Genome Biology and Evolution Advance Access published December 14, 2012 doi:10.1093/gbe/evs119

Research Article

The Missing Link of Jewish European Ancestry: Contrasting the Rhineland and the Khazarian Hypotheses

Eran Elhaik[1,2]

[1] Department of Mental Health, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA, 21208.

[2] McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA, 21208.

Running head: The Missing Link of Jewish European Ancestry

Keywords: Jewish genome, Khazars, Rhineland, Ashkenazi Jews, population isolate, Eastern European Jews, Central European Jews, Population structure,

Please address all correspondence to Eran Elhaik at eelhaik@jhsph.edu

Phone: 410-502-5740. Fax: 410-502-7544.

Abstract

The question of Jewish ancestry has been the subject of controversy for over two centuries and has yet to be resolved. The "Rhineland Hypothesis" depicts Eastern European Jews as a "population isolate" that emerged from a small group of German Jews who migrated eastward and expanded rapidly. Alternatively, the "Khazarian Hypothesis" suggests that Eastern European Jew descended from the Khazars, an amalgam of Turkic clans that settled the Caucasus in the early centuries CE and converted to Judaism in the 8th century. Mesopotamian and Greco-Roman Jews continuously reinforced the Judaized Empire until the 13th century. Following the collapse of their empire, the Judeo-Khazars fled to Eastern Europe. The rise of European Jewry is therefore explained by the contribution of the Judeo-Khazars. Thus far, however, the Khazar's contribution has been estimated only empirically, as the absence of genome-wide data from Caucasus populations precluded testing the Khazarian Hypothesis. Recent sequencing of modern Caucasus populations prompted us to revisit the Khazarian Hypothesis and compare it with the Rhineland Hypothesis. We applied a wide range of population genetic analyses to compare these two hypotheses. Our findings support the Khazarian Hypothesis and portray the European Jewish genome as a mosaic of Caucasus, European, and Semitic ancestries, thereby consolidating previous contradictory reports of Jewish ancestry. We further describe major difference among Caucasus populations explained by early presence of Judeans in the Southern and Central Caucasus. Our results have important implications on the demographic forces that shaped the genetic diversity in the Caucasus and medical studies.

Introduction

Contemporary Eastern European Jews comprise the largest ethno-religious aggregate of modern Jewish communities, accounting for nearly 90% of over 13 million Jews worldwide (United Jewish Communities 2003). Speculated to have emerged from a small Central European founder group and thought to have maintained high endogamy, Eastern European Jews are considered a "population isolate" and invaluable subjects in disease studies (Carmeli 2004), although their ancestry remains debatable between geneticists, historians, and linguists (Wexler 1993; Brook 2006; Sand 2009; Behar et al. 2010). Recently, several large-scale studies have attempted to chart the genetic diversity of Jewish populations by genotyping Eurasian Jewish and non-Jewish populations (Conrad et al. 2006; Kopelman et al. 2009; Behar et al. 2010). Interestingly, some of these studies linked Caucasus populations with Eastern European Jews, at odds with the narrative of a Central European founder group. Because correcting for population structure and using suitable controls are critical in medical studies, it is vital to examine the hypotheses pertaining to explain the ancestry of Eastern and Central European Jews. One of the major challenges for any hypothesis is to explain the massive presence of Jews in Eastern Europe, estimated at eight million people at the beginning of the 20th century. We investigate the genetic structure of European Jews, by applying a wide range of analyses — including 3 population test, principal component, biogeographical origin, admixture, identity by descent, allele sharing distance, and uniparental analyses — and test their veracity in light of the two dominant hypotheses depicting either a sole Middle Eastern ancestry or a mixed Middle Eastern-Caucasus-European ancestry to explain the ancestry of Eastern European Jews.

The "Rhineland Hypothesis" envisions modern European Jews to be the descendents of the Judeans - an assortment of Israelite-Canaanite tribes of Semitic origin (Figures 1,2) (Supplementary Note 1). It proposes two mass migratory waves: the first occurred over the two hundred years following the Muslim conquest of Palestine (638 CE) and consisted of devoted Judeans who left Muslim Palestine for Europe (Dinur 1961). Whether these migrants joined the existing Judaized Greco-Roman communities is unclear, as is the extent of their contribution to

the Southern European gene pool. The second wave occurred at the beginning of the 15th century by a group of 50,000 German Jews who migrated eastward and ushered an apparent hyper-baby-boom era for half a millennia (Atzmon et al. 2010). The Rhineland Hypothesis predicts a Middle Eastern ancestry to European Jews and high genetic similarity among European Jews (Ostrer 2001; Atzmon et al. 2010; Behar et al. 2010).

The competing "Khazarian Hypothesis" considers Eastern European Jews to be the descendants of Khazars (Supplementary Note 1). The Khazars were a confederation of Slavic, Scythian, Hunnic-Bulgar, Iranian, Alans, and Turkish tribes who formed in the central-northern Caucasus one of most powerful empires during the late Iron Age and converted to Judaism in the 8th century CE (Figures 1-2) (Polak 1951; Brook 2006; Sand 2009). The Khazarian, Armenian, and Georgian populations forged from this amalgamation of tribes (Polak 1951) followed by relative isolation, differentiation, and genetic drift *in situ* (Balanovsky et al. 2011). Biblical and archeological records allude to active trade relationships between Proto-Judeans and Armenians in the late centuries BCE (Polak 1951; Finkelstein and Silberman 2002), that likely resulted in a small scale admixture between these populations and a Judean presence in the Caucasus. After their conversion to Judaism, the population structure of the Judeo-Khazars was further reshaped by multiple migrations of Jews from the Byzantine Empire and Caliphate to the Khazarian Empire (Figure 1). Following the collapse of their Empire and the Black Death (1347-1348) the Judeo-Khazars fled eastwards (Baron 1993), settling in the rising Polish Kingdom and Hungary (Polak 1951) and eventually spreading to Central and Western Europe. The Khazarian Hypothesis posits that European Jews are comprised of Caucasus, European, and Middle Eastern ancestries. Moreover, European Jewish communities are expected to be different from one another both in ancestry and genetic heterogeneity. The Khazarian Hypothesis also offers two explanations for the genetic diversity in Caucasus groups first by the multiple migration waves to Khazaria during the 6th to 10th centuries and second by the Judeo-Khazars who remained in the Caucasus.

Genetic studies attempting to infer the ancestry of European Jews yielded inconsistent results. Some studies pointed to the genetic similarity between European Jews and Caucasus populations like Adygei (Behar et al. 2003; Levy-Coffman 2005; Kopelman et al. 2009), while some pointed to the similarity to Middle Eastern populations such as Palestinians (Hammer et al. 2000; Nebel et al. 2000), and others pointed to the similarity to Southern European populations like Italians (Atzmon et al. 2010; Zoossmann-Diskin 2010). Most of these studies were done in the pre-genomewide era using uniparental markers and included different reference populations, which makes it difficult to compare their results. More recent studies employing whole genome data reported high genetic similarity of European Jews to Druze, Italian, and Middle Eastern populations (Atzmon et al. 2010; Behar et al. 2010).

Although both the Rhineland and Khazarian Hypotheses depict a Judean ancestry and are not mutually exclusive, they are well distinguished, as Caucasus and Semitic populations are considered ethnically and linguistically distinct (Patai and Patai 1975; Wexler 1993; Balanovsky et al. 2011). Jews, according to either hypothesis, are an assortment of tribes who accepted Judaism, migrated elsewhere, and maintained their religion up to this date and are, therefore, expected to exhibit certain differences from their neighboring populations. Because, according to both hypotheses, Eastern European Jews arrived in Eastern Europe roughly at the same time (13th and 15th centuries), we assumed that they experienced similar low and fixed admixture rates with the neighboring populations, estimated at 0.5% per generation over the past 50 generations (Ostrer 2001). These relatively recent admixtures have likely reshaped the population structure of all European Jews and increased the genetic distances from the Caucasus or Middle Eastern populations. Therefore, we do not expect to achieve perfect matching with the surrogate Khazarian and Judean populations but rather to estimate their relatedness.

Materials and Methods

Data collection. The complete data set contained 1,287 unrelated individuals of 8 Jewish and 74 non-Jewish populations genotyped over 531,315 autosomal single nucleotide polymorphisms (SNPs). An LD-pruned data set was created by removing one member of any pair of SNPs in strong LD (r^2>0.4) in windows of 200 SNPs (sliding the window by 25 SNPs at a time) using *indep-pairwise* in *PLINK* (Purcell et al. 2007). This yielded a total of 221,558 autosomal SNPs that were chosen for all autosomal analyses except the identical by descent analysis that utilized the complete data set. Both data sets were obtained from http://www.evolutsioon.ut.ee/MAIT/jew_data/ (Behar et al. 2010). MtDNA and Y-chromosomal data were obtained from previously published data sets as appear in Behar et al. (2010). These markers were chosen to match the phylogenetic level of resolution achieved in previously reported data sets and represent a diversified set of markers. A total of 11,392 samples were assembled for mtDNA (6,089) and Y-chromosomal (5,303) analyses from 27 populations (Tables S1-2).

Terminology. In common parlance, Eastern and Central European Jews are practically synonymous with Ashkenazi Jews and are considered a single entity (Tian et al. 2008; Atzmon et al. 2010; Behar et al. 2010). However, the term is misleading, for the Hebrew word "Ashkenaz" was applied to Germany in medieval rabbinical literature - contributing to the narrative that modern Eastern European Jewry originated on the Rhine. We thus refrained from using the term "Ashkenazi Jews." Jews were roughly subdivided into Eastern (Belorussia, Latvia, Poland, and Romania) and Central (Germany, Netherland, and Austria) European Jews. In congruence with the literature that considers "Ashkenazi Jews" distinct from "Sephardic Jews," we excluded the later. Complete population notation is described in Table S3.

Choice of surrogate populations. As the ancient Judeans and Khazars have been vanquished and their remains have yet to be sequenced, in accordance with previous studies (Levy-Coffman

2005; Kopelman et al. 2009; Atzmon et al. 2010; Behar et al. 2010), contemporary Middle Eastern and Caucasus populations were used as surrogates. Palestinians were considered proto-Judeans because they are assumed to share a similar linguistic, ethnic, and geographic background with the Judeans and were shown to share common ancestry with European Jews (Bonné-Tamir and Adam 1992; Nebel et al. 2000; Atzmon et al. 2010; Behar et al. 2010). Similarly, Caucasus Georgians and Armenians were considered proto-Khazars because they are believed to have emerged from the same genetic cohort as the Khazars (Polak 1951; Dvornik 1962; Brook 2006).

The 3 population test. The *f3* statistics uses allele frequency differences to assess the presence of admixture in a population *X* from two other populations *A* and *B*, so that *f3*(*X*; *A*, *B*) (Reich et al. 2009). If *X* is a mixture of *A* and *B,* rather than the result of genetic drift*, f3* would be negative. A significant negative *f3* indicates that the ancestors of group *X* experienced a history of admixture subsequent to their divergence from *A* and *B*. The *f3* statistics were calculated with the *threepop* program of *TreeMix* (Pickrell and Pritchard 2012) with k=500 over the set of 221,558 SNPs. This test differs from ADXMITURE (Alexander, Novembre, and Lange 2009), which reports the proportions of admixture with the most likely ancestor.

