U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

PhD
3.º
CICLO
FCUP
I3S
IBE
2018

U.PORTO

Human Migrations and Population Dynamics in the Western Mediterranean: from the Neolithic to the present

João Carlos Pimenta

FC

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

i3S
INSTITUTO DE INVESTIGAÇÃO
E INOVAÇÃO EM SAÚDE
UNIVERSIDADE DO PORTO

IBE
CSIC UPF
INSTITUT de
BIOLOGIA
EVOLUTIVA

Human Migrations and Population Dynamics in the Western Mediterranean: from the Neolithic to the present

João Carlos Ramos de Azevedo Pimenta
Tese de Doutoramento apresentada à
Faculdade de Ciências da Universidade do Porto
Biologia
2018

Human Migrations and Population Dynamics in the Western Mediterranean: from the Neolithic to the present

João Carlos Ramos de Azevedo Pimenta

Doutoramento em Biologia
Departamento de Biologia
2018

Orientadores

David Comas, diretor do Departamento de Ciências Experimentais e da Saúde e Professor Catedrático da Universidade Pompeu Fabra, Barcelona.

António Amorim, Professor Catedrático, Faculdade de Ciências, Universidade do Porto.

Miguel Arenas, Investigador Principal, Departamento de Bioquímica, Genética e Imunologia, Universidade de Vigo

Esta dissertação foi apoiada financeiramente pela Fundação para a Ciência e a Tecnologia (FCT) através da atribuição de uma bolsa de Doutoramento (SFRH/BD/97200/2013).

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

Agradecimentos

Finda mais esta jornada, quero agradecer a todos aqueles que contribuíram de uma forma ou de outra para que esta tese ficasse completa.

Primeiro gostaria de agradecer aos meus orientadores David Comas, António Amorim e Miguel Arenas.

Ao David que me aceitou como seu doutorando no (já longínquo!!!) ano de 2012 e sem o qual esta Tese não seria possível. Obrigado pela sua dinâmica, motivação e rigor científico que foram essenciais para a qualidade do trabalho apresentado.

Ao professor Amorim que me recebeu no seu grupo ainda durante o meu mestrado. Obrigado pelo seu incansável apoio, tempo e paciência, por toda a ajuda e conselhos que foram sempre uma fonte de inspiração. A maior parte da minha vida científica e académica foram passados debaixo da sua 'asa' e o seu exemplo será sempre uma influência na forma como penso a ciência.

Ao Miguel, que entrou a meio desta aventura mas que foi decisivo para que este trabalho chegasse até ao fim. Obrigado pela tua dedicação e amizade, e por me teres ensinado tanto sobre tantos assuntos em tão pouco tempo. Sei que fui o teu primeiro aluno de doutoramento (dos muitos que se seguirão) e espero ter ido ao encontro das tuas expectativas!!!

Obrigado aos membros do grupo de Genética Populacional e Evolução (passados e presentes) pelo apoio, amizade e ótimo ambiente de trabalho. Em especial às minhas antigas orientadores, Sandra e Alexandra, ao Rui, Nádia, Verónica, Manuela, Luísa e professora Maria João. Gostaria também de agradecer ao João Alves pela sua disponibilidade na resolução de problemas informáticos nas fases iniciais deste trabalho.

Aos membros do grupo Human Genome Diversity do IBE, muito obrigado pela forma como me receberam e me integraram durante o tempo que passei em Barcelona. Particularmente à Lara Arauna por me ter ensinado a trabalhar com dados genome-wide que foram a base de todo o meu trabalho.

À minha família, em especial aos meus pais agradeço a ajuda e apoio em todos os momentos. Obrigado ao meu irmão que mesmo estando tão longe está sempre presente, à Nilmara por trazeres sempre a alegria brasileira e sertaneja, à Dona

Glória e Sr. Américo por me receberem sempre tão bem na sua casa, à Diana e Tiago pela vossa amizade e às minhas sobrinhas emprestadas que são uma constante fonte de alegria em todos os momentos.

E como para último se deve deixar o melhor. Obrigado Nádia! Pela tua força e companheirismo durante esta caminhada. Sem ti nada disto seria possível e é a ti que dedico esta tese!!!

INDEX

Resumo

O oeste Mediterrânico é uma região com uma complexa história demográfica desde a chegada do Neolítico há aproximadamente 7,500 anos. Esses eventos demográficos deixarão uma marca indelével nos genomas das populações humanas modernas. Durante a transição Neolítica, as populações humanas experimentaram desenvolvimentos culturais e tecnológicos que levaram à revolução agrícola e que remodelaram profundamente a estrutura genética, linguística e cultural das populações mundiais. Eventos mais recentes, como a expansão Muçulmana, durante o primeiro milénio da era moderna, também originou um profundo impacto demográfico nas populações do oeste Mediterrânico que potencialmente afetou a estrutura genética das populações.

Apesar da genética populacional e evolutiva das populações Neolíticas Europeias ter sido amplamente estudada, ainda pouco se sabe sobre a expansão Neolítica no Norte de África. Poder-se-á esperar que as diferentes condições ambientais e geológicas em ambas as margens do Mar Mediterrâneo tenham levado a expansões contrastantes. Por isso, na primeira parte desta Tese, comparámos a expansão Neolítica na Europa e no Norte de África, tendo também em conta possível migração entre as regiões pelo Estreito de Gibraltar, usando uma densa base de dados genéticas de populações espacialmente distribuídas ao longo do Norte de África e da Europa. Aplicámos a técnica de Aproximação Bayesiana Computacional baseada num extenso número de simulações espacialmente explícitas para selecionar entre cenários alternativos de migração através do Estreito de Gibraltar e para estimar parâmetros genéticos populacionais em ambas as expansões (Paleolítica e Neolítica). Os resultados sugerem que, apesar de tecnologicamente mais avançadas, as populações Neolíticas não expandiram mais rapidamente do que as populações Paleolíticas, o que poderá ser interpretado como sendo consequência de um estilo de vida mais sedentário. Detetamos migrações Neolíticas recíprocas entre a Península Ibérica e o Norte de África através do Estreito de Gibraltar. Contraintuitivamente, estabelecemos que as expansões Neolíticas estudadas apresentam níveis similares de capacidades de carga e migração, e que ocorreram a velocidades semelhantes, sugerindo um processo démico similar de substituição das populações caçadoras-recoletoras. De um modo geral, as expansões Neolíticas ao longo de ambas as margens Mediterrânicas não foram muito diferentes, talvez porque estas populações partilhavam habilidades técnicas e padrões de estilo de vida similares.

Na segunda parte desta Tese, estudámos as causas da estrutura genética e dinâmicas migratórias passadas na Península Ibérica, que é uma região bem delimitada geograficamente com um rica e complexa história humana. Avaliámos o fluxo e a estrutura genética ao longo da Península Ibérica com modelos espaciais explícitos baseados numa base de dados genómica georreferenciada composta por populações espacialmente bem distribuídas. Determinámos que a estrutura genética e dinâmica migratória dentro da Península Ibérica não está relacionada com o formato geográfico da região, mas em vez disso, encaixa nas atuais e antigas determinações territoriais politicas e linguísticas. Nas regiões a Norte, detetamos diversas barreiras relevantes ao fluxo genético que se assemelham aos reinos católicos durante a ocupação Muçulmana. Contrastantemente, nas regiões do sul, a *Reconquista* não consegue explicar inteiramente os padrões genéticos observados e portanto sugerimos que migrações posteriores à *Reconquista* desempenharam um papel importante na paisagem genética observada. De facto, os nossos resultados também sugerem a existência de migrações sexualmente enviesadas nas regiões Mediterrânica e do sul da Espanha. Adicionalmente, encontrámos sinais de diferenciação genética entre Portugal e Espanha, o quais associámos com a divisão política entre as duas regiões ocorrida há quase mil anos atrás.

De um modo geral, os estudos apresentados nesta Tese permitem-nos traçar um quadro abrangente da expansão Neolítica no norte de África e sul da Europa, ao mesmo tempo que fornece informações relevantes sobre a dinâmica migratória através Estreito de Gibraltar. Além disso, fornece uma imagem detalhada da estrutura genética espacial da Península Ibérica e as suas possíveis causas demográficas.

Palavras-Chave: expansão Neolítica, estrutura genética espacial, aproximação Bayesiana computacional, Estreito de Gibraltar, barreiras geográficas.

Abstract

The western Mediterranean is a region with a complex demographic history since the arrival of the Neolithic to the region approximately 7,500 years ago. These demographic events left an indelible imprint in genomes of modern human populations. During the Neolithic transition, human populations underwent cultural and technological developments that led to an agricultural revolution, which profoundly reshape the genetic structure, languages and culture of populations worldwide. More recent events, such as the Muslim expansion during the first millennia of the current era, also led to profound demographic impacts on populations from the western Mediterranean that potentially affected the genetic structure of populations.

Although the population genetics and evolution of European Neolithic populations have been extensively studied, little is known regarding the Neolithic expansion in North Africa with respect to Europe. One could expect that the different environmental and geological conditions at both shores of the Mediterranean Sea could have led to contrasting expansions. In the first part of this Thesis, we compared the Neolithic expansion in Europe and North Africa, accounting for possible migration between them through the Strait of Gibraltar, using a dense population genetic dataset spatially distributed along the North of Africa and Europe. We applied approximate Bayesian computation based on extensive spatially explicit computer simulations to select among alternative scenarios of migration through the Strait of Gibraltar and to estimate population genetics parameters in both expansions. Our results suggest that, despite being more technologically advanced, Neolithic populations did not expand faster than Palaeolithic populations, which could be interpreted as a consequence of a more sedentary lifestyle. We detected reciprocal Neolithic migration between the Iberian Peninsula and North Africa through the Strait of Gibraltar. Counterintuitively, we found that the studied Neolithic expansions presented similar levels of carrying capacity and migration, and occurred at comparable speeds, suggesting a similar demic process of substitution of hunter-gatherer populations. Altogether, the Neolithic expansion through both Mediterranean shores was not so different, perhaps because these populations shared similar technical abilities and lifestyle patterns.

In the second part of this Thesis, we studied the causes of genetic structure and past migratory dynamics in the Iberian Peninsula, which is a well-delimited geographic region with a rich and complex human history. We evaluated the gene flow and genetic structure throughout the Iberian Peninsula with spatially explicit modelling based on a georeferenced genome-wide dataset composed well spatially distributed populations. We found that the genetic structure and migratory dynamics within the Iberian Peninsula do not correlate with the geographical shape of the landscape and rather fits with political and linguistic territorial determinations. In northern regions we found several relevant barriers to gene flow that present a close resemblance to Catholic kingdoms during the Muslim occupation. By contrast, in southern regions the *Reconquista* could not entirely explain the observed genetic structure and there we suggest that migrations posterior to the *Reconquista* played an important role shaping the observed genetic landscape. Indeed, our findings also suggest the existence of sex-bias migrations in Mediterranean and southern regions of Spain. Additionally, we found signatures of genetic differentiation between Portugal and Spain that we associate with the political split of both regions that occurred almost a thousand years ago.

Altogether, the studies presented in this Thesis allow us to draw a comprehensive picture of the Neolithic process in North Africa and Europe, while providing relevant insight into the migratory dynamics across the Strait of Gibraltar. Additionally, it provides a detailed picture of the spatial genetic structure of the Iberian Peninsula and its possible migratory and demographic causes.

Keywords: Neolithic expansion, spatial genetic structure, approximate Bayesian computations, Strait of Gibraltar, geographic barriers.

Figure Index

Chapter 1.

Chapter 2.

Chapter 3.

Table Index

Chapter 2.

Chapter 3.

Abbreviations

2D	Two-dimensional
AMH	Anatomically modern human
ABC	Approximate Bayesian Computation
aDNA	Ancient DNA
AND	Andalusia
BAS	Basque
BCE	Before the current era
CE	Current era
CEU	Northern and Western Europeans currently in Utah
CNA	Central North Africa
CSP	Central Spain
ENA	Eastern North Africa
ESP	East Spain
EUR	European populations
FIN	Finland
F_{st}	Wright's fixation index
fullMIG	Reciprocal migration between both regions
GAC	Globular Amphorae culture
GAL	Galicia
GBR	Great-Britain
GIS	Geographic Information System
GLM	General linear model
G_{NEO}	Population growth rate Neolithic
G_{PALEO}	Population growth rate Palaeolithic
HG	Hunter-gatherer
IBE	Iberian populations
I_{rate}	Interbreeding rate
K_{NEOEUR}	Carrying capacity Neolithic in Europe
K_{NEONA}	Carrying capacity Neolithic in North Africa
K_{NEOPER}	Carrying capacity Neolithic in Rest of the World
K_{PALEO}	Carrying capacity Palaeolithic
LBK	Linearbandkeramic culture
LD	Linkage disequilibrium

LDD	Long-distance dispersal
LGM	Last Glacial Maximum
LIS	Lisbon
MAF	Minor allele frequency
M_{NEOEUR}	Migration rate Neolithic in Europe
M_{NEONA}	Migration rate Neolithic in North Africa
M_{NEOPER}	Migration rate Neolithic in Rest of the World
M_{nsSG}	Migration rate at Strait of Gibraltar - North to South route
M_{PALEO}	Migration rate Palaeolithic
MRCA	Most recent common ancestor
M_{snSG}	Migration rate at Strait of Gibraltar - South to North route
mtDNA	Mitochondrial DNA
N_{ANC}	Ancestral size Palaeolithic
N_{ANCNEO}	Ancestral size Neolithic
NE-SW	Northeast - southwest
noMIG	Lack of migration between both regions
nsMIG	Migration allowed only from the Iberia Peninsula to North Africa
PAR	Pseudoautosomal regions
PCA	Principal Component Analysis
PLS	Partial least square
POR	Porto
SE-NW	Southeast - northwest
snMIG	Migration allowed only from North Africa to the Iberia Peninsula
SS	Summary statistics
SYR	Syrian
T_{NEO}	Initial Neolithic Expansion
T_{OOA}	Out of Africa
TSI	Tuscany
T_{start}	Initial Expansion of modern humans
WNA	Western North Africa
ya	Years ago
YRI	Sub-Saharan (Yoruba)

Chapter 1.

General Introduction

1.1. Neolithic Revolution

Until the end of the Pleistocene, anatomically modern human (AMH) populations lived as hunter-gatherers (HG), exploiting aquatic and land resources. Then human populations started a transition to an agriculture and pastoralist lifestyle, the Neolithic Revolution, which fundamentally changed cultural and demographic features of human populations. Domestication on plants and animals emerged independently in multiple regions around the world, including Levant, China, New Guinea, Ethiopia, North America, Mesoamerica and South America, during a chronological window comprised between 12,000 and 5,500 years ago (ya) (Figure 1.1) (Diamond and Bellwood 2003). With the exception of dog domestication, which most likely occurred in a single event 20,000-40,000 ya and was HG way of life compatible (Botigué et al. 2017), the Neolithic transition started with the domestication of wild plants, such as cereals and legumes, and continued later with the domestication of animal species (Scanes 2018).

The adaptation to Agriculture led to an increment of food production per region of productive land, which made possible to feed a larger number of individuals, gradually increasing the population density. Moreover, food productive populations adopted a sedentary lifestyle, storing food reserves and constructing villages, which allowed the development of more complex technologies and stratified communities (Diamond and Bellwood 2003). However, archaeological evidence show that the initial transition to a farmer lifestyle led to a decreased stature and poorer health

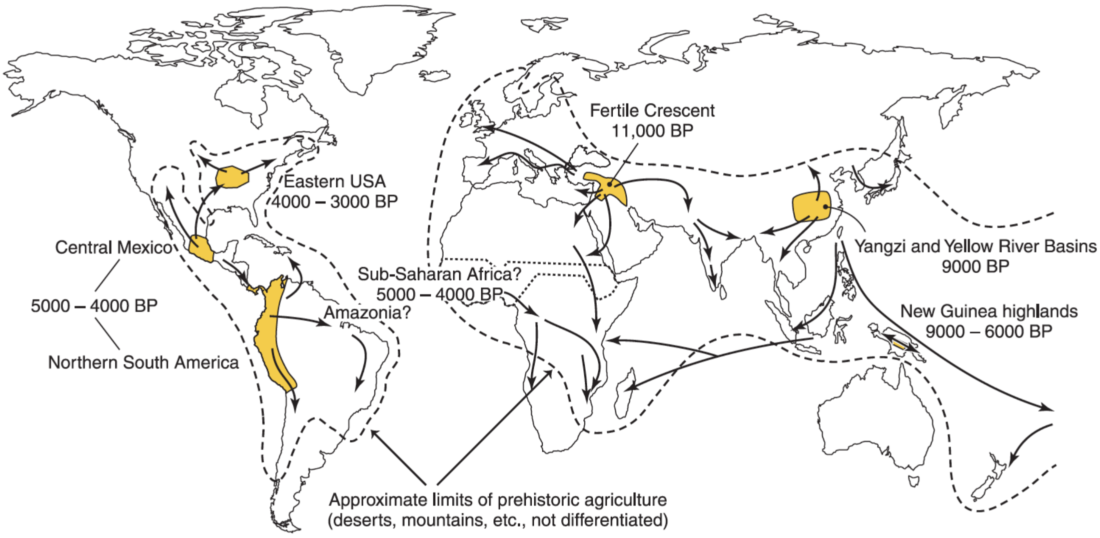


Figure 1.1. Independent regions of origin of the Neolithic transition through the world (from Diamond and Bellwood 2003. Reprinted with permission from AAAS).

conditions (Cohen and Armelagos 1984). Indeed, in a review by Mummert et al. (2011), authors showed that independently of the region around the world or the type of crops they farmed, newly farming communities were shorter and less healthy than contemporary HG populations. These conditions were most likely caused by nutritional deficiencies due to a diet based on a reduced number of crop species and a substantial decreased of meat consumption (Richards 2002), therefore poor on essential nutrients, in opposition to a more varied and rich diet from HGs populations. Indeed, there was a decreased in adult average height around 4% between Palaeolithic and Neolithic (Hermanussen 2003), which was only recovered during the 20th century. Additionally, the growing population density associated with farmer settlements and poor sanitary conditions would have helped the spread of infectious diseases (Omran 1971).

Counterintuitively, an increased population density accompanied the transition to a sedentary lifestyle, despite the harsh life conditions and the nutritional stress that farmer populations faced. Palaeodemographic data from cemeteries show a major demographic shift by an abrupt increase of the proportion of the juvenile skeletons during this crucial period, a process commonly called as Neolithic Demographic Transition (Bocquet-Appel 2002). This demographic growth has been attributed to a substantial increase in female fertility, due to the reduction of the duration of the reproductive cycle, possible caused by a combination of several demographic features that resulted from the adoption of a sedentary lifestyle (Bocquet-Appel 2011). Despite the unprecedented birth rates estimated for famer populations in comparison with forager populations (Bocquet-Appel and Naji 2006), the arrival of the Neolithic did not produce a major growth in world’s population, due to an increase of the mortality rate (Cohen and Armelagos 1984). The appearance of villages with high population density, with poor sanitary conditions, associated nutritional deficiencies, would have helped the spread of infectious diseases which incremented child mortality and reduced the life expectancy (Bocquet-Appel 2008). Indeed, the disease burden associated with the adoption of agriculture resulted in the first described Epidemiological transition (Armelagos and Harper 2005).

The process by which the farmer industry gradually replaced the HG lifestyle all around the world has been one of the most studied subjects by different fields of science. Dates on Archaeological data from Neolithic sites (Lemmen et al. 2011; Bocquet-Appel et al. 2012) show a spatiotemporal pattern that is in agreement with a radiation from the independent founder regions that introduce local differences in lifestyle, technology and population density. Therefore, the expansion dynamics of the Neolithic along these spatial gradients could be modelled by diffusion processes. There are three main processes proposed for the adoption of the Neolithic industry: *cultural diffusion*, which assumes that the spread of the Agriculture occurred by information transmission through social learning, with little or non population movement (Edmonson 1961); *demic diffusion*, which encompasses the spread of Agriculture within a range expansion of farmers, that gradually replaced local HG populations (Ammerman and Cavalli-Sforza 1973); and *independent origin*, an autonomous and convergent uprising of the Neolithic industry in different locations with the domestication of different plant and animal species (Diamond and Bellwood 2003).

The role of cultural and demic diffusion in the expansion of the Neolithic industry throughout the world has been a subject of debate for several decades. Initial studies in Europe pointed to a more relevant role of the cultural component for the expansion of farming (Edmonson 1961). In contrast, works by Ammerman and Cavalli-Sforza (1971; 1973) highlighted the role of the demic diffusion, advocating that this diffusion process was the most relevant for the spread of the Neolithic in Europe. Additionally, authors also predicted that this “wave of advance model” (Ammerman and Cavalli-Sforza 1971, pag.687) would lead to genetic clines caused by the interbreeding between individuals from the Mesolithic and the Neolithic. However, the model created by Ammerman and Cavalli-Sforza did not exclude the cultural diffusion. Indeed, authors emphasized that demic and cultural diffusions had a variable importance for the Neolithic expansion across Europe. In recent years, a new approach was developed to integrate both diffusion processes in one unified expansion model to evaluate the relative importance of the demic and cultural diffusion in the Neolithic transition (Fort 2012).

1.1.1. Neolithic Transition in Europe

The Neolithic transition in Europe comprises two main steps: the uprising of Agriculture and the farmer lifestyle in the Fertile Crescent region in the southwest Asia, approximately 12,000 ya (Figure 1.2) (Asouti and Fuller 2013; Arranz-Otaegui et al. 2016; Broushaki et al. 2016), and the gradual expansion of the Neolithic thread into Europe introduced through the region of Anatolia (Turkey) (Omrak et al. 2016). The spread of the Neolithic in Europe was object of extensive analysis due the richness of available archaeological sites and ancient DNA (aDNA). Indeed, radiocarbon dating results show that the oldest evidence of the presence of farming communities in the southeastern Europe could be traced back to around 9,000 ya

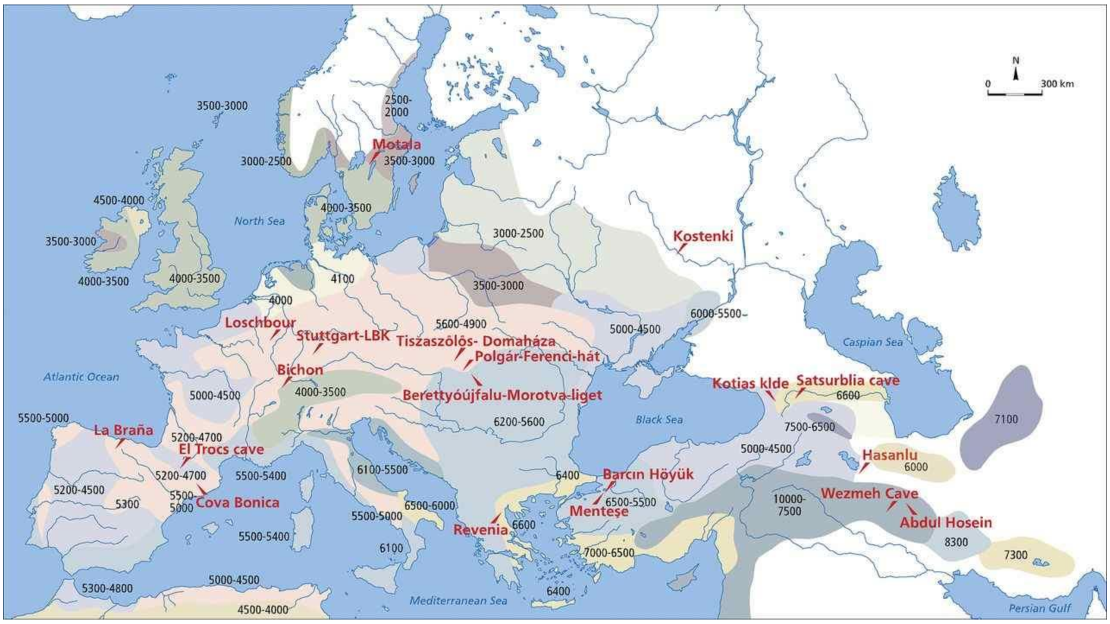


Figure 1.2. Map of the Neolithic transition across Europe. Different colours represent the isochrones of the estimation of the earliest time of arrival of farming in years before the current era (BCE) (from Broushaki et al. 2016. Reprinted with permission from AAAS).

(Weninger et al. 2014) and that these communities presented a close cultural resemblance to Levant Neolithic package (Özdogan 2014). Recent genetic analysis, identified Early farmers from Levant and Anatolia as the source population of European Neolithic groups (Mathieson et al. 2015; Kılınç et al. 2016; Lazaridis et al. 2016; Omrak et al. 2016). These first Neolithic communities from Anatolia (10,000 ya) were characterized by small-scale cultivation and herding but extensively relying on foraging practices (Kılınç et al. 2016). Indeed, this region faced a complex Neolithization process demonstrating some cultural continuity from Epipaleolithic threads, regional variations of the Neolithic package and a possible contribution of local HG populations to the gene-pool of Anatolian Neolithic populations (Kılınç et al. 2017). From this region the Neolithic spread into Europe from two main routes proposed by archaeological data: the Danubian route, which follow the Danube and Rhine River valleys into the northwest and central Europe and was characterized by the Linearbandkeramic culture (LBK); and the Mediterranean route, which follow a southern maritime route through the northern coast of the Mediterranean Sea and was associated with the adoption of the Impreso and Cardium-Impresso culture groups (Price 2000).

The pace of the dispersal of the Neolithic has also attracted the interest of researchers and led to the development of various dispersal models, such as leapfrog and wave of advance. The spread of the Neolithic happened rapidly and since its arrival to southeastern regions 9,000 ya, all Europe had adopted the farmer lifestyle within the next 4,000 years (Fort 2015) and early estimates pointed to a spread velocity as fast as 1 km per year (Ammerman and Cavalli-Sforza 1984; Pinhasi et al. 2005), which favoured a demic diffusion model for the Neolithic expansion. However, results on archaeological data showed that the cultural diffusion had a far more relevant role than previously thought, which led to the development of new models integrating demic and cultural diffusion, emphasising the contact between local HGs and foreign farmers (Fort 2012; Fort 2015). In recent decades, the development of aDNA analyses complemented and redefined the way we understand prehistoric migrations and the demographic dynamics of the Neolithization of Europe. Indeed, integrating both Archaeological data and aDNA results show very complex picture of the adoption of Agriculture.

1.1.1.1. Neolithic expansion in Southern Europe and Western Mediterranean

At the same time that Neolithic spread into central regions of Europe through the Danubian (continental) route, farming lifestyle, coming from Anatolia and Near East, expanded in southern regions of Europe until it reached the Iberian Peninsula, following a maritime route through the northern margins of the Mediterranean Sea (Figure 1.2). The spread of the Neolithic into southern and south-western Europe is characterized by the expansion of the Impreso and Cardium-Impresso culture groups that expanded at an estimated rate of more than 10 km per year, which can be best explained by a pioneer maritime colonization (Zilhao 2001). Indeed, archaeological (Bocquet-Appel et al. 2009) and genetic (Fernández et al. 2014) results suggest that the first movements of Neolithic farmers into Europe were the result of maritime migrations that colonize southeastern regions (i.e. Greece), via Cyprus and Aegean Islands, from 8,000 to 9,000 ya, mostly by a leapfrog pattern of expansion that established small communities that explored favourable agricultural environments in the Greek coast and Aegean islands. From there, the Neolithic expanded towards north, to Balkans and central regions of Europe, and towards west into the Italic Peninsula (Zeder 2008). While archaeological data attesting the Neolithic presence in Italy are scarce, results point to the introduction of farming in southeastern regions of the peninsula by seafaring colonization at least 8,000 ya and to a major discontinuity between Mesolithic and Neolithic cultures (Biagi 2003), but without a complete replacement of HG populations (Coppa et al. 2007). Recent genomic analysis on modern southern Italy showed a high proportion of ancestry retrieved from Early Neolithic farmers from Anatolia and Europe and only modest levels of HG ancestry, suggesting that farming largely replaced Mesolithic HG lifestyle, which is in agreement with archaeological findings (Sarno et al. 2017).

The initial stage of the Neolithic transition in southern France is characterized by the arrival of small groups of seafarers from the Italic Peninsula around 7,500 to 8,000 ya. Indeed archaeological evidence show close cultural resemblance between Italian Early farmers and southern France seafarers, mainly with ceramic and lithic systems and the presence of typical obsidian tools from Palmarola, Sardinian and Lipari (Bernabeu Aubán et al. 2017). After these initial settlements in the Mediterranean coast the Neolithic spread inwards occupying new ecosystems and allowing the diversification of food resources (Perrin et al. 2017). Interesting, at present, there is still lack of unequivocally evidence of contact between Mesolithic HG and Early farmers in southern France. Indeed, dating in archaeological sites show a gap of a few centuries between the occupation of HG and farmers population groups at the same location or a geographic gap of a few hundreds of kilometres between contemporary Late Mesolithic and Early Neolithic remains. However, the special distribution of a particular type of arrowhead (“Montclus arrowheads”) show the possibility of potential contacts between both population groups at the fringe of the geographic range of Early Neolithic farmers (Perrin 2013).

1.1.1.2 Neolithic Transition in the Iberian Peninsula

As for other Western Mediterranean regions the arrival of the Neolithic to the Iberian Peninsula, between 7,500 and 8,000 ya, is largely connected with the spread of the Cardial culture (Zilhao 2001; Rowley-Conwy 2011). First farmers arrived from the Italic Peninsula by pioneer maritime colonization, and first settlements were established in coastal enclaves in peripheral regions of Iberia, where the conditions were favourable for farming (Zilhao 2001; Zeder 2008). Archaeological data support a scenario of multiple points of origin of the Neolithic in the Mediterranean coast, suggesting a simultaneous initiation of farming in several regions of the Iberian Peninsula (Isern et al. 2014; Martins et al. 2015). Such a scenario could be the result of an expansion of farming by land throughout the European Mediterranean coast associated with the arrival of seafarer farmers from the western regions of North Africa, or a rapid expansion of European groups of farmers through sea along the Mediterranean Coast (Bernabeu Aubán et al. 2015). More recently, Isern et al (2017) modelling the voyaging process of the spread of Neolithic in the Iberian Peninsula showed that archaeological data fit better a leapfrog model, with long distance migrations and multiple points of entry throughout the Iberian coast (Figure 1.3).

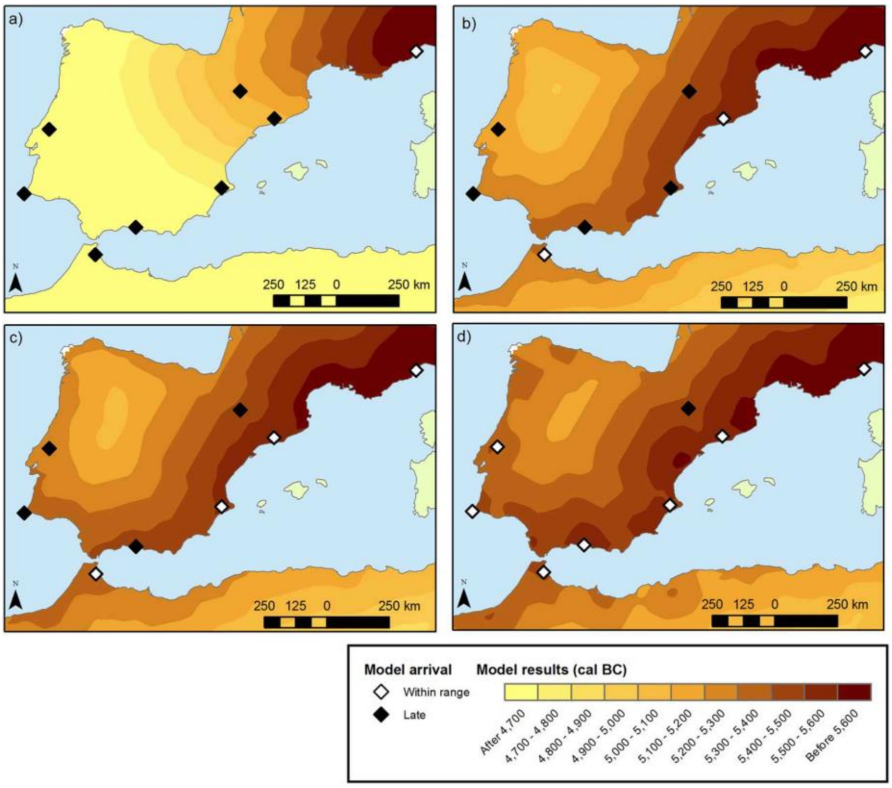


Figure 1.3. Comparison of different models for the arrival of the Neolithic to the Iberian Peninsula using archaeological data**.** The model with the best fit to the observed data (model D) included long distance migrations (leapfrog). White diamonds represent the coastal areas of the observed arrival estimates considered in a model (from Isern et al 2017. Reprinted with permission from PNAS).

Moreover, authors suggested that the rapid spread of farming was associated to an effective growth rate of at least 2.3%, which could be achieved by an increased fertility in farming population (as discussed in the section 1.1) and by interbreeding and/or acculturation with Mesolithic HG populations.

The wide distribution of Early Neolithic sites throughout the Iberian Peninsula coast, from the northeastern Spain to Atlantic façade in Portugal, indicates a rapid spread of the Neolithic industry, which has been associated with a demic diffusion process for the transition (Zilhao 2001; Zilhão 2003). However, the Neolithic expansion throughout the Iberia is still a much-debated subject. It is generally accepted that the transition was heterogeneous and different regions endured varying degrees of demic and cultural diffusion of farming (de Pablo and Puche 2009). Indeed, the spread of farming seems to have enclosed a partial replacement of indigenous populations by colonizing farmers as well as the adoption and adaptation of Neolithic threads by Mesolithic HG groups (Zeder 2008). Additionally, results based on radiocarbon dating suggests, for some regions, a time gap between the occupation of Mesolithic groups and the establishment of Neolithic communities and therefore the absence of direct contact between both population groups (Pardo Gordó et al. 2018).

Only in the last two decades, results based on aDNA for mtDNA and Y-Chromosome started to provide a complementary view to archaeological data concerning the Late Mesolithic and the Neolithic periods and the population dynamics between HG and farmer groups during the transition to agriculture in the Iberian Peninsula (Chandler et al. 2005; Lacan et al. 2011; Gamba et al. 2012; Hervella et al. 2012; Olalde et al. 2015; Szécsényi-Nagy et al. 2017). However, the limited number of human remains from the Early Neolithic period in the Iberian Peninsula led to a lack of available data to understand the genetic affinities of Iberian first farmers and the expansion of Cardial culture in the region. Nevertheless, the growing number of aDNA studies made possible to have a more detailed picture of the farming transition in the Iberian Peninsula.

Recent aDNA results show a heterogeneous expansion of the Neolithic threads in Iberia, suggesting a more complex scenario of the interaction between indigenous Mesolithic populations and incoming farmers. In general, inferred mtDNA lineages on Early Neolithic individuals from Iberia comprises typical HG haplogroups from southwestern Europe (such as U5b, H, and N*) and haplogroups associated with early farmers from Europe (such as K, J, T2, HV, V and X), which leads to an intermediate position between these two populations groups in a PCA despite some regional variation (Lacan et al. 2011; Gamba et al. 2012; Szécsényi-Nagy et al. 2017). Results on northeast Iberia supports the hypothesis of the arrival of the Early Neolithic through this region and posterior expansion following the Iberian coast line and rivers into the central regions (Lacan et al. 2011; Gamba et al. 2012; Haak et al. 2015; Olalde et al. 2015; Szécsényi-Nagy et al. 2017). Indeed, this region received the largest influx of Neolithic migrants arriving along Mediterranean route, attested by the largest proportion of farmer lineages in northeastern Iberian than in other regions of the Iberia and that between all Iberian regions, northeast Iberia has the largest genetic similarity with Anatolia Early Neolithic farmers (Szécsényi-Nagy et al. 2017). Southeastern and southwestern Iberian results (Szécsényi-Nagy et al. 2017), besides the lower proportion Neolithic lineages, showed a haplogroup composition less diverse than their northeastern counterpart. This drastic increase of Neolithic lineages in such a narrow temporal window (Zilhao 2001) suggests a rapid expansion of farming as well as a demic process of diffusion despite the early interbreeding with indigenous HG populations. In clear contrast, early Neolithic samples taken from central Portugal (Chandler et al. 2005), central Iberia (Haak et al.

2015; Alt et al. 2016; Szécsényi-Nagy et al. 2017) and northern Iberia (Hervella et al. 2012) harbour a much larger proportion and diversity of HG haplogroups, and shared less Early Neolithic lineages, therefore, presenting a greater genetic similarity with southwestern HG groups. All these aDNA results taken together demonstrate that the Neolithic revolution arrived Iberian Peninsula from southeastern Europe through a fast pioneer maritime migration, in similar fashion to central Europe through the continental route. However, the impact of the contact with indigenous HG was much more relevant during the Iberian Neolithic transition when compared with the similar period in central Europe, resulting that different early Neolithic Iberian communities had different proportions of HG ancestry. Indeed, this apparent trend is accentuated with the geographic distance to northeastern Iberia. The further the Neolithic expanded into central and northern inner regions and southern coastal parts, the larger the HG indigenous component, suggesting a relevant role of the cultural process of diffusion during the Neolithic transition in the Iberian Peninsula.

The transition from Early Neolithic to Middle/Late has been receiving a growing amount of attention due to the interesting population dynamics in this period. An increase in HG ancestry in farming populations in central and northern regions of Iberia as been described by recent aDNA results for the middle Neolithic and early Chalcolithic (Günther et al. 2015; Mathieson et al. 2015; Lipson et al. 2017; Szécsényi-Nagy et al. 2017). Despite the scarcity of available data, results for La Mina site (north-central Iberia) show a HG ancestry proportion of 23% for the Middle Neolithic, larger than contemporaneous populations from central Europe, and 27% for the Chalcolithic (Lipson et al. 2017). In line with these results, Szécsényi-Nagy et al. (2017) demonstrated that the increased of HG ancestry through time from the early Neolithic Iberia to Chalcolithic, in northeast and central Iberian.

1.1.2. Neolithic Transition in North Africa

In opposition to Europe, Neolithic archaeological and genomic data from the transition period in North Africa are still scarce. The dynamics of the farming expansion and population relationship between indigenous HG and Neolithic migrants is still far from being completely understood. Nevertheless, most results on archaeological data agree that agriculture was introduced in North Africa during the Early Neolithic period by migrants from the Near East (Linstädter et al. 2012) and northern coast of the Mediterranean (Zilhão 2014) by mainly demic diffusion. However, the spread of farming throughout the vast area of North Africa could also be the result of native development of Neolithic threads by Epipaleolithic communities (Barton et al. 2008; Mulazzani et al. 2016). Indeed, archaeological data show that Epipaleolithic communities in northeastern Africa, before the arrival of Neolithic threads from the Near East, had developed subsistence and economic technologies based on the exploitation of natural resources (hunting, fishing and gathering of plants and molluscs) (Di Lernia 2001; Garcea 2004), developing in some regions strategies of resource management, like the delayed consumption and the use of pottery (Di Lernia 2001; Garcea 2004; Mulazzani et al. 2016). Morales et al. (2015) demonstrated that Capsian populations exploited both acorns and pine nuts, highly nutritious and storable food plants, which could have played a relevant role in human subsistence, helping the development of a sedentary lifestyle without domesticated crops. In northeastern Africa the earliest date recorded of exploitation of domesticated plants is at north Sudan (in Nubia at R12 site) around 7,000 ya, with evidence of the presence of phytoliths of Near Eastern cereals (wheat and barley) (Madella et al. 2014). In Egypt, at the Kom K and Kom W sites, radiocarbon dates indicate the presence of domesticated plants (wheat and barley) since 6,500 ya (Wendrich et al. 2010). On the other hand, the introduction of domesticated animals in northeastern Africa was most likely associated with an 8,200 ya arid event (Barich 2014). Archaeological evidence indicate the establishment of a pastoralist economy based on neareastern caprines (Vermeersch et al. 2015) and goats (Gautier 2014) in Egypt, at least 8,000 ya. However, earlier evidence of management of autochthonous animal species, such as Barbary sheep (Di Lernia 2001) and aurochs (Barich 2010), by Epipaleolithic populations have been reported between 10,000 and 9,000 ya, which suggests that the introduction of domesticated animals in the region could have occurred by adaptation of local foragers (cultural diffusion) without or with little migration of Neolithic farmers.

While the Neolithic could have been introduced in Libya by land from Egypt (Garcea 2004), how agriculture gradually expanded throughout North Africa reaching the westernmost regions of Morocco is still a matter of debate, despite several results pointing to a maritime pioneer migration (Zilhao 2001; Linstädter 2008; Linstädter et al. 2012). The earliest evidence of exploitation of domesticated plants in the entire region of North Africa was found in northern Morocco coast (at Kaf Taht el-Ghar and Ifri Oudadane sites) where charred seeds of domesticated plants were dated to around 7,600 ya (Ballouche and Marinval 2003; Linstädter et al. 2012; Morales et al. 2013). To date, this is the only evidence of domesticated plants species in North African context between Libya and Morocco. Evidence of animal domestication (sheep), also in Morocco at Ifri Oudadane and Tingitana Peninsula, are dated to around 7,400 ya (Zilhão 2014; Martínez-Sánchez et al. 2017). Interestingly, similar radiocarbon dates on domesticated plant remains in both margins of western Mediterranean (Morocco and Iberian Peninsula) suggest a synchronous arrival of the Neolithic threads to both regions (Morales et al. 2013). Moreover, Zilhão (2014) suggested that the similarities concerning the cardial ceramic styles present at both shores of the Strait of Gibraltar are associated with an earlier arrival of the Neolithic to Iberia and indicate that after the diffusion of agriculture in Iberia Peninsula the Neolithic spread into northwestern Africa, attesting the Strait of Gibraltar as a bridge between both continents.

To date, only one study has focused on the analysis of aDNA of individuals from the early and late Neolithic sites in North Africa. Fregel et al. (2018) identified, in early Neolithic samples (~7,000 ya) from Morocco, an endemic ancestry component that is only present in modern North African populations. Interestingly, these early Neolithic groups shared genetic affinities with pre-Pottery Anatolian famers and Near Eastern Natufians HG, which suggest early westward migrations of pre-pottery groups. Additionally, analyses on late Neolithic samples (~5,000 ya) from Morocco showed the presence of an Iberian component that could be the result of gene flow across the Strait of Gibraltar, attesting the presence of pre-historic European ancestry in North African populations before the Roman conquest (Fregel et al. 2018).

Concluding, the spread of the Neolithic culture throughout North Africa is still a very fragmented and incongruent story. While it seems that at northeastern Africa indigenous populations had an active role during the Neolithic transition, in western regions their role is still undefined. Moreover, besides the study mentioned in the previous paragraph, no comprehensive genetic study has been conducted to support the results retrieved from archaeological sites until now.

1.2. Genetic Diversity in Human Populations

Throughout history, human populations have interacted, due to migrations, invasions, slavery or commercial trades, and the records of these historical events led to the geographic patterns of genetic diversity in human populations. In the last few decades, results on several types of genetic markers provided large amounts of information that complemented other sources of information, such as history, archaeology, anthropology and linguistics. More recently, the emergence of genome-wide technologies (The International HapMap Consortium 2005; 2007) associated with development of new bioinformatics tools, made possible to study the human genome diversity in a much finer scale, providing evidence of past admixture events and barriers to gene flow, of human adaptation to new environments and to new ways of subsistence or cultural practices (Schraiber and Akey 2015).

Several diversity projects have used genetic data to describe the migratory routes and geographic structure of AMH populations. In a landmark study, Rosenberg et al. (2002) showed that, at a global and continental scale, the genetic structure of human populations was generally weak and the largest differences were found within populations. Moreover, authors identified five major geographical regions: sub-Saharan Africa, Eurasia (composed of Europe, the Middle East, Central and South Asia), East Asia, Oceania, and America. Subsequent analyses using genome-wide data allowed the refinement of the analysis of the genetic ancestry of human populations, gradually identifying substructure within major regions (Li et al. 2008; Tishkoff et al. 2009; Shriner et al. 2014). More recently, Baker et al. (2017), using 282 populations distributed worldwide, identified 21 ancestries on present-day populations and that 97.3% of the individuals show a mixed ancestry. The genetic structure of human populations is highly correlated with cultural features (such as linguistics) and geography (Shriner et al. 2014; Baker et al. 2017). Indeed, a recent study using spatial explicit analysis on a large African and Eurasian dataset identified several topographic and marine features as barriers to gene flow, while numerous regions of high genetic similarity resembled to historical migrations and cultural boundaries (Peter et al. unpublished data).

