North African Jewish and non-Jewish populations form distinctive, orthogonal clusters

C.L. Campbell[a,1], P.F. Palamara[b,1], M. Dubrovsky[c,1], L. R. Botigué[d], M. Fellous[e], G. Atzmon[f,g], C. Oddoux[a], A. Pearlman[a], L. Hao[h], B. M. Henn[i], E. Burns[f], C. D. Bustamante[i], D. Comas[d], E. Friedman[c], I. Pe'er[b], H. Ostrer[a,g,2]

a) Department of Pathology, Albert Einstein College of Medicine, Bronx, NY 10461

b) Department of Computer Science, Columbia University, New York, NY 10027

c) The Susanne Levy Gertner Oncogenetics Unit, the Danek Gertner Institute of Human Genetics, Chaim Sheba Medical Center, 52621, Tel-Hashomer, and the Sackler School of Medicine, Tel-Aviv University, Tel-Aviv Israel

d) Institute of Evolutionary Biology (CSIC-UPF), Universitat Pompeu Fabra, 08003 Barcelona, Spain

e) Cochin Institute, INSERM 567 75014 Paris, France

f) Department of Medicine, Albert Einstein College of Medicine, Bronx, NY 10461

g) Department of Genetics, Albert Einstein College of Medicine, Bronx, NY 10461

h) Center for Genome Informatics, New Jersey Medical School, University of Medicine and Dentistry of New Jersey, Newark, NJ 07101

i) Department of Genetics, Stanford University, Stanford, CA 94305

[1]contributed equally to the work

[2]Corresponding author: Harry Ostrer, M.D.
Department of Pathology
Albert Einstein College of Medicine
1300 Morris Park Avenue
Bronx, NY 10461
tel: 718 430-8605
fax: 718 430-2623
email: harry.ostrer@einstein.yu.edu

Category: Biological Sciences; Population Biology

Abstract

North African Jews constitute the second largest Jewish Diaspora group. Yet, their relatedness to each other, to European, Middle Eastern, and other Jewish Diaspora groups and to their former North African non-Jewish neighbors has not been well-defined. Here, genome wide analysis of five North African Jewish groups (Moroccan, Algerian, Tunisian, Djerban and Libyan) and comparison with other Jewish and non-Jewish groups demonstrated distinctive North African Jewish population clusters with proximity to other Jewish populations and variable degrees of Middle Eastern, European and North African admixture. Two major sub-groups were identified by principal component, neighbor joining tree, and identity by descent (IBD) analysis — Moroccan/Algerian and Djerban/Libyan that varied in their degree of European admixture. These populations showed a high-degree of endogamy and were part of a larger Ashkenazi and Sephardic Jewish group. By principal component analysis, these North African groups were orthogonal to contemporary populations from North and South Morocco, Western Sahara, Tunisia, Libya and Egypt. Thus, this study is compatible with the history of North African Jews — founding during Classical Antiquity with proselytism of local populations, followed by genetic isolation with the rise of Christianity and then Islam, and admixture following the emigration of Sephardic Jews during the Inquisition.

\body

Introduction

Jews lived in multiple communities in North Africa for over 2000 years (1). Successive waves of migration from the Middle East and Europe as well as conversion and admixture of local populations (mostly thought to be Berber and here termed, "Maghrebi") contributed to the formation of Jewish communities (2). Although termed "Sephardic," the formation of these communities antedated the presence of Jews in the Iberian Peninsula with significant admixture occurring only following the expulsions from Spain and Portugal 1492 and 1497, respectively (2). From that time up to their migration to current Israel starting in the 1940s and more massively in the 1950s, each of these populations lived in relative seclusion and was endogamous (3, 4). This led to their developing the characteristics of genetic isolates, such as high frequencies of founder mutations for Mendelian disorders and limited repertoires of mitochondrial and Y chromosomal haplotypes (5-7).

The relatedness of these Jewish groups to each other, to European and Middle Eastern Jews, and to their non-Jewish North African neighbors has been addressed in only a fragmentary fashion in prior studies (8-14). Most studies were limited to one or two North African groups. One study challenged the story line of Judean migrants, Berber tribesmen, and Sephardic Jewish refugees contributing to the formation of these groups by demonstrating shared ancestry between Libyan Jews and Yemenite and Ethiopian Jews — groups that are thought to have limited Middle Eastern Jewish ancestry (15).

Previously, using genome wide SNP and CNV data, we demonstrated that Sephardic (Greek and Turkish), Ashkenazi (Eastern European), and Mizrahi (Iranian, Iraqi and Syrian) Jews

with origins in Europe and the Middle East were more related to each other than to their non-Jewish contemporary neighbors (16). We showed that this relatedness could be explained on the basis of sharing DNA segments identical by descent within and between populations. Here, we build on this understanding of the Jewish Diasporas by extending our analyses to members of the Jewish communities in Morocco, Algeria, Tunisia, Djerba, Libya, Ethiopia, Yemen, and Georgia and to members of non-Jewish communities from the same regions. We present a comprehensive population genetic analysis of North African Jews, a group that comprises the third major group of World Jewry, following European and Middle Eastern Jews. In addition, we extend these analyses to Georgian, Yemenite, and Ethiopian Jews, thus developing a more comprehensive genetic map for Jewish population genetics.

Results

North African Jewish populations form distinctive clusters with genetic proximity to each other and to European and Middle Eastern Jewish groups. SNP data were generated for 509 unrelated individuals (60.5% female) from the 15 Jewish populations (Table 1). These SNP data were merged with selected datasets from the Human Genome Diversity Panel (HGDP) to examine the genetic structure of Jewish populations in both global and regional contexts (Figure 1 and SI Appendix Figure 1). The first 2 principal components of worldwide populations showed that the North African Jewish populations clustered with the European and Middle Eastern Jewish groups and European non-Jewish groups, but not with the North African non-Jewish groups, suggesting origins distinctive from the latter (Figure 1A). Georgian Jews formed part of this cluster, whereas Yemenite and Ethiopian Jews did not. When compared only to the

European, Middle Eastern and North African Jewish and non-Jewish populations, the North African Jewish populations formed a common, but distinctive, cluster, observed by principal components 1 and 2, that was overlapping with the Greek and Turkish Sephardic Jewish cluster (Figure 1B). Each of these Jewish groups, in turn, formed a distinctive cluster, observed by principal components 1 and 3, that demonstrated a west to east cline with Algerian and Moroccan Jews in proximity to the Sephardic Jewish populations. The Tunisian Jews exhibited two apparent clusters, one, with proximity to Libyan and Djerban Jews and the other proximal to the Moroccan and Algerian Jews (Figure 1C).

The neighbor joining tree supported this clustering of the Jewish populations with the previously described European/Syrian and Middle Eastern branches being discernible (Figure 2). The European Turkish, Greek and Italian Jews shared a common branch with Ashkenazi and Syrian Jews forming connections to this branch. The North African populations added a sub-branch to the European/Syrian branch. In turn, this bifurcated into Moroccan-Algerian and Tunisian-Djerban-Libyan sub-branches. As reported previously, the Middle Eastern Jewish branch included the Iranian and Iraqi Jews and the non-Jewish Adygei. This branch was observed now to include Georgian Jews. The Yemenite and Ethiopian Jews were on distinctive branches with the Yemenite Jews on a branch between Palestinians and Bedouins. The robustness of this phylogenetic tree was demonstrated by the fact that a majority of branches was supported by greater than 90% of bootstrap replications.

