

SCIENTIFIC REPORTS

OPEN

Genetic structure in the paternal lineages of South East Spain revealed by the analysis of 17 Y-STRs

María Saiz¹, María Jesus Alvarez-Cubero², José Antonio Lorente^{1,3}, Juan Carlos Alvarez¹ & Luis Javier Martinez-Gonzalez³

The genetic data of 17Y chromosome short tandem repeats in 146 unrelated donor residents in the provinces of Granada, Málaga, and Almería (GMA) were analyzed to determine the genetic legacy of the male inhabitants of the former Kingdom of Granada. A total of 139 unique haplotypes were identified. Observed allele frequencies and haplogroup frequencies were also analyzed. By AMOVA and STRUCTURE analysis, the populations of the 3 provinces could be treated genetically as a single population. The most frequent haplogroup was R1b1b2 (58.22%). By network analysis of all individuals, we observed a distribution according to haplogroup assignment. To improve the characterization of GMA population, it was compared with those of North Africa, the Iberian Peninsula, and southern Europe. In our analysis of allele frequencies and genetic distances, the GMA population lay within the Spanish population group. Further, in the STRUCTURE analysis, there was no African component in the GMA population, confirming that, based on our genetic markers, the GMA population does not reflect any male genetic influence of the North African people. The presence of African haplogroups in the GMA population is irrelevant when their frequency is compared with those in other European populations.

For the nearly 800 years of the Arab invasion of the Iberian Peninsula, North African groups spread throughout all the territory except for the Basque Country, Galicia, Cantabria, Asturias, and most of the Pyrenees; but, their influence was greater in the south of the Peninsula¹. The former Kingdom of Granada comprised the present-day provinces of Granada, Málaga, and Almería, in addition to parts of Cádiz, Jaén, Córdoba, and Sevilla, with Granada as the capital. During the 14th and 15th centuries, it was one of the most prosperous cities in Europe, with almost 165,000 inhabitants².

The first evidence of African invaders originated from 711, when Syrian Berbers entered the Iberian Peninsula and conquered the Granada region, known as *Iliberir*. In 1013, the Zirid dynasty gained control of the region and established *Ilbira* in 1025. To prevent future invasions, the kingdom expanded and occupied the entire territory of the Kingdom of Granada^{3,4}. With the expansion of the Kingdom in 1090, a large part of the Iberian Peninsula came under control of the Zirid dynasty, known as *Al-Ándalus*. After losing the battle of Navas de Tolosa, in 1212, *Al-Ándalus* became subdued and was reduced to the Nasrid Kingdom of Granada. The Nasrid dynasty was the longest surviving Muslim dynasty in the Iberian Peninsula. Ultimately, with the confiscation of the city of Granada by *Los Reyes Católicos*, *Fernando II*, and *Isabel I* in 1492, the Kingdom of Granada came to an end^{3,4}.

Although the Muslims signed capitulation agreements to follow the religion of the kingdom, they were forced to convert to Christianity or emigrate. Once the Morisco properties were expropriated and expelled, it became necessary to repopulate the region by sending new inhabitants from various regions of the Peninsula. The repopulation began in 1571 and lasted until 1595, by which time a total of 12,546 families repopulated 270 areas³. On the

¹University of Granada, Laboratory of Genetic Identification, Department of Legal Medicine, Toxicology and Physical Anthropology, Faculty of Medicine, PTS. Avd. Investigación 11, 18016, Granada, Spain. ²University of Granada, Department of Biochemistry and Molecular Biology III and Immunology, Faculty of Medicine, PTS. Avd. Investigación 11, 18016, Granada, Spain. ³GENYO, Centre for Genomics and Oncological Research: Pfizer/University of Granada/Andalusian Regional Government, PTS. Avenida de la Ilustración, 114, 18016, Granada, Spain. Correspondence and requests for materials should be addressed to J.A.L. (email: jose.lorente@genyo.es)

December 9, 1609, *Felipe III* signed the expulsion warrant of all Moriscos from Spain^{3,4}. In 1833, after 314 years of existence and with the separation of the provinces of Almería and Málaga, the former Kingdom of Granada ended^{3,4}.

Thus, during the establishment of the Kingdom of Granada and throughout its existence, people from diverse religions and regions coexisted. The coexistence of Visigoths, Syrians, Saharans, Moroccans, Jews, and Christians is reflected in the architecture, culture, and folklore of the present-day cities of Granada, Málaga, and Almería.

The study of mtDNA and Y chromosomes has identified geographic regions with a genetic influence of North Africans of 8% to 10%⁵⁻⁷. Crossbreeding studies that are based on the characterization of ALU sequences have found traces of sub-Saharan genes in north Mediterranean populations, suggesting continuous contact between both coasts⁸. That genetic traits and certain specific haplotypes have been detected along the northern coast of the Mediterranean Sea confirms the hypothesis that gene flow in this region was linked to the first trans-Mediterranean sailings and remained homogeneous while the slave trade lasted, until the late 17th century, rather than reflecting Islamic expansion (S. VII to S. XV)⁸. Y chromosome studies have described the genetic structure in the Iberian Peninsula, calculating the various contributions of Muslims and Jews to the current population of the Peninsula⁶, focusing on the E3b2 haplogroup, which is common in northern Africa and present in 5.6% of the Peninsula⁹. The distribution of the Y chromosome haplogroup E-M81 in the Iberian Peninsula suggests genetic flow of North Africa during this period⁶. High levels of patrilineal descent from North Africa and Sephardic Jews in the current population of the Iberian Peninsula have been detected⁶ and contributed to the higher genetic diversity in southwestern Europe¹⁰.