Principal component analysis (PCA). Although the commonly used "multi-population" PCA has many attractive properties, it should be practiced with caution to avoid biases due to the choice of populations and varying sample sizes (Price et al. 2006; McVean 2009). To circumvent these biases, we developed a simple "dual-population" framework consisting of three "outgroup" populations that are available in large sample sizes and are the least admixed — Mbuti and Biaka Pygmies (South Africa), French Basques (Europe), and Han Chinese (East Asia) — and two populations of interest, all of equal sample sizes. The cornerstone of this framework is that it minimizes the number of significant PCs to four or fewer (Tracy-Widom test, p<0.01) and maximizes the portion of explained variance to over 20% for the first two PCs. PCA calculations were carried out using smartpca of the *EIGENSOFT* package (Patterson, Price, and Reich 2006).

Convex hulls were calculated using Matlab "convhull" function and plotted around the cluster centroids. Relatedness between two populations of interest was estimated by the commensurate overlap of their clusters. Small populations (<7 samples) were excluded from the analysis.

Estimating the biogeographical origins of population. Novembre et al. (2008) proposed a PCA-based approach, accurate to a few hundred kilometers within Europe, to identify the current biogeographical origin of a population. Although this approach has no implied historical model, it correlates genetic diversity with geography and can thus be a useful tool to study biogeography. To decrease the bias caused by multiple populations of uneven sizes (Patterson, Price, and Reich 2006; McVean 2009), we adopted the dual-population framework with three outgroup populations and two populations of interest: a population of known geographical origin during the relevant time period shown to cluster with the population in question (e.g., Armenians) and the population in question (e.g., Eastern European Jews). The first four populations were used as a training set for the population in question. PCA calculations were carried out as described above. The rotation angle of PC1-PC2 coordinates was calculated as described by Novembre et al. (2008). Briefly, in each figure the PC axes were rotated to find the angle that maximizes the summed correlation of the median PC1 and PC2 values of the training populations with the latitude and longitude of their countries. Latitudinal and longitudinal data were obtained from the literature or by the country's approximate centroid. Geodesic distances were calculated in kilometers using the Matlab function "distance."

Admixture analysis. A structure-like approach was applied in a *supervised* learning mode as implemented in *ADMIXTURE* (Alexander, Novembre, and Lange 2009). *ADMIXTURE* provides an estimation of the individual's ancestries from the allele frequencies of the designated ancestral populations. *ADMIXTURE*'s bootstrapping procedure with default parameters was used to calculate the standard errors. We observed low (<0.05) standard errors in all our analyses. Populations were sorted by their mean African and Asian ancestries. In this analysis, the three Netherland Jews were grouped with Eastern European Jews.

Identity by descent (IBD) analysis. To detect IBD segments, we ran *fastIBD* ten times using different random seeds and combined the results as described by Browning and Browning (2011). Segments were considered to be IBD only if the *fastIBD* score of the combined analysis was less than e^{-10}. This low threshold corresponds to long shared haplotypes (≥1 cM) that are likely to be IBD. Short gaps (<50 indexes) separating long domains were assumed to be false-negative and concatenated (Browning and Browning 2011). Pairwise-IBD segments between European Jews and different populations were obtained by finding the maximum total IBD sharing between each European Jew and all other individuals of a particular population.

Allele sharing distances (ASD). ASD was used for measuring genetic distances between populations as it is less sensitive to small sample sizes than other methods. Pairwise ASD was calculated using PLINK (Purcell et al. 2007), and the average ASD between populations *I* and *J*, was computed as:

$$W^{IJ} = (\sum_{i \in I} \sum_{i \in J} W^{ij}) / nm, \quad (1)$$

where W^{ij} is the distance between individuals *i* and *j* from populations *I* and *J* of sizes *n* and *m*, respectively. To verify that these ASD differences are significant, a bootstrap approach was used with the null hypothesis: H_0: ASD (p_1, p_2) = ASD (p_1, p_3), where the ASD between populations p_1 and p_2 is compared to the ASD between populations p_2 and p_3 (Supplementary Note 2). To compare continental Jewish communities, individuals were grouped by their continent and the comparison was carried as described.

Uniparental analysis. To infer the migration patterns of European Jews, we integrated haplogroup data from over 11,300 uniparental chromosomes with geographical data. The haplogroup frequencies were compared between populations to obtain a measure of distance between populations. Pairwise genetic distances between population haplogroups (Tables S1-2)

were estimated by applying the Kronecker function as implemented in *Arlequin* version 3.1 (Excoffier, Laval, and Schneider 2005). In brief, similarity between populations was defined as the fraction of haplogroups that the two populations shared as measured by the Kronecker function $\delta_{xy}(i)$:

$$d_{xy} = \sum_{i=1}^{L} \delta_{xy}(i), \quad (2)$$

which equals 1 if the haplogroup frequency of the *i*-th haplogroup is non-zero for both populations and equals 0 otherwise. In other words, populations sharing the same exact haplogroups or their mutual absence are considered more genetically similar than populations with different haplogroups. For brevity, we considered only haplogroups with frequencies higher than 0.5%. This measure has several desirable properties that make it an excellent measure for estimating genetic distance between populations, such as a simple interpretation in terms of homogeneity and applicability to both mtDNA and Y-chromosomal data.

Results

To confirm that the Rhineland and Khazarian Hypotheses indeed portray distinct ancestries, we assessed the degree of background admixture between Caucasus and Semitic populations. We calculated the *f3* statistics between Palestinians and six Caucasus and Eurasian populations using African San as outgroup, for example, *f3*(*Palestinians*, *San*, *Armenians*). The *f3* results for Turks (–0.0013), Armenians and Georgians (–0.0019), Lezgins and Adygei (–0.0015), and Russians (–0.0011) indicated a minor but significant admixture (–26<*Z-score*<–13) between Palestinians and the populations tested. Because Armenians and Georgians diverged from Turks 600 generations ago (Schonberg et al. 2011) we can assume that the lion's share of their admixture derived from that ancestry and within the expected levels of background admixture typical to the region rather than recent admixture with Semitic populations. Therefore, similarities between European Jews and Caucasus populations will unlikely to be due to a shared Semitic ancestry.

PCA was next used to identify independent dimensions that capture most of the information in the data. PCA was applied using two frameworks: the "multi-population" carried for all populations (Figure 3) and separately for Eurasian populations along with Pygmies and Han Chinese (Figure S2) and our novel "dual-population" framework (Figure S3). In all analyses, the studied samples aligned along the two well-established geographic axes of global genetic variation: PC1 (sub-Saharan Africa versus the rest of the Old World) and PC2 (east versus west Eurasia) (Li et al. 2008). Our results reveal geographically refined groupings, such as the nearly symmetrical continuous European rim extending from Western to Eastern Europeans, the parallel Caucasus rim, and the Near Eastern populations (Figure S1) organized in Turk-Iranian and Druze clusters (Figure 3). Middle Eastern populations form a gradient along the diagonal line between Bedouins and Near Eastern populations that resembles their geographical distribution. The remaining Egyptians and the bulk of Saudis distribute separately from Middle Eastern populations.

European Jews are expected to cluster with native Middle Eastern or Caucasus populations according to the Rhineland or Khazarian Hypotheses, respectively. The results of all PC analyses (Figures 3, S2-3) show that over 70% of European Jews and almost all Eastern European Jews cluster with Georgian, Armenian, and Azerbaijani Jews within the Caucasus rim (Figures 3, S3). Nearly 15% of Central European Jews cluster with Druze and the rest cluster with Cypriots. All European Jews cluster distinctly from the Middle Eastern cluster. A strong evidence for the Khazarian Hypothesis is the clustering of European Jews with the populations that reside on opposite ends of ancient Khazaria: Armenians, Georgians, and Azerbaijani Jews (Figure 1). Because Caucasus populations remained relatively isolated in the Caucasus region and because there are no records of Caucasus populations mass-migrating to Eastern and Central Europe prior to the fall of Khazaria (Balanovsky et al. 2011), these findings imply a shared origin for European Jews and Caucasus populations.

To assess the ability of our PCA-based approach to identify the biogeographical origins of a population, we first sought to identify the biogeographical origin of Druze. The Druze religion originated in the 11th century, but the people's origins remain a source of much confusion and debate (Hitti 1928). We traced Druze biogeographical origin to the geographical coordinates: 38.6±3.45° N, 36.25±1.41° E (Figure S4) in the Near East (Figure S1). Half of the Druze clustered tightly in Southeast Turkey, and the remaining was scattered along northern Syria and Iraq. These results are in agreement with Shlush et al.'s (2008) findings using mtDNA anaysis. The inferred geographical positions of Druze were used in the subsequent analyses.

The geographical origins of European Jews varied for different reference populations (Figure 4, S5), but all the results converged to Southern Khazaria along modern Turkey, Armenia, Georgia, and Azerbaijan. Eastern European Jews clustered tightly compared to Central European Jews in all analyses. The smallest deviations in the geographical coordinates were obtained with Armenians for both Eastern (38±2.7° N, 39.9±0.4° E) and Central (35±5° N, 39.7±1.1° E) European Jews (Figure 4). Similar results were obtained for Georgians (Figure S5). Remarkably, the mean coordinates of Eastern European Jews are 560 kilometers from Khazaria's southern border (42.77° N, 42.56° E) near Samandar - the capital city of Khazaria from 720 to 750 CE (Polak 1951).

The duration, direction, and rate of gene flow between populations determine the proportion of admixture and the total length of chromosomal segments that are identical by descent (IBD). Admixture calculations were carried out using a *supervised* learning approach in a structure-like analysis. This approach has many advantages over the *unsupervised* approach that not only traces ancestry to *K* abstract unmixed populations under the assumption that they evolved independently (Chakravarti 2009; Weiss and Long 2009) but also is problematic when applied to study Jewish ancestry, which can be dated only as far back as 3,000 years (Figure 2). Moreover, the results of the *unsupervised* approach vary based on the particular populations used for the analysis and the choice of *K*, rendering the results incomparable between studies. Admixture was

calculated with a reference set of seven populations representing genetically distinct regions: Pygmies (South Africa), French Basque (West Europe), Chuvash (East Europe), Han Chinese (East Asia), Palestinians (Middle East), Turk-Iranians (Near East), and Armenians (Caucasus) (Figure 5). The ancestral components grouped all populations by their geographical regions with European Jews clustering with Caucasus populations. As expected, Eastern and Western European ancestries exhibit opposite gradients among European populations. The Near Eastern-Caucasus ancestries are dominant among Central (38%) and Eastern (32%) European Jews followed by Western European ancestry (30%). Among non-Caucasus populations, the Caucasus ancestry is the largest among European Jews (26%) and Cypriots (31%). These populations also exhibit the largest fraction of Middle Eastern ancestry among non-Middle Eastern populations. As both Caucasus and Middle Eastern ancestries are absent in Eastern European populations, our findings suggest that Eastern European Jews acquired these ancestries prior to their arrival to Eastern Europe. Although the Rhineland Hypothesis explains the Middle Eastern ancestry by stating that Jews migrated from Palestine to Europe in the 7th century, it fails to explain the large Caucasus ancestry, which is nearly endemic to Caucasus populations.

