

## Clarifying Mitochondrial DNA Subclades of T2e from Mideast to Mexico

Felice L Bedford\*, Doron Yacobi, Gary Felix and Federico M Garza

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

### Abstract

We report on two of the oldest mitochondrial DNA clusters in existence with Jewish affiliation. Both are in haplogroup T2e1. Four unrelated individuals from the Mexico mtDNA project were found to have the control region mutations that characterize a Sephardic signature previously reported (motif 16114T-16192T within T2e). Full genomic sequencing found the identical coding region mutations as Sephardic individuals which provides genetic evidence for founders of Northern Mexico that were both female and Sephardic Jewish. This is in contrast to a more common finding of European male, but local female founders and additionally lends biological support to anecdotes and historical reports of Crypto-Jewish founding of the Coahuila, Nuevo León, and Tamaulipas regions of Mexico and influx to Southern Texas, USA. The haplotype is nested in an old tree with mutations at positions 2308 and 15499, presently of uncertain geographic origin. The second cluster, a Bulgarian Sephardic founding lineage (9181G within T2e) previously reported, was found here in a population of largely Americans of European descent, but only among Jewish individuals. The non-synonymous mutation in ATPase 6 was found among both Ashkenazi and Sephardic Jews from diverse regions of Czech Republic, Lithuania, the Netherlands, Poland, and Romania. Full genomic sequencing found great coding region variability with several haplotypes and suggested a Near East origin at least 3000 years old. This predates the split between Jewish groups, but more recent admixture between Sephardim and Ashkenazim cannot be ruled out. Together the two Jewish-affiliated clusters account for all the genetic distance found in branch T2e1 and much of T2e. The findings suggest reexamination of the origins of mitochondrial DNA haplogroup T2e as Levantine or early back migration to the Near East. New subclades of T2e are identified.

**Keywords:** Mitochondrial DNA; Haplogroup T; Sephardim; Population genetics; Jewish; Mexico

### Introduction

The ancient migration of populations from one geographic region to another can be traced through individuals living right now. DNA bears witness to events thousands of years old. This includes mitochondrial DNA (mtDNA) inherited only from the mother, which therefore, allows specifically for the maternal origins of a population to be identified. Haplogroup T within mitochondrial DNA refers to a well-known collection of genetic mutations believed to point to a woman living in the Near East at least 25,000 years ago, who is the common maternal ancestor of about 10% of contemporary Europeans and Americans of European descent [1-5].

A small branch (*subclade*) of haplogroup T, known as T2e [6,7], comprises only a fraction (0.2%) of those from England and some other Northern European locales but is much more prevalent in Egypt, Saudi Arabia and Italy [8,9]. A notable feature of subclade T2e is that it houses two different motifs on distinct branches believed to be affiliated with peoples of Sephardic Jewish descent. The Sephardim refer to Jews that settled in Spain and Portugal following the Diaspora [10], which led to 1500 hundred years of genetic and cultural isolation from other Jewish groups, such as the Ashkenazi (Europe) and Mizrahi (Near East). One of the motifs was reported by Behar et al. [11] as a founding lineage for the Sephardic Bulgarian community. Bulgaria was a part of the Ottoman Empire that welcomed Jews from the Iberian region following their expulsion from Spain in 1492. Inclusion is defined by a 9181G mutation in the coding region within subclade T2e. Behar et al. [11] sequenced the complete mtDNA for one person from this group and deposited the sequence to the National Center for Biotechnology Information (NCBI) Genbank, the central registry for DNA sequences. Originally labeled as T2f to highlight its distinctiveness, the motif meets all the inclusion criteria for T2e. The second motif was identified by Bedford [8]; it was found to be a rare subclade harbored by Sephardics descending from the Ottoman Empire and also appeared, quite surprisingly, in a number of individuals from Northern Mexico and South Texas in the USA. These are regions where suspicion of *converso* Sephardic origin is high [12]. The motif is identified by a pair of control region mutations, 16114T and 16192T, within subclade T2e. One person with this Sephardic signature (ancestral origin Salonica, Ottoman Empire) was fully sequenced and deposited in Genbank.

However, questions about T2e subclades remain. T2e has had the technological good fortune of being distinguishable from other branches of T2 by mutations only in the control regions of the mitochondrial DNA (16153A in control region 1 and 150T in control region 2). There are no additional mutations from the coding region that define the subclade. This means that a sample can be definitively classified as belonging or not belonging to T2e without further sequencing beyond the first control region. Because the vast majority of the extensive worldwide accumulated databases on mitochondrial DNA continue to be from the first control region, perfect control region identification has permitted accurate analyses of T2e thus far, including the creation of phylogenetic trees showing the relation between different patterns of mutations (haplotypes), and the assessment of the frequency of the subclade in different geographic regions [8,9,13,14]. Despite this advantage enjoyed by subclade T2e, there are issues that cannot be addressed without consideration of the coding region. In particular, the two Sephardic haplotypes require inspection of the coding region to clarify the significance of these subclades.

For the 16114T-16192T cluster, it remains uncertain if people of Mexican descent who harbor these mutations truly share a common origin with the known Sephardim who have the same mutations. The resemblance could instead be superficial, in which the pair of control region mutations observed merely arose independently in distinct lineages. Several studies have found that sequences with the same distinctive control region haplotypes have appreciable variation in the coding region [11,13,15]. The true basis for the assessment of common origin comes from the coding region, in part because it is larger than the control regions, and in part because surviving mutations accumulate much more slowly than in control regions. This permits temporally distant connections to be revealed. If the Sephardic and

\*Corresponding author: Felice L Bedford, University of Arizona, PO Box 210068, Tucson, AZ 85721, USA E-mail: [Bedford@u.arizona.edu](mailto:Bedford@u.arizona.edu)

Received September 03, 2013; Accepted October 10, 2013; Published October 17, 2013

Citation: Bedford FL, Yacobi D, Felix G, Garza FM (2013) Clarifying Mitochondrial DNA Subclades of T2e from Mideast to Mexico. J Phylogen Evolution Biol 1: 121. doi:10.4172/2329-9002.1000121

Copyright: © 2013 Bedford FL, et al. 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.

Mexican samples were found to have completely different coding region mutations, this would negate the hypothesis of Sephardic origin for the Mexicans and point instead to the independent emergence of the same control region mutations on two very different branches. On the other hand, finding an exact coding region match would provide firm genetic evidence that there are surviving Catholic Mexicans today who descend from Spanish and Portuguese Jewish women fleeing Iberia, presumably in attempt to escape persecution for practicing Judaism.

