Population genetic landscape of Basques

André Flores Bello

TESI DOCTORAL UPF / 2019

DIRECTOR DE LA TESI

Dr. David Comas

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

Universitat
Pompeu Fabra
Barcelona

Science knows no country, because knowledge belongs to humanity, and is the torch which illuminates the world.

Louis Pasteur

A todas las personas que de alguna forma me han acompañado a lo largo de este viaje de cuatro años.

A David. Por ser la brújula que me ha permitido seguir el sendero hasta la meta. Por la confianza, la paciencia y el apoyo durante todo este tiempo siempre que lo he necesitado. Gracias por ayudarme a crecer cuando las cosas no salían como uno esperaba.

A Francesc y Elena. Por ser el norte y sur con su ayuda a lo largo de estos cuatro años.

A todo el grupo. Tanto aquellos que ya han volado del nido como los que han continuado hasta el final acompañándome a crecer profesionalmente. Gracias por ayudarme a crear el mapa sobre el que trazar la ruta a seguir. Ha sido un verdadero privilegio poder aprender con vosotros.

A todo el Akelarre. Cada uno de vosotros habéis sido un pilar fundamental durante este viaje de crecimiento profesional y personal. Gracias por ser el asiento en primera clase durante el viaje.

A meus país e máis meu irmá, que ainda que ás veces a distancia non fora doada, nunca me deixaron só nesta viaxe. Moitísimas grazas por ser a estrela que me indica o camiño cando non hai luz.

A Juan. Gracias por ser las alas del avión que me ha permitido terminar el viaje.

Abstract

Basque people have been the focus of many studies in the last decades due to cultural and genetic characteristics that place them as an outlier population within the European landscape. Despite the existence of large amount of studies, several controversies exist in the population genetics field: their highest frequencies of Rh-negative, a strong genetic differentiation, a genetic substructure, and their connections with the main ancient European groups along history. Most of the previous studies focused on Basques have suffered from methodological flowbacks and weak study designs. In this thesis, these controversies are elucidated by using a more refined and accurate methodology, together with an exhaustive ethnolinguistic representation of the Basque population and surrounding groups. The main results show a high frequency of Rh-negative in agreement with previous studies, an internal genetic heterogeneity within Basques, together with a clear differentiation from the external populations, probably marked by a genetic continuity from the Bronze Age.

Resumen

La población vasca ha sido objecto de estudio durante las últimas décadas, debido a características culturales y genéticas que parecen definirlos como una población aislada dentro del contexto europeo. A pesar de la existencia de numerosos estudios, diversas controversias han surgido en el área de genética de poblaciones: una de las frecuencias más altas de Rh-negativo, una fuerte diferenciación genética, una subestructura poblacional y la relación con las principales poblaciones antiguas en Europa. Los estudios previos han estado sujetos a una metodología y un muestreo limitados. En esta tesis, se abordan dichas controversias con una metodología más refinada y precisa, además de una representación exhaustiva de la población vasca y de poblaciones circundantes desde un punto de vista etnolingüístico. Los principales resultados muestran: una alta frecuencia de Rh-negativo similar a la observada en trabajos anteriores, una heterogeneidad genética interna, junto con una clara diferenciación de las poblaciones externas, marcada posiblemente por una continuidad genética desde la Edad de Bronce.

Preface

Human population genetics is the field focused on the analysis of the genetic variation among human populations caused by the demographic processes. Together with other areas of knowledge such as history, linguistics and anthropology, they can provide insights into the roots of the origins, evolution and demographic history of human populations. The technological and analytical advances in genetics and genomics during the last decade have prompted the fast growth of the human population genetics. This increase of the studies in the field has been reflected in the large number of modern and ancient samples that are currently available.

In Europe, the genetic analyses have proved a relative homogeneous scenario where the genetic distances among populations are explained by their geographic closeness. In this general context, different outlier populations stood out breaking the general genetic pattern of the continent. Due to their peculiarities, isolated populations have been the focus of many fields. Anthropologists have studied the specific cultural traits that usually come together with the isolation. In medicine, the focus has been the genetic screening for rare monogenetic and complex diseases. Population geneticists have focused on the study of the genetic structure and the processes of genome variation.

Basques became an important object of attention. Several cultural and genetic traits pointed them as one of the most isolated and ancient population. A still lively debate emerged around Basque people’s genetics related to: a high frequency of Rh-negative, a genetic differentiation from the external populations, an internal heterogeneity and an unclear origin.

In this thesis, the main genetic controversies of Basques are deep disentangled, by overcoming the limitations in the previous studies by improving the ethno-linguistic representation, and using recent technical, methodological and computational advances in order to give finer and more powerful conclusions.

CONTENTS

1. INTRODUCTION

1.1. The interest in the Basques

From the beginnings of the revolution of human population genetics with Cavalli-Sforza (Cavalli-Sforza and Bodmer 1971; Edwards and Cavalli-Sforza 1965), the general context of the modern Europe has been described as a relatively homogeneous genetic scenario, compared to the higher genetic distances observed between populations in other continents (Novembre et al. 2008; Jorde et al. 2000; Nelis et al. 2009; Lao et al. 2008; Chikhi et al. 1998). In this general background, Basques, Sardinians, Sami and Icelanders stood out in the analyses and separating from the expected pattern of the other populations in the continent (Sajantila et al. 1995; Helgason et al. 2000; Calafell and Bertranpetit 1994b; Cavalli-Sforza 1997; Renfrew 2010; Cavalli-Sforza, Menozzi, and Piazza 1994). At this moment, a general interest arose within the scientific community in understanding the singular genetic aspects of these groups of people and detect new groups with these characteristics, becoming the focus of different fields during the next decades until today.

Between the European isolates, Basques positioned as one of the most intriguing unknowns, many times subject of speculation. Different disciplines have been in agreement with the peculiarity and importance of the Basque people (Rosario and Rebato 1997; López and Alonso 2013; Mitxelena Elissalt 1981; Bertranpetit et al. 1995; Collins 1990; Igartua and Zabaltza 2016; Laska et al. 2009), even before the existence of human population genetics (Wells

1920). It has been also suggested a specific Basque morphometric, specially related to craniometrical traits (Wells 1920; de la Rúa 1992). Archeology, prehistory and linguistics started to stress the curious evidences around them. Their unique non-Indo-European language in Western Europe, the Euskara, without any direct connection to the extant languages is the most important cultural trait of the Basque people. The archaeological remains in their area, dating from Palaeolithic times, evidenced that it acted as one of the most densely populated glacial refuge during the Last Glacial Maximum (LGM) (Wren and Burke 2019; Gamble et al. 2006; Villar and Prósper 2005; Straus 1990; de las Heras et al. 2012). The advances in Genetics during the firsts 50’s until the consolidation of human population genetics in the 80’s, allowed to study the connection between groups of people from a new perspective (Bodmer 2015; R. L. Cann 2006; Charlesworth and Charlesworth 2017). The higher frequencies of Rhesus (Rh) negative and O blood groups detected in Basques from the first studies (Mourant 1947, 1954; Etcheverry 1947), together with the following works based on other new genetic markers (Mourant, Kopeć, and Domaniewska-Sobczak 1976; Cavalli-Sforza and Bodmer 1971), gave rise to think about the unique and independent history of this group: Basques were defined as direct descendants from the European Mesolithic hunter-gatherers that inhabited the region during the LGM and remained isolated afterwards, avoiding the connection with the following people movements, and therefore maintaining their non-Indo-European language and ancestral Pre-Neolithic genetic traits (Bertranpetit and Cavalli-Sforza 1991; Calafell and Bertranpetit

1994a). This was the beginning of a still open controversial debate. The new genetic insights and more available techniques in the last decades made possible deeper and specific studies. Thus, the interest around Basques was then faced by using higher number and different DNA markers, such as uniparental markers, array and sequence data. However, despite the large amount of studies focused on Basques, several controversies started to emerge. The causes of the higher frequencies of Rh-negative remained unclear. The differentiation between Basques and the surrounding European populations became diffuse, and the evidences of a possible internal heterogeneity appeared (Manzano et al. 1996). Furthermore, the direct connection with the Mesolithic hunter-gatherers started to be replaced with the idea of a possible genetic continuity with posterior incoming groups of people, i.e. Neolithic farmers and Bronze Age herders (Günther et al. 2015; Olalde et al. 2019). Studies with opposite conclusions came to light (Laayouni, Calafell, and Bertranpetit 2010; Rodríguez-Ezpeleta et al. 2010). These controversies lied on a weak study design without an appropriate ethno-linguistic context, together with the limited genetic methods used. Most of the studies have been performed based on classical markers, which may not reflect the actual demographic processes due to the effect of natural selection. Moreover, just a few studies included all those groups where Euskara is spoken nowadays, or was spoken before in the history, in order to have a complete representation of the historical Basque area (Martínez-Cruz et al. 2012; Behar et al. 2012).

1.2. The Basque people: a western Pyrenees culture

Basque people are described as the ethno-linguistic groups that have been historically distributed in the Franco-Cantabrian region alongside the western edge of the Pyrenees towards the Atlantic area, maintaining a related history, plus cultural and linguistic traits. This area between the Garonne and the Ebro, has been historically associated to the concept of *Vasconia* or *Euskal Herria,* used in different moments in the history to design the Basque-related area, where Euskara was spoken (Montero 2008). The distribution of the area has been determined by the historical records and the present-day distribution of the Euskara. Thus, it is complicated to precisely define its limits along history. It would be represented somehow by the seven present provinces of: Bizkaia (Biscay), Gipuzkoa, Araba (Alava) and Nafarroa (Navarre) in the southern side of the Pyrenees, in Spain; and Lapurdi (Labourd), Nafarroa Beherea (Lower Navarre) and Zuberoa (Soule), in the northern part, in France. See Figure 1. Nevertheless, the influence of the Euskara has been suggested beyond this geographic delimitation, pointing that probably the area inhabited by Basque-related people could have been wider in some periods (Nuñez Astrain 2004).

Figure 1. Map of the seven Basque provinces. BIZ, Bizkaia; ARA, Araba; GIP, Gipuzkoa; NAF, Nafarroa; LAP, Lapurdi; NBE, Nafarroa Beherea; ZUB, Zuberoa.

A common specific continuity of culture and language cannot be assured from prehistorically times due to the limited records until the arrival of the Romans to the Franco-Cantabrian region (Villar and Prósper 2005). However, it is widely known the important people settlement during the Palaeolithic and Mesolithic comparing to the main rest of Europe (Pala et al. 2012; Sazzini, Sarno, and Luiselli 2014). In addition, it is also evident the conservation of cultural traits and their pre-Indo-European language during post-Neolithic times (i.e. Chalcolithic and Bronze Age onwards), which are associated with the expansion of the Indo-European languages through Europe (Haak et al. 2015), and the Latin during the expansion of the Romans (Nuñez Astrain 2004; Collins 1990; Montero 2008). The first descriptions of the Basques from the

Greek and Roman historians and geographical writers described them as tribes with a common language, and some cultural traits that they do not explained in depth, but they were related, from the point of view of their refine society, to an overdue agricultural and pastoral lifestyles organized in small family-centered settlements (F. Ruiz 2018; Montero 2008). Even in the first literary reference to the Basques, by the Greek geographer Strabo (64 before common era, BCE, to 24 common era, CE) in his work *Geographica* (Jones 1917), singular cultural features specific for any tribe are highlighted. He limited to barely explain the lifestyle in general for the region, with diffuse definition of the geographical and cultural limits between those tribes. This allows to understand this region as a widely populated area with several groups of people spread alongside, quite separated, but connected by common language and cultural traits, with a remarkable cultural continuity in the pre-Roman periods completely unlinked to any political administration (Collins 1990). The Roman period in the Basque area involves the arrival of the first political concepts and boundaries. It implies the beginning of the consolidation of the pre-Roman tribes as a whole entity with a delimited geographic area and the posterior reorganizations in the following centuries (Collins 1990; Montero 2008).

This cultural continuity should not be misled understanding a stereotyped hermetic society without contact to other external groups of people along history, but the evolution of a common culture and language in the region together with the coexistence with external sources and the assimilation of new traits. In fact, bringing forward the content of the next chapter, despite the preservation of the Euskara, the cultural influence during the Neolithic and the Indo-European incomers in the area is reflected in the archaeological remains, with typical tools and buildings of the Indo-European societies (Carretero Díaz et al. 2008; Montero 2008). The cultural sharing during the Roman Period is the clearest example of that, with the emergence of urbanized areas in many parts of the region, as well as the economic activity (Montero 2008; Collins 1990; Zubillaga Gil 1990; Thierry 2010; Ceberio Rodríguez 2009; Urteaga Artigas 1988). The influence of other languages in the Euskara, specially Latin, is also a fact that reinforces those historical connections (Nuñez Astrain 2004; Zuazo 2010).

1.3. A travel time: from prehistory to history

"The scant, fragmentary and frequently contradictory evidence for the earliest recorded periods of Basque history is almost too frail to bear any substantial construction and can reduce the aspiring chronicler to the rank of novelist"
(Collins 1990).

Following the words from Collins, the controversy around Basques is also tangible in the historical records. Thus, to not err on the side of novelist, a brief review throughout the available information of the history of the Basque people is shown in this chapter to give a historical context to complement the genetic results of the thesis, from the obscure prehistory, following with the arrival of the Romans and ending with the unification of the Basque people as a political entity in the already advanced Middle Age.

1.3.1. The prehistory of the Basque people

The prehistory of the Basques encompasses approximately from the Upper-Palaeolithic (c. 30,000 BCE) with the first evidences of presence of *Homo sapiens*, to the arrival of the Romans in the 2nd century BCE, when the first written records in the area where produced (Montero 2008). The historical information related to these periods is scanty and it is based on the archaeological remains found in the Franco-Cantabrian region.

The Upper-Palaeolithic period in the Franco-Cantabrian region (c. 30,000 - 10,000 BCE) is marked by the presence of the hunter-gatherer people, who obtained most of the food by foraging, and the peak of the Last Glacial Maximum (LGM) (Montero 2008; Sazzini, Sarno, and Luiselli 2014; de las Heras et al. 2012). During this final glacial period much of the northern hemisphere was uninhabitable, covered by large ice sheets (Mithen 2004). In Europe, the northern populations were forced to move towards the southern regions, where the glacial refugia maintained the proper conditions to survive (Sazzini, Sarno, and Luiselli 2014; Pala et al. 2012; Barbujani and Bertorelle 2001) Figure 2. Those glacial refugia were distributed along the Mediterranean area, especially in the Franco-Cantabrian region, as well as the Italian and Balkan peninsulas.

Here, the new incomers established contact with the southern people that already dwelled these European regions (Sazzini, Sarno, and Luiselli 2014). This new merge of people remained in the glacial refugia until the climatic conditions started to improve and the ice sheet started to melt, leading to the beginning of a slow but gradual expansion (Achilli et al. 2004; Barbujani and Bertorelle 2001). Based on the archaeological data recovered, the Franco-Cantabrian region draws attention to have been one of the most densely populated area in Europe during this dramatic period (Sazzini, Sarno, and Luiselli 2014). Due to the climatic conditions, the hunter-gatherers in the region were mainly relegated to the life in caves. This period outstands in the region because of the emergence of the first evidences of human art in the paintings of important caves, such as the one of Altamira and Izturitz (de las Heras et al. 2012).

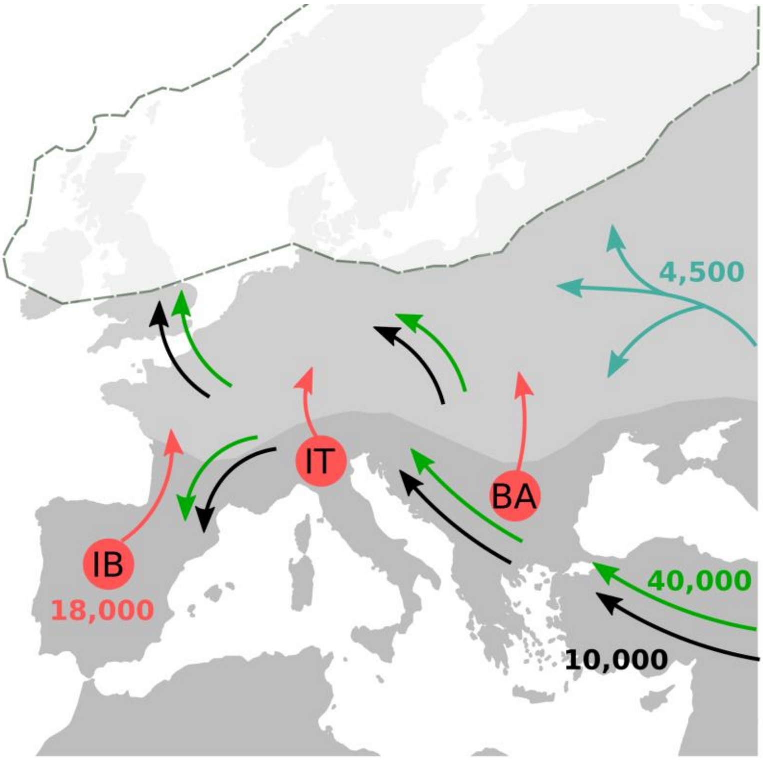


Figure 2. Schematic demographic processes in Europe from Palaeolithic to Late Neolithic. The Palaeolithic colonization c. 40,000 years before present (ybp) in green. The representation of the glacial refugia in red during the LGM, and the ice sheets in white along the northern area of Europe. The post-glacial expansions starting around c. 18,000 ypb. The Neolithic expansion c. 10,000 ybp in black. And the arrival of the steppe herders in central Europe c. 4,500 ybp. IB, Iberian Peninsula; IT, Italy; BA, Balkans. Modified from: (Barbujani and Bertorelle 2001; Sazzini, Sarno, and Luiselli 2014).

The Mesolithic (c. 10,000 – 5,000 BCE) is characterized by a transition period. The end of the LGM boosted the movements beyond the glacial refugia that had been started during the end of the Palaeolithic (Figure 2). The hunter-gatherers left behind the caves, but maintained the same foraging lifestyle. This wide geographical re-peopling of Europe, from the admixed and reduced populations of the glacial refugia, should have been a critical factor in reshaping the genetic landscape and culture of Europe, compared to the previous widespread Palaeolithic patterns (Achilli et al. 2004; Rootsi 2006; Pala et al. 2012). It has been suggested the Franco-Cantabrian region as the major source in this process, and the posterior isolation of the remaining people in the region that ended becoming the known Basque people. These hypotheses pointed to them as ancient European proxies. Although it had occurred, the archaeological remains seem to show that, in spite of a possible lower assimilation of the subsequent cultural waves in some areas or periods, they followed similar steps as in other European populations at the end (Montero 2008).

During the Neolithic period (c. 5,000 – 2,000 BCE) an important, but gradual, cultural revolution took place in the region with the arrival of the agriculture and animal domestication, that had been born in the Near East around ~9,000 BCE (Gangal, Sarson, and Shukurov 2014) (Figure 2). The people started a producing lifestyle, leaving the hunting and gathering. However, this process was more gradual than in the general European context. The foraging still remained the main lifestyle in many parts of the area (Cava 1990). Another important event in the Neolithic was the emergence of the pottery, related to the Cardial Ware culture (Alday Ruiz 2009). It was not a simultaneous process; it seemed to have been higher in the southern part of the region, while in the northern Atlantic basin the evidences point to a smaller impact (Cava 1990; Alday Ruiz 2009).

The Bronze Age and Iron Age (c. 2,000 BCE– Roman Period) are characterized by a cultural continuity, with a very slow, scarce and irregular introduction of new cultural traits from the incoming Indo-European people from the steppes (Gorrochategui and Yarritu 1990; Montero 2008) (Figure 2). In this period, the dichotomy between the Eastern and Western part in the assimilation of the innovations is still visible in the archaeological remains. In the general context of Europe, this period is a crucial chapter since it is related to the expansion of the Indo-European languages (Haak et al. 2015). Unlike in most of the rest of the continent, the Euskara overcame to the influence of the new widespread languages. The survival of the Euskara in this time, as the unique non-Indo-European language in Western Europe, is still quite an enigma.

1.3.2. The history of the Basque people

The presence of the Romans, between the 2nd century BCE and the 5th century CE, involve the beginning of the history of Basques. The first literary references related to Basque people appear in this period, from important historians and geographers like Ptolemy, Pliny the Elder or the already mentioned Strabo (Collins 1990; Kurlansky 1999). They described six groups of people or tribes living in the region, from east to west: *Aquitani*, in the northern area of the Pyrenees; and *Vascones*, *Varduli*, *Caristi*, *Autrigones* and

Berones in the southern (Collins 1990; Montero 2008) (Figure 3) The descriptions of these tribes were fairly diffuse, every writer gave them different distributions, names and characteristics, but they were defined with similar ethnic and linguistic profiles. So that at the end of the 5th century the term *Vascones* is started to be used to refer in general to all the people inhabiting the area (Montero 2008).

Figure 3. Approximate distribution of the pre-Roman tribes. Dotted lines indicates the area of the correspondent tribe. Modified from: (Martínez-Cruz et al. 2012).

This period has been suggested to be characterized collaborative relationship between the tribes and the Romans (Montero 2008; Amela Valverde 2000; Collins 1990; Kurlansky 1999). This situation led to a political, economic and infrastructural rise in the region with the construction of the first road network, connecting the region to the external areas with military and commercial purposes. The first cities were founded, starting with *Pompaelo* (current Pamplona) in the 1st century BC (Amela Valverde 2000). Thus, the emergence of a political-administrative structure started in the region. From the 3rd century onwards the Roman Empire would start the crisis that would lead to the fall of the Western Roman Empire in the 5th century (Goldsworthy 2009).

The beginning of the Early Middle Ages, were marked by a complicated period based on confrontation and the pressure from the groups that settled around them. Franks in the north and Visigoths in the south, both did not stop, trying to control the territory. However, they just could sporadically dominate some areas (Collins 1990; Montero 2008; Waldman and Mason 2006). Around the 7th century, the first political crystallization as an autonomous region of the Basques from both sides of the Pyrenees took place with the formation of the Duchy of Vasconia (Waldman and Mason 2006). The bellicose situation continued with the arrival of the Arabs in the Iberian Peninsula in the year 711 and the replacement of the Visigoths. Arabs swept deep into the Iberian Peninsula, but they barely influenced in the north (Waldman and Mason 2006; Ruiz 2018).

During the next years, synchronically to The *Reconquista*, several politic situations took place related to the Basques. The pressures in the south with the Arabs and in the north with the Franks led to the dilution of the Duchy of Vasconia in the year 824 (Waldman and

Mason 2006; Collins 1990). It was separated into Gascony, that merge with the Duchy of Aquitaine leaving behind the connection to Basque people, and the Kingdom of Pamplona (and later would be Kingdom of Navarre) (Waldman and Mason 2006), that finally included a similar area as the present-day Basque-related region. Along this period, an internal political delimitation into communities within the Kingdom of Navarra took place, shaping a similar distribution as the actual provinces (Waldman and Mason 2006). The following years after were based on a complex political situation in the Iberian Peninsula related to The *Reconquista*, with the formation and union of the Kingdom of Castile and Aragon in the 15th century (Thomas 2005) until the assimilation of Biscay, Alava, Gipuzkoa and Navarre to the Spanish territories and Labourd, Soule and Lower Navarre to the French ones.

1.4. The Euskara and its dialects

1.4.1. The distribution

The Euskara is an isolated non-Indo-European language with no close relation to any present language in the world (Hualde, Lakarra, and Trask 1996; Comrie, Polinsky, and Matthews 2003; Villar and Prósper 2005; Zuazo 2010). It is spoken in the historically Basque-related territory, i.e. the seven present-day regions that have been mentioned above, but the scholars suggest that it had been spoken in a larger area in the past due to recorded evidences from inscriptions, classical texts and linguistic remnants in those areas (Caro Baroja 1973; Gorrochategui Churruca 1984;

Luchaire 1877; Mitxelena Elissalt 1976; Rohlfs 1935; Zuazo 2010; Nuñez Astrain 2004). However, the current Euskara-speaking area does not encompass completely the seven provinces (Figure 4).

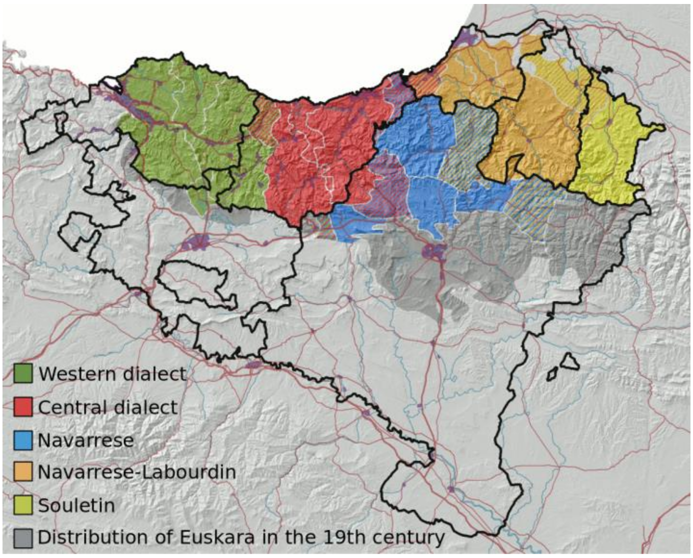


Figure 4. Present distribution of the Euskara dialects based on Koldo Zuazo's classification. Modified from (Igartua and Zabaltza 2016).

The Euskara was spoken in the western part of Biscay until the Middle Ages. Furthermore, the use of Euskara has been dramatically reduced since the 19th century in the southern area of Alava and the southeastern part of Navarre, including the valley of Roncal in the north (Abasolo 2002; Nuñez Astrain 2004; Borrás 2008) (Figure 4). In the present-day Euskara-speaking territory, the 33.9% of the population speaks the language, and the 19.1% is passive speaker. All of them are bilingual, also speaking Spanish or French in the southern or northern areas respectively. The other 47% speaks Spanish or French. (Eusko Jaurlaritza 2016).

Beyond the seven provinces, the Euskara was spoken in almost all the area of Aquitania in the northern side of the Pyrenees, below the Garonne River, at the arrival of the Romans. Furthermore, in the southern part, there are also evidences of a pre-Roman language related to Euskara in the northern areas of La Rioja, Burgos and Aragon (Nuñez Astrain 2004; López-Mugartza Iriarte 2008) (Figure 5). In the areas surrounding the seven provinces, there was actually bilingualism with other languages at that time. In Aquitania, the Euskara coexisted with the Gallo, whereas in the northern Iberian Peninsula it lived together with the Celtiberian and Iberian, in the western and eastern areas respectively.

Figure 5. Distribution of the archaic Euskara in the 1st century suggested by Luis Núñez Astrain. Source: (Nuñez Astrain 2004).

1.4.2. The dialects

"Everyone knows how the manner of speaking almost changes from one house to the next in Euskal Herria"
(Leizarraga, 1571)

These words, from the Basque priest Joanes Leizarraga (1506-1601), in the prologue of his translation of the New Testament to Euskara (Zuazo 2010), reflect clearly, simply and directly the internal dialectal diversity within the Euskara.

The Euskara has been actually the coexistence among five different dialects (Figure 4) distributed along the seven mentioned regions, at least from their origins, that are still quite disputed as it will be discussed in the next point. Thus, the normative Euskara or *Euskara batua* (from bat-, one; batu-, to unite) is the result of the process of unification, completed in 1968, of these dialects to be used as register in the administration, education and media (Zuazo 2010; Nuñez Astrain 2004).

It is widely, and wrongly, assumed that the dialects of the Euskara are abundant and different enough to be considered independent languages, although it is true that there is variation among them and, in some cases, they are more distant. They are actually several shared linguistic varieties among some dialects that overlap along the geography, and not a strict linguistic differentiation by areas (Zuazo 2010; Abaitua 2018). Thus, the distribution and classification of the dialects has been fairly diffuse and difficult to define. Different scholars have established different number and distributions of the dialects, reinforcing the difficulties of their classification. In fact, just focusing on two of the most important works defining the dialects, they are very different. In the first classification, by the historian Arnaut Oihenart (1592-1667), he described four dialects, related to the pre-Roman tribes in the area (Domene Verdú 2011). On the other side, Louis Lucien Bonaparte (1769- 1821), linguist and nephew of Napoleon Bonaparte, differentiated eight dialects and mentioned that the number could change depending on the characteristics considered (Bonaparte 1869). However, the number of dialects that modern scholars acknowledge in the present are five, although with internal variations (Zuazo 2010; Abaitua 2018): Western or Biscayan dialect, Central or Gipuzkoan dialect, Navarrese dialect, Navarrese-Labourdin dialect, and Souletin dialect (Figure 4). Another specific dialect was considered until the end of the 20th century, the Roncalese, but it has already been extinct (Estornés Lasa 1983). Despite the classification, only the dialects in the extremes are considered as the most specific and differentiated, while among central dialects the distances are small and are mainly related to the influences from the Spanish and French languages. Furthermore, the distances are smaller as closer are the dialects. Thus, a West-East/Gipuzkoan-Souletin cline cloud be defined through the Euskara-speaking area (Zuazo 2010; Abaitua 2018).

1.4.3. The origins

The origin and survival of the Euskara to the expansion of the Indo-European languages, plus the Latin and the following derived languages until today is still a lively open question, subject of many interpretations. Unraveling this issue could shed light on the suggested isolation of Basques during the time, possibly related to the cultural barrier that could suppose the language.

The study of the origins of the Euskara and its connections to other languages is quite complex, due to the scant written records during its history until the 16th century (Nuñez Astrain 2004; Waldman and Mason 2006; Igartua and Zabaltza 2016; Villasante 1979). Some inscription evidences in Aquitaine that date from 1st to 3th centuries CE, show a link to the contemporary Euskara (Nuñez Astrain 2004). Iit has been demonstrated that this ancient Aquitaine was actually an older form of the Euskara at that time (Igartua and Zabaltza 2016). The scholars have failed trying to find connections to any other language in the world: the ancient non-Indo-European Iberian (Bedialauneta Marcos 1992), the Caucasian group of languages, the Finno-Ugric languages, Berber, Paleo-Siberian or the Na-Dene languages of North America (Igartua and Zabaltza 2016; Hualde, Lakarra, and Trask 1996; Trask 1997). Thus, there is no evidence to confirm the origin of the Euskara, or a link to any other language in the history beyond to that with the ancient form of Aquitaine. Despite this, the Euskara has not been hermetic along history and it is well known the influence from its surrounding languages. Evidences have been found for Celtiberian, Latin and its romance languages, such as Gascon, Navarro-Aragonese, French or Spanish. Arabic influence is also observed in present Euskara, which was probably introduced by romance languages. One of the most accepted explanations to the survival of the Euskara from those historical external influences is the good contributions with the Romans, together with the higher linguistic distances from the Latin and the support of the Church. This helped Basques to raise a good organization and economy, allowing them to resist the following pressures of Goths, Franks and Arabs. Although it was barely written until the end of the Middle Ages it was still being spoken in the Basque area.

Regarding the dialects, there are two main streams of opinions about its origin. The first one suggests a relationship “tribes-dioceses-dialects”. Based on this thought, the dialects would be very old; pre-dating the arrival of the Romans, and their origin and distribution would be related to the demarcation of the pre-Roman tribes and the medieval dioceses, which would have been quite similar (Mujika Zufiría 1914). The second one, which is the most accepted, propose an origin around the Middle Ages (Mitxelena Elissalt 1981; Igartua and Zabaltza 2016). This idea relies on five points: i) the high unity of the Euskara do not reflect an old fragmentation, ii) the abundant common post-Roman linguistic innovations in all the dialects show that they have evolved similarly in recent times, iii) the large similarities among the central dialects evince more recent times of divergence, iv) the number of innovations are quite small, and v) the limits of the tribes and dioceses do not match with the delimitation of the dialects (Zuazo 2010). Thus, the cause of the emergence of the dialects during the medieval period could have been the different processes of administrative reorganization together with the influences of the closest surrounding languages.

1.5. Population Genetics: a powerful tool for studying diversity and history

“[h]istory can get contaminated over time. But DNA does not lie”
Rene J. Herrera (2008) (Wolinsky 2008)

The principal aim of human population genetics has been the study of the genetic variation across populations in order to depict its frequency, distribution, origin and causes, to understand the connections between them, trying to reconstruct their history. Although collaterally it has been always playing a crucial role into the medicine field by detecting new variants, studying different demographic scenarios and its genetic consequences or more recently defining the specific population genetic panels that could help in disease screening programs and genetic counseling (Templeton 2019; Burke et al. 2001; Wolinsky 2008; Bodmer 2015).

Herrera’s words, although fairly extreme, reflect somehow the importance and power of population genetics in historical researches. Genetics can supply data and new interpretations to elucidate unsolved unknown and controversies, such as the case of the Basques. Furthermore, just as cultural and historical studies have more room for speech and overinterpretation, the genetic evidences separates from ideological, essentialist, deterministic and reductionist conclusions by considering several possible representation and establishing probabilistic relationships between them (de Chadarevian 2010). However, an interdisciplinary approach is required. Historical, linguistic, anthropological and archaeological data are needed to contextualize, to test hypotheses and to help the interpretation of the results (Samida and Feuchter 2016; Wolinsky 2008; Cavalli-Sforza, Menozzi, and Piazza 1994).

1.5.1. Methods in human population genetics: a brief review

From classical markers as the basic tools in human population genetics during the early and middle 20th century, the methodology and data have experienced an astonishing development opening new chances and challenges for studying human populations (Bodmer 2015). The evolution of genome-wide data, with available samples from worldwide geographic areas, together with the advances in the statistical methodology to infer demographic events, has marked the beginning of a new era in population genetics, making possible to include thousands of markers and samples instead of the reduced and limited resources in the last century (Stoneking and Krause 2011; Schraiber and Akey 2015; Wangkumhang and Hellenthal 2018).

The classical genetic markers used to study the human genetic diversity were actually protein particles detected through indirect methods, based on immunological reactions (Stone et al. 1993; Bodmer 2015; Jobling et al. 2013). They were discovered and widely used during the 20th century, outstanding the red blood groups systems, that were the first defined human polymorphisms (Hirschfeld and Hirschfeld 1919; Fisher and Taylor 1940; Mourant, Kopeć, and Domaniewska-Sobczak 1976). Among the different blood group systems described after the ABO by Karl Landsteiner in 1900 (Owen 2000), the Rh system (Levine, Newark, and Stetson 1939) drew attention by its biomedical implications due to its importance in hemolytic reactions from transfusion incompatibilities, and its connection to the hemolytic disease of the newborn (HDN), when the immunological system of a mother lacking the Rh factor (d; Rh-negative), reacts to the presence of the factor in her child (D, Rh-positive) (Fisher 1947). The emerging studies enabled constructing the first human evolutionary tree using polymorphism frequency of just five blood group systems (Edwards and Cavalli-Sforza 1965; Cavalli-Sforza and Edwards 1963). Later, the methods evolved to electrophoretic analyses of gene product and the detection of DNA variation through Southern blotting analyses, specially focused on the detection of restriction fragment length polymorphisms (RFLPs) (Jobling et al. 2013; Solomon and Bodmer 1979).

The first big impulse of human genetic diversity was the development of the polymerase chain reaction (PCR) (Sanger, Nicklen, and Coulson 1977). This method led to the introduction of the DNA marker studies and allowed the improvement of approaches in the field during the next decades (Bodmer 2015). The problematic of the possible selection that could have been affecting the gene products of the classical markers, and therefore the results of the analyses, had been overcome. The first systematic application was the analysis of the variability of short tandem repeats (STRs) in the populations (Weber and May 1989).

The publication of “The History and Geography of Human Genetics” (Cavalli-Sforza, Menozzi, and Piazza 1994) became the reference par excellence in the field, compiling and analyzing the human genetic variation accumulated over the last fifty years. This prompted to the development of essential methods in the present, such as the implementation of Principal Components Analysis (PCA) in population genetic variation.

The achievement of the Human Genome Project (Lander et al. 2001; Venter et al. 2001) prompted the emergence of several projects analyzing the human genetic variation in a genome-wide context. It was unraveled that the actual genetic variation when studying differences among humans relies on just a 0.1% of the genome. It was also defined the concept of single nucleotide polymorphism (SNP) when more than 1% of a population presents the same variant of a position with two distinct alleles. SNPs are less variable than STRs, but they are more stable due to a lower mutation rate. Their large number in the genome made that the genetic information per individual was no longer limiting. Thus, the analysis of this variation become the essence of population genetics (Gardiner 2002; LaFramboise 2009).

After the development of Sanger sequencing (Sanger, Nicklen, and Coulson 1977), high-throughput genome wide SNP typing, together with next- and third-generation sequencing methods were developed. The reduced cost and time-processing of these methods massively increased the production of high-throughput data. All these advances made possible the freely accessibility to such an amount of diverse genomic data that could be used as reference panels, through several projects that emerged in the last decade: Human Genome Diversity Project (HGDP) (H. M. Cann et al. 2002), HapMap Project (The International HapMap Consortium 2003, 2005) and the 1000 Genomes (Laska et al. 2009). This came together with the development of more powerful computational methods to manage the large amount of genomic data produced. The populations were not studied based on just some specific genetic markers anymore, but they were helpful as starting point to carry analyses along the genome and in a wide population context (Li et al. 2008; Novembre et al. 2008; Lao et al. 2008; Bryc et al. 2010).

The revolution of ancient DNA (aDNA) with the publication of the first three ancient hominid nuclear genome sequences in 2010, meant a transformation in human population genetics (Rasmussen et al. 2010; Green et al. 2010; Reich et al. 2010). The following ancient genomes that have been available in the last years allow to describe the different demographic processes in the past (Lazaridis et al. 2014, 2016; Haak et al. 2015; Günther et al. 2015; Olalde et al. 2019; Mittnik et al. 2018; Rasmussen et al. 2010; Posth et al. 2018; Shinde et al. 2019; Schaefer and Shapiro 2019). The different main ancient populations and cultures in the world are being genetically described and, comparing them with present-day samples, allow to deeper describe and reconstruct the history of the modern populations. However, the aDNA presents several drawbacks. The damage of the DNA molecules due to conservation during long periods, together with the contamination of the samples because of the direct contact with the environment, and the length of the molecules are some of the limiting factors that are important in aDNA analyses (Stoneking and Krause 2011; Pickrell and Reich 2014; Schaefer, Shapiro, and Green 2016).

The huge impact of the methodological advances during the last decades has been reflected even in the basic analysis of F_{ST}. Its first definitions (Wright 1951) have been reconsidered in order to asses more accurate and powerful results (Nei 1973; Weir and Cockerham 1984; Slatkin 1991; Hudson, Slatkin, and Maddison 1992; Weir and Hill 2002; Holsinger and Weir 2009), with new computational methods that enable working with the new large amount of data, implementing internal jackknife permutations to obtain confidence intervals and p-values of the pairwise comparisons, and allowing to better rely on the results (Pembleton, Cogan, and Forster 2013).

Clustering sequence analyses experienced their major leap with the development of network-based methods (Bandelt, Forster, and Rohl 1999), overcoming the limitations of trees where branches diverge but never coalesce (Templeton 2019). Networks allow to study the intraspecific genetic diversity of sequences, plus the evolutionary relationships in a more accurate and robust way, enabling extensive interconnections between individuals without assuming or excluding population trees (Templeton 2019; Mardulyn 2012).

When studying demographic history of populations with large datasets, PCA is useful to synthesize in a smaller dimension the genetic variability from the multiple independent markers to describe population structure, drift and admixture populations (Cavalli-Sforza, Menozzi, and Piazza 1994; Schraiber and Akey 2015). Thus, when samples are plotted according to their loading on the PCs, they will sparse according to their genetic closeness. This is a fundamental first step to understand the general genetic scenario of our data (e.g. if our target population separates from the other populations included in the analysis, it would be evidencing a high genetic differentiation and a possible isolation). Nevertheless, to support the information suggested in a PCA, further methods have been developed to specifically probe for those evidences (Novembre and Stephens 2008; McVean 2009). Global ancestry methods, such as ADMIXTURE, STRUCTURE and FRAPPE are based on model-based clustering approaches (Alexander, Novembre, and Lange 2009; Pritchard, Stephens, and Donnelly 2000; Tang et al. 2005). They complement the PCA looking for groups of individuals that share common underlying allele frequencies. However, in order to formally test for admixture between populations, specific methods are required. *F-statistics* encompasses a series of statistical methods that allow testing for the presence of admixture or not, evaluating phylogenetic relationships based on the sharing of allele frequencies between populations along tree branches (Patterson et al. 2012; Wangkumhang and Hellenthal 2018; Peter 2016). Moreover, recent explicit spatial statistical approaches, such as SpaceMix (Bradburd, Ralph, and Coop 2016) and EEMS (Petkova, Novembre, and Stephens 2016), have emerged to test for isolation-by-distance and infer migration rates, giving a visual spatial context of corridors and barriers to complement the genetic diversity patterns observed in the basic PCA.

These methods do not consider linkage disequilibrium (LD) and can led to a remarkably loss of genetic information that could be retained and considered for more robust and specific analyses. LD between markers can be highly informative by considering haplotype information, rather than pair of independent SNPs (Wangkumhang and Hellenthal 2018; Gattepaille and Jakobsson 2012). Thus, several methods that consider the individual haplotype information, retrieved through LD and phasing, have been developed in the last decade to analyse admixture patterns in the populations. These are ChromoPainter, fineSTRUCTURE and GLOBETROTTER (Lawson et al. 2012; Hellenthal et al. 2014). These haplotype-based methods enable analyzing the individuals in a more robust and finer structure resolution, classifying them in genetically homogeneous groups beyond the geographic labels, and allowing to infer the individual shared ancestries to describe demographic processes.

ChromoPainter (Lawson et al. 2012) depicts individually the chromosomes of each recipient sample, without population specification, as a combination of chunks in LD (chromosome haplotypes) from all the other samples, treated as donor. It uses a Hidden Markov Model (HMM) in order to assemble this information in a coancestry matrix, that reflects the total length and count of the shared haplotypes among individuals, estimated from the chunks of the recipient genome that coalesce with one of the two haplotype lineages from a specific donor before that of any other individual. ChromoPainter can also be run by grouping individuals in clusters or populations that can be defined as recipient, donor or both. In this case, the coancestry matrix will reflect just the information of the individuals included in a recipient population. FineSTRUCTURE (Lawson et al. 2012) uses the coancestry matrix from ChromoPainter to perform a clustering analysis, through a Monte Carlo Markov Model Chain (MCMC), in order to assign individuals in genetically homogeneous clusters or populations that a priori were not known. GLOBETROTTER (Hellenthal et al. 2014) uses the coancestry matrix from

ChromoPainter, together with the clusters defined in fineSTRUCTURE, in order to infer admixture patterns and dates. It is used to first model each recipient cluster or population as a mixture of the haplotypes from the donors, in order to obtain the ancestral proportion of each recipient based on the donors. This is performed through a non-negative-least-squares regression (NNLS). Then, GLOBETROTTER can be used to infer specific admixture events and their respective dates, based on the estimation of the decay on the coancestry curves, that represent the distribution of the relative probability that a pair of chunks could be assigned to two surrogates (donors in ChromoPainter) separated by a given distance. The rate of decay is related to the time since the admixture event, being that recombination through generation will reduce the lengths of the admixed chunks. The strong advantage of this method is that the original sources that have been involved in the admixture events, and are not sampled because they do not longer exist, can be reconstructed from the surrogates to explain the target (recipient).

