

Does Mitochondrial DNA T2e1 Reflect Traces of Judaism on the “Down-Low”?

Felice L Bedford*

University of Arizona, PO Box 210068, Tucson, AZ 85721, USA

*Corresponding author: Felice L Bedford, University of Arizona, PO Box 210068, Tucson, AZ 85721, USA, E-mail: bedford@u.arizona.edu

Received date: October 23, 2018; Accepted date: November 07, 2018; Published date: November 15, 2018

Copyright: ©2018 Bedford FL. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

The genetic origins of European Jewish women remain controversial and new insights may be obtained through the less studied “other European Jews”, the Sephardim. The branch *T2e1* of the mitochondrial DNA phylogenetic tree contains two Sephardic branches distant from one another. To investigate further, the largest analysis of *T2e1* to date is presented with 40 new complete *T2e1* mitochondrial DNA sequences (mitogenomes) from European-descended individuals, doubling the number available. Additionally, nearly 8000 control region sequences from 37 studies were re-considered for *T2e1* specifically, from areas high in *T2e* (Near East, Italy) and areas low in *T2e* but Sephardic related (Spain, Portugal). These revealed unexpectedly an absence of *T2e1* in the Near East and a high number in Spain. By contrast, numerous *T2e*-Other sequences were in the Near East and Italy with considerable overlap. The mitogenomes elucidated new branches of *T2e1* at varying distances from the Sephardic branches including the closest, an Irish-specific clade (4107T), the most populous (nonsynonymous 14180C with unknown biological effect) from Norway (also Sweden, Germany, British Isles), and the root node of *T2e1a* (2308G). The findings suggest an origin or early arrival of *T2e1* in Europe whilst for *T2e*, strengthens the role of Near East and its early free travel with the Mediterranean. Quiet Judaism (*down-low*) may have produced *T2e1* branches where Jewish identity was lost, such as from Sephardics who took the Northern route after the Iberian expulsion. The Scandinavian branches may reflect Nordic conquest of British Isles and solve a mystery of a *T2e1* hot spot in Iceland. The Near East-Mediterranean origins of *T2e* resemble that reported for Ashkenazi maternal lineages; other Sephardic haplogroups are needed to assess generality of this common origin. Additional insight may be obtained with the first Jews to enter Europe, the Romaniots, the other *other* European Jews.

Keywords Mitochondrial DNA; Population genetics; Sephardim; Jewish genetics; Haplogroup T; Nonsynonymous variants

Introduction

The origins of founding Jewish women in Europe remain controversial [1,2]. The majority of genetic studies to address this issue have investigated the Ashkenazi population. The Ashkenazim are the most recognizable “Jews of Europe” who flourished in Eastern Europe with presence in Western Europe both before and after their rapid population expansion. However, the Sephardim are also Jews of Europe. They resided for over a thousand years in Iberia until the forced expulsion at the close of the 15th century. Moreover, the Sephardic population may be more representative of European Jewish DNA because it was not subject to the severe bottleneck of the Ashkenazi population. By one dramatic estimate, all Ashkenazi individuals today can be traced to a scant 350 individuals sometime between the 13th and 16th century [3]. The estimate is based on length of shared segments in autosomal complete genomes. In contrast, no comparable bottleneck occurred amongst Sephardim, when 150,000-250,000 left Spain in 1492, transporting with them their varied genes and liturgies alike. Thus, while in the 21st century the Ashkenazim are by far the world’s largest Jewish group in numbers of people, the Sephardim may be the largest Jewish group in numbers of varied genes. Detailed consideration of Sephardic DNA may bring different insights into the origin of European Jewish women.

A promising candidate for scrutiny is a branch of the mitochondrial DNA (mtDNA) phylogenetic tree known by label “*T2e1*” within which

two notable Sephardic subclades are found. One of these subclades originated in Iberia amongst the Sephardic Jewish population. It is found in Sephardim [4-7] who emigrated from Spain following the 1492 expulsion to Ottoman Empire cities, such as Salonika (now Thessalonica, Greece) and Constantinople/Konstantiniyye (Istanbul, Turkey). The signature has more recently turned up in crypto-Jewish villages in the Bragança district of Northeast Portugal [8]. Interestingly, we have also shown that this branch of *T2e1* crossed the Atlantic in the early days of the founding of Northeast Mexican communities [5] where this otherwise hard-to-find mtDNA subclade thrives [9]. Note that this Sephardic maternal lineage is an exception to the Iberian man-local woman union believed to originate the majority of the modern Mexican population [10]. Importantly, it also provides genetic evidence for historical accounts that Iberian founders of Northeast Mexico were *converso*, forced converts from the Jewish faith in order to avoid government persecution. This branch was recognized as “*T2e1a1a*” beginning with build 16 (currently on build 17) of the standardized phylogenetic tree for mtDNA, Phylotree [11], and is located within the *2308G* branch, that is, *T2e1a*.

The second Jewish group in *T2e1* was first reported by Behar et al. as a founding lineage among Sephardim in Bulgaria [7], also an Ottoman Empire region. Thereafter, we showed it is widely distributed amongst the Ashkenazim in the Czech Republic, Poland and Lithuania as well as Sephardic individuals in regions other than Bulgaria [5]. It may well underlie nearly all of the small *T2e* presence in the Ashkenazi population. We have thus far only found the motif in Jewish groups, though its defining variant is in the coding region of mtDNA (*9181G*) which makes it impossible to identify in sequences from only the

control region, which are more numerous in the literature. Also beginning with Phylotree build 16, it became known as *T2e1b*. Jewish *T2e1b* has no nesting relation with Jewish *T2e1a1a*, other than sharing the defining variants up to and including those of *T2e1*. In fact, they may be a very large distance apart on the *T2e1* branch of the phylogenetic tree according to a pairwise comparison of mutational differences of all *T2e* haplotypes; the comparison suggested that the most recent common ancestor of the two Jewish groups may be "as old as *T2e1* itself" [5].