Likewise, the 9181G cluster also requires coding region testing, but for a different reason. Unlike the Ottoman Sephardic/Mexican signature with control region mutations at positions 16114 and 16192, the 9181 cluster cannot be identified by means of any control region mutations. Inspection of just the first control region shows only the ancestral T2e haplotype, a very common non-identifying result found in more than half of all T2e samples [8,14]. The second control region adds only 41T, also a common, possibly unstable, variant within T2e. Without further coding region exploration, questions about the 9181 cluster cannot be addressed despite hundreds of thousands of records of control region mtDNA in multiple databases. How prevalent is this founding lineage outside the Sephardic Bulgarian community where it was discovered? Who possesses it? Finally, coding region testing can lead to advances in how these two branches, 16114T-16192T and 9181G, connect both temporally and spatially to the larger T2e tree within which they are nested.

Consequently, the present work reports on the two Jewish clusters within mitochondrial DNA subclade T2e by testing the coding region of appropriately chosen samples.

## Materials and Methods

To identify Mexican samples with the Sephardic signature 16114T-16192T haplotype, the database of the 890 participants of the Mexican mtDNA project at Family Tree DNA administered by one of us (Gary Felix) was mined. The Mexican DNA project is open to the public for anyone with Mexican roots interested in genetic testing (<http://www.familytreedna.com/public/GenealogyofMexicoDNAProject/default.aspx>.) All participants whose control region data showed they had the defining haplotype within T2e were issued invitations to participate in the present research study and undergo further testing to fully sequence their mitochondrial DNA.

To identify samples with the Sephardic Bulgarian founding lineage 9181G, a single sample with the mutation was first found in a customer at Family Tree DNA (the “kernel”). The American-based Family Tree DNA company (FTDNA) offers genetic testing services to individuals and is used largely by Americans of European descent. The data from individuals ordering tests from FTDNA is being increasingly used as a scientific resource. More than a quarter of the entries on Genbank are from FTDNA, some submitted by customers themselves. Other sequences from FTDNA testing are submitted to Genbank in connection with research studies using data from consenting FTDNA customers; most significantly, a study on the reassessment of the mtDNA reference standard by Behar et al. [16] led to a deposit of over 4000 sequences. In the present study, the kernel was used to search for other samples that differ from it by 0 to 3 mutations in individuals who all already had undergone full mitochondrial genomic testing. These people were contacted with inquires as to 1) whether they had 9181G, and 2) their deep maternal ancestry. Anyone responding affirmatively to possessing the mutation at position 9181 was then used to find other matches within 0 to 3 genetic differences and the process continued in iterative fashion until all unique matches were contacted. This process allowed the maximum number of different haplotypes bearing 9181G to be identified, as well as scouting for maximum total number of existing

9181G sequences. Matches were invited to officially participate in this research study, including permission to deposit sequences in Genbank.

To relate the results of both the 16114T-16192T and the 9181G clusters to the larger branch structure of T2e, all T2e samples currently on Genbank were inspected for any overlapping mutations with either of the current two clusters. T2e sequences were downloaded from Genbank in FASTA format and assessed for differences from the rCRS (see Discussion). In addition, we searched the Pike analysis of haplogroup T mitochondrial full genomic sequences from FTDNA, not all of which appear on Genbank [13]. Current FTDNA records were also searched directly for overlap in a manner similar to that used for the 9181 cluster. Finally, overlap was sought across literatures that do not always post in Genbank.

Because a potentially incomplete T2e sequence from Morocco appearing on Genbank [17] was pivotal to the branch structure, clarification was sought on resequencing of this sample. Network 4.6.1.1 by Fluxus was used to assist creation of phylogenetic trees. Any ambiguous branch assignment was resolved in favor of branch structure and nomenclature already existing on Phylotree Build 15 [7]. Testing was completed by Family Tree DNA, whose detailed methods are available elsewhere [5]. Information on maternal ancestry was obtained wherever possible, usually directly from the participants in this study.

Time to most recent common ancestor was assisted by considering all possible pairs within a cluster of interest and by counting the total number of differing positions along the coding region of mtDNA (from 600 to 16000) for the pair. The maximum genetic distance for the cluster was specified by the highest count among the pairs. The maximum genetic distance was determined for each of the Jewish clusters (16114T-16192T and 9181G), for each of two larger branches from which the Sephardic signature cluster 16114T-16192T is a sub-branch, and the 9181G cluster and the larger branch of 16114T-16192T taken together. For further comparison, the maximum genetic distance was also determined for all of T2e and for T2e1. These genetic distances allowed the creation of a relative timeline within subhaplogroup T2e that is stable under the greatly differing assumptions of the mutation rate currently in the scientific literature.

