Localizing the admixture signatures of Yiddish speakers to primeval villages in ancient Ashkenaz lands

Ranajit Das[1], Paul Wexler[2], Mehdi Pirooznia[3], and Eran Elhaik[1*]

[1] University of Sheffield, Department of Animal and Plant Sciences, Sheffield, UK S10 2TN

[2] Tel Aviv University, Department of Linguistics, Tel-Aviv, Israel 69978

[3] Johns Hopkins University, Department of Psychiatry and Behavioral Sciences, Baltimore, MD, USA 21205

* Please address all correspondence to Eran Elhaik at e.elhaik@sheffield.ac.uk

Keywords: Yiddish, GPS, villages, population structure

Abstract

Yiddish is a language over one thousand years old which incorporates a Slavic grammar and phonology, a largely covert Slavic lexicon, and Hebrew and German elements primarily in the lexicon. The popular view claims a German origin for Yiddish, whereas the opposing view posits a Slavic origin with strong Iranian and weak Turkic substrata. One of the major difficulties in deciding between these hypotheses is the unknown geographical origin of Yiddish speaking Ashkenazic Jews. To address this question, we adopted genetic admixture- and haplogroup-based approaches that identified the geographical origin of sole Yiddish speakers and other subgroups of Ashkenazic Jews. Nearly all the Yiddish speakers were localized along major primeval trade routes in northeastern Turkey adjacent to primeval villages, some of whose names may be derived from "Ashkenaz." Our results imply that Yiddish was created by Irano-Turko-Slavic Jewish merchants along the Silk Roads as a cryptic trade language and over time acquired a significant genuine and invented Hebrew component. Ashkenazic Jews probably originated from a Slavo-Iranian confederation, which the Jews call "Ashkenazic" (i.e. "Scythian"), though they were probably speakers of Persian and/or Ossete. This is compatible with linguistic evidence suggesting that Yiddish is a Slavic language that, beginning in the 9th century, underwent relexification to a minority of German and Hebrew and a majority of newly coined Germanoid and Hebroid elements.

Background

Paramount geographical movements, due to voluntary migration or forced resettlement, are often reflected in a language's lexicon as a new stratum of words and phrases that may replace or modify archaic terms. In an analogy to species' struggle to survive, Darwin remarked that "a struggle for life is constantly going on among the words and grammatical forms in each language" (1871). This parallelism between the history of a language and the history of its speakers and the expectation that such insights will highlight the geographical origins of populations has attracted much attention from geneticists and linguists (Cavalli-Sforza 1997; Kitchen et al. 2009; Balanovsky et al. 2011; Bouckaert et al. 2012). Major deviations from this parallelism are explicable by admixture or migration followed by extreme isolation (Ramachandran et al. 2005). In such cases, the language's lexicon may represent various stratum of words from different languages the migrating people have encountered, deeming phylogenetic-based approaches inapplicable. For that reason, it has been proposed to look at linguistic and genetic data in parallel and attempt integrative analyses (Brandt et al. 2014).

One of the last European languages whose linguistic and geographical classifications remain unclear even after three centuries of research is Yiddish (Weinreich 2008), the native language of the Ashkenazic Jewish community, whose own origin remains controversial (Costa et al. 2013; Elhaik 2013). The Yiddish language, spoken since the 9th century, is an amalgam of Hebrew, German, Slavic and other elements written in Aramaic characters (Weinreich 2008). Because of its many radical deviations from native German norms, its alleged cognate language, Yiddish has been rudely labeled both by native and non-native speakers as "bad German" and in Slavic languages as a "jargon" (Weinreich 2008). Part of the problem in deciphering its origin is that over the centuries Yiddish speakers have invented a huge number of "Germanoid" (German-like) and "Hebroid" (Hebrew-like) components based on Slavic or Iranian models. For example, the written modern-day Hebrew phrase *paxot o joter* (literally "<u>less</u> or <u>more</u>") imitates the same written Ashkenazic Hebroid phrase, derived from Upper Sorbian and Iranian languages, but not Old Semitic Hebrew. The overwhelming majority of the world's languages use "more or less." This expression appeared during the Middle Ages, long after the death of spoken Hebrew and possibly a millennium before the appearance of modern-day "Hebroid." These and other features made the origin of Yiddish word strata and its relationship to other languages multilayered, porous, ephemeral, and difficult to localize.

Two major linguistic hypotheses were invoked to elucidate the origins of Yiddish and that of Yiddish speakers, the kernel group of Ashkenazic Jews in Eastern, Central, and Western Europe (Table 1). The "Rhineland hypothesis" envisions modern Yiddish speaking Ashkenazic Jews to be the descendants of the ancient Judaeans. The presence of Jews in Western and, later, Eastern Europe is explained, in an oversimplified manner, by two allegedly mass migratory waves: The "Roman Exile" that followed the destruction of Herod's temple (70 A.D.) introduced a massive Jewish population to Roman lands (King 2001). Yiddish is assumed to have developed when French and Italian Jews migrated to the Rhineland (and Franconia) in the 9th century and replaced their Romance speech with local German dialects (Weinreich 2008). It was these Jews who allegedly created the so-called Ashkenazic culture, named after the Medieval Hebrew term for the German lands. The second migration waved took place in the 13th century, when German Jews allegedly migrated into monolingual Slavic lands.

The competing "Irano-Turko-Slavic" hypothesis considers Ashkenazic Jews to be the descendants of a heterogeneous Iranian population, which later mixed with Eastern and Western Slavs and possibly some Turks in the territory of the Khazar Empire around the 8th century A.D. The name "Ashkenaz" is the Biblical Hebrew adaptation of the Iranian tribal name which is rendered in Assyrian and Babylonian documents of the 7th century B.C. as *aškūza*, called in English by the Greek equivalent "Scythian" (Wexler 2010). Already by the 1st century, most of the Jews in the world resided in the Iranian Empire (Baron 1952). These Jews were descended either from Judaean emigrants or, more likely, from local converts to Judaism. As evident from the Talmud and non-Jewish historical sources (Gil 1974), Iranian Jews were extremely active in international trade. Over time, many of them moved north to the Khazar Empire to expand their mercantile operations.

Consequently, some of the Turkic Khazar rulers and the numerous Eastern Slavs in the Khazar Empire have converted to Judaism in order to participate in the lucrative Silk Road trade between Germany and China, which was essentially a Jewish monopoly. Yiddish emerged at that time as a secret language for trade based on Slavic and even Iranian patterns of discourse. When these Jews began settling in Western and Eastern Slavic lands, Yiddish went through a relexification process, eventually replacing the Eastern Slavic and the newly acquired Sorbian vocabularies with a German vocabulary (Wexler 2011a).

The work of Cavalli-Sforza and other investigators have already established the strong relationship between geography, genetics, and languages (Luigi Luca Cavalli-Sforza 1994; Eller 1999; Balanovsky et al. 2011; Everett 2013), implying that the geographical origin of Yiddish should correspond to that of Yiddish speakers. However, the admixed nature of both Yiddish (Table 1) and the Ashkenazic Jewish genome (Elhaik 2013) preclude using traditional phylogenetic methods to localize their geographical origins. Moreover, the high genetic heterogeneity of Ashkenazic Jews (Bray et al. 2010; Elhaik 2013) may obscure the geographical origin of Yiddish speakers. We have, therefore, adopted genetic admixture-based approaches to find the geographical origins of sole Yiddish speakers and other Ashkenazic Jewish subgroups and test whether they are compatible with either hypothesis or neither one. Assuming the history of Yiddish speakers and Yiddish is parallel (Weinreich 2008), at least in part, localizing the genomic admixture signature of Yiddish speakers may also unveil the birthplace of Yiddish. Due to the changes in the population structure of Yiddish speakers over the past millennia, we do not expect our biogeographical predictions to perfectly agree with the predictions made by either hypotheses. This is the first study that analyzes genetic data of Yiddish speakers, and it is carried out at a most timely manner as individuals who speak solely Yiddish are increasingly difficult to find.