However, contrary to what might be expected based on historical data that favor a gradient of North African genetic influence from south to north, most such influence has been found in Galicia and northern Castilla (>20%)⁶. The main gradient of the frequencies of North African genes descend from west to east¹¹. Furthermore, recent studies based on autosomal SNPs¹¹ and Y-chromosome lineages¹² reveal that Andalusian population does not specially cluster with North African populations more than other Iberian populations¹³. After the Reconquest, the Moors were distributed homogeneously throughout the Peninsula, but their final expulsion in 1609 was absolute in certain regions of Spain, Valencia, and western Andalusia, whereas in Galicia and Extremadura, the population dispersed and integrated into society⁶.

Results

Y-STR allele frequencies. Allele frequencies and forensic summary statistics for the 16 loci are summarized in Table 1. The most informative markers were DYS385 and DYS458, and the least informative marker was DYS393. The combined discriminatory power was $1-6.17168 \cdot 10^{-8}$. A total of 139 unique haplotypes were detected, 133 of which were described once, 5 that were described twice, and 1 that was described 3 times. The Y chromosome haplotype diversity was 0.9974, and the discriminatory capacity was 95.21% [(N haplotypes/n)-100].

Population substructure. By AMOVA, there was no significant genetic substructure in the GMA population ($p > 0.05$). All variations were within populations (99.93%) rather than between populations (0.07%) (Table 2).

Y chromosome haplogroups. Based on the 17 Y chromosome STR markers, most men in the GMA population carried the R1b1b2 haplogroup (58.22%). The second most frequent haplogroup was E1b1b1; many sub-haplogroups were detected (11.64%), of which E1b1b1b was the most frequent (4.79%). J2a4 had a proportion of 6.16%. Other subhaplogroups from J included J1c3 (4.11%) and J1-M267, J2b1, and J2b2f (0.68%). The G2a-P15 haplogroup was observed in 5.48% of samples (Table 3).

Network analyses were performed for all haplogroups ($n = 79$) (Fig. 1) and those in the R1b1b2 haplogroups ($n = 85$) (Supplementary Fig. 1), derived from the data on 10 markers with mutation rates under 0.0025 (DYS19, DYS389I, DYS390, DYS391, DYS392, DYS393, DYS385, DYS438, DYS437, and DYS448). Within both networks, no portioning of populations was observed by province, and individuals from the 3 provinces were distributed randomly throughout the network.

Individuals were then distributed by haplogroup (Fig. 1). The central node of the network contained the G2a, J2a, and R1b1b2 haplogroups. All individuals from the R1b1b2 haplogroup converged in 2 nodes (8 and 9 individuals), after which the remainder of the groups was created.

Surname analysis. Twenty repeated surnames were identified among the 108 surnames of the 146 individuals (58 individuals), most of which were associated with the R1b1b2-M269 haplogroup (28/58). Haplogroups from lineages E (E1b1b1c-M123, E1b1b1a1b-V13, E1b1b1b1b-M81, E1b1b1*-M35, E1b1b1a1d-V65, E1b1a1g-U175, and E1b1b1a1c-V22), I (I2a-P37.2 and I2b1-M233), J (J2a4-L26, J2b2f-L283, and J1c3-P58), and G2a-P15 were also common between surnames. Table 4 shows the composition of Y chromosome haplogroups for the 4 most frequent surnames in the sample and for double and triple surnames. R1b1b2-M269 was the most common haplogroup in this surname set, although other common European and North African haplogroups were detected. In addition, 6 Spanish surnames of Arab origin were detected, all of which were singletons—5 were linked to Y chromosome R1b1b2-M269, whereas the remainder was associated with the J1c3-P58 lineage. Network analysis on these 6 individuals did not reveal any one of them to be a specific ancestor of the other subjects (data not shown).

Unrelated men who have the same surname are significantly more likely to share haplotypes^{14,15}. However, in the GMA population, none of the individuals who shared haplotypes had the same surname. Further, except for Martín and Gomez, the remaining individuals who shared haplotypes had unique surnames. Six of the 108 different surnames in the GMA population were Arab in origin; only Silla was associated with the J1c3-P58 lineage, a characteristic of the haplogroup in the population of the Arabian Peninsula.

Population cross-comparisons. Y chromosome allele frequencies from the GMA population and published frequencies (Supplementary Table 1) were used to perform correspondence analysis (Fig. 2). The first 2 main components, together accounting for >50% of the total variance, suggested proximity of the GMA population to Spanish and European populations, clearly occupying the same genetic space.

