



OPEN

## Middle eastern genetic legacy in the paternal and maternal gene pools of Chuetas

J. F. Ferragut<sup>1</sup>✉, C. Ramon<sup>1</sup>, J. A. Castro<sup>1</sup>, A. Amorim<sup>2,3,4</sup>, L. Alvarez<sup>2,3,5</sup> & A. Picornell<sup>1</sup>

Chuetas are a group of descendants of Majorcan Crypto-Jews (Balearic Islands, Spain) who were socially stigmatized and segregated by their Majorcan neighbours until recently; generating a community that, although after the seventeenth century no longer contained Judaic religious elements, maintained strong group cohesion, Jewishness consciousness, and endogamy. Collective memory fixed 15 surnames as a most important defining element of Chueta families. Previous studies demonstrated Chuetas were a differentiated population, with a considerable proportion of their original genetic make-up. Genetic data of Y-chromosome polymorphism and mtDNA control region showed, in Chuetas' paternal lineages, high prevalence of haplogroups J2-M172 (33%) and J1-M267 (18%). In maternal lineages, the Chuetas hallmark is the presence of a new sub-branching of the rare haplogroup R0a2m as their modal haplogroup (21%). Genetic diversity in both Y-chromosome and mtDNA indicates the Chueta community has managed to avoid the expected heterogeneity decrease in their gene pool after centuries of isolation and inbreeding. Moreover, the composition of their uniparentally transmitted lineages demonstrates a remarkable signature of Middle Eastern ancestry—despite some degree of host admixture—confirming Chuetas have retained over the centuries a considerable degree of ancestral genetic signature along with the cultural memory of their Jewish origin.

Jewish communities in the Balearic Islands date back to the fifth century AD<sup>1</sup>. With the Christian conquest of Majorca in 1229, their physical survival was guaranteed, despite social and religious pressures forcing their conversion to Christianity between 1391 and 1435. Consequently, there were officially no more Jews in Majorca nearly 60 years before the Edict of Expulsion by the Catholic Kings in 1492. Many of these converted Jews were integrated in the general population; however, a few families remained in the ghetto and secretly adhered to Judaism, forming a Crypto-Jewish community which was persecuted by the Inquisition (fifteenth–seventeenth centuries)<sup>2</sup>. The last “Autos de Fe” in 1691 put a stop to their hidden Jewish religious practices, and this population of convicts and their descendants came to be known as Chuetas, a word probably derived from the Catalan for Jew<sup>3</sup>, with their social stigma and segregation (imposed by their Majorcan neighbours) continuing until the mid-twentieth century. There was a definitive point of inflection when Majorca opened to tourism, as the arrival of newcomers (Spaniards or foreigners) who had no knowledge of the status of Chuetas led to a decrease in anti-Chueta prejudice. Therefore, Chuetas were an isolated population with very scarce intermarriage with the Majorcan host population until recently<sup>4</sup>. One of the most important defining elements of this group is that they bear one of the 15 surnames of converso lineages (Aguiló, Bonnín, Cortès, Fortesa, Fuster, Martí, Miró, Picó, Pinya, Pomar, Segura, Tarongí, Valentí, Valleriola, and Valls) targeted by the inquisitorial sentences for Crypto-Judaism in the last quarter of the seventeenth century<sup>5</sup>. Some of these surnames are common in other Spanish regions, where they are not related to Judaism. In Majorca, however, they have been fixed in the collective memory by their identification as Chueta families.

The rise of modern Population Genetics in the second half of the twentieth century enabled geneticists to endeavour to define the origins and relatedness of Jewish people<sup>6</sup>. Since then, populations with Jewish origin have been analysed by means of uniparental and recombining markers<sup>7–12</sup>, and more recently also through

<sup>1</sup>Institut Universitari d'Investigació en Ciències de la Salut (IUNICS) i Laboratori de Genètica, Departament de Biologia, Universitat de les Illes Balears, Carretera de Valldemossa, km 7.5, 07122 Palma de Mallorca, Illes Balears, Spain. <sup>2</sup>IS-Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Rua Alfredo Allen, 208, 4200-135 Porto, Portugal. <sup>3</sup>PATIMUP-Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Rua Júlio Amaral de Carvalho, 45, 4200-135 Porto, Portugal. <sup>4</sup>Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre, s/n, 4169-007 Porto, Portugal. <sup>5</sup>TellmeGen, Calle Arquitecto Mora, 5-4, 46010 Valencia, Spain. ✉email: jf.ferragut@uib.es

genome-wide SNP arrays<sup>13–18</sup>. The combined analysis of millions of polymorphic markers along the genome have led to greater precision in the clustering of different Jewish groups and to the ability to estimate the Middle Eastern, European, and African components in each group. These analyses reflect that each of today's Jewish populations is the result of the blending of Middle Eastern and host populations (European, Asian, or African). Regarding haploid markers, on the one hand, male lineages indicate that most Jewish communities share a common Middle Eastern ancestral origin, and remained relatively isolated from neighbouring non-Jewish communities during the Diaspora. On the other hand, mtDNA studies lead to conclude that there are differences in the demographic history of the widespread communities resulting from the Jewish Diaspora in terms of maternal ancestries, indicating different maternal founder effects.

Converted Jews have also been subject of study, either due to their contribution to the host population genetic pool<sup>19–21</sup>, or owing to their isolation and differentiation from their neighbours, such as the Portuguese Crypto-Jew communities in Belmonte and Bragança<sup>22–24</sup> or the Chuetas in Majorca, the subject of the present study. Chuetas have been previously studied by means of autosomal and X-chromosome markers<sup>25–29</sup>. The results conducted to date demonstrate that Chuetas are a differentiated population that has kept a considerable proportion of its original genetic make-up, especially clear in some markers where Chuetas show polymorphic frequencies of alleles that are very rare in neighbour populations, but not in Middle Eastern populations<sup>30</sup>. However, a certain degree of admixture from and with the host population has also been detected<sup>26,29</sup>.

In this study, we focus on haploid markers in order to investigate the ancestry and demographic history of the maternal and paternal founding lineages of the Chueta population, and to analyse whether cultural isolation has led to the reduction of genetic diversity in mtDNA and Y-chromosome lineages in this population.

## Materials and methods

**Population sampling.** For the study of the Y-chromosome, samples from 146 unrelated males were obtained: 100 from the Chueta population, and 46 from Majorca (Balearic Islands, Spain), included as the host population of Chuetas. The Majorcan samples constitute a subset of a larger sample previously genotyped for 12 Y-STRs<sup>31</sup>. For mtDNA analyses, 183 samples were used: 104 non-maternally related individuals from the Chueta population (some of these samples were used in a preliminary analysis in a conference contribution<sup>32</sup>) and 79 Majorcans. All participants provided appropriate informed consent statements, approved by the "Direcció General de R + D + I" (Government of the Balearic Islands, Spain), and the University of the Balearic Islands (procedure AAEE24/2014), following the procedures approved by the Ethics Committee of the University of Porto (N102/CEUP/2012). Anonymity of the recruited samples was preserved during the study. All methods were carried out in accordance with the guidelines and regulations of the Declaration of Helsinki.

**DNA extraction.** DNA was extracted by standard phenol–chloroform method and quantified on a NanoVue Plus spectrophotometer (GE Healthcare Life Sciences, Cambridge, UK).

**Genotyping analyses.** *Y-chromosome.* Seventeen Y-chromosome STR markers were amplified using the Y-filer kit (Applied Biosystems, Foster City, CA, USA), following the manufacturer's recommendations. Thirty-eight SNPs were typed to define the major male lineages. Thirty-two of them were genotyped using SNaPshot kit (Applied Biosystems) in five multiplexes as previously described<sup>21,33–35</sup> (Fig. 1). M1 and M269 were genotyped with conventional PCR followed by agarose gel electrophoresis; S116, M17, and M18 were genotyped by Sanger sequencing; and DYS458.2 was used to determine the J1 chromosomes<sup>36</sup>.

