The genetic history of the Israelite nation

Robert W. Carter

In an earlier paper, I detailed the many intermarriages that are documented between the Jews in the Old Testament and the people groups around them. It was clear from this analysis that they began as a mixed people group and continued to mix with outsiders through the end of New Testament times. This has profound implications on the genetic patterns we should expect to be found among them today. Here, I outline multiple relevant lines of genetic evidence that demonstrate this mixing. Most of this genetic information is for events more recent than many of the episodes in the biblical narrative, but they illustrate the nature of populations that live in proximity to other people groups. It is clear that modern-day Judaism carries a core set of ancient lineages tracing back to the Middle East, plus a large fraction of newer genes brought in by more recent mixing. Is this core set of genes enough to conclude they indeed trace back to biblical times? Indeed so. This exercise should inform us about how ‘races’ form and what racial differences should exist among the people living on Earth today.

The Bible gives us a clear and detailed history of the Jewish nation. In fact, the biblical text may give us a more detailed ethnography for the Jews than we have for any other ancient nation. Detailed genealogical tables exist, and an unbroken history from Adam to Abraham and beyond is given to us with little to no ambiguity. But given that contemporary Jews descend from the biblical patriarch Jacob, how much of their genetics can be traced back to him? With thousands of years of history that includes migration, destruction, restoration, a ‘Diaspora’ that has never ended, and much intermarriage with non-Jews, what should the state of their genetics be today?

Ancient Jewish communities (most of which are still in existence) were founded in Africa (Algeria, Libya, Morocco, Tunisia, Ethiopia, SE Africa), Asia (the Levant, Asia Minor, Yemen, Baghdad, Iran, India, Burma), Europe (Spain, Italy, Central Europe), and possibly other places. Some of these are more ancient (e.g. Yemen) than others (e.g. the Ashkenazi). Some were also founded by fewer people (e.g. Ethiopia). But all of them claim descent from the Jews of ancient Israel. Some of these groups grew large, but others remained quite small, causing significant levels of inbreeding over time. For example, the NW African colonies may have been settled along Phoenician trade routes. They managed to hold on during many centuries of conflict, but never managed to grow very large. Thus, high rates of endogamy exist in those small pockets. They did receive an influx of Sephardic Jews after Ferdinand and Isabella expelled the Jews from Spain in 1492, but even today one can see evidence for the two different groups in the genomes of those people.¹

Modern genetics has advanced considerably over the past decade. With data from hundreds of thousands of individuals now available in public databases, we can draw historical inferences that would have been impossible just a short time ago. And statistical tools have been developed to take account of the overwhelming amount of data coming out of our sequencing machines. These tools allow us to see historical events in unprecedented detail. To that end, it is now possible to trace the ancestry of different sections of an individual’s genome. There are some geographic areas where discriminating between two countries is difficult due to high levels of historical connectivity (France vs Germany, for example), but there are other locations where ancestry is obvious due to the presence of unique and informative groups of genetic variants (among people living in specific valleys in the Swiss Alps, for example) and continental-scale ancestry is easy to determine across most of the genome.

Thus, it should be possible to determine the ancestral source population(s) of modern Jews. This has nothing to do with evolution, and the deep-time assumption at the base of all evolutionary storytelling does not generally apply. Ancestry is a statistical question: how much and what parts of a person’s genome can be localized geographically, based on what we know about the distribution of his or her genetic variants in the contemporary human population? In short, it is now possible to definitively locate Y chromosomes, mitochondrial DNA lineages, and different stretches of autosomal DNA to specific geographic regions, statistically. This did not have to be true. We did not know it would be possible until after the data started to come in. But now that there are thousands of complete human genomes available, the case for geographical ancestry has become quite strong.

Multiple questions arise when considering the Jewish people. Are there ‘Jewish’ genes? Are the Jews a genetically distinct population? Can one trace the dispersed Jewish people back to their homeland in Israel? Is there any evidence that the 12 tribes of Israel came from a single family, with a single patriarch? Questions like these are fascinating, and can
now be answered, but with multiple caveats. To get at these answers, we must first ask what we should expect to find based on biblical history. We must then look at the genetic data and assess post-biblical history, which might include confounding factors like multiple outbreeding events with non-Jewish people, high levels of local endogamy leading to strong genetic drift, and founder effects. Even though the picture is complex, we would hope that these multiple sources of information will tell us the same thing.