1.6. The genetic landscape of the Basques

“Despite the numerous methodological and technological advances that we have witnessed over the last few years, the demographic history of the Basque people is still far from being completely elucidated”

(López and Alonso 2013)

These words, from just six years ago, put in the context the unsolved situation and complicated scenario related to the genetic panorama of Basque people, and the lively controversy that has arisen during the last decades related to their genetic differentiation, heterogeneity and origins.

1.6.1. Classical markers: simple, yet interesting

The genetic interest in Basques emerged when Etcheverry observed, for the first time in 1945, a remarkable high frequency of the Rh-negative blood group (Etcheverry 1945). This high frequency impressed the scholars. Due to the association between the Rh-negative variant and the HDN, as mentioned above, a high frequency of this variant was not expected to be found in a population. Thus, many subsequent studies focused in the analysis of blood group systems arose. They confirmed Basques as the population with the highest frequency of Rh-negative in the world and clearly differentiated from the surrounding populations (Mourant 1947, 1954; Van Der Heide, Magnée, and Van Loghem 1951; Chalmers, Ikin, and Mourant 1948, 1949) (Figure 6).

New other studies including more genetic markers appeared, showing the same pattern of differentiation (Calafell and Bertranpetit 1994b; Bertranpetit and Cavalli-Sforza 1991; Calafell and Bertranpetit 1994a) (Figure 6). These results, together with the cultural, archaeological and linguistic interest of Basques led to the assumption that they would have locally descended, from ancient Rh-negative pre-Neolithic Europeans that remained isolated during the LGM, and that during the following periods avoided the geneflow with the upcoming population migrations (Van Der Heide, Magnée, and Van Loghem 1951; Bertranpetit and Cavalli-Sforza 1991; Mourant, Kopeć, and Domaniewska-Sobczak 1976; Calafell and Bertranpetit 1994a). The controversies of a pre- or post-Palaeolithic origin and the existence of an internal heterogeneity started to arise in further studies by analyzing other polymorphic genetic markers (Aguirre et al. 1991; Calafell and Bertranpetit 1994a; Calderón et al. 1998a; Manzano et al. 2002; Manzano, Orue, and de la Rúa 1996).

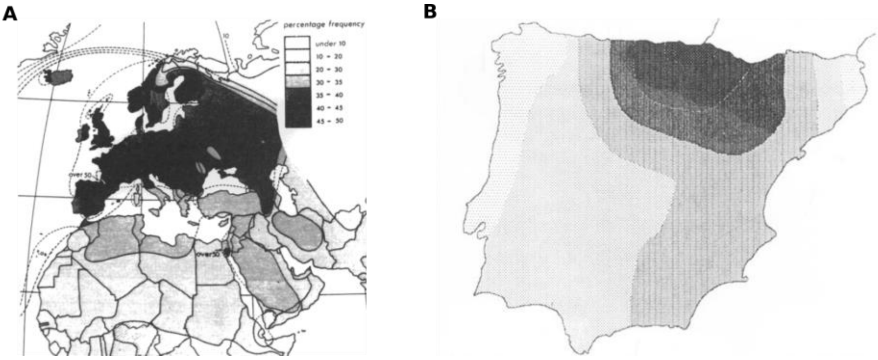


Figure 6. Synthetic maps from previous studies. (A) Distribution of the Rh-negative negative from the analysis by Mourant. He highlighted the Basques with a frequency over 50%.. Source: (Mourant, Kopeć, and Domaniewska-Sobczak 1976). (B) Map of the Iberian Peninsula with the result of a principal compoent analysis using the frequencies from 54 alleles of 2o human loci. Source: (Bertranpetit and Cavalli-Sforza 1991).

Moreover, evidences of genetic connection to unexpected populations were reported. Some studies proposed a genetic closeness with North African and Caucasian populations (Arnaiz-Villena et al. 1997; Piazza et al. 1988; Arnaiz-Villena et al. 1995;

Calderón et al. 1998b). This suggestions were mainly prompted by the widely extended hypotheses of connections between the Euskara and the North African and Caucasian languages those years (Hualde, Lakarra, and Trask 1996; Igartua and Zabaltza 2016).

Nonetheless, caution should be taken when analyzing functional markers since similar selective pressures in different populations might provide similar results, masking the population history of the populations under study. This fact, plus the limited number of classical markers analyzed in each study, challenges the reliability of the demographic conclusions provided by classical genetic markers.

1.6.2. DNA markers: a parental quest

The analyses with DNA markers have not been much prominent in Basques as the ones preformed with classical markers. Some works showed a pattern of homogeneity through Europe (Pérez-Lezaun et al. 1996). Otherwise, other analyses came out showing a clear genetic differentiation of Basques from the European and North African populations, together with an internal substructure within the region (Pérez-Miranda et al. 2005; Iriondo, Barbero, and Manzano 2003; Bauduer, Feingold, and Lacombe 2005; Young et al. 2011). An internal heterogeneity within Basques was evidenced, suggesting three main western, central and eastern geographic clusters, instead a differentiation affected by the Pyrenees (Iriondo, Barbero, and Manzano 2003). However, other studies did not support this internal differentiation and reported a pattern of homogeneity (Comas et al. 1998; Young et al. 2011). The analysis of the minisatellite locus MS205 drew attention by suggesting some specific lineages to the Basque population pre-dating Neolithic times, reflecting an ancient local continuity (Alonso and Armour 1998). Furthermore, the previously suggested relationship with North African populations was not generally supported by DNA markers (Comas et al. 1998; Zlojutro et al. 2006; Perez-Miranda et al. 2004; Pérez-Miranda et al. 2005; Young et al. 2011).

The studies based on DNA markers stood out with the analyses of uniparental markers, trying to reconstruct the history of the population by describing their paternal and maternal linages. In general, all major mitochondrial DNA (mtDNA) haplogroups in Basques are shared with the rest of Europe (Behar et al. 2012; Bertranpetit et al. 1995; Chikhi et al. 1998; Comas et al. 1997; Sajantila et al. 1995), but some peculiarities have been reported.

It has been suggested a pre-Neolithic origin in Basques, together with an expansion and resettlement of Europe after the LGM from the Basque area. The mtDNA haplogroup V is considered as representative of the European Palaeolithic hunter-gatherers, and it was reported the highest frequencies and diversities in the Basque region and the Saami. This suggested a population expansion along Europe after the LGM, from the Franco-Cantabrian glacial refuge, through a southwest-northeast cline (Torroni et al. 1998; Barral-Arca et al. 2016; Cardoso et al. 2013; Soares et al. 2010; Pala et al.

2012). The same proposal of an expansion was reported by the highest frequency in Basques of the most common European mitochondrial haplogroup H (specially H1 and H3) (See Figure 7), which has been related to the expansion of the Near Eastern Neolithic farmers (Richards et al. 1996; Côrte-Real et al. 1996; Barral-Arca et al. 2016; Achilli et al. 2004; Pereira et al. 2005; Soares et al. 2010; Szécsényi-Nagy et al. 2017; Pala et al. 2012). Besides the post-LGM expansions, a genetic continuity since pre-Neolithic times was observed in Basques (Behar et al. 2012; Martínez-Cruz et al. 2012; Cardoso et al. 2013). Moreover, is has been also suggested a possible connection between Berbers, Basques and Saami by the identification of the mtDNA haplogroup U5b1b, pointing to a second hunter-gatherer migration to the south through the Strait of Gibraltar (Achilli et al. 2005; Soares et al. 2010; Pala et al. 2012). It is also interesting the presence of the ancestral and specific phylogeny of the haplogroup U8a in Basques (González et al. 2006). Conversely, other studies have refused the idea of a pre-Neolithic Basque continuity and the post-LGM resettlement of Europe from the region. By contrast, other studies have detected low frequencies and polymorphisms of the European pre-Neolithic mtDNA haplogroups V and H along the region, refusing the hypothesis of a human refuge-expansion after the LGM (García et al. 2011; Alfonso-Sánchez et al. 2008). Regarding the internal genetic landscape, analysis of mtDNA evidenced an heterogeneity within Basques related to geography (Martínez-Cruz et al. 2012), following the previous result from Iriondo and his colleagues (Iriondo, Barbero, and Manzano 2003). In addition, a gene flow from the surrounding Franco-Cantabrian groups was observed, explained by the ancient extension of the Euskara beyond the present-day Basque area (Behar et al. 2012; Martínez-Cruz et al. 2012).

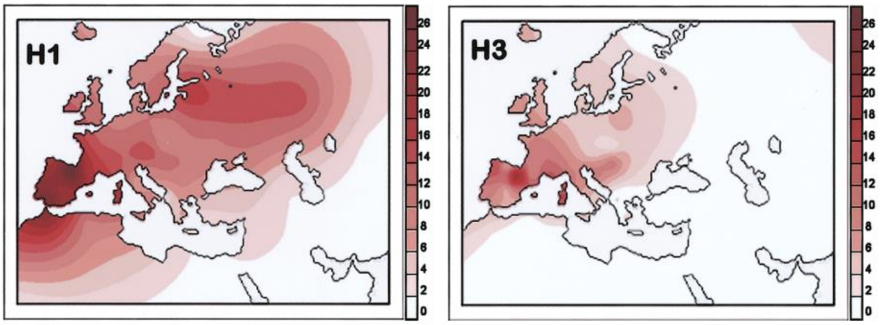


Figure 7. Spatial distribution of mtDNA H1 and H3 haplgroups. The higher frequencies are observed in the Franco-Cantabrian region. Source: (Achilli et al. 2004).

Concerning the Y-chromosome (Ychr), a reduced genetic structure has been reported though the Iberian Peninsula mainly represented, Basque inclusive with low diversities (Flores et al. 2004; Bosch et al. 2001; Alonso et al. 2005; Martínez-Cruz et al. 2012). Otherwise, an old-Pre-Neolithic substrate in the Pyrenean populations was shown by other studies (Lopez-Parra et al. 2009; Hernández et al. 2017). The distribution of some Ychr haplogroups, as R1a and R1b, together with the mtDNA haplogroup V, also seemed to support the post-LGM expansion (Semino et al. 2000; López and Alonso 2013; Soares et al. 2010). Moreover, the sublinage R1b-DF27, with a possible origin in the Bronze Age, is present in Basques at frequencies of ~70%, in contrast to the ~50% in the rest of Spain and the <20% in Europeans (Figure 8), evidencing the influence of the Bronze Age/Iron Age steppe ancestry even in the Basque area, contrary to the previous evidences of Pre-Neolithic or Neolithic continuity (Solé-Morata et al. 2017; Villaescusa et al. 2017; Valverde et al. 2016).

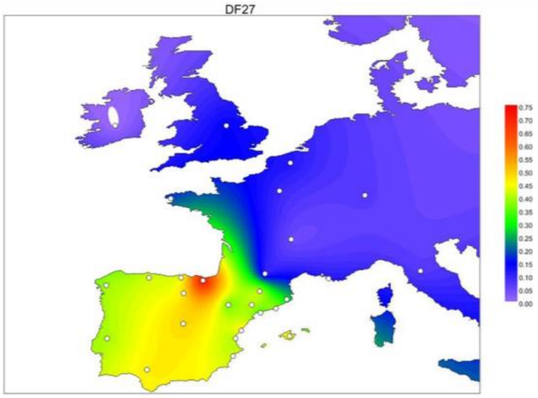


Figure 8. Spatial distribution of the Ychr R1b-Df27 haplogroup. Basques present the higher frequency of the haplogroup. Source: (Solé-Morata et al. 2017).

1.6.3. Genome-wide studies: two sides of the same coin

The genome-wide data studies focused on Basques are very scant. Li et al. performed an analysis of 650,000 SNPs in a worldwide context, observing the differentiation of Basques as a whole from the other populations (Li et al. 2008). On the contrary, another study showed a homogeneity pattern analyzing just 123 SNPs in a 1-Mb chromosome 22 (Garagnani et al. 2009).

However, a very clear case that reflects the controversy around the genetic landscape of Basques is the publication of two papers, the same year in the *Human Genetics* journal, showing completely opposite results (Laayouni, Calafell, and Bertranpetit 2010; Rodríguez-Ezpeleta et al. 2010).

First, Laayouni et al. performed the analysis of ~300,000 SNPs genotyped by pools of DNA from 30 individuals for each of 10 Spanish populations, including Basques. A clear homogeneity of Spanish Basque with the other Iberian groups was detected. However, they clearly separated from the French Basques and the other European samples (Figure 9).

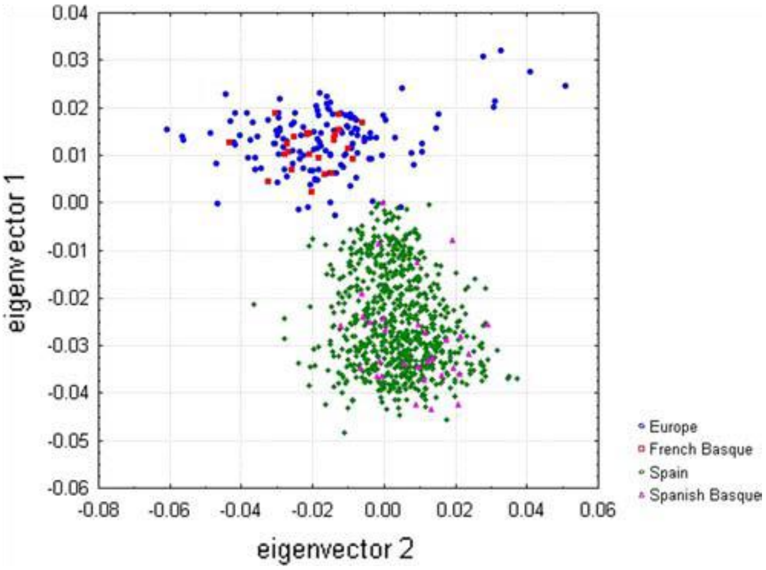


Figure 9. PCA performed with 109 highly informative SNPs from genotyping data, including Basque, Spanish French and other European samples. Image from: (Laayouni, Calafell, and Bertranpetit 2010). Basque Spanish fall with Spanish samples, saparted from the Basque French that fall within Europe.

In contrast, Rodríguez-Ezpeleta, et al. analysed ~60,000 SNPs from 83 Spanish Basques individually genotyped. In this case, a homogeneous scenario was observed within the Basques, together with a genetic differentiation from the other external population (Figure 10).

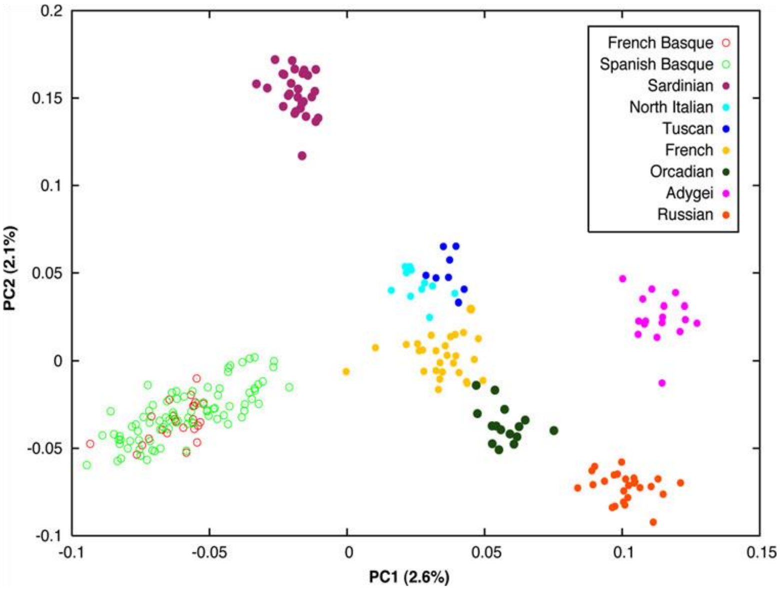


Figure 10. PCA performed with 61,504 SNPs including Basques in an European context. Source: (Rodríguez-Ezpeleta et al. 2010). Basque Spanish and French appear sepparated from the other samples forming a cluster.

The differences found in both studies might be explained by the different number of SNPs analyzed, but mainly by the different techniques used (pooling individuals versus individual data), which point to differences due to technical issues.

1.6.4. aDNA analyses: connecting with the ancestors

The analysis of mtDNA from ancient samples allowed to directly clarify the presence, or not, of the previously dated haplogroups in ancient samples from a specific period. A study analyzing ancient samples along the Franco-Cantabrian region, dating between the Neolithic and the Bronze Age period, suggested that, contrary to the analyses with modern samples (Torroni et al. 1998; Behar et al. 2012), the haplogroup V in ancient Basques samples were actually in small frequencies, not supporting the idea of a pre-Neolithic continuity (Izagirre and de la Rúa 1999). However, the latest studies with ancient mtDNA have supported the continuity, by detecting less impact in the Iberian Peninsula, and especially in the Franco-Cantabrian region, the Neolithic mitochondrial linages compared to Central Europe (Palencia-Madrid et al. 2017), in the line of the present-day analyses (Behar et al. 2012). However, the Franco-Cantabrian refuge during the LGM and the posterior expansion have not been supported by the evidences from ancient mtDNA (Szécsényi-Nagy et al. 2017). Evidences of Neolithic influences in the region, although low, were proposed by the contradictory observation of the representation of the mtDNA haplogroup J (related to the expansion of Neolithic farmers from Near East) in modern and ancient samples from the Basque area. Moreover, the analyses with ancient mtDNA showed again a possible connection between the Basques and Northwest Africa, before the arrival of the Arabs to the Iberian Peninsula (Alzualde et al. 2005; Soares et al. 2010). In addition, considering several modern and ancient samples from the Palaeolithic to the Bronze Age suggested a genetic pattern of heterogeneity among the haplogroups of mtDNA along the Franco-Cantabrian region originated during the Palaeolithic (Hervella et al. 2012), that followed by different drift effect along the region could explain the differences observed in the ancient mtDNA analyses depending on the samples studied (López and Alonso 2013).

Analyses with genome-wide sequence from ancient data helped to shed light on the ancient scenario in the region. In 2015, an analysis was performed including the first genome-wide data from Chalcolithic and Bronze Age human remains in the Iberian Peninsula (Günther et al. 2015) (Figure 11). This study supported a remarkable influence of Neolithic migrations in the Basque area and a subsequent isolation, refusing the idea of a possible Pre-Neolithic isolation as widely shown before.

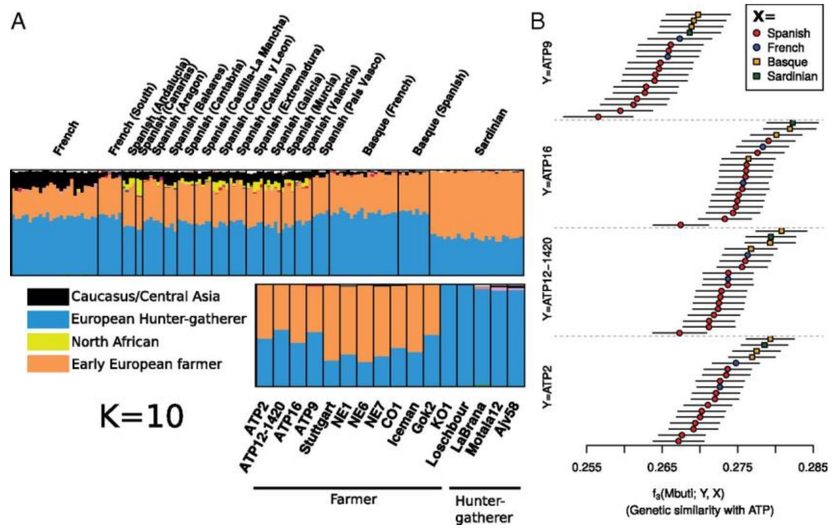


Figure 11. Observed connection between ancient farmers from Atapuerca and present-day Basques. (A) ADMIXTURE analysis including modern and ancient samples. Basques, Sardinians and farmers present similar genetic components. (B) Genetic closeness observed with outgroup f_3-stadistics. Basques and Sardinians present the highest shared drift with farmers. Source: (Günther et al. 2015)

1.6.5. Latest genetic results

The latest results that gave information about Basques, do not specifically focus on them, but in the Iberian Peninsula. They have started to support, with more accurate methods, the idea of the genetic differentiation of Basques in a general context. Bycroft et al., applying haplotype-based methods with modern samples, observed Basques clearly differentiated from the surrounding Iberian groups (Bycroft et al. 2019) (Figure 12).

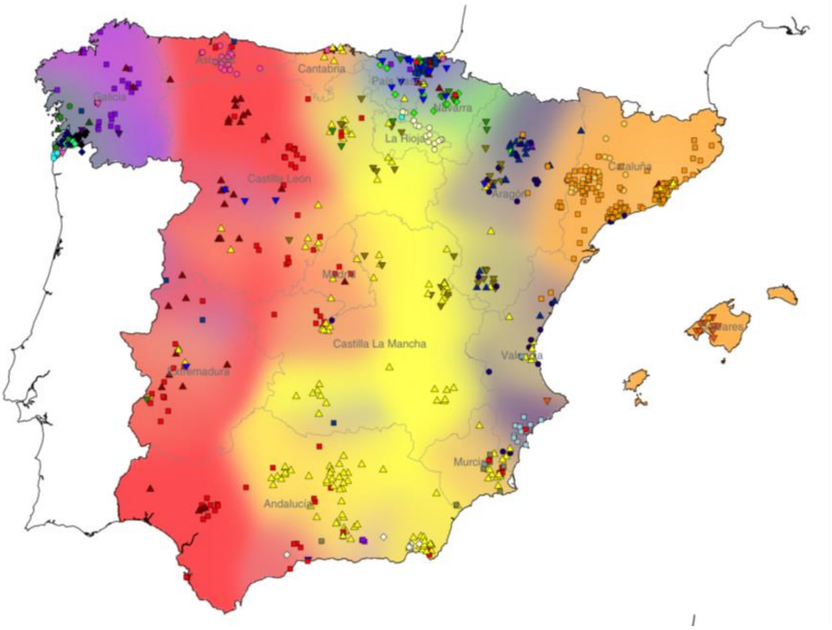


Figure 12. Spatial density distribution for the clusters defined in the fineSTRUCTURE tree. Basques appear represented by a specific cluster differentiated from the surrounding groups. Source: (Bycroft et al. 2019).

In addition, a recent paper analysing 271 ancient Iberian samples covering the periods from Mesolithic to 2000 CE suggested, based on PCA projection and qpAdm modeling, that Basques could be better described as a typical Iberian Iron Age population, reflecting the genetic influence of the steppe ancestry also in Basques together (in fact, as mentioned above thy present the highest frequencies of R1b Ychr haplogroup), without the evidences of the posterior admixture events that were observed in the rest of the Iberian Peninsula (Olalde et al. 2019).

As evident while reading this chapter, in spite of all these analyses mentioned along, there has been a lively controversy around Basques due to the opposite and numerous results that have been emerging during the last decades. These main points about Basque people could be summarized into four: i) the high frequency of Rh-negative, ii) their distinctiveness from the external European genetic context, iii) their internal heterogeneity, and iv) their most plausible origin.

Those previous genetic works have been mainly affected by a weak study design, without a correct ethnolinguistic, geographic and genetic representation of Basques. They have just included some samples from the Basque area and considering them as a single population. Many studies have performed the analyses just including a small number of genetic markers (Rodríguez-Ezpeleta et al. 2010), performing genotyping by pooling individuals (Laayouni, Calafell, and Bertranpetit 2010) or using genome-wide data from biomedical arrays that could lead to an ascertainment bias in the analysis (Bycroft et al. 2019).

2. OBJECTIVES

The aim of the thesis is to shade light on the four main peculiar points of Basques, by using more powerful and refined methodology to overcome previous limitations.

In the first project, the classical question about Basques, related to the highest frequency of the Rh-negative variant is studied, but improving the methodology through recent technical approaches by using high throughput sequencing instead of serology. Besides looking for the frequencies of the different variants of the Rh system, we explore the general framework of the genetic scenario of the Rh system in Basques.

In the second project, the other three points about Basques are assessed overcoming the previous limitations. An exhaustive representation of the Basque people in an ethnolinguistic, geographic and historical context is considered. Furthermore, a specific array designed for population genetics is used in order to avoid the ascertainment bias that could affect the analyses. Moreover, finer and more robust methodology is applied, combining allele and haplotype-based methods in a wide and fine scale context.

3. RESULTS

3.1. Sequence diversity of the Rh blood group system in Basques

André Flores-Bello, David Mas-Ponte, Miruna E. Rosu, Elena Bosch, Francesc Calafell, David Comas

European Journal of Human Genetics. 2018; 26, 1859–1866.

DOI: 10.1038/s41431-018-0232-1

Flores-Bello A, Mas-Ponte D, Rosu ME, Bosch E, Calafell F, Comas D. Sequence diversity of the Rh blood group system in Basques. Eur J Hum Genet. 2018 Dec 1;26(12):1859–66. DOI: 10.1038/s41431-018-0232-1

3.2. The genomic history of the Franco-Cantabrian region: deciphering the genetics of the Basques

(in preparation)

André Flores-Bello[1], Frédéric Bauduer[2], Jasone Salaberria[3], Bernard Oyharçabal[3], Lluis Quintana-Murci[4], David Comas[1]

[1]Departament de Ciències de la Salut i de la Vida, Institut de Biologia Evolutiva (CSIC-UPF), Universitat Pompeu Fabra, Barcelona, Spain

[2]Laboratoire Maladies Rares: Génétique et Métabolisme (MRGM), EA 4576, University of Bordeaux I, Bordeaux, France

[3]Centre de Recherche sur la Langue et les Textes Basques, Centre National de la Recherche Scientifique, UMR5478 IKER, Bayonne, France

[4]Human Evolutionary Genetics Unit, Centre National de la Recherche Scientifique URA31012, Institut Pasteur, Paris, France

Abstract

Basques have been historically distributed along the Western Pyrenees, in the Franco-Cantabrian region, between the current Spanish and the French territories. They have been in the limelight of many researches due to the several singular cultural and biological traits that, with high controversy, seemed to place them as a heterogeneous, isolated and unique population within Europe, highlighting their own non-Indo-European language, Euskara, and its five main dialects. Due to the limitations of previous studies about Basque genetic structure, there is still a lively debate about their history and their presumed singularity. To overcome the previous limitations, we have analyzed, at a micro-geographical level, genome-wide data of all Basque and surrounding groups that do not speak Euskara. Overall, a total of ~200 individuals from 18 different geographical areas were included. We have also considered other external modern and ancient samples from publicly available datasets in order to have external references. Our results, mainly focused on haplotype-based methods, show a clear differentiation of Basques from other European populations, as previously suggested. Even at a local scale, we detect a genetic differentiation of Basques compared to surrounding Spanish and French groups, with a genetic cline through the non-Basque Franco-Cantabrians. Moreover, this internal genetic heterogeneity shows a significant correlation with geography. Finally, our analyses including ancient samples show that the ancestry of Basques is mostly defined by a genetic continuity since the Bronze Age period,

which challenges the previous ideas of the Palaeolithic/Neolithic composition of Basques.

Introduction

The Franco-Cantabrian region has drawn big attention in multiple disciplines due to several factors that depict its important big role in the European human history. This region encompassed the Northern border of Spain and the Southern France through the Pyrenees. It is known that this region was one of the most densely populated glacial refugia in Europe, during the Last Glacial Maximum in the Palaeolithic [1] and it is related to very important archaeological discoveries, especially the first human cave paintings [1]. However, the most interesting peculiarity is the presence of the Basque people. They have been historically distributed along the western edge of the Pyrenees between the Spanish and the French territories currently organized in seven provinces: Gipuzkoa, Bizkaia and Araba in the southern side of the Pyrenees; and Zuberoa, Lapurdi and Nafarroa Beherea located in the northern side. Basques have stood out due to their cultural and biological traits that suggested their singularity and isolation. Their main cultural trait is the Euskara and its dialects, which is the unique non-Indo-European language in Western Europe, that has been pointed as a possible cultural barrier along history. Beyond the current distribution of the Euskara, it was historically present in a larger area where it is supposed to have coexisted with other languages. This would include the neighboring Northern Spanish areas and Southern half of Aquitaine in France [2]. Moreover, Basques have been widely studied during the last decades, trying to depict their distinctiveness from a biological point of view with the analysis of classical markers [3], microsatellite data [4], uniparental markers [5] and genome-wide data [6]. However, in spite of the numerous studies, a clear conclusion about the genetic scenario of the Basques has not been assessed yet, and a strong controversy is still alive behind it. Some studies suggested a clear differentiation of Basques from the surrounding populations, as well as an internal heterogeneity within, whereas other studies resulted in the opposite conclusions [6], [7]. The same situation is observed in relation to their origin, from previous studies pointing them as Pre-Neolithic remnants[8], [9] to the latest results that define a genetic continuity from earlier periods [10]–[12]. These remarkable contradictory results are related to the weak study design in these studies. The poor number of samples used in the analyses to represent the Basque groups and the neighboring areas in the region has supposed the major limiting factor. Furthermore, the methodology previously used was quite limited, mostly based on blood type frequencies or a reduced number of genetic markers [13]. In this study, we shed light on the distinctiveness, heterogeneity and origin of Basques people, overcoming the limitations burdened in the previous studies and giving more powerful conclusions. To that, we present here a high-resolution representation of the whole Franco-Cantabrian area to allow fine-scale results, by using genome-wide data genotyped with the Axiom™ Human Origins Array, that is optimized for human genetic studies [14]. Besides the Euskara-speaking area, Spanish and French surrounding groups that do not speak this language were also included. Furthermore, these newly reported data was merged with publicly available external modern and ancient samples to contextualize the region in a large and genetically diverse scenario. Finally, beside the allele frequency methods, haplotype-based approaches were performed. These methods analyse the samples in a more robust and refined way, beyond the geographic labels, and allow us to infer the shared ancestries among the analysed groups [15], [16]. For the first time, fine-scale and wide-scale analyses from genome-wide data have been performed covering the whole Franco-Cantabrian region, combining allele frequency and haplotype-based methods.

Materials and methods

Sample collection and genotyping

A total of 190 unrelated individual samples were collected from 18 microgeographical areas in the present-day Spanish and French territories of the Franco-Cantabrian region, including Basque-speaking, Spanish-speaking, and French-speaking groups (Fig.1A). Written informed consent was obtained from the participants, who were interviewed to determine their speaking dialect and to confirm their four grandparents were born in the same microgeographical area (Fig.1 and Table S1). The collection procedures were approved by the CCPPRB (Comité Consultatif de Protection des Personnes dans la Recherche Biomédicale d’Aquitaine) and the Ethics Committees at Universitat Pompeu Fabra, Institut Pasteur, Université Michel de Montaigne Bordeaux 3. The dataset includes seven areas where Euskara is spoken at present [5], [9] (Central

Gipuzkoa, Southwestern Gipuzkoa, Bizkaia, and Northwestern Nafarroa in the southern side of the Pyrenees; the admixed zone of Lapurdi/Baztan, Lapurdi Nafarroa, and Zuberoa in the northern side), besides three regions where it was spoken until the end of the 19th century [17], [18] (Central Western Nafarroa, Roncal, and Araba). For the sake of clarity, the previous ten regions where Eukara was spoken in historical times will be referred as “Basque” groups hereafter. The dataset also includes five surrounding Spanish-speaking areas where Euskara was never spoken but with evidence of influence from the Euskara up to medieval period (Cantabria, Northern Burgos, La Rioja, Northern Aragon and Western Bizkaia), and three French-speaking areas where Gascon was previously spoken (Bigorre, Béarn and Chalosse) [18]–[21], here described as Spanish and French Peri-Basques, respectively. DNA was extracted from blood samples using standard methods, and genotyped with the Axiom™ Genome-Wide Human Origins Array (~629 K variants) [14]. Genotype calling was performed by using the software Axiom^{TM} Analysis Suite 3.1.51.0 through the Affymetrix Best Practices Workflow. All samples passed the genotype calling process with an average quality control call rate of 99.8%. After exporting those variants that properly passed the recommended thresholds, 597,638 single-nucleotide polymorphisms (SNPs) remained in the dataset.

Data Quality Control and merging data

PLINK 1.9 [22] was used to apply the quality control filters after excluding uniparental markers and the X chromosome. No sample was excluded after applying a quality control filtering to remove individuals with more than 10% of missing SNPs. A total of 434,664 SNPs remained after removing SNPs missing in more than 5% of the samples, with an extreme deviation from Hardy-Weinberg equilibrium ($p < 10^{-5}$), and a minor allele-frequency (MAF) below 0.05. For those analyses that required linkage equilibrium between SNPs, linkage disequilibrium (LD) pruning was performed using a windows size of 200 SNPs, with a shifting of 25 SNPs and a maximum pairwise LD threshold (r^2) of 0.5. A total of 171,275 SNPs remained after LD pruning. Two individual samples (one from the Lapurdi/Baztan zone and one from Zuberoa) were discarded from the dataset due to their high relatedness to other samples of the same group (PI_HAT>0.125).

Our Franco-Cantabrian samples were merged with public available data genotyped with the same array: samples from Catalonia, Valencia and Balearic Islands [23]; North African [24]; plus modern and ancient Western Eurasian samples [24]. Basque samples from Lazaridis et al. [24] where included in the analyses in the Basque category as Spanish and French Basques. Samples from Eivissa were removed from most of the analyses since it has been recently reported to be an isolated population differentiated from the mainland context [23]. Thus, a final dataset containing 1,549 samples was used to perform the analyses.

Allele frequency analyses

PCA eigenvectors were computed by using the SmartPCA v13050 from EIGENSOFT v6.0.1 [25]. For the PCA including ancient samples, ancient samples were projected in the PCA.

Weir and Cockerham's pairwise F_{ST} indexes were calculated between the different groups by using the R package stAMPP [26], with confidence intervals after 1,000 bootstrap repetitions, and represented in a multidimensional scaling (MDS) plot.

In order to explore the population structure in the region, an analysis of the molecular variance (AMOVA) was performed with the R package poppr [27]; 1,000 permutations as described by Excoffier et al. [28] were carried out with ade4 R package [29], in order to obtain an empirical distribution under the null hypothesis.

To test the correlation between the genetic and geographic distances among groups, isolation by distance (IBD) analysis was performed. Geographic Distance Matrix Generator v1.2.3 software [30] was used to calculate a geographic distance matrix between the groups included in the analysis. Then, a Mantel test was performed between the genetic and geographic distance matrices with the function mantel.rtest from the ade4 R package [29].

In order to test and visualize the spatial population structure, effective migration and diversity rates were modeled in the Franco-Cantabrian region dataset using the estimated effective migration surfaces (EEMS) software [31]. We set nDemes=300, and ran four simultaneous EEMS analyses with four different random seeds. Next, we checked the best converged MCMC chain with the highest final log-posterior value, and we continued that chain with four new runs. We repeated this process three more times, taking the best chain after the previous set of runs in order to be definitely confident that the MCMC chain had converged with the optimal log-posterior value. For every set of four runs we performed 10 million iterations each, with 5 million discarded as burn-in and sampling every 50,000 iterations.

Model-based individual ancestries were defined by using ADMIXTURE v1.3.0 [32] in order to explore patterns of population structure. The unsupervised method was run for 10 independent iterations, both for the present-day and ancient samples analyses, and with K ancestral clusters ranging from 2 to 12. Then, PONG v1.4.7 [33] was used to combine and plot the result of the 10 different iterations in a consensus graph.

Outgroup f_3-statistics were performed with the qp3Pop program from AdmixTools package v4.1 [14] to calculate the shared drift between each potential ancient group and all the modern groups in the dataset. The analysis was performed in the form f_3(Mbuti; Ancient, Modern). We used qpGraph program from AdmixTools package v4.1 [14] to model the modern European samples to estimate the three major ancestral proportions in Europe as in Haak et al [34]: Western Pre-Neolithic hunter-gatherer, European

Neolithic farmer, and Yamnaya from Post-Neolithic steppe. We performed 100 replicates for each target modern population, then the average and the 95% confidence intervals of the replicates were calculated.

Haplotype-based analyses

Runs of homozygosity (ROH)

In order to test for inbreeding, ROHs were estimated with PLINK 1.9 [22], including some external populations. The ROH analysis was performed in sliding windows of 50 SNPs across the genome, allowing for 1 heterozygous call and 5 missing calls per window. The minimum hit rate of all scanning windows containing the SNP to be included in a ROH was set to 0.05. Only ROHs with a minimum length of 500 kb and containing at least 50 SNPs were included in the analysis. A minimum of 1 SNP per 50 kb was set as minimum density threshold and a maximum gap of 100 kb between two consecutive SNPs in a tract. The Canary and Balearic Islands [24] samples were removed from the dataset for this analysis.

Phasing

SHAPEIT v2 [35] was used to phase the data, with the population-averaged genetic map from HapMap phase II [36] and the available 1000 genomes data as reference panel [37]. The data was first aligned to the reference and the mismatched SNPs were removed, then the proper phase inference was performed.

ChromoPainter

Haplotype sharing between individual samples was estimated with ChromoPainter [38]. ChromoPainter depicts individually the haplotypes of each “recipient” sample, without population specification, as a haplotype combination of all the other samples, treated as “donor”. Thus, it estimates the total number and length of haplotype fragments (chunks) in the recipient’s genome that shares the most common recent ancestor with every donor sample. We first run ChromoPainter to estimate the global mutation probability and the switch rate parameters. Thus, the expected-maximization (EM) algorithm implemented in ChromoPainter was used over the chromosomes 1, 4, 17 and 20 for all individuals, with 10 iterations and parameters -in -iM. Then, the inferred values were averaged across the four chromosomes and individuals, weighting by the number of SNPs per chromosome. It resulted in 0.000661 and 222.54421 for the global mutation (M) and switch rate (n), respectively. Finally, ChromoPainter was run for all chromosomes and individuals fixing the previously estimated parameters to obtain the final count and length sharing coancestry matrices. We then used ChromoCombine to sum the matrices across chromosomes and obtain the copying profile for each individual, as well as the C parameter (C = 0.281698702) needed for running fineSTRUCTURE.

FineSTRUCTURE

FineSTRUCTURE v2.1.0 [38] was run to cluster the data obtained by ChromoPainter into genetic homogeneous groups following

Leslie et al. [15] documentation, using as input the chunkcount coancestry matrix of the total number of chunks copied among individuals. We performed 2 million MCMC iterations, with 1 million burn-in iterations and sampling values from the posterior probability every 10,000 iterations. FineSTRUCTURE tree was built with the default parameters -m T. The analysis was repeated for three different seeds to check consistency among the trees. Important differences were not observed among them, excepting the relocation of some individuals in other genetically close clusters.

GLOBETROTTER

GLOBETROTTER [39] was used following the recommended protocol to estimate the ancestry profile of all the modern samples in the dataset and test for admixture events in the Franco-Cantabrian, Spanish, and French groups. To perform this analysis, we first ran ChromoPainter, but now classifying the individuals based on the previous clustering results from fineSTRUCTURE. All of them were included in the analysis as donors and recipients.

Next, in order to describe the surrogate groups for the different clusters, we used the non-negative least squares (NNLS) method included in GLOBETROTTER to infer the haplotype sharing proportions between the different clusters. Thus, we set prop.ind parameter to 1, null.ind to 0 and num.mixing.iterations to 0. A different NNLS analysis was performed for each cluster as recipient and the others as donors, disallowing self-copying from the recipient itself.

Then, we ran GLOBETROTTER to check if some admixture events could be detected in our target clusters, using the others as surrogates. The copying vectors and the painting samples from the second run of ChromoPainter were used to run this analysis. As recommended, we have set null.ind parameter to 1 when testing for plausible admixture events. This accounts for uncommon patterns in the linkage disequilibrium decay that could show false signals of admixture. Then, null.ind parameter was set to 0 when estimating admixture proportions, dates and sources. We inferred the confidence intervals for the date estimates by performing 100 bootstrap iterations, for one-date and two-date admixture models.

Results

The differentiation of Basques in the European/Mediterranean genetic landscape

In a PCA analysis including modern samples from Western Eurasia and North Africa, the Basque samples fall in the periphery of Europe in a similar pattern as Sardinia, with a cline formed by the Peri-Basque groups towards the rest of the European samples (Fig. 2A). When the global genetic ancestry components of these populations were inferred through an ADMIXTURE analysis (similar values of lowest cross-validation error were obtained at K=6 and K=7, Fig.2B, Fig. S1) a differential genetic pattern is observed in Basques. In K=6, they present mainly two components: a major component (green), which is also a main component of the rest of European samples, and is also present at low frequencies in the Middle East/Caucasus and North Africa; plus a minor component (pink), maximized in Central/Eastern Europe, in similar proportion as in the surrounding populations. The other components found in the rest of European groups are not present in Basques (frequencies <1%). The Peri-Basque samples, present a similar pattern as Basques but with low frequencies of the external components not present in Basques. From K=7 onwards a new specific component appears, maximized in Basques and present in more than 50% in Peri-Basques. This component is also present in considerable proportions in Spanish and French samples, while is almost absent in the other external European samples.

The haplotype-based analyses performed by fineSTRUCTURE highlight the differentiation of Basques (Fig. 3, Fig. S2 and Table S2). First, the Basque groups cluster together within the large European branch, but in a differentiated external cluster (Fig. 3A, Fig. 3B -right-, Fig. S2A and Fig. S2B), which points to low haplotype sharing between this cluster and the other external European groups, and a clear internal and specific genetic profile of Basques. Secondly, the Peri-Basque groups also exhibit a differentiation from the other external populations, clustering internally with Europeans, but forming a specific branch with some exceptions, i.e. Cantabrian samples CAN clustering with Spanish samples (Fig. 3A, Fig. 3B -left-, Fig. S2A and Fig. S2B). In order to discard putative artifacts due to the overrepresentation of Franco-Cantabrian samples in the fineSTRUCTURE analysis, a random sampling of the Franco-Cantabrian region (50 samples including Basques and Peri-Basques) was performed with similar results (data not shown). Furthermore, the ancestry profile calculated in the NNLS analysis confirms a specific genetic scenario for the Basques, mirroring the results above (Fig. 3D and Fig. S3). Basques share haplotypes exclusively with the internal groups in the Franco-Cantabrian region and the Peri-Basques mainly share haplotypes internally with the groups in the region, but also with the non-Franco-Cantabrian Spanish and French groups, acting as a cline between Basques and the surrounding external populations. The intermediate ancestry profile observed in Peri-Basques suggests gene flow between the Franco-Cantabrian region and the external groups. Therefore, potential admixture scenarios were tested in Peri-Basques by using GLOBETROTTER. Single admixture events were detected for all the Peri-Basque targets, with close dates during The *Reconquista*. Similar sources were described in each case: a major source (~70%) mainly represented by Franco-Cantabrian ancestry, plus small proportions of Spanish ancestry; and a minor source (~30%) that represents an external Spanish-like population.