Allele	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS393	DYS391	DYS439	DYS635	DYS392	GATA_H4	DYS437	DYS438	DYS448	Allele	DYS385 A/B
8								0.007						0.007		10.14	0.014
9								0.068				0.014		0.110		11.11	0.034
10								0.548	0.082		0.007	0.034		0.274		11.13	0.041
10,11									0.014							11.14	0.336
11								0.363	0.301		0.336	0.411		0.055		11.15	0.062
12	0.014	0.130					0.171	0.014	0.466		0.062	0.473		0.534		11.16	0.007
13	0.027	0.589				0.144	0.719			0.137	0.541	0.062		0.021		11.18	0.007
14	0.041	0.274			0.048	0.644	0.110				0.041	0.007	0.459			12.12	0.034
15	0.486	0.007			0.103	0.158					0.014		0.479			12.13	0.007
16	0.349				0.144	0.034							0.062			12.14	0.062
16.2					0.007											12.15	0.014
17	0.055				0.370	0.021										12.16	0.007
17.2					0.034											13.13	0.007
18	0.027				0.178										0.166	13.14	0.055
18.2															0.007	13.15	0.021
19					0.068										0.503	13.16	0.048
20					0.048					0.075					0.207	13.18	0.014
21			0.021							0.205					0.090	13.19	0.021
22			0.068							0.116						14.14	0.041
23			0.301							0.486					0.028	14.15	0.014
24			0.541							0.110						14.16	0.034
25			0.068							0.007						14.18	0.007
27				0.021												14.19	0.007
28				0.096												15.15	0.014
29				0.432												15.16	0.007
30				0.349												15.17	0.014
31				0.082												15.18	0.007
32				0.021												15.19	0.007
																16.17	0.014
																16.18	0.007
																18.18	0.007
PD	0.635	0.561	0.607	0.676	0.790	0.538	0.441	0.563	0.667	0.690	0.589	0.603	0.556	0.624	0.667		0.864
PIC	0.573	0.495	0.546	0.619	0.766	0.498	0.398	0.480	0.611	0.652	0.518	0.523	0.456	0.570	0.624		0.859
MP	0.365	0.439	0.393	0.325	0.210	0.462	0.559	0.437	0.333	0.310	0.411	0.397	0.444	0.376	0.333		0.136

Table 1. Allele frequencies and forensic summary statistics of the Yfiler STR loci found in the GMA population sample.

Structure design	Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
3 Groups:	Among populations	2	1.032	0.00036 Va	0.07
Group 1 = Granada Group 2 = Málaga Group 3 = Almería	Within populations	143	71.413	0.49939 Vb	99.93
	Total	145	72.445	0.49975	Fixation index FST: 0.00071

Table 2. AMOVA design and results from 146 individuals (70 from Granada, 42 from Málaga and 34 from Almería).

Two main groups were observed in the analysis of Y chromosome allele frequencies (Fig. 2). One comprised the Iberian Peninsula and Mediterranean populations. Due to the frequencies of marker DYS438 (alleles 12 and 13) and marker DYS392 (allele13), the Basque Country population¹⁶ lay far from the rest of the Spanish populations^{16,17}. The other group was composed of North African populations^{18–23}; the Algerian population²⁰ resided outside of this group.

The genetic distances confirm the results of the genetic frequency analysis (Fig. 3). When Rst values were visualized with Surfer, the populations with the closest affinity to the GMA population (populations within the blue gradient lines) were from other regions of the Iberian Peninsula^{24,25} and nearby Mediterranean populations^{26–31}.

_IJ M429	E1b1a1g U175	E1b1b1 M35	E1b1b1a1b V13	E1b1b1a1cV22		E1b1b1a1d V65	E1b1b1bM81	E1b1b1cM123
0.0068	0.0068	0.0068	0.0205	0.0137		0.0137	0.0479	0.0137
F-M89	G2a-P15	H1a-M82	I2a-P37.2	I2a2a-M223		I2a2-S154	J1*-M267*	J1c3-P58
0.0068	0.0548	0.0068	0.0205	0.0274		0.0068	0.0068	0.0411
J2a4 L26	J2b1 M205	J2b2f-L283	P-M45	Q1a MEH2	Q1b M378	R1b1b2-M269	T1a1a1 P77	T1a2 L131
0.0616	0.0068	0.0068	0.0068	0.0137	0.0068	0.5822	0.0068	0.0068

Table 3. Y chromosome haplogroup frequencies in the GMA population.