Although they cluster with Caucasus populations (Figure 5), Eastern and Central European Jews share a large fraction of Western European and Middle Eastern ancestries, both absent in Caucasus populations, excepting Armenians who share 15±2% Middle Eastern ancestry (Figure S6). According to the Khazarian Hypothesis, the Western European ancestry was imported to Khazaria by Greco-Romans Jews, whereas the Middle Eastern ancestry alludes to the contribution of both early Israelite Proto-Judeans as well as Mesopotamian Jews (Polak 1951; Koestler 1976; Sand 2009). Central and Eastern European Jews differ mostly in their Middle Eastern (30% and 25%, respectively) and Eastern European ancestries (3% and 12%, respectively), probably due to late admixture.

Druze exhibit a large Turkic ancestry (83%) in accordance with their Near Eastern origin (Figure S4). Druze and Cypriot appear similar to European Jews in their Middle Eastern and Western

European ancestries, though they differ largely in the proportion of Caucasus ancestry. These results can explain the genetic similarity between European Jews, Southern Europeans, and Druze reported in studies that excluded Caucasus populations (Price et al. 2008; Atzmon et al. 2010; Zoossmann-Diskin 2010). Overall, our results portray the European Jewish genome as a mosaic of Caucasus, Western European, Middle Eastern, and Eastern European ancestries in decreasing proportions.

To glean further details of the genomic regions contributing to the genetic similarity between European Jews and the perspective populations, we compared their total genomic regions shared by IBD. If European Jews emerged from Caucasus populations, the two would share longer IBD regions than with Middle Eastern populations. The IBD analysis exhibits a skewed bimodal distribution embodying a major Caucasus ancestry with a minor Middle Eastern ancestry (Figure 6), consistent with the admixture results (Figure 5). The total IBD regions shared between European Jews and Caucasus populations (9.5 cM on average) are significantly larger than regions shared with Palestinians (5.5 cM) (Kolmogorov-Smirnov goodness-of-fit test, $p<0.001$). To the best of our knowledge, these are the largest IBD regions ever reported between European Jews and non-Jewish populations. The decrease in total IBD between European Jews and other populations combined with the increase in distance from the Caucasus support the Khazarian Hypothesis.

We next estimated the level of endogamy among Eurasian Jewish communities and compared their genetic distances to non-Jewish neighbors, Caucasus, and Middle Eastern populations. Our results expand the previous report of high endogamy in Jewish populations (Behar et al. 2010) and narrow the endogamy to regional Jewish communities (Table 1, left panel). Jews are significantly more similar to members of their own community than to other Jewish populations ($P<0.01$, bootstrap *t*-test), with the conspicuous exception of Bulgarian, Turkish, and Georgian Jews. These results stress the high heterogeneity among Jewish communities across Eurasia and even within communities, as in the case of the Balkan and Caucasus Jews.

When compared to non-Jewish populations, all Jewish communities were significantly ($P<0.01$, bootstrap *t*-test) distant from Middle Eastern populations and, with the exception of Central European Jews, significantly closer to Caucasus populations (Table 1, right panel). Similar findings were reported by Behar et al. (2010) although they were dismissed as "a bias inherent in our calculations." The close genetic distance between Central European Jews and Southern European populations can be attributed to a late admixture. The results are consistent with our previous findings in support of the Khazarian Hypothesis. As the only commonality among all Jewish communities is their dissimilarity from Middle Eastern populations (Table 1), grouping different Jewish communities without correcting for their country of origin, as is commonly done, would increase their genetic heterogeneity.

Finally, we carried uniparental analyses on mtDNA and Y-chromosome comparing the haplogroup frequencies between European Jews and other populations. The Rhineland Hypothesis depicts Middle Eastern origins for European Jews' both paternal and maternal ancestries, whereas the Khazarian Hypothesis depicts a Caucasus ancestry along with Southern European and Near Eastern contributions of migrates from Byzantium and the Caliphate, respectively. Because Judaism was maternally inherited only since the 3rd century CE (Patai and Patai 1975), the mtDNA is expected to show a stronger local female-biased founder effect compared to the Y-chromosome. Haplogroup similarities between European Jews and other populations were plotted as heat maps on the background of their geographical locations (Figure 7). The pairwise distances between all studied populations are shown in Figure S7.

Our results shed light on sex-specific processes that, although not evident from the autosomal data, are analogous to those obtained from the biparental analyses. Both mtDNA and Y-chromosomal analyses yield high similarities between European Jews and Caucasus populations rooted in the Caucasus (Figure 7) in support of the Khazarian Hypothesis. Interestingly, the maternal analysis depicts a specific Caucasus founding lineage with a weak Southern European

ancestry (Figure 7a), whereas the paternal ancestry reveals a dual Caucasus-Southern European origin (Figure 7b). As expected, the maternal ancestry exhibits a higher relatedness scale with narrow dispersal compared with the paternal ancestry.

Dissecting uniparental haplogroups allows us to delve further into European Jews' migration routes. As the results do not specify whether the Southern Europe-Caucasus migration was ancient or recent nor indicate the migration's direction i.e., from Southern Europe to the Caucasus or the opposite, there are four possible scenarios. Of these, the only historically supportable scenarios are ancient migrations from Southern Europe toward Khazaria (6^{th}-13^{th} centuries) and more recent migrations from the Caucasus to Central and Southern Europe (13^{th}-15^{th} centuries) (Polak 1951; Patai and Patai 1975; Straten 2003; Brook 2006; Sand 2009). A westward migration from the diminished Khazaria toward Central and Southern Europe would have exhibited a gradient from the Caucasus toward Europe for both matrilineal and patrilineal lines. Such a gradient was not observed. By contrast, Judaized Greco-Roman male-driven migration directly to Khazaria is consistent with historical demographic migrations and could have created the observed pattern. Moreover, we found little genetic similarity between European Jews and populations eastward to the Caspian Sea and southward to the Black Sea, delineating the geographical boundaries of Khazaria (Table 1, Figure 1).

Discussion

Eastern and Central European Jews comprise the largest group of contemporary Jews, accounting for nearly 90% of over 13 million worldwide Jews (United Jewish Communities 2003). Eastern European Jews made up over 90% of European Jews before World War II. Despite of their controversial ancestry, European Jews are an attractive group for genetic and medical studies due to their presumed genetic history (Ostrer 2001). Correcting for population structure and using suitable controls are critical in medical studies, thus it is vital to determine whether European Jews are of Semitic, Caucasus, or other ancestry.

Though Judaism was born encased in theological-historical myth, no Jewish historiography was produced from the time of Josephus Flavius (1st century CE) to the 19th century (Sand 2009). Early historians bridged the historical gap simply by linking modern Jews directly to the ancient Judeans (Figure 2); a paradigm that was later embedded in medical science and crystallized as a narrative. Many have challenged this narrative (e.g., Koestler 1976; Straten 2007), mainly by showing that a sole Judean ancestry cannot account for the vast population of Eastern European Jews in the beginning of the 20th century without the major contribution of Judaized Khazars and by demonstrating that it is in conflict with anthropological, historical, and genetic evidence (Patai and Patai 1975; Baron 1993; Sand 2009).

With uniparental and whole genome analyses providing ambiguous answers (Levy-Coffman 2005; Atzmon et al. 2010; Behar et al. 2010), the question of European Jewish ancestry remained debated between the supporters of the Rhineland and Khazarian Hypotheses. Although both theories oversimplify complex historical processes they are attractive due to their distinct predictions and testable hypotheses. We showed that the hypotheses are also genetically distinct and that the miniscule Semitic ancestry in Caucasus populations cannot account for the similarity between European Jews and Caucasus populations. The recent availability of genomic data from Caucasus populations allowed testing the Khazarian Hypothesis for the first time and prompted us to contrast it with the Rhineland Hypothesis. To evaluate the two hypotheses, we carried out a series of comparative analyses between European Jews and surrogate Khazarian and Judean populations posing the same question each time: are Eastern and Central European Jews genetically closer to Khazarian or Judean populations? Under the Rhineland Hypothesis, European Jews are also expected to exhibit high endogamy across all their communities and be more similar to Middle Eastern populations compared to their neighboring non-Jewish populations, whereas the Khazarian Hypothesis predicts the opposite scenario. We emphasize that these hypotheses are not exclusive and that some European Jews may have other ancestries.

Our PC, biogeographical estimation, admixture, IBD, ASD, and uniparental analyses were consistent in depicting a Caucasus ancestry for European Jews. Our first analyses revealed tight genetic relationship of European Jews and Caucasus populations and pinpointed the

biogeographical origin of European Jews to the south of Khazaria (Figures 3,4). Our later analyses yielded a complex ancestry with a slightly dominant Near Eastern-Caucasus ancestry, large Southern European and Middle Eastern ancestries, and a minor Eastern European contribution; the latter two differentiated Central and Eastern European Jews (Figures 4, 5 and Table 1). While the Middle Eastern ancestry faded in the ASD and uniparental analyses, the Southern European ancestry was upheld probably attesting to its later time period (Table 1 and Figure 7).

We show that the Khazarian Hypothesis offers a comprehensive explanation for the results, including the reported Southern European (Atzmon et al. 2010; Zoossmann-Diskin 2010) and Middle Eastern ancestries (Nebel et al. 2000; Behar et al. 2010). By contrast, the Rhineland Hypothesis could not explain the large Caucasus component in European Jews, which is rare in Non-Caucasus populations (Figure 5), and the large IBD regions shared between European Jews and Caucasus populations attesting to their common origins. Our findings thus reject the Rhineland Hypothesis and uphold the thesis that Eastern European Jews are Judeo-Khazars in origin. Consequently, we can conclude that the conceptualization of European Jews as a "population isolate," which is derived from the Rhineland Hypothesis, is incorrect and most likely reflects sampling bias in the lack of Caucasus non-Jewish populations in comparative analyses.

A major difficulty with the Rhineland Hypothesis, in addition to the lack of historical and anthropological evidence to the multi-migration waves from Palestine to Europe (Straten 2003; Sand 2009), is to explain the vast population expansion of Eastern European Jews from 50 thousand (15^{th} century) to 8 million (20^{th} century). The annual growth rate that accounts for this population expansion was estimated at 1.7-2%, one order of magnitude larger than that of Eastern European non-Jews in the 15^{th}-17^{th} centuries, prior to the industrial revolution (Straten 2007). This growth could not possibly be the product of natural population expansion, particularly one subjected to severe economic restrictions, slavery, assimilation, the Black Death and other plagues, forced and voluntary conversions, persecutions, kidnappings, rapes, exiles, wars, massacres, and pogroms (Koestler 1976; Straten 2003; Sand 2009). Because such an

unnatural growth rate, over half a millennia and affecting only Jews residing in Eastern Europe, is implausible - it is explained by a miracle (Atzmon et al. 2010; Ostrer 2012). Unfortunately, this divine intervention explanation poses a new kind of problem - it is not science. The question how the Rhineland Hypothesis, so deeply rooted in supernatural reasoning, became the dominant scientific narrative is debated among scholars (e.g., Sand 2009).