In order to explore the genetic differentiation of Basques, an analysis of runs of homozygosity (ROH) was performed. Basques show the highest total number (NROH) and total length (SROH) of ROHs by individual, followed by Sardinians, which have been reported as having high inbreeding [40], and Peri-Basques in an intermediate position between Basques and Sardinians and the rest of European samples (Fig. 4C, Fig. S4, Fig. S5). These results point to some degree of inbreeding of Basques compared to surrounding populations at a level even higher than the one reported in Sardinia [40].

The genetic ancestral legacy of Basques

The PCA projection of ancient samples shows Basques closer to Pre-Neolithic hunter-gatherers and Neolithic farmers, but also to some Steppe herders (Fig. S6A). The ADMIXTURE analysis (K=4, lowest cross-validation error, Fig S6B) show Basques and Peri-Basques with the lowest proportions of a Levant- and Iran-related Neolithic components, together with a slightly higher proportion of the Anatolian/European farmer component compared to other European populations. When testing for the shared drift with ancient samples, outgroup f_3-statistics show high shared drift between Basques and the three major ancient components in Europe (Fig. S7). We modeled our Franco-Cantabrian groups and other European populations with these ancient samples by using qpGraph (Fig. 5A). The model fitted for each tested European population, with Z-scores highly close to 0 for all the 100 permutations calculated (Fig. S8). The inferred mixture proportions of the three ancient components do not show differences of Basques compared to the general European context (Fig. 5B), and no internal differences regarding the proportions of the ancient components were not observed when modeling the Franco-Cantabrian groups individually (Fig. S9). Basques show percentages similar to the ones reported in Olalde et al. [12] (~11 % of WHG, ~54 % of EN, and ~35% of Steppe component), suggesting that the differentiation of Basques cannot be explained by differences in the ancestral components at least previous to the Bronze Age.

The genetic panorama within the Franco-Cantabrian region

A genetic heterogeneity within the Franco-Cantabrian region, including Basques, is observed in the analyses. Focusing in the studied area, the PC separates the Franco-Cantabrian groups, with the Basques in the extreme, from the Spanish and French non-Franco-Cantabrians through a cline represented by the Peri-Basques (Fig. 4A); whereas the second PC separates the regions from the Spanish and the French areas. The PCA show a clear structure and clustering of the Basque groups despite the analysis at a micro-geographical level. In order to have an external reference of similar size and sampling density at a micro-geographical scale, the Catalan samples from Biagini et al. [23] were compared to our dataset. The PCA of Catalans do not show any geographical structure compared to Basques and Peri-Basques (Fig S10), reinforcing the idea of strong genetic clustering in Basques. To quantify and compare the genetic differentiation between the Franco-Cantabrian groups and the Catalan ones, F_{ST} distances for each pairwise combination were estimated and shown in a MDS plot (Fig. S11). Again, Franco-Cantabrians show a clear internal differentiation with distances in a range of 10^{-2}, whereas Catalans showed no evidences of genetic structure or extreme internal differentiation with distances in the range of 10^{-3}. Moreover, we tested for heterogeneity within the region by performing AMOVA analyses at different strata in the geography. Though the explained variance was not so high in the analyses, all the results were statistically significant, pointing to an internal differentiation of the region and especially in the Basque groups (Fig. S12). In fact, the same analysis in Catalonia shows much lower explained variance in the different strata (Fig. S13).

The PCA including French and Spanish samples (Fig. 4A) show two groups of Basque samples: the Western Basque group, which includes all the Spanish Basques in the present study, with the exception of Roncal (RON), together with the samples labeled as “Basque_Spanish” and “Spanish_Pais_Vasco_IBS” from Lazaridis et al. [24]; and the Eastern Basque group including our French Basque samples plus RON and the samples labeled as “Basque_French” in Lazaridis et al. [24]. A similar East-West pattern is shown in the Peri-Basques samples, clustered in two groups: the Spanish Peri-Basques (Western) with the exception of Northern Aragon (NAR), which is clustered with the French Peri-Basques (Eastern). The analysis of genetic components performed in the region with ADMIXTURE mirrors the results above (Fig. 4B). At K=2, the lowest cross-validation error, one component (green) is maximized in the Basques. This component is also present in the Peri-Basque groups at substantial proportions, whereas it is found at low proportions in the external samples, especially in the French ones. At K=3 and K=4, internal different components appear within Basques. In K=3, the Basque-like component split in two specific components: a Western component (blue) mostly present in Spanish Basques; and an Eastern component (green) mostly present in French Basques and RON.

These components are barely presented in the non-Franco-Cantabrian samples. In the Peri-Basque groups, a heterogeneity between the samples from the French and the Spanish side of the region is also observed. Finally, at K=4 another ancestral component arises, maximized in Alava and the surrounding groups (pink), especially the Spanish Peri-Basque groups. Thus, we could summarize the components in non-Franco-Cantabrian (orange), Eastern/French Basque-like (green), central Basque-like (blue), and peripheral Spanish Basque-like (pink). The distribution of these components along the samples, it is basically evidencing the connection among the groups observed in the other analyses and the correlation between genetics and geography.

To formally test for the correlation between the genetic and geographic distances in the region, we performed an isolation by distance (IBD) analysis. We applied a Mantel test between the F_{ST} values and the geographic distances in kilometers, resulting in a positive and clear statistically significant result ($R^2 = 0.242$, p-value = 0.0163) (Fig. S10B -left-). Then, the spatially explicit statistical method, EEMS, showed a very well defined internal pattern of barriers (Fig. S10A -left-); both between the Basque and Peri-Basque area, as well as within them. In fact, the pattern of the corridors with higher migration rate is actually mirroring the observed relationships in the PCA and the ADMIXTURE analysis, between groups with overlapping standard deviations (Fig. S10C -left-, Fig. 4A and Fig. 4B). The same analyses were performed for the Catalan samples, showing a non-significant and negative trend for the Mantel test (R^2 = -0.151, p-value = 0.749), besides the absence of barriers in the EEMS analysis (Fig. S10A, S10B and Fig. S10C -right-), despite being a geographical region of similar size.

In order to refine the relationships between the populations within the area, haplotype-based methods were applied, and similar patterns of internal heterogeneity were observed (Fig. 3). Besides the differentiation of Basque cluster in the fineSTRUCTURE tree (Fig. 3 and Fig. S2), several internal clusters in the region can be defined, mostly related to geography and language. On the one hand, three big clusters can be discerned in the Basque branch (Fig. 3B -right- and 3C): one that encompasses the Central Basque area (NNO, ZMX, plus GUI and GSO), an Eastern Basque cluster (including the French groups NLA and SOU, and RON from the Spanish side) and a Southwestern Basque cluster (ALA, BBA and NCO). On the other hand, three clusters can be differentiated in the Peri-Basque branch (Fig. 3B -left- and 3C): a Western Peri-Basque cluster (BOC, besides a cluster including RIO and BUR mixed in two branches), an Eastern Peri-Basque cluster (BIG, NAR from the Spanish side of the Pyrenees, plus a branch including CHA and BEA), and a specific branch for CAN that falls within the external Spanish cluster. Similar clusters are observed when performing the analysis reducing the dataset at two different levels: for the Franco-Cantabrian samples together with the non-Franco-Cantabrian Spanish and French samples, and for the Franco-Cantabrian samples exclusively (Fig. S2). The differences found in Basques are also evident in the ancestry profiles calculated in the NNLS analysis

(Fig. 3D, Fig. S14). The French and Spanish Peri-Basque groups have different proportions from Basque, Peri-Basque, and external ancestries; whereas Basques are mainly a mixture of Basque components with some Peri-Basque ancestries. In fact, the three main different Basque clusters are defined in the NNLS analysis, with a Central group defined exclusively by Basque ancestry; the Eastern group defined by Basque ancestry with ~25% of Peri-Basque ancestry; and the Southwestern group defined by Basque ancestry, ~25% of Peri-Basque ancestry, and some traces of Spanish and French ancestries. It is noteworthy that besides these traces of Spanish and French ancestry in the Southwestern group, no traces of external clusters are found in any of the Basque samples analyzed, pointing to an internal haplotype sharing in the region without external contributions.

Discussion

We have focused our study in three direct questions: Is the genomic panorama of the Franco-Cantabrian region similar to the external European populations? How is the genomic scenario within the region? Which could be their genomic history?

First, the genetic distinctiveness of the Basques is evident in all the analysis performed. In the European/Mediterranean genetic context, the results show a similar scenario as previously reported in other studies. A peripheral distribution in the PCA (Fig. 2B) and a remarkable differentiation in the proportions of the plausible ancestral components in the ADMIXTURE (Fig. 2B) place them in the already known wide panorama related to allele frequency studies [24]. However, the most interesting results are related to the fine-scale analysis in a Spanish-French context, that show us a very detailed panorama of differentiation. The distribution in the PCA (Fig. 4A), as well as the inferred ancestral components in the ADMIXTURE analysis (Fig. 4B), finally clarify the first question overcoming the limited representation of the Franco-Cantabrian region [6], [7]. These results are also supported by the haplotype-based methods. The external branching of Basques together with the Peri-Basques clustering more internally in the fineSTRUCTURE tree (Fig. 3A, Fig. 3B, Fig. 3C and Fig. S2), plus the higher shared ancestry of Peri-Basques with external groups (Fig. 3D) completely mirror the allele frequency results. Therefore, Basque people are defined as an isolated group with a clear genetic distinctiveness, whereas Peri-Basques can be explained as a peri-isolated transition between the Basques and the external Spanish and French populations. This mirrors the distinctive continuous pattern of genomic variation as intersection between open and isolated populations [41]. Isolated populations are characterized by large number and length of ROHs, due to high levels of inbreeding and/or small effective population sizes [41]. The results in the ROH analysis clearly show a pattern of a continuous inbreeding in the Franco-Cantabrian region, and especially in Basques. Sardinia is an already studied population reported to be highly inbred and to carry many long ROH [40], [42]. Central Western Europeans and Sub-Saharan Africans represent large populations, characterized by short and few ROHs that show the basal level of ROHs [42], [43]. The highest individual values of total ROH number and length are observed in Basques, followed by Sardinians and Peri-Basques (Fig. 4C and Fig. S4). Moreover, the average of total number and length, calculated by ROH length category, in the different groups allows to understand the ROH values in a temporal context. Old inbreeding is related to high length and number values in the smaller categories due to the recombination during the time, thereby more recent inbreeding is reflected by high values in longer categories. The shorter categories (0.5-1 Mb and 1-2 Mb) are related to the basal ROH in human populations inherited from the Out-of-Africa, thus they are ubiquitous and the most frequent in all the populations [43]. Similar values have been reported for African and European non-inbred populations. Thus, greater number and total length of ROH in these categories are echoing scenarios of small and isolated areas with high levels of old inbreeding [43]. Our Franco-Cantabrian samples present a similar pattern as Sardinians. In the first two categories, 0.5-1 Mb and 1-2 Mb, both Spanish and French Basques present the highest values, followed by Sardinians. Despite Peri-Basques values are lower, they are higher than in the external Spanish and French groups. In the categories for runs 2-4 Mb and 4-8 Mb, an intermediate scenario is observed among the groups, without clear outstanding results in the Franco-Cantabrian region. However, the total proportion of samples representing these categories in the external groups is very small compared to the Franco-Cantabrian one, even below 30 % (Fig. S5). This shows that these categories are more common in the isolated groups, Basques and Sardinians, and the peri-isolated Peri-Basques, while in the external groups the values observed could be more related to cryptically inbred outliers [43]. In the last category of >8 Mb, Spanish and French Basques, together with Spanish Peri-Basques are still represented, , but together with Spanish and Central Western Europeans. All of them represented by a small proportion of samples. Therefore, a pattern of genetic isolation, together with old and continuous strong inbreeding can be detected in the Franco-Cantabrians, more accentuated in Basques groups.

Second, the internal heterogeneity in the Basques is evident from allele frequency methods to more robust and refined approach with ChromoPainter, fineSTRUCTURE and GLOBETROTTER. It is striking to find a scenario of these characteristics in such a small continental area. When comparing with a region with similar extension and sampling density, these exceptional results are clearly evident (Fig. S10, Fig. S11, Fig. S12 and Fig. S13). In fact, our results remind to the internal scenario in Sardinia, which is actually a more wider and geographically isolated island [44]. All our results suggest that, generally, three main groups can be defined within the Franco-Cantabrian region: Basques, Western Spanish Peri-Basques and Eastern Peri-Basques (Fig. 4A and Fig. S10). This result is much more evident in the fineSTRUCTURE tree showing three separated clusters, plus CAN, for the groups and evidencing the closeness of Peri-Basques with the external populations (Fig. 3). Furthermore, the shared ancestry profile with the Franco-Cantabrian region observed in the Peri-Basques principally comes from the geographically closest groups (Fig. 3D and Fig. S14). Thus, Western Peri-Basques show shared ancestry with Spanish Basque and Peri-Basque groups, whereas the Eastern Peri-Basques share with the French Basques and Peri-Basques. If we focus within Basques, a second level of differentiation can be defined. Three major Basque groups are evinced, that actually can be related to the three different components observed in the internal ADMIXTURE analysis (Fig. 4B). A Western Spanish Basque cluster that mainly shares ancestries with the Spanish Basque and Peri-Basque groups, an Eastern, mainly French, Basque peripheral area that shares ancestries from the surrounding French Basque and Peri-Basque groups, and a central Basque region with shared ancestries exclusively from Basque groups (Fig. 3D and Fig. S14). Despite small proportions, the Spanish Basque peripheral cluster is the only one that presents shared ancestry with outsider groups, probably due to a higher gene flow between these parts of the region. In fact, these genetics clusters observed here, actually reflect the historical connections between the Basques and the surrounding groups. The core cluster has been usually defined as the most differentiated and unique group in previous studies during the last decades, due to have maintained less contact with the surrounded populations [45], [46]. The Western Basque peripheral cluster, is adjacent to the Spanish Peri-Basques so the gene flow and influence with external sources should have been higher, leading to those small proportions of non-Franco-Cantabrian ancestries. Indeed, ALA and NCO do not speak Euskara anymore, and in BBA the influence of the Spanish language has been very high, reducing the number of Euskara speakers [18], [20]. In the Eastern Basque peripheral cluster, in spite of being adjacent to the French Peri-Basques, the limits of the French Basque region have been more stable than in the Spanish area [18]. Besides observing a clear heterogeneity in the region, we can definitely correlate this with the geography. This correlation might go hand in hand with the languages and the distribution of Euskara dialects as they mainly suit with the different Basque groups [19] (Table S1). It is clear that the Euskara could have acted as a cultural barrier, due to the strong differences with the surrounding Romance languages. Nevertheless, it is more difficult to support that the dialects were the direct cause of shaping the genetic heterogeneity in the region. The classification and differences among the dialects of the Euskara are very diffuse. They are actually the overlapping of linguistic varieties along the territory with similar patterns among the adjacent areas. Nonetheless, a dialectal Western-Eastern discontinuity is evident [2], [19], [21] and it is somehow supported by the clusters and genetic affinities observed among our Basque groups. Therefore, is more plausible to consider a scenario where geography is the determining factor of the connection among the groups along the region, leading to differential levels of gene flow and the diversification of the dialects among the groups during the history, diachronically or not. Thus, we could explain the observed genetic affinity between the close groups from the Spanish and French areas along the Pyrenees: NAR and the French Peri-Basques, RON and the French Basque groups, and Northwestern Nafarroa with the western part of Lapurdi. This suggests that the internal heterogeneity in Basques is more complex than the Northern-Southern geographic and administrative limits between the present-day Spanish and French territories. They are diffuse and do not shape the genetic scenario in the region. Instead, concentric stripes are evident along the Western Pyrenees in the region, through a genetic cline from the core to the external groups, clarifying previous results with classical markers where strict vertical groups were described for Basques [45], [47].

Third, we do not detect specific differences in the Franco-Cantabrian region related to the main three ancestral components in Europe. Several studies, based on classical and uniparental markers, have placed Basques, as Pre-Neolithic remnants that avoided the latter Neolithic and Bronze Age migrations, with low gene flow from early farmers and steppe herders [48], [49]. However, the latest results with genome-wide data and including ancient samples abandoned those conclusions, pointing to the origin of Basques as the result of gene flow between the Neolithic and the upcoming historical groups [11], [12]. Olalde, et al. [12] suggested a genetic continuity since the Bronze Age - Iron Age for Basques, and lack of admixture with the posterior movements of people such as Roman (Italian-Greek) and North African. Despite this result has been very striking because of using many ancient samples, this was previously suggested from fine-scale studies using uniparental markers [5]. When explicitly modeling our samples for proportions of Pre-Neolithic Hunter-gatherer, Neolithic Farmer and Bronze Age Steppe components with qpGraph, no unexpected results are observed in the region (Fig. 5), following the Northern-Southern European expected cline of ancient components [34]. Furthermore, the shared ancestry profiles inferred with NNLS (Fig. 3D) do not show any remarkable evidence of connection between Basques and post-Bronze Age potential ancestries considering the modern samples as proxies (i.e. Greek or Italian ancestries as representative for Roman gene flow, or North African from the Arab expansion). Our results confirm, in a wider European panorama and with fine-scale representation of the region, that all the Basques were influenced by the two major migration waves in Europe until the Bronze Age period, similarly as in the surrounding populations. Then, it was followed by an isolation, based on low geneflow with posterior population movements, despite having cultural contact with at least the Romans. This strong genetic continuity of Basques since Bronze also support that the expansion of the steppe ancestry do not always prompt shifts to Indo-European languages like in the rest of Western Europe as previously noticed [12], [34]. Furthermore, it does not strictly means that they would not had experienced a previous isolation to that time, in fact the results of ROHs evince a signal of ancient inbreeding in the region, higher than in Sardinia which is suggested to be isolated after Neolithic times.

Conclusion

In this study, we disentangle the main controversies around Basques overcoming the limitations in the previous studies. The representation of the entire region with samples from the main Franco-Cantabrian areas allows assessing fine-scale results and describing the internal scenario in the region. Furthermore, the use of genome-wide data genotyped with the Human Origins Array allows studying the demography with higher number of genetic markers that have been specifically selected to avoid the population ascertainment bias in the analysis. Finally, our methodology includes more robust and specific techniques that complement the allele frequency and the haplotype-based approaches to face the objectives from different perspectives and go beyond the available geographic labels. Thus, the results show that Franco-Cantabrians present a specific genetic profile that place Basques as an isolated population in the wide Western-Eurasian context and, the here called Peri-Basques as peri-isolated groups forming a cline towards the external populations. Furthermore, a clear heterogeneity correlated with geography, and potentially language, is observed within the Franco-Cantabrian region. Finally, the specific genetic scenario observed in the region seem to has started after a post-Bronze Age genetic isolation, due to the lack or very low gene flow with the following feasible components that are found in the surrounding populations. Thus, the origin of the genetic panorama in the region could be an isolation of the groups after the Bronze Age, probably encouraged by the geography and language, together with a continuous inbreeding, especially in Basques. Disentangling the controversies around Basques is useful not only in human population genetics, but also in archaeology and linguistics. In the case of using Basques as proxy for ancestral European gene pool, it should be taking into account that their genetics actually depicts the genomic situation of Western Europe in the Bronze Age - Iron Age period, but not before. Furthermore, it would be more appropriate to use those groups from the central area of the region as genetic representatives, since they remain more genetically distant to the external populations than the other peripheral Basque groups. This should not be misunderstood in the sense of the population as a whole; at the end, Basque people are connected by a common strong culture and history more than a genetic uniformity, in the sight of the observed internal heterogeneity. Finally, our results could also be helpful to enlighten the anthropological and chronological connections between the ancient remains, as well as the historical linguistic relationships around the Euskara.

References

[1] A. W. G. Pike *et al.*, "U-Series Dating of Paleolithic Art in 11 Caves in Spain," *Science*, vol. 336, no. 6087, pp. 1409–1413, Jun. 2012.

[2] L. Nuñez Astrain, *El euskera arcaico : extensión y parentescos*. Txalaparta, 2004.

[3] A. E. Mourant, "The Blood Groups of the Basques," *Nature*, vol. 160, no. 4067, pp. 505–506, Oct. 1947.

[4] A. M. Pérez-Miranda *et al.*, "Microsatellite data support subpopulation structuring among Basques," *J. Hum. Genet.*, vol. 50, no. 8, pp. 403–414, 2005.

[5] B. Martínez-Cruz *et al.*, "Evidence of pre-roman tribal genetic structure in basques from uniparentally inherited markers," *Mol. Biol. Evol.*, vol. 29, no. 9, pp. 2211–2222, 2012.

[6] N. Rodríguez-Ezpeleta *et al.*, "High-density SNP genotyping detects homogeneity of Spanish and French Basques, and confirms their genomic distinctiveness from other European populations," *Hum. Genet.*, vol. 128, no. 1, pp. 113–117, 2010.

[7] H. Laayouni, F. Calafell, and J. Bertranpetit, "A genome-wide survey does not show the genetic distinctiveness of Basques," *Hum. Genet.*, vol. 127, no. 4, pp. 455–458, 2010.

[8] J. Bertranpetit, J. Sala, F. Calafell, P. A. Underhill, P. Moral, and D. Comas, "Human mitochondrial DNA variation and the origin of Basques," *Ann. Hum. Genet.*, vol. 59, no. Pt 1, pp. 63–81, 1995.

[9] D. M. Behar *et al.*, "The Basque paradigm: Genetic evidence of a maternal continuity in the Franco-Cantabrian region since pre-neolithic times," *Am. J. Hum. Genet.*, vol. 90, no. 3, pp. 486–493, 2012.

[10] N. Solé-Morata *et al.*, "Analysis of the R1b-DF27 haplogroup shows that a large fraction of Iberian Y-chromosome lineages originated recently in situ," *Sci. Rep.*, vol. 7, no. 1, pp. 1–13, Aug. 2017.

[11] T. Günther *et al.*, "Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques," *Proc. Natl. Acad. Sci.*, vol. 112, no. 38, pp. 11917–11922, 2015.

[12] I. Olalde *et al.*, "The genomic history of the Iberian Peninsula over the past 8000 years," *Science (80-.).*, vol. 363, no. 6432, pp. 1230–1234, Mar. 2019.

[13] S. López and S. Alonso, “Genetics and the History of the Basque People.” 25-Aug-2013.

[14] N. Patterson *et al.*, “Ancient admixture in human history.,” *Genetics*, vol. 192, no. 3, pp. 1065–93, Nov. 2012.

[15] S. Leslie *et al.*, “The fine-scale genetic structure of the British population,” *Nature*, vol. 519, no. 7543, pp. 309–314, Mar. 2015.

[16] L. M. Gattepaille and M. Jakobsson, “Combining markers into haplotypes can improve population structure inference,” *Genetics*, vol. 190, no. 1, pp. 159–174, 2012.

[17] K. Zuazo, “The basque country and the Basque language,” in *Towards a History of the Basque Language*, 1996, p. 5.

[18] C. C. Abasolo, “Las fronteras de la lengua vasca a lo largo de la historia,” *Rev. Filol. románica*, vol. 19, 2002.

[19] J. Abaitua, “Patrones geolingüísticos, áreas dialectales y cronologías absolutas del EHHA,” *Fontes linguae Vascon. Stud. Doc.*, vol. 50, no. 126, pp. 283–322, Aug. 2018.

[20] H. K. Borrás, “El euskera en tierras del romance: Rioja alavesa, La Rioja, Burgos, Encartaciones,” *Huarte San Juan . Filol. y Didáctica la Leng.*, vol. 10, 2008.

[21] K. Zuazo, *El euskera y sus dialectos*. Irún: Alberdania S. L., 2010.

[22] S. Purcell *et al.*, “PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses,” *Am. J. Hum. Genet.*, vol. 81, no. 3, pp. 559–575, Jul. 2007.

[23] S. A. Biagini, N. Solé-Morata, E. Matisoo-Smith, P. Zalloua, D. Comas, and F. Calafell, “People from Ibiza: an

unexpected isolate in the Western Mediterranean," *Eur. J. Hum. Genet.*, Feb. 2019.

[24] I. Lazaridis *et al.*, "Genomic insights into the origin of farming in the ancient Near East," *Nature*, vol. 536, no. 7617, pp. 419–424, Aug. 2016.

[25] N. Patterson, A. L. Price, and D. Reich, "Population Structure and Eigenanalysis," *PLOS Genet.*, vol. 2, no. 12, p. e190, Jul. 2006.

[26] L. W. Pembleton, N. O. I. Cogan, and J. W. Forster, "StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations," *Mol. Ecol. Resour.*, vol. 13, no. 5, pp. 946–952, Jul. 2013.

[27] Z. N. Kamvar, J. F. Tabima, and N. J. Grünwald, "Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction," *PeerJ*, vol. 2, p. e281, Jul. 2014.

[28] L. Excoffier, P. E. Smouse, and J. M. Quattro, "Analysis of Molecular Variance Inferred from Metric Distances among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data," *Genetics*, vol. 131, no. 2, pp. 479–491, Jul. 1992.

[29] S. Dray and A.-B. Dufour, "The ade4 Package: Implementing the Duality Diagram for Ecologists," *J. Stat. Softw.*, vol. 22, no. 1, pp. 1–20, Jul. 2007.

[30] "Geographic Distance Matrix Generator (version 1.2.3)." Center for Biodiversity and Conservation, American Museum of Natural History, New York (USA), 14-Feb-2014.

[31] D. Petkova, J. Novembre, and M. Stephens, "Visualizing spatial population structure with estimated effective migration surfaces," *Nat. Genet.*, vol. 48, no. 1, pp. 94–100, Jan. 2016.

[32] D. H. Alexander, J. Novembre, and K. Lange, "Fast model-based estimation of ancestry in unrelated individuals," *Genome Res.*, vol. 19, no. 9, pp. 1655–1664, Sep. 2009.

[33] A. A. Behr, K. Z. Liu, G. Liu-Fang, P. Nakka, and S. Ramachandran, "pong: fast analysis and visualization of latent clusters in population genetic data," *Bioinformatics*, vol. 32, no. 18, pp. 2817–2823, Sep. 2016.

[34] W. Haak *et al.*, "Massive migration from the steppe was a source for Indo-European languages in Europe," *Nature*, vol. 522, no. 7555, pp. 207–211, Jun. 2015.

[35] J. O'Connell *et al.*, "A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness," *PLoS Genet.*, vol. 10, no. 4, p. e1004234, Apr. 2014.

[36] R. A. Gibbs *et al.*, "The International HapMap Project," *Nature*, vol. 426, no. 6968, pp. 789–796, Dec. 2003.

[37] T. 1000 G. P. Consortium, "A global reference for human genetic variation," *Nature*, vol. 526, no. 7571, pp. 68–74, Jul. 2015.

[38] D. J. Lawson, G. Hellenthal, S. Myers, and D. Falush, "Inference of Population Structure using Dense Haplotype Data," *PLoS Genet.*, vol. 8, no. 1, p. e1002453, Jan. 2012.

[39] G. Hellenthal *et al.*, "A genetic atlas of human admixture history.," *Science*, vol. 343, no. 6172, pp. 747–751, Feb. 2014.

[40] C. Di Gaetano *et al.*, "Sardinians genetic background explained by runs of homozygosity and genomic regions under positive selection," *PLoS One*, vol. 9, no. 3, 2014.

[41] P. Anagnostou *et al.*, "Overcoming the dichotomy between open and isolated populations using genomic data from a large European dataset," *Sci. Rep.*, vol. 7, p. 41614, Aug. 2017.

[42] F. C. Ceballos, P. K. Joshi, D. W. Clark, M. Ramsay, and J. F. Wilson, "Runs of homozygosity: windows into population history and trait architecture," *Nat. Rev. Genet.*, vol. 19, no. 4, pp. 220–234, Aug. 2018.

[43] M. Kirin, R. McQuillan, C. S. Franklin, H. Campbell, P. M. Mckeigue, and J. F. Wilson, "Genomic runs of homozygosity record population history and consanguinity," *PLoS One*, vol. 5, no. 11, 2010.

[44] C. W. K. Chiang *et al.*, "Genomic history of the Sardinian population," *Nat. Genet.*, vol. 50, no. 10, pp. 1426–1434, Aug. 2018.

[45] M. Iriondo, M. C. Barbero, and C. Manzano, "DNA polymorphisms detect ancient barriers to gene flow in

basques," *Am. J. Phys. Anthropol.*, vol. 122, no. 1, pp. 73–84, 2003.

[46] C. Manzano, A. I. Aguirre, M. Iriondo, M. Martín, L. Osaba, and C. de la Rúa, "Genetic polymorphisms of the Basques from Gipuzkoa: genetic heterogeneity of the Basque population," *Ann. Hum. Biol.*, vol. 23, no. 4, pp. 285–296, 1996.

[47] F. Calafell and J. Bertranpetit, "Mountains and genes: population history of the Pyrenees," *Hum. Biol.*, vol. 66, no. 5, pp. 823–42, Oct. 1994.

[48] N. Izagirre and C. de la Rúa, "An mtDNA analysis in ancient Basque populations: implications for haplogroup V as a marker for a major paleolithic expansion from southwestern europe.," *Am. J. Hum. Genet.*, vol. 65, no. 1, pp. 199–207, Jul. 1999.

[49] A. M. Lopez-Parra *et al.*, "In search of the Pre- and Post-Neolithic Genetic Substrates in Iberia: Evidence from Y-Chromosome in Pyrenean Populations," *Ann. Hum. Genet.*, vol. 73, no. 1, pp. 42–53, Aug. 2009.

[50] R. Collins, *The Basques*. Basil Blackwell, 1990.

Main Figures

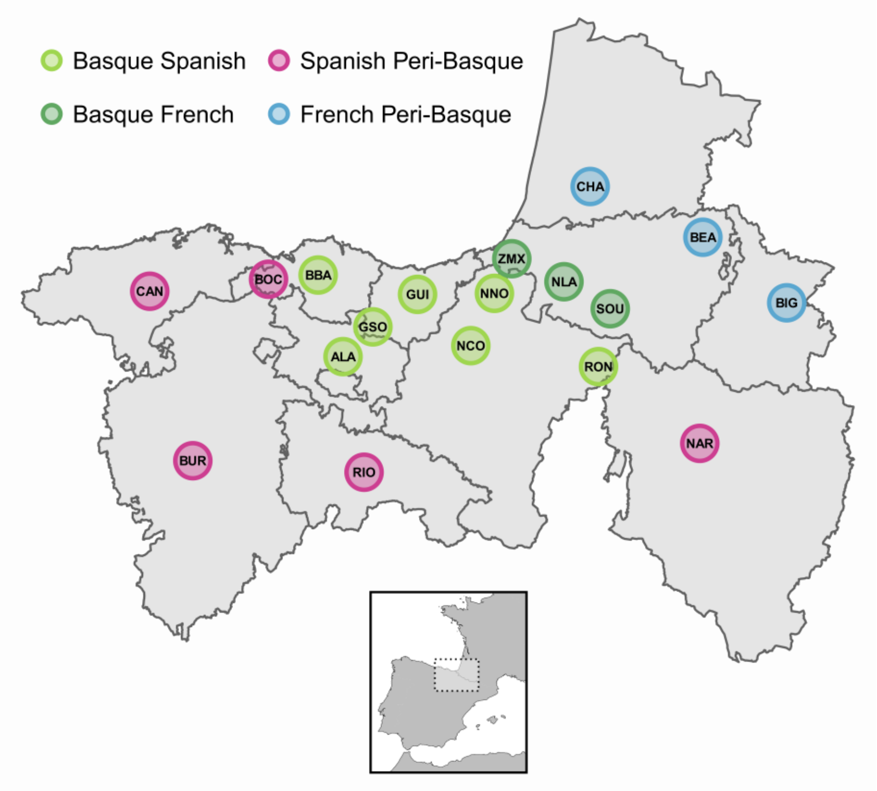


Figure 1. Geographic distribution of the Franco-Cantabrian region and the areas included in the study. The colors represent the main language spoken in the corresponding area. ALA, Araba; BBA, Bizkaia; BOC, Western Bizkaia; GUI, Gipuzkoa; GSO, Southwestern Gipuzkoa; NNO, Northwestern Nafarroa; NCO, Central Western Nafarroa; ZMX, Lapurdi/Baztan; NLA, Lapurdi Nafarroa; SOU, Zuberoa; RON, Roncal; CAN, Cantabria; BUR, Northern Burgos; RIO, Rioja; NAR, Northern Aragon; CHA, Chalosse; BEA, Béarn; BIG, Bigorre.

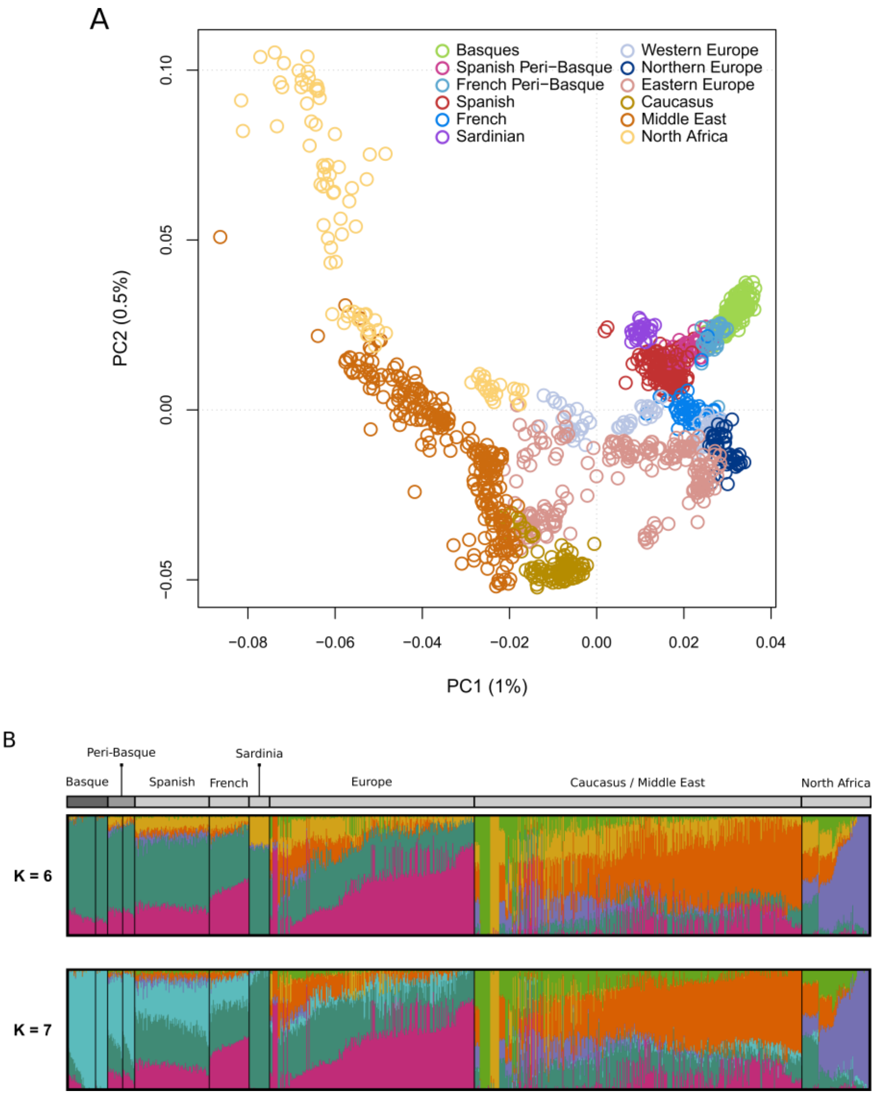


Figure 2. Contextualizing the Franco-Cantabrian region. (A) Principal component analysis including all the modern samples. (B) ADMIXTURE results at K=7, which is the one with the lowest cross-validation error. The other Ks considered in the ADMIXTURE analysis are available Fig. S1.

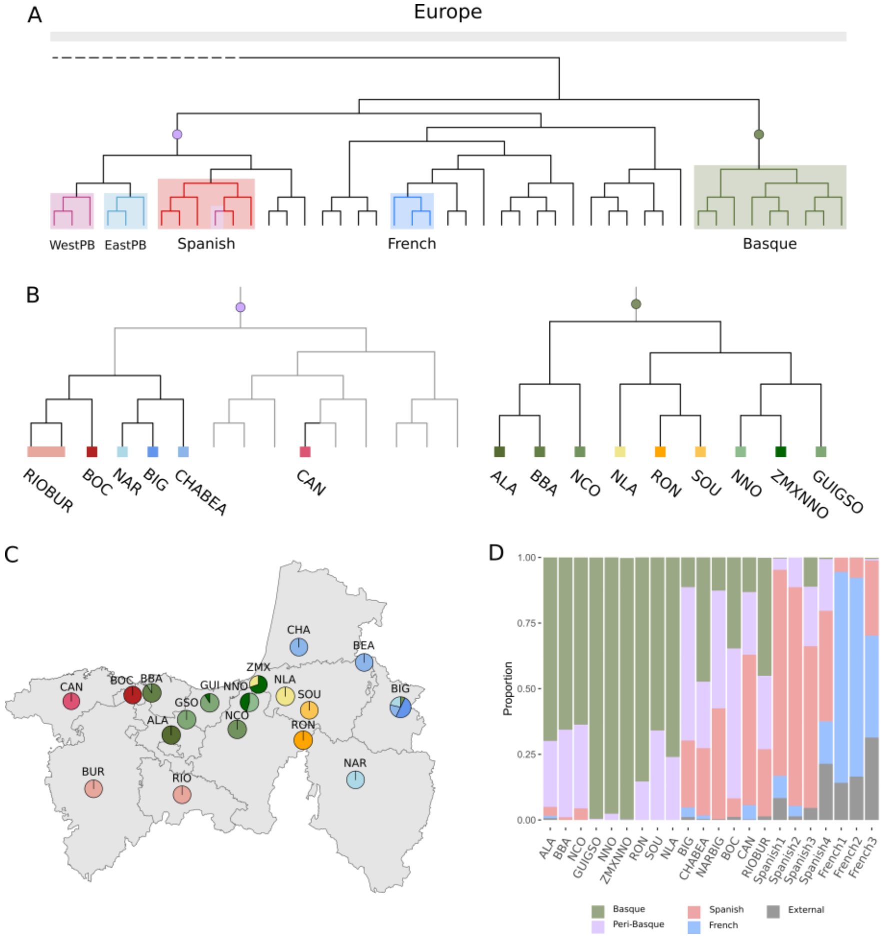


Figure 3. Haplotype-based analyses. (A) European branch of the fineSTRUCTURE tree inferred for the whole dataset. The complete tree is shown in Figure S2A. The Franco-Cantabrian clusters are highlighted in colors, besides the Spanish and French ones. WestPB, West Peri-Basque; EastPB, East Peri-Basques. (B) Internal Franco-Cantabrian clusters. (C) Geographical distribution of the haplotypes inferred in the fineSTRUCTURE tree. Each pie chart represents the proportion of the clusters in the corresponding region. (D) Inferred proportions of shared ancestry using the NNLS method from GLOBETROTTER. Those outsider proportions that are not Spanish or French were compiled together as “external”.

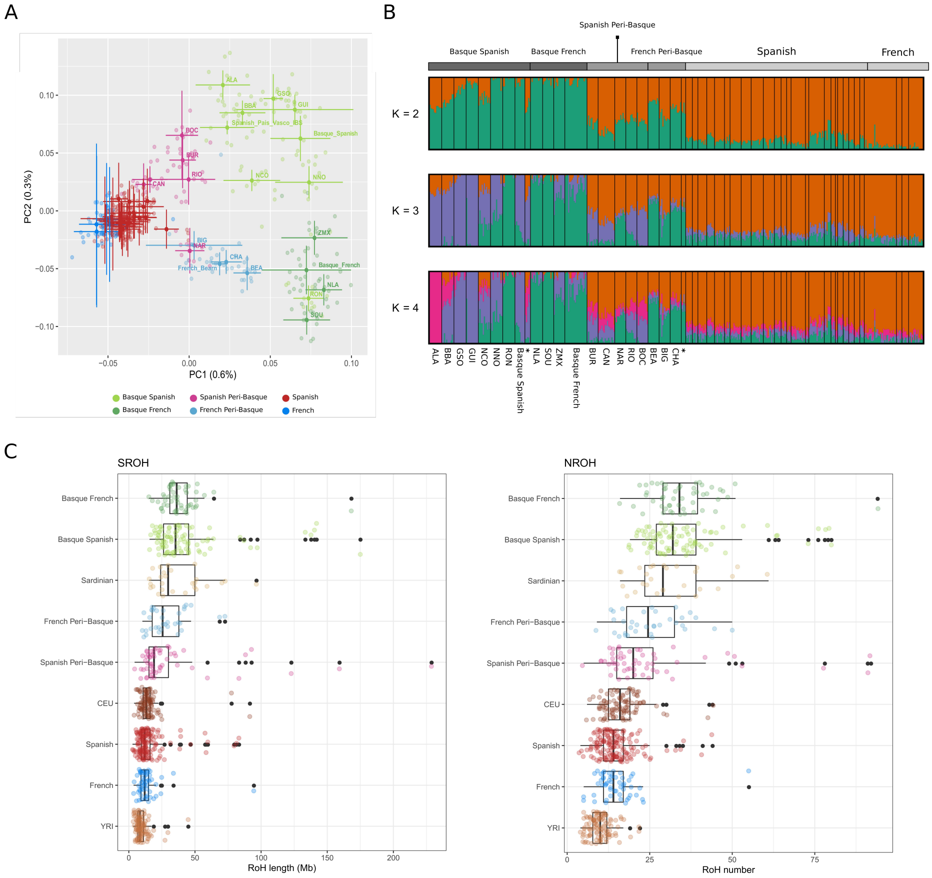


Figure 4. Population stratification of the Franco-Cantabrian region. (A) PCA using the samples from the Iberian Peninsula and France. The PCA averages and standard deviation of the different geographic groups were plotted. (B) ADMIXTURE results from K=2 to K=4. The lowest cross-validation error in the analysis was K=2. The first * “Spanish_Pais_Vasco_IBS” and the second “French_Bearn”. (C) Whisker plots representing the Individual information of SROH -left- and NROH -right- for each group from France and Iberian Peninsula. Sardinian, Central Western Europeans (CEU) and sub-Saharan Africans (YRI) were also included as reference populations with well-known demographic histories.

