Sephardic Origins Revealed for Rare Skin Disorder, Recessive Dystrophic Epidermolysis Bullosa, in Individuals Carrying the Unique c.6527insC Mutation

Emily Mira Warshauer[1], Paul A. Maier[2], Goran Runfeldt[2], Ignacia Fuentes[3,4,5], María J. Escámez[6], Laura Valinotto[7,8], Mónica Natale[7], Graciela Manzur[9], Nuria Illera[6], Marta García[6], Marcela del Río[6], Ángeles Mencia[6], Almudena Holguín[6], Fernando Larcher[6,10], Garrett Hellenthal[11], Adam Brown[12], Liliana Consuegra[13], Carolina Rivera[13,14], Inês Nogueiro[15], Jean Tang[16], Anthony Oro[16,17], Peter Marinkovich[17], Francis Palisson[4,5], Matthias Titeux[18,19], Alain Hovnanian[18,19,20], Eli Sprecher[21,22], Karl Skorecki[23,24], David Norris[1,25], Anna Bruckner[25], Igor Kogut[1,25], Ganna Bilousova[1,25], Dennis R. Roop[1,25]

[1]Charles C. Gates Center for Regenerative Medicine, University of Colorado School of Medicine, Anschutz Medical Campus, Aurora, CO

[2]FamilyTreeDNA, Gene by Gene, Houston, Texas, USA

[3]Centro de Genética y Genomica, Facultad de Medicina Clínica Alemana, Universidad del Desarrollo, Santiago, Chile

[4]Fundacion DEBRA Chile, Santiago, Chile

[5]Departamento de Biología Celular y Molecular, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile

[6]Universidad Carlos III de Madrid, Departamento de Bioingeniería (UC3M), Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER-ISCIII), Instituto de Investigación Sanitaria Fundación Jiménez Díaz (IIS-FJD), Madrid, Spain

[7]Center for research in Genodermatosis and Epidermolysis Bullosa (*CEDIGEA*), Argentina

[8]National Scientific and Technical Research Council (*CONICET*), Argentina

[9]Head of the Dermatology Unit, Hospital de Clínicas, UBA, Buenos Aires, Argentina

[10]Epithelial Biomedicine Division, Centre for Energy, Environment and Technology Research (CIEMAT), Madrid

[11]UCL Genetics Institute (UGI), Department of Genetics, Evolution and Environment, UCL, London, United Kingdom

[12]Avotaynu Foundation, Englewood, NJ

[13]Fundacion DEBRA Colombia, Bogotá, Colombia

[14]Department of Medical Genetics, Pediatric Hospital, Fundacion Cardioinfantil-Universidad del Rosario, Bogotá, Colombia

[15]Institute of Molecular Pathology and Immunology of the University of Porto Porto, Portugal ; Faculty of Sciences, University of Porto Porto, Portugal

[16]Department of Dermatology, Stanford University, Stanford, California

[17]Program in Epithelial Biology and Department of Dermatology, Stanford University School of Medicine, Stanford, California

[18]Imagine Institute, Paris Cité University, 75015 Paris, France

[19]Laboratory of Genetic Skin Diseases, INSERM U1163, 75015 Paris, France

[20]Department of Genetics, Necker Hospital for Sick Children, AP-HP, 75015 Paris, France

[21]Department of Dermatology, Tel-Aviv Sourasky Medical Center, Tel Aviv, Israel

[22]Department of Human Molecular Genetics, Sackler Faculty of Medicine, Tel-Aviv University, Tel Aviv, Israel

[23]Department of Genetics & Developmental Biology, Rappaport Faculty of Medicine & Research Institute, Rambam Health Care Campus, Technion-Israel Institute of Technology, Haifa, Israel

[24]Department of Nephrology, Rappaport Faculty of Medicine & Research Institute, Rambam Health Care Campus, Technion-Israel Institute of Technology, Haifa, Israel

[25]Department of Dermatology, University of Colorado School of Medicine, Anschutz Medical Campus, Aurora, CO

Abstract

Background: Recessive Dystrophic Epidermolysis Bullosa (RDEB) is a rare and severe blistering skin disorder caused by loss-of-function mutations in the type VII collagen gene (*COL7A1*). The *COL7A1* c.6527insC mutation is curiously prevalent amongst RDEB individuals and is found worldwide in Europe and the Americas. Previous research has suggested the possibility of a Sephardic Jewish origin of the mutation, however RDEB individuals are not known to have predominant Jewish ancestry.

Methods: In this study, a global cohort of RDEB individuals with the c.6527insC founder mutation from Spain, France, Argentina, Chile, Colombia, and the USA were investigated by autosomal genotyping, pairwise identical-by-descent matching and a local ancestry analysis. Age estimation analysis was performed to determine when Jewish founders introduced the c.6527insC mutation into Iberian and Native American populations (~900 CE and 1492 CE, respectively).

Results: Sephardic ancestry was identified at the haplotype spanning the c.6527insC mutation in 85% of the individuals, despite mixed ancestry elsewhere in the genome and no known recent Sephardic ancestry. Identical-by-descent matching between this RDEB subpopulation and a known crypto-Jewish community in Belmonte, Portugal was also ascertained, providing support for crypto-Jewish ancestry in this RDEB subpopulation.

Conclusion: The identification of this unique RDEB subpopulation unified by the single most prevalent c.6527insC mutation holds great potential to facilitate promising new RDEB therapies using CRISPR Cas 9 gene and base editing. The identification of a single guide RNA allowing efficient and safe editing of this variant would represent a unique drug to treat a large cohort of patients with the same founder mutation.

Background

Recessive Dystrophic Epidermolysis Bullosa (RDEB) is a rare and severe blistering skin disorder caused by mutations in the type VII collagen gene (*COL7A1*) [1–3]. RDEB exhibits pronounced mutational heterogeneity and most *COL7A1* mutations recur quite infrequently [4], yet the *COL7A1* c.6527insC mutation is intriguingly prevalent [5]. While the exact mechanism underlying the c.6527insC mutation has not been fully elucidated, it is known that premature termination codons (PTCs) mediate mRNA decay, leading to null alleles and a deficiency of functional type VII collagen [1, 6]. The great majority of RDEB individuals with the c.6527insC mutation exhibit a very severe phenotype, except for rare cases with a mild phenotype [7]. The recent investigation of several RDEB Hispanic populations from Spain and other Spanish-speaking populations in Europe and the Americas suggested common ancestry, inherited at least in part through Sephardic migration, among individuals carrying the c.6527insC mutation [8].

Jewish people have faced complex waves of migrations during their extensive and dynamic history, resulting in elaborate population genetic patterns, and providing insight into general patterns of health and disease in society [9, 10]. The majority of Sephardic Jews emigrated

from Spain during the time of the Spanish Inquisition to other European countries, North Africa, the Middle East and the Americas, along with a number of pathogenic mutations. Many who did not leave Spain during the Inquisition converted to Catholicism while continuing to observe Judaism in secrecy, creating crypto-Jewish communities. The presence of surviving crypto-Jewish descendants has been established in an earlier study of paternal lineages across the Iberian Peninsula [11].

RDEB individuals carrying the c.6527insC mutation with otherwise unknown Jewish ancestry from Spain, France, Argentina, Chile, Colombia, and the USA were investigated in detail and considerable Sephardic ancestry at the region of the c.6527insC mutation was unambiguously identified. The findings in this study strongly support crypto-Jewish roots as part of RDEB history.