1.2.1. Genetic Structure in Europe

The genetic landscape of present-day European populations and the demographic processes that shaped the patterns of genetic structure have always attracted great interest of researchers from the different fields of knowledge. Early studies on genetic gradients of European populations using PCA showed a southeastern-northwestern axis of variation (Cavalli-Sforza and Edwards 1963; Menozzi et al. 1978; Piazza et al. 1995) (Figure 1.4A), with a high correlation with geography (Menozzi et al. 1978). Classical results from genome-wide data confirmed isolation-by-distance and *allele surfing* patterns phenomena, derived from the geographic and environmental conditions [i.e., Last Glacial Maximum (LGM)], as a cause of the observed genetic variation in Europe (Lao et al. 2008; Novembre et al. 2008; François et al. 2010; Arenas, François, et al. 2013) (Figure 1.4B). In recent decades, the growing number of aDNA studies on European populations allowed us to understand that present-day populations reflect past demographic events.

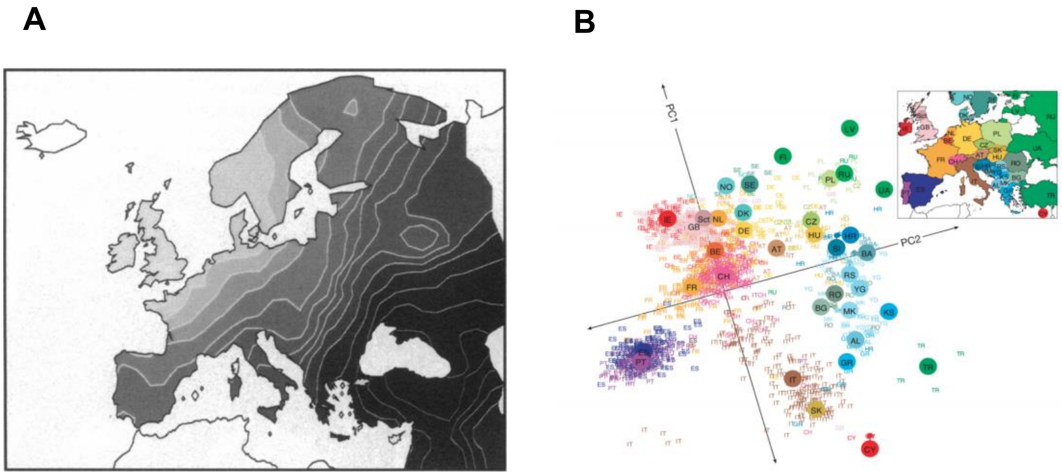


Figure 1.4. Maps of genetic differentiation in Europe**.** (A) Synthetic map of the genetic variation for the first principal component using 95 classical genetic markers (from Piazza et al 1995. Copyright 1995 National Academy of Sciences); (B) Principal component analysis based on a SNP dataset of 1,387 individuals distributed across Europe. In small letters are represent all individuals that composed the dataset and the large circles represent the median position of each country (from Novembre et al 2008).

Europe has a complex demographic history since the arrival of AMH around 45,000 ya (Mellars 2004; Benazzi et al. 2011). Interestingly, initial European settlers, between 45,000 and 37,000 ya, had a very limited contribution to present-day populations (Fu et al. 2014; Fu et al. 2015; Fu et al. 2016). Only after 37,000 ya do Europeans individuals, from the Upper Palaeolithic, carry genetic components shared with most of the present-day populations (Seguin-Orlando et al. 2014; Fu et al. 2016). Moreover, until 14,000 ya European populations do not carry any genetic signature of admixture with an external population (Fu et al. 2016). During this period of time, Palaeolithic individuals endured the LGM that caused a severe range contraction, with populations seeking refuge in southern regions (Stewart and Stringer 2012), and most likely causing a bottleneck, which is reflected by the low genomic diversity estimated in Mesolithic populations (Skoglund et al. 2014). Starting around 14,000 ya, local European populations admixed with migrants coming from the Near East (Fu et al. 2016). Curiously, the genetic landscape of Mesolithic HG populations indicates the presence of some geographical structure, with western (Sánchez-Quinto et al. 2012; Olalde et al. 2014) and central (Lazaridis et al. 2014; Jones et al. 2015) populations appearing distinct from Scandinavian and eastern populations (Haak et al. 2015). Indeed, Scandinavia Mesolithic populations show evidence of two postglacial migrations (one from the south and another one from the northeast following the Atlantic coast of Norway) resulting in a higher genetic diversity in Scandinavia Mesolithic individuals than in western and central Mesolithic populations (Günther et al. 2018).

During the Neolithic revolution, agriculture was introduced in Europe by migrations of early Anatolian farmers, reaching Scandinavia and the British Islands around 6,000 ya. The genetic diversity of European populations was partially and heterogeneously replaced by an earlier farmer ancestral component. Several studies had highlighted the strong genetic differentiation between the HG ancestry and the Neolithic component (Skoglund et al. 2012; Gamba et al. 2014; Günther et al. 2015; Olalde et al. 2015; Mathieson et al. 2018). Indeed, while Mesolithic ancestry falls outside the scope of the genetic diversity of present-day European populations, being closest to modern northeastern populations, early Neolithic individuals are genetically similar to present day southern Europeans, closely resemble modern day Sardinians (Haak et al. 2015).

The Neolithic was not the last massive pre-historic migration that had great impact on the genetic structure of European populations. During the late Neolithic and Bronze Age movements of Yamnaya pastoralists from the Russian steppe introduced an eastern component that replaced most ancestry of northern European populations and partially replaced the ancestry of central and western Europeans (Allentoft et al. 2015; Haak et al. 2015; Mathieson et al. 2018). This migration of herders from Eurasian steppe is thought to be responsible for the expansion of the Neolithic Corded Ware culture into central Europe and the spread of Indo-European languages (Haak et al. 2015).

Another demographic relevant cultural expansion associated with the Bronze Age is the spread of the Bell Beaker culture (Czebreszuk 2004). The oldest radiocarbon dates of this complex were found in Atlantic Iberia from around 4,750 ya, suggesting an Iberian origin of the Bell Beaker culture (Cardoso 2014). However, the origin of this culture is still very contentious (Jeunesse 2015). A recent aDNA study that analysed a widespread distribution of Bell Beaker associated individuals found a dual model of the complex expansion (Olalde et al. 2018). While the divergence between the ancestral origin of Iberian and Central European Bell Beaker populations, suggests that the expansion of this culture between these two regions is associated with the spread of ideas (acculturation), in other regions (such as Britain) the spread was accompanied by a genomic transformation associated with migrations (Olalde et al. 2018).

At the end of the Bronze Age the genetic patterns already resemble the modern-day European populations. However, during the last 3,000 years the European history is marked by several migrations. Extensive studies on British Isles’ populations using both ancient and modern DNA revealed demographic events occurring during the

Iron Age (Martiniano et al. 2016; Schiffels et al. 2016), the Roman empire (Leslie et al. 2015; Martiniano et al. 2016), the Anglo-Saxon (Leslie et al. 2015; Schiffels et al. 2016) and Viking invasions (Leslie et al. 2015). The mainland European populations were also affected by several small and large migrations from both within and outside Europe, that regionally affected the genetic diversity and structure (Botigue et al. 2013; Ralph and Coop 2013; Hellenthal et al. 2014; Busby et al. 2015; Finocchio et al. 2018; Veeramah et al. 2018). Nevertheless, the HG recolonization of Europe after the LGM, the expansion of the Neolithic by near eastern populations and the migration of Yamnaya pastoralist from the steppe are the three most influential demographic events to shape the genetic structure of modern-day European populations.

1.2.1.1. Genetic Structure in the Iberian Peninsula

The Iberian Peninsula has been continuously inhabited by AMH since their arrival to the region around 41,000 ya (Mellars 2004), having one of the most complex demographic histories among European populations. Indeed, during the LGM the region was one of three major refuges in southern Europe for human populations and a reservoir of biodiversity, from where AMH recolonize Europe when the ice sheets gradually melted down. Moreover (as detailed in Section 1.1.1.2), the Neolithic arrival to the Iberian Peninsula was very heterogeneous, with coastal regions having a larger proportion of Neolithic lineages, favouring a demic diffusion process of agriculture mediated by pioneer maritime migrants, while in central and northern regions, the acculturation process had a more relevant role. On the other hand, the steppe migration during the late Neolithic and early Bronze Age had little influence on the genetic makeup of Iberian populations (Morales et al. 2013; Szécsényi-Nagy et al. 2017; Olalde et al. 2018). Indeed, among European populations the Iberians (Günther et al. 2015) and Sardinians groups show the largest affinity with early Neolithic farmers (Skoglund et al. 2012; Lazaridis et al. 2014; Günther et al. 2015). These results indicate that after the end of the Bronze Age the genetic ancestry on the Iberian Peninsula did not suffer another major turnover, in line with the rest of Europe, despite, in recent millennia, several population groups occupied the region, such as, Phoenicians, Greeks, Romans, German tribes, and more recently, the region had a centuries-long period of Muslim rule (Carr 2000).

The genetic structure of the Iberian Peninsula reflects its complex demographic history and admixture backgrounds, being one of the regions of Europe with the highest genetic diversity (Wang et al. 2012). Previous results on uniparental markers and genome-wide data only presented a fragmented picture of the genetic structure in Iberia because analysis were performed on small fractions of the genome (mtDNA and Y chromosome) (Adams et al. 2008; Alvarez et al. 2010; Pardiñas et al. 2012; Pardiñas et al. 2014; Santos et al. 2014) or on a small number of populations (Gayan et al. 2010; Rodríguez-Ezpeleta et al. 2010). Nevertheless, results show low levels of genetic differentiation throughout the Iberian Peninsula (Gayan et al. 2010; Santos et al. 2014). Despite the global homogeneity, evidence on uniparental markers showed slightly larger levels of genetic differentiation along the Mediterranean-Atlantic cline than between north-south populations (Santos et al. 2014) and evidence of genetic structure were usually only reported at a local level (Alvarez et al. 2010; Pardiñas et al. 2012; Pardiñas et al. 2014). Marks of genetic differentiation were also found along the political border between Portugal and Spain, between northeasternmost regions of Portugal and the region of Zamora in Spain (Alvarez et al. 2010; Pinto et al. 2015).

The Basque Country population is an interesting case within the Iberian Peninsula. Basques speak a non-Indo-European language that does not belong to any other language family in the world and they have been considered as a genetic isolated among their neighbour populations (Adams et al. 2008; Rodríguez-Ezpeleta et al. 2010; Behar et al. 2012) despite some contradictory results (Garagnani et al. 2009; Laayouni et al. 2010).

Until recently, the fine-scale structure of Iberian Peninsula remained unexplored. Bycroft et al (unpublished data) using a haplotype-based approach performed an extensive analysis on the genetic structure of Iberia. Authors found an east-west preferential direction for genetic differentiation and north-south direction of gene flow, confirming previous assumptions produced by uniparental markers (Santos et al. 2014). These patterns of genetic structure were associated with the population movements during the *Reconquista*, the period between 711 CE and 1492 CE, in which catholic kingdoms took control of the Peninsula after the Muslim invasion (Carr 2000). Also, in northern regions, the genetic clusters resembled the linguistic boundaries associated with the Catholic kingdoms existing at the critical period of the *Reconquista*. Moreover, authors indicate that despite their genetic particularities (Adams et al. 2008; Rodríguez-Ezpeleta et al. 2010; Behar et al. 2012), Basques show evidence of admixture with their neighbour populations at least in the last 1,000 years, which is in agreement with previously described results for the Y chromosome (Hurles et al. 1999).

Another interesting feature relevant for the study of the genetic structure of modern-day Iberian populations is the influence of sub-Saharan Africa or North Africa migrations into the region. Results show that only a small proportion of the Iberian gene pool can be traced back to sub-Saharan populations (Moorjani et al. 2011), while the North African ancestry demonstrated incoherent results in comparison to the demographic history. Migrations between both regions, especially across the Strait of Gibraltar have been described at least since the arrival of the Neolithic to the western Mediterranean (Zilhão 2014), but there has been suggestions that migration occurred even in pre-Neolithic times (Currat et al. 2010). Results on uniparental markers (Regueiro et al. 2015; Hernández and Calderón 2017) and genome-wide data (Botigue et al. 2013; Arauna et al. 2017) showed evidence of gene flow between the two regions, but also confirmed the Gibraltar Strait as a genetic barrier. Interestingly, the presence of the North African specific mtDNA haplogroup U6 only on northwest of the Iberian Peninsula (Gonzalez et al. 2003), as well as, larger proportions of typical North African Y chromosome lineages in northern regions than in southern areas (Adams et al. 2008) are in line with genome-wide fine scale results, showing highest amounts of North Africa ancestry in western regions of Iberia (Portugal and Galicia), which possibly invalidates the hypothesis of the Muslim occupation as the only explanation for the presence of North Africa ancestry in the Iberian Peninsula. Speculatively, this pattern could be the result of late migrations within Iberia (Bycroft et al. unpublished data) or even enforced migrations of Muslim populations to northern and western regions post-*Reconquista* (Carr 2000).

1.2.2. Genetic Structure in Africa

AMH have existed in Africa longer than in any other place in the world, continuously inhabiting the region for the last 200,000 years, while maintaining large effective population sizes and therefore high levels of genetic diversity (Jakobsson et al. 2008; Schlebusch et al. 2012). Indeed, the ancestral components found outside Africa are only a subset of the diversity within it (Tishkoff et al. 2009). Africa contains more than 2,000 ethnolinguistic groups that largely belong to four independent families: Niger-Congo, Khoisan, Afro-Asiatic and Nilo-Saharan. The majority of modern-day sub-Saharan populations speak one of the more than 500 Bantu related languages, included in the Niger-Congo major family (Tishkoff et al. 2009). The Bantu Expansion is associated with the spread of agriculture, from an independent Neolithic origin in western Africa around 4,000 ya (Figure 1.1), which introduced farming to eastern Africa 2,000 ya and reaching southern regions of Africa around 1,800 ya (de Filippo et al. 2012; Li et al. 2014; Patin et al. 2017). Modern-day populations from southern regions of Africa are highly homogeneous (Tishkoff et al. 2009) suggesting that this massive migration resulted in the replacement of indigenous HG populations, favouring a demic diffusion model (Schlebusch et al. 2012; Li et al. 2014). During the Neolithic expansion the Bantu derived ancestry spread through an area of around 500,000 km^2, being the largest ancestral component present in modern-day African genomes (Tishkoff et al. 2009; Russell et al. 2014). Therefore, our current knowledge concerning the genetic structure of sub-Saharan populations before the Bantu Expansion is scarce. A recent aDNA study (Skoglund et al. 2017) showed evidence of a geographical pattern in HG populations stretching from Ethiopia to South Africa, before the expansion of agriculture, and an ancient genetic structure in western HG populations. Moreover, authors described the profound demographic effect of the spread of farming in eastern and southern African populations, emphasizing the heterogeneity of the interactions with local HG groups (Skoglund et al. 2017). Only few populations from central, eastern and southern regions retain large proportion of HG ancestry. Indeed, HG populations from central Africa rainforest (previously known as Pygmies) derived most of their ancestry from Palaeolithic groups being that admixture with farmer populations occurred only in the last millennium (Verdu et al. 2009; Patin et al. 2014). In the same line, Khoisan populations from southern Africa demonstrate evidence of admixture with west African (Bantu speaking) groups in the last two millennia, which has been linked to the Neolithic expansion in the region (Pickrell et al. 2012), while Hadza HG populations from eastern Africa, besides the admixture event associated with the Bantu expansion, derived a small portion of their ancestry from populations related to Khoisan groups, suggesting ancient migrations between these two regions (Pickrell et al. 2012). Nowadays, sub-Saharan populations exhibit a genetic structure that broadly reflect ethnolinguistic similarities and geography (Busby et al. 2016)

1.2.2.1. Genetic Structure in North Africa

North Africa has a distinctive geographic position, separated from the rest of Africa by the Sahara Desert and from Europe by the Mediterranean Sea; the region is only connected to the Arabian Peninsula by a small stretch of land in the Sinai Desert. This geographic situation led to a unique demographic history in the region, with North African populations clustering closely with Eurasians and limited ancestral components deriving from sub-Saharan populations (Rosenberg et al. 2002; Henn et al. 2012; Arauna et al. 2017). Since the arrival of AMH to the region around 135,000 ya, North African populations endured several climatic changes, involving dry periods during interglacial eras and humid conditions during glacial eras (Castaneda et al.

2009; Drake et al. 2013), causing the Sahara Desert to expand and contract, affecting human settlements in the region (Drake and Breeze 2016).

Modern-day North African populations constitute a very heterogeneous group with ancestry derived from at least three sources with origin in regions outside North Africa: a back-to-Africa migration from Eurasia at least 12,000 ya, a Near East gene-flow (1,400 ya) and migrations from sub-Saharan Africa (1,200 ya) (Henn et al. 2012; Fadhlaoui-Zid et al. 2013; Arauna et al. 2017; Font-Porterias et al. 2018). Indeed, an autochthonous Maghrebi ancestral component is found in an increasing frequency from western to eastern regions of North Africa (Henn et al. 2012), even reaching the Canary Islands (Rodríguez-Varela et al. 2017), that is associated with an ancient back-to-Africa gene-flow. Interestingly, early Neolithic individuals (around 7,000 ya) from Morocco had a large proportion of this endemic Maghrebi ancestry, displaying a strong genetic affinity with Late Stone age individuals from the same region and attesting the population continuity in the region (Fregel et al. 2018). By contrast, genome-wide studies reveal a Near East ancestry gradient decreasing from the east to the west of North Africa, responsible for the genetic affinities with Middle Eastern and west Eurasian populations (Henn et al. 2012). Moreover, a similar pattern is observed across northeastern African populations, where the genetic similarity with Near Eastern groups decreases with increasing geographic distance towards the south across Sudan and South Sudan (Hollfelder et al. 2017). Dates estimated for these admixture events largely coincide with the timing of the Muslim expansion (Arauna et al. 2017; Hollfelder et al. 2017). On the other hand, the sub-Saharan ancestry in North Africa is very limited and occurred only in recent times. southwestern North Africa populations present the largest proportions of sub-Saharan ancestry but varying greatly across individuals, suggesting recent and/or ongoing gene-flow (Henn et al. 2012).

1.3. Spatially explicit approaches applied to human evolution

The evolutionary history of AMH is marked by several events that affected the demography of human settlements, modifying the genetic diversity and reshaping the genetic structure of populations. Evidence of these events is present in the genomes of modern-day populations and, together with archaeological and historical data, allow a better understanding of human evolution. The advent of new sequencing technologies (Metzker 2010) led to a substantial increase of genetic data available for analysis, which gave rise to new methodological challenges. Indeed, in recent years, the development of innovative methods and theories to interpret these massive amounts of data have been an exciting field of population genetics, allowing to build more complicated and realistic models of the human evolutionary history (Schraiber and Akey 2015). Nevertheless, most of these recently developed methodologies largely underused geographic data (Schraiber and Akey 2015; Novembre and Peter 2016). Furthermore, describing the patterns of spatial structure and genetic differentiation is crucial to shed light into the demographic features of modern populations and how they relate to the geographic landscape. Several promising approaches have been developed to visualize patterns of genetic differentiation, which represent the records of the evolutionary history. Two recently introduced approaches based on Bayesian Whishart framework, EEMS (Petkova et al. 2015) and SpaceMix (Bradburd et al. 2016), are able to work with thousands of genetic markers and thousands of individuals, describing patterns of genetic differentiation between populations as a function of the geographic distance. EEMS uses pairwise genetic dissimilarities between individuals to estimate effective migration rate surfaces under a stepping-stone migration model and assuming that populations are in equilibrium. Therefore, inferred surfaces show patterns of genetic differentiation, identifying corridors and barriers to gene flow. On the other hand, SpaceMix uses a covariance matrix to produce a geogenetic map, in which geographic distances between populations are proportional to genetic distances. For instance, larger geogenetic distance between populations implies the presence of a barrier to gene flow. Another approach based on the Wishart distribution describes the genetic differentiation between populations by modelling a landscape feature and testing its effect on gene flow (Hanks and Hooten 2013), while other methodologies modulate F_{st} between populations: to detect deviations from an isolation-by-distance model, identifying barriers to gene flow (Duforet-Frebourg and Blum 2014) or to study anisotropic patterns of genetic differentiation (Jay et al. 2013).

Another strategy that is gaining popularity to study the evolutionary history of humans through the structure and genetic diversity of modern-day populations is the application of spatially explicit computer simulations. Computer simulations allow us to understand which evolutionary parameters influence the processes of particular systems. The development of more realistic scenarios and sophisticated computational frameworks led to the application of computer simulations in a wide number of population genetic fields (Arenas 2012; see also Benguigui and Arenas 2014) and with diverse applications such as hypothesis testing (Pierron et al. 2011; Arenas et al. 2012; Alves et al. 2016; Silva et al. 2017), comparison of different analytical approaches (Arenas et al. 2008; Westesson and Holmes 2009), investigate interactions among evolutionary processes (Arenas et al. 2012; Arenas, François, et al. 2013), and estimate demographic and evolutionary parameters (Beaumont 2010). Concerning the human evolutionary history, spatially explicit computer simulations have contributed decisively for our current understanding of the patterns of genetic diversity through the statistical testing of competing evolutionary models, as well as the estimation of evolutionary parameters specified by those models.

There are two main approaches commonly used in population and evolutionary genetics to simulate the human demographic history, the forward in time simulation of the evolutionary history of a whole population (forward-time) and the coalescent (backward-time) simulation of a sample of a population. In the forward-time approach, simulations evolve a population history from the past to the present (Peck 2004; Carvajal-Rodriguez 2008; Carvajal-Rodriguez 2010). This approach considers all the ancestral information of the population, taking into account diverse processes such as interactions between individuals (Peng et al. 2007; Sellman et al. 2018), population admixture (Rasteiro et al. 2012; Amorim et al. 2017), selection (Ray and Excoffier 2010; Currat et al. 2016; Lewandowska et al. 2017) and complex scenarios of migration (Arenas, François, et al. 2013; Mona et al. 2014; Alves et al. 2016). Consequently, the computational cost of forward-time simulations is usually high in comparison with coalescent simulations, although there have been improvements (Padhukasahasram et al. 2008). On the other hand, the coalescent (Kingman 1982) simulates the genealogical history of a sample from the present to the past (backwards in time) until its most recent common ancestor (MRCA) (see reviews, Nordborg 2008; Wakeley 2008). The coalescent allows the simulation of a limited number of demographic models, such as changes in population sizes (Slatkin 2001), population structure and migration (Hudson 1998; Arenas and Posada 2007; Kelleher et al. 2016), recombination (Hudson 1983; Arenas 2013) and selection (Kaplan et al. 1988; Arenas and Posada 2010). Since coalescent simulations are only based on the evolutionary history of a sample, the population history is not simulated and consequently these simulations are usually computationally faster than forward-time simulations. Because of that efficiency, coalescent simulations have been used to study the evolutionary history of humans with approaches based on a huge number of computer simulations (Fagundes et al. 2007; Laval et al. 2010; Arenas and Posada 2014; Kanitz et al. 2018).

Sometimes, forward-time and backward-time simulations can be complementary approaches, allowing fast simulations of complex evolutionary scenarios (Ewing and

Hermisson 2010; Nicolas Ray et al. 2010). Examples of this combination are several methods to simulate genetic data accounting for spatially explicit processes. In spatially explicit computer simulations, the first step is the simulation of the whole population (forward-time simulation) considering the demographic history of the population through temporal and spatial (geographic) models. For evolutionary studies, two-dimensional (2D) spatial features can be considered to generate models more realistic than those based on a lower number of dimensions (Duning 1995; Epperson et al. 2010), because 2D models can explicitly consider environmental changes, geographic features and range expansions, range contractions, habitat fragmentations and range shifts (Ray and Excoffier 2009; Arenas et al. 2012; Arenas, François, et al. 2013; Alves et al. 2016). Indeed, spatially explicit computer simulations allow implement complex processes of migration (i.e., anisotropic migration and long-distance dispersal migration) and admixture between populations (see for a review, Benguigui and Arenas 2014). Demographic simulations are performed under a 2D landscape that can be obtained from a Geographic Information System (GIS) framework and where the habitat is divided into a grid of demes (Bivand 2000). Next, a population expansion starts from a chosen deme under demographic parameters such as an effective population size at the onset of the expansion, a population growth rate, carrying capacity (number of individuals that each deme can sustain) and a predefined number of generations to be simulated (Currat et al. 2004). Population movements occur between demes based on a given migration rate and usually following the 2D stepping-stone migration model (Kimura and Weiss 1964) (Figure 1.5A)

After the forward in time simulation of the history of the

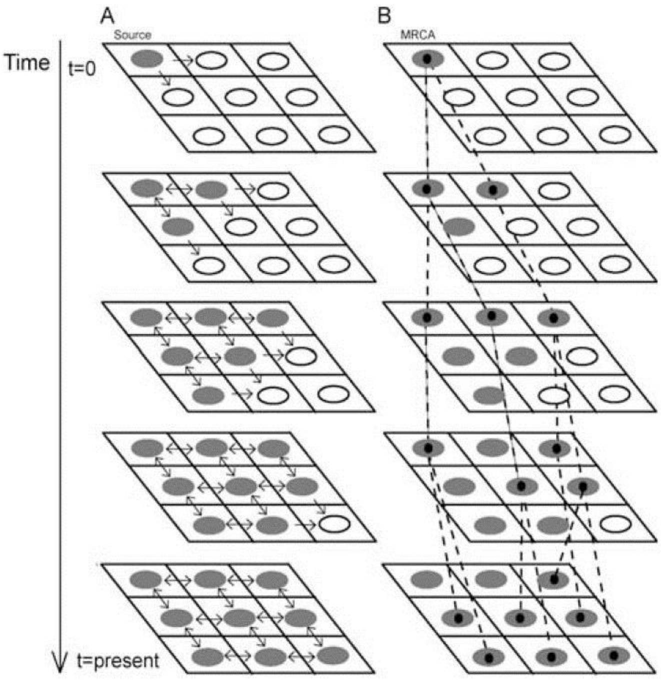


Figure 1.5. Forward in time simulations coupled with coalescent simulations. (A) Representation of the progress of the forward in time simulation through the space following a 2D stepping stone migration model. (B) Representation of the coalescent simulation to reconstruct the evolutionary history of a sample collected in the present (from Benguigui and Arenas 2014).

whole population, the backward in time (coalescent) simulation can be performed to obtain the evolutionary history of an user-specified sample (accounting for the demographic history previously simulated with the forward-time approach (Currat et al. 2004)**)** (Figure 1.5B). Finally, genetic data are simulated by assigning a random sequence to the root of the coalescent history (MRCA) and then evolving that sequence forward-time following the simulated coalescent history (Yang 2006; Arenas 2012).

1.3.1. Modelling the evolution of modern humans

By modelling evolutionary scenarios, we can obtain theoretical patterns of genetic diversity that can be evaluated with observed data. Using this modelling we can confront different evolutionary scenarios against the real data, and select which scenario is more likely to explain the observed patterns of genetic diversity. However, the presence of multiple factors acting alone or together make complex the final structure and genetic diversity of populations, and this is an important challenge to reconstruct the evolutionary history of AMH.

In the previous section, it was shown that spatially explicit computer simulations allow to generate theoretical patterns of genetic diversity under various evolutionary scenarios, which combine demographic and genetic parameters, and environmental and geographic information (Benguigui and Arenas 2014). While these models do not attempt to replicate precisely the past evolutionary history, they can incorporate factors that could have the largest influence shaping modern-day diversity patterns. There has been an increasingly use of spatially explicit computer simulators applied to evolutionary genetics. Several simulators are freely available (Benguigui and Arenas 2014). Generally, spatially explicit simulators can be divided into individual-based or deme-based models. Individual-based models are more versatile and can be more realistic than deme-based models, because, in theory, they can accommodate any possible evolutionary scenario, since the demographic and evolutionary history of each individual is simulated. By contrast, deme-based models consider groups of individuals (demes) as a unit, but allowing migration of individuals between them. The main advantage of deme-based models is their low computational costs (with respect to individual-based models).

In recent years, spatially explicit computer simulations have been applied in studies of human evolutionary history to compare alternative scenarios (e.g. range expansions, climate changes or admixture) and estimate evolutionary parameters (e.g. migration rates or interbreeding rates). Concerning the origin of AMH and the posterior world colonization, results from spatially explicit simulations favored a unique East African origin (Ray et al. 2005; Liu et al. 2006) and demonstrated that human settlements were largely affected by climatic changes during the expansion of AMH (Eriksson et al. 2012). Moreover, recently, Alves et al. (2016) found that the dispersion of AMH throughout Eurasia followed a model of long-distance dispersal (LDD) (especially towards previously occupied regions) and that this dispersion process had a larger influence on the genetic structure of Eurasian populations than the severe range contraction during the LGM.

Studies on the admixture between Neanderthal and arriving AMH showed that the proportion of interbreeding was smaller than 2%, which was attributed to a strong barrier to gene flow between both species (Currat and Excoffier 2004; 2011). Interestingly, a recent study by Kolodny and Feldman (2017) demonstrated that the replacement of Neanderthal (and posterior extinction) was mainly determined by continuous migration of AMH individuals from Africa into Eurasia, despite potential influence of selection and environmental factors in the population dynamics between both species.

Classical PCA applied to identify genetic gradients in Europe (Cavalli-Sforza and Edwards 1963; Menozzi et al. 1978; Piazza et al. 1995) showed a genetic gradient along the southeast - northwest (SE-NW) axis, that was interpreted as the result of demic diffusion of the Neolithic throughout Europe. Using spatially explicit simulations, Currat and Excoffier (2005) found that both Palaeolithic and Neolithic expansions were the cause of the genetic gradients observed in European populations, due to *allele surfing* in the front of the wave of expansion (Klopfstein et al. 2006; Hallatschek et al. 2007). However, François et al (2010), repeating these simulations with more detailed evolutionary models, demonstrated that these expansions do not necessarily lead to a PC gradient along the expansion wave. Indeed, those authors found a PC1 diversity gradient perpendicular to the direction of the dispersion axis, which was associated to *allele surfing* (Excoffier and Ray 2008). More recently, Arenas et al (2013) performed an extensive analysis on PC gradients by simulating several scenarios of admixture between Palaeolithic and Neolithic populations and range contractions towards southern Europe or towards the Iberian Peninsula during the LGM. Results showed that the PC1 gradient is perpendicular (NE-SW) to the expansion axis (SE-NW) when the expansion is recent (e.g. Neolithic) and that could be explained by *allele surfing*. By contrast, the PC1 gradient followed the direction of the expansion (SE-NW) when the expansion is old (e.g.

Palaeolithic) and that could be explained by *isolation by distance*. Scenarios of Palaeolithic populations after the LGM led to gradients perpendicular to the re-expansion axis, which was interpreted as *allele surfing*. Overall, authors demonstrated that a SE-NW gradient (gradient derived from real data) is obtained if the Neolithic had little impact on the gene pool of modern-day Europeans and/or there was a range contraction caused by the LGM that crucially affected the genetic patterns of Europe. Interestingly, a recent study by Silva et al (2017) developed a new spatially explicit method and investigated population continuity since the arrival of the Neolithic using a dataset composed of both ancient and modern individuals from Germany and France. Results showed a matrilineal continuity in French populations but not in the German populations, in line with the heterogenic Neolithic transition and the later Neolithic expansion of steppe groups from the east observed in previous studies (Lazaridis et al. 2014; Haak et al. 2015; Szécsényi-Nagy et al. 2017). Additionally, Rasteiro et al (2012) studied the role of sex-biased migration during the Neolithic transition by applying an individual-based spatially explicit simulator. Authors simulated several scenarios of the Neolithic expansion with different degrees of admixture with Palaeolithic populations and different models of post-marital residence. Results showed that the genetic diversity patterns in Europe fitted better a patrilocality model of arriving farmers than matrilocality or bilocality.

Improvements to current spatially explicit simulators should still be made to include more complex spatially explicit scenarios, such as the variation of demographic parameters through time (i.e., variable growth rate), complex migration rates (e.g. anisotropic migration, sex-biased dispersal and LDD) or complex admixture events (e.g. Neolithic transition), that can shape the patterns of genetic diversity (Ray and Excoffier 2010; Arenas et al. 2012; Rasteiro et al. 2012; Arenas, François, et al. 2013; Mona et al. 2014; Alves et al. 2016; Amorim et al. 2017; Branco et al. 2018). Moreover, the implementation of more realistic substitution models of evolution, especially to integrate the analysis of codons (Arenas and Posada 2010; 2012), proteins (Arenas, Dos Santos, et al. 2013; Arenas et al. 2015; Echave et al. 2016), genome-wide sequences (Arbiza et al. 2011; Liberles et al. 2013; Arenas and Posada 2014) and sequences of ancient individuals (Metzker 2010; Pedersen et al. 2014; Ebenesersdóttir et al. 2018; Valdiosera et al. 2018; Veeramah et al. 2018), will be important to produce more robust inferences about the human evolutionary history. In recent years, the development of analytical approaches like Approximate Bayesian Computation (ABC) (Beaumont 2010) allowed the application of computer simulations to perform model selection and parameters estimation.

1.4. Approximate Bayesian Computation (ABC) applied to human evolution

ABC is a computational methodology rooted in Bayesian statistics, and a likelihood-free approach, that has been applied in population genetics in the last 20 years. Traditionally, the likelihood function was fundamental for statistical inference in model-based approaches since it specifies the probability of the observed data given a set of values for the parameters and under a model of the underlying process. However, for complex models a Bayesian analysis may not be performed because the likelihood function cannot be designed (or its computation is intractable). The ABC methodology somehow allows bypassing the computation of the likelihood function providing a very useful alternative.

ABC is intuitively easy to understand. Generally, independent simulations based on prior information are performed to generate thousands (or millions) of simulated data. Then simulations are evaluated with respect of the genetic patterns observed in the real data, usually quantified with summary statistics (SS), through complex statistical approaches to obtain posterior distributions (estimates). ABC can be used for both model selection and parameters estimation.

1.4.1. Strategies, goals and pitfalls of Approximate Bayesian Computation

This section presents a brief description of the procedure used in ABC, including the steps of choice of SS, rejection and regression approaches, model selection and parameters estimation. Indeed, for reviews on ABC see (Beaumont 2010; Bertorelle et al. 2010; Csilléry et al. 2010; Sunnåker et al. 2013).

1.4.1.1. Choice of Summary Statistics

The ABC approach is usually based on the comparison of SS calculated from simulated data with SS calculated from observed (real) data. However, the choice of informative SS is still a rather arbitrary and complex area of ABC, since there is no a general rule about the number and features of required SS, despite the importance of this step to make inferences about evolutionary processes (Beaumont et al. 2002; Marjoram et al. 2003). The reason for this is that different datasets can be explained by different SS and, moreover, different SS can be used to perform different analysis based on the same dataset. Counterintuitively, this is an advantage of ABC, allowing a free selection of SS that can be informative for a particular dataset and for the desired estimates. As a general rule, for a consistent ABC analysis, the selected SS should contain most of the relevant information of the observed and simulated data for the desired estimations and the number of SS should be as small as possible to avoid unnecessary computation. Ideally, the posterior distribution of a parameter given the selected SS should be identical to the posterior distribution given by the dataset itself (this is, without SS). Performing inferences based on SS implies wasting some (hopefully little) useful information but considering the full data is computationally intractable.

1.4.1.2. Estimation of posterior distributions

In ABC, the posterior distribution of a parameter simulated according to a prior distribution can be estimated from the distance between SS of the simulated and observed data. Although many methods have already been developed to infer posterior distributions in ABC, the traditional rejection and regression methods remain as the most frequently used methods in studies based on ABC.

The ABC rejection approach

The rejection approach considers a small set of simulated data (retained data, simulated data with SS closer to the SS derived from the real data) to perform the estimation. Hence, all the simulated data (with SS far from the SS of the real data) are not considered for the estimation (rejected data). The proportion of retained data is given as an input. Because of the criterion of retained/rejected simulated data, this approach was called as a rejection approach.

The ABC regression approach

Beaumont et al (2002) inserted the regression step in the rejection algorithm. The regression step was inserted in the ABC as an attempt to mitigate the bias introduced by a non-zero threshold, reducing the variance of the posterior estimates. The slope of the regression line between a parameter and the vector of SS is estimated using only retained simulations and assigning more weight to those simulations that generate SS closer to the observed data (Figure 1.6). The regression step is used to correct sampled parameters in the direction of the observed data SS, reducing the difference between simulated and observed SS and, therefore, requiring fewer simulations. Nevertheless, some problems can arise with regression methods, especially if a model is misspecified, since adjusted SS could fall outside the prior distribution, causing an extrapolation rather than an interpolation of the posterior distribution (Beaumont et al. 2002).

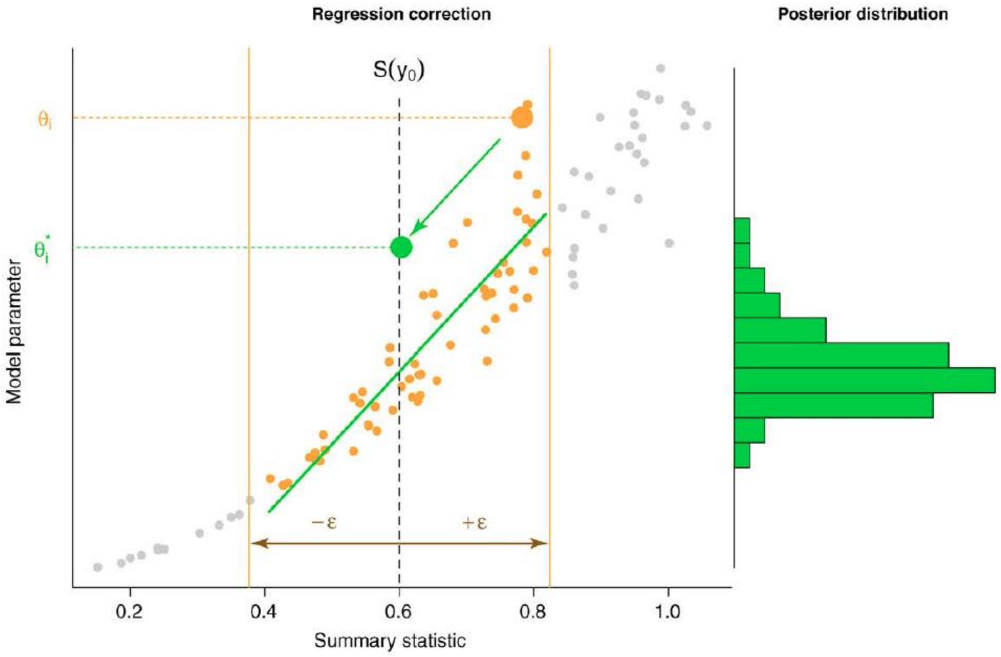


Figure 1.6. Representation of the linear regression adjustment in the ABC framework**.** Initially, a parameter value is sampled from the prior distribution, to simulate a dataset under a given model. From the simulated dataset, a SS is estimated and then compared with the SS for the observed data, using an error threshold. A simulation is accepted if it falls within the define threshold (vertical orange lines in the main plot). Then, parameters values from the accepted simulations are adjusted using a linear regression model (green line) and the posterior distribution is estimated using these adjusted parameter values (histogram on the right side) (from Csilléry et al 2010)

Recently, Blum and François (2010) suggested an improvement to the regression-based methods by introducing a nonlinear regression based on a feed-forward neural networks model. Indeed, Leuenberger and Wegmann (2010) reformulated the regression method with a general linear model (GLM). This method assumes that SS are endogenous response variables created within the model, while the regression method (Beaumont et al. 2002) considers SS as explanatory variables of the model. GLM never produces posterior distributions outside the prior, but assumes a normal distribution for the prior and it is computationally more expensive than the regression approach (Bertorelle et al. 2010).

1.4.1.3. Applying ABC for selection among alternative scenarios

An important application of the ABC approach is the selection among alternative scenarios, which encloses the comparison of different evolutionary scenarios and the process of inference of the goodness-of-fit between the different scenarios and the observed data. Generally, the different evolutionary scenarios are run for the same number of simulations, so that the amount of sampling from the prior probability for each scenario is equal, and then an ABC algorithm is used to infer the scenario with the highest posterior probability with the real data.

Pritchard et al. (1999) describe the use of ABC to estimate posterior probabilities for distinguishing between alternative evolutionary scenarios. Performing the ABC rejection algorithm, the probability of a given scenario is proportional to the frequency of simulations that fall within a threshold region (Pritchard et al. 1999). However, here it is also possible to introduce a regression approach to improve the rejection approach (Fagundes et al. 2007; Beaumont 2008). This approach is used to obtain the relation between categorical variables under a multinomial logistic regression, estimating the posterior probability of a scenario as a function of the accepted SS (Fagundes et al. 2007; Beaumont 2008; Cornuet et al. 2008).

1.4.1.4. Applying ABC for parameters estimation

Parameter estimation is usually the last step of an ABC analysis and it is performed for the simulated scenario with the highest posterior probability. As mentioned previously, the parameter values of the accepted (retained) simulations are used to estimate the posterior distribution of the parameter values corresponding to the real data. This estimation can be performed and evaluated by several methods.

Similarly to model selection, the simplest method to infer the posterior distribution of a parameter is the rejection method, since retained simulations constitute a sample of the posterior distribution of the observed parameter (Pritchard et al. 1999). However, if too many simulations are retained due to a highly permissive threshold, the estimated posterior will be similar to the prior distribution, which is disadvantageous for parameter estimation. On the other hand, if the threshold is very restrictive (i.e. approximates to 0), the direct approach rebuilds accurately the posterior distribution of the parameter, but a very large number of simulations are needed, which is computationally expensive. As mentioned in Section 1.4.1.2, Beaumont et al (2002) introduced a regression based approach, in which parameter estimation could be improved without increasing the number of simulations using a local linear regression model. Improving the accuracy of parameter estimation is a field of ABC that has received major attention from geneticists and mathematicians in recent years, and so new methods are continuously proposed (e.g. Leuenberger and Wegmann 2010; Blum and François 2010; Blum et al. 2013).

1.4.2. Examples of approximate Bayesian computation applied to analyse human evolution

ABC has been widely applied to the study of the human evolutionary history. The ability to accurately choose between several evolutionary scenarios and to estimate parameter values using ABC methods is evolving rapidly as new techniques are developed (Arenas 2015; Sisson et al. unpublished data). Below I briefly describe some relevant examples of application of ABC to decipher the human evolutionary history.

In a landmark study, Fagundes et al (2007) applied an ABC approach using extensive computer simulations to evaluate the likelihood of different scenarios for the origin of AMH. Authors showed that a scenario of single sub-Saharan origin with gradual expansion and posterior replacement of archaic humans had a better fitting with the observed data than a multiregional or assimilation scenario. Moreover, the origin of AMH was estimated to be around 141,000 ya, the out-of-Africa around 51,000 ya and the colonization of America around 10,500 ya. Posterior studies addressed the expansion patterns of AMH across the globe.

Eriksson et al (2012) constructed an scenario with climate conditions over the last 120,000 years, to evaluate the effect of climate on the modern patterns of genetic diversity. Results showed a clear correlation between climate and population demography of modern hunter-gatherer populations worldwide, and the estimated arrival times of AMH to different continents were consistent with the archaeological findings.

More recently, Alves et al (2016) analysing the expansion of AMH across Eurasia, demonstrated that the real data supported better scenarios with LDD events that models without it. Interestingly, results also showed an absence of LDD events ahead of the front wave expansion (this is, to places not yet colonized) during the settlement of Eurasia but demonstrated that these events were crucial for modulating the genetic structure of modern Eurasian populations.