Pairwise F_{ST} analysis indicated that each of the North African Jewish populations was distinct and by bootstrap analysis statistically different from all of the others (SI Appendix

Tables 1 (top) and 2). Although F_{ST} may be sensitive to small samples sizes, these population differences were confirmed by ANOVA on the PCA eigenvectors ($p<0.05$, SI Appendix Table 3), with the exception of Algerian and Moroccan Jews, who were found to overlap. These differences were also confirmed by permutation testing of between-group IBS for all pair-wise comparisons of the 15 Jewish populations (SI Appendix Table5). In addition, the exact test for population differentiation also indicated that these populations were significantly different ($p<0.001$), again with the exception of Algerian and Moroccan Jews ($p=0.25$). These findings demonstrated that the most differentiated of the North African Jewish populations was Djerban (average F_{ST} to all other Jewish populations 0.026). The smallest F_{ST} was between Greek and Turkish Sephardic Jews ($F_{ST} = 0.0024$) who were close, in turn, to Italian, Algerian, Moroccan and Ashkenazi Jews. The second smallest F_{ST} observed was between Algerian and Moroccan Jews ($F_{ST} = 0.0027$). As a point of reference, the average pairwise F_{ST} between Jews and non-Jews (excluding African and Asian reference populations) was 0.019. Thus, North African Jews were identifiable as a third major group along with Middle Eastern Jews and European/Syrian Jews, albeit with a higher degree of relatedness to European Jews.

North African Jewish populations showed a high-degree of endogamy and IBD sharing between Jewish groups. We studied the frequency of identical by descent (IBD) haplotypes shared by unrelated individuals within and across the groups analyzed. When IBD within populations was examined, the non-Jewish Tunisian Berbers exhibit the highest level of haplotype sharing, suggesting a small effective population size and high levels of endogamy (Figure 3A) (17, 18). With the exception of this Tunisian cohort, the Jewish populations

generally showed higher IBD sharing than non-Jewish groups, indicating greater genetic isolation.

The relationships of the Jewish communities were outlined further by the IBD sharing across populations (SI Appendix Tables 1 (bottom) and 4 and Figure 3B), as the Jewish groups generally demonstrated closer relatedness with other Jewish communities than with geographically near non-Jewish populations. In particular, North African Jewish communities show some of the highest levels of cross-population IBD sharing for the average pair of individuals (SI Appendix Figure 5). A strong degree of relatedness was observed across individuals from the Djerban, Tunisian and Libyan Jewish communities. Noticeable proximity was also found between Jewish Algerian samples and other North African Jewish cohorts, such as Moroccan, Tunisian, Libyan and Djerban Jews, and across individuals from the Tunisian and Moroccan Jewish groups. Among non-Jewish North African groups, Algerians, South Moroccans and West Saharan samples were found to share on average a smaller proportion of their genome IBD to other cohorts.

North African Jewish and non-Jewish populations vary in their proportions of European and Middle Eastern ancestry. By STRUCTURE analysis, the North African Jewish groups demonstrated inferred North-African – Middle Eastern ancestry with varying inferred European ancestry (Figure 4 and SI Appendix Figure 2). The proportion of inferred European ancestry increased from east to west, with Moroccan Jews demonstrating the highest proportion. In contrast, the neighboring non-Jewish North African populations demonstrated substantially higher inferred North African ancestry and less European ancestry. In addition, these

neighboring non-Jewish populations showed inferred Sub-Saharan ancestry that was not demonstrable by STRUCTURE analysis in the North African Jewish populations. Reflecting the high degree of relatedness and shared ancestry among Tunisian, Djerban and Libyan Jews, a novel (red) component was observed at K=6 & 7. Similarly, reflecting the high degree of relatedness and shared ancestry among Iranian and Iraqi Jews a novel (green) component was observed at K=7.

Using Xplorigin to perform ancestry deconvolution for a subset of the populations, the Maghrebi (Tunisian non-Jewish), European (Basque) and Middle-Eastern (Palestinian) ancestry components of North African Jewish communities were compared to the corresponding non-Jewish groups (Figure 5). A stronger signal of European ancestry was found in the genomes of Jewish samples, with a decreased fraction of Maghrebi origins, whereas the Middle-Eastern component was comparable across groups. In Jewish groups geographical proximity to the Iberian Peninsula correlates with an increase in European ancestry and a decrease in Middle-Eastern ancestry, while the Maghrebi component is only mildly reduced. Differences in ancestry proportions were found to be significant ($p<0.05$), except for the Maghrebi component of non-Jewish Northern Moroccan compared to non-Jewish Algerian samples, and the European component of Jewish Moroccan compared to Jewish Algerian samples.

Given the distinctive genetic identity of the Basque population compared to other European populations (19), we also run the Xplorigin analysis using 48 randomly selected HGDP Russian haplotypes as a reference for the European ancestral component of the analyzed populations. The results of such analysis suggested that while the estimated European

components using Basque or Russian reference genomes are correlated, short haplotype frequencies found in the Russian samples are less representative of the analyzed groups' European ancestry (SI Appendix Figure 6 and Figure 5).

In addition to genome-wide proportions, this ancestry painting analysis was intersected with regions that harbor long-range IBD haplotypes. The ancestry of these loci likely reflects more recent demographic trends, since long IBD haplotypes are co-inherited from more recent common ancestors than the average genomic locus. In Jewish populations the ancestry proportions in corresponding IBD regions highlighted mild, but in some cases significant, deviations from genome-wide averages (SI Appendix Figure 3, detail in SI Appendix Table 6), whereas stronger differences were observed in the recent ancestry for the corresponding non-Jewish communities. In these groups recently co-inherited regions exhibited significantly increased European ancestry, with significantly decreased Maghrebi ancestry, when compared to genome-wide averages. This phenomenon was generally stronger for loci shared IBD with individuals from Jewish communities. This effect may be interpreted in several ways: (i) This may be due to the inherently higher European ancestry of Jewish segments planted into the genomes of non-Jewish populations. (ii) Alternatively, the difference in genome-wide ancestries between Jewish and non-Jewish groups alone could explain this observation in the case of recent symmetric gene flow in both directions. However, this second scenario alone is inconsistent with the data, as it would imply a comparable decrease of European ancestry in regions IBD to non-Jewish populations to be observed in Jewish genomes. (iii) Finally, the observed increase of European ancestry could be similarly explained by European segments newly planted in both populations. This explanation is also unlikely, as it would result in a

comparable increase of European ancestry in Jewish genomes, which is instead observed to only mildly increase compared to genome-wide averages. The increase in European ancestry is stronger in IBD regions of length between 3 and 4 cM, compared to regions at least 4 cM long (SI Appendix Table 7). This is compatible with European admixture occurring several generations before present, through ancestors that resided in the Iberian Peninsula.