Results

Our search for the geographical origins of the admixture signature of Yiddish speaker genomes was focused on Eurasia, with particular consideration in the area covering the geographical origins predicted by each hypothesis (Table 1). This area encompasses German lands, South Russia, and the area between ancient Judea and the western regions of the former Iranian (Sassanian) Empire.

Biogeographical mapping of Eurasian population.

All biogeographical inferences were carried out using the geographic population structure (GPS) tool reported to predict worldwide individuals with high sensitivity (0.75) and specificity (0.99) (Elhaik et al. 2014). Briefly, GPS infers the geographical coordinates of an individual by matching its admixture proportions with those of reference populations known to reside in a certain geographical region for a substantial period of time. Whereas a population movement followed by a gene exchange with other populations modifies the admixture signature, isolation and segregation preserve the original admixture signature of the migratory population. GPS predictions should therefore be interpreted as the last place that admixture has occurred, termed here *geographical origins*. For an individual of mixed origins, the inferred coordinates represent the mean geographical locations of one's immediate ancestors.

Prior to applying GPS to elucidate the geographical origins of Yiddish speakers, we sought to estimate its accuracy on Eurasian populations. For that, we analyzed over 600 individuals belonging to 35 populations (Figure 1a). The geographical distributions of their admixture components are shown in Figure 1b. The genetic admixture diversity within each population was assessed by comparing their genetic admixture distances (d) (Figure S1), calculated as the minimal Euclidean distances between individuals and members of a population of interest. The median genetic distances for all populations were very small, ranging from 0 to 5.26 with a mean of 2.1. We then applied GPS using the leave-one-out procedure at the population level. Assignment accuracy was determined for each individual based on whether the predicted geographical coordinates were within 500 kilometers (km) or 250km from the political boundaries of the individual's country or regional locations. GPS correctly assigned 83% and 78% of the individuals within less than 500km and 250km from their countries, respectively (Figure 2, Table S2). Within the area covered by the two linguistic hypotheses and harbored by

554 individuals belonging to 31 populations, the accuracy was 2% higher. As expected, the prediction accuracy within that area was even higher (97% and 94% of the individuals were assigned within less than 500km and 250km of their countries, respectively) for speakers of geographically localized languages (Abkhazians, Armenians, Bulgarians, Danes, Finns, Georgians, Greeks, Romanians, Germans, and Palestinians), which also include some of the putative basal components of Yiddish (Romance, Slavic, and German). These results illustrate the tight relationships between genome, geography, and language and delineate the expected assignment accuracy for Yiddish speakers.

Biogeographical mapping of Yiddish speakers

The Yiddish speaker genomes exhibit a mixture of three major components: Mediterranean ($\bar{X}$=52%), Southwest Asian ($\bar{X}$=24%), and Northern European ($\bar{X}$=16%) (Figure 1). Although the Subsaharan African component is extremely rare among northern Eurasians, a small and consistent portion ($\bar{X}$~2%) was found in Yiddish speakers and the ancient pre-Scythian.

GPS positioned nearly all 186 Yiddish speakers and five descendants of priestly lineages on the southern coast of the Black Sea in northeastern Turkey (Figure 3). There we located four primeval villages that bear names that may derive from "Ashkenaz"—İşkenaz (or Eşkenaz) at (40° 9'N, 40° 26'E) and Eşkenez (or Eşkens) at (40° 4'N, 40° 8'E) in the province of Trabzon (or Trebizond), Aşhanas (today Üzengili) at (40° 5', 40° 4'E) in the province of Bayburt, and Aschuz (or Hassis/Haza, 30 B.C.–A.D. 640) (Bryer and Winfield 1985; Roaf et al. 2015) in the province of Tunceli–all of which are in close proximity to major trade routes. The Turkish toponyms/ethnonyms are very suggestive of a Jewish trading presence, but given the poor state of Turkish toponymic studies, we cannot say for sure. There are no other placenames anywhere in the world derived from this ethnonym. Instead, the many Jewish "way stations" on the trade routes throughout Afro-Eurasia are named after the root "Jew," but these may be places named by non-Jews. Most of the Yiddish speakers in this region were positioned within about 200km from at least one of these villages, which are also located close to ancient Khazaria's southern borders (Brook 2006). These individuals were also proximate to Iranian Jews, positioned within less than 200km from eastern Turkey in Malatya Province, a medieval Jewish center. Three Yiddish speakers were positioned to the eastern shores of the Black Sea within Bulgaria and Romania. No individual was positioned in Germany or proximate to the ancient pre-Scythian.

A comparison of the genetic distances between Yiddish speakers and the reference population (Figure S2) confirmed that Yiddish speakers are significantly (Kolmogorov-Smirnov goodness-of-fit test, p<0.01) closer to Turks ($\tilde{d}$=9%), Armenians ($\tilde{d}$=11.4%), and Romanians ($\tilde{d}$=12.26). The genetic distance to Germans was $\tilde{d}$=26.85%.

The admixture and GPS results for Yiddish and non-Yiddish speakers were very similar. On average, these two cohorts have the same admixture components (Figure S3), and their geographical origins follow similar trends (Figure S4, S5). That all Ashkenazic Jews were predicted away from their parental birth countries (Figure S5) implies arrival by migration and limited gene exchange with Western and Central European populations.

Haplogroup analysis of Yiddish speakers

Yiddish speakers belong to 46 and 69 unique Y chromosomal and mtDNA haplogroups, respectively. The most common (frequency ≥ 5%) low-frequency mtDNA haplogroups explain less of the variation compared to the Y haplogroups. More specifically, the most common mtDNA haplogroups K1a, H1, J1, HV, K2a, and N1 are present in 62% of the individuals compared with 84% of the individuals that belong to the most common Y haplogroups E1b, J1a, J2a, R1a, R1b, Q1b, and J2b. The five most common high-resolution mtDNA (K1a1b1a [16%], K1a9 [6.5%], K2a2a [4.8%], HV1b2 [4.3%], and N1 [4.3%]) and Y (R1a1a2a2 [6.4%], E1b1b1b2a1a [4.8%], J1a1a1a1a1 [4.3%], J1a1a1a1a1a3 [3.7%], and Q1b1a [3.7%]) haplogroups are present in about a third of the samples. Yiddish speakers exhibit a higher haplogroup diversity compared to non-Yiddish speakers (Figures 3, S3) and belong to maternal lineages like H7, I, T2, and V alongside the paternal Q1b – all are rare or absent in non-Yiddish speakers (Table S3). Nearly all the common high-resolution haplogroups appear more frequently in Jews than non-Jews, although none are found uniquely in Ashkenazic Jews or Jews in general and three of them appear less frequently in Ashkenazic Jews compared to other groups (Figure S6).

The most common Y haplogroups dominate the area between the Black and Caspian Seas and represent the major lineages among populations inhabiting Western Asian regions, including Turkey, Iran, Afghanistan, and the Caucasus (Yardumian and Schurr 2011; Cristofaro et al. 2013; Tarkhnishvili et al. 2014). By contrast, the mtDNA haplogroups indicate a more diffused origin and include haplogroups common in Africa (e.g., L2), Central Asian (e.g., I1 and J), Europe (e.g., H) North Eurasia (e.g., T and U), Northwest Eurasia (e.g., V), and Northeast Eurasia (e.g., X and W) (Jobling, Hurles, and Tyler-Smith 2013). High genetic diversity was also observed among descendants of Cohen and Levite lineages in the Y (I2, J1a1a1a1a1, R1a1a2a2) and mtDNA haplogroups (K1a1b1a, N1, HV1b2, K1a, J1c5).