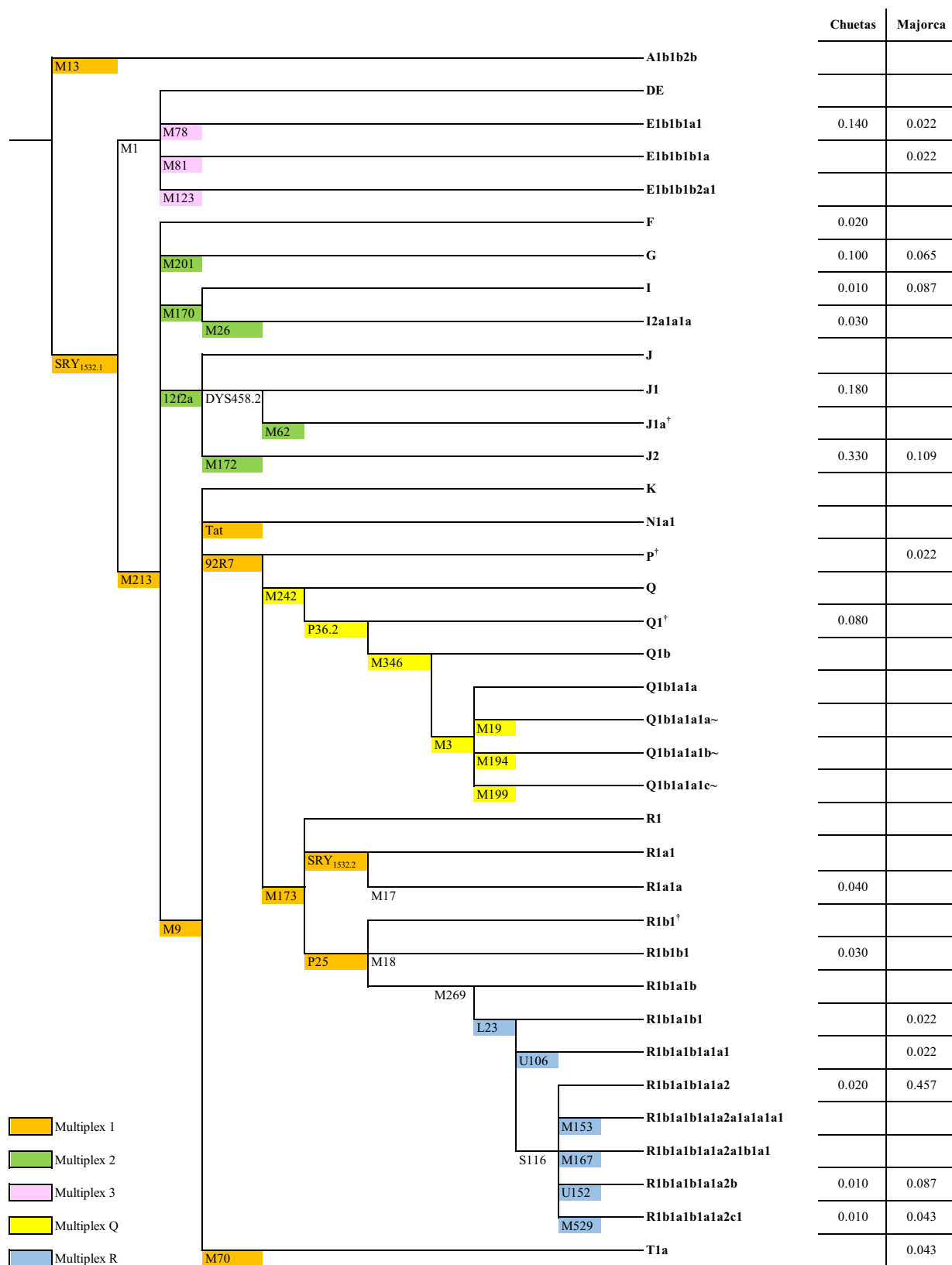
Y-STR amplification products and Y-SNP purified minisequencing products were separated in an ABI PRISM 3130 Genetic Analyser, and electropherograms were analysed using GeneMapper ID software v3.2 and Peak Scanner software (Applied Biosystems). Y-STR alleles were designated according to ISFG recommendations<sup>37</sup>, and Y-SNP haplogroups according to the latest ISOGG update (2019).

**Mitochondrial DNA.** The mitochondrial DNA control region, or D-loop (16024–576), was amplified with two overlapping fragments using mtDNA-specific primers (L15997, H016 and L16555, H639)<sup>38</sup>. The amplified product was purified with the MBS Spin PCRapace kit (Invitek, Berlin, Germany), and subsequent sequencing reactions were carried out using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems), following the manufacturers' procedures. Finally, products were run in an ABI prism 3130 analyser.

Sequences were assembled and compared to the revised Cambridge Reference Sequence (rCRS) (NC\_012920) using Geneious software version 7.1.3 (Biomatters, Ltd., Auckland, New Zealand). Haplogroups were classified following the updated mtDNA phylogeny, PhyloTree, mtDNA tree Build 17 (<http://www.phylotree.org/>) using HaploGrep2 tool<sup>39,40</sup>, and assigned haplotypes were validated by EMPOP (<http://empop.org/>) curators.

Moreover, the entire mtDNA molecule was sequenced for six Chueta samples belonging to the modal haplogroup R0a + 60.1 T. Amplification of the whole molecule was carried out by 19 overlapping fragments, and the sequencing strategy used 31 smaller fragments, primers, and specifications as described in Ramos et al.<sup>41,42</sup>. When Chueta specific mutations were identified in the complete mtDNA molecule, the status of such nucleotide positions was interrogated in the rest of the samples belonging to the haplogroup.

**Data analysis.** Haplogroup frequencies were estimated by gene counting. Allele and haplotype frequencies, molecular diversity indices, Analysis of Molecular Variance (AMOVA), as well as the corresponding non-differentiation p-values, and Y-chromosome pairwise  $F_{ST}$  genetic distances were assessed using Arlequin v3.5.1.2<sup>43</sup>. For easier visualisation of the genetic distances observed, a multidimensional scaling (MDS) plot of the pairwise  $F_{ST}$  matrix was represented using SPSS v.15.0 (SPSS, Inc., Chicago, IL, USA).



**Figure 1.** Phylogenetic tree of the 38 Y-SNPs typed, and haplogroup frequencies for Chuetas and Majorca populations. Haplogroups were named in accordance with the latest Update of ISOGG 2019. Haplogroups labelled with † are named as in their original description: J1a-M62, P-92R7, and R1b1-P25 (Brion et al.<sup>35</sup>), and Q1-P36.2 (Roewer et al.<sup>35</sup>).

	N	K	UH	HD	h	MPD
<b>Yfiler</b>						
Chuetas	100	53	39	0.965 ± 0.008	0.611 ± 0.311	10.387 ± 4.774
Majorca	46	44	44	0.998 ± 0.005	0.606 ± 0.312	10.308 ± 4.783
<b>Y-SNPs</b>						
Chuetas	100	13	3	0.827 ± 0.023	0.106 ± 0.059	4.033 ± 2.029
Majorca	47	12	5	0.771 ± 0.058	0.121 ± 0.067	4.614 ± 2.302

**Table 1.** Y chromosome molecular diversity indices for haplotypes and haplogroups in Chuetas and Majorcans. *N* number of individuals, *K* number of haplotypes/haplogroups, *UH* unique haplotypes/haplogroups, *HD* haplotype diversity ± SD, *h* gene diversity over loci, *MPD* mean pairwise differences.

Population	N	K (% K)	S	HD ± SD	Π ± SD	θ <sub>k</sub> (95% CI)
<b>D-loop (16,024–576)</b>						
Chuetas	104	50 (48.08%)	110	0.950 ± 0.015	0.010 ± 0.005	37.216 (25.061; 55.103)
Majorca	79	67 (84.81%)	120	0.995 ± 0.003	0.009 ± 0.005	205.807 (117.911; 374.184)
<b>HVRI (16,024–16,365) and HVRII (72–300)</b>						
Chuetas	104	46 (44.23%)	74	0.948 ± 0.014	0.013 ± 0.007	31.003 (20.769; 46.042)
Bragança Jews <sup>24</sup>	57	35 (61.40%)	61	0.967 ± 0.012	0.014 ± 0.008	37.616 (22.328; 63.866)
Belmonte Jews <sup>22</sup>	30	2 (6.67%)	6	0.129 ± 0.115	0.001 ± 0.001	0.279 (0.065; 1.097)
Bulgarian Jews <sup>22</sup>	71	46 (64.79%)	70	0.982 ± 0.007	0.012 ± 0.007	55.477 (34.470; 90.159)
Turkish Jews <sup>22</sup>	123	85 (69.11%)	109	0.985 ± 0.005	0.013 ± 0.007	120.394 (82.638; 176.996)
Libyan Jews <sup>22</sup>	83	36 (43.37%)	63	0.922 ± 0.019	0.013 ± 0.007	23.631 (15.104; 36.717)
Moroccan Jews <sup>22</sup>	148	80 (54.05%)	92	0.979 ± 0.005	0.012 ± 0.006	70.307 (50.484; 97.941)
Tunisian Jews <sup>22</sup>	36	25 (69.44%)	43	0.971 ± 0.015	0.012 ± 0.007	34.939 (18.044; 69.325)
Iranian Jews <sup>22</sup>	82	43 (52.44%)	76	0.971 ± 0.008	0.016 ± 0.008	35.821 (23.098; 55.513)
Iraqi Jews <sup>22</sup>	134	48 (35.82%)	79	0.950 ± 0.009	0.016 ± 0.008	26.344 (18.151; 37.920)

**Table 2.** Diversity indices results calculated for the complete D-loop and for the HVRI + HVRII fragment for inter-population comparison. *K* Number of different haplotypes, *S* number of polymorphic sites, *HD* haplotype diversity, *Π* nucleotide diversity averaged over loci, *θ<sub>k</sub>* theta estimator based on the number of different lineages.

To attain statistical significance for frequencies of putative Jewish founding lineages, Bayesian 0.90 credible region (90% CR) was calculated using SAMPLING software (Macaulay, personal communication). Furthermore, another criterion established by Behar et al.<sup>22</sup> was to consider haplogroups with a frequency greater than 5% and TMRCA prior to 2 Kya as founder lineages.

Median joining networks of Y-STR haplotypes were constructed using Network 4.6.1.1 ([www.fluxus-technology.com](http://www.fluxus-technology.com))<sup>44</sup>.