Ethiopian and Yemenite Jewish populations form distinctive clusters, whereas Georgian Jews do not. As noted, by PCA, Georgian Jews formed part of the Jewish cluster, whereas Yemenite and Ethiopian Jews did not. ANOVA on the PCA Eigenvalues and pairwise F_{ST} analysis indicated that the Ethiopian, Yemenite and Georgian Jewish populations were distinct and statistically different from all of the others (SI Appendix Tables 1 (top), 2 and 3). Ethiopian Jews fell outside of the main three Jewish genetic clusters with the highest average genetic differentiation when compared to all other Jewish groups (F_{ST} = 0.047). They were most closely related to non-Jewish Libyans and South Moroccans (F_{ST} = 0.019) and then to the other North African and Middle Eastern non-Jewish populations. Their closest (yet still quite distant) Jewish neighbors were Yemenite Jews (F_{ST} = 0.038). Likewise, they showed little IBD sharing with other Jewish populations (Figure 3). By STRUCTURE analysis, their ancestry appeared to be of North African, Middle Eastern and sub-Saharan origin with little European contribution (Figure 4).

Despite forming a cluster on principal component analysis and neighbor joining tree that appeared intermediate to Jews and Middle Eastern non-Jews, the Yemenite Jews were genetically closest to Egyptians by F_{ST} (0.008), followed by Middle Eastern non-Jews, then Turkish and Greek Jews (F_{ST} = 0.010 and 0.012, respectively); however, their mean F_{ST} to all other Jewish populations was similar to that of all other Jewish populations with the exception

of Ethiopian Jews (SI Appendix Figure 4). Their mean levels of IBD sharing with Jewish populations were comparable to the mean levels of IBD sharing of other Jewish populations, except Ethiopian Jews (SI Appendix Figure 4). By STRUCTURE analysis, their inferred ancestry was predominantly Middle Eastern and North African with little European contribution.

As noted, the Georgian Jewish cluster overlaps the overall Jewish cluster and the Georgian Jewish sub-branch on the neighbor joining tree was intermediate to those of Iranian and Iraqi Jews and the Adygei. The Georgian Jews had a low F_{ST} when compared to Sephardic Jews (mean F_{ST} = 0.009) despite their similarity with Iranian and Iraqi Jews in the neighbor joining tree and PC analysis. By pairwise IBD sharing with other Jewish populations, the Georgian Jews fell within the pattern of Jewish relatedness. By STRUCTURE analysis, their inferred ancestry was predominantly Middle Eastern and European with little North African contribution. Notably, they shared a small proportion of the novel K=7 (green) component that was observed in the Iranian and Iraqi Jewish populations. This may be due to either small sample size or genetic drift and founder effect.

Discussion

This study supports and expands the classification of Jewish populations that has been developed by us and others (14-16). It defined North African Jews as a distinct branch with significant relatedness to European Jews and Middle Eastern Jews with both being part of a larger Jewish cluster. Within this branch are two sub-branches, the highly endogamous and related Djerban, Libyan and Tunisian Jews and the more European Algerian and Moroccan Jews. With these methods, the Tunisian Jews could be differentiated into two sub-clusters by PCA

that were more Libyan/Djerban-related and more Moroccan/Algerian-related and the Moroccan and Algerian Jews could be demonstrated to be no different from a single population by F_{ST} and the exact test of population differentiation. All of these populations were differentiated from the current non-Jewish populations in these countries reflecting distinctive genetic histories.

These observations are consonant with the history of Jews in North Africa, which stretch back to the earliest recorded history of the region (1-4, 20). Israelite traders may have been among the earliest Phoenician traders who colonized the African coast and established Carthage. The first evidence for Jews in North Africa is from 312 BCE when King Ptolemy Lagi of Egypt settled Jews in the cities of Cyrenaica in current-day Tunisia. The later Pax Romana facilitated communication among the Jewish communities of the Mediterranean Basin and assured establishment of Judaism in the two African provinces of Proconsular (Libya and Tunisia) and Caesarean (Algeria and Morocco) – reflecting the current sub-branches. Following the destruction of the Second Temple in Jerusalem by Roman Emperor Titus in 70 CE, thirty thousand Jews were deported to Carthage in current day Tunisia. Josephus reported the presence of 500,000 Jews in Cyrenaica in the 1st century CE. Jewish communities have been identified from the synagogue remains at Carthage and at least thirteen other sites and Saint Augustine wrote about Jewish communities at Utica, Simittra, Thusurus and Oea. Thus, Jewish communities originated in pre-Classical Antiquity and expanded during Classical Antiquity to grow quite large and cover a significant proportion of North Africa.

As demonstrated from these analyses, both admixture and isolation with endogamy contributed to the formation of these groups. Judaism is thought to have spread among the

indigenous Berbers of North Africa through proselytism, although the degree is unknown. This was at a time when proselytism to Judaism was quite common. Significant admixture occurred with the Sephardic Jews of Spain following their expulsion during the Inquisition in 1492 and later, when tens of thousands migrated to the Mahgreb (Morocco, Algeria and Tunisia) and fewer going elsewhere. This is reflected in the higher proportion of European ancestry among Moroccan and Algerian Jews and the greater genetic proximity of these groups and high degree of IBD segment sharing with Sephardic Greek, Turkish and Italian Jews. Admixture occurred with North African non-Jewish populations. Yet, as demonstrated by the apparent unrelatedness of the Jewish and non-Jewish North Africans, these events were not recent.

Isolation began for Jews when the Roman Emperor, Constantine, converted to Christianity and made this the state religion of the Roman Empire. In the process, Jews were deprived of their right to convert pagans or accept proselytes. Most Christian communities of North Africa disappeared following the Arab conquest of the 8th century CE. The Jewish communities remained, but were subject to religious and civil suppression. The Jews tended to live in their own special quarters and the area as a whole was fragmented into small tribal states. The resulting high degree of endogamy within the North African Jewish populations is reflected in the high degree of within-population IBD sharing, the patterns of within-population rare Mendelian disorders, and the identification of a shared inferred ancestral component among the Djerban, Tunisian and Libyan Jews by STRUCTURE, neighbor joining tree and, to a less obvious degree, by PCA. Such a uniquely shared inferred ancestral component was also observed among the Middle Eastern/Caucasian — Iranian, Iraqi and Georgian Jews.

These results are in agreement with previous population genetic studies of North African Jews, yet significantly expand their observations by using larger numbers of populations and some novel contemporary methods. Earlier studies based on blood group markers and serum proteins differentiated North African Jews from other Jewish groups and from non-Jewish North Africans (8-13). A more recent study identified a distinctive signature for Libyan Jews (15). Here, this signature was confirmed and shown to be shared by Djerban and Tunisian Jews. A global study of Jewish population genetics from 2010 that partitioned most Jewish genomes into Ashkenazi–North African–Sephardic, Caucasus–Middle Eastern, and Yemenite subclusters demonstrated that an Ethiopian subcluster was close to the local population, in accordance with what is observed here (14). A previous study of monoallelic matrilineal inheritance demonstrated limited mitochondrial lineages in Tunisian and Libyan, but not in Moroccan, Jews which is observed in this study as a high degree of extended IBD sharing among the more endogamous Tunisian and Libyan populations (5).

