


# Evaluating Causes of Current Genetic Gradients of Modern Humans of the Iberian Peninsula

David Ferreiro<sup>1,2</sup>, Bernabé Núñez-Estévez<sup>1,2</sup>, Mateo Canedo<sup>1,2</sup>, Catarina Branco<sup>1,2</sup>, and Miguel Arenas <sup>1,2,\*</sup>

<sup>1</sup>CINBIO, Universidade de Vigo, Spain

<sup>2</sup>Universidade de Vigo, Departamento de Bioquímica, Xenética e Immunoloxía, Spain

\*Corresponding author: E-mail: marenas@uvigo.es.

Accepted: 6 April 2021

## Abstract

The history of modern humans in the Iberian Peninsula includes a variety of population arrivals sometimes presenting admixture with resident populations. Genetic data from current Iberian populations revealed an overall east–west genetic gradient that some authors interpreted as a direct consequence of the *Reconquista*, where Catholic Kingdoms expanded their territories toward the south while displacing Muslims. However, this interpretation has not been formally evaluated. Here, we present a qualitative analysis of the causes of the current genetic gradient observed in the Iberian Peninsula using extensive spatially explicit computer simulations based on a variety of evolutionary scenarios. Our results indicate that the Neolithic range expansion clearly produces the orientation of the observed genetic gradient. Concerning the *Reconquista* (including political borders among Catholic Kingdoms and regions with different languages), if modeled upon a previous Neolithic expansion, it effectively favored the orientation of the observed genetic gradient and shows local isolation of certain regions (i.e., Basques and Galicia). Despite additional evolutionary scenarios could be evaluated to more accurately decipher the causes of the Iberian genetic gradient, here we show that this gradient has a more complex explanation than that previously hypothesized.

**Key words:** genetic gradients, modern humans, spatially explicit simulations, Iberian Peninsula, range expansion.

## Significance

The spatial distribution of current genetic variants of modern humans throughout the Iberian Peninsula presents an overall east-west orientation that has been traditionally interpreted (although without a formal evaluation) as a direct consequence of the expansion of Catholic Kingdoms toward the south during the *Reconquista*. In order to test this hypothesis, we performed extensive computer simulations to qualitatively evaluate the influence of diverse evolutionary scenarios (including Paleolithic and Neolithic range expansions) on the genetic gradient. We found that the Neolithic range expansion can already cause the orientation of the observed genetic gradient and a subsequent *Reconquista* process could promote local isolation in certain regions. We conclude that the observed genetic gradient presents a more complex explanation than that previously proposed.

## Introduction

The Iberian Peninsula (IP) is a well-delimited geographic region that presents a vast human history. Briefly, the IP was inhabited by Neanderthals, who coexisted with modern human populations, before their extinction around the Middle Paleolithic (Galván et al. 2014). It is assumed that Paleolithic hunter–

gatherer populations reached the IP approximately 35,500 years ago (ya), according to the oldest archeological evidence found in the Altamira Caves (de las Heras et al. 2012), through a range expansion from the Middle East that crossed the Pyrenees Mountains (Garate et al. 2015). A few studies proposed that some small Paleolithic populations

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

from Northern Africa could also have reached the IP through the Strait of Gibraltar (Curat et al. 2010), although there is not yet enough archeological evidence supporting this scenario. Indeed, the IP was a refuge region for European Paleolithic populations during the Last Glacial Maximum (Achilli et al. 2004). Later, Neolithic populations that expanded from the Middle East arrived to the IP about 6,000–5,500 ya (Broushaki et al. 2016). It is not totally clear if Neolithic populations arrived to the IP only through the Pyrenees Mountains or if there were also maritime arrivals (Linstädter et al. 2012). Posterior relevant expansions in the IP can be those made by the Roman Empire (~2,200 ya, although this expansion presented a very small population size compared with the indigenous population, McEvedy and Jones 1978) and various Germanic tribes. Later, approximately 1,300 ya, the Muslim expansion rapidly covered the IP (excepting northern regions and most of the Pyrenees) and remained in the IP during near 800 years with some genetic admixture (Adams et al. 2008). Next, around the middle of the eighth century, the Catholic Kingdoms conquered the IP from the north toward the south in a historical period usually called as the *Reconquista* (supplementary fig. S1 and Supplementary Material, Supplementary Material online). As a consequence of this complex admixture of cultures (Adams et al. 2008; Botigue et al. 2013), the IP is one of the European regions with the greatest genetic diversity (Wang et al. 2012).

The current genetic gradient along the IP describes an overall (with small deviations) east-west (E-W) orientation (Romón et al. 2016; Bycroft et al. 2019; Pimenta et al. 2019) that is especially clear in the northern IP. Some studies proposed that the direct cause of this observed genetic gradient is the *Reconquista* (Bycroft et al. 2019; Pimenta et al. 2019) due to the north to south (N-S) expansion of Catholic Kingdoms (i.e., N-S gene flow) and some isolation between Catholic Kingdoms (supplementary fig. S1, Supplementary Material online). However, this hypothesis has not yet been formally evaluated and it is tempting to presume that other evolutionary processes could also lead to the observed genetic gradient. For example, the Paleolithic and Neolithic range expansions in the IP (occurring from the east to the west) could result in a genetic gradient with E-W orientation due to serial founder effects (SFEs) according to the direction of those expansions (Barbujani et al. 1995). Indeed, the long E-W distance of the IP could also favor the cited gradient through isolation by distance (IBD; Kanitz et al. 2018). In general, we believe that additional evolutionary processes could produce the E-W orientation of the currently observed genetic gradient of modern humans in the IP. Hence, in this study, we evaluated diverse evolutionary processes (we selected those becoming more influential to shape genetic gradients considering previous studies) using extensive spatially explicit computer simulations that mimic such evolutionary scenarios (e.g., Francois et al. 2010; Petkova et al. 2016; Branco and Arenas 2018). We found that several evolutionary scenarios, including (but not

only) those proposed in previous studies, can produce the overall E-W genetic gradient observed in the IP.