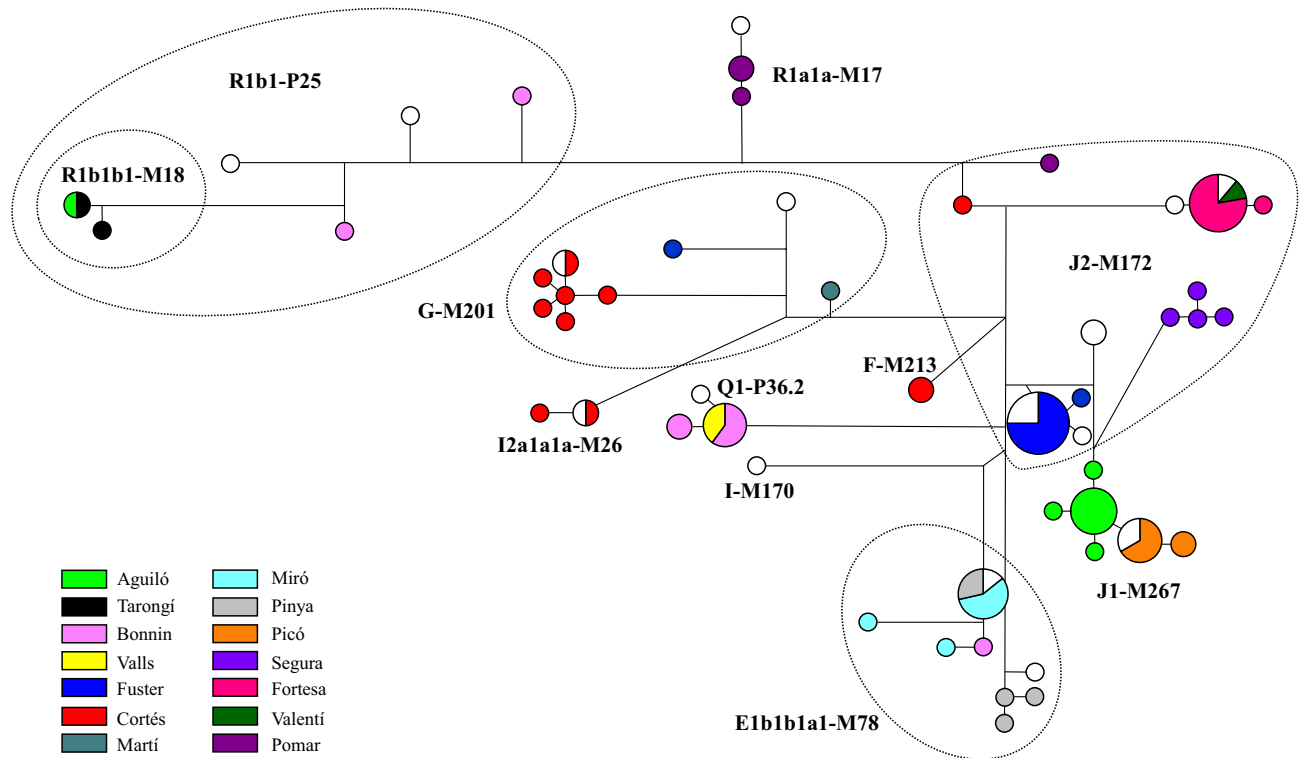
In order to visualise the distribution of mtDNA haplogroup R0a frequencies in the Mediterranean geographic context, an isometric spatial frequency distribution map was constructed with the program Surfer 9 (Golden Software, <http://www.goldensoftware.com/products/surfer>).

## Results and discussion

**Genetic diversity.** *Haplotype diversity.* Allele frequencies and gene diversities of each Y-STR of the populations under study are shown in Supplementary Table 1. The DYS385 locus showed the highest gene diversity (GD) values (>0.83) in both Chuetas and Majorcans, as expected, due to its duplicated structure, while especially low values were found at DYS389I (0.224) and DYS392 (0.288).

Amongst the 146 males analysed, 97 Y-STR haplotypes were observed (Table 1 and Supplementary Table 2), 81 of which were observed only once (singletons). The frequencies of the most common haplotypes were 12% and ~4% in Chuetas and Majorcans, respectively. No haplotype was shared between Chuetas and Majorcans. Haplotype diversity in Chuetas (0.965) was considerably lower than in Majorcans (0.998), Bragança Jews (0.987)<sup>23</sup>, or other populations in the literature<sup>45,46</sup>. These results are in accordance with the historical and demographic data of this population and with the reduced genetic diversity in some markers found in previous genetic studies<sup>29,47</sup>.

Complete mitochondrial control region haplotypes for Majorcan and Chueta populations are presented in Supplementary Tables 3 and 4. In the 104 samples from the Chueta population, 50 (48.08%) different haplotypes were identified; meanwhile, in the 79 Majorcans, 67 (84.81%) different haplotypes were found. Estimated diversity parameters are summarised in Table 2. Notably, theta *k* values (θ<sub>k</sub>) in Chuetas were much lower than in Majorcans, but within the same range as those reported for other non-Ashkenazi Jews<sup>22,24</sup>. Therefore, the estimated number of putative female founders in Chuetas is similar to the one estimated in most of these Jewish groups.



**Figure 2.** Median-Joining network of Chueta haplotypes. Colour code indicates the Chueta surname of the sample, the colour white is used for anonymised individuals. Smaller circles are singletons and size is proportional to haplotype frequency. Haplogroup assignment is indicated in each group of haplotypes.

**Haplogroup composition.** Regarding the Y-chromosome, Chueta samples were classified into 13 different haplogroups according to the 38 SNPs genotyped (Table 1 and Fig. 1). The six most frequent haplogroups ( $\geq 4\%$ ) (E1b1b1a1-M78 (hereafter E1b-M78), G-M201, J1-DYS458.2 (used in this study as synonymous of J1-M267), J2-M172, Q1-P36.2, and R1a1a-M17) accounted together for up to 87%, with J2-M172 as the most frequent (33%). Three of the seven remaining haplogroups occurred in only one individual. For the Majorca population, 12 haplogroups were found, the most common being R1b-S116 (46%). High diversity values, based on haplogroup frequencies, were found in Chuetas ( $0.827 \pm 0.023$ ) in comparison with other Iberian populations and Majorca ( $0.771 \pm 0.058$ ), but in the same range as the Sephardic Jewish population<sup>19</sup>.

AMOVA analyses at both haplogroup and haplotype (Y-STRs) level showed significance when taking Chuetas and Majorcans as a whole ( $F_{ST} = 0.196$ ;  $p < 10^{-5}$ ), indicating the differentiation of the Chuetas with respect to their host population. SAMPLING software identified three differential haplogroups between Chuetas and Majorcans: R1b1a1a2-M269 (hereafter R1b-M269), J1-DYS458.2, and J2-M172, pointing towards the lack of R1b-M269 and the presence of J1-DYS458.2 and J2-M172 as Chuetas' putative founding lineages. The frequency in Chuetas of haplogroups rarely found in neighbouring populations—E1b-M78, Q1-P36.2, G-M201, and R1a1a-M17 (14, 10, 8, and 4%, respectively)—could also mean that they might have been present in the original Jewish Majorcan gene pool.

Network analyses were performed for the main haplogroups (Supplementary Fig. 1). Within networks, Majorcan and Chueta individuals did not share haplotypes. R1b-M269, found mainly in Majorcans, showed the highest diversity (0.998), while the main haplogroups in Chuetas manifested much lower haplotype diversities [ranging between 0.684 (J-12f2a) and 0.833 (R1a1-M17)], except for G-M201 (0.978).

Although studying the relationship between haplogroups and surnames was not an initial aim of our work on the Y-chromosome in Chuetas and, therefore, the sampling was not designed for it, with a median joining network we assigned and compared the haplotypes and haplogroups found with the surnames of the individuals (Fig. 2). Fourteen of the 15 Chueta surnames are represented, but not the surname Valleriola, which left no descendants. When sorted and analysed by surnames, the majority of these surname sets contained a highly reduced haplogroup diversity ( $h = 0.000$ – $0.222$ ), contrarily to that observed in Majorcan non-Chuetas surnames included in other studies<sup>48</sup>. The surnames Bonnin and Pomar revealed two different haplogroups with intermediate  $h$  values ( $\sim 0.500$ ) and Cortés was the most diverse with 4 haplogroups and an  $h$  value of 0.691; although the sample size would need to be enlarged to confirm these results. Putative founding haplogroups in the Chueta population, J1-M267 and J2-M172, were found to be associated to the surnames Picó and Aguiló (J1-M267) and Segura, Fortesa, and Fuster (J2-M172). The haplotypes carried by the individuals of most surnames show a star-like distribution with only one or two mutational steps between them. Foundation of each Chueta surname by one or a very few individuals in the Christian conversions (fourteenth–fifteenth centuries) could explain these results. In a few cases, the same haplotype is shared by different surnames or, contrarily, individuals within a surname

Haplogroup	Chueta	Majorca	Haplogroup	Chueta	Majorca	Haplogroup	Chueta	Majorca
D1j	0.0096	–	K1a	0.0385	0.0380	T2b23	0.0192	–
H*	–	0.0633	K1a1b1a	0.0385	–	T2b5a1	0.0096	–
H1	0.0865	0.1772	K1a4a1a + 195	0.0096	0.0127	T2c1d	0.0481	0.0127
H2a2a	0.0385	0.0253	K1b1a1 + 199	0.0096	–	U1a1a	0.0481	–
H3	0.0288	0.0380	K1b1 + 16,093	–	0.0127	U2e1'2'3	–	0.0127
H4a1a	–	0.0253	K1c	0.0096	–	U2e1e	–	0.0127
H6a1	0.0096	0.0380	K2a5	–	0.0127	U2e2a2	–	0.0127
H11a	0.0096	0.0253	K2b1a1	0.0096	0.0506	U3	0.0192	–
HV0	–	0.0759	L2a1b + 143	–	0.0127	U3a	0.0096	–
HV0 + 195	0.0192	0.0380	L3d1b2	–	0.0127	U4b3	–	0.0127
HV15	–	0.0127	L3e2b + 152	0.0481	–	U5a2	0.0096	–
HV4a2a	–	0.0127	M1a1	0.0192	–	U5b1d2	0.0096	–
I	–	0.0127	M5a1	0.0192	–	U5b1f1a	0.0192	–
I1c1	0.0096	–	N1b1	–	0.0127	U5b2a2	–	0.0127
I2'3	–	0.0127	R0a + 60.1 T	0.2019	–	U5b2b1a1	–	0.0127
J1b1a1	–	0.0127	T	0.0288	0.0127	U5b2b3	–	0.0253
J1b1b	–	0.0127	T1a	0.0577	–	U5b3	0.0096	0.0253
J1c	–	0.0127	T1a1'3	0.0096	–	U6a	0.0096	–
J1c2o	0.0096	–	T2	0.0096	0.0127	V + 16,298	–	0.0127
J1d1	0.0096	–	T2a1b	–	0.0253	X2c	–	0.0127
J2a1a1	0.0288	–	T2b	0.0096	–	Unclassified	–	0.0127
J2b1a	0.0096	0.0127						

**Table 3.** Haplogroup frequencies in Chueta and Majorcan populations.

belong to very distant Y-lineages, although the scarcity of historical documents with the Christian names that converted Jews adopted does not allow us to assess the different scenarios that could explain these cases further.