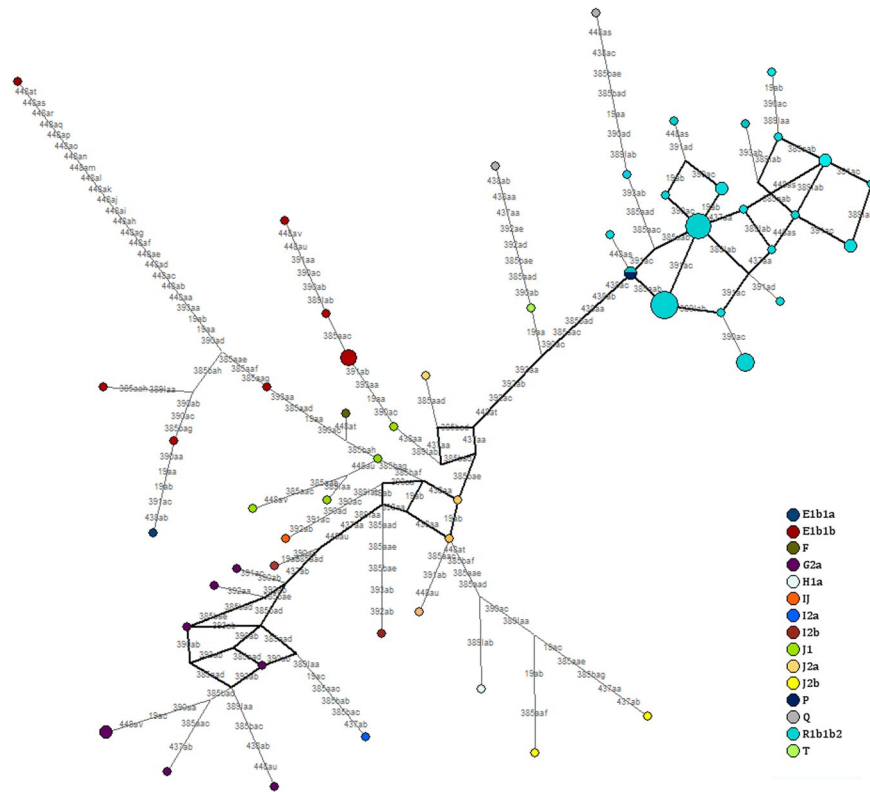


Figure 1. Network for Y chromosome haplogroups.

The Libyan and Tunisian populations had the biggest differences compared with the GMA populations (darker red gradient). Studies in the Libyan population support the importance of migratory movements that lead to admixture between the original Berber inhabitants and neighboring and more distant populations, despite a solid Berber genetic background remaining³². Berber populations inhabited the Iberian Peninsula for almost 250 years³ but were completely expelled, and very few genetic substrata can be detected in the GMA population, as evidenced by the high genetic differences between the GMA population and those with a significant Berber genetic influence.

The Rst pairwise distances between the GMA and European and African populations indicated that prior to Bonferroni adjustments; all pairwise comparisons generated statistically significant differences.

Discussion

The populations of the provinces of Granada, Malaga, and Almeria behaved as a single population with regard to their genetic structure. These results are consistent with the historical and cultural expectations, because these 3 provinces share the same origin and proximity over a large area. We are aware that 146 individuals is a small sample size for 3 provinces. Reduced sample sizes in population and evolutionary studies may underscore the allelic frequencies of certain alleles, underestimating the number of alleles that are detected^{33,34}. However, studies that have compared different sample sizes have revealed that, after increasing sample size, newly detected alleles are rare³⁴. In this case, nearly all alleles represented in the allelic ladder are represented in the population and all alleles have frequencies that are similar to those in nearby populations²⁴.

Based on the analysis of 17 Y chromosome STR markers, most men in the GMA population carried the R1b1b2 (haplogroup 58.22%), which is the most common European haplogroup, increasing in frequency east to west. This gradient indicates that this haplogroup spread throughout Europe from a single source in the Near East during the Neolithic period³⁵. The second most frequent haplogroup was E1b1b1, but many subhaplogroups were seen (11.64%), of which E1b1b1b was the most frequent (4.79%). The highest frequencies of this haplogroup were

	Doubletons	Trios	Rodríguez	García	Sánchez	Fernández
n	20	18	4	5	5	6
E1b1b1*-M35(xM78.M81.M123)	0.05					
E1b1b1a1b-V13	0.05	0.056				
E1b1b1b-M81	0.1					0.167
E1b1a1g-U175			0.25			
E1b1b1a1c-V22			0.25			
E1b1b1c-M123	0.05	0.056				
E1b1b1a1d-V65		0.056				
I2a-P37.2	0.05	0.056			0.2	
I2b1-M223	0.05	0.111			0.2	
F-M89		0.056				
G2a-P15	0.05				0.2	
H1a-M82			0.25			
J2a4-L26	0.05	0.056		0.2		
J2b2f-L283						0.167
J1c3-P58						0.333
Q1a-MEH2	0.05					
R1b1b2-M269	0.45	0.556	0.25	0.8	0.4	0.333
T1a1b-P77	0.05					

Table 4. Y chromosome haplogroup frequencies in surnames with more than 1 occurrence in the Andalusian sample.