The existence of these two distant Jewish groups within *T2e1* raises the question: What other populations comprise *T2e1* and how do they relate to the Jewish subclades and each other? Can these provide insight into the origin and dispersal of Sephardic mtDNA?

Besides the two well-researched branches, little is known about *T2e1*. The mother clade *T2e* is found in relatively high numbers in parts of the Near East, including Saudi Arabia and northern Egypt and widely distributed throughout Europe with highest frequency within Italy [4]. By the time it reaches England, it is outnumbered twenty fold by the more common European *T2b* branch of *T2* [4], yet appears to be a hotspot in Iceland [12]. We have suggested that the origin of *T2e* is either Near Eastern or Southern European/ Mediterranean [4] whilst Pala et al. [13] concluded it was definitely European.

Previous investigations of *T2e1* have lumped it together with *T2e* overall. Consequently, we sought to clarify the constituency of specifically subclade *T2e1*, through both complete mitochondrial sequences (mitogenomes) and sequences from the control region only. For mitogenomes, we more than double the number of *T2e1* samples with 40 new sequences from individuals of European descent. For the lower resolution, but much higher number, control region only sequences, we re-evaluate the *T2e* data in the literature by keeping separate track of *T2e1* from the remainder of *T2e* primarily in regions where *T2e* is especially prevalent (Near East, Italy) and of special interest to Sephardim (Spain, Portugal).

Materials and Methods

High resolution

To obtain *T2e1* mitogenomes, the *T2* mitochondrial DNA project at Family Tree DNA (familytreedna.com/groups/t2) was searched for potential participants. The large predominately United States customer database at Family Tree DNA (FTDNA) has proved to be a valuable research resource [4,5,14-17]. Sequences from the control region of mtDNA harboring both the variant that identifies *T2e*, i.e. *16153A* in the first hypervariable segment (*HVS1*), and the variant that identifies most of the *T2e1* branch, *41T* in the second hypervariable segment (*HVS2*), were selected for further consideration. Any branches of *T2e1* with a mutation back to ancestral *41C* ("back mutation" [18]) were also sought including further instances of *T2e1a1a* [5] and potential new ones by looking for overlapping variants with known *T2e1* sequences. Individuals identified as possessing mtDNA *T2e1* who also met additional criteria of having had the entire mitochondrial DNA sequenced and not previously deposited to Genbank were invited to participate in the present research project.

Sequencing of the *16569* mtDNA base pairs and *T2* haplogroup assignment was provided by FTDNA, a division of Gene by Gene, LTD (Houston, TX, <https://www.genebygene.com/#>), using Illumina (San Diego, CA) MiSeq platform with greater than 1000X coverage on average. Early customers at FTDNA had DNA sequenced instead with

Sanger sequencing [15]. Participants' data was not considered separately by the methodology used for their sequencing. Data files in FASTA format were obtained from participants and checked for variants with MToolBox hosted on MSeqDR (<https://mseqdr.org/mtoolbox.php>). Geographic origins and date on the deep maternal line ("mother's mother's mother going back as far as you can go on just the mother's side") was obtained through self-report. To screen for sequences from the same immediate family, participants with same or similar sequences were queried further and a criterion of no common maternal relative within four generations of another participant was used to determine inclusion. To screen for very recent migrations, self-reported European locales were only listed if they exceeded 100 years before present.

Participants provided consent directly to FTDNA for their sequencing. In addition, permission was given directly to Genbank on an individual basis for anonymous public posting of their mtDNA sequence; whilst we prepared sqn files and assisted with submission, each participant with a submitted file was contacted individually by NCBI and required to consent directly.

Variants found in the coding region were checked for whether they produced any change to the protein instruction or caused a premature stop codon (nonsynonymous changes). A rooted phylogenetic tree was reconstructed from all unrelated *T2e1* mitogenomes along with our previously published *T2e1* sequences and relevant Genbank entries with the assistance of Network v. 5 by Fluxus Engineering, MtPhyl and Dendroscope.

Low resolution

To reassess the frequency and geographic concentration of *T2e* with *T2e1* considered separately from the remainder of *T2e*, *T2e* sequences primarily from the control region were inspected predominately from the Near East, Italy, Iberia, and Iceland.

For the Near East, the two studies previously reviewed [4] with high prevalence of *T2e* from Saudi Arabia and Egypt [S1,S2] were added to from six other mtDNA studies of Egypt, Iraq, Qatar Jordan, and Yemen [S3-S8] and from among the 1114 additional sequences of the Genographic study for Lebanon, Jordan, and Palestine [S9]. One study of Arabian mitogenomes was available [S10]. To our knowledge, these ten studies reflect the bulk of publicly available mtDNA sequences from these Near East regions. *T2e* sequences for Italy were mined from the 395 subjects in the North Central Italy study [S11] that was reviewed previously to report the high *T2e* frequency in Italy [4]. It was added to by eight other studies of mtDNA throughout mainland Italy [S12-S19] which included mitogenomes and two for the islands of Sicily and Sardinia [S17,S20]. For Iberia, a total of 14 studies were inspected [S21-S34]. Finally, Icelandic *T2e* modern and ancient sequences were assessed [S35-S36] in this *T2e* hotspot, as well as ancient Crete [S37] and two studies of Greece [S38-S41].

