Annals of Human Biology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/iahb20

Genetic variation patterns of β-thalassemia in Western Andalusia (Spain) reveal a structure of specific mutations within the Iberian Peninsula

Luis J. Sánchez-Martínez, Candela L. Hernández, Juan N. Rodríguez, Jean M. Dugoujon, Andrea Novelletto, Paloma Ropero, Luisa Pereira & Rosario Calderón

To cite this article: Luis J. Sánchez-Martínez, Candela L. Hernández, Juan N. Rodríguez, Jean M. Dugoujon, Andrea Novelletto, Paloma Ropero, Luisa Pereira & Rosario Calderón (2021) Genetic variation patterns of β-thalassemia in Western Andalusia (Spain) reveal a structure of specific mutations within the Iberian Peninsula, Annals of Human Biology, 48:5, 406-417, DOI: 10.1080/03014460.2021.2000641

To link to this article: https://doi.org/10.1080/03014460.2021.2000641

Published online: 27 Dec 2021.

ANNALS OF HUMAN BIOLOGY
2021, VOL. 48, NO. 5, 406–417
https://doi.org/10.1080/03014460.2021.2000641

RESEARCH PAPER

Genetic variation patterns of β-thalassemia in Western Andalusia (Spain) reveal a structure of specific mutations within the Iberian Peninsula

Luis J. Sánchez-Martínez[a], Candela L. Hernández[a], Juan N. Rodríguez[b], Jean M. Dugoujon[c], Andrea Novelletto[d], Paloma Ropero[e], Luisa Pereira[f] and Rosario Calderón[a]

[a]Departamento de Biodiversidad, Ecología y Evolución, Facultad de Biología, Universidad Complutense, Madrid, Spain; [b]Servicio de Hematología y Hemoterapia, Hospital Juan Ramón Jiménez, Huelva, Spain; [c]CNRS UMR 5288 Laboratoire d'Anthropologie Moléculaire et d'Imagerie de Synthèse (AMIS), Université Paul Sabatier Toulouse III, Toulouse, France; [d]Dipartimento di Biologia, Università Tor Vergata, Rome, Italy; [e]Servicio de Hematología y Hemoterapia, Hospital Clínico San Carlos, Madrid, Spain; [f]i3S, Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal

ABSTRACT

Background: Analyses of the genomic variation in the western Mediterranean population are being used to reveal its evolutionary history and to understand the molecular basis of particular diseases.
Aim: To observe the β-thalassemia mutational spectrum in western Andalusia, Spain, in the context of the Mediterranean. In addition, associations between disease and neutral gene variants within the β-globin gene (HBB) were also evaluated.
Subjects and methods: This study included 63 unrelated individuals diagnosed with β-thalassemia. In addition, 97 unrelated, healthy subjects of the same territory were also analysed as proxies of the normal genetic background. Allele associations and population genetic structure analyses were performed using different methodologies.
Results: Data have revealed a rather restricted spectrum of β-thalassemia mutations in the analysed sample. Although the detected variants fit well with the Mediterranean pattern, certain singularities support a structure of some specific β-thalassemia alleles. The *IVSI-1 (G > A)* shows a strong regionalisation. The spatial correlogram revealed a typically narrow wave structure, presumably linked to genetic isolation and genetic drift.
Conclusions: The long history of endemic malaria in the study territory, the rather high consanguinity rates among its autochthonous population, and other demographic features have been used here to understand the western Andalusian β-thalassemia molecular portrait.

ARTICLE HISTORY
Received 11 February 2021
Revised 15 September 2021
Accepted 13 October 2021

KEYWORDS
Haemoglobin disorders IVSI-1 mutation β-globin gene western Mediterranean malaria

Introduction

Haemoglobin diseases represent some of the most common pathologies of the human population; ∼7% of people are carriers of mutations in certain genes in the α- and β-globin chains that cause abnormal synthesis or structure of the tetramers of normal haemoglobin A (HbA, α_2 β_2) in adults (Weatherall 2010). At present, 733 mutations causing α- or β-thalassemia are registered in the *IthaGenes* (Ithanet) database, and nearly 419 have been identified at the β-globin gene locus (HBB). Most of these mutations are point mutations either within the HBB gene or in the nearest flanking sequence as a result of the substitution of one single nucleotide by another (SNPs, single nucleotide polymorphisms). Only a small fraction of mutations have been identified as deletions or insertions of one or a few nucleotide base pairs in a DNA sequence. These changes in the globin chains cause clinical manifestations of congenital haemolytic anaemias, with a variable degree of severity.

Thalassemias represent a wide and heterogeneous group of recessive diseases that are the result of partial or total deletions of globin chains. Depending on the type of affected chain, we can distinguish, mainly due to their epidemiological importance, α- and β-thalassemias (Cao and Kan 2013). The occurrence of complete deletions in the HBB gene is considered rare or very rare in human populations.

Currently, population-based studies on this varied group of blood disorders (haemoglobinopathies) are not exclusively clinical but were performed to design and apply health policies (Williams and Weatherall 2012; Yasmeen et al. 2016). Additionally, the anthropogenetic interest in α- and β-thalassemias is due to the geographically patterned nature of particular mutations, with some alleles acting as authentic genetic markers of specific regions or in well-defined human populations (Currat et al. 2002; Modell and Darr 2002; Williams and Weatherall 2012; Taher et al. 2018).

Population genetic studies dealing with β-thalassemia have been carried out in many continental areas (Cao and

CONTACT Rosario Calderón rcalfer@bio.ucm.es Departamento de Biodiversidad, Ecología y Evolución, Facultad de Biología, Universidad Complutense, Madrid, Spain

Supplemental data for this article can be accessed here.

Galanello 2010). These surveys have repeatedly revealed that although a high number of β-thalassemia molecular variants have been characterised in many present-day human populations worldwide, a restricted representation of those variants (i.e. 4–10 mutations) reach informative frequencies (>5%) in the affected populations sampled.

The Mediterranean is a special target region, as β-thalassemia disease is prevalent in many of the coastal (littoral) populations. Data on β-thalassemia mutations across the Mediterranean have revealed clearly visible genetic variation patterns, and several fair examples have been found for *CD39 (C > T)* and *IVS I-110 (G > A)* gene variants. The former exhibits the highest frequencies on the western and central sides of the basin (Falchi et al. 2005; Monni et al. 2018); however, the latter is mostly concentrated on its easternmost extreme side (Tadmouri et al. 1998; Murad et al. 2018). Signatures of β-thalassemia blood disease in the Mediterranean seem to be linked to ancient times. Viganó et al. (2017) documented the earliest case of the *CD39* mutation detected to date by analysing aDNA samples from skeletal remains buried in Punic and Roman necropolises (~2,000 years BP) on Sardinia Island.

Within the Iberian Peninsula, molecular studies on β-thalassemia are still very few in Spain, and they are limited to particular geographical areas. The first population genetic data were published from the 1980s onward (Amselem et al. 1988). Villegas et al. (2001) analysed mutated samples from different Spanish regions and observed that the five most frequent β-thalassemia mutations in decreasing order were *CD39, IVS I-1, CD 8/9, IVS I-6,* and *IVS I-110*. Other studies performed in southwestern Spain (Ropero et al. 1994; Villegas et al. 1996; Ribeiro et al. 1997) interestingly revealed that *IVS I-1* was the most frequent mutation registered among autochthonous affected individuals.

The territory of Huelva, in southwestern Spain, represents a special "*hot spot*" within the Iberian Peninsula due to its confirmed long history of malaria among its native population (Madoz 1847; Pellicer 1969; Sousa et al. 2014). There are very old references about malaria in Spain, and mentions of this infectious disease are known through the Romanisation of Iberia (between the second century BC and the very beginning of the fifth century AD). The first field studies in Spain on malaria started in 1899 and were performed by the Scottish medical doctor Ian MacDonald, in the "Minas del Rio Tinto," the important mining area of southwestern Andalusia (Huelva) (Macdonald 1904) where extensive wetlands have prevailed until recently. The survey performed by doctors Huertas and Mendoza (1903) on the epidemiology of malaria and its prophylaxis in the Extremadura region (Caceres province) also deserves attention.

Tartessians already exploited mineral deposits of the Rio Tinto. When the Tartessian kingdom collapsed around the second half of the sixth century BC, a period of the Carthaginian rule followed until Roman dominion began by the second century BC. Roman colonisation was intense in the Guadalquivir Valley (*Baetica*), where many cities were founded, and Rio Tinto mineral deposits were deeply exploited by Romans using a new mining technique (further details are provided in Calderón et al. 2006).

Both the prehistory and history of Huelva have been largely documented through numerous sites. Important archaeological sites of human settlements (e.g. Phoenicians and Tartessians, eighth century BC) were established there (Fernández-Jurado et al. 1997) with fair evidence of high metal trade activity (e.g. silver) with Greeks settled in other neighbouring Mediterranean areas.