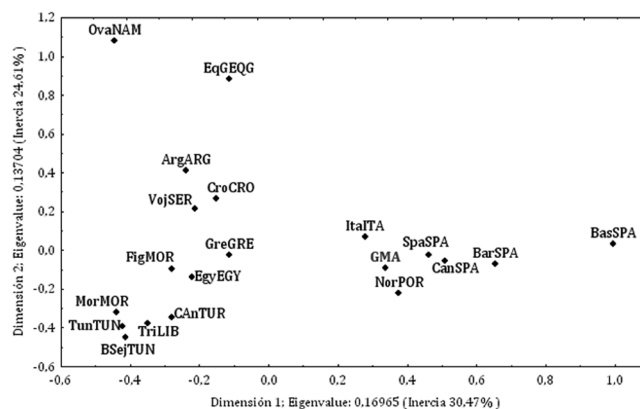


Figure 2. Correspondence analysis between GMA population and European and African populations (MorMOR, Morocco; FigMOR, Figuig Oasis from Morocco; ArgARG, Oran Area from Algeria; BSejTUN, Berbers from Sejenane from Tunisia; TunTUN, Tunisia; EgvEGY, Egypt; TriLIB, Tripoli Region from Libya; CAnTUR, Central Anatolia from Turkey; BasSPA, Basque Country from Spain; CanSPA, Cantabria from Spain; BarSPA, Barcelona from Spain, SpaSPA, Spain; NorPOR, North of Portugal; ItaITA, Italy; GreGRE, Greece; CroCRO, Croatia; VojSER, Vojvodina from Serbia; HolHOL, Holland; EqGEQG, Equatorial Guinea; OvaNAM, Ovambo from Namibia).

detected in the north of Africa, at >80% in the Moroccan Berbers³⁶. However, this haplogroup is rare in Europe, except in the Iberian Peninsula (5% of the individuals)^{6,36}. Studies based on whole Y-chromosome sequences attribute these frequencies to genetic drift acting on this low-frequency variant³⁷. J2a4 had a proportion of 6.16%. The other subhaplogroups from J that were observed were J1c3 (4.11%), typically seen in Arabian Peninsula populations (40% to 75% male lineages)³⁸ and J1-M267, J2b1, J2b2 (0.68%)—both lineages (J1 and J2) are detected primarily in southeast Europe. The G2a-P15 haplogroup was observed in 5.48% of samples (Table 2).

Based on the Y chromosome haplogroups, there was a large difference between the North African and Spanish populations—specifically the GMA population (Fig. 4). The main haplogroup in North African populations was E, which was also found in the Spanish population but at lower frequencies and similar to those in other European populations³⁶. Recent studies based on the analysis of haplogroup E1b1b-M81 suggest that Andalusian Y chromosomes may derive from a single common ancestor imported from North Africa that originated in the Andalusian present-day haplotypes¹². Analyzing the second most common Y chromosome haplogroup among North African populations (haplogroup J), the GMA population had high values of this haplogroup. However, a detailed evaluation of J subhaplogroups showed that the prevalent haplogroup in the GMA population was lineage J2a4, found along the northern coast of the Mediterranean and the Caucasus, whereas the predominant

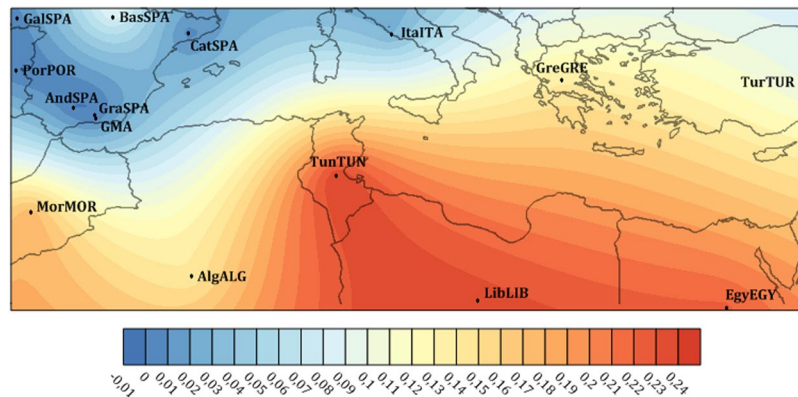


Figure 3. Map of Y STR RST genetic distances between the GMA population and Mediterranean and North African populations plotted with Surfer 13 using the Kriging method dark blue corresponds to lower values of genetic distances and dark red to higher values of genetic distances (MorMOR, Morocco; AlgALG, Oran Area from Algeria; TunTUN, Tunisia; EgyEGY, Egypt; LibLIB, Libya; TurTUR, Turkey; BasSPA, Basque Country from Spain; GalSPA, Galicia from Spain, CatSPA, Catalonia from Spain, AndSPA, Andalusia from Spain; GraSPA, Granada from Spain; PorPOR, Portugal; ItaITA, Italy; GreGRE, Greece).

