukaryotes, and the human reference sequence GRCh38 to construct the *MALT* database using default parameters (Giffin et al., 2020), with an index step size of 6. For each Erfurt individual, we screened all merged, de-duplicated sequences, applying a minimum complexity filter with a threshold of 0.3. We ran *MALT* (version 0.3.8), using the parameters `-mode BlastN, -alignmentType SemiGlobal, -minPercentIdentity 0.85, -topPercent 1, -minSupport 1, -maxAlignmentsPerQuery 100`. We then screened the results using the *HOPS* workflow (Hubler et al., 2019) to determine whether there was evidence of authentic DNA from a set of 348 pathogens of interest ([Data S2](#), Table 4). We assessed authenticity using a standard three step screening pipeline considering (1) the edit distance distribution of all sequences that align to the pathogen of interest; (2) the presence of C-to-T (or G-to-A) sequence damage (which is characteristic of authentic ancient DNA); and (3) the edit distance distribution of the subset of aligned sequences that contain C-to-T damage (Hubler et al., 2019).

QUANTIFICATION AND STATISTICAL ANALYSES

Detecting genetic relatives

We used *READ* (Monroy Kuhn et al., 2018) to identify 1st- and 2nd-degree relatives. We identified five individuals with 1st-degree relationships ([Figure S1A](#)). Three individuals — I14850 (female), I14853 (male), and I14898 (female) — were all identified as 1st-degree relatives of one another, and we denoted them as Family A ([Figure 1B](#)). To determine their relationships, we first used information on age at death and mitochondrial DNA haplogroup. Individuals I14853 and I14898's estimated ages were 14–17 and 10–13 years old, respectively ([Data S2](#), Table 1), and hence they must have been siblings. I14850 was estimated to die in her 40s, and all three individuals shared the same mtDNA haplogroup. These constraints are consistent with I14850 being either the mother or a sibling who has survived to adulthood. To distinguish between these two possibilities, we calculated the mismatch rate along the genome between I14850 and I14853 (who had the higher coverage among the two sibs). We compared the mismatch rate to that of (i) unrelated individuals; and (ii) two sets of sequences sampled of the same individual. In (ii), we only used SNPs covered by at least two sequences, and we generated each duplicate by sampling a single sequence (without replacement) for each SNP. We calculated the mismatch rate in blocks of 20 Mb and normalized it to the number of SNPs. [Figure S1B](#) shows that the mismatch rate between I14850 and I14853 is consistently intermediate between that of unrelated and duplicate individuals, as expected for a parent-child relationship. We thus conclude that I14850 is the mother of I14853 and I14898. Two other individuals — I13869 (female) and I14904 (male) — were 1st-degree relatives and were designated as Family B. As they carried different mtDNA haplogroups, the only possible relationship is that of a father and a daughter.

Individual I14855 was found by *READ* to be a 2nd-degree relative of I14854 and I14848. The latter two individuals had a low mismatch rate but were not classified as 2nd-degree relatives, possibly due to their very low coverage (15k and 13k SNPs, respectively). I14854 and I14855 carried the same mtDNA haplogroup, but not I14848. The uncertainty ranges for the two identified relationships included a 1st-degree relationship and no relationship ([Figure S1A](#)).

Investigating genetic ancestry using PCA

We used *smartPCA* (Patterson et al., 2006), with the option "Isqproject", which enables projection of individuals with high missingness when the PCs were learned from modern individuals. The following populations from the Human Origins dataset were used: Armenian, Iranian, Turkish, Albanian, Bergamo, Bulgarian, Cypriot, Greek, Italian_South, Maltese, Sicilian, Italian_North, English, French, Icelandic, Norwegian, Orcadian, Scottish, BedouinA, BedouinB, Jordanian, Palestinian, Saudi, Syrian, Abkhasian, Adygei, Balkar, Chechen, Georgian, Kumyk, Lezgin, Ossetian, Jew_Ashkenazi, Jew_Georgian, Jew_Iranian, Jew_Iraqi, Jew_Moroccan, Jew_Tunisian, Jew_Libyan, Jew_Turkish, Jew_Yemenite, Basque, Spanish, Spanish_North, Druze, Lebanese, Belarusian, Croatian, Czech, Estonian, Hungarian, Lithuanian, Ukrainian, Canary_Islander, Sardinian, Finnish, Mordovian, Russian, and Polish.

We projected the Erfurt individuals on the PC space learned by the West-Eurasian populations. Down-sampling experiments showed that 50k or more SNPs provide reproducible assignment of individuals to Erfurt subgroups ([Data S1](#), section 5). We consequently excluded eight EAJ individuals with <50k SNPs from all PC analyses.

We also ran PCA with a larger sample size of MAJ, as follows. We merged the Human Origins dataset with whole-genomes of $n = 544$ modern AJ sampled in the USA and Israel (Lencz et al., 2018). These genomes were generated in Phase 2 of The Ashkenazi Genome Consortium (TAGC), and relatives were removed (Fridman et al., 2020). The merged dataset had about 470k SNPs. We used

the same Human Origins populations as above to learn the PC space, except that here we removed the Human Origins MAJ population. We then projected both EAJ and the TAGC MAJ on the PC space. Due to the large MAJ sample size, we plotted the positions of these individuals as a 2-dimensional kernel density plot (Figure S2A) using the function *stat_density_2days()* from the package *ggplot2* in R. The density was scaled to 1 using the argument *contour_var = "ndensity"*. To evaluate the effect of coverage on placement in PC space, we down-sampled a randomly selected subset of $n = 525$ MAJ individuals to match the set of SNPs covered by the EAJ genomes. Each of the 25 (non-low-coverage) EAJ genomes was used to match $525/25 = 21$ MAJ individuals. We then selected a single allele at random from each down-sampled MAJ individual and projected the MAJ and EAJ individuals as above (Figure S2B). See Data S1, section 5 for details on the PCA that included MAJ of Western and Eastern European origin.

Admixture modeling with ADMIXTOOLS

For the f_4 and *qpWave* analyses, we used the Human Origins dataset merged with whole-genomes of $n = 544$ modern AJ (Lencz et al., 2018). We used ADMIXTOOLS (Patterson et al., 2012) version 5.1 for running the analyses. We computed f_4 -statistics with default parameters. In the *qpWave* and *qpAdm* analyses, we only used transversion SNPs (about 110k) to avoid bias due to ancient DNA damage. In all population-level analyses, we (1) omitted one of each pair of first-degree relatives, keeping the individual with the higher coverage; and (2) included the low-coverage individuals (<50k SNPs), except in analyses that required the Erfurt-EU and Erfurt-ME group affiliation.