Y chromosomes

A direct paternal lineage from the patriarch Jacob is generally thought to define Jewishness. Yet, early analyses with limited data concluded that more than one Y chromosome haplogroup exists among modern Jews. Does this contradict the biblical narrative? Actually, no. When comparing the Y chromosome haplotypes found in Jewish and non-Jewish Middle Easterners, it is clear that they share a common pool of sequences. The two together are distinct from non-Jewish groups in Europe, North Africa, and South Africa. But the majority of non-Middle Eastern Jews also trace back to a Middle Eastern Y chromosome source, meaning all three groups (Middle Eastern Jews, Middle Eastern non-Jews, and European Jews) come from the same paternal stock. Hammer et al. (2000) conclude that a major portion of the diversity of Jewish Y chromosomes:

"... traces to a common Middle Eastern source population several thousand years ago. The implication is that this source population included a large number of distinct paternal and maternal lineages, reflecting genetic variation established in the Middle East at that time. In turn, this source diversity has been maintained within Jewish communities, despite numerous migrations during the Diaspora and long-term residence as isolated subpopulations in numerous geographic locations outside of the Middle East [emphasis added]."

The multiplicity of distinct ancestral paternal lineages is due to the extensive mixing between Jews and non-Jews throughout their history plus the possibility of founder effects at different places in Jewish history. In other words, it is possible that a ‘non-Jewish’ Y chromosome is found at higher-than-expected frequencies in specific Jewish populations.

Figure 1. An unrooted neighbour-joining phylogenetic tree I created from the Y chromosome data in the Simons Genome Diversity Project. The majority of Jewish men belong to haplogroup J (there is no intended association between the ‘J’ and ‘Jewish’), but Jewish men also belong to many different haplogroups, including R, E, and others. Tree drawn using MEGA 7.0. The scale bar represents approximately 700 mutations.
today simply because it happened to be represented in one of the men who founded a small colony.

Among the Sephardic Jews, Y chromosomes from a diverse background can be found. The majority come from haplogroup J (strongly associated with Middle Eastern peoples). But others are from E1b1a and E1b1b (common in Africa and other places), R1a (up to 30% in Ashkenazi men), R1b (the most common lineage in Europe), Q (Asia), I (Europe, but rare), and G (mainly Western Asia).6 The distribution of haplogroups found among the Spanish Sephardim was similar to a Jewish population in Turkey (sometimes included within the term “Sephardim”). This pattern also held for Jewish men from NE Portugal. Nearly one third of the Y chromosomes were European (R1b, common in Eastern Europe), and over one half were ‘Middle Eastern’ (37% J and 16% T).7 Note that the ‘Middle Eastern’ chromosomes come from diverse lineages. Thus, all Jewish men do not all trace back to ‘Jacob/Israel’ and which lineage might represent him is a matter of debate. See figure 1 for an explanation of these groups and how they are related to one another. There is some debate among biblical creationists as to where the ‘root’ of the tree should be placed, but all of the people with whom I have discussed the subject, to date, agree that ‘Y chromosome Noah’ should be somewhere near the center of the starburst pattern. The evolutionary root would be located about halfway along the A1 branch. But root placement has nothing to do with the relative relationships among the groups.

The claims above were made on limited data (e.g. the studies often used only a few tandem repeats that have a high rate of back-mutation), not fully sequenced chromosomes.10 This type of data is only useful for analyzing recent events. Re-examining such data caused Tofanelli et al. (2014) to conclude that, while one might be able to draw conclusions about the founding of various Jewish groups, the fact that multiple diverse lineages exist in these populations and that these lineages exist in non-Jewish groups as well “highlights the lack of support for using them either as markers of Jewish ancestry or Biblical tales”.11

Yet, these comments were premature. Further analysis of fully sequenced Y chromosomes indicates that the Ashkenazi individuals who carry the ‘European’ Y chromosomes of haplogroup R are not, in fact, related to the European branches in the Y chromosome family tree.12 Even though this family might be common in Europe, the Jewish versions were found to branch off earlier and to be more closely related to Middle Eastern men who carry haplogroup R. The clear implication is that this is an old, pre-Diaspora, possibly non-Jacob-descent Y chromosome family which is now found within the greater Jewish population. The majority of lines really do go back to a Middle Eastern setting. But would we expect them to go back to a single person? In fact, no. We would not, based on the known history of the nation of Israel.2

Aaron’s Y Chromosome?