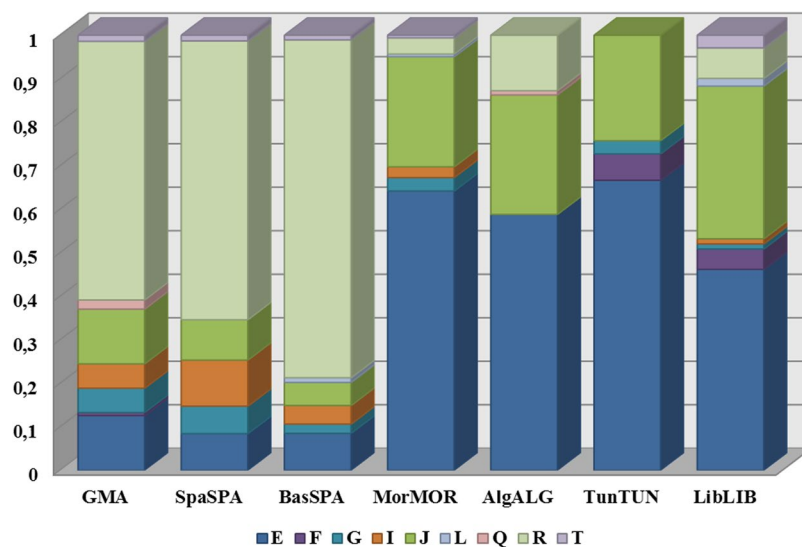


Figure 4. Histograms of Y chromosome haplogroup frequencies in the GMA, Spanish, and North African populations (BasSPA, Basque Country from Spain; SpaSPA, Spain; MorMOR, Morocco; AlgALG, Oran Area from Algeria; TunTUN, Tunisia; LibLIB, Libya; TurTUR, Turkey).

subhaplogroup in the North African populations were J1 lineages, existing at low frequencies in subpopulations of southern Europe.

The analyses of allele frequencies and genetic distances confirm the results of the haplogroup analyses. Mediterranean populations were genetically closest to the GMA population, especially the Spanish populations. The genetic data revealed that no significant African component remained in the genetic legacy of the population of the southern Iberian Peninsula compared to other Iberian and European populations, despite North African people living in the region for almost 800 years. Similar results have been recently reported. An analysis of Y-chromosome lineages in the Andalusian population demonstrated that Andalusia and other Iberian populations are related to North African populations on a larger scale than other European regions but that if South European populations, which historically have been influenced by North African populations, are included in the analysis, this influence is not significant enough¹².

Surnames are used to enhance the genetic signals of a population structure^{39,40}. In most western societies, they are transmitted through the male line and inherited as alleles^{41,42}, which is their transmission should closely match the inheritance of the Y chromosome. The spectrum of surnames in this study is supported by history. Most of the surnames originated in the north and center of the Peninsula but are more frequently observed in the South.

The analysis of Y chromosomes in the GMA population indicated that the male influence of the North African inhabitants that could have remained in the population in the southern peninsula did not influence the genetic

legacy of the population in this region more intensely than in other Iberian populations. After the Reconquest of the region by the Catholic Monarchs, the region was repopulated with entire families from the rest of the peninsula³. Although many of the Moriscos who inhabited the region were converted to Christianity and although mixed marriages might have formed, the Y chromosome lineages indicate that this phenomenon occurred at such a low frequency that its small influence prevented it from surviving the 600 years that have passed since the dissolution of the Kingdom of Granada.

Methods

Population samples. A total of 146 buccal cell swab samples from unrelated individuals were collected. Subjects were selected according to their self-declared affiliation to the provinces of Granada, Málaga, and Almería and their residence in the region for at least 3 generations (Supplementary Fig. 1). This study was approved by the Ethics Committee of the University of Granada (Approval Number: 885). All subjects gave their informed consent per the Declaration of Helsinki. All methods were performed in accordance with the relevant guidelines and regulations of the University of Granada.

Surnames study. The first surnames of all 146 unrelated individuals were queried and annotated to establish the relationship between the distribution of haplotypes and haplogroups and the observed surnames. Surnames were compared with an available list of Spanish surnames of Arab origin⁴³.

DNA extraction. Genomic DNA was extracted using phenol/chloroform/isoamyl alcohol and proteinase K. The DNA was purified on Amicon[®] 100 units (Millipore) and quantified on an 0.8% agarose gel.

Y-STR genotyping. One hundred forty-six samples were amplified using the AmpF/STR[®] Yfiler[®] PCR Amplification kit (Applied Biosystems, Foster City, CA), per the manufacturer. Alleles were separated and detected on an Applied Biosystems ABI 310 genetic analyzer. Fragment sizes were analyzed using GeneMapper ID-X v1.1 (Applied Biosystems, Foster City, CA). The alleles were named according to the number of repeated units, based on the sequenced allelic ladder (ISFG recommendations).

All individuals with complete genotypes were deposited into the YHRD database⁴⁴ (accession numbers YA004154–YA004156).

Statistical analyses. Y STR allele frequencies, polymorphism information content (PIC), power of discrimination (PD), and matching probability (MP) were calculated for each locus using PowerStats v12⁴⁵. Y chromosome haplotypes were calculated with Arlequin v3.5.1.3. Analysis of molecular variance (AMOVA) was performed with Arlequin v3.5.1.3 to determine any possible population substructure.