Upon determination a sequence belonged to the *T2e* branch, it was then assigned to one of three categories: *T2e1*, not *T2e1* (*T2e1*-Other), and *T2e1* status unknown. To make the assignments, sequences were checked for the presence of the defining *41T* in *HVS2* and for any *T2e1* branches with back mutations to *41C*. When a study did not test *HVS2*, tested a partial *HVS2* region that did not include position *41* (e.g. *70-390*), or was unclear, inferences as to *T2e1* status was made where possible by matching any variants present to sequences from studies where position *41* was tested.

A maximum parsimony median joining network of control regions was created for all *T2e* sequences where *T2e1* status could be established.

Results

T2e1 mitogenomes (high resolution)

A total of 1146 individuals from the *T2* project had the entire mtDNA sequenced, of which 77 were found to be from a *T2e* branch, or 6.7%. This proportion of *T2e* to total *T2* is nearly double that found in the 76,000 participant National Geographic (NG) project (3.7%) [18] (Fisher's exact test NG *vs* current dataset $p < .0001$) and may reflect greater representation of Western Europeans in the NG databases. Of the 77 *T2e* sequences, 40 were identified as *T2e1*, or roughly half (51.9%), comparable to Genbank's 44.4% (32 of 72). A maximum parsimony phylogenetic reconstruction of the *T2e1* haplotypes in the present study is presented in Figure 1. Named additions and changes to the phylogenetic tree of *T2e1* are suggested in Table 1. New branches are detailed as follows.

14180C in Scandinavia, British Isles, Germany

The most frequent new branch of *T2e1* comprised 22.5% of the sequences (nine unrelated individuals), all with variant *14180C* from the coding region. Sequences originating in Norway, Germany, Sweden, and Denmark harbored only this variant whilst four unique regional-specific subclades branched from this stem: British Isles, Germany, and two for Norway (Figure 1). What amounts to a distinct fifth subclade of the *14180C* group was deposited previously in Genbank [14], a lone sequence with geographic location unspecified (not shown). The multiple branches and nesting from these regions is notable.

Ireland in the large *2308G* (*T2e1a*) branch

Adding to the diversity of previously elaborated *T2e1a* [4,5] was a new Irish-specific clade found amongst three unrelated individuals that migrated to US, New Zealand, and Canada. It has a fascinating placement in the phylogenetic tree. The branch shares the 15499T variant with the lineages of Ottoman Sephardic-Mexican *converso* (*T2e1a1a*) and Colonial America-England-Netherlands (*T2e1a1b*). It is also notable that its nearest neighbor is a sequence from Morocco [19] (corrected version, Ana M. González Matilla, personal communication, December 14, 2011, [5]. The Irish and Moroccan sequences share *14180C* and bifurcate to *4107T* (Irish) and *736T* (Moroccan). Note *14180C* is the same variant found for the populous Northern branch described above. While it would be intriguing if they were all part of the same branch, the most parsimonious reconstruction of the phylogenetic relation is that the *14180C* variant arose twice as independent events. We suggest a possible reason for its persistence in two distinct lines in the Discussion.

2308G root

We also report the first ancestral *T2e1a* sequence; that is, harbors *2308G* without any further variants besides those leading up to its

inclusion in *T2e1a*. The maternal origins traced to the current border region of Poland and the Czech Republic.

Colonial American branch of *2308G*

One further twig was found from the Colonial American branch we reported on previously [4,5]. The sequence contains heteroplasmy (a mixture of nucleotide bases) at a position in the control region. The nearness of that branch to the Ottoman Sephardic-Mexican *converso* *T2e1a1a* branch remains notable.

Is there a *200G* clade?

Also shown in Figure 1 is a tentative branch linking a few haplotypes with *200G*. However, these may be unrelated and the stability of *200G* in the context of *T2e1* awaits clarification.

Remaining clades

A total of nine additional haplotypes were found in one or two individuals from Italy, Switzerland, Germany, Sweden, and England and one additional frequent haplotype in Norway.

Overall, there were a total of 40 mitogenomes belonging to mtDNA haplogroup *T2e1* of which 19 were unique sequences. The sequences formed a star-like distribution emanating from the central ancestral *T2e1* node that has 41T. No ancestral *T2e1* sequence was found. The closest arguably were from Sweden (*16189Y*) and the British Isles (*200G*). The 19 unique sequences had a total of 36 mutated positions, 9 from *HVS1*, 6 from *HVS2* and 21 from the coding region. Note that had only the control region been sequenced, 42% would erroneously have appeared as the ancestral sequence which is consistent with previous studies in Haplogroup T [4,14]. In the coding region, 11 of the 21 variants, approximately half, were nonsynonymous with changes in ND2, ND4, ND5, ND6, ATP6, and CO1. There were additionally 3 RNA changes.

Literature sequences (Low resolution)

Of the total 1553 mostly control region sequences from Saudi Arabia, Egypt, Yemen, Lebanon, Qatar, Iraq and Jordan, 28 were from *T2e* haplogroup. Re-examination of the sequences unexpectedly did not find any belonging to branch *T2e1*. A total of 20 were ruled out as *T2e1* based on apparent direct testing of position 41 and 6 sequences inferred not to have 41T based on matches to these and sequences from Italy, Spain, and Greece. The remaining two were ancestral (*16126T 16153A 16294T 16296T*) from Jordan which did not allow the definitive exclusion of *T2e1* but the only other Jordanian ancestral sequence (from a different study) was definitively not *T2e1*. The geographic location of ancestral *T2e* in the Levant is of interest and higher resolution testing required for clarification.

There were 10 unique haplotypes of *T2e*-Other, several found in more than one of the Arabian locales.