A surprising percentage of men across Judaism who claim to belong to the Levitical priesthood (the Cohanim) share a common Y chromosome within a subgroup of haplogroup J.13 This is despite the fact that they have been separated in some cases for approximately 2,500 years. This Y chromosome type has been named the Cohen Modal Haplotype (CMH), and men who carry it are more closely related to each other than they are to the other Jews in the communities in which they live. The CMH is rare in non-Levitical Jews, is found in 50% of Levites, and occurs in a much higher proportion of the Cohanim.14 Hammer et al. (2009) discovered that the Cohanim carry many different Y chromosome types, but most are at low frequencies (figure 2).15

This leads us to a fascinating anecdote that combines history, genetics, and oral traditions. The Lemba are a tribe from SE Africa with distinct customs and language. Their oral tradition claims they are Jewish, descendants of men involved in long-distance trade. But they are Bantu speakers and are surrounded by Bantu-speaking peoples. Analysis of their Y chromosomes revealed that more than half are of Middle Eastern origin.16 Mitochondrial DNA, however,

![Figure 2. The frequencies of various Y chromosome haplogroups among the Ashkenazi, Cohanim vs non-Cohanim. The number and letter combination after the main haplotype designation (e.g. J-M578) references the name of a characteristic mutation that helps to define that lineage. These naming conventions have changed much over the years. (From Hammer et al.13).](image-url)
reveals no evidence of West Asian roots. This is consistent with the tradition that they were descended from Jewish traders coming down the coast from the north. But the history is further confirmed by the fact that the CMH is found within a certain subset of the Lemba. So not only did a group of Jewish men become marooned in southernmost Africa, and not only did they marry local women and settle down thousands of miles south of the equator and thousands of miles more from their ancestral homeland, but it appears that a descendent of Aaron was among them.

Yet the CMH has been disputed of late. Specifically, since it is also found in non-Jewish populations, some argue that the occurrence of the CMH is nothing more than a statistical anomaly. Of course, having Jewish Y chromosomes spread far and wide is expected based on biblical history. But the original studies were also performed on only a few Y chromosome markers. When the number of markers was increased, it was learned that the Cohanim carry multiple unrelated Y chromosome types, several of which predominate. Due to the random nature of paternity in small populations, it is possible that the original was lost or that an introgression occurred, and a subsequent random fluctuation brought a new lineage to the fore.

This may have been the case among the living Samaritans of Israel. Their genealogy is restricted to four main families, but the Cohen family that claims the Samaritan priesthood carries haplotype E. In contrast, the other main Samaritan Y chromosomes belonged to haplogroup J. Since it is conceivably possible that something similar could have occurred among the Jews at some point in history, it is not possible to definitively state that the CMH is Aaron’s Y chromosome. However, it is still likely, especially when one considers the fact that the men who carry it trace their tradition of being among the Cohanim to a time prior to the Diaspora, which began before 500 BC.

Mitochondrial DNA

Among European Jews, Y chromosomes and mtDNA yield different patterns. More than one half of the mtDNA of Ashkenazi Jews can be traced back to three unrelated women from the same haplogroup K (figure 3). This represents a major population expansion and a significant founder event. These mitochondrial lines are almost non-existent in non-Jewish populations. But note that nearly half of the mitochondrial lineages do not belong to one of these three matriarchs. Just under one half belong to haplogroup H, which is more associated with Europeans.

More detailed analyses, however, showed that the common H and K lineages (~40% of the Ashkenazim) as well as most of the minor lineages cluster deeply among non-Jewish Europeans. This would seem to indicate a significant

Figure 3. An unrooted neighbour-joining phylogenetic tree created from the mitochondrial chromosome data in the 1,000 Genomes Project. Common mitochondrial lineages found among modern Jewish females include haplogroups K and H. The scale bar represents approximately 14 mutations.
number of the founding women among the Ashkenazi were converts. However, this does not support the contested hypothesis that the Ashkenazi came from a mass conversion of people in the Khazar Empire of the Asian steppes (see below), for the common mitochondrial lineages cluster with European lineages, to the exclusion of the others.

