
No Evidence from Genome-wide Data of a Khazar Origin for the Ashkenazi Jews

DORON M. BEHAR,^{1,2†*} MAIT METSPALU,^{2,3,4†} YAEL BARAN,⁵ NAAMA M. KOPELMAN,⁶ BAYAZIT YUNUSBAYEV,^{2,7} ARIELLA GLADSTEIN,⁸ SHAY TZUR,¹ HOVHANNES SAHAKYAN,^{2,9} ARDESHIR BAHMANIMEHR,⁹ LEVON YEPISKOPOSYAN,⁹ KRISTIINA TAMBETS,² ELZA K. KHUSNUTDINOVA,^{2,7,10} ALENA KUSHNIAREVICH,² OLEG BALANOVSKY,^{11,12} ELENA BALANOVSKY,^{11,12} LEJLA KOVACEVIC,^{13,14} DAMIR MARJANOVIC,^{13,15} EVELIN MIHAILOV,¹⁶ ANASTASIA KOUVATSI,¹⁷ COSTAS TRIANTAPHYLIDIS,¹⁷ ROY J. KING,¹⁸ ORNELLA SEMINO,^{19,20} ANTONIO TORRONI,¹⁹ MICHAEL F. HAMMER,⁸ ENE METSPALU,³ KARL SKORECKI,^{1,21} SAHARON ROSSET,²² ERAN HALPERIN,^{5,23,24} RICHARD VILLEMS,^{2,3,25} AND NOAH A. ROSENBERG^{26*}

¹Molecular Medicine Laboratory, Rambam Health Care Campus, Haifa, Israel.

²Estonian Biocentre, Evolutionary Biology Group, Tartu, Estonia.

³Department of Evolutionary Biology, University of Tartu, Tartu, Estonia.

⁴Department of Integrative Biology, University of California, Berkeley, CA.

⁵Blavatnik School of Computer Science, Tel-Aviv University, Tel-Aviv, Israel.

⁶Porter School of Environmental Studies, Department of Zoology, Tel-Aviv University, Tel-Aviv, Israel.

⁷Institute of Biochemistry and Genetics, Ufa Research Center, Russian Academy of Sciences, Ufa, Russia.

⁸ARL Division of Biotechnology, University of Arizona, Tucson, AZ.

⁹Laboratory of Ethnogenomics, Institute of Molecular Biology, National Academy of Sciences, Yerevan, Armenia.

¹⁰Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa, Russia.

¹¹Vavilov Institute for General Genetics, Russian Academy of Sciences, Moscow, Russia.

¹²Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow, Russia.

¹³Institute for Genetic Engineering and Biotechnology, Sarajevo, Bosnia and Herzegovina.

¹⁴Faculty of Pharmacy, University of Sarajevo, Sarajevo, Bosnia and Herzegovina.

¹⁵Genos doo, Zagreb, Croatia.

¹⁶Estonian Genome Center, University of Tartu, Tartu, Estonia.

¹⁷Department of Genetics, Development and Molecular Biology, Aristotle University of Thessaloniki, Thessaloniki, Greece.

¹⁸Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA.

¹⁹Dipartimento di Biologia e Biotechnologie "Lazzaro Spallanzani," Università di Pavia, Pavia, Italy.

²⁰Centro Interdipartimentale "Studi di Genere," Università di Pavia, Pavia, Italy.

²¹Ruth and Bruce Rappaport Faculty of Medicine and Research Institute, Technion-Israel Institute of Technology, Haifa, Israel.

²²Department of Statistics and Operations Research, School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv, Israel.

²³Department of Molecular Microbiology and Biotechnology, George Wise Faculty of Life Science, Tel-Aviv University, Tel-Aviv, Israel.

²⁴International Computer Science Institute, Berkeley, CA.

²⁵Estonian Academy of Sciences, Tallinn, Estonia.

²⁶Department of Biology, Stanford University, Stanford, CA.

[†]These authors contributed equally to this work.

*Correspondence to: Doron M. Behar, Molecular Medicine Laboratory, Rambam Medical Center, Efron 9 St., Haifa 31096, Israel. E-mail: d_behar@rambam.health.gov.il. Noah A. Rosenberg, Department of Biology, 371 Serra Mall, Stanford University, Stanford, CA 94305-5020, USA. E-mail: noahr@stanford.edu.

Human Biology, December 2013, v. 85, no. 6, pp. 859–xx.

Copyright © 2014 Wayne State University Press, Detroit, Michigan 48201-1309

KEY WORDS: ANCESTRY, JEWISH GENETICS, POPULATION STRUCTURE, SINGLE-NUCLEOTIDE POLYMORPHISMS.

Abstract The origin and history of the Ashkenazi Jewish population have long been of great interest, and advances in high-throughput genetic analysis have recently provided a new approach for investigating these topics. We and others have argued on the basis of genome-wide data that the Ashkenazi Jewish population derives its ancestry from a combination of sources tracing to both Europe and the Middle East. It has been claimed, however, through a reanalysis of some of our data, that a large part of the ancestry of the Ashkenazi population originates with the Khazars, a Turkic-speaking group that lived to the north of the Caucasus region ~1,000 years ago. Because the Khazar population has left no obvious modern descendants that could enable a clear test for a contribution to Ashkenazi Jewish ancestry, the Khazar hypothesis has been difficult to examine using genetics. Furthermore, because only limited genetic data have been available from the Caucasus region, and because these data have been concentrated in populations that are genetically close to populations from the Middle East, the attribution of any signal of Ashkenazi-Caucasus genetic similarity to Khazar ancestry rather than shared ancestral Middle Eastern ancestry has been problematic. Here, through integration of genotypes from newly collected samples with data from several of our past studies, we have assembled the largest data set available to date for assessment of Ashkenazi Jewish genetic origins. This data set contains genome-wide single-nucleotide polymorphisms in 1,774 samples from 106 Jewish and non-Jewish populations that span the possible regions of potential Ashkenazi ancestry: Europe, the Middle East, and the region historically associated with the Khazar Khaganate. The data set includes 261 samples from 15 populations from the Caucasus region and the region directly to its north, samples that have not previously been included alongside Ashkenazi Jewish samples in genomic studies. Employing a variety of standard techniques for the analysis of population-genetic structure, we found that Ashkenazi Jews share the greatest genetic ancestry with other Jewish populations and, among non-Jewish populations, with groups from Europe and the Middle East. No particular similarity of Ashkenazi Jews to populations from the Caucasus is evident, particularly populations that most closely represent the Khazar region. Thus, analysis of Ashkenazi Jews together with a large sample from the region of the Khazar Khaganate corroborates the earlier results that Ashkenazi Jews derive their ancestry primarily from populations of the Middle East and Europe, that they possess considerable shared ancestry with other Jewish populations, and that there is no indication of a significant genetic contribution either from within or from north of the Caucasus region.

The Ashkenazi Jewish population has long been a subject of intense scholarly interest from the standpoint of such fields as anthropology, demography, history, medicine, and, more recently, genetics. As a result of the availability of high-throughput genetic data covering the whole of the human genome, the last several

years have seen major advances in the potential of population genetics to contribute to the study of population relationships and genetic origins (Cavalli-Sforza and Feldman 2003; Lawson and Falush 2012; Novembre and Ramachandran 2011). For the Ashkenazi Jewish population, genetic studies by several different investigators making use of a variety of genetic markers, genotyping platforms, analytical tools, and independently collected samples have converged on a series of remarkably similar results. First, it is possible to assess whether an individual has Ashkenazi Jewish ancestry, not only for subjects who identify as having exclusively Ashkenazi Jewish ancestors in recent generations but also, in many cases, for subjects who report only one or two Ashkenazi Jewish grandparents (Bauchet et al. 2007; Guha et al. 2012; Need et al. 2009; Price et al. 2008; Seldin et al. 2006; Tian et al. 2008). Second, Ashkenazi Jewish individuals share relatively long stretches of the genome, compared with both their genomic sharing with individuals from other populations and levels of within-population genomic sharing in these other populations (Atzmon et al. 2010; Campbell et al. 2012; Guha et al. 2012; Henn et al. 2012). Third, relatively little observable genetic difference exists between representatives of eastern and western Ashkenazi Jewish populations, suggesting that, genetically, the Ashkenazi Jewish population approximates a single large community (Guha et al. 2012). Fourth, considering the Ashkenazi Jewish population in relation to other populations, Ashkenazi Jews show the greatest genetic similarity to Sephardi Jews and, to a lesser extent, North African Jews (Atzmon et al. 2010; Behar et al. 2010; Campbell et al. 2012; Kopelman et al. 2009).

The issue of the geographic origin of the Ashkenazi Jews has been a source of considerable discussion, repeatedly addressed in the historical literature for over a century (see Efron this issue), and it has similarly not escaped the attention of population geneticists. Competing theories include a hypothesis that Ashkenazi Jews descend largely from the Khazar Khaganate, a conglomerate of mostly Turkic tribes that ruled about 1,400–1,000 years ago in what is now southern Russia with the capital Atil in the Volga delta on the northwestern banks of the Caspian Sea (Figure 1). According to this hypothesis, a portion of the Khazar population, among whom at least some had converted to Judaism, migrated north and west into Europe from their ancestral lands to become the ancestors of some or all of the Ashkenazi Jewish population. This hypothesis can be regarded as an alternative view to a perspective that the Ashkenazi Jewish population originated in the west rather than the east, with Jewish migrations north into Europe from Italy through France. Historical scholarship has provided considerable documentary evidence that Jews did indeed live along this latter route during the period of their entry into central Europe (Baron 1957; Ben-Sasson 1976; De Lange 1984; Mahler 1971), and the discussion can be viewed as an attempt to evaluate the relative magnitudes of possible eastern and western contributions.

The genetic perspective on Ashkenazi Jewish origins has pointed to a complex and multilayered construction of the Ashkenazi community giving rise to its contemporary shape. Most major genome-wide population-genetic studies of Ashkenazi Jews have detected evidence that the population has elements of

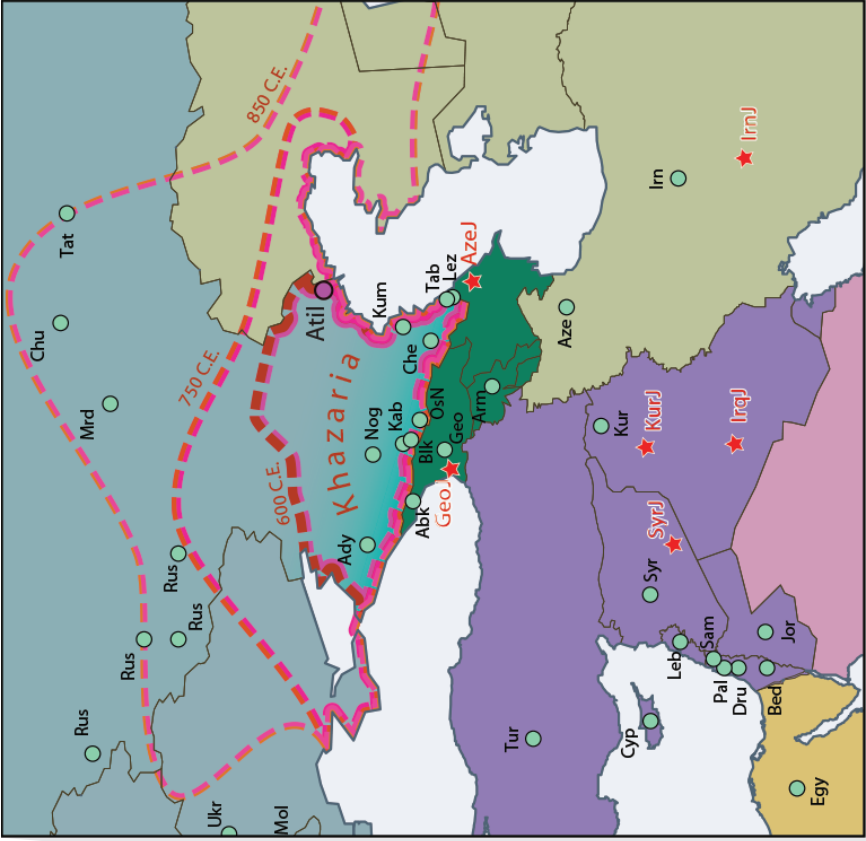


Figure 1. Geographical map of the populations included in this study. The crude borders of the Khazar Khaganate at three stages of its expansion, along with its capital at Atil, are shown. The Khazar Khaganate (~650–1,000 CE), one of the largest states of medieval Eurasia, extended from the Volga region in the north to the Northern Caucasus and Crimea in the south, and from present-day Ukraine in the west to the western borders of present-day Kazakhstan and Uzbekistan in the east (Golden et al. 2007). Abbreviations, as well as exact geographic locations, are detailed in Supplemental File S1.

ancestry both from Europe and from the Middle East (Atzmon et al. 2010; Behar et al. 2010; Campbell et al. 2012; Kopelman et al. 2009). Ashkenazi Jews have been placed intermediately between non-Jewish Europeans and non-Jewish Middle Easterners in a variety of analyses, including multidimensional scaling and principal components analyses (PCA), Bayesian clustering, and population trees. In one of the largest of these studies, encompassing 1,287 subjects from 14 Jewish and 69 non-Jewish populations, we found clear signatures of a Levantine ancestry component for Ashkenazi Jews, a component that was partially shared with other Jewish populations (Behar et al. 2010). These genome-wide results have supported earlier mitochondrial DNA and Y-chromosomal studies, which found that most lineages in the Ashkenazi Jewish population along the male and female lines trace primarily to the Levant, with the remaining lineages likely representing European contributions (Behar et al. 2003, 2004, 2006; Hammer et al. 2000, 2009; Nebel et al. 2001; Ritte et al. 1993; Santachiara Benerecetti et al. 1993).

Aware of uncertainties in the historical scholarship, genomic studies have also attempted to address the potential Khazar contribution to the Ashkenazi Jewish population, facing the fundamental problem that no contemporary population is identified, either by self-identification or by historians, as Khazars or Khazar descendants. For example, Behar et al. (2003) suggested that a specific R1a1 Y-chromosomal lineage, comprising 50% of the Ashkenazi Levites and observable in non-Jewish eastern Europeans, could represent either a European contribution or a trace of the lost Khazars. Similarly, based on autosomal markers, Kopelman et al. (2009), Need et al. (2009), and Guha et al. (2012) detected a small but measurable signal of similarity between Ashkenazi Jews and a sample of the Adygei population from the North Caucasus region. In each of these studies, the possible signal of Caucasus ancestry was relatively small compared to that observed from Europe and the Middle East. However, although no gross signal of Caucasus ancestry has been apparent, it is noteworthy that all of the major genetic studies were able to base their conclusions only on a limited representation of the Caucasus region, thereby leaving open the possibility that such a signal might be detectable in a larger Caucasus sample.

