



OPEN ACCESS

EDITED BY
Miguel Arenas,
University of Vigo, Spain

REVIEWED BY
Veronica Fernandes,
Universidade do Porto, Portugal
Hannah Moots,
The University of Chicago, United States

*CORRESPONDENCE
Joan Fibla,
✉ joan.fibla@udl.cat
Oscar Lao,
✉ oscar.lao@ibe.upf-csic.es

†PRESENT ADDRESS
Oscar Lao, Institut de Biologia Evolutiva
(UPF-CSIC), Parc de Recerca Biomèdica
de Barcelona, Barcelona, Spain

SPECIALTY SECTION
This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

RECEIVED 16 November 2022
ACCEPTED 12 December 2022
PUBLISHED 10 January 2023

CITATION
Fibla J, Maceda I, Laplana M,
Guerrero M, Álvarez MM, Burgueño J,
Camps A, Fàbrega J, Felisart J, Grané J,
Remón JL, Serra J, Moral P and Lao O
(2023), The power of geohistorical
boundaries for modeling the genetic
background of human populations: The
case of the rural catalan Pyrenees.
Front. Genet. 13:1100440.
doi: 10.3389/fgene.2022.1100440

COPYRIGHT
© 2023 Fibla, Maceda, Laplana,
Guerrero, Álvarez, Burgueño, Camps,
Fàbrega, Felisart, Grané, Remón, Serra,
Moral and Lao. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

The power of geohistorical boundaries for modeling the genetic background of human populations: The case of the rural catalan Pyrenees

Joan Fibla^{1,2*}, Iago Maceda³, Marina Laplana¹,
Montserrat Guerrero⁴, Miguel Martín Álvarez⁵,
Jesús Burgueño⁴, Agustí Camps⁶, Jordi Fàbrega⁷,
Josefina Felisart⁸, Joan Grané⁹, José Luis Remón¹⁰,
Jordi Serra¹¹, Pedro Moral¹² and Oscar Lao^{3*†}

¹Department of Basic Medical Sciences, University of Lleida, Lleida, Spain, ²Institute of Biomedical Research of Lleida (IRBLleida), Lleida, Spain, ³CNAG-CRG, Centre for Genomic Regulation, The Barcelona Institute of Science and Technology, Barcelona, Spain, ⁴Department of Geography and Sociology, University of Lleida. Pl. Víctor Siurana, Lleida, Spain, ⁵Genome Data Science, Institute for Research in Biomedicine (IRB Barcelona), Barcelona Institute for Science and Technology, Barcelona, Spain, ⁶Hospital Sant Bernabé, Berga, Spain, ⁷Fundació Sant Hospital La Seu d'Urgell, La Seu d'Urgell, Spain, ⁸Hospital d'Olot i Comarcal de la Garrotxa, Olot, Girona, Spain, ⁹Hospital de Campdevàrol, Campdevàrol, Girona, Spain, ¹⁰Servei d'atenció primària Lleida Nord, Gerència Territorial Alt Pirineu i Aran, Institut Català de la Salut, Tremp, Spain, ¹¹Laboratori d'Anàlisis Clíniques, Hospital Comarcal del Pallars, Tremp, Spain, ¹²Department of Evolutionary Biology, Ecology and Environmental Sciences, Faculty of Biology, Biodiversity Research Institute, University of Barcelona, Barcelona, Spain

The genetic variation of the European population at a macro-geographic scale follows genetic gradients which reflect main migration events. However, less is known about factors affecting mating patterns at a micro-geographic scale. In this study we have analyzed 726,718 autosomal single nucleotide variants in 435 individuals from the catalan Pyrenees covering around 200 km of a vast and abrupt region in the north of the Iberian Peninsula, for which we have information about the geographic origin of all grand-parents and parents. At a macro-geographic scale, our analyses recapitulate the genetic gradient observed in Spain. However, we also identified the presence of micro-population substructure among the sampled individuals. Such micro-population substructure does not correlate with geographic barriers such as the expected by the orography of the considered region, but by the bishoprics present in the covered geographic area. These results support that, on top of main human migrations, long ongoing socio-cultural factors have also shaped the genetic diversity observed at rural populations.

KEYWORDS

genetic variation, rural population, inbreeding, population differentiation, bishopric boundaries, rural catalan Pyrenees

Introduction

The current genetic variation of European human populations at a macro-geographic scale is mostly driven by isolation by distance patterns (Lao et al., 2008) and relatively modern massive migrations (Hellenthal et al., 2014). These results have been also observed at a country-based geographic scale (Lao et al., 2013; Leslie et al., 2015; Bycroft et al., 2019; Raveane et al., 2019). Nevertheless, despite this general trend, some populations appear as genetic outliers, mostly related to geographically isolated rural and religious areas (Xue et al., 2017), as well as regions showing peculiar orographic events, such as islands (Biagini et al., 2019; Font-Porterías et al., 2019; Marcus et al., 2020). Self-identified ethnicity (i.e., Romani people (Font-Porterías et al., 2019)) and ethnolinguistic background (such as Basques (Flores-Bello et al., 2021)) can also promote genetic barriers at different scales.

A basic question is whether there are socio-cultural factors in Europe that remained immutable over time and affected mate preferences at a microgeographic scale. This question is complex, as the recent demography of European populations has been shaped by a progressive regional concentration (Martí-Henneberg, 2005). This phenomenon is mainly explained by the continuous migration from rural areas towards industrialized urban areas during the last century (Brown, 2011). Consequently, micro-geographic patterns can be expected to be currently blurred at urban areas, and patterns shaped by cultural factors present at rural areas obscured by their recent depopulation.

Studying a relatively isolated European rural geographic region showing geohistorical singularities that maintain over time political and cultural borders could help understanding the role of different socio-cultural factors in mating patterns during the last centuries. Within this context, the catalan Pyrenees appear as a good candidate. First, the catalan Pyrenees constitute a vast and abrupt region in the north of the Iberian Peninsula, delimited to the east by the Mediterranean Sea, to the west by the Aigüestortes National Park and the Noguera Ribagorçana river valley, to the north by the border with France and Andorra and to the south by the central depression of Catalonia. A marked orography is distributed on a west-east axis along the Prepyrenees massifs (Figure 1).

Most of the Pyrenean valleys run from north to south, except for the maritime ends and the Puigcerdà-La Seu d'Urgell axis. The Pyrenean valleys, of river origin, run in parallel to the watershed, with narrow gorges, which does not facilitate communications between valleys (Gimenez, 2005). Along its entire extension, north-south communication axes are imposed, following the fluvial courses, which force its residents to move to the lower regions to ensure intra-Pyrenees relations. Furthermore, communications between valleys have been severely limited over time. It is at the end of the 19th and beginning of 20th century in which the construction of roads begins, first following the course of the rivers in a north-south axis, and much later the first road connections between the valleys were established. The limitations in the connections are strongly marked on the western slope, valleys of the Noguera

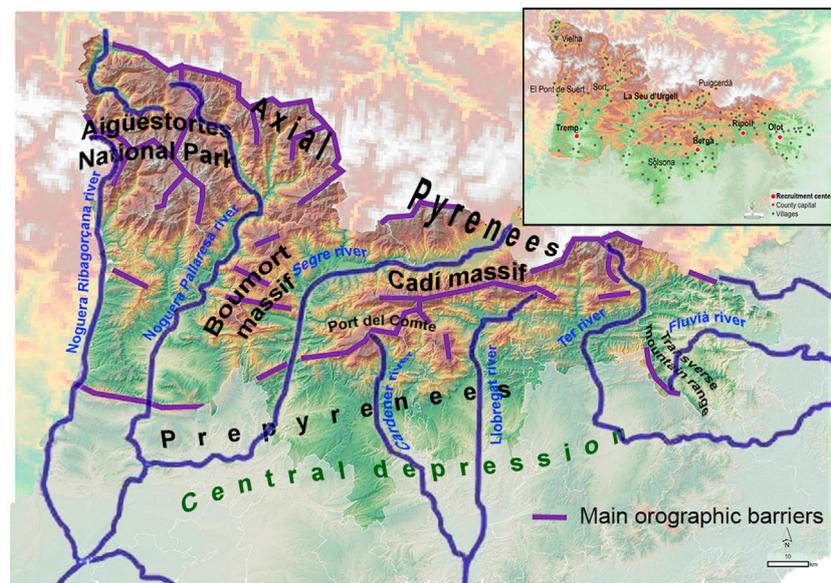


FIGURE 1

Orographic map of the catalan Pyrenean region considered in this study, depicting the most relevant orographic elements as massifs, rivers and depressions. Labeled by purple line the main orographic barriers defined by mountains higher than 1,500 m. Insert: location of villages (gray dotted points), comarca capitals (red dotted points), and recruitment centers (white-red circles).

Pallaresa and Segre rivers, while on the eastern slope, much less steep, the communications network is denser and earliest (Font, 1993) (see [Supplementary Figure S1](#)).