In 2010, Ray et al (2010) performed an ABC analysis to evaluate alternative models of settlement of the American Continent, showing that a single wave and two wave scenarios were not supported and that recurrent gene flow after the initial settlement between Asia and America fitted better the observed data. Moreover, authors estimated that the initial colonization occurred 13,000 ya and involved a small number of individuals.

The Neolithic transition in Europe was also extensively studied using the ABC approach. Initial results supported a demic diffusion model for the expansion of the

Neolithic when analysing early Central European farmers (Haak et al. 2010). This genetic discontinuity between Mesolithic and Neolithic groups was further supported by posterior results for the Iberian Peninsula (Gamba et al. 2012; Sánchez-Quinto et al. 2012) and the Sardinia (Modi et al. 2017). In an interesting study by Tassi et al (2017) concerning the late Neolithic Indo-European culture (Globular Amphorae culture - GAC) from Central Europe, in which several different scenarios of migration from the Pontic steppe were tested, results showed that individuals from GAC did not have special affinity with steppe related individuals in contrast with individuals from the Corded Ware culture. To date Vallée et al (2016) performed the only study concerning the Neolithic expansion outside Europe. Authors constructed a very complex model to simulate the Neolithic expansion and the mobility between islands in southeast Asia. Results showed that the dispersion of Asian ancestry individuals throughout the island system was correlated with higher migration rates and birth rates than Papuan related individuals were correlated with the expansion of mainland Asian ancestry, and that mainland Asian ancestry was already present in western Island Southeast Asia before the Neolithic expansion.

The ABC framework has been also applied to estimate divergence times and the origin of the structure of modern population groups. In 2010, Ghirotto et al (2010) reconstructed the demographic history of Sardinia by comparing aDNA and modern human DNA from two regions (Ogliastra and Gallura) under a series of demographic scenarios with an ABC approach. They demonstrated a genetic continuity since the Bronze Age in the Ogliastra region but not in the Gallura. Indeed, for this latter region, evidence showed a larger effective and growth rate than Ogliastra region, which was partially attributed to gene flow from the mainland Italy. In 2012, Veeramah et al (2012) applied ABC to characterize the origin of modern sub-Saharan populations by building several models of divergence of the ancestors of modern populations. Results showed that the Khoisan population diverge first from the ancestor of Pygmies groups around 110,000 ya, while the Pygmy divergence occurred around 50,000 ya.

Altogether, these studies show the potential of the ABC framework to shed light into complex human evolutionary questions. The advent of next generation sequencing technologies is producing large amounts of genome-wide data that can be highly informative about the evolutionary history (Excoffier et al. 2013) and would make possible to build more realistic analyses of human populations. Indeed, ABC is a simplistic but highly efficient approach to answer complex human evolutionary questions and constitutes an exciting field of population genetics in years to come.

1.5. Objectives

The Mediterranean Basin is a region with a complex demographic history and an intricate population dynamics during and after the uprising of the Neolithic. While the transition to farming is well described for European populations, archaeological and genetic data for North Africa is still scarce, resulting in an incomplete and incongruent story for the Neolithic transition in the region. Additionally, despite some results have shown some degree of contact between populations from the Iberian Peninsula and North Africa across the Strait of Gibraltar, the population dynamics between both regions is still poorly understood. Consequences of migrations between regions are imprinted in the genetic landscape of populations, especially in the Iberian Peninsula, which had most of its area under Islamic administration less than a 1,000 ya. In order to shed light on the causes and genetic consequences of the evolutionary history of modern humans of the Mediterranean Basin the present Thesis work aims to:

- Evaluate the Neolithic expansion over North Africa with respect to the Neolithic expansion over Europe, especially in terms of speed of expansion, migration rate and carrying capacity.
- Characterize the amount, direction and implication of migration across the Strait of Gibraltar since the Neolithic expansion at both shores of the Mediterranean Sea.
- Investigate the genetic structure of current humans of the Iberian Peninsula at fine-scale and interpret it concerning geography and historical events.

Chapter 2 presents the analysis of Neolithic migration between the Iberian Peninsula and North Africa through the Strait of Gibraltar and the evaluation of the Neolithic transition in North Africa (for which very little was known) with respect to the Neolithic transition in Europe. This study was performed applying an ABC method.

Chapter 3 presents a fine-scale analysis of the patterns of genetic differentiation throughout the Iberian Peninsula landscape. Additionally, this chapter includes the analysis of possible sex biases by comparing patterns estimated from autosomes data with patterns estimated from X chromosome data. This study applied a georeferenced dataset and recently developed methods oriented to analyse spatial patterns of genetic differentiation.

1.6. References

Adams SM, Bosch E, Balaresque PL, Ballereau SJ, Lee AC, Arroyo E, López-Parra AM, Aler M, Grifo MSG, Brion M, 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.

Allentoft ME, Sikora M, Sjögren KG, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L, et al. 2015. Population genomics of Bronze Age Eurasia. Nature 522:167–172.

Alt KW, Zesch S, Garrido-Pena R, Knipper C, Szécsényi-Nagy A, Roth C, Tejedor-Rodríguez C, Held P, García-Martínez-De-Lagrán Í, Navitainuck D, et al. 2016. A community in life and death: The late neolithic megalithic tomb at Alto de Reinoso (Burgos, Spain). PLoS One 11.

Alvarez L, Santos C, Ramos A, Pratdesaba R, Francalacci P, Aluja MP. 2010. Mitochondrial DNA patterns in the Iberian Northern plateau: Population dynamics and substructure of the Zamora province. Am. J. Phys. Anthropol. 142:531–539.

Alves I, Arenas M, Currat M, Hanulova AS, Sousa VC, Ray N, Excoffier L. 2016. Long-distance dispersal shaped patterns of human genetic diversity in Eurasia. Mol. Biol. Evol. 33:946–958.

Ammerman AJ, Cavalli-Sforza L. 1984. The Neolithic Transition and the Genetics of Populations in Europe. (Press PU, editor.). Princeton University Press

Ammerman AJ, Cavalli-Sforza LL. 1971. Measuring the Rate of Spread of Early Farming in Europe. Man 6:674.

Ammerman AJ, Cavalli-Sforza LL. 1973. A population model for the diffusion of early farming in Europe. The Explanation of Culture Change. (Refrew C, editor.). Duckworth, London

Amorim CEG, Hofer T, Ray N, Foll M, Ruiz-Linares A, Excoffier L. 2017. Long-distance dispersal suppresses introgression of local alleles during range expansions. Heredity (Edinb). 118:135–142.

Arauna LR, Mendoza-Revilla J, Mas-Sandoval A, Izaabel H, Bekada A, Benhamamouch S, Fadhlaoui-Zid K, Zalloua P, Hellenthal G, Comas D, et al. 2017. Recent historical migrations have shaped the gene pool of Arabs and Berbers in North Africa. Mol. Biol. Evol. 34:318–329.

Arbiza L, Patricio M, Dopazo H, Posada D. 2011. Genome-wide heterogeneity of nucleotide substitution model fit. Genome Biol. Evol. 3:896–908.

Arenas M. 2012. Simulation of molecular data under diverse evolutionary scenarios. PLoS Comput. Biol. 8.

Arenas M. 2013. The importance and application of the ancestral recombination graph. Front. Genet. 4.

Arenas M. 2015. Advances in Computer Simulation of Genome Evolution: Toward More Realistic Evolutionary Genomics Analysis by Approximate Bayesian Computation. J. Mol. Evol. 80:189–192.

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.

Arenas M, Posada D. 2007. Recodon: Coalescent simulation of coding DNA sequences with recombination, migration and demography. BMC Bioinformatics 8.

Arenas M, Posada D. 2010. Coalescent simulation of intracodon recombination. Genetics 184:429–437.

Arenas M, Posada D. 2012. Simulation of coding sequence evolution. In: Codon Evolution. Oxford University Press. p. 126–132.

Arenas M, Posada D. 2014. Simulation of genome-wide evolution under heterogeneous substitution models and complex multispecies coalescent histories. Mol. Biol. Evol. 31:1295–1301.

Arenas M, Ray N, Currat M, Excoffier L. 2012. Consequences of range contractions and range shifts on molecular diversity. Mol. Biol. Evol. 29:207–218.

Arenas M, Sánchez-Cobos A, Bastolla U. 2015. Maximum-likelihood phylogenetic inference with selection on protein folding stability. Mol. Biol. Evol. 32:2195–2207.

Arenas M, Dos Santos HG, Posada D, Bastolla U. 2013. Protein evolution along phylogenetic histories under structurally constrained substitution models. Bioinformatics 29:3020–3028.

Arenas M, Valiente G, Posada D. 2008. Characterization of reticulate networks based on the coalescent with recombination. Mol. Biol. Evol. 25:2517–2520.

Armelagos GJ, Harper KN. 2005. Genomics at the origins of agriculture, part one. Evol. Anthropol. 14:109–121.

Arranz-Otaegui A, Colledge S, Zapata L, Teira-Mayolini LC, Ibáñez JJ. 2016. Regional diversity on the timing for the initial appearance of cereal cultivation and domestication in southwest Asia. Proc. Natl. Acad. Sci. 113:14001–14006.

Asouti E, Fuller DQ. 2013. A Contextual Approach to the Emergence of Agriculture in Southwest Asia: reconstructing Early Neolithic Plant-Food Production. Curr. Anthropol. A world J. Sci. man 54:299–345.

Baker JL, Rotimi CN, Shriner D. 2017. Human ancestry correlates with language and reveals that race is not an objective genomic classifier. Sci. Rep. 7.

Ballouche A, Marinval P. 2003. Données palynologiques et carpologiques sur la domestication des plantes et l´agriculture dans le Néolthique ancient du Maroc septentrional. (Site de Kaf Taht El-Ghar). Rev. d'Archéométrie 27:49–54.

Barich BE. 2010. Antica Africa: alle origine delle società. Roma: L'Erma di Bretschneider

Barich BE. 2014. Northwest Libya from the early to late Holocene: New data on environment and subsistence from the Jebel Gharbi. Quat. Int. 320:15–27.

Barton N, Bouzouggar A, Humphrey L, Berridge P, Collcutt S, Gale R, Parfitt S, Parker A, Rhodes E, Schwenninger JL. 2008. Human burial evidence from Hattab II Cave and the question of continuity in late Pleistocene-Holocene mortuary practices in northwest Africa. Cambridge Archaeol. J. 18:195–214.

Beaumont M. 2010. Approximate Bayesian computation in evolution and ecology. Annu. Rev. Ecol. Evol. Syst. 41:379–406.

Beaumont MA. 2008. Joint determination of topology, divergence time, and immigration in population trees. In: Simulations, Genetics and Human Prehistory. p. 135–154.

Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. Genetics 162:2025–2035.

Behar DM, Harmant C, Manry J, Van Oven M, Haak W, Martinez-Cruz B, Salaberria J, Oyharabal B, Bauduer F, Comas D, 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.

Benazzi S, Douka K, Fornai C, Bauer CC, Kullmer O, Svoboda J, Pap I, Mallegni F, Bayle P, Coquerelle M, et al. 2011. Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. Nature 479:525–528.

Benguigui M, Arenas M. 2014. Spatial and temporal simulation of human evolution. Methods, frameworks and applications. Curr. Genomics 15:245–255.

Bernabeu Aubán J, Manen C, Pardo Gordó S. 2017. Spatial and Temporal Diversity During the Neolithic Spread in the Western Mediterranean: The First Pottery Productions. In: García-Puchol O, Salazar-García DC, editors. Times of

Neolithic Transition along the Western Mediterranean. p. 373–398.

Bernabeu Aubán J, Michael Barton C, Pardo Gordó S, Bergin SM. 2015. Modeling initial Neolithic dispersal. The first agricultural groups in West Mediterranean. Ecol. Modell. 307:22–31.

Bertorelle G, Benazzo A, Mona S. 2010. ABC as a flexible framework to estimate demography over space and time: Some cons, many pros. Mol. Ecol. 19:2609–2625.

Biagi P. 2003. A review of the Late Mesolithic in Italy and its implications for the Neolithic transition. In: Ammerman AJ, Biagi P, editors. The widening harvest. The Neolithic transition in Europe: looking back, looking forward. Archaeological Institute of America. p. 133–155.

Bivand RS. 2000. Using the R statistical data analysis language on GRASS 5.0 GIS database files. Comput. Geosci. 26:1043–1052.

Blum MGB, François O. 2010. Non-linear regression models for Approximate Bayesian Computation. Stat. Comput. 20:63–73.

Blum MGB, Nunes MASN, Prangle D, Sisson S a. 2013. A Comparative Review of Dimension Reduction Methods in Approximate Bayesian Computation. Stat. Sci. 28:189–208.

Bocquet-Appel J-P. 2008. The Neolithic demographic transition, population pressure and cultural change. Comp. Civilizations Rev.:36–49.

Bocquet-Appel J-P, Naji S. 2006. Testing the Hypothesis of a Worldwide Neolithic Demographic Transition. Curr. Anthropol. 47:341–365.

Bocquet-Appel JP. 2011. When the world's population took off: The springboard of the neolithic demographic transition. Science (80-.). 333:560–561.

Bocquet-Appel JP, Naji S, Vander Linden M, Kozlowski J. 2012. Understanding the rates of expansion of the farming system in Europe. J. Archaeol. Sci. 39:531–546.

Bocquet-Appel JP, Naji S, Linden M Vander, Kozlowski JK. 2009. Detection of diffusion and contact zones of early farming in Europe from the space-time distribution of 14C dates. J. Archaeol. Sci. 36:807–820.

Bocquet-Appel J. 2002. Paleoanthropological Traces of a Neolithic Demographic Transition. Curr. Anthropol. 43:637–650.

Botigue LR, Henn BM, Gravel S, Maples BK, Gignoux CR, Corona E, Atzmon G, Burns E, Ostrer H, Flores C, et al. 2013. Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. Proc. Natl. Acad. Sci. 110:11791–11796.

Botigué LR, Song S, Scheu A, Gopalan S, Pendleton AL, Oetjens M, Taravella AM, Seregély T, Zeeb-Lanz A, Arbogast RM, et al. 2017. Ancient European dog genomes reveal continuity since the Early Neolithic. Nat. Commun. 8.

Bradburd GS, Ralph PL, Coop GM. 2016. A Spatial Framework for Understanding Population Structure and Admixture. PLoS Genet. 12.

Branco C, Velasco M, Benguigui M, Currat M, Ray N, Arenas M. 2018. Consequences of diverse evolutionary processes on american genetic gradients of modern humans. Heredity (Edinb).

Broushaki F, Thomas MG, Link V, López S, van Dorp L, Kirsanow K, Hofmanová Z, Diekmann Y, Cassidy LM, Díez-del-Molino D, et al. 2016. Early Neolithic genomes from the eastern Fertile Crescent. Science (80-.). 353:499–503.

Busby GB, Band G, Si Le Q, Jallow M, Bougama E, Mangano VD, Amenga-Etego LN, Enimil A, Apinjoh T, Ndila CM, et al. 2016. Admixture into and within sub-Saharan Africa. Elife 5.

Busby GBJ, Hellenthal G, Montinaro F, Tofanelli S, Bulayeva K, Rudan I, Zemunik T, Hayward C, Toncheva D, Karachanak-Yankova S, et al. 2015. The role of recent admixture in forming the contemporary West Eurasian genomic landscape. Curr. Biol. 25:2518–2526.

Bycroft C, Fernández-Rozadilla C, Ruiz-Ponte C, Quintela-García I, Carracedo Á, Donnelly P, Myers S. unpublished data. https://www.biorxiv.org/content/early/2018/03/12/250191. last accessed June 5, 2018.

Cardoso JL. 2014. Absolute chronology of the Beaker phenomenon North of the Tagus estuary: demographic and social implications. Trab. Prehist. 71:56–75.

Carr R. 2000. Spain: a history. Oxford University Press

Carvajal-Rodriguez A. 2008. Simulation of Genomes: A Review. Curr. Genomics 9:155–159.

Carvajal-Rodriguez A. 2010. Simulation of Genes and Genomes Forward in Time. Curr. Genomics 11:58–61.

Castaneda IS, Mulitza S, Schefuss E, dos Santos RAL, Damste JSS, Schouten S. 2009. Wet phases in the Sahara/Sahel region and human migration patterns in North Africa. Proc. Natl. Acad. Sci. 106:20159–20163.

Cavalli-Sforza LL, Edwards AWF. 1963. Analysis of human evolution. In: Geerts SJ, editor. Genetics Today; Proceedings of the 11th International Congress of Genetics. Pergamon Press. p. 923–993.

Chandler HUDO (Reino U, Sykes BUDO (Reino U, Zilhao JUDL (Portugal). 2005.

Using ancient DNA to examine genetic continuity at the Mesolithic-Neolithic transition in Portugal. Actas del III Congr. del Neolítico en la Península Ibérica:781–786.

Cohen MN, Armelagos GJ. 1984. Paleopathology at the origins of agriculture. Academic Press

Coppa A, Cucina A, Lucci M, Mancinelli D, Vargiu R. 2007. Origins and spread of agriculture in Italy: A nonmetric dental analysis. Am. J. Phys. Anthropol. 133:918–930.

Cornuet JM, Santos F, Beaumont MA, Robert CP, Marin JM, Balding DJ, Guillemaud T, Estoup A. 2008. Inferring population history with DIY ABC: A user-friendly approach to approximate Bayesian computation. Bioinformatics 24:2713–2719.

Csilléry K, Blum MGB, Gaggiotti OE, François O. 2010. Approximate Bayesian Computation (ABC) in practice. Trends Ecol. Evol. 25:410–418.

Currat M, Currat M, Nunes J, Sanchez-Mazas A, Gerbault P, Di D. 2016. Forward-in-Time, Spatially Explicit Modeling Software to Simulate Genetic Lineages Under Selection. Evol. Bioinforma. 11:27.

Currat M, Excoffier L. 2004. Modern humans did not admix with Neanderthals during their range expansion into Europe. PLoS Biol. 2.

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, Excoffier L. 2011. Strong reproductive isolation between humans and Neanderthals inferred from observed patterns of introgression. Proc. Natl. Acad. Sci. 108:15129–15134.

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. SPLATCHE: A program to simulate genetic diversity taking into account environmental heterogeneity. Mol. Ecol. Notes 4:139–142.

Czebreszuk J. 2004. Similar But Different. Bell Beakers in Europe. Adam Mickiewicz University

Diamond J, Bellwood P. 2003. Farmers and their languages: The first expansions. Science (80-.). 300:597–603.

Drake NA, Breeze P, Parker A. 2013. Palaeoclimate in the Saharan and Arabian Deserts during the Middle Palaeolithic and the potential for hominin dispersals. Quat. Int. 300:48–61.

Drake N, Breeze P. 2016. Climate change and modern human occupation of the

Sahara from MIS 6-2. In: Jones SC, Stewart BA, editors. Africa from MIS 6-2. Dordrecht: Springer. p. 103–122.

Duforet-Frebourg N, Blum MGB. 2014. Nonstationary patterns of isolation-by-distance: Inferring measures of local genetic differentiation with bayesian kriging. Evolution (N. Y). 68:1110–1123.

Duning JB. 1995. Spatially explicit population models: current forms and future uses. Ecol. Appl. 5:3–11.

Ebenesersdóttir SS, Sandoval-Velasco M, Gunnarsdóttir ED, Jagadeesan A, Guðmundsdóttir VB, Thordardóttir EL, Einarsdóttir MS, Moore KHS, Sigurðsson Á, Magnúsdóttir DN, et al. 2018. Ancient genomes from Iceland reveal the making of a human population. Science 360:1028–1032.

Echave J, Spielman SJ, Wilke CO. 2016. Causes of evolutionary rate variation among protein sites. Nat. Rev. Genet. 17:109–121.

Edmonson MS. 1961. Neolithic Diffusion Rates. Curr. Anthropol. 2:71–102.

Epperson BK, McRae BH, Scribner K, Cushman SA, Rosenberg MS, Fortin MJ, James PMA, Murphy M, Manel S, Legendre P, et al. 2010. Utility of computer simulations in landscape genetics. Mol. Ecol. 19:3549–3564.

Eriksson A, Betti L, Friend AD, Lycett SJ, Singarayer JS, von Cramon-Taubadel N, Valdes PJ, Balloux F, Manica A. 2012. Late Pleistocene climate change and the global expansion of anatomically modern humans. Proc. Natl. Acad. Sci. U. S. A. 109:16089–16094.

Ewing G, Hermisson J. 2010. MSMS: A coalescent simulation program including recombination, demographic structure and selection at a single locus. Bioinformatics 26:2064–2065.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust Demographic Inference from Genomic and SNP Data. PLoS Genet. 9.

Excoffier L, Ray N. 2008. Surfing during population expansions promotes genetic revolutions and structuration. Trends Ecol. Evol. 23:347–351.

Fadhlaoui-Zid K, Haber M, Martínez-Cruz B, Zalloua P, Elgaaied AB, Comas D. 2013. Genome-wide and paternal diversity reveal a recent origin of human populations in north africa. PLoS One 8.

Fagundes NJR, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L. 2007. Statistical evaluation of alternative models of human evolution. Proc. Natl. Acad. Sci. 104:17614–17619.

Fernández E, Pérez-Pérez A, Gamba C, Prats E, Cuesta P, Anfruns J, Molist M, Arroyo-Pardo E, Turbón D. 2014. Ancient DNA Analysis of 8000 B.C. Near

Eastern Farmers Supports an Early Neolithic Pioneer Maritime Colonization of Mainland Europe through Cyprus and the Aegean Islands. PLoS Genet. 10.

de Filippo C, Bostoen K, Stoneking M, Pakendorf B. 2012. Bringing together linguistic and genetic evidence to test the Bantu expansion. Proc. R. Soc. B Biol. Sci. 279:3256–3263.

Finocchio A, Trombetta B, Messina F, D'Atanasio E, Akar N, Loutradis A, Michalodimitrakis EI, Cruciani F, Novelletto A. 2018. A finely resolved phylogeny of Y chromosome Hg J illuminates the processes of Phoenician and Greek colonizations in the Mediterranean. Sci. Rep. 8:7465.

Font-Porterias N, Solé-Morata N, Serra-Vidal G, Bekada A, Fadhlaoui-Zid K, Zalloua P, Calafell F, Comas D. 2018. The genetic landscape of Mediterranean North African populations through complete mtDNA sequences. Ann. Hum. Biol. 45:98–104.

Fort J. 2012. Synthesis between demic and cultural diffusion in the Neolithic transition in Europe. Proc. Natl. Acad. Sci. 109:18669–18673.

Fort J. 2015. Demic and cultural diffusion propagated the Neolithic transition across different regions of Europe. J. R. Soc. Interface 12:20150166–20150166.

François O, Currat M, Ray N, Han E, Excoffier L, Novembre J. 2010. Principal component analysis under population genetic models of range expansion and admixture. Mol. Biol. Evol. 27:1257–1268.

Fregel R, Mendez FL, Bokbot Y, Martin-Socas D, Camalich-Massieu MD, Avila-Arcos MC, Underhill PA, Shapiro B, Wojcik GL, Rasmussen M, et al. 2018. Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. Proc. Natl. Acad. Sci.:191569.

Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, Patterson N, Rohland N, Lazaridis I, Nickel B, et al. 2015. An early modern human from Romania with a recent Neanderthal ancestor. Nature 524:216–219.

Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, Johnson PLF, Aximu-Petri A, Prüfer K, De Filippo C, et al. 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature 514:445–449.

Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, Furtwängler A, Haak W, Meyer M, Mittnik A, et al. 2016. The genetic history of Ice Age Europe. Nature 534:200–205.

Gamba C, FernÚndez E, Tirado M, Deguilloux MF, Pemonge MH, Utrilla P, Edo M, Molist M, Rasteiro R, Chikhi L, et al. 2012. Ancient DNA from an Early Neolithic Iberian population supports a pioneer colonization by first farmers. Mol. Ecol.

21:45–56.

Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, Domboróczki L, Kővári I, Pap I, Anders A, et al. 2014. Genome flux and stasis in a five millennium transect of European prehistory. Nat. Commun. 5:5257.

Garagnani P, Laayouni H, González-Neira A, Sikora M, Luiselli D, Bertranpetit J, Calafell F. 2009. Isolated populations as treasure troves in genetic epidemiology: The case of the Basques. Eur. J. Hum. Genet. 17:1490–1494.

Garcea EAA. 2004. An alternative way towards food production: The perspective from the Libyan Sahara. J. World Prehistory 18:107–154.

Gautier A. 2014. Animal remains from the hidden Valley Neolithic site, Farafra oasis. In: Barich BE, Lucarini G, Hamdan MA, Hassan FA, editors. From Lake to Sand. The Archaeology of Farafra Oasis, Western Desert, Egypt. Firenze: All'Insegna del giglio. p. 369–376.

Gayan J, Galan J, Gonzalez-Perez A, Saez M, Martinez-Larrad M, Zabena C, Rivero M, Salinas A, Ramirez-Lorca R, Moron F, et al. 2010. Genetic structure of the Spanish population. BMC Genomics 11:326.

Ghirotto S, Mona S, Benazzo A, Paparazzo F, Caramelli D, Barbujani G. 2010. Inferring genealogical processes from patterns of bronze-age and modern DNA variation in Sardinia. Mol. Biol. Evol. 27:875–886.

Gonzalez AM, Brehm A, Pérez JA, Maca-Meyer N, Flores C, Cabrera VM. 2003. Mitochondrial DNA affinities at the atlantic fringe of Europe. Am. J. Phys. Anthropol. 120:391–404.

Günther T, Malmström H, Svensson EM, Omrak A, Sánchez-Quinto F, Kılınç GM, Krzewińska M, Eriksson G, Fraser M, Edlund H, et al. 2018. Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. PLoS Biol. 16.

Günther T, Valdiosera C, Malmström H, Ureña I, Rodriguez-Varela R, Sverrisdóttir ÓO, Daskalaki EA, Skoglund P, Naidoo T, Svensson EM, et al. 2015. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. Proc. Natl. Acad. Sci. 112:11917–11922.

Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, Adler CJ, der Sarkissian CSI, Brandt G, Schwarz C, Nicklisch N, et al. 2010. Ancient DNA from European early Neolithic farmers reveals their near eastern affinities. PLoS Biol. 8.

Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K, et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522:207–211.

Hallatschek O, Hersen P, Ramanathan S, Nelson DR. 2007. Genetic drift at expanding frontiers promotes gene segregation. Proc. Natl. Acad. Sci. 104:19926–19930.

Hanks EM, Hooten MB. 2013. Circuit theory and model-based inference for landscape connectivity. J. Am. Stat. Assoc. 108:22–33.

Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, Myers S. 2014. A genetic atlas of human admixture history. Science (80-.). 343:747–751.

Henn BM, Botigué LR, Gravel S, Wang W, Brisbin A, Byrnes JK, Fadhlaoui-Zid K, Zalloua PA, Moreno-Estrada A, Bertranpetit J, et al. 2012. Genomic ancestry of North Africans supports back-to-Africa migrations. PLoS Genet. 8.

Hermanussen M. 2003. Stature of early Europeans. Hormones 2:175–178.

Hernández CL, Calderón R. 2017. Matrilineal Heritage in Southern Iberia Reveals Deep Genetic Links between Continents. Coll Antropol 41:1–10.

Hervella M, Izagirre N, Alonso S, Fregel R, Alonso A, Cabrera VM, de la Rúa C. 2012. Ancient DNA from Hunter-Gatherer and Farmer Groups from Northern Spain Supports a Random Dispersion Model for the Neolithic Expansion into Europe. PLoS One 7:e34417.

Hollfelder N, Schlebusch CM, Günther T, Babiker H, Hassan HY, Jakobsson M. 2017. Northeast African genomic variation shaped by the continuity of indigenous groups and Eurasian migrations. PLoS Genet. 13.

Hudson RR. 1983. Properties of a neutral allele model with intragenic recombination. Theor. Popul. Biol. 23:183–201.

Hudson RR. 1998. Island models and the coalescent process. Mol. Ecol. 7:413–418.

Hurles ME, Veitia R, Arroyo E, Armenteros M, Bertranpetit J, Pérez-Lezaun a, Bosch E, Shlumukova M, Cambon-Thomsen a, McElreavey K, et al. 1999. Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. Am. J. Hum. Genet. 65:1437–1448.

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.

Isern N, Zilhão J, Fort J, Ammerman AJ. 2017. Modeling the role of voyaging in the coastal spread of the Early Neolithic in the West Mediterranean. Proc. Natl.

Acad. Sci. 114:897–902.

Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung HC, Szpiech ZA, Degnan JH, Wang K, Guerreiro R, et al. 2008. Genotype, haplotype and copy-number variation in worldwide human populations. Nature 451:998–1003.

Jay F, Sjödin P, Jakobsson M, Blum MGB. 2013. Anisotropic isolation by distance: The main orientations of human genetic differentiation. Mol. Biol. Evol. 30:513–525.

Jeunesse C. 2015. The dogma of the Iberian origin of the Bell Beaker: attempting its deconstruction. J. Neolit. Archaeol. 16:158–166.

Jones ER, Gonzalez-Fortes G, Connell S, Siska V, Eriksson A, Martiniano R, McLaughlin RL, Gallego Llorente M, Cassidy LM, Gamba C, et al. 2015. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat. Commun. 6.

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.

Kaplan NL, Darden T, Hudson RR. 1988. The coalescent process in models with selection. Genetics 120:819–829.

Kelleher J, Etheridge AM, McVean G. 2016. Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. PLoS Comput. Biol. 12.

Kimura M, Weiss GH. 1964. The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance. Genetics 49:561–576.

Kingman JFC. 1982. The coalescent. Stoch. Process. their Appl. 13:235–248.

Kılınç GM, Koptekin D, Atakuman Ç, Sümer AP, Dönertaş HM, Yaka R, Bilgin CC, Büyükkarakaya AM, Baird D, Altınışık E, et al. 2017. Archaeogenomic analysis of the first steps of Neolithization in Anatolia and the Aegean. Proc. R. Soc. B Biol. Sci. 284.

Kılınç GM, Omrak A, Özer F, Günther T, Büyükkarakaya AM, Bıçakçı E, Baird D, Dönertaş HM, Ghalichi A, Yaka R, et al. 2016. The Demographic Development of the First Farmers in Anatolia. Curr. Biol. 26:2659–2666.

Klopfstein S, Currat M, Excoffier L. 2006. The fate of mutations surfing on the wave of a range expansion. Mol. Biol. Evol. 23:482–490.

Kolodny O, Feldman MW. 2017. A parsimonious neutral model suggests Neanderthal replacement was determined by migration and random species drift. Nat. Commun. 8.

Laayouni H, Calafell F, Bertranpetit J. 2010. A genome-wide survey does not show the genetic distinctiveness of Basques. Hum. Genet. 127:455–458.

Lacan M, Keyser C, Ricaut F-X, Brucato N, Duranthon F, Guilaine J, Crubezy E, Ludes B. 2011. Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. Proc. Natl. Acad. Sci. 108:9788–9791.

Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, Caliebe A, Balascakova M, Bertranpetit J, Bindoff LA, Comas D, et al. 2008. Correlation between Genetic and Geographic Structure in Europe. Curr. Biol. 18:1241–1248.

Laval G, Patin E, Barreiro LB, Murci LQ. 2010. Formulating a historical and demographic model of recent human evolution based on resequencing data from noncoding regions. PLoS One 5.

Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, Fernandes D, Novak M, Gamarra B, Sirak K, et al. 2016. Genomic insights into the origin of farming in the ancient Near East. Nature 536:419–424.

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M, et al. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513:409–413.

Lemmen C, Gronenborn D, Wirtz KW. 2011. A simulation of the Neolithic transition in Western Eurasia. J. Archaeol. Sci. 38:3459–3470.

Di Lernia S. 2001. Dismantling dung: Delayed use of food resources among Early Holocene foragers of the Libyan Sahara. J. Anthropol. Archaeol. 20:408–441.

Leslie S, Winney B, Hellenthal G, Davison D, Boumertit A, Day T, Hutnik K, Royrvik EC, Cunliffe B, Lawson DJ, et al. 2015. The fine-scale genetic structure of the British population. Nature 519:309–314.

Leuenberger C, Wegmann D. 2010. Bayesian computation and model selection without likelihoods. Genetics 184:243–252.

Lewandowska M, Jędrychowska-Dańska K, Zamerska A, Płoszaj T, Witas HW. 2017. The genetic profile of susceptibility to infectious diseases in Roman-Period populations from Central Poland. Infect. Genet. Evol. 47:1–8.

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, et al. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science (80-.). 319:1100–1104.

Li S, Schlebusch C, Jakobsson M. 2014. Genetic variation reveals large-scale population expansion and migration during the expansion of Bantu-speaking peoples. Proc. R. Soc. B Biol. Sci. 281.

Liberles DA, Teufel AI, Liu L, Stadler T. 2013. On the need for mechanistic models in computational genomics and metagenomics. Genome Biol. Evol. 5:2008–2018.

Linstädter J. 2008. The Epipalaeolithic-Neolithic-Transition in the Mediterranean region of Northwest Africa. Quartär 55:41–62.

Linstädter J, Medved I, Solich M, Weniger GC. 2012. Neolithisation process within the Alboran territory: Models and possible African impact. Quat. Int. 274:219–232.

Lipson M, Szécsényi-Nagy A, Mallick S, Pósa A, Stégmár B, Keerl V, Rohland N, Stewardson K, Ferry M, Michel M, et al. 2017. Parallel palaeogenomic transects reveal complex genetic history of early European farmers. Nature 551:368–372.

Liu H, Prugnolle F, Manica A, Balloux F. 2006. A Geographically Explicit Genetic Model of Worldwide Human-Settlement History. Am. J. Hum. Genet. 79:230–237.

Madella M, García-Granero JJ, Out WA, Ryan P, Usai D. 2014. Microbotanical evidence of domestic cereals in Africa 7000 years ago. PLoS One 9.

Marjoram P, Molitor J, Plagnol V, Tavaré S. 2003. Markov chain Monte Carlo without likelihoods. Proc. Natl. Acad. Sci. 100:15324–15328.

Martínez-Sánchez RM, Vera-Rodríguez JC, Pérez-Jordà G, Peña-Chocarro L, Bokbot Y. 2017. The beginning of the Neolithic in northwestern Morocco. Quat. Int.

Martiniano R, Caffell A, Holst M, Hunter-Mann K, Montgomery J, Müldner G, McLaughlin RL, Teasdale MD, Van Rheenen W, Veldink JH, et al. 2016. Genomic signals of migration and continuity in Britain before the Anglo-Saxons. Nat. Commun. 7.

Martins H, Xavier Oms F, Pereira L, Pike AWG, Rowsell K, Zilhão J. 2015. Radiocarbon dating the beginning of the Neolithic in Iberia: New results, new problems. J. Mediterr. Archaeol. 28:105–131.

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

Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, et al. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. Nature 528:499–503.

Mellars P. 2004. Neanderthals and the modern human colonization of Europe. Nature 432:461–465.

Menozzi P, Piazza A, Cavalli-Sforza L. 1978. Synthetic maps of human gene frequencies in Europeans. Science (80-.). 201:786–792.

Metzker ML. 2010. Sequencing technologies the next generation. Nat. Rev. Genet. 11:31–46.

Modi A, Tassi F, Susca RR, Vai S, Rizzi E, Bellis G De, Lugliè C, Gonzalez Fortes G, Lari M, Barbujani G, et al. 2017. Complete mitochondrial sequences from Mesolithic Sardinia. Sci. Rep. 7.

Mona S, Ray N, Arenas M, Excoffier L. 2014. Genetic consequences of habitat fragmentation during a range expansion. Heredity (Edinb). 112:291–299.

Moorjani P, Patterson N, Hirschhorn JN, Keinan A, Hao L, Atzmon G, Burns E, Ostrer H, Price AL, Reich D. 2011. The history of african gene flow into Southern Europeans, Levantines, and Jews. PLoS Genet. 7.

Morales J, Mulazzani S, Belhouchet L, Zazzo A, Berrio L, Eddargach W, Cervi A, Hamdi H, Saidi M, Coppa A, et al. 2015. First preliminary evidence for basketry and nut consumption in the Capsian culture (ca. 10,000-7500BP): Archaeobotanical data from new excavations at El Mekta, Tunisia. J. Anthropol. Archaeol. 37:128–139.

Morales J, Pérez-Jordà G, Peña-Chocamo L, Zapata L, Ruiz-Alonso M, López-Sáez JA, Linstädter J. 2013. The origins of agriculture in North-West Africa: macro-botanical remains from Epipalaeolithic and Early Neolithic levels of Ifri Oudadane (Morocco). J. Archaeol. Sci. 40:2659–2669.

Mulazzani S, Belhouchet L, Salanova L, Aouadi N, Dridi Y, Eddargach W, Morales J, Tombret O, Zazzo A, Zoughlami J. 2016. The emergence of the Neolithic in North Africa: A new model for the Eastern Maghreb. Quat. Int. 410:123–143.

Mummert A, Esche E, Robinson J, Armelagos GJ. 2011. Stature and robusticity during the agricultural transition: Evidence from the bioarchaeological record. Econ. Hum. Biol. 9:284–301.

Nordborg M. 2008. Coalescent Theory. In: Handbook of Statistical Genetics: Third Edition. Vol. 2. p. 843–877.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, et al. 2008. Genes mirror geography within Europe. Nature 456:98–101.

Novembre J, Peter BM. 2016. Recent advances in the study of fine-scale population structure in humans. Curr. Opin. Genet. Dev. 41:98–105.

Olalde I, Allentoft ME, Sánchez-Quinto F, Santpere G, Chiang CWK, DeGiorgio M, Prado-Martinez J, Rodríguez JA, Rasmussen S, Quilez J, et al. 2014. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature 507:225–228.

Olalde I, Brace S, Allentoft ME, Armit I, Kristiansen K, Booth T, Rohland N, Mallick S, Szécsényi-Nagy A, Mittnik A, et al. 2018. The Beaker phenomenon and the genomic transformation of northwest Europe. Nature 555:190–196.

Olalde I, Schroeder H, Sandoval-Velasco M, Vinner L, Lobón I, Ramirez O, Civit S, Borja PG, Salazar-García DC, Talamo S, et al. 2015. A common genetic origin for early farmers from mediterranean cardial and central european LBK cultures. Mol. Biol. Evol. 32:3132–3142.

Omrak A, Günther T, Valdiosera C, Svensson EM, Malmström H, Kiesewetter H, Aylward W, Storå J, Jakobsson M, Götherström A. 2016. Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool. Curr. Biol. 26:270–275.

Omran AR. 1971. The epidemiologic transition. Milbank Mem. Fund Q. 49:509–538.

Özdogan M. 2014. A new look at the introduction of the Neolithic way of life in Southeastern Europe. Changing paradigms of the expansion of the Neolithic way of life. Doc. Praehist. 41:33–49.

de Pablo JFL, Puche MG. 2009. Climate change and population dynamics during the Late Mesolithic and the Neolithic transition in Iberia. Doc. Praehist. 36:67–96.

Padhukasahasram B, Marjoram P, Wall JD, Bustamante CD, Nordborg M. 2008. Exploring population genetic models with recombination using efficient forward-time simulations. Genetics 178:2417–2427.

Pardiñas AF, Roca A, García-Vazquez E, López B. 2012. Assessing the Genetic Influence of Ancient Sociopolitical Structure: Micro-differentiation Patterns in the Population of Asturias (Northern Spain). PLoS One 7.

Pardiñas AF, Roca A, García-Vazquez E, Lõpez B. 2014. Evaluation of large-scale genetic structure in complex demographic and historical scenarios: The mitochondrial DNA and Y-chromosome pools of the Iberian Atlantic façade. Am. J. Phys. Anthropol. 153:617–626.

Pardo Gordó S, Puchol OG, Castillo A, McClure S, Cabanilles J, Ripoll M, Balaguer L, Bernabeu Aubán J, Benito J, Kennett DJ, et al. 2018. Taphonomic processes inconsistent with indigenous Mesolithic acculturation during the transition to the Neolithic in the Western Mediterranean. Quat. Int.:DOI: 10.1016/j.quaint.2018.05.008.

Patin E, Lopez M, Grollemund R, Verdu P, Harmant C, Quach H, Laval G, Perry GH, Barreiro LB, Froment A, et al. 2017. Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. Science (80-.). 356:543–546.

Patin E, Siddle KJ, Laval G, Quach H, Harmant C, Becker N, Froment A, Régnault B, Lemée L, Gravel S, et al. 2014. The impact of agricultural emergence on the genetic history of African rainforest hunter-gatherers and agriculturalists. Nat. Commun. 5.

Peck SL. 2004. Simulation as experiment: A philosophical reassessment for biological modeling. Trends Ecol. Evol. 19:530–534.

Pedersen JS, Valen E, Velazquez AMV, Parker BJ, Rasmussen M, Lindgreen S, Lilje B, Tobin DJ, Kelly TK, Vang S, et al. 2014. Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. Genome Res. 24:454–466.

Peng B, Amos CI, Kimmel M. 2007. Forward-time simulations of human populations with complex diseases. PLoS Genet. 3:0407–0420.

Perrin T. 2013. Potentialites de contacts entre mésolithiques et néolithiques dans le sud de la France. In: Jaubert J, Fourment N, Depaepe P, editors. Transitions, ruptures et continuite en Prehistoire, Evolution des techniques - Comportements funeraires - Neolithique ancien. p. 357–372.

Perrin T, Manen C, Valdeyron N, Guilaine J. 2017. Beyond the sea... The Neolithic transition in the southwest of France. Quat. Int.

Peter BM, Petkova D, Novembre J. unpublished data. https://www.biorxiv.org/content/early/2018/03/26/233486. last accessed June 5, 2018.

Petkova D, Novembre J, Stephens M. 2015. Visualizing spatial population structure with estimated effective migration surfaces. Nat. Genet. 48:94–100.

Piazza A, Rendine S, Minch E, Menozzi P, Mountain J, Cavalli-Sforza LL. 1995. Genetics and the origin of European languages. Proc. Natl. Acad. Sci. 92:5836–5840.

Pickrell JK, Patterson N, Barbieri C, Berthold F, Gerlach L, Güldemann T, Kure B, Mpoloka SW, Nakagawa H, Naumann C, et al. 2012. The genetic prehistory of southern Africa. Nat. Commun. 3.

Pierron D, Chang I, Arachiche A, Heiske M, Thomas O, Borlin M, Pennarun E, Murail P, Thoraval D, Rocher C, et al. 2011. Mutation rate switch inside eurasian mitochondrial haplogroups: Impact of selection and consequences for dating settlement in europe. PLoS One 6.

Pinhasi R, Fort J, Ammerman AJ. 2005. Tracing the origin and spread of agriculture in Europe. PLoS Biol. 3:1–9.

Pinto JC, Pereira V, Marques SL, Amorim A, Alvarez L, Prata MJ. 2015. Mirandese language and genetic differentiation in Iberia: A study using X chromosome markers. Ann. Hum. Biol. 42:20–25.

Price T. 2000. Europe's first farmers. Cambridge University Press

Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol. Biol. Evol. 16:1791–1798.

Ralph P, Coop G. 2013. The Geography of Recent Genetic Ancestry across Europe. PLoS Biol. 11.

Rasteiro R, Bouttier PA, Sousa VC, Chikhi L. 2012. Investigating sex-biased migration during the neolithic transition in Europe, using an explicit spatial simulation framework. Proc. R. Soc. B Biol. Sci. 279:2409–2416.

Ray N, Currat M, Berthier P, Excoffier L. 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. 2009. Inferring Past Demography Using Spatially Explicit Population Genetic Models. Hum. Biol. 81:141–157.

Ray N, Excoffier L. 2010. A first step towards inferring levels of long-distance dispersal during past expansions. Mol. Ecol. Resour. 10:902–914.

Ray N, Wegmann D, Fagundes NJR, Wang S, Ruiz-Linares A, Excoffier L. 2010. A statistical evaluation of models for the initial settlement of the american continent emphasizes the importance of gene flow with Asia. Mol. Biol. Evol. 27:337–345.