Methods

Participant selection and sample collection of individuals with the c.6527insC mutation

A total of 132 RDEB homozygous and compound heterozygous patient samples from Spain, France, Argentina, Chile, Colombia and the United States with the c.6527insC mutation in *COL7A1* were used in this study. Their genotypes were previously identified using various sequencing technologies and subsequently confirmed by Sanger sequencing (Table S1) [5, 12–14]. Out of these, 126 were used in the final analysis (see "Autosomal genotyping and kinship estimation" below). In addition, five Sephardic individuals (two kept in the final analysis) from the endogamous community of Belmonte, Portugal were included to evaluate the genetic relationship of RDEB-carrying individuals with a uniquely preserved crypto-Jewish community with thriving modern Sephardic people.

Autosomal genotyping and kinship estimation

Autosomal genotyping was performed by Gene by Gene, Ltd., on a customized version of the Infinium Global Screening Array-24 v3.0 BeadChip, and analyzed with the Family Finder autosomal DNA test [15]. This array includes approximately 700,000 SNPs for the 22 pairs of autosomal chromosomes and chromosome X. We used KING v2.2.4 [16] to estimate kinship coefficient between each pair of 137 samples initially considered in the study. Individuals were removed from further analysis if they were third degree relatives or closer, resulting in a final total of 126 c.6527insC-carriers and two Belmonte individuals.

Local ancestry analysis

Local population ancestry was estimated for the 126 unrelated individuals carrying one or more copies of the c.6527insC mutation. Samples were first imputed to the Illumina Human OmniExpress BeadChip array SNP set for compatibility with all reference data. We used IMPUTE2 [17] and the 1000 Genomes [18] reference panel with a union of SNPs from both chip types, and an imputation probability threshold of 0.9. Next, we used Eagle v2.4.1 [19] for long-range phasing of haplotypes, and to impute any remaining no-calls. For a reference panel, we used an extensive collection (104,521) of FamilyTreeDNA (hereafter "FTDNA") customers spanning each major worldwide population. With a genotyping rate of 0.988 and all SNPs present in the phasing panel, we phased and imputed 637,645 SNPs across

chromosomes 1 through 22. One positional Burrows–Wheeler transform (PBWT) phasing iteration using the default auto-selection process.

A reference panel for local ancestry classification was constructed to represent population ancestries from each major continent, as well as various Jewish and non-Jewish populations in Europe. We included nine reference populations in total: Sub-Saharan African, Native American, East Asian, South Asian, North European, South European, Ashkenazi Jewish, Sephardic Jewish from Turkey, and Sephardic Jewish from Morocco (Table 1) [18, 20–23]. Sample size disparities can adversely influence machine learning accuracy, so we randomly chose an approximately equal number of samples from each population. The two smallest sample sizes available were Turkish Sephardic (n=53) and Moroccan Sephardic (n=38), and inordinately small sample sizes can also be problematic. Therefore, we compromised on a sample size of 53 for all populations and used all available samples of Moroccan Sephardic.

We used a method similar to RFmix [24] for local classification but adapted by FTDNA for their proprietary tool myOrigins v3.0 [23]. This method for classifying phased segments into populations was found to be more accurate than RFmix, and the subsequent step of smoothing out classification errors with a conditional random field is identical to that described in RFmix. We classified each 500 SNP segment, using a sliding window to move each overlapping segment across each chromosome in increments of 200 SNPs. A hidden Markov model (HMM) was used to correct switch errors implicit in imperfect phasing, by predicting the true (hidden) diploid phase of each maternal and paternal segment, given the observed order of population labels [23]. The transition probabilities used were as follows: no strand flip (0.850; e.g., 1/2 to 1/2), strand flip (0.000; e.g., 1/2 to 2/1), partial overlap (0.128; e.g., 1/2 to 1/3), partial overlap and strand flip (0.017; e.g., 1/2 to 3/1), other (0.005; e.g., 1/2 to 3/4). Results were processed and plotted using custom scripts with R v4.1.2 [25].

Pairwise identical-by-descent matching

We used the Family Finder algorithm to perform a SNP-wise comparison for each pair of all 128 individuals to find shared segments, defined as runs of SNPs sharing at least one allele. Seed segments were formed from at least 900 adjacent matching SNPs, and extended in both directions, to merge with adjacent segments. Unqualified segments were then discarded, if they did not contain at least 480 SNPs, a genetic distance of 2 centimorgans (cM), or a density of 105 SNPs/cM. We used the linkage map (build 37) from the 1000 Genomes Project [18]. Given the potentially very distant identical-by-descent (IBD) matches expected amongst descendants of a centuries-old Jewish diaspora, we used custom filters to determine whether two individuals "match." A match is defined as two people with at least 5 cM in common and allowing for shared segments ≥ 2 cM. We summarized pairwise matches overall, and those with segments overlapping the c.6527insC mutation, in R v4.1.2 [25].

Triangulation

We used triangulation as an additional validation to more conservatively assess IBD matching between all 128 samples. Triangulation [26] is the transitive property of shared DNA: segments that are shared IBD between three or more individuals from a recent common ancestor should match pairwise between all three, at an overlapping genetic

location. Any segments with a total overlap between group members of < 2 cM were removed. We constructed a dendrogram based on a hierarchical structure analysis of total shared triangulated DNA to find clusters of more closely related individuals. This assumes all individuals have recent common ancestry through just one genealogical line, through which they inherited the c.6527insC mutation. We used the hclust and dendrogram functions in base R, with the complete linkage method to find similar clusters [25]. Distances were computed as $1 - (cM/cM_{max})$.

Age estimates for c.6527insC-containing haplotypes

We used two independent methods to estimate the date that c.6527insC was introduced, first from Sephardic people into Iberia, and later into the Americas. First, we applied the "Gamma" method [27, 28]. The basic theory is that, under the Haldane recombination model [29], the length between recombination breakpoints is a random Poisson process, and exponentially distributed with rate of 1.0. This simple model does not account for interference (inhibition of one crossover by another) so is not perfectly accurate. However, at shorter genetic distances, the likelihood (and effect) of multiple recombination events is negligible. The logic is as follows: a mutation is inherited within a haplotype, which is iteratively broken into smaller segments with each successive meiosis. The mean distance from mutation to one of its two flanking recombination breakpoints is $1/\tau$ Morgans, or $100/\tau$ cM, where τ is the time to most recent common ancestor (TMRCA) in units of generations. Note: the TMRCA for a group of mutation-carriers may be (substantially) younger than the TMRCA of the mutation itself, since it only considers a small subset of all mutation-carriers, past and present. Therefore, the length of the entire segment is distributed as $Gamma(2, \tau)$, with expected length $2/\tau$ Morgans, or $200/\tau$ cM.

The method of Gandolfo[28] corrects for the bias of small sample size of segment lengths l_i, with a bias correction factor, $b(n) = (2n - 1)/2n$. Using this correction, they estimate the TMRCA as:

$$\hat{\tau} = \frac{b(n)2n}{\sum_{i=1}^{n} l_i}$$

Exact confidence intervals are then calculated from $\tau/\hat{\tau}$. This calculation assumes that all mutation-carriers are unrelated ("independent") since their TMRCA, i.e., no close cousin relationships, which would break the distributional assumption of independent and identically distributed samples. A second "correlated" version of the calculation accounts for the possibility of reticulating history, using the bias correction factor $b(n*) = 2n/(1 + (n - 1)\rho)$, where ρ is the mean pairwise correlation between segment lengths, and $\bar{l}$ and s^2 are respectively the mean and variance of segment lengths:

$$\hat{\rho} = \frac{n\bar{l} - s^2(2n + 1)}{n\bar{l} + s^2(n - 1)}$$

For the Sephardic-Spanish time estimate (T_1), we used the population segment edges derived from myOrigins v3.0. Segment boundaries were defined as start and end points of a Sephardic segment spanning the mutation at position 48,611,297 on chr3 [30]. For compound heterozygotes, we only used the haplotype carrying the c.6527insC mutation. This haplotype was determined by IBD matching to c.6527insC-homozygous individuals and selecting the best match (i.e., having the fewest nucleotide mismatches). For the American time estimate (T_2), we used the maximum of pairwise IBD matching segments derived from Family Finder. Only segments from New World individuals were used for T_2.