Upon analysing mtDNA results (Table 3), we found the Majorcan population had haplogroup H as the most common one (39%), as was expected of a typical European population<sup>47,49</sup>. Haplogroup H together with haplogroups U, K, and HV (frequencies ranging from 13 to 14%) accounted for 78% of total diversity. Other haplogroups found in this population were I, J, L, N, T, V, and X. In contrast with the Majorcan population, in the Chueta samples, haplogroup H only accounted for 17% of total diversity. The modal haplogroup in Chuetas (~20%) was found to be the Middle Eastern haplogroup R0a + 60.1 T, followed by haplogroups T, K, U, and J (19%, 12%, 8%, and 6%, respectively). Together these 5 haplogroups represented 64% of all variation. The remaining lineages were observed at frequencies ranging from ~1 to 5% (Table 3).

All in all, the mitochondrial haplogroup composition indicated that Chuetas are statistically different from their host population (Majorca) ( $F_{ST}$  p-values <  $10^{-5}$ ). By means of the SAMPLING analyses, and also with the Behar et al.<sup>22</sup> criterion, two haplogroups showed up as the founder lineages in the Chueta population: R0a + 60.1 T (~20%) and T1a (~6%).

The presence in Chuetas of haplogroups rarely found in neighbouring populations—L3eb + 152, U1a1a, and K1a1b1a (with frequencies of 5%, 5%, and 4%, respectively)—could also mean that they might have been present in the original Jewish Majorcan gene pool.

**Phylogenetic relation with other populations.** *Y-chromosome.* The most frequent subclade of haplogroup R in Europe is R1b-M269, with frequencies ranging from 41 to 83%<sup>50</sup>. Ancient DNA studies carried out in recent years<sup>51–54</sup> have shown that this lineage is associated with the spread of Steppe ancestry during the Bronze Age. In Iberia former Y-chromosome lineages were nearly completely replaced by haplogroup R1b-M269<sup>53</sup>, and a west-to-east gene flow from Iberia could have introduced these haplogroups into Western Mediterranean islands<sup>54</sup>. Majorca showed similar values (63%) to the rest of Iberian populations<sup>19,46</sup>, whereas the Chueta population (4%), had similar values to Middle Eastern and North African populations<sup>55,56</sup>. In Jewish populations, values range between 5% and 11.5%<sup>19,57,58</sup>, with the highest frequencies in Sephardim. These results support some degree of Iberian admixture in Sephardic Jews<sup>19</sup> and important gene flow from the host population in Bragança Crypto-Jews (with an R1b-M269 frequency of 28%), as suggested by Nogueiro et al.<sup>23</sup>.

Haplogroup J-12f2a, thought to have originated in the Fertile Crescent<sup>59</sup>, shows an East to West gradient in Europe. Although the phylogeography of haplogroup J is complex, its radiation seems to be concentrated mainly in the Bronze Age, an essential period for the establishment of the modern European genetic pool<sup>60</sup>. It contains two major branches, one of them, J2-M172, has its high frequencies in the Levant and is the most frequent sub-haplogroup in Europe, mainly throughout the Mediterranean basin. It was long considered to have spread across Europe with the demic diffusion of Neolithic agriculturalists<sup>61,62</sup>, but ancient DNA studies contradict this hypothesis, since in early Middle Eastern farmers, and also in European Neolithic remains, haplogroup J2-M172 is only detected sporadically<sup>63</sup>. Signs of population movements from the East, mainly maritime, such

as the Phoenicians<sup>7,56,63</sup> have been linked to J2<sup>56,60</sup> and also to the other main branch, J1-M267. This lineage has its maximum frequency in Arabia, but also high frequencies in the Middle East and in Jewish groups<sup>58,64</sup>. Some of its derived lines have purportedly been associated with Arabisation in North Africa<sup>46,62</sup> while others have been implicated with different expansions of Middle Eastern populations through the Mediterranean Sea<sup>56</sup>. In most Jewish populations, J2-M172 reaches considerable frequencies<sup>9,14,19,57</sup> (Fig. 3), but in Europeans it is about 10%. In the Chueta population, J2-M172 was the modal haplogroup with a frequency of 33%, while in its host population it was found in 10% of males. This value is similar to other Jewish populations and higher than the average Sephardic percentage (22%). Haplogroup J1-M267, not found in Majorca, was the second most frequent in Chuetas, with similar values (18%) to those in other Jews (Fig. 3)<sup>14,19</sup>.

Other Y-haplogroups with differential frequency between Chuetas and their host population were E1b-M78, Q1-P36.2, G-M201, and R1a1a-M17. E1b-M78 seems to have originated in north-eastern Africa and several lines of evidence suggest that some E1b-M78 derived lines have been involved in trans-Mediterranean migrations directly from Africa to Europe<sup>65</sup>. Chuetas showed a frequency for this haplogroup of 14%. In other populations with Jewish origin, it has been found to range from 3.5% (Bragança Crypto-Jews) to 15% (North African Jews)<sup>19,23,57,58,61</sup>; and in Middle East non-Jewish populations, from 10.0 to 17.0%<sup>57</sup>. Haplogroup Q1-P36.2 (xM346) is practically absent in Europe and Africa<sup>19,66</sup>. In Chuetas it showed a frequency of 8%, while in Jewish populations, percentages ranging from < 1% to 5% have been found<sup>9,19,58,66</sup>, with one branch typical of Ashkenazi Jews<sup>66</sup>. One of the main sub-branches of haplogroup G-M201, G2a, was the predominant male lineage in early European farmers, although the important Y-chromosome turnover in the Bronze Age nearly completely replaced it with the R1b lineage<sup>52,53</sup>. Nowadays haplogroup G-M201 is most common in Caucasus where the maximum frequencies are observed (> 70%), but it also occurs in the rest of the Middle East and south-western European countries at frequencies ranging from 5 to 15%<sup>67</sup>; while in North Africa it is far lower than in any European Mediterranean population<sup>46</sup>. In Jewish populations, Moroccan and Sephardic Jews have the highest values, 16% and 19%, respectively<sup>14,23,57</sup> (Fig. 3). The existence of gene flow between Sephardim and Iberians (with frequencies reaching 5.0%) has been suggested<sup>19,23</sup>, although the direction of the introgression could not be determined. Lastly, the main subclade of the typical Eastern Europe R1a-M240<sup>68</sup> is R1a1a-M17<sup>69</sup>, which is found in high percentages in Ashkenazim (up to 14.5%) and is also present in other non-Ashkenazi Jewish populations, but at a much lower frequency (4.4%)<sup>9,70</sup>. In Chuetas, R1a1a-M17 was found in 4% of the individuals analysed, but not in the Majorcan samples. It could be interesting to assess whether the Chueta R1a1a chromosomes belong to the subclade shared by Ashkenazim and other Middle Eastern populations (M582), but absent in Eastern Europeans<sup>70</sup>.

All in all, the Y-chromosome haplogroup profile in Chuetas is clearly dissimilar from their neighbouring population, Majorca, and quite similar to the haplogroup frequencies found in Sephardic and other Jewish populations, evidencing a considerable frequency of J2 + J1 haplogroups and low values of R1b (Fig. 3). Therefore, it seems likely that the differential presence in the Chueta population of both haplogroups (and also of E1b-M78, Q1-P36.2, G-M201, and R1a1a-M17) results from their well-known historical Jewish origin and/or by admixture with other Jewish groups, especially with North Africans, due to commercial contact between both communities<sup>71</sup>.

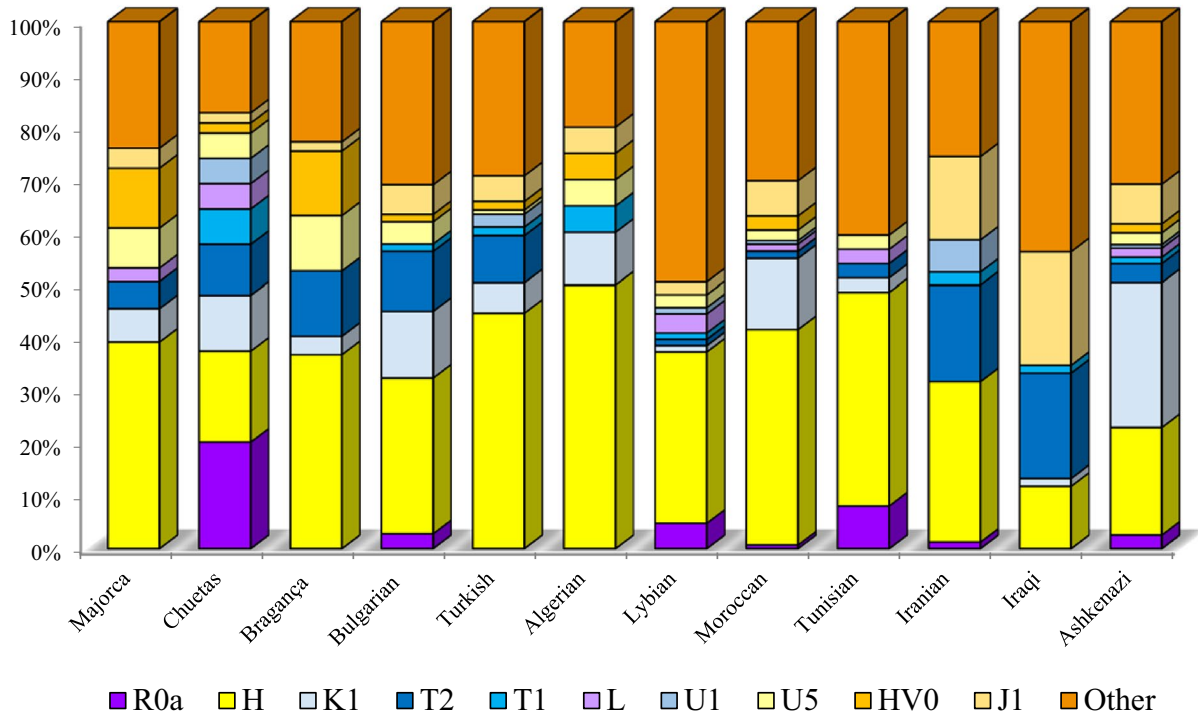
A comparison with available populations in the literature, based on both STRs and SNPs, (Fig. 4) positioned Chuetas with other Jewish and Middle Eastern populations, far from their geographical neighbours, evidencing that Chuetas maintain, in male lineages, substantial relics of their Jewish ancestry.