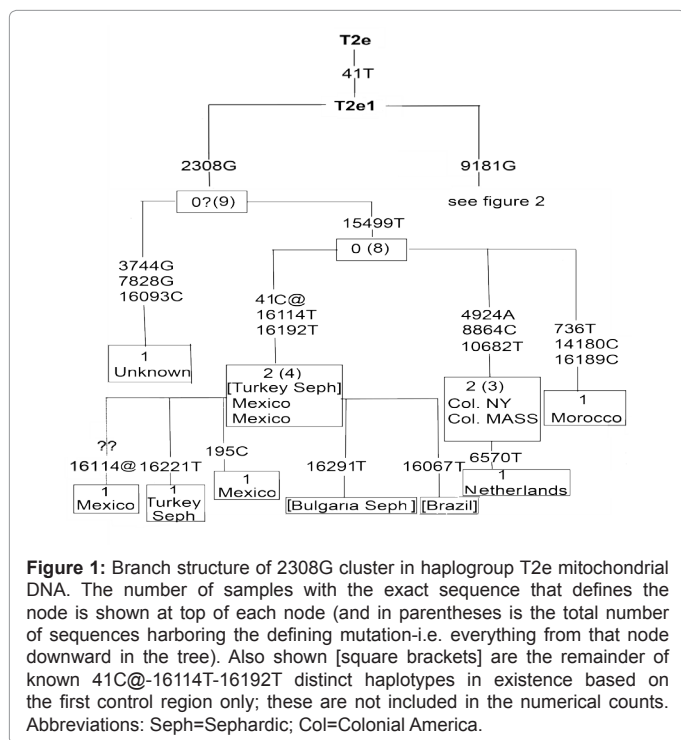
## Results

### Mexico

Three unrelated participants of the Mexican mtDNA project were found to have the 16114T-16192T Sephardic signature within T2e. All agreed to further testing. The geographic locations of the maternal ancestors dating back to the 19<sup>th</sup> century were provided by the participants as Mier, Tamaulipas Mexico; San Diego, Texas, USA; and Comales, Tamaulipas Mexico. These locations from Northern Mexico and South Texas are consistent with the previous report by Bedford [8] of Latinos found to have the signature. Two additional individuals from the Mexican mtDNA project were found to have the 16192T mutation within T2e, but not 16114T. A previous search of over a quarter of a million mtDNA records did not find a single sequence within T2e possessing only one of the pair of required mutations for the Sephardic signature [8]. The present 16192T transition without 16114T could reflect a newly discovered branch of the Sephardic signature or could reflect an independent emergence in a different lineage of T2e. While 16192T has thus far proved rare within T2e (occurring only in the Sephardic signature), it emerges not infrequently in other branches of haplogroup T. The present 16192T-only sequences belonged to a mother and daughter who trace their most distant maternal ancestor to Aramberri, Nuevo León, Mexico. They agreed to participate and one of their samples was tested further. The 5 samples in the present study had no additional private mutations in the first control region (HVS-1).

Full genomic sequencing confirmed all 3 of the exact HVS-1 matches to the Sephardic signature (16114T-16192T) belonged to haplogroup T2 with its defining mutations at positions 11812 and 14233. This result was expected under all hypotheses. In addition, all 3 samples were found to have additional private mutations: 2308G and 15499T. These are the identical mutations found for the Sephardic sample descended from Salonica, Ottoman Empire that was sequenced previously [8]. Neither of these two coding region mutations is defining for the Sephardic signature however, as both are found in T2e samples that do not share the critical defining pair of control region mutations (See “Larger 2308 branch” below). No additional coding region mutations were present in either the previously sequenced Salonican sample or any of the 3 current Northeast Mexican/Texan samples. One of the present samples had an additional mutation in the second control region (the Comales sample), at position 195, a not uncommonly varying site.

The finding of an identical coding region match to the known Sephardic sequence disproves the hypothesis that the control region match between Sephardim and Mexicans was a coincidental one.



Instead, the data favor a shared common lineage.

The fourth sample tested, which harbored only one of the pair of Sephardic identifying control region mutations, was also found to have the identical private coding region mutations of 2308G and 15499T as the other samples. No additional private mutations were found. Thus, the 16192T-only motif within T2e reflects a newly discovered branch of the Ottoman-Mexican Sephardic signature.

For results related to geographic origins, 2 of the participants were found to have a most recent common ancestor 11 generations removed along the deep maternal line: Inez Benavidez, born circa 1655 in Monterrey, Nuevo León, Mexico. Indeed, Inez Benavidez’s mother, Clara Flores Cerda, has over 21,000 descendants, with 151 named people predicted to have the T2e Sephardic signature haplotype between the years 1709 and 1887 bearing 23 different surnames and 7 different geographic locations (home.earthlink.net/~shharmembers/tcmdna.pdf). The deep maternal line may be traceable further [18,19] to Maria-Ines de la Cerda y Castro born in 1590, one of the founding families of the Texas-Coahuila- Nuevo León region, and to Josefa de Castro living in Spain 2 generations before that. Claims of early lineages can be difficult to substantiate. Several levels could use further proof, including the parentage and birth year of Clara Flores de Cerda despite numerous anecdotal reports that Josefa Fernández de Castro y de la Cerda was her mother. Table 1 shows the putative maternal line from the most recent common ancestor to its origins in Spain.

### Larger 2308G branch

The results of comparison of the present T2e motif 16114T-16192T-2308G-15499T to Genbank found 5 sequences in T2e with overlapping mutations, 2 of which were from our previous study [8] (one with Sephardic signature and one with both 2308G and 15499T and 3 additional coding region mutations). The oldest sequence of T2e deposited in Genbank was in the year 2006 belonging to a Moroccan individual reported to have only the mutation at position 2308 but not at 15499, a back mutation at the defining T2 position 11812 from base G to A, 2 additional coding region mutations, 16189C in the first control region, and the absence of 41T (i.e., 41C) in the second control region [17]. Since this sequence was critical to the branch structure of the entire 2308G cluster, clarification of this pioneering sequence was sought. Resequencing of fragments found that the sample does not have a back mutation at position 11812 and moreover, also possesses mutations at 15499 and 41 (Ana M. González Matilla, personal communication, Dec 14, 2011). This correction allows a coherent branch structure of the 2308G cluster to be created.

Results of a constructed maximum parsimonious phylogenetic tree

G #	Name (Female lineage)	Birth year	Birth location	Comment
17	Josepha de Castro	?	Spain, probably	
16	Isabel de la Castro y Figueroa	1565/1582	Spain	Birth dates vary (death before 1651)
15	Inés de la Cerda y Castro	1590/1601	?	Birth dates vary; Had one other daughter: Catarina, who died in Parras, Coahuila, Mexico
14	Ana Maria de la Cerda (y Porcallo?)	1630	Mexico, probably	Died in Monterrey, Nuevo León Mexico.in 1707; Father’s name Santos-Coy, not Porcallo; Had one other daughter: Jacinta
13	Josefa Fernández de Castro y de la Cerda	1652/1643?	Saltillo, Coahuila, Mexico	Birth dates vary; Had 3 other daughters: Juana, Maria, & Teresa
12	Clara Flores de la Cerda	1672 (1634 ?)	Cerralvo, Nuevo León Mexico	Birth year for her and husband vary substantially.
11	Inéz Báez de Benavides y Flores	1685	Cerralvo, Nuevo León Mexico	MRCA for 2 of our participants with T2e1a1a motif 41C@-16114T-16192T

**Abbreviations:** MRCA=Matrilineal most recent common ancestor G#=Generation number counting back from present generation.

**Table 1:** A possible Mexican founding lineage for the mtDNA T2e Ottoman-Mexican Sephardic signature Names of other daughters whose descendents would also harbor this signature are noted.

with this corrected Moroccan sequence, the new Mexican samples, and the previous Genbank sequences are shown in Figure 1. Position 16189 was down weighted to 0 due to its frequent appearance in independent lines within subclade T2e and was not used to define nodes.