Second, the sum of all these mountain territories accounts for approximately a quarter of the surface of Catalonia, while the current population (70,000 people) barely exceeds 1% of the total catalan population. Throughout history, the demography of the Pyrenean region has undergone marked variations. The times of maximum splendor go back to the Middle Ages, followed by successive demographic oscillations. In Modern times, the maximum population was reached in the 1880s, followed by a progressive population decline until today (Gallart, 2002). Due to the orography of the region, a distinctive feature of the Pyrenees settlements is its scattering, with villages mainly centered around the rivers and lower part of the valleys (see insert [Figure 1](#)).

Finally, from a political point of view there are six Pyrenean *comarques* of Catalonia, corresponding to Alt Urgell, Alta Ribagorça, Val d'Aran, Pallars Jussà, Pallars Sobirà and Cerdanya and the northern part of another four more, Solsonès, Berguedà, Ripollès and Garrotxa ([Supplementary Figure S3F](#)). The Iberian peninsula, and in particular the ancient Pyrenees, are considered to have been culturally and genetically heterogeneous (Martin and Vaquer, 1995; Olalde et al., 2019). Before the mid-sixth millennium BC, the region was inhabited by Mesolithic hunter-gatherer communities (Oms et al., 2016). Several pre-Roman Pyrenean tribes mentioned by ancient sources are already distributed roughly following a West-East axis, such as the Airenosini (Pallars, Alt Urgell), Bergistani (Alt Urgell, Berguedà) and Castellani (Ripollès, Garrotxa). It is estimated that the Roman and Visigoth people only represented a 2.2%–4.4% of the Pyrenean population, while the Islamic conquest lasted only 80 years (Riu, 1995). Administrative-social boundaries have remained stable practically since the Middle Ages. The oldest administrative divisions correspond to the limits established between bishoprics, with the bishopric of Urgell (Diocesis Urgellensis) being the oldest (531 A.D.), followed by the Diocese of Vic (Diocesis Vicensis, 561 AD), Girona (Diocesis Gerundensis, 4th century), Lleida (Diocesis Ilerdensis, 12th century) and Solsona (Diocesis Celsonensis, 16th century) ([Supplementary Figure S2A](#)). Other political-administrative divisions have been superimposed on these administrative divisions, such as the Middle-Age counties (11th century) ([Supplementary Figure S2B](#)), *Vegueries* (12th–18th centuries) ([Supplementary Figure S3C](#)), *Corregimientos* (18th century) ([Supplementary Figure 2D](#)), Provinces and *Comarques* (19th century) ([Supplementary Figure S2E](#)) (Burgueño, 1991, 1992). Despite these multiple denominations, the orographic backbone has imposed that they present great coincidences in their limits, so we could consider that these divisions have perpetuated the limits

established from the old bishoprics with few variations ([Supplementary Figure S2A](#)).

In a previous study using whole genome sequencing on a limited sample size of 29 individuals (Maceda et al., 2021), we showed that populations from the catalan Pyrenees were interesting for understanding the genetic diversity of Spanish rural populations. More importantly, we observed the presence of micro-population substructure in this area, despite just covering populations in a region of around 200 km. Nevertheless, the lack of a continuous geographic sampling over the whole area of that study and the low sample size prevented identifying the atomic units of such substructure, as well as testing the factors that shaped the distribution of their genetic variation in the last millennia.

In the present study we have conducted a large genotyping study on 435 samples, analyzing and interpreting their genetic diversity in orographic and cultural terms.

Methods

Sample recruitment

We have collected samples from five reference hospital centers belonging to three health regions expanding the catalan Pyrenees. Recruitment period was extended from July 2014 to October 2016. In total, we recruited 435 samples (51.3% females) of subjects born and with all their grandparents born in the Pyrenean region (see insert [Figure 1](#)).

The data set represented an oldest extract of the population, with an average age of 55.6 years and equal proportion of sexes ([Table 1](#)). Therefore, this sample should be considered moderately affected by the demographic changes that occurred during the 19th and 20th centuries. All subjects signed an informed consent, and the study had the approval of the Ethics Committee of the University of Barcelona.

For each participant, we picked up demographic data such as sex, age and place of birth of the participant and place of birth their father and mother and for all their four grandparents. Birth places were assigned to geographic coordinates using *Vissir3*, the official viewer of the Cartographic and Geological Institute of Catalonia (Vidal, 2012). Assignment of centroid location of participants was done by calculating the mean longitude and latitude of grandparents (see [Supplementary Table S1](#)). Regional cartographic base maps and the terrain elevation model were from the Cartographic and Geological Institute of Catalonia. Maps were made with ArcGis 8.1 software (ESRI, Redlands, CA: Environmental Systems Research Institute), using facilities at the service of Cartography & SIG, of the University of Lleida (UdL).

TABLE 1 Distribution of samples according to recruitment center, region and bishopric.

Classification by	Area of recruitment	n (% female)	Age (mean; 95%CI)	Age (median; range)
Recruitment center (Health region)				
	Hospital Comarcal del Pallars (Alt Pirineu-Aran)	86 (53.5%)	56.9; 54–59.9	56.5; 19–87
	Fundació Sant Hospital La Seu d'Urgell (Alt Pirineu-Aran)	112 (65.2%)	49.2; 46.7–51.8	50.0; 15–76
	Hospital de Campdevàno (Catalunya Central)	86 (52.9%)	52.1; 48–56.2	55.0; 15–93
	Hospital Sant Bernabé (Catalunya Central)	45 (37.8%)	57.8; 52.9–62.7	59.0; 15–86
	Hospital d'Olot i Comarcal de la Garrotxa (Girona)	106 (40.0%)	62.9; 60.1–65.7	63.0; 17–90
Pyrenean region	West	198 (60.1%)	52.8; 50.9–54.8	87; 15–52.9
	Mid	85 (52.9%)	54; 50.3–57.7	93; 18–53.7
	East	152 (38.8%)	61.8; 59.5–64.1	90; 17–62.5
Bishoprics	Lleida	12 (33.3%)	54.3; 45.1–63.6	83; 29–54.1
	Urgell	182 (62.1%)	52.7; 50.7–54.7	87; 15–52.8
	Solsona	92 (52.2%)	54.9; 51.4–58.4	93; 18–54.8
	Vic	30 (36.7%)	59.2; 54.8–63.5	80; 39–59.1
	Girona	119 (39.5%)	61.9; 59.1–64.7	90; 17–62.8
Full GENPIR data set		435 (51.3%)	55.6; 54.0–57.1	57.0; 15–93

Genotyping

We collected 1 ml of peripheral blood from each participant in EDTA-coated tubes. All participants were genotyped using Applied Biosystems Axiom™ Precision Medicine Diversity Array (PMDA) on GeneTitan™ Multi-Channel Instrument (Thermo Fisher, CA, United States) at *Centro Nacional de Genotipado* facilities (CEGEN, nodo Santiago de Compostela, Spain). Genotype clustering was analyzed using the Axiom Analysis Suite 6.0 software adjusting for BestWorkflow settings. Quality control (QC) was tuned to sample calling rate >0.95, marker calling rate >0.90, and Hardy-Weinberg equilibrium (HWE) >10⁻⁵. After removing Indel and copy number variation (CNV), and QC, 780,310 SNPs (726,718 at autosomes, 49,319 at chromosome X, 406 at chromosome Y, 3,457 at pseudoautosomal region and 410 at mitochondrial genome) on 435 subjects were available for subsequent analyses (thereafter GENPIR data set).

Relatedness among individual participants

We estimated relatedness among samples by using SNPRelate R package (Zheng et al., 2012). Identical by descent (IBD) analysis reveals a degree of relatedness among several individuals (Supplementary Figure S3A). In order to consider unrelated individuals for further analyses, we selected one individual from each pair of individuals having a

kinship ≥ 0.09 (Supplementary Figure S3B). This yields a final set of 397 unrelated individuals.

Data phasing and IBD tract estimation

Data was phased using the Michigan-imputation server-1.5.7 pipeline (Das et al., 2016) with the eagle-2.4 phasing (Loh et al., 2016), using TopMed reference panel (Version R2 on GRC38) as reference.

IBD tracts between pairs of chromosomes from different individuals and homozygote by descent (HBD) fragments between chromosomes from the same individual were inferred with Beagle 4.0 (IBD_Beagle) (Browning and Browning, 2013), using the human recombination map provided by the same software (Browning and Browning, 2013).

Merging with reference populations

To make comparisons with other populations, GENPIR genotype data was merged with genotype data from samples of the 1,000 Genomes Phase 3 project (1KG3) (2,504 individuals) (Auton et al., 2015) and Human Genome Diversity Panel (HGDP) (929 individuals) (Bergström et al., 2020). After merging with the GENPIR dataset, a total of 3,868 individuals and 738,061 single nucleotide variants (SNVs) were selected with a

genotyping rate of 0.989 (690830 at autosomes, 46,947 at chromosome X, 280 at chromosome Y). We also applied unrelatedness criteria to the merged data set resulting in a final worldwide set of 3,805 unrelated individuals that were used for further analyses.