The second age estimate method we applied was Alder v1.03 [31], which uses patterns of linkage disequilibrium (LD) decay to compute time since an admixture event. Alder builds upon previous methods [32, 33] which show that admixture LD scales with time since admixture (t), genetic distance (d), and the initial difference (D_0) in allele frequencies between mixing populations:

$$D_t \approx e^{-td} D_0$$

The method fits a least-squares curve to patterns of exponential LD decay, to solve for the number of generations since admixture, and then calculates confidence intervals by jackknifing each chromosome.

We used Alder to estimate T_1 as the admixture date between Sephardic and Iberian references, with Old World RDEB individuals as the admixed population. We then estimated T_2 as the admixture date between Native American references and Old World RDEB individuals, with New World RDEB individuals as the admixed population. For both estimations, we considered a range of genetic distances between 0.05 and 50 cM. We converted g generations with generation time to years before present as $\times (g + 1)$, to account for the unknown age of participants.

Importantly, each of the events we tested (T_1 and T_2) actually consists of two potentially different time periods: (a) the TMRCA of an IBD segment, and (b) the potentially later time of population admixture. For example (T_2), the ancestral population of Sephardic-Iberian mutation donors could have lived in Europe for some time prior to colonization of the Americas, and subsequent intermixing with Native Americans. Gamma gets at (a) whereas Alder gets at (b). However, (a) and (b) should converge on the same time period if the population size of donors was small. Although we hereafter refer to T_1 and T_2 as each being a single time period, we note the possibility that Gamma could overestimate the admixture time compared to Alder.

Results

Local ancestry analysis

myOrigins v3.0 analysis found a strong signal of Sephardic population ancestry spanning the c.6527insC mutation (Fig. 2). Out of 164 haplotypes carrying the mutation (38

homozygotes; 88 heterozygotes), 136 (83%) were found to be Sephardic, and 142 (87%) were found to be Jewish more generally. In contrast, only 15% (95% CI: 8–22%) of the entire genome was Sephardic, and only 22% (95% CI: 13–30%) of the entire genome was Jewish (yellow line in Fig. 2b). Many heterozygotes did not have overlapping Sephardic ancestry on their second haplotype (Fig. S1), which included a range of other RDEB-causing mutations (Table S1), suggesting that the c.6527insC mutation in particular shows this pattern. As expected, the main difference between New and Old-World individuals was a high prevalence of Native American segments in the New World, but also higher Iberian ancestry. No other Sephardic spikes of ancestry were observed elsewhere in the genome (Fig. 2b). This is consistent with our two-step model of Sephardic heritage for this mutation (Fig. 1b).

Pairwise identical-by-descent matching

Out of 8,128 possible pairs of matching participants ($128 \times 127/2$), 1,203 pairs (15%) were found to have matching segments after applying a 5 cM minimum IBD threshold (Fig. 3). Given the potentially long time elapsed since the cohort shared a common ancestor, and the wide variance of IBD overlap after just a few generations, many null matches were expected. Out of 15% of matching pairs, 1,035 of 1,203 (86%) matched at a segment spanning the c.6527insC mutation. We performed a one-sample Wilcoxon ranked-sign test to determine that 1,035 matches spanning c.6527insC is significantly higher than the genome-wide mean of 30 or 3% ($W = 0$, $p < 0.005$, $n = 100$). The Chilean samples showed the highest incidence of pairwise matching, probably owing to higher post-Columbian shared ancestry and/or endogamy. Notably, the two included Belmonte Sephardic samples (with no mutation; last two columns of Fig. 3) matched many mutation-carrying samples (with no previously known Sephardic ancestry).

Triangulation

We found a general pattern of autosomal clustering within countries of origin for most study participants (Fig. 4). Proximate countries in the New World (Argentina, Chile, and Colombia) and Old World (Spain, Portugal, France) were found on many adjacent branches of the hierarchical tree, suggesting shared post-Columbian ancestry between geographically close individuals. Interestingly, one French individual was found nested in a clade of Chilean individuals (BP52019) whereas 1 Chilean (MK40783) and 2 Argentinean samples were found nested in Spanish clades (MK65370 and MK65374). Most New World (particularly Chilean) clades were much more recently related than clades in other countries, consistent with our IBD results.

Age estimates for c.6527insC-containing haplotypes

Both Gamma and Alder methods of age estimation supported a pre-Columbian introduction of c.6527insC-containing haplotypes into Iberia (T_1), and subsequent post-Columbian introduction by this admixed population into the Americas (T_2; Table 2). Given the evidence from IBD and triangulation analysis of recent shared ancestry in the Americas, we favor the "correlated" genealogy approach of the Gamma method, however we report "independent" values for completeness. Generation times have varied considerably between eras and

cultures, and we report several calendar dates based on generation times of 20, 25, and 30 years per generation. However, here we assume 30 years, because recent estimates suggest a value between 26–30 [34].

The Gamma method estimated T_1 to occur in 990 (95% CI: 371–1610) CE, whereas Alder estimated an earlier time of 767 (95% CI: 308–1226) CE. Although there is a 200-year difference in these mean estimates, their 95% confidence intervals are highly overlapping (Fig. S2), and both estimates coincide with the so-called Golden Age of Jews in Spain. As noted in the methods, the Gamma method estimates the TMRCA of an IBD segment, whereas Alder estimates the potentially later time of population admixture. Convergent estimates by these two approaches likely implies a small donor population size.

The Gamma method estimated T_2 to occur in 1702 (95% CI: 1539–1859) CE, whereas Alder estimated an earlier time of 1665 (95% CI: 1650–1679) CE. These estimates are more consistent and have tighter confidence intervals, but are noticeably later than the post-Columbian era beginning in 1492 CE. This leaves the possibility that Jews carrying the c.6527insC mutation in Europe first assimilated there before migrating to the Americas generations later. This is consistent with our hypothesis (Fig 1b) and supported by the fact that more European than Native American segments flank the c.6527insC segment (Fig. 2a). A one-tailed binomial test confirmed that flanking segments to the mutation were more European than expected by chance ($p=2.62\times10^{-12}$).

Conclusion

Genetics, culture, history and religion have unified the Jewish people since their origins in the Middle East more than 5,000 years ago [35]. The c.6527insC mutation in modern Jews originally dates to ~1300 BCE with successive admixture amongst ancient Iberians occurring in ~900 CE and with Native Americans ~1492 CE (Fig. 1a and 1b). It is still present in Spain today. Endogamy historically predominated in Jewish populations, giving rise to a high prevalence of genetic diseases. Previously unknown as prevalent in Jewish populations, RDEB was only recently recognized to have an association with Sephardic communities [8]. The RDEB population in this study from Spain, France, Argentina, Chile, Colombia, and the USA, previously unknown to have Jewish origins, unambiguously demonstrate considerable Sephardic ancestry at the region surrounding the c.6527insC mutation (Fig. 2a & 2b). An unequivocal Sephardic lineage in this population can be explained by shared ancestors who underwent forced conversions to Catholicism on the Iberian Peninsula during the Spanish Inquisition. Crypto-Jewish communities emerged at this time of massive persecution so families could secretly maintain their Jewish faith while outwardly professing adherence to Catholicism. Descendants of crypto-Jews frequently lost their Jewish identity over time, however genetic diseases present in these communities consistently survived. The c.6527insC mutation undoubtedly has Jewish origins, arriving from the Middle East ~900 CE (estimates ranging from 767–1101 CE; Table 2) during the "golden age", when Jewish life was flourishing on the Iberian Peninsula (Fig. 1b). Thus, our results are broadly consistent with our hypothesized time course (Fig. 1b). The mixing of RDEB individuals with Native American individuals has estimated dates ranging from

1665–1725 CE (Table 2), during a post-Columbian era that saw the emergence of many crypto-Jewish communities.