One recent study (Elhaik 2013), making use of part of our data set (Behar et al. 2010), focused specifically on the Khazar hypothesis, arguing that it has strong genetic support. This claim was built on a series of analyses similar to those performed in our original study that initially reported the data. However, the reanalysis relied on the provocative assumption that the Armenians and Georgians of the South Caucasus region could serve as appropriate proxies for Khazar descendants (Elhaik 2013). This assumption is problematic for a number of reasons. First, because of the great variety of populations in the Caucasus region and the fact that no specific population in the region is known to represent Khazar descendants, evidence for ancestry among Caucasus populations need not reflect Khazar ancestry. Second, even if it were allowed that Caucasus affinities could represent Khazar ancestry, the use of the Armenians and Georgians as Khazar proxies is particularly poor, as they represent the southern part of the Caucasus region, while the Khazar Khaganate

was centered in the North Caucasus and further to the north. Furthermore, among populations of the Caucasus, Armenians and Georgians are geographically the closest to the Middle East, and are therefore expected a priori to show the greatest genetic similarity to Middle Eastern populations. Indeed, a rather high similarity of South Caucasus populations to Middle Eastern groups was observed at the level of the whole genome in a recent study (Yunusbayev et al. 2012). Thus, any genetic similarity between Ashkenazi Jews and Armenians and Georgians might merely reflect a common shared Middle Eastern ancestry component, actually providing further support to a Middle Eastern origin of Ashkenazi Jews, rather than a hint for a Khazar origin.

Here, we examine Ashkenazi Jewish origins by assembling new and previously reported data from the three regions relevant to the origins of the Ashkenazi population, namely, Europe, the Middle East, and the region historically associated with the Khazar Khaganate. The data set, which contains 222 individuals from 13 populations covering the full Caucasus region, as well as 39 individuals from two populations in the region of the Khazar Khaganate located to the north of the Caucasus, is the largest available genome-wide sample set overlapping the Khazar region (Figure 1). Our study is the first to integrate genomic data spanning the Khazar region together with a large collection of Jewish samples. With the inclusion of the new data from the region of the Khazar Khaganate, each of a series of approaches, including PCA, spatial ancestry analysis (SPA), Bayesian clustering analysis, and analyses of genetic distance and identity-by-descent (IBD) sharing continues to support the view that Ashkenazi Jewish ancestry derives from the Middle East and Europe, and not from the Caucasus region.

Materials and Methods

Sample Set. All samples reported herein were derived from buccal swabs or blood cells collected with informed consent according to protocols approved by the National Human Subjects Review Committee in Israel and Institutional Review Boards of participating research centers. Individual population assignments follow self-identifications as members of one of the Jewish or non-Jewish populations, at the level of all four grandparents (Supplemental File S1, p. 1).

A total of 1,774 samples, including 352 that are newly reported, were assembled, incorporating 88 non-Jewish populations from Arabia, Central Asia, East Asia, Europe, the Middle East, North Africa, Siberia, South Asia, and sub-Saharan Africa. The sample collection contains 222 samples representing 13 populations specifically from the Caucasus region and 39 samples representing two populations from the Volga region north of the north Caucasus (Supplemental File S1, p. 2) (Behar et al. 2010; International HapMap 3 Consortium 2010; Li et al. 2008; Yunusbayev et al. 2012). A total of 202 samples from 18 Jewish populations spanning the range of the Jewish Diaspora were considered, including 84 novel samples and 118 samples that were previously reported (Behar et al. 2010). The aim of using

such a broad data set was to enable interpretation of the analyses of the Ashkenazi Jewish samples in the context of worldwide populations and to specifically allow contrasts of Ashkenazi Jews with populations from three geographic sources that have potentially contributed to their ancestry: Europe, the Middle East, and the geographic regions considered to have been part of the Khazar Khaganate.

It is important to clarify the conceptual difference between sampling contemporary European, Middle Eastern, and Jewish populations as representing descendants of past populations and suggesting that certain samples might represent specifically the ancient Khazar Khaganate, which disappeared ~1,000 years ago with no apparent modern population representing documented direct Khazar descendants. As it is not possible to rely on known direct descendants of the Khazars, we can regard populations presently residing in regions considered to comprise the Khazar Khaganate merely as potential proxies for Khazar ancestry. Under this assumption, we have employed populations in three geographic regions as possible proxies: South Caucasus (Abkhasian, Armenian, Azeri, and Georgian), North Caucasus (Adygei, Balkar, Chechen, Kabardin, Kumyk, Lezgin, Nogai, North Ossetian, and Tabasaran), and the Volga region north of the North Caucasus region (Chuvash and Tatar). Among these three regions, the one considered to best overlap with the center of the Khazar Khaganate is the Volga region, followed by the North Caucasus region. Supplemental File S1 lists all included regions and populations, the color and three-letter codes representing each population throughout the various analyses, and the publication in which the samples were first used. In addition, when possible, the geographic coordinates assigned for each of the non-Jewish populations are reported.

Genotyping of the New Samples. Following the manufacturer's protocol, samples were molecularly analyzed using the Illumina iScan System and the Illumina HumanOmniExpress BeadChip process. Genotype data were evaluated using Illumina GenomeStudio, version 2011.1, making use of genome build GRCh37/hg19.

Quality Control and Assembly of the Data Set. The previously reported data were obtained using five overlapping Illumina genotyping arrays (Human610-Quad, HumanHap650Y, Human660W-Quad, HumanOmniExpress-12v1 730K, and HumanOmni1-Quad), following the manufacturer's protocols, and they were evaluated using GenomeStudio, version 2011.1, with the latest available manifest files. The raw data from the previously published and new samples were first combined by array version and then lifted using the Liftover tool at the UCSC Genome Browser (Kent et al. 2002) to reflect physical positions of human genome build 37 (GRCh37). Marker rs numbers were matched with dbSNP hg19 build 135 using Snap (Johnson et al. 2008), and the strand was set according to the 1000 Genomes Project. AT and GC markers were removed in order to minimize potential strand errors during the merging of the data from the different Illumina arrays.

After data were merged, the combined data set was filtered using PLINK

(Purcell et al. 2007) to include only (a) single-nucleotide polymorphisms (SNPs) with genotyping success rate >99.5% and minor allele frequency >1%, and (b) individuals with genotyping success rate >96.5%. The stringent genotyping success filter ensures that missing data do not reflect markers that were absent in some of the arrays used less frequently in our panel. After filtering, the data contained 270,898 autosomal SNPs in 1,774 individuals.

We tested for cryptic relatedness in our data set using the King program (Manichaikul et al. 2010), finding one cryptic pair of first-degree relatives (both Kurdish Jews) and eight pairs of second-degree cryptic relatives (Supplemental File S1). Given the known strong founder effect in some Jewish groups, these pairs were not removed in some of the analyses.

Population Groups. Regional population groupings were used for analyses of genetic distance and IBD. Where appropriate, some populations were placed into multiple groupings:

1. Middle Eastern Jewish: Azerbaijani, Georgian, Iranian, Iraqi, Kurdish, and Uzbekistani (Bukharan)
2. Sephardi Jewish: Bulgarian and Turkish
3. North African Jewish: Algerian, Libyan, Moroccan, and Tunisian
4. Levantine: Bedouin, Cypriot, Druze, Jordanian, Lebanese, Palestinian, Samaritan, and Syrian
5. East European non-Jewish: Belarusian, Estonian, Lithuanian, Polish, Romanian, and Ukrainian
6. West and South European non-Jewish: French, Italian, and Spanish
7. North Caucasus: Adygei, Balkar, Chechen, Kabardin, Kumyk, Lezgin, North Ossetian, and Tabasaran
8. South Caucasus: Abkhasian, Armenian, Azeri, and Georgian
9. Caucasus: groups 7 and 8 combined
10. West Turkic: Azeri, Balkar, Chuvash, Kumyk, Nogai, and Tatar
11. East Turkic: Altaian, Kyrgyz, Turkmen, Tuvinian, Uygur, and Uzbek

Jewish groups include “Jewish” in the name; all others are non-Jewish.

A Marker Subset Pruned by Linkage Disequilibrium Patterns. For certain analyses, we thinned the data set to minimize the possible effects of linkage disequilibrium (LD). We used PLINK (Purcell et al. 2007) to calculate an LD score (r^2) for each pair of SNPs in 200-SNP windows, excluding one SNP from the pair if $r^2 > 0.4$. The window was advanced by 25 SNPs at a time. This procedure yielded a reduced set of 171,126 SNPs.

Phasing. Beagle, version 3.3.2 (Browning and Browning 2007), with default parameters was used to phase and impute missing genotypes in the full set of 1,774 samples and 270,898 SNPs. The genotyping error rate was low (6.5×10^{-4}),

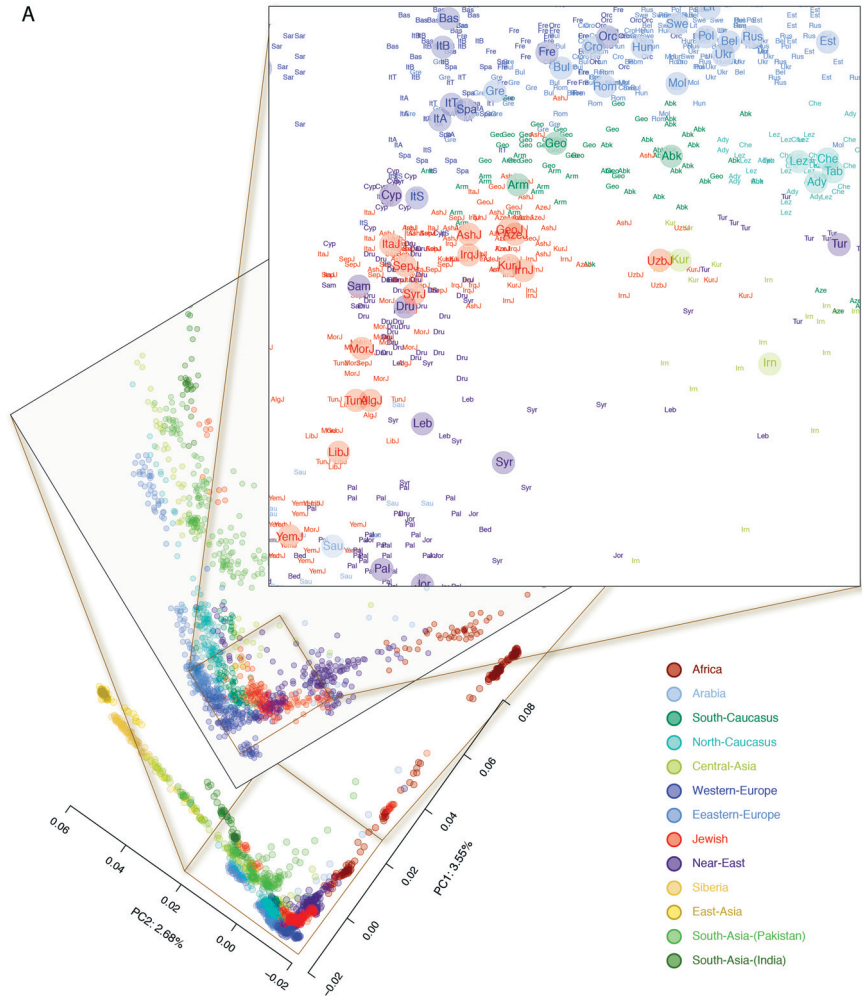
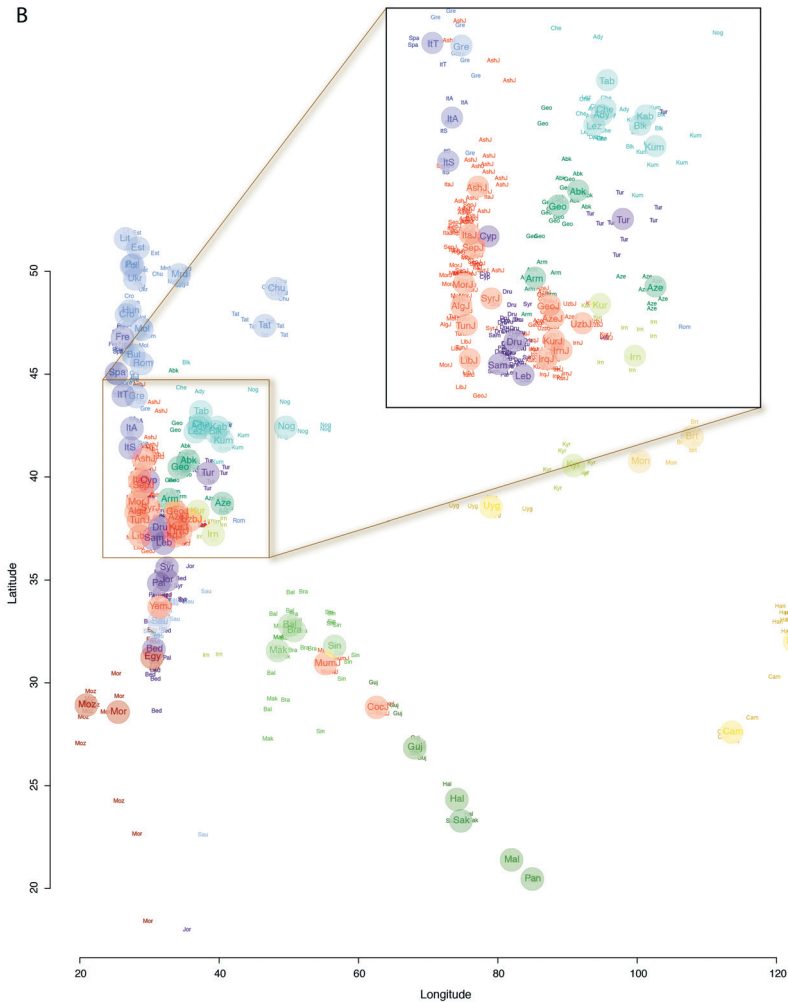


Figure 2. Principal components analysis (A, above) and spatial ancestry analysis (B, opposite). (A) The first two principal components, shown at three stages of magnification of the same plot of all individuals included in this study. (B) Scatter plot results of the inferred locations of all individuals in relation to the actual geographic locations of the reported populations. Each letter code (see Supplemental File S1) corresponds to one individual, and the color indicates the geographic region of origin. Median coordinate points for populations are shown as circles. For a high-resolution version of this figure, please visit the following web address: <http://digitalcommons.wayne.edu/cgi/viewcontent.cgi?filename=3&article=2036&context=humbiol&type=additional>



with a maximum of 0.032 across individuals, so relatively few positions were imputed. Positions 20,000,000–40,000,000 of chromosome 6, encompassing the anomalous HLA region, were discarded from the phased data. The phased data were used for both SPA and analyses of IBD.