Feature extraction analyses

Principal component analysis (PCA)

PCA was done on the unrelated set of GENPIR and worldwide merged datasets using plink1.9 software (Chang et al., 2015), after performing QC that comprised: excluding markers with minor allele frequency (MAF) <0.05 and pruning to-*indep-pairwise* 200 25 0.3. After QC passing, PCA analysis was done on 3,805 subjects (2026 males, 1779 females) and 175,277 autosomal markers. PCA QC control was also performed to a data set obtained after extracting subjects belonging to GENPIR and European populations including Finland (FIN, $n = 99$), Russia (RUS, $n = 25$); Great Britain and Scotland (GBR, $n = 91$); Orcadian (Scotland) (ORC, $n = 15$); Utah residents (CEPH) with Northern and Western European ancestry (CEU, $n = 93$); French Basques (FRB, $n = 23$); France (FRE, $n = 28$); Iberian Peninsula (Spain) (IBS, $n = 107$); Bergamo (Italy) (BGT, $n = 12$) and Toscana (Italy) (TSI, $n = 115$), (1,005 samples, 492 males and 513 females; 152,717 autosomal SNVs) in order to perform independent PCA analyses. European populations of Adygey and Sardinia were not included in the PCA analysis as they are genetically isolated populations which can bias the results.

Admixture

ADMIXTURE v1.30 (Alexander et al., 2009) was used to analyze the population structure and calculate the corresponding cross-validation error value (CV) of the considered clusters ($K = 2-4$). ADMIXTURE analysis was done on merged GENPIR and European populations (1,005 samples, 492 males and 513 females; 152,717 autosomal SNVs) and applying the same QC control as for PCA analysis. Distribution of derived European components from $K = 3$, classified as North (N) and Central (C), as from $K = 4$, classified as Southwest (SW) and Southeast (SE), were compared among Pyrenean regions (West, Mid, East) by pairwise *t*-test.

Multiple correspondence analysis

The bishopic origin of the grandparents from a given sample was projected in two dimensions using a multiple correspondence analysis (Greenacre, 2007) as implemented in the package FactoMiner (Lê et al., 2008) in R software (Team R core, 2020).

SMACOF

Given the regionality of the sampled dataset, the presence of population substructure must be relatively small. Since the

IBD statistic is sensitive towards subtle hidden genetic relationships between individuals, we computed an IBD matrix with Beagle (Browning and Browning, 2013). We projected the IBD_Beagle matrix in a lower dimensional space after transforming it into dissimilarity by means of a non-metric ordinal SMACOF analysis (Leeuw and Mair, 2009) implemented in R software.

Estimating ancestry proportions across geographic space

We evaluated the genetic ancestry across geographic space on GENPIR population autosomal data set by tess3r software (Caye et al., 2018), applying the following parameters: method = “projected.ls”, ploidy = 2, max.iteration = 200, rep = 10, keep = “best”, tolerance = $1e-05$. We kept the most highly supported model (i.e., “best” based on cross-entropy scores) within each of the 10 replicates. Two subjects were excluded lacking centroid assignment. We also excluded markers with $MAF < 0.05$ as well as after pruning to-*indep-pairwise* 200 25 0.3.

Mantel test and procrustes

The relationship between genetic variation and cultural or political units while controlling by the putative confounder effect of geography was tested by means of a partial Mantel test (Lichstein, 2007) implemented in the R package vegan (Oksanen et al., 2020).

A Procrustes analysis (Peres-Neto and Jackson, 2001) between geography and the coordinates estimated by PCA or SMACOF using genetic variation was computed using the R package vegan.

MCLUST

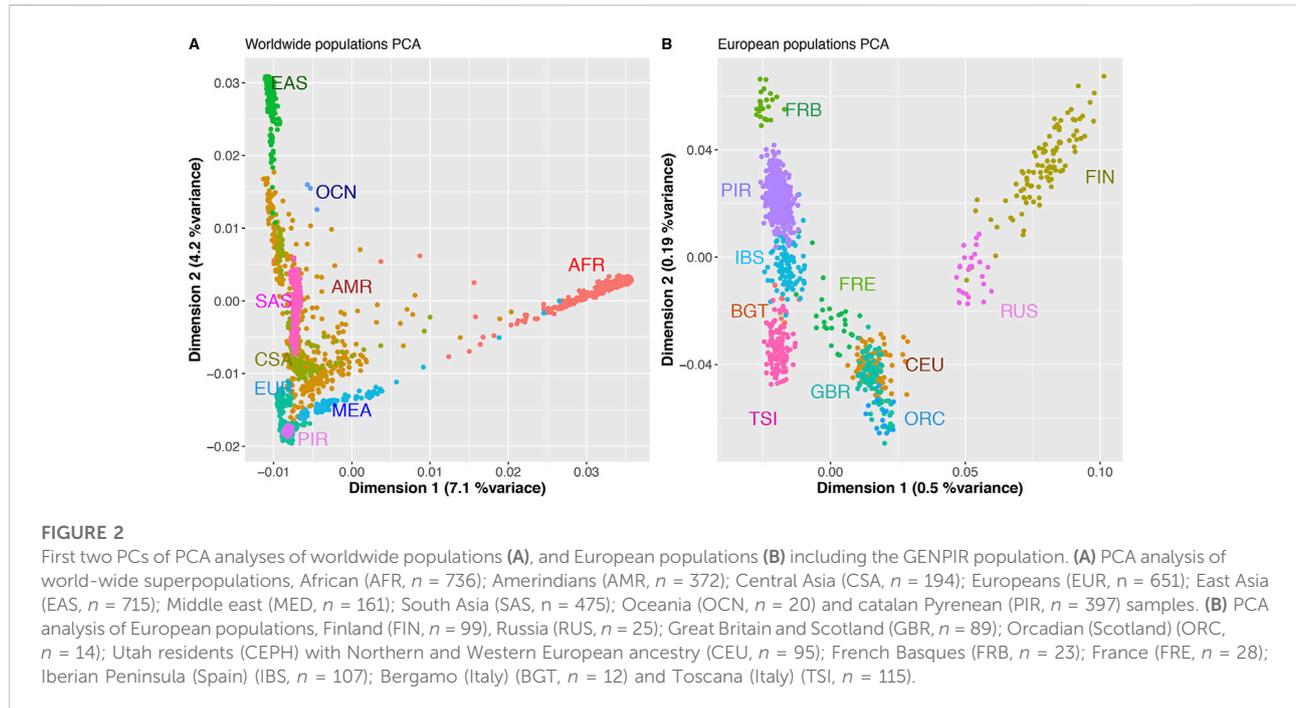
Unsupervised clustering of the individuals using the first two dimensions computed by SMACOF was conducted by means of the MCLUST package (Scrucca et al., 2016) implemented in R. Mclust models the points as a finite mixture of Gaussian distributions. Ascertainment of the best number of clusters was conducted based on BIC metrics.

Results and discussion

The genetic diversity of the catalan Pyrenees in the context of worldwide populations

We first analyzed the genetic relationships of the sampled individuals at worldwide level using a Principal Component Analysis (PCA) (Figure 2, Supplementary Figure S4).

The first two PCA components (PCs), explaining 11.3% of variance, show that the considered catalan Pyrenees samples are genetically related to other samples with European



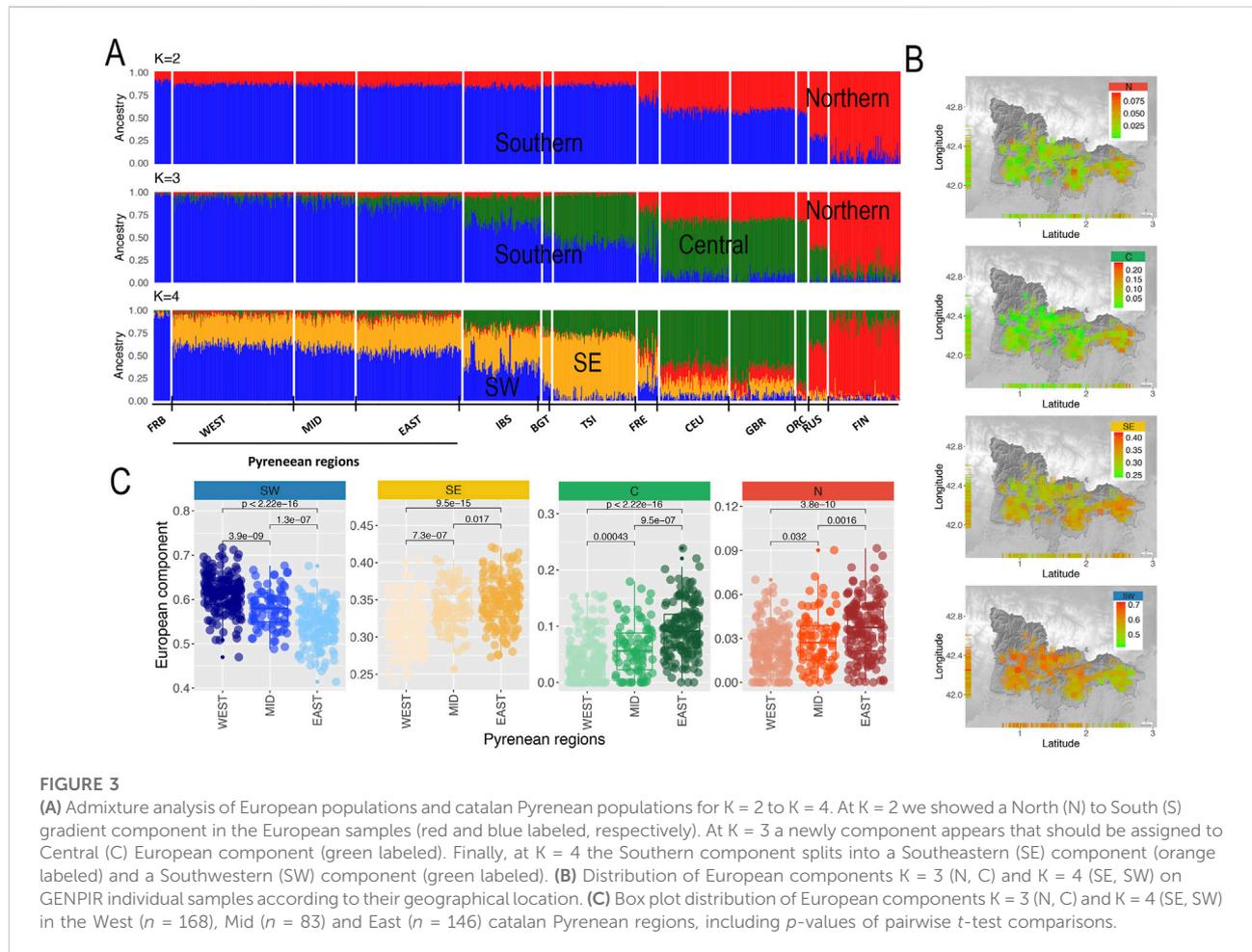
ancestry (Figure 2A). Within Europe, the catalan Pyrenees samples fall within the first two PCs (explaining 0.7%) close to samples from Southern Europe, close to individuals from Iberian and French Basque origin (Figure 2B). Scatterplots of the top 4 PCs with proportional variance explained included are also presented in Supplementary Figure S4.