The observations for Georgian and Ethiopian Jews meet historical expectations – Georgian Jews are an outgrowth from the Iranian and Iraqi Jewish communities and Ethiopian Jews are an ancient community that had relatively few, if any Jewish founders from elsewhere and existed in isolation over two thousand years. Nonetheless, the low F_{ST} between Sephardic and Georgian Jews suggests that the latter may have had significant contact with Turkish or Syrian Jews. The observations for the Yemenite Jews are even more surprising. Like the Ethiopian Jews, this population was founded over 2,000 years ago and was thought to be comprised mostly of local proselytes – this is reflected in the distinctive clustering of the population away from other Jewish groups and the mostly Middle Eastern ancestry present in

this group. Yet, the observation of comparable F_{ST} and IBD sharing with other Jewish communities implies significant common Jewish founders in the absence of more recent genetic flow into the community. Thus, although Jewishness was transmitted by the flow of ideas and genes, both appear to have been under selection for long periods of time.

Material and Methods

Recruitment and genotyping of Jewish populations. Recruitment of Jews of European and Middle Eastern origin was described previously (16, 21), and followed a New York University School of Medicine Institutional Review Board-approved protocol (07-333 "Origins and Migrations of Jewish People"). Recruitment of North African, Ethiopian, and Georgian Jews occurred at Sheba Medical Centre in Tel Hashomer, Israel following a local ethics committee and an Israeli Ministry of Health Institutional Review Board approved protocol. Recruitment of non-Jewish individuals from seven different North African locations (North Morocco, South Morocco, Western Sahara, Algeria, Tunisia, Libya and Egypt) was reported previously and followed a Universitat Pompeu Fabra, Ethics Committee-approved IRB protocol (17). In every case, subjects provided informed consent. Jewish subjects were included only if all 4 grandparents came from the same Jewish community. Subjects were excluded if they were known first or second degree relatives of other participants based on IBD analysis. Cryptic relatives were defined as pairs of individuals sharing more than a total of 800 cM and more than 10 IBD segments longer than 10 cM. These values are conservative cutoffs for the

exclusion of first-degree cousins or closer relatives (22, 23). Such pairs were used for further analysis by removing one individual for each detected cryptic relationship.

Analytical methods. DNA preparation and genotyping using the Affymetrix Genome-Wide Human SNP Array 6.0 was performed as previously described (16). Our Jewish data set was then merged with selected data from the HGDP world populations (24) (run using Illumina HumanHap650K Beadchips) to enable comparisons with non-Jewish world-wide populations. Following merging of these data sets from two different platforms and filtering out low call-rate (<5%) and symmetric SNPs (alleles A-G or C-T), there were 163,199 SNPs remaining. A further filtering step was then applied, keeping only markers with no failures, and resulted in a set of 46,324 SNPs, which were used for principal component and F_{ST} calculations. Principal component analysis was performed using the SMARTPCA program from the EIGENSOFT package (v3.0) (25). First, outlier removal was performed using the default parameters (samples greater than 6 standard deviations from the mean in any of the top 10 eigenvectors removed over 5 iterations). Differences between subgroups were assessed using analysis of variance (ANOVA) of the top three eigenvectors.

F_{ST} values were calculated for each population pair using Genepop (26). Confidence intervals were estimated using a bootstrap test in which markers were sampled with replacement for 500 iterations. These data were also used to generate a consensus tree using the neighbor-joining method implemented in PHYLIP (v3.69) (27). An exact test of population differentiation (28) was also carried out using Genepop.

Population structure was inferred using the program STRUCTURE (v2.3) (29, 30). A subset of 5,113 markers were chosen that had high total absolute differences in allele frequencies among pairs of populations, and low linkage disequilibrium. The main parameters for STRUCTURE included 30,000 burn-in and data collection iterations, a separate alpha estimate for each population, and assumption of an admixture model with correlated allele frequencies. Ten repeats of the program were run for each K-value (3-7), and the results were combined into a single Q matrix using CLUMPP (v1.1.2) (31) to resolve label-switching issues.

For IBD discovery, a total of 598,260 SNPs were used. Shared IBD segments were detected using the GERMLINE software package (32) version 1.5. All individuals were computationally phased using Beagle (33) version 3.0.1 and then processed with GERMLINE using the parameters "-min_m 3 -err_hom 0 -err_het 1 -bits 64 -g_extend". To circumvent the problem of phase errors for long-range haplotypes, the genotype extension mode of the GERMLINE algorithm creates a dictionary of short, locally phased haplotypes for all individuals. Pairs of matching short haplotypes are then extended only considering mutually homozygous markers until a specified density of mismatching sites is encountered. Matching segments are reported if their length passes a minimum required centiMorgan length (3 cM for this analysis).

Ancestry deconvolution was performed on a subset of North African groups using the Xplorigin software package, as described previously (23, 34). Briefly, the Xplorigin algorithm builds a database of short haplotype frequencies from three reference populations. This database is then used to probabilistically infer local ancestry for the analyzed samples, assuming each sample results from the admixture of the reference populations. Samples of

North African origins were analyzed with respect to their Maghrebi, Middle Eastern and European ancestry using 36 non-Jewish Tunisian Berber, 48 Palestinian and 48 Basque reference haplotypes, respectively. Basques have experienced very low levels of gene flow from non-European populationscompared to other neighboring European populations, and, for this reason, are good proxies of the European gene pool (17, 19). The analyzed non-Jewish sample comprised 19 Algerians, 17 Libyans, 18 Northern Moroccans, 16 Southern Moroccans and were compared to 24 Jewish Algerians, 37 Jewish Libyans and 38 Jewish Moroccan individuals. A total of 163,199 SNP markers overlapped between different platforms and were used for the ancestry analysis. Genome-wide ancestry proportions were obtained by averaging the inferred ancestry proportions for all analyzed sites within individual groups. Ancestry deconvolution of shared haplotypes was limited to regions detected to be IBD across pairs of individuals from the analyzed groups. Due to phase uncertainty, the reported values were obtained by averaging the ancestral proportions of both maternal and paternal chromosomes in genomic regions delimited by IBD segments. Permutation testing was performed by randomly dropping IBD segments on the genomes of sharing pairs then recomputing ancestry proportions, 10,000 times. We report as significant IBD ancestry values for which 97.5% of permutations result in lower/higher ancestry score.

The significance of the differences of each ancestral component across populations was assessed using a re-sampling procedure. For each individual in a population, we estimated the genome-wide ancestry proportions resulting from the Xplorigin analysis. For each population, we then created 100,000 datasets by randomly sampling with replacement individuals from the

original group. For each such random dataset, we computed the average ancestry of all ancestral components, and compared all cross-population differences.