Y chromosome haplogroups were determined with the 23-Haplogroup Beta program (online software)^{46,47} and YPredictor by Vadim Urasin v15.0.

Network analyses were performed for individuals representing all haplogroups reducing the sample size to seventy nine individuals and individuals R1b1b2 (n = 85) to determine the most common ancestor.

Correspondence analysis with Y-STR allele frequencies from published populations (Supplementary Table 1) was performed with Statistica v9.1. Y-STR haplotype data were used to calculate pairwise genetic distances (Rst) and the corresponding p-values. The genetic distances were then used to construct a contour map with Surfer13.

References

- Brion, M., Salas, A., González-Neira, A., Lareu, M. V. & Carracedo, A. Insights into Iberian population origins through the construction of highly informative Y-chromosome haplotypes using biallelic markers, STRs, and the MSY1 minisatellite. *Am. J. Phys. Anthropol.* **122**, 147–161 (2003).
- Chandler, T. *Four Thousand Years of Urban Growth: An Historical Census*. (St. David's University Press, 1987).
- Bueno, P. *El reino de Granada (De orígenes a 1936)*. (Don Quijote Editorial, 2004).
- Garzón, M. *Historia de Granada. I* (Gráficas del Sur, 1980).
- Pino-Yanes, M. *et al.* North African influences and potential bias in case-control association studies in the Spanish population. *PLoS One* **6**, e18389 (2011).
- 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**, 725–36 (2008).
- Maca-Meyer, N. *et al.* Y chromosome and mitochondrial DNA characterization of Pasiegos, a human isolate from Cantabria (Spain). *Ann. Hum. Genet.* **67**, 329–339 (2003).
- González-Pérez, E. *et al.* Population relationships in the Mediterranean revealed by autosomal genetic data (Alu and Alu/STR compound systems). *Am. J. Phys. Anthropol.* **141**, 430–9 (2010).
- Flores, C. *et al.* Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. *Eur. J. Hum. Genet.* **12**, 855–63 (2004).
- Botigué, L. R. *et al.* Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. *Proc. Natl. Acad. Sci. USA* **110**, 11791–6 (2013).
- Bycroft, C. *et al.* Patterns of genetic differentiation and the footprints of historical migrations in the Iberian Peninsula. *bioRxiv* 250191, <https://doi.org/10.1101/250191> (2018).
- Rey-González, D. *et al.* Micro and macro geographical analysis of Y-chromosome lineages in South Iberia. *Forensic Sci. Int. Genet.* **29**, e9–e15 (2017).
- Larmuseau, M. H. D. & Ottoni, C. Mediterranean Y-chromosome 2.0—why the Y in the Mediterranean is still relevant in the postgenomic era. *Ann. Hum. Biol.* **45**, 20–33 (2018).
- King, T. E. & Jobling, M. A. Founders, drift, and infidelity: The relationship between y chromosome diversity and patrilineal surnames. *Mol. Biol. Evol.* **26**, 1093–1102 (2009).
- Bowden, G. R. *et al.* Excavating past population structures by surname-based sampling: The genetic legacy of the vikings in Northwest England. *Mol. Biol. Evol.* **25**, 301–309 (2008).
- Núñez, C. *et al.* Highly discriminatory capacity of the PowerPlex[®] Y23 System for the study of isolated populations. *Forensic Sci. Int. Genet.* **17**, 104–7 (2015).