Regueiro M, Garcia-Bertrand R, Fadhlaoui-Zid K, Álvarez J, Herrera RJ. 2015. From Arabia to Iberia: A Y chromosome perspective. Gene 564:141–152.

Richards MP. 2002. A brief review of the archaeological evidence for Palaeolithic and Neolithic subsistence. Eur. J. Clin. Nutr. 56:1270–1278.

Rodríguez-Ezpeleta N, Álvarez-Busto J, Imaz L, Regueiro M, Azcárate MN, Bilbao R, Iriondo M, Gil A, Estonba A, Aransay AM. 2010. High-density SNP genotyping detects homogeneity of Spanish and French Basques, and confirms their genomic distinctiveness from other European populations. Hum. Genet. 128:113–117.

Rodríguez-Varela R, Günther T, Krzewińska M, Storå J, Gillingwater TH, MacCallum M, Arsuaga JL, Dobney K, Valdiosera C, Jakobsson M, et al. 2017. Genomic

Analyses of Pre-European Conquest Human Remains from the Canary Islands Reveal Close Affinity to Modern North Africans. Curr. Biol. 27:3396–3402.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. 2002. Genetic structure of human populations. Science (80-.). 298:2381–2385.

Rowley-Conwy P. 2011. Westward Ho! The Spread of Agriculturalism from Central Europe to the Atlantic Author(s): Curr. Anthropol. 52:S431–S451.

Russell T, Silva F, Steele J. 2014. Modelling the spread of farming in the bantu-speaking regions of africa: An archaeology-based phylogeography. PLoS One 9.

Sánchez-Quinto F, Schroeder H, Ramirez O, Ávila-Arcos MC, Pybus M, Olalde I, Velazquez AM V, Marcos MEP, Encinas JMV, Bertranpetit J, et al. 2012. Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. Curr. Biol. 22:1494–1499.

Santos C, Fregel R, Cabrera VM, Álvarez L, Larruga JM, Ramos A, López MA, Pilar Aluja M, González AM. 2014. Mitochondrial DNA and Y-chromosome structure at the mediterranean and Atlantic façades of the Iberian Peninsula. Am. J. Hum. Biol. 26:130–141.

Sarno S, Boattini A, Pagani L, Sazzini M, De Fanti S, Quagliariello A, Ruscone GAG, Guichard E, Ciani G, Bortolini E, et al. 2017. Ancient and recent admixture layers in Sicily and Southern Italy trace multiple migration routes along the Mediterranean. Sci. Rep. 7.

Scanes CG. 2018. Chapter 6 – The Neolithic Revolution, Animal Domestication, and Early Forms of Animal Agriculture. Available from: http://dx.doi.org/10.1016/B978-0-12-805247-1/00006-X

Schiffels S, Haak W, Paajanen P, Llamas B, Popescu E, Loe L, Clarke R, Lyons A, Mortimer R, Sayer D, et al. 2016. Iron Age and Anglo-Saxon genomes from East England reveal British migration history. Nat. Commun. 7.

Schlebusch CM, Skoglund P, Sjödin P, Gattepaille LM, Hernandez D, Jay F, Li S, De Jongh M, Singleton A, Blum MGB, et al. 2012. Genomic variation in seven Khoe-San groups reveals adaptation and complex African history. Science (80-.). 338:374–379.

Schraiber JG, Akey JM. 2015. Methods and models for unravelling human evolutionary history. Nat. Rev. Genet. 16:727–740.

Seguin-Orlando A, Korneliussen TS, Sikora M, Malaspinas AS, Manica A, Moltke I, Albrechtsen A, Ko A, Margaryan A, Moiseyev V, et al. 2014. Genomic structure in Europeans dating back at least 36,200 years. Science (80-.). 346:1113–1118.

Sellman S, Tsao K, Tildesley MJ, Brommesson P, Webb CT, Wennergren U, Keeling MJ, Lindström T. 2018. Need for speed: An optimized gridding approach for spatially explicit disease simulations. PLoS Comput. Biol. 14.

Shriner D, Tekola-Ayele F, Adeyemo A, Rotimi CN. 2014. Genome-wide genotype and sequence-based reconstruction of the 140,000 year history of modern human ancestry. Sci. Rep. 4.

Silva NM, Rio J, Currat M. 2017. Investigating population continuity with ancient DNA under a spatially explicit simulation framework. BMC Genet. 18.

Sisson SA, Fan Y, Beaumont MA. 2018. Overview of Approximate Bayesian Computation. arXiv:1802. Available from: https://arxiv.org/pdf/1802.09720

Skoglund P, Malmström H, Omrak A, Raghavan M, Valdiosera C, Günther T, Hall P, Tambets K, Parik J, Sjögren KG, et al. 2014. Genomic diversity and admixture differs for stone-age Scandinavian foragers and farmers. Science (80-.). 344:747–750.

Skoglund P, Malmström H, Raghavan M, Storå J, Hall P, Willerslev E, Gilbert MTP, Götherström A, Jakobsson M. 2012. Origins and genetic legacy of neolithic farmers and hunter-gatherers in Europe. Science (80-.). 336:466–469.

Skoglund P, Thompson JC, Prendergast ME, Mittnik A, Sirak K, Hajdinjak M, Salie T, Rohland N, Mallick S, Peltzer A, et al. 2017. Reconstructing Prehistoric African Population Structure. Cell 171:59–71.

Slatkin M. 2001. Simulating genealogies of selected alleles in a population of variable size. Genet. Res. 78:49–57.

Stewart JR, Stringer CB. 2012. Human evolution out of Africa: The role of refugia and climate change. Science (80-.). 335:1317–1321.

Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C. 2013. Approximate Bayesian Computation. PLoS Comput. Biol. 9.

Szécsényi-Nagy A, Roth C, Brandt G, Rihuete-Herrada C, Tejedor-Rodríguez C, Held P, García-Martínez-De-Lagrán Í, Arcusa Magallón H, Zesch S, Knipper C, et al. 2017. The maternal genetic make-up of the Iberian Peninsula between the Neolithic and the Early Bronze Age. Sci. Rep. 7.

Tassi F, Vai S, Ghirotto S, Lari M, Modi A, Pilli E, Brunelli A, Susca RR, Budnik A, Labuda D, et al. 2017. Genome diversity in the neolithic globular amphorae culture and the spread of Indo-European languages. Proc. R. Soc. B Biol. Sci. 284.

The International HapMap Consortium. 2005. A haplotype map of the human genome. Nature 437:1299–1320.

The International HapMap Consortium. 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851–861.

Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O, et al. 2009. The genetic structure and history of Africans and African Americans. Science (80-.). 324:1035–1044.

Valdiosera C, Günther T, Vera-Rodríguez JC, Ureña I, Iriarte E, Rodríguez-Varela R, Simões LG, Martínez-Sánchez RM, Svensson EM, Malmström H, et al. 2018. Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia. Proc. Natl. Acad. Sci.:201717762.

Vallée F, Luciani A, Cox MP. 2016. Reconstructing demography and social behavior during the neolithic expansion from genomic diversity across island southeast asia. Genetics 204:1495–1506.

Veeramah KR, Rott A, Groß M, van Dorp L, López S, Kirsanow K, Sell C, Blöcher J, Wegmann D, Link V, et al. 2018. Population genomic analysis of elongated skulls reveals extensive female-biased immigration in Early Medieval Bavaria. Proc. Natl. Acad. Sci.:201719880.

Veeramah KR, Wegmann D, Woerner A, Mendez FL, Watkins JC, Destro-Bisol G, Soodyall H, Louie L, Hammer MF. 2012. An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. Mol. Biol. Evol. 29:617–630.

Verdu P, Austerlitz F, Estoup A, Vitalis R, Georges M, Théry S, Froment A, Le Bomin S, Gessain A, Hombert JM, et al. 2009. Origins and Genetic Diversity of Pygmy Hunter-Gatherers from Western Central Africa. Curr. Biol. 19:312–318.

Vermeersch PM, Linseele V, Marinova E, Van Neer W, Moeyersons J, Rethemeyer J. 2015. Early and Middle Holocene Human Occupation of the Egyptian Eastern Desert: Sodmein Cave. African Archaeol. Rev. 32:465–503.

Wakeley J. 2008. Coalescent Theory: An Introduction. Roberts and Company Publishers

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.

Wendrich W, Taylor RE, Southon J. 2010. Dating stratified settlement sites at Kom K and Kom W: Fifth millennium BCE radiocarbon ages for the Fayum Neolithic.

Nucl. Instruments Methods Phys. Res. Sect. B Beam Interact. with Mater. Atoms 268:999–1002.

Weninger B, Clare L, Gerritsen F, Horejs B, Krauß R, Linstädter J, Özbal R, Rohling EJ. 2014. Neolithisation of the Aegean and Southeast Europe during the 6600 – 6000 calBC period of Rapid Climate Change. Doc. Praehist. XLI:1–32.

Westesson O, Holmes I. 2009. Accurate detection of recombinant breakpoints in whole-genome alignments. PLoS Comput. Biol. 5.

Yang Z. 2006. Computational Molecular Evolution. Oxford University Press

Zeder MA. 2008. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. Proc. Natl. Acad. Sci. 105:11597–11604.

Zilhao J. 2001. Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. Proc. Natl. Acad. Sci. 98:14180–14185.

Zilhão J. 2003. The Neolithic Transition in Portugal and the Role of Demic Diffusion in the Spread of Agriculture across West Mediterranean Europe. In: Ammerman AJ, Biagi P, editors. The Widening Harvest. The Neolithic Transition in Europe: Looking Back, Looking Forward. Archaeological Institute of America. p. 207–223.

Zilhão J. 2014. Early prehistoric navigation in the Western Mediterranean: Implications for the Neolithic transition in Iberia and the Maghreb. Eurasian Prehistory 11:185–200.

Chapter 2.

Evaluating the Neolithic Expansion at Both Shores of the Mediterranean Sea

João Pimenta, Alexandra M. Lopes, David Comas, António Amorim and Miguel Arenas

Molecular Biology and Evolution, 2017, 34(*12*):3232–3242.
(doi:10.1093/molbev/msx256)

2.1. Introduction

The Neolithic transition in Europe and North Africa started in the Near East around 10,000 years ago (ya) and generated a revolution in the living-style with the emergence of farmers after previous hunter-gatherer Palaeolithic populations (Zeder and Hesse 2000; Skoglund et al. 2012; Broushaki et al. 2016; Gallego-Llorente et al. 2016; Lazaridis et al. 2016; Meiklejohn et al. 2017). Two contrasting hypotheses were postulated for the expansion of Agriculture, the *demic diffusion* model (Ammerman and Cavalli-Sforza 1984; Sokal et al. 1991) and the *cultural diffusion* model (Fort 2012). The first one supports a migratory flux of Neolithic populations across Europe, which partially replaced the indigenous hunter-gatherer populations and implies gene flow between hunter-gatherers and farmers. The second model predicts the cultural adoption of the Neolithic lifestyle by neighbour populations with limited gene flow from Near Eastern populations (Sampietro et al. 2007; Fort 2012; Fu et al. 2012). Agriculture spread rapidly along the Southeast of Europe (Zeder 2008), leading to a pronounced cultural revolution, and left profound genetic marks in European populations (Currat and Excoffier 2005; Bramanti et al. 2009). In contrast to Europe, the *tempo and mode* of the Neolithic expansion throughout the North of Africa is still obscure, mostly because of a lack of archaeological and biological data (Barich 2014). Most evidence of the presence of domesticated species in North Africa dates around 8,000 ya and older (Linseele et al. 2014; Madella et al. 2014), suggesting an early Neolithic presence in the region, which is in agreement with the Mediterranean expansion route (Zeder 2008; Isern et al. 2017). Beforehand, one could predict that the transition to a farmer lifestyle in North Africa could have been slower than in Europe due to the arid conditions of the Sahara desert. It is also noteworthy that the region suffered frequent and drastic climatic changes during the Palaeolithic, with wet and arid periods (Adkins et al. 2006; McGee et al. 2013), which could have caused major demographic and cultural shifts (Clarke et al. 2016). However, North Africa experienced a long African Humid Period (AHP) between 12,000 and 5,000 ya (Adkins et al. 2006; Dunne et al. 2012; McGee et al. 2013), coincident with the Neolithic expansion period.

Importantly, the Mediterranean Sea was a geographical barrier to migration between Neolithic populations of North Africa and Europe, generating great cultural differences between both regions, exemplified by the expansion of Indo-European languages in Europe (Haak et al. 2015) and the expansion of Afro-Asiatic languages in North Africa (Ruhlen 1991). However, the geographical proximity at the westernmost region of both continents (Strait of Gibraltar) could have provided a migration corridor between the two shores (Bosch et al. 2001; González-Pérez et al. 2003; Rhouda et al. 2009;

Botigué et al. 2013), although still little is known about the specific direction and amount of migration through this Strait during the Neolithic.

Here we aim to evaluate and compare the Neolithic expansion on both shores of the Mediterranean Sea accounting for the role of the Strait of Gibraltar as a potential migration corridor between them. Using an approximate Bayesian computation (ABC) method based on extensive spatially explicit computer simulations, we investigated the population genetics and evolution of the Neolithic expansion in Europe and North Africa. First, we selected the evolutionary scenario of migration through the Strait of Gibraltar that fitted best with the real data and second, under the previously selected evolutionary scenario, we estimated population genetics parameters for the Neolithic expansion at each shore of the Mediterranean Sea. For this purpose, we built an X-chromosome dataset that combines data from several SNP dataset panels and includes samples from Europe, North Africa, Middle East and Sub-Saharan Africa. The use of the X-chromosome SNPs allowed us to perform an ABC analysis in a computational cost-benefit manner, and avoids errors of haplotype reconstruction since phasing is directly inferred in male individuals.

2.2. Material and Methods

2.2.1. Data and Genotyping

A dataset composed by 20 populations with 846 individuals (580 males and 266 females) was built with published data (Figure 2.1 and Table S2.1; Supplementary Material). The dataset included 11 European populations, Finland (FIN), Northern and Western Europeans currently in Utah (CEU), Great-Britain (GBR), Tuscany (TSI), East Spain (ESP) and Central Spain (CSP) populations retrieved from 1,000 Genomes Phase 3 (The 1000 Genomes Project Consortium 2012); Basque (BAS), Andalusia (AND) and Galicia (GAL) populations retrieved from Henn et al. (2012) and 1,000 Genomes Phase 3; and Porto (POR) and Lisbon (LIS) populations retrieved from Lopes et al. (2013). It also included 7 North African populations retrieved from Henn et al. (2012) and a Sub-Saharan [Yoruba (YRI)] populations from 1,000 Genomes Phase 3. Finally, a Syrian population retrieved from Arauna et al. (2017).

A quality control filter was applied to the X chromosome data using PLINK 1.07 (Purcell et al. 2007) and SNPs with missing rate higher than 10%, SNPs that failed the Hardy-Weinberg equilibrium under a threshold of 0.05 and SNPs with Minor Allele Frequency (MAF) lower than 0.05 were excluded to avoid biases. In addition, individuals with a missing rate higher than 10% and those with an identity by state, for the autosomes, larger than 85%, were also excluded. Moreover, SNPs within both pseudoautosomal regions (PAR1 and PAR2) and those that were heterozygous in the X specific region in male individuals were also excluded. A total of 16,939 SNPs were maintained to proceed with the phasing step.

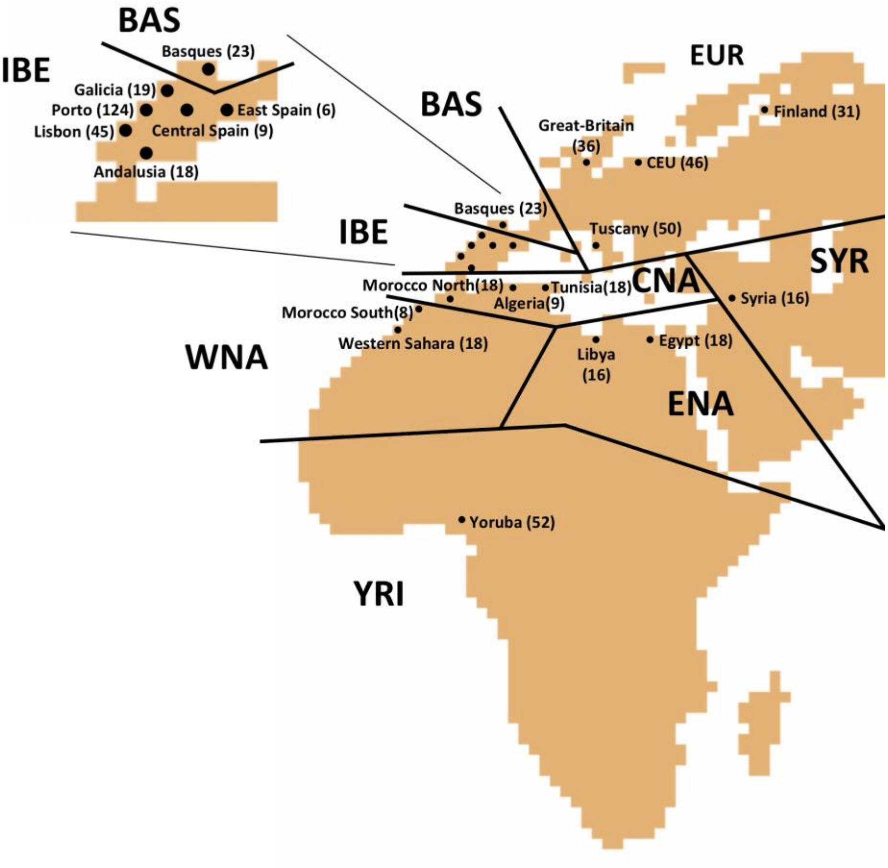


Figure 2.1. Location and sample size (580 male individuals) of the 20 populations that composed the studied dataset. Populations were divided in 8 geographic groups abbreviated as EUR (European populations), BAS (Basque Country population), IBE (Iberian populations), WNA (western North Africa populations), CNA (central North Africa populations), ENA (eastern North Africa populations), SYR (Syrian population) and YRI (Yoruba population).

2.2.2. Phasing

As a sanity check, the dataset was phased using the SHAPEIT2 software (Delaneau et al. 2013), the HapMap PhaseII b37 (The International Hapmap Consortium 2003) with the genetic map and the 1000 genomes dataset as the reference panel (The 1000 Genomes Project Consortium 2012). The dataset was phased after an alignment with the reference panel and the removal of SNPs that did not align. Next, all genotypes were phased simultaneously with a total of 100 iterations, half of them for burn-in and pruning stages. A total of 16,917 SNPs were kept for subsequent analyses. For the ABC analysis, only male samples were considered, since X chromosome haplotypes can be directly (without biases) identified.

2.2.3. Spatially explicit computer simulations

Palaeolithic and Neolithic expansions were simulated with the evolutionary framework SPLATCHE2 (Ray et al. 2010). SPLATCHE2 is a well-established spatially explicit computer simulator that has been widely used to investigate human evolution (Benguigui and Arenas 2014). This program is based on three major steps. The first one consists of a forward-in-time simulation of the whole population under demographic parameters such as population size at the onset of the simulation, population growth rate and number of generations. This demographic simulation is performed over a two-dimensional (2D) landscape (grid of demes) that considers a carrying capacity per deme (number of individuals that can be sustained by the resources of the deme) and where individuals can move between demes according to a migration rate and under a 2D stepping-stone migration model (Kimura and Weiss 1964). The landscape used for this study was obtained by a geographic information system (GIS) and includes 4,814 demes with deme size 150 × 150 km following Alves et al. (2016) (Figure 2.1). A land bridge was added to connect Great Britain with mainland Europe, allowing the colonization of these islands. The second step is a backward-in-time (coalescent) simulation that generates the evolutionary history of a user-specified sample accounting for the history of the whole population stored in the previous step (further details in Currat et al. 2004). In the third step, the program simulates the genetic data of the sample. To perform this task, a randomly selected sequence is assigned to the root node of the previously simulated coalescent history and is evolved going forward-in-time over such coalescent history, according to a substitution model of evolution, to generate genetic sequences for the sampled nodes (Yang 2006; Arenas 2012). Altogether, the simulated genetic data is influenced by the user-specific population genetics parameters (i.e., large population size can lead to high genetic diversity) (Ray et al. 2010).

We performed computer simulations following settings described in recent studies (François et al. 2010; Arenas et al. 2013; Alves et al. 2016). The Palaeolithic expansion was assumed to start from the Ethiopian region (Pritchard et al. 2000) at T_{START} generations ago with an ancestral population size of N_{ANC} haploid individuals. At each generation, individuals migrated to the adjacent demes at a migration rate M_{PALEO} and the population density for each deme was determined by the population growth rate G_{PALEO} and the carrying capacity K_{PALEO} (Currat et al. 2004). A range contraction was simulated to mimic the *last glacial maximum* (LGM) period in Africa and Eurasia described by Alves et al. (2016). Briefly, this range contraction consisted of 30 consecutive events that progressively reduced the habitable regions (i.e., at the North of Europe or increasing deserts size; Figure S2.1, Supplementary Material) of human settlements (Arenas et al. 2012). The range contraction started at 25,000 ya and took 3,000 years to reach its maximum. Next, individuals were only living at the refugia (South of Europe, North Africa and South of Africa; see Figure S2.1) during 4,000 years and finally re-expanded at 18,000 ya (Alves et al. 2016).

In addition, we simulated a Neolithic expansion that started from the Middle East (Zeder 2008) at T_{NEO} generations ago and with an ancestral size at the onset of the expansion of N_{ANCNEO} haploid individuals and a population growth rate G_{NEO}. Here, the map was divided into three different geographic regions: Europe, North Africa and peripheral regions (i.e., Arabian Peninsula and Sub-Saharan Africa) (Figure 2.2A). Each of these regions had a specific migration rate (M_{NEOEUR}, M_{NEONA} and M_{NEOPER}) and carrying capacity (K_{NEOEUR}, K_{NEONA} and K_{NEOPER}), actually these are the only parameters implemented in SPLATCHE2 with spatio-temporal variation. The contribution of the Palaeolithic to the final genetic pool was considered with the parameter interbreeding

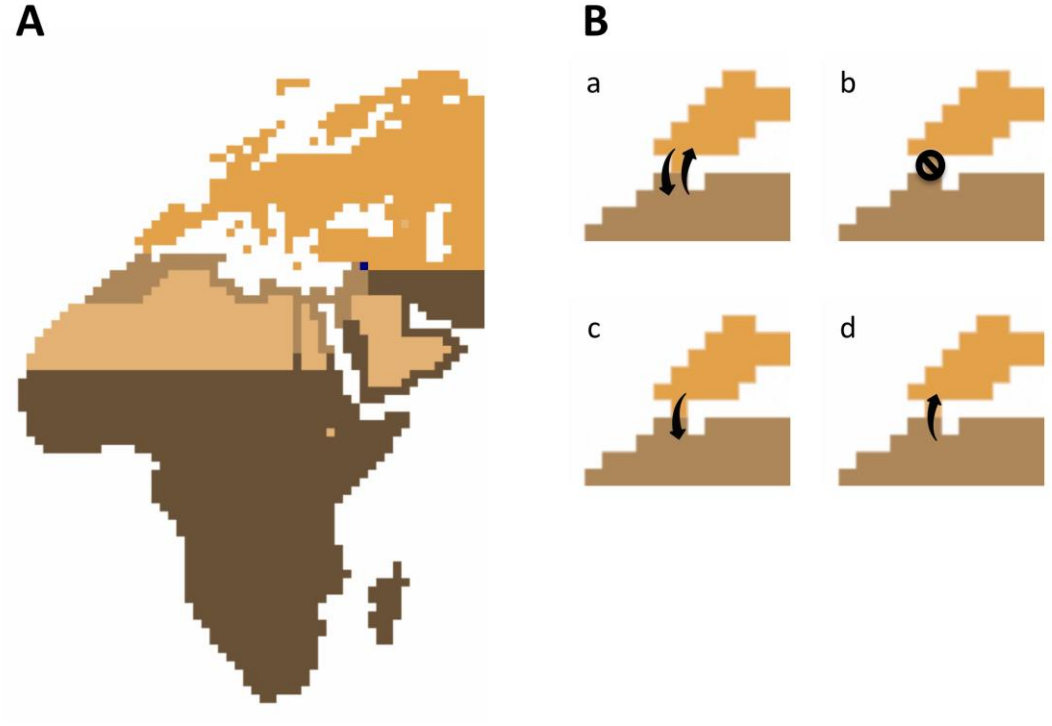


Figure 2.2. Landscape features for the spatially explicit computer simulations of the Neolithic expansion. (A) 2D map of the three different regions considered for the Neolithic expansion: Europe, North Africa and peripheral regions. Uninhabited regions (i.e., deserts) are shown in light brown. The origin of the simulated Neolithic expansion is shown in blue. (B) Alternative evolutionary scenarios of migration through the Strait of Gibraltar: (a) reciprocal migration between both regions (*fullMIG*), (b) lack of migration between both regions (*noMIG*), (c) migration allowed only from the Iberia Peninsula to North Africa (*nsMIG*), (d) migration allowed only from North Africa to the Iberia Peninsula (*snMIG*).

rate *Irate* (Barbujani et al. 1995; Currat and Excoffier 2004; Currat and Excoffier 2005; Currat et al. 2008; François et al. 2010; Currat and Excoffier 2011; Arenas et al. 2013), which is assumed to be spatio-temporal invariable in the simulator.

All the parameters were drawn from Uniform prior distributions (Table S2.2; Supplementary Material) based on previous studies (Arenas et al. 2013; Alves et al. 2016) and were estimated with approximate Bayesian computation (ABC; details below).

Next, based on the previously simulated demographic history, we simulated genetic data for the sample of 580 male individuals of the 20 populations. Each individual included an X chromosome sequence of 16,917 SNPs divided in 14 independent blocks (Table S2.3; Supplementary Material). These blocks were determined by the presence of particular regions without recombination based on published data (Tillmar et al. 2008) and based on estimated recombination hotspots for the CEU, POR and YRI populations (chosen arbitrarily) with the program LDhat 2.2 (McVean et al. 2004). LDhat was run following the recommendations provided in the software documentation. We applied a sliding window of 2,000 SNPs with an overlap of 500 SNPs and performed 6 independent runs (2 per population) with 10 million iterations each (Figure S2.2A; Supplementary Material). We specified 2,000 iterations per sampling and a block penalty of 20 for changes in recombination rate ρ ($\rho=4N_er$). Following the documentation, 10% of iterations were discarded as burn-in. Convergence between the independent runs was assessed considering a recombination rate threshold of 20, however most SNPs presented a difference lower than 1 indicating a very high convergence (Figure S2.2BCD). To perform the computer simulations, the MAF value was set to be not lower than 0.05 for every block, in line with the SNP filter applied to the real dataset, as described above.

2.2.4. Alternative Scenarios

In order to properly study the Neolithic expansion over Europe and over North Africa, we explored genetic signatures of migration through the Strait of Gibraltar during the Neolithic, assuming absence or little migration between these regions before that period (e.g., Timmermann and Friedrich 2016). Although some studies suggested Palaeolithic migration through this Strait (Currat et al. 2010), other studies described or assumed lack of this migration (Ferembach 1985; Bocquet-Appel and Demars 2000; Liu et al. 2006; Timmermann and Friedrich 2016). We believe that if Palaeolithic migration through the Strait took place, the corresponding genetic signatures could be insignificant due to the small number of migrants and evolutionary processes like genetic drift. Next, Neolithic migration through the Strait of Gibraltar could affect Neolithic populations at both shores of the Mediterranean Sea by exchanging genetic material between them. Therefore, first, we investigated the fitting of alternative evolutionary scenarios to the real data with ABC (details below). One of these scenarios considered migration in both directions between Iberian Peninsula and North Africa (*fullMIG*, Figure 2.2Ba), with migration from North to South and from South to North through the Strait of Gibraltar with rates M_{nsSG} and M_{snSG}, respectively. The other scenario considered lack of migration (*noMIG*, Figure 2.2Bb).

Next, since *fullMIG* better fitted with the real data (see Results), we designed two additional scenarios to compare both directions of migration through the Strait of Gibraltar. Both scenarios allowed migration but under only one direction (anisotropic migration): a scenario of migration from the Iberian Peninsula to North Africa (*nsMIG*, Figure 2.2Bc) at M_{nsSG} rate and, other scenario of migration from North Africa to the Iberian Peninsula (*snMIG*, Figure 2.2Bd) with rate M_{snSG}.

Finally, we investigated model selection under some additional combinations of evolutionary scenarios: (*i*) the four scenarios together (*fullMIG vs noMIG vs nsMIG vs snMIG*) and (*ii*) the scenario of bidirectional migration against each scenario of unidirectional migration (*fullMIG vs nsMIG* and, *fullMIG vs snMIG*). However here note that *fullMIG* includes a different prior distribution for the migration rate towards each direction, whose sampling can sometimes lead to situations (large anisotropic migration) near to *nsMIG* or *snMIG* and this must be considered in the analysis of these scenarios.

2.2.5. Approximate Bayesian Computation for Selection of Evolutionary Scenarios and Parameter Estimation

The real data was analysed with ABC (Beaumont et al. 2002; Beaumont 2010), a statistical approach to obtain posterior probabilities in the selection among alternative scenarios and the estimation of parameters. Next, we describe the different steps of our ABC method.

Computer Simulations. For each evolutionary scenario (of the 4 above described), we performed a total of 100,000 spatially explicit computer simulations according to the prior distributions described in Table S2.2 (Supplementary Material).

Summary Statistics. Summary statistics (SS) from real and simulated datasets were obtained with ARLEQUIN ver.3.5 (Excoffier and Lischer 2010). To perform this task, the 20 populations were classified into eight geographical groups (described in Figure 2.1) that allowed the estimation of diverse group-based SS. This grouping is important because the ABC performance reduces quickly as the number of statistics grows (Beaumont 2010). The computed SS included: number of alleles (k), heterozygosity (H), number of segregating sites (s) and, F_{ST} between groups and populations. Next, we selected a small set (36) of the most informative SS for the estimations (Table S2.4; Supplementary Material). This was done by testing the contribution of the different SS to the estimations from 100 simulated datasets (hereafter *test datasets*) that are independent of the 100,000 simulated datasets used to generate the ABC method. We did not apply partial least square (PLS) components to reduce the number of SS (Wegmann et al. 2009) because our selected SS led to accurate estimations and because PLS components can generate SS without biological meaning and thus being difficult to interpret, and also can generate biases when comparing components that were separately computed (Aeschbacher et al. 2012).

Selection of Evolutionary Scenarios. As noted above, we generated 100,000 simulated datasets for each of the 4 migration scenarios. In addition, we simulated 100 *test datasets* for each scenario to evaluate the accuracy of the method. We applied 4 ABC methods to perform the selection of the best-fitting evolutionary scenario to the real data: the *rejection* method developed by Pritchard et al. (1999); and the methods *rejection*, *multinomial logistic regression* and *neural network*, implemented in the *abc* package of R (Csilléry et al. 2012). Following the authors recommendation we retained 1% of simulations that were closer to the observed data (Csilléry et al. 2012). Before analysing the real data, the accuracy of the ABC methods was assessed with the *test datasets* (considering them as *true* data) and by cross-validation based on 100 permutations and a tolerance of 1% (following the recommendations described in the package *abc* (Csilléry et al. 2012)). Although traditionally the goodness of fit analysis is only applied to the preferred scenario, we performed goodness of fit analyses (based on principal component analysis (PCA) and histogram of null distributions from the selected SS (Csilléry et al. 2012)) to the four investigated evolutionary scenarios.

Parameters Estimation. We performed the parameters estimation under the evolutionary scenario (*fullMIG*), which is the scenario that fitted best with the real data. We applied the *rejection* method implemented in the *abc* package (which generated the most accurate inferences for the studied dataset) by retaining 1,000 simulated datasets of the closest simulations. The robustness of the parameter estimation was assessed with the 100 *test datasets* simulated under the *fullMIG* scenario, considering them as *true* data. Then we computed the distance between the true value and the estimated value (median, mean and mode) from each of the 100 *test datasets*.

2.3. Results

We first describe the selection between alternative evolutionary scenarios of migration through the Gibraltar strait. After that, we present the parameter estimation under the best fitting evolutionary scenario, with a focus on those parameters that are specific for the Neolithic expansion over each shore of the Mediterranean Sea.

2.3.1. Migration through the Strait of Gibraltar occurred in Both Directions

The selection of the evolutionary scenario (*fullMIG vs noMIG*) presented very high posterior probabilities to identify the correct scenario. We found that all the applied ABC methods were able to identify the correct scenario with probabilities that ranged from 0.78 (*rejection* method implemented in the *abc* package) to 0.98 (*multinomial logistic regression* and *neural network* methods) (Table S2.5; Supplementary Material).
Next, real data clearly fitted better with the *fullMIG* scenario (posterior probabilities around 0.99 under all the ABC methods excepting the rejection method of *abc* package (0.85) (Figure S2.3A). Goodness of fit analyses also indicated this finding (Figure S2.4; Supplementary Material). Altogether the results suggest reciprocal migration through the Strait of Gibraltar.

Given that the real data fitted better with the presence of migration, we next investigated the direction of the migration through the Strait of Gibraltar. The validation of the selection of the evolutionary scenario (*nsMIG vs snMIG*) also presented very high posterior probabilities to identify the correct scenario. We found that all the ABC methods performed acceptably to recognize the correct scenario with probabilities that ranged from 0.85 (*rejection* method implemented in the *abc* package) to 1.00 (*Pritchard's rejection* method) (Table S2.6; Supplementary Material).
Next, the fitting of real data with both evolutionary scenarios, *nsMIG* and *snMIG*, suggested migration in both directions, although with a higher probability for the *nsMIG* scenario (Figure S2.3B). The results varied among ABC methods but led to the same overall conclusion.

Next, we describe the results (evaluation of model selection and analysis of real data) from the two additional comparisons of evolutionary scenarios, (*i*) the four evolutionary scenarios together (*fullMIG vs noMIG vs nsMIG vs snMIG*) and (*ii*) the evolutionary scenario of bidirectional migration against each evolutionary scenario of unidirectional migration (*fullMIG vs nsMIG* and, *fullMIG vs snMIG*).

Concerning the evaluation of the model selection, when the four scenarios are evaluated together (Table S2.7; Supplementary Material), the scenario *noMIG* was recognized with the highest probability (reaching 0.85) when applying the *multinomial logistic regression* and the *neural network* methods, while the three other scenarios presented lower probabilities (which is expected since *fullMIG* can sometimes be near to *nsMIG* or *snMIG*, see Materials and methods) but showed the highest probability with the correct scenario. Concerning comparisons between the scenario of bidirectional migration with each scenario of unidirectional migration, we found that all the ABC methods performed acceptably to recognize the correct scenario with probabilities between 0.65 and 0.85 (Tables S2.8 and S2.9; Supplementary Material). Again despite of *fullMIG* can sometimes be near to *nsMIG* or *snMIG* (Materials and methods).

Concerning the fitting with real data, the highest probability corresponded to the *fullMIG* scenario, as expected followed by the scenarios with unidirectional migration since these scenarios are very near to the *fullMIG* scenario (Figure 2.3, see also Figure S2.4). We found similar results when comparing the scenario of bidirectional migration with each scenario of unidirectional migration (Figures S2.5 and S2.6; Supplementary Material).

2.3.2. Similar Neolithic Expansion at Both Shores of the Mediterranean Sea

The estimation of the population genetics parameters was validated with the *tests datasets* of the best-fitting evolutionary scenario (*fullMIG*). These estimations presented a performance with overall small estimation errors that fell within the 50% HPDI with respect to the true value (Table S2.10; supplementary material), although for some parameters related with very ancestral populations (i.e., ancestral Palaeolithic population size) the precision was low, as expected and also found in other studies (Wegmann and Excoffier 2010; Alves et al. 2016) The estimated population genetics parameters for the real data are shown in Table 2.1 and in Figure S2.7 (Supplementary Material). Again, some parameters (especially those related with Palaeolithic populations and ancestral population sizes) presented wide posterior distributions, which is expected due to loss of very ancestral genetic signatures (Fagundes et al. 2007; Wegmann and Excoffier 2010; Alves et al. 2016) However we preferred to account for this variation instead of fixing parameters' values since previous studies

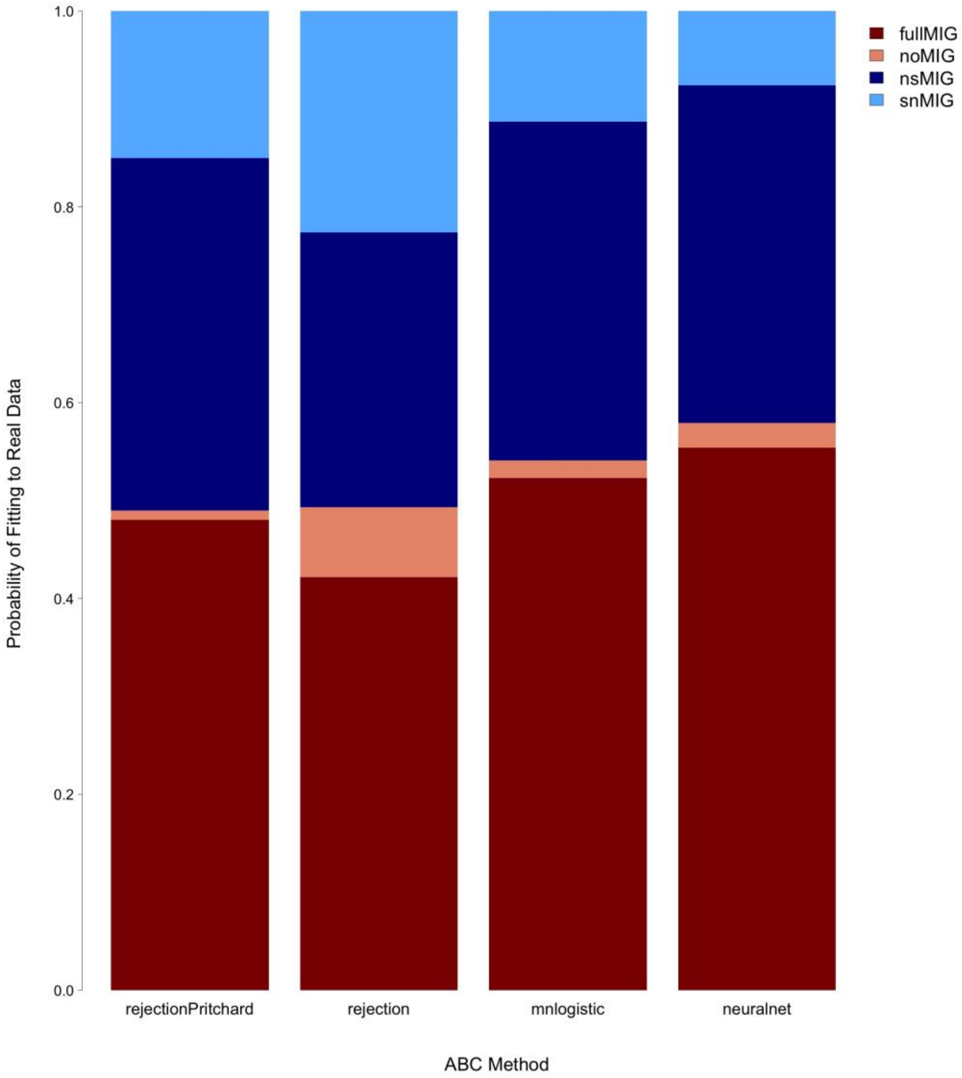


Figure 2.3. Fitting of the four alternative evolutionary scenarios of migration through the Gibraltar Strait (*fullMIG*, *noMIG*, *nsMIG* and *snMIG*) to the real data. Posterior probabilities of evolutionary scenarios were estimated with four different ABC methods (the *rejection* method by Pritchard and the *rejection*, *multinomial logistic regression* and *neural network* methods implemented in the *abc* package (Csilléry et al. 2012)).

disagree on specific parameters values (i.e., Benguigui and Arenas 2014). Regarding the Palaeolithic population, our estimated date for the onset of the expansion of Anatomically Modern Humans (AMH) in Africa is similar (≈100 kya) to that obtained by Alves et al. (2016), Voight et al. (2005) and Fagundes et al. (2007), and like in those studies we also obtained a wide posterior distribution. The estimated ancestral size (≈15,000) was slightly larger but still in agreement with previous analyses based on other genetic markers and also with wide posterior distributions (Fagundes et al. 2007; Alves et al. 2016). The estimation of the onset of the out-of-Africa expansion (≈70 kya but with a wide posterior distribution) is very similar to that obtained by Alves et al. (2016) and within the intervals obtained by Fagundes et al. (2007). The carrying capacity estimated for the Palaeolithic population (≈1,500) was slightly larger than in other studies (Hamilton et al. 2007; Deshpande et al. 2009) but falls within the estimations by Alves et al. (2016). The population growth rate per generation (≈0.70) fell within the range estimated in other studies for hunter-gatherer populations (between 0.2 and 0.9) (Alves et al. 2016). The migration rate for Palaeolithic populations (≈0.23) also fell within the range estimated by Alves et al. (2016). The estimated interbreeding rate was ≈0.011, which corresponds to approximately 71% of admixture between Palaeolithic and Neolithic populations (obtained with a polynomial regression based on Arenas et al. (2013)) and falls within the range described in other studies (Bramanti et al. 2009; Lazaridis et al. 2014; Haak et al. 2015; Lazaridis et al. 2016).

Regarding the Neolithic population, the estimated time for the onset of the Neolithic expansion (around 10 kya) was in agreement with previous findings (Goring-Morris and Belfer-Cohen 2011). As expected, the estimated Neolithic ancestral size (around 23,000) and carrying capacity (≈6,200) were higher than for the Palaeolithic population (Table 2.1). However, the population growth rate and migration rate for Neolithic and Palaeolithic populations were similar. We did not find significant differences (at the 95% HPDI) between the migration rate from the Iberian Peninsula to the North of Africa and the migration rate in the opposite direction. Indeed and as expected, the migration rate through the Strait (≈0.15) was much lower than the migration rate in other regions, which is indicative of the barrier to gene flow generated by the Strait. Importantly, the Neolithic migration rate through Europe and through the North of Africa presented similar levels (≈0.23). Indeed, the Neolithic carrying capacity for European and North African populations also exhibited similar levels (Table 2.1).

2.4. Discussion

The change from the hunter-gatherer lifestyle to the more sedentary farmers of the Neolithic was a major transition in human history (Bocquet-Appel 2011). The Near Eastern Neolithic revolution expanded to several regions, including Europe and North of Africa. However, little is known about the Neolithic expansion in North of Africa and its comparison with the expansion in Europe. Here, we evaluate the Neolithic progression in Europe, a region where the expansion from Anatolia was fast and agriculture was adopted in ≈3,000 years (Bocquet-Appel et al. 2012), and in North Africa, a region still poorly studied. To address this aim we analysed the entire X chromosome (16,917 SNPs) from 580 male individuals that belong to 20 populations from Europe, Africa and Near East.

Table 2.1. Population genetics parameters estimated with ABC. For each parameter, we present the median, mean and mode of the estimated posterior distribution. Parameters in bold correspond to comparisons between European and North African Neolithic populations. In italic we indicate the most accurate measure (median, mean or mode) according to Table S10. 95% HPDI indicates the 95% highest posterior probability.