The Spanish Inquisition is among the fiercest examples of a system enforcing religious intolerance and propagating genetic homogeneity, with consequences clearly still relevant to our world today [36]. It is estimated that persecuted Jewish individuals during the Inquisition represented approximately one-third of early Spanish immigrants and the propagation of a rare genetic disease, such as RDEB, in Iberian exile communities can be attributed, at least in part, to a founder effect [37, 38]. To date, there is scarce knowledge about Sephardic crypto-Jewish descendants, due to limited data from the original source of Sephardic Jews on the Iberian Peninsula combined with the challenge of identifying the existent communities today [39]. Communities in Portugal (Belmonte and Bragança) and Mallorca (Chueta) have known crypto-Jewish ancestry. Interestingly, individuals from the Belmonte community were shown to distantly match RDEB individuals carrying the c.6527insC mutation, substantiating crypto-Jewish ancestry (Fig. 3 & 4).

The migration patterns of RDEB individuals with the c.6527insC mutation notably resemble the Sephardic Diaspora map after 1492 CE (Fig. 1a). While Sephardic ancestry predominates overwhelmingly at the region surrounding the c.6527insC mutation in this RDEB population, traces of Ashkenazi ancestry are also evident (Fig. 2a). The combination of both Sephardic and Ashkenazi ancestry at this region, suggests that the mutation predates Sephardic and Ashkenazi admixture during the Middle Ages [10]. Jewish people have lived on the Iberian Peninsula as far back as the Roman Empire and the c.6527insC mutation can be traced back even further to pre-Roman communities on the Iberian Peninsula [14].

Waves of Jewish people, including RDEB ancestors, migrated to the Americas from Europe after 1492 CE and greatly shaped the modern day Latin American population structure [37]. The European, Native American and Sephardic admixture represented by the c.6527insC mutation reflects the migration of Jewish-Iberian people escaping the Iberian Peninsula during the Inquisition [38, 40]. The c.6527insC mutation is widespread in the Americas and recently was identified as the most frequent RDEB variant in Brazil [41]. While many individuals did leave for the Americas, the majority in this RDEB subpopulation are Spanish, indicating a strong example of Jewish families remaining in Spain throughout the duration of the Inquisition. Many of those individuals who did leave Spain, settled in remote areas of Europe and the Americas where they were less likely to be discovered and exposed. A number of mutations were brought to the Americas via these routes, including the pathogenic growth hormone receptor (GHR) mutation in Laron syndrome, observed in the isolated Lojano community in Ecuador, a population known to have the influence of Sephardic crypto-Jewish ancestry [38].

One particular region in North America well-known to attract immigrants who wished to maintain their Jewish practices under less scrutiny in the New World while demonstrating a veneer of Catholicism was the isolated San Luis Valley in northern New Mexico and southern Colorado [42]. This region of the USA supports a population structure with greater frequencies of genetic diseases well-known in Jewish populations, including Pemphigus Vulgaris, Bloom Syndrome and BRCA1/BRCA2 associated breast cancer, suggesting those

fleeing the Inquisition may have brought Jewish founder mutations to contemporary Hispanic populations [40, 43]. The 185delAG mutation in BRCA1 was identified at a surprisingly high frequency in non-Jewish Colorado families originating in the San Luis Valley and similar to the RDEB population, were found to have Jewish ancestry [43, 44].

Interestingly, one of the RDEB individuals is heterozygous for c.6527insC and c.7485+5G>A, a mutation predominantly found in the San Luis Valley, harboring multiple RDEB mutations potentially integrated in crypto-Jewish communities. Further studies will elucidate the extent of RDEB mutations associated with Jewish ancestry.

This unique group of RDEB individuals carrying the c.6527insC mutation unambiguously exhibits collective Sephardic ancestry, and may also represent the largest set of RDEB individuals ever reported with a single mutation in an otherwise remarkably heterogeneous disease. The recognition of this unique RDEB sub-population highlights the patterns of this rare genetic disease and illuminates the genetic architecture of the Sephardic Jewish population. Furthermore, recognition of the RDEB subpopulation unified by the single most prevalent c.6527insC mutation will enhance the efficient implementation of CRISPR Cas 9 gene and base editing therapies. The identification of a single guide RNA allowing efficient and safe editing of this variant would represent a unique drug to treat a large cohort of patients with the same founder mutation. At present, the therapies accomplished in preclinical settings, including the *ex vivo* and *in vivo* correction of c.6527insC by multiple gene editing strategies, cultivate great optimism for the future of promising treatments for RDEB and other rare diseases [45–50].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Data Availability Statement

Data are available upon reasonable request.

References

1. Hovnanian A, Rochat A, Bodemer C, et al. Characterization of 18 new mutations in COL7A1 in recessive dystrophic epidermolysis bullosa provides evidence for distinct molecular mechanisms underlying defective anchoring fibril formation. Am J Hum Genet 1997; 61: 599–610. 1997/10/27. DOI: 10.1086/515495. [PubMed: 9326325]

2. Has C, Bauer JW, Bodemer C, et al. Consensus re-classification of inherited epidermolysis bullosa and other disorders with skin fragility. Br J Dermatol 2020 2020/02/06. DOI: 10.1111/bjd.18921.

3. Bardhan A, Bruckner-Tuderman L, Chapple ILC, et al. Epidermolysis bullosa. Nat Rev Dis Primers 2020; 6: 78. 20200924. DOI: 10.1038/s41572-020-0210-0. [PubMed: 32973163]

4. van den Akker PC, Jonkman MF, Rengaw T, et al. The international dystrophic epidermolysis bullosa patient registry: an online database of dystrophic epidermolysis bullosa patients and their COL7A1 mutations. Hum Mutat 2011; 32: 1100–1107. 2011/06/18. DOI: 10.1002/humu.21551. [PubMed: 21681854]

5. Escamez MJ, Garcia M, Cuadrado-Corrales N, et al. The first COL7A1 mutation survey in a large Spanish dystrophic epidermolysis bullosa cohort: c.6527insC disclosed as an unusually recurrent mutation. Br J Dermatol 2010; 163: 155–161. 2010/02/27. DOI: 10.1111/j.1365-2133.2010.09713.x. [PubMed: 20184583]

6. Uitto J, Hovnanian A and Christiano AM. Premature termination codon mutations in the type VII collagen gene (COL7A1) underlie severe recessive dystrophic epidermolysis bullosa. Proc Assoc Am Physicians 1995; 107: 245–252. [PubMed: 8624860]

7. Chacon-Solano E, Leon C, Carretero M, et al. Mechanistic interrogation of mutation-independent disease modulators of RDEB identifies the small leucine-rich proteoglycan PRELP as a TGF-beta antagonist and inhibitor of fibrosis. Matrix Biol 2022; 111: 189–206. 20220630. DOI: 10.1016/j.matbio.2022.06.007. [PubMed: 35779740]

8. Warshauer EM, Brown A, Fuentes I, et al. Ancestral patterns of recessive dystrophic epidermolysis bullosa mutations in Hispanic populations suggest sephardic ancestry. Am J Med Genet A 2021; 185: 3390–3400. 2021/08/27. DOI: 10.1002/ajmg.a.62456. [PubMed: 34435747]

9. Behar DM, Metspalu E, Kivisild T, et al. Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. PLoS One 2008; 3: e2062. 2008/05/01. DOI: 10.1371/journal.pone.0002062. [PubMed: 18446216]