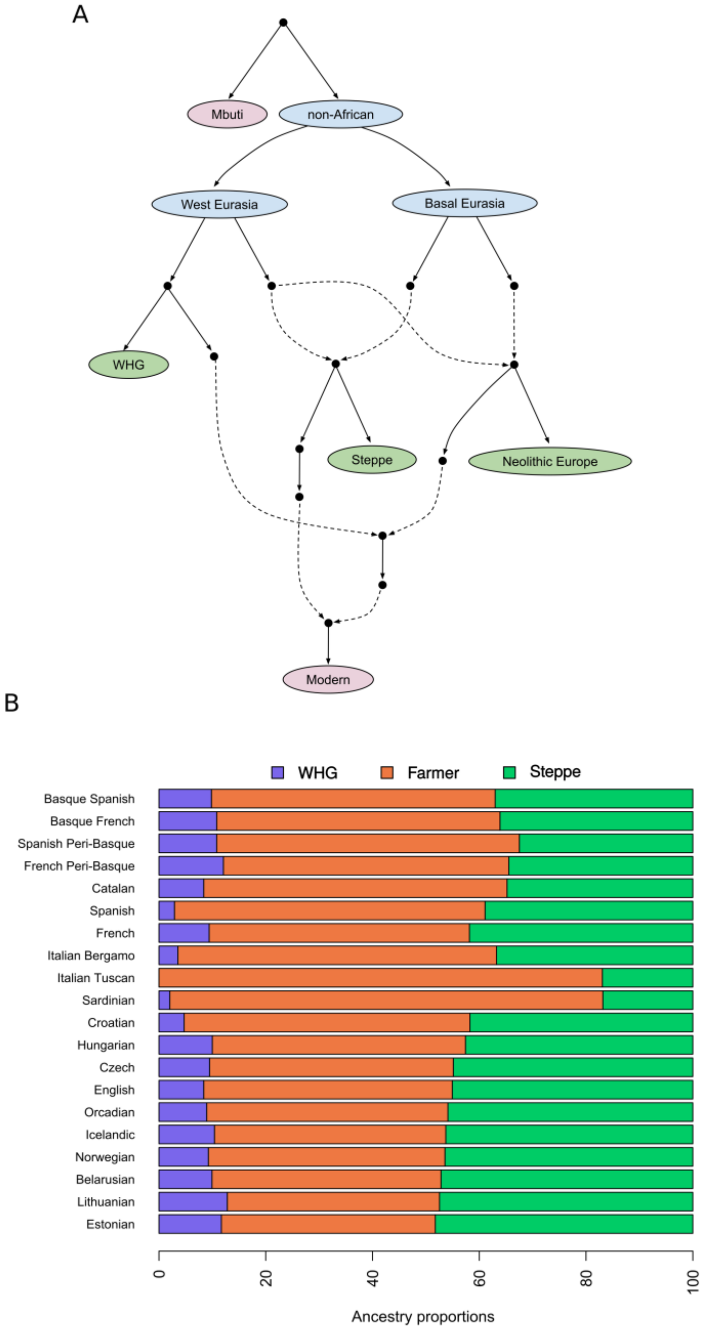
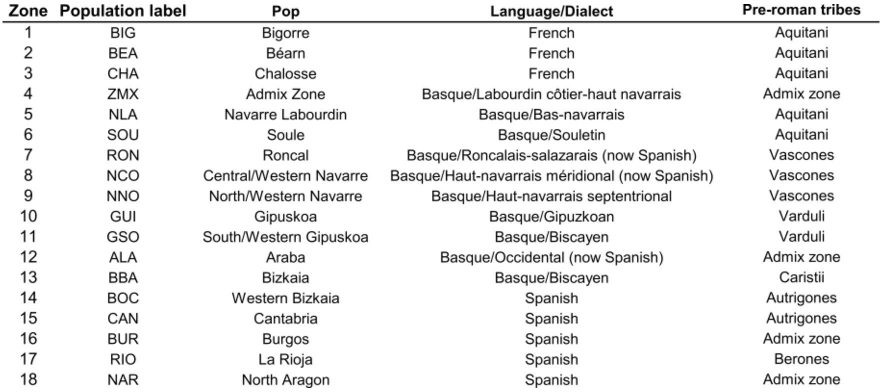


Figure 5. Modeling the Franco-Cantabrian ancient history. (A) qpGraph model used to calculate the mixture proportions. Dashed lines represent admixture. Modern samples are colored in pink, ancient samples

in green, main ancestral Eurasian groups in blue and intermediate reconstructed ancestral groups as a black dot. (B) Mixture proportions in different European populations.

Supplementary Data

Table S1. Information of the newly reported samples. A total of 190 samples from 18 areas of the Franco-Cantabrian region were genotyped with the Axiom™ Genome-Wide Human Origins Array. All these individuals had their four grandparents from the same geographical area. We included the main areas were Euskara is spoken nowadays, besides the surrounding areas where Spanish and French is used instead.



Zone	Population label	Pop	Language/Dialect	Pre-roman tribes
1	BIG	Bigorre	French	Aquitani
2	BEA	Béarn	French	Aquitani
3	CHA	Chalosse	French	Aquitani
4	ZMX	Admix Zone	Basque/Labourdin côtier-haut navarrais	Admix zone
5	NLA	Navarre Labourdin	Basque/Bas-navarrais	Aquitani
6	SOU	Soule	Basque/Souletin	Aquitani
7	RON	Roncal	Basque/Roncalais-salazarais (now Spanish)	Vascones
8	NCO	Central/Western Navarre	Basque/Haut-navarrais méridional (now Spanish)	Vascones
9	NNO	North/Western Navarre	Basque/Haut-navarrais septentrional	Vascones
10	GUI	Gipuskoa	Basque/Gipuzkoan	Varduli
11	GSO	South/Western Gipuskoa	Basque/Biscayen	Varduli
12	ALA	Araba	Basque/Occidental (now Spanish)	Admix zone
13	BBA	Bizkaia	Basque/Biscayen	Caristii
14	BOC	Western Bizkaia	Spanish	Autrigones
15	CAN	Cantabria	Spanish	Autrigones
16	BUR	Burgos	Spanish	Admix zone
17	RIO	La Rioja	Spanish	Berones
18	NAR	North Aragon	Spanish	Admix zone

Table S2. Composition of the fineSTRUCTURE clusters based on the tree. The first column is the clusters names used for the second step of ChromoPainter and the NNLS. The second column indicates the label of the populations within the cluster. The third column is the number of samples from each population within the repective cluster.

Cluster name	**Population name**	**N**
ALA	ALA	11
GUIGSO	Basque_Spanish	3
	BBA	1
	GSO	11
	GUI	10
BBA	BBA	10
	Spanish_Pais_Vasco_IBS	5
ZMXNNO	Basque_French	2
	GUI	1
	NNO	5
	ZMX	7
NCO	BIG	1
	NCO	11
NNO	Basque_French	1
	Basque_Spanish	6
	NNO	6
RON	RON	11
NLA	Basque_French	14
	NLA	11
	ZMX	3
SOU	Basque_French	3
	SOU	10
Spanish2	Alacant	2
	Camp_de_Tarragona	3
	Castello	2
	Catalunya_Central	1
	French_Southwest	2
	Lleida	1
	Mallorca	1
	Penedes	1
	Pirineu	2
	Valencia	2
Spanish1	Alacant	2
	Barcelones	5
	Camp_de_Tarragona	5
	Castello	3
	Catalunya_Central	8
	Eivissa	4
	French_Southwest	1
	Girona	10
	Lleida	8
	Mallorca	8
	Menorca	3
	Penedes	10
	Peri_Barcelona	10
	Pirineu	3
	Spanish_Aragon_IBS	1
	Spanish_Baleares_IBS	3
	Spanish_Cataluna_IBS	4
	Spanish_Galicia_IBS	1
	Terres_de_lEbre	5
	Valencia	3
	Alacant	2
	Castello	2
	Spanish_Andalucia_IBS	4

Spanish3	Spanish_Aragon_IBS	3
	Spanish_Castilla_la_Mancha_IBS	5
	Spanish_Castilla_y_Leon_IBS	3
	Spanish_Cataluna_IBS	1
	Spanish_Extremadura_IBS	2
	Spanish_Murcia_IBS	4
	Spanish_Valencia_IBS	5
	Valencia	4
Spanish4	CAN	1
	Pirineu	1
	Spanish_Baleares_IBS	1
	Spanish_Canarias_IBS	2
	Spanish_Castilla_y_Leon_IBS	2
	Spanish_Extremadura_IBS	3
	Spanish_Galicia_IBS	4
RIOBUR	BUR	10
	RIO	10
CAN	CAN	9
	Spanish_Cantabria_IBS	5
NARBIG	BIG	2
	French_Bigorri	1
	NAR	10
	Spanish_Aragon_IBS	2
BOC	BOC	10
CHABEA	BEA	10
	CHA	10
	French_Bigorri	3
BIG	BIG	7
French2	French_North	2
	French_NorthWest	5
	French_Southwest	20
FrenchUK	English_Cornwall_GBR	5
	English_Kent_GBR	5
	French_NorthWest	1
	French_Southwest	5
	Scottish_Argyll_Bute_GBR	4
	South_Corsica	5
French1	French_North	4
	French_NorthWest	6
	French_Southwest	1
Italian_North	French_NorthWest	1
	Italian_Bergamo	12
	Italian_Tuscan	8
ItalianSicily	Italian_EastSicilian	5
	Italian_South	1
	Italian_WestSicilian	6
MediterraneanIslands	Cypriot	6
	Greek_Coriell	1
	Italian_South	3
	Maltese	8
Sardinian	Sardinian	27
Ice_Orc	Icelandic	12
	Orcadian	12
Hun_Czech	Czech	10
	Hungarian_Coriell	10

	Hungarian_Metspalu	9
	Ukrainian_West	1
NE_Europe	Belarusian	10
	Chuvash	10
	Estonian	10
	Finnish_FIN	7
	Lithuanian	10
	Mordovian	10
	Russian	22
	Saami_WGA	1
	Ukrainian_East	6
	Ukrainian_West	2
NorwCro	Croatian	10
	Norwegian	11
SE_Europe	Albanian	6
	Bulgarian	9
	Greek_Comas	14
	Greek_Coriell	4
	Hungarian_Metspalu	1
	Romanian	10
	Turkish_Istanbul	3
AshkenaziJew	Ashkenazi_Jew	7
WesternMiddleEast	Cypriot	2
	Druze	1
	Jordanian	1
	Lebanese	6
	Lebanese_Christian	9
	Lebanese_Muslim	11
	Saudi	1
	Syrian	2
	Turkish_Adana	2
Caucasus1	Armenian	10
	Assyrian	11
	Georgian_Jew	5
	Greek_Coriell	1
	Syrian	1
	Turkish_Adana	3
	Turkish_Kayseri	2
Turkish2	Turkish	4
	Turkish_Adana	4
	Turkish_Aydin	7
	Turkish_Balikesir	6
	Turkish_Istanbul	6
	Turkish_Kayseri	8
Iranian2	Iranian	8
	Iranian_Lor	10
	Iranian_Mazandarani	10
	Iranian_Persian	9
	Turkish_Adana	1
	Turkish_Istanbul	1
TurkishJew	Turkish_Jew	8
Turkish1	Turkish_Trabzon	9
	Abkhasian	9
	Adygei	16
	Balkar	10
	Georgian_Megrels	10

Caucasus2	Nogai	9
	North_Ossetian	10
	Chechen	9
	Kumyk	8
	Lezgin	9
MiddleEastJew	Georgian_Jew	2
	Iranian_Jew	9
	Iraqi_Jew	6
MiddleEast1	BedouinA	9
	Iranian_Bandari	2
	Jordanian	8
	Lebanese	2
	Palestinian	5
	Saudi	2
	Syrian	5
	Yemen	3
MiddleEast2	BedouinA	8
	Egyptian_Comas	11
	Egyptian_Metspalu	6
	Yemen	1
BedouinA	BedouinA	7
MiddleEast3	BedouinA	1
	BedouinB	19
	Saudi	5
Druze	Druze	36
Iranian1	Iranian_Bandari	6
	Iranian_Persian	1
Palestinian	Palestinian	30
NorthAfrica1	Algerian	5
	Egyptian_Metspalu	1
	Libyan	5
	Moroccan	10
	Tunisian	8
	Yemen	2
YemenJew	Yemenite_Jew	8
NorthAfricaJew	Libyan_Jew	9
	Moroccan_Jew	6
	Tunisian_Jew	7
Mozabite	Mozabite	21

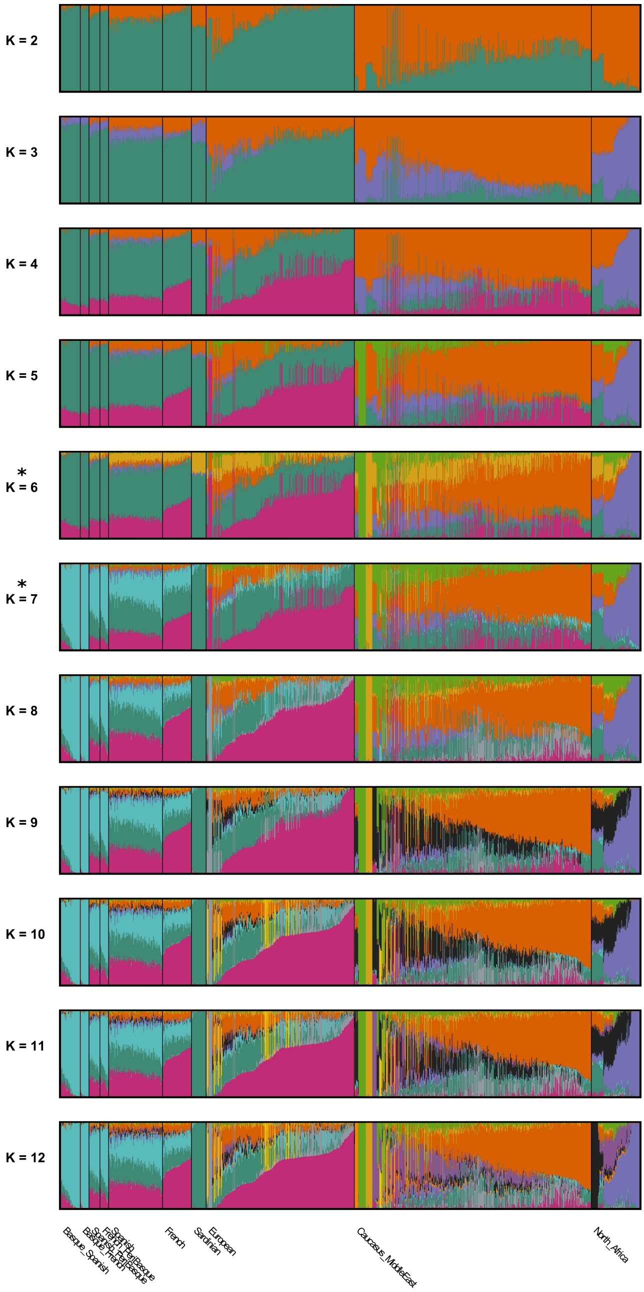


Figure S1. ADMIXTURE results from K=2 to K=12, including all the modern samples from the dataset. The lowest cross-validation error in the analysis was K=7.

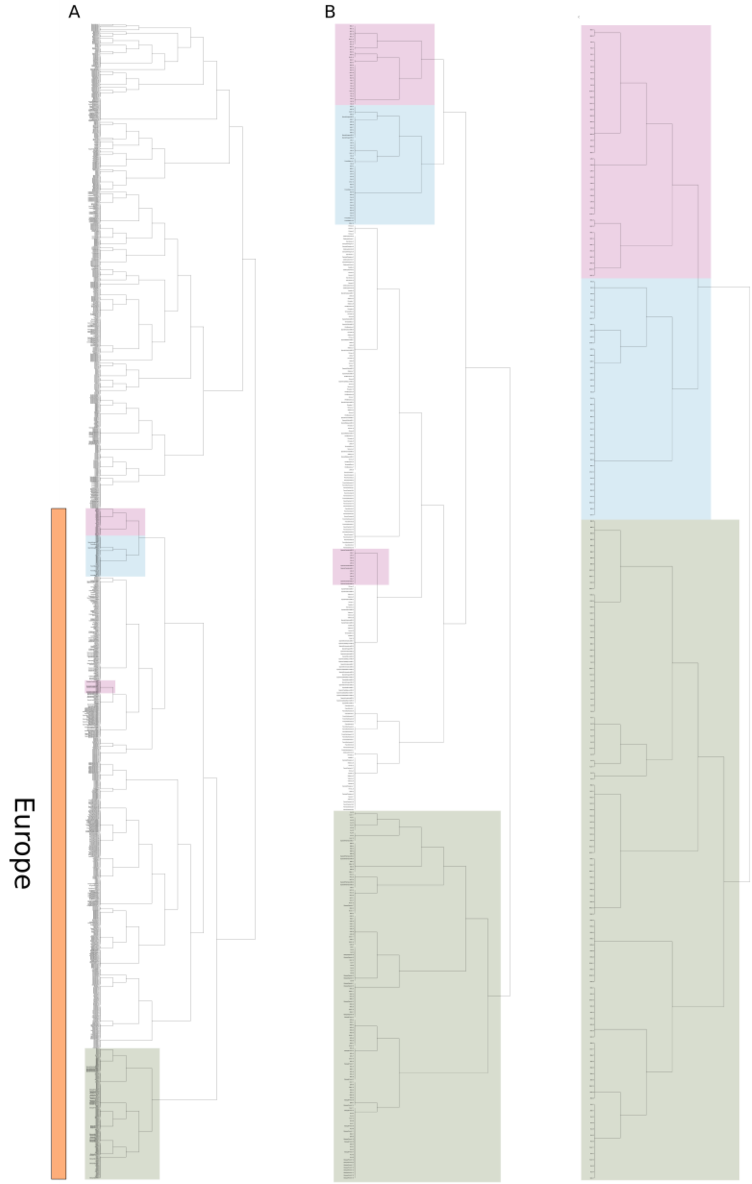


Figure S2. Trees inferred from the different levels of ChromoPainter and fineSTRUCTURE analyses: (A) the whole dataset, (B) Iberian Peninsula and France, plus (C) the Franco-Cantabrian region separately. Our target samples are distributed in similar clusters in the three analyses. The Franco-Cantabrian clusters are highlighted with colors: Basques in green, Western Peri-Basques (Spanish) in pink and Eastern Peri-Basques (French and NAR) in blue.

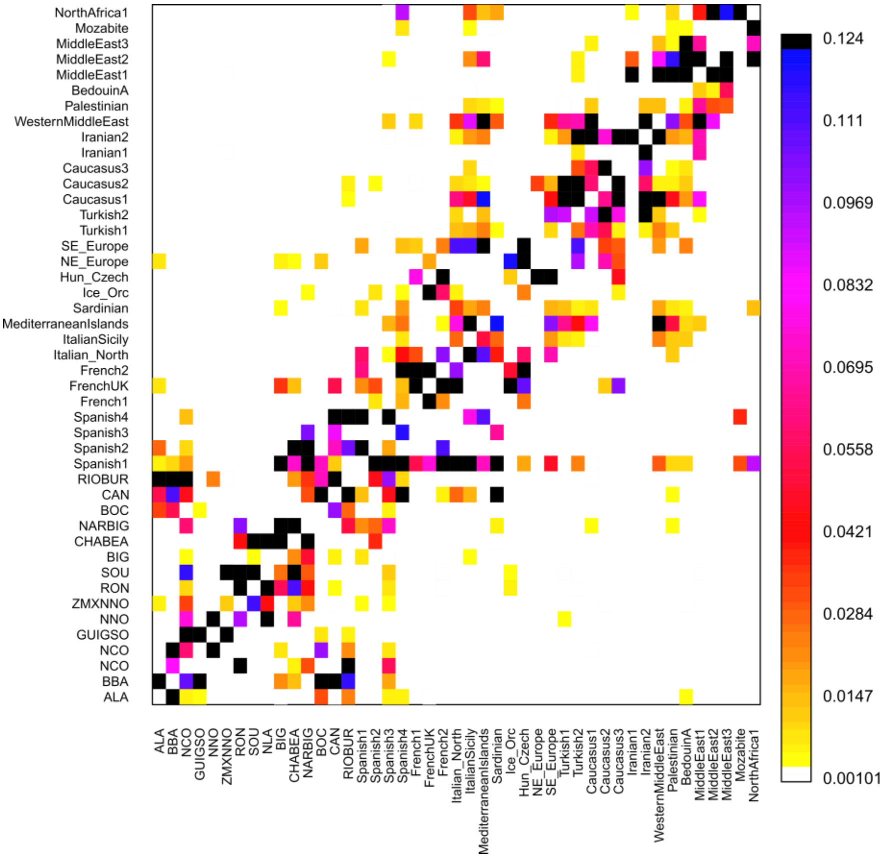


Figure S3. Heatmap showing the inferred ancestry proportions with GLOBETROTTER between the clusters defined from the fineSTRUCTURE tree. The color scale is fitted for the quantile 5 and 95. Donors and recipients are represented by rows and columns, respectively.

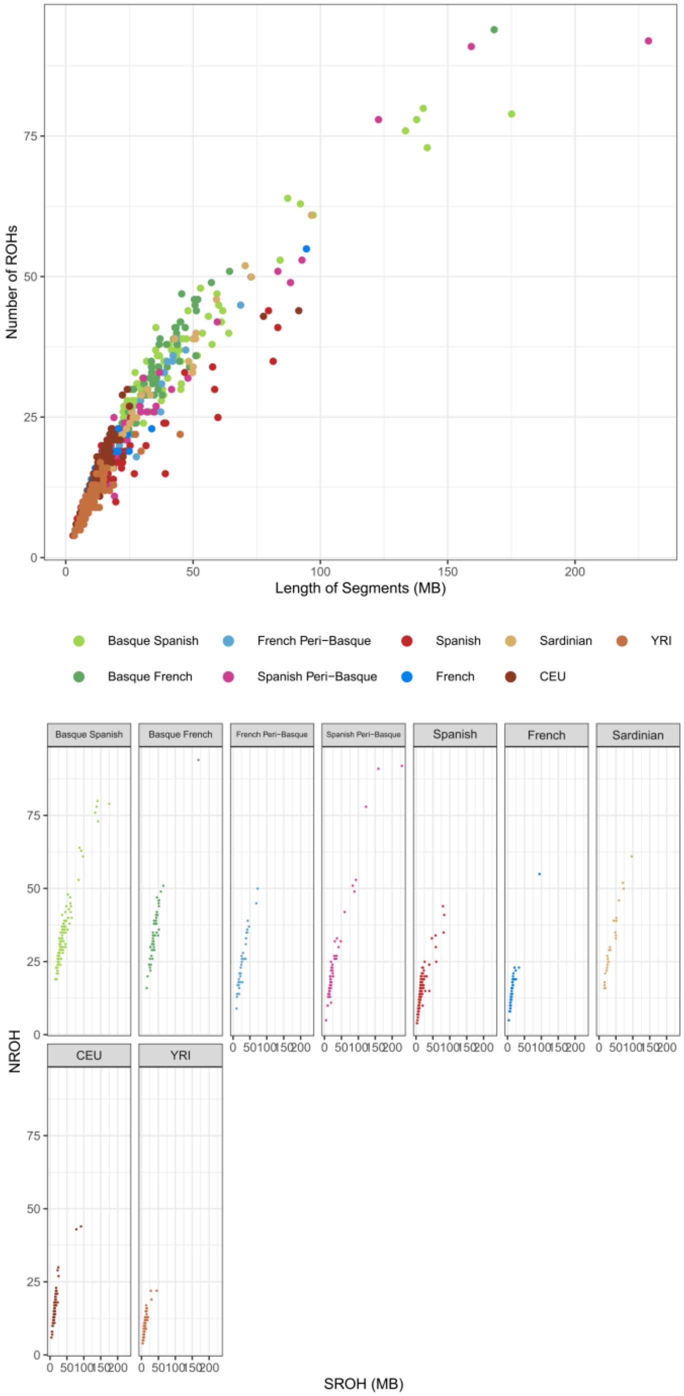


Figure S4. Runs of homozygosity plots: NROH vs SROH. Comparison between the total length and number of runs of homozygosity for each individual sample.

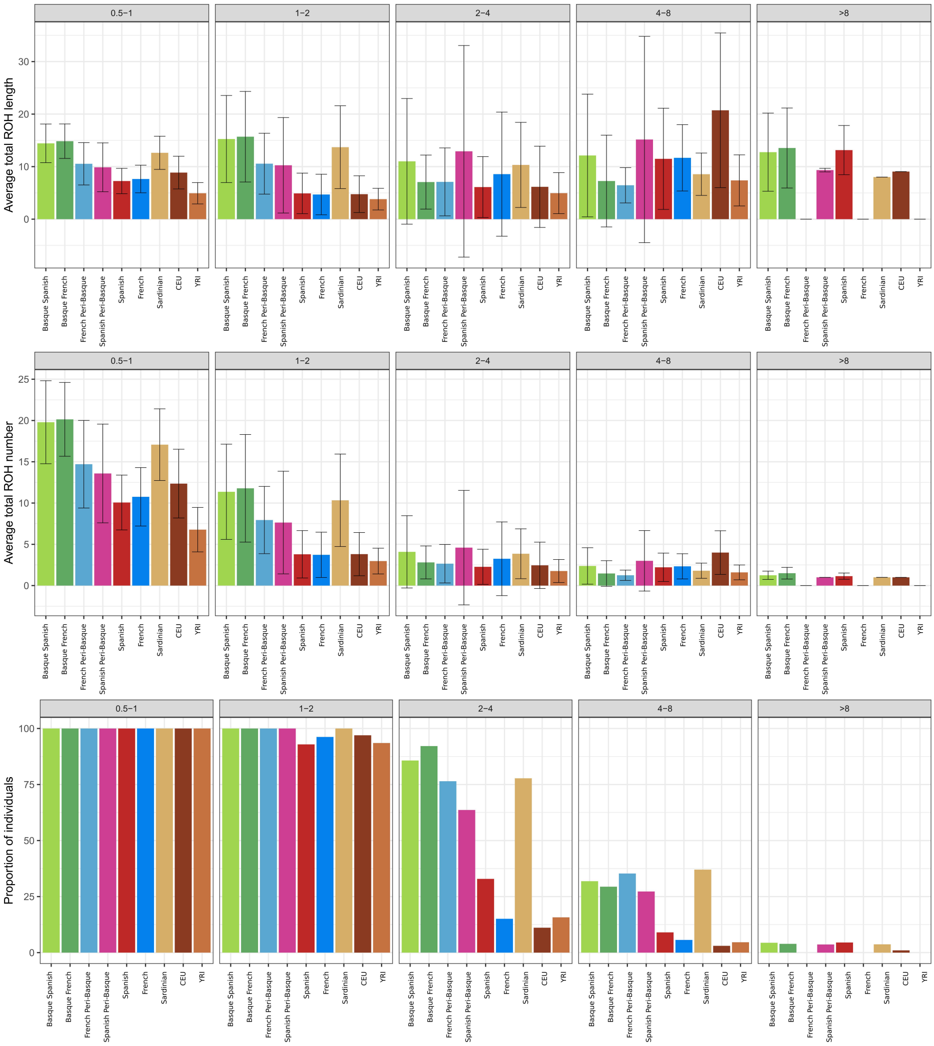


Figure S5. Runs of homozygosity plots: ROH length categories. Average of the total length and number of runs of homozygosity for different length categories. The proportion of the total samples for the different groups represented in each category is also shown in the last figure.

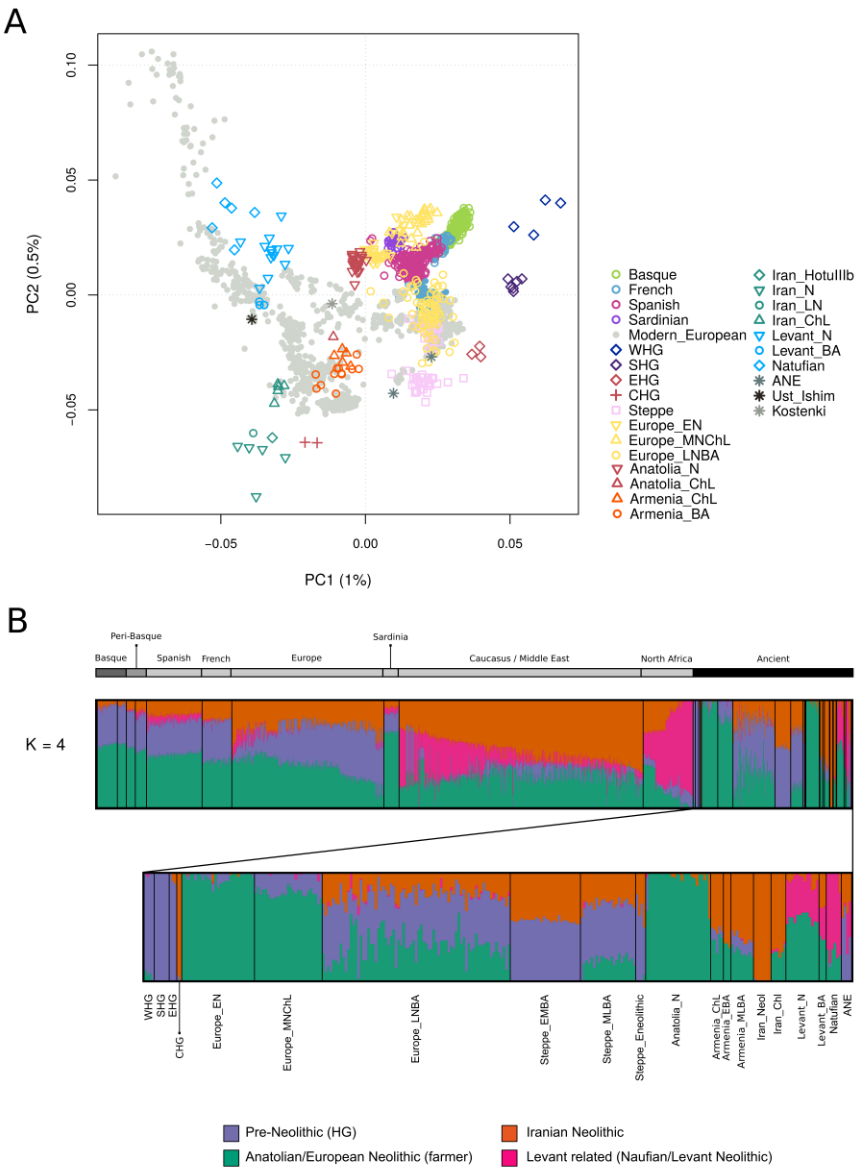
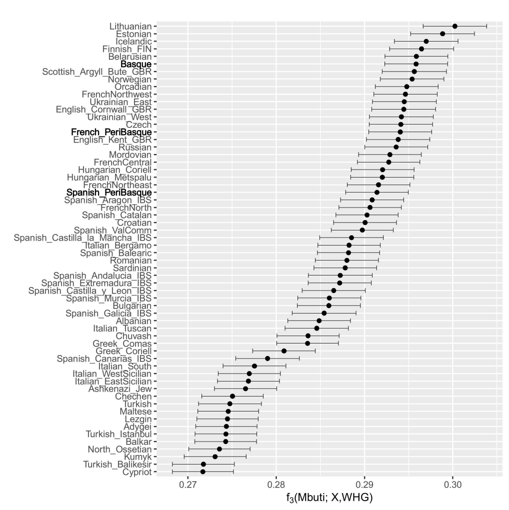
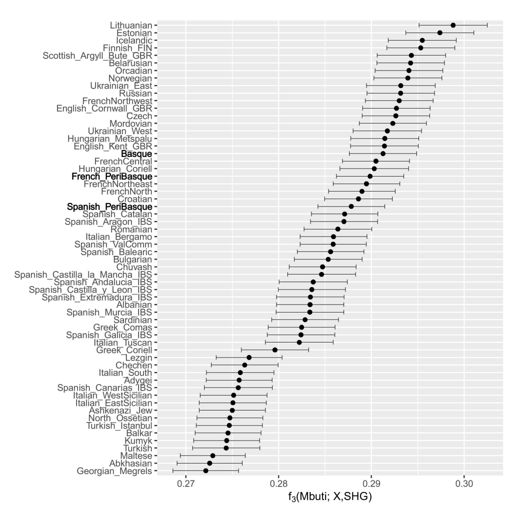
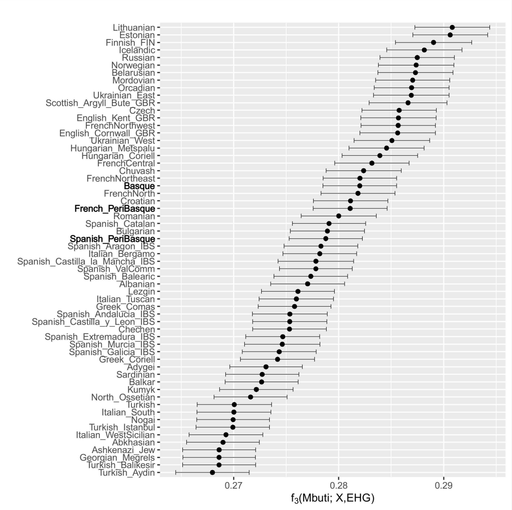
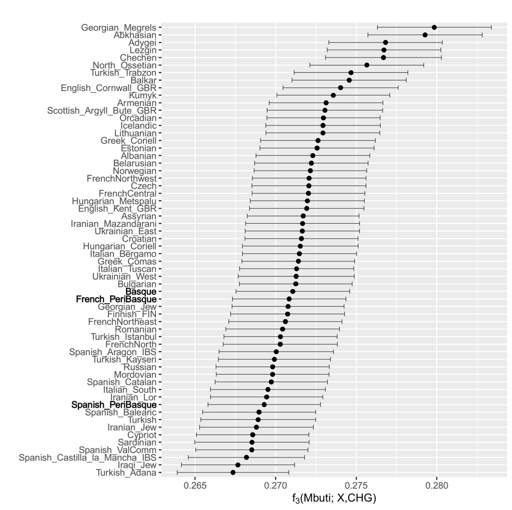


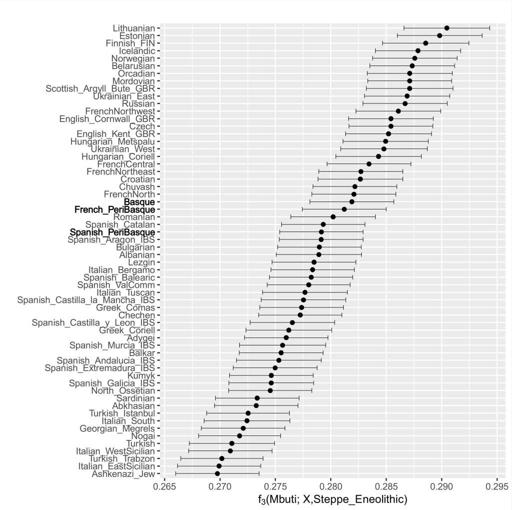
Figure S6. Franco-Cantabrians in the ancient context. (A) Principal component analysis projecting ancient samples. (B) Unsupervised admixture analysis including ancient samples. Both K=6 and K=7 showed the lowest cross validation errors with similar values.

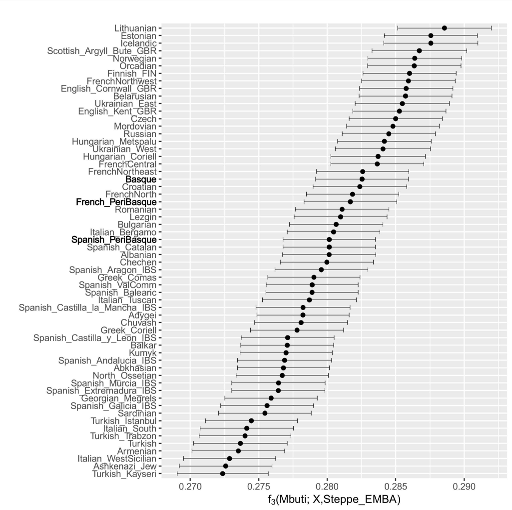

Lithuanian
Estonian
Icelandic
Finnish_FIN
Belarusian
Basque
Scottish_Argyll_Bute_GBR
Norwegian
Orcadian
FrenchNorthwest
Ukrainian_East
English_Cornwall_GBR
Ukrainian_West
Czech
French_PeriBasque
English_Kent_GBR
Russian
Mordovian
FrenchCentral
Hungarian_Coriell
Hungarian_Metspalu
FrenchNortheast
Spanish_PeriBasque
Spanish_Aragon_IBS
FrenchNorth
Spanish_Catalan
Croatian
Spanish_ValComm
Spanish_Castilla_la_Mancha_IBS
Italian_Bergamo
Spanish_Balearic
Romanian
Sardinian
Spanish_Andalucia_IBS
Spanish_Extremadura_IBS
Spanish_Castilla_y_Leon_IBS
Spanish_Murcia_IBS
Bulgarian
Spanish_Galicia_IBS
Albanian
Italian_Tuscan
Chuvash
Greek_Comas
Greek_Coriell
Spanish_Canarias_IBS
Italian_South
Italian_WestSicilian
Italian_EastSicilian
Ashkenazi_Jew
Chechen
Turkish
Maltese
Lezgin
Adygei
Turkish_Istanbul
Balkar
North_Ossetian
Kumyk
Turkish_Balikesir
Cypriot
0.27
0.28
0.29
0.30
f_3(Mbuti; X,WHG)

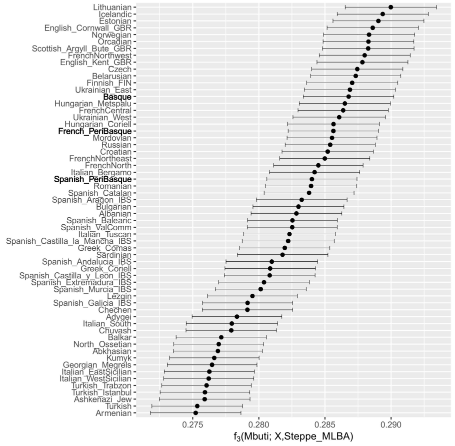

Lithuanian
Estonian
Icelandic
Finnish_FIN
Scottish_Argyll_Bute_GBR
Belarusian
Orcadian
Norwegian
Ukrainian_East
Russian
FrenchNorthwest
English_Cornwall_GBR
Czech
Mordovian
Ukrainian_West
Hungarian_Metspalu
English_Kent_GBR
Basque
FrenchCentral
Hungarian_Coriell
French_PeriBasque
FrenchNortheast
FrenchNorth
Croatian
Spanish_PeriBasque
Spanish_Catalan
Spanish_Aragon_IBS
Romanian
Italian_Bergamo
Spanish_ValComm
Spanish_Balearic
Bulgarian
Chuvash
Spanish_Castilla_la_Mancha_IBS
Spanish_Andalucia_IBS
Spanish_Castilla_y_Leon_IBS
Spanish_Extremadura_IBS
Albanian
Spanish_Murcia_IBS
Sardinian
Greek_Comas
Spanish_Galicia_IBS
Italian_Tuscan
Greek_Coriell
Lezgin
Chechen
Italian_South
Adygei
Spanish_Canarias_IBS
Italian_WestSicilian
Italian_EastSicilian
Ashkenazi_Jew
North_Ossetian
Turkish_Istanbul
Balkar
Kumyk
Turkish
Maltese
Abkhasian
Georgian_Megrels
0.27
0.28
0.29
0.30
f_3(Mbuti; X,SHG)

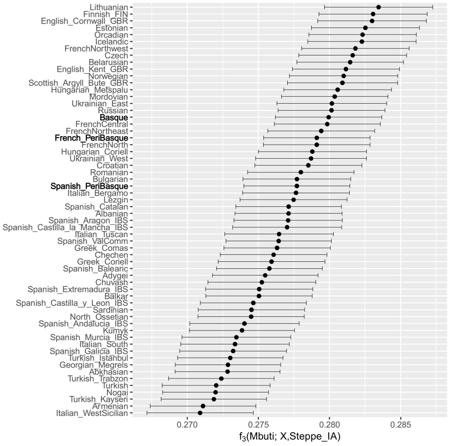

Lithuanian
Estonian
Finnish_FIN
Icelandic
Russian
Norwegian
Belarusian
Mordovian
Orcadian
Ukrainian_East
Scottish_Argyll_Bute_GBR
Czech
English_Kent_GBR
FrenchNorthwest
English_Cornwall_GBR
Ukrainian_West
Hungarian_Metspalu
Hungarian_Coriell
FrenchCentral
Chuvash
FrenchNortheast
Basque
FrenchNorth
Croatian
French_PeriBasque
Romanian
Spanish_Catalan
Bulgarian
Spanish_PeriBasque
Spanish_Aragon_IBS
Italian_Bergamo
Spanish_Castilla_la_Mancha_IBS
Spanish_ValComm
Spanish_Balearic
Albanian
Lezgin
Italian_Tuscan
Greek_Comas
Spanish_Andalucia_IBS
Spanish_Castilla_y_Leon_IBS
Chechen
Spanish_Extremadura_IBS
Spanish_Murcia_IBS
Spanish_Galicia_IBS
Greek_Coriell
Adygei
Sardinian
Balkar
Kumyk
North_Ossetian
Turkish
Italian_South
Nogai
Turkish_Istanbul
Italian_WestSicilian
Abkhasian
Ashkenazi_Jew
Georgian_Megrels
Turkish_Balikesir
Turkish_Aydin
0.27
0.28
0.29
f_3(Mbuti; X,EHG)