Process	Population genetics parameter	Median	Mean	Mode	95% HPDI
Time	Initial Expansion of modern humans (T_{START})[a]	100,313	*100,182*	88,041	81,048–119,050
	Out of Africa (T_{OOA})[a]	66,313	65,864	*75,516*	50,899–79,300
	Initial Neolithic Expansion (T_{NEO})[a]	10,425	*10,279*	11,321	8,174–11,900
Population size and demographics	Ancestral size Palaeolithic (N_{ANC})	13,966	*13,428*	16,687	5,726–19,625
	Ancestral size Neolithic (N_{ANCNEO})	26,917	27,354	*21,314*	6,272–49,117
	Population growth rate Palaeolithic (G_{PALEO})	0.671	*0.662*	0.730	0.414–0.890
	Population growth rate Neolithic (G_{NEO})	0.671	0.665	*0.836*	0.416–0.888
Migration	Migration rate Palaeolithic (M_{PALEO})	0.205	0.191	*0.233*	0.075–0.248
	Migration rate Neolithic in Europe (M_{NEOEUR})	**0.210**	**0.199**	***0.232***	**0.092–0.248**
	Migration rate Neolithic in North Africa (M_{NEONA})	**0.192**	**0.183**	***0.231***	**0.074–0.248**
	Migration rate Neolithic in Rest of the World (M_{NEOPER})	0.167	0.163	*0.223*	0.056–0.247
	Migration rate at Strait of Gibraltar - North to South route (M_{nsSG})	0.149	0.151	*0.138*	0.056–0.245
	Migration rate at Strait of Gibraltar - South to North route (M_{snSG})	0.153	0.152	*0.122*	0.055–0.245
Carrying capacity	Carrying capacity Palaeolithic (K_{PALEO})	1,477	1,414	*1,845*	571–1,970
	Carrying capacity Neolithic in Europe (K_{NEOEUR})	**5,173**	**4,996**	***6,315***	**2,285–6,928**
	Carrying capacity Neolithic in North Africa (K_{NEONA})	**5,034**	**4,841**	***6,155***	**2,280–6,920**
	Carrying capacity Neolithic in Rest of the World (K_{NEOPER})	4,681	*4,655*	6,384	2,224–6,863
Admixture	Interbreeding rate (I_{rate})	0.015	0.017	*0.011*	0.004–0.042

[a] Time is shown in years, assuming a 25 year generation time.

Migration through the Strait of Gibraltar during the Neolithic is abundantly verified by archaeological and genetic data (Hervella et al. 2016), even in studies based on uniparental markers (Bosch et al. 2001; Gonzalez et al. 2003; González-Pérez et al. 2003; Cruciani et al. 2004; Ennafaa et al. 2009; Rhouda et al. 2009) and genome-wide data (Botigué et al. 2013; Arauna et al. 2017). Although we also detected migration through this strait, we found that the Strait of Gibraltar was a barrier to gene flow since the estimated migration rate through this strait was lower than the estimated migration rates in Europe or in the North of Africa (Table 2.1). Our results support a reciprocal gene flow through the Strait of Gibraltar, which suggests European influence on the genetic diversity of North African populations (and *vice versa*), increasing their heterogeneity. Concerning North Africa, this is in agreement with the recent study by Arauna et al. (2017), which showed that North African populations present, in general, high levels of genetic heterogeneity and low geographical structure. Note that North Africa has a peculiar geographic landscape that contributed for its complex genetic diversity, limited by the Sahara desert and the Mediterranean Sea and connected to the Middle East through the Sinai and Arabian Peninsulas. Both the Sahara desert and the Mediterranean Sea operated as permeable barriers, not completely blocking the contact between North Africa and the rest of the African continent and Europe (Henn et al. 2012; Botigué et al. 2013; Arauna et al. 2017). Those migrations, together with the particular above described migration through the Strait of Gibraltar, played an important role on current genetic diversity of North African populations (Henn et al. 2012; Arauna et al. 2017). By contrast, Southern Europe might have presented a lower number of different migration routes, leading to a higher population structure (Hofmanová et al. 2016; Baker et al. 2017).

The parameter estimation showed that Neolithic migration and carrying capacity in Europe and North Africa were very similar, even presenting comparable speeds of expansion on both regions. This finding fits well with the existence of: *i*) maritime migrations that could potentiate a regular contact between both shores (Fernández et al. 2014; Paschou et al. 2014) and *ii*) favourable climatic conditions in both Southern Europe and North Africa that facilitated the adoption of the farmer lifestyle (Dunne et al. 2012; Manning and Timpson 2014; Olalde et al. 2015). Archaeological data shows that the introduction of the Neolithic in the Iberian Peninsula involved a pioneer expansion of the first farmers by sea (Zilhão 2001) and that this rapid expansion most likely occurred through long-distance dispersal (LDD) events along the coast (Isern et al. 2017). Furthermore, the archaeological evidence in North Africa and the Iberian Peninsula suggests a simultaneous arrival of the Neolithic by seafarers based on similar pottery and manufacturer techniques in both regions (Cortés Sánchez et al.

2012; Linstadter et al. 2012). Also, these results suggest a regular contact between western Mediterranean populations, which could be favoured by their geographical proximity (through the Strait of Gibraltar) and are in agreement with our selected evolutionary scenario of reciprocal migration between both shores.

Regarding the speed of the Neolithic range expansion, we performed a set of computer simulations based on the estimated posterior distributions (within the 95% HPDI) to estimate the dates of Neolithic arrival to the Iberian Peninsula and to perform comparisons with other studies (Isern et al. 2017). We found that the timing based on empirical data (Cortés Sánchez et al. 2012; Linstadter et al. 2012; Bernabeu Aubán et al. 2015; Isern et al. 2017) fell within our estimated time intervals, which indeed may provide an additional validation of our models. Next, our models generated a similar time of arrival of Agriculture at the Iberian Peninsula and Western North Africa, which is in agreement with radiocarbon dates of Neolithic plants found in both regions (Morales et al. 2013; Isern et al. 2017).

The transition from hunter-gatherer to farmer lifestyle is still shrouded in controversy with two main hypotheses proposed for the way it took place: cultural or demic diffusion. Archaeological and classical genetic analyses of the European Neolithic expansion suggested a more influential role of the demic diffusion at the continental level (Cavalli-Sforza et al. 1993; Cavalli-Sforza et al. 1995), while the cultural diffusion had a secondary role, despite possible regional variations (Fort 2012; Rasteiro and Chikhi 2013; Fort et al. 2015). Recent whole-genome data (Lazaridis et al. 2014; Lazaridis et al. 2016) reinforces the idea of the Neolithic demic diffusion from the Near East and differential admixture with hunter-gatherers in Europe. The spread of the Neolithic in Europe was estimated to be around 1.0 km per year (Pinhasi et al. 2005), which is in agreement with predictions for the demic diffusion model. Our results suggest a similar scenario for the Neolithic expansion in North Africa considering that this region presented a similar migration rate to the corresponding one in Europe (Table 2.1). However, despite our estimates of a similar general velocity of the Neolithic front-wave in Europe and in North Africa, the Neolithic transition of Northern African populations could have been more heterogeneous. Archaeological evidence denotes a longer period of coexistence of hunter-gatherers and farmers in North Africa than in the Iberian Peninsula, which could be attributed to harsher environmental conditions for crops and cattle production and could lead to a slower Neolithization process by cultural assimilation (Linstadter et al. 2012). Indeed, genetic data shows some discontinuity between both shores of the Mediterranean (Botigué et al. 2013) despite the regular contact between the two regions, at least since the emergence of the Neolithic, as indicated by our results and archaeological studies (Linstadter et al. 2012; Mulazzani et al. 2016). Our results follow those based on archaeological data (in terms of timing), which might suggest that the Neolithic cultural component was more important for the contact between both shores than a genetic exchange.

Interestingly, we did not find significant differences between migration rates of Palaeolithic and Neolithic populations despite the larger carrying capacities and ancestral sizes presented in the Neolithic. These results, which are in agreement with previous studies (e.g., Pinhasi et al. 2005), could be caused by a space-competition between Mesolithic and Neolithic populations leading to a slowdown of the front wave transition of the Neolithic (Fort and Méndez 1999; Isern and Fort 2010; Isern et al. 2012).

Although the X chromosome is usually ignored in population genetics studies, it encloses some useful features (in comparison to genetic markers on the autosomes, Y chromosome and mtDNA) that we took advantage while performing this study. The X chromosome has a lower recombination rate than the autosomes due to its particular inheritance pattern, with recombination in the X chromosome non-pseudoautosomal region only occurring in females, resulting in the presence of long chromosomal segments that share a common history. Furthermore, haplotypes can be directly inferred in male individuals. In contrast to the Y chromosome and mtDNA, genetic drift has a lower effect on X chromosome markers while having a larger number of loci with potential different histories (Schaffner 2004). Additionally, male haplotypes of a given generation are the result of meiotic recombination in females in the previous generation and half of the X chromosomes in females in the following generation, meaning that for studies involving a large number of generations, problems concerning sex-bias are mostly avoided indicating that results would not be substantially different using autosomal markers. Indeed, Goldberg and Rosenberg (2015) demonstrated that, under a model of constant migration rate with no sex-bias, the admixture fractions for both autosomes and X chromosome are very similar when considering a large number of generations. Interestingly, a recent study based on three X chromosome regions with little or no recombination where European populations were analysed (Delser et al. 2017), showed evidence of pre-Neolithic expansions at 15-18 kya and diversity in one of the studied regions supported the existence of a Neolithic expansion around 10 kya. This is in line with our model for a post LGM recolonization of Europe and a Near East Neolithic expansion, respectively. In conclusion, our study and other recent studies

(Goldberg and Rosenberg 2015; Delser et al. 2017) showed that the X chromosome per se can provide a large amount of useful information for population genetics analyses aimed to dissect the complex genetic history of modern humans, not merely as a complement for other genetic markers.

2.5. Acknowledgments

We would like to thank Lara R. Arauna (Universitat Pompeu Fabra) for technical advice with the datasets used. IPATIMUP integrates the i3S Research Unit, which is partially supported by FCT in the framework of the project “Institute for Research and Innovation in Health Sciences” (POCI-01-0145-FEDER-007274). JP, AML and MA and are funded by the Portuguese Government through the FCT fellowship SFRH/BD/97200/2013 and research contracts IF/01262/2014 and IF/00955/2014, respectively. MA was also supported by the “Ramón y Cajal” grant RYC-2015-18241 from the Spanish Government. DC was supported by the Spanish MINECO grant CGL2013-44351-P.

2.6. References

Adkins J, deMenocal P, Eshel G. 2006. The “African humid period” and the record of marine upwelling from excess 230Th in Ocean Drilling Program Hole 658C. Paleoceanography 21.

Aeschbacher S, Beaumont MA, Futschik A. 2012. A novel approach for choosing summary statistics in approximate Bayesian computation. Genetics 192:1027–1047.

Alves I, Arenas M, Currat M, Hanulova AS, Sousa VC, Ray N, Excoffier L. 2016. Long-distance dispersal shaped patterns of human genetic diversity in Eurasia. Mol. Biol. Evol. 33:946–958.

Ammerman AJ, Cavalli-Sforza L. 1984. The Neolithic Transition and the Genetics of Populations in Europe. (Press PU, editor.). Princeton University Press

Arauna LR, Mendoza-Revilla J, Mas-Sandoval A, Izaabel H, Bekada A, Benhamamouch S, Fadhlaoui-Zid K, Zalloua P, Hellenthal G, Comas D, et al. 2017. Recent historical migrations have shaped the gene pool of Arabs and Berbers in North Africa. Mol. Biol. Evol. 34:318–329.

Arenas M. 2012. Simulation of molecular data under diverse evolutionary scenarios. PLoS Comput. Biol. 8.

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.

Arenas M, Ray N, Currat M, Excoffier L. 2012. Consequences of range contractions and range shifts on molecular diversity. Mol. Biol. Evol. 29:207–218.

Baker KH, Gray HWI, Ramovs V, Mertzanidou D, Akin Peksen Ç, Bilgin CC, Sykes N, Hoelzel AR. 2017. Strong population structure in a species manipulated by humans since the Neolithic: the European fallow deer (Dama dama dama). Heredity (Edinb). 119:16–26.

Barbujani G, Sokal RR, Oden NL. 1995. Indo-European origins - A computer simulation test of five hypotheses. Am. J. Phys. Anthropol. 96:109–132.

Barich BE. 2014. Northwest Libya from the early to late Holocene: New data on environment and subsistence from the Jebel Gharbi. Quat. Int. 320:15–27.

Beaumont MA. 2010. Approximate Bayesian Computation in Evolution and Ecology. Annu. Rev. Ecol. Evol. Syst. 41:379–406.

Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. Genetics 162:2025–2035.

Benguigui M, Arenas M. 2014. Spatial and temporal simulation of human evolution. Methods, frameworks and applications. Curr. Genomics 15:245–255.

Bernabeu Aubán J, Michael Barton C, Pardo Gordó S, Bergin SM. 2015. Modeling initial Neolithic dispersal. The first agricultural groups in West Mediterranean. Ecol. Modell. 307:22–31.

Bocquet-Appel J-P. 2011. When the world’s population took off: the springboard of the Neolithic Demographic Transition. Science 333:560–561.

Bocquet-Appel J-P, Demars PY. 2000. Neanderthal contraction and modern human colonization of Europe. Antiquity 74:544–552.

Bocquet-Appel JP, Naji S, Vander Linden M, Kozlowski J. 2012. Understanding the rates of expansion of the farming system in Europe. J. Archaeol. Sci. 39:531–546.

Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, Bertranpetit J. 2001. High-Resolution Analysis of Human Y-Chromosome Variation Shows a Sharp Discontinuity and Limited Gene Flow between Northwestern Africa and the Iberian Peninsula. Am. J. Hum. Genet 68:1019–1029.

Botigué LR, Henn BM, Gravel S, Maples BK, Gignoux CR, Corona E, Atzmon G, Burns E, Ostrer H, Flores C, 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.

Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, Tambets K, Antanaitis-Jacobs I, Haidle MN, Jankauskas R, Kind C-J, et al. 2009. Genetic Discontinuity

Between Local Hunter-Gatherers and Central Europe's First Farmers. Science (80-.). 326:137–140.

Broushaki F, Thomas MG, Link V, López S, Dorp L Van, Kirsanow K, Hofmanová Z, Diekmann Y, Cassidy LM, Díez-del-molino D, et al. 2016. Early Neolithic genomes from the eastern Fertile Crescent. Science (80-.). 353:499–503.

Cavalli-Sforza L, Menozzi P, Piazza A. 1995. The History and Geography of Human Gene. (Press PU, editor.).

Cavalli-Sforza LL, Menozzi P, Piazza A. 1993. Demic expansions and human evolution. Science (80-.). 259:639–646.

Clarke J, Brooks N, Banning EB, Bar-Matthews M, Campbell S, Clare L, Cremaschi M, di Lernia S, Drake N, Gallinaro M, et al. 2016. Climatic changes and social transformations in the Near East and North Africa during the "long" 4th millennium BC: A comparative study of environmental and archaeological evidence. Quat. Sci. Rev. 136.

Cortés Sánchez M, Jiménez Espejo FJ, Simón Vallejo MD, Gibaja Bao JF, Carvalho AF, Martinez-Ruiz F, Gamiz MR, Flores JA, Paytan A, López Sáez JA, et al. 2012. The Mesolithic-Neolithic transition in southern Iberia. Quat. Res. 77:221–234.

Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, Colomb EB, Zaharova B, et al. 2004. Phylogeographic Analysis of Haplogroup E3b (E-M215) Y Chromosomes Reveals Multiple Migratory Events Within and Out Of Africa. Am. J. Hum. Genet. 74:1014–1022.

Csilléry K, François O, Blum MGB. 2012. Abc: An R package for approximate Bayesian computation (ABC). Methods Ecol. Evol. 3:475–479.

Currat M, Excoffier L. 2004. Modern humans did not admix with Neanderthals during their range expansion into Europe. PLoS Biol. 2.

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, Excoffier L. 2011. Strong reproductive isolation between humans and Neanderthals inferred from observed patterns of introgression. Proc. Natl. Acad. Sci. U. S. A. 108:15129–15134.

Currat M, Poloni ES, Sanchez-Mazas A. 2010. Human genetic differentiation across the Strait of Gibraltar. BMC Evol. Biol. 10:237–255.

Currat M, Ray N, Excoffier L. 2004. SPLATCHE: A program to simulate genetic diversity taking into account environmental heterogeneity. Mol. Ecol. Notes 4:139–142.

Currat M, Ruedi M, Petit RJ, Excoffier L. 2008. The hidden side of invasions: Massive introgression by local genes. Evolution (N. Y). 62:1908–1920.

Delaneau O, Howie B, Cox AJ, Zagury JF, Marchini J. 2013. Haplotype estimation using sequencing reads. Am. J. Hum. Genet. 93:687–696.

Delser PM, Neumann R, Ballereau S, Hallast P, Batini C, Zadik D, Jobling MA. 2017. Signatures of human European Palaeolithic expansion shown by resequencing of non recombining X-chromosome segments. Eur. J. Hum. Genet. 25:485–492.

Deshpande O, Batzoglou S, Feldman MW, Cavalli-Sforza LL. 2009. A serial founder effect model for human settlement out of Africa. Proc. R. Soc. B Biol. Sci. 276:291–300.

Dunne J, Evershed RP, Salque MM, Cramp L, Bruni S, Ryan K, Biagetti S, di Lernia S. 2012. First dairying in green Saharan Africa in the fifth millennium BC. Nature 486:390–394.

Ennafaa H, M Cabrera V, K Abu-Amero K, M González A, B Amor M, Bouhaha R, Dzimiri N, Ben Ammar-Elgaaïed A, M Larruga J. 2009. Mitochondrial DNA haplogroup H structure in North Africa. BMC Genet. 10:8.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10:564–567.

Fagundes NJ, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L. 2007. Statistical evaluation of alternative models of human evolution. Proc Natl Acad Sci U S A 104:17614–17619.

Ferembach D. 1985. On the origin of the iberomaurusians (Upper palaeolithic: North Africa). A new hypothesis. J. Hum. Evol. 14:393–397.

Fernández E, Pérez-Pérez A, Gamba C, Prats E, Cuesta P, Anfruns J, Molist M, Arroyo-Pardo E, Turbón D. 2014. Ancient DNA Analysis of 8000 B.C. Near Eastern Farmers Supports an Early Neolithic Pioneer Maritime Colonization of Mainland Europe through Cyprus and the Aegean Islands. PLoS Genet. 10:e1004401.

Fort J. 2012. Synthesis between demic and cultural diffusion in the Neolithic transition in Europe. Proc. Natl. Acad. Sci. U. S. A. 109:18669–18673.

Fort J, Crema ER, Madella M. 2015. Modeling Demic and Cultural Diffusion: An Introduction. Hum. Biol. 87:141–149.

Fort J, Méndez V. 1999. Time-Delayed Theory of the Neolithic Transition in Europe. Phys. Rev. Lett. 82:867–870.

François O, Currat M, Ray N, Han E, Excoffier L, Novembre J. 2010. Principal component analysis under population genetic models of range expansion and admixture. Mol Biol Evol 27:1257–1268.

Fu Q, Rudan P, Pääbo S, Krause J. 2012. Complete mitochondrial genomes reveal neolithic expansion into Europe. PLoS One 7.

Gallego-Llorente M, Connell S, Jones ER, Merrett DC, Jeon Y, Eriksson A, Siska V, Gamba C, Meiklejohn C, Beyer R, et al. 2016. The genetics of an early Neolithic pastoralist from the Zagros, Iran. Sci. Rep. 6:31326.

Goldberg A, Rosenberg NA. 2015. Beyond 2/3 and 1/3: The complex signatures of sex-biased admixture on the X chromosome. Genetics 201:263–279.

González-Pérez E, Via M, Esteban E, López-Alomar A, Mazieres S, Harich N, Kandil M, Dugoujon J-M, Moral P. 2003. Alu insertions in the Iberian Peninsula and north west Africa--genetic boundaries or melting pot? Coll Antropol 27:491–500.

Gonzalez AM, Brehm A, Pérez JA, Maca-Meyer N, Flores C, Cabrera VM. 2003. Mitochondrial DNA affinities at the atlantic fringe of Europe. Am. J. Phys. Anthropol. 120:391–404.

Goring-Morris AN, Belfer-Cohen A. 2011. Neolithization Processes in the Levant: The Outer Envelope. Curr. Anthropol. 52:S195–S208.

Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K, et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522:207–211.

Hamilton MJ, Milne BT, Walker RS, Burger O, Brown JH. 2007. The complex structure of hunter-gatherer social networks. Proc. Biol. Sci. 274:2195–2202.

Henn BM, Botigué LR, Gravel S, Wang W, Brisbin A, Byrnes JK, Fadhlaoui-Zid K, Zalloua PA, Moreno-Estrada A, Bertranpetit J, et al. 2012. Genomic ancestry of North Africans supports back-to-Africa migrations. PLoS Genet. 8:e1002397.

Hervella M, Svensson EM, Alberdi A, Günther T, Izagirre N, Munters AR, Alonso S, Ioana M, Ridiche F, Soficaru A, et al. 2016. The mitogenome of a 35,000-year-old Homo sapiens from Europe supports a Palaeolithic back-migration to Africa. Sci. Rep. 6:25501.

Hofmanová Z, Kreutzer S, Hellenthal G, Sell C, Diekmann Y, Díez-del-Molino D, van Dorp L, López S, Kousathanas A, Link V, et al. 2016. Early farmers from across Europe directly descended from Neolithic Aegeans. Proc. Natl. Acad. Sci. 113:6886–6891.

Isern N, Fort J. 2010. Anisotropic dispersion, space competition and the slowdown of the Neolithic transition. New J. Phys. 12.

Isern N, Fort J, Vander Linden M. 2012. Space Competition and Time Delays in Human Range Expansions. Application to the Neolithic Transition. PLoS One 7.

Isern N, Zilhão J, Fort J, Ammerman AJ. 2017. Modeling the role of voyaging in the coastal spread of the Early Neolithic in the West Mediterranean. Proc. Natl. Acad. Sci. 114:897–902.

Kimura M, Weiss GH. 1964. The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance. Genetics 49:561–576.

Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, Fernandes D, Novak M, Gamarra B, Sirak K, et al. 2016. Genomic insights into the origin of farming in the ancient Near East. Nature 536.

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Sudmant PH, Schraiber JG, Castellano S, Kirsanow K, Economou C, et al. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513:409–413.

Linseele V, Van Neer W, Thys S, Phillipps R, Cappers R, Wendrich W, Holdaway S. 2014. New archaeozoological data from the fayum “neolithic” with a critical assessment of the evidence for early stock keeping in egypt. PLoS One 9.

Linstadter J, Medved I, Solich M, Weniger GC. 2012. Neolithisation process within the Alboran territory: Models and possible African impact. Quat. Int. 274:219–232.

Liu H, Prugnolle F, Manica A, Balloux F. 2006. A geographically explicit genetic model of worldwide human-settlement history. Am. J. Hum. Genet. 79:230–237.

Lopes AM, Aston KI, Thompson E, Carvalho F, Gonçalves J, Huang N, Matthiesen R, Noordam MJ, Quintela I, Ramu A, et al. 2013. Human Spermatogenic Failure Purges Deleterious Mutation Load from the Autosomes and Both Sex Chromosomes, including the Gene DMRT1. PLoS Genet. 9.

Madella M, Garc??a-Granero JJ, Out WA, Ryan P, Usai D. 2014. Microbotanical evidence of domestic cereals in Africa 7000 years ago. PLoS One 9.

Manning K, Timpson A. 2014. The demographic response to holocene climate change in the Sahara. Quat. Sci. Rev. 101:28–35.

McGee D, deMenocal PB, Winckler G, Stuut JBW, Bradtmiller LI. 2013. The magnitude, timing and abruptness of changes in North African dust deposition over the last 20,000 yr. Earth Planet. Sci. Lett. 371–372:163–176.

McVean G a T, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. 2004. The fine-scale structure of recombination rate variation in the human genome. Science 304:581–584.

Meiklejohn C, Merrett DC, Reich D, Pinhasi R. 2017. Direct dating of human skeletal material from Ganj Dareh, Early Neolithic of the Iranian Zagros. J. Archaeol. Sci. Reports 12:165–172.

Morales J, Pérez-Jordà G, Peña-Chocarro L, Zapata L, Ruíz-Alonso M, López-Sáez JA, Linstädter J. 2013. The origins of agriculture in North-West Africa: Macro-botanical remains from Epipalaeolithic and Early Neolithic levels of Ifri Oudadane (Morocco). J. Archaeol. Sci. 40:2659–2669.

Mulazzani S, Belhouchet L, Salanova L, Aouadi N, Dridi Y, Eddargach W, Morales J, Tombret O, Zazzo A, Zoughlami J. 2016. The emergence of the Neolithic in North Africa: A new model for the Eastern Maghreb. Quat. Int. 410:123–143.

Olalde I, Schroeder H, Sandoval-Velasco M, Vinner L, Lobon I, Ramirez O, Civit S, Borja PG, Salazar-Garcia DC, Talamo S, et al. 2015. A common genetic origin for early farmers from mediterranean cardial and central european LBK cultures. Mol. Biol. Evol. 32:3132–3142.

Paschou P, Drineas P, Yannaki E, Razou A, Kanaki K, Tsetsos F, Padmanabhuni SS, Michalodimitrakis M, Renda MC, Pavlovic S, et al. 2014. Maritime route of colonization of Europe. Proc. Natl. Acad. Sci. U. S. A. 111:9211–9216.

Pinhasi R, Fort J, Ammerman AJ. 2005. Tracing the origin and spread of agriculture in Europe. PLoS Biol. 3:1–9.

Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol 16:1791–1798.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575.

Rasteiro R, Chikhi L. 2013. Female and Male Perspectives on the Neolithic Transition in Europe: Clues from Ancient and Modern Genetic Data. PLoS One 8.

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.

Rhouda T, Martínez-Redondo D, Gómez-Durán A, Elmtili N, Idaomar M, Díez-Sánchez C, Montoya J, López-Pérez MJ, Ruiz-Pesini E. 2009. Moroccan mitochondrial genetic background suggests prehistoric human migrations across the Gibraltar Strait. Mitochondrion 9:402–407.

Ruhlen M. 1991. A Guide to the World’s Languages. (Press SU, editor.).

Sampietro ML, Lao O, Caramelli D, Lari M, Pou R, Marti M, Bertranpetit J, Lalueza-Fox

C. 2007. Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. Proc Biol Sci 274:2161–2167.

Schaffner SF. 2004. The X chromosome in population genetics. Nat. Rev. Genet. 5:43–51.

Skoglund P, Malmström H, Raghavan M, Storå J, Hall P, Willerslev E, Gilbert MTP, Götherström A, Jakobsson M, Menozzi P, et al. 2012. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. Science 336:466–469.

Sokal RR, Oden NL, Wilson C. 1991. Genetic evidence for the spread of agriculture in Europe by demic diffusion. Nature 351:143–145.

The 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1 , 092 human genomes. Nature 491:56–65.

The International Hapmap Consortium. 2003. The International HapMap Project. Nature 426:789–796.

Tillmar AO, Mostad P, Egeland T, Lindblom B, Holmlund G, Montelius K. 2008. Analysis of linkage and linkage disequilibrium for eight X-STR markers. Forensic Sci. Int. Genet. 3:37–41.

Timmermann A, Friedrich T. 2016. Late Pleistocene climate drivers of early human migration. Nature 538:92–95.

Voight BF, Adams AM, Frisse L a, Qian Y, Hudson RR, Di Rienzo A. 2005. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. Proc. Natl. Acad. Sci. U. S. A. 102:18508–18513.

Wegmann D, Excoffier L. 2010. Bayesian inference of the demographic history of chimpanzees. Mol. Biol. Evol. 27:1425–1435.

Wegmann D, Leuenberger C, Excoffier L. 2009. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. Genetics 182:1207–1218.

Yang Z. 2006. Computational Molecular Evolution. (Press OU, editor.).

Zeder MA. 2008. Domestication and early agriculture in the Mediterranean Basin: origins, diffusion, and impact. Proc. Natl. Acad. Sci. U. S. A. 105:11597–11604.

Zeder MA, Hesse B. 2000. The initial domestication of goats (Capra hircus L.) in the Zagros mountains 10,000 years ago. Science (80-.). 287:2254–2257.

Zilhão J. 2001. Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. Proc. Natl. Acad. Sci. U. S. A. 98:14180–14185.

2.7. Supplementary Material

Evaluating the Neolithic Expansion at Both Shores of the Mediterranean Sea

João Pimenta, Alexandra M. Lopes, David Comas, António Amorim and Miguel Arenas

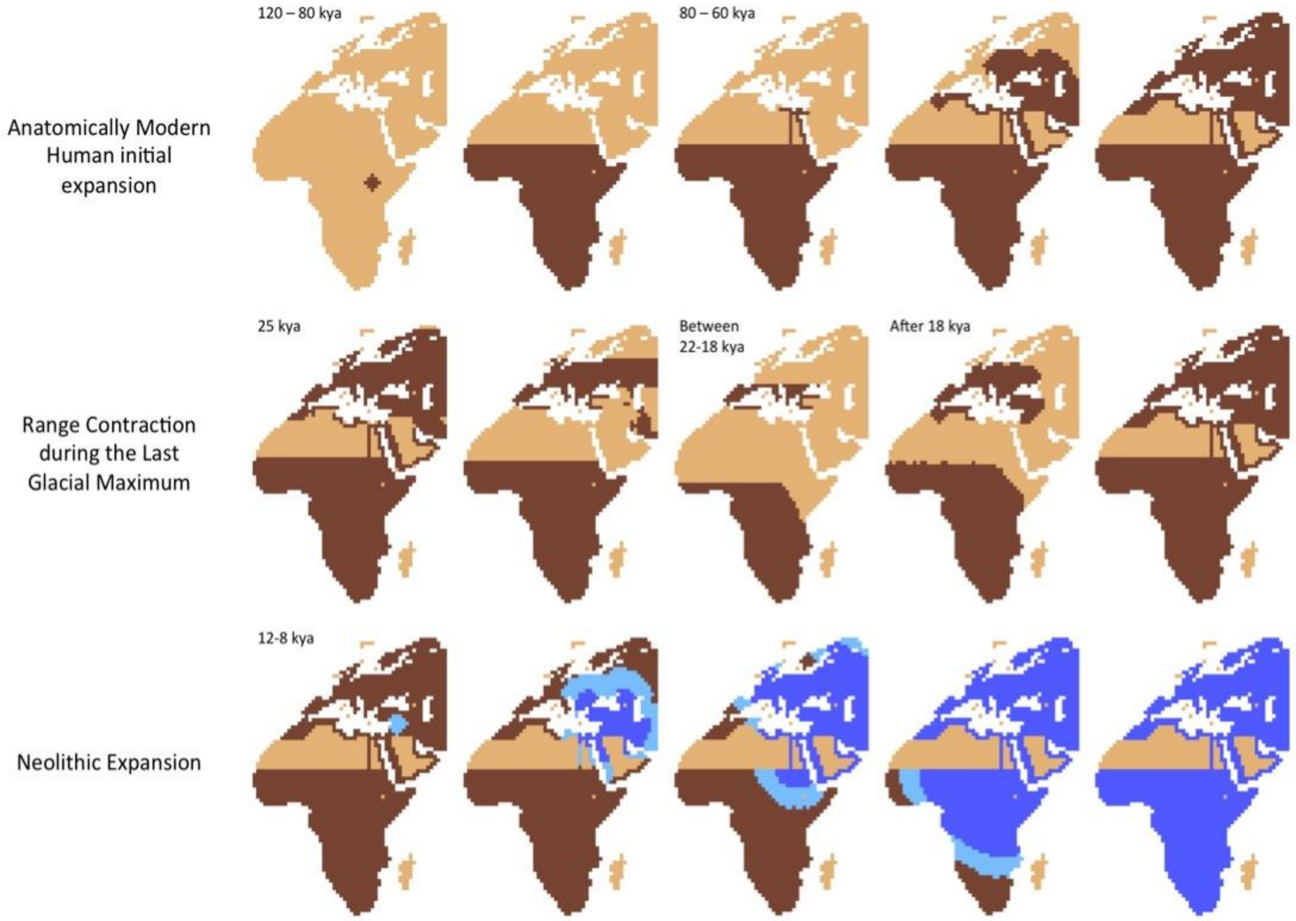


Figure S2.1. Illustrative example of the performed spatially explicit computer simulations. The first row shows the expansion of Anatomically Modern Humans (AMH, brown) throughout Sub-Saharan Africa, followed by the out-of-Africa along the Nile valley and the East African coast and the posterior colonization of Middle East and Europe. The second row shows the range contraction of these populations due to the Last Glacial Maximum (the snapshot in the middle shows refugia areas in Europe, North Africa and Sub-Saharan Africa due to the extension of the Sahara Desert (Alves et al. 2016)) and subsequent re-expansion since 18 kya. The third row shows the Neolithic expansion (blue) over Europe and Africa. The light blue colour indicates regions where Palaeolithic and Neolithic populations coexist.

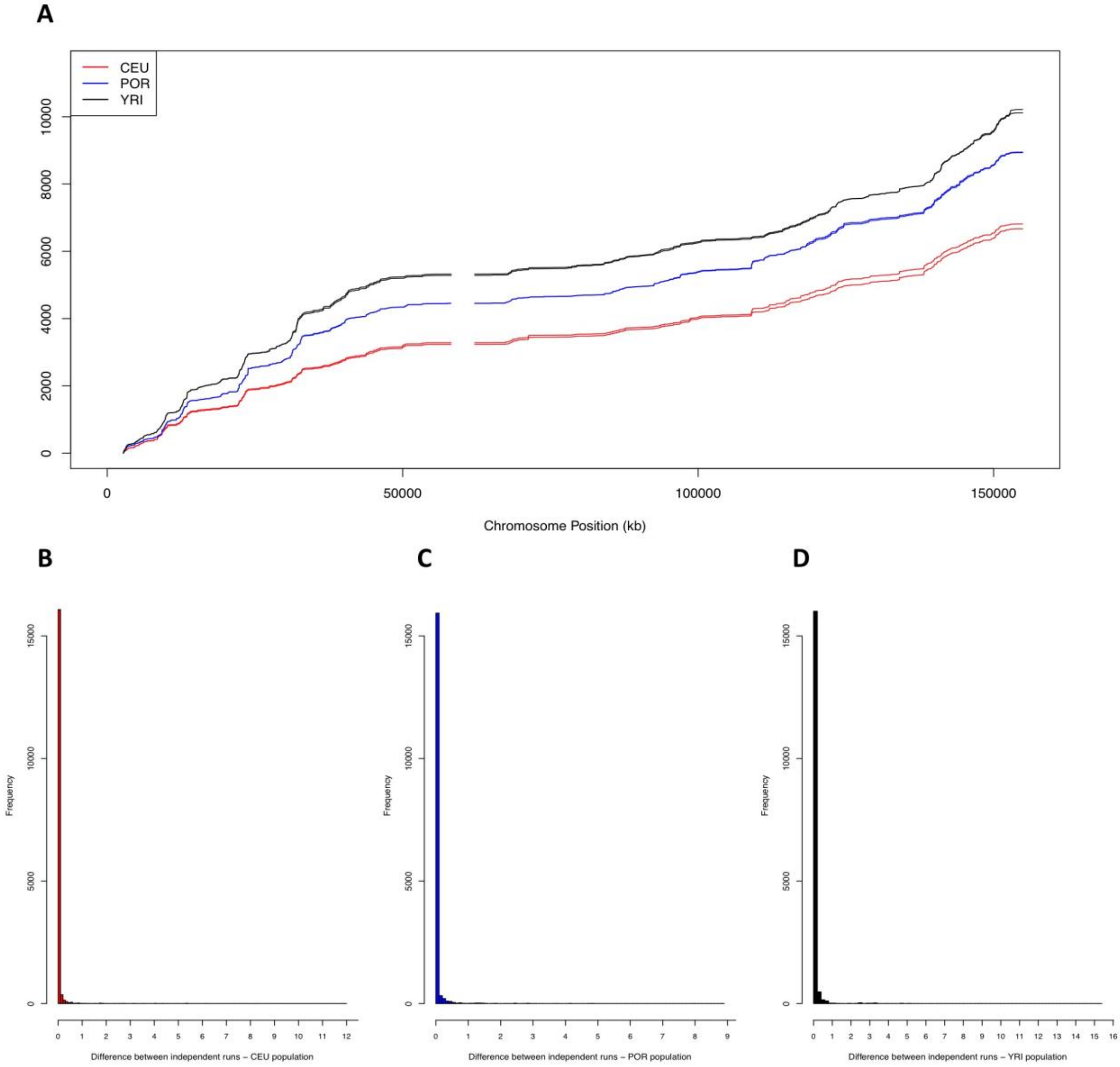


Figure S2.2. Recombination rate throughout the X chromosome. *(A)* Cumulative plots of the recombination rate ($\rho=4N_er$) estimates for the X chromosome based on 6 independent runs performed with LDhat 2.2 (McVean et al. 2004). To check for consistency the recombination rate was estimated for 3 different populations [Northern and Western Europeans currently in Utah (CEU), Porto (POR) and Yoruba (YRI)] using a block penalty of 20 (2 independent runs for each population; in red, blue and black, respectively). The line break in the plot, between 58,000 kb and 62,000 kb, corresponds to the centromeric region of the X chromosome. *(B)* Histogram of the absolute differences per position of ρ between two independent runs for CEU population. *(C)* Histogram of the absolute differences per position of ρ between two independent runs for POR population. *(D)* Histogram of the absolute differences per position of ρ between two independent runs for YRI population. Note that these ρ differences are usually below 1, suggesting a high convergence between the independent runs.
]

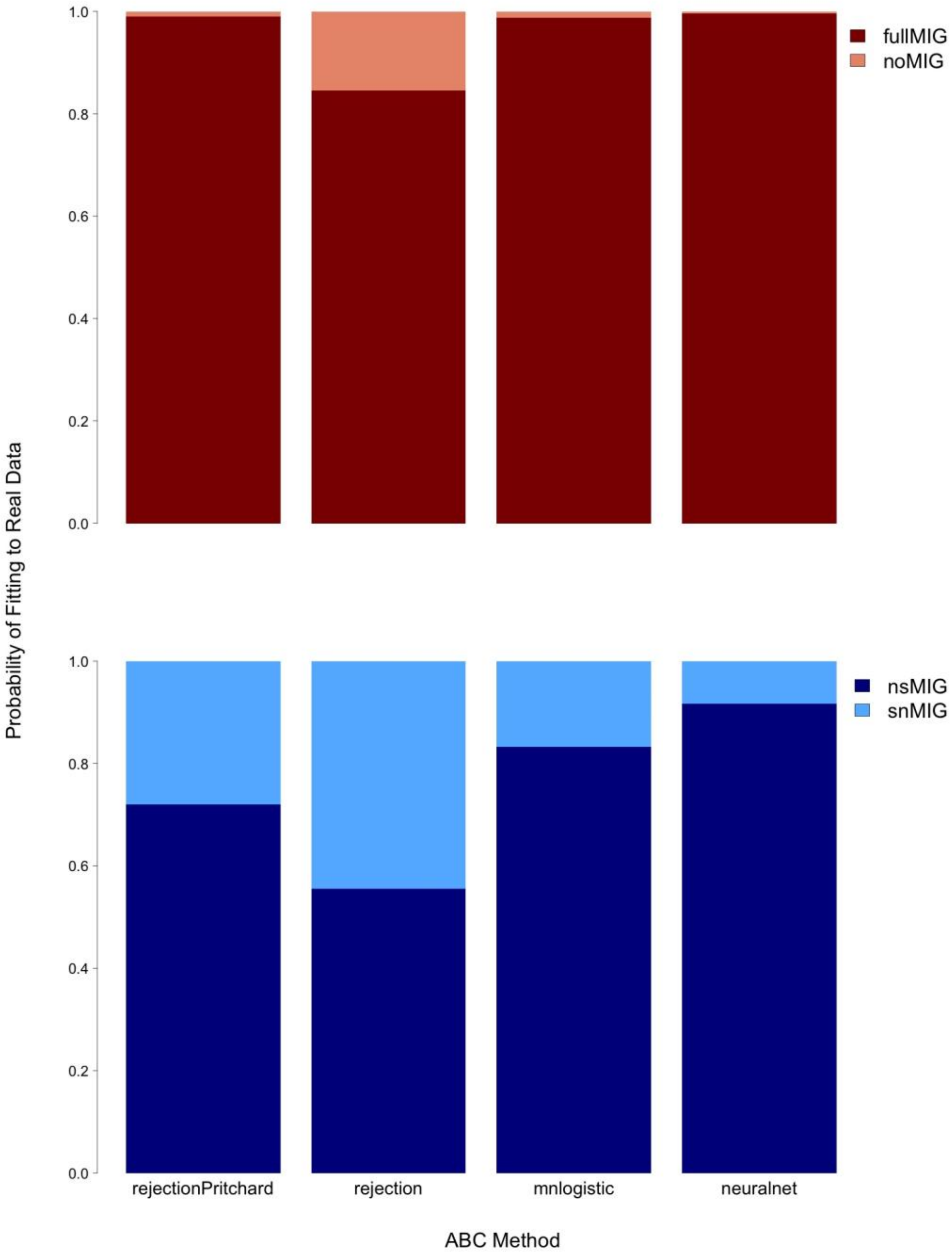


Figure S2.3. Fitting of the alternative evolutionary scenarios of migration through the Gibraltar Strait to the real data. (A) Fitting of *fullMIG* and *noMIG* evolutionary scenarios to the real data. (B) Fitting of *nsMIG* and *snMIG* evolutionary scenarios to the real data. Posterior probabilities of evolutionary scenarios were estimated with four different ABC methods (the *rejection* method by Pritchard and the *rejection*, *multinomial logistic regression* and *neural network* methods implemented in the *abc* package (Csilléry et al. 2012)).

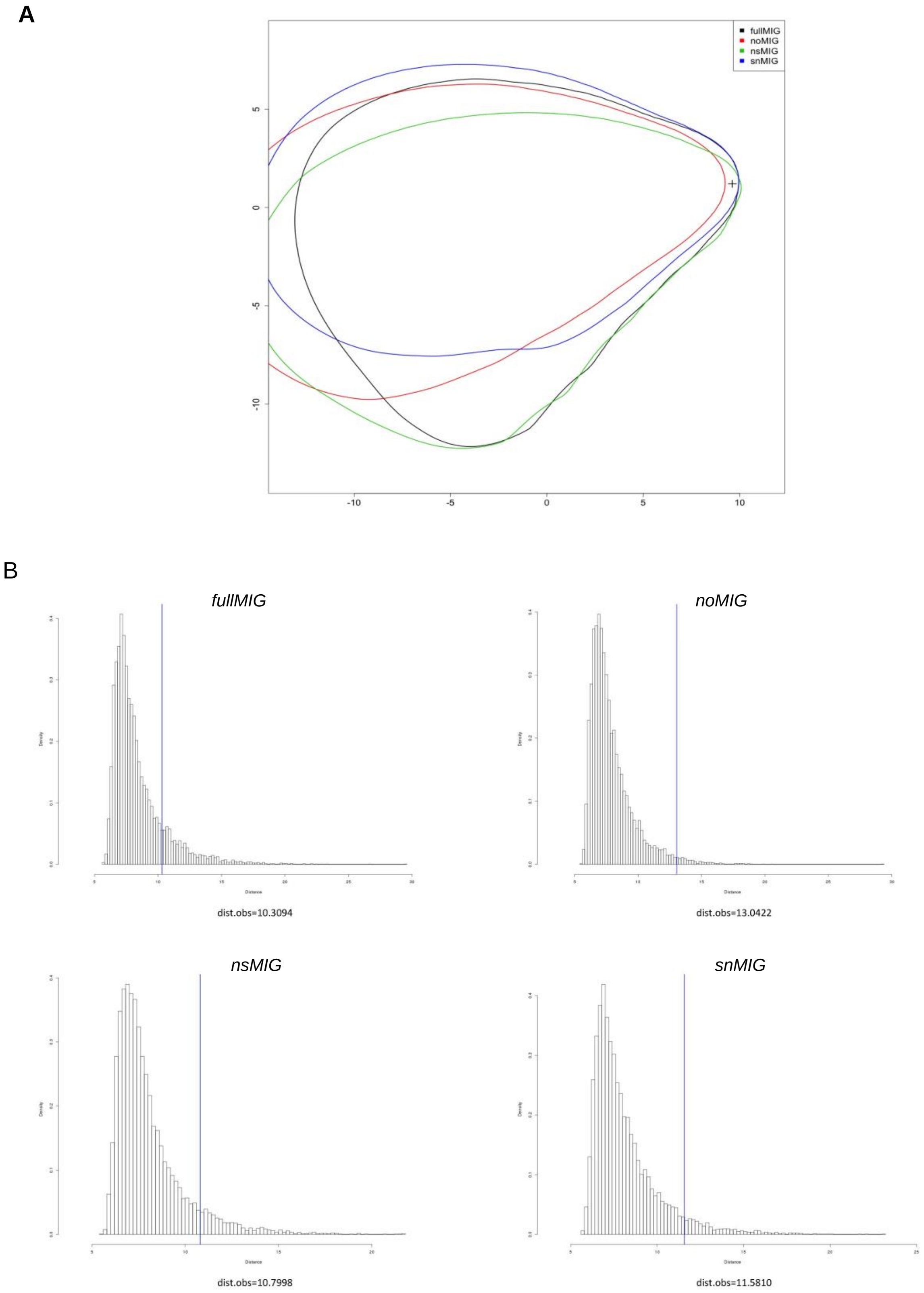


Figure S2.4. Goodness of fit for the four evolutionary scenarios with real data. Although the goodness of fit is usually applied to only the best fitting scenario before turning to parameter inference and in order to check that the preferred scenario provides a good fitting to the real data (Csilléry et al. 2012), here we analysed all the investigated scenarios to provide an additional evaluation of the fitting with the real data, but it is expected that less fitting scenarios can generate poorer goodness of fit. A: Evaluation based on the first two components of a principal component analysis where all the selected SS for the ABC analysis are considered under a p-value of 5x10-4. The cross represents the real data. B: Evaluation based on histograms of simulated

SS. The blue vertical line indicates the real SS. The preferred model (fullMIG) falls within the 90% of the distribution. These analyses followed the settings and recommendations indicated in the documentation of the package abc (Csilléry et al. 2012).