10. Oddoux C, Guillen-Navarro E, Ditivoli C, et al. Mendelian diseases among Roman Jews: implications for the origins of disease alleles. J Clin Endocrinol Metab 1999; 84: 4405–4409. 1999/12/22. DOI: 10.1210/jcem.84.12.6268. [PubMed: 10599695]

11. Adams SM, Bosch E, Balaresque PL, et al. The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. Am J Hum Genet 2008; 83: 725–736. 2008/12/09. DOI: 10.1016/j.ajhg.2008.11.007. [PubMed: 19061982]

12. Natale MI, Manzur GB, Lusso SB, et al. Analysis of COL7A1 pathogenic variants in a large cohort of dystrophic epidermolysis bullosa patients from Argentina reveals a new genotype-phenotype correlation. Am J Med Genet A 2022; 188: 3153–3161. 2022/08/19. DOI: 10.1002/ajmg.a.62957. [PubMed: 35979658]

13. Cuadrado-Corrales N, Sanchez-Jimeno C, Garcia M, et al. A prevalent mutation with founder effect in Spanish Recessive Dystrophic Epidermolysis Bullosa families. BMC Med Genet 2010; 11: 139. 2010/10/06. DOI: 10.1186/1471-2350-11-139. [PubMed: 20920254]

14. Sanchez-Jimeno C, Cuadrado-Corrales N, Aller E, et al. Recessive dystrophic epidermolysis bullosa: the origin of the c.6527insC mutation in the Spanish population. Br J Dermatol 2013; 168: 226–229. 2012/07/05. DOI: 10.1111/j.1365-2133.2012.11128.x. [PubMed: 22757647]

15. Hu R, Maier P, Runfeldt G, et al. FAMILY FINDER MATCHING 5.0 Matching Algorithm and Relationship Estimation. 2021.

16. Manichaikul A, Mychaleckyj JC, Rich SS, et al. Robust relationship inference in genome-wide association studies. Bioinformatics 2010; 26: 2867–2873. 20101005. DOI: 10.1093/bioinformatics/btq559. [PubMed: 20926424]

17. Howie BN, Donnelly P and Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 2009; 5: e1000529. 20090619. DOI: 10.1371/journal.pgen.1000529. [PubMed: 19543373]

18. Siva N 1000 Genomes project. Nat Biotechnol 2008; 26: 256. DOI: 10.1038/nbt0308-256b.

19. Loh PR, Palamara PF and Price AL. Fast and accurate long-range phasing in a UK Biobank cohort. Nat Genet 2016; 48: 811–816. 20160606. DOI: 10.1038/ng.3571. [PubMed: 27270109]

20. Eichstaedt CA, Antao T, Pagani L, et al. The Andean adaptive toolkit to counteract high altitude maladaptation: genome-wide and phenotypic analysis of the Collas. PLoS One 2014; 9: e93314. 20140331. DOI: 10.1371/journal.pone.0093314. [PubMed: 24686296]

21. Mallick S, Li H, Lipson M, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature 2016; 538: 201–206. 20160921. DOI: 10.1038/nature18964. [PubMed: 27654912]

22. Pagani L, Lawson DJ, Jagoda E, et al. Genomic analyses inform on migration events during the peopling of Eurasia. Nature 2016; 538: 238–242. 20160921. DOI: 10.1038/nature19792. [PubMed: 27654910]

23. Maier P, Hu R, Runfeldt G, et al. MYORIGINS 3.0: Combining Global and Local Methods for Determining Population Ancestry. 2021.

24. Maples BK, Gravel S, Kenny EE and Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am J Hum Genet 2013; 93: 278–288. 2013/08/06. DOI: 10.1016/j.ajhg.2013.06.020. [PubMed: 23910464]

25. Team RC. R: a language and environment for statistical computing, https://www.r-project.org (2022).

26. Ferrando-Bernal M, Morcillo-Suarez C, de-Dios T, et al. Mapping co-ancestry connections between the genome of a Medieval individual and modern Europeans. Sci Rep 2020; 10: 6843. . DOI: 10.1038/s41598-020-64007-2. [PubMed: 32321996]

27. McPeek MS and Strahs A. Assessment of linkage disequilibrium by the decay of haplotype sharing, with application to fine-scale genetic mapping. Am J Hum Genet 1999; 65: 858–875. DOI: 10.1086/302537. [PubMed: 10445904]

28. Gandolfo LC, Bahlo M and Speed TP. Dating rare mutations from small samples with dense marker data. Genetics 2014; 197: 1315–1327. 20140530. DOI: 10.1534/genetics.114.164616. [PubMed: 24879464]

29. Haldane JBS. The combination of linkage values, and the calculation of distances between the loci of linked factors. Journal of Genetics 1919; 8: 299–309.

30. The Human Gene Mutation Database.

31. Loh PR, Lipson M, Patterson N, et al. Inferring admixture histories of human populations using linkage disequilibrium. Genetics 2013; 193: 1233–1254. 20130214. DOI: 10.1534/genetics.112.147330. [PubMed: 23410830]

32. Moorjani P, Patterson N, Hirschhorn JN, et al. The history of African gene flow into Southern Europeans, Levantines, and Jews. PLoS Genet 2011; 7: e1001373. 20110421. DOI: 10.1371/journal.pgen.1001373. [PubMed: 21533020]

33. Patterson N, Moorjani P, Luo Y, et al. Ancient admixture in human history. Genetics 2012; 192: 1065–1093. 20120907. DOI: 10.1534/genetics.112.145037. [PubMed: 22960212]

34. Moorjani P, Sankararaman S, Fu Q, et al. A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. Proc Natl Acad Sci U S A 2016; 113: 5652–5657. 20160502. DOI: 10.1073/pnas.1514696113. [PubMed: 27140627]

35. Atzmon G, Hao L, Pe'er I, et al. Abraham's children in the genome era: major Jewish diaspora populations comprise distinct genetic clusters with shared Middle Eastern Ancestry. Am J Hum Genet 2010; 86: 850–859. 2010/06/22. DOI: 10.1016/j.ajhg.2010.04.015. [PubMed: 20560205]

36. Drelichman M, Vidal-Robert J and Voth HJ. The long-run effects of religious persecution: Evidence from the Spanish Inquisition. Proc Natl Acad Sci U S A 2021; 118 2021/08/15. DOI: 10.1073/pnas.2022881118.

37. Velez C, Palamara PF, Guevara-Aguirre J, et al. The impact of Converso Jews on the genomes of modern Latin Americans. Hum Genet 2012; 131: 251–263. 2011/07/27. DOI: 10.1007/s00439-011-1072-z. [PubMed: 21789512]

38. Goncalves FT, Fridman C, Pinto EM, et al. The E180splice mutation in the GHR gene causing Laron syndrome: witness of a Sephardic Jewish exodus from the Iberian Peninsula to the New World? Am J Med Genet A 2014; 164A: 1204–1208. 2014/03/26. DOI: 10.1002/ajmg.a.36444. [PubMed: 24664892]

39. Nogueiro I, Teixeira JC, Amorim A, et al. Portuguese crypto-Jews: the genetic heritage of a complex history. Front Genet 2015; 6: 12. 2015/02/24. DOI: 10.3389/fgene.2015.00012. [PubMed: 25699075]

40. Ostrer H The origin of the p.E180 growth hormone receptor gene mutation. Growth Horm IGF Res 2016; 28: 51–52. 2015/08/19. DOI: 10.1016/j.ghir.2015.08.003. [PubMed: 26277320]

41. Ota VK, Mariath LM, Torrelio RMF, et al. Genetic Profiling of Epidermolysis Bullosa in a Large Brazilian Cohort. Clin Exp Dermatol 2025 20250724. DOI: 10.1093/ced/llaf303.