Acknowledgements

This study was supported, in part, by the Lewis and Rachel Rudin Foundation, the Iranian-American Jewish Federation of New York, the U.S.-Israel Binational Science Foundation, the US National Institutes of Health (5 U54 CA121852), and Ruth and Sidney Lapidus. LRB and DC were supported by MCINN grant CGL2010-14944/BOS. The authors thank Anmol Tiwari for his contributions to the permutation testing, Alon Keinan, Yongzhao Shao, and Simon Gravel for useful technical discussions; Lanchi U, Malka Sasson, Bridget Riley, Jidong Shan, and David Reynolds for technical contributions; and Lawrence Schiffman, Aron Rodrigue and Harvey Goldberg for informative historical discussions.

References

1. Baron SW (1952) *A social and religious history of the Jews* (Columbia University Press, New York,) 2d Ed pp v. <1-18 >.
2. Stillman NA (1979) *The Jews of Arab lands : a history and source book* (Jewish Publication Society of America, Philadelphia) 1st Ed pp xxx, 473 p., 412 leaves of plates.
3. Hirschberg HZ (1974) *A history of the Jews in North Africa* (Brill, Leiden) 2d rev. Ed.
4. Chouraqui A (1968) *Between East and West; a history of the Jews of North Africa* (Jewish Publication Society of America, Philadelphia,) [1st Ed pp xxii, 376 p.
5. Behar DM*, et al.* (2008) Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. *PLoS One* 3(4):e2062.
6. Ostrer H (2001) A genetic profile of contemporary Jewish populations. *Nat Rev Genet* 2(11):891-898.
7. Manni F*, et al.* (2005) A Y chromosomal portrait of the population of Djerba (Tunisia) to elucidate its complex demographic history. *Bull Memoires Soc Anthro Paris* 17(1-12).
8. Bonne-Tamir B, Ashbel S, & Bar-Shani S (1978) Ethnic communities in Israel: the genetic blood markers of the Moroccan Jews. *Am J Phys Anthropol* 49(4):465-471.
9. Bonne-Tamir B, Ashbel S, & Modai J (1977) Genetic markers in Libyan Jews. *Hum Genet* 37(3):319-328.
10. Carmelli D & Cavalli-Sforza LL (1979) The genetic origin of the Jews: a multivariate approach. *Hum Biol* 51(1):41-61.
11. Karlin S, Kenett R, & Bonne-Tamir B (1979) Analysis of biochemical genetic data on Jewish populations: II. Results and interpretations of heterogeneity indices and distance measures with respect to standards. *Am J Hum Genet* 31(3):341-365.
12. Kobyliansky E, Micle S, Goldschmidt-Nathan M, Arensburg B, & Nathan H (1982) Jewish populations of the world: genetic likeness and differences. *Ann Hum Biol* 9(1):1-34.
13. Livshits G, Sokal RR, & Kobyliansky E (1991) Genetic affinities of Jewish populations. *Am J Hum Genet* 49(1):131-146.
14. Behar DM*, et al.* (2010) The genome-wide structure of the Jewish people. *Nature* 466(7303):238-242.
15. Rosenberg NA*, et al.* (2001) Distinctive genetic signatures in the Libyan Jews. *Proc Natl Acad Sci U S A* 98(3):858-863.
16. Atzmon G*, et al.* (2010) Abraham's children in the genome era: major Jewish diaspora populations comprise distinct genetic clusters with shared Middle Eastern Ancestry. *Am J Hum Genet* 86(6):850-859.
17. Henn BM*, et al.* (2012) Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. *PLoS genetics* 8(1):e1002397.
18. Wright S (1931) Evolution in Mendelian Populations. *Genetics* 16(2):97-159.
19. Rodriguez-Ezpeleta N*, et al.* (2010) High-density SNP genotyping detects homogeneity of Spanish and French Basques, and confirms their genomic distinctiveness from other European populations. *Hum Genet* 128(1):113-117.
20. Gubbay L & Levy A (1992) *The Sephardim: Their Glorious Tradition from the Babylonian Exile to the Present Day.* (Carnell Limited, London).
21. Oddoux C*, et al.* (1999) Mendelian diseases among Roman Jews: implications for the origins of disease alleles. *J Clin Endocrinol Metab* 84(12):4405-4409.
22. Kenny EE*, et al.* (2012) A genome-wide scan of Ashkenazi Jewish Crohn's disease suggests novel susceptibility loci. *PLoS genetics* 8(3):e1002559.

23. Velez C*, et al.* (2011) The impact of Converso Jews on the genomes of modern Latin Americans. *Hum Genet*.
24. Rosenberg NA*, et al.* (2002) Genetic structure of human populations. *Science* 298(5602):2381-2385.
25. Patterson N, Price AL, & Reich D (2006) Population structure and eigenanalysis. *PLoS genetics* 2(12):e190.
26. Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular ecology resources* 8(1):103-106.
27. Felsenstein J (1989) PHYLIP - phylogeny inference package (version 3.2). *Cladistics* 5:164-166.
28. Raymond M & Rousset F (1995) An exact test for population differentiation. *Evolution* 49(6):1280-1283.
29. Falush D, Stephens M, & Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164(4):1567-1587.
30. Pritchard JK, Stephens M, & Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155(2):945-959.
31. Jakobsson M & Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23(14):1801-1806.
32. Gusev A*, et al.* (2009) Whole population, genome-wide mapping of hidden relatedness. *Genome research* 19(2):318-326.
33. Browning SR & Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* 81(5):1084-1097.
34. Bonnen PE*, et al.* (2010) European admixture on the Micronesian island of Kosrae: lessons from complete genetic information. *Eur J Hum Genet* 18(3):309-316.

Figure Legends

Figure 1. Principal component analysis of Jewish populations combined with other HGDP groups in global (A) and regional contexts (B, C). Dense regions with many overlapping populations are circled for the purpose of illustration, with a list of the groups adjacent.

Figure 2. Neighbor-joining tree showing the relationship of European, Jewish, Middle Eastern, and North African populations, using F_{ST} as the distance metric. The tree was rooted using the reference mixed Central and Southern African population as an out-group. Major population groups are labeled at the right, with the red bar within the Jewish group denoting North African Jews. 500 bootstrap iterations were tested to assess the robustness of the tree (the labels at the nodes represent the number of iterations (%) in which that configuration was seen). Populations labeled with an asterisk (*) cluster outside of their expected groups.