Detecting admixture with qpWave

We ran *qpWave* with the option "allsnps:YES", which means that in each f_4 test, the program uses all SNPs that were non-missing in all four populations, but a different set of SNPs can be analyzed for each underlying f_4 -statistic. We ran *qpWave* tests separately against reference ("right") European and reference Middle Eastern populations. In our *qpWave* analyses with European reference populations, we chose the right populations as modern populations that represent main European ancestries: Russian, Norwegian, French, Spanish, Bulgarian, and Italian_North, with Primate_Chimp as an outgroup (first right population). Middle Eastern populations are closely related, and we therefore chose the right populations as follows. We ran f_4 tests of the form $f_4(\text{EAJ}, \text{MAJ}; \text{ME1}, \text{ME2})$ where ME1 and ME2 represent all possible pairs of populations from: BedouinA, BedouinB, Palestinian, Lebanese, Syrian, Jordanian, Egyptian, Saudi, and Druze. We used populations that were involved in tests with p value < 0.05: Druze, Lebanese, Jordanian, and BedouinA. We again used Primate_Chimp as an outgroup. In tests with South-Italians, we used individuals of Sicilian and Italy_South together as one group.

Inferring admixture models with qpAdm

Here too, we used the option "allsnps:YES". The reference populations (right populations) for the *qpAdm* analyses were: Mbuti, Ami, Basque, Biaka, Bougainville, Chukchi, Eskimo_Naukan, Han, Iranian, Ju_hoan_North, Karitiana, Papuan, Sardinian, She, Ulchi, and Yoruba. Mbuti was used as the outgroup (provided to ADMIXTOOLS as the first in the list of reference populations) in all analyses. In robustness tests, we replaced Mbuti with Ami as the outgroup. As in the *qpWave* analyses, in models with South-Italians we used individuals of Sicilian and Italy_South together as one group. In models with ancient Germans, we used individuals from (Veeramah et al., 2018), not including individuals with elongated skulls or with Southern European ancestry. The ancient Levant (Canaanite) individuals (Agranat-Tamir et al., 2020) were from Bronze-Age Megiddo (Megiddo_MLBA) and the ancient Rome individuals (Antonio et al., 2019) were from Late Antiquity (Italy_LA.SG) and Imperial Rome (Italy_Imperial.SG).

For the analyses at the individual level, we used all SNPs, as the coverage of many individuals was already low. To guarantee that using all SNPs did not bias the results, we repeated the analyses at the population level with all SNPs instead of just transversions and verified that the results remained qualitatively unchanged (Data S1, section 7). We included first-degree relatives in the individual-level analyses, but omitted the low-coverage individuals (<50k SNPs). For individuals for which the Eastern-EU ancestry proportion was inferred to be negative (Figure 3), we re-ran *qpAdm* with only Southern EU and Middle Eastern sources.

Inferring admixture time with DATES

We attempted to estimate the time of Eastern European gene flow into Erfurt-EU using DATES (Chintalapati et al., 2022) (Data S1, section 8). However, our simulations showed that DATES estimates in our setting were unreliable (Data S1, section 8).

Maximal post-medieval gene flow into AJ

The *qpWave* test comparing EAJ and MAJ gave $p = 0.15$ (Table S1), consistent with these groups being a clade with respect to reference European populations. To estimate the maximal degree of post-medieval gene flow into AJ that would still be consistent with this result, we used simulations. Specifically, we simulated AJ groups that have experienced increasing magnitudes of admixture with Eastern European sources. We then tested, using *qpWave* (with respect to the same European populations as described above), which simulated group is still inferred as a clade with modern AJ. The Eastern European sources included 30 individuals from Belarusians, Lithuanians, Ukrainians, and Russians. These individuals were removed from the subsequent *qpWave* analyses. The AJ source was 30 MAJ individuals that were sequenced in Phase 1 of The Ashkenazi Genome Consortium (Carmi et al., 2014a) and were not included in the MAJ dataset that was used for the original *qpWave* analyses. For each admixture scenario, we simulated 30 individuals using the procedure described in Data S1, section 6. We used admixture times of $G = 5, 10, 15, 20$ generations.

The *qpWave* *p* values reported in Table S2 show that the maximal proportion of the AJ gene pool that could have been replaced by East-EU admixture and still remain consistent with being a clade with MAJ is about 2%–4%. In our simulations, all gene flow was assumed to occur over a single generation. With continuous gene flow, a replacement of a proportion *m* of the gene pool per generation over 20 generations would lead to a total replacement of $1 - (1 - m)^{20}$ of the total ancestry. Equating to 4% and solving for *m* gives $m = 1 - (1 - 0.04)^{1/20} = 0.2\%$.

Correlation between EAJ grave location and group affiliation

We used the Mantel test to investigate the correlation between group affiliation and the distances between the graves in the cemetery. We calculated the distances based on the approximate coordinates of the skeletons' heads in the cemetery map (Figure 1A). For group affiliation, we set the distance between Erfurt-ME and Erfurt-EU to 1 and the distance within each group to zero. We used the function *mantel.rtest* from the *ade4* package in *R*. We also ran the Mantel test by replacing the group affiliation distance with the distances in the PC1-PC2 space, based on the PCA of Figure 2 of the main text and using Euclidean distances.

Aligning the mtDNA sequences of modern and ancient K1a1b1a carriers

To obtain the sequences of modern K1a1b1a carriers, we used *n* = 544 genomes of Phase 2 of The Ashkenazi Genome Consortium (after removing related individuals) (Lencz et al., 2018). We first discarded sites with GQ < 40 and indels. Some sites had heterozygous genotypes (possibly due to heteroplasmy). We encoded these sites as having the alternate allele if the allele was observed at more than 50% of the reads and otherwise encoded them as having the reference allele. We used *bcftools consensus* (Li, 2011) with parameter -H A using the rCRS reference sequence to generate an alignment of the sequences of all individuals. We then used *HaploGrep2* (Weissensteiner et al., 2016) to call mitochondrial haplogroups and focused on the 107 carriers of K1a1b1a. For the ancient individuals, haplogroups were called as described above and we only considered the 11 K1a1b1a carriers. Their median coverage was 304x (range: 47–447x; Data S2, Table 2). We discarded indels and otherwise performed no additional filtering. We again used *bcftools consensus* to generate an alignment of the sequences, but this time with the RSRS reference sequence (Behar et al., 2012).