The geographical and ancestral origins of Yiddish speakers

The GPS analysis raises two concerns: first that the Turkish "Ashkenaz" region may be the centric location of other regions rather than the place where the Yiddish speaker admixture signature was formed; second, in the absence of "Ashkenazic" Turks it is impossible to compare the genetic similarity between the two populations to assess common origins implied by the GPS analysis.

To overcome these problems we derived the admixture signatures of "native" populations corresponding to the geographic coordinates of interest from the global distributions of admixture components (Figure 1b) and compared the genetic distances between Yiddish speakers and these populations. Short, as opposed to longer, genetic distances suggest common geographical and genetic origins. Using "native" populations generated directly from the genetic admixture model has more advantages. First, it minimizes the effect of outliers in modern day populations. Second, it circumvents, to a certain degree, the problem of comparing Yiddish speakers with modern day populations that may have experienced various levels of gene exchange or genetic drift past their mixture with

Yiddish speakers. Third, since the genetic model relies on modern day populations, the simulated individuals would be of a similar time period.

We generated the admixture signatures of 100-200 "native" individuals from six areas associated with the origin of Yiddish speakers or Yiddish (Figure 3, Table 1): Germany, Ukraine, Khazaria, Turkish "Ashkenaz," Israel, and Iran (Figure 4a, 4c). We first tested the genetic affinity of these "native" populations with their closest populations by calculating *d*. For Israelites, we used Palestinian and Bedouins, and for Khazars, we used Armenians, Georgians, Abkhazians, Chechens, and Ukrainians (Figure 5b). The average median *d* between the native and modern day populations was 4, slightly higher than the value calculated within modern day populations (Figure S1). We next tested the geographical affinity of these populations by calculating their GPS coordinates (Figure 4d). Most of the individuals were largely mapped to their correct geographical origins, with the exception of the heterogeneous Iranians and the Khazars, likely due to the shared genetic and geographic background of Iranians, Turks, and southern Caucasus populations (Shapira 1999; Brook 2006).

Yiddish speakers predicted in our earlier analysis (Figure 3) largely overlapped with "native" "Ashkenazic" Turks and a few Khazarian Caucasians and Iranians mapped in their vicinity. A comparison of the genetic distances between Yiddish speakers and the "native" populations (Figure 4e), confirmed that Yiddish speakers are significantly (Kolmogorov-Smirnov goodness-of-fit test, p<0.01) closer to each other ($\tilde{d}$=1.2%), followed by "native" Khazarian Caucasians ($\tilde{d}$=5.3%), "Ashkenazic" Turks ($\tilde{d}$=7.7%), Iranian ($\tilde{d}$=12%), Israelites ($\tilde{d}$=13%), Germans ($\tilde{d}$=17.7%), and Ukrainians ($\tilde{d}$=18.9%). Similar results were obtained for non-Yiddish speakers (Figures S7, S8). That $\tilde{d}$ is relatively high between Yiddish speakers and its closest matches compared with other populations (Figure S1) can be explained by the inappropriate proxies for the Khazars and the absence of key founding populations from our model. In a similar manner, we compared the proportion of individuals geographically closest to Yiddish speakers. We found that most of the Yiddish speakers are geographically closer to "native" Khazars (76%), followed by Iranian (13%) and Turk Ashkenaz (11%) as are non-Yiddish speakers. Interestingly, all the descendants of priestly lineages were geographically closer to “native” “Ashkenazic” Turks (Figure 4f).

These results support and expand on our previous findings (Figure 3) suggesting a single-geographical origin for Ashkenazic Jews in an area covering northeastern Turkey, Armenia, Georgia, and the southern regions of the Khazarian Empire along major trace routes. Altogether, our results highlight the genetic similarity between Iranian, Turks, Yiddish speakers, and perhaps Slavs (Figure 4e), as expected by the lexicographical similarity between Yiddish and their languages according to the Irano-Turko-Slavic hypothesis (Table 1).

Discussion

Every language is the creative product of a community and a co-creator of behavior and values, but Yiddish has experienced especially extreme peregrinations as the millennia-old vernacular of Ashkenazic Jews. The questions of Ashkenazic Jews and Yiddish origins have been some of the most debatable questions in genetics, history, and linguists over the past 300 years. While Yiddish is clearly a blend of at least three languages — German, Slavic, and Hebrew — the exact proportion remains unsettled (Table 1) since the cyclical migration and resettlement of Yiddish speakers that influenced their genetic background also restructured the lexical and grammatical strata of the language, thus concealing its geographical origin(s).

Weinreich (2008) emphasized the truism that the history of Yiddish mirrors the history of its speakers. The strong relationship between linguistics, genetics, and geography alongside the availability of well-annotated genomic data from Ashkenazic Yiddish speakers prompted us to adopt admixture-based approaches to investigate the origins of Yiddish speakers and assess the findings under the light of two linguistic hypotheses advocating either a mostly German or an Irano-Turko-Slavic origin. Since languages, like populations and genomes, are not created *ex nihilo* but rather borrow elements from other languages, usually following demographic changes, our approach may also reveal the birthplace of Yiddish. Considering the biogeographical and ancestral origins of Yiddish speakers, the actual birthplace of their parents, and their paternal and maternal haplogroups allowed us to reconstruct the origins of Yiddish speakers and Yiddish.

The similarities between for Yiddish and non-Yiddish speakers raise the question whether they should be studied together. The major differences in haplogroup frequencies persuaded us to focus our investigation on Yiddish speakers and discuss the larger Ashkenazic Jewish cohort, when possible.

Evaluating the evidence for the geographical origin of Yiddish speakers

Intriguingly, our biogeographical analysis positioned nearly all Ashkenazic Jews in the vicinity of the ancient Scythian-inhabited territory, in close proximity to primeval villages that may derive their names from “Ashkenaz” (Figure 3). The Trabzon district, where most of these villages were found, was once inhabited by a Jewish community (Holo 2009) and a center of commercial and coastal trade prior and sporadically through the early 10th century. Most of the Yiddish speakers were localized between Trabzon and Amisus (today Samsun), found ~300km east of Trabzon, where a widespread Jewish settlement existed during the early centuries A.D. Primeval Iraqi Jewish communities that have proliferated by 600 A.D., like Sarari, Nisibis, and Argiza could be found ~300km south to the Bayburt province (Gilbert 1993). Remarkably, our findings echo Harkavy’s, who wrote in 1867 that "the first Jews who came to the southern regions from Russia did not originate in Ashkenaz [Germany], as many writers tend to believe, but from the Greek cities on the shores of the Black Sea [historically, the Trabzon district resided in the Greek Kingdom of Pontus (Bryer and Winfield 1985)], and from Asia via the mountains of the Caucasus.” (Harkavy 1867).

Thus far only few studies attempted to trace the geographical origins of Ashkenazic Jews. Our results are in general agreement with two small scale studies: the first positioned 20 Eastern (38±2.7° N, 39.9±0.4° E) and Central (35±5° N, 39.7 ±1.1° E) European Jews south of the Black Sea (Elhaik 2013), some 100km away from the province of Tunceli.