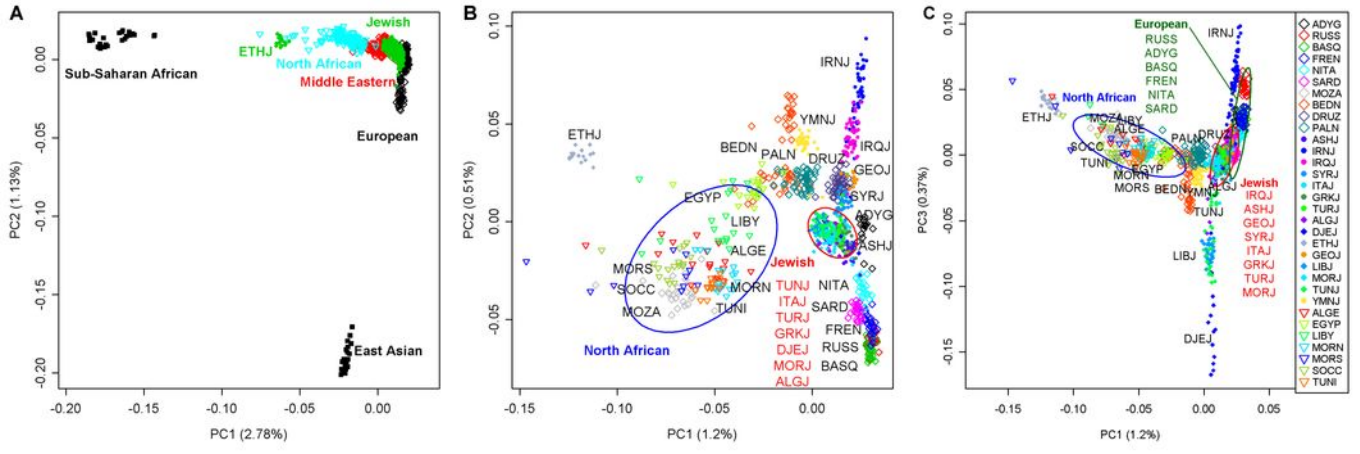
Figure 3. Genome-wide IBD sharing for the average pair of individuals within (A) and across populations (B, C). With the exception of non-Jewish Tunisian samples, IBD sharing is higher within Jewish groups, reflecting higher levels of endogamy. Jewish populations exhibit higher sharing with other Jewish populations than with geographically near groups, The average total sharing across Jewish populations is generally higher than the sharing across other population pairs, and pairs of North African Jewish populations (dark to red color bars) share more segments IBD than most other Jewish pairs.

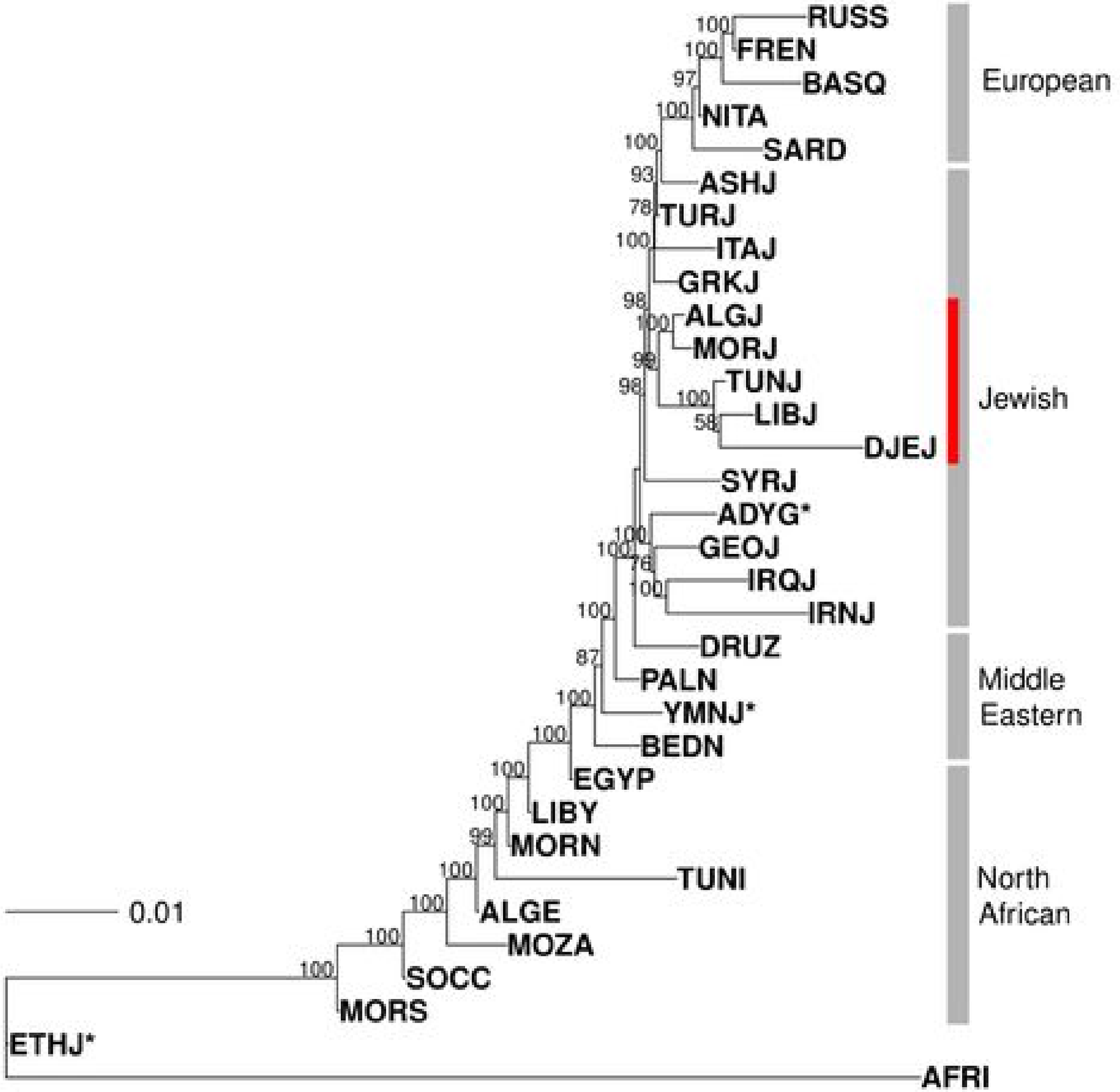
Figure 4. STRUCTURE results for Jewish populations combined with Middle Eastern, European, East Asian and African populations from HGDP. K values of 3 through 7 are shown; each represents an alignment of 10 independent runs. Vertical bars represent individuals, which are grouped by their known populations, and further combined into general regional groups, illustrated by the top bar.