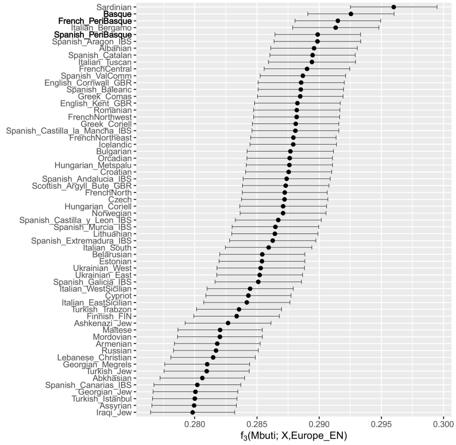

Georgian_Megrels
Abkhasian
Adygei
Lezgin
Chechen
North_Ossetian
Turkish_Trabzon
Balkar
English_Cornwall_GBR
Kumyk
Armenian
Scottish_Argyll_Bute_GBR
Orcadian
Icelandic
Lithuanian
Greek_Coriell
Estonian
Albanian
Belarusian
Norwegian
FrenchNorthwest
Czech
FrenchCentral
Hungarian_Metspalu
English_Kent_GBR
Assyrian
Iranian_Mazandarani
Ukrainian_East
Croatian
Hungarian_Coriell
Italian_Bergamo
Greek_Comas
Italian_Tuscan
Ukrainian_West
Bulgarian
Basque
French_PeriBasque
Georgian_Jew
Finnish_FIN
FrenchNortheast
Romanian
Turkish_Istanbul
FrenchNorth
Spanish_Aragon_IBS
Turkish_Kayseri
Russian
Mordovian
Spanish_Catalan
Italian_South
Iranian_Lor
Spanish_PeriBasque
Spanish_Balearic
Turkish
Iranian_Jew
Cypriot
Sardinian
Spanish_ValComm
Spanish_Castilla_la_Mancha_IBS
Iraqi_Jew
Turkish_Adana
0.265
0.270
0.275
0.280
f_3(Mbuti; X,CHG)

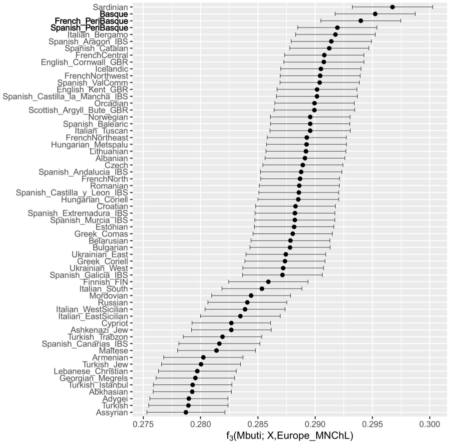

Lithuanian
Estonian
Finnish_FIN
Icelandic
Norwegian
Belarusian
Orcadian
Mordovian
Scottish_Argyll_Bute_GBR
Ukrainian_East
Russian
FrenchNorthwest
English_Cornwall_GBR
Czech
English_Kent_GBR
Hungarian_Metspalu
Ukrainian_West
Hungarian_Coriell
FrenchCentral
FrenchNortheast
Croatian
Chuvash
FrenchNorth
Basque
French_PeriBasque
Romanian
Spanish_Catalan
Spanish_PeriBasque
Spanish_Aragon_IBS
Bulgarian
Albanian
Lezgin
Italian_Bergamo
Spanish_Balearic
Spanish_ValComm
Italian_Tuscan
Spanish_Castilla_la_Mancha_IBS
Greek_Comas
Chechen
Spanish_Castilla_y_Leon_IBS
Greek_Coriell
Adygei
Spanish_Murcia_IBS
Balkar
Spanish_Andalucia_IBS
Spanish_Extremadura_IBS
Kumyk
Spanish_Galicia_IBS
North_Ossetian
Sardinian
Abkhasian
Turkish_Istanbul
Italian_South
Georgian_Megrels
Nogai
Turkish
Italian_WestSicilian
Turkish_Trabzon
Italian_EastSicilian
Ashkenazi_Jew
0.265
0.270
0.275
0.280
0.285
0.290
0.295
f_3(Mbuti; X,Steppe_Eneolithic)

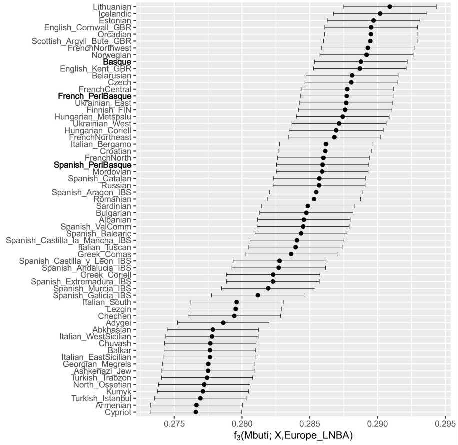

Lithuanian
Estonian
Icelandic
Scottish_Argyll_Bute_GBR
Norwegian
Orcadian
Finnish_FIN
FrenchNorthwest
English_Cornwall_GBR
Belarusian
Ukrainian_East
English_Kent_GBR
Czech
Mordovian
Russian
Hungarian_Metspalu
Ukrainian_West
Hungarian_Coriell
FrenchCentral
FrenchNortheast
Basque
Croatian
FrenchNorth
French_PeriBasque
Romanian
Lezgin
Bulgarian
Italian_Bergamo
Spanish_PeriBasque
Spanish_Catalan
Albanian
Chechen
Spanish_Aragon_IBS
Greek_Comas
Spanish_ValComm
Spanish_Balearic
Italian_Tuscan
Spanish_Castilla_la_Mancha_IBS
Adygei
Chuvash
Greek_Coriell
Spanish_Castilla_y_Leon_IBS
Balkar
Kumyk
Spanish_Andalucia_IBS
Abkhasian
North_Ossetian
Spanish_Murcia_IBS
Spanish_Extremadura_IBS
Georgian_Megrels
Spanish_Galicia_IBS
Sardinian
Turkish_Istanbul
Italian_South
Turkish_Trabzon
Turkish
Armenian
Italian_WestSicilian
Ashkenazi_Jew
Turkish_Kayseri
0.270
0.275
0.280
0.285
0.290
f_3(Mbuti; X,Steppe_EMBA)

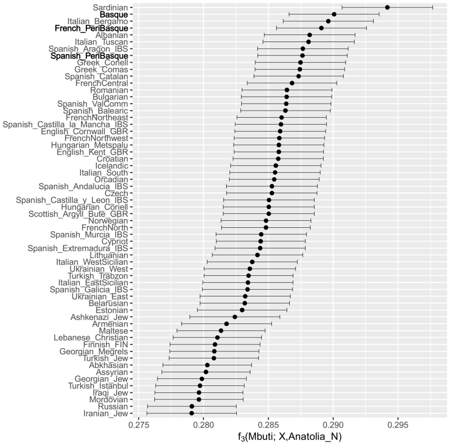

Lithuanian
Icelandic
Estonian
English_Cornwall_GBR
Norwegian
Orcadian
Scottish_Argyll_Bute_GBR
FrenchNorthwest
English_Kent_GBR
Czech
Belarusian
Finnish_FIN
Ukrainian_East
Basque
Hungarian_Metspalu
FrenchCentral
Ukrainian_West
Hungarian_Coriell
French_PeriBasque
Mordovian
Russian
Croatian
FrenchNortheast
FrenchNorth
Italian_Bergamo
Spanish_PeriBasque
Romanian
Spanish_Catalan
Spanish_Aragon_IBS
Bulgarian
Albanian
Spanish_Balearic
Spanish_ValComm
Italian_Tuscan
Spanish_Castilla_la_Mancha_IBS
Greek_Comas
Sardinian
Spanish_Andalucia_IBS
Greek_Coriell
Spanish_Castilla_y_Leon_IBS
Spanish_Extremadura_IBS
Spanish_Murcia_IBS
Lezgin
Spanish_Galicia_IBS
Chechen
Adygei
Italian_South
Chuvash
Balkar
North_Ossetian
Abkhasian
Kumyk
Georgian_Megrels
Italian_EastSicilian
Italian_WestSicilian
Turkish_Trabzon
Turkish_Istanbul
Ashkenazi_Jew
Turkish
Armenian
0.275
0.280
0.285
0.290
f_3(Mbuti; X,Steppe_MLBA)

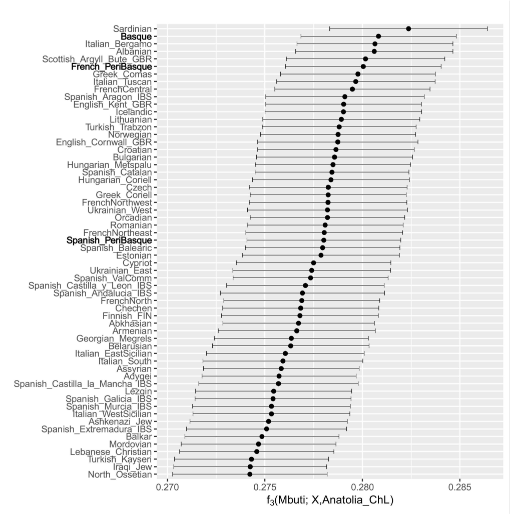

Lithuanian
Finnish_FIN
English_Cornwall_GBR
Estonian
Orcadian
Icelandic
FrenchNorthwest
Czech
Belarusian
English_Kent_GBR
Norwegian
Scottish_Argyll_Bute_GBR
Hungarian_Metspalu
Mordovian
Ukrainian_East
Russian
Basque
FrenchCentral
FrenchNortheast
French_PeriBasque
FrenchNorth
Hungarian_Coriell
Ukrainian_West
Croatian
Romanian
Bulgarian
Spanish_PeriBasque
Italian_Bergamo
Lezgin
Spanish_Catalan
Albanian
Spanish_Aragon_IBS
Spanish_Castilla_la_Mancha_IBS
Italian_Tuscan
Spanish_ValComm
Greek_Comas
Chechen
Greek_Coriell
Spanish_Balearic
Adygei
Chuvash
Spanish_Extremadura_IBS
Balkar
Spanish_Castilla_y_Leon_IBS
Sardinian
North_Ossetian
Spanish_Andalucia_IBS
Kumyk
Spanish_Murcia_IBS
Italian_South
Spanish_Galicia_IBS
Turkish_Istanbul
Georgian_Megrels
Abkhasian
Turkish_Trabzon
Turkish
Nogai
Turkish_Kayseri
Armenian
Italian_WestSicilian
0.270
0.275
0.280
0.285
f_3(Mbuti; X,Steppe_IA)

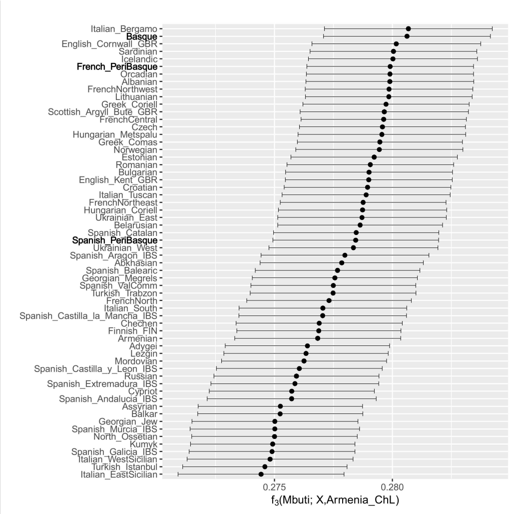

Sardinian
Basque
French_PeriBasque
Italian_Bergamo
Spanish_PeriBasque
Spanish_Aragon_IBS
Albanian
Spanish_Catalan
Italian_Tuscan
FrenchCentral
Spanish_ValComm
English_Cornwall_GBR
Spanish_Balearic
Greek_Comas
English_Kent_GBR
Romanian
FrenchNorthwest
Greek_Coriell
Spanish_Castilla_la_Mancha_IBS
FrenchNortheast
Icelandic
Bulgarian
Orcadian
Hungarian_Metspalu
Croatian
Spanish_Andalucia_IBS
Scottish_Argyll_Bute_GBR
FrenchNorth
Czech
Hungarian_Coriell
Norwegian
Spanish_Castilla_y_Leon_IBS
Spanish_Murcia_IBS
Lithuanian
Spanish_Extremadura_IBS
Italian_South
Belarusian
Estonian
Ukrainian_West
Ukrainian_East
Spanish_Galicia_IBS
Italian_WestSicilian
Cypriot
Italian_EastSicilian
Turkish_Trabzon
Finnish_FIN
Ashkenazi_Jew
Maltese
Mordovian
Armenian
Russian
Lebanese_Christian
Georgian_Megrels
Turkish_Jew
Abkhasian
Spanish_Canarias_IBS
Georgian_Jew
Turkish_Istanbul
Assyrian
Iraqi_Jew
0.280
0.285
0.290
0.295
0.300
f_3(Mbuti; X,Europe_EN)

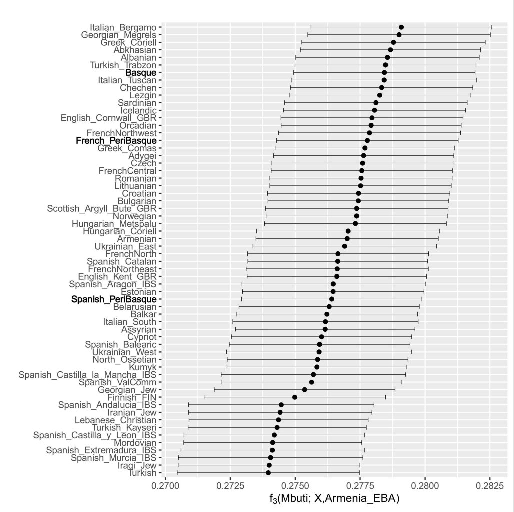

Sardinian
Basque
French_PeriBasque
Spanish_PeriBasque
Italian_Bergamo
Spanish_Aragon_IBS
Spanish_Catalan
FrenchCentral
English_Cornwall_GBR
Icelandic
FrenchNorthwest
Spanish_ValComm
English_Kent_GBR
Spanish_Castilla_la_Mancha_IBS
Orcadian
Scottish_Argyll_Bute_GBR
Norwegian
Spanish_Balearic
Italian_Tuscan
FrenchNortheast
Hungarian_Metspalu
Lithuanian
Albanian
Czech
Spanish_Andalucia_IBS
FrenchNorth
Romanian
Spanish_Castilla_y_Leon_IBS
Hungarian_Coriell
Croatian
Spanish_Extremadura_IBS
Spanish_Murcia_IBS
Estonian
Greek_Comas
Belarusian
Bulgarian
Ukrainian_East
Greek_Coriell
Ukrainian_West
Spanish_Galicia_IBS
Finnish_FIN
Italian_South
Mordovian
Russian
Italian_WestSicilian
Italian_EastSicilian
Cypriot
Ashkenazi_Jew
Turkish_Trabzon
Spanish_Canarias_IBS
Maltese
Armenian
Turkish_Jew
Lebanese_Christian
Georgian_Megrels
Turkish_Istanbul
Abkhasian
Adygei
Turkish
Assyrian
0.275
0.280
0.285
0.290
0.295
0.300
f_3(Mbuti; X,Europe_MNChL)

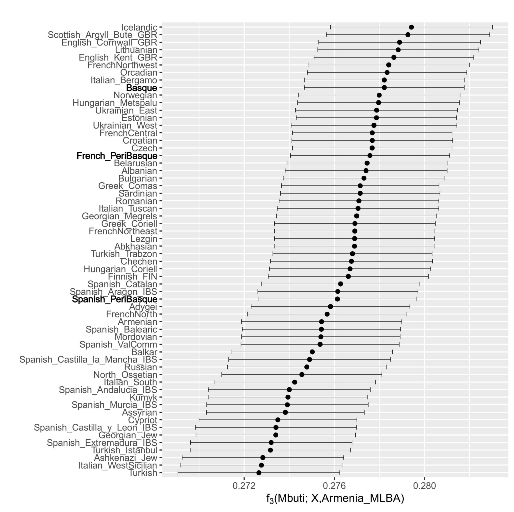

Lithuanian
Icelandic
Estonian
English_Cornwall_GBR
Orcadian
Scottish_Argyll_Bute_GBR
FrenchNorthwest
Norwegian
Basque
English_Kent_GBR
Belarusian
Czech
FrenchCentral
French_PeriBasque
Ukrainian_East
Finnish_FIN
Hungarian_Metspalu
Ukrainian_West
Hungarian_Coriell
FrenchNortheast
Italian_Bergamo
Croatian
FrenchNorth
Spanish_PeriBasque
Mordovian
Spanish_Catalan
Russian
Spanish_Aragon_IBS
Romanian
Sardinian
Bulgarian
Albanian
Spanish_ValComm
Spanish_Balearic
Spanish_Castilla_la_Mancha_IBS
Italian_Tuscan
Greek_Comas
Spanish_Castilla_y_Leon_IBS
Spanish_Andalucia_IBS
Greek_Coriell
Spanish_Extremadura_IBS
Spanish_Murcia_IBS
Spanish_Galicia_IBS
Italian_South
Lezgin
Chechen
Adygei
Abkhasian
Italian_WestSicilian
Chuvash
Balkar
Italian_EastSicilian
Georgian_Megrels
Ashkenazi_Jew
Turkish_Trabzon
North_Ossetian
Kumyk
Turkish_Istanbul
Armenian
Cypriot
0.275
0.280
0.285
0.290
0.295
f_3(Mbuti; X,Europe_LNBA)

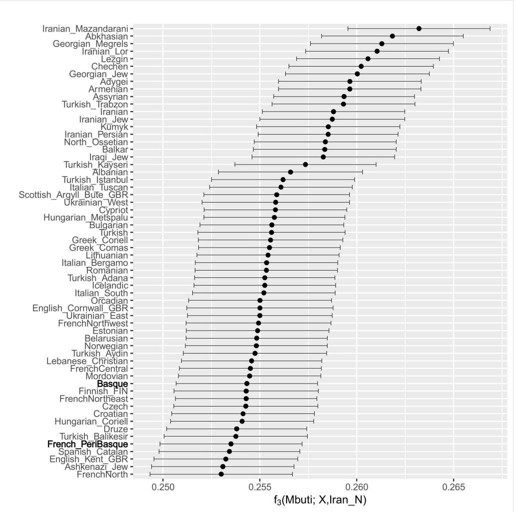

Sardinian
Basque
Italian_Bergamo
French_PeriBasque
Albanian
Italian_Tuscan
Spanish_Aragon_IBS
Spanish_PeriBasque
Greek_Coriell
Greek_Comas
Spanish_Catalan
FrenchCentral
Romanian
Bulgarian
Spanish_ValComm
Spanish_Balearic
FrenchNortheast
Spanish_Castilla_la_Mancha_IBS
English_Cornwall_GBR
FrenchNorthwest
Hungarian_Metspalu
English_Kent_GBR
Croatian
Icelandic
Italian_South
Orcadian
Spanish_Andalucia_IBS
Czech
Spanish_Castilla_y_Leon_IBS
Hungarian_Coriell
Scottish_Argyll_Bute_GBR
Norwegian
FrenchNorth
Spanish_Murcia_IBS
Cypriot
Spanish_Extremadura_IBS
Lithuanian
Italian_WestSicilian
Ukrainian_West
Turkish_Trabzon
Italian_EastSicilian
Spanish_Galicia_IBS
Ukrainian_East
Belarusian
Estonian
Ashkenazi_Jew
Armenian
Maltese
Lebanese_Christian
Finnish_FIN
Georgian_Megrels
Turkish_Jew
Abkhasian
Assyrian
Georgian_Jew
Turkish_Istanbul
Iraqi_Jew
Mordovian
Russian
Iranian_Jew
0.275
0.280
0.285
0.290
0.295
f_3(Mbuti; X,Anatolia_N)

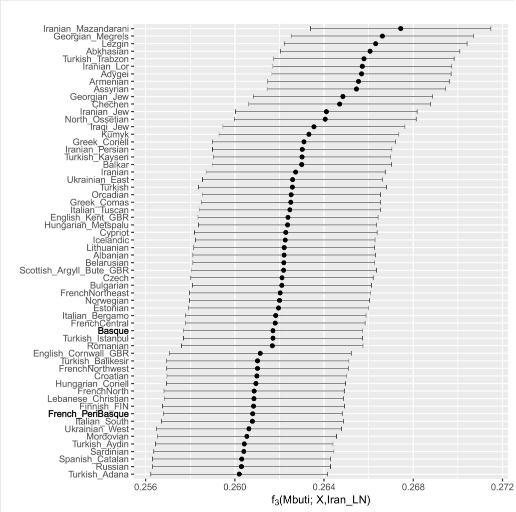

Sardinian
Basque
Italian_Bergamo
Albanian
Scottish_Argyll_Bute_GBR
French_PeriBasque
Greek_Comas
Italian_Tuscan
FrenchCentral
Spanish_Aragon_IBS
English_Kent_GBR
Icelandic
Lithuanian
Turkish_Trabzon
Norwegian
English_Cornwall_GBR
Croatian
Bulgarian
Hungarian_Metspalu
Spanish_Catalan
Hungarian_Coriell
Czech
Greek_Coriell
FrenchNorthwest
Ukrainian_West
Orcadian
Romanian
FrenchNortheast
Spanish_PeriBasque
Spanish_Balearic
Estonian
Cypriot
Ukrainian_East
Spanish_ValComm
Spanish_Castilla_y_Leon_IBS
Spanish_Andalucia_IBS
FrenchNorth
Chechen
Finnish_FIN
Abkhasian
Armenian
Georgian_Megrels
Belarusian
Italian_EastSicilian
Italian_South
Assyrian
Adygei
Spanish_Castilla_la_Mancha_IBS
Lezgin
Spanish_Galicia_IBS
Spanish_Murcia_IBS
Italian_WestSicilian
Ashkenazi_Jew
Spanish_Extremadura_IBS
Balkar
Mordovian
Lebanese_Christian
Turkish_Kayseri
Iraqi_Jew
North_Ossetian
0.270
0.275
0.280
0.285
f_3(Mbuti; X,Anatolia_ChL)

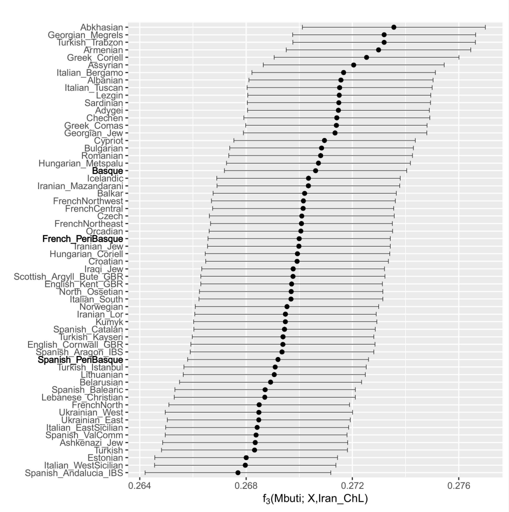

Italian_Bergamo
Basque
English_Cornwall_GBR
Sardinian
Icelandic
French_PeriBasque
Orcadian
Albanian
FrenchNorthwest
Lithuanian
Greek_Coriell
Scottish_Argyll_Bute_GBR
FrenchCentral
Czech
Hungarian_Metspalu
Greek_Comas
Norwegian
Estonian
Romanian
Bulgarian
English_Kent_GBR
Croatian
Italian_Tuscan
FrenchNortheast
Hungarian_Coriell
Ukrainian_East
Belarusian
Spanish_Catalan
Spanish_PeriBasque
Ukrainian_West
Spanish_Aragon_IBS
Abkhasian
Spanish_Balearic
Georgian_Megrels
Spanish_ValComm
Turkish_Trabzon
FrenchNorth
Italian_South
Spanish_Castilla_la_Mancha_IBS
Chechen
Finnish_FIN
Armenian
Adygei
Lezgin
Mordovian
Spanish_Castilla_y_Leon_IBS
Russian
Spanish_Extremadura_IBS
Cypriot
Spanish_Andalucia_IBS
Assyrian
Balkar
Georgian_Jew
Spanish_Murcia_IBS
North_Ossetian
Kumyk
Spanish_Galicia_IBS
Italian_WestSicilian
Turkish_Istanbul
Italian_EastSicilian
0.275
0.280
f_3(Mbuti; X,Armenia_ChL)

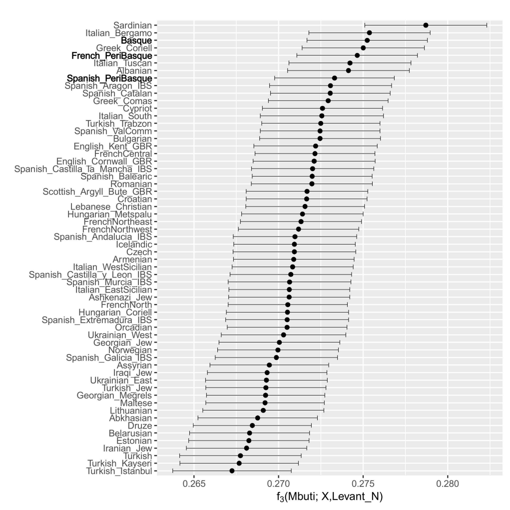

Italian_Bergamo
Georgian_Megrels
Greek_Coriell
Abkhasian
Albanian
Turkish_Trabzon
Basque
Italian_Tuscan
Chechen
Lezgin
Sardinian
Icelandic
English_Cornwall_GBR
Orcadian
FrenchNorthwest
French_PeriBasque
Greek_Comas
Adygei
Czech
FrenchCentral
Romanian
Lithuanian
Croatian
Bulgarian
Scottish_Argyll_Bute_GBR
Norwegian
Hungarian_Metspalu
Hungarian_Coriell
Armenian
Ukrainian_East
FrenchNorth
Spanish_Catalan
FrenchNortheast
English_Kent_GBR
Spanish_Aragon_IBS
Estonian
Spanish_PeriBasque
Belarusian
Balkar
Italian_South
Assyrian
Cypriot
Spanish_Balearic
Ukrainian_West
North_Ossetian
Kumyk
Spanish_Castilla_la_Mancha_IBS
Spanish_ValComm
Georgian_Jew
Finnish_FIN
Spanish_Andalucia_IBS
Iranian_Jew
Lebanese_Christian
Turkish_Kayseri
Spanish_Castilla_y_Leon_IBS
Mordovian
Spanish_Extremadura_IBS
Spanish_Murcia_IBS
Iraqi_Jew
Turkish
0.2700
0.2725
0.2750
0.2775
0.2800
0.2825
f_3(Mbuti; X,Armenia_EBA)

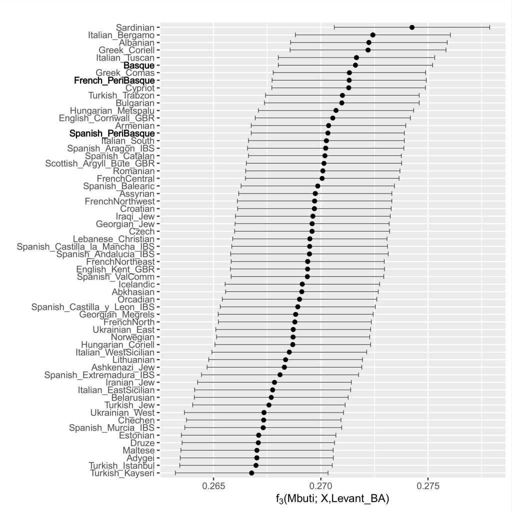

Icelandic
Scottish_Argyll_Bute_GBR
English_Cornwall_GBR
Lithuanian
English_Kent_GBR
FrenchNorthwest
Orcadian
Italian_Bergamo
Basque
Norwegian
Hungarian_Metspalu
Ukrainian_East
Estonian
Ukrainian_West
FrenchCentral
Croatian
Czech
French_PeriBasque
Belarusian
Albanian
Bulgarian
Greek_Comas
Sardinian
Romanian
Italian_Tuscan
Georgian_Megrels
Greek_Coriell
FrenchNortheast
Lezgin
Abkhasian
Turkish_Trabzon
Chechen
Hungarian_Coriell
Finnish_FIN
Spanish_Catalan
Spanish_Aragon_IBS
Spanish_PeriBasque
Adygei
FrenchNorth
Armenian
Spanish_Balearic
Mordovian
Spanish_ValComm
Balkar
Spanish_Castilla_la_Mancha_IBS
Russian
North_Ossetian
Italian_South
Spanish_Andalucia_IBS
Kumyk
Spanish_Murcia_IBS
Assyrian
Cypriot
Spanish_Castilla_y_Leon_IBS
Georgian_Jew
Spanish_Extremadura_IBS
Turkish_Istanbul
Ashkenazi_Jew
Italian_WestSicilian
Turkish
0.272
0.276
0.280
f_3(Mbuti; X,Armenia_MLBA)

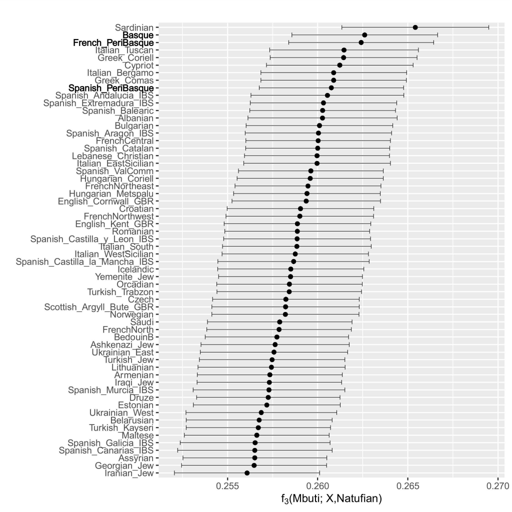

Iranian_Mazandarani
Abkhasian
Georgian_Megrels
Iranian_Lor
Lezgin
Chechen
Georgian_Jew
Adygei
Armenian
Assyrian
Turkish_Trabzon
Iranian
Iranian_Jew
Kumyk
Iranian_Persian
North_Ossetian
Balkar
Iraqi_Jew
Turkish_Kayseri
Albanian
Turkish_Istanbul
Italian_Tuscan
Scottish_Argyll_Bute_GBR
Ukrainian_West
Cypriot
Hungarian_Metspalu
Bulgarian
Turkish
Greek_Coriell
Greek_Comas
Lithuanian
Italian_Bergamo
Romanian
Turkish_Adana
Icelandic
Italian_South
Orcadian
English_Cornwall_GBR
Ukrainian_East
FrenchNorthwest
Estonian
Belarusian
Norwegian
Turkish_Aydin
Lebanese_Christian
FrenchCentral
Mordovian
Basque
Finnish_FIN
FrenchNortheast
Czech
Croatian
Hungarian_Coriell
Druze
Turkish_Balikesir
French_PeriBasque
Spanish_Catalan
English_Kent_GBR
Ashkenazi_Jew
FrenchNorth
0.250
0.255
0.260
0.265
f_3(Mbuti; X,Iran_N)

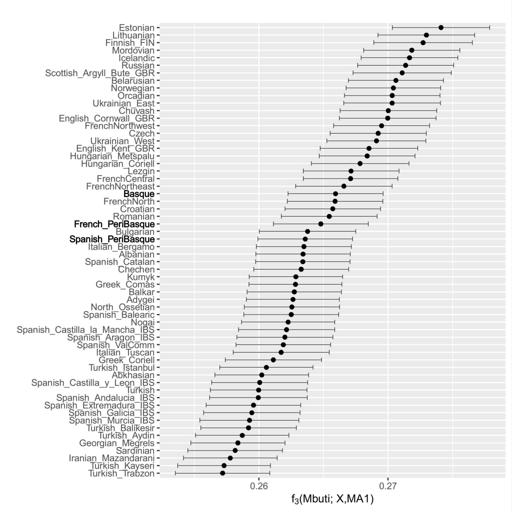

Iranian_Mazandarani
Georgian_Megrels
Lezgin
Abkhasian
Turkish_Trabzon
Iranian_Lor
Adygei
Armenian
Assyrian
Georgian_Jew
Chechen
Iranian_Jew
North_Ossetian
Iraqi_Jew
Kumyk
Greek_Coriell
Iranian_Persian
Turkish_Kayseri
Balkar
Iranian
Ukrainian_East
Turkish
Orcadian
Greek_Comas
Italian_Tuscan
English_Kent_GBR
Hungarian_Metspalu
Cypriot
Icelandic
Lithuanian
Albanian
Belarusian
Scottish_Argyll_Bute_GBR
Czech
Bulgarian
FrenchNortheast
Norwegian
Estonian
Italian_Bergamo
FrenchCentral
Basque
Turkish_Istanbul
Romanian
English_Cornwall_GBR
Turkish_Balikesir
FrenchNorthwest
Croatian
Hungarian_Coriell
FrenchNorth
Lebanese_Christian
Finnish_FIN
French_PeriBasque
Italian_South
Ukrainian_West
Mordovian
Turkish_Aydin
Sardinian
Spanish_Catalan
Russian
Turkish_Adana
0.256
0.260
0.264
0.268
0.272
f_3(Mbuti; X,Iran_LN)

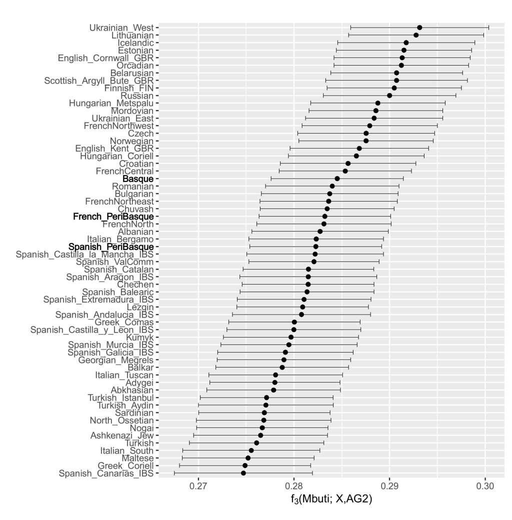

Abkhasian
Georgian_Megrels
Turkish_Trabzon
Armenian
Greek_Coriell
Assyrian
Italian_Bergamo
Albanian
Italian_Tuscan
Lezgin
Sardinian
Adygei
Chechen
Greek_Comas
Georgian_Jew
Cypriot
Bulgarian
Romanian
Hungarian_Metspalu
Basque
Icelandic
Iranian_Mazandarani
Balkar
FrenchNorthwest
FrenchCentral
Czech
FrenchNortheast
Orcadian
French_PeriBasque
Iranian_Jew
Hungarian_Coriell
Croatian
Iraqi_Jew
Scottish_Argyll_Bute_GBR
English_Kent_GBR
North_Ossetian
Italian_South
Norwegian
Iranian_Lor
Kumyk
Spanish_Catalan
Turkish_Kayseri
English_Cornwall_GBR
Spanish_Aragon_IBS
Spanish_PeriBasque
Turkish_Istanbul
Lithuanian
Belarusian
Spanish_Balearic
Lebanese_Christian
FrenchNorth
Ukrainian_West
Ukrainian_East
Italian_EastSicilian
Spanish_ValComm
Ashkenazi_Jew
Turkish
Estonian
Italian_WestSicilian
Spanish_Andalucia_IBS
0.264
0.268
0.272
0.276
f$_3$(Mbuti; X,Iran_ChL)


Sardinian
Italian_Bergamo
Basque
Greek_Coriell
French_PeriBasque
Italian_Tuscan
Albanian
Spanish_PeriBasque
Spanish_Aragon_IBS
Spanish_Catalan
Greek_Comas
Cypriot
Italian_South
Turkish_Trabzon
Spanish_ValComm
Bulgarian
English_Kent_GBR
FrenchCentral
English_Cornwall_GBR
Spanish_Castilla_la_Mancha_IBS
Spanish_Balearic
Romanian
Scottish_Argyll_Bute_GBR
Croatian
Lebanese_Christian
Hungarian_Metspalu
FrenchNortheast
FrenchNorthwest
Spanish_Andalucia_IBS
Icelandic
Czech
Armenian
Italian_WestSicilian
Spanish_Castilla_y_Leon_IBS
Spanish_Murcia_IBS
Italian_EastSicilian
Ashkenazi_Jew
FrenchNorth
Hungarian_Coriell
Spanish_Extremadura_IBS
Orcadian
Ukrainian_West
Georgian_Jew
Norwegian
Spanish_Galicia_IBS
Assyrian
Iraqi_Jew
Ukrainian_East
Turkish_Jew
Georgian_Megrels
Maltese
Lithuanian
Abkhasian
Druze
Belarusian
Estonian
Iranian_Jew
Turkish
Turkish_Kayseri
Turkish_Istanbul
0.265
0.270
0.275
0.280
f$_3$(Mbuti; X,Levant_N)


Sardinian
Italian_Bergamo
Albanian
Greek_Coriell
Italian_Tuscan
Basque
Greek_Comas
French_PeriBasque
Cypriot
Turkish_Trabzon
Bulgarian
Hungarian_Metspalu
English_Cornwall_GBR
Armenian
Spanish_PeriBasque
Italian_South
Spanish_Aragon_IBS
Spanish_Catalan
Scottish_Argyll_Bute_GBR
Romanian
FrenchCentral
Spanish_Balearic
Assyrian
FrenchNorthwest
Croatian
Iraqi_Jew
Georgian_Jew
Czech
Lebanese_Christian
Spanish_Castilla_la_Mancha_IBS
Spanish_Andalucia_IBS
FrenchNortheast
English_Kent_GBR
Spanish_ValComm
Icelandic
Abkhasian
Orcadian
Spanish_Castilla_y_Leon_IBS
Georgian_Megrels
FrenchNorth
Ukrainian_East
Norwegian
Hungarian_Coriell
Italian_WestSicilian
Lithuanian
Ashkenazi_Jew
Spanish_Extremadura_IBS
Iranian_Jew
Italian_EastSicilian
Belarusian
Turkish_Jew
Ukrainian_West
Chechen
Spanish_Murcia_IBS
Estonian
Druze
Maltese
Adygei
Turkish_Istanbul
Turkish_Kayseri
0.265
0.270
0.275
f$_3$(Mbuti; X,Levant_BA)


Sardinian
Basque
French_PeriBasque
Italian_Tuscan
Greek_Coriell
Cypriot
Italian_Bergamo
Greek_Comas
Spanish_PeriBasque
Spanish_Andalucia_IBS
Spanish_Extremadura_IBS
Spanish_Balearic
Albanian
Bulgarian
Spanish_Aragon_IBS
FrenchCentral
Spanish_Catalan
Lebanese_Christian
Italian_EastSicilian
Spanish_ValComm
Hungarian_Coriell
FrenchNortheast
Hungarian_Metspalu
English_Cornwall_GBR
Croatian
FrenchNorthwest
English_Kent_GBR
Romanian
Spanish_Castilla_y_Leon_IBS
Italian_South
Italian_WestSicilian
Spanish_Castilla_la_Mancha_IBS
Icelandic
Yemenite_Jew
Orcadian
Turkish_Trabzon
Czech
Scottish_Argyll_Bute_GBR
Norwegian
Saudi
FrenchNorth
BedouinB
Ashkenazi_Jew
Ukrainian_East
Turkish_Jew
Lithuanian
Armenian
Iraqi_Jew
Spanish_Murcia_IBS
Druze
Estonian
Ukrainian_West
Belarusian
Turkish_Kayseri
Maltese
Spanish_Galicia_IBS
Spanish_Canarias_IBS
Assyrian
Georgian_Jew
Iranian_Jew
0.255
0.260
0.265
0.270
f$_3$(Mbuti; X,Natufian)


Estonian
Lithuanian
Finnish_FIN
Mordovian
Icelandic
Russian
Scottish_Argyll_Bute_GBR
Belarusian
Norwegian
Orcadian
Ukrainian_East
Chuvash
English_Cornwall_GBR
FrenchNorthwest
Czech
Ukrainian_West
English_Kent_GBR
Hungarian_Metspalu
Hungarian_Coriell
Lezgin
FrenchCentral
FrenchNortheast
Basque
FrenchNorth
Croatian
Romanian
French_PeriBasque
Bulgarian
Spanish_PeriBasque
Italian_Bergamo
Albanian
Spanish_Catalan
Chechen
Kumyk
Greek_Comas
Balkar
Adygei
North_Ossetian
Spanish_Balearic
Nogai
Spanish_Castilla_la_Mancha_IBS
Spanish_Aragon_IBS
Spanish_ValComm
Italian_Tuscan
Greek_Coriell
Turkish_Istanbul
Abkhasian
Spanish_Castilla_y_Leon_IBS
Turkish
Spanish_Andalucia_IBS
Spanish_Extremadura_IBS
Spanish_Galicia_IBS
Spanish_Murcia_IBS
Turkish_Balikesir
Turkish_Aydin
Georgian_Megrels
Sardinian
Iranian_Mazandarani
Turkish_Kayseri
Turkish_Trabzon
0.260
0.270
f$_3$(Mbuti; X,MA1)


Ukrainian_West
Lithuanian
Icelandic
Estonian
English_Cornwall_GBR
Orcadian
Belarusian
Scottish_Argyll_Bute_GBR
Finnish_FIN
Russian
Hungarian_Metspalu
Mordovian
Ukrainian_East
FrenchNorthwest
Czech
Norwegian
English_Kent_GBR
Hungarian_Coriell
Croatian
FrenchCentral
Basque
Romanian
Bulgarian
FrenchNortheast
Chuvash
French_PeriBasque
FrenchNorth
Albanian
Italian_Bergamo
Spanish_PeriBasque
Spanish_Castilla_la_Mancha_IBS
Spanish_ValComm
Spanish_Catalan
Spanish_Aragon_IBS
Chechen
Spanish_Balearic
Spanish_Extremadura_IBS
Lezgin
Spanish_Andalucia_IBS
Greek_Comas
Spanish_Castilla_y_Leon_IBS
Kumyk
Spanish_Murcia_IBS
Spanish_Galicia_IBS
Georgian_Megrels
Balkar
Italian_Tuscan
Adygei
Abkhasian
Turkish_Istanbul
Turkish_Aydin
Sardinian
North_Ossetian
Nogai
Ashkenazi_Jew
Turkish
Italian_South
Maltese
Greek_Coriell
Spanish_Canarias_IBS
0.27
0.28
0.29
0.30
f$_3$(Mbuti; X,AG2)

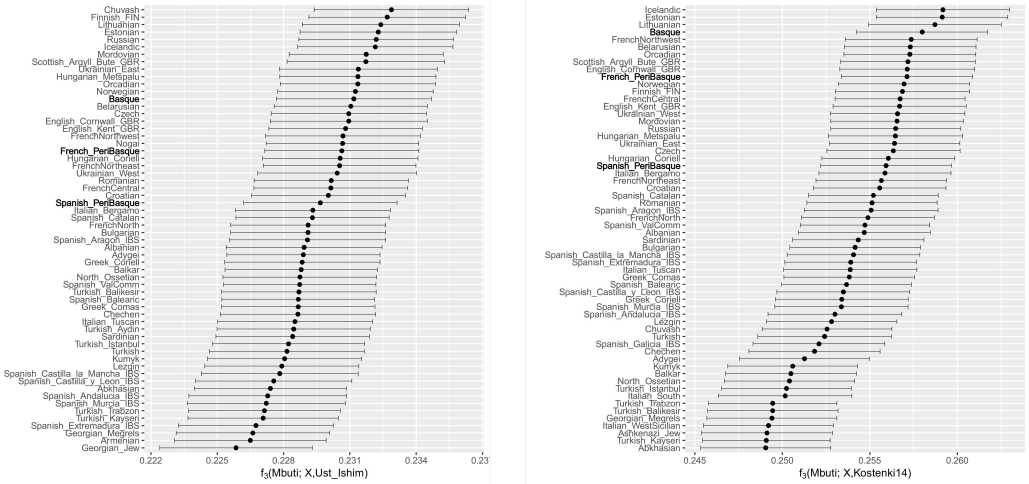
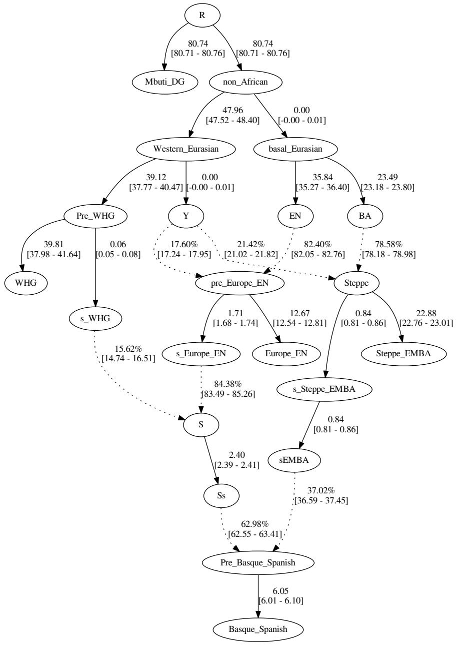


Figure S7. Outgroup f_3-statistics results. The analysis was performed in the form f_3(Mbuti; Ancient, Modern). For each ancient group, the shared drift was calculated with every modern populations. The larger values of f_3 evidence a higher shared drift in these pairs of populations.