Over the last decade, the extent of human genetic diversity in Andalusia, the large southernmost region of Spain, has been thoroughly investigated. Andalusians originating on the western and eastern sides of the region from Huelva and Granada Provinces, respectively, have been studied for different genetic polymorphisms and genomic regions (Ambrosio et al. 2012; Fortes-Lima et al. 2014; Hernández et al. 2015, 2017, 2019, 2020, among others). Primary conclusions that have been drawn from these studies include the following: (i) Western Andalusians from Huelva have a high incidence and diversification of African maternal lineages when compared to their eastward relatives in Granada and other Iberian populations; (ii) The diversity of maternal and paternal lineages in the region indicates that these populations have integrated multiple migrations from the Mediterranean space; and (iii) Transcontinental ancient contacts between North African and Iberian populations surrounding the maritime region of the Gibraltar Strait seem to date from the early Holocene onwards. Therefore, data would support the existence of an ancient, frequently denied bridge connecting the Maghreb and Andalusia with conspicuous signals of the African genome across the Atlantic façade of the Iberian Peninsula.

This work aims to provide molecular results of a study of β-thalassemia in a well-defined sample from southwestern Spain (Huelva Province). The obtained results will reflect the inherent characteristics of the studied territory and population. Moreover, particular attention is paid to the diversity of the HBB gene in the context of the presence and frequency of neutral mutations shared in both mutated and control samples with the same population origins, searching for allele associations. The spectrum of detected β-thalassemia mutations has been combined with those found in other Iberian, Mediterranean non-Iberian, and southwestern Asian populations for analysis of genetic affinities and structure. In this respect, special attention is paid to the northwestern African populations, the nearest neighbour of the coastal lands of the southern Iberian Peninsula.

Subjects and methods

The studied area and sampled population

The wide region of Andalusia (87,268 km^2) is administratively divided into seven provinces, among which is that of Huelva, the autochthonous population that is the focus of our study. Huelva Province shapes the westernmost end of the southern coast of Spain and borders southern Portugal (the Algarve Portuguese region). On its north side, Huelva is geographically adjacent to the Extremadura region, with Seville

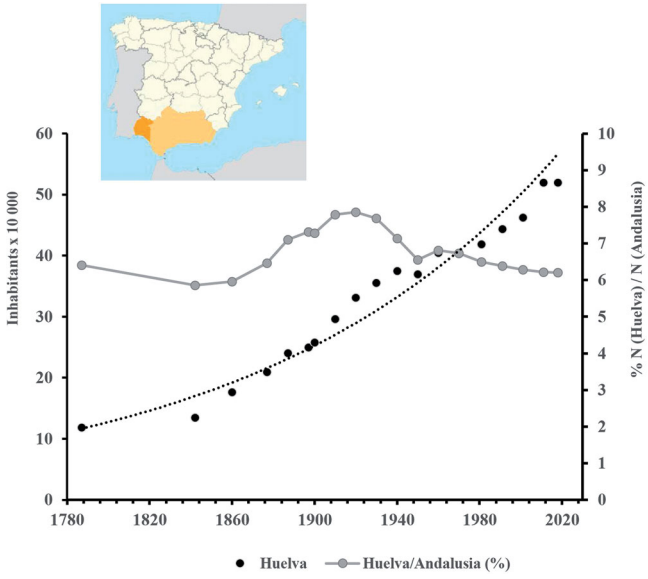

Figure 1. Secular trend of demographic size (N) registered in the province of Huelva. The proportion of the Huelva demographic size regarding the Andalusia region as a whole is plotted in grey colour. On the top, a map of the Iberian Peninsula showing the location of Huelva Province within Andalusia, Spain.

and Cadiz Provinces to the east and the Atlantic Ocean to the south (see Figure 1). The main rivers of Huelva Province are the Odiel and the Tinto, which join to form "*the ria*" in Huelva city, the provincial capital.

Since the eighteen century until the present, the population growth rate of the province of Huelva has followed the same pattern as the whole Andalusia region, that is, of increasing over time [Instituto Nacional de Estadística, INE; http://www.ine.es]. Nevertheless, it has repeatedly registered the smallest demographic size (34.40 inhabitants/km^2 as average) of all the Andalusian Provinces [http://www.ine.es], a singular condition that has remained largely stable over the years (Figure 1).

This study included 63 unrelated subjects of Huelva origin who were diagnosed with β-thalassemia at the Service of Haematology and Hemotherapy of Juan Ramón Jiménez Hospital located in Huelva city. One member of this group (JNR) took responsibility for the blood sample collection. The β-thalassemia sample set was composed of 41 females and 22 males (mean age: 56 ± 19 years; range: 10–91 years). Haematological and serological parameters of each individual were recorded, including mean levels of red blood cell count (RBC), haemoglobin concentration, packed cell volume (PCV), mean corpuscular volume (MCV), mean corpuscular haemoglobin (MCH), serum ferritin, HbA2 and foetal haemoglobin (HbF) (see for details Cao and Kan 2013). The obtained results were included in files designed for each donor. Clinical diagnosis relied on microcytic anaemia combined with normal ferritin serum values and increased levels of haemoglobin A2 (HbA2).

In addition to the recruited β-thalassemia samples, β-globin gene variation was also studied in 97 unrelated, healthy subjects of Huelva origin (39% females, 61% males; mean age: 56 ± 12 years, range: 34–84) as proxies of the normal genetic background in the geographic area. These control individuals were randomly drawn from the global, general stock of samples collected by us in several fieldwork campaigns in Huelva province. This DNA material, which was also obtained from total blood, has been used in our published studies on human genomic diversity in western Andalusia.

Informed consent was obtained from all participants, as well as details on the geographic origin of their families and personal medical history. The Bioethics Committees at the Complutense University of Madrid and the Juan Ramon Jimenez Hospital of Huelva approved the research protocols used for this study.

SNP genotyping within the β-globin gene (HBB)

DNA samples from individuals clinically diagnosed with β-thalassemia were genotyped at the Molecular Anthropology Laboratory (Complutense University of Madrid) and the Service of Haematology and Hemotherapy of San Carlos Clinic Hospital (Madrid). In both thalassemia and normal subjects, DNA was isolated from blood samples by standard isopropanol precipitation.

The β-globin gene was amplified using the following two pairs of primers: β1D: 5′-CCT AAG CCA GTG CCA GAA G-3′ (from nt −160 to + −142) and CD2: 5′-GAC CTC CCA CAT TCC CTT TT-3′ (from nt +1659 to + 1643) (all nt positions are provided relative to the CAP site = nt1 from NCBI GenBank®). For both reactions, we used 35 cycles and a melting temperature of 55°. PCR products were purified by using SPEEDTOOLS PCR Clean-Up kit (Biotools) and then treated with the ABI PRISM^{TM} BigDye® Terminator V1.1 Cycle Sequencing Ready Reaction Kit (PE Applied BioSystems, Foster City, CA, USA) with the β1D and CD2 primers for sequencing following the manufacturer's instructions. The sequences were analysed on the ABI PRISM^{TM} 310 Genetic Analyser (PE Applied Biosystems).

Each mutation was confirmed manually, and they were scored relative to the β-globin reference sequence (NG_000007.3) by using BioEdit (Hall 1999). Nomenclature for polymorphisms was set according to the online HbVar database (http://globin.bx.psu.edu/hbvar/menu.html) and ITHANET (http://www.ithanet.eu/). Due to technical errors during the genotyping process, two samples were discarded from the total sample set ($n = 63$).

Statistical and genetic data analysis

Allele frequencies were estimated by direct gene counting (Hartl and Clark 2006). The Hardy Weinberg equilibrium (HWE) has been calculated using the HWChisq function from the R Hardy-Weinberg library. None of the loci showed significant deviations from HWE in either the thalassemia or control samples.

An updated database of β-thalassemia studies in contemporaneous Mediterranean populations was constructed (see Table S1). Country codes used are those officially published in www.nationsonline.org/oneworld/country_code_list.htm. Populations were coherently selected following two criteria: geography and the minimum number of chromosomes analysed (≥35). Five different population groups were

established: Iberian Peninsula ($n = 9$), Central and Eastern Europe (European Mediterranean) ($n = 11$), Near/Middle East ($n = 5$), Northern Africa ($n = 4$), and Southwestern Asia ($n = 8$). Only those most frequent β-thalassemia gene variants among Mediterranean populations were compiled. Therefore, the database entailed mutation data from 68,180 β-thalassemia independent chromosomes from 37 populations.

Variation patterns of mutation frequencies were graphically observed by means of stacked histograms at different geographical scales. Likewise, surface interpolation maps of the frequencies of β-thalassemia mutations *CD39, IVS I-110,* and *IVS I-1* across the Mediterranean were obtained using *ArcGIS v.10.3* software using Ward's linkage algorithm. The existence or absence of a particular spatial genetic structure associated with these selected β-thalassemia mutations was further proven through a classical spatial autocorrelation analysis, which executes one allele at a time. Moran's *I* autocorrelation coefficient, which is an average correlation among populations as a function of distance *k* (Epperson 2003), was estimated using PASSaGE v0.2 software (Rosenberg and Anderson 2011).