Given these results, we run an unsupervised ADMIXTURE analysis focusing on the European region for K ranging from one to four in order to quantify the European ancestry proportions in GENPIR samples, revealing a minimum cross validation error value for $K = 2-3$ (Supplementary Figure S5).

As shown in Figure 3A, for $K = 2$ a North (red) to South (blue) clinal component was observed that was extended to an additional Central (green) component at $K = 3$. Finally, for $K = 4$, Southern component was split into a Southeastern (orange) and Southwestern (blue) component. Considering these four major components maximizing in present-day Southwestern European populations as “Northern (N)”, “Southeastern (SE)”, “Southwestern (SW)” and “Central (C)” (Figure 3A), we estimated their proportion in our samples. Figure 3B, C shows a decreasing gradient for the contribution of N, C and SE components being highest at the East, intermediate at the Mid and lowest at the West Pyrenean regions. In contrast, the SW European component shows an increasing gradient being lowest at the East, intermediate at the Mid and highest at the West Pyrenean regions. These results agree with the longitudinal gradient observed on the same region by other studies using classical markers (Calafell and Bertranpetit,

1994), as well as with the main axes of genetic differentiation within the Iberian Peninsula (Bycroft et al., 2019). Therefore, all these results support that the catalan Pyrenees fall within the genetic continuum within the Iberian Peninsula. These overall patterns have been previously interpreted in terms of ancient migrations (Calafell and Bertranpetit, 1994), as well as on the more recent migrations of Christian kingdoms and linguistic differences during the middle ages (Bycroft et al., 2019). In particular, the West to East pattern observed in the Iberian Peninsula has been explained by the admixture between Paleolithic and Neolithic populations and further geographic isolation, independently of which was the cause of that isolation (i.e., Catholic Kingdoms or linguistic differences) (Ferreiro et al., 2021).

At a local level, the presence of a higher proportion of SW European ancestry in the West region of the catalan Pyrenees supports that this region has been more accessible for migrations than other populations from the Iberian Peninsula. In addition, the higher contribution of N, C and SE European components at the East region of the catalan Pyrenees reflects a major permeability of this Pyrenean region, where geographical barriers are much weaker and with a much richer communications network, that has allowed contact with the nearby regions of northern and eastern Europe. We noted that the SE component, being significant on the Mid and East Pyrenean regions, reflects potential migratory events from Southeast Mediterranean coasts.



Population substructure is present at a microscale level within the catalan Pyrenees

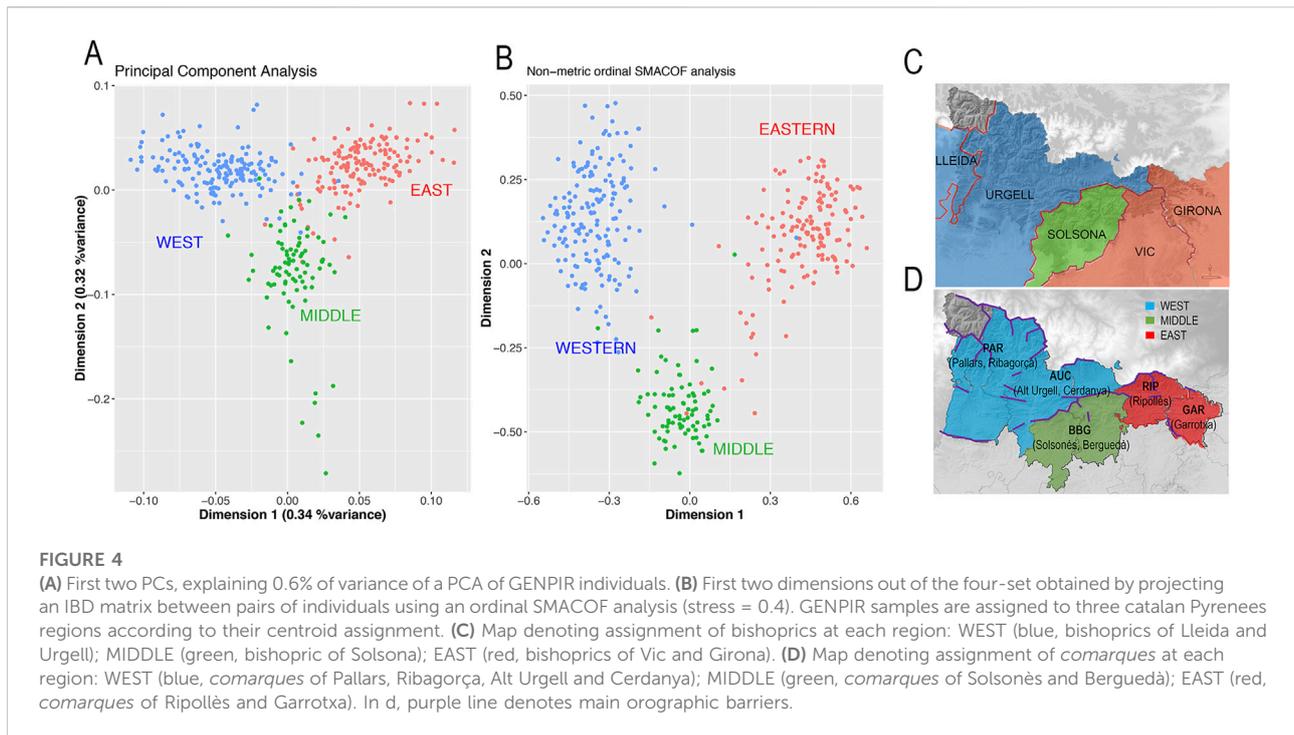
Given the observed differences in ancestry among the GENPIR samples, we conducted a PCA only using GENPIR samples (Figure 4A; 0.5% of variance).

We observed that the genetic map defined by the two PCs supports the presence of a non-random distribution of GENPIR individuals. Individuals from West, East and Middle Pyrenean regions tend to be grouped into three prominent clusters (Figure 4A). Since shared haplotype tracts are more sensitive for identifying subtle population substructure (Lawson et al., 2012), we wondered if the pattern observed by the PCA could be enhanced by projecting the IBD matrix between pairs of individuals estimated from Beagle 4.0 (Browning and Browning, 2013). We applied an ordinal SMACOF approach, in which the distance between pairs of individuals is interpreted in rank terms rather than by its magnitude. This decision was taken due to the fact that the overall

genetic differences between all individuals are expected to be limited. Therefore, the presence of few outliers showing a higher degree of relationship (even after excluding highly related individuals) could distort the projection towards these relationships rather than describing the patterns observed among all samples. The first two dimensions from the SMACOF analysis considering 15 dimensions (stress = 0.4) produce a similar picture as observed by PCA (Figure 4B). However, it enhances the differences between the three different clusters of populations, supporting micro substructure in the Pyrenees.

Furthermore, this micro-population substructure can be interpreted in geographic terms. In particular, a procrustes analysis between the geographic coordinates of the samples and the genetic map defined by the first two PCA dimensions showed a symmetric correlation of 0.61 (p -value after 10,000 permutations $<9.999e-05$).