Estimating the time to the most recent common ancestor (TMRCA) of the K1a1ba1 carriers

To estimate the TMRCA of the K1a1ba1 lineage, we used a Bayesian coalescent analysis, as implemented in *BEAST 2* (v2.6.6) (Bouckaert et al., 2019). The input was the mtDNA sequence alignment of the 11 EAJ carriers and the 107 MAJ carriers. We set the dates of the EAJ individuals to 650 years ago. For modeling mutations, we used the HKY model with Gamma distributed rates (four categories), and the strict clock model. For the population size prior, we used the coalescent Bayesian skyline with four segments. We also ran *BEAST* with a coalescent exponential population prior, but convergence was poor and we did not further consider this prior. For the skyline model, we ran 10 chains for 100 million steps each. For each chain, we used 10% of the steps as burn-in and recorded the parameters every 100,000 steps, resulting in a total of 9000 samples. All other parameters were set to their default values. We combined the samples from all chains using *LogCombiner* and visualized the results using *Tracer* and *FigTree*.

Detecting runs of homozygosity in the ancient genomes

To identify runs of homozygosity (ROH) in the ancient genomes, we used the Python package *hapROH* version 0.1a8 (Ringbauer et al., 2021), using 5,008 global haplotypes from the 1000 Genomes Project as the reference panel and the pseudo-haploid genotypes of the ancient genomes as input. As recommended for datasets with genotype data for 1240k SNPs (Ringbauer et al., 2021), we applied our method to 16 Erfurt individuals covered in at least 400k SNPs and called ROHs longer than 4 cM. These individuals include two parent-child pairs. We used the default parameters and post-processing of *hapROH*, which are optimized for this data type (described in detail in (Ringbauer et al., 2021)). We report the sum of the lengths of all ROHs longer than 4 cM in each individual in Data S2, Table 1.

We manually inspected the ROH results by examining the positions of putatively heterozygous sites (Figure S4). Given that the called genotypes are pseudo-haploid (representing the allele of a randomly selected read), information on heterozygosity requires the original sequencing reads. For each SNP, we obtained the read counts for each allele from the processed BAM files using *samtools mpileup* (Li, 2011). Due to the low coverage, high quality diploid genotype calls could not be made. We identified putatively heterozygous sites as SNPs with at least one read supporting each of the reference and alternate alleles. While not all heterozygous sites could be detected using this approach (due to the low coverage), their depletion at inferred ROH segments is evident (Figure S4). There was no correlation between ROH counts and coverage (Data S1, section 10).

To contextualize the ROH results, we repeated the ROH analysis in other ancient populations. We used the Allen Ancient DNA resource (V50.0, Oct 10 2021; <https://reich.hms.harvard.edu>) and extracted previously published genomes from individuals who lived in the past ≈2000 years. We downloaded the data in *eigenstrat* format. As above, we used *hapROH* with the recommended settings (Ringbauer et al., 2021) and applied it to pseudo-haploid genotypes. The populations we studied included Hungary Lango-bard (Amorim et al., 2018); Germany Early Medieval (Veeramah et al., 2018); Italy Imperial, Late Antiquity, and Medieval (Antonio et al., 2019); and Denmark Viking (Margaryan et al., 2020). The sum of the lengths of ROH segments in individuals belonging to these populations is shown in Figure S4.

Ashkenazi founder alleles in Erfurt Jews

We defined founder alleles as those having minor allele frequency $>0.5\%$ in MAJ and $<0.01\%$ in non-Finnish Europeans (using *gnomAD* [Karczewski et al., 2020]). To exclude variants that may have a Middle Eastern source (which is not covered by *gnomAD*), we used 221 Middle Eastern individuals from the Human Origins dataset (Data S1, section 6, Table 1 therein), and excluded alleles that appeared more than once among these individuals. Finally, we removed SNPs that were genotyped in fewer than three EAJ individuals (after removing first-degree relatives), leaving a total of 216 SNPs.

Among the EAJ individuals, we excluded all children from families A and B, as well as one individual who was not genotyped in any of the founder SNPs, and was thus uninformative. Within the remaining 29 EAJ individuals, we detected 15 founder alleles in 11 individuals (Data S2, Table 1). All variants except one appeared in just a single individual. The remaining variant appeared in three individuals, and thus the total number of copies of these alleles was 17.

Binomial simulations of founder allele counts

To determine whether the number of founder alleles present in EAJ is as expected if EAJ had already experienced the AJ bottleneck, we used binomial simulations. In each run and for each founder allele, we drew an EAJ allele count as a binomial variable with n equals to the number of EAJ individuals that were genotyped in that SNP, and p equals to the allele frequency in MAJ (based on *gnomAD*). Note that we used the number of EAJ individuals (and not twice the number) as our genotypes are pseudo-haploid. In each run, we recorded the number of alleles (out of 216) that appeared in at least one individual. The distribution of the number of observed alleles across 10,000 runs is shown in Figure S5C. The results remained similar when modeling reference allele bias (Data S1, section 11).

Detecting IBD and ROH segments in modern genomes

We used sequencing data of $n = 637$ MAJ from the two phases of The Ashkenazi Genome Consortium (Carmi et al., 2014a; Lencz et al., 2018), after relatives and duplicates were removed (Fridman et al., 2020). To detect IBD sharing in MAJ, we used *IBDseq* (Browning and Browning, 2013) with default parameters. We detected runs of homozygosity in $n = 574$ genomes (Phase 2 data) using the software *bcftools/ROH* (Narasimhan et al., 2016). To meaningfully compare ROH in ancient and modern individuals, we first down-sampled the modern data to the 1240k SNPs used in the ancient DNA analysis. We used the same sex-averaged genetic map as in the analysis of the ancient data (provided with the *hapROH* reference panel), and set the hidden Markov model transition probabilities to parameters optimized for 1240k data (see (Ringbauer et al., 2021); $\text{toA} = 6.7\text{e-}8$, $\text{toHW} = 5\text{e-}9$). To obtain allele frequencies, we used the diploid data of the 574 modern individuals (using the VCF field MLEAF). As in the ancient data, we merged gaps of length up to 0.5 cM between two ROHs (both at least 2 cM long, one at least 4 cM). We manually inspected the ROH results by determining whether inferred ROH segments are depleted of heterozygous sites (Figure S4), considering only bi-allelic SNPs in the 1240k set. We finally removed nine individuals with total ROH length >50 cM (in segments longer than 4 cM), as these individuals likely have closely related parents.