The most parsimonious explanation for our findings is that Eastern European Jews are of Judeo-Khazarian ancestry forged over many centuries in the Caucasus. Jewish presence in the Caucasus and later Khazaria was recorded as early as the late centuries BCE and reinforced due to the increase in trade along the Silk Road (Figure 1), the decline of Judah (1^{st}-7^{th} centuries), and the uprise of Christianity and Islam (Polak 1951). Greco-Roman and Mesopotamian Jews gravitating toward Khazaria were also common in the early centuries and their migrations were intensified following the Khazars' conversion to Judaism (Polak 1951; Brook 2006; Sand 2009). The eastward male-driven migrations (Figure 7) from Europe to Khazaria solidified the exotic Southern European ancestry in the Khazarian gene pool, (Figure 5) and increased the genetic heterogeneity of the Judeo-Khazars. The religious conversion of the Khazars encompassed most of the Empire's citizens and subordinate tribes and lasted for the next 400 years (Polak 1951; Baron 1993) until the invasion of the Mongols (Polak 1951; Dinur 1961; Brook 2006). At the final collapse of their empire (13^{th} century), many of the Judeo-Khazars fled to Eastern Europe and later migrated to Central Europe and admixed with the neighboring populations.

Historical and archeological findings shed light on the demographic events following the Khazars' conversion. During the half millennium of their existence (740-1250 CE), the Judeo-Khazars sent offshoots into the Slavic lands, such as Romania and Hungary (Baron 1993), planting the seeds of a great Jewish community to later rise in the Khazarian diaspora. We hypothesize that the settlement of Judeo-Khazars in Eastern Europe was achieved by serial founding events, whereby populations expanded from the Caucasus into Eastern and Central Europe by successive splits, with daughter populations expanding to new territories following changes in socio-political conditions (Gilbert 1993). These events may have contributed to the higher homogeneity observed in Jewish communities outside Khazaria's borders (Table 1).

After the decline of their Empire, the Judeo-Khazars refugees sought shelter in the emerging Polish Kingdom and other Eastern European communities where their expertise in economics, finances, and politics were valued. Prior to their exodus, the Judeo-Khazar population was estimated to be half a million in size, the same as the number of Jews in the Polish-Lithuanian kingdom four centuries later (Polak 1951; Koestler 1976). Some Judeo-Khazars were left behind, mainly in the Crimea and the Caucasus, where they formed Jewish enclaves surviving into modern times. One of the dynasties of Jewish princes ruled in the 15th century under the tutelage of the Genovese Republic and later of the Crimean Tartars. Another vestige of the Khazar nation are the "Mountain Jews" in the North Eastern Caucasus (Koestler 1976).

The remarkable close proximity of European Jews and populations residing on the opposite ends of ancient Khazaria, such as Armenians, Georgians, Azerbaijani Jews, and Druze (Figures 3, S2-3, 5), supports a common Near Eastern-Caucasus ancestry. These findings are not explained by the Rhineland Hypothesis and are staggering due to the uneven demographic processes these populations have experienced in the past eight centuries. The high genetic similarity between European Jews and Armenian compared to Georgians (Figures 5 and 6, Table 1) is particularly bewildering because Armenians and Georgians are very similar populations that share a similar genetic background (Schonberg et al. 2011) and long history of cultural relations (Payaslian 2007). We identified a small Middle Eastern ancestry in Armenians that does not exist in Georgians and is likely responsible to the high genetic similarity between Armenians and European Jews (Figure S6). Because the Khazars blocked the Arab approach to the Caucasus, we suspect that this ancestry was introduced by the Judeans arriving at a very early date to Armenia and was absorbed into the populations, whereas Judeans arriving to Georgia avoided assimilation (Shapira 2007). Similarly, the relatedness between European Jews and Druze reported here and in the literature (Behar et al. 2010) is explained by Druze Turkish-Southern Caucasus origins. Druze migrated to Syria, Lebanon, and eventually to Palestine between the 11th and 13th centuries during the Crusades, a time when the Jewish population in Palestine was at minimum. The genetic similarity between European Jews and Druze therefore supports the Khazarian Hypothesis and should not be confused with a Semitic origin, which can be easily distinguished from the non-Semitic origin (Figure 5). We emphasize that testing the Middle Eastern origin of

European Jews can only be done with indigenous Middle Eastern groups. Overall, the similarity between European Jews and Caucasus populations underscores the genetic continuity that exists among Eurasian Jewish and non-Jewish Caucasus populations.

This genetic continuity is not surprising. The Caucasus gene pool proliferated from the Near Eastern pool due to an Upper Paleolithic (or Neolithic) migration and was shaped by significant genetic drift, due to relative isolation in the extremely mountainous landscape (Balanovsky et al. 2011; Pagani et al. 2011). Caucasus populations are therefore expected to be genetically distinct from Southern European and Middle Eastern populations (Figure 5) but to share certain genetic similarity with Near Eastern populations such as Turks, Iranians, and Druze. In all our analyses, Middle Eastern samples clustered together or exhibited high similarity along a geographical gradient (Figure 3) and were distinguished from Arabian Peninsula Arab samples on one hand and from Near Eastern - Caucasus samples on the other hand.

Our study attempts to shed light on the forgotten Khazars and elucidate some of the most fascinating questions of their history. Although the Khazars' conversion to Judaism is not in dispute, there are questions as to how widespread and established the new religion became. Despite the limited sample size of European Jews, they represent members from the major residential Jewish countries (i.e., Poland and Germany) and exhibit very similar trends. Our findings support a large-scale migration from South-Central Europe and Mesopotamia to Khazaria that reshaped the genetic structure of the Khazars and other Caucasus populations in the central and upper Caucasus. Another intriguing question touches upon the origins of the Khazars, speculated to be Turk, Tartar, or Mongol (Brook 2006). As expected from their common origin, Caucasus populations exhibit high genetic similarity to Iranian and Turks with mild Eastern Asian ancestry (Figure 5, Figure S7). However, we found a weak patrilineal Turkic contribution compared to Caucasus and Eastern European contributions (Figure 7). Our findings thus support the identification of Turks as the Caucasus' ancestors but not necessarily the predominant ancestors. Given their geographical position, it is likely the Khazarian gene pool was also influenced by Eastern European populations that are not represented in our dataset.

Our results fit with evidence from a wide range of fields. Linguistic findings depict Eastern European Jews as descended from a minority of Israelite-Palestinian Jewish emigrates who intermarried with a larger heterogeneous population of converts to Judaism from the Caucasus, the Balkans, and the Germano-Sorb lands (Wexler 1993). Yiddish, the language of Central and Eastern European Jews, began as a Slavic language that was re-lexified to High German at an early date (Wexler 1993). Our findings are also in agreement with archeological, historical, linguistic, and anthropological studies (Polak 1951; Patai and Patai 1975; Wexler 1993; Brook 2006; Kopelman et al. 2009; Sand 2009) and reconcile contradicting genetic findings observed in uniparental and biparental genome data. The conclusions of the latest genome-wide studies (Atzmon et al. 2010; Behar et al. 2010) that European Jews had a single Middle Eastern origin are incomplete as neither study tested the Khazarian Hypothesis, to the extent done here. Finally, our findings confirm both oral narratives and the canonical Jewish literature describing the Khazar's conversion to Judaism (e.g., "Sefer ha-Kabbalah" by Abraham ben Daud [1161 CE], and "The Khazars" by Rabbi Jehudah Halevi [1140 CE]) (Polak 1951; Koestler 1976).

Although medical studies were not conducted using Caucasus and Near Eastern populations to the same extent as with European Jews, many diseases found in European Jews are also found in their ancestral groups in the Caucasus (e.g., Cystic fibrosis and α-thalassaemia), the Near East (e.g., Factor XI deficiency, type II), and Southern Europe (e.g., Non-syndromic recessive deafness) (Ostrer 2001), attesting to their complex multi-origins.

Because our study is the first to directly contrast the Rhineland and Khazarian Hypotheses, a caution is warranted in interpreting some of our results due to small sample sizes and availability of surrogate populations. To test the Khazarian Hypothesis, we used a crude model for the Khazar's population structure. Our admixture analysis suggests that certain ancestral elements in the Caucasus genetic pool may have been unique to the Khazars. Therefore, using few contemporary Caucasus populations as surrogates may capture only certain shades of the Khazarian genetic spectrum. Further studies are necessary to test the magnitude of the Judeo-Khazar demographic contribution to the presence of Jews in Europe (Polak 1951; Dinur 1961; Koestler 1976; Baron 1993; Brook 2006). These studies may yield a more complex demographic

model than the one tested here and illuminate the complex population structure of Caucasus populations. Irrespective of these limitations, our results were robust across diverse types of analyses, and we hope that they will provide new perspectives for genetic, disease, medical, and anthropological studies.

Conclusions

We compared two genetic models for European Jewish ancestry depicting a mixed Khazarian-European-Middle Eastern and sole Middle Eastern origins. Contemporary populations were used as surrogate to the ancient Khazars and Judeans, and their relatedness to European Jews was compared over a comprehensive set of genetic analyses. Our findings support the Khazarian Hypothesis depicting a large Caucasus ancestry along with Southern European, Middle Eastern, and Eastern European ancestries, in agreement with recent studies and oral and written traditions. We conclude that the genome of European Jews is a tapestry of ancient populations including Judaized Khazars, Greco-Romans Jews, Mesopotamian Jews, and Judeans and that their population structure was formed in the Caucasus and the banks of the Volga with roots stretching to Canaan and the banks of the Jordan.

Supplemental Materials

Supplementary Note 1
Supplementary Note 2
Supplementary figures
Table S1 – Mt Haplogroups
Table S2 – Y Haplogroups
Table S3 – Studied Populations
Table S4 – ASD analysis for Non-Jews
Table S5 – ASD analysis for Jews
Table S6 – Population Summary of ASD analysis
Table S7 – ASD results

Acknowledgment

I am grateful to Brian and Sharon Browning for their help with the IBD analysis, to my colleagues for their valuable comments. I thank two anonymous reviewers for their constructive comments and suggestions.

Reference

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655-1664.

Atzmon G, Hao L, Pe'er I, et al. [11 co-authors]. 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.

Balanovsky O, Dibirova K, Dybo A, et al. [24 co-authors]. 2011. Parallel Evolution of Genes and Languages in the Caucasus Region. Mol. Biol. Evol. 28:2905-2920.

Baron SW. 1993. A Social and Religiolls History of the Jews. New York: Columbia University Press.

Behar DM, Thomas MG, Skorecki K, et al. [12 co-authors]. 2003. Multiple origins of Ashkenazi Levites: Y chromosome evidence for both Near Eastern and European ancestries. Am. J. Hum. Genet. 73:768-779.

Behar DM, Yunusbayev B, Metspalu M, et al. [21 co-authors]. 2010. The genome-wide structure of the Jewish people. Nature. 466:238-242.

Bonné-Tamir B, Adam A. 1992. Genetic diversity among Jews : diseases and markers at the DNA level. New York: Oxford University Press.

Brook KA. 2006. The Jews of Khazaria. Lanham, MD: Rowman & Littlefield.

Browning BL, Browning SR. 2011. A fast, powerful method for detecting identity by descent. Am. J. Hum. Genet. 88:173-182.