The second reported an Eastern Turkish origin (41° N, 30° E) to 29 Ashkenazic Jews (Behar et al. 2013), some 628km east to the mean geographical coordinates we obtained (40° 71' N, 37° 47' E). These findings are at odds to the authors' claims of a Middle Eastern origin for Ashkenazic Jews. We emphasize that Palestinians and Bedouins are the only populations localized to Israel (Figure 2).

Evaluating the evidence for the ancestral origins of Yiddish speakers

Although our biogeographical results are well localized, the exact identity of the ancestral groups of Ashkenazic Jews remains unclear. The term "Ashkenaz," which denotes "Scythians" in the Hebrew Bible, is already a tantalizing clue to the large Iranian-origin group that inhabited the central Eurasian steppes, though it cannot be considered evidence of a Scythian origin due to the lack of records about Scythian culture and the obsolescence of Scythian language about five hundred years earlier. It is more likely that Ashkenazic Jews called themselves "Scythians" because this was a popular name in the Bible and in the Caucasus-Ukraine area even long after the disappearance of the Scythians. Ashkenazic Jews may have even considered themselves related to the Scythians based on a shared Irano-Turkish origin, as evident from the proximity of Yiddish speakers to Iranian Jews, however they probably were not Scythians. Irano-Turkish Jews were speakers of Persian, Ossete, or other forms of Iranian, which became extinct during the 10th century. This conclusion is further corroborated by the large geographical distance between the Yiddish speakers and the ancient pre-Scythian individual (Figure 3). Our autosomal analyses confirmed that Yiddish speakers are genetically closest to Turks (Figure S1) and "native" southern Caucasians alongside Iranians, similarly to non-Yiddish speakers (Figure 4e). These results support an Irano-Turkish origin for Ashkenazic Jews.

The inheritance patterns of the mtDNA chromosomes are directly related to the question of Ashkenazic origin. It has been reported that there are four major founding mtDNA lineages that account for ~40% of Ashkenazic mtDNA variation (K1a1b1a [20%], K1a9 [6%], K2a2a1 [5%], and N1b2 (N1b1b) [9%]) (Costa et al. 2013). The authors interpreted these results as a major contribution of Judaized women in the formation of Ashkenazic communities. These haplogroups were among the five most common high-resolution haplogroups in our cohort and accounted for a 37.6% and 39.5% of the mtDNA variation among Yiddish and non-Yiddish speaker, respectively, in general agreement with Costa et al. (2013). Interestingly, Tian and colleagues (2015) provided the first evidence for a significant genetic contribution from Chinese and Central Asians to Eastern European Ashkenazic Jews between around 640 and 1,400 years ago.

A comparison of the haplogroup diversity among Yiddish and non-Yiddish speakers provides additional information about the ancestors of Ashkenazic Jews. The limited haplogroup diversity among non-Yiddish speakers likely indicates the loss of rare haplogroups inherited from Ashkenazic Jews' ancestors, likely through genetic drift since they are uncommon in Europe. These haplogroups reflect a residue of the vast heterogeneity of the founding populations who have considered themselves "Ashkenazic" and established a proselytization center in "Ashkenaz" lands where they have anointed Levites and Cohens to Judaize their slaves and neighboring populations. This is exemplified first by the high haplogroup diversity of priests' lineages, all of which were predicted to be geographically closer to simulated "Ashkenazic" Turk. Second, the Northern Asiatic Q1b1a Y haplogroup, one of the most common haplogroups among Yiddish speakers (3.7%), is completely absent among non-Yiddish speakers. Finally, the mitochondrial haplogroup L2a1 is found in five maternal lineages, where 80% of the mothers speak solely Yiddish (Table S3). A search in the Genographic public dataset found 229 individuals with that haplogroup, 169 of which provided their maternal descent. Of those, nine were Jews, mostly of Ashkenazic descent, four were Europeans, and the remaining of African origin.

Many of our findings have been reported in other studies utilizing genome-wide data but were interpreted as being in support of a Judaean ancestry, although the data do not support such contention. For example, Behar and colleague (Behar et al. 2010) concluded a Middle Eastern origin for Ashkenazic Jews clustered between Turkish and southern Caucasus populations. Even after tracing Ashkenazic Jews to eastern Turkey, Behar and colleagues interpreted the results as evidence to a Middle Eastern and European ancestries (Behar et al. 2013) rather than a Turko-Iranian or Khazarian ancestry. Atzmon and colleague have also interpreted the high similarity between Ashkenazic Jews, Turks and Iranian as an evidence for a Middle Eastern ancestry, citing a miracle to explain the high admixture among Ashkenazic Jews (Atzmon et al. 2010). Some of the common features of such studies include nullification of an alternative hypotheses (e.g., Atzmon et al. 2010; Behar et al. 2013) or using an inadequate alternative hypothesis (e.g., Behar et al. 2010), reliance on inaccurate tools like spatial ancestry analysis (SPA) (Yang et al. 2012; Elhaik et al. 2014) as in the Behar et al.'s (2013) study, and developing models based on fictional events (e.g., Atzmon et al. 2010; Carmi et al. 2014). We emphasize that none of these studies have demonstrated that Ashkenazic Jews are the closest to Israelite or German populations.

The geographical and ancestral origins we found are in agreement with some of the predictions of the Irano-Turko-Slavic hypothesis (Table 1) and imply that the migration of Ashkenazic Jews to Europe was followed by social isolation and avoidance of intermarriages, which retained their northeastern Turkish admixture signature. We cannot rule out the possibility that a limited gene exchange between non-Yiddish speaking Ashkenazic Jews and Europeans took place and reduced the haplogroup diversity in that group, compared to the Yiddish speaking cohort. We can conclude that Socio-religious practices compounded with a unique language can be more effective means of genetic isolation than geographical barriers (Elhaik 2012). These conclusions may not apply to the larger Ashkenazic Jewish community that include mixed couples of non-Ashkenazic or non-Jewish origins. A search in the Genographic dataset showed that this community is twice the size of the Ashkenazic Jewish cohort studied here.

Evaluating the evidence for the Rhineland hypothesis

The Rhineland hypothesis is not supported by our analyses (Figures 3, 4) and suffers from several weaknesses. First, it relies on an unsubstantiated event purported to explain how Judaeans arrived in Eastern Europe from Judea or Roman-Palestine (Sand 2009). Second, it assumes a supernatural event that inflated the population size from fifty thousand (15th century) to five million (20th century) (Atzmon et al. 2010; Ostrer 2012). Ironically, mysticism, superstition, and other supernatural elements have likely been introduced to Ashkenazic Jews by Judaized pagans (Wexler 1993). Third, it ignores the small size of the Jewish population in Middle Ages Germany that was on the order of hundreds or thousands, which makes them unlikely to exact a strong cultural influence on the numerous Irano-Turko-Slavic Ashkenazic Jews (Polak 1951). Crucially, much of the "German" component that buttresses the Rhineland hypothesis is actually "Germanoid" elements that deviate from native German norms and were invented by Yiddish speakers, mainly based on Slavic and, to a lesser extent, on Iranian models (Wexler 2012). It is also unclear why Semitic Hebrew, which had been dead for nearly a millennium, would be revived in the 9th century and for what purpose other than having a cryptic language and Hebrew/Hebroid lexicon for the marketplaces of Afro-Eurasia.