42. Alberro S Crypto-Jews and the Mexican Holy Office in the Seventeenth Century. The Jews and the expansion of Europe to the West 1450–1800 2001; 2.

43. Mullineaux LG, Castellano TM, Shaw J, et al. Identification of germline 185delAG BRCA1 mutations in non-Jewish Americans of Spanish ancestry from the San Luis Valley, Colorado. Cancer 2003; 98: 597–602. 2003/07/25. DOI: 10.1002/cncr.11533. [PubMed: 12879478]

44. Makriyianni I, Hamel N, Ward S, et al. BRCA1:185delAG found in the San Luis Valley probably originated in a Jewish founder. J Med Genet 2005; 42: e27. 2005/05/03. DOI: 10.1136/jmg.2004.029785. [PubMed: 15863659]

45. Garcia M, Bonafont J, Martinez-Palacios J, et al. Preclinical model for phenotypic correction of dystrophic epidermolysis bullosa by in vivo CRISPR-Cas9 delivery using adenoviral vectors. Mol Ther Methods Clin Dev 2022; 27: 96–108. 20220916. DOI: 10.1016/j.omtm.2022.09.005. [PubMed: 36212909]

46. Bonafont J, Mencia A, Chacon-Solano E, et al. Correction of recessive dystrophic epidermolysis bullosa by homology-directed repair-mediated genome editing. Mol Ther 2021; 29: 2008–2018. 20210218. DOI: 10.1016/j.ymthe.2021.02.019. [PubMed: 33609734]

47. Bonafont J, Mencia A, Garcia M, et al. Clinically Relevant Correction of Recessive Dystrophic Epidermolysis Bullosa by Dual sgRNA CRISPR/Cas9-Mediated Gene Editing. Mol Ther 2019; 27: 986–998. 2019/04/02. DOI: 10.1016/j.ymthe.2019.03.007. [PubMed: 30930113]

48. Mencía Á, Chamorro C, Bonafont J, et al. Deletion of a Pathogenic Mutation-Containing Exon of COL7A1 Allows Clonal Gene Editing Correction of RDEB Patient Epidermal Stem Cells. Mol Ther Nucleic Acids 2018; 11: 68–78. 2018/06/03. DOI: 10.1016/j.omtn.2018.01.009. [PubMed: 29858091]

49. Peking P, Koller U, Duarte B, et al. An RNA-targeted therapy for dystrophic epidermolysis bullosa. Nucleic Acids Res 2017; 45: 10259–10269. DOI: 10.1093/nar/gkx669. [PubMed: 28973459]

50. Chamorro C, Mencia A, Almarza D, et al. Gene Editing for the Efficient Correction of a Recurrent COL7A1 Mutation in Recessive Dystrophic Epidermolysis Bullosa Keratinocytes. Mol Ther Nucleic Acids 2016; 5: e307. 20160405. DOI: 10.1038/mtna.2016.19. [PubMed: 27045209]

What is already known on this topic:

The origins of RDEB mutations have not been precisely identified and common Sephardic ancestry has been suggested.

What this study adds:

Our study elucidates common Sephardic ancestry for RDEB individuals carrying the c.6527insC mutation in Spain, France, Argentina, Chile, Colombia, and the USA.

How this study might affect research, practice or policy:

We report the most comprehensive study to date of RDEB individuals carrying a single mutation (c.6527insC) with a unique shared history, findings which hold great potential to accelerate promising new RDEB therapies.
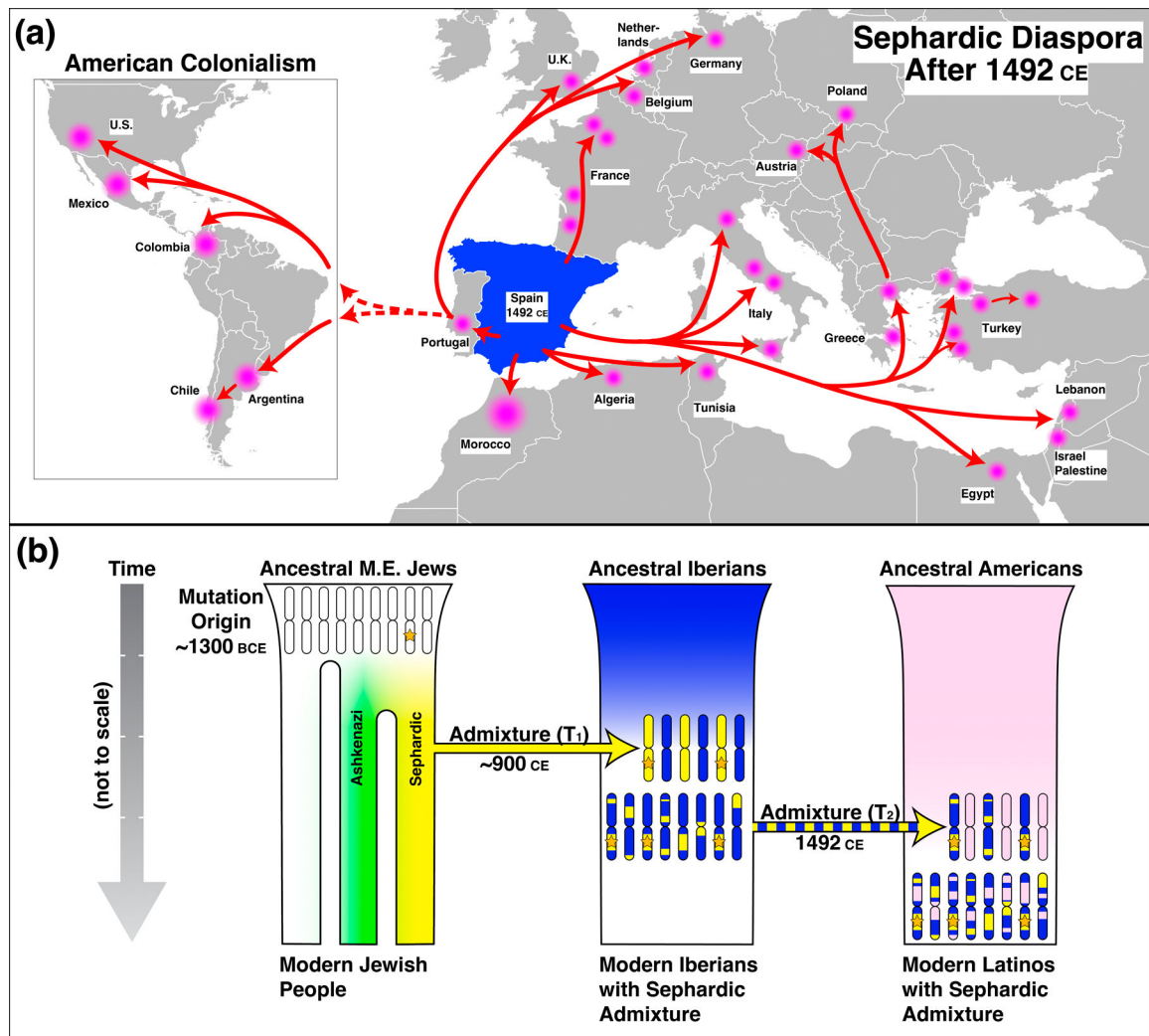
Figure 1.