Principal Components Analysis. The Smartpca program (Patterson et al. 2006) was used to run PCA on the LD-pruned individual data set, and the first three principal components (PCs) were extracted (Figure 2A; Supplemental Figures S1, S2). No standardization or transformation of genotypes was performed before running Smartpca. To present the results at the population level, we show

the population median for PC coordinates. PCA results were plotted using R (R Development Core Team 2012).

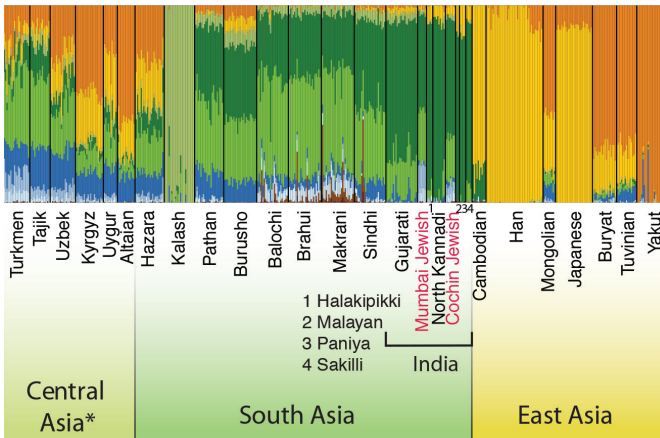
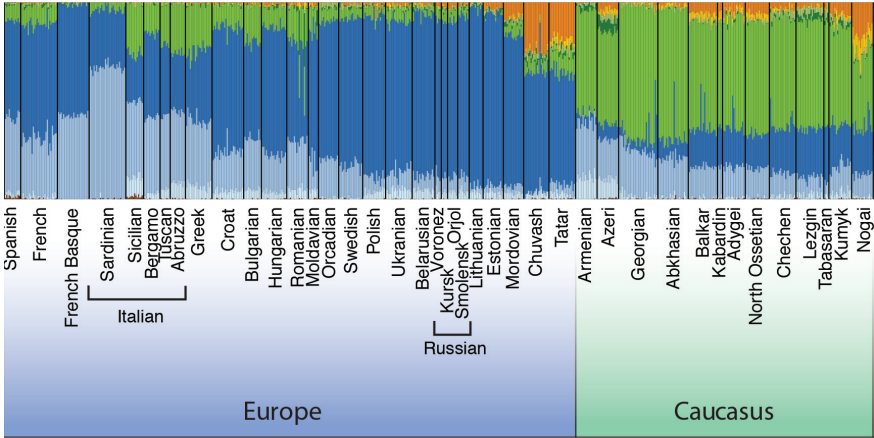
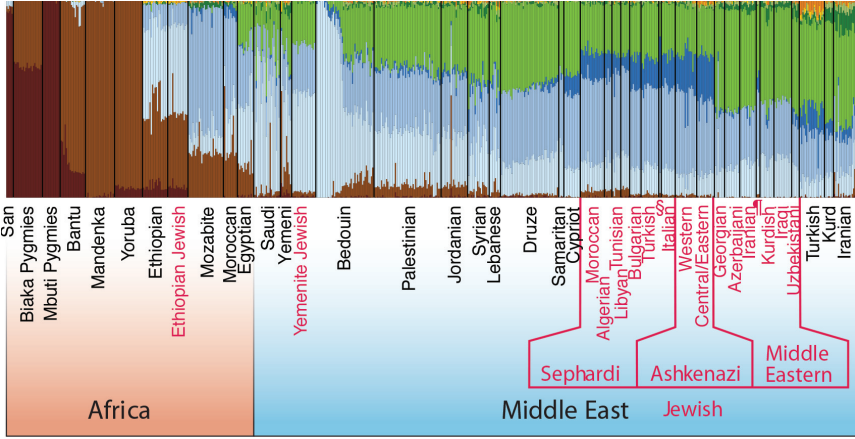
Spatial Ancestry Analysis. The Loco-LD localization method (Baran et al. 2013) was used with the phased unpruned data to geographically localize the Jewish samples among the west Eurasian samples (Figure 2B). Loco-LD is an extension of SPA, a recently developed model-based approach for the inference of spatial genetic diversity (Yang et al. 2012). The major improvement that Loco-LD introduces is a correction for LD between proximal markers. Loco-LD infers a spatial genetic model by utilizing training samples for which both genotypes and estimated geographic locations are given, and it then uses this model to localize additional samples.

With the current data set, we trained the Loco-LD model on the non-Jewish samples and then used the model to localize the Jewish samples. Specifically, the model was trained on samples from western Eurasian populations whose locations are known (Supplemental File S1, p. 4). From each training population, half of the samples were used for training. The inferred parameters of the model were then used to localize the rest of the western Eurasian sample. Thus, the samples localized by Loco-LD include the other half of the samples from populations of known locations and samples from populations whose locations are treated as unknown, among them the Jewish samples. We plotted the results using R (R Development Core Team 2012); for clarity, we also show median coordinates at the population level.

Admixture Analysis. For analyses with Admixture (Alexander et al. 2009), a Structure-like program that distributes individuals across a set of K groups inferred from unsupervised mixture-based clustering of multilocus genotypes, we used the LD-pruned unphased data. We ran Admixture at $K = 2$ to $K = 20$ clusters, considering 100 replicates for each K (Supplemental Figure S3).

Admixture includes a cross-validation procedure to help choose the “best” K , defined as the K for which the model has the best predictive accuracy (Supplemental Figure S4). The approach masks subsets of genotypes and uses the estimated ancestry proportions and allele frequencies under the model to predict the masked genotypes. On the basis of the cross-validation error distribution, the genetic structure in our sample set is best described at $K = 10$ (Figure 3). To assess the convergence of individual Admixture runs at each K , we monitored the maximum

Figure 3 (opposite). Population structure inferred by Admixture analysis: Admixture plots at $K = 10$ (see “Materials and Methods” for choice of K). Each individual is represented by a vertical (100%) stacked column of genetic membership proportions. The Jewish groups are highlighted in red. The Ashkenazi Jews are grouped as western (France, Germany, Netherlands) and central and eastern (Austria, Belarus, Hungary, Latvia, Lithuania, Poland, Romania, Russia). See Supplemental Figure S3 for plots at other values of K . *Includes Altaians from southern Siberia; §Belmonte Jewish; ¶Syrian Jewish.



difference in log-likelihood (LL) scores in fractions of runs with the highest LL scores at that value of K . We assume that a global LL maximum was reached at a given K if, say, the 10% of the runs with the highest LL score had minimal (≤ 1 LL unit) variation in LL scores. According to this reasoning, the global LL maximum was reached in runs at $K = 2-17$, excluding $K = 6, 12, 13,$ and 16 (Supplemental Figure S5). We verified our LL-differences approach using the Clumpp program (Jakobsson and Rosenberg 2007), confirming that indeed all the runs whose LL scores differed by less than 3 from the highest LL score resulted in nearly identical membership proportions (Clumpp score ≥ 0.9999) (Supplemental Figures S6, S7).

Judging from the cross-validation error distribution and our assessment of K values in which a global maximum likelihood solution was likely reached, we chose $K = 10$ as the best single representation of the Admixture genetic structure of the sample. For convenience, we plotted the runs with the highest LL score (Figure 3, Supplemental Figure S3); a nearly identical plot would have resulted had we used any of the runs yielding LL scores within 3 of the best run (as verified by Clumpp). To facilitate visual inspection of the Admixture plot at $K = 10$, we correlated population-specific average cluster memberships treated as arrays and plotted, for each Jewish group, the 20 most similar populations (Figure 4).

Analysis of Allele-Sharing Distance. We calculated allele-sharing distance (ASD) (Gao and Martin 2009) using the unphased unpruned SNP set (Figure 5). We calculated ASD between Ashkenazi Jews and our 11 regional groups. Three separate analyses using different Ashkenazi Jewish groupings were considered: all Ashkenazi Jews (Figure 5a), western Ashkenazi Jews only (Supplemental Figure S8a), and eastern Ashkenazi Jews only (Supplemental Figure S8b). For each computation, we calculated the mean ASD between pairs of individuals, one Ashkenazi Jewish individual and one from the regional group, considering all possible pairs.

To determine whether differences in ASD were statistically significant, we adopted a two-dimensional bootstrap approach (Behar et al. 2010) (Supplemental Table S1). Briefly, we tested a null hypothesis that a difference between two mean ASD values is not significant, by estimating the variance of this difference using a bootstrap approach and performing a standard normal test with the estimated variance (Behar et al. 2010). To compare ASD patterns observed with Ashkenazi Jews with those seen with other populations, we repeated the full ASD analysis three times, replacing Ashkenazi Jews with Cypriots, Druze, and Palestinians.

Figure 4 (opposite). Correlation of population-level mean membership proportions. Each plot, based on the proportions inferred in Figure 3, shows the populations with the highest correlation of membership proportions (up to 20 populations, if the number of populations with a high value of the correlation is large). The name of a plotted population starts at the value of the correlation coefficient. For a high-resolution version of this figure, please visit the following web address: <http://digitalcommons.wayne.edu/cgi/viewcontent.cgi?filename=4&article=2036&context=humbiol&type=additional>

For these analyses, Cypriots, Druze, and Palestinians were excluded from their respective regional groups.

Identity-by-Descent Sharing. IBD was analyzed using Germline, version 1.5.1 (Gusev et al. 2009), on the phased unpruned data. We ran Germline with default parameters (`-min_m 3 -bits 128 -err_hom 4 -err_het 1`) to detect pairwise IBD sharing for all pairs of study samples. Following previous work (Gusev et al. 2012), we searched for genomic regions in which sparse SNP coverage yields false-positive IBD calls and excised them from the Germline-estimated IBD segments; we divided the genome into nonoverlapping 1 Mb blocks and excised blocks with <100 SNPs. We then kept only the shared IBD segments whose length, following excisions, exceeded 3 Mb. Finally, we discarded from the analysis chromosomes 6, 11, and 12, which presented a high level of excessive false-positive sharing, similar to effects observed previously (Gusev et al. 2012).

For each population group G , we computed the mean length of IBD sharing with Ashkenazi Jews as

$$\sum_{i=1}^{N_0} \sum_{j=1}^N I_{ij} / (N_0 N),$$

where N_0 is the number of Ashkenazi Jewish samples, N is the sample size of group G , and I_{ij} is the total length of shared IBD segments between samples i and j . To test hypotheses about differential levels of sharing between different groups and the Ashkenazi Jews, we used a Wilcoxon signed rank test. Specifically, let G_1 and G_2 be two population groups. We wish to test the hypothesis that one of these groups shares more IBD segments with the Ashkenazi Jewish group than does the other. For each Ashkenazi Jewish sample i , we compute s_{i1} and s_{i2} , the mean IBD sharing between sample i and all samples in G_1 and in G_2 , respectively. Our null hypothesis is that for every randomly chosen Ashkenazi sample, $P(s_{i1} < s_{i2}) = P(s_{i2} < s_{i1})$. The two-tailed Wilcoxon signed rank p -value was computed for each pair (G_1, G_2) (Supplemental Table S2).

Results

Principal Components Analysis. Figure 2A presents the first two PCs of genetic variation at three levels of magnification, with samples color coded by geographic region. The two plots at lower magnification indicate that PC placement of most Jewish populations, including the Ashkenazi Jews, is far from such geographically distant populations as East Asians, South Asians, and sub-Saharan Africans. In the highest-magnification plot, focusing on the Jewish populations, samples are represented by a three-letter code according to Supplemental File S1, and color-coded circles indicate population-level PC coordinate medians. The plot possesses a geographic structure, with Middle Eastern populations at the

bottom and European populations at the top, arranged with Southern Europeans on the left and Eastern Europeans on the right.

The Ashkenazi Jewish samples produce a relatively tight cluster that overlaps with some Jewish and non-Jewish populations. Among the Jewish populations, Ashkenazi Jews fall closest to Italian Jews, Middle Eastern Jews, North African Jews, and Sephardi Jews, positioned continuously with other Middle Eastern non-Jewish populations along PC1. Among non-Jewish populations, Ashkenazi Jews lie nearest to Armenians, Cypriots, Druze, Greeks, and Sicilians. Four Ashkenazi Jews fall outside the main Ashkenazi cluster and lie closer to Europeans.

Samples representing the geographic region associated with the Khazar Khaganate are widely spread along the plot. Populations from the northern Volga region (Chuvash and Tatar; see Figure 1) are located far from Jewish, Middle Eastern, and Southern European populations and do not appear in the highest-magnification plot focused on the Jewish samples. Populations from the North Caucasus are largely placed in the upper right part of this plot, falling between Turks, Azeris, and Eastern Europeans (Figure 2A). The four South Caucasus populations are less closely clustered than the North Caucasus populations, with Azeris overlapping Iranians and Turks, and Abkhazians appearing closer to Eastern Europeans, Kurds, and Turks. The Georgians and Armenians fall close to each other, with Georgians placed between Eastern European populations, North Caucasus populations, Southern Europeans, and the cluster of Ashkenazi Jews, Middle Eastern Jews, Sephardi Jews, Cypriots, and Druze; Armenians lie somewhat closer to this latter cluster, particularly to the tight cluster containing Azerbaijani, Georgian, Iranian, and Kurdish Jews. Within the Khazar region, the farther south a population is, the closer it lies to Middle Eastern non-Jewish populations. No particular similarity of Ashkenazi Jews with Volga or North Caucasus populations is evident; further, the South Caucasus populations fall closer to non-Ashkenazi Middle Eastern Jewish populations than to Ashkenazi Jews.

Spatial Ancestry Analysis. Figure 2B presents the results of spatial localization with Loco-LD, showing only the samples whose spatial ancestry was inferred with respect to samples of presumed known spatial ancestry. Each sample, placed according to its estimated geographic coordinates, is color coded and represented by a three-letter code according to Supplemental File S1. As in the PCA figure, a plot at low magnification indicates placement far from most Jews of populations from a number of distant geographic regions, including Siberians and South Asians.