Several things are apparent. Note first that with the correction of the Moroccan sequence, this leaves only one branch within this 2308G cluster that does not have 41T in the second control region, namely the Sephardic signature sequence. Although the stability at position 41 in T2e has yet to be completely determined (for example, a subclade recently labeled T2e5 also contains some motifs with and without 41T), to be consistent with previously published phylogeny and nomenclature [7] (PhyloTree.org), the Sephardic signature cluster is shown as a subclade of T2e1. T2e1 has been defined by the presence of a transition from C to T at position 41. Thus, the Sephardic signature is now clarified as having a back mutation at position 41, a new criterion for inclusion along with the previously established 16114T and 16192T. A total of 5 individuals with these control region mutations to date have been fully sequenced and we have labeled this subclade here as T2e1a1a in keeping with naming conventions.

As can also be seen in the tree, there are 3 branches bearing the 15499T mutation. One is the Sephardic signature as discussed. Another is the (corrected) Moroccan sequence for an individual from Northern Morocco of reported Arab origin. We have tentatively labeled this branch held by a single individual as T2e1a1c but additional sequences are required for confirmation. There was an influx of Sephardic Jews into Morocco following the expulsion from Spain adding to an ancient Jewish population already present. There is no particular reason to suspect this sample has Sephardic origins, but in any event, both the Arab and Sephardic clusters together point to an old Near Eastern origin for more than one branch of this 2308G tree. The third branch is a puzzling one with 2 samples that date to colonial times in the USA, from New York and Massachusetts. The latter is traceable to England and the former contains Dutch surnames in an extended family tree available from the Sorenson Foundation. A third member of this cluster reports origins in the Netherlands, of apparent Mennonite ethnicity. There are historical connections between the Sephardim in Netherlands and English visitors who returned to England with Sephardic customs— and perhaps genetics? There were also early Sephardic settlers in colonies of the United States, although Massachusetts initially did not allow Jewish residents. These connections may be worth noting for any potential relevance that may surface in the future. We have labeled this branch T2e1a1bl, with inclusion criteria of 3 additional coding regions and 2 control region mutations; refinements may be possible with future samples within this branch.

There is thus far only one confirmed sample possessing 2308G without 15499T. This was originally thought to characterize the Moroccan sequence, but the correction as stated above has found that sequence to also possess 15499T. The remaining 2308G-only sequence is a relatively new one to Genbank deposited by Behar/FTDNA in connection with the aforementioned article (see Materials and Methods) on the reconstructed sapiens reference standard and depositing of 4265 sequences in Genbank [16]. It has a control region mutation of 16093C, an old mutation that can be seen within T2e sequences from Northern Egypt [20,21] Italy [22], Iceland [23] British Isles[8] and an ancient sample from Germany more than 5300 years before present [24]. These may not all reflect a monophyletic clade. The 16093 position is involved in at least one reticulation in a T2e network [24] and relatively frequent heteroplasmy has also been found here. The role of mutation 16093C in haplogroup T2e has yet to be determined. The 2308-only sequence possesses 2 additional private coding region mutations. We

have tentatively labeled the branch T2e1a2 and additional samples are required for confirmation. No information as to the ancestral origins of this sequence was provided by Behar/FTDNA; Geographic and ethnic origin for this sequence might have provided a clue as to the origins of the entire cluster because it is closest to the topmost 2308 node. Also informative would be an ancestral 15499T sequence without further mutations, but such a sequence has not yet been reported. Likewise, an ancestral 2308G sequence without further mutations has yet to be discovered. An ancestral 2308G sequence may have been so found by Pike et al. [13] but some of their reported sequences were listed as “simplified”. Since the source was from a customer testing with FTDNA, the sequence may have been the 2308G-only (i.e. lacking 15499T) sequence but with the 3 additional mutations just described that was deposited by Behar et al. [11] to Genbank.

The Sephardic signature cluster can be seen within the larger 2308G tree. There is ambiguity about the placement of the newly found 16192T-only sequence. The parsimonious placement based on mutations alone would indicate that it is a branch which precedes the emergence of the signature with 16192T-16114T mutations, which until the present report, have only been found in tandem. However, the fact that the 16192T sequence has surfaced in only one mother-daughter pair from a derived population may point instead to a recent back mutation at 16114 (from T back to C) occurring only after the establishment of the 16192T-16114T Sephardic signature. Maternal lineage for this family beyond 3 generations was unobtainable. To be conservative, we have currently not used this 16192T-only sequence to define a superordinate node on the tree.

Due to a number of distinct haplotypes, the 2308 branch of T2e is an old one (see below, “Time to most recent common ancestor”).

### The 9181 cluster

A total of 24 extended matches were found to the kernel sequence described in Materials and Methods. Of these potential matches, 15 responded to inquires wherein it was determined that 6 had the mutation of interest at position 9181, just like the kernel sequence, and 9 did not. For the total of 7 individuals with the defining mutation (kernel plus 6 matches), all 7 reported Jewish ancestry along the deep maternal line. This proved to be in sharp contrast to the 9 individuals without 9181G, none of whom knew of any Jewish ancestry (Fisher exact test,  $p < .0001$ ). Details of the Jewish ancestry among the 9181G group were 2 Sephardic (Netherlands, Romania), 4 Ashkenazi (Czech Republic, Lithuania Poland, unknown) and 1 unknown Jewish. Three of the sequences were identical (Ashkenazi Czech Republic, Ashkenazi Poland, Sephardic Netherlands) with mutations of only 9181G in the coding region and 41T in the second control region. Two sequences had an additional coding region mutation of 15787C, one of whom also showed heteroplasmy at a second control region site, G215R. The 6<sup>th</sup> sequence had two completely different coding region mutations along with 9181G. The final sequence had the 9181G mutation along with heteroplasmy at a commonly varying site, C195Y. Note the threshold for detecting heteroplasmy was 25% for sequencing before April 2013 and 20% after April 2013. This leaves a small chance that one of the sequences which was sequenced before the transition date also has heteroplasmy by the newer lower threshold.

A9181G is a non-synonymous mutation with a serine to glycine amino acid change (AGC to GGC) in the ATP synthase F0 subunit 6 of the mitochondrion. This usually highly conserved position in many species has no known disease association (mitomap.org), but has been theorized to be deleterious with an increased production of reactive oxygen speciation (ROS) and resulting in cell death in some nutrient

environments[25]. The other transition found more than once in this cluster, T15787C, is a synonymous mutation in cytochrome B.