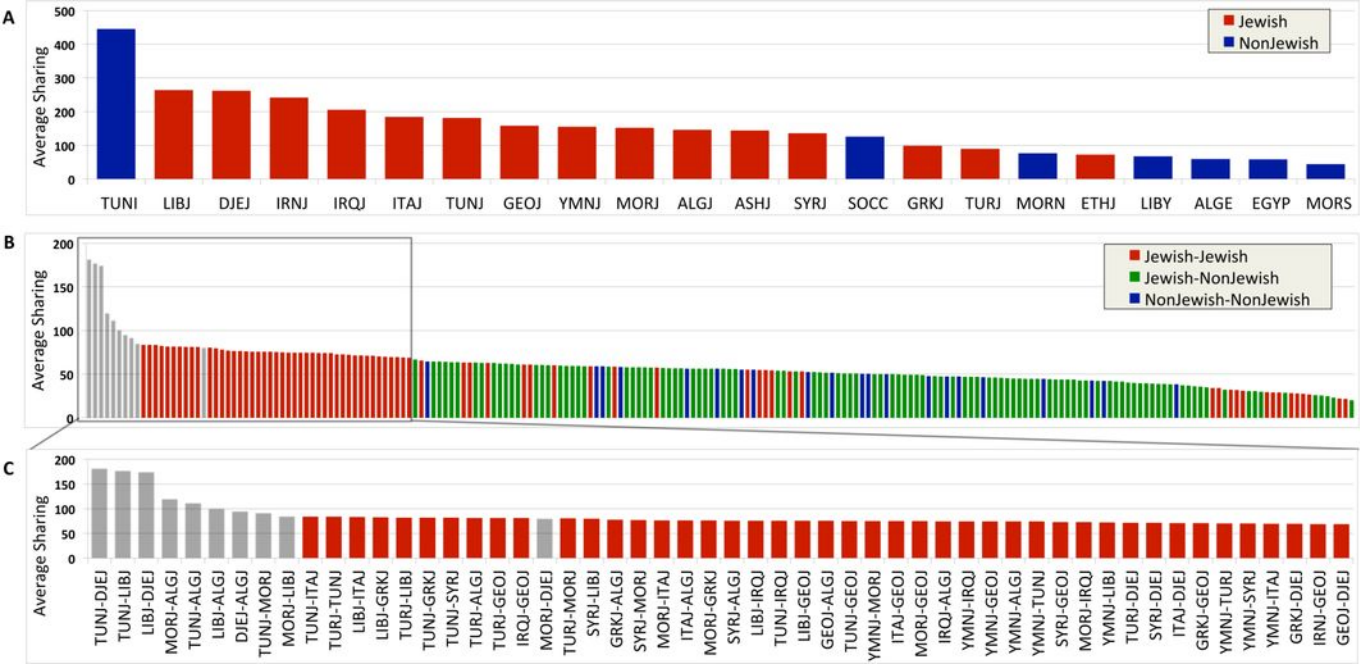
Figure 5. Ancestry deconvolution. The genome-wide ancestry of North African Jewish and non-Jewish populations is compared with respect to European (Basque), Maghrebi (Tunisian non-Jewish) and Middle Eastern (Palestinian) origins. Jewish populations exhibit increased European and decreased Maghrebi ancestry compared with corresponding non-Jewish groups. The Middle Eastern component is comparable across all groups.

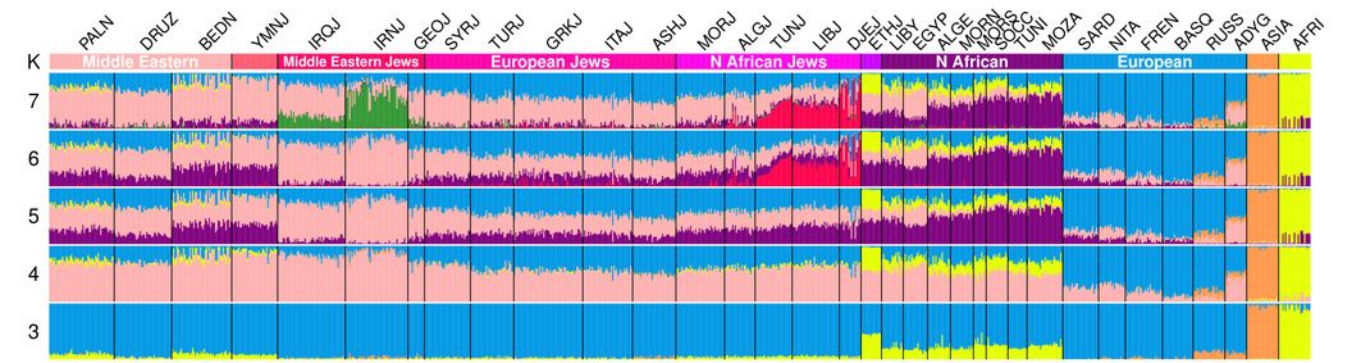
Table Legends

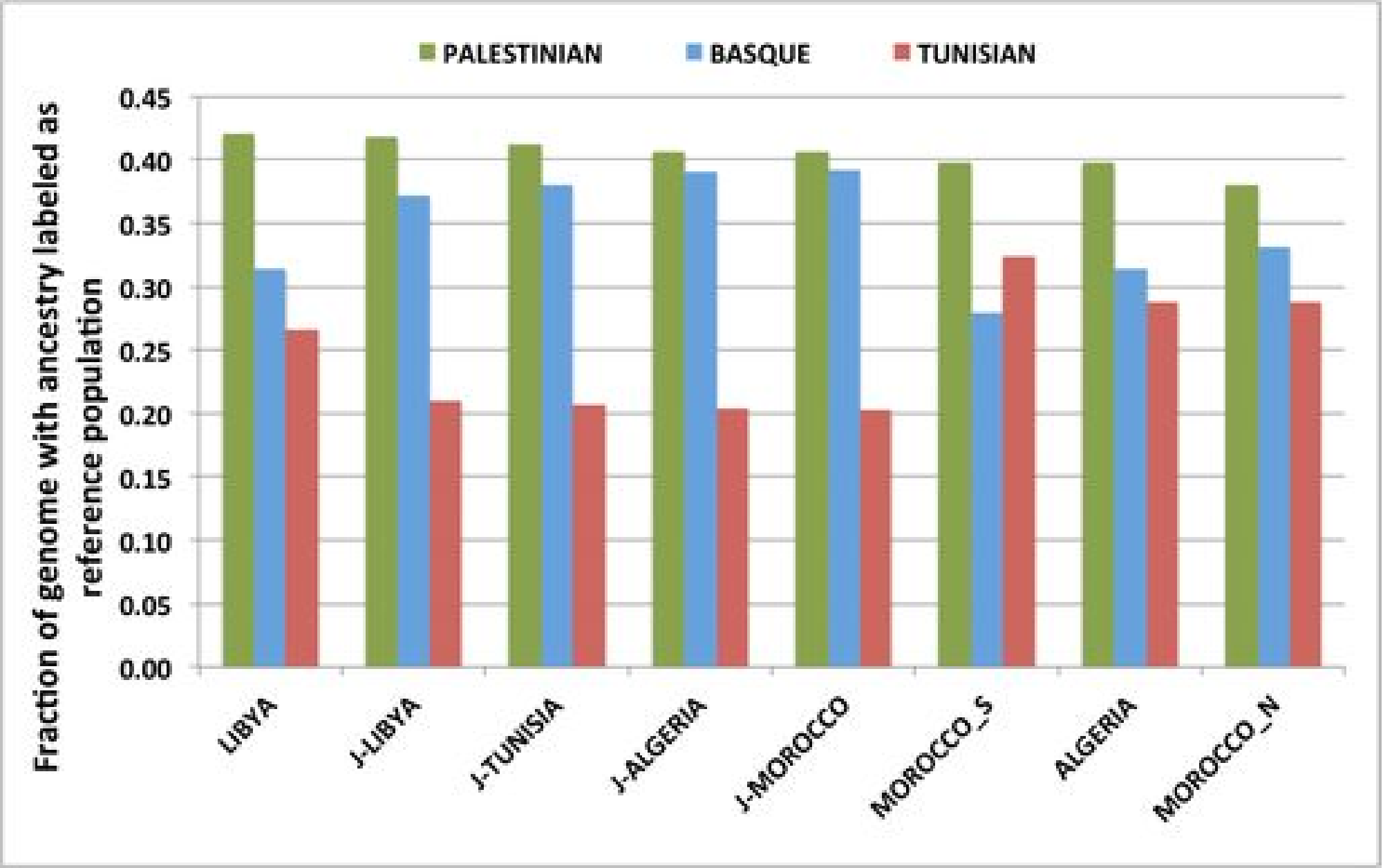
Table 1. Summary of populations included in this study. An asterisk (*) next to the population name indicates samples that were genotyped and reported in Atzmon, et al. (16).