Carmeli DB. 2004. Prevalence of Jews as subjects in genetic research: Figures, explanation, and potential implications. Am. J. Med. Genet. A. 130A:76-83.

Chakravarti A. 2009. Being human: kinship: race relations. Nature. 457:380-381.

Conrad DF, Jakobsson M, Coop G, Wen X, Wall JD, Rosenberg NA, Pritchard JK. 2006. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. Nat. Genet. 38:1251-1260.

Dinur BZ. 1961. Israel in Exile (in Hebrew). Tel Aviv, Israel: Devir and Mosad Bialik.

Dvornik F. 1962. The Slavs in European history and civilization. New Brunswick, NJ: Rutgers University Press.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol. Bioinform. Online. 1:47-50.

Finkelstein I, Silberman NA. 2002. The Bible unearthed : archaeology's new vision of ancient Israel and the origin of its sacred texts. New York, NY: Simon and Schuster.

Gilbert M. 1993. The atlas of Jewish history. New York: William Morrow and Company.

Hammer MF, Redd AJ, Wood ET, et al. [12 co-authors]. 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.

Hitti PK. 1928. The origins of the Druze people and religion, with extracts from their sacred writings. New York: Columbia University Press.

Koestler A. 1976. The thirteenth tribe : the Khazar empire and its heritage. New York: Random House.

Kopelman NM, Stone L, Wang C, Gefel D, Feldman MW, Hillel J, Rosenberg NA. 2009. Genomic microsatellites identify shared Jewish ancestry intermediate between Middle Eastern and European populations. BMC Genet. 10:80.

Levy-Coffman E. 2005. A mosaic of people: the Jewish story and a reassessment of the DNA evidence. J. Genetic Genealogy. 1:12-33.
Li JZ, Absher DM, Tang H, et al. [11 co-authors]. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science. 319:1100-1104.
McVean G. 2009. A genealogical interpretation of principal components analysis. PLoS Genet. 5:e1000686.
Nebel A, Filon D, Weiss DA, Weale M, Faerman M, Oppenheim A, Thomas MG. 2000. High-resolution Y chromosome haplotypes of Israeli and Palestinian Arabs reveal geographic substructure and substantial overlap with haplotypes of Jews. Hum. Genet. 107:630-641.
Novembre J, Johnson T, Bryc K, et al. [12 co-authors]. 2008. Genes mirror geography within Europe. Nature. 456:98-101.
Ostrer H. 2001. A genetic profile of contemporary Jewish populations. Nat. Rev. Genet. 2:891-898.
Ostrer H. 2012. Legacy: A Genetic History of the Jewish People. OUP USA.
Pagani L, Ayub Q, Macarthur DG, et al. [13 co-authors]. 2011. High altitude adaptation in Daghestani populations from the Caucasus. Hum. Genet.:1-11.
Patai R, Patai J. 1975. The myth of the Jewish race. New York: Scribner.
Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. Pp. e190. PLoS Genet.
Payaslian S. 2007. The history of Armenia : from the origins to the present. New York: Palgrave Macmillan.
Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. arXiv. 1206.2332v1.
Polak AN. 1951. Khazaria - The History of a Jewish Kingdom in Europe (in Hebrew). Tel-Aviv, Israel: Mosad Bialik and Massada Publishing Company.
Price AL, Butler J, Patterson N, et al. [20 co-authors]. 2008. Discerning the ancestry of European Americans in genetic association studies. PLoS Genet. 4:e236.
Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38:904-909.
Purcell S, Neale B, Todd-Brown K, et al. [11 co-authors]. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559-575.
Reich D, Thangaraj K, Patterson N, Price AL, Singh L. 2009. Reconstructing Indian population history. Nature. 461:489-494.
Sand S. 2009. The invention of the Jewish people. London: Verso.
Schonberg A, Theunert C, Li M, Stoneking M, Nasidze I. 2011. High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. Eur. J. Hum. Genet. 19:988-994.
Shapira DDY. 2007. Armenian and Georgian sources on the Khazars: a re-evaluation *In:* Golden, PB, Ben-Shammai, H., and Róna-Tas, A., editor. The World of the Khazars: New Perspectives - Selected Papers from the Jerusalem 1999 International Khazar Colloquium. Leiden, Boston: Brill. p. 291-306.

Shlush LI, Behar DM, Yudkovsky G, Templeton A, Hadid Y, Basis F, Hammer M, Itzkovitz S, Skorecki K. 2008. The Druze: a population genetic refugium of the Near East. PLoS One. 3:e2105.
Straten JV. 2007. Early Modern Polish Jewry The Rhineland Hypothesis Revisited. Hist. Methods. 40:39-50.
Straten JV. 2003. Jewish Migrations from Germany to Poland: the Rhineland Hypothesis Revisited. Mankind Quarterly. 44:367-384.
Tian C, Plenge RM, Ransom M, et al. [11 co-authors]. 2008. Analysis and application of European genetic substructure using 300 K SNP information. PLoS Genet. 4:e4.
United Jewish Communities. 2003. National Jewish Population Survey, 2000-01. New York: United Jewish Communities.
Weiss KM, Long JC. 2009. Non-Darwinian estimation: my ancestors, my genes' ancestors. Genome Res. 19:703-710.
Wexler P. 1993. The Ashkenazic Jews: A Slavo-Turkic People in Search of a Jewish Identity. Colombus, OH: Slavica.
Zoossmann-Diskin A. 2010. The origin of Eastern European Jews revealed by autosomal, sex chromosomal and mtDNA polymorphisms. Biol Direct. 5:57.

Tables

Table 1. **Genetic distances (ASD) between regional and continental Jewish communities (left panel) and between regional Jewish communities and their non-Jewish neighboring populations, Caucasus, and Middle Eastern populations (right panel).** Bold entries are significantly smaller throughout each panel. The geographically nearest non-Jewish population were considered neighboring populations. The distances in the last two columns are between a Jewish and one Caucasus (Armenians of Georgians) or Middle Eastern (Palestinians, Bedouins, or Jordanians) population that exhibited the lowest mean ASD.

Regional Jewish community	Jewish populations				Non-Jewish populations			
	Self	European	Asian	African	Neighboring population [Δ]		Caucasus[&]	Middle Eastern[§]
Eastern European	**0.2318**	0.2328	0.2381	0.2446	Hungarian	0.2346	**0.2340**	0.2387
Central European	**0.2312**	0.2326	0.2378	0.2445	Italians	**0.2335**	0.2338	0.2385
Bulgarian	0.2326	0.2331	0.2376	0.2439	Romanian	0.2347	**0.2337**	0.2380
Turkish	0.2336	0.2336	0.2376	0.2439	Turkish	0.2353	**0.2337**	0.2379
Iraqi	**0.2303**	0.2351	0.2375	0.2447	Iranian	0.2363	**0.2338**	0.2381
Georgian	0.2304	0.2345	0.2372	0.2442	Georgian	0.2332	**0.2332**	0.2378
Azerbaijani	**0.2304**	0.2365	0.2386	0.2465	Lezgins	0.2367	**0.2352**	0.2398
Iranian	**0.2310**	0.2364	0.2391	0.2434	Iranian	0.2414	**0.2361**	0.2383

Figure legends

Figure 1

Map of Eurasia. A map of Khazaria and Judah is shown with the state of origin of the studied groups. Eurasian Jewish and non-Jewish populations used in all analyses are shown in square and round bullets, respectively (see Table S3). The major migrations that formed Eastern European Jewry according to the Khazarian and Rhineland Hypotheses are shown in yellow and brown, respectively.

Figure 2

An illustrated timeline for the relevant historical events. The horizontal dashed lines represent controversial historical events explained by the different hypotheses, whereas solid black lines represent undisputed historical events.

Figure 3

Scatter plot of all populations along the first two principal components. For brevity, we show only the populations relevant to this study. The inset magnifies Eurasian and Middle Eastern individuals. Each letter code corresponds to one individual (Table S3). A polygon surrounding all of the individual samples belonging to a group designation highlights several population groups.

Figure 4

Biogeographical origin of European Jews. First two principal components were calculated for Pygmies, French Basque, Han Chinese (black), Armenians (blue), and Eastern or Central European Jews (red) - all of equal size. PCA was calculated separately for Eastern and Central

European Jews and the results were merged. Using the first four populations as training set, Eastern (squares) and Central (circles) European Jews were assigned to geographical locations by fitting independent linear models for latitude and longitude as predicted by PC1 and PC2. Each shape represents an individual. Major cities are marked in cyan.

Figure 5

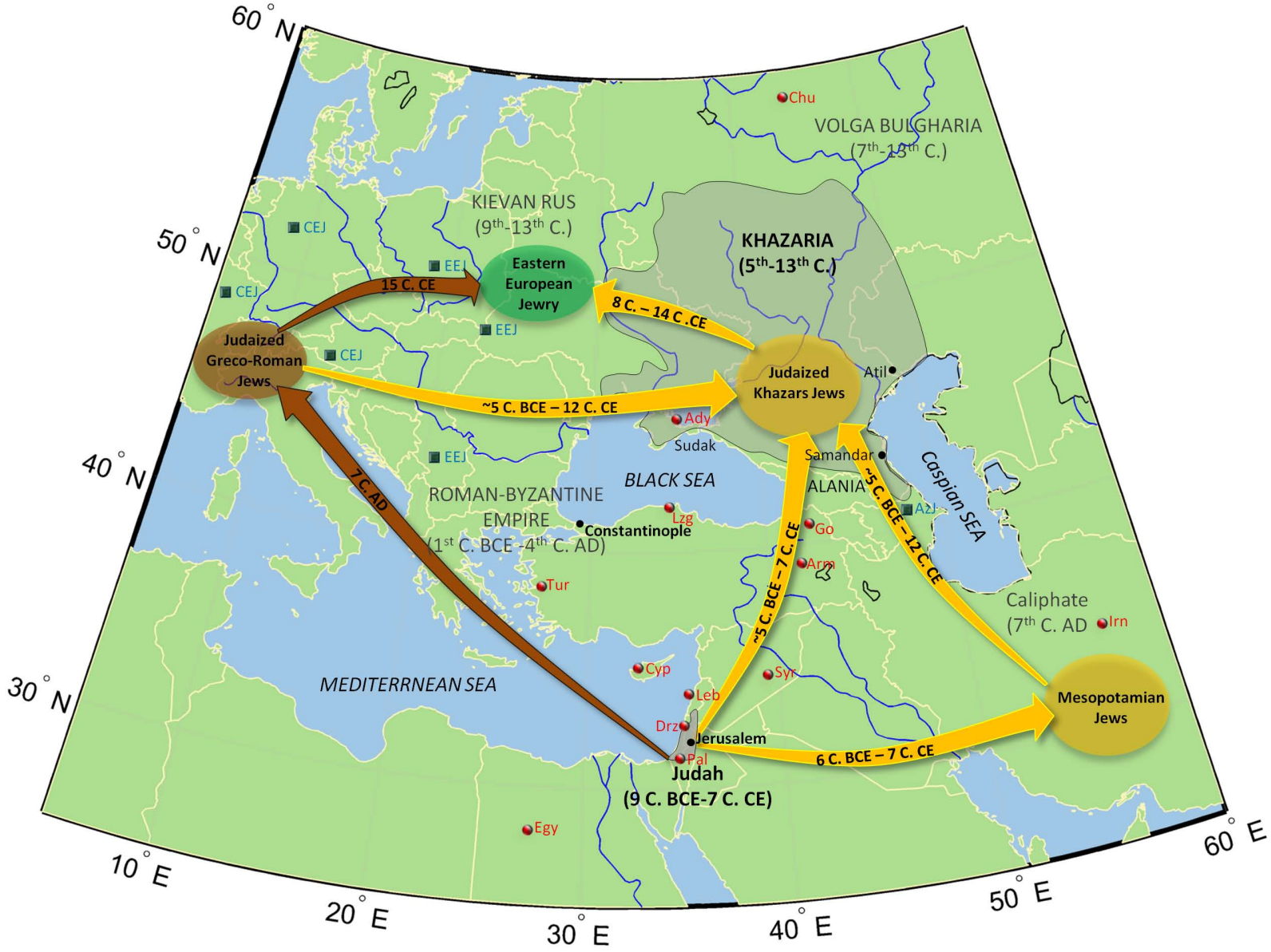
Admixture analysis of European, Caucasus, Near Eastern, and Middle Eastern populations. The *x*-axis represents individuals from populations sorted according to their ancestries and arrayed geographically roughly from North to South. Each individual is represented by a vertical stacked column (100%) of color-coded admixture proportions of the ancestral populations.