## Results

### Genetic Gradient of the IP Estimated from Real Data

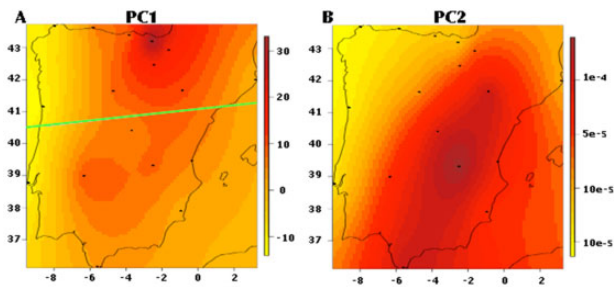
The genetic gradient that we obtained from the real data showed an overall E-W orientation (from the Basque region toward the east and toward the west; fig. 1A), especially in the north, and it also highlighted the well-known specific genetic profile of Basques (fig. 1A). Indeed, the PC2 map showed some genetically isolated regions, including Galicia and Castile-La Mancha (fig. 1B). We believe that the well-known genetic differentiation of the Basques population (Behar et al. 2012; Günther et al. 2015) is not reflected in the PC2 map because its genetic signature is already extracted by the PC1. Next, in order to evaluate if the Basques population and populations of its neighboring regions (Navarra and La Rioja) are crucial for the current genetic gradient observed in the IP, we performed an analysis excluding samples of those populations from the real data set. We found that excluding those populations produces genetic gradients with E-W orientation (supplementary fig. S4 and Supplementary Material, Supplementary Material online) that are similar to the gradient derived from the real data with all the populations (fig. 1A).

### Influence of the Sample Size on the Simulated Genetic Gradients

We obtained similar genetic gradients for a variable number of individuals and sequence length (supplementary figs. S5 and S6, and Supplementary Material, Supplementary Material online). As expected, we also found that increasing the sample size (number of individuals and sequence length) reduces the variability among genetic gradients from data simulated under the same evolutionary scenario (supplementary figs. S5 and S6, Supplementary Material online).

### A Neolithic Range Expansion Produces a Genetic Gradient with Orientation Similar to the Real Genetic Gradient

We found that the pure Paleolithic range expansion produces a genetic gradient with a NW-SE orientation (fig. 2A). Indeed, the PC2 maps derived from this scenario showed genetic isolation of Basques and southeast IP (supplementary fig. S7A and Supplementary Material, Supplementary Material online). Next, the scenarios of admixture between Paleolithic and Neolithic populations showed that increasing the Neolithic contribution leads to a genetic gradient with an overall E-W orientation (fig. 2). In these scenarios of admixture, the PC2 maps generally exhibited genetic isolation in the east and west coasts of the IP (supplementary figs. S7B and C). Indeed, a pure Neolithic expansion produced a genetic gradient with a clear E-W orientation (fig. 2D). A similar orientation



**FIG. 1.**—Estimated genetic gradient of the IP based on real data. The genetic gradient was obtained by applying PCA to the real data from Pimenta et al. (2019) and Bycroft et al. (2019). The vertical and horizontal axes indicate the geographic coordinates of the map and the black points throughout the map mean sample locations (supplementary fig. S2, Supplementary Material online). (A) Genetic gradient obtained from the PC1. The green line results from connecting the geographical centroids of regions with positive and negative PC1 values. Note the relevant influence of Basques on the gradient. (B) PC2 map that is traditionally considered to identify genetically isolated regions (Novembre and Stephens 2008; Francois et al. 2010), such as Galicia (clear yellow) and Castile-La Mancha (dark orange).

(that fitted very well with the orientation of the gradient derived from the real data) was found in simulations with origin of the Neolithic range expansion in regions of the Mediterranean coast (supplementary fig. S8 and Supplementary Material, Supplementary Material online). The PC2 map also exhibited some genetic isolation in the east and west coasts of the IP (supplementary fig. S7D, Supplementary Material online).

When migration through long-distance dispersal (LDD, see Materials and Methods) was considered in Paleolithic and Neolithic populations, we obtained a huge variability among the genetic gradients of the 100 data sets simulated under the corresponding scenario (supplementary fig. S9 and Supplementary Material, Supplementary Material online).

#### Geographic Isolation Caused by Catholic Kingdoms and Different Languages Maintains the Orientation of the Neolithic Genetic Gradient to Be Similar to the Real Genetic Gradient

Our results showed that all the evolutionary scenarios considering a Neolithic expansion with geographic isolation, regardless the cause of that isolation (i.e., Catholic Kingdoms or linguistic differences), produce a genetic gradient with an overall E-W orientation (fig. 3). Concerning genetically isolated regions indicated in PC2 maps, the evolutionary scenario based on the geographic delimitation of Catholic Kingdoms indicated some genetic isolation in the regions of Galicia and Valencia (supplementary fig. S10A and Supplementary Material, Supplementary Material online). As expected, the previous scenario additionally accounting for geographic isolation of the Basque region produced isolation of the own