In the higher-magnification visualization, Ashkenazi Jews form a linear cluster in the latitudinal dimension and are closest to Italian Jews, North African Jews, Sephardi Jews, Cypriots, and Sicilians. Among populations of the Khazar region, as in PCA, the Chuvash and Tatar from the Volga region are absent from the magnified plot, and most North Caucasus and South Caucasus populations appear relatively far from the Jewish populations. Armenians fall closer to the tight cluster of six Middle Eastern Jewish populations, including two from the Caucasus region

(Azerbaijani Jews and Georgian Jews), than to Ashkenazi Jews. Notable again is the lack of any particular similarity between any population group representing one of the three Khazar regions and Ashkenazi Jews. The general pattern seen in PCA is also observed with Loco-LD: South Caucasus populations lie closer than North Caucasus and Volga populations to the Middle East, and the Jewish populations closest to Armenians are non-Ashkenazi Middle Eastern Jewish groups. Indeed, in relation to other non-Jewish populations, the Armenian median point is nearly equidistant from Middle Eastern and other South Caucasus populations, indicating a general genetic proximity with Middle Eastern populations.

Admixture. The estimated population structure from Admixture with $K = 10$ identifies several clusters corresponding to populations that are geographically distant from most Jewish populations, including two clusters centered on sub-Saharan Africa, two primarily visible in Central Asia and East Asia, and one for the Kalash population from Pakistan (Figure 3). Many of the remaining clusters, which we indicate numerically, are spread across broad regions from Europe to South Asia, and it is possible to interpret the placement of Jewish populations in terms of their membership proportions in these clusters.

The Jewish populations separate into five groups with distinct Admixture patterns (Figure 3). First, Indian Jews share similar cluster memberships with other populations of India. A clear link to the Middle East, however, is visible in the presence of clusters $k3$ (light blue) and $k4$ (intermediate blue), both of which appear in the Middle East, and the absence of $k5$ (dark blue), which contributes to Indus Valley populations but is largely missing from the Middle East. Second, Ethiopian Jews are similar to their geographic neighbors, with membership proportions that are largely indistinguishable from Amhara and Tigray Semitic-speaking Ethiopians. Third, Yemenite Jews separate from other Jews, with increased membership in cluster $k3$. Fourth, Caucasus (Azerbaijani and Georgian) and Middle Eastern (Iranian, Iraqi, Kurdish, and Uzbekistani) Jews form a group, with similar membership proportions in clusters $k3$, $k4$, and $k7$ (dark green). Finally, the fifth and largest Jewish group unites Ashkenazi, North African, and Sephardi Jews. While these populations do differ slightly in the proportions of clusters $k2$ (light red), $k4$, and $k5$, their genetic similarity is striking. Minimal distinction is visible between the Western and Eastern Ashkenazi Jews, but a minutely elevated membership is visible in the Eastern Ashkenazi group for the largely East Asian clusters $k9$ (yellow) and $k10$ (orange).

To complement the visual assessment of clustering patterns, we next identified populations with patterns most similar to Jewish groups by quantitatively correlating population-specific mean membership proportions, treated as arrays (Figure 4). For the Jewish populations included in a large group containing Ashkenazi, North African, and Sephardi Jews, most of the populations with the highest similarity of cluster membership coefficients are other Jewish populations. Considering 10 Jewish populations included in the group (Algerian, Belmonte, Bulgarian, Eastern Ashkenazi, Italian, Libyan, Moroccan, Tunisian, Turkish, Western Ashkenazi), the non-Jewish populations that appear on lists of populations with the most similar

cluster memberships are French Basques, Bulgarians, Cypriots, Druze, Greeks, Jordanians, Lebanese, Palestinians, Samaritans, Spanish, Syrians, and Italians from Abruzzo, Bergamo, Sicily, Sardinia, and Tuscany. Notably absent from this list are any of the populations from the Khazar region. In Figure 4, only Jewish populations from the Caucasus (Azerbaijani and Georgian) and the Middle East (Iranian, Iraqi, Kurdish, Syrian, and Uzbekistani) have greatest clustering similarity with Caucasus non-Jewish populations. For these groups, with the exception of the 19th and 20th most similar populations to the Uzbekistani Jews, the list of the closest populations within the Khazar region ordered by clustering similarity includes only South Caucasus populations.

Genetic Distance Analysis. The mean ASD differences from 11 population groups were similar for comparisons examining all Ashkenazi Jews, Western Ashkenazi Jews, and Eastern Ashkenazi Jews (Supplemental Figure S8, Supplemental Table S1). We thus concentrate here on the analysis of all Ashkenazi Jews. The lowest mean ASD was to Sephardi Jews (0.2721), followed by Western and Southern Europeans, South Caucasus populations, Eastern Europeans, and North African Jews. The mean ASD from the Middle East was 0.2767, among the largest values. Figure 5a shows a density plot of the pairwise ASD values between Ashkenazi Jewish individuals and individuals in each group, producing the same patterns as those seen with the mean values.

Figure 5b–d shows density plots of pairwise ASD between Cypriots, Druze, and Palestinians and individuals in each of the 11 groups (excluding the tested population from the relevant group). These plots identify the South Caucasus among the closest of the 11 groups to Cypriots, Druze, and Palestinians, as was observed for Ashkenazi Jews. This pattern reflects signals from other analyses of close genetic proximity to the South Caucasus in Middle Eastern populations that are not unique to Ashkenazi Jews.

Supplemental Table S1 shows differences between mean ASDs and p -values for the null hypothesis that no difference exists. Except for the smallest differences, most differences are statistically significant. For example, mean ASD between Ashkenazi and Sephardi Jews is smaller than mean ASD between Ashkenazi Jews and Western and Southern Europeans by a nonsignificant 0.00043 ($p = 0.18$); this ASD is smaller than the mean ASD between Ashkenazi Jews and the South Caucasus by 0.0013, but the larger difference is significant ($p = 0.0018$).

Identity-by-Descent Sharing. Figure 6 reports the mean genomic sharing between Ashkenazi Jews and the 11 population groups, and Supplemental Table S2 gives p -values for tests of the null hypotheses of equal mean IBD sharing with Ashkenazi Jews for pairs of population groups. The greatest level of sharing was observed with Sephardi Jews, considerably greater than with other populations. Substantial sharing with Eastern Europeans was also observed, though at a much lower level. Sharing with most other populations was lower still, and with Caucasus populations, the level of sharing was similar to that observed for the

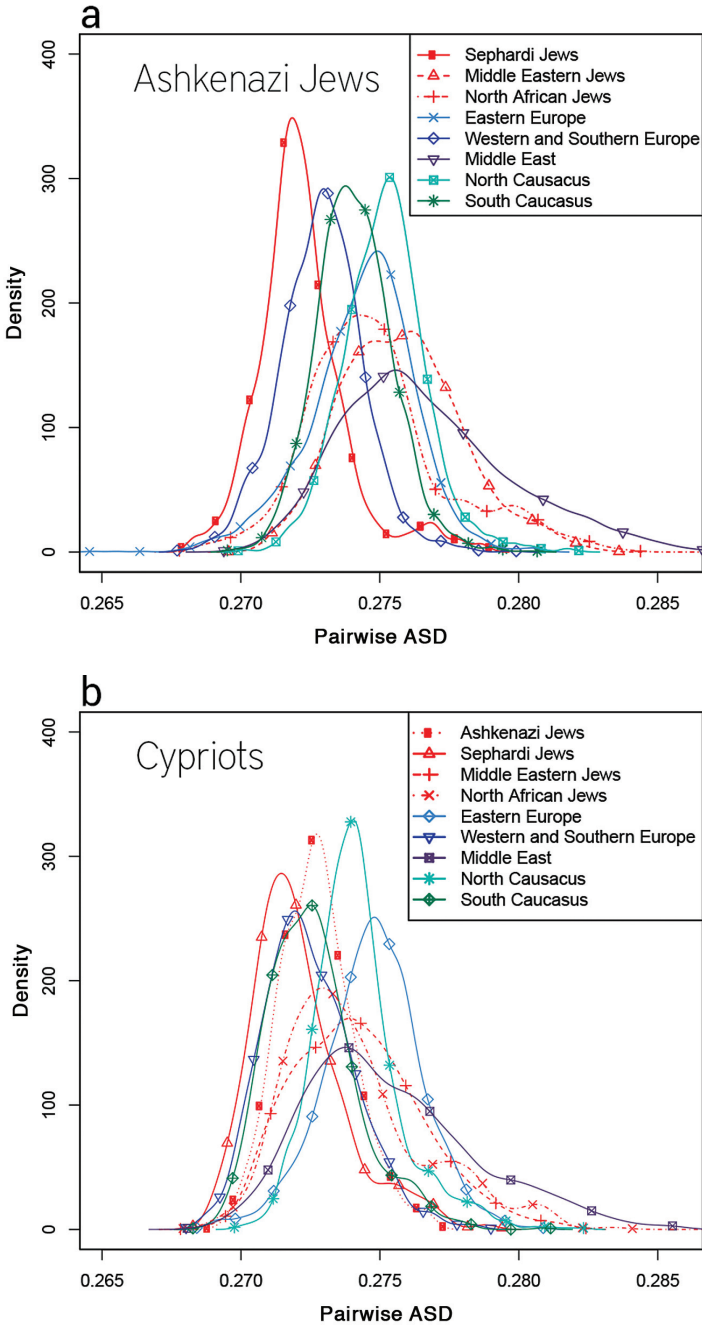
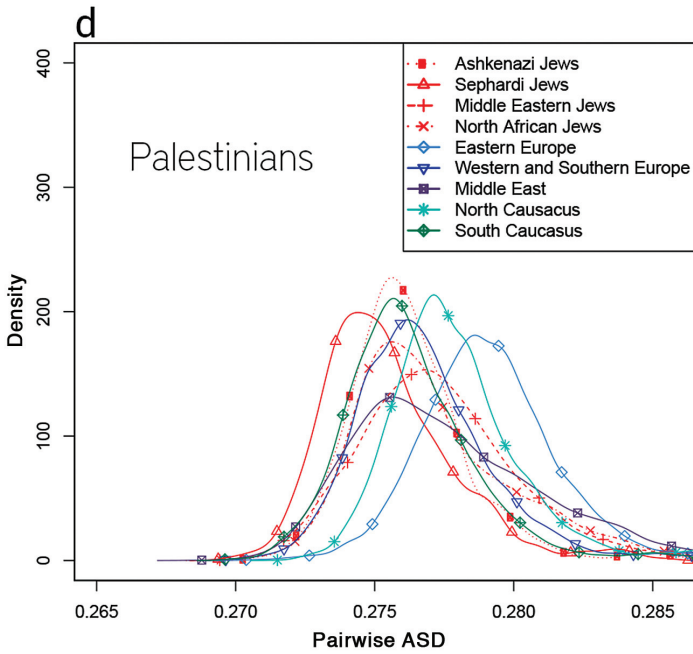
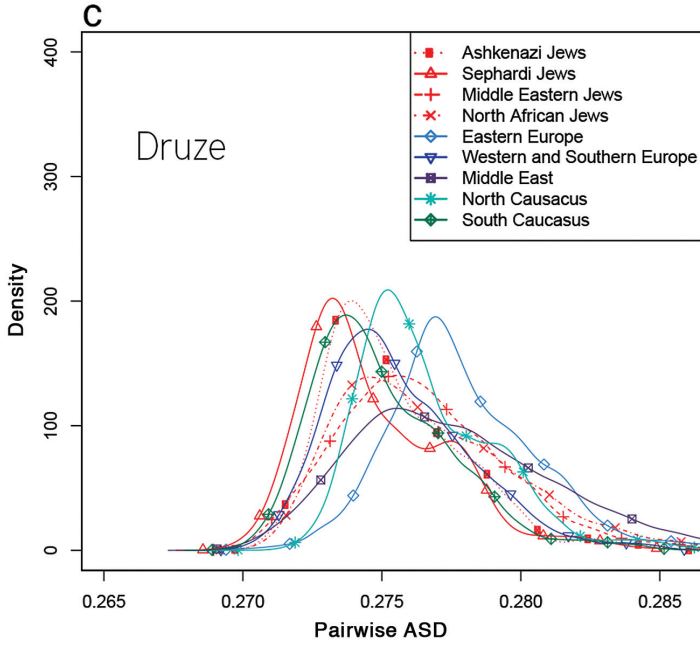


Figure 5. Density plot of pairwise genetic distances between individuals from a specific population and individuals from population groups.



Middle East. In accordance with the results from other analyses, the IBD sharing of Caucasus populations with Ashkenazi Jews was relatively low.

Discussion

This work has been the first to assemble extensive genome-wide data from all three regions that have been proposed as ancestral sources for the Ashkenazi Jewish population (Figure 1). The collection of samples from contemporary European, Middle Eastern, and Jewish populations is straightforward, as multiple forms of documentation, including the cultural identities of the populations themselves, link the modern populations to ancestral groups living at the time of the early history of the Ashkenazi Jews. By contrast, obtaining samples representing Khazars, for whom no direct link to extant populations has been established, mandates careful consideration. Recognizing this problem, we proceeded by including as many samples as possible from a region encompassing the geographic range believed to correspond to the Khazar Khaganate. After assembly of the data set, we focused our analysis on the geographic origin of the Ashkenazi Jewish population, employing a variety of analyses of population-genetic structure.

Population-Genetic Structure and Ashkenazi Jews. Our sample set representing the geographic region of the Khazar Khaganate can be split into three subsets of populations (Figure 1): from the South Caucasus region (Abkhasian, Armenian, Azeri, Georgian), North Caucasus region (Adygei, Balkar, Chechen, Kabardin, Kumyk, Lezgin, Nogai, North Ossetian, Tabasaran), and Volga region in the most northerly reaches of the Khazar expanse (Chuvash, Tatar). Under the hypothesis of a strong Khazar contribution to the Ashkenazi Jewish population, we might have expected PCA (Figure 2A) to place the Ashkenazi Jews in tight overlap with populations representing the Khazar region. Instead, Ashkenazi Jews were positioned alongside other Jewish samples, between Southern Europeans and samples from the Middle East, and did not substantially overlap populations from the Khazar region. The three Khazar subsets are themselves differentiated in PCA, with the Volga and North Caucasus populations, which approximate the Khazar region more closely than do the South Caucasus populations, positioned most distantly from Ashkenazi Jews.

Whereas PCA is an unsupervised approach for placing samples in a low-dimensional space, treating all populations as having unknown coordinates a priori, Loco-LD represents a supervised approach in which Jewish populations are placed in a spatial diagram in relation to non-Jewish samples whose geographic locations are treated as known (Figure 2B). Loco-LD confirms and sharpens the lack of evidence for the Khazar hypothesis observed in PCA, placing the Ashkenazi Jewish sample in close proximity to Italian Jews, North African Jews, Sephardi Jews, and Mediterranean non-Jewish populations such as Cypriots and Italians. Of the three Khazar subsets, the two northern groups are again distant from the Ashkenazi Jews. Among the four South Caucasus populations, the Armenian and Azeri populations

in particular lie closer to non-Jewish Middle Eastern populations, including Druze, Iranians, Kurds, and Lebanese, than to Ashkenazi Jews. Strikingly, the Ashkenazi Jewish population shows no overlap even with the South Caucasus groups, and moreover, it is apparent that the South Caucasus Armenian population is genetically closer to Middle Eastern Jewish populations than to Ashkenazi Jews.