**Mitochondrial DNA.** The origin in terms of location and timescale of the Chuetas' modal haplogroup R0a has been under debate in recent years due to the geographic distribution of its frequencies<sup>72–74</sup>. As shown in Fig. 5, this haplogroup is practically absent in Europe (0–2%), although some exceptions are found, such as Cappadocia village in Italy (14.6%)<sup>75</sup>. The highest frequencies of haplogroup R0a are found in the Arabian Peninsula and the Horn of Africa, reaching values as high as ~ 25% in Soqatra Island in Yemen<sup>73</sup>. Frequencies in Jewish groups<sup>8,22,73,76</sup> on the whole show similar frequencies to their host populations (Fig. 5), but not in the case of Chuetas, who present a frequency of R0a of 20.2% while it is absent in their host population. The first dating of the haplogroup (~ 19 Kya) suggested an Arabian origin<sup>72</sup>. Later studies dated the haplogroup earlier, ~ 22.5 Kya<sup>73</sup> and ~ 30 Kya<sup>74</sup>. Both studies discuss whether the origin could be in the Horn of Africa or the Arabian Peninsula. Phylogeographic differences in the regional distribution of R0a and the fact that the most ancient reservoir of R0a variation is found in Arabia led the authors to conclude an Arabian origin of the haplogroup. Two main branches characterise this haplogroup: R0a1 (~ 26 Kya) and R0a2'3 (~ 21 Kya). R0a2'3 is defined by the insertion 60.1 T and most of the Jewish populations where information is available show this branch. Only Yemenite Jews show both<sup>22</sup>.

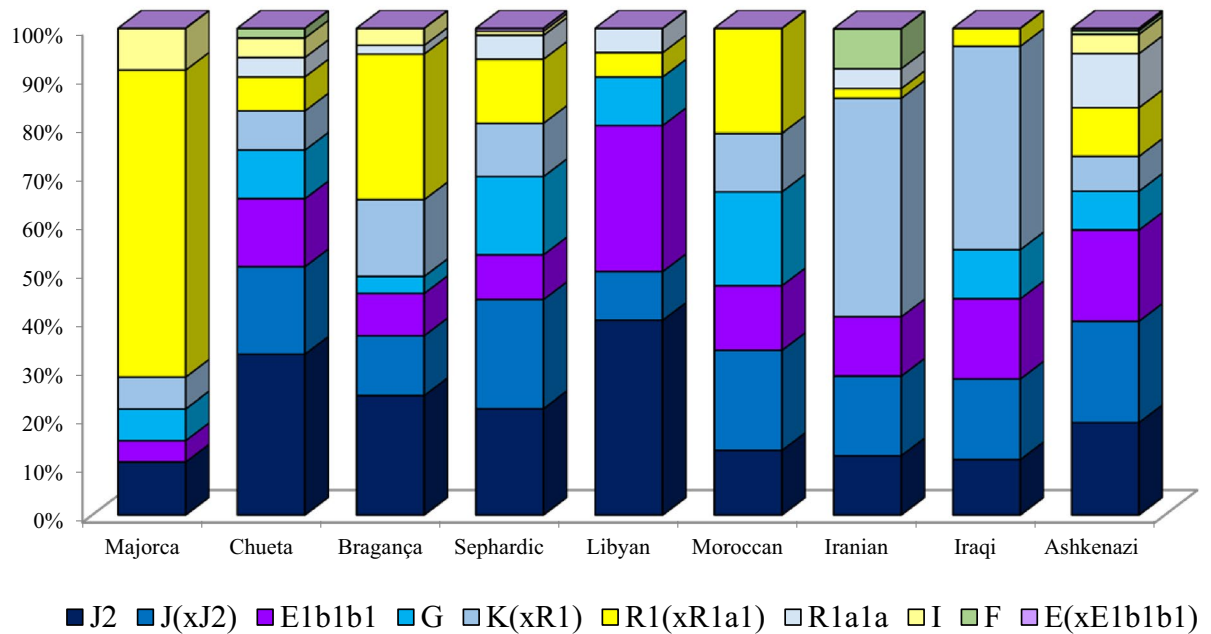
Taking into account the high prevalence of this haplogroup in the Chueta population, it was considered important to delve into the phylogeny of their possible main maternal founder. Thus, the complete mtDNA genome was obtained for 6 out of the 21 R0a + 60.1 T samples, which classified the Chueta samples as R0a2m. The R0a2m branch, dated by Gandini et al.<sup>74</sup> to ~ 1.4 Kya, is found in just three samples in the literature, two of them from Jewish origin and the other with unknown ethnicity. In Fig. 6, it can be seen that all six complete Chueta molecules share an additional mutation (A13858G), whereas one sample has a G15734A private mutation. In all the rest of R0a + 60.1 T Chueta samples, we have checked the defining mutation of the R0a2m branch (A4767G). Besides, the two specific Chueta positions have also been examined, revealing that A13858G was present in all 21 individuals, suggesting a new R0a2m sub-branch in this population, while the G15734A mutation remained as a private variant.

The time of the appearance of haplogroup JT can be estimated at ~ 58 Kya, before the settlement of the Fertile Crescent. It has been suggested that haplogroups J and T diverged during the settlement ~ 40 Kya and ~ 30 Kya,

a) mtDNA

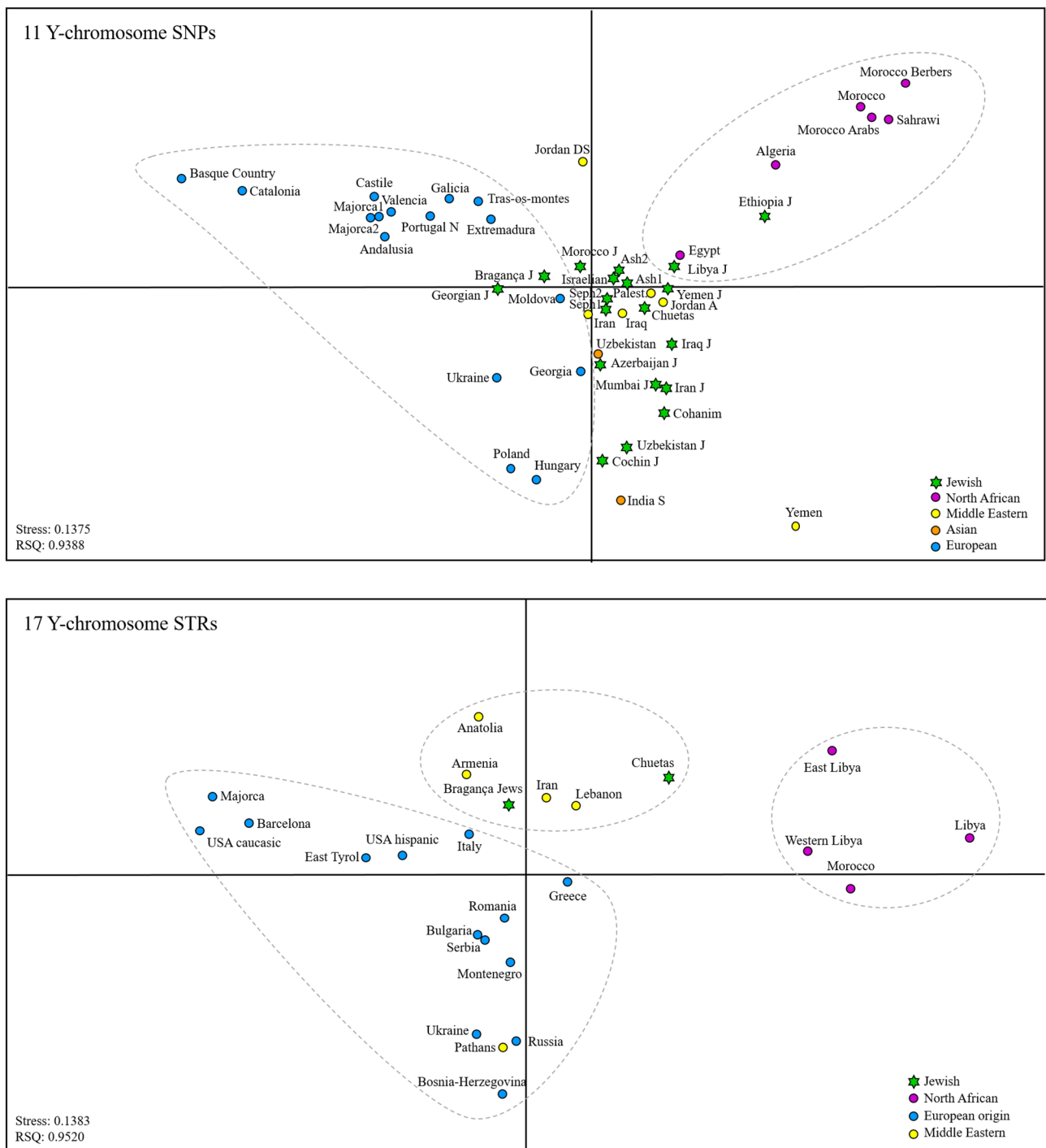


b) Y-chromosome



**Figure 3.** Haplogroup frequencies of mtDNA (a) and Y-chromosome (b) in Chuetas, Majorcans, and different populations with Jewish origin, based on data from the literature (Supplementary Table 5). Y-chromosome haplogroups are defined with the following SNPs: M172 (J2), 12f2a (J(xJ2)), M35 (E1b1b1), M201 (G), M9(xM173) K(xR1), M173(xM17) R1(xR1a1), M17 (R1a1), M170 (I), M213 (F), M1(xM35) E(xE1b1b1).