Sephardic Local Ancestry Spans c.6527insC mutation. (A) Results of myOrigins V.3.0 analysis showing local and overall population ancestry for chromosome 3. All 126 unrelated carriers of the c.6527insC mutation are shown for the haplotype that is most Sephardic at position 48,611,297. Turkish and Moroccan Sephardic ancestry are merged into one group for display due to lower sample sizes and genetic similarity. Country of origin is denoted by rainbow colours, and genotype is denoted by circles. (B) Genome-wide population ancestry aggregated across all 126 samples. Both haplotypes are included for homozygotes, and one haplotype (best match to homozygotes) is included for heterozygotes. A peak of 83% Sephardic ancestry is found at chr3, position 48,611,297.
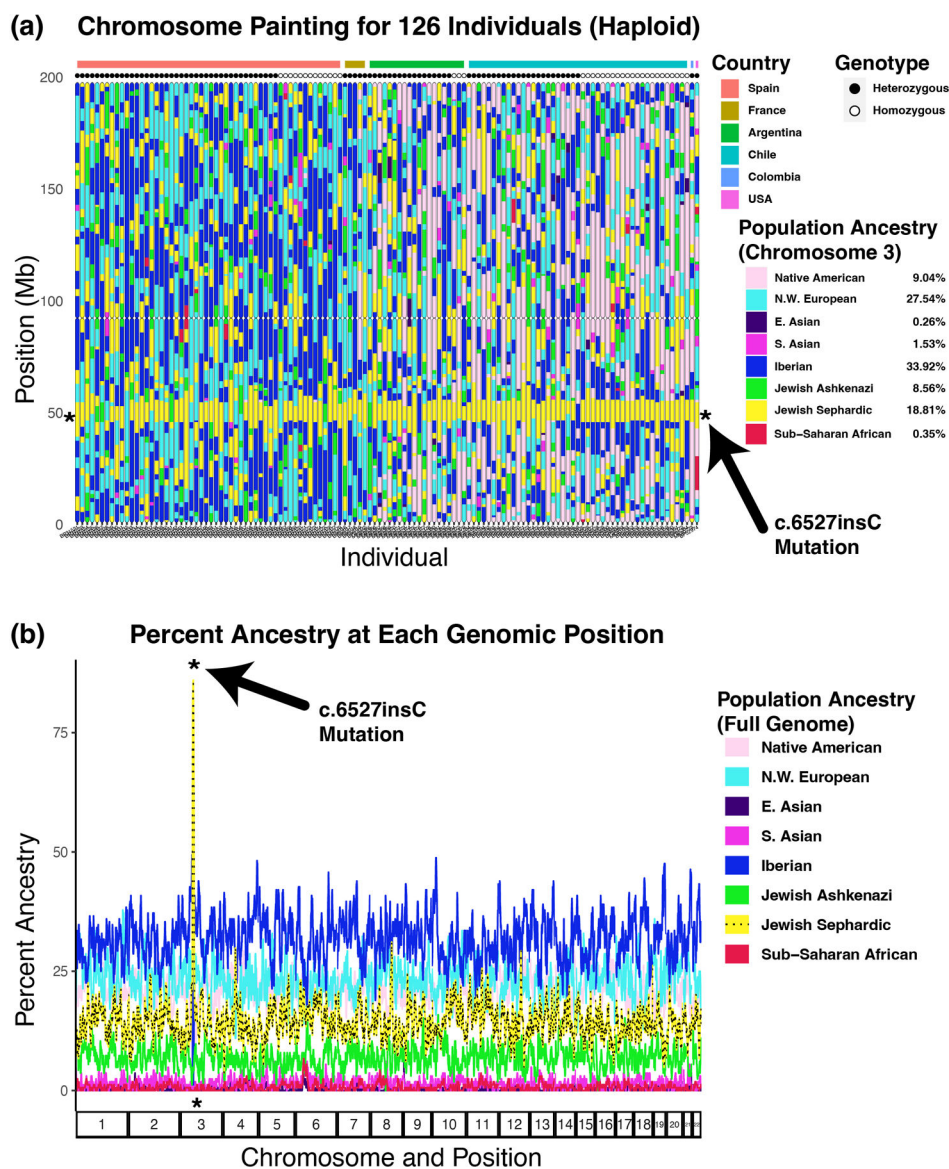
Figure 2.

Sephardic diaspora and the c.6527insC mutation. (A) Geographic representation of Sephardic Diaspora during post-Columbian history (after 1492 CE). Routes and locations are only approximate and likely incomplete. (B) Conceptual historical model of the c.6527insC mutation, with hypothesised times to be tested. First (T 0), the mutation arose in a single ancestor during the Iron Age.14 Next (T 1), a group of Sephardic individuals with high prevalence of the mutation migrated to Iberia where they admixed with local Spaniards. This may have occurred during the Golden Age of Jews in Spain.8 14 Finally (T 2), many post-Columbian diaspora ensued following the Spanish Inquisition, including converso Sephardic individuals who migrated to the Americas. Importantly, this model predicts the mutation to be overlapped by primarily Sephardic DNA segments in both Old and New World individuals.
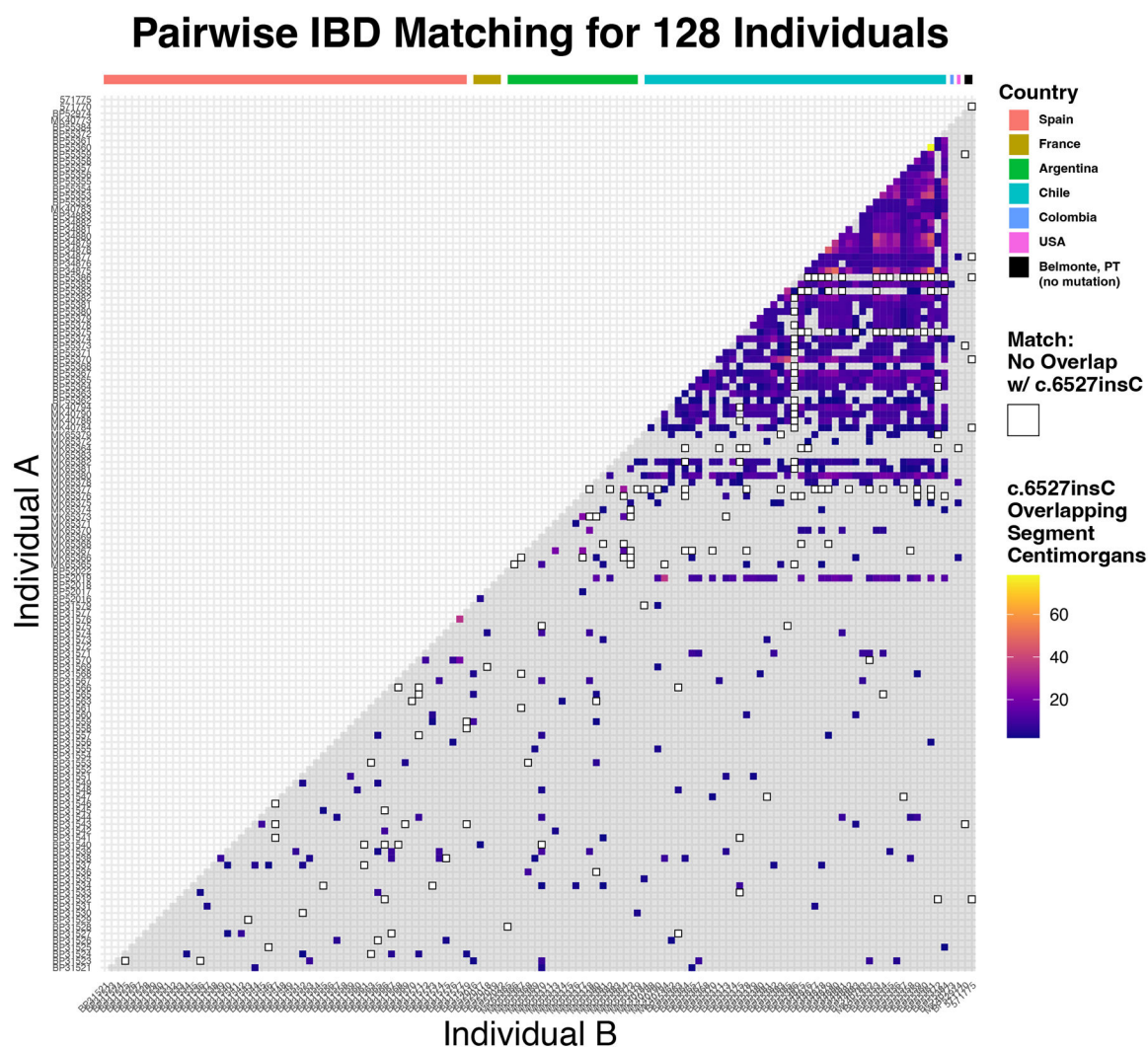
Pairwise IBD Matching for 128 Individuals