17. Sánchez, C. *et al.* Haplotype frequencies of 16 Y-chromosome STR loci in the Barcelona metropolitan area population using Y-Filer™ kit. *Forensic Sci. Int.* **172**, 211–217 (2007).
18. Aboukhalid, R. *et al.* Haplotype frequencies for 17 Y-STR loci (AmpFISTR® Y-filer™) in a Moroccan population sample. *Forensic Sci. Int. Genet.* **4**, e73–e74 (2010).
19. Palet, L. *et al.* Y-STR genetic diversity in Moroccans from the Figuiu oasis. *Forensic Sci. Int. Genet.* **4**, e139–41 (2010).
20. Robino, C. *et al.* Analysis of Y-chromosomal SNP haplogroups and STR haplotypes in an Algerian population sample. *Int. J. Legal Med.* **122**, 251–255 (2008).
21. Frigi, S. *et al.* Data for Y-chromosome haplotypes defined by 17 STRs (AmpFLSTR® Yfiler™) in two Tunisian Berber communities. *Forensic Sci. Int.* **160**, 80–83 (2006).
22. Arredi, B. *et al.* A Predominantly Neolithic Origin for Y-Chromosomal DNA Variation in North Africa. *Am. J. Hum. Genet.* **75**, 338–345 (2004).
23. Triki-Fendri, S. *et al.* Population genetics of 17 Y-STR markers in West Libya (Tripoli region). *Forensic Sci. Int. Genet.* **7**, e59–61 (2013).
24. Ambrosio, B. *et al.* Y-STR genetic diversity in autochthonous Andalusians from Huelva and Granada provinces (Spain). *Forensic Sci. Int. Genet.* **6**, e66–e71 (2012).
25. García, O. *et al.* Data for 27 Y-chromosome STR loci in the Basque Country autochthonous population. *Forensic Sci. Int. Genet.*, <https://doi.org/10.1016/j.fsigen.2015.09.010> (2015).
26. Robino, C. *et al.* Development of an Italian RM Y-STR haplotype database: Results of the 2013 GEFI collaborative exercise. *Forensic Sci. Int. Genet.* **15**, 56–63 (2015).
27. Piglionica, M. *et al.* Population data for 17 Y-chromosome STRs in a sample from Apulia (Southern Italy). *Forensic Sci. Int. Genet.* **7**, e3–4 (2013).
28. Alves, C., Gomes, V., Prata, M. J., Amorim, A. & Gusmão, L. Population data for Y-chromosome haplotypes defined by 17 STRs (AmpFISTR YFiler) in Portugal. *Forensic Sci. Int.* **171**, 250–255 (2007).
29. Carvalho, M. *et al.* Y-chromosome STR haplotypes in two population samples: Azores Islands and Central Portugal. *Forensic Sci. Int.* **134**, 29–35 (2003).
30. Kovatsi, L., Saunier, J. L. & Irwin, J. A. Population genetics of Y-chromosome STRs in a population of Northern Greeks. *Forensic Sci. Int. Genet.* **4**, e21–2 (2009).
31. Katsaloulis, P., Tsekoura, K., Vouropoulou, M. & Miniati, P. Genetic population study of 11 Y chromosome STR loci in Greece. *Forensic Sci. Int. Genet.* **7**, e56–8 (2013).
32. Triki-Fendri, S. *et al.* Paternal lineages in Libya inferred from Y-chromosome haplogroups. *Am. J. Phys. Anthropol.* **157**, 242–51 (2015).
33. Edwards, A., Hammond, H. A., Jin, L., Caskey, C. T. & Chakraborty, R. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* **12**, 241–253 (1992).
34. Restrepo, T. *et al.* Database sample size effect on minimum allele frequency estimation: Database comparison analysis of samples of 4652 and 560 individuals for 22 microsatellites in Colombian population. *Forensic Sci. Int. Genet. Suppl. Ser.* **3**, e13–e14 (2011).
35. Balaesque, P. *et al.* A predominantly neolithic origin for European paternal lineages. *PLoS Biol.* **8** (2010).
36. Alvarez, L. *et al.* Y-chromosome variation in South Iberia: insights into the North African contribution. *Am. J. Hum. Biol.* **21**, 407–9 (2009).
37. Solé-Morata, N. *et al.* Whole Y-chromosome sequences reveal an extremely recent origin of the most common North African paternal lineage E-M183 (M81). *Sci. Rep.* **7**, 1–11 (2017).
38. Kitchen, A., Ehret, C., Assefa, S. & Mulligan, C. J. Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East. *Proc. Biol. Sci.* **276**, 2703–10 (2009).
39. King, T. E., Ballereau, S. J., Schürer, K. E. & Jobling, M. A. Genetic signatures of coancestry within surnames. *Curr. Biol.* **16**, 384–388 (2006).
40. King, T. E. & Jobling, M. A. What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. *Trends Genet.* **25**, 351–360 (2009).
41. Jobling, M. A. & Tyler-Smith, C. Fathers and sons: The Y chromosome and human evolution. *Trends Genet.* **6**, 799–803 (1995).
42. Jobling, M. A. In the name of the father: Surnames and genetics. *Trends Genet.* **17**, 353–357 (2001).
43. Calvo Baeza, J. M. *Apellidos españoles de origen árabe* (Darek-Nyumba, 1990).
44. Willuweit, S. & Roewer, L. Y chromosome haplotype reference database (YHRD): Update. *Forensic Sci. Int. Genet.* **1**, 83–87 (2007).
45. Tereba, A. Tools for Analysis of Population Statistics. *Profiles DNA* (2001).
46. Athey, T. W. Haplogroup Prediction from Y-STR Values Using an Allele-Frequency Approach. *J. Genet. Geneal.* **2**, 34–39 (2006).
47. Athey, T. W. Haplogroup Prediction from Y-STR Values Using a Bayesian-Allele-Frequency Approach. *J. Genet. Geneal.* **1**, 1–7 (2005).

Acknowledgements

The content of this article is part of the Ph.D. thesis of María Saiz which was conducted at the University of Granada under the doctoral programme “Biomedicine”. The authors thank all of the participants who donated buccal swabs and all those who helped in the sample collection—namely, María Luisa Aceituno Villalva, Leticia Olga Rubio Lamia, and Verónica Delgado López. In addition, the authors want to thank Xiomara Gálvez for the technical assistance in the laboratory.

Author Contributions

M.S. originated the study, conducted laboratory analysis, analysed data, drafted figures and draft the manuscript. M.A.-C. collaborate in the analyses of the samples. L.J.M.-G. helped with the statistical analyses. J.C.A. helped draft the manuscript. J.A.L. helped draft and reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-41580-9>.

Competing Interests: The authors declare no competing interests.

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019