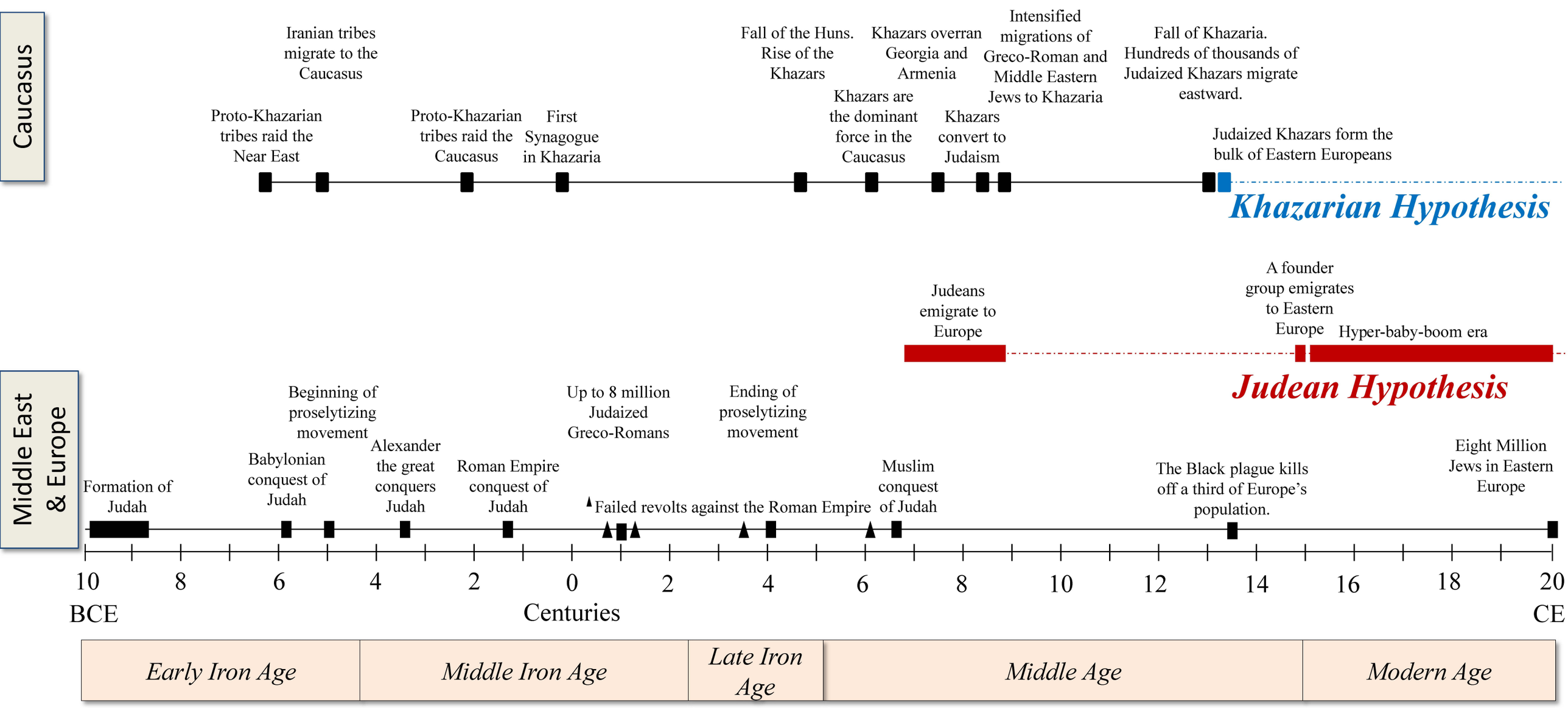
Figure 6

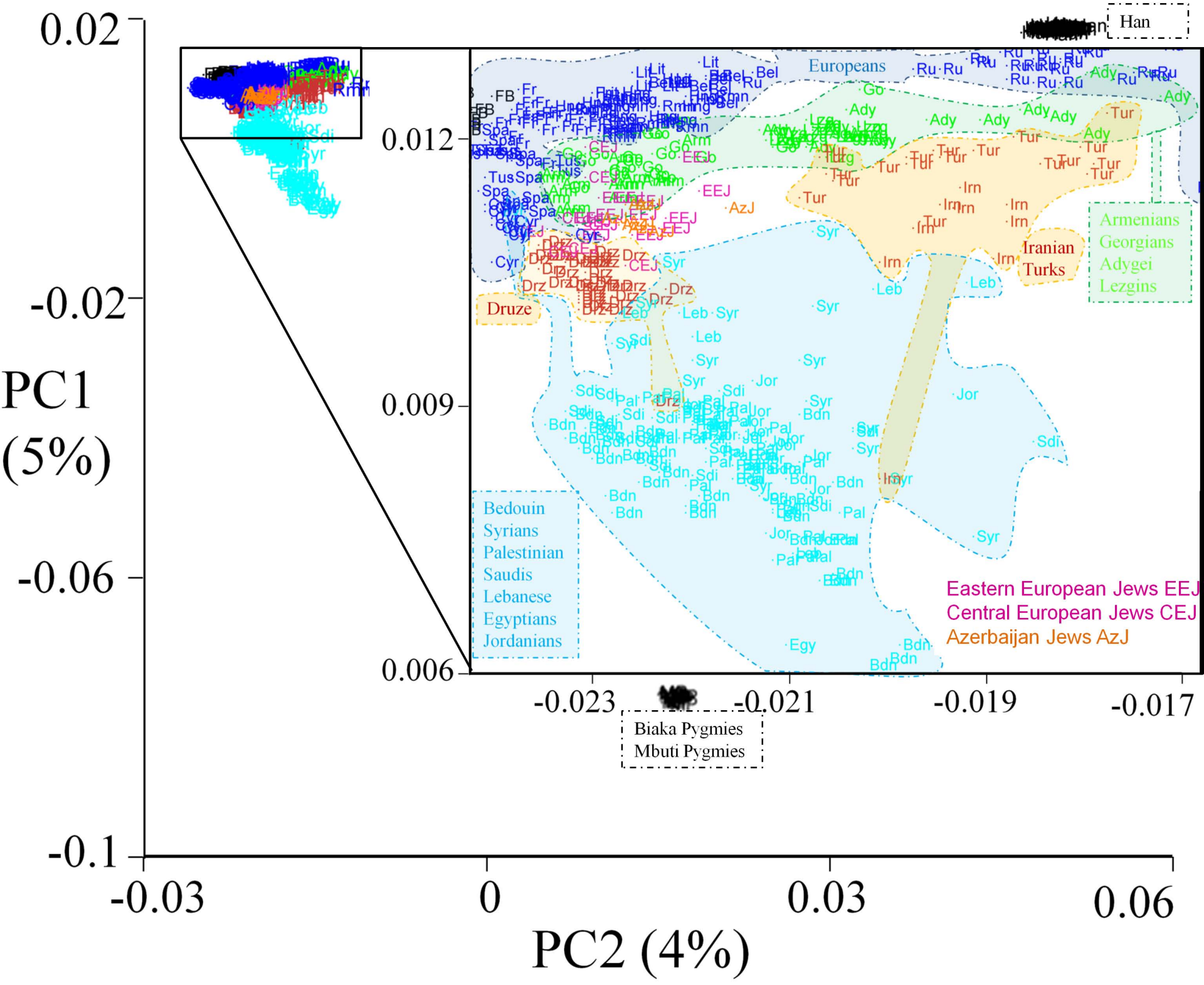
Proportion of total IBD sharing between European Jews and different populations. Populations are sorted by decreasing distance from the Caucasus. The maximal IBD between each European Jew and an individual from each population are summarized in box plots. Lines pass through the mean values.