ARLEQUIN v 3.5.1.2 (Excoffier and Lischer 2010) was used for Analysis of Molecular Variance (AMOVA), which aims to hierarchically explore population genetic differentiation by means of the *F* parameter and its fixation indices: F_{ST} (genetic differences within populations), F_{CT} (between population groups) and F_{SC} (among populations within groups) (Holsinger and Weir 2009).

Lineal discriminant analysis (LDA) and multidimensional scaling (MDS), the latter of which is based on F_{ST} genetic distances, were mainly used to determine the genetic clustering of the population panels used here as references for comparisons. To perform the mentioned statistical multivariate procedures, free open access R software was used (R Core Team 2020).

Table 1. Types and frequencies of β-thalassemia mutations in western Andalusia (Huelva province).

Mutations	Number Chromosomes	Homozygous/ Heterozygous	Frequency (%)
IVS I-1 (G > A)	30	1/28	52.63
CD39 (CAG > GAG)	17	0/17	29.82
IVS I-110 (G > A)	5	0/5	8.77
IVS II-745 (C > G)	4	0/4	7.02
IVS I-6 (T > C)	1	0/1	1.75

The number of chromosomes and genotypes identified in the analysed affected sample are also shown.

Results

We did not find any deleterious variants in five of the samples from patients previously diagnosed with β- thalassemia. In this set, we performed an analysis of α-thalassemia conditions by using a reverse hybridisation assay for the rapid and simultaneous detection of the 21 most frequent α-globin mutations with the commercial Alpha-Globin StripAssay Kit (ViennaLab Diagnostic GmbH, Vienna, Austria) (Puehringer et al. 2007). Only one of these individuals was finally classified as thalassemia, harbouring a deletion of 3.7 kb in α-thalassemia genes ($--\alpha^{3.7}/\alpha\alpha$).

The sequencing of the gene revealed a total of five pathogenic mutations, listed in sequence order: *IVS I-1 (G > A), IVS I-6 (T > C), IVS I-110 (G > A), CD39 (CAG > GAG),* and *IVS II-745 (C > G).* In addition to these deleterious variations, neutral polymorphisms (*CAP +20 C > T, CD 2 CAC > CAT [His > His], IVS II-16, IVS II-74 T > G, IVS II-81 C > T, IVS II-666 C > T*) were also detected in both patients and control groups. Last, one of the patients harboured a neutral mutation on one chromosome at position 71,839 (A > C) that, to our knowledge, has not been previously described elsewhere.

Figure 2 shows α-globin and HBB gene cluster arrangements on chromosome 16 and chromosome 11, respectively. The structure of the HBB gene as well as the position of the 11 most common β-thalassemia mutations (in black bold) identified in Mediterranean populations are also shown. The position of the seven neutral mutations found in both

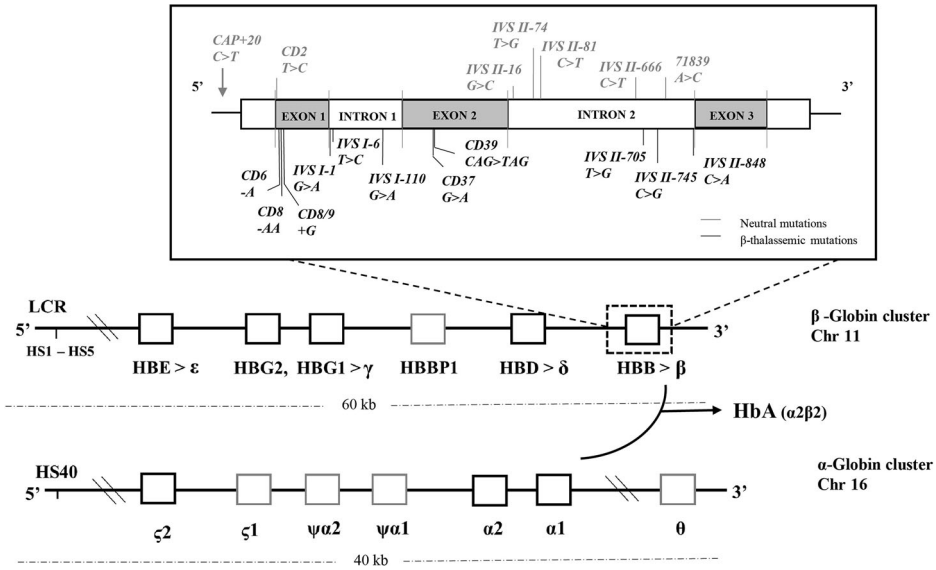

Figure 2. Schematic structure of α-globin and β-globin gene families. Boxes outlined in black represent functional genes (expressed genes) in each cluster. The β-globin gene (HBB) is zoomed showing its exon-intron structure as well as the genomic locations of the most common β-thalassemia mutations (in black bold) found in the Mediterranean. The set of neutral mutations identified in the present study are highlighted in grey bold.

Table 2. Frequencies of neutral mutations detected along the β-globin gene for the groups of patients and controls with family origins in western Andalusia (Huelva).

	Patients ($n = 48$)			Controls ($n = 97$)		
Mutations	Homozygous/ Heterozygous	Number alleles	Frequency (%)	Homozygous/ Heterozygous	Number alleles	Frequency (%)
CAP +20 C > T	0/4	4	4.17	0/0	0	0.00
CD 2 CAC > CAT [His > His]	39/8	86	89.58	62/32	156	80.41
IVS II-16	39/8	86	89.58	63/30	156	80.41
IVS II-74 T > G	9/23	41	42.17	17/48	82	42.27
IVS II-81 C > T	1/8	10	10.42	1/29	31	15.98
IVS II-666 C > T	38/9	85	88.54	61/32	154	79.38
71839	0/1	1	1.04	0/0	0	0.00

affected and control samples is displayed (in grey bold) on the upper side of the HBB gene.

The molecular analysis of the 56 affected individuals revealed five mutations, and their estimated frequencies in decreasing order are shown in Table 1. *IVSI-1 (G > A)* was the most common allele (52.6%), followed by *CD39 (CAG > TAG)* (29.8%). Both gene variants account for over 80% of the total affected alleles. Other β-thalassemia mutations identified but at frequencies <10% were *IVS I-110 (G > A)* (8.8%), *IVS II-745 (C > G)* (7%), and *IVS I-6 (T > C)* (1.8%). Among the total number of patients examined, only one individual was identified that carried *IVSI-1* on both chromosomes, whereas the rest of the patients were heterozygous (thalassemia trait phenotypes), which is normally asymptomatic. The unusually high representation of the β-thalassemia *IVS I-1 (G > A)* allele in southwestern Spain is consistent with other earlier fieldworks from in the region (52.23%: Ropero et al. 1994; Benito et al. 1996). In Portugal, *IVS I-1* exhibits a positive cline in the north-south direction: 17% (N) vs. 34% (S) (Cabeda et al. 1999; Faustino et al. 1999).

Table 2 presents the number and types of non-disease-causing mutations (neutral mutations) that have been identified within the β-globin gene locus in both affected and unaffected (healthy) Andalusian study samples. The data show similar patterns (shared genetic background) in both population groups, which is fair evidence of the good strategy adopted in the sampling process.

A further prospective analysis aimed to detect possible associations between disease and neutral mutations allowed us to determine that four individuals (8.33% of the total sample) having the *CAP + 20 C > T* neutral mutation also carried the *IVS II-745 C > G* disease mutation (7.02% of the total sample). However, *CAP + 20 (C > T)* was not identified in the control sample, which comprised 97 individuals ($n = 194$ chromosomes). The application of the 2-sample test for equality of proportions for the neutral mutations of Table 2 concludes that there are significant differences (p values< 0.05) in the frequency of the *CAP + 20* mutation between both groups. However, similar comparisons for the rest of the mutations were not significant. Given the low frequencies of *CAP + 20 C > T* and *IVS II-745 C > G* in human populations, it would be highly unlikely that they would be found by random chance in the genome of any subject and even far more unlikely to be associated. By using association rules analysis together with the *Apriori* algorithm (Aggarwal 2015) our data have revealed that when an individual harbours the β-thalassemia *IVS II-745 (C > G)* gene variant, the probability that a *CAP + 20 (C > T)* neutral mutation is present in his (her) genome is multiplied by 12 times. This outcome is in accordance with that observed in other similar β-thalassemia population studies, mainly in Mediterranean populations (Orkin et al. 1982; Ropero et al. 2013; Cherry et al. 2016; among others).