Demographic inference using IBD and ROH segments

Complete details are provided in Data S1, section 12. Briefly, we searched for a model that would fit the total count of IBD segments between all pairs of individuals in 11 length bins between 4 and 15 cM (Figure 4B). The model of Figure 4A is characterized by five parameters. For each proposed set of values for these parameters, we computed the expected number of IBD segments based on theory developed in (Carmi et al., 2014b; Palamara et al., 2012; Ringbauer et al., 2017). For each length bin, and conditioning on the TMRCA, the theory provides the expected number of segments shared between two haplotypes in a chromosome of a given length. We then summed over all possible values of the TMRCA, with the probability of each TMRCA computed using coalescent theory given the currently assumed demographic model. We then summed over all chromosomes and multiplied by $\binom{2n}{2} - n$ to sum over all (non-ROH) pairs of haplotypes coming from n diploid individuals. The likelihood of the count in each length bin was computed by assuming a Poisson distribution for the number of segments with mean equal to the theoretical expectation. Finally, we constructed a composite likelihood by multiplying the likelihood across length bins. We searched for the parameters that maximize the log of the composite likelihood using the *R* package *DEoptim* (Mullen et al., 2011). To compute confidence intervals, we used parametric bootstrap. To generate each bootstrap replicate, we drew a new count for the number of segments, for each length bin, as a Poisson variable with mean equals to the real count.

When using ROH to infer the parameters of the demographic models, we used the same approach, based on counts of ROH segments across length bins, except for the following differences. First, the number of haplotype pairs was n . Second, we assumed no coalescence in the past two generations (reflecting no sib mating). Third, for modeling ROH in EAJ, we considered segments of length up to 40 cM. We assumed that the EAJ genomes were sampled 26 generations ago, and computed the coalescence probabilities accordingly. To model consanguinity, we assumed that a fraction of the individuals are children of couples who are first cousins. To estimate the parameters of a model with a narrower bottleneck than expected based on the modern data, we estimated N_a and N_b using ROH in EAJ, after fixing all other parameters to their values from Figure 4A. To infer a longer bottleneck, we estimated T_b using ROH in EAJ, after fixing all other parameters. In both cases, we then used the modern IBD data to re-estimate d and N_c .

To estimate demographic parameters using both modern IBD and ancient ROH, we computed their expected counts based on the same demographic model, accounting for the sampling of the ROH segments 26 generations ago. We then computed the log-(composite) likelihood per haplotype pair separately for IBD and ROH, and computed the joint log-likelihood as their sum. This guarantees that each data type contributes roughly equally to the likelihood. Finally, to estimate the parameters of the two-population model, we used the same approach as above, except that we computed the coalescence probabilities taking into account that two lineages can coalesce during the period of population split only if both descended from the same sub-population.

We used parametric bootstrap for model selection, using simulations to generate the empirical null distribution of the increase in log-likelihood when fitting a two-population model to data simulated under a single-population model (Data S1, section 13). Specifically, we simulated segment counts under the single-population model as Poisson variables with means equal to the theoretical expectations. We then compared the increase in the log-likelihood when fitting the data to a two-population (vs a single-population) model to the increase observed in the real data (Data S1, section 13).

Defining pathogenic founder variants

We started with Supplementary Table 4 from an earlier paper analyzing AJ sequencing data (Carmi et al., 2014a). We excluded 11 variants that were not present in modern AJ (based on *gnomAD*) or had higher frequency in other populations in *gnomAD*, leaving 62 variants. Among them, 47 were present in our reference panel and could be imputed. None of the pathogenic variants appeared in the list of AJ founder alleles defined above, as most of them were not enriched. Of the enriched variants, four had frequency >0.01% in Europeans, one was not genotyped in the modern Middle Eastern individuals, and one did not appear in *gnomAD*.

Imputation using pseudo-haploid ChromoPainter (PHCP)

We developed a method for imputation of pseudo-haploid genotypes based on the previously reported PHCP model (Agranat-Tamir et al., 2020; Wasik et al., 2021). Briefly, PHCP (Pseudo-Haploid ChromoPainter) is an extension of the ChromoPainter model (Lawson et al., 2012), which itself is based on the Li-Stephens (Li and Stephens, 2003) hidden Markov model (HMM). PHCP models a target ancient sequence as a mosaic of modern haplotypes (“donors”). For each SNP, the hidden state of the HMM is a pair of modern haplotypes from which the target is “copied.” The observed (haploid) ancient allele is assumed to derive from each of these two haplotypes with equal probability. Transitions between donor haplotypes along the target sequence are assumed to be due to ancestral recombinations, and emissions model the combined effect of mutations, gene conversion events, and genotyping and other errors that lead to imperfect copying of the donor haplotypes. The full transition and emission probabilities were previously described in (Agranat-Tamir et al., 2020), where we used the population labels of the inferred donors of each target to learn about the target’s ancestry composition.

To impute an ancient target genome, we used the forward-backward algorithm. We computed, for each SNP of the target, the marginal posterior probability of each possible pair of donor haplotypes. For SNPs in the reference panel that were not enriched or were not covered in the target, we assigned to the target the marginal probabilities of the nearest covered SNP (in cM). For each SNP, we divided all haplotype pairs into three classes, based on the diploid genotype they imply for the target (AA/AB/BB). For each of the three possible genotypes, we defined their marginal probability as the sum of the marginal probabilities of all pairs of donors in that genotype’s class. In downstream analyses, we used for each SNP the most likely genotype.

To impute the EAJ genomes, we used $n = 702$ MAJ genomes (both phases of the Ashkenazi Genome Consortium (Carmi et al., 2014a; Lencz et al., 2018); without removing relatives or other individuals). We considered only autosomal chromosomes. We set the parameters for PHCP to $N_e = 64.57$, $\theta = 0.0014$ (Agranat-Tamir et al., 2020). To reduce running time, we used, for each chromosome, a set of 200 donor haplotypes that were most informative for that chromosome. The donor haplotypes were ranked based on the number of SNPs where they share an alternate allele with the target (Wasik et al., 2021) (and scaled by the number of SNPs where they have an alternate allele, as in (Wasik et al., 2021), although we used all available sites and not just the rare variants). The PHCP imputation software is available at <https://github.com/ShamamW/PHCPImpute>.

Imputation using GLIMPSE

We used GLIMPSE (v1.0.0) (Rubinacci et al., 2021) with $n = 702$ MAJ individuals as the reference panel (as for PHCP). Diploid genotype calls were generated using *bcftools mpileup* (v1.10.2) (Li, 2011). We imputed all autosomal bi-allelic SNPs and indels in MAJ. However, genotype likelihoods were used only for bi-allelic SNPs (generated by *mpileup*) as input to build the phasing and imputation model. Indels were imputed without genotype likelihood information, due to their more severe reference bias.