Taking into account the orography of the region, the Cadi Massif constitutes a clear geographic barrier detaching the Western to the Middle and Eastern regions while no clear geographic barrier appears to detach the Middle to the Eastern



regions (Supplementary Figure S1). When applying actual (*comarques*) and historical (bishoprics) administrative-religious boundaries, the West cluster comprises Pallars, Ribagorça, Cerdanya and Alt Urgell *comarques* as well as Lleida and Urgell bishoprics; the Middle cluster comprises Solsonès and Berguedà *comarques* and Solsona bishopric and the East cluster includes Ripollès and Garrotxa *comarques* as well as Vic and Girona bishoprics (Figures 4C, D).

Micro-population substructure within the catalan Pyrenees is better explained by geohistorical organization (religious) structures

Given these results supporting a geographic interpretation of the genetic variation, we wondered whether models incorporating geographic information could identify genetic barriers present between the samples. The ancestry map generated by tess3r at $K = 2$ to $K = 4$ (Figure 5) identifies a strong longitudinal component that agrees with ADMIXTURE analyses and procrustes, and splits individuals following a geographic partition that matches the genetic differentiation observed in previous analyses.

Interestingly, these divisions do not fully agree on orographic events, such as the presence of mountains, that could act as genetic barriers. Particularly, the division that is observed between East and West at $K = 2$ is set in a region lacking

clear geographic barriers. Therefore, as previously considered, other factors correlating with geography must be shaping the observed genetic differentiation across the region. We also point out that at $K = 3$ the algorithm mainly reproduces the West, Middle and East clustering, which we previously observed with both PCA and SMACOF analysis.

We wondered if such genetic clusters identified in previous analyses were shaped by current administrative divisions (*comarques*), or if they were influenced by other long term ongoing cultural factors such as the limits defined by middle age-founded bishoprics. We applied a Mclust analysis using the first two dimensions from SMACOF in order to identify clusters showing a similar probability distribution function. The optimal number of clusters identified by Mclust was six (BIC = 406.33) that accurately reflect a hardest effect of geographic barriers that is shaped by administrative divisions, both *comarques* and bishopric (Figure 6). Cluster I includes subjects from the *comarques* of Pallars and bishopric of Lleida, cluster II subjects from the *comarques* of Pallars and Alt Urgell as well as bishopric of Urgell, cluster III mainly includes subjects from the comarca of Alt Urgell and bishopric of Urgell, cluster IV subjects from comarca of Solsones and Berguedà as well as bishopric of Solsona, cluster V subjects from comarca of Ripollès and bishopric of Vic and cluster VI mainly includes subjects from comarca of Garrotxa and bishopric of Girona (Figure 6).

Interestingly, such bishopric division from the SMACOF bi-plot is also observed within the geographic region

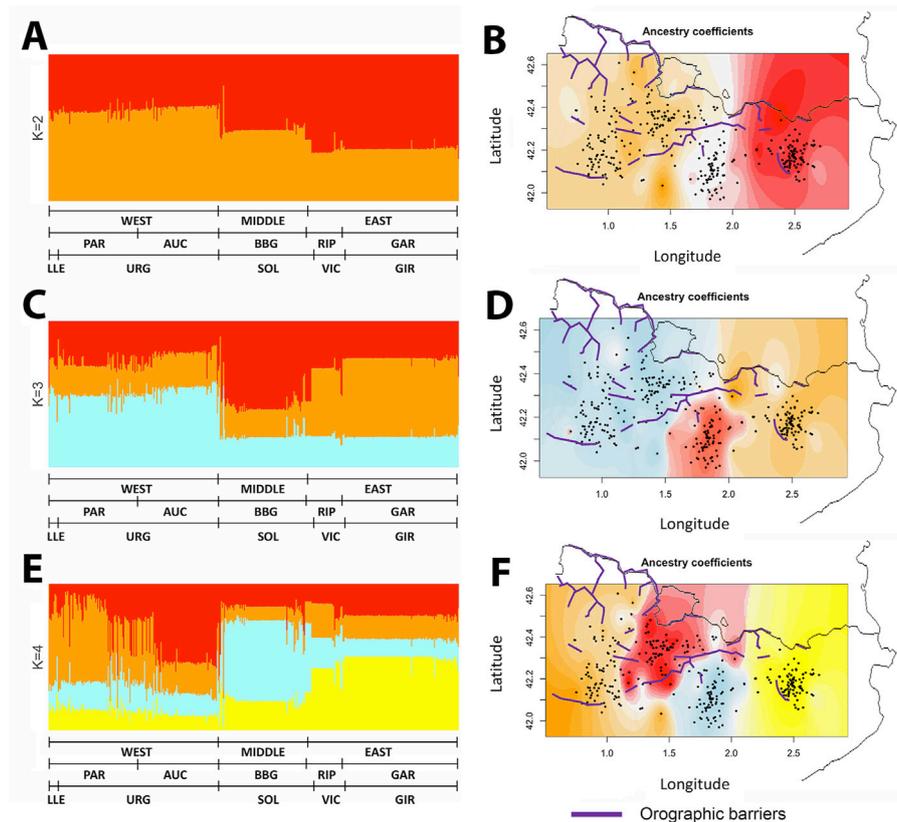


FIGURE 5

Spatial distribution analysis of the ancestry of GENPIR individuals. Barplot (A) (C), (E) and Geographical (B) (D), (F) pattern distribution of ancestry components of GENPIR samples assuming $K = 2$ to $K = 4$. In (A) (C) and (E), GENPIR samples are assigned to WEST, MIDDLE or EAST catalan Pyrenees regions, *comarques* (PAR, Pallars-Ribagorça; AUC, Alt Urgell-Cerdanya; BBG, Berguedà and Solsonès; RIP, Ripollès and GAR, Garrotxa) and bishoprics (LLE, Lleida; URG, Urgell; SOL, Solsona; VIC, Vic and GIR, Girona), according to their centroid assignment. In (B) (D) and (F), individual values of the ancestry components of GENPIR samples for $K = 2$ to 4 are plotted over the geographic location of each subject. Purple line denotes main orographic barriers.