Some of the confusion contributing to the establishment of this hypothesis stems from the erroneous association of the term "Ashkenaz" with "German lands, Germans (Jews and non-Jews)" in the late 11th century, contemporaneous with the rise of Yiddish (Wexler 2011b). Ashkenazic began with the meaning of "Scythian" (though the Jews were probably not of Scythian descent primarily (Figure 3); the term was rather a symbol of their Iranian ethnicity); in the 10th century in Baghdad we find it in the meaning of "Slavic" and by the early 1100s in Europe it assumes the meaning of German/Yiddish, and later the German non-Jews and the German lands. In the 10th century a Moroccan Karaite philologist knew that the Ashkenazic people descended from Khazars and "Germans" - meaning that they came from the Khazar Empire and spoke Yiddish. The author of a Hebrew-Persian dictionary from Urgench (present-day Uzbekistan) in the early 14th century called his native land "Ashkenaz." In the early 20th century, Caucasian Jews were still known by their Lesgian neighbors as "Ashkenazic"(Byhan 1926). The surname Ashkenazic was also occasionally found among the Crimean Krimchaks (Weinreich 2008).

Reconstructing the origin of Yiddish speakers and Yiddish

The most parsimonious explanation for our findings is that Yiddish speaking Ashkenazic Jews have originated from mixed Slavo-Turko-Iranian populations residing in "Ashkenaz" lands centered between the Black and Caspian Seas (Figures 3, 4) that espoused Judaism in a variety of venues by the end of the first millennium A.D. At the height of the Khazar Empire (8th-9th centuries), Hebrew as a native language had been dead for some five-six centuries. In the Empire, Slavic and Iranian had become a major *lingua franca*s. At this time Iranian Jews had brought to the Khazar Empire an Iranianized Judaism, together with the Talmud, as well as written Talmudic Aramaic, Biblical Hebrew, written Hebroid, and spoken Eastern Aramaic and Iranian (Wexler 2010). The Khazars chose Judaism in order to be able to profit from the transit trade across their territories. They appear not to have participated very much as merchants abroad. The Judaization, carried out by lay merchants, of the Khazar élite and the presence of the international Jewish merchants on the international Silk Roads between China and Europe prompted the Irano-Turkic- Slavo Jewish merchants to create Yiddish for use in Europe, Loterā'i for use in Iran (this was a cryptic language first cited in 10thcentury Azerbaijan, which survives to present day), and the many variants of cryptic Hebrew and Hebroid lexicon for the use of Jewish merchants throughout Afro-Eurasia (Wexler 2010). This is evident in both genetic and linguistic evidence, by the biogeographical proximity of Yiddish speakers to Iranians and Iranian Jews (Figures 3, 4), the large diversity of maternal and paternal haplogroups in Yiddish speakers (Figure 3), and the existence of over 250 terms in the semantic field of "buying and selling" in Yiddish, most of which were Hebroidisms, Germanoidisms, and Slavisms, with only a handful of authentic German terms (Wexler 2011a). While Hebrew could serve as the basis of the international cryptic trade lexicon, it could not serve as a full-fledged language since no one could speak the language.

In the 9th century a Persian postal official in the Baghdad Caliphate, ibn Khordāδbeh, described the Iranian Jewish traders, who by then may have already become a tribal confederation of Slavic, Iranian, and Turkic converts to Judaism, as conversant in the main components of Yiddish: Slavic, German, Iranian, Hebrew, in addition to several other languages. The total number of languages given was six, but some of his language names were most likely abbreviations of sets of languages, e.g. *'andalusijja* probably denoted Andalusian Arabic, Berber and various forms of Ibero-Romance. In addition to Slavic Yiddish, the only surviving type of Yiddish in our day, there were probably also Iranian and Turkic "Yiddishes" that were invented in the Khazar Empire for use in diverse geographical locales. We can surmise this on the basis of a major Iranian and a minor Turkic component of Yiddish which not only can be identified for Slavic Yiddish, but actually predicted for it.

When the Khazar Empire lost its prominence and the Jewish monopoly on the Silk Road ended (c. 11th century), the relexification process was gradually abandoned (Wexler 2002). At that point, Slavic Yiddish became the first and only spoken and written language of the European Ashkenazic Jews (Iranian remained the language of the Central Asian and Iranian Ashkenazic Jews—and both groups continued to call themselves "Ashkenazic" up to the present) and began to absorb more German influence post-relexificationally (Wexler 2011a). This process, however, was not accompanied by massive gene transfers between Jews and non-Jews (Figure 3), likely due to the severe restrictions set on mixed marriages by the Middle Ages Christian authorities (Sand 2009). If one examines the "German" and "Hebrew" component of contemporary Yiddish, one can still see the enormity of the Germanoid and Hebroid component in comparison to genuine Germanisms and Hebraisms. To take one example, Yiddish *unterkojfn* 'to bribe' has German components ('under' + 'to buy'), but the combination and meaning are impossible in all forms of German, past or present[31]. The fusing of diverse Iranian, Turkic, and first and foremost Slavic, components in Yiddish is probably a reflection of what happened in other mixed tribal confederations in Europe in the 8th-10th centuries.

Further evidence that Yiddish is a relexified Slavic language can be found in the many customs and their names concerning the Jewish religion, which were probably introduced by Slavic converts to Judaism. For example, the Yiddish term *trejbern* 'to remove the forbidden parts of the animal to render the meat kosher' is from Slavic, see e.g. Ukrainian *terebyty* 'to peel, shell; clean a field' (the Yiddish meaning is obviously innovative). Other Ashkenazic custom of distinctly non-Jewish origin include breaking a glass at a wedding ceremony (Slavic and Iranian) (Wexler 1993). A striking fact that is hardly ever appreciated is that Yiddish *košer* 'kosher' is not a Hebraism, as is widely believed (it appears centuries after the demise of colloquial Semitic Hebrew), but the source of the term is a common Iranian word meaning 'to slaughter an animal', see e.g. Ossete *kušart* 'animal slaughtered for food'. Apparently, it was Yiddish speakers who "Hebroidized" the Iranianism with the legitimate Biblical Hebrew *kašer* which meant only 'fit, suitable' but had no connection to food. Many of the Arabic-speaking Jews to this day do not use the Hebrew/Hebroid term at all.

Limitations

Our study has three limitations. First, because our study is the first to analyze the genomes of Yiddish speaking Ashkenazic Jews, a caution is warranted in interpreting some of our results due to the choice of data, method, and samples. Second, DNA samples were genotyped on the GenoChip (Elhaik et al. 2013), which is relatively small in size and does not allow carrying out haplotype- or identity-by-descent (IBD) analyses, although previous IBD findings support our findings (Atzmon et al. 2010; Elhaik 2013). By contrast, GenoChip's is able to detect 80-90% of known haplogroups. Finally, GPS infers the geographical origins of an individual by averaging over the origins of all its ancestors, raising doubts as to whether the reported area is the actual origin or middle point of several origins. We have accounted for that by carrying a separate analysis that confirmed the high genetic similarity between Yiddish speakers, modern Turks (Figure S2), and simulated "native" Turks (Figure 4).

Conclusions

Language is the atom of a community, the molecule that binds its history, culture, behavior, and identity, and the compound that unites its geography and genetics together. The origin of Yiddish Speaking Ashkenazic Jews in Europe is among the most enigmatic and underexplored pages in the history of the region. The linguistic approaches utilized to answer this question have thus far provided inconclusive results. The popular view of

Yiddish origins claims that the language is a dialect of High German that arose in the 9-10th centuries when Romance-speaking Jews settled in the Rhineland (Weinreich 2008). The absence of local Rhineland German dialect features subsequently prompted linguists to relocate its birthplace to Bavaria (King 2001) despite the fact that Yiddish grammar and phonology are Slavic (with some Irano-Turkic input) and only some of the lexicon is German. This suggests Yiddish is a unique Slavic language, relexified to a minority of German and Hebrew and a majority of newly coined Germanoid and Hebroid elements, based mainly on Slavic and Iranian models. To shed light on this question, we carried a genetic analysis aimed to find the origins of Yiddish speakers.