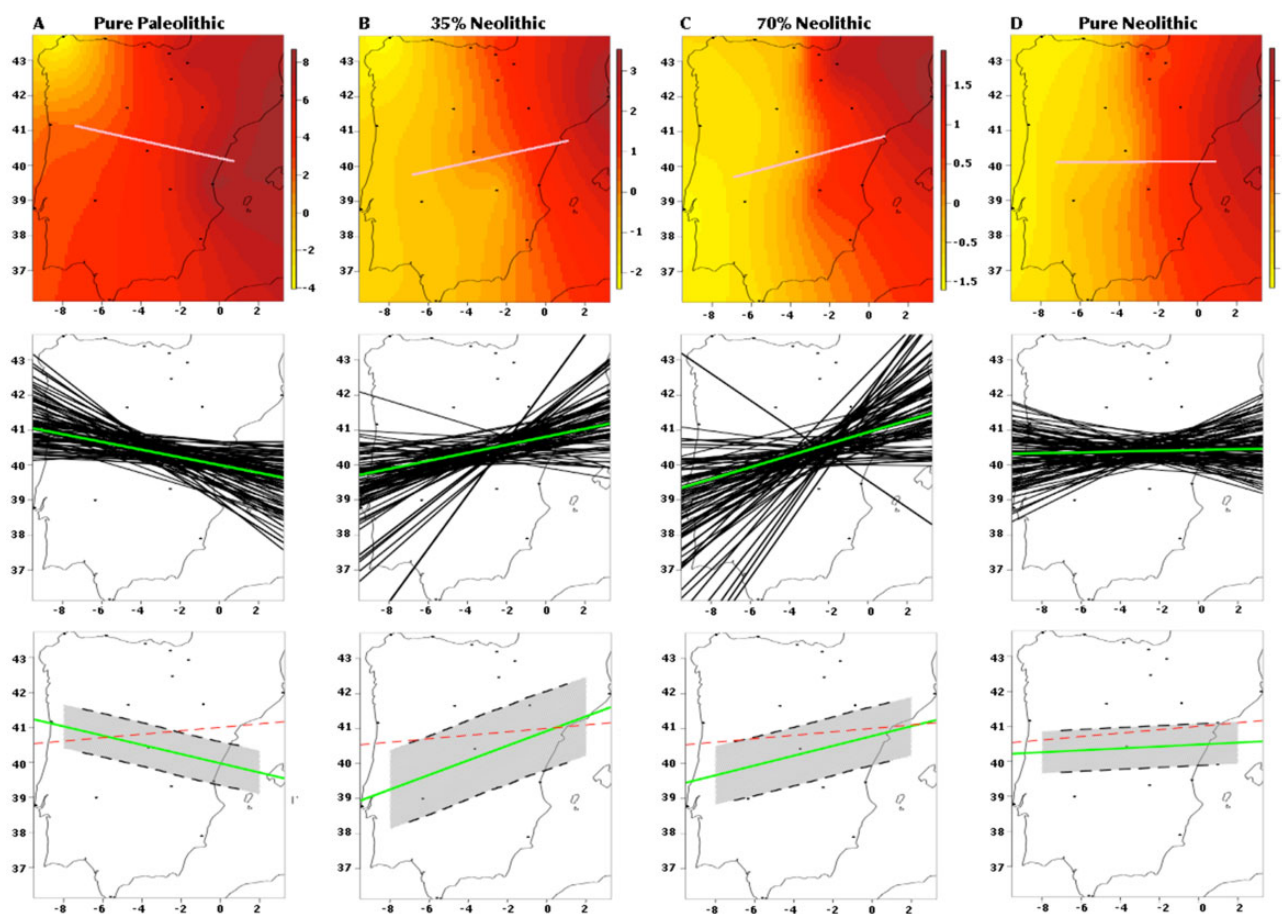
Basque region (supplementary fig. S10B) and, if the different languages present in the IP are additionally considered in the evolutionary scenario, then we found genetically isolated regions in the northwest and southeast of the IP (supplementary fig. S10C, Supplementary Material online).

Similar to other evolutionary scenarios considering a proportion of migration through LDD, we found that geographic isolation based on the Catholic Kingdoms and linguistic differences with the presence of LDD leads to a strong variability of genetic gradients among simulations of the scenario, thus disabling the identification of a representative genetic gradient for such a scenario (supplementary fig. S11 and Supplementary Material, Supplementary Material online).

## Discussion

The identification of the causes of current genetic gradients of modern humans remains a relevant topic of debate (Novembre and Stephens 2010; Branco and Arenas 2018). Concerning the particular case of the IP, several studies found that the current genetic gradient presents an overall E-W orientation that is particularly clear in the north of the IP (Romòn et al. 2016; Bycroft et al. 2019; Pimenta et al. 2019) and some of those studies directly attributed this gradient to the *Reconquista* considering the north to south expansion of the Catholic Kingdoms (Bycroft et al. 2019; Pimenta et al. 2019). However, this attribution has not been properly evaluated. Note that this interpretation acknowledges historical records but also, from a genetic perspective, one should consider that the *Reconquista* involved a small number of generations. Taking advantage from previous studies that evaluated the influence of diverse evolutionary scenarios on current genetic gradients of modern humans from different regions of the world using spatially explicit computer simulations of genetic data and principal component analysis (PCA; Francois et al. 2010; Arenas et al. 2013; Petkova et al. 2016; Branco and Arenas 2018; Branco et al. 2018, 2020) here we investigated some relevant evolutionary scenarios of the IP that could have affected the current genetic gradient of this region.