Analyzing the pathogenic founder variants detected in Erfurt Jews

After imputing the EAJ genomes with PHCP and GLIMPSE, 16 pathogenic variants were either directly genotyped or successfully imputed (Data S2, Table 6). We defined a set of high-confidence pathogenic variants present in EAJ based on the following criteria: (i) the PHCP probability for having at least one alternate allele was >97% and (ii) the alternate allele was detected by GLIMPSE with probability >50%. One variant (*CLRN1*, NP_001182723.1:p.N48K) was detected by PHCP with probability >97% in an individual that was not imputed with GLIMPSE due to its low coverage (Data S2, Table 6); we considered this variant as high-confidence. We also

defined a set of low-confidence pathogenic founder variants as follows: (i) *PHCP* marginal probability >97% and *GLIMPSE* probability <50%; or (ii) *PHCP* probability in the range 50%–97%; or (iii) *PHCP* marginal probability <50% and *GLIMPSE* probability >50% (Data S2, Table 6).

For each high-confidence variant detected in EAJ, we determined whether the corresponding gene is present in pre-conception carrier screening (PCS) panels. We considered four PCS sequencing-based panels aimed at the AJ population: Genpath (<https://www.genpathdiagnostics.com/hcp/womens-health/carrier-screening/ashkenazi-jewish-cancer-screening/>), SEMA4 (<https://sema4.com/elements/expandedcarrierscreen/>), fulgent (<https://www.fulgentgenetics.com/beacon-ashkenazi-jewish-female-carrier-screening>) and Baylor genetics (<https://www.baylorgenetics.com/geneaware/>). All data were downloaded in November 2020. For each variant, we indicate in Table 1 the number of panels in which the gene is included.

Supplemental figures

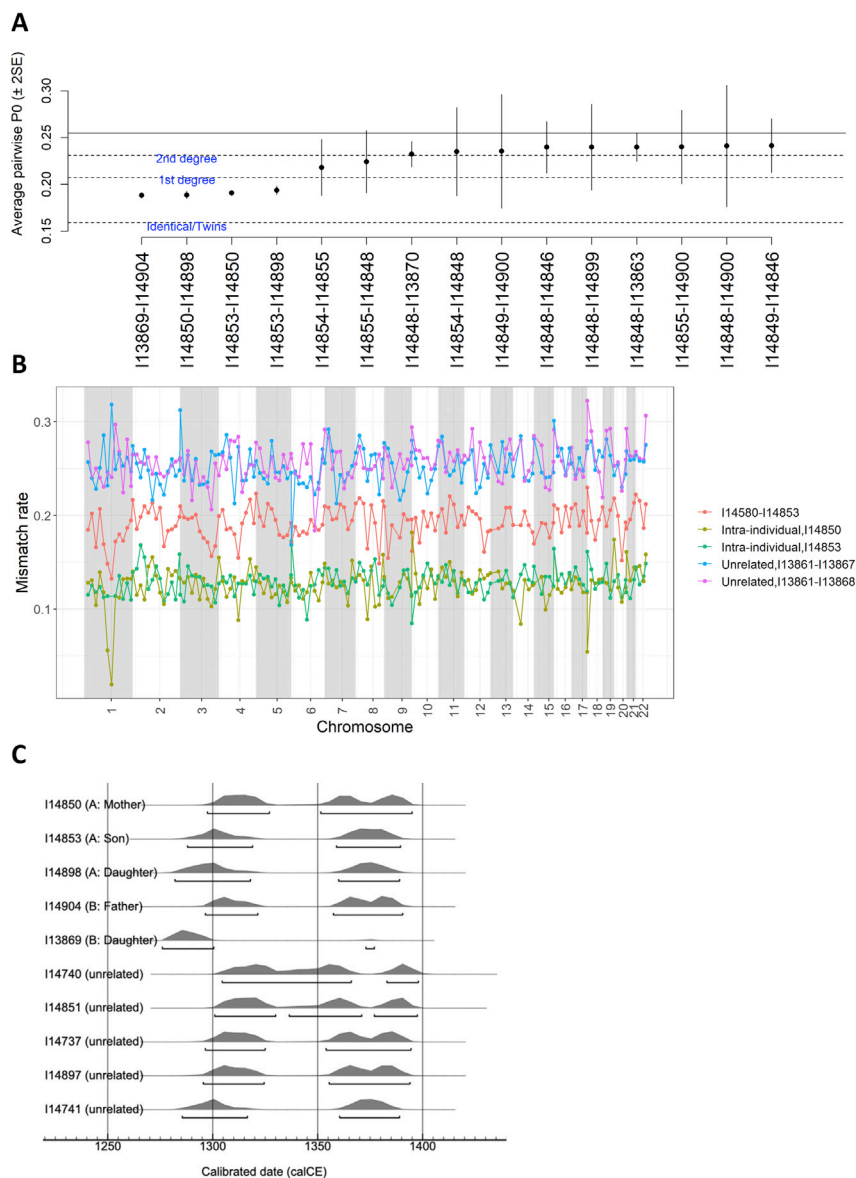


Figure S1. Genetic families and radiocarbon dating, related to Figure 1

(A) The output of *READ* for the first 15 pairs of EAJ individuals with the lowest proportion of non-matching alleles (P0). Each vertical line corresponds to a single pair. Seven of the pairs have a point estimate of having either a first- or a second-degree relationship. The y axis shows the mean P0 (across 1Mb genomic windows) between pairs of individuals. The horizontal solid line corresponds to the median P0 in the entire Erfurt sample, including pairs that are not shown. The horizontal dashed lines correspond to the cutoffs for first-degree relatives, second-degree relatives, and unrelated individuals (Monroy Kuhn et al., 2018). The vertical lines for each pair represent two standard errors of the mean (across genomics windows). Two pairs of individual, I14855 and I14854, and I14855 and I14848, were estimated to be second-degree relatives, although the confidence intervals also include a first-degree relationship and no relationship. The value of P0 for I14854 and I14848 was slightly above the cutoff for a second-degree relationship. All three samples had low coverage (<40 k SNPs). (B) The mismatch rate along the genome (in blocks of 20 Mb) between I14850 and I14853 is shown in red. For comparison, we show the mismatch rate for two pairs of unrelated individuals (blue and pink; legend) and the mismatch rate between two sets of randomly selected reads from the same individual (I14850 or I14853 in yellow and green, respectively). Data from blocks with fewer than 100 covered SNPs is not shown. The mismatch rate between I14850 and I14853 is intermediate between that of the unrelated individuals and the within-individual comparisons, indicating that they share exactly one chromosome along their entire genome, as

(legend continued on next page)

expected for a parent-child relationship. (C) The output of OxCal for ten samples that underwent radiocarbon dating. For each individual, the underlying bars denote intervals that have a cumulative probability of 95.4%. The estimated dates are almost equally likely to pre-date or post-date the 1349 pogrom. One individual — I13869 — was inferred to be much more likely to date to the end of the 13th century. However, there was a small peak in the late 14th century, and this individual is the daughter of I14904 (Figure 1B), who is more likely to date to the 14th century. This implies that I13869 might also date to the 14th century, despite this event having a smaller probability.