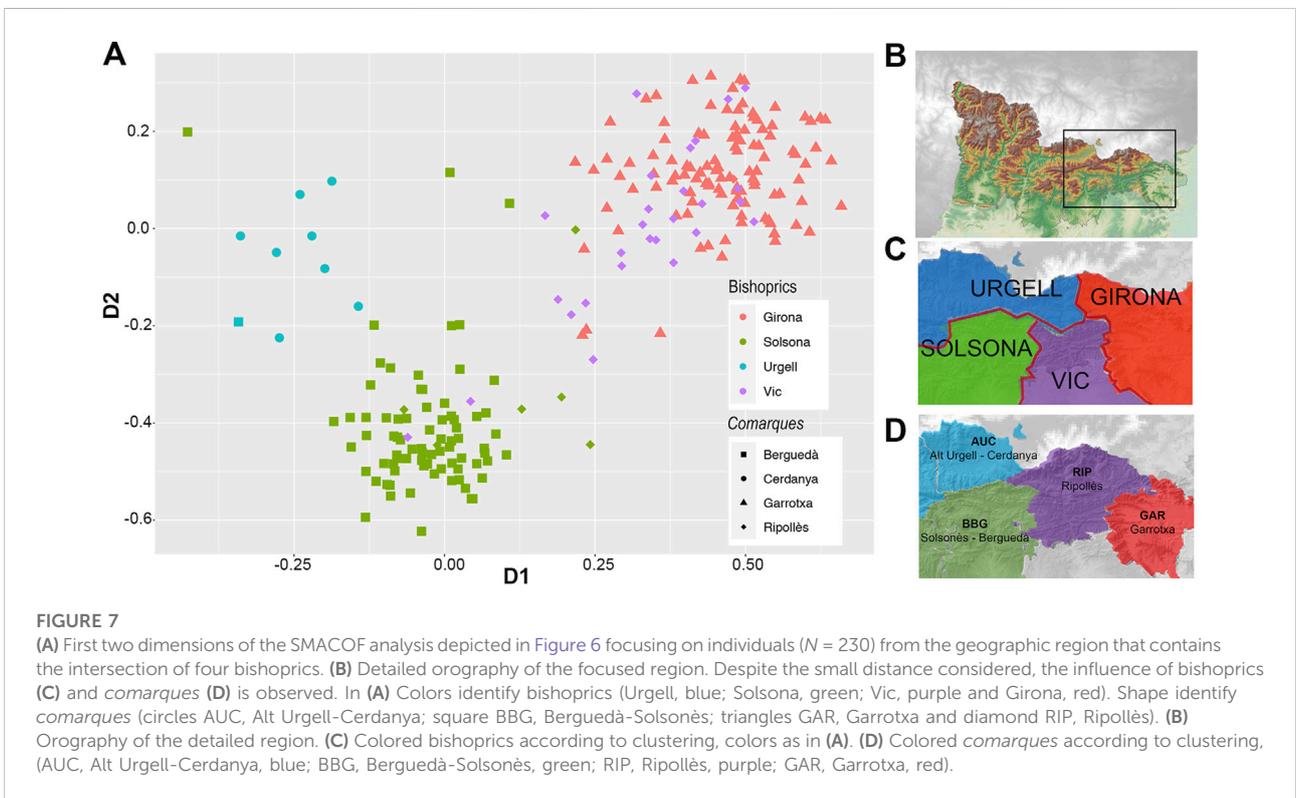
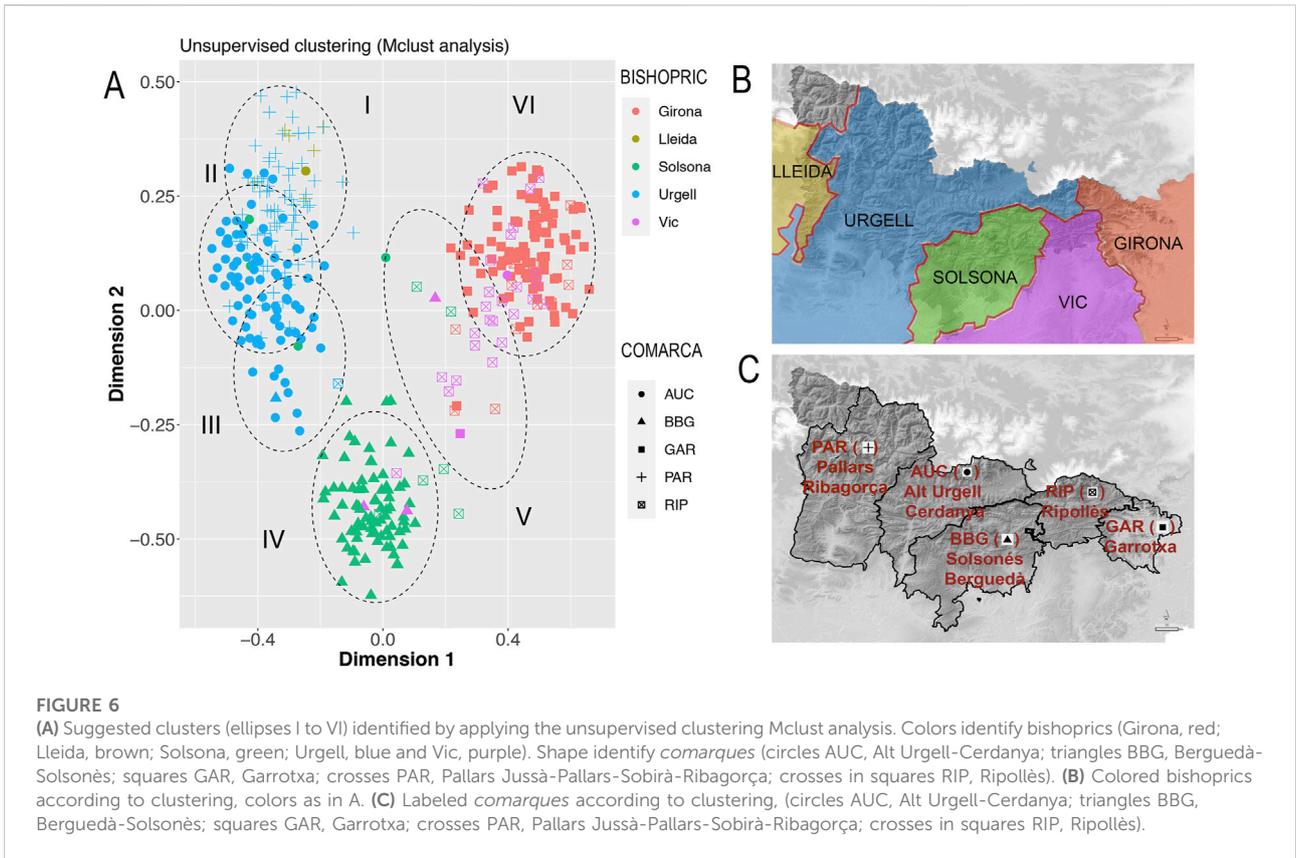
(depicting ~70 km far away) that contains the intersection of four bishoprics (Figure 7).

Since bishoprics and *comarques* depend on geography, the identified geographic clusters by tess3r could reflect cultural and/or political barriers. We conducted a partial Mantel test between the Euclidean distance estimated from the first two PCs, and the assignment to the same administrative division or bishoprics, controlled by the geographic distance. Pearson's correlation between PC based distances and bishoprics after controlling by geography in a partial Mantel test was 0.434 (p -value $< 1e-04$ using 9,999 permutations). This degree of association is higher than the observed between genetics and administrative divisions of *comarques* (Pearson's correlation = 0.324, p -value $< 1e-04$ using 9,999 permutations). When considering a classification based on bishoprics and administrative divisions at the same time, Pearson's correlation decreased to 0.294 (p -value $< 1e-04$ using 9,999 permutations). Similar results supporting higher correlation between bishopric divisions and genetic diversity

were obtained when using the first two dimensions from the SMACOF analysis conducted with the IBD distance matrix. The degree of correlation between genetics and bishoprics after controlling by geographic distance was 0.563 (p -value $< 1e-04$ using 9,999 permutations). This correlation decreases to 0.382 (p -value $< 1e-04$ using 9,999 permutations) when considering administrative divisions and it is similar when considering both bishopric and administrative divisions ($R = 0.361$, p -value $< 1e-04$ using 9,999 permutations). Overall, these results suggest that bishoprics play a role in shaping the genetic variation of GENPIR samples by modifying their mating patterns.

We supported such a hypothesis based on genetic data by taking advantage of the sampling strategy conducted in this study, which incorporated the geographic birthplace of the four grandparents (Figure 8, Supplementary Table S1).

Figure 8 shows the grandparent to *propositus* network in which each *propositus* is linked to their four grandparents generating a complex network centered on bishopric



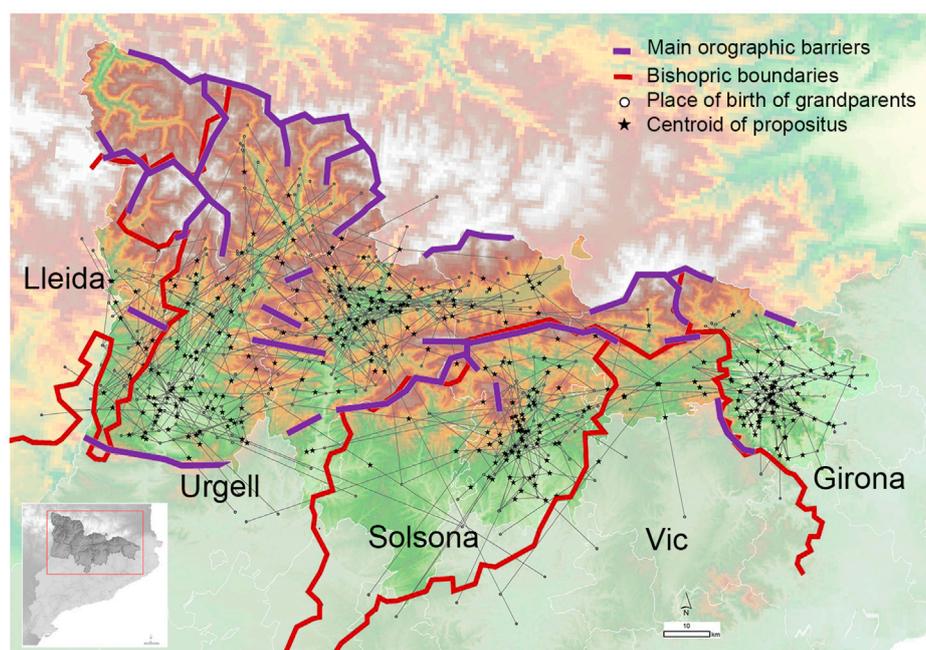


FIGURE 8

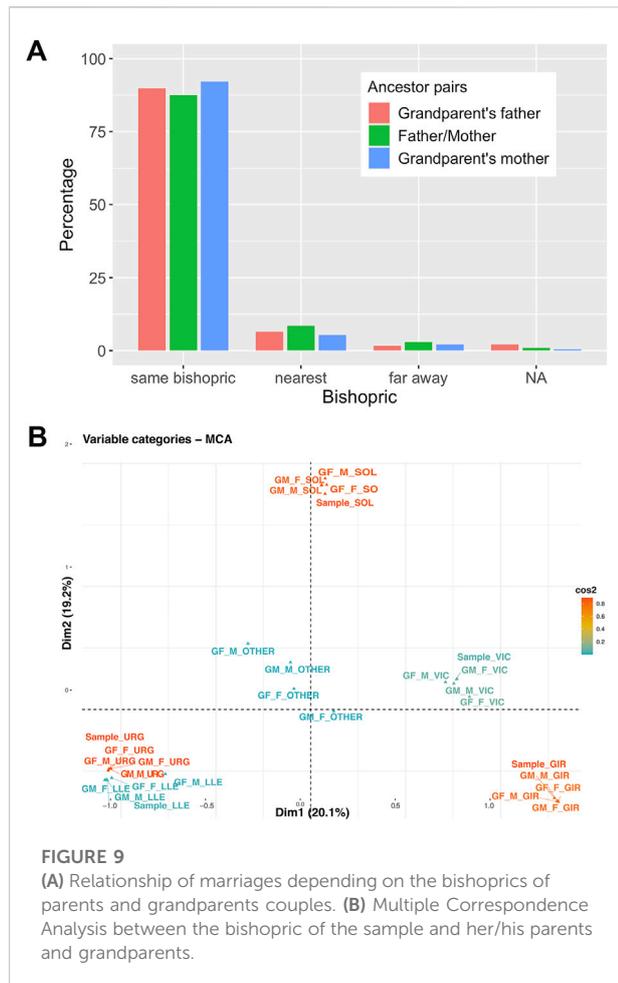
Geographic location of the individuals with regards to the mating relationship of grandparents. Each *propositus* is linked to their grandparents generating a complex network centered on bishoprics. Red lines delimit the bishopric boundaries. Purple lines indicate the presence of orographic events preventing direct routes of contact between regions.