Our results indicated that, of course, the geographic delimitations among Catholic Kingdoms that expanded in the *Reconquista*, as well as regions of the IP presenting different languages, can favor the presence of an E-W genetic gradient. As expected, we found that a geographic isolation of the Basque region should have occurred to explain the observed genetic isolation of that region. These findings partially support the hypothesis of the previously cited studies (Bycroft et al. 2019; Pimenta et al. 2019) that interpreted the observed E-W genetic gradient as only consequence of the *Reconquista*. Next, we also found that other evolutionary scenarios can cause the observed overall E-W genetic gradient. The pure Paleolithic range expansion produced a genetic gradient with SE-NW orientation that does not fit well with the

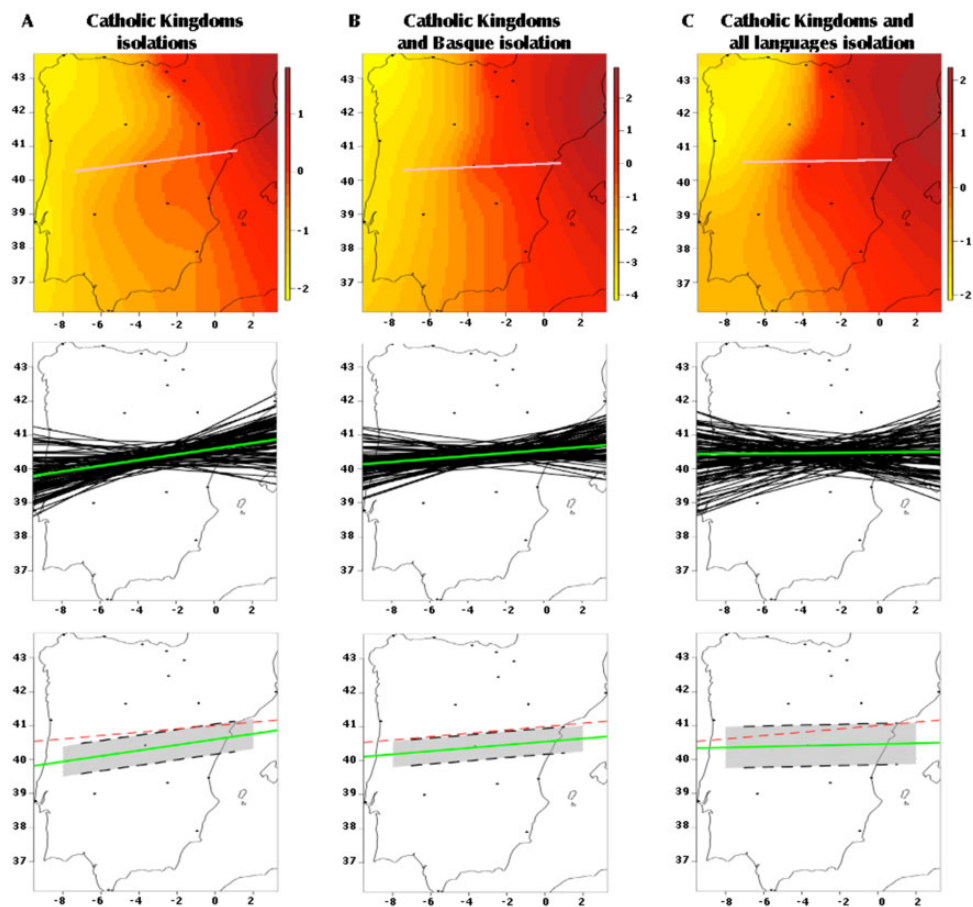


**Fig. 2.**—Genetic gradients caused by Paleolithic and Neolithic range expansions with and without population admixture among them. The plots show the genetic gradient estimated from a pure Paleolithic population (A) and different amounts of population admixture [35% (B), 70% (C), and 100% (D)] between Paleolithic and Neolithic populations. In the first row, we provide the genetic gradient (PC1 map) from an illustrative simulation of the corresponding scenario. The line represents the orientation of the genetic gradient from connecting the geographical centroids of regions with positive and negative PC1 values. In the second row, we present the orientation of 100 genetic gradients (100 simulated data sets, black lines) and their median (green line). At the bottom, we present the orientation of 100 genetic gradients (100 simulated data sets) by a green line that represents the linear regression from 2,000 random points sampled from the 100 genetic gradients under a 90% of confidence (gray area delimited by two dashed lines). The red dashed line represents the orientation of the gradient produced by the real data (fig. 1). The vertical and horizontal axes indicate the geographic coordinates of the map and the black points throughout the map indicate the sample locations ([supplementary fig. S2, Supplementary Material](#) online).

observed E-W gradient. Nevertheless, if the Neolithic range expansion is considered (especially if its contribution to the genetic pool is large and if it expanded from locations in the Mediterranean coast) the derived genetic gradient presents an orientation similar to the orientation of the genetic gradient obtained from the real data. This genetic gradient caused by the Neolithic expansion does not include local geographic regions presenting prominent genetic isolation (i.e., Basques; Günther et al. 2015; Romòn et al. 2016; Pimenta et al. 2019) as observed in the real genetic gradient (although PC2 maps from the simulated Neolithic expansion suggest a few regions, such as coasts, with some genetic isolation). This can be expected since the relevant isolation of Basques is independent of the Neolithic expansion (Behar et al. 2012) and thus, one needs to model the isolation of Basques (or other

geographic regions, for example by considering political borders imposed by Catholic Kingdoms or geographic delimitations of languages) to obtain simulated genetic gradients with the isolation of those specific regions. This intuitive reasoning, that still must be validated as we performed in this study, is in agreement with our results. We found that accounting for diverse (political and linguistic) migration barriers allows for obtaining genetically isolated regions more similar to those present in the genetic gradient derived from the real data. For example, accounting for barriers caused by the Catholic Kingdoms and different languages resulted in an isolated migration corridor along the Atlantic coast (current Galicia, Portugal, which are regions with remarkable linguistic relationships; [supplementary fig. S10C, Supplementary Material](#) online) that can fit with the real genetic gradient (fig. 1 and