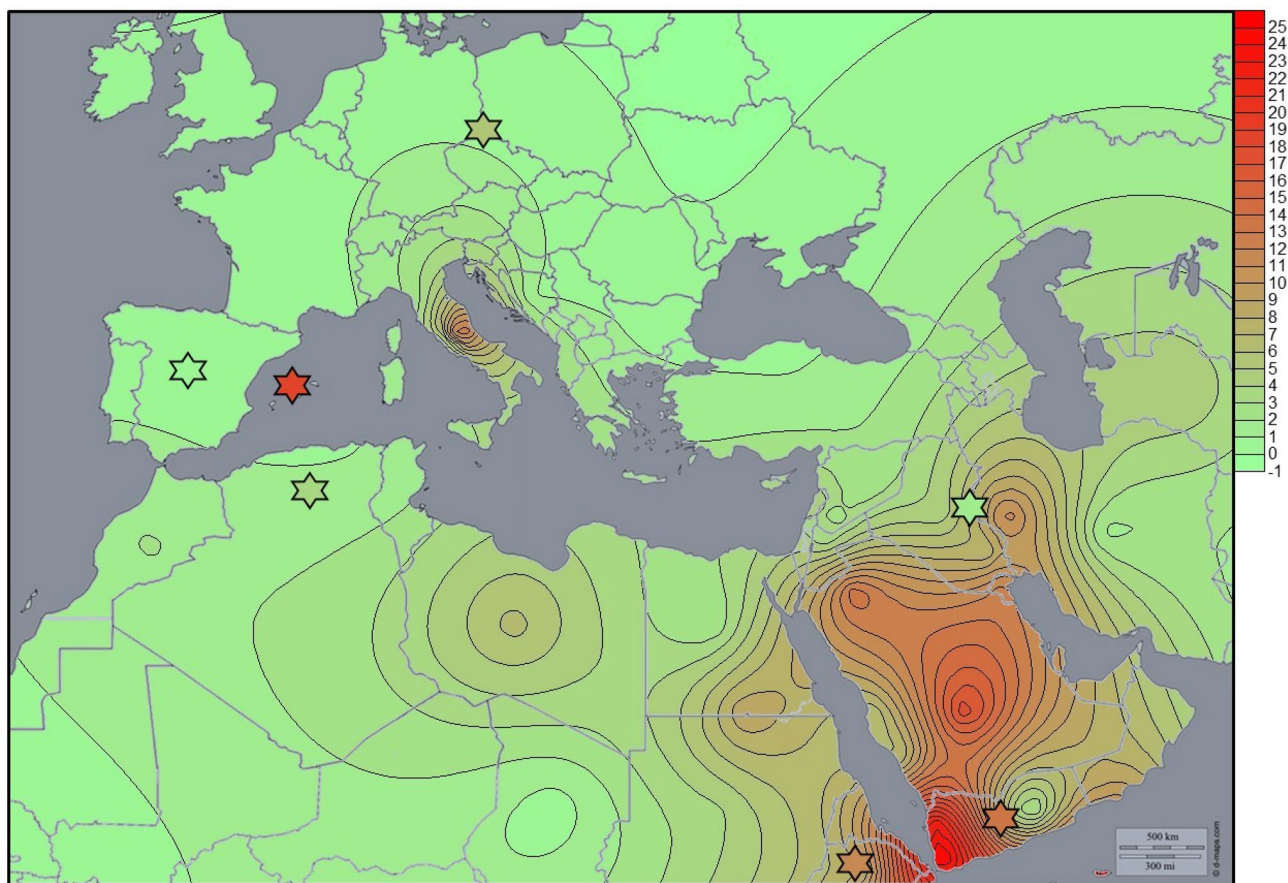




**Figure 4.** Multidimensional Scaling plot performed with (a) 11 SNPs: M1(xM35) E(xE1b1b1), M35 (E1b1b1), M213 (F), M201 (G), M170 (I), 12f2a (J(xJ2)), M172 (J2), M9(xM173) K(xR1), M173(xM17) R1(xR1a1), M17 (R1a1) and a final category for other haplogroups not included in these 10 SNPs and (b) Y-filer STRs. Circles defining each population are coloured following the legend code. Jewish populations are labelled with a Star of David. Populations and references are in Supplementary Table 5.

respectively<sup>77</sup>. Haplogroup J has higher frequencies in the Middle East and Arabia than in Europe (13–20% vs. 9%), while haplogroup T shows the opposite behaviour (10% in Europe and 8% in the Middle East)<sup>77,78</sup>. No significant differences were found between modern and ancient Majorcan populations regarding haplogroup frequencies since at least the Bronze Age<sup>79</sup>. T1a (5.8%) is considered one of the founders of the Chueta population, and originated in the Near East ~ 17 Kya, although most of its sub-branches seem to be European<sup>77</sup>.

Regarding the other mtDNA haplogroups with differential frequencies between Chuetas and their host population, Haplogroup U is the second most frequent in modern Europeans and was predominant in pre-agricultural



**Figure 5.** Isofrequency map of haplogroup R0a based on data from the literature (Supplementary Table 5) generated by Surfer v.8 (Golden Software Inc., Golden, Colorado) using a base map image (<http://d-maps.com>). Jewish populations are indicated with a Star of David.

Europe<sup>80</sup>. The U1a sub-haplogroup is dated at ~13–15 Kya and is present in Southwest and South Asia, the Caucasus, and Europe. Five U1a1a1 samples were found in the Chueta population, whereas no U1 sub-haplogroup was observed in the host population. Haplogroup K origin has been dated to ~36 Kya and, although the place of origin is still under discussion<sup>81,82</sup>, a Levantine origin seems the most likely. Sub-haplogroup K1a1b1a (dated to ~4.4 Kya<sup>82</sup>), found in four Chueta samples, is a founder lineage in Ashkenazim<sup>81</sup> and also present in Sephardic communities<sup>22</sup>, but not in non-European Jews, which can be seen as evidence of its European origin<sup>82</sup>.

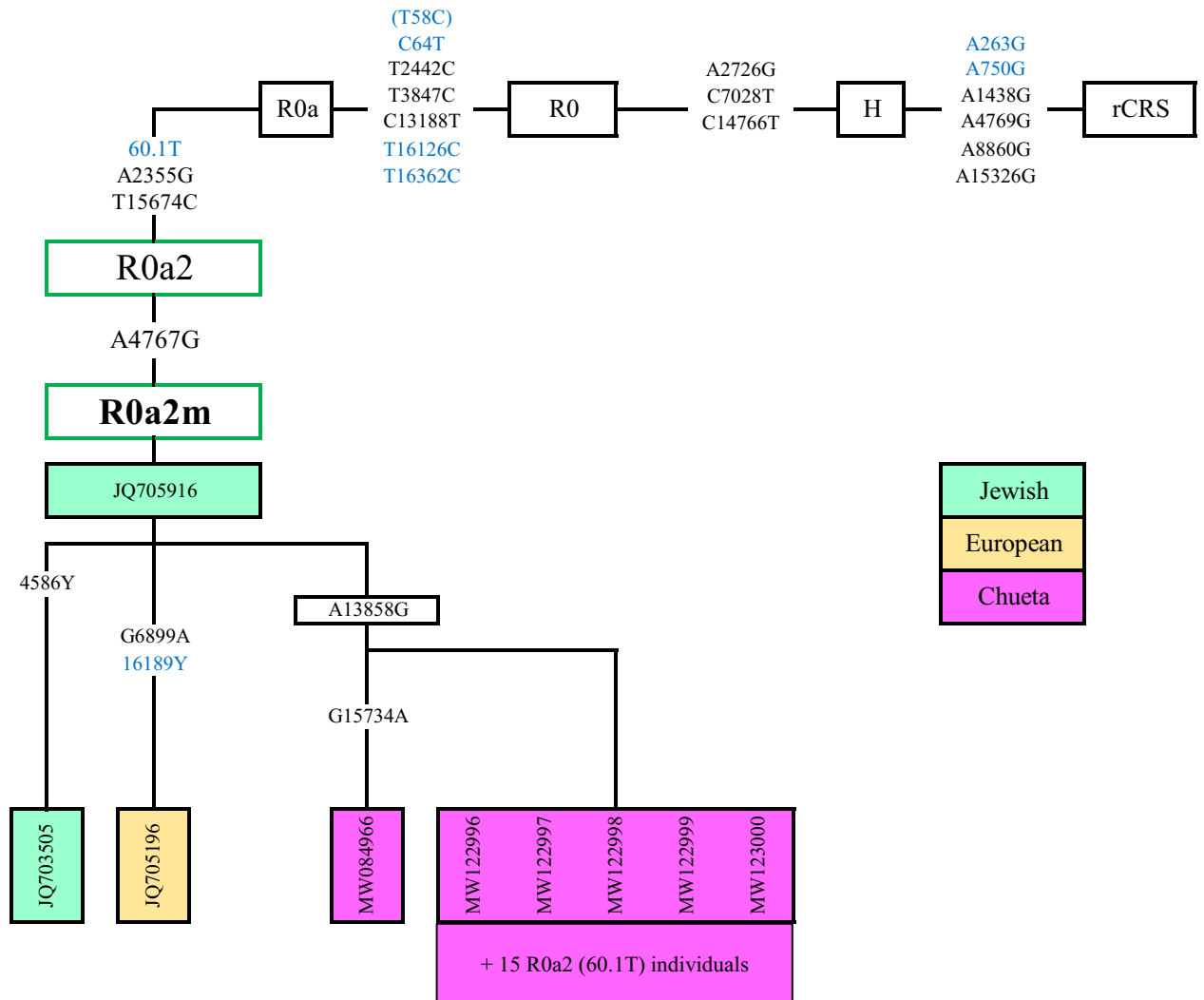
Haplogroup L3e is widespread in Africa but practically absent in Eurasia (except in neighbouring areas due to genetic exchange). Its origin is situated in Central or Eastern Africa about 46 Kya<sup>83</sup> and one of its most frequent lineages in West-Central Africa is L3e2b (7%)<sup>84</sup>. Five Chueta samples showed haplogroup L3e2b, while in the host population just one sample presented one African haplogroup L3, but belonging to another subclade, L3d.