Overlap to the coding region mutations found here, including 9181G, to sequences preexisting on Genbank in T2e was limited to the one sequence from a Sephardic Bulgarian that was described earlier in the discovery of the founding mothers of the Sephardic Bulgarian community [11]. In the control region, two Greek sequences [9] harbored 195C as was also found here, but they appear to be from a different branch of T2e without a 41T mutation or any of the coding region mutations. These were assumed to reflect an independent event. One additional T2e sequence with 9181G was reported in the medical literature [25] which described a patient with a pathological heteroplasmic mutation at 4936T. The present 7 sequences along with the one on Genbank (Sephardic Bulgarian) and the pathological sequence from the medical literature were used to create a phylogenetic tree. The tree is presented in Figure 2. New labels for branches include T2e1b and T2e1b1 and are summarized for both this cluster and the previous 2308G cluster in Table 2.

### Time to most recent common ancestor

The maximum genetic distance (MGD) between the most varied pair of the 9181 cluster (Figure 2) was calculated as 6 based on 9 samples producing 36 different pairs. The MGD for the whole 2308 cluster (Figure 1) was slightly older at 7 (10 samples, 45 pairs) and the next oldest node of 15499T along with 2308G, proved to be the same age as the 9181 cluster (MGD=6). All coding region mutations were counted in the calculation based on the assumption that each arose independently. The Sephardic Ottoman-Mexican branch is a relatively recent subclade within the 2308G cluster with an MGD of 0, and differing by a maximum of 2 mutations in the first control region. While only a few full genomic sequences have been obtained for this signature, we are confident that the coding region will show minimal if any variation in further samples.

Haplogroup	Mutation positions
<b>T2e1a</b>	2308G
<b>T2e1a1</b>	15499T
<b>T2e1a1a</b>	41C@-16114T-16192T
T2e1a1a1	16221T
T2e1a1a2 *	16067T
T2e1a1a3 *	16291T
T2e1a1a4	195C
T2e1a1a5?	16114C@ ?
<b>T2e1a1b</b>	247A-4924A-8864C-10682T-16183C-16189C
T2e1a1b1	6570T
T2e1a1c	736T-14180C-16189C
T2e1a2	3744G-7828G-16093C
<b>T2e1b</b>	9181G
<b>T2e1b1</b>	15787C
T2e1b1a	13696G-13803T-13945G
T2e1b1b	215R**
T2e1b2	195Y**
T2e1b3	10321C-13148T
T2e1b4	4936Y**

**Table 2:** Newly defined subclades within mitochondrial DNA haplogroup T2e1. Bold labels refer to haplotypes with at least 3 independent sequences. For geographic origins of sequences, refer to Figures 1 and 2. @ refers to mutation back to ancestral base \* Not verified by full genomic sequencing \*\* Found thus far only with hetero plasmly at this position.

For all of T2e, the MGD is 13, based on 69 samples (present study plus Genbank; pairs=22346), and for branch T2e1, the MGD=11 (sample size=21, pairs=210). Thus, the 9181 and the 2308 branches of T2e can be estimated as having a time to most recent common ancestor that is approximately 46% (6/13) and 54% (7/13) respectively of the age of the larger T2e clade.

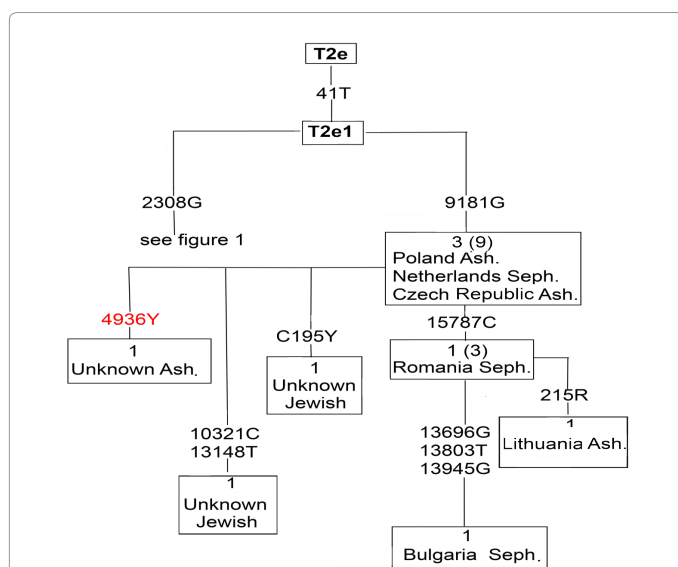
Finally, it is interesting that when the Jewish 9181 cluster was taken together with the 2308 cluster of which the Sephardic signature sequence is a subset, the MGD was found to be 11, which is the same age as all of T2e1. It would be also be the same age as T2e total, save for one sequence from Azerbaijan [15], which increased the MGD for T2e from 11 to 13. A common ancestor for the 9181 and 2308 branches is as old as T2e1 itself.

### Discussion

The work reports on two Jewish signatures in haplogroup T2e that are in nested trees among the oldest of any Jewish-specific mitochondrial groups ever reported.

One of these is defined by 16114T-16192T in the first control region and, as became evident here, a back mutation at position 41 in the second control region. We have found that individuals of Mexican descent share this rare mitochondrial genetic haplotype with Sephardic Jews. A match between these groups was reported previously [8], but was based solely on the control region of the mitochondrial DNA. By finding here an identical match in the coding region from full mitochondrial genomic sequencing, we have ruled out the possibility of a superficial resemblance from unrelated mutational events. Instead, the two groups are part of the same phylogenetic clade and share a common origin.