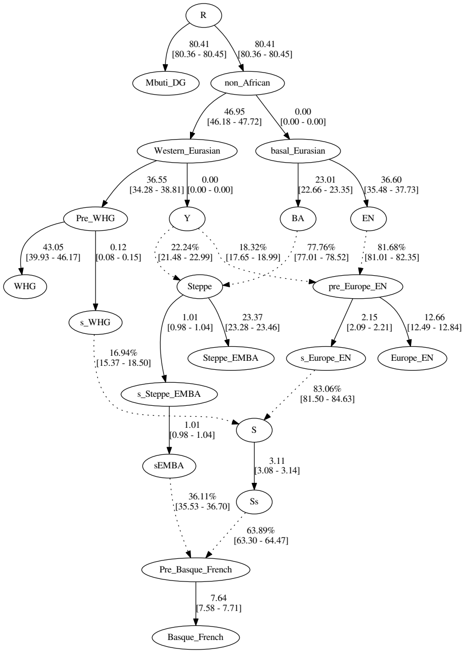
```
Basque_Spanish    Eur   Bas   Ste   Bas   -0.001069   -0.001110   -0.000042   0.000247   -0.168
Basque_Spanish    Eur   Bas   Ste   Bas   -0.001068   -0.001110   -0.000042   0.000247   -0.172
Basque_Spanish    Eur   Bas   Ste   Bas   -0.001068   -0.001110   -0.000042   0.000247   -0.170
Basque_Spanish    Eur   Bas   Ste   Bas   -0.001008   -0.001110   -0.000102   0.000247   -0.414
Basque_Spanish    Eur   Bas   Ste   Bas   -0.001069   -0.001110   -0.000042   0.000247   -0.169
Basque_Spanish    Eur   Bas   Ste   Bas   -0.001040   -0.001110   -0.000071   0.000247   -0.286
Basque_Spanish    Eur   Bas   Ste   Bas   -0.001007   -0.001110   -0.000104   0.000247   -0.419
Basque_Spanish    Eur   Bas   Ste   Bas   -0.001067   -0.001110   -0.000043   0.000247   -0.174
Basque_Spanish    Eur   Bas   Ste   Bas   -0.001069   -0.001110   -0.000041   0.000247   -0.168
Basque_Spanish    Eur   Bas   Ste   Bas   -0.000973   -0.001110   -0.000137   0.000247   -0.555
Basque_Spanish    Eur   Bas   Ste   Bas   -0.001069   -0.001110   -0.000041   0.000247   -0.167
Basque_Spanish    Eur   Bas   Ste   Bas   -0.001067   -0.001110   -0.000044   0.000247   -0.176
Basque_Spanish    Eur   Bas   Ste   Bas   -0.001067   -0.001110   -0.000043   0.000247   -0.175
```

[95% confidence intervals] - 100 different seeds



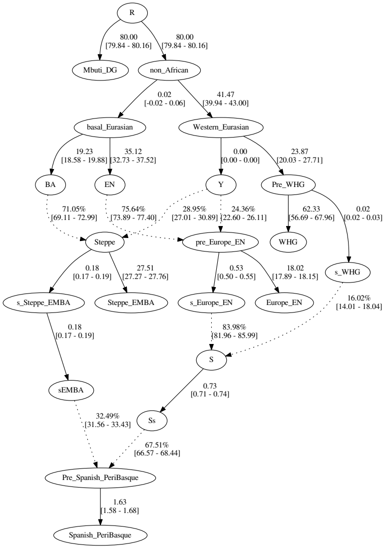
```
Basque_French    Eur   Bas   Ste   Bas   0.001256   0.001217   -0.000039   0.000301   -0.130
Basque_French    Eur   Bas   Ste   Bas   0.001350   0.001217   -0.000133   0.000301   -0.441
Basque_French    Eur   Bas   Ste   Bas   0.001373   0.001217   -0.000156   0.000301   -0.519
Basque_French    Eur   Bas   Ste   Bas   0.001391   0.001217   -0.000175   0.000301   -0.580
Basque_French    Eur   Bas   Ste   Bas   0.001317   0.001217   -0.000100   0.000301   -0.332
Basque_French    Eur   Bas   Ste   Bas   0.001255   0.001217   -0.000038   0.000301   -0.127
Basque_French    Eur   Bas   Ste   Bas   0.001332   0.001217   -0.000116   0.000301   -0.384
Basque_French    Eur   Bas   Ste   Bas   0.001256   0.001217   -0.000039   0.000301   -0.129
Basque_French    Eur   Bas   Ste   Bas   0.001316   0.001217   -0.000100   0.000301   -0.331
Basque_French    Eur   Bas   Ste   Bas   0.001317   0.001217   -0.000100   0.000301   -0.333
Basque_French    Eur   Bas   Ste   Bas   0.001257   0.001217   -0.000040   0.000301   -0.134
Basque_French    Eur   Bas   Ste   Bas   0.001258   0.001217   -0.000041   0.000301   -0.136
```

[95% confidence intervals] - 100 different seeds



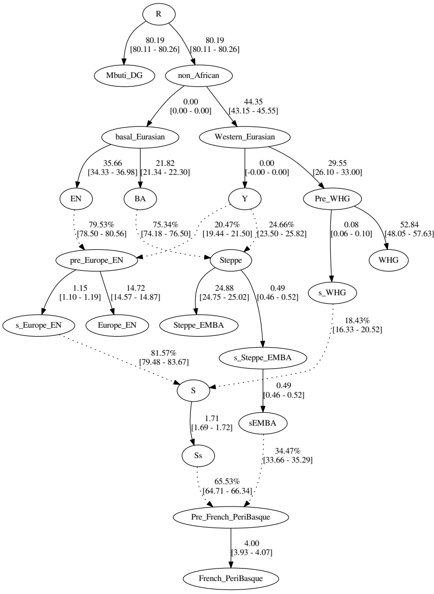
```
Spanish_PeriBasque    WHG   Ste   Eur   Spa   -0.000475   -0.000292    0.000183   0.000439    0.416
 Spanish_PeriBasque   Eur   Spa   Ste   Spa   -0.004365   -0.004386   -0.000021   0.000233   -0.089
 Spanish_PeriBasque   Eur   Spa   Ste   Spa   -0.004263   -0.004386   -0.000123   0.000233   -0.527
 Spanish_PeriBasque   Eur   Spa   Ste   Spa   -0.004264   -0.004386   -0.000122   0.000233   -0.523
 Spanish_PeriBasque   Eur   Spa   Ste   Spa   -0.004366   -0.004386   -0.000021   0.000233   -0.088
Spanish_PeriBasque    WHG   Ste   Eur   Spa   -0.000475   -0.000292    0.000182   0.000439    0.415
Spanish_PeriBasque    WHG   Ste   Eur   Spa   -0.000474   -0.000292    0.000182   0.000439    0.414
 Spanish_PeriBasque   Eur   Spa   Ste   Spa   -0.004365   -0.004386   -0.000021   0.000233   -0.091
 Spanish_PeriBasque   Eur   Spa   Ste   Spa   -0.004365   -0.004386   -0.000021   0.000233   -0.090
Spanish_PeriBasque    WHG   Ste   Eur   Spa   -0.000472   -0.000292    0.000180   0.000439    0.409
Spanish_PeriBasque    WHG   Ste   Eur   Spa   -0.000476   -0.000292    0.000184   0.000439    0.419
 Spanish_PeriBasque   Eur   Spa   Ste   Spa   -0.003338   -0.004386   -0.001048   0.000233   -4.487
 Spanish_PeriBasque   Eur   Spa   Ste   Spa   -0.004283   -0.004386   -0.000104   0.000233   -0.444
Spanish_PeriBasque    WHG   Ste   Eur   Spa   -0.000467   -0.000292    0.000174   0.000439    0.397
 Spanish_PeriBasque   Eur   Spa   Ste   Spa   -0.004365   -0.004386   -0.000021   0.000233   -0.088
Spanish_PeriBasque    WHG   Ste   Eur   Spa   -0.000475   -0.000292    0.000183   0.000439    0.417
Spanish_PeriBasque    WHG   Ste   Eur   Spa   -0.000479   -0.000292    0.000186   0.000439    0.424
 Spanish_PeriBasque   Eur   Spa   Ste   Spa   -0.004282   -0.004386   -0.000104   0.000233   -0.446
 Spanish_PeriBasque   Eur   Spa   Ste   Spa   -0.004270   -0.004386   -0.000116   0.000233   -0.497
Spanish_PeriBasque    WHG   Ste   Eur   Spa   -0.000471   -0.000292    0.000179   0.000439    0.408
Spanish_PeriBasque    WHG   Ste   Eur   Spa   -0.000476   -0.000292    0.000183   0.000439    0.417
Spanish_PeriBasque    WHG   Ste   Eur   Spa   -0.000474   -0.000292    0.000182   0.000439    0.415
Spanish_PeriBasque    WHG   Ste   Eur   Spa   -0.000473   -0.000292    0.000181   0.000439    0.412
Spanish_PeriBasque    WHG   Ste   Eur   Spa   -0.000474   -0.000292    0.000181   0.000439    0.413
Spanish_PeriBasque    WHG   Ste   Eur   Spa   -0.000476   -0.000292    0.000184   0.000439    0.418
 Spanish_PeriBasque   Eur   Spa   Ste   Spa   -0.004279   -0.004386   -0.000108   0.000233   -0.461
 Spanish_PeriBasque   Eur   Spa   Ste   Spa   -0.004364   -0.004386   -0.000022   0.000233   -0.094
```

[95% confidence intervals] - 100 different seeds



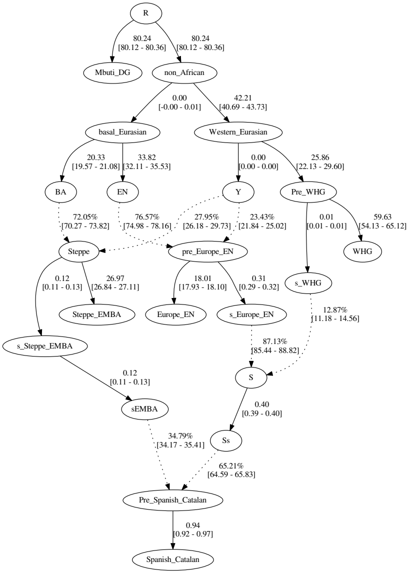
```
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002495   -0.002527   -0.000032   0.000261   -0.125
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002495   -0.002527   -0.000032   0.000261   -0.124
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002433   -0.002527   -0.000095   0.000261   -0.364
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002432   -0.002527   -0.000095   0.000261   -0.365
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002432   -0.002527   -0.000096   0.000261   -0.367
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002495   -0.002527   -0.000032   0.000261   -0.123
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002432   -0.002527   -0.000095   0.000261   -0.366
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002495   -0.002527   -0.000033   0.000261   -0.125
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002429   -0.002527   -0.000099   0.000261   -0.379
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002436   -0.002527   -0.000091   0.000261   -0.350
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002429   -0.002527   -0.000098   0.000261   -0.377
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002414   -0.002527   -0.000114   0.000261   -0.435
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002432   -0.002527   -0.000096   0.000261   -0.366
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002420   -0.002527   -0.000107   0.000261   -0.410
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002487   -0.002527   -0.000040   0.000261   -0.153
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002433   -0.002527   -0.000095   0.000261   -0.363
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002422   -0.002527   -0.000105   0.000261   -0.405
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002433   -0.002527   -0.000094   0.000261   -0.361
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002429   -0.002527   -0.000098   0.000261   -0.376
French_PeriBasque   Eur   Fre   Ste   Fre   -0.002443   -0.002527   -0.000084   0.000261   -0.322
```

[95% confidence intervals] - 100 different seeds



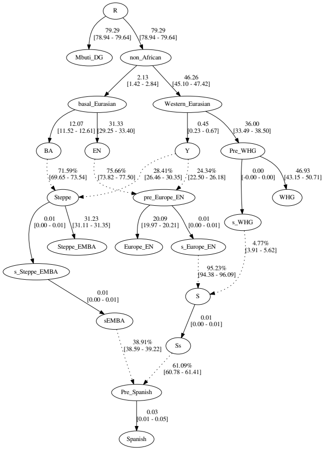
Spanish_Catalan	Eur	Spa	Ste	Spa	-0.005597	-0.005620	-0.000022	0.000224	-0.100
Spanish_Catalan	WHG	Ste	Eur	Spa	0.000345	0.000559	0.000214	0.000432	0.496
Spanish_Catalan	Eur	Spa	Ste	Spa	-0.005597	-0.005620	-0.000023	0.000224	-0.101
Spanish_Catalan	Eur	Spa	Ste	Spa	-0.005596	-0.005620	-0.000023	0.000224	-0.105
Spanish_Catalan	WHG	Ste	Eur	Spa	0.000342	0.000559	0.000217	0.000432	0.503
Spanish_Catalan	WHG	Ste	Eur	Spa	0.000338	0.000559	0.000221	0.000432	0.512
Spanish_Catalan	WHG	Ste	Eur	Spa	0.000344	0.000559	0.000215	0.000432	0.498
Spanish_Catalan	WHG	Ste	Eur	Spa	0.000342	0.000559	0.000218	0.000432	0.504
Spanish_Catalan	WHG	Ste	Eur	Spa	0.000348	0.000559	0.000212	0.000432	0.490
Spanish_Catalan	Eur	Spa	Ste	Spa	-0.005598	-0.005620	-0.000022	0.000224	-0.099
Spanish_Catalan	WHG	Ste	Eur	Spa	0.000342	0.000559	0.000217	0.000432	0.504
Spanish_Catalan	WHG	Ste	Eur	Spa	0.000343	0.000559	0.000216	0.000432	0.501
Spanish_Catalan	Eur	Spa	Ste	Spa	-0.005598	-0.005620	-0.000022	0.000224	-0.098
Spanish_Catalan	WHG	Ste	Eur	Spa	0.000341	0.000559	0.000218	0.000432	0.506
Spanish_Catalan	WHG	Ste	Eur	Spa	0.000341	0.000559	0.000218	0.000432	0.505
Spanish_Catalan	WHG	Ste	Eur	Spa	0.000339	0.000559	0.000220	0.000432	0.510
Spanish_Catalan	WHG	Ste	Eur	Spa	0.000351	0.000559	0.000208	0.000432	0.483
Spanish_Catalan	WHG	Ste	Eur	Spa	0.000341	0.000559	0.000219	0.000432	0.507
Spanish_Catalan	Eur	Spa	Ste	Spa	-0.005478	-0.005620	-0.000142	0.000224	-0.631
Spanish_Catalan	WHG	Ste	Eur	Spa	0.000337	0.000559	0.000223	0.000432	0.516
Spanish_Catalan	WHG	Ste	Eur	Spa	0.000344	0.000559	0.000215	0.000432	0.499
Spanish_Catalan	WHG	Ste	Eur	Spa	0.000344	0.000559	0.000216	0.000432	0.499
Spanish_Catalan	Eur	Spa	Ste	Spa	-0.005596	-0.005620	-0.000023	0.000224	-0.104

[95% confidence intervals] - 100 different seeds



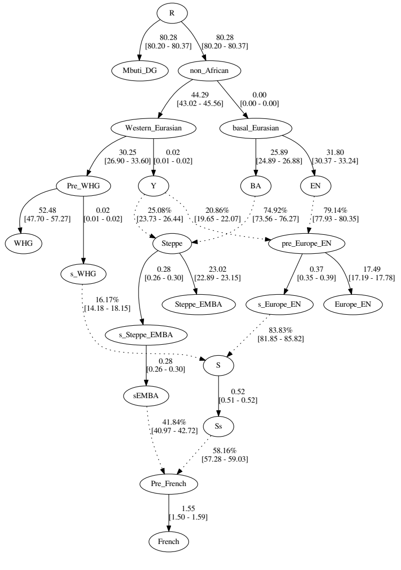
Spanish	Eur	Ste	Ste	Spa	-0.044692	-0.044716	-0.000024	0.000468	-0.052
Spanish	Eur	Ste	Ste	Spa	-0.044689	-0.044716	-0.000026	0.000468	-0.057
Spanish	Eur	Ste	Ste	Spa	-0.044692	-0.044716	-0.000024	0.000468	-0.051
Spanish	Eur	Spa	Ste	Spa	-0.005475	-0.005496	-0.000022	0.000225	-0.098
Spanish	Eur	Spa	Ste	Spa	-0.005450	-0.005496	-0.000046	0.000225	-0.206
Spanish	Eur	Spa	Ste	Spa	-0.005408	-0.005496	-0.000088	0.000225	-0.392
Spanish	Eur	Spa	Ste	Spa	-0.005455	-0.005496	-0.000041	0.000225	-0.185
Spanish	Eur	Ste	Ste	Spa	-0.044683	-0.044716	-0.000033	0.000468	-0.071
Spanish	Eur	Ste	Ste	Spa	-0.044687	-0.044716	-0.000029	0.000468	-0.061
Spanish	Eur	Spa	Ste	Spa	-0.005474	-0.005496	-0.000022	0.000225	-0.099
Spanish	Eur	Ste	Ste	Spa	-0.044688	-0.044716	-0.000028	0.000468	-0.060
Spanish	WHG	Ste	Eur	Spa	0.000175	0.000516	0.000341	0.000432	0.789
Spanish	Eur	Ste	Ste	Spa	-0.044689	-0.044716	-0.000027	0.000468	-0.057
Spanish	Eur	Spa	Ste	Spa	-0.005482	-0.005496	-0.000014	0.000225	-0.064
Spanish	Eur	Ste	Ste	Spa	-0.044688	-0.044716	-0.000027	0.000468	-0.059
Spanish	Eur	Spa	Ste	Spa	-0.005468	-0.005496	-0.000028	0.000225	-0.125
Spanish	Eur	Spa	Ste	Spa	-0.005480	-0.005496	-0.000017	0.000225	-0.076
Spanish	Eur	Ste	Ste	Spa	-0.044691	-0.044716	-0.000024	0.000468	-0.052
Spanish	Eur	Ste	Ste	Spa	-0.044691	-0.044716	-0.000025	0.000468	-0.053
Spanish	Eur	Spa	Ste	Spa	-0.005450	-0.005496	-0.000047	0.000225	-0.209
Spanish	Eur	Ste	Ste	Spa	-0.044691	-0.044716	-0.000025	0.000468	-0.054
Spanish	WHG	Ste	Eur	Spa	0.000183	0.000516	0.000333	0.000432	0.770
Spanish	WHG	Ste	Eur	Spa	0.000182	0.000516	0.000334	0.000432	0.773
Spanish	Eur	Spa	Ste	Spa	-0.005419	-0.005496	-0.000077	0.000225	-0.344
Spanish	Eur	Ste	Ste	Spa	-0.044690	-0.044716	-0.000026	0.000468	-0.055
Spanish	Eur	Spa	Ste	Spa	-0.005350	-0.005496	-0.000146	0.000225	-0.652
Spanish	Eur	Spa	Ste	Spa	-0.005443	-0.005496	-0.000054	0.000225	-0.239
Spanish	WHG	Ste	Eur	Spa	0.000185	0.000516	0.000331	0.000432	0.765
Spanish	Eur	Spa	Ste	Spa	-0.005444	-0.005496	-0.000053	0.000225	-0.234
Spanish	Eur	Spa	Ste	Spa	-0.005445	-0.005496	-0.000052	0.000225	-0.231
Spanish	Eur	Spa	Ste	Spa	-0.005463	-0.005496	-0.000034	0.000225	-0.150
Spanish	Eur	Ste	Ste	Spa	-0.044690	-0.044716	-0.000026	0.000468	-0.056
Spanish	Eur	Spa	Ste	Spa	-0.005416	-0.005496	-0.000081	0.000225	-0.360
Spanish	WHG	Ste	Eur	Spa	0.000162	0.000516	0.000354	0.000432	0.819
Spanish	Eur	Spa	Ste	Spa	-0.005425	-0.005496	-0.000072	0.000225	-0.319
Spanish	Eur	Spa	Ste	Spa	-0.005446	-0.005496	-0.000050	0.000225	-0.225
Spanish	Eur	Spa	Ste	Spa	-0.005406	-0.005496	-0.000090	0.000225	-0.402
Spanish	Eur	Spa	Ste	Spa	-0.005479	-0.005496	-0.000018	0.000225	-0.079
Spanish	Eur	Spa	Ste	Spa	-0.005448	-0.005496	-0.000049	0.000225	-0.217
Spanish	Eur	Spa	Ste	Spa	-0.005481	-0.005496	-0.000016	0.000225	-0.069
Spanish	Eur	Spa	Ste	Spa	-0.005465	-0.005496	-0.000031	0.000225	-0.139
Spanish	Eur	Spa	Ste	Spa	-0.005470	-0.005496	-0.000026	0.000225	-0.116
Spanish	Eur	Spa	Ste	Spa	-0.005461	-0.005496	-0.000036	0.000225	-0.160
Spanish	Eur	Spa	Ste	Spa	-0.005483	-0.005496	-0.000013	0.000225	-0.060
Spanish	WHG	Ste	Eur	Spa	0.000181	0.000516	0.000335	0.000432	0.775
Spanish	Eur	Spa	Ste	Spa	-0.005472	-0.005496	-0.000024	0.000225	-0.108
Spanish	Eur	Spa	Ste	Spa	-0.005474	-0.005496	-0.000022	0.000225	-0.098
Spanish	WHG	Ste	Eur	Spa	0.000179	0.000516	0.000337	0.000432	0.779

[95% confidence intervals] - 100 different seeds



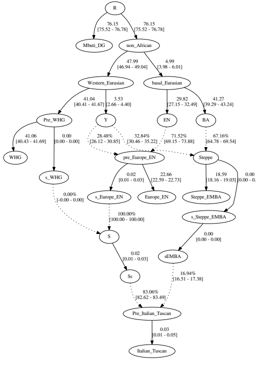
French	Eur	Fre	Ste	Fre	-0.005919	-0.006046	-0.000126	0.000230	-0.548
French	Eur	Fre	Ste	Fre	-0.006012	-0.006046	-0.000033	0.000230	-0.145
French	Eur	Fre	Ste	Fre	-0.005937	-0.006046	-0.000109	0.000230	-0.472
French	Eur	Fre	Ste	Fre	-0.006012	-0.006046	-0.000034	0.000230	-0.146
French	Eur	Fre	Ste	Fre	-0.005819	-0.006046	-0.000227	0.000230	-0.986
French	Eur	Fre	Ste	Fre	-0.005935	-0.006046	-0.000111	0.000230	-0.482
French	Eur	Fre	Ste	Fre	-0.005938	-0.006046	-0.000107	0.000230	-0.466
French	Eur	Fre	Ste	Fre	-0.005938	-0.006046	-0.000107	0.000230	-0.465
French	Eur	Fre	Ste	Fre	-0.005939	-0.006046	-0.000107	0.000230	-0.465
French	Eur	Fre	Ste	Fre	-0.006012	-0.006046	-0.000034	0.000230	-0.147
French	Eur	Fre	Ste	Fre	-0.006010	-0.006046	-0.000036	0.000230	-0.156
French	Eur	Fre	Ste	Fre	-0.005939	-0.006046	-0.000107	0.000230	-0.462
French	Eur	Fre	Ste	Fre	-0.005945	-0.006046	-0.000101	0.000230	-0.437
French	Eur	Fre	Ste	Fre	-0.005938	-0.006046	-0.000108	0.000230	-0.467
French	Eur	Fre	Ste	Fre	-0.005936	-0.006046	-0.000110	0.000230	-0.475
French	Eur	Fre	Ste	Fre	-0.005914	-0.006046	-0.000132	0.000230	-0.571
French	Eur	Fre	Ste	Fre	-0.005938	-0.006046	-0.000108	0.000230	-0.469
French	Eur	Fre	Ste	Fre	-0.005929	-0.006046	-0.000116	0.000230	-0.505
French	Eur	Fre	Ste	Fre	-0.005921	-0.006046	-0.000125	0.000230	-0.540
French	Eur	Fre	Ste	Fre	-0.005937	-0.006046	-0.000109	0.000230	-0.471
French	Eur	Fre	Ste	Fre	-0.005927	-0.006046	-0.000118	0.000230	-0.514
French	Eur	Fre	Ste	Fre	-0.006011	-0.006046	-0.000034	0.000230	-0.149
French	Eur	Fre	Ste	Fre	-0.005940	-0.006046	-0.000106	0.000230	-0.460
French	Eur	Fre	Ste	Fre	-0.005939	-0.006046	-0.000107	0.000230	-0.464
French	Eur	Fre	Ste	Fre	-0.006012	-0.006046	-0.000034	0.000230	-0.145
French	Eur	Fre	Ste	Fre	-0.005938	-0.006046	-0.000108	0.000230	-0.466

[95% confidence intervals] - 100 different seeds



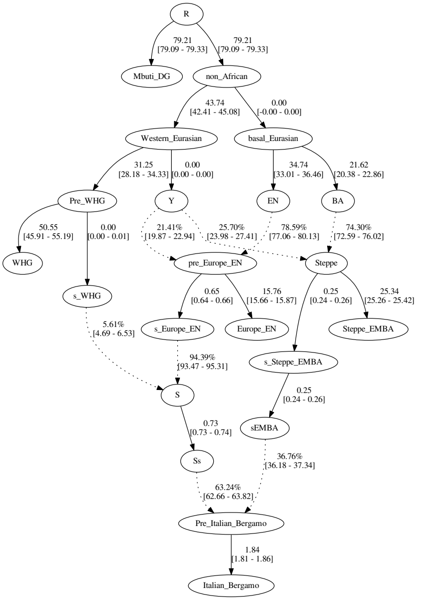
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002774	0.003628	0.000854	0.000526	1.623
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002840	0.003628	0.000788	0.000526	1.497
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002689	0.003628	0.000939	0.000526	1.783
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002583	0.003628	0.001045	0.000526	1.985
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002840	0.003628	0.000788	0.000526	1.496
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002714	0.003628	0.000914	0.000526	1.736
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002841	0.003628	0.000787	0.000526	1.496
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002670	0.003628	0.000958	0.000526	1.820
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002537	0.003628	0.001091	0.000526	2.073
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002832	0.003628	0.000796	0.000526	1.512
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002579	0.003628	0.001049	0.000526	1.992
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002622	0.003628	0.001006	0.000526	1.911
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002735	0.003628	0.000893	0.000526	1.696
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002802	0.003628	0.000826	0.000526	1.570
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002739	0.003628	0.000889	0.000526	1.688
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002579	0.003628	0.001050	0.000526	1.994
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002839	0.003628	0.000789	0.000526	1.498
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002267	0.003628	0.001361	0.000526	2.585
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002816	0.003628	0.000812	0.000526	1.543
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002841	0.003628	0.000788	0.000526	1.496
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002833	0.003628	0.000796	0.000526	1.511
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002797	0.003628	0.000831	0.000526	1.579
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002716	0.003628	0.000912	0.000526	1.732
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002611	0.003628	0.001017	0.000526	1.931
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002570	0.003628	0.001058	0.000526	2.010
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002763	0.003628	0.000866	0.000526	1.644
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002794	0.003628	0.000835	0.000526	1.585
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002789	0.003628	0.000840	0.000526	1.595
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002663	0.003628	0.000966	0.000526	1.834
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002721	0.003628	0.000907	0.000526	1.723
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002781	0.003628	0.000847	0.000526	1.609
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002417	0.003628	0.001212	0.000526	2.301
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002784	0.003628	0.000844	0.000526	1.604
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002688	0.003628	0.000940	0.000526	1.786
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002842	0.003628	0.000786	0.000526	1.493
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002734	0.003628	0.000894	0.000526	1.698
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002550	0.003628	0.001078	0.000526	2.048
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002694	0.003628	0.000934	0.000526	1.774
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002552	0.003628	0.001076	0.000526	2.043
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002619	0.003628	0.001010	0.000526	1.918
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002656	0.003628	0.000972	0.000526	1.846
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002614	0.003628	0.001014	0.000526	1.926
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002646	0.003628	0.000982	0.000526	1.865
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002838	0.003628	0.000790	0.000526	1.501
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002714	0.003628	0.000915	0.000526	1.737
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002376	0.003628	0.001252	0.000526	2.378
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002647	0.003628	0.000981	0.000526	1.864
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002828	0.003628	0.000801	0.000526	1.521
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002837	0.003628	0.000791	0.000526	1.503
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002526	0.003628	0.001102	0.000526	2.093
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002711	0.003628	0.000918	0.000526	1.743
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002464	0.003628	0.001165	0.000526	2.212
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002720	0.003628	0.000908	0.000526	1.725
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002783	0.003628	0.000845	0.000526	1.606
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002587	0.003628	0.001041	0.000526	1.977
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002836	0.003628	0.000792	0.000526	1.505
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002703	0.003628	0.000925	0.000526	1.758
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002839	0.003628	0.000789	0.000526	1.499
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002565	0.003628	0.001064	0.000526	2.020
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002580	0.003628	0.001049	0.000526	1.992
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002831	0.003628	0.000798	0.000526	1.515
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002306	0.003628	0.001323	0.000526	2.512
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002657	0.003628	0.000971	0.000526	1.844
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002562	0.003628	0.001067	0.000526	2.026
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002533	0.003628	0.001096	0.000526	2.081
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002728	0.003628	0.000901	0.000526	1.711
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002417	0.003628	0.001212	0.000526	2.302
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002755	0.003628	0.000873	0.000526	1.659
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002838	0.003628	0.000790	0.000526	1.500
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002215	0.003628	0.001414	0.000526	2.685
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002590	0.003628	0.001038	0.000526	1.972
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002661	0.003628	0.000967	0.000526	1.837
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002783	0.003628	0.000845	0.000526	1.605
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002416	0.003628	0.001212	0.000526	2.303
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002152	0.003628	0.001476	0.000526	2.803
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002605	0.003628	0.001023	0.000526	1.943
Italian_Tuscan	WHG	Ste	Eur	Ita	0.002558	0.003628	0.001070	0.000526	2.033

[95% confidence intervals] - 100 different seeds



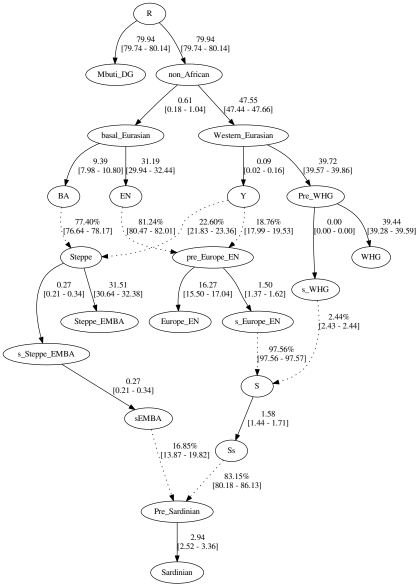
```
 Italian_Bergamo    Eur    Ita    Ste    Ita   -0.005182  -0.005210  -0.000028   0.000285  -0.098
Italian_Bergamo    WHG     Ste    Eur     Ita    0.002035   0.002412   0.000377   0.000495   0.762
Italian_Bergamo    WHG     Ste    Eur     Ita    0.002026   0.002412   0.000386   0.000495   0.781
 Italian_Bergamo    Eur    Ita    Ste    Ita   -0.005182  -0.005210  -0.000028   0.000285  -0.097
Italian_Bergamo    WHG     Ste    Eur     Ita    0.002036   0.002412   0.000377   0.000495   0.761
Italian_Bergamo    WHG     Ste    Eur     Ita    0.002035   0.002412   0.000377   0.000495   0.763
Italian_Bergamo    WHG     Ste    Eur     Ita    0.002053   0.002412   0.000359   0.000495   0.726
 Italian_Bergamo    Eur    Ita    Ste    Ita   -0.005182  -0.005210  -0.000028   0.000285  -0.099
 Italian_Bergamo    Eur    Ita    Ste    Ita   -0.005181  -0.005210  -0.000028   0.000285  -0.099
Italian_Bergamo    WHG     Ste    Eur     Ita    0.002033   0.002412   0.000379   0.000495   0.767
Italian_Bergamo    WHG     Ste    Eur     Ita    0.002036   0.002412   0.000376   0.000495   0.760
Italian_Bergamo    WHG     Ste    Eur     Ita    0.002034   0.002412   0.000379   0.000495   0.765
Italian_Bergamo    WHG     Ste    Eur     Ita    0.002037   0.002412   0.000375   0.000495   0.758
 Italian_Bergamo    Eur    Ita    Ste    Ita   -0.005181  -0.005210  -0.000029   0.000285  -0.102
 Italian_Bergamo    Eur    Ita    Ste    Ita   -0.005181  -0.005210  -0.000028   0.000285  -0.100
Italian_Bergamo    WHG     Ste    Eur     Ita    0.001991   0.002412   0.000422   0.000495   0.852
 Italian_Bergamo    Eur    Ita    Ste    Ita   -0.005180  -0.005210  -0.000030   0.000285  -0.105
 Italian_Bergamo    Eur    Ita    Ste    Ita   -0.005179  -0.005210  -0.000030   0.000285  -0.107
Italian_Bergamo    WHG     Ste    Eur     Ita    0.002050   0.002412   0.000363   0.000495   0.732
```

[95% confidence intervals] - 100 different seeds



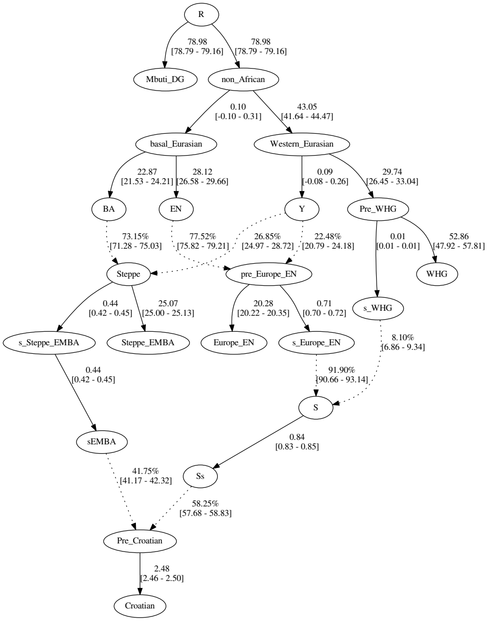
```
 Sardinian    Ste    Sar    Ste    Sar    0.054448   0.054491   0.000043   0.000538   0.081
Sardinian    Eur    Ste    Ste    Sar   -0.052910  -0.052931  -0.000021   0.000544  -0.038
 Sardinian    Ste    Sar    Ste    Sar    0.054441   0.054491   0.000050   0.000538   0.094
 Sardinian    Ste    Sar    Ste    Sar    0.054448   0.054491   0.000044   0.000538   0.081
 Sardinian    Ste    Sar    Ste    Sar    0.054446   0.054491   0.000045   0.000538   0.084
Sardinian    Eur    Sar    Eur     Sar    0.022626   0.022600  -0.000026   0.000310  -0.082
Sardinian    Eur    Ste    Ste    Sar   -0.052889  -0.052931  -0.000042   0.000544  -0.077
 Sardinian    Ste    Sar    Ste    Sar    0.054448   0.054491   0.000043   0.000538   0.080
Sardinian    Eur    Sar    Eur     Sar    0.022632   0.022600  -0.000032   0.000310  -0.102
Sardinian    Eur    Sar    Eur     Sar    0.022635   0.022600  -0.000035   0.000310  -0.113
 Sardinian    Ste    Sar    Ste    Sar    0.054448   0.054491   0.000044   0.000538   0.082
Sardinian    Eur    Sar    Eur     Sar    0.022624   0.022600  -0.000024   0.000310  -0.076
 Sardinian    Ste    Sar    Ste    Sar    0.054449   0.054491   0.000043   0.000538   0.079
Sardinian    Eur    Ste    Ste    Sar   -0.052897  -0.052931  -0.000034   0.000544  -0.063
 Sardinian    Ste    Sar    Ste    Sar    0.054447   0.054491   0.000045   0.000538   0.083
 Sardinian    Ste    Sar    Ste    Sar    0.054449   0.054491   0.000043   0.000538   0.080
 Sardinian    Ste    Sar    Ste    Sar    0.054441   0.054491   0.000050   0.000538   0.093
Sardinian    Eur    Ste    Ste    Sar   -0.052904  -0.052931  -0.000027   0.000544  -0.050
 Sardinian    Ste    Sar    Ste    Sar    0.054446   0.054491   0.000046   0.000538   0.085
Sardinian    Eur    Ste    Ste    Sar   -0.052899  -0.052931  -0.000032   0.000544  -0.059
Sardinian    Eur    Sar    Eur     Sar    0.022632   0.022600  -0.000031   0.000310  -0.101
```

[95% confidence intervals] - 100 different seeds



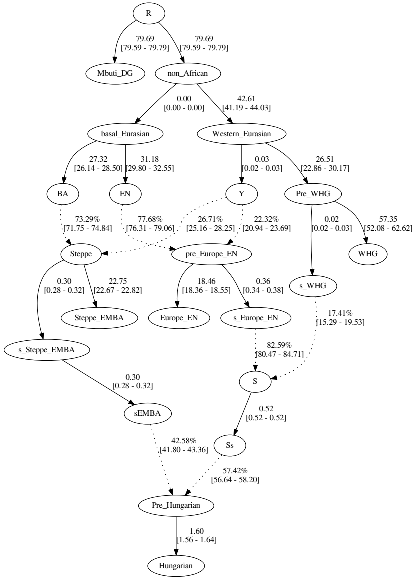
```
 Croatian    Eur    Cro    Ste    Cro   -0.003671  -0.003692  -0.000021   0.000324  -0.065
Croatian    WHG     Ste    Eur     Cro    0.002378   0.002685   0.000307   0.000543   0.565
 Croatian    Eur    Cro    Ste    Cro   -0.003671  -0.003692  -0.000021   0.000324  -0.066
Croatian    WHG     Ste    Eur     Cro    0.002377   0.002685   0.000308   0.000543   0.566
Croatian    WHG     Ste    Eur     Cro    0.002379   0.002685   0.000306   0.000543   0.563
Croatian    WHG     Ste    Eur     Cro    0.002369   0.002685   0.000316   0.000543   0.582
 Croatian    Eur    Cro    Ste    Cro   -0.003669  -0.003692  -0.000023   0.000324  -0.071
Croatian    WHG     Ste    Eur     Cro    0.002377   0.002685   0.000307   0.000543   0.566
 Croatian    Eur    Cro    Ste    Cro   -0.003670  -0.003692  -0.000022   0.000324  -0.067
Croatian    WHG     Ste    Eur     Cro    0.002379   0.002685   0.000306   0.000543   0.564
 Croatian    Eur    Cro    Ste    Cro   -0.003670  -0.003692  -0.000022   0.000324  -0.069
 Croatian    Eur    Cro    Ste    Cro   -0.003460  -0.003692  -0.000233   0.000324  -0.718
Croatian    WHG     Ste    Eur     Cro    0.002372   0.002685   0.000313   0.000543   0.576
Croatian    WHG     Ste    Eur     Cro    0.002378   0.002685   0.000307   0.000543   0.564
Croatian    WHG     Ste    Eur     Cro    0.002384   0.002685   0.000301   0.000543   0.555
Croatian    WHG     Ste    Eur     Cro    0.002376   0.002685   0.000309   0.000543   0.568
```

[95% confidence intervals] - 100 different seeds



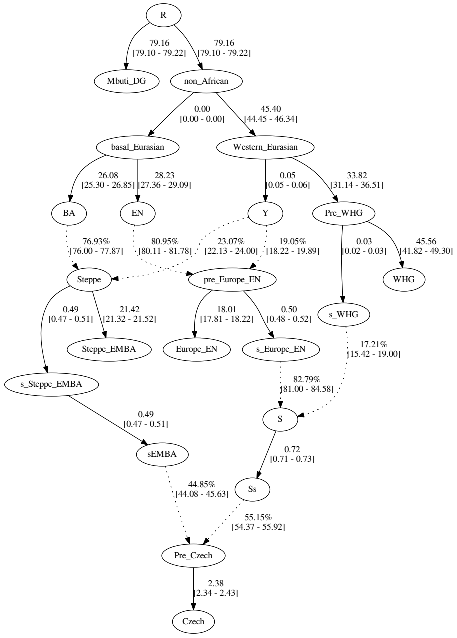
```
Hungarian    Eur    Hun    Ste    Hun   -0.005719  -0.005751  -0.000032   0.000253  -0.125
Hungarian    Eur    Hun    Ste    Hun   -0.005642  -0.005751  -0.000108   0.000253  -0.429
Hungarian    Eur    Hun    Ste    Hun   -0.005642  -0.005751  -0.000109   0.000253  -0.429
Hungarian    Eur    Hun    Ste    Hun   -0.005629  -0.005751  -0.000122   0.000253  -0.483
Hungarian    Eur    Hun    Ste    Hun   -0.005718  -0.005751  -0.000033   0.000253  -0.130
Hungarian    Eur    Hun    Ste    Hun   -0.005610  -0.005751  -0.000140   0.000253  -0.555
Hungarian    Eur    Hun    Ste    Hun   -0.005608  -0.005751  -0.000143   0.000253  -0.565
Hungarian    Eur    Hun    Ste    Hun   -0.005642  -0.005751  -0.000109   0.000253  -0.430
Hungarian    Eur    Hun    Ste    Hun   -0.005643  -0.005751  -0.000107   0.000253  -0.425
Hungarian    Eur    Hun    Ste    Hun   -0.005643  -0.005751  -0.000108   0.000253  -0.427
Hungarian    Eur    Hun    Ste    Hun   -0.005719  -0.005751  -0.000032   0.000253  -0.126
Hungarian    Eur    Hun    Ste    Hun   -0.005619  -0.005751  -0.000132   0.000253  -0.521
Hungarian    Eur    Hun    Ste    Hun   -0.005618  -0.005751  -0.000133   0.000253  -0.527
Hungarian    Eur    Hun    Ste    Hun   -0.005608  -0.005751  -0.000142   0.000253  -0.563
Hungarian    Eur    Hun    Ste    Hun   -0.005719  -0.005751  -0.000032   0.000253  -0.127
Hungarian    Eur    Hun    Ste    Hun   -0.005638  -0.005751  -0.000113   0.000253  -0.447
Hungarian    Eur    Hun    Ste    Hun   -0.005642  -0.005751  -0.000109   0.000253  -0.431
Hungarian    Eur    Hun    Ste    Hun   -0.005641  -0.005751  -0.000109   0.000253  -0.433
Hungarian    Eur    Hun    Ste    Hun   -0.005632  -0.005751  -0.000119   0.000253  -0.471
Hungarian    Eur    Hun    Ste    Hun   -0.005641  -0.005751  -0.000109   0.000253  -0.432
Hungarian    Eur    Hun    Ste    Hun   -0.005644  -0.005751  -0.000106   0.000253  -0.421
Hungarian    Eur    Hun    Ste    Hun   -0.005642  -0.005751  -0.000108   0.000253  -0.428
```