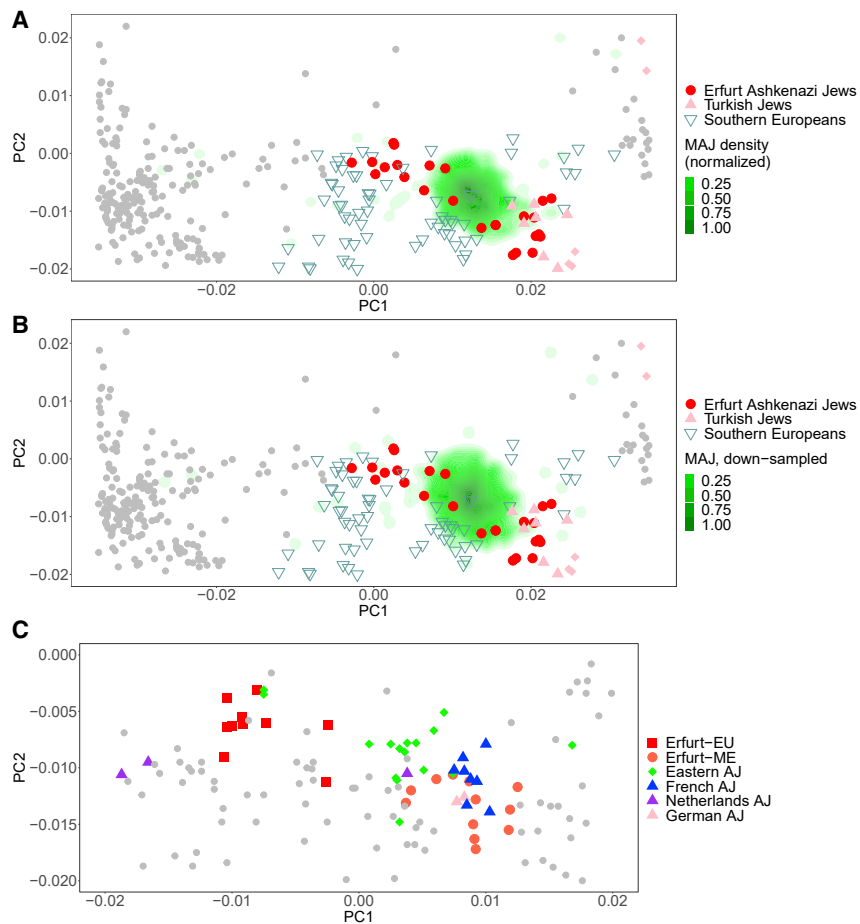


Figure S2. PCA plots of modern and ancient AJ, related to Figure 2

We learned the PCs using all West-Eurasian individuals from the Human Origins dataset. The figure shows just the subset of the space relevant for within-AJ structure. (A) Projection of a large modern AJ sample onto PC space. The plot is similar to the inset of Figure 2 of the main text, with two differences. (1) We projected both Erfurt and modern AJ individuals. (2) We did not include the seven modern AJ samples that were part of the Human Origins dataset. For modern AJ, we used The Ashkenazi Genome Consortium (TAGC) dataset ($n = 544$), down-sampled to the approximately 470k Human Origins SNPs. As the modern AJ sample is very large, we do not plot individual points but rather their density (STAR Methods). As in Figure 2, the Erfurt genomes show higher variability on the PC1 axis (the European/Middle Eastern cline) than modern AJ. (B) To demonstrate that the higher variability in EAJ is not due to their lower coverage compared to MAJ, we down-sampled each of the TAGC individuals to match the SNPs covered in an EAJ genome. We implemented this by arbitrarily ordering (randomly selected) 525 MAJ and the 25 non-low-coverage EAJ genomes, and sequentially matching MAJ and EAJ samples, cycling over the EAJ samples until covering all modern genomes. For each down-sampled MAJ genome, we further used a single, randomly-selected allele. Down-sampling did not qualitatively change the results. (C) To test whether the two subgroups of Erfurt correspond to modern AJ of Eastern European or Western European origin, we merged the Erfurt data with that of Behar et al. (2013). The merged dataset included about 246k SNPs. We projected both EAJ and MAJ onto the PC plane. The results show that French and German MAJ overlap with Erfurt-ME and that Eastern MAJ have more EU ancestry, although most of them still cluster primarily next to Erfurt-ME.