The finding provides genetic evidence for a Sephardic origin detectable in the modern people of parts of Mexico. This meshes well with historical analysis of the regions of Mexico and South Texas where individuals bearing this Ottoman-Mexican Sephardic signature



**Figure 2:** Branch structure of 9181G cluster in haplogroup T2e mitochondrial DNA based on full genomic sequencing. The number of samples with the exact sequence that defines the node is shown at top of each node (and in parentheses is the total number of sequences harboring the defining mutation—i.e. everything from that node downward in the tree). Known pathological mutation shown in red. Abbreviations: Seph.=Sephardic Jewish; Ash.=Ashkenazi Jewish.

are found. Some of our samples can be traced to Monterrey, Mexico. Monterrey is the center of Nuevo León, the northern region developed by Luis de Carvajal y de la Cueva who was a Portuguese New Christian, or convert from Judaism to Christianity [26]. It has been reported that the founding of Monterrey was with Sephardic inhabitants brought in from Portugal [27], initially the only region of Mexico open to many New Christians, and after, with residents from Mexico City who wished to practice their original religion and hoped to avoid detection by relocating further north [12]. Other samples with the O-M Sephardic signature are from nearby areas from the rest of the state of Nuevo León and the same history applies. Remaining Mexican samples we have found are from the states of Tamaulipas and Coahuila, northern regions also believed to harbor Jewish residents; [28] for example, it has been said that "Secret Jews colonized the states of Nuevo Leon, Coahuila, Tamaulipas and good ole Texas, USA in the 1640's-1680s and thereafter. The majority of Texas's Spanish-speaking immigrants came from Nuevo Leon, Tamaulipas, and Coahuila (the old Nuevo Reyno de Leon) beginning in the 1680s" (Anne de Sola Cardoza, <http://www.sefarad.org/publication/lm/011/texas.html>). Latino Texan individuals bearing this signature, including Roma Texas in the United States, fall into this area as well. Roma Texas is a border town along the Rio Grande that was part of Mexico until 1836, is across from Tamaulipas Mexico and is 95 miles from Monterrey. Conversos from the South Texas Rio Grande Valley are now well known [12]. Northern Mexico and South Texas form a region with a notable Crypto-Jewish history.

While genetic evidence of Jewish 1DNA in Latino populations of the New World has been reported before, the emphasis has been on paternal lineages carried through the Y chromosome and on autosomal DNA, which can also result from exclusively paternal origins. Velez et al. [29] found evidence of Y chromosome haplotypes that trace to European males with Jewish Levantine origins, but mtDNA from local women. This European male-local female connection is a not uncommon pattern reported in the founding of new communities. The present work highlights a European and Jewish maternal origin in addition to the paternal one. Different findings may be location-specific; the Valez et al. study investigated Latino communities in Ecuador and Colorado, USA. The particular regions of Northern Mexico that surfaced with the Ottoman-Mexican Sephardic signature in the present study may be a European female founder hotspot. The first ship in the founding of Nuevo León was described in historical doctrines from Spain which decreed that 60 of the ship's 100 laborers consist of married men with their wives and children [28]. Several of the registered passenger colonists from that 1580 voyage vanished or were executed in "New Spain" by the Inquisition for practicing secret Judaism, which suggests they may have left little DNA trace. However, others survived and additional ships with women and girls most certainly followed.

The present T2e haplotype reflecting maternal European origins is not an isolated incident. The FTDNA Mexico project has uncovered several mitochondrial DNA European haplogroups: H, I, J, K, T and U, V and X. Any connections to Sephardic mitochondrial DNA await further investigation.

It also remains to be determined if the present signature within T2e reflects a founding lineage for Northern Mexico in the Beharian operational sense of constituting greater than 5% of the genetic variation in the modern community [30]. The present signature does meet the criterion of founder type in the sense of a motif in one locale (Iberia) brought to a derived population (Mexico) [6] that presumably originated with only a small fraction of genetic variation present in the source population [31]. While there is much talk of the genetic founders

of a community, it is rare to actually be able to name them. Maria-Clara Flores de Cerda is the mother of more than 20,000 known named descendants. She has descendants living today in whom we still find the Ottoman-Mexican mitochondrial T2e motif with 41@-2308G-15499T-16114T-16192T, as reported here. Maria-Ines de la Cerda y Castro may have brought this otherwise rare Sephardic signature to Mexico from Iberia, but hard proof and details may be lost to history. It gives pause for thought that bits of DNA occurring in a vanishingly small Ottoman Sephardic Jewish population with a big history nonetheless lives and thrives in a community of Catholic Latinos in a new world.

The origin of the Ottoman-Mexican signature appears to be Iberian owing to its exclusive presence in peoples affiliated with Iberia. We have preferred an interpretation that the motif originated and expanded specifically within Sephardim living in Iberia. The fact that it is found in greater numbers in Sephardic and *converso* exiles than in modern Iberians points in this direction. In addition, the mother clade, T2e, from which this signature developed is present in numbers in Sephardim that are closer to those of parts of the Near East, Egypt, and Italy than to its host populations of Spain and Portugal [8]. Sephardic origin would place the time of emergence of the cluster to between 2100 and 500 years ago, the times of arrival of Jews to Iberia during the Diaspora and the time of exit during the forced expulsion, respectively. This estimate is well within the range of proposed mutation rates for the control regions in the scientific literature, which vary considerably. Proposing the most distant common ancestor occurred closer to the beginning of the range (2000 YBP) to allow for the variability of this motif within the Sephardic Ottoman population sets the molecular clock to one mutation every 5810 years (based on rho statistic).

Origin and timing of the larger tree within which this signature is found has proved more elusive. All individuals with the Ottoman-Mexican Sephardic signature also possess coding region mutations of 2308G and 15499T. We have found three branches to possess these two coding region mutations, the signature just described, a second consisting of a lone sample from Morocco, and a third with a curious constituency of Netherlands, colonial America via England and colonial America, uncertain source. Whether these can be traced to the Near East, and perhaps even Judaism, requires further investigation but remains a seductive possibility. Origin of the topmost node 2308G, which currently consists of a single individual, likewise remain unknown. While the Ottoman-Mexican Sephardic signature cluster is new, this entire 2308 cluster is not. We found it to have a genetic distance of 7 between the most divergent of the samples bearing 2308G compared to 13 for all of T2e and 11 for T2e1 (the branch of T2e possessing 41T). Exactly how old this makes the cluster depends on the estimate for T2e, which continues to vary considerably in the literature. The 2308G branch is greater than half the age of T2e.

While we cannot prove that the "Jewishness" of the old 2308 cluster extends beyond the relatively new Ottoman-Mexican clade within, this is not so for the other old cluster reported in this article. It is defined by a coding region mutation of 9181G and was identified by Behar et al. [11] as a founding lineage for Sephardic Ottoman Bulgaria [11]. Through the present full genomic sequencing of seven new samples, along with one on Genbank, and one in the medical literature, several striking aspects became apparent 1) All of those harboring 9181G report Jewish maternal ancestry 2) There is great variability in the additional coding region mutations (Genetic Distance=6) and 3) The mutation is found in both those of Ashkenazi and of Sephardic maternal origin. Taken together, these findings suggest that this is a surprisingly old clade that may well predate the split between Jewish groups. This would be

consistent with both historical and genetic evidence of the common origin of most different Jewish groups [32].