boundaries. We can also observe that grandparent and parent couples come from the same bishopric (over 80% of couples) or to a nearest bishopric (near 10% of couples) (Figure 9A). The overall relationship between bishopric and mating was summarized by a multiple correspondence analysis (MCA) using the bishopric of each grandparent, parent and the considered sample. The relationship defined by the bishopric origin of the grandparents and samples is summarized by a triangle in the first two dimensions (explaining 39.3% of total inertia; Figure 9B) of the MCA. The first vertex contained relationships from the bishoprics of Urgell and Lleida. The second vertex was determined by the relationship between members from the bishopric of Solsona and the third one by the bishopric of Girona. The grandparents and current samples from the bishopric of Vic fall within the axis defined by Solsona and Girona, indicating a higher number of bishopric-admixed grandparent couples with the bishoprics of Solsona and Girona (Figure 9B). The center of the triangle depicted relationships where at least one of the grandparents is out of the bishoprics. Overall, the observed mating relationships agree with the geographic proximity between bishoprics (Lleida and Urgell, Vic between Solsona and Girona), as expected given the dependence between geography and these geohistorical

organizations and highlight the bishopric-based mating choice of the individuals from GENPIR.

The ecclesiastical division of the modern era responds to human spheres of multisecular relationships. For this reason, it is not surprising that the bundles of family relationships roughly coincide with the bishoprics: a central Pyrenean space –bishopric of Urgell—with abundant interactions (although it is capable of being subdivided into *comarques*); an eastern area (Girona); and a southern territory or central Catalonia (bishopric of Solsona, perhaps undifferentiated from that of Vic).

Traditionally, the main force used for explaining biases in mating choice in Spanish rural areas compared to urban areas has been geography (Fuster and Colantonio, 2004). Poor transportation facilities and limited communications - which in turn depend on orography- enhance the practice of consanguineous marriages and inbreeding in Spain. In particular, in small rural and dispersed localities, a limited number of suitable local partners conditioned marriage, and marrying a distant relative was a likely option when few other partners were available. This situation was even more common when the population started increasing due to medical advances, but mobility remained restricted (Gamella and Nez-Negrillo, 2019). Furthermore, in a sociocultural context, aunt-nephew and uncle-niece marriages could have been



practiced in rural areas in order to maintain or enlarge the family inheritance (Fuster and Colantonio, 2004). Similarly, the regional practice in Northern and Northwestern Spain from the 18th to the 20th century of a complex household system could have favored first cousin mating in these rural areas (Fuster and Colantonio, 2004). Overall, these socio-cultural processes could explain the overall patterns of isolation by distance that we detected in our different analyses. However, our study also supports that bishopric can act on shaping the genetic variation of rural areas. This observation can be interpreted in the socio-cultural context of the catalan Pyrenees. In this region, the old urban centers of cities and towns with episcopal seats are the first location of markets or fairs. Cities such as La Seu d'Urgell (bishopric of Urgell), Besalú (bishopric of Girona), Berga (bishopric of Solsona) or Vic (bishopric of Vic) have documented markets or fairs since XII and XIII centuries. In rural areas, markets and fairs act as a space for sociability promoting personal contacts between relatives and acquaintances. Social relationships were established or renewed, labor contracts were established, and marriages were arranged (Gallart, 2004).

Conclusion

Humans, as social animals, are subject to rules that emanate from power, economic, political or religious, which beyond allowing the proper functioning of social structures, also modulate interpersonal relationships. The role of the church has been prominent throughout history, acting as a central and determining factor of interpersonal relationships in past societies. In particular, the Catholic Church, with its rigid pyramidal structure, has left an imprint on societies that go beyond religious aspects. As our results show, the genetic structure of the rural populations of the catalan Pyrenees, in addition to the geographical factor, has also been modulated by administrative-social factors, which have established limits to the social relationships of individuals, being the limits established by the bishoprics one of the determining factors.

In this context, the bishoprics can be considered as the most representative territorial organization frameworks, historically very stable and cohesive, and necessarily well adapted to the strong geographical conditions that mark the Pyrenean environment.

A singular orography that has determined the establishment of administrative-religious limits, perpetuated unalterably for centuries, and the rigid social structure enforced by the Catholic Church over long periods of time, have shaped the genetic structure of the catalan Pyrenees population. It remains to be determined to what extent this structure may have generated interpopulation differences able to be reflected in differences on genetic susceptibility to diseases.

Data availability statement

The data presented in the study are deposited in the Figshare repository, accession numbers https://figshare.com/articles/dataset/Genotype_files/21404643 and https://figshare.com/articles/dataset/Code_for_figures_and_statistics/21406935.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of the University of Barcelona. The patients/participants provided their written informed consent to participate in this study.

Author contributions

JF, PM and OL contributed to the study conception and design. Material preparation, recruitment and data collection

were performed by AC, JFa, JFe, JG, JR and JS. Analysis and interpretation of the data was performed by JF, IM, ML, MMA and OL. Geohistorical contextualization was performed by MG, JB, JF and OL. The first draft of the manuscript was written by JF and OL and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the *Diputació of Lleida* grant to JF. OL and IM acknowledge the support from Spanish Ministry of Science and Innovation to the EMBL partnership, the Centro de Excelencia Severo Ochoa, CERCA Program/Generalitat de Catalunya, Spanish Ministry of Science and Innovation through the Instituto de Salud Carlos III, Generalitat de Catalunya through Departament de Salut and Departament d'Empresa i Coneixement, Co-financing with funds from the European Regional Development Fund by the Spanish Ministry of Science and Innovation corresponding to the Programa Operativo FEDER Plurirregional de España (POPE) 2014–2020 and by the Secretaria d'Universitats i Recerca, Departament d'Empresa i Coneixement of the Generalitat de Catalunya corresponding to the Programa Operatiu FEDER de Catalunya 2014–2020. OL gratefully acknowledges the financial support from Ministerio de Economía, Industria y Competitividad BFU 201568759-P and Generalitat de Catalunya (Government of Catalonia)—GRC 2017 SGR 937. IM gratefully acknowledges the financial support from the Government of Catalonia | Agència de Gestió d'Ajuts Universitaris i de Recerca (Agency for Management of University and Research Grants)—GRC 2014 SGR 615. MMA was supported by a FPU15/01251 from Ministerio de Ciencia, innovación y Universidades. MMA acknowledges the financial support from Precipita-FECYT (FBG 309307 project). The samples analyzed in this study are part of a much larger

References

- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi:10.1101/gr.094052.109
- Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi:10.1038/nature15393
- Bergström, A., McCarthy, S. A., Hui, R., Almarri, M. A., Ayub, Q., Danecek, P., et al. (2020). Insights into human genetic variation and population history from 929 diverse genomes. *Science* 367, eaay5012. doi:10.1126/science.aay5012
- Biagini, S. A., Solé-Morata, N., Matisoo-Smith, E., Zalloua, P., Comas, D., and Calafell, F. (2019). People from ibiza: An unexpected isolate in the western mediterranean. *Eur. J. Hum. Genet.* 27, 941–951. doi:10.1038/s41431-019-0361-1
- Brown, D. L. (2011). "Migration and rural population change: Comparative views in more developed nations," in *International handbook of rural demography international handbooks of population*. Editors C. KJ, 35–48. doi:10.1007/978-94-007-1842-5_4
- Browning, B. L., and Browning, S. R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* 194, 459–471. doi:10.1534/genetics.113.150029
- Burgueño, J. (1991). Del corregiment a la comarca de muntanya. Les divisions territorials al Pirineu (1). *Treballs de la Societat Catalana de Geografia* 28/29, 55–89. Available at: <http://hdl.handle.net/10459.1/47855>.
- Burgueño, J. (1992). Del Corregiment a la Comarca de muntanya. Les divisions territorials al Pirineu (11). *Treballs de la Societat Catalana de Geografia* 32, 73–101. Available at: <http://hdl.handle.net/10459.1/47854>.
- Bycroft, C., Fernandez-Rozadilla, C., Ruiz-Ponte, C., Quintela, I., Carracedo, Á., Donnelly, P., et al. (2019). Patterns of genetic differentiation and the footprints of historical migrations in the Iberian Peninsula. *Nat. Commun.* 10, 551. doi:10.1038/s41467-018-08272-w
- Calafell, F., and Bertranpetit, J. (1994). Mountains and genes: Population history of the Pyrenees. *Hum. Biol.* 66, 823–842.

sample whose collection was supported by the project CGL 2011–27866 of the Ministerio de Ciencia y Tecnología, the Fundació Moret i Marguà to PM.