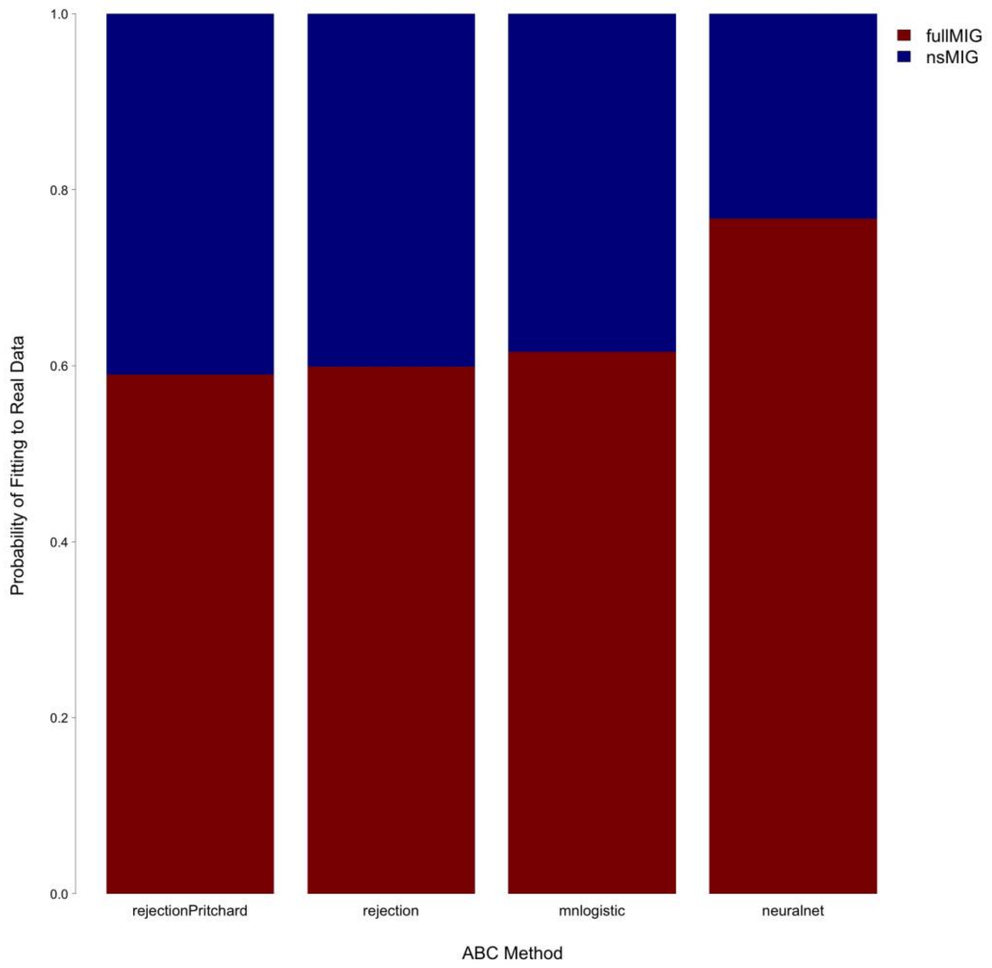


Figure S2.5. Fitting of the alternative evolutionary scenarios of migration through the Gibraltar Strait, *fullMIG* and *nsMIG,* to the real data. Posterior probabilities of evolutionary scenarios were estimated with four different ABC methods (the *rejection* method by Pritchard and the *rejection*, *multinomial logistic regression* and *neural network* methods implemented in the *abc* package (Csilléry et al. 2012)).

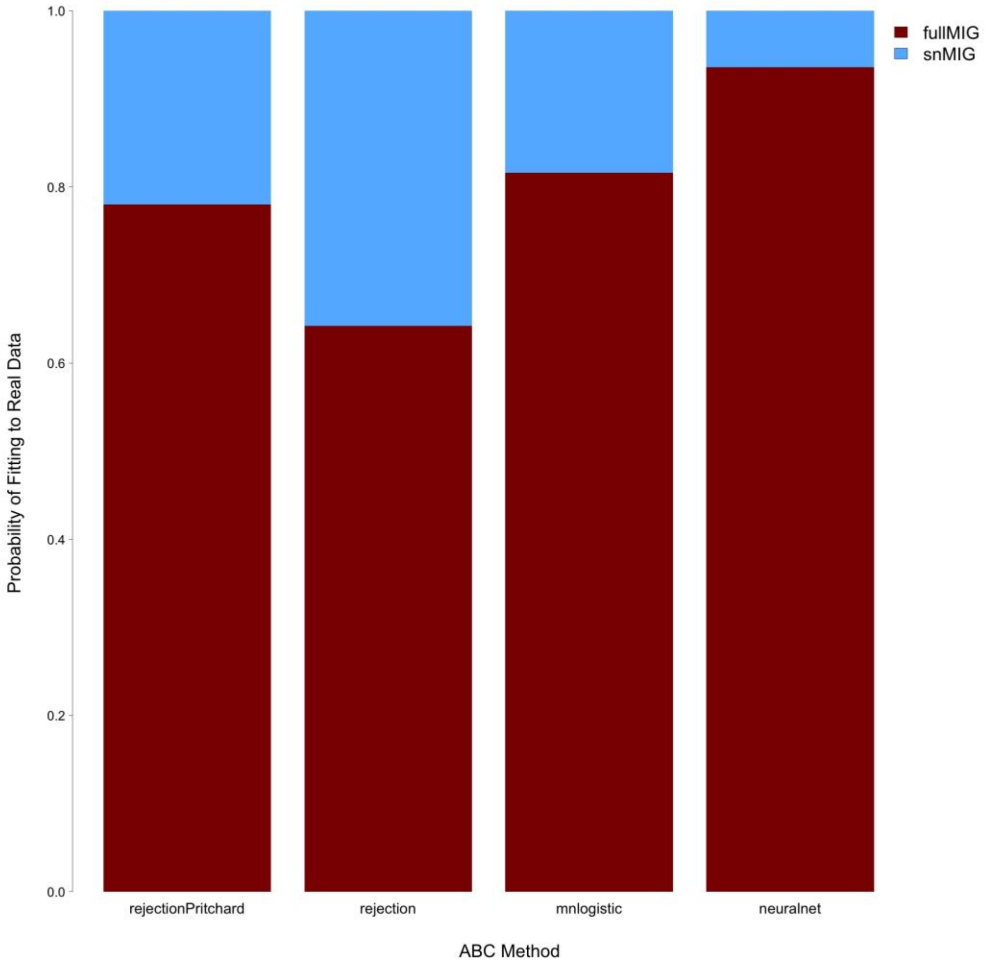


Figure S2.6. Fitting of the alternative evolutionary scenarios of migration through the Gibraltar Strait, *fullMIG* and *snMIG*, to the real data. Posterior probabilities of evolutionary scenarios were estimated with four different ABC methods (the *rejection* method by Pritchard and the *rejection*, *multinomial logistic regression* and *neural network* methods implemented in the *abc* package (Csilléry et al. 2012)).

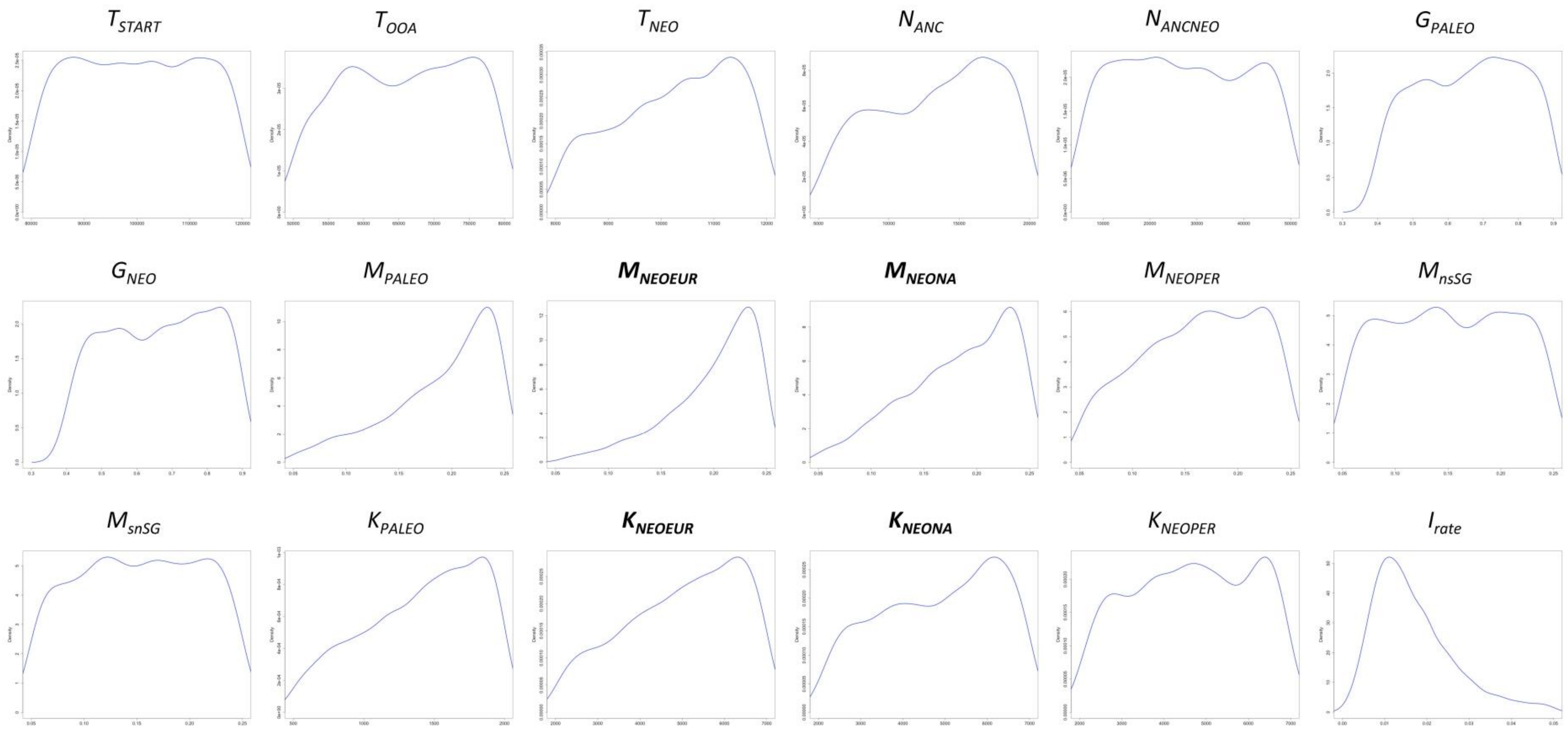


Figure S2.7. Posterior distributions of the parameters of the evolutionary scenario that fitted the data best (*fullMIG*). We retained 1,000 simulations of the closest to the observed data and used the rejection method implemented in the *abc* package (Csilléry et al. 2012). The more relevant parameters for this study are shown in bold: Neolithic migration in Europe (M_{NEOEUR}), Neolithic migration in North of Africa (M_{NEONA}), Neolithic carrying capacity in Europe (K_{NEOEUR}), Neolithic carrying capacity in North of Africa (K_{NEONA}).

Table S2.1 Details of the real dataset analysed in the present study.

Population	Total number of samples	Sample size	Reference
Europe			
Finland	79	31	1000 Genomes Project Consortium et al. 2012
Great-Britain	76	36	1000 Genomes Project Consortium et al. 2012
Northern and Western Europeans from Utah	87	46	1000 Genomes Project Consortium et al. 2012
Tuscany, Italy	96	50	1000 Genomes Project Consortium et al. 2012
Basque Country, Spain	25	23	Henn et al. 2012 and 1000 Genomes Project Consortium et al. 2012
East Spain	14	6	1000 Genomes Project Consortium et al. 2012
Central Spain	15	9	1000 Genomes Project Consortium et al. 2012
Galicia, Spain	22	19	Botigué et al. 2013 and 1000 Genomes Project Consortium et al. 2012
Porto, Portugal	124	124	Lopes et al. 2013
Lisbon, Portugal	45	45	Lopes et al. 2013
Andalusia, Spain	20	18	Botigué et al. 2013 and 1000 Genomes Project Consortium et al. 2012
North Africa			
Western Sahara	18	18	Henn et al. 2012
Morocco South	16	8	Henn et al. 2012
Morocco North	18	18	Henn et al. 2012
Algeria	19	9	Henn et al. 2012
Tunisia	18	18	Henn et al. 2012
Libya	17	16	Henn et al. 2012
Egypt	18	18	Henn et al. 2012
Sub-Saharan Africa			
Yoruba, Nigeria	100	52	1000 Genomes Project Consortium et al. 2012
Near East			
Syria	19	16	Arauna et al. 2017

Table S2.2. Prior distributions for the population genetics parameters included in all the studied evolutionary scenarios.

Parameter	Distribution	Presence/Absence in the scenarios
Initial Expansion of modern humans (T_{START})	Uniform (80,000–120,000)	All scenarios
Out of Africa (T_{OOA})	Uniform (50,000–80,000)	All scenarios
Initial Neolithic Expansion (T_{NEO})	Uniform (8,000–12,000)	All scenarios
Ancestral size Palaeolithic (N_{ANC})	Uniform (5,000–20,000)	All scenarios
Ancestral size Neolithic (N_{ANCNEO})	Uniform (5,000–50,000)	All scenarios
Population growth rate Palaeolithic (G_{PALEO})	Uniform (0.3–0.9)	All scenarios
Population growth rate Neolithic (G_{NEO})	Uniform (0.3–0.9)	All scenarios
Migration rate Palaeolithic (M_{PALEO})	Uniform (0.05–0.25)	All scenarios
Migration rate Neolithic in Europe (M_{NEOEUR})	Uniform (0.05–0.25)	All scenarios
Migration rate Neolithic in North Africa (M_{NEONA})	Uniform (0.05–0.25)	All scenarios
Migration rate Neolithic in Rest of the World (M_{NEOPER})	Uniform (0.05–0.25)	All scenarios
Migration rate at Strait of Gibraltar - North to South route (M_{nsSG})	Uniform (0.05–0.25)	All scenarios, except *snMIG*
Migration rate at Strait of Gibraltar - South to North route (M_{snSG})	Uniform (0.05–0.25)	All scenarios, except *nsMIG*
Carrying capacity Palaeolithic (K_{PALEO})	Uniform (500–2,000)	All scenarios
Carrying capacity Neolithic in Europe (K_{NEOEUR})	Uniform (2,000–7,000)	All scenarios
Carrying capacity Neolithic in North Africa (K_{NEONA})	Uniform (2,000–7,000)	All scenarios
Carrying capacity Neolithic in Rest of the World (K_{NEOPER})	Uniform (2,000–7,000)	All scenarios
Interbreeding rate (I_{rate})	Uniform (0.00–0.05)	All scenarios

Table S2.3. Considered blocks of linked loci in the X chromosome. The SNP data of the X chromosome was divided in 14 independent blocks based on recombination rates estimated with LDhat 2.2 (Auton and McVean 2007) and published results on STR recombination rates (Tillmar et al. 2008). A new block was considered for each region delimited by a recombination hotspot. A total of 4 blocks were based on Tillmar et al. 2008, which unravel these regions in the X chromosome with very little or without recombination.

Block number	Begin (kb)	End (kb)	Definition
1	2,714	9,198	SNP recombination rate
2	9,198	9,335	Tillmar et al. 2008
3	9,335	22,243	SNP recombination rate
4	22,243	32,009	SNP recombination rate
5	32,009	58,317	SNP recombination rate
6	62,058	64,579	SNP recombination rate
7	64,579	66,918	Tillmar et al. 2008
8	66,918	86,301	SNP recombination rate
9	86,301	108,922	SNP recombination rate
10	108,922	133,258	SNP recombination rate
11	133,258	133,499	Tillmar et al. 2008
12	133,499	149,313	SNP recombination rate
13	149,313	149,472	Tillmar et al. 2008
14	149,472	154,925	SNP recombination rate

Table S2.4. Chosen Summary Statistics (SS) for the ABC analyses and their values in the real dataset. The dataset of 20 populations was divided in 8 geographic groups (Figure 2.1) to reduce the number of SS estimated with Arlequin ver3.5 (Excoffier and Lischer 2010). Note that for any ABC analysis, the chosen SS are the same for simulated and real data (Beaumont et al. 2002). This set of SS was selected to provide the highest amount of information to perform selection between alternative evolutionary scenarios and parameters estimation. Their selection was based on their fitting with *test datasets*.

Summary Statistic	Description	Value in the real dataset
FST_BAS–EUR	Average number of pairwise differences between population from Basque Country and European populations	0.01169
FST_IBE–EUR	Average number of pairwise differences between populations from Iberia Peninsula and European populations	0.00497
FST_IBE–BAS	Average number of pairwise differences between populations from Iberia Peninsula and Basque Country population	0.01030
FST_WNA–EUR	Average number of pairwise differences between populations from Western North Africa and European population	0.04298
FST_WNA–BAS	Average number of pairwise differences between populations from Western North Africa and Basque Country population	0.04238
FST_WNA–IBE	Average number of pairwise differences between populations from Western North Africa and Iberia Peninsula populations	0.03817
FST_CNA–EUR	Average number of pairwise differences between populations from Central North Africa and European populations	0.02016
FST_CNA–BAS	Average number of pairwise differences between populations from Central North Africa and Basque Country population	0.02460
FST_CNA–IBE	Average number of pairwise differences between populations from Central North Africa and Iberia Peninsula populations	0.01682

FST_CNA–WNA	Average number of pairwise differences between populations from Central North Africa and Western North Africa populations	0.01445
FST_ENA–EUR	Average number of pairwise differences between populations from East North Africa and European populations	0.01804
FST_ENA–BAS	Average number of pairwise differences between populations from East North Africa and Basque Country population	0.02503
FST_ENA–IBE	Average number of pairwise differences between populations from East North Africa and Iberia Peninsula populations	0.01678
FST_ENA–WNA	Average number of pairwise differences between populations from East North Africa and Western North Africa populations	0.00823
FST_ENA–CNA	Average number of pairwise differences between populations from East North Africa and Central North Africa populations	0.00463
FST_SYR–EUR	Average number of pairwise differences between population from Syria and European populations	0.01384
FST_SYR–BAS	Average number of pairwise differences between population from Syria and Basque Country population	0.02327
FST_SYR–IBE	Average number of pairwise differences between population from Syria and Iberia Peninsula populations	0.01523
FST_SYR–WNA	Average number of pairwise differences between population from Syria and Western North Africa populations	0.02537
FST_SYR–CNA	Average number of pairwise differences between population from Syria and Central North Africa populations	0.01253
FST_SYR–ENA	Average number of pairwise differences between population from Syria and East North Africa	0.00521

	populations	
FST_YRI–EUR	Average number of pairwise differences between population from Yoruba and European populations	0.13169
FST_YRI–BAS	Average number of pairwise differences between population from Yoruba and Basque Country populations	0.13369
FST_YRI–IBE	Average number of pairwise differences between population from Yoruba and Iberia Peninsula populations	0.13059
FST_YRI–WNA	Average number of pairwise differences between population from Yoruba and Western North Africa populations	0.08242
FST_YRI–CNA	Average number of pairwise differences between population from Yoruba and Central North Africa populations	0.10672
FST_YRI–ENA	Average number of pairwise differences between population from Yoruba and East North Africa populations	0.09754
FST_YRI–SYR	Average number of pairwise differences between population from Yoruba and Syrian population	0.11840
FST_AND–BAS	Average number of pairwise differences between Andalusian population and Basque Country population	0.00744
FST_SAH–AND	Average number of pairwise differences between Western Saharan population and Andalusian population	0.03002
FST_MORN–GAL	Average number of pairwise differences between North Moroccan population and Galician population	0.01188
FST_MORN–AND	Average number of pairwise differences between North Moroccan population and Andalusian population	0.00973
FST_ALG–MORN	Average number of pairwise differences between Algerian population and North Moroccan population	0.00906

FST_TUN–LIS	Average number of pairwise differences between Tunisian population and Lisbon population	0.03857
FST_YRI–LIS	Average number of pairwise differences between Yoruba population and Lisbon population	0.12640

Table S2.5. Power of the model selection method to retrieve the *fullMIG* and *noMIG* scenarios under the four different ABC methods. The identification of the correct scenario is shown in bold. The Pritchard's rejection method was analysed with the 100 *test datasets*. The other three methods, *rejection*, *multinomial logistic regression* and *neural network* of the *abc* package, were analysed with a cross-validation process based on 100 permutations and a tolerance of 0.01 (following the recommendations described in the documentation of the package *abc* (Csilléry et al. 2012)).

	fullMIG				***noMIG***			
	Pritchard rej.	**Rejection**	**Mnlogistic**	**Neuralnet**	**Pritchard rej.**	**Rejection**	**Mnlogistic**	**Neuralnet**
fullMIG	**0.940**	**0.778**	**0.976**	**0.979**	0.060	0.222	0.024	0.021
noMIG	0.040	0.151	0.020	0.017	**0.960**	**0.849**	**0.980**	**0.983**

Table S2.6. Power of the model selection method to retrieve the *nsMIG* and *snMIG* scenarios under the four different ABC methods. The identification of the correct scenario is shown in bold. The Pritchard's rejection method was analysed with the 100 *test datasets*. The other three methods, *rejection*, *multinomial logistic regression* and *neural network* of the *abc* package, were analysed with a cross-validation process based on 100 permutations and a tolerance of 0.01 (following the recommendations described in the documentation of the package *abc* (Csilléry et al. 2012)).

	nsMIG				***snMIG***			
	Pritchard rej.	**Rejection**	**Mnlogistic**	**Neuralnet**	**Pritchard rej.**	**Rejection**	**Mnlogistic**	**Neuralnet**
nsMIG	**1.000**	**0.861**	**0.991**	**0.985**	0.000	0.139	0.009	0.015
snMIG	0.000	0.146	0.018	0.021	**1.000**	**0.854**	**0.982**	**0.979**

Table S2.7. Power of the model selection method to retrieve all four scenarios (*fullMIG vs noMIG vs nsMIG vs snMIG*) under the four different ABC methods. The identification of the correct scenario is shown in bold. The Pritchard's rejection method (PritRej) was analysed with the 100 *test datasets*. The other three methods, *rejection* (Rej), *multinomial logistic regression* (Mnl) and *neural network* (Neu) of the *abc* package, were analysed with a cross-validation process based on 100 permutations and a tolerance of 0.01 (following the recommendations described in the documentation of the package *abc* (Csilléry et al. 2012)).

	fullMIG				***noMIG***				***nsMIG***				***snMIG***			
	PritRej	Rej	Mnl	Neu	PritRej	Rej	Mnl	Neu	PritRej	Rej	Mnl	Neu	PritRej	Rej	Mnl	Neu
fullMIG	**0.580**	**0.372**	**0.562**	**0.566**	0.000	0.095	0.008	0.011	0.110	0.258	0.193	0.191	0.310	0.275	0.238	0.233
noMIG	0.010	0.079	0.017	0.010	**0.750**	**0.554**	**0.850**	**0.849**	0.210	0.233	0.095	0.110	0.030	0.134	0.037	0.031
nsMIG	0.220	0.237	0.177	0.171	0.140	0.261	0.095	0.081	**0.640**	**0.432**	**0.724**	**0.741**	0.000	0.069	0.004	0.007
snMIG	0.190	0.281	0.239	0.235	0.030	0.158	0.035	0.044	0.000	0.070	0.009	0.005	**0.780**	**0.491**	**0.716**	**0.716**

Table S2.8. Power of the model selection method to retrieve *fullMIG* and *nsMIG* scenarios under the four different ABC methods. The identification of the correct scenario is shown in bold. The Pritchard's rejection method was analysed with the 100 *test datasets*. The other three methods, *rejection*, *multinomial logistic regression* and *neural network* of the *abc* package, were analysed with a cross-validation process based on 100 permutations and a tolerance of 0.01 (following the recommendations described in the documentation of the package *abc* (Csilléry et al. 2012)).

	fullMIG				***nsMIG***			
	Pritchard rej.	**Rejection**	**Mnlogistic**	**Neuralnet**	**Pritchard rej.**	**Rejection**	**Mnlogistic**	**Neuralnet**
fullMIG	**0.850**	**0.634**	**0.815**	**0.799**	0.150	0.366	0.185	0.201
nsMIG	0.230	0.318	0.198	0.167	**0.770**	**0.682**	**0.802**	**0.833**

Table S2.9. Power of the model selection method to retrieve *fullMIG* and *snMIG* scenarios under the four different ABC methods. The identification of the correct scenario is shown in bold. The Pritchard's rejection method was analysed with the 100 *test datasets*. The other three methods, *rejection*, *multinomial logistic regression* and *neural network* of the *abc* package, were analysed with a cross-validation process based on 100 permutations and a tolerance of 0.01 (following the recommendations described in the documentation of the package *abc* (Csilléry et al. 2012)).

	fullMIG				***snMIG***			
	Pritchard rej.	**Rejection**	**Mnlogistic**	**Neuralnet**	**Pritchard rej.**	**Rejection**	**Mnlogistic**	**Neuralnet**
fullMIG	**0.670**	**0.662**	**0.768**	**0.734**	0.330	0.338	0.232	0.266
snMIG	0.180	0.340	0.252	0.280	**0.820**	**0.660**	**0.748**	**0.720**

Table S10. Distribution of distance between true and estimated parameter values. The table shows boxplots with the distance between the true and estimated (median, mean and mode) parameter value. This was done with the 100 *test datasets* (*fullMIG*) and for all the studied parameters (T_{START}, T_{OOA}, T_{NEO}, N_{ANC}, N_{ANCNEO}, G_{PALEO}, G_{NEO}, M_{PALEO}, M_{NEOEUR}, M_{NEONA}, M_{NEOPER}, M_{nsSG}, M_{snSG}, K_{PALEO}, K_{NEOEUR}, K_{NEONA}, K_{NEOPER}, I_{rate}). As a reference, the dashed line indicates that the simulated (true) and estimated values are equal (error = 0).

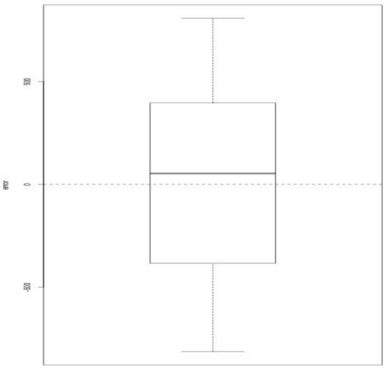
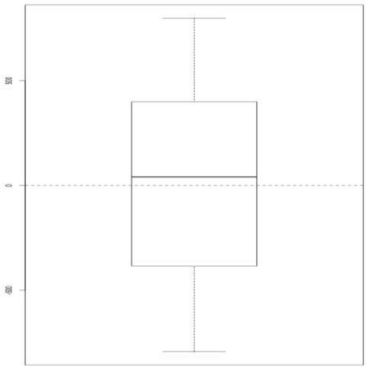
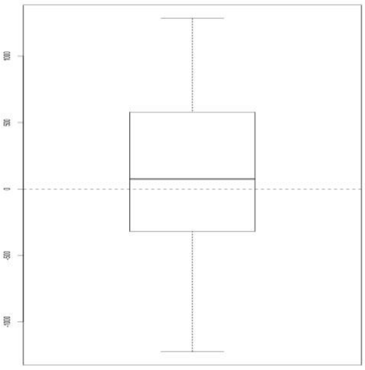
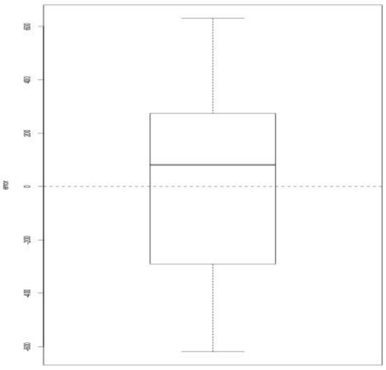
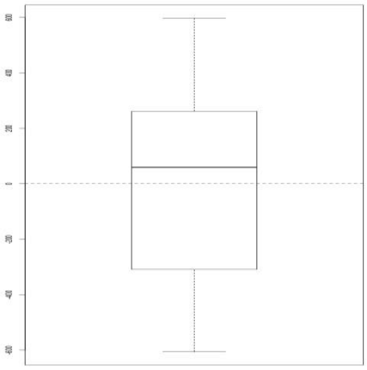
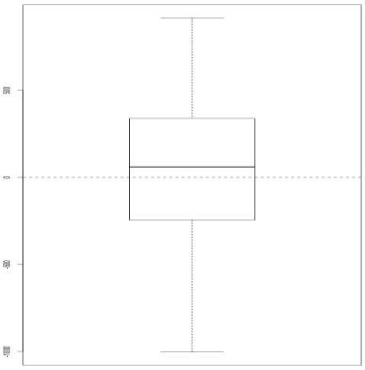
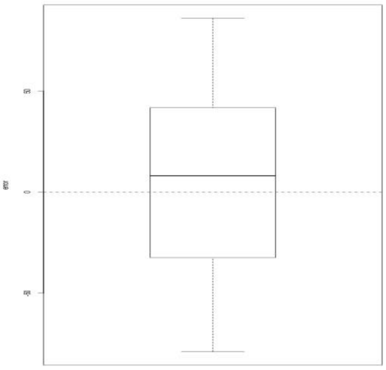
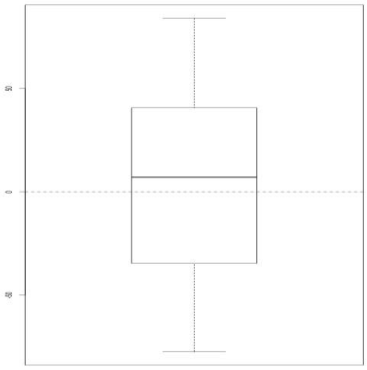
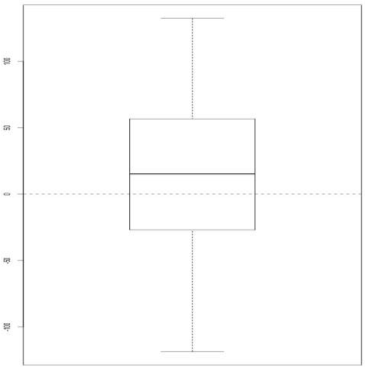
Parameter	**Error Estimation**		
	Median	**Mean**	**Mode**
Initial Expansion of modern humans (T_{START})			
Out of Africa (T_{OOA})			
Initial Neolithic Expansion (T_{NEO})			

Table S10 cont.

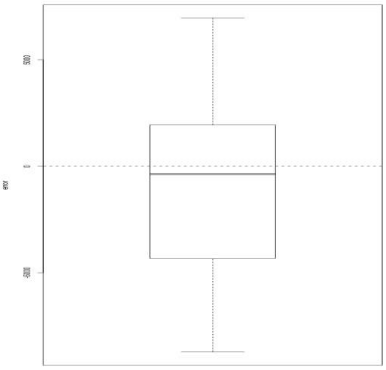
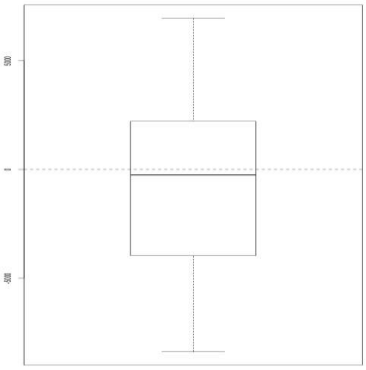
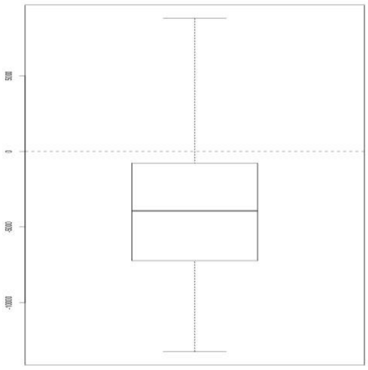
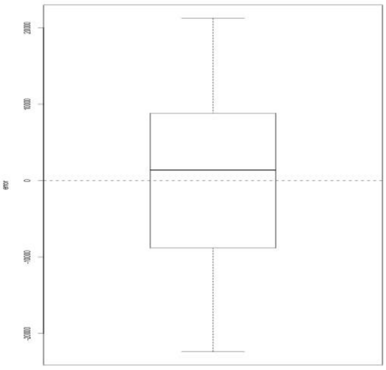
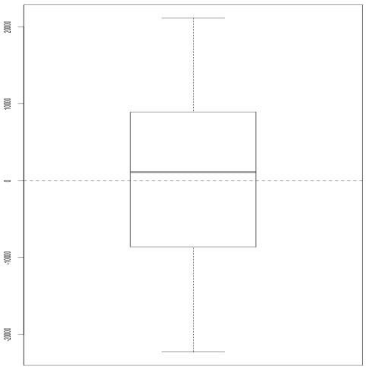
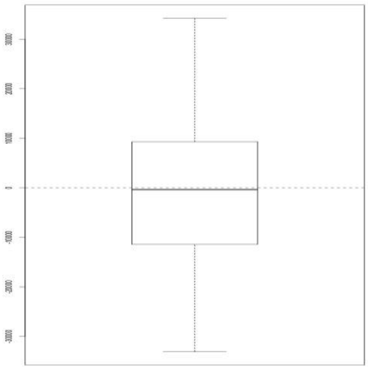
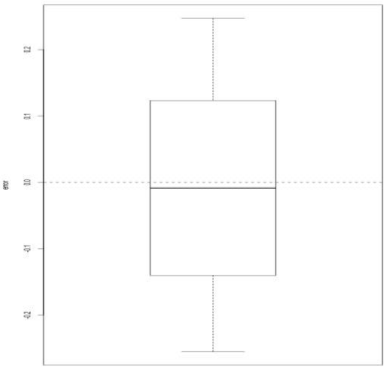
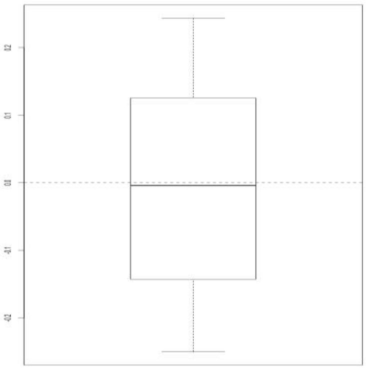
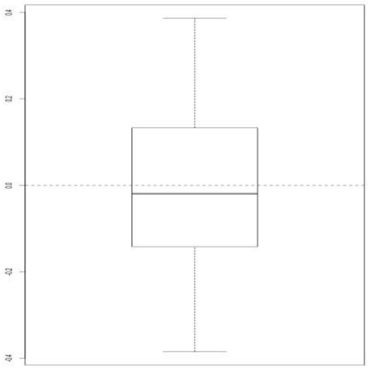
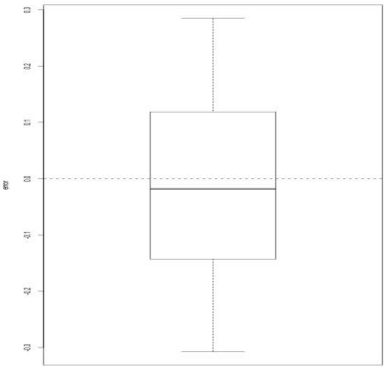
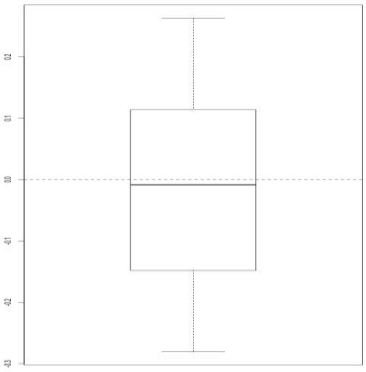
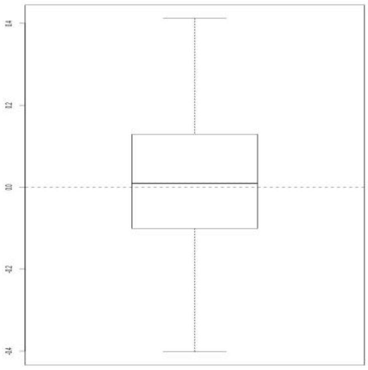
Parameter	Error Estimation		
	Median	Mean	Mode
Ancestral size Palaeolithic (N_{ANC})			
Ancestral size Neolithic (N_{ANCNEO})			
Population growth rate Palaeolithic (G_{PALEO})			
Population growth rate Neolithic (G_{NEO})			

Table S10 cont.

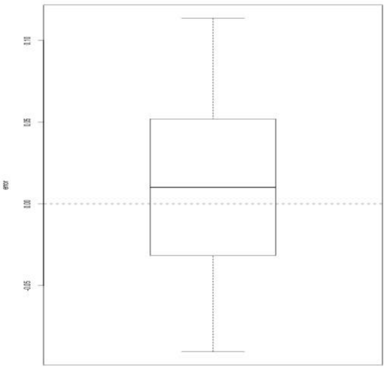
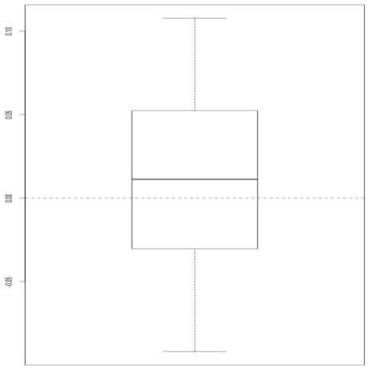
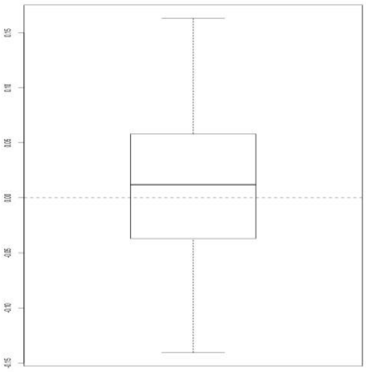
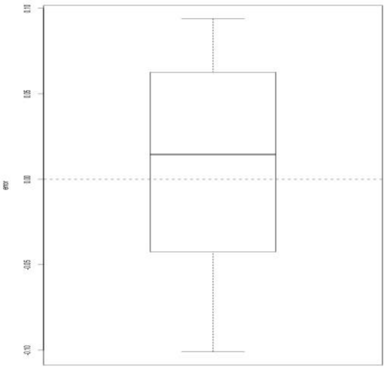
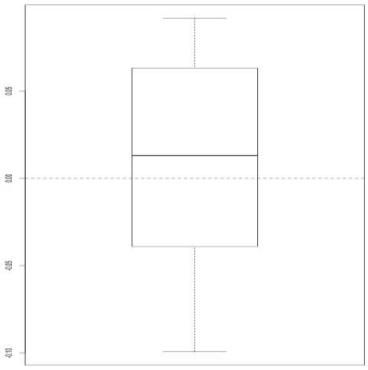
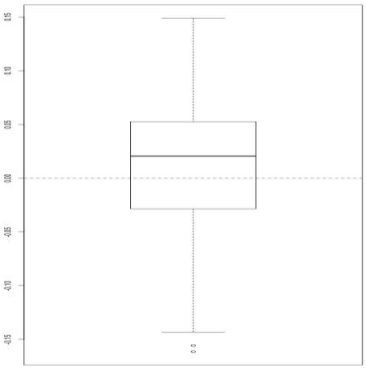
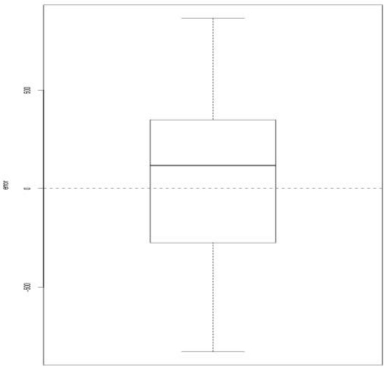
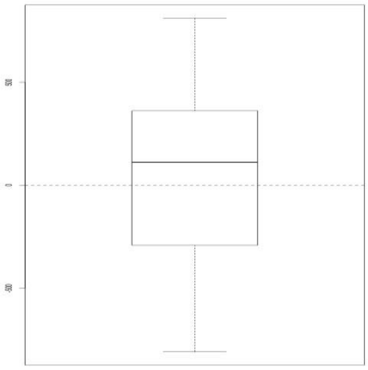
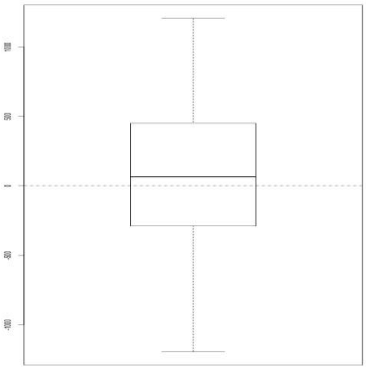
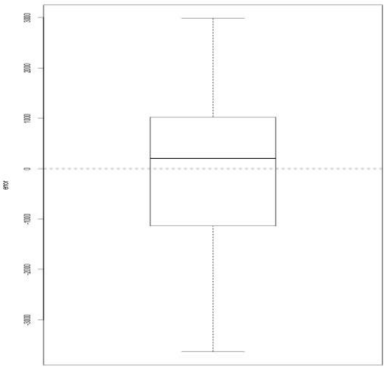
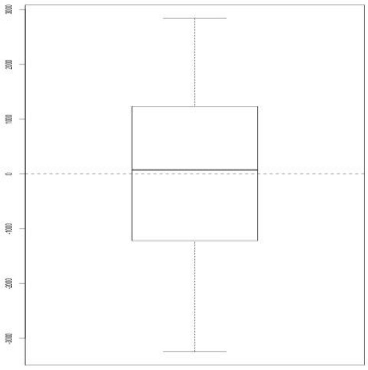
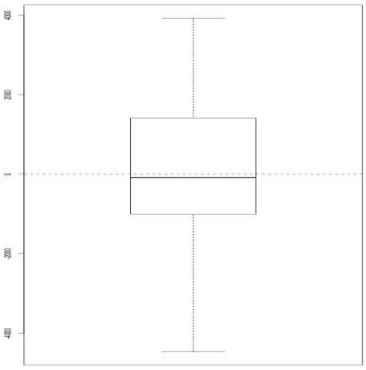
Parameter	Error Estimation		
	Median	Mean	Mode
Migration rate at Strait of Gibraltar - North to South route (M_{nsSG})			
Migration rate at Strait of Gibraltar - South to North route (M_{snSG})			
Carrying capacity Palaeolithic (K_{PALEO})			
Carrying capacity Neolithic in Europe (K_{NEOEUR})			

Table S10 cont.