Haplogroup	Defining Variants	No.	Comment
Additions			
<i>T2e1c</i>	14180C	>3	Scandinavia, Germany
<i>T2e1c1</i>	200G-2887C-16525G	2	England, Ireland
<i>T2e1a1c1</i>	4017G	3	*Ireland. Closest new branch to the Sephardic signature
<i>T2e1d</i>	194T-11710T-12762T	3	Norway
<i>T2e1e</i>	9932A-5396T	1	Italy, plus 1 on Genbank since 2007 (Italy)
<i>T2e1f</i>	10586A-11167G-16296T	2	Norway
Changes			
<i>T2e1a1c</i>	14180C (16189C)		* Current criterion 16189C

Table 1: Additions and changes to *T2e1* phylogeny. Suggestions based on guidelines for recognized branches on Phylotree. Variants are with respect to the rCRS *Assumes corrected Morocco sequence with 41T-736T-2308G-14180C-15499T-16189C.

Out of 2033 total sequences from Italy (North, Central, South, Sicily), 19 were *T2e*. This tabulation does not include Croatians residing in Italy [S13], the Pala et al. study [S18] that specifically targeted JT haplogroups, Gasparre et al. [S15], also a non-random selection, nor the recent Sardinian mitogenome [S20]. Classification of the 19 *T2e* sequences was less certain than for the Near East, but there was clearly one *T2e1* sequence (a mitogenome) and the non-random studies added two *T2e1* sequences (and 13 *T2e* -Other) [S15,S18]. Classification was based on direct testing of position 41 (7 sequences), inferred direct testing from primer descriptions (5), matches to Near East (3), probable matches (3), and unclassifiable (1). Despite some classification ambiguity, Italy, in contrast to the Near East, harbors *T2e1* sequences. There were numerous unique *T2e*-Other haplotypes, including 8 in the random 19 sequences.

A total of 28 *T2e* sequences were found from Spain (23) and Portugal (5) out of 3,029 and 1,208 samples respectively. The counts do not include studies targeting the known Crypto-Jewish community in Portugal. Eight of the 28 were definitely *T2e1*, 13 were definitely or very likely Not *T2e1*, and the remainder could not be classified. Of the *T2e1* sequences, interestingly half were the Sephardic signature *T2e1a1a* [4]. It was the most prevalent haplotype and was found in the studies examined from central Portugal [S32] and in the northeast among linguistic isolates of Miranda do Douro [S30] and from Spain, including Barges in the North [S22]. Of the remaining *T2e1* sequences, all in Spain, three were ancestral *T2e1* in the control region (Andalusia,

Navarra, and Zamora, Bajo Duero) and one sequence had 16207A (Zaragosa).

By contrast to the Near East, Italy, and Iberia, all sequences of the Iceland *T2e* hotspot, both modern (2 haplotypes) and ancient (1 additional haplotype), were found to be *T2e1*.

A median joining network of control region variation is shown in Figure 2 for all sequences for which *T2e1* status could be determined.

Comparison across populations: For *T2e*-Other, the Near East had the greatest number of unique haplotypes followed by Italy, with Iberia a distant third. Some of the haplotypes were regional specific: Sicily only, Yemen only, Central Italy and ancient Crete, Italy and Spain (including Basque) but not the Near East, and Saudi Arabia and Lebanon but not Mediterranean European. Notable, however, was the overlap in *T2e*-Other sequences between the Near East and Mediterranean Europe (Figure 2). The most populous Near Eastern branch found in Saudi Arabia, Yemen, Qatar, and Iraq with multiple variants also appears to overlap with a sequence from Greece. There was a shared haplotype between Egypt and Central Italy, another amongst Jordan, Central Italy and Greece. Spain also shared motifs with one branch found both in Spain and Saudi Arabia and one appearing in Spain, Italy, and Egypt. Overlap of variants in the literature from Near East and Mediterranean Europe to the largely US based FTDNA *T2* project was minimal (not shown) suggesting the overlap amongst Mediterranean Europe and the Near East was not simply because of widespread dispersal throughout Europe. For the few *T2e1* literature haplotypes, a 16207T variant present in both Spain and Italy was common and found in several samples from FTDNA *T2* project originating from Germany, England, and early US (all low-resolution). When *T2e1* is extracted from the remainder of *T2e*, the geographic prevalence reverses to: highest in Spain, then Italy, and non-existent in the Near East.

Discussion

The new findings bear on the origins of mtDNA haplogroup *T2e* and its most frequent branch *T2e1* (Figure 3). Strikingly, the high frequency of *T2e* in the Near East [4,13] was not from the *T2e1* branch. *T2e1* was not found in Saudi Arabia, Northern Egypt, Yemen, Iraq, Lebanon, Qatar and most likely not Jordan. There was, in fact, a reversal of the frequency of *T2e1* when considered separately from *T2e* overall from a presence in Near East greater than Italy greater than Spain to Spain greater than Italy and greater than (not found) in the Near East. Although the failure to detect the clade in the Near East does not prove its absence, it does increase the importance of Europe for the rise of *T2e1*. If *T2e1* originated in Europe, one candidate is Italy where numerous different haplotypes in mother clade *T2e* were found. Note *T2e1* comprised no more than 10% of the *T2e* haplotypes in Italy with lower than expected diversity. If *T2e1* originated in Italy, expansion of the branch occurred within the recipients of migration rather than in Italy itself.