Founding lineages in Chuetas, R0a + 60.1 T (~20%) and T1a (~6%), are different from those of other Sephardic populations, which also show important dissimilarity between each other. For instance, while HV0b is found as a founder in the Portuguese Crypto-Jewish communities from Belmonte and Bragança<sup>22–24</sup>, this haplogroup is absent or very uncommon in Chuetas and other Sephardic groups; additionally, haplogroup K1a1b1a, which is a founder (8.5%) in the Iberian Exile Jewish community from Bulgaria, has a lower frequency in other populations with Sephardic origin (4% in Chuetas and 0.8% in Turkish Jews)<sup>22</sup>.

## Conclusions

Genetic diversity in both paternal and maternal lineages in the Chueta population was higher than expected for a small, endogamous population. Comparable high diversity values were observed in the Portuguese Crypto-Jewish communities of Bragança<sup>23,24</sup>. These results reveal that demographic processes more complex than the loss of genetic diversity expected under conditions of extreme inbreeding and drift, have shaped the gene pool of both isolated populations. Ongoing data from recombinant markers, together with classical genealogical studies, will help to explain what mating strategies were undertaken by these communities to avoid the expected reduction of diversity, and also whether other factors, such as high heterogeneity in founder populations, could have contributed to the diversity observed.

The composition of the Chuetas' uniparentally transmitted lineages indicates a remarkable signature of Middle Eastern ancestry. In recent years, archaeogenetic research has shed light on the history of European and Middle Eastern populations, revealing a greater degree of population movements and interactions in the past than previously considered. In this context, from the analysis of current populations, it is difficult to infer what the genetic



**Figure 6.** Phylogenetic tree of the completed mtDNA sequences belonging to haplogroup R0a2m. Mutations relative to the revised Cambridge Reference Sequence (rCRS) (NC\_012920) are shown. Mutations coloured in blue correspond to the D-loop. The origin of the samples in the haplogroup are labelled in colours. In green, two Jewish samples from the Czech Republic and Ukraine (JQ705916 and JQ705196); and in yellow, one sample from Poland with unknown ethnicity (JQ703505). Finally, Chueta samples, with their corresponding GenBank accession numbers, are labelled in pink. The remaining 15 R0a + 60.1 T Chueta samples were found to be R0a2m (A4767G). Presence of the A13858G mutation was confirmed in all, whereas the G15734A mutation remained as a private variant of the MW084966 sample.

profile of the parental populations that gave rise to the Chuetas was exactly, and therefore precise inferences of the original source of the haplogroups found in this population, since there are no reasons to assume or assess the degree of genetic continuity. However, the most reasonable explanation for the differences found between Chuetas and their host population seems to be the Jewish origin of the Chuetas, considering the unquestionable historical evidence relating them with the Jewish populations who settled in Majorca long ago in the past. Therefore, our results would confirm that Chuetas have kept not only the cultural memory of their Jewish origin over centuries, but also a substantial degree of ancestral genetic signature.

In terms of paternal lineages, the results show that most Jewish communities are more similar to each other and to Middle Eastern populations than to their host populations. The Chueta population has the same behaviour, which can be observed by the high prevalence of haplogroups J2-M172 and J1-M267, and the lack of R1b-M269. Haplogroup distribution in Chuetas is very similar to other Sephardic communities, although in their gene pool there might be signatures of other Jewish communities' contribution, such as North African and Ashkenazim, which can be inferred from the presence of haplogroups such as E1b1b1a1-M78, Q1-P36.2, and R1a1a-M17.

The hallmark in the maternal gene pool in Chuetas is the presence of a new sub-branching of the rare Middle Eastern haplogroup R0a + 60.1 T as their modal haplogroup, and the low frequency of H. The presence of other haplogroups found in Jewish/Middle Eastern populations (K1a1b1a and U1a1a1) is also noteworthy. Current Jewish populations do not usually share modal maternal lineages, unlike the situation on the paternal side; not even populations with a supposed common ancestry, such as the Sephardic branch<sup>8,22</sup>. Various scenarios could explain this dissimilarity found in maternal founder lineages in distinct current populations with Sephardic

origin: differences in gene flow from and admixture with other populations resulting from the contrasting history of each community; a lack of homogeneity in maternal lineages of the original Sephardic groups that settled in different areas of the Iberian Peninsula (and Balearic Islands); or genetic drift in the current populations, resulting in a lack of lineages that do not allow us to infer the original mtDNA composition of the Sephardic Jews that lived in Spain and Portugal in the middle ages from the surviving lines. Tests based on genome-wide data that enable a greater genetic resolution, such as analysis of Runs of Homozygosity (ROH), together with ancient DNA analysis, could help to solve this question.

Received: 24 April 2020; Accepted: 19 November 2020

Published online: 08 December 2020

## References

1. Cortés, G. *Historia de los judíos mallorquines y sus descendientes cristianos*. (ed. Font, M.) (Palma, 1985).
2. Braunstein, B. *The Chuetas of Majorca. Conversos and the Inquisition of Majorca* (Ktav Pub Inc., New York, 1936).
3. de Muntaner, P. *Martí: Una família del brazo noble mallorquín durante el siglo XVII en Homenaje a Guillem Rosselló Bordoy*. Vol. 2. (Govern de les Illes Balears, Palma 2002).
4. Porqueres, E. *Lendogàmia dels xuetes de Malloca. Identitat i matrimoni en una comunitat de conversos (1435–1750)*. (ed. Lleopard, M.) (Palma, 2001).
5. Forteza, M. *Els descendents dels jueus conversos de Mallorca* (Moll, Palma, 1972).
6. Ostrer, H. & Skorecki, K. The population genetics of the Jewish people. *Hum. Genet.* **132**(2), 119–127 (2013).
7. Hammer, M. F. *et al.* Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. *Proc. Natl. Acad. Sci. U. S. A.* **97**(12), 6769–6774 (2000).
8. Thomas, M. G. *et al.* Founding mothers of Jewish communities: Geographically separated Jewish groups were independently founded by very few female ancestors. *Am. J. Hum. Genet.* **70**(6), 1411–1420 (2002).
9. Behar, D. M. *et al.* Contrasting patterns of Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations. *Hum. Genet.* **114**(4), 354–365 (2004).
10. Behar, D. M. *et al.* MtDNA evidence for a genetic bottleneck in the early history of the Ashkenazi Jewish population. *Eur. J. Hum. Genet.* **12**(5), 355–364 (2004).
11. Listman, J. B. *et al.* Identification of population substructure among Jews using STR markers and dependence on reference populations included. *BMC Genet.* **11**, 48 (2010).
12. Tian, J. Y. *et al.* A genetic contribution from the Far East into Ashkenazi Jews via the ancient Silk Road. *Sci. Rep.* **5**, 8377 (2015).
13. Atzmon, G. *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.* **86**(6), 850–859 (2010).
14. Behar, D. M. *et al.* The genomewide structure of the Jewish people. *Nature* **466**(7303), 238–242 (2010).
15. Behar, D. M. *et al.* No evidence from genome-wide data of a Khazar origin for the Ashkenazi Jews. *Hum. Biol.* **85**(6), 859–900 (2013).
16. Waldman, Y. Y. *et al.* The genetic history of Cochin Jews from India. *Hum. Genet.* **135**(10), 1127–1143 (2016).
17. Xue, J. *et al.* The time and place of European admixture in Ashkenazi Jewish history. *PLoS Genet.* **13**(4), e1006644 (2017).
18. Gladstein, A. L. & Hammer, M. F. Substructured population growth in the Ashkenazi Jews inferred with approximate bayesian computation. *Mol. Biol. Evol.* **36**(6), 1162–1171 (2019).
19. Adams, S. M. *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.* **83**(6), 725–736 (2008).
20. Velez, C. *et al.* The impact of Converso Jews on the genomes of modern Latin Americans. *Hum. Genet.* **131**(2), 251–263 (2012).
21. Marques, S. L. *et al.* Y chromosome diversity in a linguistic isolate (Mirandese, NE Portugal). *Am. J. Hum. Biol.* **28**(5), 671–680 (2016).
22. Behar, D. M. *et al.* Counting the founders: The matrilineal genetic ancestry of the Jewish Diaspora. *PLoS ONE* **3**(4), e2062 (2008).
23. Nogueiro, I. *et al.* Phylogeographic analysis of paternal lineages in NE Portuguese Jewish communities. *Am. J. Phys. Anthropol.* **141**(3), 373–381 (2010).
24. Nogueiro, I. *et al.* Echoes from Sepharad: Signatures on the maternal gene pool of crypto-Jewish descendants. *Eur. J. Hum. Genet.* **23**(5), 693–699 (2015).
25. Picornell, A. *et al.* Genetics of the Chuetas (Majorcan Jews): A comparative study. *Hum. Biol.* **69**(3), 313–328 (1997).
26. Tomás, C. *et al.* Genetic variability at nine STR loci in the Chueta (Majorcan Jews) and the Balearic populations investigated by a single multiplex reaction. *Int. J. Leg. Med.* **113**(5), 263–267 (2000).
27. Cambra, A. *et al.* MICA-HLA-B haplotype diversity and linkage disequilibrium in a population of Jewish descent from Majorca (the Balearic Islands). *Hum. Immunol.* **70**(7), 513–517 (2009).
28. Ferragut, J. F. *et al.* Genetic diversity of 38 insertion–deletion polymorphisms in Jewish populations. *Forensic Sci. Int. Genet.* **21**, 1–4 (2016).
29. Ferragut, J. F. *et al.* Genetic portrait of Jewish populations based on three sets of X-chromosome markers: Indels, Alu insertions and STRs. *Forensic Sci. Int. Genet.* **31**, e5–e11 (2017).
30. Nevo, S. *et al.* Orosomuroid (ORM1) polymorphism in Arabs and Jews of Israel: More evidence for a Middle Eastern origin of the Jews. *Hum. Biol.* **68**(2), 217–229 (1996).
31. Rodríguez, V. *et al.* Genetic substructure in western Mediterranean populations revealed by 12 Y-chromosome STR loci. *Int. J. Leg. Med.* **123**(2), 137–141 (2009).
32. Ferragut, J. F. *et al.* Founding mothers of Chueta population. *Forensic Sci. Int. Genet. Suppl. Ser.* **5**, e492–e494 (2015).
33. Brion, M. *et al.* Hierarchical analysis of 30 Y-chromosome SNPs in European populations. *Int. J. Leg. Med.* **119**(1), 10–15 (2005).
34. Gomes, V. *et al.* Digging deeper into East African human Y chromosome lineages. *Hum. Genet.* **127**(5), 603–613 (2010).
35. Roewer, L. *et al.* Continentwide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans. *PLoS Genet.* **9**(4), e1003460 (2013).
36. Myres, N. M. Y chromosome short tandem repeat DYS458.2 non-consensus alleles occur independently in both binary haplogroups J1–M267 and R1b3–M405. *Croat. Med. J.* **48**(4), 450–459 (2007).
37. Gusmão, L. *et al.* DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis. *Int. J. Leg. Med.* **120**(4), 191–200 (2006).
38. Marques, S. L. *et al.* Portuguese mitochondrial DNA genetic diversity—An update and a phylogenetic revision. *Forensic Sci. Int. Genet.* **15**, 27–32 (2015).
39. van Oven, M. PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Sci. Int. Genet. Suppl. Ser.* **5**, e392–e394 (2015).