Furthermore, the same four subjects presented similar genotypes (heterozygous) for the mutation arrays: *CAP +20:* ***C > T****; CD2:* ***T > C****; IVS I-1:* ***G****; IVS I-6:* ***T****; IVS I-110:* ***G****; CD39:* ***C****; IVS II-16:* ***G > C****; IVS II-81:* ***C > T****; IVS II-666:* ***C > T****; IVS II-745:* ***C > G****; 71839:* ***A.*** This finding suggests a shared origin, and when we referred to the family background of these people, they were all born in a small area (1270 km^2) within Huelva Province known as "la Comarca del Condado." Likewise, the analysis of the control population sample allowed us to observe that 61 individuals out of 97 analysed (62.89% of the total sample) shared the haplotype *CAP +20:* ***C****; CD2:* ***C****; IVS I-1:* ***G****; IVS I-6:* ***T****; IVS I-110:* ***G****; CD39:* ***C****; IVS II-16:* ***C****; IVS II-81:* ***C****; IVS II-666:* ***T****; IVS II-745:* ***C****; 71839:* ***A***. The high population frequency of this haplotype is a particularity and is likely associated with the study population and zone. The allelic associations and statistical values in both patients and control samples are provided in Table S2.

Haematological and serological parameters, as clinical indicators of β-thalassemia diagnosis, were obtained and results are shown in Table S3. *CD39,* one of the most severe mutations in terms of biomedical impact significantly influenced some values when compared to clinically derived effects from the rest of the identified β-thalassemia mutations in the present study. This outcome seems to give support to the fact that although the type of mutation is what marks the underlying degree of severity (Thein 2013), other secondary factors could influence the phenotype, determining the effects of varying severity associated with the disease (Joly et al. 2014). In addition, special interest arises from the only identified *IVS I-1* homozygous individual, which showed significantly low levels of RBC and HbA_2 against high rates of MCV, MCH, and ferritin serum.

The extent of the relationship among haematological and serological parameters associated with β-thalassemia patients was evaluated by means of a heatmap, and the constructed correlation matrix is shown in Figure S1. The data suggest that haematocrit values are positively and significantly related (Spearman's correlation, $r > 0.70$) to haemoglobin concentration and RBC. The same is true between MCV and MCH ($r = 0.90$). Nevertheless, most of the other pairwise correlation assessments tended to be less intense, both in a

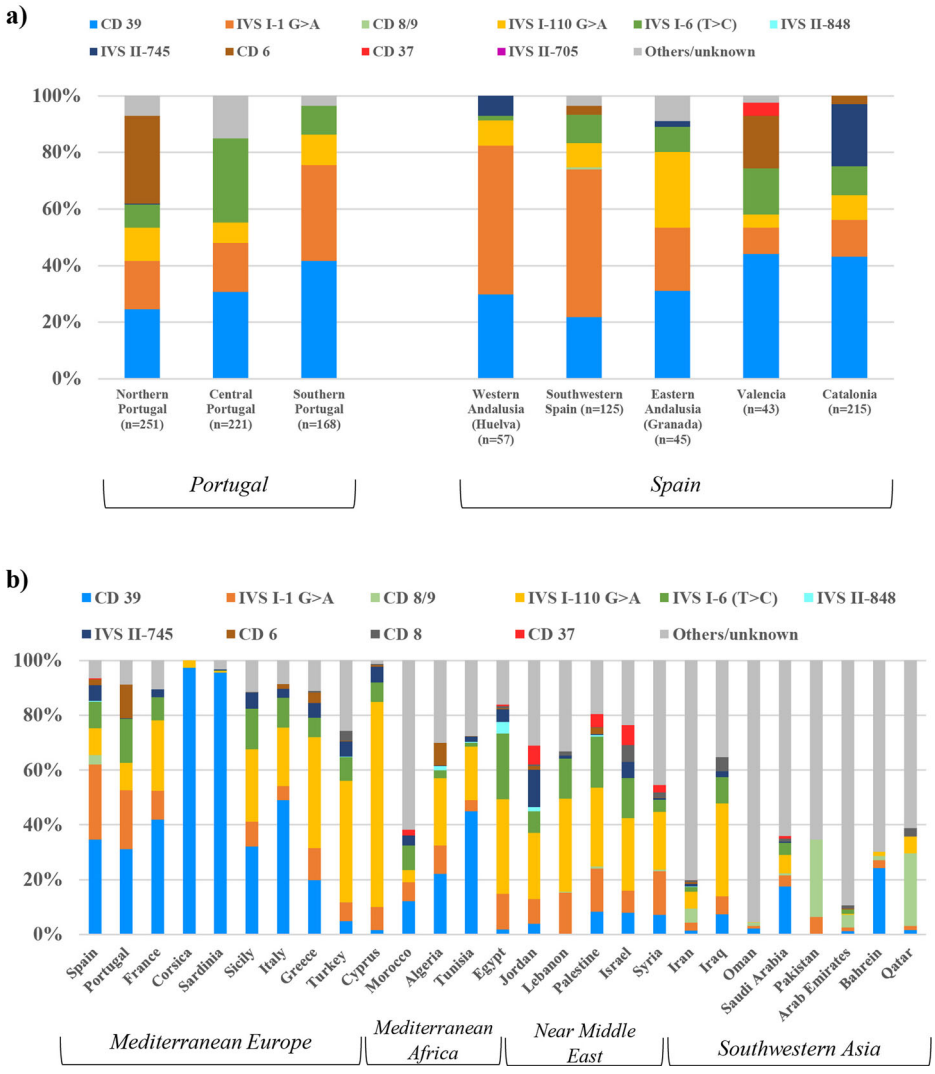

Figure 3. β-thalassemia mutation frequencies (in percentages) and their distributions across the Iberian Peninsula (a), the Mediterranean and the southwest of Asia (b). Information about the populations used in Table S1.

positive and negative sense. When applying a discriminant analysis (DA) to the five characterised β-thalassemia mutations together with their associated haematimetric values, the results strongly emphasise that no clear differences appear among mutations. The DA (plot not shown) displays a clear overlap among affected individuals carrying specific variants. The success rate of the constructed discriminant equations, by applying a cross-validation method, was 37.5%.

Figure 3(a) depicts variation patterns in β-thalassemia alleles across the Iberian Peninsula, differentiating Spain and Portugal. As expected, *CD39* was the most frequent gene variant in most of the sampled Iberian populations (∼45% as average). Nevertheless, in southwestern Spain, the predominance of *IVSI-1* (50–53%) among its autochthonous population is fairly demonstrated. In the adjacent Portuguese Algarve, the frequency of the *IVS I-1* allele is somewhat lower (34%).

Portugal shows interesting differences in β-thalassemia allele profiles among its three major geographical regions in accordance with its history. Northern Portugal is the most distinctive, with high frequencies of *CD6* mutations (31%) and rather modest values of *CD39* (24.6%) (Ribeiro et al. 1997; Cabeda et al. 1999). To our knowledge, the β-thalassemia *CD6* gene variant has not been detected in central and southern Portugal, although its presence has been found in polymorphic frequencies in southwestern Spain (3.05%) (Ropero et al. 1994). This finding should be interpreted in terms of genetic signals of population connectivity from ancient times being long maintained through shared borders from medieval times. Curiously, in eastern Spanish Mediterranean areas (e.g. Valencia, Balearic islands), the *CD6* mutation is well represented (12.5–18.6%) (Pérez et al. 1998; López-Escribano et al. 2013).

Figure 3(b) shows geographical patterns of spatial distributions of β-thalassemia mutations across the Mediterranean, where a large and rich amount of genetic data are available today. The stacked histogram shows a general view of how strong the genetic contrast is among the geographic groupings established here. Comparatively, β-thalassemia allele profiles in southwestern Asian populations are very distinct from those exhibited by Mediterranean populations, even when referencing the most Eastern Mediterranean populations.

CD39 and *IVSI-110* mutations are paradigmatic because of their peculiar genetic geography. Both the west and the

central Mediterranean are distinguished by the high and widespread prevalence of *CD39*, with the exception of cases from Sardinia and Corsica Islands, where the mentioned disease allele reaches modal frequencies (>90%: Rosatelli et al. 1992; Falchi et al. 2005). In contrast, the *IVSI-110* registers its maximum representation in the eastern Mediterranean (e.g. Turkey, Greece, Cyprus, Egypt, and Near/Middle East) with frequencies ranging between 34 and 45% (Tadmouri et al. 2001; Kalleas et al. 2012; Aydinok et al. 2018). The *IVS I-110* mutation concentrates the highest values in Cyprus (~75%) (Colah et al. 2010).

IVSI-110 seems to be evolutionarily older than the *CD39* mutation. The study of Chen et al. (1990) showed the nearly exclusive association of *IVSI-110* with ancestral haplotype I (see for further details Orkin et al. 1982). Nevertheless, Tadmouri et al. (2001) who analysed β-thalassemia samples from the Anatolian Peninsula, observed that the *IVSI-110* allele was associated with six different haplotypes, among which haplotype I was predominant. The haplotype diversity, greater than any other disease mutation in the β-globin gene, indicates the existence of recombination events among haplotypes, a process that requires an extended number of generations. Furthermore, the authors suggested that the geographical origin of the *IVSI-110* allele could be in Turkey, coinciding with the beginning and spread of agriculture during the Neolithic, which was introduced in southeastern Europe *via* Anatolia by the middle of 7000 BC (Hofmanová et al. 2016; Mathieson et al. 2018). The subsequent spread of the *IVSI-110* allele throughout the rest of the Mediterranean could be strongly influenced by the numerous settlements of Greek and Phoenician colonies scattered along its coasts (Cao et al. 1989; Perrin et al. 1998), including those of Iberia.