We traced nearly all the individuals to major primeval trade routes in northeastern Turkey adjacent to primeval villages, whose names may be derived from "Ashkenaz." We have also found that Yiddish speakers are genetically closest to Turks and southern Caucasian populations, indicating a common origin in "Ashkenaz" lands. Our findings imply that Yiddish (and some other initially secretive Jewish trade languages) was created by Slavo-Iranian Jewish merchants plying the Silk Roads between Germany, North Africa, and China, while its unique Hebroid component was diffused among all Jewish trading venues (and was to become a major component of [Slavic] Palestinian/Israeli "Hebrew" beginning in the late 19th century). The Silk Roads also explain the few Mongolian and Chinese influences in Yiddish and the existence of rare Asian maternal haplogroups in Ashkenazic Jews (Tian et al. 2015). We have rejected the Rhineland hypothesis purporting that "Ashkenaz" corresponds to Germany. In fact, the changing meanings of *Ashkenaz(ic)* reflect language shifts; the original meaning of *Ashkenaz(ic)* was (a) "Scythian" (Iranian), which in the Khazar Empire came to denote (b) "Slavic" in the 10th century (Slavs were a major part of the Khazar population and some became Jewish), and finally (c) "German/Yiddish" (relexified Slavic) in 12th-century Europe. Applications of our approach may provide clues to the origin of two dozen other worldwide languages that went through relexification, like the Romani and Creole languages (Horvath and Wexler 1997).

Methods

Sample collection

Genetic data of Ashkenazic Jews. The National Geographic Society's Genographic Project contains genetic and demographic data from over 320,000 anonymous participants (https://genographic.nationalgeographic.com/). Participants were genotyped on the GenoChip microarray that includes nearly 150,000 highly informative Y-chromosomal, mitochondrial, autosomal, and X-chromosomal markers (Elhaik et al. 2013). All participants provided written informed consent for the use of their DNA in genetic studies. Jews represent ~4% of individuals in the database, of which 55% have self-identified as Ashkenazic Jews and 5% as Sephardi Jews.

We accessed the Genographic Project's database through http://geno-web.nationalgeographic.com/geno2/dist/. Our search in this database (January 2015) retrieved 367 individuals who reported having two Ashkenazic Jewish parents, 187 of which had reported having two parents who spoke solely Yiddish were considered "Yiddish speakers" and the remaining "non-Yiddish speakers." Of the latter, 29 individuals had parents who did not speak Yiddish, but their results were similar to those of their cohort. Demographic and genetic data (Table S3) were stripped from information that could lead to identification. The mtDNA notation corresponds to build B16 and the Y haplogroup notation corresponds to the 2015 tree. The mutations associated with the mtDNA and Y chromosomal haplogroups (2015 tree and B16 build, respectively) are listed in Tables S4 and S5, respectively. Haplogroup assignment was done by the Genographic Project. Plink (1.07) was used to test the relatedness among Yiddish speakers using the --genome flag. The average PiHat was 1.8% and maximum PiHat was 5.14% indicating the absence of close relatives in our data.

Genetic data of an ancient pre-Scythian individual. Raw reads for the ancient pre-Scythian Iron Age individual were generated by Gamba et al. (2014). Reads were processed through our standardized variant calling pipeline (Pirooznia et al. 2014). Briefly, reads were aligned to the human reference assembly (UCSC hg19 -

http://genome.ucsc.edu/), allowing two mismatches in the 30-base seed. Alignments were then imported to binary bam format sorted and indexed. Optical duplicates were removed. High quality alignments with a minimum mapping quality score of 20 were selected. The Genome Analysis Toolkit (GATK) (McKenna et al. 2010) (2.6) was used by employing genotype likelihoods calculation model to generate both SNP and small indel calls for the data using the GATK Unified Genotyper function. Variants were filtered for a minimum confidence score of 30 and minimum mapping quality of 20. Additional variant recalibration step was conducted and filters were applied for base quality score, strand bias, mapping quality rank sum, read position rank sum, and homopolymer stretches. SNP clusters (>3 SNPs per 10 bp window) were excluded. Finally, calls were converted to plink format. Overall we obtained over 388,000 high confidence SNPs, of which we analyzed over 58,000 that overlapped with the GenoChip microarray.

Genetic data of reference populations. To curate the reference population dataset and demonstrate the validity of our approach, we studied 602 unrelated individuals representing 35 populations and subpopulations with ~16 samples per population. About 250 individuals from 19 populations and subpopulations were obtained from the Genographic Project and the 1000 Genomes Project that were genotyped on the GenoChip microarray (Elhaik et al. 2014). Bedouins and Turks were obtained from (Behar et al. 2010) and Palestinians were obtained from the HGDP dataset (Conrad et al. 2006). The remaining individuals were obtained from Yunusbayev et al. (2011). From this dataset, we selected 202 individuals from 13 populations for which localized geographical origin and sufficient data (>4 samples) were available (Table S1). Eight Iranian Jews were obtained from Behar et al. (2013). From all these datasets, we analyzed only the ~100,000 autosomal markers that overlapped with the GenoChip markers.

Curating a reference population data set

Biogeographical analysis was carried using the Geographic Population Structure (GPS) tool, shown to be highly accurate compared to alternative approaches like spatial ancestry analysis (SPA), that in turn, is slightly more accurate than principal component analysis

(PCA) -based approach for biogeography (Yang et al. 2012; Elhaik et al. 2014). GPS finds the geographical origin of a sample by matching its admixture signature with reference samples of known geographical origin. To infer the geographical coordinates (latitude and longitude) of an individual given K admixture proportions, GPS requires a reference population set of N populations with both K admixture proportions and two geographical coordinates (longitude and latitude). All supervised admixture proportions were calculated as in (Elhaik et al. 2014).

Detailed annotation for subpopulations was unavailable for most populations (Figure S1), though they exhibited fragmented subpopulation structure (Figure 1). To determine the number of subpopulations in each populations, we adopted a similar approach to Elhaik et al.'s (2014). Let $N\alpha$ denote the number of samples per population α; if $N\alpha$ was less than four individuals, the population was left unchanged. For other populations, we used k-means clustering routine with five replications implemented in Matlab. Let X_{ij} be the admixture proportions of individual i in component j. For each population, we ran k-means clustering for $k \in 2$, using $N\alpha\times9$ matrix of admixture proportions (X_{ij}) as input. At each iteration, we calculated the ratio of the mean square and sum of squares between the groups. If this ratio was <0.9 and there are more than three samples in each cluster, then we accepted the k-component model, whereas smaller clusters were removed.

To bolster the accuracy of GPS inferences beyond what has been previously reported (Elhaik et al. 2014), we have updated the reference panel to comprise of highly localized Eurasian populations. For that, we applied GPS to all Eurasian individuals (Table S1) using the leave-one-out procedure at the population level. This approach is more rigorous than the leave-one-out individual procedure and ensures that the reference panel will not be biased by outliers that do not fit with the genetic profile of the region. Individuals predicted to reside within the political borders of their countries or less than 200km outside of them were retained and were used to recompile the reference population set using the technique described above. This procedure was repeated until the rate of correctly assigned individuals exceeded 80%. Due to their extreme geographical locations Germans and Altai could not satisfy the filtering criteria and were supplemented to the final reference panel using the admixture proportions calculated in a previous round. Overall, we included 26 populations, with some appear as two subpopulations, in our reference population set (Figure 2). These populations were considered hereafter as *reference populations*.