Multiple Jewish populations seem to stem from independent founder events with a limited number of female lineages that have not increased much in diversity since their founding. In those same communities, Y chromosomes show much more diversity. This is counter-intuitive on some levels because one might expect outsider females to be over-represented in the ancestral Jewish population over outsider males. Yet, on the one hand, this might explain why so many Jewish mitochondrial sequences trace back to European founders. On the other hand, it clearly shows that paternal ancestry has been important to the Jewish people for thousands of years, as the Bible indicates, despite the fact that ‘Jewishness’ has been traced through the mother since the 2nd century.

**Genome-wide comparisons**

More detailed data sets have confirmed the conclusions of earlier studies, showing us that there exists among modern Jews a core set of genes and haplogroups that trace directly back to the Middle East. Also, significant sections of IBD (identity by descent) and other measures indicate that the various Jewish populations are more closely related to each other than they are to outsiders, despite the fact that some of these groups have been separated from one another for millennia.

When comparing the various Jewish groups to one another and to non-Jewish outgroups, interesting patterns emerge. For instance, since they started from a small group of Jewish settlers and interbred extensively with non-Jewish outsiders, the Ethiopian Jews unsurprisingly cluster with Ethiopians in genetic comparisons. But they cluster with the Semitic-speaking Ethiopians, not Cushitic-speaking Ethiopians. The Jews of Yemen, another old and isolated population, are more closely related to Bedouins and Saudis than they are to the Yemenis. Indian Jews, like Ethiopian Jews, are closely related to their host population, indicating significant mixing over the years.

But mixing can also be detected on larger scales. The following is based on evolutionary rates and dates, but it is an interesting conceptual framework that can help us to understand large-scale historic processes. The average southern European has 1–3% African ancestry, most of which probably traces back to the Arab migrations after the fall of the Roman Empire. People from the regions surrounding Israel carry 4–15% African ancestry, but these links probably trace back to more recent times (maybe 1,000 years ago).

Multiple Jewish subpopulations also carry African DNA (3–5%), with an estimated mixing date of 70 generations ago. This is much older mixing than the other groups, going back more than 2,000 years. This means the African ancestry was almost certainly among them prior to the Diaspora. Yet even this is much later than the biblical date for the founding of Israel (it is only about halfway to Abraham). Genetic evidence being what it is, all we can say is that the Jews in late Old Testament times already had a small proportion of African ancestry. This may not turn up in Y chromosome or mitochondrial studies due to genetic drift (where rare lineages tend to be lost) and founder events, but it has carried over to the rest of the genome.

**Disease**

Due to high levels of inbreeding in their close-knit communities, Jewish groups carry a surprising number of otherwise rare, recessive genetic diseases. For example, a major variant of the Creutzfeldt–Jakob disease has been associated with an origin in Spain and subsequent spreading caused by the 1492 expulsion of the Jews from Spain. Bloom Syndrome is another rare, autosomal, recessive disorder most common in Ashkenazi Jews. But it is also found among Spanish-speaking individuals in Central America. Most likely, the individuals living in the Americas that are carriers had a Sephardic Jewish ancestor from Spain (confirmed by ancestry in several families). Disease studies are fascinating in their own right, but since there are specific genetic anomalies associated with Jewish populations, this tells us that outbreeding with non-Jews was always limited.

**Alternative theories**

The Khazar hypothesis

Once Need et al. discovered that one can statistically separate people with Jewish ancestry from a general population of Europeans, the race to discover ‘Jewish genes’ was on. Even though there are clear patterns, the idea that there are exclusive Jewish genes was recently contradicted by Elhaik. He makes six major points, all of which are questionable. I will restate them here:

1. It is unlikely that an ‘Israelite gene’ ever existed since Iron Age Israelite tribes exchanged genes with their neighbouring tribes.
2. Had there been a paternal ‘Israelite gene’ on the Y chromosome, it would have been lost due to the transition to matrilineal descent.
3. Had there been a maternal ‘Israelite gene’ on the mitochondrial chromosome, it would have been lost due to the initial period of patrilineal descent.
4. Had there been an autosomal ‘Israelite gene’, it would have been lost due to the high rates of movements into the religion.
5. Had there been an autosomal ‘Israelite gene’, it would not be unique to Jews due to the high rates of movements out of the religion.
6. Had there been an autosomal ‘Israelite gene’ that survived to modern days, it would have been extremely rare and undetectable by popular search approaches that prioritize findings common to a large fraction of Jews.