Figures 4(a–c) depict contour maps for the three selected β-thalassemia mutations: *CD39, IVSI-110,* and *IVSI-1*. As expected, the *CD39* surface map (Figure 4(a)) exhibits a decreasing frequency in a west-east direction across the Mediterranean and shows a consistent pattern of regional genetic structure. The spatial autocorrelation analysis fairly confirms this, with the largest positive Moran's *I* values ($I = 0.83 \pm 0.12$; *p* values = 0.00) observed between pairs of populations at the shortest distances and subsequent positive correlations declining very sharply until reaching the *x*-intercept. At distances ~2000 km, all correlations become negative and generally significant with no tendency for *I* values to persist, decreasing further as distance classes increase. Such profiles are generally interpreted as signals of subtle clines in allele frequencies, which would strongly suggest a typical model of isolation by distance (Legendre and Fortin 1989; Manel et al. 2003).

Comparatively, the *IVSI-110* surface map (Figure 4(b)) yielded a longitudinally opposed pattern compared with that shown by the *CD39* mutation. Modal frequencies are strongly concentrated in the highly diversified eastern Mediterranean, with a descending trend towards the western side of the basin. Therefore, a repeating correlogram shape as above is mapped but with some punctual differences. For the three first distance classes, Moran's *I* correlations had smaller but significant positive values ($I = 0.51$–0.27; *p* values $\leq$ 0.01), with the tendency to be moderately sharp when

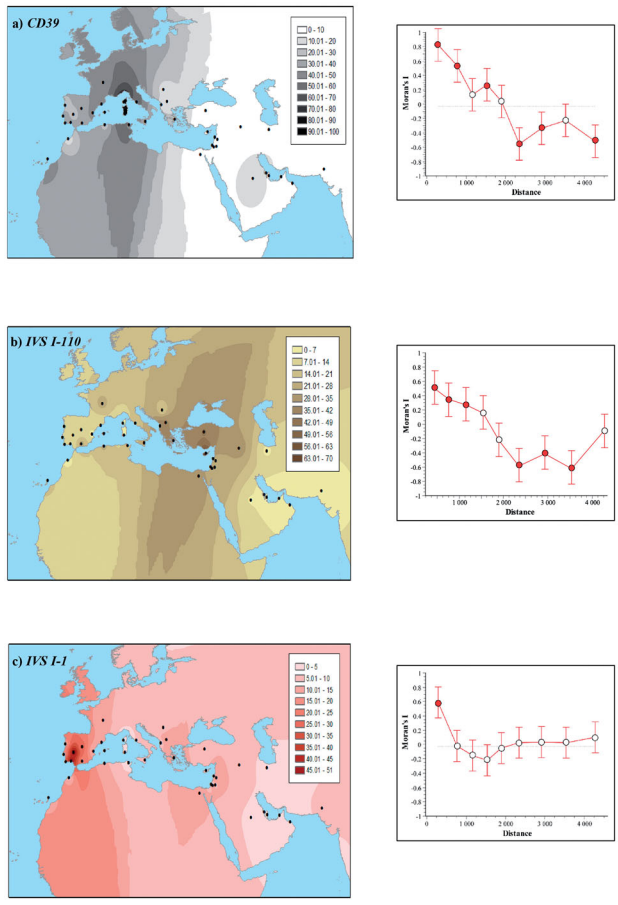

Figure 4. Surface maps of β-thalassemia mutations *CD39* (a), *IVSI-110* (b), and *IVSI-1* (c). Spatial autocorrelation analysis correlograms with significant values of Moran's *I* after Bonferroni correction are shown. Significant points are indicated as filled circles (*p* values = 0.05).

approaching zero. At distances larger than 2000 km, the correlogram profile provides support to that direction in the spatial differentiation of *IVS I-110* frequencies across the Mediterranean.

Figure 4(c) reveals the strong regionalisation of the *IVS I-1* mutation in the southwest of the Iberian Peninsula irrespective of its wide presence throughout Europe and northern Africa. The correlogram produced from the corresponding surface map turned out to be very different from the other two exemplified cases. The existence of one single point, which is identified by the largest positive Moran's *I* correlation value ($I = 0.59 \pm 0.11$; *p* values= 0.00) at the shortest distance class (0–500 km), just where the mutation frequency increases considerably, represents a distinctive feature. The rest of the relationships among populations were slightly negatively correlated and encompassed steady oscillations around zero as geographic distance classes were greater. These types of correlograms, typically known as a "*narrow wave*" spatial structure (Legendre and Fortin 1989), are presumably linked to evolutionary processes related to genetic isolation, inbreeding, and genetic drift.

The spatial population genetic structure of β-thalassemia alleles was further examined by means of hierarchical AMOVA (see Table 3). When the whole Mediterranean population dataset ($n = 29$) was considered as one group, a highly

Table 3. Genetic structure of Mediterranean populations based on mutational frequencies of β-thalassemia.

	Fixation index			
Population groups	F_{ST}	*P-value*	F_{CT}	*P-value*
Mediterranean Basin (19)[a]	0.2797	0.0000***		
Western Mediterranean (10) vs. Eastern Mediterranean (9)	0.3306	0.0000***	0.1541	0.0097**
Iberian Peninsula (9)	0.0842	0.0000***		
Iberian Peninsula (9) vs. Maghreb (3)	0.1332	0.0000***	0.0505	0.0322*
Iberian Peninsula (9) vs. Northwestern Africa[b] (2)	0.1546	0.0000***	0.0716	0.0752 (ns)
Southern Iberia[c] (4) vs. Maghreb (3)	0.1770	0.0000***	0.1050	0.0459*
Southern Iberia (4) vs. Northwestern Africa (2)	0.1918	0.0000***	0.1222	0.0635 (ns)
Southern Spain[d] (3) vs. Maghreb (3)	0.1952	0.0000***	0.1206	0.0889 (ns)
Southern Spain (3) vs. Northwestern Africa (2)	0.2014	0.0000***	0.1235	0.2101 (ns)

[a]Number of population per region
[b]Morocco + Algeria
[c]Southern Portugal + Western Andalusia (present study) + Eastern Andalusia (Granada) + Extremadura region
[d]Western Andalusia (present study)+Eastern Andalusia (Granada) +Extremadure region
ns: not significant; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$

significant genetic structure was observed ($F_{ST} = 0.2797$, p values < 0.001). This finding is in accordance with the proven genetic differentiation between western/central vs. eastern Mediterranean populations ($F_{CT} = 0.1541$, p values< 0.01). In contrast, when genetic analyses were performed following a latitudinal orientation north/south (e.g. southern Iberia/northwestern Africa), F_{CT} values were, in general, nonsignificant or the level of significance was low. These results again suggest the genetic affinity among human populations settled on the opposite coasts of the western extreme of the Mediterranean.

Likewise, genetic clustering among Mediterranean populations and their neighbouring southeastern Asian populations has been explored by means of lineal discriminant analysis (LDA) and multidimensional scaling (MDS). The LDA in Figure 5(a) shows three well-structured population clusters, which fairly show the differential weight of particular β-thalassemia gene variants among Mediterranean regions and their neighbours. Linear discriminant 1 (LD1) separates all Mediterranean populations from those of southwestern Asia, with the *CD8/9* mutation and "other mutations" showing the highest correlation values, −0.570 and −0.926, respectively. The latter concerns mutations locally regionalised in southwestern Asia. LD2, however, seems to strongly distinguish western/central Mediterranean populations from their relatives on the eastern side. The group of populations, which are located on the negative direction of axis 1, are dominantly conditioned by the *CD39* mutation ($r = 0.533$), while the eastern Mediterranean population, on the positive side of LD2, reveals high correlation values associated with the *IVS I-110* ($r = 0.707$) and *CD37* ($r = 0.607$) mutations.

The MDS genetic map, which was constructed using F_{ST} genetic distances (Figure 5(b)), illustrates topologies similar to those above for the major Mediterranean population groups, with the *CD39* and *IVSI-110* mutations being responsible for their relevant genetic structure. The privative presence of punctual β-thalassemia alleles in southwestern Asian and Near/Middle Eastern countries makes these populations genetically different from their neighbourhood (Hassan et al. 2010; Moatter et al. 2012).