The geographical distributions of the reference populations (Figure 1a) were calculated based on the geographical locations and admixture proportion of the reference populations (Figure 2) using the Matlab function TriScatteredInterp that performs linear interpolation of two dimensional datasets. This allowed us to evaluate the admixture proportion of any coordinates within the geographical range of the reference populations (Figure 4d).

Calculating the biogeographical origin of a test sample and genetic admixture distances. GPS coordinates for a test sample were calculated as previously described (Elhaik et al. 2014). Briefly, given a sample of unknown geographical origin and nine admixture proportions that correspond to nine putative ancestral populations, GPS converts the genetic distances between the test sample and the nearest M=10 reference populations to geographic distances. All maps were plotted using the R package rworldmap (South 2011). The Silk Road and trade route maps were plotted according to the maps available from the Stanford Program on International and Cross-cultural Education (SPICE) interactive resource http://virtuallabs.stanford.edu/silkroad/SilkRoad.html. The geographical coordinates of the Turkish place names were obtained from the Geographical Names website (http://www.geographic.org/geographic_names/).

We defined genetic admixture distance (*d*) as the minimal Euclidean distance between the admixture proportions of an individual to those of all individuals of a certain population.

Competing interests

EE is a consultant to DNA Diagnostic Centre.

Authors' contributions

EE initiated the study and carried all analyses with RD and MP. EE and PW wrote the paper.

Acknowledgment

E.E was supported in part by Genographic grant (GP 01-12), The Royal Society International Exchanges Award (IE140020), MRC Confidence in Concept Scheme (xxx), and NSF Standard Grant (#1456634). We thank the many public participants for donating their DNA sequences for scientific studies and The Genographic Project's public database for providing us with their data. We also thank Dr. Ahmet Reyiz Yılmaz for his contribution to the study.

REFERENCES

Atzmon, G.L. HaoI. 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. American Journal of Human Genetics. **86**:850-859.

Balanovsky, O.K. DibirovaA. Dybo et al. 2011. Parallel evolution of genes and languages in the Caucasus region. Molecular Biology and Evolution. **28**:2905-2920.

Baron, S. W. 1952. A Social and Religious History of the Jew, 2nd ed. Columbia University Press, New York.

Behar, D. M., M. Metspalu, Y. Baran, N. M. Kopelman, B. Yunusbayen, A. Gladstein, S. Tzur, H. Sahakyan, A. Bahmanimehr, and L. Yepiskoposyan. 2013. No Evidence from Genome-Wide Data of a Khazar Origin for the Ashkenazi Jews. Human Biology. **85**:859-900.

Behar, D. M.B. YunusbayevM. Metspalu et al. 2010. The genome-wide structure of the Jewish people. Nature. **466**:238-242.

Bouckaert, R., P. Lemey, M. Dunn, S. J. Greenhill, A. V. Alekseyenko, A. J. Drummond, R. D. Gray, M. A. Suchard, and Q. D. Atkinson. 2012. Mapping the origins and expansion of the Indo-European language family. Science. **337**:957-960.

Brandt, G., A. Szécsényi-Nagy, C. Roth, K. W. Alt, and W. Haak. 2014. Human paleogenetics of Europe–The known knowns and the known unknowns. Journal of Human Evolution.

Bray, S. M., J. G. Mulle, A. F. Dodd, A. E. Pulver, S. Wooding, and S. T. Warren. 2010. Signatures of founder effects, admixture, and selection in the Ashkenazi Jewish population. Proceedings of the National Academy of Sciences of the United States of America. **107**:16222-16227.

Brook, K. A. 2006. The Jews of Khazaria. Rowman & Littlefield, Lanham, MD.

Bryer, A., and D. Winfield. 1985. The Byzantine monuments and topography of the Pontos. Dumbarton Oaks Research Library and Collection, Washington, D.C.

Byhan, A. 1926. Kaukasien, Ost- und Nordrussland, Finnland: I. Die kaukasischen Völker. Pp. 659-1022 *in* G. Buschan, ed. Illustrierte Völkerkunde. Strecker und Schroeder, Stuttgart.

Carmi, S.K. Y. HuiE. Kochav et al. 2014. Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins. Nature Communications. **5**:4835.

Cavalli-Sforza, L. L. 1997. Genes, peoples, and languages. Proceedings of the National Academy of Sciences of the United States of America. **94**:7719-7724.

Conrad, D. F., M. Jakobsson, G. Coop, X. Wen, J. D. Wall, N. A. Rosenberg, and J. K. Pritchard. 2006. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. Nature Genetics. **38**:1251-1260.

Costa, M. D.J. B. PereiraM. Pala et al. 2013. A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. Nature Communications. **4**:2543.

Darwin, C. 1871. The descent of man and selection in relation to sex. John Murray London, England.

Elhaik, E. 2013. The Missing Link of Jewish European Ancestry: Contrasting the Rhineland and the Khazarian Hypotheses. Genome Biology and Evolution. **5**:61-74.

Elhaik, E. 2012. Empirical Distributions of F_{ST} from Large-Scale Human Polymorphism Data. PLoS ONE. **7**:e49837.

Elhaik, E.E. GreenspanS. Staats et al. 2013. The GenoChip: a new tool for genetic anthropology. Genome Biol Evol. **5**:1021-1031.

Elhaik, E.T. TatarinovaD. Chebotarev et al. 2014. Geographic population structure analysis of worldwide human populations infers their biogeographical origins. Nature Communications. **5**.

Eller, E. 1999. Population substructure and isolation by distance in three continental regions. American Journal of Physical Anthropology. **108**:147-159.

Everett, C. 2013. Evidence for direct geographic influences on linguistic sounds: the case of ejectives. PLoS ONE. **8**:e65275.

Gamba, C.E. R. JonesM. D. Teasdale et al. 2014. Genome flux and stasis in a five millennium transect of European prehistory. Nature Communications. **5**:5257.

Gil, M. 1974. The Rādhānite Merchants and the Land of Rādhān. Journal of the Economic and Social History of the Orient. **17**:299-328.

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

Harkavy, A. E. 1867. The Jews and the language of the Slavs Vilnius: Menahem Rem (in Hebrew).

Hijmans, R. J. 2014. geosphere: Spherical Trigonometry.

Holo, J. 2009. Byzantine Jewry in the Mediterranean economy. Cambridge University Press, Cambridge, UK.

Horvath, J., and P. Wexler. 1997. Relexification: prolegomena to a research program. Pp. 11-71 *in* J. Horvath, and P. Wexler, eds. Relexification in creole and non-creole languages, Wiesbaden: Harrassowitz.

Jobling, M., M. Hurles, and C. Tyler-Smith. 2013. Human evolutionary genetics: origins, peoples & disease. Garland Science.

King, R. D. 2001. The Paradox of Creativity in Diaspora: The Yiddish Language and Jewish Identity. Studies in the Linguistic Sciences. **31**:213-229.

Kitchen, A., C. Ehret, S. Assefa, and C. J. Mulligan. 2009. Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East. Proceedings of the Royal Society. Series B: Biological Sciences. **276**:2703-2710.

Loh, P.-R., M. Lipson, N. Patterson, P. Moorjani, J. K. Pickrell, D. Reich, and B. Berger. 2013. Inferring admixture histories of human populations using linkage disequilibrium. Genetics. **193**:1233-1254.

Luigi Luca Cavalli-Sforza, P. M., Alberto Piazza. 1994. The history and geography of human genes. Princeton University Press.

McKenna, A.M. HannaE. Banks et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research. **20**:1297-1303.

Ostrer, H. 2012. Legacy: A Genetic History of the Jewish People. OUP USA.