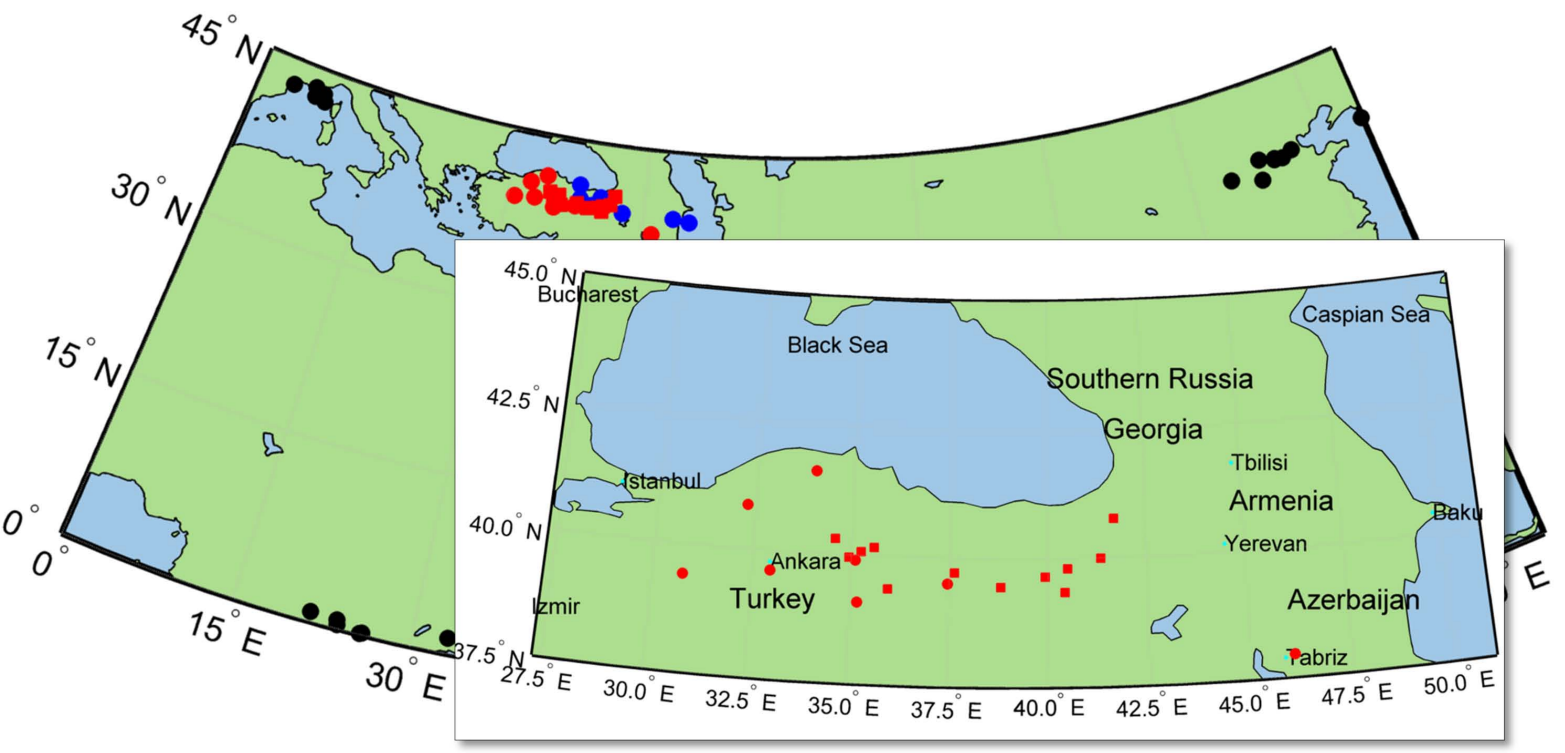
Figure 7

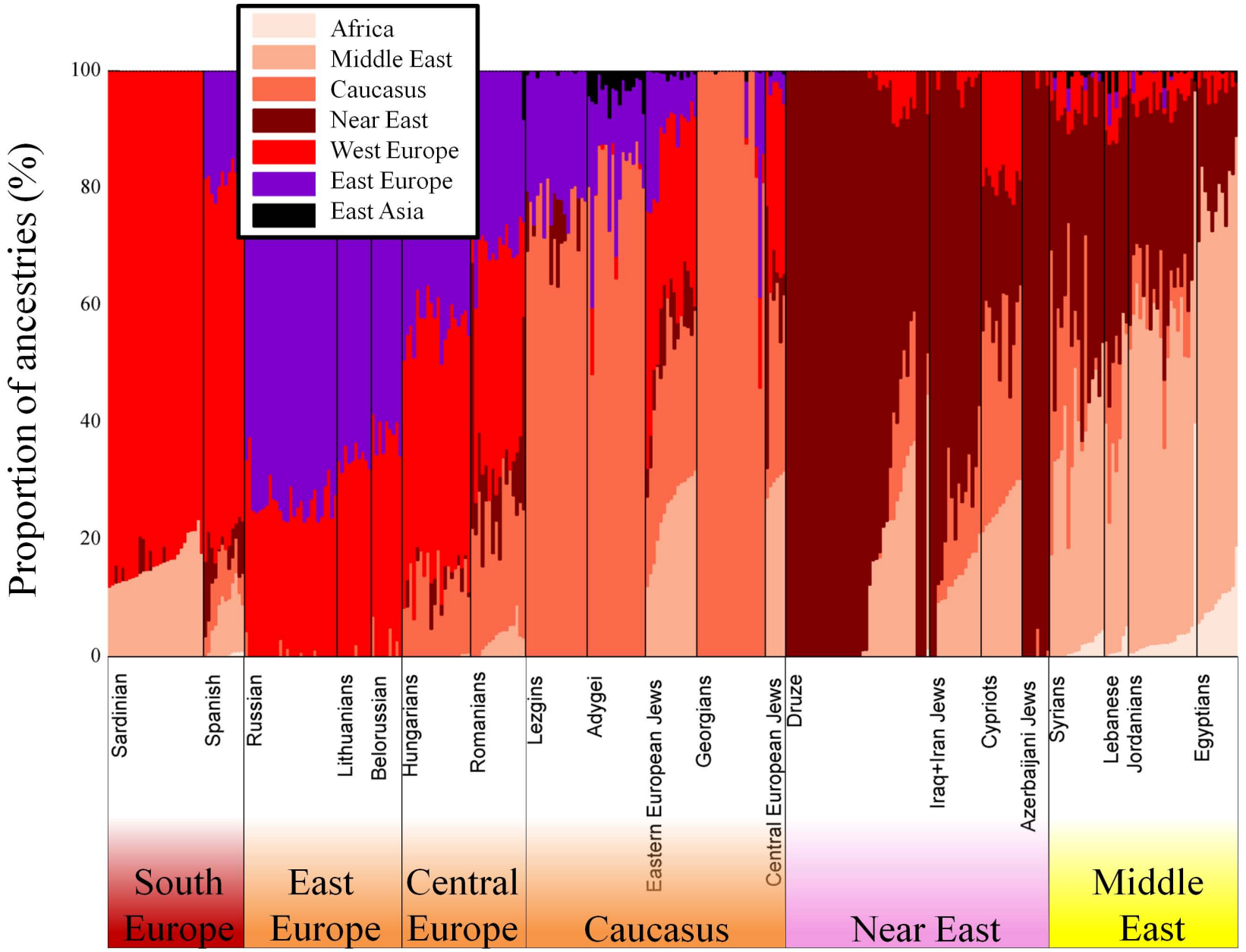
Pairwise genetic distances between European Jews and other populations measured across a) mtDNA and b) Y-chromosomal haplogroup frequencies. The values of $1-\delta_{xy}$ are color coded in a heat map with darker colors indicating higher haplogroup similarity with European Jews.

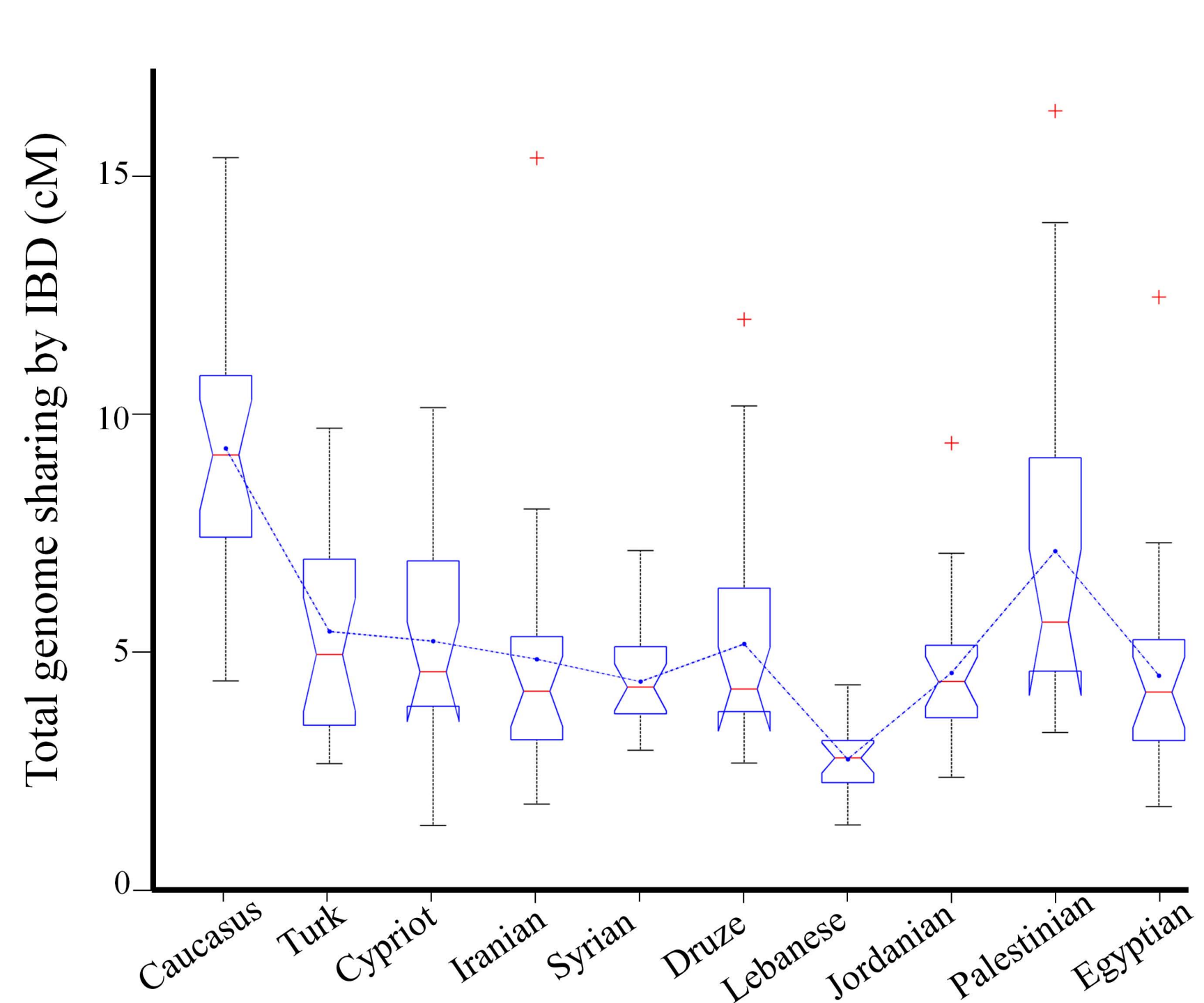

60° N
50° N
40° N
30° N
10° E
20° E
30° E
40° E
50° E
60° E
Chu
VOLGA BULGHARIA
(7th-13th C.)
KIEVAN RUS
(9th-13th C.)
KHAZARIA
(5th-13th C.)
CEJ
EEJ
Eastern European Jewry
15 C. CE
8 C. – 14 C .CE
Judaized Greco-Roman Jews
Judaized Khazars Jews
Atil
~5 C. BCE – 12 C. CE
Ady
Sudak
Samandar
Caspian SEA
BLACK SEA
ALANIA
AZJ
ROMAN-BYZANTINE EMPIRE
(1st C. BCE -4th C. AD)
Constantinople
Lzg
Go
Arm
7 C. AD
~5 C. BCE – 7 C. CE
~5 C. BCE – 12 C. CE
Tur
Caliphate
(7th C. AD
Irn
Cyp
Syr
MEDITERRNEAN SEA
Leb
Mesopotamian Jews
Drz
Jerusalem
Pal
6 C. BCE – 7 C. CE
Judah
(9 C. BCE-7 C. CE)
Egy