**FIG. 3.**—Genetic gradients caused by geographic isolation promoted by the Catholic Kingdoms and languages. The plots show the genetic gradient of evolutionary scenarios that include the Neolithic range expansion and geographic isolation between Catholic Kingdoms (A), geographic isolation between Catholic Kingdoms and geographic isolation of the region of the Basque language (B) geographic isolation between Catholic Kingdoms and geographic isolation of regions with all the different languages (Spanish, Basque, Catalan, and Galician) (C). In the first row, we provide the genetic gradient (PC1 map) from an illustrative simulation of the corresponding scenario. The line represents the orientation of the genetic gradient from connecting the geographical centroids of regions with positive and negative PC1 values. In the second row, we present the orientation of 100 genetic gradients (100 simulated data sets, black lines) and their median (green line). At the bottom, we present the orientation of 100 genetic gradients (100 simulated data sets) by a green line that represents the linear regression from 2,000 random points sampled from the 100 genetic gradients under a 90% of confidence (gray area delimited by two dashed lines). The red dashed line represents the orientation of the gradient produced by the real data (fig. 1). The vertical and horizontal axes indicate the geographic coordinates of the map and the black points throughout the map indicate the sample locations (supplementary fig. S2, Supplementary Material online).

Romòn et al. 2016; Bycroft et al. 2019; Pimenta et al. 2019). Interestingly, we also found that the isolation of geographic regions by Catholic Kingdoms and languages maintained the E-W genetic gradient caused by the Neolithic expansion, thus fitting with the orientation of the real genetic gradient. As suggested by other authors (Romòn et al. 2016; Bycroft et al. 2019; Pimenta et al. 2019), we believe that this can be expected because the barriers between those regions are mainly oriented from the north toward the south (fig. 1) inducing isolation between east and west regions.

Some migration through LDD could have occurred in the IP (Alves et al. 2016). However, we found that data simulated under LDD result in genetic gradients that were highly variable

among simulations of the studied evolutionary scenario. In other words, we could not identify a genetic gradient representative of an evolutionary scenario of migration with LDD. However, this could be expected because the LDD model implemented in the simulator considers LDD events toward every direction and the parameters of this model are only designed to accommodate the length of the migration events (Ray and Excoffier 2010) (Supplementary Material, Supplementary Material online). In this concern, we believe that the LDD model implemented in the simulator should be improved by acquiring additional parameters related with the orientation of LDD events, although we also realize that over-parameterization should be avoided. In practice, our results

did not allow making conclusions about the influence of LDD on the genetic gradient of the IP and further works are required in this concern.

An interesting issue is how the Neolithic range expansion can produce a genetic gradient with E-W orientation (observed in the real genetic gradient). Since the orientation of this genetic gradient overall follows the direction of the range expansion, we believe that SFEs (Barbujani et al. 1995) could be an important genetic process behind the orientation of this gradient. Indeed, IBD (Wright 1943), caused by the long E-W distance of the IP, could also favor this E-W orientation of the genetic gradient.

Altogether, our results indicate that the *Reconquista* (including the north to south expansion of Catholic Kingdoms and the presence of geographic regions with different languages) could have favored the maintenance of the E-W genetic gradient and genetically isolated regions currently observed in the IP. However, we also found that the Neolithic expansion can clearly produce the orientation of the observed genetic gradient. In all, despite further works evaluating the impact of additional evolutionary scenarios on genetic gradients of the IP would be useful to better identify the reasons of the currently observed genetic gradient, our findings suggest that the genetic gradient observed in the IP can have more complex causes than those previously proposed.

## Materials and Methods

We evaluated the genetic gradients caused by diverse evolutionary scenarios (presented below) using extensive spatially explicit computer simulations of genetic data. Next, we estimated genetic gradients from the simulated and real data using PCA. Finally, we qualitatively compared the genetic gradient obtained from the real data with the genetic gradients obtained from data simulated under each evolutionary scenario. The applied methodologies are described below and are based on a variety of previous studies (Paschou et al. 2007; Francois et al. 2010; Branco and Arenas 2018; Branco et al. 2018).

### Estimation of the Genetic Gradient from Real Data

Several studies have already analyzed the genetic gradient of modern humans in the IP using diverse approaches and obtaining similar results (Bycroft et al. 2019; Pimenta et al. 2019); however, here we estimated the genetic gradient of real data (details below) using PCA as a sanity check to later perform proper comparisons between the genetic gradients derived from real and simulated data. Despite some progresses made in the field (Petkova et al. 2016), PCA still constitutes a powerful and straightforward approach to estimate genetic gradients (Paschou et al. 2007). In this concern, one can highlight the Cavalli-Sforza et al. studies (Cavalli-Sforza

et al. 1993) applying PCA to estimate genetic gradients of modern humans in diverse regions of the world. Here, we performed PCA with the *prcomp* function of the *R* statistical package (R Core Team 2018) in agreement with other studies (Arenas et al. 2013; Branco et al. 2018, 2020; Novembre and Stephens 2008). Note that the first principal component (PC1 map) provides the genetic gradient (Novembre and Stephens 2008) and the PC2 map (which presents less information than PC1) has been traditionally used to identify genetically isolated regions (e.g., Novembre and Stephens 2008; Francois et al. 2010; Arenas et al. 2013). The studied real data consisted of the complete autosomal genome (173,760 informative single nucleotide polymorphisms [SNPs]) of 1,492 individuals collected from 17 populations (the same locations that we use in the simulations, see [supplementary fig. S2](#) and [Supplementary Material, Supplementary Material](#) online, and also see next subsection) provided in previous genetic studies of the IP (Bycroft et al. 2019; Pimenta et al. 2019). In order to explore the influence of the Basque Country and its neighboring regions on the PC maps, we also analyzed the same real data but excluding the samples of the populations of 1) the Basque country; 2) the Basque Country and Navarra; and 3) the Basque Country, Navarra, and La Rioja.