Figure 3.
Pairwise IBD matching. Family finder pairwise matching with relevant thresholds applied between all pairs of 128 samples (including two Belmonte samples). Any IBD match spanning the c.6527insC mutation is shown by plasma colours denoting the segment length in cM. If a pair matches but without spanning the mutation, an empty box is shown. Rainbow colours denoting country of origin are identical to those in figure 1, except that Belmonte samples are shown in black. IBD, identical-by-descent.
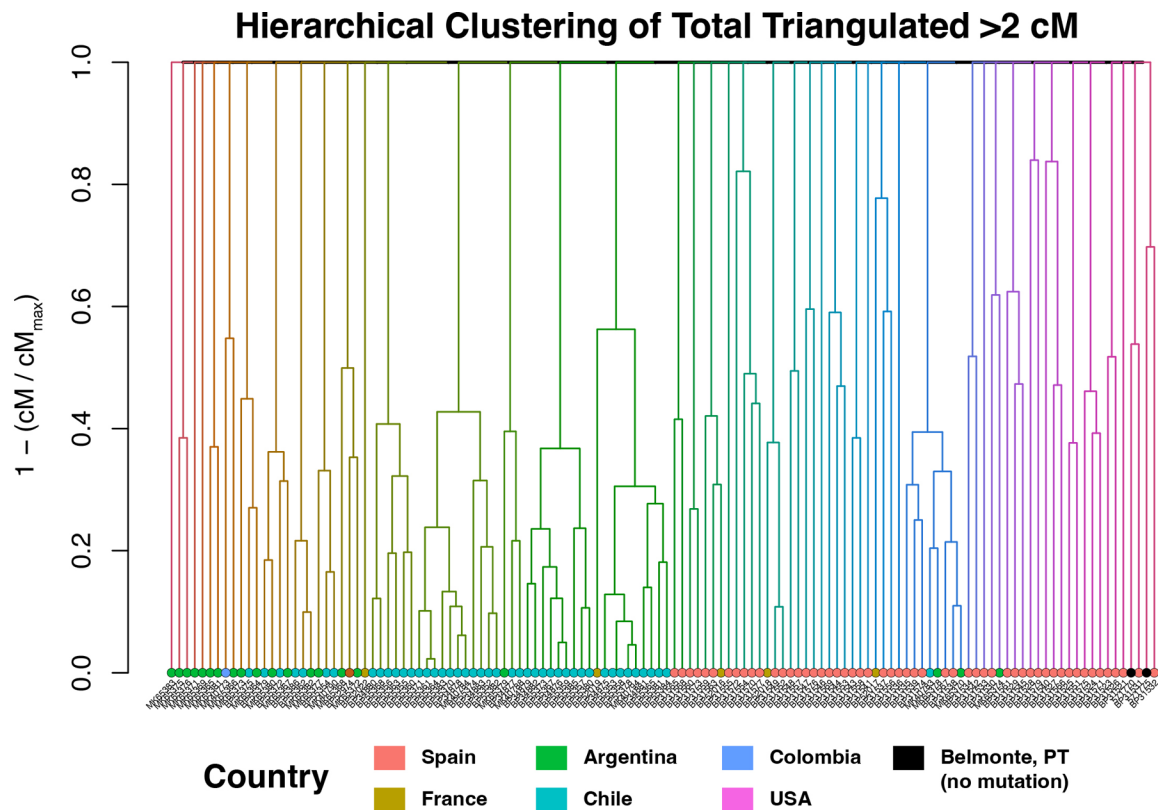
Figure 4.

Autosomal relationships of study participants. Hierarchical clustering tree showing proximity of the relationship for all 128 individuals in the study. Clustering was performed using total cM for all triangulated DNA segments, that is, those matching between three or more individuals. The vertical axis is normalised between 0 and 1. Rainbow colours for circles denote country of origin and are identical to those in figure 3. Rainbow colours for tree branches denote clusters.

Table 1.

Reference populations used for local ancestry classification, and their sources.

Target Population[*]	Proxy Population[†]	N[‡]	Citation[§]
Sub-Saharan African	Nigerian	53	Siva 2008
Native American	Native American groups in South America: Cachi, Colla, Quechua, unknown tribal affiliation in Ecuador, Peru, and Bolivia	53	Eichstaedt et al. 2014; Pagani et al. 2016; Mallick et al. 2016; Maier et al. 2021
East Asian	South Han Chinese	53	Maier et al. 2021
South Asian	Dravidian	53	Siva 2008
North European	British	53	Maier et al. 2021
South European	Iberian	53	Maier et al. 2021
Ashkenazi Jewish	Ashkenazi Jewish	53	Maier et al. 2021
Sephardic Jewish in Turkey	Sephardic Jewish in Turkey	53	Maier et al. 2021
Sephardic Jewish in Morocco	Sephardic Jewish in Morocco	38	Maier et al. 2021

[*] Continental or regional population of interest

[†] Specific reference population used in local ancestry analysis

[‡] Sample size; all non-Jewish reference populations were capped to the largest Jewish sample size

[§] Study from which reference samples were drawn

Table 2.

Age estimates for admixture of c.6527insC mutation into Iberia and Americas.

Admixture Event[*]	Method[†]	Assumption[‡]	Generations BP[§]	Calendar Date (20 Y/G)[¶]	Calendar Date (25 Y/G)[¶]	Calendar Date (30 Y/G)[¶]
Sephardic + Spanish	Gamma	'Independent' genealogy	29.7 (26.4–33.5)	1408 (1333–1475) CE	1254 (1160–1338) CE	1101 (988–1201) CE
Sephardic + Spanish	Gamma	'Correlated' genealogy	33.4 (12.8–54.1)	1334 (922–1748) CE	1162 (647–1679) CE	990 (371–1610) CE
Sephardic + Spanish	Alder	-	40.9 (25.6–56.2)	1186 (880–1492) CE	976 (594–1359) CE	767 (308–1226) CE
American + Span-Seph	Gamma	'Independent' genealogy	8.9 (7.5–10.6)	1824 (1790–1853) CE	1774 (1732–1810) CE	1725 (1674–1767) CE
American + Span-Seph	Gamma	'Correlated' genealogy	9.7 (4.5–15.1)	1809 (1701–1914) CE	1755 (1620–1886) CE	1702 (1539–1859) CE
American + Span-Seph	Alder	-	10.9 (10.4–11.4)	1784 (1774–1794) CE	1725 (1712–1737) CE	1665 (1650–1679) CE

[*] Which modeled admixture event, either T_1 or T_2

[†] Age estimation method

[‡] Assumption of Gamma method

[§] Mean estimated age in generations before present, with 95% CI

[¶] Mean estimated age in calendar year (see assumed generation time), with 95% CI