The age of the 9181G cluster is estimated here as at least 46% the age of mother clade T2e. Recent thorough analyses of the entire T haplogroup lead to an estimate to most recent common ancestors for T2e of 8,900 years by Pala et al. [9] considering just the coding regions (and up to 11,000 YBP with control regions); Pike et al. [13] concluded the age of T2e was approximately 26,000 years old. Even using the lowest age estimate would localize the 9181G cluster then to at least 4100 years before present (YBP). While this is certainly before the split between Jewish groups, it is also *too* old. It predates Jewish tribes and may even predate the 3000-year-old Israelites, depending on exact confidence intervals of time estimates. If time estimates are to be taken seriously, this suggests that 9181G should be found in present-day Near East (Jewish and non-Jewish) populations, which thus far has not been reported.

However, 9181 is a hidden cluster because there are no identifying mutations in the control region. Individuals bearing this signature would appear to have the ancestral T2e motif if the first control region alone were inspected. Failure to detect this signature among current Near East populations may be due to relatively infrequent testing of the coding region for Near East populations. (Note we do not think there is sufficient evidence that the presence or absence of 41T in the second control region is a reliable marker for the presence or absence of 9181G in the coding region.) In the population we used to obtain samples, there are numerous underrepresented geographic regions, including southern Europe, Mediterranean, the Caucasus and certainly non-Jewish Near East regions. We can only rule out from the present study that non-Jewish individuals of Central, Northern, and Eastern Europe are not part of the 9181G clade.

The present study also leaves open the possibility that the appearance of this mutation in both Sephardic and Ashkenazi Jewish groups is due to a more recent admixture between the Jewish groups rather than predating their split. For instance, Bulgaria and Romania are two of the ancestral sites of individuals with this mutation and these countries housed both Ashkenazi and Sephardic populations at least since the late 15<sup>th</sup> century. While different Jewish groups have been predominately concentrated in different regions following the Diaspora, they nonetheless crossed paths in lower numbers throughout Europe.

The 9181 mutation from A to G is a non-synonymous mutation involved in ATP synthesis. Adenine at position 9181 is ordinarily highly conserved in mammals and fish. The transition to G is therefore potentially pathological. While there have been no recorded disease associations (mitomap.org), Zhang et al. [25] have argued its effects could be deleterious and lead to increased destructive reaction oxygen species generation. They also note a greater number of patients in their mitochondrial disease clinic that possess the 9181G mutation than they believe to be present elsewhere. Other nearby mutations within the same ATPase6 gene has lead to disease [33]. We have found the 9181G mutation to be part of a thriving cluster, far from rare. It is possible that any deleterious effects of this mutation do not present a problem until older ages when it would not decrease reproductive fitness. It is also possible the mutation is problematic only in combination with other mitochondrial mutations or in selected environments. Since we have found this mutation to be present only in Jewish individuals, the greater number of patients observed at their clinic with the mutation compared to other populations could be due to a greater Jewish representation at their clinic either for uninteresting regional population reasons or because of linkage disequilibrium with other mitochondrial mutations

found among Jewish individuals. Despite being highly conserved, it is also possible a mutation affecting ATP has a counteracting advantage for some environments, like the hypothesized climate-selected mutations present in all of T haplogroups [34]. Further investigation of this ancient mutation may lead to health-related implications for multiple Jewish groups.

Assuming the existence of a molecular clock and currently accepted mutations rates, the age and Jewish affiliation of the 9181G cluster places this clade in the Near East. This is again consistent with both historical records of Jewish origins in the Levant as well as genetic evidence that both Sephardic and Ashkenazi groups have DNA that does not closely match that of the European populations that hosted them for hundreds of years; instead, the DNA can be characterized as falling between Europe and the Near East [32]. It may also be significant that both of the old, highly variable clusters from the present work, 9181G which is Near Eastern and 2308G which contains Near Eastern-derived branches, together have a common ancestor that is as old as T2e1 itself. They account for much of the genetic variability currently seen for all of T2e. The remaining haplotype that adds great variability and age to the T2e subclade is a single sample from Azerbaijan [15], a region in the Caucasus which can also have genetic Near East input. We suggest these considerations should reopen the question of geographic origin of the T2e branch of haplogroup T.

Pala et al. [9] concluded that the origin of T2e was European, just like T2b, the most frequently occurring branch of T2 (the next highest node in the tree). They proposed the Near Eastern presence of T2e was due to back migrations. Bedford [8] similarly suggested Saudi Arabia was likely a recipient of migration rather than the center of expansion but noted a high occurrence of T2e in parts of Saudi Arabia, Egypt, and Italy. This finding suggested that closer scrutiny of these regions be required to determine the origin of T2e. In addition, the work showed the very different geographic distribution patterns of T2e and T2b from the Near East to Europe. For instance, in England and Ireland, T2b comprises more than 4% of the population and T2e less than 0.5%, a pattern that is reversed in the (albeit small) samples of Western Saudi Arabia where T2b falls to 1.4% and T2e climbs to nearly 3%. Vastly different migration patterns would be unexpected if both T2b and T2e have the same origin and age of divergence from the rest of T2, although differing origin locales within Europe may provide a sufficient explanation for the differences.

T2e has been found in a sample of ancient Minoans in Crete dated 3700-4900 [35], Northern Germany from 5300 years ago [24], and in Iceland from pre-Christian remains 1000 years before present (specifically T2e1, indicated by the presence of 41T) [36], where T2e is still prominent today. While interesting, all of these are too recent to help constrain the origin of T2e, already widely dispersed by 5500 years past. Two old subclades of T2e with ties to the Near East, 9181G and 2308G, were investigated in this study in Americans and Mexicans. At present, there is insufficient evidence to distinguish between the possibilities that they stem from a straight-forward origination of T2e in the Near East or from a branch of T2e that migrated back to the Near East early from an origin in southern Europe or the Mediterranean before making its way back to Europe and eventually the New World. And if widely accepted assumptions of a molecular clock and mutation rates change, then other more European-centered alternatives are also back on the table. It may be notable that these two subclades split from each other early, with their common ancestor as old as all of T2e1, a large branch of T2e. Also, might the great similarity of the genes of Yemenite Jews to Saudi Arabians [32]-the latter also with a spike in T2e

occurrence-be revealing? Could the patchiness of T2e's distribution in the Near East-present in some locales but absent in adjoining ones-be reflective of its expansion in specific ethnic groups that were not welcome in the entire country?