Since simulating this large number of individuals and SNPs is computationally intractable, we investigated whether the sample size (number of individuals and sequence length) could affect the genetic gradients. In particular, we explored the influence of different sequence lengths (500, 1,000, 3,000, and 5,000 SNPs) and number of individuals sampled from each population (20, 35, 50, and 100 individuals) by computer simulations (details are shown in the following subsection) and we found that these parameters (at these levels) do not qualitatively affect genetic gradients (see Results).

### Spatially Explicit Computer Simulations and Study Evolutionary Scenarios

Despite spatially explicit computer simulations are computationally slow compared with other simulation approaches (i.e., the standard coalescent), these simulations are highly realistic by explicitly taking into account evolutionary, geographic, and environmental information (Dunning et al. 1995; Epperson et al. 2010; Benguigui and Arenas 2014). We performed spatially explicit computer simulations of genetic data with the framework *SPLATCHE3* (Currat et al. 2019), a simulator that has been widely used for human evolutionary studies (e.g., Currat and Excoffier 2004, 2005; Ray 2005; Francois et al. 2010; Alves et al. 2016; Pimenta et al. 2017; Branco and Arenas 2018; Quilodr an et al. 2020). Basically, this framework consists of the following three main steps: 1) A forward-in-time simulation (from the past to the present) of the entire population upon a user-specified landscape, demographic and migration model. The applied demographic and migration models are presented in the [Supplementary Material](#),

Supplementary Material online, and can also be found in the documentation of the framework *SPLATCHE3*. 2) A coalescent (backward in time) simulation of a user-specified sample of individuals (usually belonging to the present). 3) Assignment of a genetic sequence (arbitrarily simulated) to the most recent common ancestor of the previously simulated coalescent tree followed by a forward-in-time simulation of molecular evolution of that sequence along the branches of the coalescent tree to obtain a multiple sequence alignment corresponding to the specified sample. Therefore, for example, two geographically distant individuals presenting a low migration rate could require a long time to coalesce (backward in time), leading to long branches, accumulation of many mutation events and, finally, long genetic distance between the sequences of these individuals. Further details about the steps implemented in *SPLATCHE* can be found in Currat et al. (2004, 2019) and Ray et al. (2010).

Using a Geographic Information System, we obtained a two-dimensional (2D) map of 775 demes, where each deme corresponds to  $50 \times 50 \text{ km}^2$  (supplementary fig. S2, Supplementary Material online). We selected 17 sample locations well distributed throughout the IP (supplementary fig. S2, Supplementary Material online) following the locations of real data analyzed in previous studies (Bycroft et al. 2019; Pimenta et al. 2019), which is convenient to perform proper comparisons between estimates from simulated and real data. We simulated 20 individuals per location and each individual included 500 SNPs under a minor allele frequency of 0.03 to avoid rare genetic variants (Arenas et al. 2013). Therefore, each simulated data set consisted of  $17 \times 20 = 340$  sequences with 500 SNPs for each sequence. As indicated in the previous subsection, we also explored if other sample sizes (numbers of individuals per location and SNPs) could affect genetic gradients, which was not the case (see Results). Individuals migrated among demes according to a 2D stepping-stone migration model (Kimura and Weiss 1964) (supplementary fig. S3 and Supplementary Material, Supplementary Material online), although we also studied some scenarios where a proportion of individuals migrated under an LDD model (Ray et al. 2010), following other studies (Branco et al. 2018, 2020; Illera et al. 2019; Arenas et al. 2020). We assumed that Paleolithic and Neolithic populations expanded throughout the IP from the Eastern Pyrenees (Garate et al. 2015) (supplementary fig. S3, Supplementary Material online), but also we explored Neolithic range expansions with origin in locations at the Mediterranean coast (Catalonia and Valencia) following Isern et al. (2014). Indeed, according to a previous study (Alves et al. 2016), we considered that demes with mean height above 2,000 m cannot allocate individuals by setting their carrying capacity to 0. Additional population genetics parameters used for the simulations followed previous studies and are described in the Supplementary Material, Supplementary Material online.

We studied the following evolutionary scenarios (classified in supplementary table S1 and Supplementary Material, Supplementary Material online) that could have a major impact on genetic gradients of the IP considering previous findings in other regions of the world (Francois et al. 2010; Branco et al. 2018; Branco and Arenas 2018).

1. *Paleolithic and Neolithic range expansions*. We simulated scenarios considering as follows: a) a pure Paleolithic range expansion (without posterior Neolithic range expansion), b) a pure Neolithic range expansion (without Paleolithic populations), and c) Paleolithic and Neolithic range expansions with admixture between both populations. In the later case, we investigated several levels of population admixture since studies disagree concerning the level of admixture that could have occurred between these populations (e.g., Chikhi et al. 2002; Kanitz et al. 2018). In particular, we explored interbreeding rates of 0, 0.04, 0.068, and 0.1 that correspond to 100% (note that this is equal to a pure Neolithic range expansion), 70%, 50%, and 35% of Neolithic contribution to the final genetic pool, respectively.
2. *Paleolithic and Neolithic range expansions with LDD*. We simulated scenarios of Paleolithic and/or Neolithic range expansions with a proportion of individuals (4% according to previous estimates for modern humans, Alves et al. 2016; Arenas et al. 2020) migrating with LDD. Further details about the parameters of the LDD model (Ray and Excoffier 2010) are described in the Supplementary Material, Supplementary Material online.
3. *The Reconquista, Catholic Kingdoms and geographic distribution of languages*. The simulation of a range expansion mimicking a pure *Reconquista* (without previous populations in the Iberian Peninsula) cannot be performed because the number of generations of the *Reconquista* is too small for a realistic demographic settlement of the entire IP and the simulation of populations that only exist since the *Reconquista* is too unrealistic. Therefore, these evolutionary scenarios consisted of a Neolithic range expansion (previously described) followed by the next situations: a) geographic isolation among the Catholic Kingdoms considering the timeline of the *Reconquista* (supplementary fig. S1, Supplementary Material online), b) geographic isolation among the Catholic Kingdoms considering the timeline of the *Reconquista* (previous situation) and geographic isolation of the Basque region considering its different languages and ancestry according to (Behar et al. 2012; Günther et al. 2015), and c) geographic isolation among Catholic Kingdoms considering the *Reconquista* and geographic isolation of the Basque region (previous situation) and, geographic isolation among all the other regions presenting different languages (supplementary fig. S1, Supplementary Material online). The isolation was simulated by setting the migration rate to 0 or 0.1 (half of the migration rate present in the rest of demes) in the demes of