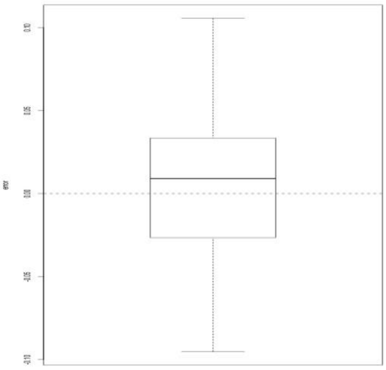
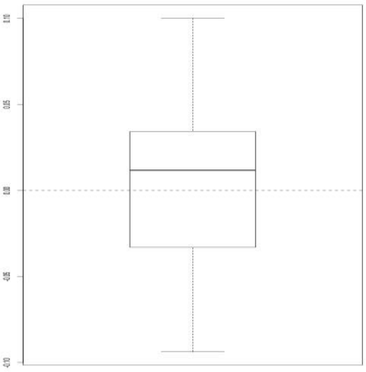
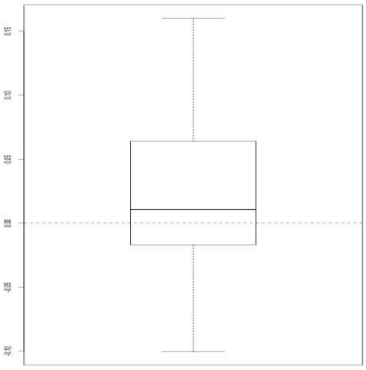
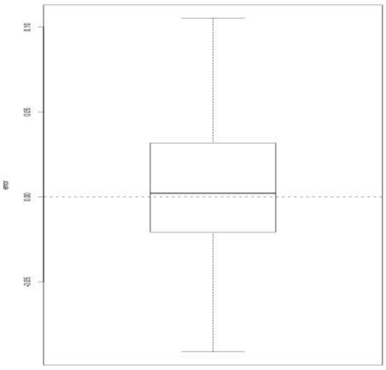
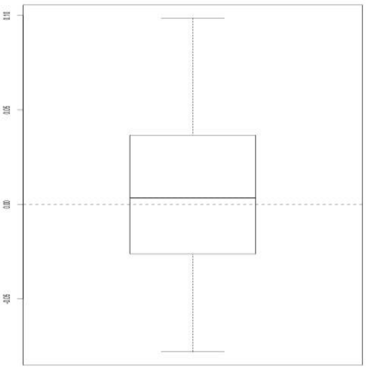
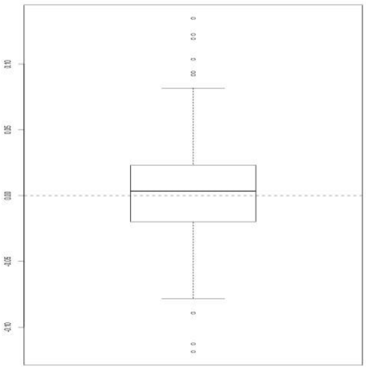
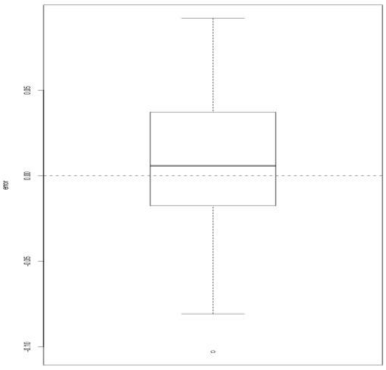
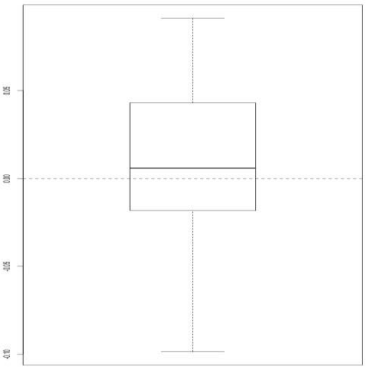
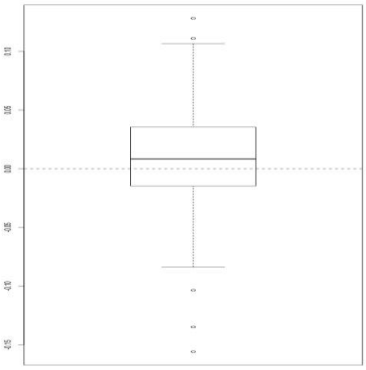
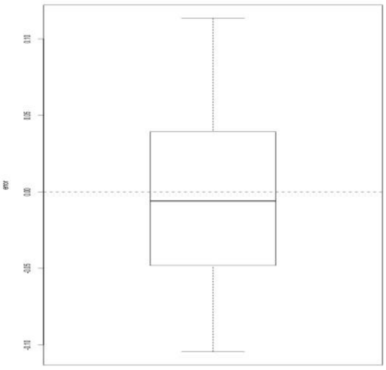
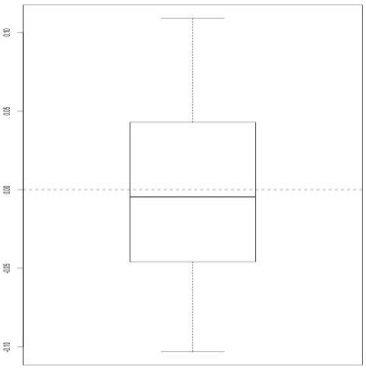
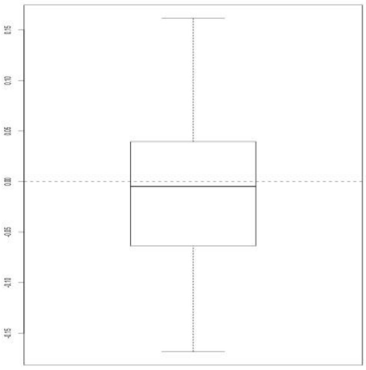
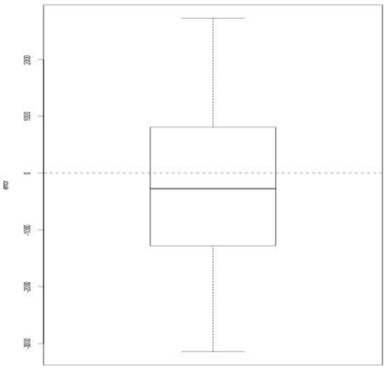
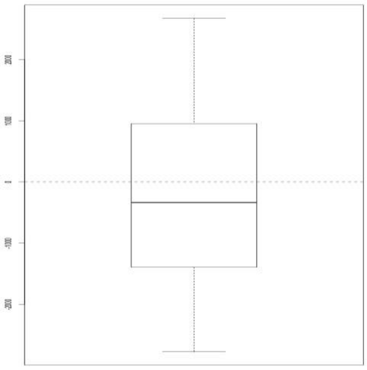
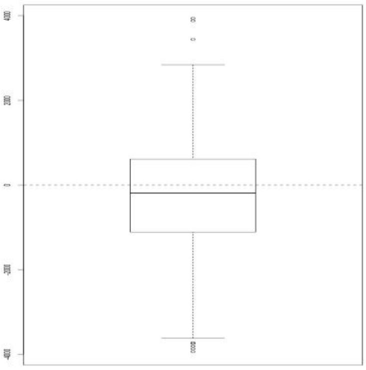
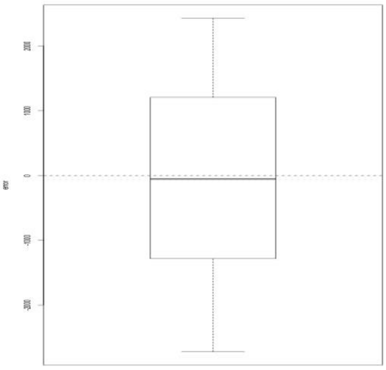
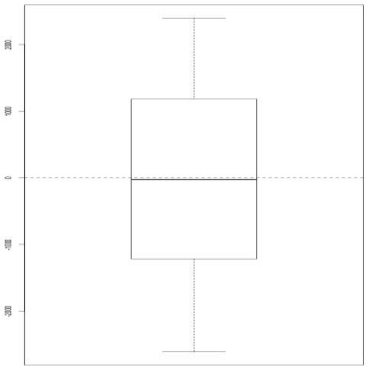
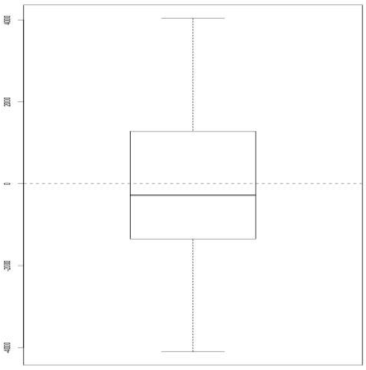
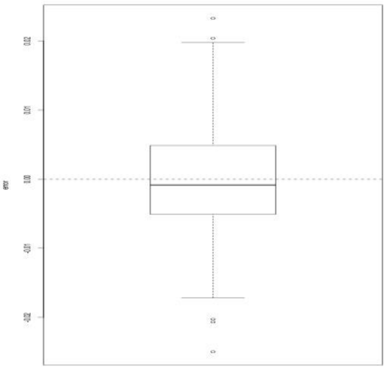
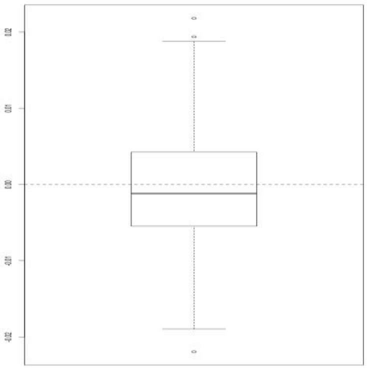
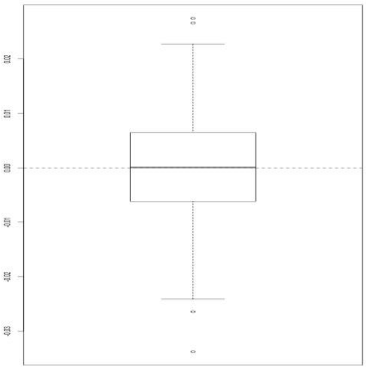
Parameter	Error Estimation		
	Median	**Mean**	**Mode**
Migration rate Palaeolithic (M_{PALEO})			
Migration rate Neolithic in Europe (M_{NEOEUR})			
Migration rate Neolithic in North Africa (M_{NEONA})			
Migration rate Neolithic in Rest of the World (M_{NEOPER})			

Table S10 cont.

Parameter	**Error Estimation**		
	Median	**Mean**	**Mode**
Carrying capacity Neolithic in North Africa (K_{NEONA})			
Carrying capacity Neolithic in Rest of the World (K_{NEOPER})			
Interbreeding rate (I_{rate})			

2.8. References cited in the Supplementary Material

Alves I, Arenas M, Currat M, Hanulova AS, Sousa VC, Ray N, Excoffier L. 2016. Long-distance dispersal shaped patterns of human genetic diversity in Eurasia. Mol. Biol. Evol. 33:946–958.

Arauna LR, Mendoza-Revilla J, Mas-Sandoval A, Izaabel H, Bekada A, Benhamamouch S, Fadhlaoui-Zid K, Zalloua P, Hellenthal G, Comas D, et al. 2017. Recent historical migrations have shaped the gene pool of Arabs and Berbers in North Africa. Mol. Biol. Evol. 34:318–329.

Auton A, McVean G. 2007. Recombination rate estimation in the presence of hotspots. Genome Res. 17:1219–1227.

Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. Genetics 162:2025–2035.

Botigué LR, Henn BM, Gravel S, Maples BK, Gignoux CR, Corona E, Atzmon G, Burns E, Ostrer H, Flores C, 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.

Csilléry K, François O, Blum MGB. 2012. Abc: An R package for approximate Bayesian computation (ABC). Methods Ecol. Evol. 3:475–479.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10:564–567.

Henn BM, Botigué LR, Gravel S, Wang W, Brisbin A, Byrnes JK, Fadhlaoui-Zid K, Zalloua PA, Moreno-Estrada A, Bertranpetit J, et al. 2012. Genomic ancestry of North Africans supports back-to-Africa migrations. PLoS Genet. 8:e1002397.

Lopes AM, Aston KI, Thompson E, Carvalho F, Gonçalves J, Huang N, Matthiesen R, Noordam MJ, Quintela I, Ramu A, et al. 2013. Human Spermatogenic Failure Purges Deleterious Mutation Load from the Autosomes and Both Sex Chromosomes, including the Gene DMRT1. PLoS Genet. 9.

McVean G a T, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. 2004. The fine-scale structure of recombination rate variation in the human genome. Science 304:581–584.

The 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1 , 092 human genomes. Nature 491:56–65.

Tillmar AO, Mostad P, Egeland T, Lindblom B, Holmlund G, Montelius K. 2008. Analysis of linkage and linkage disequilibrium for eight X-STR markers. Forensic Sci. Int. Genet. 3:37–41.

Chapter 3.

Spatially Explicit Analysis Reveals Complex Human Genetic Gradients in the Iberian Peninsula

João Pimenta, Alexandra M. Lopes, Angel Carracedo, Miguel Arenas, António Amorim, David Comas

3.1. Introduction

The advent of high-throughput genotyping technologies led to a revolution in our understanding of human history and population structure (The 1000 Genomes Project Consortium 2012; Fu et al. 2013). Results on genome-wide autosomal studies showed that among European populations, genetic differentiation was, in general, highly correlated with geography (Lao et al. 2008; Novembre et al. 2008). However, genetic differentiation has also been associated with political-cultural boundaries and geographic barriers that limit the gene flow between populations and promote genetic structuration (Barbujani and Sokal 1990; Adams et al. 2008; Lao et al. 2008).

In recent years, the development of innovative methodologies that integrate geographic information with genetic data has given rise to robust analyses of the variation of the population structure across a landscape (Arenas et al. 2013; Hanks and Hooten 2013; Duforet-Frebourg and Blum 2014; Bradburd et al. 2016; Petkova et al. 2016). Essentially, these methods allow to study patterns of spatially explicit genetic differentiation by modulating genetic distances between samples or populations as a function of the geographic distance (Bradburd et al. 2016; Petkova et al. 2016). Previous studies showed that, at a global scale, several major geographic features often lead to strong genetic differentiation (Peter et al. unpublished data), while at a regional/local scale, besides geographic characteristics, cultural barriers often additionally play a role modulating the genetic structure of populations (Messina et al. 2016; Uren et al. 2016; Jeong et al. 2017; Peter et al. unpublished data).

The Iberian Peninsula nowadays comprises two major countries, Portugal and Spain, and it is separated from the rest of Europe by a range of mountains (the Pyrenees) and from Africa by a small stretch of water (the Strait of Gibraltar). The history of the Iberian Peninsula is characterized by multiple migrations and settlements of different population groups. The region was a refuge for populations fleeing from glaciers advance during the Last Glacial Maximum (LGM) (Torroni et al. 2001; Achilli et al. 2004) and was probably a reservoir for the repopulation of Europe after the end of the LGM (Pereira et al. 2005). Only in the last two millennia, Iberia was occupied by Phoenicians, Greeks, Romans, German tribes and, more recently, most of the Peninsula was under Islamic administration, which lasted from the beginning of the 8th till the end of the 15th century when Catholic kingdoms concluded the control of the territory in what has been called the *Reconquista* (Carr 2000; Prag and Quinn

2011). Modern Iberian populations reflect this complex admixture of cultures and genetic backgrounds (Adams et al. 2008; Botigue et al. 2013), being one of the regions with the highest genetic diversity in Europe (Wang et al. 2012). On the other hand, results based on autosomal markers showed a general homogeneity of Iberian populations (Gayán et al. 2010) despite some local differentiation observed for mitochondrial DNA and Y-Chromosome (Brion et al. 2004; López-Oarra et al. 2009; Alvarez et al. 2010; Pardiñas et al. 2012).

Here we extended those works investigating the genetic structure, genetic gradients and migratory dynamics of humans in the Iberian Peninsula at a fine-scale level. We applied spatially explicit models to analyse a comprehensive genome-wide SNP dataset of 1,204 individuals belonging to 26 different populations of the area.

3.2. Materials and Methods

3.2.1. Data and Genotyping

We compiled genome-wide SNP data of individuals belonging to 26 populations from publicly available data to generate two datasets: A global dataset, to provide an exploratory analysis on patterns of genetic variation and ancestral components in the Iberian Peninsula at the continental level, that included all the compiled populations (Table 3.1): 17 Iberian populations (2 Portuguese populations retrieved from Lopes et al. (2013) and 15 Spanish populations retrieved from Botigue et al. (2013), Henn et al. (2012) and Fernandez-Rozadilla et al. (2013)), 4 European populations (French Basques from the Human Genome Diversity Panel (Li et al. 2008), northern Europeans (CEU), Tuscans (TSI) and Finns (FIN) obtained from the 1000 Genomes Project (The 1000 Genomes Project Consortium 2012)). Out of Europe it included 4 north African populations (Algeria, Morocco North, Morocco South and Tunisia) retrieved from Henn et al. (2012) and 1 Sub-Saharan (Yoruba) from the 1000 Genomes Project (The 1000 Genomes Project Consortium 2012). The second dataset (hereafter, Iberian dataset), to provide a fine-scale analysis of the genetic structure and heterogeneity in the Iberian Peninsula, was composed of the 17 Iberian populations previously indicated (Figure 3.1).

Next, a quality control filter was applied using PLINK 1.9 (Purcell et al. 2007). We excluded SNPs with missing genotype rate >10%, that failed Hardy-Weinberg equilibrium under a threshold of 0.05 and SNPs with a minor allele frequency lower than 0.05. In addition, we excluded individuals with a missing rate >10% and that shared an identity-by-state >85%. After the quality control filter, the global dataset and Iberian dataset presented a total of 1,204 and 746 individuals, respectively (Table 3.1).

Table 3.1. Location, sample size and reference of the dataset analysed in the present study.

Population	Sample size	Reference
Iberia Peninsula		
Basque Country, Spain	57	Henn et al. 2012; Fernandez-Rozadilla et al. 2013
La Rioja	10	Fernandez-Rozadilla et al. 2013
Navarra	12	Fernandez-Rozadilla et al. 2013
Catalonia	93	Fernandez-Rozadilla et al. 2013
Aragon	24	Fernandez-Rozadilla et al. 2013
Valencian Community	16	Fernandez-Rozadilla et al. 2013
Murcia	13	Fernandez-Rozadilla et al. 2013
Andalusia	68	Botigue et al. 2013; Fernandez-Rozadilla et al. 2013
Extremadura	14	Fernandez-Rozadilla et al. 2013
Castilla - La Mancha	34	Fernandez-Rozadilla et al. 2013
Madrid	2	Fernandez-Rozadilla et al. 2013
Castilla y Leon	36	Fernandez-Rozadilla et al. 2013
Cantabria	10	Fernandez-Rozadilla et al. 2013
Asturias	11	Fernandez-Rozadilla et al. 2013
Galicia	177	Botigue et al. 2013; Fernandez-Rozadilla et al. 2013
Porto	124	Lopes et al. 2013
Lisbon	45	Lopes et al. 2013
Europe		
Finland	79	1000 Genomes Project Consortium et al. 2012
Northern and Western Europeans from Utah	87	1000 Genomes Project Consortium et al. 2012
Tuscany, Italy	96	1000 Genomes Project Consortium et al. 2012
Basque Country, France	24	Li et al. 2008
North Africa		
Morocco South	16	Henn et al. 2012
Morocco North	18	Henn et al. 2012
Algeria	19	Henn et al. 2012
Tunisia	18	Henn et al. 2012
Sub-Saharan Africa		
Yoruba, Nigeria	101	1000 Genomes Project Consortium et al. 2012

3.2.2. Analysis of population structure of the Iberian Peninsula

In order to explore the presence of population stratification in the global and in the Iberian datasets, we performed a Principal Component Analysis (PCA) with the *smartPCA* algorithm implemented in the EIGENSOFT 5.0.1 package (Patterson et al. 2006). Additionally, we applied ADMIXTURE v1.3.0 (Alexander et al. 2009) under unsupervised mode. Results were presented in plots using Distruct1.1 (Rosenberg 2004). In order to mitigate the effect of SNPs in high linkage disequilibrium (LD), we applied an additional filter to remove SNPs with excess of LD, using PLINK (Purcell et al. 2007). We utilized a sliding window of 50 SNPs, a shift step of 5 and a LD threshold of 0.1, hence keeping only the most informative SNPs. Finally, a total of 34,748 and 50,103 SNPs for the global and the Iberian datasets, respectively, were considered.

3.2.3. Analysis of spatial structure in the Iberian Peninsula

For the spatially explicit investigation, we applied the same pruning step (equal window size and shift step) to the Iberian dataset, but relaxing the LD threshold to 0.5 to reduce the amount of redundant data while maximizing the number of SNPs available for the analysis. We kept a total of 174,001 and 4,792 SNPs for autosomes and X chromosome, respectively, for posterior analysis.

In order to investigate patterns of isolation by distance in the Iberian Peninsula we applied the Bayesian framework SpaceMix (Bradburd et al. 2016) to the Iberian dataset for the autosomes. This analysis provides genetic relationships between populations as a “geogenetic map” in which geographical distances between populations are proportional to genetic distances. Following the software documentation, we ran 10 independent short chains of 10^6 iterations each, followed by a long chain of 10^7 iterations based on the estimates of the last iteration of the short chains with the best posterior probability. A sample was taken every 10^4 iterations, keeping a total of 1,000 draws to estimate the posterior distribution of each parameter. Initial population locations were randomly taken from a uniform distribution of -180 to 180 and -90 to 90 for longitude and latitude, respectively.

In order to further identify patterns of spatial structure and genetic heterogeneity within the Iberian Peninsula, we applied the software EEMS (Petkova et al. 2016) to the Iberian dataset, for both autosomes and X chromosome. We estimated migration rate surfaces to visualize corridors and barriers to gene flow. Basically, EEMS considers the stepping-stone migration model to infer migration rates through a

Bayesian inference (Petkova et al. 2016). The method applies a dense triangular grid that fills the entire landscape and assigns each individual to the geographic neighbour deme of each population to finally provide a map quantifying genetic dissimilarities (migration rate). We also estimated a matrix of genetic dissimilarities between all 746 individuals with the *bed2cliffs* method implemented in the EEMS package. For all the analyses we specified a total of 1,000 demes and we performed 5 independent runs with $1.1\text{x}10^7$ iterations, a thinning interval of 1,000 iterations and 10^7 iterations for burn-in (ensuring converged iterations among different runs). Indeed, also following the software documentation, we adjusted migration and diversity parameters to specify acceptance rates around 20-30%.

3.3. Results

We first present the results for the population stratification and global ancestry inference for the global dataset and the Iberian dataset. Then, we present the patterns of genetic differentiation and spatial genetic gradients in the Iberian Peninsula along with the fine-scale spatially explicit analysis.

3.3.1. High genetic similarity and mainly European ancestry in the Iberian Peninsula

PCA results derived from the global dataset showed a clear genetic separation between Europeans, North Africans and Sub-Saharan African populations for the first PC (PC1), while PC2 distinguished between northern, central and southern Europeans (Figure S3.1A). In both PCs, Iberian populations cluster together. Interestingly, PC3 shows a cluster composed mainly by Iberian Basques and French Basques, both disconnected from the rest of the Iberian Peninsula (Figure S3.1A). On the other hand, PCA results derived from the Iberian dataset presented some genetic differentiation within the Iberian Peninsula (Figure 3.1). PC1 separated Portugal and Galicia samples from other Iberians, whereas PC2 separated most Spanish samples from Portuguese and some Basque samples (Figure 3.1). PC3, PC4, and PC5 showed the inner diversity of Iberian populations while highlighting the global homogeneity (Fig S3.1B).

Unsupervised clustering analysis on the global dataset showed that most ancestry in Iberians was caused by European contributions, followed by contributions from Africa (Figure S3.2A). When considering K=4, (the model with the lowest cross-validation error) (Figure S3.2B), most Iberian individuals present three main ancestral components. Two components could be associated with European ancestry, while the third one correlated with North African ancestry. It is noteworthy that for Basque individuals the North African ancestral component present in the rest of Iberian samples is shown at a vestigial level (Figure S3.2A).

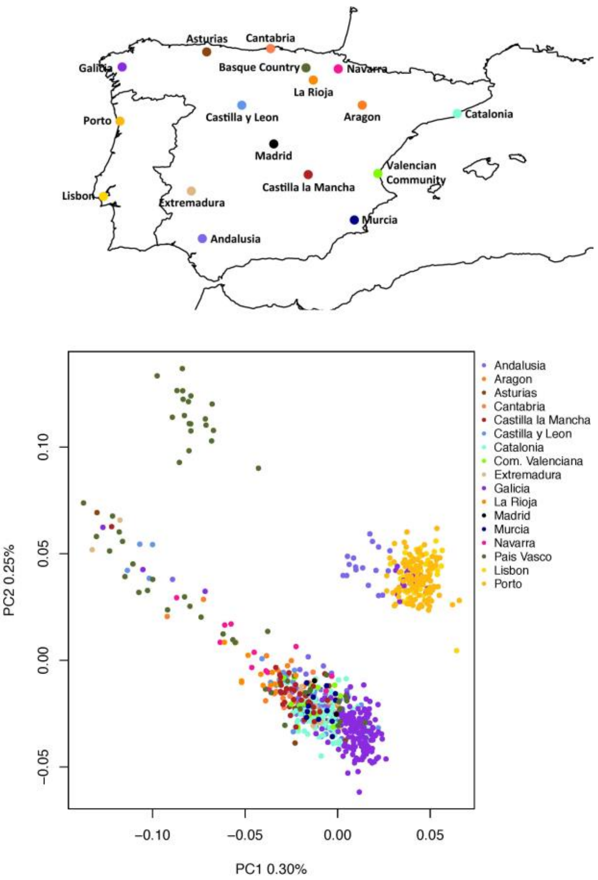


Figure 3.1. Geographic location of the 17 Iberian populations of the studied dataset and population structure in the Iberian Peninsula. First two principal components of the Iberian dataset. Iberian samples and French Basques are colour-coded whereas the rest of samples are shown in grey.

3.3.1. Subtle genetic structure in the Iberian Peninsula based on the spatially explicit analysis

We modulated the geographic structure of the Iberian Peninsula with the Bayesian framework included in SpaceMix (see Methods). The results presented a geogenetic map with 5 population groups based on the 95% confidence surfaces (Figure 3.2). The largest genetic divergence observed was between Portuguese and Basque Country populations. Results highlighted some genetic isolation of Galician populations with respect to all the other populations of the Iberian Peninsula, as well as a genetic differentiation between populations of the Basque region (Basque Country, La Rioja and Navarra). The rest of Spanish regions (Northern, Central and Mediterranean populations) presented genetically close proximity, suggesting high genetic similarity. The stability of the posterior probability trace and the correlation between inferred and observed genetic covariance indicated that the applied model could accurately describe the observed data (Figure S3.3ABC).

Concerning patterns of genetic differentiation within the Iberian Peninsula (obtained with EEMS), results for autosomes presented a dual structure in the Iberian Peninsula, with several areas of high genetic differentiation splitting the northern regions and a much more homogeneous southern Iberia (Figure 3.3). Galicia

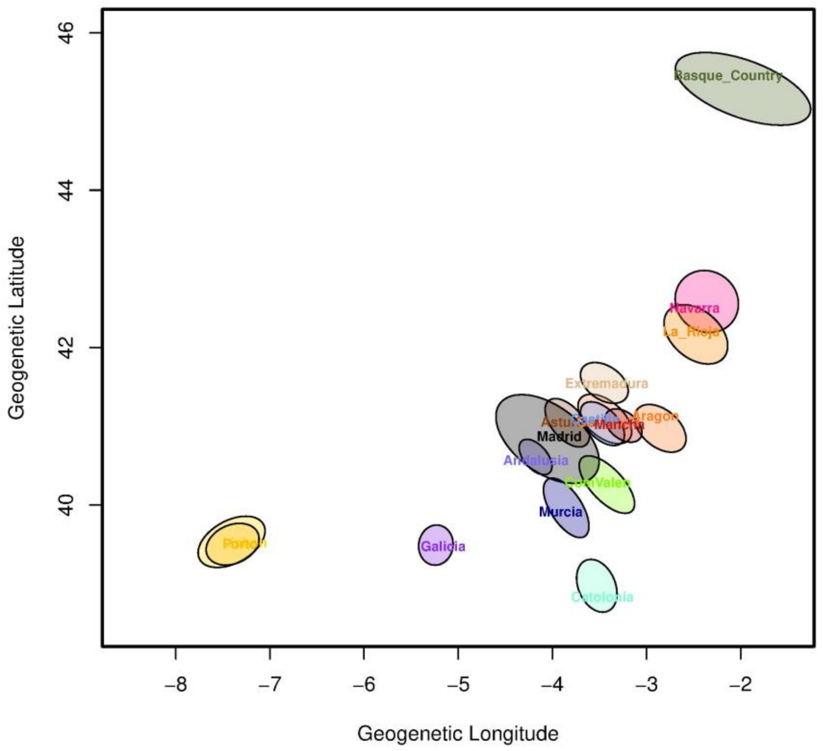


Figure 3.2. Inferred geogenetic locations of the Iberian populations. The figure shows the 95% confidence surfaces. Some names are abbreviated as Castilla (Castilla y Leon), Mancha (Castilla la Mancha) and ComValen (Valencian Community). Inferences performed with the framework SpaceMix**.**

(northwesternmost region of Iberia) presented some isolation from the rest of the Peninsula through barriers with Asturias (northeast) and, in a lower extent, with the north of Portugal (south) (Figure S3.4). Indeed, populations from the Basque region appeared almost genetically isolated from the rest of the Peninsula. Another detected strong barrier is placed between north and northeast regions (between Aragon and Catalonia) (Figure S3.4). Concerning the opposite pattern (high genetic similarity), we detected two regions with a high effective migration rate. One of them is a corridor along the entire Mediterranean coast, from the northeast to the south of the Iberian Peninsula. The other region presenting high effective migration involves the central and north coast of Portugal. Finally, another corridor enclosed several Spanish populations that connected northern regions (Asturias and Cantabria) with central regions of the Iberian Peninsula (Castilla y Leon, Madrid, Castilla la Mancha) (Figure 3.3). The correlation between estimated and observed genetic dissimilarities between and within demes (R^2 coefficients of 0.80 and 0.95, respectively) indicated the ability of the EEMS model to describe the observed data (Figure S3.5AB). Indeed, the correlation between geographic distance and genetic distance revealed that a model of isolation by distance cannot explain the population structure found in the Iberian Peninsula (Figure S3.5C).

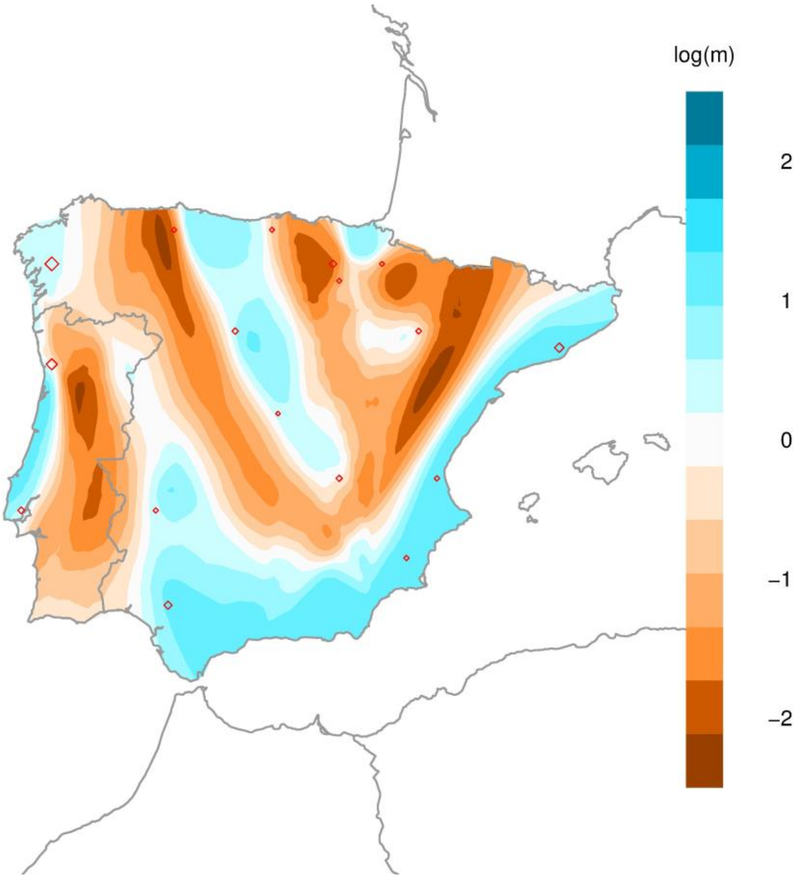


Figure 3.3. Effective Migration Map estimated for autosomes for the Iberian dataset. Plots were estimated with EEMS under a $\log_{10}$ scale and after mean centering. Blue regions indicate areas with effective migration rate higher than average, while brown colours indicate a lower effective migration between demes.

In order to detect putative sex biased population structure within the Iberian Peninsula, we compared EEMS results for autosomes with EEMS results from the X chromosome. EEMS results for the X chromosome revealed some different features concerning the genetic structure in comparison with autosomes (Figure S3.6). Northern Iberia populations showed a stronger genetic structure (lower migration rates) than central and southern Spanish populations (higher migration rates). In agreement with the data found for autosomes, the area with the highest genetic differentiation was the Basque region (Figure S3.7). Basque Country and neighbouring populations were surrounded by a strong barrier to gene flow, while a large region with high effective migration rate cluster populations from Eastern and Central regions of Iberia (Figure S3.6). Another region with high effective migration rate was identified connecting the southern regions Andalusia and Extremadura (Figure S3.6). As for autosomes, Portuguese populations are clustered together and are isolated from Spanish populations (Figure S3.6). Diagnostic plots for model fitting (Figure S3.8AB) showed that EEMS results present an excellent fitting with the data for within demes dissimilarities (R^2 = 1.00) but weak when considering between demes dissimilarities (R^2 = 0.373). Similarly to the autosomes, a significant deviation from an isolation by distance model (R^2 coefficient of 0.01) was found when the observed pairwise genetic distances between populations are related with geographic distances (Figure S3.8C).

3.4. Discussion

The presence of a genetic structure in the Iberian Peninsula has been described based on Y chromosome and mtDNA at a regional level (Adams et al. 2008; Alvarez et al. 2010; Pardiñas et al. 2012; Hernández et al. 2017). Here we extended those studies by considering a more comprehensive dataset based on genome-wide genetic information, thus analysing both autosomes and, for the first time, X chromosome. We found that the characterization of the genetic landscape of the Iberian Peninsula, using spatially explicit approaches, showed subtle structure features. Indeed, areas of strong genetic differentiation are observed in northern regions of Iberia (Figures 3.3 and S3.4), while corridors of genetic similarity mainly appear along a latitudinal orientation (Figure 3.3). This north-south genetic correlation between populations has been associated with the *Reconquista*, the long period, between 711 CE and 1492 CE, in which the Catholic kingdoms took entire control of the Peninsula after the Muslim invasion (Bycroft et al. unpublished data). In line with our results for autosomes and for the X chromosome (Figures 3.3 and S3.6), a fine-scale analysis in Spanish samples showed patterns of strong population structure in northern Iberia, which was associated to cultural and linguistic differences between the different Catholic kingdoms existing in the critical historical period of the *Reconquista*, in opposition with its southern counterpart that was much more homogeneous (Bycroft et al. unpublished data). However, this gene flow associated with the *Reconquista* cannot entirely explain the structure that here we found for the southern Iberia, due to the lack of agreement between the regions conquered by the different Catholic kingdoms (Baldinger 1963) and the barriers and corridors to gene flow found in our spatial explicit analysis. We suggest that population movements posterior to the *Reconquista* across several Iberian regions might have played a major role modulating current genetic structure, with a special flow following the Mediterranean coast. In addition, we suggest sex-biased migrations because of the absence of genetic differentiation between the regions of Catalonia and Aragon and the lack of structure in Central and Southern regions of Spain when separately analysing the X chromosome.

Populations from the Basque region showed a genetic distinctness from the other Iberian populations for both autosomes (Figures 3.1, 3.2 and S3.4) and X chromosome (Figure S3.6). This genetic differentiation could be caused by cultural aspects since Basques are characterized by their unique non-Indo-European language and limited gene flow from external Iberian populations such as north Africans, as shown in our analyses and also previously reported (Martínez-Cruz et al. 2012). Additionally, the genetic similarity found between Spanish and French Basques using both PCA (Figure S3.1A) and ancestry profiles (Figure S3.2A) could be explained by the sharing of cultural traditions between these regions.

The northwest region of the Peninsula (Galicia) presented a genetic differentiation higher than the average of other Spanish populations (Figures 3.2 and 3.3). As for Basque populations, cultural and linguistic differences could be accounted for this important genetic divergence (Mackenzie 2000). Moreover, a study on marital behaviour showed a high proportion of inbreeding that could lead to genetic differentiation (Varela et al. 2003). Interestingly, Portugal and Galicia are considered to largely share their ancestral history (Bycroft et al. unpublished data) and we also found results in this direction. The geogenetic map (Figure 3.2) demonstrates that Galicia is the closest region to Portugal, despite the larger geogenetic distance estimated between Galicia and Portugal respect to the distance between Galicia and central regions of Spain. Additionally, the estimated Effective Migration Surface for autosomes (Figure 3.3) suggests a small coastal corridor of gene flow connecting Galicia to northern regions of Portugal. These results can be attributed to the long historical relationship between Galicia and northern regions of Portugal. Indeed, before the Islamic invasion in the 8th century, both regions belonged to the Roman province of Gallaecia and later on to the kingdom of the Suebi (405 CE and 585 CE), before the annexation by the Visigoths (Barreiro Fernández et al. 1984; Mattoso 1994). Portugal became politically independent only in 1143 and expanded south very fast (Portuguese *Reconquista* ended by 1249). The establishment of a political border led to some cultural divergence but still important relationships were kept because of their geographic proximity, similar language and sociological factors. A recent study, also based on genome data but applying other approaches, showed remarkable genetic similarities between both regions (Bycroft et al. unpublished data), in agreement with our findings.

In conclusion, we found that the genetic variation thorough the Iberian Peninsula is complex and differs when comparing the North, with remarkable spatial genetic dissimilarity, and the south, with a more homogeneous genetic similarity. We suggest that geography is not the main contributor shaping the genetic landscape of the Iberian Peninsula. Instead, the observed major genetic dissimilarities seem mainly based on political and cultural barriers that influenced the relationships between populations and migratory patterns.

3.5. Acknowledgments

IPATIMUP integrates the i3S Research Unit, which is partially supported by FCT in the framework of the project “Institute for Research and Innovation in Health Sciences” (POCI-01-0145-FEDER-007274). JP and AML are funded by the Portuguese Government through the FCT fellowship SFRH/BD/97200/2013 and research contract IF/01262/2014, respectively. MA was supported by the “Ramón y Cajal” grant RYC-2015-18241 from the Spanish Government. DC was supported by the Spanish grant CGL2016-75389-P (MINECO/FEDER, UE), and “Unidad María de Maeztu” funded by the MINECO (MDM-2014-0370).

3.6. References

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, 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, Bosch E, Balaresque PL, Ballereau SJ, Lee AC, Arroyo E, López-Parra AM, Aler M, Grifo MSG, Brion M, 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.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–1664.

Alvarez L, Santos C, Ramos A, Pratdesaba R, Francalacci P, Aluja MP. 2010. Mitochondrial DNA patterns in the Iberian Northern plateau: Population dynamics and substructure of the Zamora province. Am. J. Phys. Anthropol. 142:531–539.

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.

Baldinger K. 1963. La formación de los dominios lingüísticos en la península ibérica. Gredos

Barbujani G, Sokal RR. 1990. Zones of sharp genetic change in Europe are also linguistic boundaries. Proc. Natl. Acad. Sci. U. S. A. 87:1816–1819.

Barreiro Fernández XR, Diaz-Fierros F, Fabra Barreiro G. 1984. Los Gallegos. Istmo

Botigue LR, Henn BM, Gravel S, Maples BK, Gignoux CR, Corona E, Atzmon G, Burns E, Ostrer H, Flores C, et al. 2013. Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. Proc. Natl. Acad. Sci. 110:11791–11796.

Bradburd GS, Ralph PL, Coop GM. 2016. A Spatial Framework for Understanding Population Structure and Admixture. PLoS Genet. 12:e1005703.

Brion M, Quintans B, Zarrabeitia M, Gonzalez-Neira A, Salas A, Lareu V, Tyler-Smith C, Carracedo A. 2004. Micro-geographical differentiation in Northern Iberia revealed by Y-chromosomal DNA analysis. Gene 329:17–25.

Bycroft C, Fernández-Rozadilla C, Ruiz-Ponte C, Quintela-García I, Carracedo Á, Donnelly P, Myers S. unpublished data. https://www.biorxiv.org/content/early/2018/03/12/250191. last accessed June 5, 2018.

Carr R. 2000. Spain: a history. Oxford University Press

Duforet-Frebourg N, Blum MGB. 2014. Nonstationary patterns of isolation-by-distance: Inferring measures of local genetic differentiation with bayesian kriging. Evolution (N. Y). 68:1110–1123.

Fernandez-Rozadilla C, Cazier J-B, Tomlinson IP, Carvajal-Carmona LG, Palles C, Lamas MJ, Baiget M, López-Fernández LA, Brea-Fernández A, Abulí A, et al. 2013. A colorectal cancer genome-wide association study in a Spanish cohort identifies two variants associated with colorectal cancer risk at 1p33 and 8p12. BMC Genomics 14:55.

Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, Gabriel S, Altshuler D, Shendure J, Nickerson DA, et al. 2013. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature 493:216–220.

Gayán J, Galan JJ, González-Pérez A, Sáez M, Martínez-Larrad M, Zabena C, Rivero MC, Salinas A, Ramírez-Lorca R, Morón FJ, et al. 2010. Genetic Structure of the Spanish Population. BMC Genomics 11:326.

Hanks EM, Hooten MB. 2013. Circuit theory and model-based inference for landscape connectivity. J. Am. Stat. Assoc. 108:22–33.

Henn BM, Botigué LR, Gravel S, Wang W, Brisbin A, Byrnes JK, Fadhlaoui-Zid K, Zalloua PA, Moreno-Estrada A, Bertranpetit J, et al. 2012. Genomic ancestry of North Africans supports back-to-Africa migrations. PLoS Genet. 8:e1002397.

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

Jeong C, Peter BM, Basnyat B, Neupane M, Beall CM, Childs G, Craig SR, Novembre J, Di Rienzo A. 2017. A longitudinal cline characterizes the genetic structure of human populations in the Tibetan plateau. PLoS One 12:e0175885.

Lao O, Lu TT, Nothnagel M, Junge O, Freitag-Wolf S, Caliebe A, Balascakova M, Bertranpetit J, Bindoff LA, Comas D, et al. 2008. Correlation between Genetic and Geographic Structure in Europe. Curr. Biol. 18:1241–1248.

Li JZ, Absher DM, Tang H, Southwich AM, Casto A, Ramachandran S, Cann HM, Barsh GS, Feldman MW, Cavalli-Sforza LL, et al. 2008. Worldwide Human Relationships Inferred from Genome-WidePatterns of Variation. Science (80-.). 319:1100–1104.

Lopes AM, Aston KI, Thompson E, Carvalho F, Gonçalves J, Huang N, Matthiesen R, Noordam MJ, Quintela I, Ramu A, et al. 2013. Human Spermatogenic Failure Purges Deleterious Mutation Load from the Autosomes and Both Sex Chromosomes, including the Gene DMRT1. PLoS Genet. 9:e1003349.

López-Oarra AM, Gusmão L, Tavares L, Baeza C, Amorim A, Mesa MS, Prata MJ, Arroyo-Pardo E. 2009. In search of the pre- and post-neolithic genetic substrates in Iberia: Evidence from Y-chromosome in Pyrenean populations. Ann. Hum. Genet. 73:42–53.