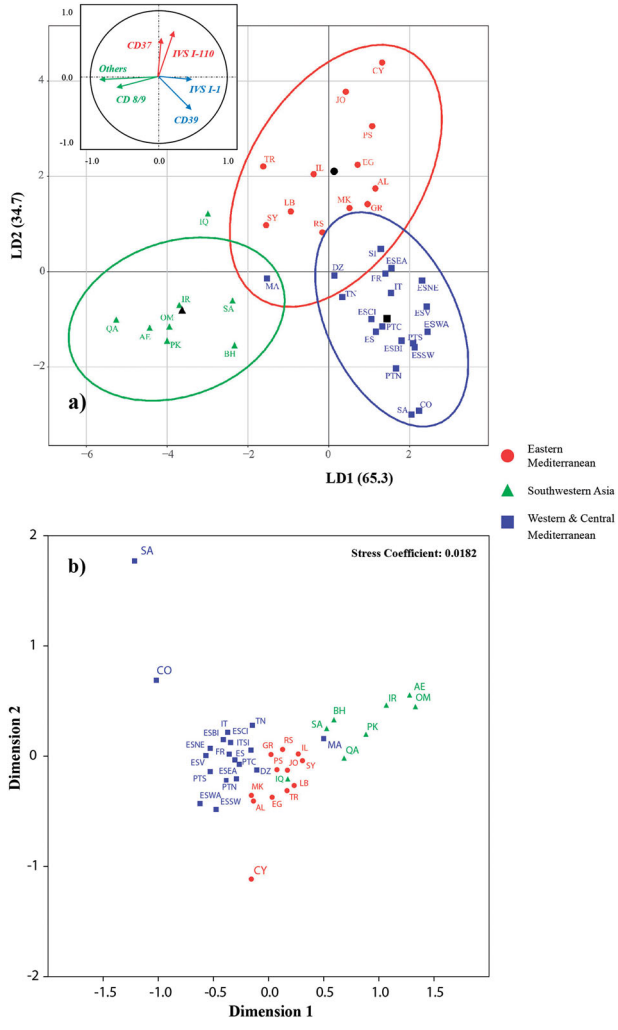

Figure 5. Lineal Discriminant (LD) (a) and Multidimensional Scaling (MDS) (b) analyses show genetic clustering patterns of Mediterranean and other neighbouring populations based on types and frequencies of β-thalassemia mutations. See information about the populations used in Table S1.

Discussion

The considerable variation at the β-globin gene locus in the Mediterranean has shown conspicuous spatial patterns for some particular mutations across the basin. This scenario is also confirmed on more restrictive geographical scales, such as the Iberian Peninsula, where an important variation in frequencies and types of β-thalassemia variants has been observed among territories and populations.

The present work has revealed a rather restricted spectrum of β-thalassemia mutations in the analysed sample

from western Andalusia (Huelva Province, Spain). Nevertheless, although the detected gene variants fit well with the Mediterranean pattern, certain singularities were perceived in the identified allele profile that supports a structure of some specific β-thalassemia alleles. This portrait is convergent when exploring human genomic diversity patterns in well-defined, non-mutated samples from the same population, mainly based on haploid polymorphisms (e.g. mtDNA).

β-thalassemia is the most frequent haemoglobinopathy in Spain, as it is still found in most Mediterranean coastal populations. Southwestern Spain is considered a key epidemiological area due to the relatively high prevalence of this genetic recessive disorder, affecting approximately 1% of births. The shown scenario parallels consistently with the long and shared secular history of endemic malaria in the zone. Huelva Province is one of the last Spanish cores where cases of severe malaria appeared until the Sixties of the past century (Sousa et al. 2014). After the Spanish Civil War, malaria showed an important rebound effect (Avellò 1950).

In this context, there is strong empirical evidence that malaria can explain the high frequencies of β-thalassemia heterozygosity as it seems to be selectively protective (positive selection) against malaria infection (Kariuki and Williams 2020). Nonetheless, in addition to malaria, other relevant factors related to specific territorial population dynamics themselves [e.g. demography, prolonged customary consanguineous marriages and their biological consequences (inbreeding levels) among others] can also influence the frequency and maintenance of recessive allele diseases in particular geographical areas (Modell et al. 2007).

Over the last three centuries, the population size of Huelva Province has only represented 6–7% of the total Andalusia population. This demography, together with the deeply rooted and sustained ruralism of its autochthonous population, has been able to influence mating system patterns, especially on related marriages, generating local genetic structure.

Consanguinity population studies in Andalusia are very scarce (Calderón et al. 2018; Gamella & Núñez-Negrillo 2019). Pinto-Cisternas et al. (1979) observed that rates of first-cousin marriages (M22) in Huelva Province in 1911–1943 were very close to those of second cousin marriages (M33), respectively: 23.08 vs. 23.58 by 1000 marriages. Consequently, the M22/M33 ratio (preferentiality index) turned out to be highly deviated (0.979) from the theoretical proportions (0.25) under Hardy-Weinberg equilibrium (random mating). The mean inbreeding coefficient,$\bar{F}F$, of the Huelva general population, was 0.00209, a figure relatively high assuming that after the Spanish Civil War (1936–1939) consanguinity started declining significantly all over Spain. Hence, consanguineous marriages in the study population would have historically signified an important component of its marital structure. This marital behaviour has been further supported through genome-wide analyses (GWAs) using a panel of 2.5 million SNPs. A longer genome with ROH (runs of homozygosity) fragments; a higher number of long ROHs (>5 Mb) as signals of recent inbreeding, and a higher genomic inbreeding coefficient, F_{ROH}, seem to distinguish Huelva genomes from other southern Iberian populations, including southern Portuguese (Calderón et al. paper in progress).

Given the cascade of signals about the strong relationships among western Mediterranean populations, based on both classical and DNA polymorphisms, it makes sense in the context of the present survey to know the extent of β-thalassemia gene variants shared by this human metapopulation. All the β-thalassemia mutations detected in the analysed sample of Andalusians from Huelva are present in northwestern Africans, although allele frequencies are significantly different for particular cases. Nevertheless, β-thalassemia haemoglobin disorder is highly heterogeneous in North Africa, representing an identity mark of its native populations. Such a wide variation in mutation diversity is not observed among Iberian populations.

IVSI-1 (G > A), the most predominant β-thalassemia allele in southwestern Spain (present study), is also present in northwestern Africans but at much lower frequencies (e.g. 8.5–12%). Hypotheses about where *IVSI-1* arose and where it started spreading have given rise to lively debate. In Iberia (Amselem et al. 1988), Sicily (Giambona et al. 2011), Tunisia (Jouini et al. 2013), Turkey (Tadmouri et al. 2001), and Lebanon (Makhoul et al. 2005), the *IVSI-1* mutation has been found to be associated with a single, unique β-globin cluster haplotype (haplotype V). Nevertheless, in Algerian and Moroccan samples (mainly from Berber origin), it has also been found to be preferentially linked with haplotype V and, to a lesser extent, to other specific haplotypes. Given this genetic background among western Maghrebian autochthonous populations, Bennani et al. (1994) and Lemsaddek et al. (2004) have suggested a local ancient origin centred in the region from where the *IVSI-1* allele there would have migrated to neighbouring areas such as the Iberian Peninsula during the long-lasting Islamic Arab/Berber occupation. Although non-haplotype data are available for the Andalusian study sample, in Portuguese, the *IVSI-1* allele mutation has been observed in association with the same haplotypes as in northwestern Africans (Osòrio 2015). In this regard, the positive selection on heterozygous β-thalassemia in past malarial populations could permit the successful spreading of the *IVSI-1* allele by migratory movements towards localised areas of the Iberian Peninsula. The history of malaria in Europe seems to be a Holocene development (Sallares 2006). The prevalence of specific environmental conditions (i.e. malaria) in Huelva has been linked to population demography features until recently, all contributing to the incidence and rise of specific β-thalassemia genes, generating local genetic structure. Along this line, it would be highly interesting to explore whether *IVSI-1* spread along the Atlantic façade of the Iberian Peninsula from a high-frequency area, as it is the westernmost extreme of Andalusia, following the south-north "*Via de la Plata*". However, the available information on *IVSI-1* mutation frequencies in the westernmost Iberian populations is still very limited.

The relatively moderate frequency of the *IVSI-1* disease allele among eastern Andalusians from Granada (22%), a proportion that represents less than half of that reported in their

relatives from Huelva, deserves attention. The territory of Granada has never been identified as a malarial area, and its demography is different from that of western Andalusia. The complex history of Granada could have had a notable genetic impact on its autochthonous genomes. There, the Nazari Kingdom (thirteenth to fifteenth centuries) extended Islamic domination for two and a half centuries longer than in the rest of Spain. After the conquest of the Nazari Kingdom in 1492 and the expulsion of the Moriscos from Spain (descendants of the Muslim population), at the beginning of the seventeen century, the influence of the Maghreb gene pool in Andalusia, particularly in Granada, was substantially modified. Complex episodes of repopulation, colonisation, and population amalgamation processes with "old Christian" people from Castile, mainly from northern Meseta, Aragón, Galice, and even Portugal, occurred in the region (Caro-Baroja 1995). Therefore, all these intricate population movements could have significant effects on the genetic composition in Eastern Andalusia.