PCA (Figure 2A) and Loco-LD (Figure 2B) both highlight two pairs of Ashkenazi Jewish samples distant from the major cluster of Ashkenazi Jews. The first pair lies within the larger European cluster and consists of two Ashkenazi Jewish samples from the Netherlands population, which was previously shown to be admixed at the level of uniparental markers (Behar et al. 2004). The other pair includes a Belarusian and a Romanian sample that are genetically outside the Jewish and European clusters, likely representing recent undocumented admixture rather than a signal of ancient Khazar origin. Because time erodes the variance of admixture across individuals in admixed populations (Verdu and Rosenberg 2011), had a Khazar contribution been made to the Ashkenazi community ~1,000 years ago, it would be common to nearly all modern Ashkenazi individuals through generations of endogamy and would not be centered on a few outliers.

We used a maximum-likelihood-based Structure-like approach as implemented in Admixture to assess the position of the Jewish groups in relation to the established genetic structure of Eurasian populations (Auton et al. 2009; Behar et al. 2010; Li et al. 2008; Metspalu et al. 2011; Yunusbayev et al. 2012) (Figure 3). In this analysis, a large group of Jewish populations, containing Ashkenazi, North African, and Sephardi Jews, produced similar patterns of membership. The similarity of the genetic membership proportions suggests a common origin of the Jewish populations in this group and limited or comparable levels of admixture with closely related host populations. Similar membership in the k_5 component might be interpreted as an admixture event between Jews and European host populations that predates the split of European and North African Jews.

Genetic structure is evident within the larger group of similar Jewish populations. North African Jews show slightly elevated membership in the k_2 component prevalent in African populations. Similarly, in the Ashkenazi Jews, the proportion of the largely European k_5 component is somewhat larger than that in the Sephardi Jews (23% vs. 16%). Within the Ashkenazi Jews from Eastern and Central Europe, we do see a signal (2.2%) of components common in East Asia that are less visible in Ashkenazi Jews from Western Europe or European Sephardi Jews (0.6%). These components also appear in Eastern Europeans and in some Middle Eastern populations, such as Yemenis, so it is difficult to attribute their minor elevation in Eastern Ashkenazi Jews to a particular origin. The most prevalent cluster in the Caucasus region is k_6 , which appears throughout Europe, the Middle East, and South Asia. Nevertheless, the Ashkenazi Jews do not stand out from other Jewish populations in possessing higher proportions of this component. Rather, Caucasus Jews and even Sephardi Jews have higher proportions of the component dominant in the Caucasus than do Ashkenazi Jews.

In brief, judging from the similarity of the membership proportion distributions

(Figure 4), Admixture demonstrates the connection of Ashkenazi, North African, and Sephardi Jews, with the most similar non-Jewish populations to Ashkenazi Jews being Mediterranean Europeans from Italy (Sicily, Abruzzo, Tuscany), Greece, and Cyprus. When subtracting the $k5$ component, which perhaps originates in Ashkenazi and Sephardi Jews from admixture with European hosts, the best matches for membership patterns of the Ashkenazi Jews shift to the Levant: Cypriots, Druze, Lebanese, and Samaritans.

Quantitative measures of genetic proximity from genetic distance analysis (Figure 5) agree with the results of the other methods. As no significant differences in genetic distance (Supplemental Figure S8, Supplemental Table S1) were noted when the Ashkenazi Jews were split into Eastern and Western groups, in agreement with the work of Guha et al. (2012), we examined Ashkenazi Jews as a single population. The lowest mean genetic distance for Ashkenazi Jews was with Sephardi Jews, followed by Western and Southern Europeans, populations of the South Caucasus, and North African Jews. Repeating the analysis by comparing the most relevant non-Jewish Middle Eastern populations to all other groups, we found that the greatest proximity of Middle Eastern populations was to the South Caucasus, again suggesting that any Ashkenazi similarity to the South Caucasus merely reflects a Middle Eastern component of their ancestry (Supplemental Table S1).

Analysis of genomic sharing, focused on IBD sharing between Ashkenazi Jews and population groups, further sharpens the results from genetic distance analysis (Figure 6). IBD analysis, which focuses on the most recent tens of generations of ancestry, is expected to generate tighter clustering of individuals within populations, between populations that have a recent common ancestral deme, or between populations that have recently experienced reciprocal gene flow (Gusev et al. 2009, 2012). Considering the IBD threshold of 3 Mb for shared segments, Ashkenazi Jews are expected to show no significant IBD sharing with any population from which they have been isolated for ≥ 20 generations. In accordance with the results from the other methods of analysis, Ashkenazi Jews show significant IBD sharing only with Eastern Europeans, North African Jews, and Sephardi Jews. Sharing was minimal with Middle Eastern populations, a not unexpected result given that the time frame for the split from Middle Eastern populations is beyond the detection power of our IBD analysis.

Conclusions: No Evidence for a Khazar Origin. Cumulatively, our analyses point strongly to ancestry of Ashkenazi Jews primarily from European and Middle Eastern populations and not from populations in or near the Caucasus region. The combined set of approaches suggests that the observations of Ashkenazi proximity to European and Middle Eastern populations in population structure analyses reflect actual genetic proximity of Ashkenazi Jews to populations with predominantly European and Middle Eastern ancestry components, and a lack of visible introgression from the region of the Khazar Khaganate—particularly among the northern Volga and North Caucasus populations—into the Ashkenazi community. We note that while we find no evidence for any significant contribution of the

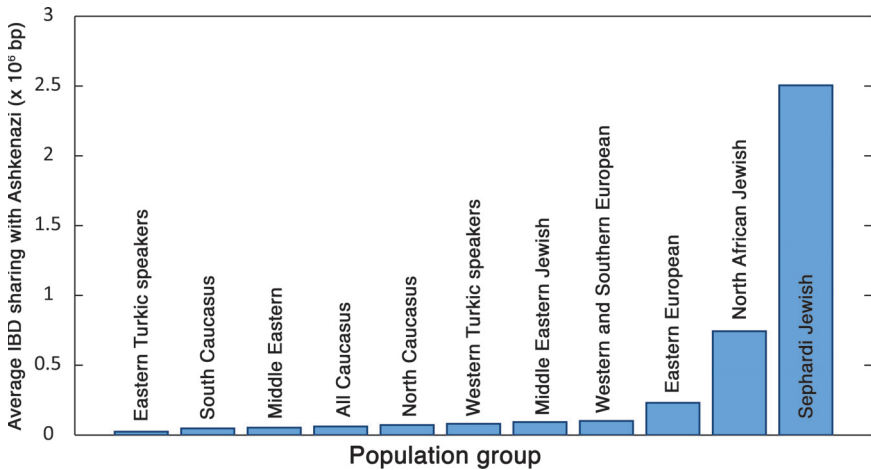


Figure 6. Average identity-by-descent sharing between different population groups and Ashkenazi Jews.

Khazar region to the Ashkenazi Jews, we cannot rule out the possibility that a level of Khazar or other Caucasus admixture occurred below the level of detectability in our study. Contemporary populations represent the outcome of many layers of minor and major demographic events that do not always leave a visible genetic signature. However, our study clearly identifies signals of Europe and the Middle East in Ashkenazi Jewish ancestry, rendering any possible undetected Khazar contribution below a minimal threshold.

Our results contrast sharply with the work of Elhaik (2013), who claimed strong support for a Khazar origin of Ashkenazi Jews. This disagreement merits close examination. Elhaik (2013) based his claim on an assumption that two South Caucasus populations, Georgians and Armenians, are suitable proxies for Khazar descendants, and on observations of similarity of these populations with Ashkenazi Jews. By assembling a larger data set containing populations that span the full range of the Khazar Khaganate, we find no evidence that a particular similarity exists between Ashkenazi Jews and any of the populations of the Khazar region; further, within the region, the newly incorporated northern populations that best overlap with the presumed center of the Khazar Khaganate are the most genetically distant from Ashkenazi Jews.

While we do observe some evidence of similarity between Ashkenazi Jews and South Caucasus populations, particularly the Armenians, it is important to assess whether this similarity could reflect Khazar origins or, rather, a shared ancestry of Ashkenazi Jews and South Caucasus populations in the Middle East. We found that the Ashkenazi Jews carry no more genetic similarity to the South Caucasus compared with many other populations from the Middle East,

Mediterranean Europe, and particularly several of the Middle Eastern Jewish populations. The South Caucasus has been previously shown (Haber et al. 2013; Yunusbayev et al. 2012) to have common genetic ancestry with much of the Middle East, as also found here. Therefore, it cannot be claimed that evidence of Ashkenazi Jewish similarity to Armenians and Georgians reflects a South Caucasus origin for Ashkenazi Jews without also claiming that the same South Caucasus ancestry underlies both Middle Eastern Jews and a large number of non-Jewish populations both from the Middle East and from Mediterranean Europe. Thus, if one accepts the premise that similarity to Armenians and Georgians represents Khazar ancestry for Ashkenazi Jews, then by extension one must also claim that Middle Eastern Jews and many Mediterranean European and Middle Eastern populations are also Khazar descendants. This claim is clearly not valid, as the differences among the various Jewish and non-Jewish populations of Mediterranean Europe and the Middle East predate the period of the Khazars by thousands of years (Baron 1957; Ben-Sasson 1976; De Lange 1984; Mahler 1971).

We take this opportunity to clarify the differences between genetic proximity of populations and ancestor–descendant relationships. Most illustrative are the results obtained from Admixture (Figure 3). This analysis clearly shows that populations throughout the vast western Eurasian region share the same genetic clusters, albeit at different frequencies. These genetic components typically represent ancient genetic forces that have shaped the current genetic landscape, and an attempt to connect them to a particular population that has likely arisen much later than their establishment is inherently problematic. Specifically, our analysis highlighted the Armenian population and, to a lesser extent, the Azeri population as the only Caucasus populations that present all genetic components also observed in Middle Eastern and North African populations. Thus, Armenians, used by Elhaik (2013) as a potential proxy for a Khazar source population, could equally well have been employed as a misleading proxy for many populations across the Middle East with similar cluster memberships, thereby producing the same problematic interpretation that each such population is ancestral to Ashkenazi Jews. The mere finding of shared cluster membership does not unambiguously attest to a demographic event responsible for the cluster and therefore cannot be further interpreted to suggest that one population is ancestral to another population simply because it is found within the same cluster. Thus, for example, it would be misleading to conclude from the Admixture analysis that Ashkenazi Jews are actually the primary source population giving rise to the Sicilians, Druze, or North African Jews with whom they share similar membership coefficients.

In summary, in this most comprehensive study to date, we have examined the three potential sources for contemporary Ashkenazi Jews, using a new sample set that covers the full extent of the Khazar realm of the sixth to tenth centuries. Analysis of this large data set does not change and, in fact, reinforces the conclusions of multiple past studies, including ours and those of other groups (Atzmon et al. 2010; Bauchet et al. 2007; Behar et al. 2010; Campbell et al. 2012; Guha et al. 2012; Haber et al. 2013; Henn et al. 2012; Kopelman et al. 2009; Seldin et al.

2006; Tian et al. 2008). We confirm the notion that the Ashkenazi, North African, and Sephardi Jews share substantial genetic ancestry and that they derive it from Middle Eastern and European populations, with no indication of a detectable Khazar contribution to their genetic origins.

Acknowledgments We acknowledge the Morrison Institute for Population and Resource Studies for travel support for D.M.B. to visit the lab of N.A.R. R.V., M.M., E.M., K.T., B.Y., A.K., and H.S. were supported by the E.U. European Regional Development Fund through the Centre of Excellence in Genomics to Estonian Biocentre, and University of Tartu and Estonian Basic Research grant SF0270177As08. M.M. acknowledges Estonian Science Foundation grant 8973 and the Baltic-American Freedom Foundation Research Scholarship program. E.H. is a faculty fellow of the Edmond J. Safra Center for Bioinformatics at Tel Aviv University. Y.B. was supported in part by a fellowship from the Edmond J. Safra Center for Bioinformatics at Tel Aviv University. E.H. was also partially supported by U.S. National Science Foundation grant III-1217615. We thank the High Performance Computing Center of the University of Tartu and Tuuli Mets from the Estonian Biocentre core facility.

Received 30 August 2013; revision accepted for publication 6 October 2013.

Literature Cited

- Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19:1,655–1,664.
- Atzmon, G., L. Hao, I. Pe'er et al. 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.
- Auton, A., K. Bryc, A. R. Boyko et al. 2009. Global distribution of genomic diversity underscores rich complex history of continental human populations. *Genome Res.* 19:795–803.
- Baran, Y., I. Quintela, A. Carracedo et al. 2013. Enhanced localization of genetic samples through linkage-disequilibrium correction. *Am. J. Hum. Genet.* 92:882–894.
- Baron, S. W. 1957. *A Social and Religious History of the Jews*. New York: Columbia University Press.
- Bauchet, M., B. McEvoy, L. N. Pearson et al. 2007. Measuring European population stratification with microarray genotype data. *Am. J. Hum. Genet.* 80:948–956.
- Behar, D. M., D. Garrigan, M. E. Kaplan et al. 2004. Contrasting patterns of Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations. *Hum. Genet.* 114:354–365.
- Behar, D. M., E. Metspalu, T. Kivisild 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., M. G. Thomas, K. Skorecki et al. 2003. Multiple origins of Ashkenazi Levites: Y chromosome evidence for both Near Eastern and European ancestries. *Am. J. Hum. Genet.* 73:768–779.
- Behar, D. M., B. Yunusbayev, M. Metspalu et al. 2010. The genome-wide structure of the Jewish people. *Nature* 466:238–242.
- Ben-Sasson, H. H., ed. 1976. *A History of the Jewish People*. Cambridge, MA: Harvard University Press.