Acknowledgments

We thank all the volunteers who donated DNA samples and who were involved in the collection of samples and genealogical data (Hospital Sant Bernabé, Berguedà; Fundació Sant Hospital, Alt Urgell; Hospital d'Olot i Comarcal, Garrotxa; Hospital de Campdevànol, Ripollès and Hospital Comarcal, Pallars).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.1100440/full#supplementary-material>

- Caye, K., Jay, F., Michel, O., and François, O. (2018). Fast inference of individual admixture coefficients using geographic data. *Ann. Appl. Stat.* 12, 586–608. doi:10.1214/17-aos1106
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4, 7. doi:10.1186/s13742-015-0047-8
- Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., et al. (2016). Next-generation genotype imputation service and methods. *Nat. Genet.* 48, 1284–1287. doi:10.1038/ng.3656
- Ferreiro, D., Núñez-Estévez, B., Canedo, M., Branco, C., and Arenas, M. (2021). Evaluating causes of current genetic gradients of modern humans of the Iberian Peninsula. *Genome Biol. Evol.* 13, evab071. doi:10.1093/gbe/evab071
- Flores-Bello, A., Bauduer, F., Salaberria, J., Oyharçabal, B., Calafell, F., Bertranpetit, J., et al. (2021). Genetic origins, singularity, and heterogeneity of Basques. *Curr. Biol.* 31, 2167–2177.e4. doi:10.1016/j.cub.2021.03.010
- Font, J. (1993). La formació històrica de la xarxa de carreteres a Catalunya (1761–1935). *Estudi General* 13, 127–152.
- Font-Porterías, N., Arauna, L. R., Poveda, A., Bianco, E., Rebato, E., Prata, M. J., et al. (2019). European Roma groups show complex West Eurasian admixture footprints and a common South Asian genetic origin. *Plos Genet.* 15, e1008417. doi:10.1371/journal.pgen.1008417
- Fuster, V., and Colantonio, S. (2004). Socioeconomic, demographic, and geographic variables affecting the diverse degrees of consanguineous marriages in Spain. *Hum. Biol.* 76, 1–14. doi:10.1353/hub.2004.0021
- Gallart, C. B. (2004). Fires i mercats: Factors de dinamisme econòmic i centres de sociabilitat (segles XI a XV). Rafael Dalmau Available at: <https://books.google.es/books?id=811IAAAACAAJ>.
- Gallart, D. M. (2002). El proceso de desertización demográfica de la montaña pirenaica en el largo plazo: Cataluña. *J. Depopulation Rural Dev. Stud.* 2, 81–99.
- Gamella, J. F., and Nez-Negrillo, A. M. (2019). The evolution of consanguineous marriages in the archdiocese of granada, Spain (1900/1979). *Hum. Biol.* 90, 97–114. doi:10.13110/humanbiology.90.2.02
- Gimenez, R. (2005). *Els Pirineus geografia de la circulació i política de transports*. Barcelona: Treballs de la Societat Catalana de Geografia, 87–107.
- Greenacre, M. (2007). *Correspondence analysis in practice*. Second Edition. New York, NY: Chapman Hall Crc Interdiscip Statistics Ser, 145–152. doi:10.1201/9781420011234.ch19
- Hellenthal, G., Busby, G. B. J., Band, G., Wilson, J. F., Capelli, C., Falush, D., et al. (2014). A genetic atlas of human admixture history. *Science* 343, 747–751. doi:10.1126/science.1243518
- Lao, O., Altena, E., Becker, C., Brauer, S., Kraaijenbrink, T., Oven, M. V., et al. (2013). Clinal distribution of human genomic diversity across The Netherlands despite archaeological evidence for genetic discontinuities in Dutch population history. *Investig. Genet.* 4, 9. doi:10.1186/2041-2223-4-9
- Lao, O., Lu, T. T., Nothnagel, M., Junge, O., Freitag-Wolf, S., Caliebe, A., et al. (2008). Correlation between genetic and geographic structure in Europe. *Curr. Biol. CB* 18, 1241–1248. doi:10.1016/j.cub.2008.07.049
- Lawson, D. J., Hellenthal, G., Myers, S., and Falush, D. (2012). Inference of population structure using dense haplotype data. *Plos Genet.* 8, e1002453. doi:10.1371/journal.pgen.1002453
- Lé, S., Josse, J., and Husson, F. (2008). FactoMineR : An R package for multivariate analysis. *J. Stat. Softw.* 25. doi:10.18637/jss.v025.i01
- Leeuw, J. D., and Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *J. Stat. Softw.* 31. doi:10.18637/jss.v031.i03
- Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T., et al. (2015). The fine-scale genetic structure of the British population. *Nature* 519, 309–314. doi:10.1038/nature14230
- Lichstein, J. W. (2007). Multiple regression on distance matrices: A multivariate spatial analysis tool. *Plant Ecol.* 188, 117–131. doi:10.1007/s11258-006-9126-3
- Loh, P.-R., Danecek, P., Palamara, P. F., Fuchsberger, C., Reshef, Y. A., Finucane, H. K., et al. (2016). Reference-based phasing using the haplotype reference consortium panel. *Nat. Genet.* 48, 1443–1448. doi:10.1038/ng.3679
- Maceda, I., Álvarez, M. M., Athanasiadis, G., Tonda, R., Camps, J., Beltran, S., et al. (2021). Fine-scale population structure in five rural populations from the Spanish Eastern Pyrenees using high-coverage whole-genome sequence data. *Eur. J. Hum. Genet.* 29, 1557–1565. doi:10.1038/s41431-021-00875-0
- Marcus, J. H., Posth, C., Ringbauer, H., Lai, L., Skeates, R., Sidore, C., et al. (2020). Genetic history from the middle neolithic to present on the mediterranean island of Sardinia. *Nat. Commun.* 11, 939. doi:10.1038/s41467-020-14523-6
- Martin, A., and Vaquer, J. (1995). “El poblament dels Pirineus a l’Holocè, del Mesolític a l’edat del Bronze,” in *Muntanyes i població. El Passat dels Pirineus des d’una perspectiva multidisciplinària*. Editors J. Bertranpetit and E. Vives (Andorra la Vella), 35–73. I Simposi de Poblament dels Pirineus.
- Martí-Henneberg, J. (2005). Empirical evidence of regional population concentration in Europe, 1870–2000. *Popul. Space Place* 11, 269–281. doi:10.1002/psp.373
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., et al. (2020). *vegan: Community ecology package*. R package version 2.5-7. Available at: <https://CRAN.R-project.org/package=vegan>.
- Olalde, I., Mallick, S., Patterson, N., Rohland, N., Villalba-Mouco, V., Silva, M., et al. (2019). The genomic history of the Iberian Peninsula over the past 8000 years. *Science* 363, 1230–1234. doi:10.1126/science.aav4040
- Oms, F. X., Martín, A., Esteve, X., Mestres, J., Morell, B., Subirà, M. E., et al. (2016). The neolithic in northeast iberia: Chronocultural phases and 14C. *Radiocarbon* 58, 291–309. doi:10.1017/rdc.2015.14
- Peres-Neto, P. R., and Jackson, D. A. (2001). How well do multivariate data sets match? The advantages of a procrustean superimposition approach over the Mantel test. *Oecologia* 129, 169–178. doi:10.1007/s004420100720
- Raveane, A., Aneli, S., Montinaro, F., Athanasiadis, G., Barlera, S., Birolo, G., et al. (2019). Population structure of modern-day Italians reveals patterns of ancient and archaic ancestries in Southern Europe. *Sci. Adv.* 5, eaaw3492. doi:10.1126/sciadv.aaw3492
- Riu, M. (1995). “El poblament dels Pirineus, segles VII–XIV,” in *Muntanyes i població. El Passat dels Pirineus des d’una perspectiva multidisciplinària*. Editors J. Bertranpetit and E. Vives (Andorra la Vella: I Simposi de Poblament dels Pirineus), 195–218.
- Scrucca, L., Fop, M., Murphy, T. B., and Raftery, A. E. (2016). Mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *R. J.* 8, 289–317. doi:10.32614/rj-2016-021
- Team Core R (2020). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing Available at: <http://www.R-project.org/>.
- Vidal, A. (2012). “Vissir3: Nuevas posibilidades de visualización e interacción con la cartografía del ICC,” in *VI jornadas de SIG libre* (Girona: Universitat de Girona). Available at: <https://dugi-doc.udg.edu/bitstream/handle/10256/41957/Art-Vissir3.pdf?sequence=1&isAllowed=y>.
- Xue, Y., Mezzavilla, M., Haber, M., McCarthy, S., Chen, Y., Narasimhan, V., et al. (2017). Enrichment of low-frequency functional variants revealed by whole-genome sequencing of multiple isolated European populations. *Nat. Commun.* 8, 15927. doi:10.1038/ncomms15927
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28, 3326–3328. doi:10.1093/bioinformatics/bts606