Genetic differentiation between western and eastern Andalusians has also been observed for the β-thalassemia *IVS I-110* allele, the anthropogenetic marker of the eastern Mediterranean. Andalusians from Granada register frequencies of the *IVSI-110* allele ∼27% (modal value within the Iberian Peninsula), whereas in Huelva Province, it varies from 9 to 12% (Molina et al. 1994; present study). The increased genetic influence of the eastern Mediterranean into the eastern Andalusian genetic structure is consistent with that indicated by other human genetic diversity studies in the region (Ambrosio et al. 2012; Hernández et al. 2015, 2017, 2019).

The information provided here on the number and types of mutations detected in a well-defined β-thalassemia sample from western Andalusia reflects the high interest in integrating multidisciplinary approaches (i.e. demographic, marital structure and environmental, among others) for a better understanding of the disease, which ultimately will help to establish well-founded health-care strategies. Likewise, the joint analysis of genetic variation associated with diseases (both common single-gene disorders and complex, multifactorial diseases) with that observed from genetic polymorphisms (with non-adaptative significance) within a population, represents modern research areas that not only uncover the implications of the types of mutations affecting human health but also provide insights into recent evolutionary histories.

Acknowledgements

We warmly thank all Andalusian blood donors participating in this study and other previously developed by this team in the region.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Ministerio de Economía, Industria y Competitividad, Gobierno de España (Projects: CGL2010-15191 and CGL2014-53985-R, both granted to RC). CLH has been supported by a research contract linked to the latter Project.

References

Aggarwal CC. 2015. Data mining: the textbook. New York (NY): Springer.
Ambrosio B, Novelletto A, Hernandez C, Dugoujon JM, Fortes-Lima C, Rodriguez JN, Calderón R. 2012. Y-STR genetic diversity in autochthonous Andalusians from Huelva and Granada provinces (Spain). Forensic Sci Int Genet. 6(2):e66–e71.
Amselem S, Nunes V, Vidaud M, Estivill X, Wong C, d'Auriol L, Vidaud D, et al. 1988. Determination of the spectrum of beta-thalassemia genes in Spain by use of dot-blot analysis of amplified beta-globin DNA. Am J Hum Genet. 43(1):95–100.
Avellò RC. 1950. La epidemia de paludismo de la postguerra. Rev San Hig Publ. 24:701–737.
Aydinok Y, Oymak Y, Atabay B, Aydoğan G, Yeşilipek A, Ünal S, Kılınç Y, et al. 2018. A National Registry of Thalassemia in Turkey: demographic and disease characteristics of patients, achievements, and challenges in prevention. TJH. 35(1):12–18.
Benito A, Villegas A, Perez-Cano R, Bernal R. 1996. Beta-thalassaemia in south-western Spain: high frequency of G->A (IVS I-1) mutation. Br J Haematol. 92(2):336–338.
Bennani C, Bouhass R, Perrin-Pecontal P, Tamouza R, Malou M, Elion J, Trabuchet G, et al. 1994. Anthropological approach to the heterogeneity of beta-thalassemia mutations in Northern Africa. Hum Biol. 66(3):369–382.
Cabeda JM, Correia C, Estevinho A, Simões C, Amorim ML, Pinho L, Justiça B. 1999. Unexpected pattern of beta-globin mutations in beta-thalassaemia patients from northern Portugal. Br J Haematol. 105(1):68–74.
Calderón R, Ambrosio B, Guitard E, González-Martín A, Aresti U, Dugoujon JM. 2006. Genetic position of Andalusians from Huelva in relation to other European and North African populations: a study based on GM and KM allotypes. Hum Biol. 78(6):663–679.
Calderón R, Hernández CL, García-Varela G, Masciarelli D, Cuesta P. 2018. Inbreeding in Southeastern Spain: the impact of geography and demography on marital mobility and marital distance patterns (1900-1969). Hum Nat. 29(1):45–64.
Cao A, Gossens M, Pirastu M. 1989. Beta thalassaemia mutations in Mediterranean populations. Br J Haematol. 71(3):309–312.
Cao C, Galanello G. 2010. Beta-thalassemia. Genet Med. 12(2):61–76.
Cao C, Kan YW. 2013. The prevention of thalassemia. Cold Spring Harb Perspect Med. 3(2):a011775.
Caro-Baroja J. 1995. Los moriscos del reino de Granada. Madrid: Istmo.
Chen LZ, Easteal S, Board PG, Kirk RL. 1990. Evolution of β -globin haplotypes in human populations. Mol Biol Evol. 7(5):423–437.
Cherry L, Calo C, Talmaci R, Perrin P, Gavrila L. 2016. β-thalassemia haplotypes in romania in the context of genetic mixing in the Mediterranean area. Hemoglobin. 40(2):85–96.
Colah R, Gorakshakar A, Nadkarni A. 2010. Global burden, distribution and prevention of β-thalassemias and hemoglobin E disorders. Expert Rev Hematol. 3(1):103–117.
Currat M, Trabuchet G, Rees D, Perrin P, Harding RM, Clegg JB, Langaney A, Excoffier L. 2002. Molecular analysis of the beta-globin gene cluster in the Niokholo Mandenka population reveals a recent origin of the beta(S) Senegal mutation. Am J Hum Genet. 70(1):207–223.
Epperson BK. 2003. Covariances among join-count spatial autocorrelation measures. Theor Popul Biol. 64(1):81–87.
Excoffier L, Lischer HEL. 2010. 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.
Falchi A, Giovannoni L, Vacca L, Latini V, Vona G, Varesi L. 2005. Beta-globin gene cluster haplotypes associated with beta-thalassemia on Corsica Island. Am J Hematol. 78(1):27–32.
Faustino P, Pacheco P, Loureiro P, Nogueira P, Lavinha J. 1999. The geographic pattern of beta-thalassaemia mutations in the Portuguese population. Br J Haematol. 107(4):903–904.

Fernández-Jurado J, García-Sanz C, Rufete P. 1997. De Tartessos a Onuba. Huelva: Diputación de Huelva.

Fortes-Lima C, Dugoujon J, Hernández C, Reales G, Calderón R. 2014. Immunoglobulin genes in Andalusia (Spain). Genetic diversity in the Mediterranean space. C R Biol. 337(11):646–656.

Gamella JF, Núñez-Negrillo AM. 2019. The evolution of consanguineous marriages in the archdiocese of Granada, Spain (1900-1979). Hum Biol. 90(2):97–114.

Giambona A, Vinciguerra M, Cannata M, Cassarà F, Fiorentino G, Leto F, Gioco PL, et al. 2011. The genetic heterogeneity of β-globin gene defects in Sicily reflects the historic population migrations of the island. Blood Cells Mol Dis. 46(4):282–287.

Hall T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser. 41:95–98.

Hartl D, Clark A. 2006. Principles of population genetics. 4th ed. Sunderland (UK): Sinauer Associates.

Hassan S, Hamza N, Al-Lawatiya F, Mohammed A, Harteveld C, Rajab A, Giordano PC. 2010. Extended molecular spectrum of beta- and alpha-thalassemia in Oman. Hemoglobin. 34(2):127–134.

Hernández CL, Soares P, Dugoujon JM, Novelletto A, Rodríguez JN, Rito T, Oliveira M, et al. 2015. Early holocenic and historic mtDNA African signatures in the Iberian Peninsula: The Andalusian Region as a Paradigm. PLOS One. 10(10):e0139784.

Hernández CL, Dugoujon JM, Novelletto A, Rodríguez JN, Cuesta P, Calderón R. 2017. The distribution of mitochondrial DNA haplogroup H in southern Iberia indicates ancient human genetic exchanges along the western edge of the Mediterranean. BMC Genet. 18(1):46.

Hernández CL, Dugoujon JM, Sánchez-Martínez LJ, Cuesta P, Novelletto A, Calderón R. 2019. 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.

Hernández CL, Pita G, Cavadas B, López S, Sánchez-Martínez LJ, Dugoujon JM, Novelletto A, et al. 2020. Human genomic diversity where the Mediterranean Joins the Atlantic. Mol Biol Evol. 37(4): 1041–1055.

Hofmanová Z, Kreutzer S, Hellenthal G, Sell S, Diekmann Y, Díez-Del-Molino D, van Dorp L, et al. 2016. Early farmers from across Europe directly descended from Neolithic Aegeans. Proc Natl Acad Sci USA. 113(25):6886–6891.

Holsinger KE, Weir BS. 2009. Genetics in geographically structured populations: defining, estimating and interpreting F(ST). Nat Rev Genet. 10(9):639–650.

Huertas F, Mendoza A. 1903. El Paludismo y su profilaxis en la provincia de Cáceres. Investigaciones y Estudios sobre Paludismo en España. Proceedings of the 4th International Congress of Medicine; Barcelona. Tipografía la Académica.

Joly P, Pondarre C, Badens C. 2014. Les beta-thalassemies: aspects molécularies, épidémiologiques, diagnostiques et cliniques. Ann Biol Clin. 72(6):639–668.