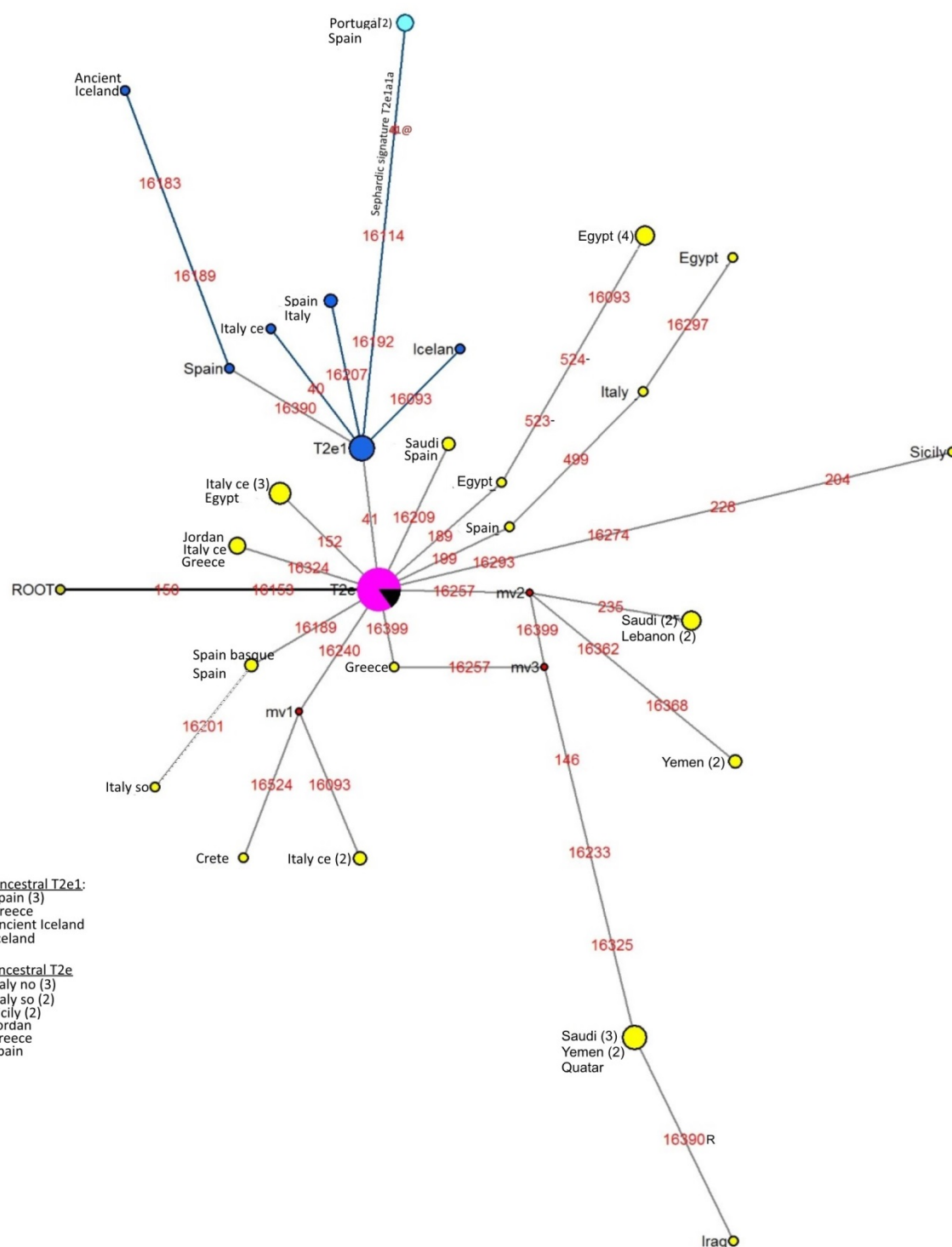
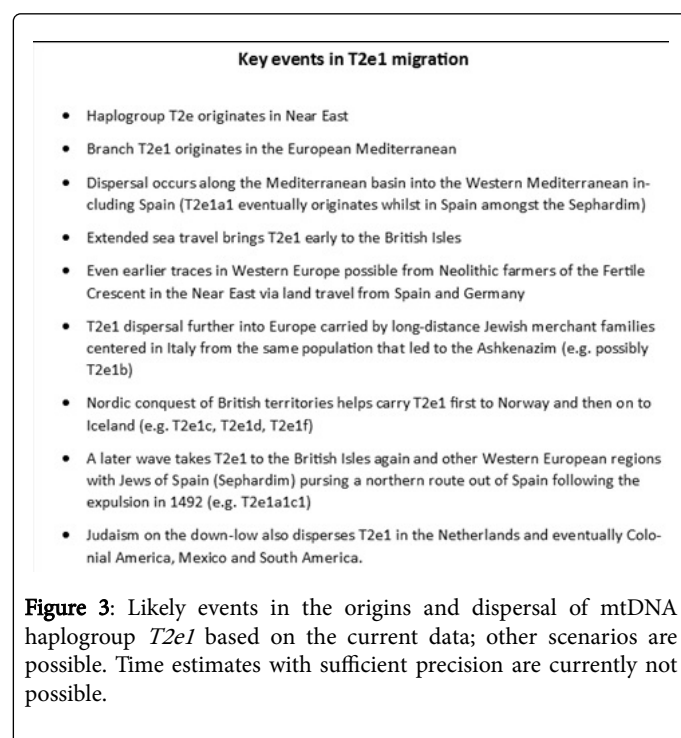


Figure 2: Median joining network for mtDNA haplogroup *T2e* from the literature in the Near East, Italy, Spain, and ancient and modern Iceland and Greece (partial) for control regions *HVS1* and *HSV2* of mitochondrial DNA. Only sequences that could be clearly classified as either *T2e1* or *T2e* - Other are included. Note: 1) *T2e1* is not present in the Near East and 2) There is overlap in sequences between the Near East and Mediterranean Europe for *T2e* - Other reflecting early gene flow. 3) Spain and ancient Iceland may share a rare haplotype bearing on the origins of a *T2e1* hotspot in Iceland. "ce"=central; "mv"=missing value where sequences for that node were not found.