[95% confidence intervals] - 100 different seeds



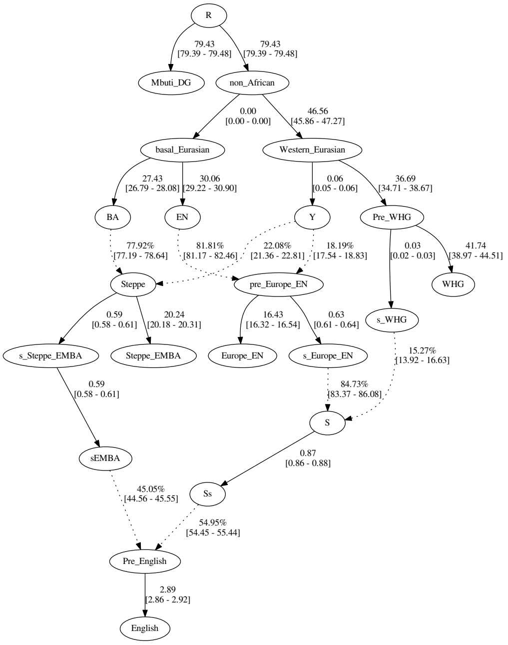
```
Czech    Eur  Cze  Ste  Cze  -0.004840  -0.004938  -0.000098  0.000320  -0.306
Czech    Eur  Cze  Ste  Cze  -0.004903  -0.004938  -0.000034  0.000320  -0.107
Czech    Eur  Cze  Ste  Cze  -0.004830  -0.004938  -0.000108  0.000320  -0.337
Czech    Eur  Cze  Ste  Cze  -0.004903  -0.004938  -0.000034  0.000320  -0.108
Czech    Eur  Cze  Ste  Cze  -0.004836  -0.004938  -0.000101  0.000320  -0.317
Czech    WHG  Eur  Ste  Cze   0.003868   0.004036   0.000168  0.000558   0.301
Czech    Eur  Cze  Ste  Cze  -0.004787  -0.004938  -0.000150  0.000320  -0.470
Czech    Eur  Cze  Ste  Cze  -0.004837  -0.004938  -0.000100  0.000320  -0.314
Czech    Eur  Cze  Ste  Cze  -0.004837  -0.004938  -0.000101  0.000320  -0.314
Czech    Eur  Cze  Ste  Cze  -0.004903  -0.004938  -0.000035  0.000320  -0.108
Czech    Eur  Cze  Ste  Cze  -0.004903  -0.004938  -0.000035  0.000320  -0.109
Czech    Eur  Cze  Ste  Cze  -0.004902  -0.004938  -0.000036  0.000320  -0.112
Czech    Eur  Cze  Ste  Cze  -0.004871  -0.004938  -0.000067  0.000320  -0.208
Czech    Eur  Cze  Ste  Cze  -0.004817  -0.004938  -0.000121  0.000320  -0.377
```

[95% confidence intervals] - 100 different seeds



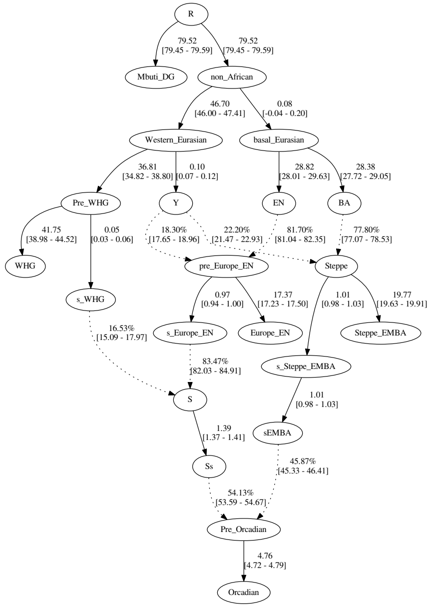
```
English  Eur  Eng  Ste  Eng  -0.005099  -0.005140  -0.000041  0.000307  -0.132
English  Eur  Eng  Ste  Eng  -0.005032  -0.005140  -0.000107  0.000307  -0.349
English  Eur  Eng  Ste  Eng  -0.005099  -0.005140  -0.000041  0.000307  -0.133
English  Eur  Eng  Ste  Eng  -0.005032  -0.005140  -0.000108  0.000307  -0.351
English  Eur  Eng  Ste  Eng  -0.005097  -0.005140  -0.000042  0.000307  -0.137
English  Eur  Eng  Ste  Eng  -0.004970  -0.005140  -0.000169  0.000307  -0.551
English  Eur  Eng  Ste  Eng  -0.005032  -0.005140  -0.000107  0.000307  -0.350
English  Eur  Eng  Ste  Eng  -0.005099  -0.005140  -0.000040  0.000307  -0.132
English  Eur  Eng  Ste  Eng  -0.005098  -0.005140  -0.000041  0.000307  -0.134
English  Eur  Eng  Ste  Eng  -0.005040  -0.005140  -0.000100  0.000307  -0.325
English  Eur  Eng  Ste  Eng  -0.005028  -0.005140  -0.000112  0.000307  -0.365
English  Eur  Eng  Ste  Eng  -0.005097  -0.005140  -0.000042  0.000307  -0.138
English  Eur  Eng  Ste  Eng  -0.004945  -0.005140  -0.000194  0.000307  -0.632
```

[95% confidence intervals] - 100 different seeds



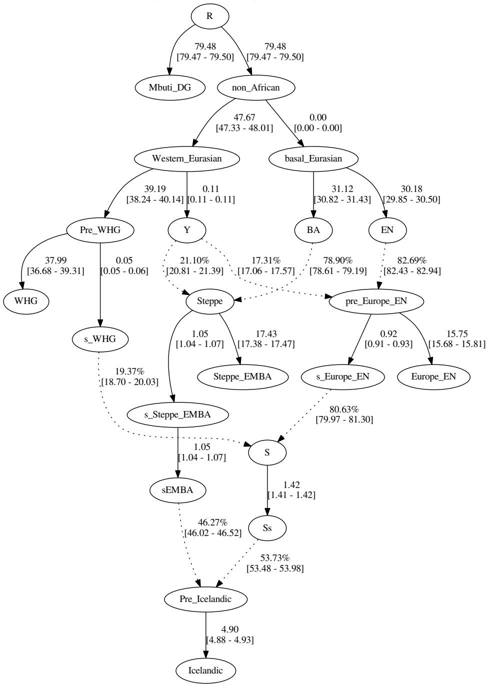
```
Orcadian  Eur  Orc  Ste  Orc  -0.002666  -0.002705  -0.000039  0.000330  -0.119
Orcadian  Eur  Orc  Ste  Orc  -0.002599  -0.002705  -0.000106  0.000330  -0.320
Orcadian  Eur  Orc  Ste  Orc  -0.002645  -0.002705  -0.000060  0.000330  -0.182
Orcadian  Eur  Orc  Ste  Orc  -0.002665  -0.002705  -0.000039  0.000330  -0.120
Orcadian  Eur  Orc  Ste  Orc  -0.002630  -0.002705  -0.000075  0.000330  -0.228
Orcadian  Eur  Orc  Ste  Orc  -0.002666  -0.002705  -0.000039  0.000330  -0.120
Orcadian  Eur  Orc  Ste  Orc  -0.002600  -0.002705  -0.000105  0.000330  -0.319
Orcadian  Eur  Orc  Ste  Orc  -0.002665  -0.002705  -0.000040  0.000330  -0.120
Orcadian  Eur  Orc  Ste  Orc  -0.002665  -0.002705  -0.000040  0.000330  -0.123
Orcadian  Eur  Orc  Ste  Orc  -0.002599  -0.002705  -0.000106  0.000330  -0.322
Orcadian  Eur  Orc  Ste  Orc  -0.002649  -0.002705  -0.000056  0.000330  -0.170
Orcadian  Eur  Orc  Ste  Orc  -0.002591  -0.002705  -0.000114  0.000330  -0.347
Orcadian  Eur  Orc  Ste  Orc  -0.002666  -0.002705  -0.000039  0.000330  -0.118
Orcadian  Eur  Orc  Ste  Orc  -0.002600  -0.002705  -0.000105  0.000330  -0.318
Orcadian  Eur  Orc  Ste  Orc  -0.002531  -0.002705  -0.000174  0.000330  -0.528
Orcadian  Eur  Orc  Ste  Orc  -0.002665  -0.002705  -0.000040  0.000330  -0.121
Orcadian  Eur  Orc  Ste  Orc  -0.002665  -0.002705  -0.000040  0.000330  -0.122
Orcadian  Eur  Orc  Ste  Orc  -0.002546  -0.002705  -0.000159  0.000330  -0.481
```

[95% confidence intervals] - 100 different seeds



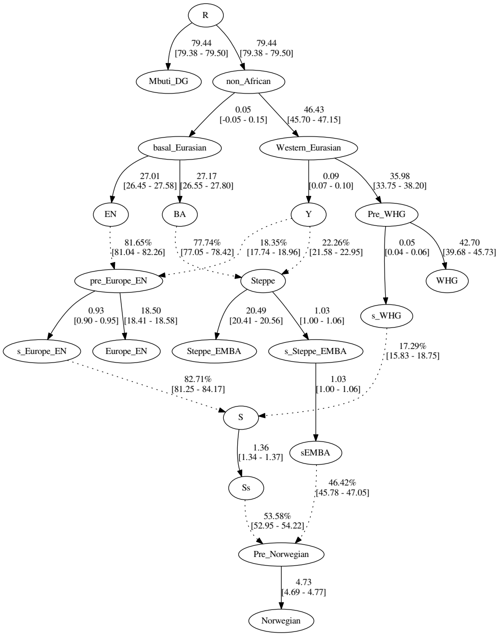
```
Icelandic  Eur  Ice  Ste  Ice  -0.003188  -0.003237  -0.000048  0.000343  -0.141
Icelandic  Eur  Ice  Ste  Ice  -0.003188  -0.003237  -0.000049  0.000343  -0.143
Icelandic  Eur  Ice  Ste  Ice  -0.003188  -0.003237  -0.000049  0.000343  -0.142
Icelandic  Eur  Ice  Ste  Ice  -0.003183  -0.003237  -0.000053  0.000343  -0.155
Icelandic  Eur  Ice  Ste  Ice  -0.003189  -0.003237  -0.000048  0.000343  -0.140
Icelandic  Eur  Ice  Ste  Ice  -0.003120  -0.003237  -0.000116  0.000343  -0.339
Icelandic  Eur  Ice  Ste  Ice  -0.003123  -0.003237  -0.000114  0.000343  -0.332
Icelandic  Eur  Ice  Ste  Ice  -0.003188  -0.003237  -0.000048  0.000343  -0.140
```

[95% confidence intervals] - 100 different seeds



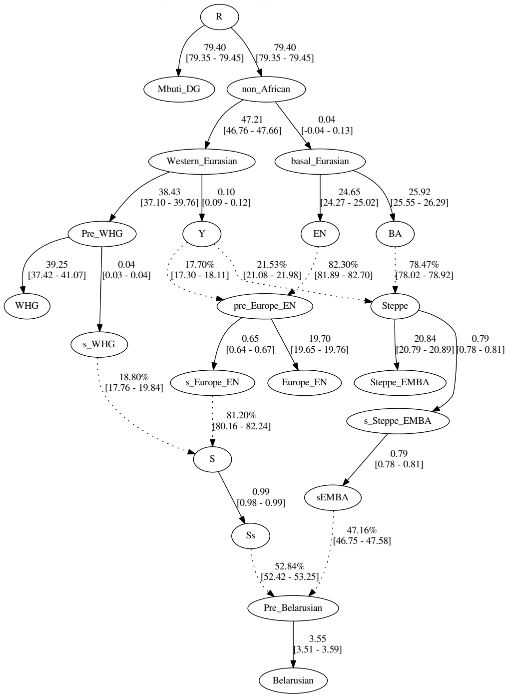
```
Norwegian    Eur   Nor   Ste   Nor   -0.002196   -0.002231   -0.000035   0.000339   -0.104
Norwegian    WHG   Eur   Ste   Nor    0.002968    0.003132    0.000164   0.000551    0.299
Norwegian    WHG   Eur   Ste   Nor    0.002968    0.003132    0.000165   0.000551    0.299
Norwegian    WHG   Eur   Ste   Nor    0.002969    0.003132    0.000164   0.000551    0.297
Norwegian    Eur   Nor   Ste   Nor   -0.002196   -0.002231   -0.000035   0.000339   -0.105
Norwegian    Eur   Nor   Ste   Nor   -0.002125   -0.002231   -0.000106   0.000339   -0.312
Norwegian    Eur   Nor   Ste   Nor   -0.002195   -0.002231   -0.000036   0.000339   -0.105
Norwegian    Eur   Nor   Ste   Nor   -0.002193   -0.002231   -0.000038   0.000339   -0.113
Norwegian    Eur   Nor   Ste   Nor   -0.002129   -0.002231   -0.000102   0.000339   -0.302
Norwegian    Eur   Nor   Ste   Nor   -0.002195   -0.002231   -0.000036   0.000339   -0.106
Norwegian    Eur   Nor   Ste   Nor   -0.002196   -0.002231   -0.000035   0.000339   -0.103
Norwegian    WHG   Eur   Ste   Nor    0.002967    0.003132    0.000165   0.000551    0.299
Norwegian    WHG   Eur   Ste   Nor    0.002968    0.003132    0.000164   0.000551    0.298
Norwegian    Eur   Nor   Ste   Nor   -0.002155   -0.002231   -0.000076   0.000339   -0.225
```

[95% confidence intervals] - 100 different seeds



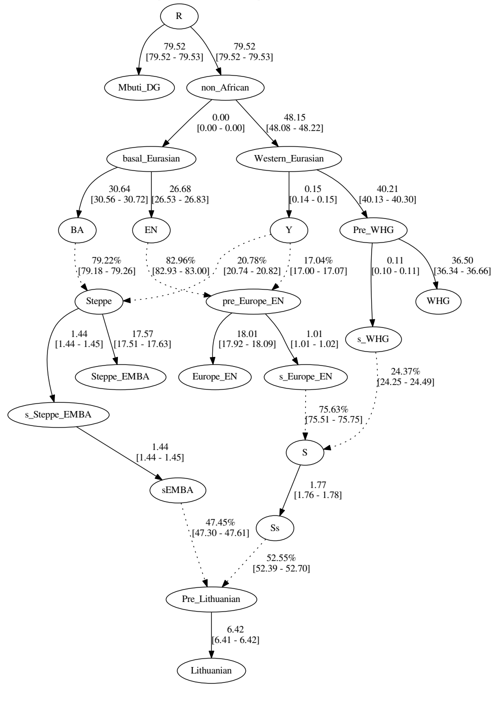
```
Belarusian   Eur   Bel   Ste   Bel   -0.003083   -0.003115   -0.000032   0.000333   -0.097
Belarusian   Eur   Bel   Ste   Bel   -0.003057   -0.003115   -0.000058   0.000333   -0.174
Belarusian   Eur   Bel   Ste   Bel   -0.003082   -0.003115   -0.000032   0.000333   -0.097
Belarusian   Eur   Bel   Ste   Bel   -0.003081   -0.003115   -0.000034   0.000333   -0.103
Belarusian   Eur   Bel   Ste   Bel   -0.003082   -0.003115   -0.000033   0.000333   -0.100
Belarusian   Eur   Bel   Ste   Bel   -0.003023   -0.003115   -0.000092   0.000333   -0.277
Belarusian   Eur   Bel   Ste   Bel   -0.002995   -0.003115   -0.000120   0.000333   -0.361
```

[95% confidence intervals] - 100 different seeds



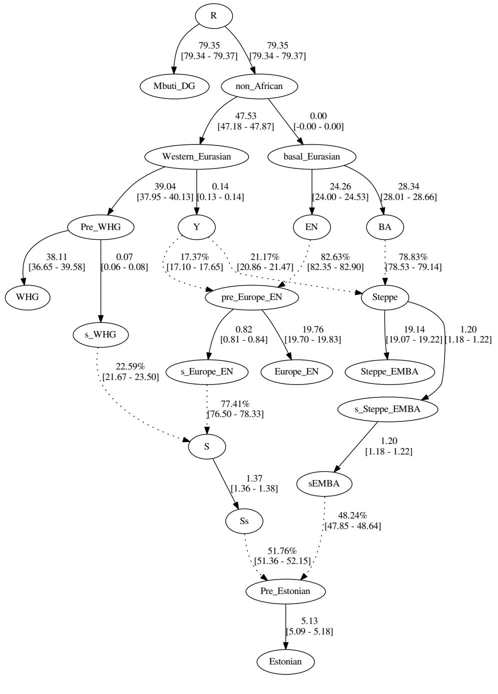
```
Lithuanian   Eur   Lit   Ste   Lit   -0.000293   -0.000336   -0.000043   0.000357   -0.119
Lithuanian   Eur   Lit   Ste   Lit   -0.000293   -0.000336   -0.000042   0.000357   -0.119
Lithuanian   Eur   Lit   Ste   Lit   -0.000279   -0.000336   -0.000057   0.000357   -0.159
Lithuanian   Eur   Lit   Ste   Lit   -0.000293   -0.000336   -0.000043   0.000357   -0.121
Lithuanian   Eur   Lit   Ste   Lit   -0.000293   -0.000336   -0.000042   0.000357   -0.118
Lithuanian   Eur   Lit   Ste   Lit   -0.000293   -0.000336   -0.000043   0.000357   -0.120
Lithuanian   Eur   Lit   Ste   Lit   -0.000292   -0.000336   -0.000044   0.000357   -0.123
Lithuanian   Eur   Lit   Ste   Lit   -0.000291   -0.000336   -0.000044   0.000357   -0.124
```

[95% confidence intervals] - 100 different seeds



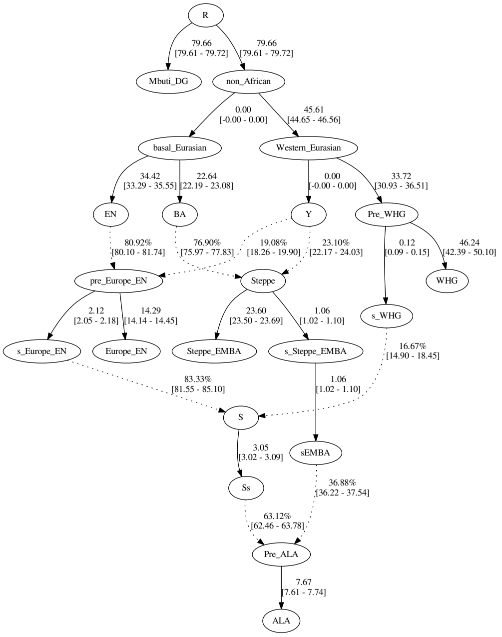
```
Estonian   Eur   Est   Ste   Est   -0.001216   -0.001253   -0.000037   0.000357   -0.104
Estonian   Eur   Est   Ste   Est   -0.001218   -0.001253   -0.000035   0.000357   -0.098
Estonian   Eur   Est   Ste   Est   -0.001218   -0.001253   -0.000035   0.000357   -0.099
Estonian   Eur   Est   Ste   Est   -0.001133   -0.001253   -0.000120   0.000357   -0.337
Estonian   Eur   Est   Ste   Est   -0.001213   -0.001253   -0.000040   0.000357   -0.112
Estonian   Eur   Est   Ste   Est   -0.001217   -0.001253   -0.000036   0.000357   -0.099
Estonian   Eur   Est   Ste   Est   -0.001217   -0.001253   -0.000036   0.000357   -0.100
Estonian   Eur   Est   Ste   Est   -0.001161   -0.001253   -0.000092   0.000357   -0.257
Estonian   Eur   Est   Ste   Est   -0.001217   -0.001253   -0.000035   0.000357   -0.099
```

[95% confidence intervals] - 100 different seeds



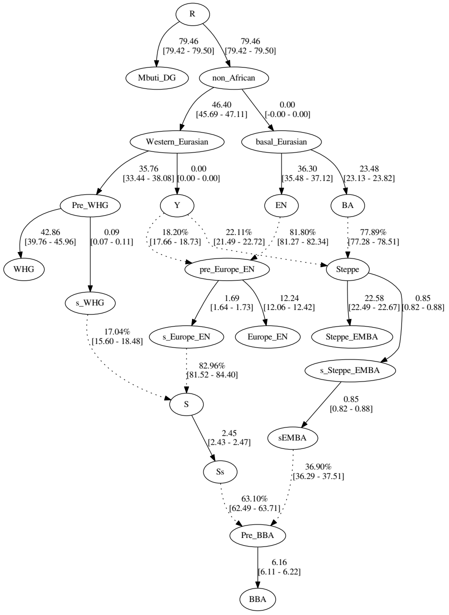
ALA	Eur	ALA	Ste	ALA	0.001706	0.001671	-0.000035	0.000374	-0.092
ALA	Eur	ALA	Ste	ALA	0.001706	0.001671	-0.000034	0.000374	-0.092
ALA	WHG	Eur	Ste	ALA	0.006816	0.006984	0.000168	0.000607	0.277
ALA	Eur	ALA	Ste	ALA	0.001708	0.001671	-0.000037	0.000374	-0.099
ALA	WHG	Eur	Ste	ALA	0.006816	0.006984	0.000169	0.000607	0.278
ALA	Eur	ALA	Ste	ALA	0.001706	0.001671	-0.000035	0.000374	-0.093
ALA	WHG	Eur	Ste	ALA	0.006815	0.006984	0.000169	0.000607	0.278
ALA	Eur	ALA	Ste	ALA	0.001707	0.001671	-0.000035	0.000374	-0.095
ALA	Eur	ALA	Ste	ALA	0.001794	0.001671	-0.000123	0.000374	-0.329
ALA	Eur	ALA	Ste	ALA	0.001804	0.001671	-0.000133	0.000374	-0.354
ALA	Eur	ALA	Ste	ALA	0.001785	0.001671	-0.000114	0.000374	-0.304
ALA	WHG	Eur	Ste	ALA	0.006816	0.006984	0.000168	0.000607	0.276
ALA	Eur	ALA	Ste	ALA	0.001783	0.001671	-0.000112	0.000374	-0.298
ALA	WHG	Eur	Ste	ALA	0.006816	0.006984	0.000168	0.000607	0.278

[95% confidence intervals] - 100 different seeds



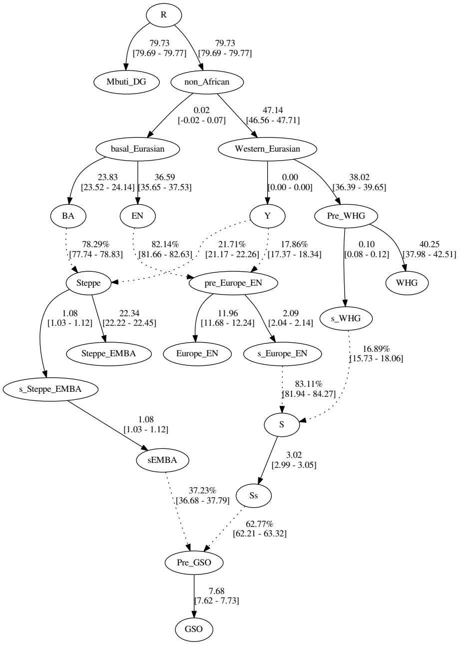
BBA	Eur	BBA	Ste	BBA	-0.000761	-0.000859	-0.000099	0.000333	-0.297
BBA	Eur	BBA	Ste	BBA	-0.000818	-0.000859	-0.000041	0.000333	-0.123
BBA	Eur	BBA	Ste	BBA	-0.000760	-0.000859	-0.000099	0.000333	-0.299
BBA	Eur	BBA	Ste	BBA	-0.000818	-0.000859	-0.000041	0.000333	-0.124
BBA	Eur	BBA	Ste	BBA	-0.000757	-0.000859	-0.000102	0.000333	-0.306
BBA	Eur	BBA	Ste	BBA	-0.000752	-0.000859	-0.000108	0.000333	-0.323
BBA	Eur	BBA	Ste	BBA	-0.000757	-0.000859	-0.000102	0.000333	-0.308
BBA	Eur	BBA	Ste	BBA	-0.000758	-0.000859	-0.000102	0.000333	-0.306
BBA	Eur	BBA	Ste	BBA	-0.000817	-0.000859	-0.000042	0.000333	-0.128
BBA	Eur	BBA	Ste	BBA	-0.000817	-0.000859	-0.000042	0.000333	-0.126
BBA	Eur	BBA	Ste	BBA	-0.000758	-0.000859	-0.000102	0.000333	-0.305
BBA	Eur	BBA	Ste	BBA	-0.000755	-0.000859	-0.000105	0.000333	-0.315
BBA	Eur	BBA	Ste	BBA	-0.000818	-0.000859	-0.000042	0.000333	-0.125
BBA	Eur	BBA	Ste	BBA	-0.000819	-0.000859	-0.000041	0.000333	-0.123
BBA	Eur	BBA	Ste	BBA	-0.000814	-0.000859	-0.000045	0.000333	-0.135
BBA	Eur	BBA	Ste	BBA	-0.000819	-0.000859	-0.000041	0.000333	-0.122
BBA	Eur	BBA	Ste	BBA	-0.000810	-0.000859	-0.000050	0.000333	-0.149

[95% confidence intervals] - 100 different seeds



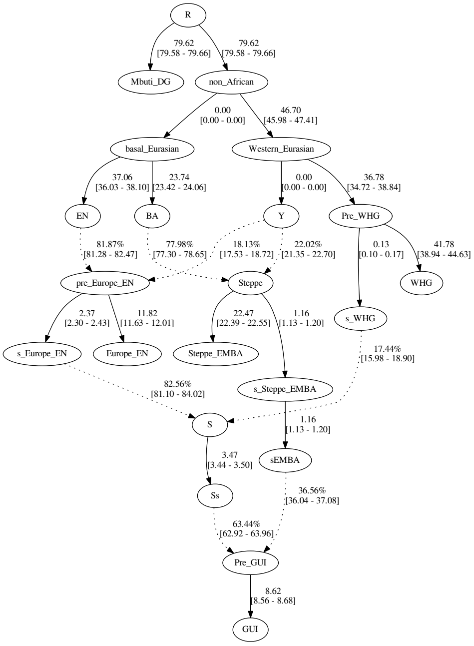
GSO	Eur	GSO	Ste	GSO	0.000944	0.000902	-0.000042	0.000351	-0.121
GSO	Eur	GSO	Ste	GSO	0.000944	0.000902	-0.000042	0.000351	-0.120
GSO	Eur	GSO	Ste	GSO	0.001020	0.000902	-0.000119	0.000351	-0.338
GSO	Eur	GSO	Ste	GSO	0.000995	0.000902	-0.000094	0.000351	-0.267
GSO	Eur	GSO	Ste	GSO	0.000954	0.000902	-0.000053	0.000351	-0.150
GSO	Eur	GSO	Ste	GSO	0.000945	0.000902	-0.000043	0.000351	-0.122
GSO	Eur	GSO	Ste	GSO	0.000945	0.000902	-0.000043	0.000351	-0.123
GSO	Eur	GSO	Ste	GSO	0.000945	0.000902	-0.000044	0.000351	-0.124
GSO	Eur	GSO	Ste	GSO	0.000947	0.000902	-0.000045	0.000351	-0.129
GSO	Eur	GSO	Ste	GSO	0.001012	0.000902	-0.000111	0.000351	-0.315
GSO	Eur	GSO	Ste	GSO	0.001017	0.000902	-0.000115	0.000351	-0.328
GSO	Eur	GSO	Ste	GSO	0.001059	0.000902	-0.000158	0.000351	-0.449
GSO	Eur	GSO	Ste	GSO	0.001001	0.000902	-0.000099	0.000351	-0.283

[95% confidence intervals] - 100 different seeds



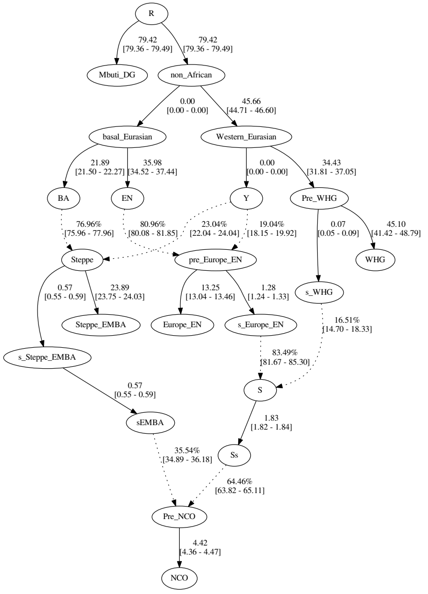
GUI	Eur	GUI	Ste	GUI	0.002238	0.002196	-0.000042	0.000359	-0.116
GUI	Eur	GUI	Ste	GUI	0.002341	0.002196	-0.000144	0.000359	-0.402
GUI	Eur	GUI	Ste	GUI	0.002238	0.002196	-0.000042	0.000359	-0.117
GUI	Eur	GUI	Ste	GUI	0.002298	0.002196	-0.000102	0.000359	-0.284
GUI	Eur	GUI	Ste	GUI	0.002315	0.002196	-0.000118	0.000359	-0.330
GUI	Eur	GUI	Ste	GUI	0.002296	0.002196	-0.000099	0.000359	-0.277
GUI	Eur	GUI	Ste	GUI	0.002238	0.002196	-0.000041	0.000359	-0.115
GUI	Eur	GUI	Ste	GUI	0.002318	0.002196	-0.000122	0.000359	-0.340
GUI	Eur	GUI	Ste	GUI	0.002310	0.002196	-0.000114	0.000359	-0.317
GUI	Eur	GUI	Ste	GUI	0.002298	0.002196	-0.000102	0.000359	-0.283
GUI	Eur	GUI	Ste	GUI	0.002244	0.002196	-0.000048	0.000359	-0.133
GUI	Eur	GUI	Ste	GUI	0.002351	0.002196	-0.000154	0.000359	-0.430
GUI	Eur	GUI	Ste	GUI	0.002237	0.002196	-0.000041	0.000359	-0.115

[95% confidence intervals] - 100 different seeds



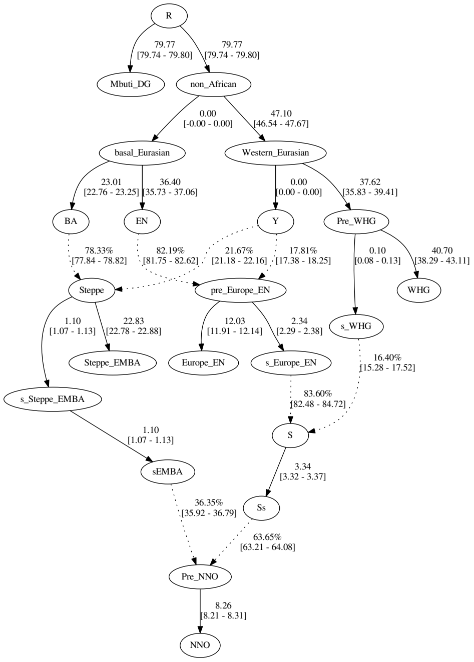
```
NCO   Eur   NCO   Ste   NCO   -0.002432   -0.002468   -0.000036   0.000348   -0.104
NCO   Eur   NCO   Ste   NCO   -0.002266   -0.002468   -0.000202   0.000348   -0.581
NCO   Eur   NCO   Ste   NCO   -0.002429   -0.002468   -0.000038   0.000348   -0.111
NCO   Eur   NCO   Ste   NCO   -0.002432   -0.002468   -0.000036   0.000348   -0.103
NCO   Eur   NCO   Ste   NCO   -0.002431   -0.002468   -0.000037   0.000348   -0.107
NCO   Eur   NCO   Ste   NCO   -0.002373   -0.002468   -0.000095   0.000348   -0.274
NCO   Eur   NCO   Ste   NCO   -0.002430   -0.002468   -0.000038   0.000348   -0.108
NCO   Eur   NCO   Ste   NCO   -0.002357   -0.002468   -0.000111   0.000348   -0.319
NCO   Eur   NCO   Ste   NCO   -0.002431   -0.002468   -0.000036   0.000348   -0.105
NCO   Eur   NCO   Ste   NCO   -0.002345   -0.002468   -0.000122   0.000348   -0.352
NCO   Eur   NCO   Ste   NCO   -0.002343   -0.002468   -0.000125   0.000348   -0.358
NCO   Eur   NCO   Ste   NCO   -0.002367   -0.002468   -0.000101   0.000348   -0.290
NCO   Eur   NCO   Ste   NCO   -0.002371   -0.002468   -0.000097   0.000348   -0.280
NCO   Eur   NCO   Ste   NCO   -0.002370   -0.002468   -0.000098   0.000348   -0.281
NCO   Eur   NCO   Ste   NCO   -0.002362   -0.002468   -0.000106   0.000348   -0.306
NCO   Eur   NCO   Ste   NCO   -0.002369   -0.002468   -0.000099   0.000348   -0.284
NCO   Eur   NCO   Ste   NCO   -0.002432   -0.002468   -0.000036   0.000348   -0.105
NCO   Eur   NCO   Ste   NCO   -0.002431   -0.002468   -0.000037   0.000348   -0.106
NCO   Eur   NCO   Ste   NCO   -0.002296   -0.002468   -0.000172   0.000348   -0.493
NCO   Eur   NCO   Ste   NCO   -0.002341   -0.002468   -0.000127   0.000348   -0.364
```

[95% confidence intervals] - 100 different seeds



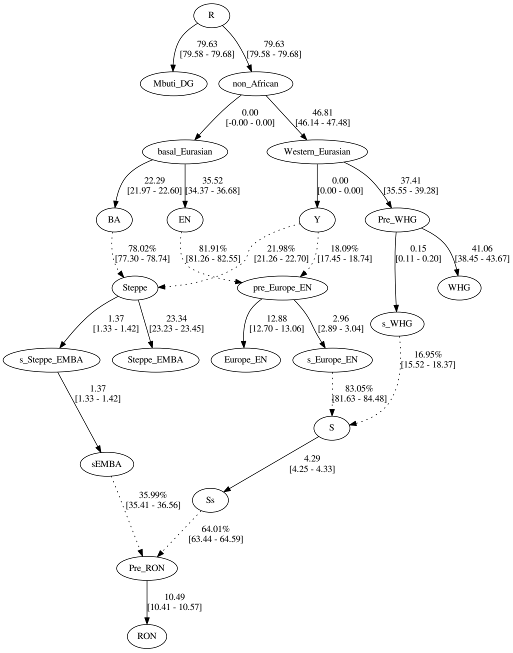
```
NNO   Eur   NNO   Ste   NNO   0.001886   0.001846   -0.000041   0.000333   -0.122
NNO   Eur   NNO   Ste   NNO   0.001946   0.001846   -0.000101   0.000333   -0.302
NNO   Eur   NNO   Ste   NNO   0.001887   0.001846   -0.000041   0.000333   -0.124
NNO   Eur   NNO   Ste   NNO   0.001886   0.001846   -0.000040   0.000333   -0.121
NNO   Eur   NNO   Ste   NNO   0.001887   0.001846   -0.000042   0.000333   -0.125
NNO   Eur   NNO   Ste   NNO   0.001887   0.001846   -0.000041   0.000333   -0.123
NNO   Eur   NNO   Ste   NNO   0.001886   0.001846   -0.000040   0.000333   -0.120
NNO   Eur   NNO   Ste   NNO   0.001948   0.001846   -0.000103   0.000333   -0.308
NNO   Eur   NNO   Ste   NNO   0.001946   0.001846   -0.000100   0.000333   -0.301
NNO   Eur   NNO   Ste   NNO   0.001956   0.001846   -0.000110   0.000333   -0.330
NNO   Eur   NNO   Ste   NNO   0.001942   0.001846   -0.000096   0.000333   -0.289
```

[95% confidence intervals] - 100 different seeds



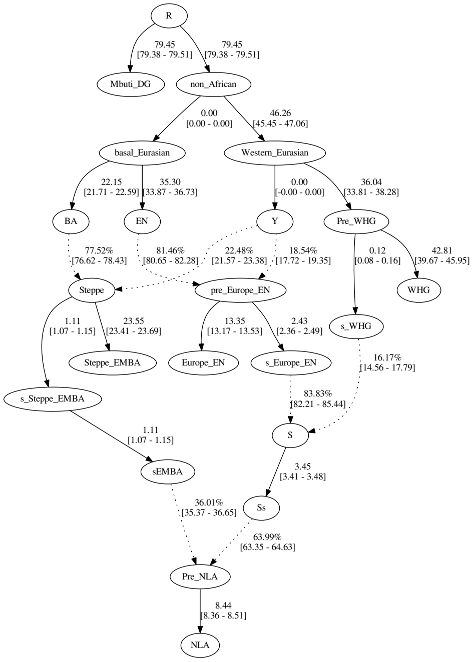
```
RON   Eur   RON   Ste   RON   0.005183   0.005147   -0.000037   0.000388   -0.095
RON   Eur   RON   Ste   RON   0.005183   0.005147   -0.000037   0.000388   -0.094
RON   Eur   RON   Ste   RON   0.005186   0.005147   -0.000040   0.000388   -0.102
RON   Eur   RON   Ste   RON   0.005185   0.005147   -0.000039   0.000388   -0.099
RON   Eur   RON   Ste   RON   0.005184   0.005147   -0.000038   0.000388   -0.097
RON   WHG   Eur   Ste   RON   0.006436   0.006595   0.000159    0.000639   0.249
RON   Eur   RON   Ste   RON   0.005355   0.005147   -0.000209   0.000388   -0.538
RON   Eur   RON   Ste   RON   0.005184   0.005147   -0.000037   0.000388   -0.096
RON   Eur   RON   Ste   RON   0.005344   0.005147   -0.000198   0.000388   -0.510
RON   Eur   RON   Ste   RON   0.005267   0.005147   -0.000120   0.000388   -0.310
RON   WHG   Ste   Eur   RON   -0.002377  -0.002296  0.000081    0.000585   0.138
RON   Eur   RON   Ste   RON   0.005184   0.005147   -0.000037   0.000388   -0.095
```

[95% confidence intervals] - 100 different seeds



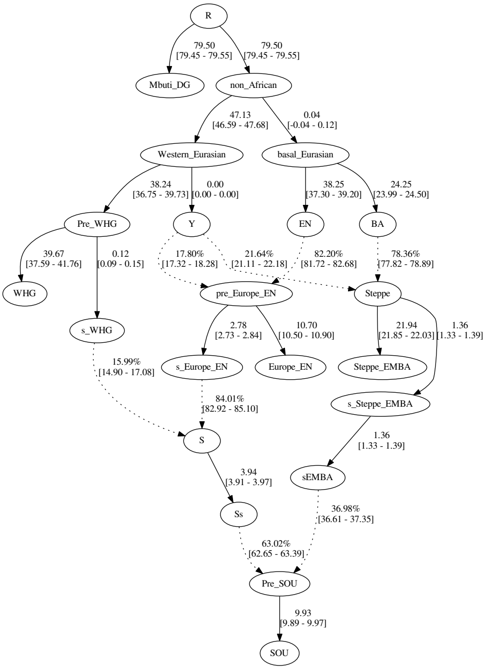
```
NLA   Eur   NLA   Ste   NLA   0.002672   0.002636   -0.000036   0.000425   -0.084
NLA   Eur   NLA   Ste   NLA   0.002961   0.002636   -0.000325   0.000425   -0.764
NLA   WHG   Eur   Ste   NLA   0.007013   0.007178   0.000165    0.000635   0.260
NLA   Eur   NLA   Ste   NLA   0.002754   0.002636   -0.000118   0.000425   -0.278
NLA   WHG   Eur   Ste   NLA   0.007014   0.007178   0.000164    0.000635   0.258
NLA   Eur   NLA   Ste   NLA   0.002672   0.002636   -0.000036   0.000425   -0.085
NLA   Eur   NLA   Ste   NLA   0.002671   0.002636   -0.000035   0.000425   -0.083
NLA   Eur   NLA   Ste   NLA   0.002689   0.002636   -0.000052   0.000425   -0.123
NLA   Eur   NLA   Ste   NLA   0.002672   0.002636   -0.000035   0.000425   -0.084
NLA   WHG   Eur   Ste   NLA   0.007014   0.007178   0.000164    0.000635   0.259
NLA   Eur   NLA   Ste   NLA   0.002842   0.002636   -0.000206   0.000425   -0.485
```