borders delimiting geographic regions (Mona et al. 2014) (further information about political and linguistic delimitations and their timeline is included in the [Supplementary Material, Supplementary Material](#) online).

We performed a total of 100 computer simulations for each evolutionary scenario.

### Estimation of Genetic Gradients from Simulated Data

The estimation of genetic gradients from simulated data was performed with PCA and following the same procedure to estimate the genetic gradient from the real data (previously shown). Since for each evolutionary scenario, we obtained 100 simulated data sets (resulting in 100 genetic gradients) and in order to show the 100 genetic gradients into a single map (Arenas et al. 2013; Branco et al. 2018, 2020), we showed the genetic gradient of every data set by a straight line with the orientation of the corresponding genetic gradient (i.e., fig. 1A). This line is built by connecting the geographical centroids of the regions on the map with positive and negative PC1 coordinates (Arenas et al. 2013; Branco et al. 2018, 2020). Next, for every evolutionary scenario, we obtained a median line from the 100 lines (based on their slopes and intercepts) that has been considered as an overall representative genetic gradient for the study evolutionary scenario (Arenas et al. 2013; Branco et al. 2018, 2020). We also included a multiple linear regression with 90% confidence interval obtained from 2,000 randomly sampled points from the PC1 lines (20 points from each gradient line) that can be used to compare roughly the orientation of genetic gradients.

### Supplementary Material

[Supplementary data](#) are available at *Genome Biology and Evolution* online.

### Acknowledgments

We thank the Supercomputing Center of Galicia (CESGA) for the use of the computer cluster. We also thank the Associate Editor and two anonymous reviewers for their insightful comments. Project supported by a 2019 Leonardo Grant for Researchers and Cultural Creators, BBVA Foundation. The BBVA Foundation accepts no responsibility for the opinions, statements, and contents included in the project and/or the results thereof, which are entirely the responsibility of the authors. This work was supported by the Ministerio de Economía y Competitividad of the Spanish Government (Grant No. RYC-2015-18241 to M.A.) and the Fundação para a Ciência e Tecnologia (FCT) of the Portuguese Government (Grant No. SFRH/BD/143607/2019 to C.B.). Funding for open access charge: Universidade de Vigo/CISUG.

### Author Contributions

M.A. conceived the original idea. M.A. and C.B. designed the analyses. D.F., B.N.E, M.C., and C.B. conducted the analyses. All authors contributed to produce the submitted manuscript.

### Data Availability

The real data are provided by published studies (Bycroft et al. 2019; Pimenta et al. 2019) and the simulated data, obtained with publicly available software (Currat et al. 2019), are stored and freely available from the repository Zenodo, <https://doi.org/10.5281/zenodo.4118237>.

### Literature Cited

- Achilli A, et al. 2004. The molecular dissection of mtDNA Haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet.* 75:910–918.
- Adams SM, et al. 2008. The genetic legacy of religious diversity and intolerance: paternal Lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *Am J Hum Genet.* 83:725–736.
- Alves I, et al. 2016. Long-distance dispersal shaped patterns of human genetic diversity in Eurasia. *Mol Biol Evol.* 33:946–958.
- Arenas M, et al. 2020. The early peopling of the Philippines based on mtDNA. *Sci Rep.* 10:4901.
- Arenas M, François O, Currat M, Ray N, Excoffier L. 2013. Influence of admixture and paleolithic range contractions on current European diversity gradients. *Mol Biol Evol.* 30:57–61.
- Barbujani G, Sokal RR, Oden NL. 1995. Indo-European origins: a computer-simulation test of five hypotheses. *Am J Phys Anthropol.* 96:109–132.
- Behar DM, et al. 2012. The Basque Paradigm: genetic evidence of a maternal continuity in the Franco-Cantabrian region since pre-Neolithic times. *Am J Hum Genet.* 90:486–493.
- Benguigui M, Arenas M. 2014. Spatial and temporal simulation of human evolution. methods, frameworks and applications. *Curr Genomics.* 15:245–255.
- Botigue LR, et al. 2013. Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. *Proc Natl Acad Sci U S A.* 110:11791–11796.
- Branco C, et al. 2018. Consequences of diverse evolutionary processes on American genetic gradients of modern humans. *Heredity* 121:548–556.
- Branco C, Arenas M. 2018. Selecting among alternative scenarios of human evolution by simulated genetic gradients. *Genes* 9:506.
- Branco C, Ray N, Currat M, Arenas M. 2020. Influence of Paleolithic range contraction, admixture and long-distance dispersal on genetic gradients of modern humans in Asia. *Mol Ecol.* 29:2150–2159.
- Broushaki F, et al. 2016. Early Neolithic genomes from the eastern fertile crescent. *Science.* 353:499–503.
- Bycroft C, et al. 2019. Patterns of genetic differentiation and the footprints of historical migrations in the Iberian Peninsula. *Nat Commun.* 10:551.
- Cavalli-Sforza L, Menozzi P, Piazza A. 1993. Demic expansion and human evolution. *Science* 259:639–646.
- Chikhi L, Nichols RA, Barbujani G, Beaumont MA. 2002. Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci U S A.* 99:11008–11013.
- Currat M, Arenas M, Quilodrán CS, Excoffier L, Ray N. 2019. SPLATCHE3: simulation of serial genetic data under spatially explicit evolutionary