- Browning, S. R., and B. L. Browning. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81:1,084–1,097.
- Campbell, C. L., P. F. Palamara, M. Dubrovsky et al. 2012. North African Jewish and non-Jewish populations form distinctive, orthogonal clusters. *Proc. Natl. Acad. Sci. USA* 109:13,865–13,870.
- Cavalli-Sforza, L. L., and M. W. Feldman. 2003. The application of molecular genetic approaches to the study of human evolution. *Nat. Genet.* 33(Supplemental):266–275.
- De Lange, N. 1984. *Atlas of the Jewish World*. New York: Facts on File.
- Elhaik, E. 2013. The missing link of Jewish European ancestry: Contrasting the Rhineland and the Khazarian hypotheses. *Genome Biol. Evol.* 5:61–74.
- Gao, X., and E. R. Martin. 2009. Using allele sharing distance for detecting human population stratification. *Hum. Hered.* 68:182–191.
- Golden, P. B., H. Ben-Shammai, and A. Róna-Tas, eds. 2007. *The World of the Khazars: New Perspectives. Selected Papers from the Jerusalem 1999 International Khazar Colloquium*. Leiden: Brill.
- Guha, S., J. A. Rosenfeld, A. K. Malhotra et al. 2012. Implications for health and disease in the genetic signature of the Ashkenazi Jewish population. *Genome Biol.* 13:R2.
- Gusev, A., J. K. Lowe, M. Stoffel et al. 2009. Whole population, genome-wide mapping of hidden relatedness. *Genome Res.* 19:318–326.
- Gusev, A., P. F. Palamara, G. Aponte et al. 2012. The architecture of long-range haplotypes shared within and across populations. *Mol. Biol. Evol.* 29:473–486.
- Haber, M., D. Gauguier, S. Youhanna et al. 2013. Genome-wide diversity in the Levant reveals recent structuring by culture. *PLoS Genet.* 9:e1003316.
- Hammer, M. F., D. M. Behar, T. M. Karafet et al. 2009. Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish priesthood. *Hum. Genet.* 126:707–717.
- Hammer, M. F., A. J. Redd, E. T. Wood et al. 2000. Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. *Proc. Natl. Acad. Sci. USA* 97:6,769–6,774.
- Henn, B. M., L. Hon, J. M. Macpherson et al. 2012. Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. *PLoS One* 7:e34267.
- International HapMap 3 Consortium. 2010. Integrating common and rare genetic variation in diverse human populations. *Nature* 467:52–58.
- Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1,801–1,806.
- Johnson, A. D., R. E. Handsaker, S. L. Pulit et al. 2008. SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 24:2,938–2,939.
- Kent, W. J., C. W. Sugnet, T. S. Furey et al. 2002. The human genome browser at UCSC. *Genome Res.* 12:996–1,006.
- Kopelman, N. M., L. Stone, C. Wang et al. 2009. Genomic microsatellites identify shared Jewish ancestry intermediate between Middle Eastern and European populations. *BMC Genet.* 10:80.
- Lawson, D. J., and D. Falush. 2012. Population identification using genetic data. *Annu. Rev. Genomics Hum. Genet.* 13:337–361.
- Li, J. Z., D. M. Absher, H. Tang et al. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319:1,100–1,104.
- Mahler, R. A. 1971. *A History of Modern Jewry, 1780–1815*. New York: Schocken Books.
- Manichaikul, A., J. C. Mychaleckyj, S. S. Rich et al. 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics* 26:2,867–2,873.
- Metspalu, M., I. G. Romero, B. Yunusbayev et al. 2011. Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. *Am. J. Hum. Genet.* 89:731–744.
- Nebel, A., D. Filon, B. Brinkmann et al. 2001. The Y chromosome pool of Jews as part of the genetic landscape of the Middle East. *Am. J. Hum. Genet.* 69:1,095–1,112.

- Need, A. C., D. Kasperavičiūtė, E. T. Cirulli et al. 2009. A genome-wide genetic signature of Jewish ancestry perfectly separates individuals with and without full Jewish ancestry in a large random sample of European Americans. *Genome Biol.* 10:R7.
- Novembre, J., and S. Ramachandran. 2011. Perspectives on human population structure at the cusp of the sequencing era. *Annu. Rev. Genomics Hum. Genet.* 12:245–274.
- Patterson, N., A. L. Price, and D. Reich. 2006. Population structure and eigenanalysis. *PLoS Genet.* 2:e190.
- Price, A. L., J. Butler, N. Patterson et al. 2008. Discerning the ancestry of European Americans in genetic association studies. *PLoS Genet.* 4:e236.
- Purcell, S., B. Neale, K. Todd-Brown et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–575.
- R Development Core Team. 2012. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ritte, U., E. Neufeld, E. M. Prager et al. 1993. Mitochondrial DNA affinity of several Jewish communities. *Hum. Biol.* 65:359–385.
- Santachiara Benerecetti, A. S., O. Semino, G. Passarino et al. 1993. The common, Near-Eastern origin of Ashkenazi and Sephardi Jews supported by Y-chromosome similarity. *Ann. Hum. Genet.* 57:55–64.
- Seldin, M. F., R. Shigeta, P. Villoslada et al. 2006. European population substructure: Clustering of northern and southern populations. *PLoS Genet.* 2:e143.
- Tian, C., R. M. Plenge, M. Ransom et al. 2008. Analysis and application of European genetic substructure using 300 K SNP information. *PLoS Genet.* 4:e4.
- Verdu, P., and N. A. Rosenberg. 2011. A general mechanistic model for admixture histories of hybrid populations. *Genetics* 189:1,413–1,426.
- Yang, W. Y., J. Novembre, E. Eskin et al. 2012. A model-based approach for analysis of spatial structure in genetic data. *Nat. Genet.* 44:725–731.
- Yunusbayev, B., M. Metspalu, M. Jarve et al. 2012. The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Mol. Biol. Evol.* 29:359–365.

Supplemental Material

Supplemental File S1. List of samples and populations used in this study. The file details the population name, region, three-letter code identifier, color, geographic coordinates, and publication in which they were first used. Supplemental File S1 is available online at <http://digitalcommons.wayne.edu/cgi/viewcontent.cgi?filename=0&article=2196&context=humbiol&type=additional>.

Supplemental Table S1. *p*-Values for the Two-Dimensional Bootstrap Approach to Determine Statistical Significance of Average Allele-Sharing Distances

<i>Ashkenazi Jews</i>	SEPHARDI JEWS	MIDDLE E. JEWS	N. AFRICAN JEWS	E. EUROPE
ASD*	0.272187427	0.275802089	0.274887323	0.274352106
Sephardi Jews	NA	-0.003614662	-0.002699896	-0.002164679
Middle Eastern Jews	4.67E-28	NA	0.000914765	0.001449982
North African Jews	4.62E-08	0.020887550	NA	0.000535217
Eastern Europe	3.18E-10	2.21E-07	0.108959242	NA
Western & Southern Europe	0.009876445	0	3.32E-07	2.33E-15
Middle East	1.92E-40	0.000569784	2.65E-05	3.65E-15
North Caucasus	1.06E-21	0.003944615	0.236643759	1.97E-05
South Caucasus	4.99E-10	3.02E-14	0.023792594	0.090980888
All Caucasus	9.31E-17	9.84E-07	0.348801917	0.024087646
Turkic West	8.05E-50	1.10E-06	1.48E-07	1.07E-45
Turkic East	0	0	2.80E-192	0
<i>East Ashkenazi Jews</i>				
	SEPHARDI JEWS	MIDDLE E. JEWS	N. AFRICAN JEWS	E. EUROPE
ASD	0.27225019	0.275776739	0.27484848	0.274218332
Sephardi Jews	NA	-0.003526549	-0.00259829	-0.001968141
Middle Eastern Jews	3.81E-30	NA	0.000928259	0.001558408
North African Jews	6.34E-08	0.021757868	NA	0.000630149
Eastern Europe	4.62E-09	1.23E-07	0.083379561	NA
Western & Southern Europe	0.007483019	0	8.78E-07	1.60E-09
Middle East	4.34E-41	0.000533368	2.84E-05	4.08E-14
North Caucasus	1.67E-21	0.001164837	0.310165398	4.19E-05
South Caucasus	9.95E-10	6.88E-15	0.023128105	0.197866372
All Caucasus	9.88E-17	1.52E-07	0.300103991	0.026193947
Turkic West	3.51E-54	9.66E-06	1.08E-06	4.15E-40
Turkic East	0	4.03E-277	4.93E-165	0
<i>West Ashkenazi Jews</i>				
	SEPHARDI JEWS	MIDDLE E. JEWS	N. AFRICAN JEWS	E. EUROPE
ASD	0.272110179	0.275833288	0.27493513	0.274516752
Sephardi Jews	NA	-0.003723108	-0.00282495	-0.002406573
Middle Eastern Jews	2.16E-23	NA	0.000898158	0.001316536
North African Jews	8.47E-08	0.022402679	NA	0.000418377
Eastern Europe	6.22E-07	0.000876763	0.215130778	NA
Western & Southern Europe	0.027607288	0	1.43E-06	2.56E-13
Middle East	1.99E-36	0.000877794	3.68E-05	1.57E-07
North Caucasus	4.74E-17	0.037151359	0.179306931	0.001237331
South Caucasus	5.33E-09	1.52E-11	0.030551707	0.136629461
All Caucasus	7.07E-14	8.67E-05	0.423042212	0.112687529
Turkic West	6.36E-32	1.88E-05	5.81E-07	2.01E-35
Turkic East	9.05E-273	5.20E-251	1.80E-181	0

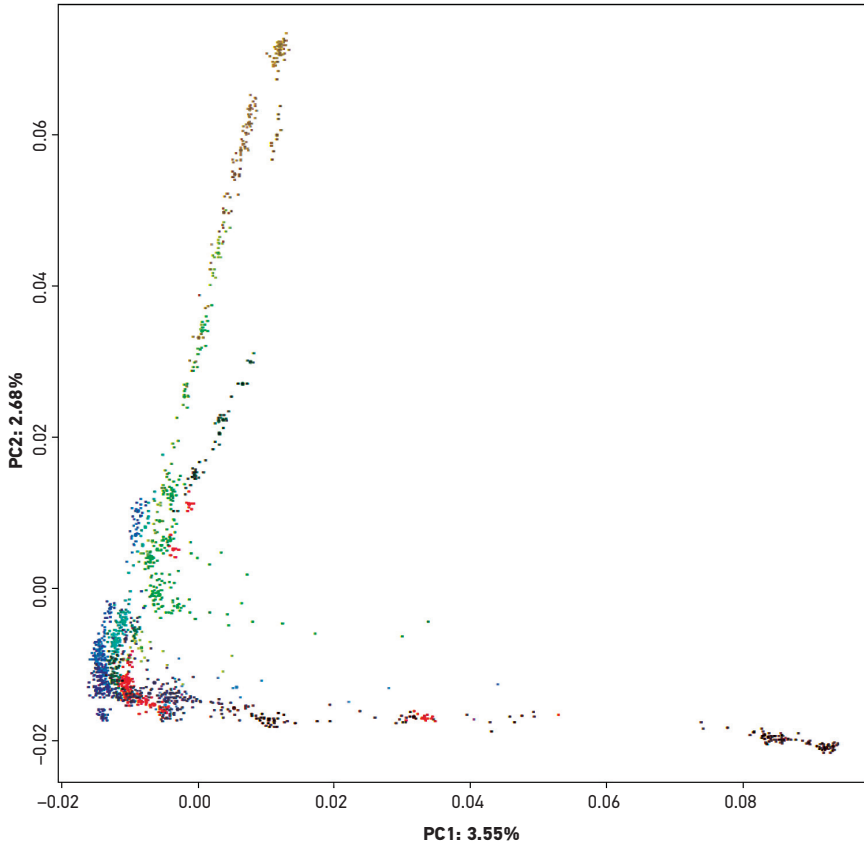
Allele-sharing distance (ASD) differences (above diagonal) and their *p*-values for the null hypothesis of no difference in true ASD (below diagonal) for comparing Ashkenazi Jews to a variety of population groupings, as described in the text.

*The first row gives the mean ASD between the populations in the columns and Ashkenazi Jews