Considerable overlap was found for *T2e*-Other motifs amongst the Near East, Italy, Spain, and Greece. This is consistent with historical accounts of early back and forth travel around the European Mediterranean basin and the Near East before the separation of Muslim and Christian - ruled territories. This also makes it difficult to pin where *T2e* originated. Diversity of haplotypes was found relatively high both in the Near East and in Italy. Interestingly, by removing *T2e1* from consideration, we have found the bulk of *T2e* is no longer in Mediterranean Europe, which may have been used to conclude European origin of *T2e* [13]. Consistent with a relatively greater Near East predominance is the recent conclusion that *T2e* is one of the JT lineages that arrived in Mediterranean Europe from the Near East before the Neolithic in early post-glacial times [20]. A *T2e* sequence found in ancient Egypt 2300 years before present (YBP) [21] is consistent with either origin. Within Mediterranean Europe, we note an overlap between an ancient Crete sequence from the Minoan civilization dated 3700 to 4900 YBP [22] and central Italy (branch *T2e6*) which may provide clues to early dispersal.



It may also be important that *T2e1* is not one of the shared haplotypes with the Near East. This may suggest a temporal origin for *T2e1* later than thought to a time postdating frequent Near East - Mediterranean travel, a spatial origin in Europe further away from the Mediterranean than thought, or the unsatisfying possibility that this clade simply did not survive in the Near East.

Phylogeny

We also report a new major branch of *T2e1* (tentative *T2e1c*), especially notable in Norway with presence in Germany, Sweden, Denmark, and the British Isles. The branch may help solve a mystery of a *T2e* hotspot in Iceland [12]. As noted in Results, Iceland's sequences are entirely *T2e1*. This raises the possibility that the Norway-prominent *T2e1* branches are what underlies Iceland's *T2e* population and reflects the ancient Nordic settlement of Iceland. High resolution sequences are needed for comparison. *T2e1c* may share the migration

path of J2a, which Helgason et al. reported to be found in Iceland, Norway, Sweden, Germany, England, and Ireland [12]. It is fascinating that they also reported the presence of Haplogroup *T2e* in DNA extracted from ancient Icelandic remains dating more than 1000 YBP [23]; two sequences were found to be *T2e*, both of which we note have 41T and are therefore *T2e1*. We have been able to locate one possible match to the sequence harboring variant 16390A (from Saldaña Spain; mitosearch.org, [4]). These are low resolution sequences but again raise the issue of a Spain connection for *T2e1*, even as far as Iceland.

A new Irish-specific subclade found is notable for its nearness to the Ottoman Sephardic - Mexican *converso* branch. Both branches harbor 2308G and 15499T variants, which comprise *T2e1a1*. The Irish clade's nearest genetic neighbor is a single (corrected) Moroccan sequence [19,5] and together they form a third bifurcating branch (tentative *T2e1a1c*) of *T2e1a1* reported in our previous work, the first being the Sephardic signature (*T2e1a1a*) and the second a mixture of the Netherlands and Colonial Americans from England (*T2e1a1b*). The relatively close genetic distance amongst these three branches suggests meaningful gene flow between Jewish and non-Jewish affiliated groups (see Jewish Demographics below).

We were also able to find the first sequence at the *T2e1a* root, a sequence from the border between Poland and the Czech Republic bearing no additional variants other than 2308G. Whilst the Polish-Czech sequence is at the root of this tree, we do not think it means that the 2308G variant originated in Eastern Europe nor that this signature was carried in a migration from Eastern Europe to Ireland, Iberia, Spain, and Denmark, the branch locations of this tree. Instead, the ancestral 2308G sequence is likely the only surviving example currently to be found, originally perhaps even from Spain, with the mother node now largely extinguished in favor of daughter node expansion. (Ashkenazim H7 provides another example of this phenomenon [16]). It instead likely reflects traces of dispersal of *T2e1a* as far as Eastern Europe, like the previously elaborated Jewish 9181G branch [5]. Both this root and the Irish clade help flesh out *T2e1a*, a major bifurcation of *T2e1*.

Variant 16207T may prove an important reflection of gene flow between southern and western Europe. Minor new branches of *T2e1* hailed from Italy, England, Ireland, Germany, Sweden, Switzerland, Norway, and Austria, not uncommon origins for our largely United States participant pool.

Deleterious, advantageous, or neutral variants?

A transition from T to C at position 14180 in the ND6 subunit of Complex 1 mtDNA was found here in two distinct branches. It is predicted to be likely damaging by PolyPhen-2 and was noted to accompany a case of hereditary optical neuropathy (LHON) in haplogroup H [24]. We suggest that it could instead have a positive effect in some haplogroups. It is well-known that all of haplogroup T harbors the nonsynonymous variant of 4917G, expected to decrease production of ATP which in turn will decrease the production of ROS [25]. In the haplogroup T context of lower ATP production and consequent lower ROS, a variant which increases ROS may be better tolerated and any increase in successful ATP production may even convey an advantage in preventing conditions in which lowered energy (e.g., cardiomyopathy, muscle weakness) is problematic.

Intriguingly, the second of the three major *T2e1* clades, Jewish *T2e1b*, also has its defining variant argued in the literature to be pathogenic and linked to increasing ROS [26,5]. The 9181G

nonsynonymous variant that defines the clade leads to a protein change that was shown directly by Zhang et al. [26] to produce increased ROS and cell death. As with *14180C* above, in the setting of the T haplogroup with lower ATP production, the *9181G* variant may not only fail to be harmful but we suggest could mitigate any low energy risks of Haplogroup T. Note the first of the three major *T2e1* clades, *T2e1a*, has an RNA variant of *2308G* whose effects are unknown.