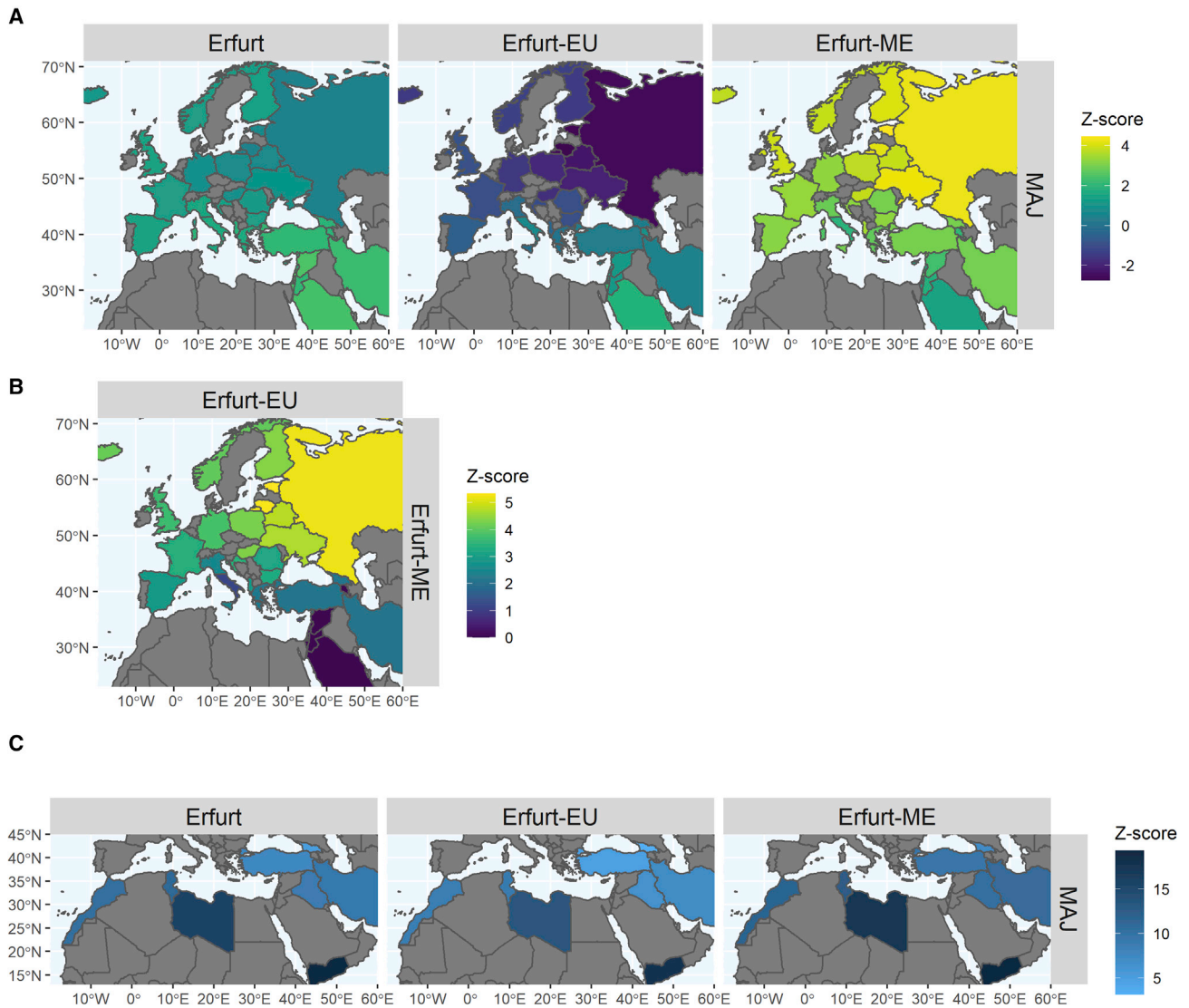


Figure S3. Results of the f_4 -statistics tests, related to Figure 3

(A) Tests of the form $f_4(\text{MAJ}, \text{EAJ}; X, \text{chimp})$, where X represents any non-Jewish West-Eurasian population. Each country on the map was colored based on the Z score for deviation from zero of the f_4 -statistic when replacing X with the local population in the Human Origins dataset. Gray represents countries that were not tested. In the middle and right columns, EAJ were replaced with Erfurt-EU and Erfurt-ME, respectively. (B) Tests of the form $f_4(\text{Erfurt-EU}, \text{Erfurt-ME}; X, \text{chimp})$. Here too, X represents any non-Jewish West-Eurasian population. (C) Tests of the form $f_4(\text{MAJ}, X; \text{EAJ}, \text{chimp})$, where X represents Jewish non-Ashkenazi populations. The location of each Jewish population on the map is represented by its origin in the diaspora. In the middle and right columns, EAJ were replaced with Erfurt-EU and Erfurt-ME, respectively. In all analyses, the Z score was positive and >3 , indicating that MAJ is the closest Jewish population to EAJ.

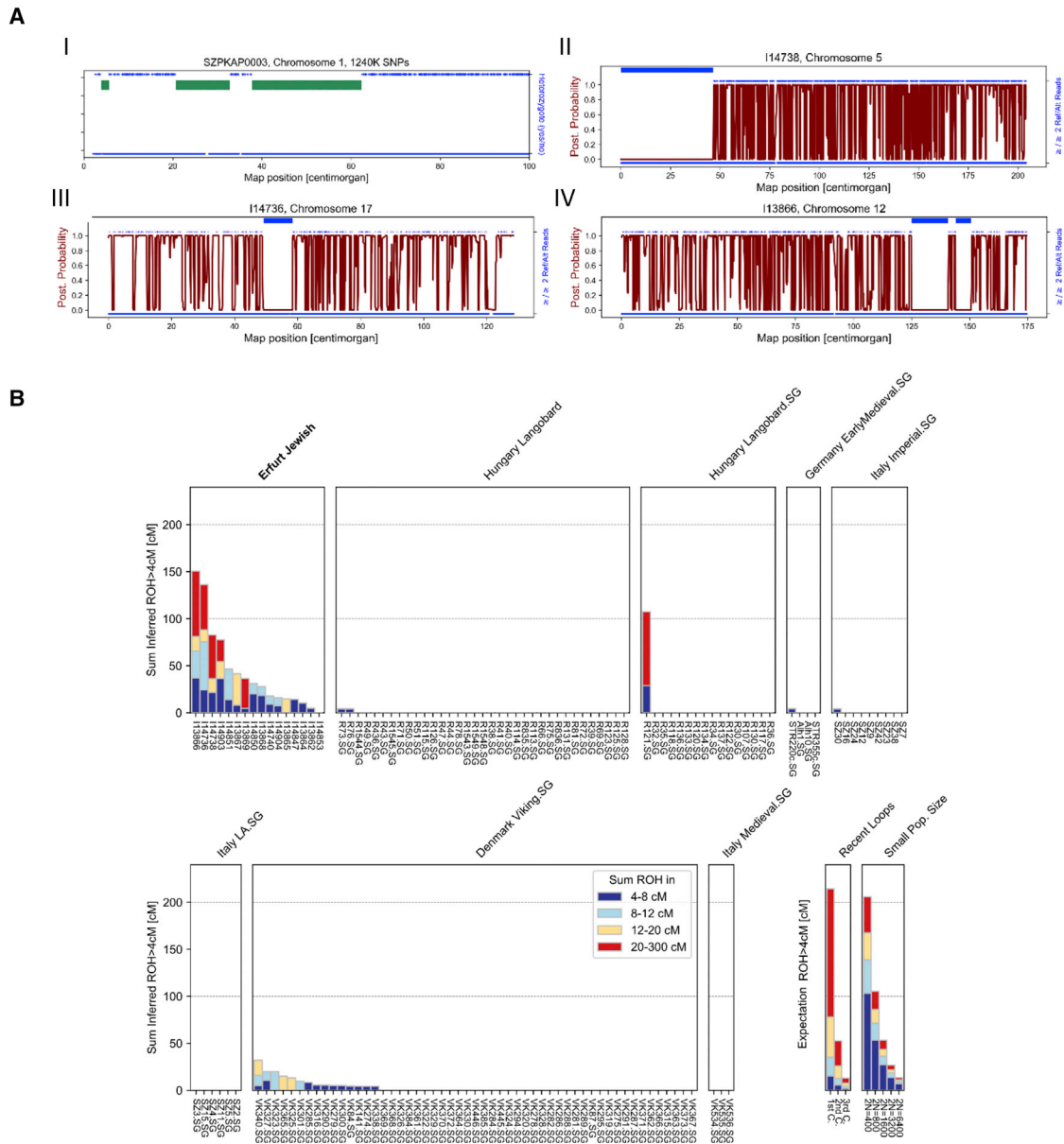


Figure S4. Visual inspection of runs of homozygosity (ROH) segments in modern and ancient genomes and a comparison to ancient European populations, related to Figure 4