Jouini L, Sahli CA, Laaouini N, Ouali F, Youssef IB, Dakhlaoui B, Othmeni R, et al. 2013. Association between clinical expression and molecular heterogeneity in β-thalassemia Tunisian patients. Mol Biol Rep. 40(11): 6205–6212.

Kalleas C, Anagnostopoulos K, Sinopoulou K, Delaki E, Margaritis D, Bourikas G, Tsatalas C, et al. 2012. Phenotype and genotype frequency of β-thalassemia and sickle cell disease carriers in Halkidiki, Northern Greece. Hemoglobin. 36(1):64–72.

Kariuki SN, Williams TN. 2020. Human genetics and malaria resistance. Hum Genet. 139(6–7):801–811.

Legendre P, Fortin M. 1989. Spatial pattern and ecological analysis. Vegetatio. 80(2):107–138.

Lemsaddek W, Picanço I, Seuanes F, Nogueira P, Mahmal L, Benchekroun S, Khattab M, Osório-Almeida L. 2004. The beta-thalassemia mutation/haplotype distribution in the Moroccan population. Hemoglobin. 28(1):25–37.

López-Escribano H, Parera MM, Guix P, Serra JM, Gutierrez A, Balsells D, Oliva-Berini E, et al. 2013. Balearic archipelago: three islands, three beta-thalassemia population patterns. Clin Genet. 83(2):175–180.

Macdonald I. 1904. Etiología del Paludismo en la Provincia de Huelva. In: Proceedings of the 5th International Congress of Medicine; Madrid. p. 122–124.

Madoz P. 1847. Diccionario geográfico, histórico y estadístico de España. Madrid: La Ilustración.

Makhoul N, Wells RS, Kaspar H, Shbaklo H, Taher A, Chakar N, Zalloua PA. 2005. Genetic heterogeneity of Beta thalassemia in Lebanon reflects historic and recent population migration. Ann Hum Genet. 69(Pt 1):55–66.

Manel S, Schwartz MK, Luikart G, Taberlet P. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends Ecol Evol. 18(4):189–197.

Mathieson I, Alpaslan-Roodenberg S, Posth C, Szécsényi-Nagy A, Rohland N, Mallick S, Olalde I, et al. 2018. The genomic history of southeastern Europe. Nature. 555(7695):197–203.

Moatter T, Kausar T, Aban M, Ghani S, Pal J. 2012. Prenatal screening for β-thalassemia major reveals new and rare mutations in the Pakistani population. Int J Hematol. 95(4):394–398.

Modell B, Darr A. 2002. Science and society: genetic counselling and customary consanguineous marriage. Nat Rev Genet. 3(3):225–229.

Modell B, Darlison M, Birgens H, Cario H, Faustino P, Giordano PC, Gulbis B, et al. 2007. Epidemiology of haemoglobin disorders in Europe: an overview. Scand J Clin Lab Invest. 67(1):39–69.

Molina M, Romero M, Abril E, Delgado I, Cano R, Garrido F, de Pablos JM, Garrido ML. 1994. Frequency of molecular alterations in heterozygous beta-thalassemia in southern Spain and their relation to the hematologic phenotype. Sangre. 39(4):253–256.

Monni G, Peddes C, Iuculano A, Ibba R. 2018. From prenatal to preimplantation genetic diagnosis of β-thalassemia. prevention model in 8748 cases: 40 years of single center experience. JCM. 7(2):35.

Murad H, Moassas F, Ghoury I, Mukhalalaty Y. 2018. Haplotype analysis of three common β-thalassemia mutations in Syrian Patients. Hemoglobin. 42(5–6):302–305.

Orkin S, Kazazian HH, Antonarakis SE, Goff SC, Boehm CD, Sexton JP, Waber PG, Giardina PJ. 1982. Linkage of beta-thalassaemia mutations and beta-globin gene polymorphisms with DNA polymorphisms in human beta-globin gene cluster. Nature. 296(5858):627–631.

Osòrio L. 2015. Human gene mutations and migratory flows—portugal and the Mediterranean. Advances in Anthropology. 5(3):164–169.

Pellicer A. 1969. Studies on thalassemia, glucose-6-phosphate dehydrogenase deficiency, and sickle cell trait in the Province of Huelva (Spain). Am J Hum Genet. 21(2):109–114.

Pérez M, Moreno I, Bolufer-Gilabert P, Lerma E, Gómez M, Montaña A, Tascón A, et al. 1998. Molecular characterization of thalassemias in the Valencia community and its relationship with the hematological phenotype. Sangre. 43(5):392–398.

Perrin P, Bouhassa R, Mselli L, Garguier N, Nigon VM, Bennani C, Labie D, Trabuchet G. 1998. Diversity of sequence haplotypes associated with beta-thalassaemia mutations in Algeria: implications for their origin. Gene. 213(1–2):169–177.

Pinto-Cisternas J, Zei G, Moroni A. 1979. Consanguinity in Spain, 1911-1943: general methodology, behavior of demographic variables, and regional differences. Soc Biol. 26(1):55–71.

Puehringer H, Najmabadi H, Law HY, Krugluger W, Viprakasit V, Pissard S, Baysal E, et al. 2007. Validation of a reverse-hybridization StripAssay for the simultaneous analysis of common alpha-thalassemia point mutations and deletions. Clin Chem Lab Med. 45(5):605–610.

R Core Team 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Ribeiro M, Gonçalves P, Cunha E, Bento C, Almeida H, Pereira J, Núñez GM, Tamagnini GP. 1997. Genetic Heterogeneity of beta-thalassemia in populations of the Iberian Peninsula . Hemoglobin. 21(3):261–269.

Ropero P, Sánchez J, González F, Armada B, Benito A, Caldeira A. 1994. Heterogeneidad molecular de la β-talasemia. Sangre. 39:365–368.

Ropero P, González F, Cela E, Beléndez C, Cervera A, Martínez-Nieto J, de la Fuente-Gonzalo F, et al. 2013. Association in cis of the mutations +20 (C > T) in the 5′ untranslated region and IVS-II-745 (C > G) on the β-globin gene. Hemoglobin. 37(2):112–118.

Rosatelli M, Dozy A, Faà V, Meloni A, Sardu R, Saba L, Kan YW, Cao A. 1992. Molecular characterization of beta-thalassemia in the Sardinian population. Am J Hum Genet. 50(2):422–426.

Rosenberg MS, Anderson CD. 2011. PASSaGE: pattern analysis, spatial statistics and geographic exegesis. Version 2. Methods Ecol Evol. 2(3): 229–232.

Sallares R. 2006. Role of environmental changes in the spread of malaria in Europe during the Holocene. Quat Int. 150(1):21–27.

Sousa A, Garcia-Barrón L, Vetter M, Morales J. 2014. The historical distribution of main malaria foci in Spain as related to water bodies. Int J Environ Res Public Health. 11(8):7896–7917.

Tadmouri G, Tüzmen S, Ozçelik H, Ozer A, Baig SM, Senga EB, Başak AN. 1998. Molecular and population genetic analyses of beta-thalassemia in Turkey. Am J Hematol. 57(3):215–220.

Tadmouri G, Garguier N, Demont J, Perrin P, Başak A. 2001. History and origin of beta-thalassemia in Turkey: sequence haplotype diversity of beta-globin genes. Hum Biol. 73(5):661–674.

Taher AT, Weatherall DJ, Cappellini MD. 2018. Thalassaemia. Lancet. 391(10116):155–167.

Thein SL. 2013. Genetic association studies in β-hemoglobinopathies. Hematology Am Soc Hematol Educ Program. 2013:354–361.

Viganó C, Haas C, Rühli FJ, Bouwman A. 2017. 2,000 Year old β-thalassemia case in Sardinia suggests malaria was endemic by the Roman period. Am J Phys Anthropol. 164(2):362–370.

Villegas A, Ropero P, González F, Armada B, Martí B, Carreño M, Sánchez J. 1996. Molecular heterogeneity of beta-thalassaemia in a Spanish population from the area centre of Spain. Eur J Haematol. 57(5): 389–391.

Villegas A, Ropero P, González F, Anguita E, Espinós D. 2001. The thalassemia syndromes: molecular characterization in the Spanish population. Hemoglobin. 25(3):273–283.

Weatherall DJ. 2010. The inherited diseases of hemoglobin are an emerging global health burden. Blood. 115(22):4331–4336.

Weatherall DJ. 2013. The role of the inherited disorders of hemoglobin, the first "molecular diseases," in the future of human genetics. Annu Rev Genomics Hum Genet. 14:1–24.

Williams TN, Weatherall DJ. 2012. World distribution, population genetics, and health burden of the hemoglobinopathies. Cold Spring Harb Perspect Med. 2(9):a011692.

Yasmeen H, Toma S, Killeen N, Hasnain S, Foron L. 2016. The molecular characterization of Beta globin gene in thalassemia patients reveals rare and a novel mutations in Pakistani population. Eur J Med Genet. 59(8):355–562.