- scenarios including long-distance dispersal. *Bioinformatics* 35:4480–4483.
- Currat M, Excoffier L. 2004. Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLoS Biol.* 2:e421.
- Currat M, Excoffier L. 2005. The effect of the Neolithic expansion on European molecular diversity. *Proc R Soc B Biol Sci.* 272:679–688.
- Currat M, Poloni E, Sanchez-Mazas A. 2010. Human genetic differentiation across the Strait of Gibraltar. *BMC Evol Biol.* 10:237.
- Currat M, Ray N, Excoffier L. 2004. Splatc: a program to simulate genetic diversity taking into account environmental heterogeneity. *Mol Ecol Notes.* 4:139–142.
- de las Heras C, Montes R, Lasheras J. 2012. Altamira: nivel gravetiense y cronología de su arte rupestre. In: de las Heras C, Lasheras J, Arrizabalaga A, de la Rasilla M, editors. *Pensando el Gravetiense: nuevos datos para la región cantábrica en su contexto peninsular y pirenaico*. España: Ministerio de Educación, Cultura y Deporte. p. 476–491.
- Dunning JB, et al. 1995. Spatially explicit population models: current forms and future uses. *Ecol Appl.* 5:3–11.
- Epperson BK, et al. 2010. Utility of computer simulations in landscape genetics. *Mol Ecol.* 19:3549–3564.
- Francois O, et al. 2010. Principal component analysis under population genetic models of range expansion and admixture. *Mol Biol Evol.* 27:1257–1268.
- Galván B, et al. 2014. New evidence of early Neanderthal disappearance in the Iberian Peninsula. *J Hum Evol.* 75:16–27.
- Garate D, Rivero O, Ruiz-Redondo A, Rios-Garaizar J. 2015. At the crossroad: a new approach to the upper Paleolithic art in the Western Pyrenees. *Quat Int.* 364:283–293.
- Günther T, et al. 2015. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proc Natl Acad Sci U S A.* 112:11917–11922.
- Illera JC, Arenas M, López-Sánchez CA, Obeso JR, Laiolo P. 2019. Gradual distance dispersal shapes the genetic structure in an Alpine Grasshopper. *Genes.* 10:590.
- Isern N, Fort J, Carvalho AF, Gibaja JF, Ibañez JJ. 2014. The Neolithic transition in the Iberian Peninsula: data analysis and modeling. *J Archaeol Method Theory.* 21:447–460.
- Kanitz R, Guillot EG, Antoniazza S, Neuenschwander S, Goudet J. 2018. Complex genetic patterns in human arise from a simple range-expansion model over continental landmasses. *PLoS One* 13:e0192460.
- Kimura M, Weiss G. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49:561–576.
- Linstädter J, Medved I, Solich M, Weniger G-C. 2012. Neolithisation process within the Alboran territory: models and possible African impact. *Quat Int.* 274:219–232.
- McEvedy C, Jones R. 1978. *Atlas of world population history*. Great Britain: Penguin Books.
- Mona S, Ray N, Arenas M, Excoffier L. 2014. Genetic consequences of habitat fragmentation during a range expansion. *Heredity* 112:291–299.
- Novembre J, Stephens M. 2008. Interpreting principal component analyses of spatial population genetic variation. *Nat Genet.* 40:646–649.
- Novembre J, Stephens M. 2010. Response to Cavalli-Sforza interview [Human Biology 82(3):245–266 (June 2010)]. *Hum Biol.* 82:469–470.
- Paschou P, et al. 2007. PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genet.* 3:e160.
- Petkova D, Novembre J, Stephens M. 2016. Visualizing spatial population structure with estimated effective migration surfaces. *Nat. Genet.* 48:94–100.
- Pimenta J, et al. 2019. Spatially explicit analysis reveals complex human genetic gradients in the Iberian Peninsula. *Sci Rep.* 9:7825.
- Pimenta J, Lopes AM, Comas D, Amorim A, Arenas M. 2017. Evaluating the Neolithic expansion at both shores of the Mediterranean Sea. *Mol Biol Evol.* 34:3232–3242.
- Quilodrán CS, Tsoupas A, Currat M. 2020. The spatial signature of introgression after a biological invasion with hybridization. *Front Ecol Evol.* 8:1–15.
- R Core Team. 2018. R version 3.5.0. R: A language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: <https://www.r-project.org/>.
- Ray N. 2005. Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations. *Genome Res.* 15:1161–1167.
- Ray N, Currat M, Foll M, Excoffier L. 2010. SPLATCHE2: a spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics* 26:2993–2994.
- Ray N, Excoffier L. 2010. A first step towards inferring levels of long-distance dispersal during past expansions. *Mol Ecol Resour.* 10:902–914.
- Romón I, et al. 2016. Mapping the HLA diversity of the Iberian Peninsula. *Hum Immunol.* 77:832–840.
- Wang C, Zöllner S, Rosenberg NA. 2012. A Quantitative comparison of the similarity between genes and geography in worldwide human populations. *PLoS Genet.* 8:e1002886.
- Wright S. 1943. Isolation by distance. *Genetics* 38:114.

**Associate editor:** Emilia Huerta-Sanchez