Mackenzie D. 2000. Encyclopedia of the Languages of Europe. (Price G, editor.). Blackwell Publishing

Martínez-Cruz B, Harmant C, Platt DE, Haak W, Manry J, Ramos-Luis E, Soria-Hernanz DF, Bauduer F, Salaberria J, Oyharçabal B, et al. 2012. Evidence of pre-roman tribal genetic structure in basques from uniparentally inherited markers. Mol. Biol. Evol. 29:2211–2222.

Mattoso J. 1994. História de Portugal - Antes de Portugal - Vol. I. Editorial.

Messina F, Finocchio A, Akar N, Loutradis A, Michalodimitrakis EI, Brdicka R, Jodice C, Novelletto A. 2016. Spatially explicit models to investigate geographic patterns in the distribution of forensic STRs: Application to the north-eastern mediterranean. PLoS One 11.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, et al. 2008. Genes mirror geography within Europe. Nature 456:98–101.

Pardiñas AF, Roca A, García-Vazquez E, López B. 2012. Assessing the Genetic Influence of Ancient Sociopolitical Structure: Micro-differentiation Patterns in the Population of Asturias (Northern Spain). PLoS One 7:e50206.

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. PLoS Genet. 2:2074–2093.

Pereira L, Richards M, Goios A, Alonso A, Albarrán C, Garcia O, Behar DM, Gölge M, Hatina J, Al-Gazali L, et al. 2005. High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. Genome Res. 15:19–24.

Peter BM, Petkova D, Novembre J. unpublished data. https://www.biorxiv.org/content/early/2018/03/26/233486. last accessed June 5, 2018.

Petkova D, Novembre J, Stephens M. 2016. Visualizing spatial population structure with estimated effective migration surfaces. Nat. Genet. 48:94–100.

Prag JRW, Quinn JC. 2011. The Hellenistic west: Rethinking the ancient mediterranean. Cambridge University Press

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575.

Rosenberg NA. 2004. DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138.

The 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1 , 092 human genomes. Nature 491:56–65.

Torroni A, Bandelt H-J, Macaulay V, Richards M, Cruciani F, Rengo C, Martinez-Cabrera V, Villems R, Kivisild T, Metspalu E, et al. 2001. A Signal, from Human mtDNA, of Postglacial Recolonization in Europe. Am. J. Hum. Genet. 69:844–852.

Uren C, Kim M, Martin AR, Bobo D, Gignoux CR, Van Helden PD, Möller M, Hoal EG, Henn BM. 2016. Fine-scale human population structure in Southern Africa reflects ecogeographic boundaries. Genetics 204:303–314.

Varela TA, Aínsua RL, Fariña J. 2003. Consanguinity in the Bishopric of Ourense

(Galicia, Spain) from 1900 to 1979. Ann. Hum. Biol. 30:419–433.

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.

3.7. Supplementary Material

Spatially Explicit Analysis Reveals Complex Genetic Gradients in the Iberian Peninsula

João Pimenta, Alexandra M. Lopes, Angel Carracedo, Miguel Arenas, António Amorim, David Comas

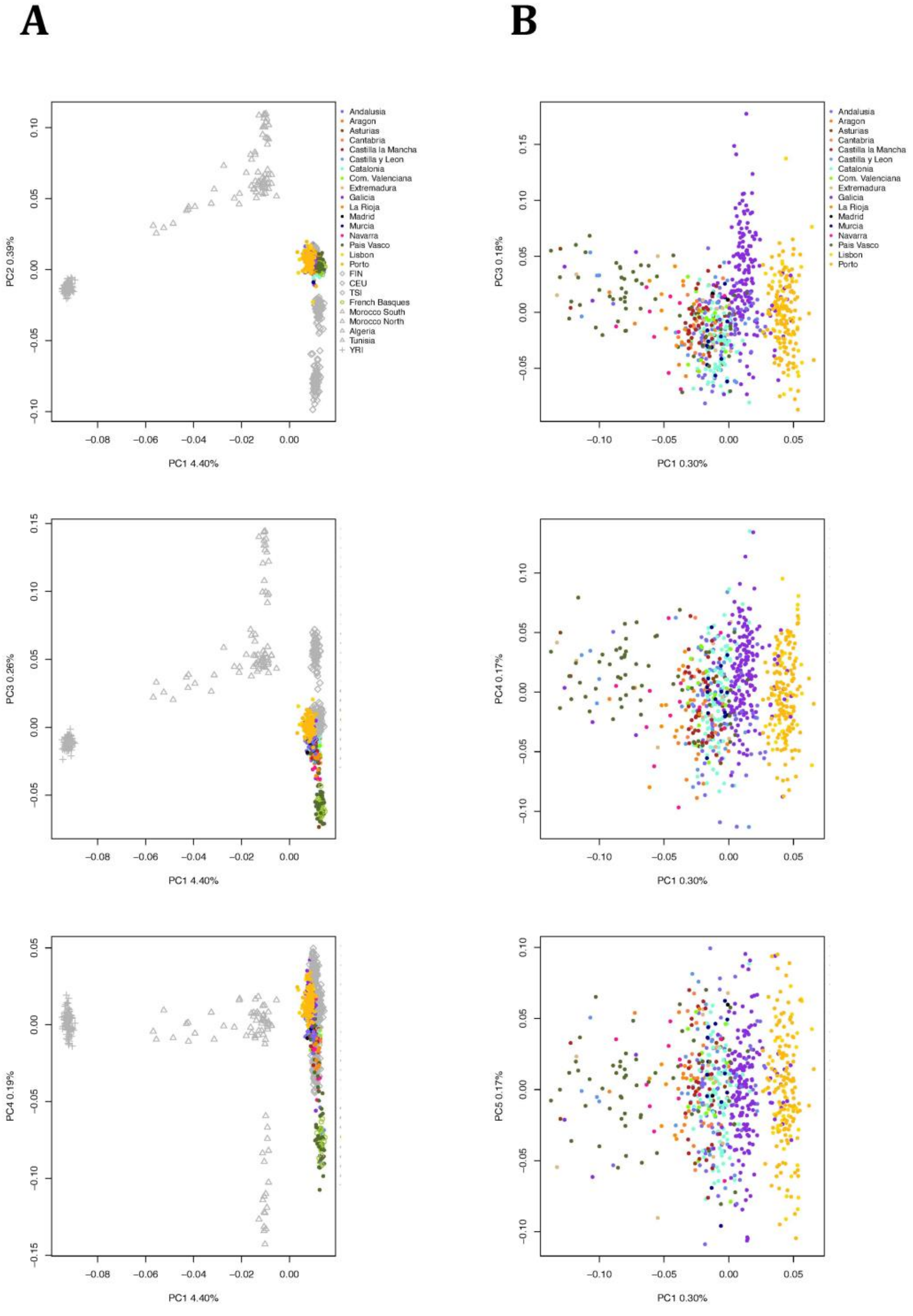


Figure S3.1. Population structure based on Principal Component Analysis. (A) PC1 against PC2, PC3, PC4 for the dataset of European and Africans samples. (B) PC1 against PC3, PC4, PC5 for the Iberian dataset.

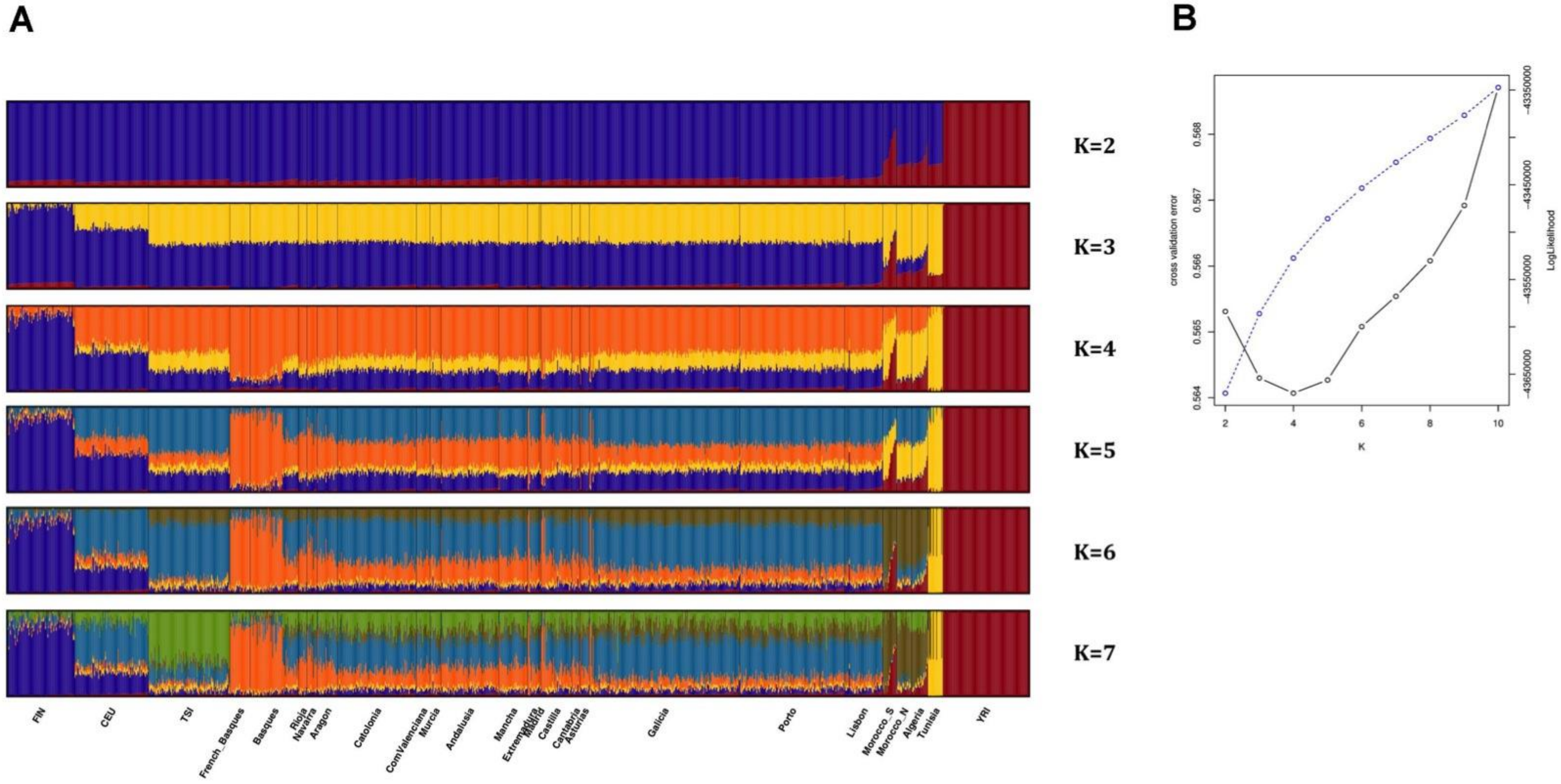


Figure S3.2. Ancestry inference estimated on the basis of model-based unsupervised clustering for the global dataset. (A) Ancestry plots for two to seven ancestral populations (K=2 to K=7) for all 26 populations included in the analysis. (B) Cross-validation and log-likelihood for the ancestry analysis for K=2 to K=10. For K=4, the majority of Iberian ancestry is derived from a component largely present in European populations (orange), while a Northern/Central European component (blue) is the second most relevant component in Iberian populations. A component associated with North Africa ancestry (yellow) is the third most relevant in Iberians and the Sub-Saharan component (red) is only present at a vestigial level.

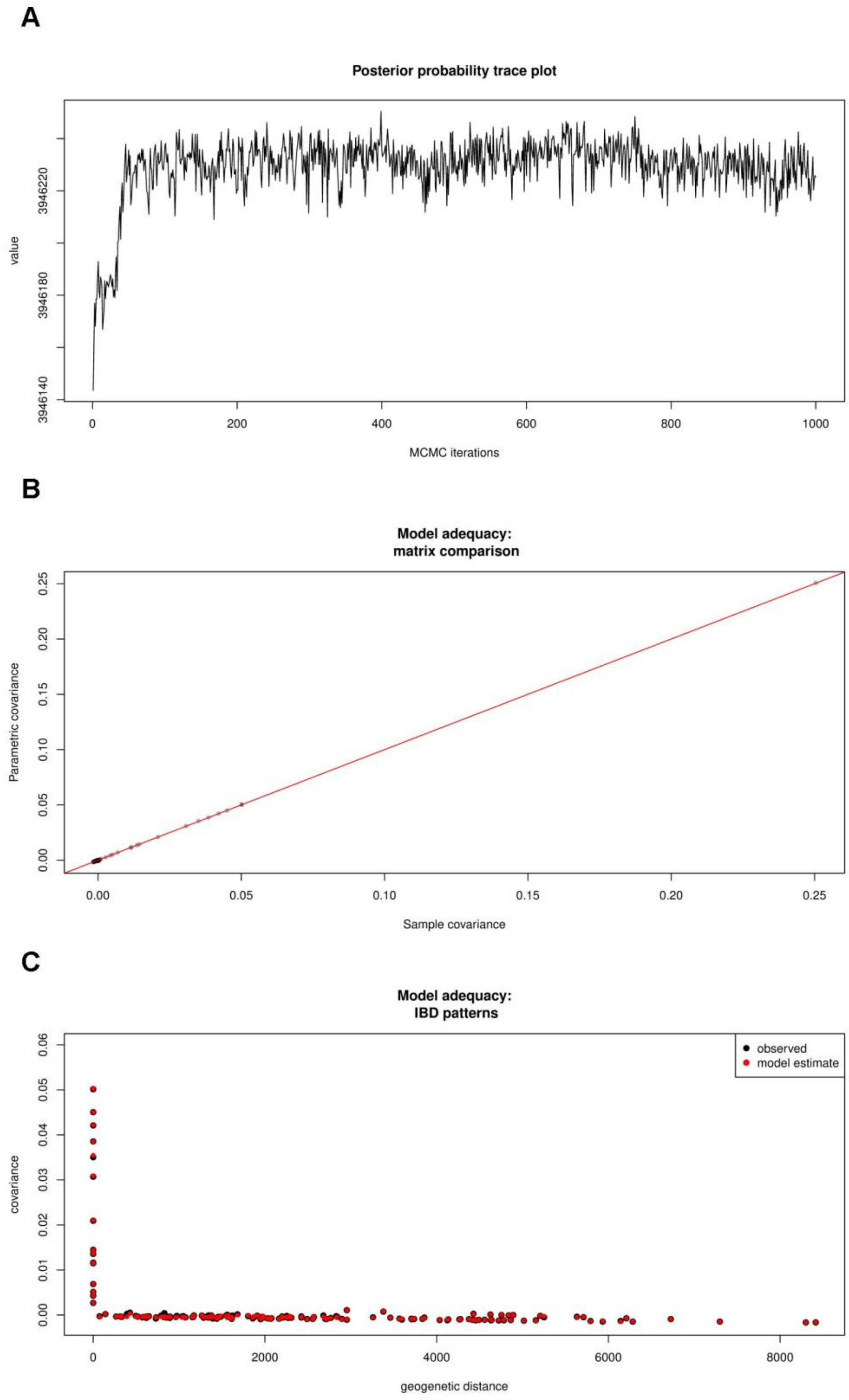


Figure S3.3. Performance and fitting of the SpaceMix model with the Iberian dataset for the autosomes. The performance of the approach was assessed by evaluating (A) the posterior probability trace derived from the MCMC chain, (B) the correlation between observed and parametric co-variance matrices and (C) the decay of co-variance with geographic distance for observed and estimated data.

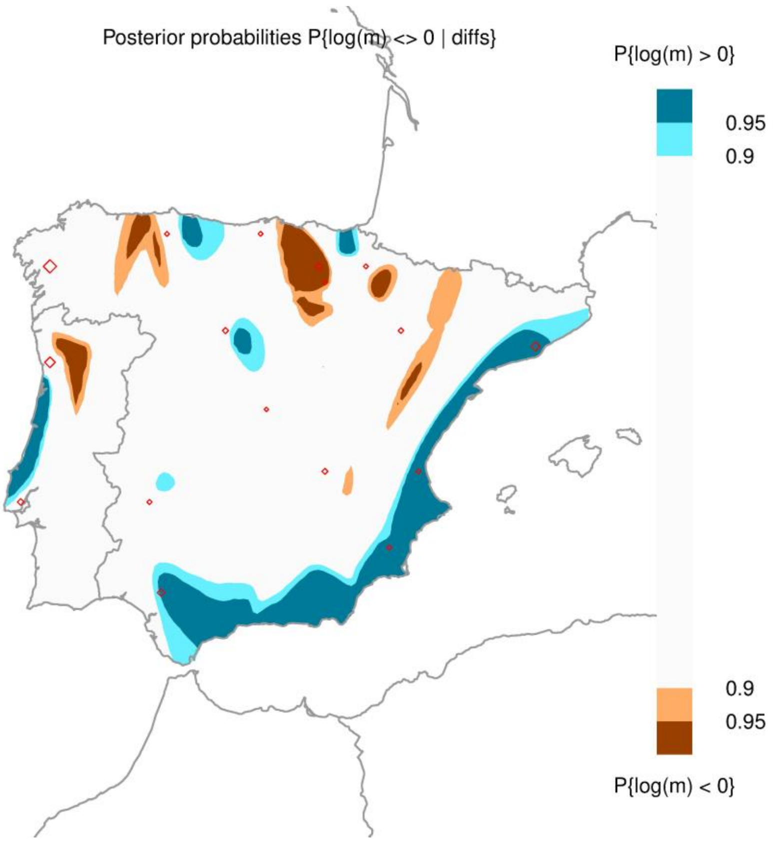


Figure S3.4. Corridors and barriers to gene flow estimated for autosomes for the Iberian dataset. This map highlights regions with a migration rate significantly high (blue) or low (brown) relative to the estimated mean.

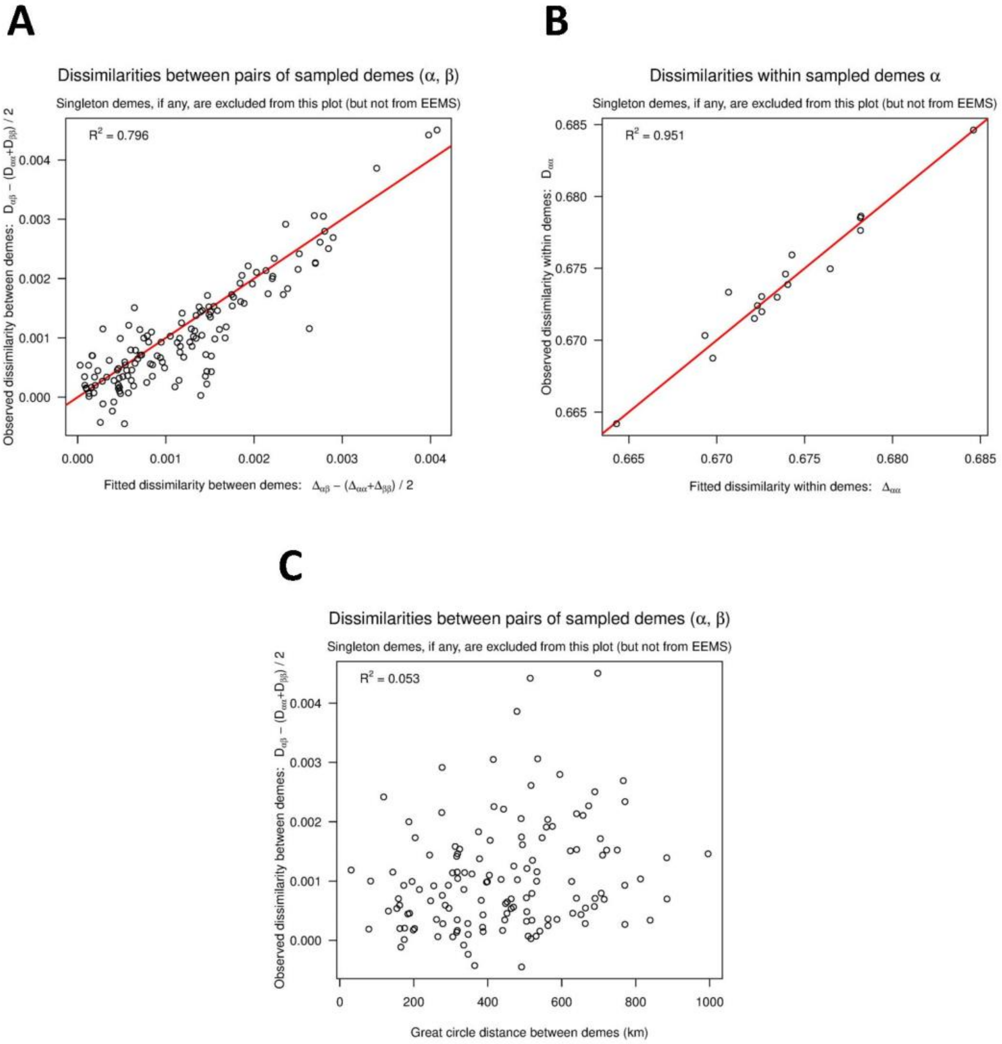


Figure S3.5. Diagnostic plots for EEMS model fitting for the autosomes. (A) Pairwise comparison of estimated and observed genetic dissimilarities between demes. (B) Pairwise comparison of estimated and observed genetic dissimilarities within demes. Note that the EEMS model presents an acceptable fitting with the observed data. (C) Scatter plot of observed genetic distances with geographic distances between populations. The R^2 coefficient was estimated for each scatterplot and shown at the top left of each plot.

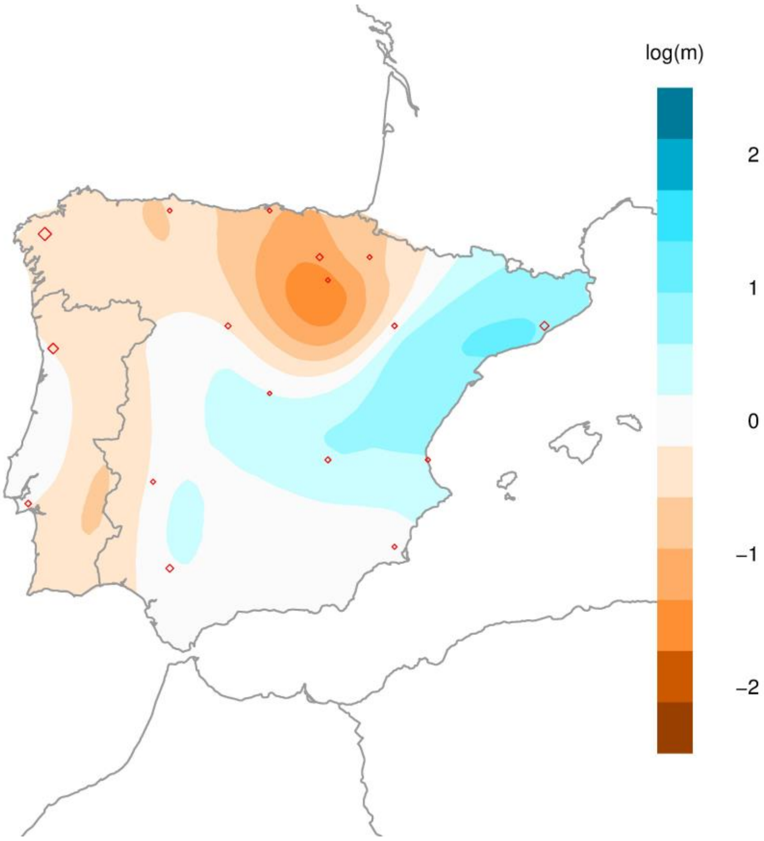


Figure S3.6. Effective migration map obtained for X chromosome based on the Iberian dataset. Plots were estimated with EEMS under a $\log_{10}$ scale and after mean centering. Regions with white colour present the mean effective migration rate. Blue regions indicate areas with migration rate higher than average, while the brown colour indicate regions with a low migration rate between demes relative to the mean.

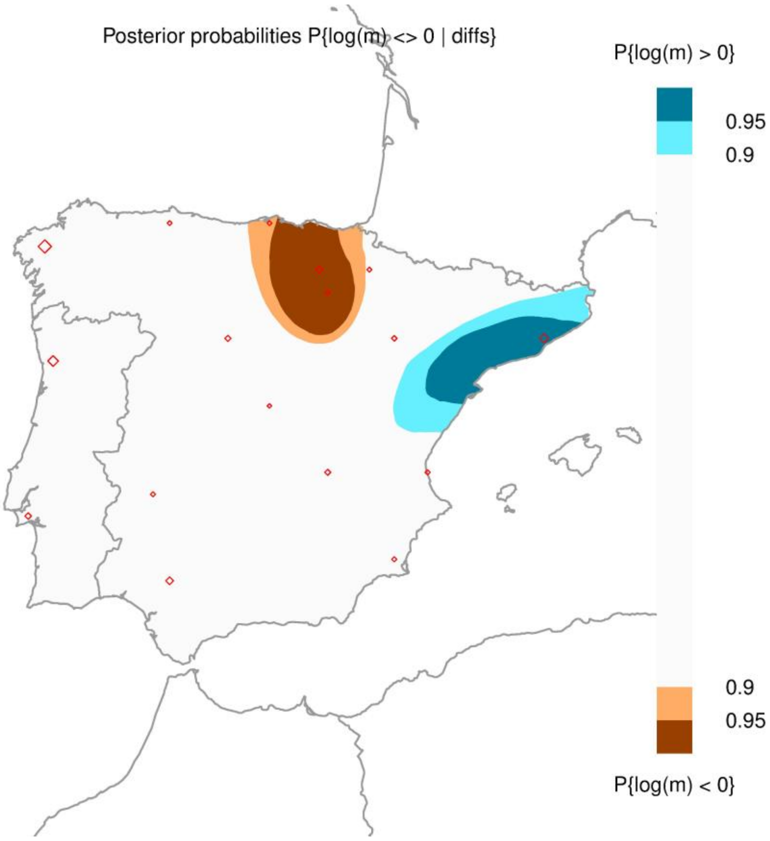


Figure S3.7. Corridors and barriers to gene flow based on the X chromosome. This map highlights regions with a migration rate significantly high (blue) or low (brown) relative to the estimated mean.

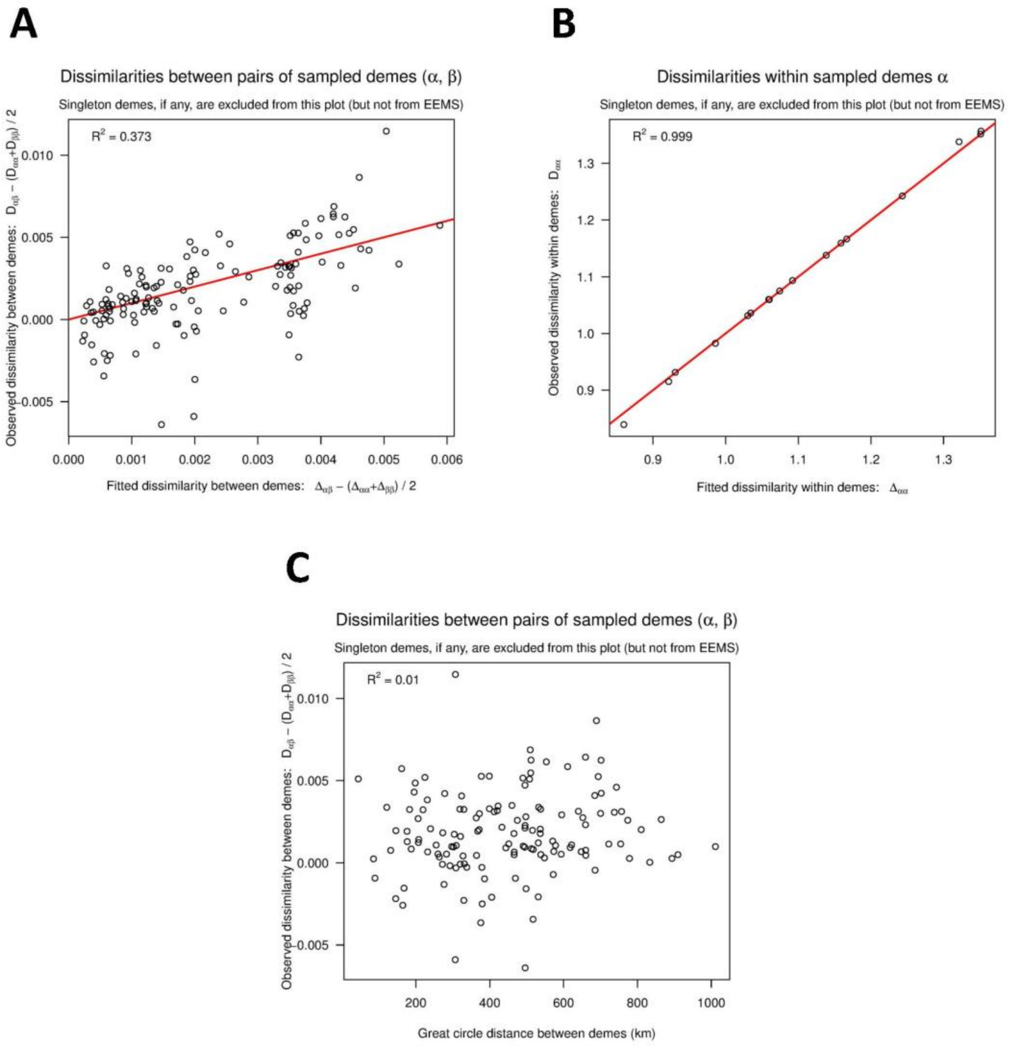


Figure S3.8. Diagnostic plots for EEMS model fitting for the X chromosome. (A) Pairwise comparison of estimated and observed genetic dissimilarities between demes. (B) Pairwise comparison of estimated and observed genetic dissimilarities within demes. Note that the EEMS model presents an acceptable fitting with the observed data. (C) Scatter plot of observed genetic distances with geographic distances between populations. The R^2 coefficient was estimated for each scatterplot and shown at the top left of each plot.

Chapter 4.

General Discussion

During the last decade, the increasing availability of genome wide datasets for modern and ancient populations has allowed the reconstruction of the evolutionary history of modern humans at unprecedented depths. It is well established that several past demographic events have reshaped the genetic, cultural and political landscape of human populations. Numerous methods have been developed to investigate, at a spatio-temporal level, the processes that caused current patterns of genetic diversity. Nevertheless, studying such processes and their effect on the population structure is a challenging problem, due to the complex evolutionary history of modern humans. In the present Thesis, I used model-based approaches to study the Neolithic transition in North Africa and Europe, the migratory dynamics between the Iberian Peninsula and North Africa (across the Strait of Gibraltar), and the spatial structure and migratory dynamics of populations within the Iberian Peninsula.

4.1. Neolithic transition in North Africa

The Neolithic transition in North Africa is still poorly understood, mainly due to the lack of attention that this region received from researchers and the poor environmental conditions for DNA conservation that makes unlikely to retrieve data from ancient human remains. In Chapter 2, we designed a complex demographic model to evaluate the Neolithic transition throughout North Africa, by comparing it to the farming expansion in southern Europe, which is fairly described, using genome-wide data for modern populations. First of all, I would like to point out a few limitations that our study presents. For instance, we were not able to include in our model aDNA data, which would allow to simulate genetic diversity of populations on several locations and at different times, giving a more comprehensive narrative to our model, since we could simulate the effects of a larger number of demographic processes. But, note that at the time we published our findings, aDNA results for the region were not yet available (Fregel et al. 2018). Second, the spatially explicit computer simulator (Ray et al. 2010), applied in our study, presents several aspects that are convenient to realistically model human evolution such as range expansions, spatial and temporal variable of the population growth rate and admixture between two populations.

Using publicly available genome-wide data for the X chromosome, our results clearly showed a similar expansion at both shores of the Mediterranean Sea, describing very similar rates of progression and demographic parameters for both regions. Indeed, our model predicted that the Neolithic arrived at approximately the same time to westernmost regions of the Mediterranean Sea (Iberian Peninsula and Morocco), which is in agreement with results for archaeological data (Linstädter 2008; Linstädter et al. 2012). We suggested that, as for Southern European regions, the spread of the Neolithic throughout North Africa most likely occurred by pioneering maritime migrations that started in the Near East, as previously suggested by results based on archaeological data (Zilhão 2014). However, while this latter assumption can explain the expansion throughout central and westernmost regions of North Africa, the adaptation to farming in eastern regions most likely had a larger cultural component, since autochthonous Epipaleolithic communities show evidence of exploitation of natural resources and resource management before the arrival of the Neolithic from the Near East (Di Lernia 2001; Garcea 2004; Mulazzani et al. 2016), which could facilitate the adaptation to farming lifestyle. Genetic results have demonstrated that the population history of North Africa is quite complex (Henn et al. 2012) and recent aDNA results from (Fregel et al. 2018) revealed an autochthonous North African component restricted to modern populations in the region (Henn et al. 2012) that can be traced back to Epipaleolithic populations from the Near East. Indeed, Fregel et al (2018) argued that the initial transition to Neolithic in Morocco largely involve the transmission of ideas rather than people, and only posterior stages of the Neolithic brought European-like ancestry to the region. This seems incongruent with the scenario presented in chapter 2. However, Fort (2012; 2015) developing models for the Neolithic transition in Europe, using archaeological data, showed that for a rapid spread of farming, as it is verified in the Mediterranean basin, both models of diffusion, cultural and demic, must have acted synergistically. Moreover, the scarcity of aDNA available for North Africa invalidates the formulation of a generalized model of transition in the region by applying this kind of data.

In all, we still have an incomplete picture of the history of the Neolithic expansion throughout North Africa, but our study provided the first insight into the spread of farming for the entire region by the direct comparison of the patterns of expansion over Europe and Northern Africa.

4.1.1. Migratory dynamics across the Strait of Gibraltar

The Mediterranean basin is an interesting region to study genetic structure and population dynamics since it constitutes a natural geographic barrier between Europe and North Africa, which promotes genetic differentiation without completely blocking the gene flow between the two regions. The Strait of Gibraltar represents the region where these continents are closest, separated by a stretch of water of only 14 km, which potentiated migratory movements between both regions. In chapter 2, we constructed and compare four different evolutionary scenarios (reciprocal migration absence of migration, migration only from the Iberian Peninsula towards North Africa and migration only from North Africa towards the Iberian Peninsula) to clarify the role of the Strait of Gibraltar as a point of contact between both continents.

For simplification, we assumed in our model that migration through the strait occurred only after the arrival of the Neolithic to the region. In this concern, previously Currat et al (2010) performing a model comparison study, demonstrated that gene flow across the Strait of Gibraltar occurred at least since the latter Palaeolithic but their estimations of migration rate were not precise and other studies pointed out the lack of evidence to support for pre-Neolithic migrations between both continents (Ferembach 1985; Bocquet-Appel and Demars 2000; Liu et al. 2006; Timmermann and Friedrich 2016).

Our results identified the Strait of Gibraltar as a corridor for gene flow between both regions but highlighted its role as a permeable barrier, since estimated migration rates for both directions were lower than migrations rates estimated for the Neolithic expansion. Our findings are clearly in agreement with previous results for uniparental (Bosch et al. 2001; Gonzalez et al. 2003; González-Pérez et al. 2003; Cruciani et al. 2004; Ennafaa et al. 2009; Rhouda et al. 2009) and genome-wide markers (Botigue et al. 2013; Arauna et al. 2017), which attest the robustness of our study to describe the migratory dynamics across the Strait of Gibraltar. Moreover, our study did not favor any single direction of migration, suggesting that both regions affected the genetic structure of the other in similar levels.

4.2. Genetic Structure of the Iberian Peninsula

Chapter 3 describes an analysis of the genetic spatial structure on the Iberian Peninsula. For this purpose we applied recently developed methods that model patterns of genetic differentiation across a landscape using georeferenced data (Petkova et al. 2015; Bradburd et al. 2016). Indeed, Iberia constitutes a suitable region to apply this type of analysis since its prehistory is marked by a series of intricate demographic events. Additionally, we performed an independent analysis of the X chromosome to gain insight into potential sex-biased patterns of genetic differentiation. We further compare our findings with results from a complementary study that described the fine-scale patterns of population structure in the Iberian Peninsula, using a Spanish cohort similar to the one included in our analysis and by applying a haplotype-based approach, called fineSTRUCTURE (Lawson et al. 2012).

Our results showed a strong genetic differentiation along an east-west cline than in the north-south direction and two contrasting regions in terms of genetic structure. Northern regions follow patterns of genetic differentiation that closely resemble the geographic delimitation of the catholic kingdoms during the early stages of Muslim occupation of the region, which was also observed by Bycroft el al (unpublished data). On the other hand, the spatial structure in central and southern regions did not fit the linguistic and geopolitical boundaries of the catholic kingdoms during and post *Reconquista*, which could suggest that posterior migrations reshaped the genetic structure of central and southern regions of the Iberian Peninsula. Interestingly, our results for these latter regions are not in full agreement with the fine-scale haplotype-based analysis performed by Bycroft el al (unpublished data), that shows a strong structure for the entire Iberian landscape. These two studies offer a different and complementary analysis on the fine-scale structure of Iberia, demonstrating that the patterns of genetic differentiation within the region can be explained by demographic events occurring during the last 1,000 years. Also, we demonstrated the possibility of sex-biased migrations due to the differences between the spatial patterns of genetic differentiation showed for autosomes and X chromosome. Altogether, our results suggest that isolation by distance cannot explain the patterns of genetic structure and that, at regional level, geographic features had not a large influence shaping those patterns.

4.3. Conclusion

Overall, the work presented in this thesis suggests, for the first time, a scenario where pioneer maritime migrants introduced the Neolithic innovation in westernmost regions of North Africa, in a similar process of dissemination of agriculture in Southern Europe. A dual model of demic and cultural diffusion offers the most parsimonious scenario for the expansion of the Neolithic throughout North Africa. Additional archaeological and genomic data will be needed to fully understand the process of Neolithization in the region. Additionally, we showed that the genetic structure of European and North Africa populations was influenced by reciprocal migration across the Strait of Gibraltar during the Neolithic, which could favor the similar population genetics patterns estimated for the Neolithic populations of both shores of the Mediterranean Sea.

Here we demonstrated that demographic events were more relevant than geographic features to shape the genetic patterns throughout the Iberian Peninsula. Indeed, the patterns of genetic differentiation seem mainly associated with political and cultural barriers related with the *Reconquista*.

Altogether, we consider that the present work of Thesis provides a significant advance in the understanding of the migratory patterns and population dynamics in the western regions of the Mediterranean Sea.

4.4 References

Arauna LR, Mendoza-Revilla J, Mas-Sandoval A, Izaabel H, Bekada A, Benhamamouch S, Fadhlaoui-Zid K, Zalloua P, Hellenthal G, Comas D. 2017. Recent Historical Migrations Have Shaped the Gene Pool of Arabs and Berbers in North Africa. Mol. Biol. Evol. 34:318–329.

Bocquet-Appel JP, Demars PY. 2000. Neanderthal contraction and modern human colonization of Europe. Antiquity 74:544–552.

Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, Bertranpetit J. 2001. High-Resolution Analysis of Human Y-Chromosome Variation Shows a Sharp Discontinuity and Limited Gene Flow between Northwestern Africa and the Iberian Peninsula. Am. J. Hum. Genet 68:1019–1029.

Botigue LR, Henn BM, Gravel S, Maples BK, Gignoux CR, Corona E, Atzmon G, Burns E, Ostrer H, Flores C, et al. 2013. Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. Proc. Natl. Acad. Sci. 110:11791–11796.

Bradburd GS, Ralph PL, Coop GM. 2016. A Spatial Framework for Understanding Population Structure and Admixture. PLoS Genet. 12.

Bycroft C, Fernández-Rozadilla C, Ruiz-Ponte C, Quintela-García I, Carracedo Á, Donnelly P, Myers S. unpublished data. https://www.biorxiv.org/content/early/2018/03/12/250191. last accessed June 5, 2018.

Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, Colomb EB, Zaharova B, et al. 2004. Phylogeographic Analysis of Haplogroup E3b (E-M215) Y Chromosomes Reveals Multiple Migratory Events Within and Out Of Africa. Am. J. Hum. Genet. 74:1014–1022.

Currat M, Poloni ES, Sanchez-Mazas A. 2010. Human genetic differentiation across the Strait of Gibraltar. BMC Evol. Biol. 10:237–255.

Ennafaa H, Cabrera VM, Abu-Amero KK, González AM, Amor MB, Bouhaha R, Dzimiri N, Elgaaïed AB, Larruga JM. 2009. Mitochondrial DNA haplogroup H structure in North Africa. BMC Genet. 10.

Ferembach D. 1985. On the origin of the iberomaurusians (Upper palaeolithic: North Africa). A new hypothesis. J. Hum. Evol. 14:393–397.

Fort J. 2012. Synthesis between demic and cultural diffusion in the Neolithic transition in Europe. Proc. Natl. Acad. Sci. 109:18669–18673.

Fort J. 2015. Demic and cultural diffusion propagated the Neolithic transition across different regions of Europe. J. R. Soc. Interface 12:20150166–20150166.

Fregel R, Mendez FL, Bokbot Y, Martin-Socas D, Camalich-Massieu MD, Avila-Arcos MC, Underhill PA, Shapiro B, Wojcik GL, Rasmussen M, et al. 2018. Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. Proc. Natl. Acad. Sci.:191569.

Garcea EAA. 2004. An alternative way towards food production: The perspective from the Libyan Sahara. J. World Prehistory 18:107–154.

González-Pérez E, Via M, Esteban E, López-Alomar A, Mazieres S, Harich N, Kandil M, Dugoujon J-M, Moral P. 2003. Alu insertions in the Iberian Peninsula and North West Africa - Genetic boundaries or melting pot? Coll. Antropol. 27:491–500.

Gonzalez AM, Brehm A, Pérez JA, Maca-Meyer N, Flores C, Cabrera VM. 2003. Mitochondrial DNA affinities at the atlantic fringe of Europe. Am. J. Phys. Anthropol. 120:391–404.

Henn BM, Botigué LR, Gravel S, Wang W, Brisbin A, Byrnes JK, Fadhlaoui-Zid K, Zalloua PA, Moreno-Estrada A, Bertranpetit J, et al. 2012. Genomic ancestry of North Africans supports back-to-Africa migrations. PLoS Genet. 8.

Lawson D, Hellenthal G, S M, Falush D. 2012. Inference of population structure using dense haplotype data. PLoS Genet. 8:e1002453.

Di Lernia S. 2001. Dismantling dung: Delayed use of food resources among Early Holocene foragers of the Libyan Sahara. J. Anthropol. Archaeol. 20:408–441.

Linstädter J. 2008. The Epipalaeolithic-Neolithic-Transition in the Mediterranean region of Northwest Africa. Quartär 55:41–62.

Linstädter J, Medved I, Solich M, Weniger GC. 2012. Neolithisation process within the Alboran territory: Models and possible African impact. Quat. Int. 274:219–232.

Liu H, Prugnolle F, Manica A, Balloux F. 2006. A Geographically Explicit Genetic Model of Worldwide Human-Settlement History. Am. J. Hum. Genet. 79:230–237.

Mulazzani S, Belhouchet L, Salanova L, Aouadi N, Dridi Y, Eddargach W, Morales J, Tombret O, Zazzo A, Zoughlami J. 2016. The emergence of the Neolithic in

North Africa: A new model for the Eastern Maghreb. Quat. Int. 410:123–143.

Petkova D, Novembre J, Stephens M. 2015. Visualizing spatial population structure with estimated effective migration surfaces. Nat. Genet. 48:94–100.

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.

Rhouda T, Martínez-Redondo D, Gómez-Durán A, Elmtili N, Idaomar M, Díez-Sánchez C, Montoya J, López-Pérez MJ, Ruiz-Pesini E. 2009. Moroccan mitochondrial genetic background suggests prehistoric human migrations across the Gibraltar Strait. Mitochondrion 9:402–407.

Timmermann A, Friedrich T. 2016. Late Pleistocene climate drivers of early human migration. Nature 538:92–95.

Zilhão J. 2014. Early prehistoric navigation in the Western Mediterranean: Implications for the Neolithic transition in Iberia and the Maghreb. Eurasian Prehistory 11:185–200.