Jewish demographics

We have found that there are several non-Jewish European sister and "cousin" clades to the Sephardic branches within *T2e1* and that *T2e1* has yet to been found in the Near East. However, we do not think a European origin can be concluded from the data. The presence of nearby parallel European branches to Jewish branches indicates that there was gene flow between the branches but does not indicate the direction of exchange. Jewish migration following arrival in Europe may have caused genetic introgression into Europe rather than the other way around.

Some *T2e1* branches may be sensible in such a view. The connection amongst the extensive *2308G* group (*T2e1a*) from Colonial America, England, Netherlands, Morocco, Ireland, Denmark, Poland, Ottoman Sephardic, Mexican *converso*, and Portuguese cryptos may fit with the two migration paths from Spain following Sephardic exile. Jews who went south into Muslim lands where they were welcome were able to openly practice Judaism. This makes the mtDNA branches easy to recognize as "Jewish clades". Exilers who instead went north through Christian Europe had to keep a much quieter existence. We suggest a name generally for less open Jewish affiliation based on American slang: *Judaism on the down-low*. They split further into populations that stayed in Portugal (including crypto-Jews), some of whom eventually went on to Mexico (*converso*), and those who quietly distributed around Europe. The Netherlands, British Isles, and Colonial America all have historically documented Sephardic presence. European sister clades to known Sephardic branches within *T2e1a* may reflect traces of post-Iberian Sephardic expulsion along the difficult northern route out of Spain. Earlier travel of Jews in the Mediterranean dispersed through merchant families through long distance trade routes also may be causal for these and other nearby European branches. With sister clades to the Sephardic ones we may be detecting traces of a genetic footprint of long forgotten Jewish ethnic and religious affiliation, perhaps even as far as Iceland.

It would appear than an estimate for the age of *T2e1* of *9176 YBP* based on thousands of mitogenomes [15] would disprove the Jewish hypothesis of relatively recent ("biblical-time") migration from the Near East. We would like to make explicit that what is critical for Jewish demographics is not if they came from the Near East but when because all Europeans came from the Near East. However, the age estimate is based on mtDNA variant rates using phylogenies. Estimates of variant rates based instead on pedigrees have found them at least three times faster [27], which would put the timeline back in the range of historical Jewish diaspora [17]. Time estimates with sufficient precision are currently not possible [4,5,16]. The best evidence for an older emergence of *T2e1* that predates the Israelites comes from a sample of *T2e1* found in skeletal remains in Germany estimated from the mid Neolithic about *5300 YBP* [28]. It bears *41T* marking it as *T2e1* specifically and interestingly, was not found to have any additional variants unlike any modern *T2e1* found to date. Thus far, there has been only one Neolithic *T2e1* finding. The *T2e1* branch of *T2e* has not been found in a sample of over 3000 modern Sardinians, a

Mediterranean island population believed to have the greatest genetic similarity in Europe to early Neolithic farmers [29]. Spatial resolution of genetic variants is now approaching its maximum with the increased testing of the entire mitochondrial DNA. Higher resolution coalescence times for genetic variation (i.e. time estimates) on par with high spatial resolution might be the next breakthrough in Jewish genomics.

The origins of *T2e1* for "the other European Jews", the Sephardim, appear to resemble that for Ashkenazi Jewish maternal lineages found by Costa et al. [2]. Both Sephardic haplogroup *T2e1* derived from Near Eastern-Mediterranean *T2e* and Ashkenazi haplogroups may trace to the same early Italy - Mediterranean population. Consistent with a common origin of Ashkenazi and Sephardic founding women are mtDNA haplogroups that are shared between both populations [1,2,5,7]. Autosomal genetic studies have found overlap between Italians and both Sephardic and Ashkenazi Jews [30]. Shared maternal origins amongst the different Jewish groups may have far-reaching implications. Even what is believed to be quintessential Ashkenazi haplotype, *K1a1b1a*, could prove to be Sephardic in origin. The relative incidence in the populations may have been deceptive; the vast presence of the *K1* haplotype in Ashkenazim but mere spattering in Sephardim, could result not from an origin with the Ashkenazim but only because of the great population decline in Sephardim, the explosion in Ashkenazim and its very narrow founder effect. Re-examination of all shared Jewish mitochondrial haplotypes may be fruitful.

Finally, the very diversity of Sephardic mtDNA that makes it attractive for uncovering the origins of European Jewish women also makes it difficult to generalize. The present haplogroup *T2e1* investigated appears to be part of a very early Jewish presence in Iberia 350 BCE to 40 CE from the same source as Jews in Central Italy with dispersal around Mediterranean by sea and as those that became the now populous Ashkenazim, as well as many, we argue, that lost their Jewish identity. While this very early Jewish presence in Spain with connections to other European Jewish populations is remarkable to have found in the genes, we do not know if this is an isolated incident for Sephardim. One might instead expect the bulk of the current Sephardic population to comprise a much later influx to Iberia from 700 CE when Jews residing in Babylonia travelled to the newly Muslim Iberia *via* land thru Muslim North Africa and Sicily.

Future investigations may find Sephardic genomics add further novel insight to origins and migrations of the Jewish people in Europe. If Sephardim are considered the other European Jews then additional richness may be uncovered by studying the very first Jews to settle in Europe, the Romaniots, the other *other* European Jews.

Acknowledgments

Many thanks are due to Robert Bedford of the Foundation for the Advancement of Sephardic Studies and Culture for providing details of Sephardic history. We also wish to thank Cristina Martinez-Labarga, Luca Pagani, Alexander Peltzer, and Marian M. de Pancorbo for clarification of literature DNA sequences, FTDNA *T2* project administrators Anne Berge, Elizabeth Kipp, and David Pike, project participants and FTDNA for their kind cooperation, and finally Ian Logan, Doron Yacobi and two anonymous reviewers for helpful comments on an earlier version of the manuscript.