(A) Examples of inferred ROH segments in one modern and three ancient AJ. Panel I demonstrates the inferred ROH segments (green bars, called with *bcftools/ROH*; STAR Methods) along a subset of chr1 in one modern AJ individual. We considered only bi-allelic SNPs included in the “1240” SNP panel. Blue dots at the top (bottom) of the panel show the positions of heterozygous (homozygous) sites. The inferred ROH segments are depleted of heterozygous sites. Panels II–IV demonstrate the inferred ROH segments (blue bars, inferred using hapROH; STAR Methods) in three chromosomes from three Erfurt individuals. The red lines show the posterior (“post.”) probability estimated by hapROH that a SNP is in a non-ROH state given the data. The blue dots show sites that were covered by at least one read (STAR Methods). A dot is plotted at the top of the panel whenever the reads covered both alleles, which suggests heterozygosity. The inferred ROH segments are again depleted of these putatively heterozygous sites. (B) ROH levels across European populations from the past two millennia. Each bar represents one individual, and individuals are grouped by population labels. We show the sum of the lengths of ROHs in four length bins (see legend). On the bottom right, we demonstrate the expected sum of ROH lengths for individuals whose parents are close relatives (first, second, and third cousins; “recent loops”), as well as for individuals from a population of a given constant effective size (N is in number of diploid individuals; “small pop. size”). See (Ringbauer et al., 2021) for details. The Erfurt samples have substantially longer ROHs compared to all other populations. SG: shotgun sequencing. The Hungary Langobard data (SNP enrichment and SG) is from (Amorim et al., 2018). The Germany Early Medieval data is from (Veeramah et al., 2018). The Italy Imperial, Late Antiquity (LA), and Medieval data is from (Antonio et al., 2019). The Denmark Viking data is from (Margaryan et al., 2020).

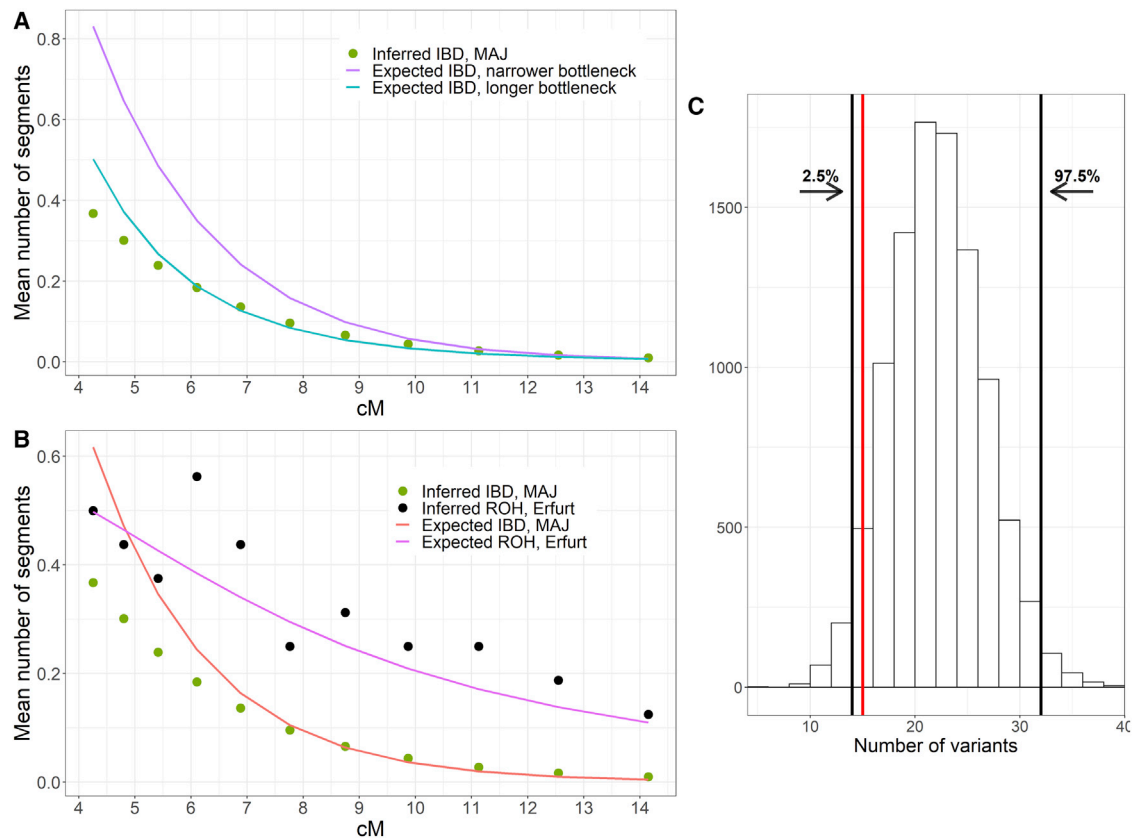


Figure S5. Evaluating demographic models and simulations of founder alleles, related to Figure 4

(A) and (B). Comparing predictions from demographic models for modern and Erfurt AJ vs observed data. (A) We plot the mean number of IBD segments (per pair of haploid genomes) across length bins in modern AJ in circles. The purple and teal lines show the expected counts (STAR Methods and Data S1, section 12) as predicted by models having a narrower or a longer bottleneck, respectively (Figure 4D; Table S5, models (E) and (F)), as compared to the model inferred using modern IBD. These models, in particular the narrower bottleneck model, do not fit the modern data well. (B) We plot the inferred number of IBD segments in MAJ and ROH segments in EAJ in green and black circles, respectively. The expectations based on the single-population joint-likelihood model, as described in Table S5, model (G), are shown in red and pink lines, respectively. The expected number of short IBD segments in MAJ is overestimated by the model. (C) Simulations for the expected number of AJ founder alleles in EAJ under modern allele frequencies. In each iteration and for each founder SNP, we drew a minor allele count as a binomial variable with n equals to the number of (pseudo-haploid) EAJ individuals that were genotyped in that SNP and p equals the allele frequency in MAJ (from gnomAD (Karczewski et al., 2020)). The figure shows the distribution of the simulated total number of founder SNPs with at least one minor allele across 10,000 runs. We show the 2.5- and 97.5-percentiles in black vertical lines, and the observed number in EAJ in a red line.