40. Weissensteiner, H. *et al.* HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* **44**(W1), W58–W63 (2016).
41. Ramos, A. *et al.* Human mitochondrial DNA complete amplification and sequencing: A new validated primer set that prevents nuclear DNA sequences of mitochondrial origin co-amplification. *Electrophoresis* **30**(9), 1587–1593 (2009).
42. Ramos, A. *et al.* Validated primer set that prevents nuclear DNA sequences of mitochondrial origin co-amplification: A revision based on the New Human Genome Reference Sequence (GRCh37). *Electrophoresis* **32**(6–7), 782–783 (2011).
43. Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**(3), 564–567 (2010).
44. Bandelt, H. J. *et al.* Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**(1), 37–48 (1999).
45. Tokdemir, M. & Tunçez, F. T. Genetic polymorphisms of 17 Y-STR loci in Eastern Turkey population. *Gene Rep.* **6**, 15–18 (2017).
46. Hernández, C. L. *et al.* Paternal lineages in southern Iberia provide time frames for gene flow from mainland Europe and the Mediterranean world. *Ann. Hum. Biol.* **46**(1), 63–76 (2019).
47. Picornell, A. *et al.* Mitochondrial DNA HVRI variation in Balearic populations. *Am. J. Phys. Anthropol.* **128**(1), 119–130 (2005).
48. Solé-Morata, N. *et al.* Y-chromosome diversity in Catalan surname samples: Insights into surname origin and frequency. *Eur. J. Hum. Genet.* **23**(11), 1549–1557 (2015).
49. Roostalu, U. *et al.* Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: The Near Eastern and Caucasian perspective. *Mol. Biol. Evol.* **24**, 436–448 (2007).
50. Myres, N. M. *et al.* A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur. J. Hum. Genet.* **19**(1), 95–101 (2011).
51. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**(7555), 207–211 (2015).
52. Olalde, I. *et al.* The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**(7695), 190–196 (2018).
53. Olalde, I. *et al.* The genomic history of the Iberian Peninsula over the past 8000 years. *Science* **363**(6432), 1230–1234 (2019).
54. Fernandes, D. M. *et al.* The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean. *Nat. Ecol. Evol.* **4**(3), 334–345 (2020).
55. Bosch, E. *et al.* High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *Am. J. Hum. Genet.* **68**(4), 1019–1029 (2001).
56. Zalloua, P. A. *et al.* Identifying genetic traces of historical expansions: Phoenician footprints in the Mediterranean. *Am. J. Hum. Genet.* **83**(5), 633–642 (2008).
57. Shen, P. *et al.* Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. *Hum. Mutat.* **24**(3), 248–260 (2004).
58. Hammer, M. F. *et al.* Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish priesthood. *Hum. Genet.* **126**(5), 707–717 (2009).
59. Cinnioglu, C. *et al.* Excavating Y chromosome haplotype strata in Anatolia. *Hum. Genet.* **114**(2), 127–148 (2004).
60. Finocchío, A. *et al.* A finely resolved phylogeny of Y chromosome Hg J illuminates the processes of Phoenician and Greek colonizations in the Mediterranean. *Sci. Rep.* **8**(1), 1–9 (2018).
61. Semino, O. *et al.* Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: Inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am. J. Hum. Genet.* **74**(5), 1023–1034 (2004).
62. Francalacci, P. & Sanna, D. History and geography of human Y-chromosome in Europe: A SNP perspective. *J. Anthropol. Sci.* **86**, 59–89 (2008).
63. Manco, L. *et al.* The Eastern side of the Westernmost Europeans: Insights from subclades within Y-chromosome haplogroup J-M304. *Am. J. Hum. Biol.* **30**(2), e23082 (2018).
64. El-Sibai, M. *et al.* Geographical structure of the Y-chromosomal genetic landscape of the Levant: A coastal-inland contrast. *Ann. Hum. Genet.* **73**(6), 568–581 (2009).
65. Cruciani, F. *et al.* Tracing past human male movements in northern/eastern Africa and western Eurasia: New clues from Y-chromosomal haplogroups E-M78 and J-M12. *Mol. Biol. Evol.* **24**(6), 1300–1311 (2007).
66. Balanovsky, O. *et al.* Phylogeography of human Y-chromosome haplogroup Q3–L275 from an academic/citizen science collaboration. *BMC Evol. Biol.* **17**(1), 18 (2017).
67. Rootsi, S. *et al.* Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus. *Eur. J. Hum. Genet.* **20**(12), 1275 (2012).
68. Kayser, M. *et al.* Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. *Hum. Genet.* **117**(5), 428–443 (2005).
69. Underhill, P. A. *et al.* Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *Eur. J. Hum. Genet.* **18**(4), 479–484 (2010).
70. Rootsi, S. *et al.* Phylogenetic applications of whole Y-chromosome sequences and the Near Eastern origin of Ashkenazi Levites. *Nat. Commun.* **4**, 2928 (2013).
71. Chacón, J. M. Los judíos mallorquines en el comercio y en las redes de intercambio valencianas y mediterráneas del medievo. *Anales de la Universidad de Alicante. Revista de Historia Medieval* **15**, 75–85 (2009).
72. Abu-Amero, K. K. *et al.* Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. *BMC Evol. Biol.* **7**(1), 32 (2007).
73. Černý, V. *et al.* Internal diversification of mitochondrial haplogroup R0a reveals post-last glacial maximum demographic expansions in South Arabia. *Mol. Biol. Evol.* **28**(1), 71–78 (2011).
74. Gandini, F. *et al.* Mapping human dispersals into the Horn of Africa from Arabian Ice Age refugia using mitogenomes. *Sci. Rep.* **6**, 25472 (2016).
75. Messina, F. *et al.* Traces of forgotten historical events in mountain communities in Central Italy: A genetic insight. *Am. J. Hum. Biol.* **27**(4), 508–519 (2015).
76. Picornell, A. *et al.* Mitochondrial DNA sequence variation in Jewish populations. *Int. J. Leg. Med.* **120**(5), 271–281 (2006).
77. Pala, M. *et al.* Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *Am. J. Hum. Genet.* **90**(5), 915–924 (2012).
78. Fernandes, V. *et al.* Genetic stratigraphy of key demographic events in Arabia. *PLoS ONE* **10**(3), e0118625 (2015).
79. Simón, M. *et al.* Dissecting mitochondrial DNA variability of balearic populations from the bronze age to the current era. *Am. J. Hum. Biol.* **29**(1), e22883 (2017).
80. Davidovic, S. *et al.* Mitochondrial super-haplogroup U diversity in Serbians. *Ann. Hum. Biol.* **19**, 1–11 (2017).
81. Behar, D. M. *et al.* The matrilineal ancestry of Ashkenazi Jewry: portrait of a recent founder event. *Am. J. Hum. Genet.* **78**(3), 487–497 (2006).
82. Costa, M. D. *et al.* A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. *Nat. Commun.* **4**, 2543 (2013).
83. Bandelt, H. J. *et al.* Phylogeography of the human mitochondrial haplogroup L3e: A snapshot of African prehistory and Atlantic slave trade. *Ann. Hum. Genet.* **65**(6), 549–563 (2001).
84. Cerezo, M. *et al.* Comprehensive analysis of Pan-African mitochondrial DNA variation provides new insights into continental variation and demography. *J. Genet. Genom.* **43**(3), 133–143 (2016).

## Acknowledgements

This work was partially supported by grant AAEE246/2014 from the Direcció General de R + D + I (Comunitat Autònoma de les Illes Balears) and European Regional Development Fund (ERDF) and FCT, Portuguese Foundation for Science and Technology PTDC/ATP-DEM/4545/2012 project financed by European Social Funds (COMPETE-FEDER). We are grateful to Maria Trinidad Garcia (from the Serveis Científicotècnics of the Universitat de les Illes Balears) for her assistance and help with capillary electrophoresis procedures, María Luz Gómez-Barbeito for her help in mtDNA sequencing and Meryl Wyn Jones for the English language corrections. We would also like to sincerely thank all the people who volunteered to participate in this study.

## Author contributions

J.F.F. collected the samples, conducted laboratory analysis, analysed data, and participated in drawing figures; C.R. participated in the discussion and reviewing the manuscript; C.R. and L.A. provided financial support; J.A.C. help with the statistical analyses; A.A. and L.A. participated in the design and coordination of the study. A.P. participated in the design and coordination of the study and wrote the manuscript with support from J.F.F.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-78487-9>.

**Correspondence** and requests for materials should be addressed to J.F.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020