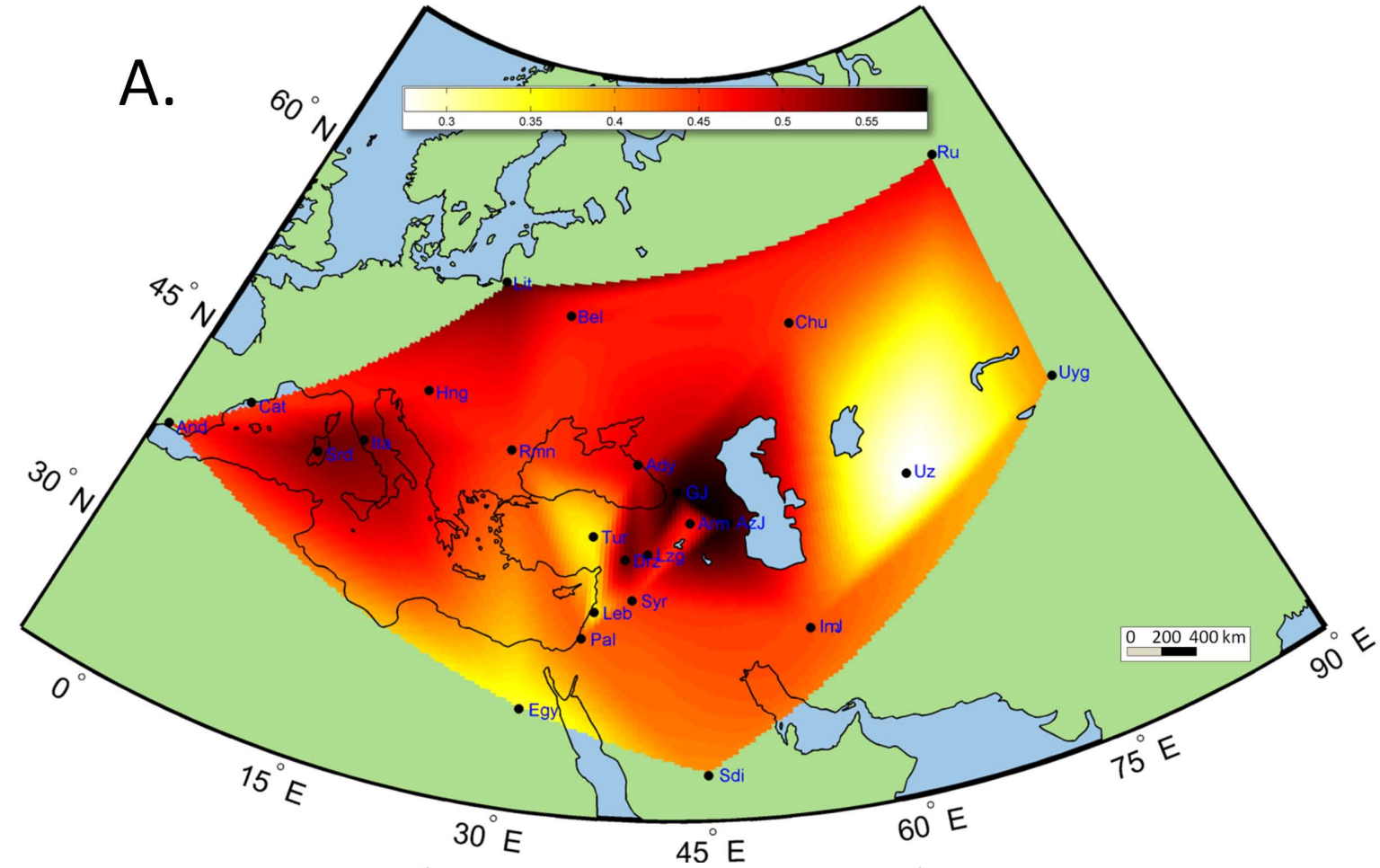

Caucasus
Iranian tribes migrate to the Caucasus
Proto-Khazarian tribes raid the Near East
Proto-Khazarian tribes raid the Caucasus
First Synagogue in Khazaria
Fall of the Huns. Rise of the Khazars
Khazars are the dominant force in the Caucasus
Khazars overran Georgia and Armenia
Khazars convert to Judaism
Intensified migrations of Greco-Roman and Middle Eastern Jews to Khazaria
Fall of Khazaria. Hundreds of thousands of Judaized Khazars migrate eastward.
Judaized Khazars form the bulk of Eastern Europeans
Khazarian Hypothesis
Judeans emigrate to Europe
A founder group emigrates to Eastern Europe
Hyper-baby-boom era
Judean Hypothesis
Middle East & Europe
Formation of Judah
Babylonian conquest of Judah
Beginning of proselytizing movement
Alexander the great conquers Judah
Roman Empire conquest of Judah
Up to 8 million Judaized Greco-Romans
Failed revolts against the Roman Empire
Ending of proselytizing movement
Muslim conquest of Judah
The Black plague kills off a third of Europe's population.
Eight Million Jews in Eastern Europe
10 8 6 4 2 0 2 4 6 8 10 12 14 16 18 20
BCE
Centuries
CE
Early Iron Age
Middle Iron Age
Late Iron Age
Middle Age
Modern Age

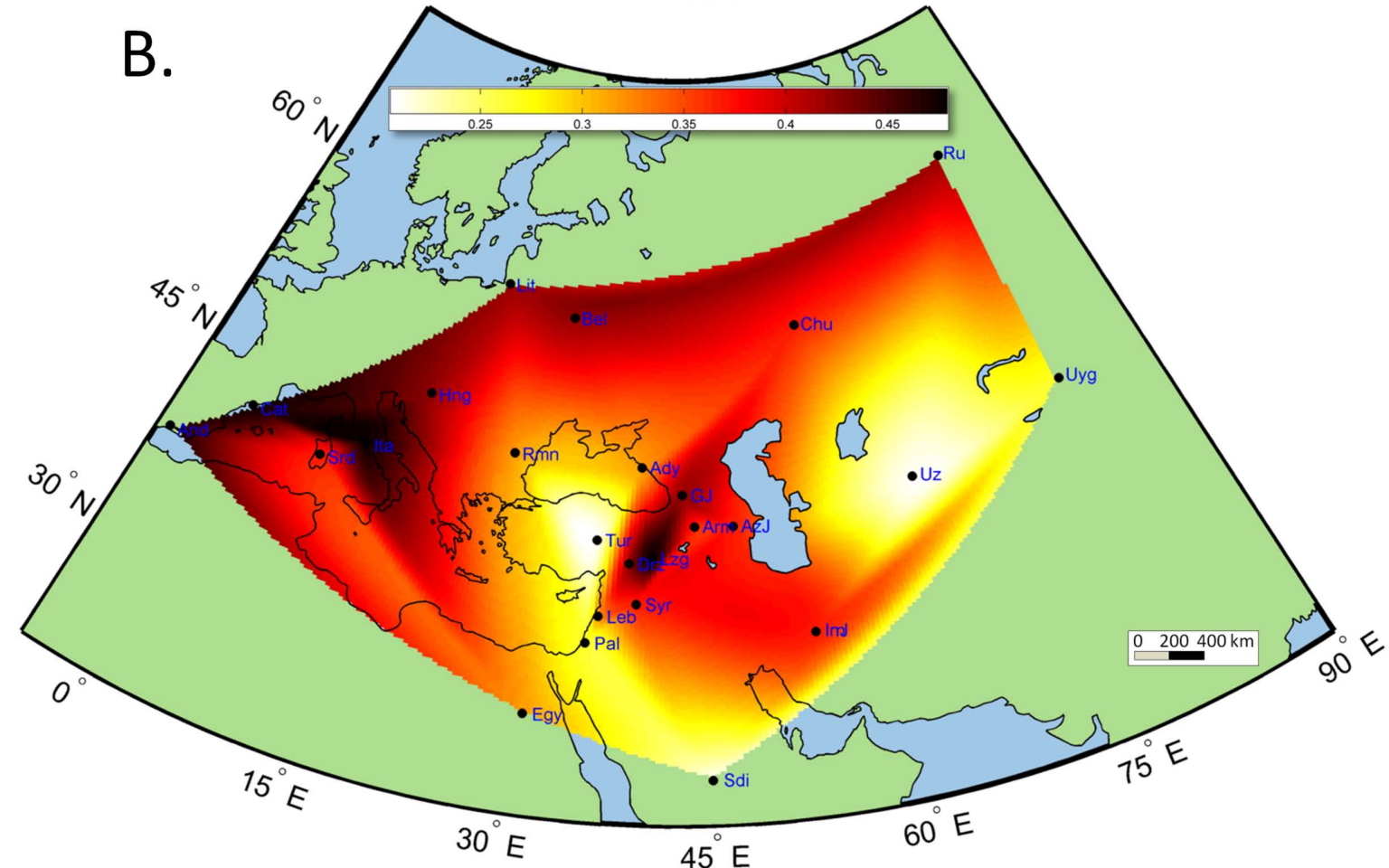

Han
Europeans
Armenians
Georgians
Adygei
Lezgins
Iranian
Turks
Druze
Bedouin
Syrians
Palestinian
Saudis
Lebanese
Egyptians
Jordanians
Eastern European Jews EEJ
Central European Jews CEJ
Azerbaijan Jews AzJ
Biaka Pygmies
Mbuti Pygmies
0.02
-0.02
-0.06
-0.1
PC1
(5%)
0.012
0.009
0.006
-0.023
-0.021
-0.019
-0.017
-0.03
0
0.03
0.06
PC2 (4%)


45° N
30° N
15° N
0°
0°
15° E
30° E
45.0° N
42.5° N
40.0° N
37.5° N
27.5° E
30.0° E
32.5° E
35.0° E
37.5° E
40.0° E
42.5° E
45.0° E
47.5° E
50.0° E
Bucharest
Black Sea
Caspian Sea
Southern Russia
Georgia
Tbilisi
Istanbul
Armenia
Baku
Yerevan
Ankara
Izmir
Turkey
Azerbaijan
Tabriz


Africa
Middle East
Caucasus
Near East
West Europe
East Europe
East Asia
Proportion of ancestries (%)
100
80
60
40
20
0
Sardinian
Spanish
Russian
Lithuanians
Belorussian
Hungarians
Romanians
Lezgins
Adygei
Eastern European Jews
Georgians
Central European Jews
Druze
Iraq+Iran Jews
Cypriots
Azerbaijani Jews
Syrians
Lebanese
Jordanians
Egyptians
South Europe
East Europe
Central Europe
Caucasus
Near East
Middle East


Total genome sharing by IBD (cM)
0
5
10
15
Caucasus
Turk
Cypriot
Iranian
Syrian
Druze
Lebanese
Jordanian
Palestinian
Egyptian


A.
0.3
0.35
0.4
0.45
0.5
0.55
60 ° N
45 ° N
30 ° N
0 °
15 ° E
30 ° E
45 ° E
60 ° E
75 ° E
90 ° E
Ru
Lit
Bel
Chu
Uyg
Hng
Cat
Apd
Srd
Ita
Rmn
Ady
GJ
Uz
Arm
AZJ
Tur
Lzg
Drz
Syr
Leb
Pal
Imj
Egy
Sdi
0 200 400 km


B.
0.25
0.3
0.35
0.4
0.45
60 ° N
45 ° N
30 ° N
0 °
15 ° E
30 ° E
45 ° E
60 ° E
75 ° E
90 ° E
Ru
Lit
Bel
Chu
Uyg
Hng
Cat
Apd
Srd
Ita
Rmn
Ady
GJ
Uz
Arm
AZJ
Tur
Lzg
Drz
Syr
Leb
Pal
Imj
Egy
Sdi
0 200 400 km