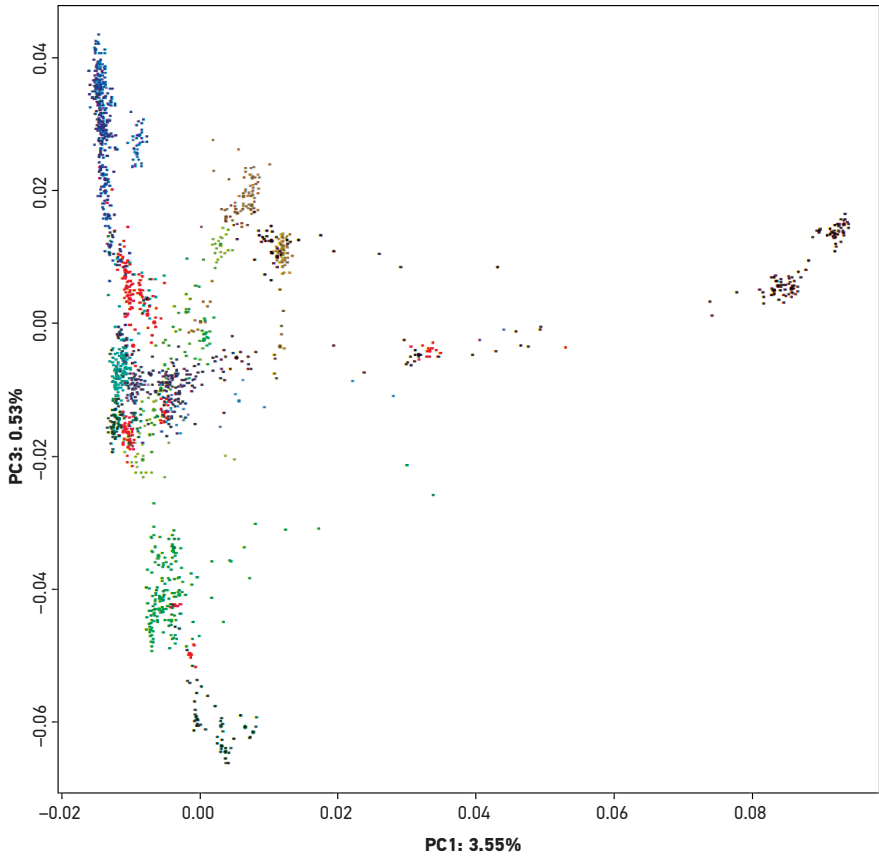
<i>W. & S. EUROPE</i>	<i>MIDDLE EAST</i>	<i>N. CAUSACUS</i>	<i>S. CAUCASUS</i>	<i>ALL CAUCASUS</i>	<i>TURKIC WEST</i>	<i>TURKIC EAST</i>
0.272904713	0.276737205	0.275180508	0.274090984	0.274730947	0.277086961	0.288030926
-0.000717286	-0.004549778	-0.002993082	-0.001903557	-0.002543521	-0.004899534	-0.015843499
0.002897376	-0.000935116	0.000621580	0.001711104	0.001071141	-0.001284873	-0.012228838
0.001982611	-0.001849882	-0.000293185	0.000796339	0.000156376	-0.002199638	-0.013143603
0.001447394	-0.002385098	-0.000828402	0.000261122	-0.000378841	-0.002734855	-0.01367882
NA	-0.003832492	-0.002275796	-0.001186272	-0.001826235	-0.004182249	-0.015126214
2.80E-75	NA	0.001556696	0.002646221	0.002006257	-0.000349757	-0.011293722
5.64E-68	2.55E-15	NA	0.001089524	0.000449561	-0.001906453	-0.012850418
1.66E-22	0	0	NA	-0.000639963	-0.002995977	-0.013939942
1.76E-56	0	0	7.93E-24	NA	-0.002356014	-0.013299979
1.22E-170	0.069159855	7.98E-35	3.02E-112	6.06E-67	NA	-0.010943965
0	3.40E-308	0	0	0	0	NA
<i>W. & S. EUROPE</i>	<i>MIDDLE EAST</i>	<i>N. CAUSACUS</i>	<i>S. CAUCASUS</i>	<i>ALL CAUCASUS</i>	<i>TURKIC WEST</i>	<i>TURKIC EAST</i>
0.272951547	0.276715439	0.275058816	0.274020694	0.274630465	0.276905653	0.287600393
-0.000701356	-0.004465249	-0.002808625	-0.001770504	-0.002380274	-0.004655462	-0.015350203
0.002825192	-0.0009387	0.000717924	0.001756045	0.001146275	-0.001128913	-0.011823654
0.001896934	-0.001866959	-0.000210335	0.000827786	0.000218016	-0.002057172	-0.012751913
0.001266785	-0.002497108	-0.000840484	0.000197637	-0.000412133	-0.002687321	-0.013382062
NA	-0.003763893	-0.002107269	-0.001069147	-0.001678918	-0.003954106	-0.014648847
2.56E-76	NA	0.001656624	0.002694745	0.002084975	-0.000190213	-0.010884954
7.07E-52	1.55E-14	NA	0.001038121	0.000428351	-0.001846837	-0.012541578
8.36E-16	0	4.44E-16	NA	-0.00060977	-0.002884959	-0.013579699
1.55E-43	0	4.44E-16	4.17E-16	NA	-0.002275188	-0.012969929
3.63E-143	0.216210581	1.15E-36	7.75E-88	3.24E-67	NA	-0.010694741
0	1.33E-248	0	0	0	0	NA
<i>W. & S. EUROPE</i>	<i>MIDDLE EAST</i>	<i>N. CAUSACUS</i>	<i>S. CAUCASUS</i>	<i>ALL CAUCASUS</i>	<i>TURKIC WEST</i>	<i>TURKIC EAST</i>
0.272847071	0.276763993	0.275330284	0.274177495	0.274854619	0.27731011	0.288560813
-0.000736891	-0.004653813	-0.003220105	-0.002067316	-0.002744439	-0.005199931	-0.016450633
0.002986217	-0.000930705	0.000503004	0.001655793	0.000978669	-0.001476822	-0.012727525
0.002088059	-0.001828863	-0.000395155	0.000757635	8.05E-05	-0.00237498	-0.013625683
0.001669682	-0.002247241	-0.000813532	0.000339257	-0.000337867	-0.002793358	-0.014044061
NA	-0.003916922	-0.002483213	-0.001330424	-0.002007548	-0.004463039	-0.015713742
2.06E-43	NA	0.001433709	0.002586498	0.001909374	-0.000546117	-0.01179682
7.32E-64	3.45E-09	NA	0.001152789	0.000475665	-0.001979826	-0.013230529
1.13E-15	0	0	NA	-0.000677124	-0.003132615	-0.014383318
6.88E-46	0	0	3.60E-18	NA	-0.002455491	-0.013706194
1.94E-133	0.051780291	5.63E-25	1.59E-51	3.26E-39	NA	-0.011250703
0	6.68E-239	0	0	0	0	NA

<i>Cypriots</i>	<i>ASHKENAZI JEWS</i>	<i>SEPHARDI JEWS</i>	<i>MIDDLE E. JEWS</i>	<i>N. AFRICAN JEWS</i>	<i>E. EUROPE</i>
ASD	0.272716316	0.272071417	0.274191394	0.274073326	0.274720827
Ashkenazi Jews	NA	0.000644899	-0.001475078	-0.001357010	-0.002004511
Sephardi Jews	0.019573857	NA	-0.002119977	-0.002001910	-0.002649411
Middle Eastern Jews	7.59E-08	4.52E-12	NA	0.000118068	-0.000529434
North African Jews	0.000302541	9.58E-06	0.390889391	NA	-0.000647501
Eastern Europe	1.14E-20	9.82E-16	0.030733627	0.04841699	NA
Western & Southern Europe	0.064618295	0.111367335	1.38E-11	1.22E-05	0
Middle East	1.19E-26	1.05E-25	3.42E-05	0.001221921	0.005542903
North Causacus	8.97E-14	4.20E-12	0.459825392	0.406910180	0.001492770
South Causacus	0.248544402	0.051580622	2.74E-09	0.000101299	0
All Causacus	8.21E-06	9.69E-07	0.003835459	0.075849150	3.38E-12
Turkic West	2.09E-105	2.40E-56	8.14E-21	5.07E-12	7.98E-27
Turkic East	0	0	6.61E-305	9.23E-216	0
<i>Druze</i>	<i>ASHKENAZI JEWS</i>	<i>SEPHARDI JEWS</i>	<i>MIDDLE E. JEWS</i>	<i>N. AFRICAN JEWS</i>	<i>E. EUROPE</i>
ASD	0.275421256	0.274786351	0.276405933	0.276628862	0.277899613
Ashkenazi Jews	NA	0.000634905	-0.000984677	-0.001207606	-0.002478357
Sephardi Jews	0.007880354	NA	-0.001619582	-0.001842510	-0.003113262
Middle Eastern Jews	9.90E-05	6.98E-09	NA	-0.000222929	-0.001493680
North African Jews	0.000359106	3.02E-06	0.286033881	NA	-0.001270751
Eastern Europe	9.39E-40	1.20E-33	6.80E-10	9.04E-05	NA
Western & Southern Europe	0.118383996	0.000335739	0.000581384	0.001296155	0
Middle East	1.59E-16	7.27E-21	1.46E-05	0.007995378	0.181316008
North Causacus	1.15E-14	3.67E-14	0.128257383	0.459013665	7.75E-14
South Causacus	0.004445469	0.162098312	1.53E-09	1.84E-06	0
All Causacus	4.14E-05	6.70E-07	0.032855827	0.032061492	0
Turkic West	1.33E-125	8.86E-69	1.79E-29	1.47E-14	1.01E-15
Turkic East	0	0	0	1.38E-263	0
<i>Palestinians</i>	<i>ASHKENAZI JEWS</i>	<i>SEPHARDI JEWS</i>	<i>MIDDLE E. JEWS</i>	<i>N. AFRICAN JEWS</i>	<i>E. EUROPE</i>
ASD	0.276267470	0.275487765	0.277288712	0.277151090	0.279170196
Ashkenazi Jews	NA	0.000779706	-0.001021241	-0.000883619	-0.002902726
Sephardi Jews	0.000904893	NA	-0.001800947	-0.001663325	-0.003682431
Middle Eastern Jews	4.07E-05	4.31E-12	NA	0.000137622	-0.001881484
North African Jews	0.005339760	2.06E-05	0.365120222	NA	-0.002019106
Eastern Europe	9.17E-57	1.81E-46	2.77E-15	2.16E-09	NA
Western & Southern Europe	0.001050343	2.40E-07	0.012225127	0.126773844	0
Middle East	1.60E-09	1.73E-14	0.100272404	0.093098477	1.96E-08
North Causacus	9.28E-30	3.24E-23	0.002259224	0.012972672	3.51E-14
South Causacus	0.408183940	0.001624064	2.62E-06	0.004868948	0
All Causacus	4.19E-13	4.60E-13	0.396015600	0.409756101	0
Turkic West	2.39E-167	3.25E-83	2.94E-35	9.08E-20	7.05E-13
Turkic East	0	0	0	1.21E-252	0

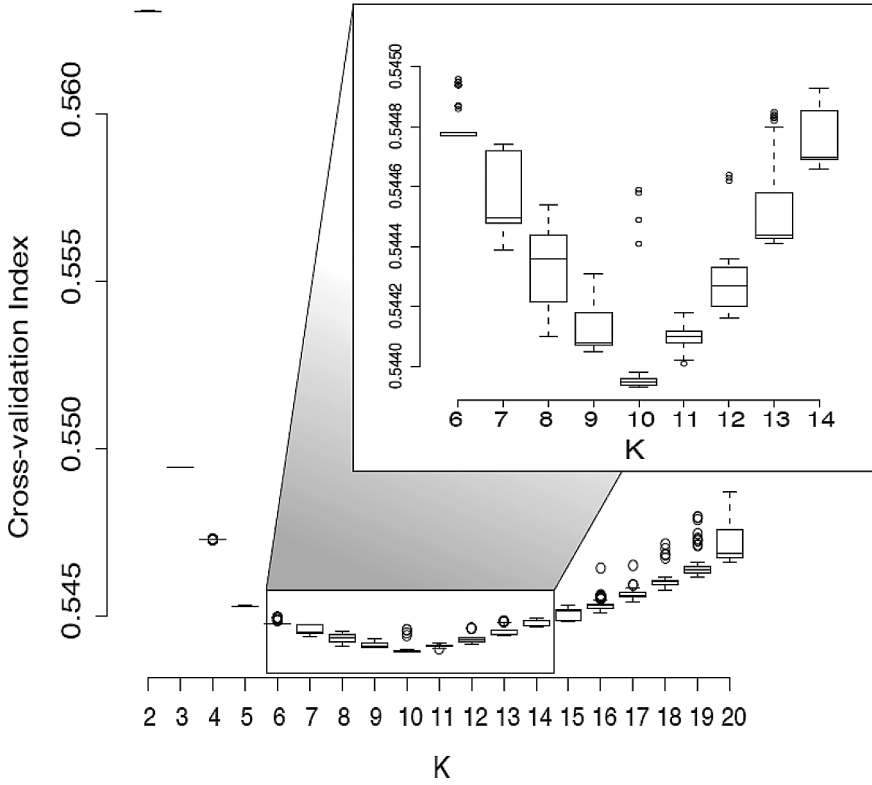
<i>W. & S. EUROPE</i>	<i>MIDDLE EAST</i>	<i>N. CAUSACUS</i>	<i>S. CAUCASUS</i>	<i>ALL CAUCASUS</i>	<i>TURKIC WEST</i>	<i>TURKIC EAST</i>
0.272437149	0.275435943	0.274165860	0.272583872	0.273513098	0.276820025	0.288299925
0.000279167	-0.002719627	-0.001449543	0.000132445	-0.000796781	-0.004103709	-0.015583609
-0.000365732	-0.003364526	-0.002094443	-0.000512455	-0.001441681	-0.004748608	-0.016228509
0.001754245	-0.001244549	2.55E-05	0.001607522	0.000678296	-0.002628631	-0.014108532
0.001636178	-0.001362617	-9.25E-05	0.001489455	0.000560229	-0.002746698	-0.014226599
0.002283679	-0.000715116	0.000554968	0.002136956	0.001207730	-0.002099197	-0.013579098
NA	-0.002998794	-0.001728711	-0.000146723	-0.001075949	-0.004382876	-0.015862777
2.15E-44	NA	0.001270083	0.002852071	0.001922845	-0.001384082	-0.012863983
3.98E-30	4.51E-09	NA	0.001581988	0.000652762	-0.002654165	-0.014134066
0.190372615	0	0	NA	-0.000929226	-0.004236153	-0.015716054
5.21E-14	0	0	4.17E-33	NA	-0.003306927	-0.014786828
1.66E-154	1.18E-10	4.30E-58	9.60E-229	9.76E-128	NA	-0.011479901
0	0	0	0	0	0	NA
<i>W. & S. EUROPE</i>	<i>MIDDLE EAST</i>	<i>N. CAUSACUS</i>	<i>S. CAUCASUS</i>	<i>ALL CAUCASUS</i>	<i>TURKIC WEST</i>	<i>TURKIC EAST</i>
0.275623002	0.277652317	0.276664686	0.275045070	0.275996398	0.279377852	0.290105409
-0.000201746	-0.002231061	-0.001243430	0.000376186	-0.000575142	-0.003956596	-0.014684153
-0.000836650	-0.002865965	-0.001878335	-0.000258719	-0.001210047	-0.004591501	-0.015319057
0.000782931	-0.001246384	-0.000258753	0.001360863	0.000409535	-0.002971919	-0.013699476
0.00100586	-0.001023455	-3.58E-05	0.001583792	0.000632464	-0.002748991	-0.013476547
0.002276611	0.000247296	0.001234927	0.002854543	0.001903215	-0.001478239	-0.012205796
NA	-0.002029315	-0.001041685	0.000577932	-0.000373396	-0.003754851	-0.014482407
4.02E-19	NA	0.000987630	0.002607247	0.001655919	-0.001725536	-0.012453092
1.34E-16	3.77E-06	NA	0.001619617	0.000668288	-0.002713166	-0.013440722
1.83E-05	0	0	NA	-0.000951328	-0.004332783	-0.015060339
0.001082960	7.66E-15	0	3.92E-52	NA	-0.003381454	-0.014109011
3.31E-125	2.86E-14	1.88E-88	1.34E-251	3.33E-169	NA	-0.010727556
0	0	0	0	0	0	NA
<i>W. & S. EUROPE</i>	<i>MIDDLE EAST</i>	<i>N. CAUSACUS</i>	<i>S. CAUCASUS</i>	<i>ALL CAUCASUS</i>	<i>TURKIC WEST</i>	<i>TURKIC EAST</i>
0.276764088	0.277682275	0.277930050	0.276234314	0.277230353	0.280335472	0.290416455
-0.000496618	-0.001414805	-0.001662579	3.32E-05	-0.000962883	-0.004068002	-0.014148985
-0.001276323	-0.002194511	-0.002442285	-0.000746549	-0.001742588	-0.004847707	-0.014928690
0.000524624	-0.000393564	-0.000641338	0.001054398	5.84E-05	-0.003046761	-0.013127743
0.000387002	-0.000531186	-0.00077896	0.000916776	-7.93E-05	-0.003184383	-0.013265365
0.002406108	0.001487921	0.001240146	0.002935882	0.001939843	-0.001165276	-0.011246259
NA	-0.000918187	-0.001165962	0.000529774	-0.000466265	-0.003571384	-0.013652367
1.58E-05	NA	-0.000247774	0.001447961	0.000451922	-0.002653197	-0.012734180
2.10E-20	0.133613450	NA	0.001695736	0.000699697	-0.002405422	-0.012486405
1.81E-05	1.25E-10	0	NA	-0.000996039	-0.004101158	-0.014182141
2.43E-05	0.019385810	0	2.36E-51	NA	-0.003105119	-0.013186102
6.39E-138	7.53E-33	1.32E-65	3.41E-275	6.82E-151	NA	-0.010080983
0	0	0	0	0	0	NA