Pirooznia, M., M. Kramer, J. Parla, F. S. Goes, J. B. Potash, W. R. McCombie, and P. P. Zandi. 2014. Validation and assessment of variant calling pipelines for next-generation sequencing. Hum Genomics. **8**:14.
Polak, A. N. 1951. Khazaria - The History of a Jewish Kingdom in Europe (in Hebrew). Mosad Bialik and Massada Publishing Company, Tel-Aviv, Israel.
Ramachandran, S., O. Deshpande, C. C. Roseman, N. A. Rosenberg, M. W. Feldman, and L. L. Cavalli-Sforza. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proceedings of the National Academy of Sciences of the United States of America. **102**:15942-15947.
Roaf, M., T. Sinclair, S. Kroll, S. J. Simpson, DARMC, R. Talbert, S. Gillies, T. Elliott, and B. J. 2015. Places: 874507 (Haza/Hassis). Pleiades.
Sand, S. 2009. The invention of the Jewish people. Verso, London.
Shapira, D. D. Y. 1999. Armenian and Georgian sources on the Khazars: a re-evaluation. Pp. 307-352 *in* P. B. Golden, Ben-Shammai, H., and Róna-Tas, A., ed. The World of the Khazars: New Perspectives - Selected Papers from the Jerusalem 1999 International Khazar Colloquium. Brill, Leiden, Boston.
South, A. 2011. rworldmap: A New R package for Mapping Global Data. Pp. 35-43. The R Journal.
Tian, J.-Y., H.-W. Wang, Y.-C. Li, W. Zhang, Y.-G. Yao, J. van Straten, M. B. Richards, and Q.-P. Kong. 2015. A genetic contribution from the Far East into Ashkenazi Jews via the ancient Silk Road. Sci. Rep. **5**.
Weinreich, M. 2008. History of the Yiddish language. Yale University Press, New Haven, CT.
Wexler, P. 1993. The Ashkenazic Jews: A Slavo-Turkic people in search of a Jewish identity. Slavica Publishers, Columbus, OH.
Wexler, P. 2011a. A covert Irano-Turko-Slavic population and its two covert Slavic languages: The Jewish Ashkenazim (scythians), Yiddish and 'Hebrew'. Zbornik Matice srpske za slavistiku. **80**:7-46.
Wexler, P. 2002. Two-Tiered Relexification in Yiddish: Jews, Sorbs, Khazars, and the Kiev-Polessian Dialect. Mouton de Gruyter, Berlin & New York.
Wexler, P. 2011b. The Myths and Misconceptions of Jewish Linguistics. Jewish Quarterly Review. **101**:276-291.
Wexler, P. 2012. Relexification in Yiddish: A Slavic Language Masquerading As a High German dialect? Pp. 212-230 *in* A. Danylenko, and S. H. Vakulenko, eds. Studien zu Sprache, Literatur und Kultur bei den Slaven: Gedenkschrift für George Y. Shevelov aus Anlass seines 100. Geburtstages und 10. Todestages. Verlag Otto Sagner, München, Berlin.
Wexler, P. 2010. Do Jewish Ashkenazim (i.e. “Scythians”) originate in Iran and the Caucasus and is Yiddish Slavic? Pp. 189-216 *in* E. Stadnik-Holzer, and G. Holzer, eds. Sprache und Leben der frühmittelalterlichen Slaven: Festschrift für Radoslav Katičić zum 80. Geburtstag. Peter Lang, Frankfurt.
Yang, W. Y., J. Novembre, E. Eskin, and E. Halperin. 2012. A model-based approach for analysis of spatial structure in genetic data. Nature Genetics. **44**:725-731.

Yunusbayev, B.M. MetspaluM. Järve et al. 2011. The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. Molecular Biology and Evolution. **29**:359-365.

Tables

Table 1

Two hypotheses regarding the origin of the Yiddish language and its lexicographical. The Rhineland hypothesis differs from the Irano-Turko-Slavic hypothesis by ignoring the Iranian component alongside the “Hebroidisms” and “Germanoidisms,” whose geographical origins are unclear. Both hypotheses, however, agree on the same three basic components: German, Slavic, and Hebrew, though they disagree on their proportions admixture.

Hypotheses	Lexicographical admixture	Geographical origin	References
Rhineland	80% German (Bavaria), 15% Hebrew, and 5% Slavic.	Southeastern Germany	(King 2001)
Irano-Turko-Slavic	Slavic (43%), German and Germanoid (35%), Hebrew and Hebroid (8%), and the remaining (14%) are Iranian, Turkic and unique Romance, Arabic (including Berberized Arabic), and Greek.	1) The Khazar’s Empire 2) Kievan Rus' (today's Ukraine) 3) Sorbian areas of Germany	(Wexler 2010)

Figure legend

Figure 1

Depicting the distributions of nine admixture components. A) Admixture proportions of all populations included in this study. For brevity, subpopulations were collapsed. The x axis represents individuals. Each individual is represented by a vertical stacked column of color-coded admixture proportions that reflects genetic contributions from nine putative ancestral populations. B) The geographical distribution of admixture proportions in Eurasia.

Figure 2

GPS predicted coordinates for individual of Eurasian populations and subpopulations. Individual label and color match their known region/state/country of origin using the following legend: AB (Abkhazian), ARM (Armenian), BDN (Bedouin), BU (Bulgarian), DA (Dane), EG (Egyptian), FIN (Finnish), GO (Georgian), GR (German), GK (Greek), IR (Iranian), ID/TSI (Italy: Sardinian/Tuscan), KR (Kurds), LE (Lebanese), Palestinian (PAL), RO (Romanian), R-A/B/C/I/K/MO/M/N/NO/T (Russians: Altaian/Balkar/Chechen/Ingush/Kumyk/Mordovian/Moscow/Nogai/North Ossetian/Tatar), PT (Pamiri from Tajikistan), TR (Turkmen), TUR (Turk), UK (United Kingdom), UR (Ukranian). Pie charts reflect the admixture proportions and geographical locations of the reference populations. *Note*: occasionally all individuals of certain populations (e.g., Altaians) were predicted to the same spot and thus appear as a single individual.

Figure 3

A map depicting the predicted location of Jewish (triangles) Yiddish speakers (orange) and Iranian Jews (yellow) alongside the ancient pre-Scythian individual (blue diamond). An inset shows the sample distribution in northern Turkey and the locations of the four villages that may derive their names from “Ashkenaz.” Large (13-23%), medium (4-10%), and small (1-4%) circles reflect the percentage of Yiddish speakers’ parents born in each region. The paternal and maternal haplogroups of the Yiddish speakers are shown at the top of the figure.

Figure 4

Comparing Yiddish speakers with "native" individuals from six populations. A) Admixture proportions of Yiddish speakers and all simulated individuals included in this analysis. The x axis represents individuals. Each individual is represented by a vertical stacked column of color-coded admixture proportions that reflects genetic contributions from nine putative ancestral populations. B) The genetic admixture distances (*d*) between the simulated individuals and their nearest modern day populations. C) The geographical coordinates from which the admixture signatures (A) were derived. D) GPS predictions for the admixture signatures of the simulated individuals of the six populations. Pie charts denote the proportion of individuals correctly predicted to the countries of origins, color-coded by the colors of the six countries (C) or white for other countries. The geographical origins of Yiddish speakers previously obtained are shown for comparison. An inset magnifies northeastern Turkey. E) The genetic admixture distances (*d*) within Yiddish speakers and between them to the simulated individuals. F) The proportion of simulated individuals that are geographically closest to Ashkenazic Jewish subgroups.