A
Jewish
ETHJ
North African
Middle Eastern
Sub-Saharan African
European
East Asian
PC2 (1.13%)
PC1 (2.78%)
B
IRNJ
YMNJ
ETHJ
BEDN
PALN
DRUZ
IRQJ
GEOJ
SYRJ
ADYG
ASHJ
EGYP
LIBY
ALGE
MORS
SOCC
MOZA
MORN
TUNI
Jewish
TUNJ
NITA
ITAJ
SARD
TURJ
FREN
GRKJ
RUSS
DJEJ
BASQ
MORJ
ALGJ
North African
PC2 (0.51%)
PC1 (1.2%)
C
European
RUSS
ADYG
BASQ
FREN
NITA
SARD
IRNJ
North African
ETHJ
MOZA
LIBY
ALGE
PALN
DRUZ
SOCC
TUNI
EGYP
MORN
MORS
BEDN
YMNJ
GEOJ
Jewish
IRQJ
ASHJ
GEOJ
SYRJ
ITAJ
GRKJ
TURJ
MORJ
TUNJ
LIBJ
DJEJ
ADYG
RUSS
BASQ
FREN
NITA
SARD
MOZA
BEDN
DRUZ
PALN
ASHJ
IRNJ
IRQJ
SYRJ
ITAJ
GRKJ
TURJ
ALGJ
DJEJ
ETHJ
GEOJ
LIBJ
MORJ
TUNJ
YMNJ
ALGE
EGYP
LIBY
MORN
MORS
SOCC
TUNI
PC3 (0.37%)
PC1 (1.2%)


RUSS
FREN
BASQ
NITA
SARD
ASHJ
TURJ
ITAJ
GRKJ
ALGJ
MORJ
TUNJ
LIBJ
DJEJ
SYRJ
ADYG*
GEOJ
IRQJ
IRNJ
DRUZ
PALN
YMNJ*
BEDN
EGYP
LIBY
MORN
TUNI
ALGE
MOZA
SOCC
MORS
ETHJ*
AFRI
0.01
European
Jewish
Middle Eastern
North African


A
Average Sharing
500
400
300
200
100
0
Jewish
NonJewish
TUNI
LIBJ
DJEJ
IRNJ
IRQJ
ITAJ
TUNJ
GEOJ
YMNJ
MORJ
ALGJ
ASHJ
SYRJ
SOCC
GRKJ
TURJ
MORN
ETHJ
LIBY
ALGE
EGYP
MORS
B
Average Sharing
200
150
100
50
0
Jewish-Jewish
Jewish-NonJewish
NonJewish-NonJewish
C
Average Sharing
200
150
100
50
0
TUNJ-DJEJ
TUNJ-LIBJ
LIBJ-DJEJ
MORJ-ALGJ
TUNJ-ALGJ
LIBJ-ALGJ
DJEJ-ALGJ
TUNJ-MORJ
MORJ-LIBJ
TUNJ-ITAJ
TURJ-TUNJ
LIBJ-ITAJ
LIBJ-GRKJ
TURJ-LIBJ
TUNJ-GRKJ
TUNJ-SYRJ
TURJ-ALGJ
TURJ-GEOJ
IRQJ-GEOJ
MORJ-DJEJ
TURJ-MORJ
SYRJ-LIBJ
GRKJ-ALGJ
SYRJ-MORJ
MORJ-ITAJ
ITAJ-ALGJ
MORJ-GRKJ
SYRJ-ALGJ
LIBJ-IRQJ
TUNJ-IRQJ
LIBJ-GEOJ
GEOJ-ALGJ
TUNJ-GEOJ
YMNJ-MORJ
ITAJ-GEOJ
MORJ-GEOJ
IRQJ-ALGJ
YMNJ-IRQJ
YMNJ-GEOJ
YMNJ-ALGJ
YMNJ-TUNJ
SYRJ-GEOJ
MORJ-IRQJ
YMNJ-LIBJ
TURJ-DJEJ
SYRJ-DJEJ
ITAJ-DJEJ
GRKJ-GEOJ
YMNJ-TURJ
YMNJ-SYRJ
YMNJ-ITAJ
GRKJ-DJEJ
IRNJ-GEOJ
GEOJ-DJEJ


PALN
DRUZ
BEDN
YMNJ
IRQJ
IRNJ
GEOJ
SYRJ
TURJ
GRKJ
ITAJ
ASHJ
MORJ
ALGJ
TUNJ
LIBJ
DJEJ
ETHJ
LIBY
EGYP
ALGE
MORN
MORS
TUNI
MOZA
SARD
NITA
FREN
BASQ
RUSS
ADYG
ASIA
AFRI
K
Middle Eastern
Middle Eastern Jews
European Jews
N African Jews
N African
European
7
6
5
4
3


PALESTINIAN
BASQUE
TUNISIAN
Fraction of genome with ancestry labeled as reference population
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
LIBYA
J-LIBYA
J-TUNISIA
J-ALGERIA
J-MOROCCO
MOROCCO_S
ALGERIA
MOROCCO_N

Population ID	Female	Male	Total	Population
ALGJ	23	1	24	Algerian Jewish
ASHJ	14	20	34	Ashkenazi Jewish*
DJEJ	0	17	17	Djerban Jewish
ETHJ	13	3	16	Ethiopian Jewish
GEOJ	4	9	13	Georgian Jewish
GRKJ	25	29	54	Greek Jewish*
IRNJ	22	27	49	Iranian Jewish*
IRQJ	25	28	53	Iraqi Jewish*
ITAJ	20	19	39	Italian Jewish*
LIBJ	31	6	37	Libyan Jewish
MORJ	32	6	38	Moroccan Jewish
SYRJ	15	21	36	Syrian Jewish*
TUNJ	24	5	29	Tunisian Jewish
TURJ	24	10	34	Turkish Jewish*
YMNJ	36	0	36	Yemini Jewish
ADYG	10	7	17	Adygei
ALGE	9	9	18	Algerian
BASQ	8	16	24	Basque
BEDN	20	27	47	Bedouin
DRUZ	32	13	45	Druze
EGYP	0	19	19	Egyptian
FREN	17	12	29	French
LIBY	1	16	17	Libyan
MORN	0	18	18	N Moroccan
MORS	5	5	10	S Moroccan
MOZA	9	19	28	Mozabite
NITA	7	14	21	N Italian
PALN	34	17	51	Palestinian
RUSS	9	16	25	Russian
SARD	12	16	28	Sardinian
SOCC	0	17	17	Saharan
TUNI	0	15	15	Tunisian
AFRI	1	24	25	Sub-Saharan African
ASIA	10	15	25	Asian