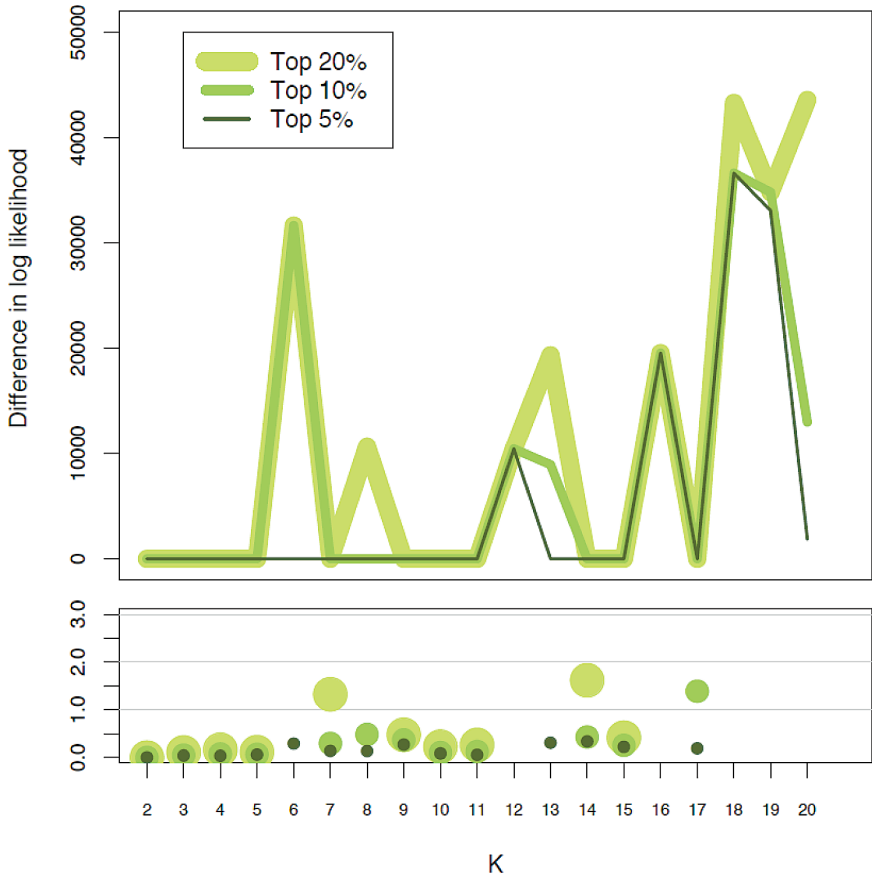
Supplemental Figure S1. Scatter plot of the first and second principal components for all samples included in the study. For a high-resolution version of this figure, please visit the following web address: <http://digitalcommons.wayne.edu/cgi/viewcontent.cgi?filename=5&article=2036&context=humbiol&type=additional>



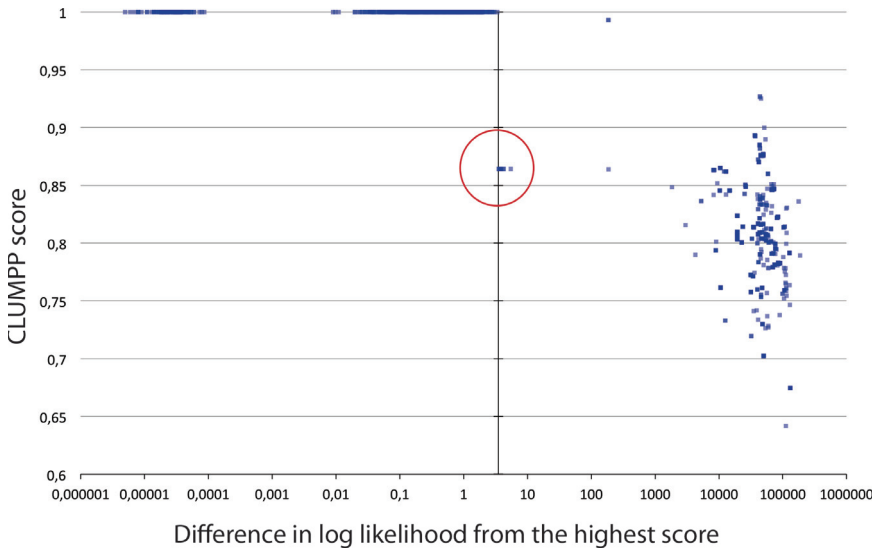
Supplemental Figure S2. Scatter plot of the first and third principal components for all samples included in the study. For a high-resolution version of this figure, please visit the following web address: <http://digitalcommons.wayne.edu/cgi/viewcontent.cgi?filename=5&article=2036&context=humbiol&type=additional>



Supplemental Figure S4. Cross-validation errors of the Admixture runs at $K = 2-20$, with magnification for $K = 6-14$.

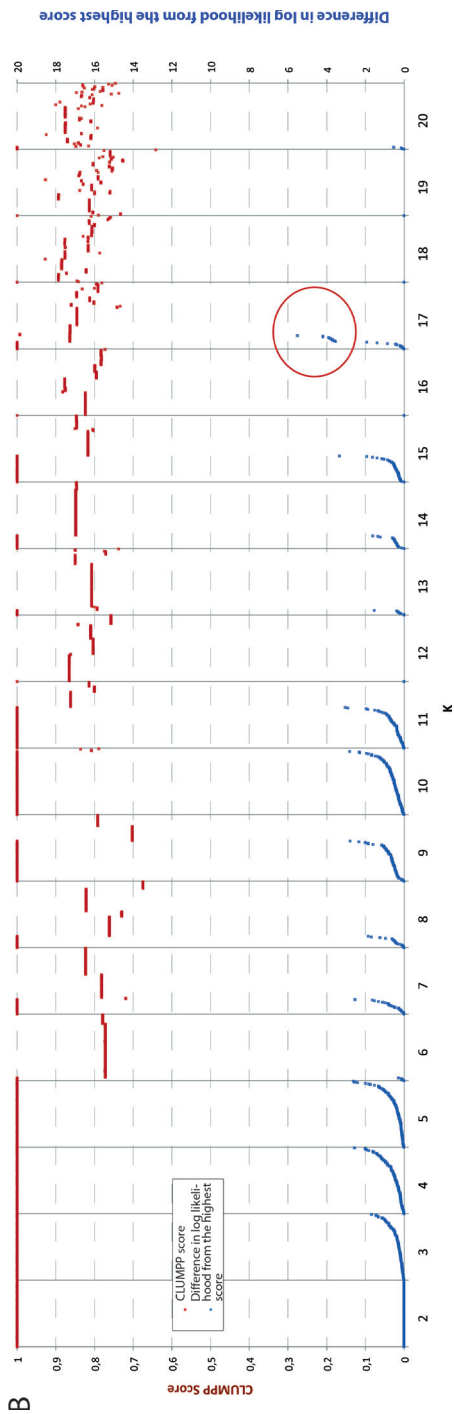
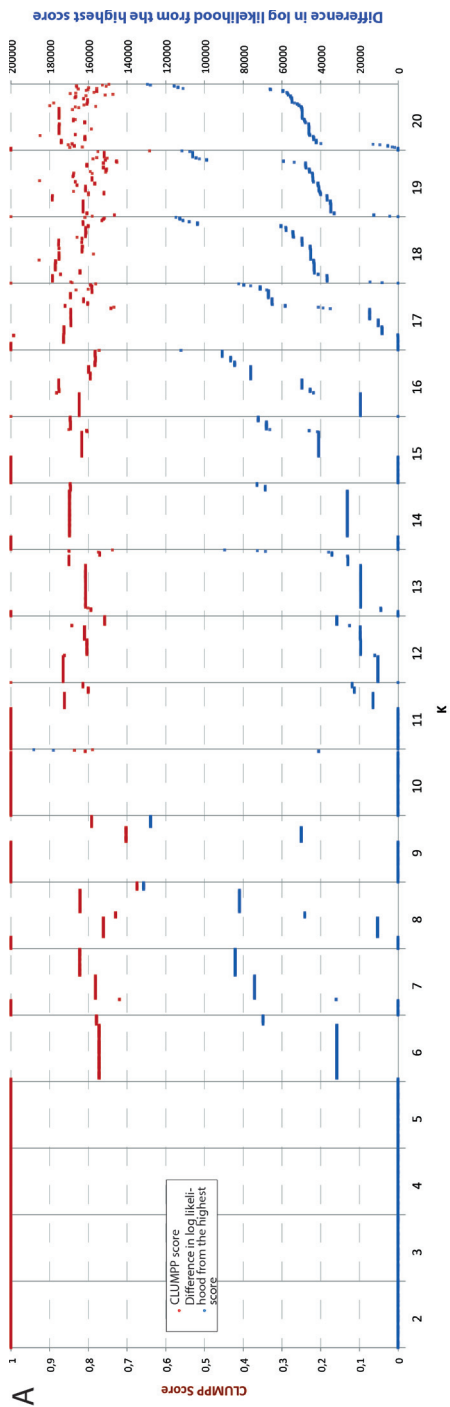


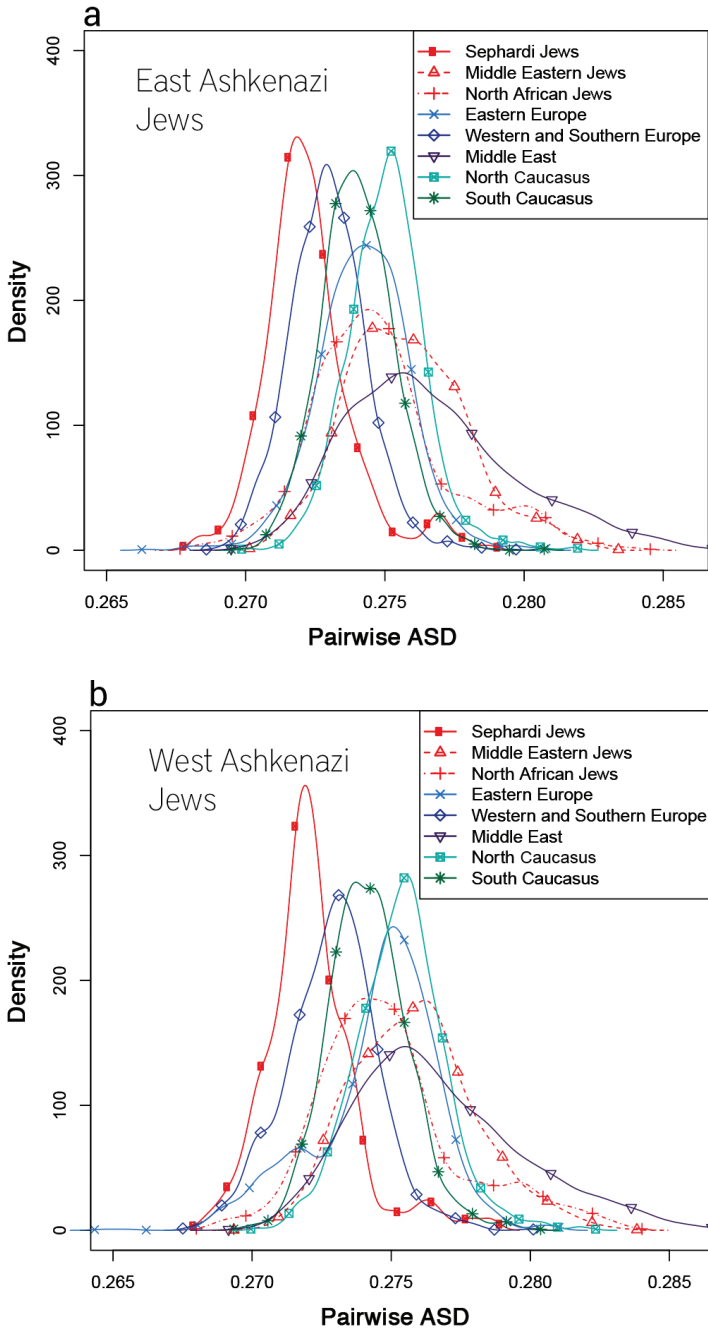
Supplemental Figure S5. Maximum difference in log-likelihood scores in fractions (0.05, 0.1, 0.2) of runs with the highest log-likelihood scores.



Supplemental Figure S6. Clumpp scores versus log-likelihood (LL) differences. The x-axis shows LL differences to the run arriving at the highest LL score as a function of K . Provided the LL difference is <100 , the Clumpp score is ≥ 0.9999 . The highlighted exceptions (red circle) come from $K = 17$, where 10 runs display low LL differences from the best run (3.5–5.5 LL units) and at the same time show a Clumpp score well below 1. Note, however, that at $K = 17$, 11 runs reach a Clumpp score of ~ 1 and are within 2 LL units from the best run. For a high-resolution version of this figure, please visit the following web address: <http://digitalcommons.wayne.edu/cgi/viewcontent.cgi?filename=7&article=2036&context=humbiol&type=additional>

Supplemental Figure S7 (opposite). Clumpp scores (right y-axis) versus log-likelihood (LL) differences (left y-axis; A, 0–200,000; B, 0–20) for different values of K (x-axis). Data points show all runs, sorted by LL-score difference to the run arriving at the highest LL score among the runs at a particular K . Similar LL scores translate to similar Clumpp scores. However, no direct relation exists between LL scores and Clumpp scores; an increase in the LL scores difference can result in higher Clumpp scores (see $K = 7$ in A). However, if the Clumpp score is ~ 1 , then the LL difference is <20 . Exceptions in this respect at $K = 17$ (see Supplemental Figure S6) are highlighted. In B the y-axis is zoomed to highlight LL score differences <20 . For a high-resolution version of this figure, please visit the following web address: <http://digitalcommons.wayne.edu/cgi/viewcontent.cgi?filename=8&article=2036&context=humbiol&type=additional>





Supplemental Figure S8. Density plot of the pairwise allele-sharing distances between individuals from a specific population and individuals from a population group.