Points 1–4 depend on the degree of mixing with other populations, plus chance. Being that the Jews have always had a certain degree of separation from other cultures, the assumption should be that they may have maintained some genetic distinction. Point 5 is a given and is well attested to in the literature. This is one of the main reasons why there are no ‘Jewish’ genes. But this still does not mean there are no genetic Jews. We may not be able to positively identify a Jew by their genes, but when one does find genes common to Jews found throughout their current distribution, this is a strong case for the antiquity of Judaism. His final point 6 depends on the status of his Khazar hypothesis.

Elhaik has attempted to claim Eastern European Jews were descended from the Turkic Khazars of central West Asia. This seems to be a resurrection of the controversial ideas of Koestler. But this theory was quickly shot down by others. Specifically, it seems that his team’s use of the Georgian Jews as a proxy for the assumed genotype of someone from the now-vanished kingdom of Khazar was a mistake, for the Georgians are not only south of the Caucasus, but they cluster tightly with the Middle Eastern Jews in genetic analyses. Instead of being evidence for a mass conversion of Jews in the religion.

Conclusions

There is no ‘Jewish gene’, although there are certainly genes common to Jews. And because of the possibility of genetic drift, we cannot be 100% certain that things like the CMH are actually the Y chromosome of Aaron. However, Jews today absolutely fit both the biblical expectations and their oral and written history since the completion of the Old Testament canon. From the detailed history of the Jewish nation preserved in the biblical narrative, it should be clear that they started as a mixed population, maintained a degree of mixing with their neighbours, and continue to mix with outsiders today. However, as a Middle Eastern tribal community, they should have Middle Eastern genetic roots, and the evidence tells us they certainly do.

What does this tell us about races in general? Slatkin and Racimo (2016) say it this way:

“We now know that present-day populations were created by a complex history of admixture and population movement. Although local genetic continuity over long periods has been documented in a few cases, these are exceptional. The general rule is that the ancestors of present-day populations lived somewhere else.”

So now we know that human history is complex and sometimes convoluted. The various people groups on this planet all have the tell-tale signs that their ancestors mixed with other people. The Jewish people, due to their rich historical background, serve as a test case that informs us what we should expect to find in other groups. Essentially every people group should represent a history of mixing that unites us all under one common banner: descendants of Adam and relatives of Jesus Christ, the Kinsman-Redeemer (Isaiah 59:20).
Acknowledgements

I thank Jonathan Sarfati for his critical review of an earlier draft of this manuscript as well as the efforts of two anonymous reviewers.

References


6. Adams, S.M. et al., The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula, Amer. J. Hum. Genet. 83:725–736, 2008. Note that the haplotype names reported in this paper (e.g. E3a and E3b) differ from those I reported in the text (in this case, E1b1a and E1b1b). Revisions in haplotype names are constantly occurring, so I strove for consistency. I chose not to use a recent convention of naming haplogroups based on lineage-specific mutations (e.g. E-P147, which defines haplogroup E1) because it would make it too difficult for the reader to cross-check the references and discussion.


13. Thomas, M.G. et al., Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors, Amer. J. Hum. Genet. 70:1411–1420, 2002.


37. For example, see Das, R. et al., Responding to an enquiry concerning the Geographic Population Structure (GPS) approach and the origin of Ashkenazic Jews—a reply to Flegontov et al., dx.doi.org/10.1093/genetics/axv026.


Robert Carter received his Bachelor of Science in Applied Biology from the Georgia Institute of Technology in 1992 and his Ph.D. in Coral Reef Ecology from the University of Miami in 2003. He has studied the genetics of fluorescent proteins in corals and sea anemones and holds one patent on a particular fluorescent protein gene. His current research involves looking for genetic patterns in the human genome and the development of a biblical model of human genetic history. He works as a speaker and scientist at CMI-US.