Conflict of Interest

No conflict of interest is declared.

References

- Behar DM, Metspalu E, Kivisild T, Achilli A, Hadid Y, et al. (2006) The matrilineal ancestry of Ashkenazi Jewry: Portrait of a recent founder event. *Am J Hum Genet* 78: 487-497.
- Costa MD, Pereira JB, Pala M, Fernandes V, Olivieri A, et al. (2013) A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. *Nat Commun* 4: 2543.
- Carmi S, Hui KY, Kochav E, Liu X, Xue J, et al. (2014) Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins. *Nat Commun* 5: 4835.
- Bedford FL (2012) Sephardic signature in haplogroup T mitochondrial DNA. *Eur J Hum Genet* 20: 441-448.
- Bedford FL, Yacobi D, Felix G, Garza F (2013) Clarifying mitochondrial DNA subclades of T2e from Mideast to Mexico. *J Phylogenetics Evol Biol* 2: 1-8.
- Bedford FL, Yacobi D (2015) On two Jewish clades in mitochondrial DNA. *Eur J Hum Genet* 23: 993-994.
- Behar DM, Metspalu E, Kivisild T, Rosset S, Tzur S, et al. (2008) Counting the founders: The matrilineal genetic ancestry of the Jewish Diaspora. *PLoS One* 3: e2062.
- Nogueiro I, Teixeira J, Amorim A, Gusmão L, Alvarez L (2015) Echoes from Sepharad: Signatures on the maternal gene pool of crypto-Jewish descendants. *Eur J Hum Genet* 23: 693-699.
- Rendon C (2017) DNA test reveals Woman with 21,795 descendants is Haplotype T. Available at: home.earthlink.net/~shharmembers/tcmtdna.pdf. (Accessed: 6th April 2017)
- Kumar S, Bellis C, Zlojutro M, Melton PE, Blangero J, et al. (2011) Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. *BMC Evol Biol* 11: 293.
- van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30: E386-394.
- Helgason A, Sigurðardóttir S, Gulcher JR, Ward R, Stefánsson K (2000) mtDNA and the origin of the icelanders: Deciphering signals of recent population history. *Am J Hum Genet* 66: 999-1016.
- Pala M, Olivieri A, Achilli A, Accetturo M, Metspalu E, et al (2012) Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *Am J Hum Genet* 90: 915-924.
- Pike DA, Barton TJ, Bauer SL, Kipp E (2010) mtDNA haplogroup T phylogeny based on full mitochondrial sequences. *J Genet Geneal* 6.
- Behar DM, Oven M, Rosset S, Metspalu M, Loogvali EL, et al. (2012) A 'Copernican' reassessment of the human mitochondrial DNA tree from its root. *Am J Hum Genet* 90: 675-684.
- Yacobi D, Bedford FL (2016) Evidence of early gen flow between Ashkenazi Jews and non-Europeans in mitochondrial DNA haplogroup H7. *J Genet Geneal* 8.
- Tofanelli S, Taglioli L, Bertoni S, Francalacci P, Klyosov A, et al. (2014) Mitochondrial and Y chromosome haplotype motifs as diagnostic markers of Jewish ancestry: A reconsideration. *Front Genet* 5.
- Behar DM, Rosset S, Smith JB, Balanovsky O, Tzur S, et al. (2007) The Genographic Project public participation mitochondrial DNA database. *PLoS Genet* 3: e104.
- Maca-Meyer N, González AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* 2: 13.
- Pereira JB, Costa MD, Vieira D, Pala M, Bamford L, et al. (2017) Reconciling evidence from ancient and contemporary genomes: A major source for the European Neolithic within Mediterranean Europe. *Proc R Soc B* 284: 20161976.
- Schuenemann VJ, Peltzer A, Welte B, Pelt P, Molak M, et al. (2017) Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. *Nat Commun* 8: 15694.
- Hughey JR, Paschou P, Drineas P, Mastropaolo D, Lotakis DM, et al. (2013) A European population in minoan bronze age crete. *Nat Commun* 4: 1861.
- Helgason A, Fox CL, Ghosh S, Sigurðardóttir S, Sampietro ML, et al. (2009) Sequences from first settlers reveal rapid evolution in icelandic mtDNA pool. *PLOS Genet* 5: e1000343.
- La Morgia C, Caporali L, Gandini F, Olivieri A, Toni F, et al. (2014) Association of the mtDNA m.4171C>A/MT-ND1 mutation with both optic neuropathy and bilateral brainstem lesions. *BMC Neurol* 14: 116.
- Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC (2004) Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* 303: 223-226.
- Zhang C, Huang VH, Simon M, Sharma LK, Fan W, et al. (2012) Heteroplasmic mutations of the mitochondrial genome cause paradoxical effects on mitochondrial functions. *FASEB J* 26: 4914-4924.
- Madrigal L, Castri L, Obando MM, Palma RV, Barrantes R, et al. (2012) High mitochondrial mutation rates estimated from deep-rooting Costa Rican pedigrees. *Am J Phys Anthropol* 148: 327-333.
- Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, et al. (2015) Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522: 207-211.
- Olivieri A, Sidore C, Achilli A, Angius A, Posth C, et al. (2017) Mitogenome diversity in Sardinians: A genetic window onto an island's past. *Mol Biol Evol* 34: 1230-1239.
- Atzmon G, Hao L, Peer I, Velez C, Pearlman A, et al. (2010) Abraham's children in the genome era: Major jewish diaspora populations comprise distinct genetic clusters with shared middle eastern ancestry. *Am J Hum Genet* 86: 850-859.