A final note is in order: We have intentionally used the revised Cambridge Reference Standard (rCRS) rather than the Reconstructed Sapiens Reference Standard (RSRS) when describing mutations. The RSRS was put forth recently to accurately capture the evolutionary phylogeny of all of mtDNA and provide an objective standard from the beginning of mitochondrial origin [16]. Our selection of the preexisting reference standard instead was not based on the convenience of a familiar sequence. Rather, choosing the rCRS allows one to subtract out those mutations that most sequences from Europe and the Mideast have in common. This is an appropriate choice for population dispersal issues that pertain to a relatively short time scale of a few thousand years rather than ancient ones. In connection with creating the RSRS, over 4000 sequences were deposited to Genbank, most without accompanying geographic origins. While these greatly help flesh out the branch structure of mitochondrial DNA haplotypes, they do so without illumination of branch origins. Perhaps study of phylogeny and population migration may be facilitated if every researcher were to seek out additional information through the adoption of a subclade.

\* New accession numbers for this article in Genbank are KF577586-KF577589, KF564292-564293, KF657641, and KF048033.1

## References

- Richards M, Côté-Real H, Forster P (1996) The emerging tree of West Eurasian mtDNAs, a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 185-203.
- Richards MB, Macaulay VA, Bandelt HJ, Sykes BC (1998) Phylogeography of mitochondrial DNA in Western Europe. *Ann Hum Genet* 62: 241-260.
- Macaulay V, Richards M, Hickey E (1999) The emerging tree of West Eurasian mtDNAs, a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 64: 232-249.
- Torroni A, Huoponen K, Francalacci P (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144: 1835-1850.
- Behar DM, Rosset S, Blue-Smith J (2007) The genographic project public participation mitochondrial DNA database. *PLoS Genetics*.
- Richards M, Macaulay V, Hickey E (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 67: 1251-1276.
- Van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30: 386-394.
- Bedford FL (2012) Sephardic signature in haplogroup T mitochondrial DNA. *Eur J Hum Genet EJHG* 20: 441-448.
- Pala M, Olivieri A, Achilli A (2012) Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *Am J Hum Genet* 90: 915-924.
- Gerber J (1994) A history of the Sephardic experience. Free Press, New York, USA.
- Behar DM, Metspalu E, Kivisild T (2008) Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. *PLoS ONE*.
- Liebman SB (1970) Faith, Flame and the Inquisition. University of Miami, USA.
- Pike DA, Barton TJ, Bauer SL, Kipp EB (2010) mtDNA Haplogroup T phylogeny based on full mitochondrial sequences. *J Genet Geneal* 6: 1-24.
- Pike DA, Genet Geneal TJ (2006) Phylogenetic networks for the human mtDNA haplogroup 2: 1-11.
- Schonberg A, Theunert C, Li M, Stoneking M, Nasidze I (2011) High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: High diversity and demographic inferences. *Eur J Hum Genet* 19: 988-994.
- Behar DM, van Oven M, Rosset S (2012) A "Copernican" reassessment of the human mitochondrial DNA tree from its root. *Am J Hum Genet* 90: 675-684.
- Maca-Meyer N, González AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* 2: 13.
- Leal GG (1995) Origen de los Fundadores de Texas. Coahuila y Nuevo León, Nuevo México.
- Cabohe (2012) La Herencia de Los Longoria. Trafford Publishing, USA.
- Rowold DJ, Luis JR, Terreros MC, Herrera RJ, Mitochondrial DNA (2007) GeneFlow indicates preferred usage of the Levant Corridor over the Horn of Africa passageway. *J Hum Genet* 52: 436-447.
- Saunier JL, Irwin JA, Strouss KM, Ragab H, Sturk KA, et al. (2009) Mitochondrial control region sequences from an Egyptian population sample. *Forensic Sci Int Genet* 3: 97-103.
- Tagliabracchi A, Turchi C, Buscemi L, Sassaroli C (2001) Polymorphism of the mitochondrial DNA control region in Italians. *Int J Legal Med* 114: 224-228.
- Helgason A, Sigureth Ardóttir 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.
- Bramanti B, Thomas MG, Haak W (2009) Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326: 137-140.
- Zhang C, Huang VH, Simon M (2012) Heteroplasmic mutations of the mitochondrial genome cause paradoxical effects on mitochondrial functions. *FASEB J Off Publ Fed Am Soc Exp Biol* 26: 4914-4924.
- Raphael DT (2001) Conquistadores and Crypto-Jews of Monterrey. Carmi House, USA.
- K AMG, Diccionario Porrúa De Historia, Biografía Y, Geografía De Mexico (1976) Cuarta Edición. Editorial Porrúa, S.A.
- Santos RG (2000) Silent heritage: The Sephardim and the colonization of the Spanish North American frontier. *San Antonio* 1492-1600.
- Velez C, Palamara PF, Guevara-Aguirre J (2012) The impact of Converso Jews on the genomes of modern Latin Americans. *Hum Genet* 131: 251-263.
- Behar DM, Metspalu E, Kivisild T (2006) The matrilineal ancestry of Ashkenazi Jewry: Portrait of a recent founder event. *Am J Hum Genet* 78: 487-497.
- Mayr E (1963) Animal species and evolution. Belknap Press of Harvard University Press, Cambridge, UK.
- Behar DM, Yunusbayev B, Metspalu M (2010) The genome-wide structure of the Jewish people. *Nature* 466: 238-242.
- Ogilvie I, Capaldi RA (1999) Mutations of the mitochondrially encoded ATPase 6 gene modeled in the ATP synthase of *Escherichia coli*. *FEBS Lett* 453: 179-182.
- 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.
- Hughey JR, Paschou P, Drineas P (2013) A European population in Minoan Bronze Age Crete. *Nat Commun* 4: 1861.
- Helgason A, Lalueza-Fox C, Ghosh S (2009) Sequences from first settlers reveal rapid evolution in Icelandic mtDNA pool. *PLoS Genet* 5: e1000343.