[95% confidence intervals] - 100 different seeds



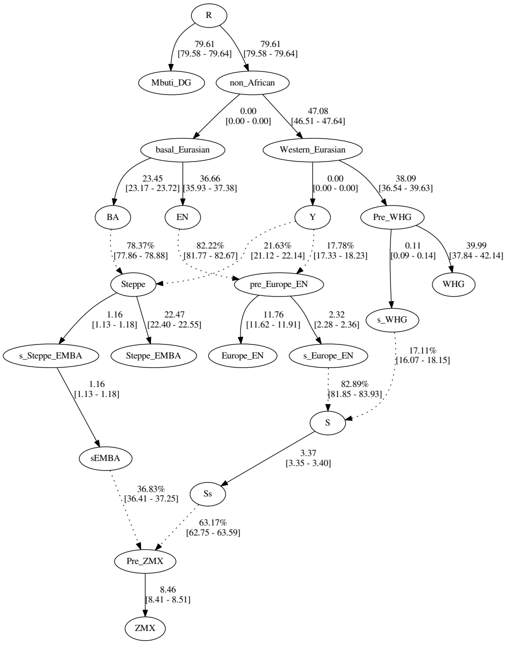
```
SOU    Eur    SOU    Ste    SOU    0.003436    0.003391   -0.000045    0.000411   -0.108
SOU    Eur    SOU    Ste    SOU    0.003436    0.003391   -0.000044    0.000411   -0.108
SOU    Eur    SOU    Ste    SOU    0.003436    0.003391   -0.000045    0.000411   -0.109
SOU    Eur    SOU    Ste    SOU    0.003436    0.003391   -0.000045    0.000411   -0.110
SOU    Eur    SOU    Ste    SOU    0.003544    0.003391   -0.000152    0.000411   -0.371
SOU    Eur    SOU    Ste    SOU    0.003549    0.003391   -0.000158    0.000411   -0.384
SOU    WHG    Ste    Eur    SOU   -0.001643   -0.001542    0.000101    0.000560    0.180
SOU    WHG    Eur    Ste    SOU    0.007659    0.007830    0.000171    0.000632    0.270
SOU    Eur    SOU    Ste    SOU    0.003466    0.003391   -0.000075    0.000411   -0.182
SOU    Eur    SOU    Ste    SOU    0.003508    0.003391   -0.000116    0.000411   -0.284
```

[95% confidence intervals] - 100 different seeds



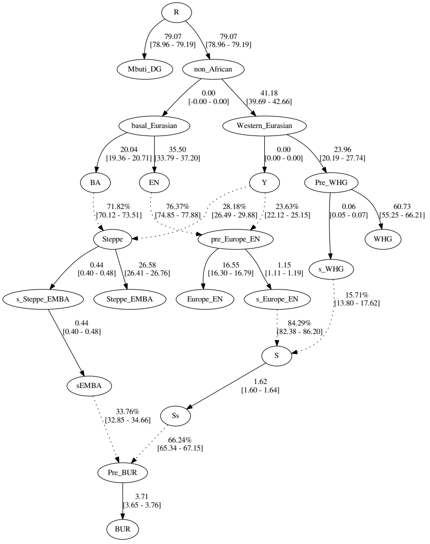
```
ZMX    Eur    ZMX    Ste    ZMX    0.002011    0.001969   -0.000042    0.000387   -0.108
ZMX    Eur    ZMX    Ste    ZMX    0.002101    0.001969   -0.000132    0.000387   -0.341
ZMX    Eur    ZMX    Ste    ZMX    0.002012    0.001969   -0.000043    0.000387   -0.110
ZMX    Eur    ZMX    Ste    ZMX    0.002011    0.001969   -0.000042    0.000387   -0.109
ZMX    Eur    ZMX    Ste    ZMX    0.002012    0.001969   -0.000043    0.000387   -0.112
ZMX    Eur    ZMX    Ste    ZMX    0.002010    0.001969   -0.000041    0.000387   -0.107
ZMX    Eur    ZMX    Ste    ZMX    0.002074    0.001969   -0.000105    0.000387   -0.272
ZMX    Eur    ZMX    Ste    ZMX    0.002068    0.001969   -0.000099    0.000387   -0.256
ZMX    Eur    ZMX    Ste    ZMX    0.002011    0.001969   -0.000042    0.000387   -0.107
ZMX    Eur    ZMX    Ste    ZMX    0.002079    0.001969   -0.000110    0.000387   -0.284
ZMX    Eur    ZMX    Ste    ZMX    0.002072    0.001969   -0.000103    0.000387   -0.267
ZMX    WHG    Ste    Eur    ZMX   -0.002316   -0.002231    0.000086    0.000552    0.155
ZMX    Eur    ZMX    Ste    ZMX    0.002013    0.001969   -0.000044    0.000387   -0.113
```

[95% confidence intervals] - 100 different seeds



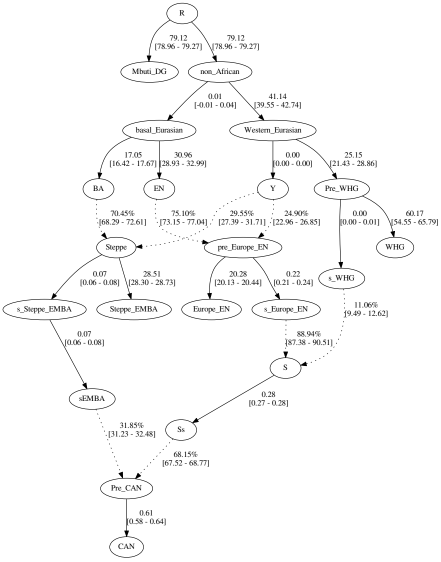
```
BUR    WHG    Ste    Eur    BUR   -0.000259   -0.000077    0.000182    0.000506    0.360
BUR    WHG    Ste    Eur    BUR   -0.000259   -0.000077    0.000181    0.000506    0.358
BUR    Eur    BUR    Ste    BUR   -0.001998   -0.002021   -0.000023    0.000333   -0.070
BUR    WHG    Ste    Eur    BUR   -0.000260   -0.000077    0.000183    0.000506    0.361
BUR    WHG    Ste    Eur    BUR   -0.000257   -0.000077    0.000180    0.000506    0.355
BUR    Eur    BUR    Ste    BUR   -0.001998   -0.002021   -0.000023    0.000333   -0.069
BUR    WHG    Ste    Eur    BUR   -0.000260   -0.000077    0.000182    0.000506    0.360
BUR    WHG    Ste    Eur    BUR   -0.000259   -0.000077    0.000182    0.000506    0.359
BUR    WHG    Ste    Eur    BUR   -0.000260   -0.000077    0.000183    0.000506    0.362
BUR    WHG    Ste    Eur    BUR   -0.000262   -0.000077    0.000184    0.000506    0.364
BUR    Eur    BUR    Ste    BUR   -0.001998   -0.002021   -0.000023    0.000333   -0.068
BUR    Eur    BUR    Ste    BUR   -0.001997   -0.002021   -0.000024    0.000333   -0.073
BUR    WHG    Ste    Eur    BUR   -0.000261   -0.000077    0.000184    0.000506    0.363
BUR    WHG    Ste    Eur    BUR   -0.000260   -0.000077    0.000182    0.000506    0.361
BUR    WHG    Ste    Eur    BUR   -0.000259   -0.000077    0.000181    0.000506    0.359
BUR    WHG    Ste    Eur    BUR   -0.000277   -0.000077    0.000199    0.000506    0.394
BUR    WHG    Ste    Eur    BUR   -0.000262   -0.000077    0.000185    0.000506    0.365
BUR    WHG    Ste    Eur    BUR   -0.000275   -0.000077    0.000198    0.000506    0.391
BUR    WHG    Ste    Eur    BUR   -0.000268   -0.000077    0.000190    0.000506    0.376
BUR    WHG    Ste    Eur    BUR   -0.000265   -0.000077    0.000188    0.000506    0.371
BUR    Eur    BUR    Ste    BUR   -0.001993   -0.002021   -0.000028    0.000333   -0.084
BUR    WHG    Ste    Eur    BUR   -0.000256   -0.000077    0.000178    0.000506    0.352
BUR    Eur    BUR    Ste    BUR   -0.001997   -0.002021   -0.000023    0.000333   -0.070
BUR    Eur    BUR    Ste    BUR   -0.001967   -0.002021   -0.000053    0.000333   -0.161
BUR    WHG    Ste    Eur    BUR   -0.000263   -0.000077    0.000185    0.000506    0.367
BUR    WHG    Ste    Eur    BUR   -0.000258   -0.000077    0.000181    0.000506    0.357
BUR    WHG    Ste    Eur    BUR   -0.000258   -0.000077    0.000181    0.000506    0.358
BUR    Eur    BUR    Ste    BUR   -0.001996   -0.002021   -0.000025    0.000333   -0.074
```

[95% confidence intervals] - 100 different seeds



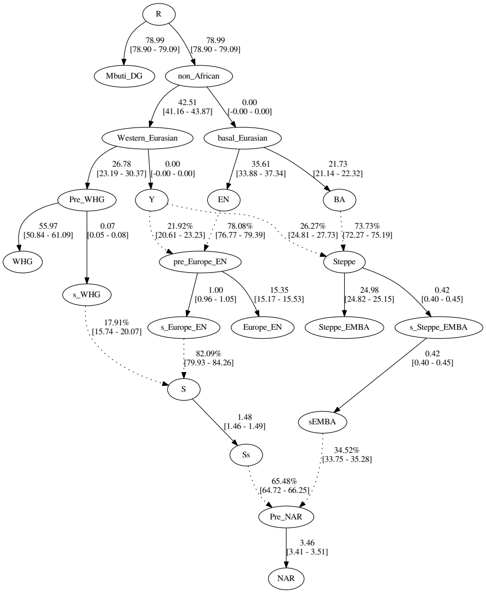
```
CAN    Eur    Ste    Eur    Ste     0.073457    0.073487    0.000030    0.000664    0.045
CAN    WHG    Ste    Eur    CAN    -0.000024    0.000219    0.000244    0.000497    0.490
CAN    WHG    Ste    Eur    CAN    -0.000022    0.000219    0.000241    0.000497    0.485
CAN    WHG    Ste    Eur    CAN    -0.000025    0.000219    0.000244    0.000497    0.491
CAN    Eur    Ste    Eur    Ste     0.073457    0.073487    0.000029    0.000664    0.044
CAN    Eur    Ste    Eur    Ste     0.073456    0.073487    0.000031    0.000664    0.046
CAN    Eur    Ste    Ste    CAN    -0.043907   -0.043929   -0.000022    0.000489   -0.045
CAN    WHG    Ste    Eur    CAN    -0.000013    0.000219    0.000232    0.000497    0.467
CAN    WHG    Ste    Eur    CAN    -0.000025    0.000219    0.000244    0.000497    0.490
CAN    WHG    Ste    Eur    CAN    -0.000024    0.000219    0.000243    0.000497    0.489
CAN    Eur    CAN    Ste    CAN    -0.004547   -0.004779   -0.000232    0.000289   -0.803
CAN    WHG    Ste    Eur    CAN    -0.000024    0.000219    0.000243    0.000497    0.488
CAN    WHG    Ste    Eur    CAN    -0.000025    0.000219    0.000244    0.000497    0.492
CAN    Eur    Ste    Ste    CAN    -0.043906   -0.043929   -0.000022    0.000489   -0.046
CAN    WHG    Ste    Eur    CAN    -0.000023    0.000219    0.000242    0.000497    0.487
CAN    WHG    Ste    Eur    CAN    -0.000028    0.000219    0.000247    0.000497    0.496
CAN    WHG    Ste    Eur    CAN    -0.000026    0.000219    0.000245    0.000497    0.492
CAN    WHG    Ste    Eur    CAN    -0.000027    0.000219    0.000246    0.000497    0.494
CAN    Eur    CAN    Ste    CAN    -0.004549   -0.004779   -0.000229    0.000289   -0.795
CAN    WHG    Ste    Eur    CAN    -0.000033    0.000219    0.000252    0.000497    0.507
CAN    WHG    Ste    Eur    CAN    -0.000027    0.000219    0.000246    0.000497    0.495
CAN    Eur    CAN    Ste    CAN    -0.004752   -0.004779   -0.000027    0.000289   -0.092
CAN    Eur    Ste    Eur    Ste     0.073456    0.073487    0.000030    0.000664    0.045
CAN    WHG    Ste    Eur    CAN    -0.000025    0.000219    0.000245    0.000497    0.492
CAN    WHG    Ste    Eur    CAN    -0.000028    0.000219    0.000247    0.000497    0.497
```

[95% confidence intervals] - 100 different seeds



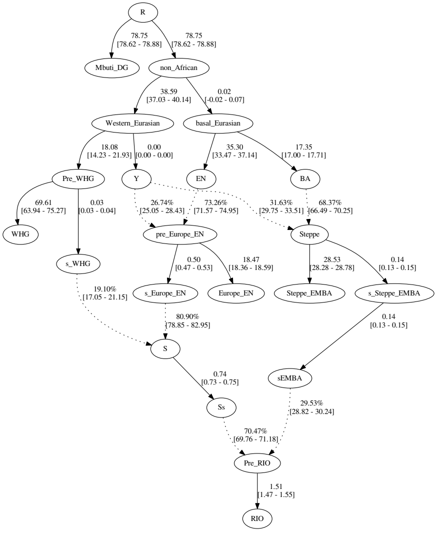
NAR	Eur	NAR	Ste	NAR	-0.002904	-0.002934	-0.000030	0.000328	-0.090
NAR	WHG	Ste	Eur	NAR	-0.000722	-0.000569	0.000154	0.000509	0.302
NAR	WHG	Ste	Eur	NAR	-0.000723	-0.000569	0.000155	0.000509	0.304
NAR	Eur	NAR	Ste	NAR	-0.002904	-0.002934	-0.000029	0.000328	-0.090
NAR	Eur	NAR	Ste	NAR	-0.002904	-0.002934	-0.000030	0.000328	-0.092
NAR	Eur	NAR	Ste	NAR	-0.002828	-0.002934	-0.000106	0.000328	-0.323
NAR	Eur	NAR	Ste	NAR	-0.002813	-0.002934	-0.000121	0.000328	-0.368
NAR	Eur	NAR	Ste	NAR	-0.002834	-0.002934	-0.000100	0.000328	-0.304
NAR	WHG	Ste	Eur	NAR	-0.000723	-0.000569	0.000154	0.000509	0.303
NAR	Eur	NAR	Ste	NAR	-0.002811	-0.002934	-0.000123	0.000328	-0.375
NAR	WHG	Ste	Eur	NAR	-0.000722	-0.000569	0.000153	0.000509	0.301
NAR	Eur	NAR	Ste	NAR	-0.002904	-0.002934	-0.000030	0.000328	-0.091
NAR	Eur	NAR	Ste	NAR	-0.002902	-0.002934	-0.000032	0.000328	-0.097
NAR	Eur	NAR	Ste	NAR	-0.002904	-0.002934	-0.000029	0.000328	-0.089
NAR	Eur	NAR	Ste	NAR	-0.002901	-0.002934	-0.000033	0.000328	-0.099
NAR	Eur	NAR	Ste	NAR	-0.002811	-0.002934	-0.000123	0.000328	-0.374
NAR	Eur	NAR	Ste	NAR	-0.002719	-0.002934	-0.000215	0.000328	-0.655
NAR	WHG	Ste	Eur	NAR	-0.000750	-0.000569	0.000182	0.000509	0.357
NAR	Eur	NAR	Ste	NAR	-0.002900	-0.002934	-0.000033	0.000328	-0.102
NAR	Eur	NAR	Ste	NAR	-0.002903	-0.002934	-0.000030	0.000328	-0.092

[95% confidence intervals] - 100 different seeds



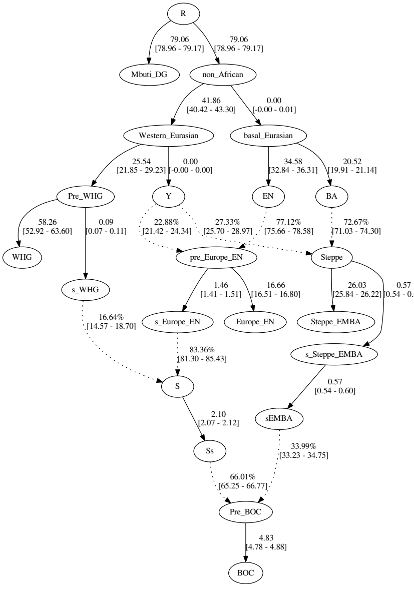
RIO	Eur	Ste	Eur	Ste	0.073219	0.073251	0.000033	0.000662	0.049
RIO	WHG	Ste	Eur	RIO	-0.001099	-0.000931	0.000168	0.000524	0.321
RIO	WHG	Ste	Eur	RIO	-0.001098	-0.000931	0.000168	0.000524	0.320
RIO	WHG	Ste	Eur	RIO	-0.001099	-0.000931	0.000168	0.000524	0.320
RIO	WHG	Ste	Eur	RIO	-0.001103	-0.000931	0.000172	0.000524	0.329
RIO	Eur	Ste	Eur	Ste	0.073219	0.073251	0.000033	0.000662	0.050
RIO	WHG	Ste	Eur	RIO	-0.001103	-0.000931	0.000172	0.000524	0.328
RIO	WHG	Ste	Eur	RIO	-0.001098	-0.000931	0.000167	0.000524	0.318
RIO	WHG	Ste	Eur	RIO	-0.001106	-0.000931	0.000175	0.000524	0.335
RIO	WHG	Ste	Eur	RIO	-0.001119	-0.000931	0.000188	0.000524	0.360
RIO	WHG	Eur	Ste	RIO	0.007797	0.007974	0.000177	0.000577	0.306
RIO	Eur	Ste	Ste	RIO	-0.044258	-0.044283	-0.000025	0.000490	-0.051
RIO	WHG	Ste	Eur	RIO	-0.001102	-0.000931	0.000171	0.000524	0.326
RIO	Eur	Ste	Ste	RIO	-0.044257	-0.044283	-0.000026	0.000490	-0.053
RIO	WHG	Ste	Eur	RIO	-0.001113	-0.000931	0.000182	0.000524	0.347
RIO	Eur	RIO	Ste	RIO	-0.003576	-0.003739	-0.000164	0.000330	-0.496
RIO	WHG	Ste	Eur	RIO	-0.001099	-0.000931	0.000168	0.000524	0.322
RIO	WHG	Ste	Eur	RIO	-0.001100	-0.000931	0.000169	0.000524	0.323
RIO	Eur	Ste	Ste	RIO	-0.044259	-0.044283	-0.000024	0.000490	-0.049
RIO	WHG	Ste	Eur	RIO	-0.001105	-0.000931	0.000174	0.000524	0.332
RIO	WHG	Ste	Eur	RIO	-0.001098	-0.000931	0.000167	0.000524	0.319
RIO	WHG	Ste	Eur	RIO	-0.001109	-0.000931	0.000178	0.000524	0.340
RIO	WHG	Ste	Eur	RIO	-0.001107	-0.000931	0.000176	0.000524	0.336
RIO	WHG	Ste	Eur	RIO	-0.001098	-0.000931	0.000167	0.000524	0.320
RIO	WHG	Ste	Eur	RIO	-0.001101	-0.000931	0.000170	0.000524	0.324
RIO	WHG	Eur	Ste	RIO	0.007796	0.007974	0.000177	0.000577	0.307
RIO	WHG	Ste	Eur	RIO	-0.001106	-0.000931	0.000175	0.000524	0.334
RIO	WHG	Ste	Eur	RIO	-0.001100	-0.000931	0.000170	0.000524	0.324
RIO	WHG	Eur	Ste	RIO	0.007791	0.007974	0.000183	0.000577	0.317
RIO	WHG	Ste	Eur	RIO	-0.001092	-0.000931	0.000161	0.000524	0.307

[95% confidence intervals] - 100 different seeds



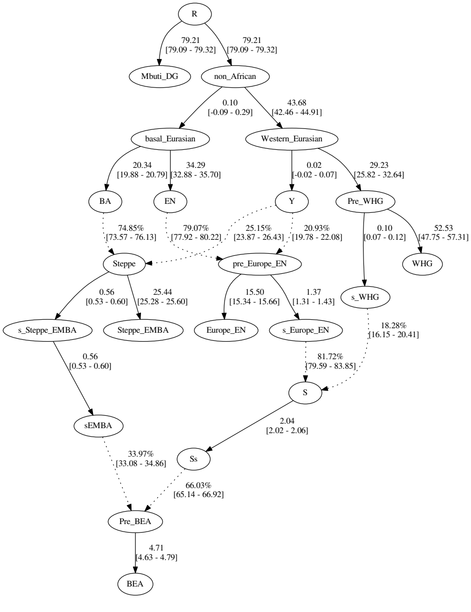
BOC	Eur	BOC	Ste	BOC	-0.000605	-0.000629	-0.000024	0.000349	-0.069
BOC	WHG	Eur	Ste	BOC	0.007855	0.008038	0.000183	0.000621	0.294
BOC	WHG	Ste	Eur	BOC	-0.000498	-0.000332	0.000166	0.000562	0.295
BOC	WHG	Ste	Eur	BOC	-0.000499	-0.000332	0.000167	0.000562	0.297
BOC	Eur	BOC	Ste	BOC	-0.000605	-0.000629	-0.000024	0.000349	-0.068
BOC	Eur	BOC	Ste	BOC	-0.000605	-0.000629	-0.000024	0.000349	-0.070
BOC	WHG	Ste	Eur	BOC	-0.000498	-0.000332	0.000166	0.000562	0.296
BOC	Eur	BOC	Ste	BOC	-0.000474	-0.000629	-0.000155	0.000349	-0.444
BOC	Eur	BOC	Ste	BOC	-0.000469	-0.000629	-0.000160	0.000349	-0.458
BOC	Eur	BOC	Ste	BOC	-0.000603	-0.000629	-0.000026	0.000349	-0.074
BOC	Eur	BOC	Ste	BOC	-0.000588	-0.000629	-0.000041	0.000349	-0.116
BOC	Eur	BOC	Ste	BOC	-0.000510	-0.000629	-0.000119	0.000349	-0.340
BOC	Eur	BOC	Ste	BOC	-0.000605	-0.000629	-0.000025	0.000349	-0.070
BOC	WHG	Ste	Eur	BOC	-0.000499	-0.000332	0.000168	0.000562	0.298
BOC	WHG	Ste	Eur	BOC	-0.000497	-0.000332	0.000165	0.000562	0.294
BOC	WHG	Eur	Ste	BOC	0.007854	0.008038	0.000183	0.000621	0.295
BOC	WHG	Eur	Ste	BOC	0.007850	0.008038	0.000188	0.000621	0.302
BOC	Eur	BOC	Ste	BOC	-0.000520	-0.000629	-0.000109	0.000349	-0.312
BOC	WHG	Ste	Eur	BOC	-0.000499	-0.000332	0.000167	0.000562	0.298
BOC	WHG	Eur	Ste	BOC	0.007854	0.008038	0.000184	0.000621	0.296
BOC	WHG	Ste	Eur	BOC	-0.000497	-0.000332	0.000166	0.000562	0.295

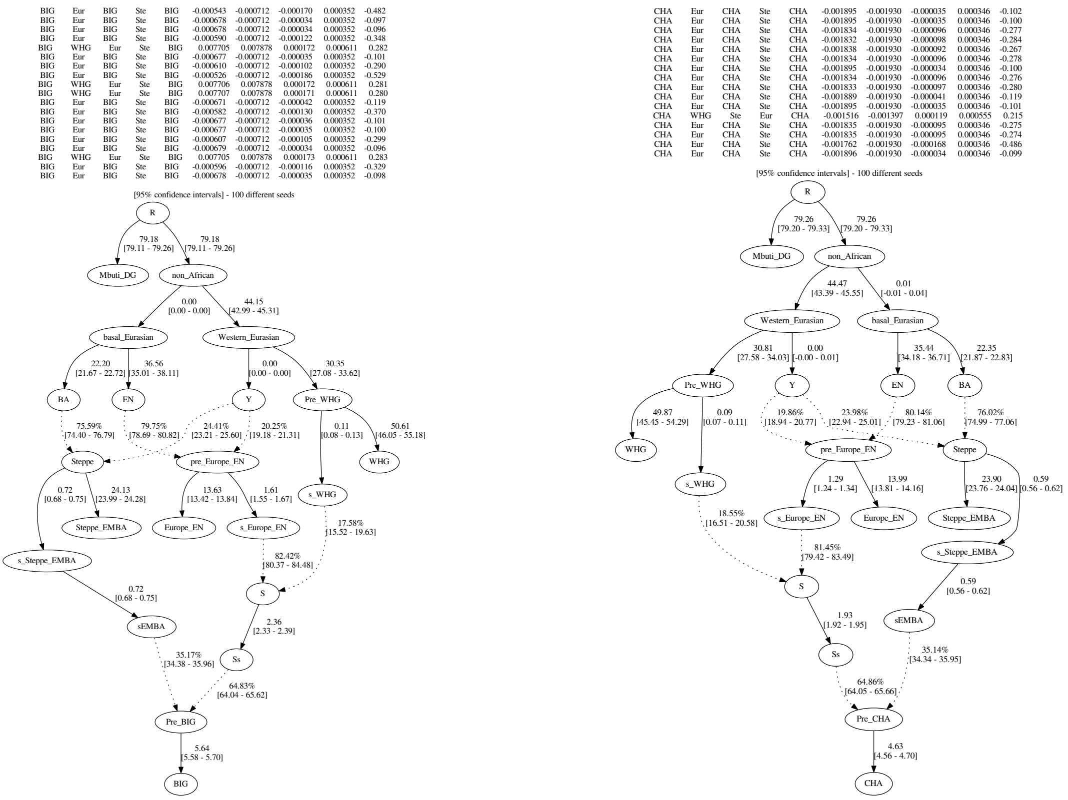
[95% confidence intervals] - 100 different seeds



BEA	Eur	BEA	Ste	BEA	-0.001016	-0.001043	-0.000028	0.000366	-0.076
BEA	WHG	Eur	Ste	BEA	0.006871	0.007033	0.000162	0.000621	0.261
BEA	WHG	Eur	Ste	BEA	0.006870	0.007033	0.000163	0.000621	0.262
BEA	Eur	BEA	Ste	BEA	-0.001014	-0.001043	-0.000029	0.000366	-0.080
BEA	Eur	BEA	Ste	BEA	-0.001015	-0.001043	-0.000028	0.000366	-0.076
BEA	WHG	Eur	Ste	BEA	0.006871	0.007033	0.000162	0.000621	0.262
BEA	WHG	Eur	Ste	BEA	0.006871	0.007033	0.000162	0.000621	0.260
BEA	WHG	Eur	Ste	BEA	0.006872	0.007033	0.000162	0.000621	0.260
BEA	Eur	BEA	Ste	BEA	-0.001016	-0.001043	-0.000027	0.000366	-0.075
BEA	Eur	BEA	Ste	BEA	-0.001015	-0.001043	-0.000028	0.000366	-0.077
BEA	WHG	Eur	Ste	BEA	0.006872	0.007033	0.000161	0.000621	0.260
BEA	Eur	BEA	Ste	BEA	-0.001014	-0.001043	-0.000029	0.000366	-0.079
BEA	WHG	Eur	Ste	BEA	0.006872	0.007033	0.000161	0.000621	0.259
BEA	Eur	BEA	Ste	BEA	-0.000936	-0.001043	-0.000108	0.000366	-0.294
BEA	Eur	BEA	Ste	BEA	-0.001012	-0.001043	-0.000032	0.000366	-0.087
BEA	Eur	BEA	Ste	BEA	-0.000912	-0.001043	-0.000131	0.000366	-0.358
BEA	Eur	BEA	Ste	BEA	-0.000947	-0.001043	-0.000096	0.000366	-0.262
BEA	Eur	BEA	Ste	BEA	-0.000935	-0.001043	-0.000109	0.000366	-0.296
BEA	Eur	BEA	Ste	BEA	-0.001014	-0.001043	-0.000030	0.000366	-0.081

[95% confidence intervals] - 100 different seeds





```
BIG     Eur     BIG     Ste     BIG     -0.000543   -0.000712   -0.000170   0.000352   -0.482
BIG     Eur     BIG     Ste     BIG     -0.000678   -0.000712   -0.000034   0.000352   -0.097
BIG     Eur     BIG     Ste     BIG     -0.000678   -0.000712   -0.000034   0.000352   -0.096
BIG     Eur     BIG     Ste     BIG     -0.000590   -0.000712   -0.000122   0.000352   -0.348
BIG     WHG     Eur     Ste     BIG      0.007705    0.007878    0.000172   0.000611    0.282
BIG     Eur     BIG     Ste     BIG     -0.000677   -0.000712   -0.000035   0.000352   -0.101
BIG     Eur     BIG     Ste     BIG     -0.000610   -0.000712   -0.000102   0.000352   -0.290
BIG     Eur     BIG     Ste     BIG     -0.000526   -0.000712   -0.000186   0.000352   -0.529
BIG     WHG     Eur     Ste     BIG      0.007706    0.007878    0.000172   0.000611    0.281
BIG     WHG     Eur     Ste     BIG      0.007707    0.007878    0.000171   0.000611    0.280
BIG     Eur     BIG     Ste     BIG     -0.000671   -0.000712   -0.000042   0.000352   -0.119
BIG     Eur     BIG     Ste     BIG     -0.000582   -0.000712   -0.000130   0.000352   -0.370
BIG     Eur     BIG     Ste     BIG     -0.000677   -0.000712   -0.000036   0.000352   -0.101
BIG     Eur     BIG     Ste     BIG     -0.000677   -0.000712   -0.000035   0.000352   -0.100
BIG     Eur     BIG     Ste     BIG     -0.000607   -0.000712   -0.000105   0.000352   -0.299
BIG     Eur     BIG     Ste     BIG     -0.000679   -0.000712   -0.000034   0.000352   -0.096
BIG     WHG     Eur     Ste     BIG      0.007705    0.007878    0.000173   0.000611    0.283
BIG     Eur     BIG     Ste     BIG     -0.000596   -0.000712   -0.000116   0.000352   -0.329
BIG     Eur     BIG     Ste     BIG     -0.000678   -0.000712   -0.000035   0.000352   -0.098
```

[95% confidence intervals] - 100 different seeds

```
CHA     Eur     CHA     Ste     CHA     -0.001895   -0.001930   -0.000035   0.000346   -0.102
CHA     Eur     CHA     Ste     CHA     -0.001895   -0.001930   -0.000035   0.000346   -0.100
CHA     Eur     CHA     Ste     CHA     -0.001834   -0.001930   -0.000096   0.000346   -0.277
CHA     Eur     CHA     Ste     CHA     -0.001832   -0.001930   -0.000098   0.000346   -0.284
CHA     Eur     CHA     Ste     CHA     -0.001838   -0.001930   -0.000092   0.000346   -0.267
CHA     Eur     CHA     Ste     CHA     -0.001834   -0.001930   -0.000096   0.000346   -0.278
CHA     Eur     CHA     Ste     CHA     -0.001895   -0.001930   -0.000034   0.000346   -0.100
CHA     Eur     CHA     Ste     CHA     -0.001834   -0.001930   -0.000096   0.000346   -0.276
CHA     Eur     CHA     Ste     CHA     -0.001833   -0.001930   -0.000097   0.000346   -0.280
CHA     Eur     CHA     Ste     CHA     -0.001889   -0.001930   -0.000041   0.000346   -0.119
CHA     Eur     CHA     Ste     CHA     -0.001895   -0.001930   -0.000035   0.000346   -0.101
CHA     WHG     Ste     Eur     CHA     -0.001516   -0.001397    0.000119   0.000555    0.215
CHA     Eur     CHA     Ste     CHA     -0.001835   -0.001930   -0.000095   0.000346   -0.275
CHA     Eur     CHA     Ste     CHA     -0.001835   -0.001930   -0.000095   0.000346   -0.274
CHA     Eur     CHA     Ste     CHA     -0.001762   -0.001930   -0.000168   0.000346   -0.486
CHA     Eur     CHA     Ste     CHA     -0.001896   -0.001930   -0.000034   0.000346   -0.099
```

[95% confidence intervals] - 100 different seeds

Figure S8. qpGraph models and inferred WHG, Europe_EN and Steppe_EMBA mixture proportions for each modern target groups. 100 permutations were performed in each case and the 95 % confidence intervals were calculated. The outlier f4-statistics with the worst Z-scores are indicated in the header of each graph. All of them were statistically significant with Z-scores between 3 and -3.

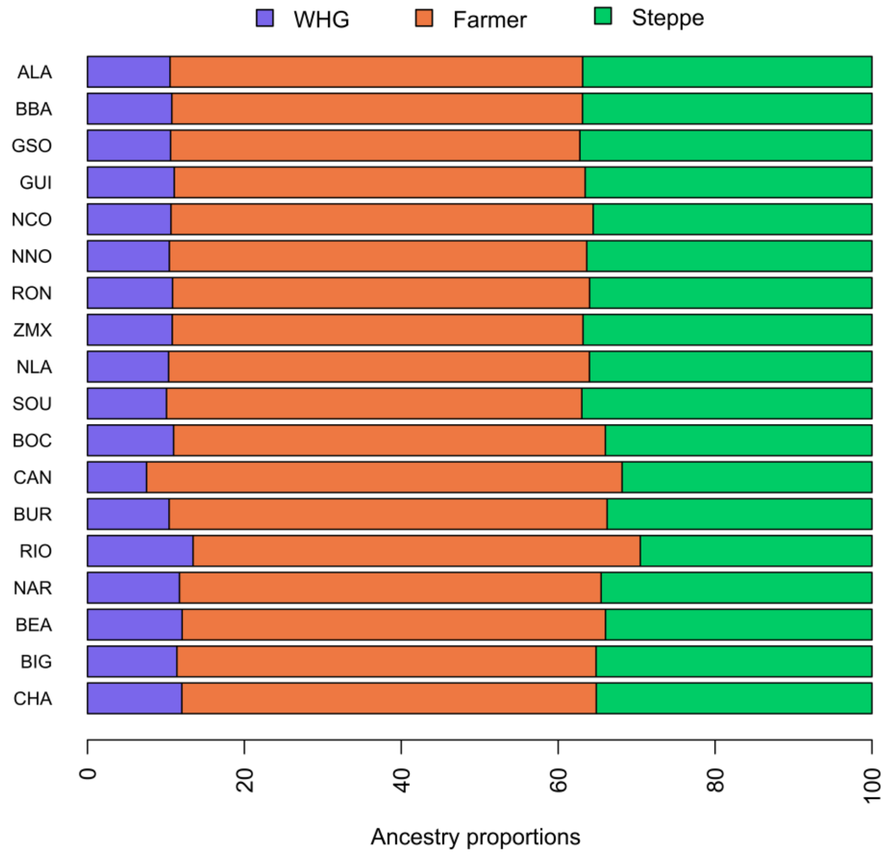


Figure S9. Admixture proportions of HG, farmer and Steppe modeled in Franco-Cantabrian groups with qpGraph.

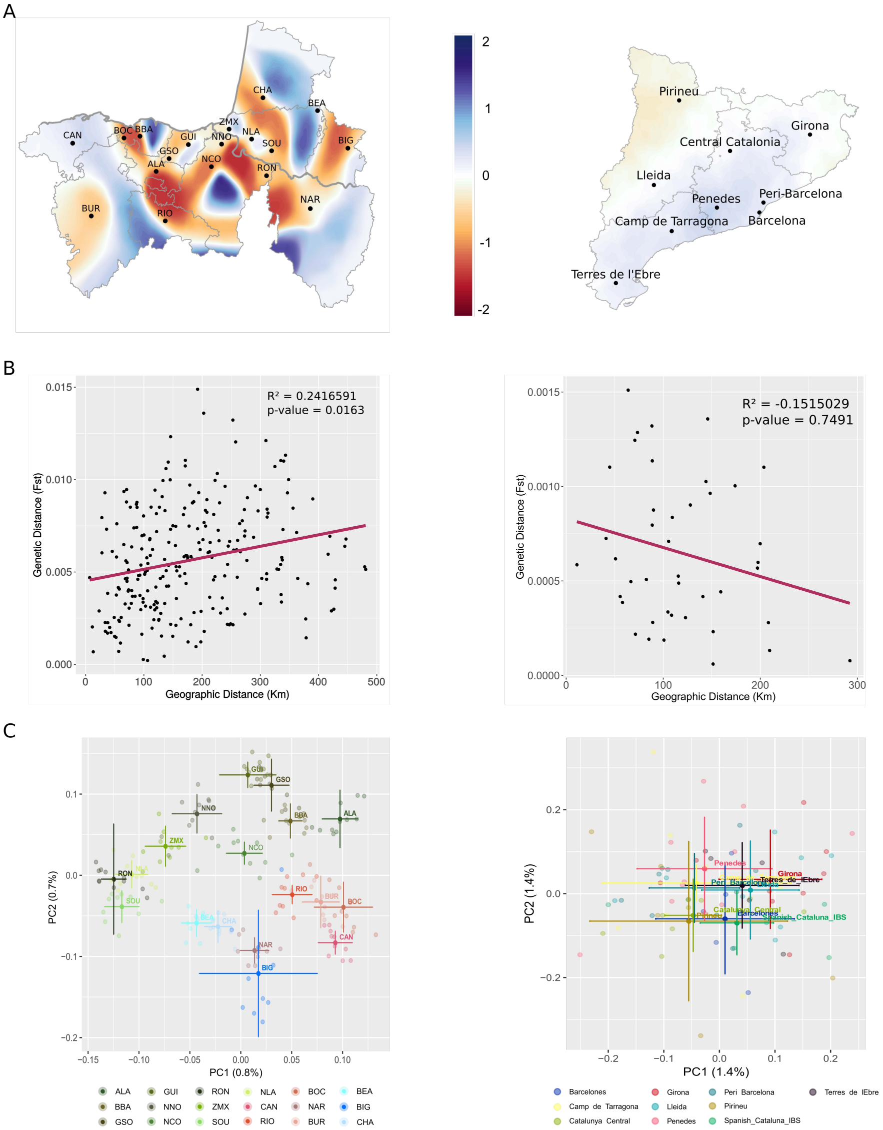
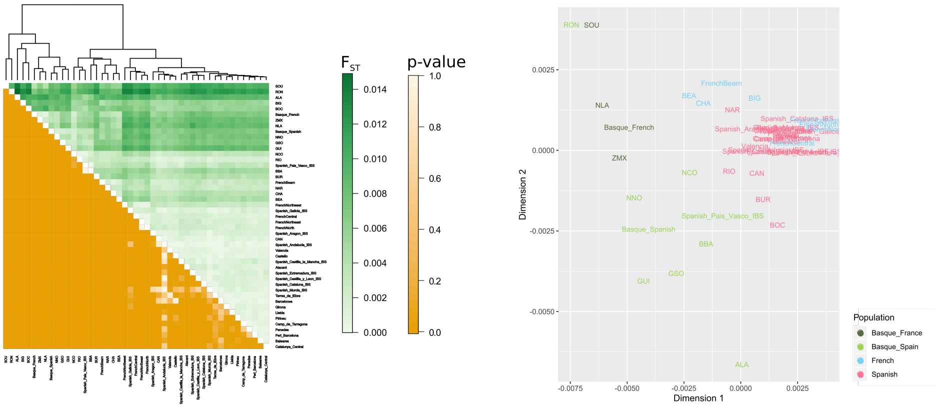
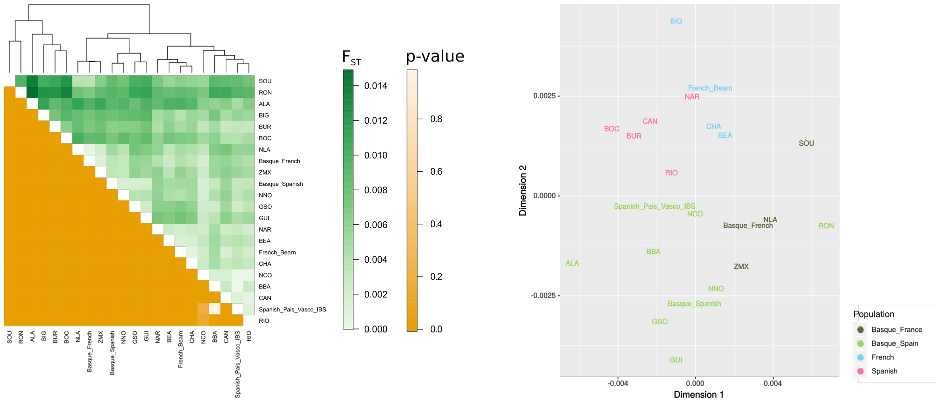


Figure S10. Comparing the Franco-Cantabrian region with Catalonia, as reference population of similar size and high-resolution sampling. In the left, the analysis in the Franco-Cantabrian region. In the right, the analysis in Catalonia. (A) EEMS analysis result. The posterior mean migration rates on the log10 scale are represented in the map. Values >0 indicate corridors, whereas <0 evince barriers. (B) The correlation between the pairwise genetic (F_{ST}) and geographic (Km) distances, together with the Mantel test result. (C) Principal component analyses.

A. Iberian Peninsula and France



B. Franco-Cantabrian region



C. Catalonia

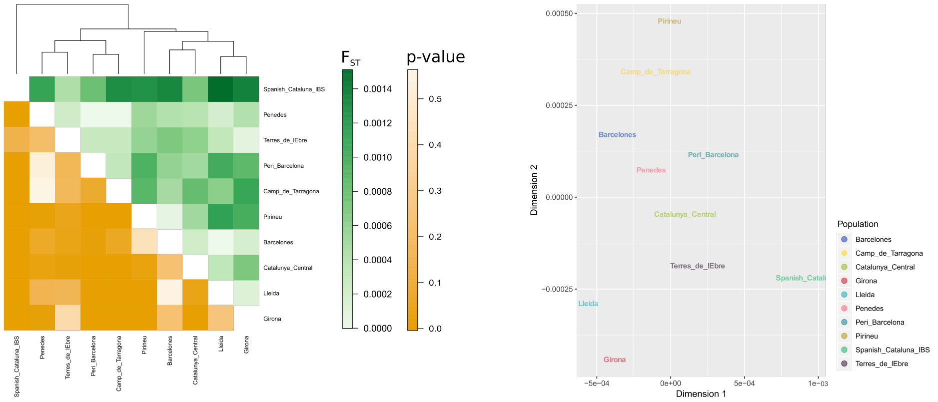
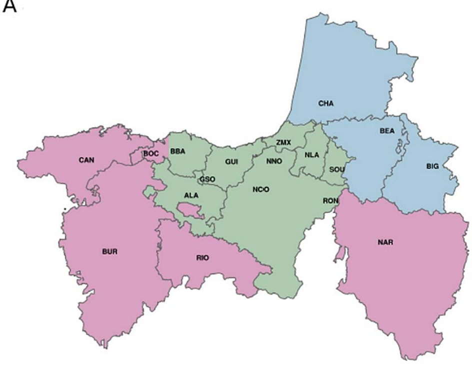


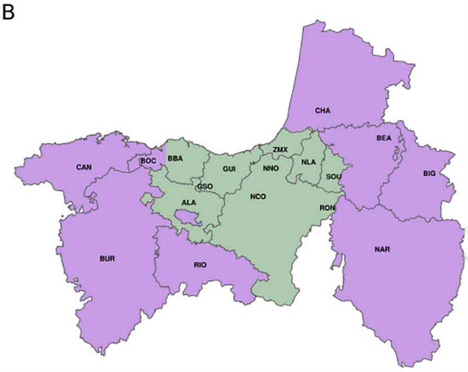
Figure S11. Pairwise genetic distance analyses at different levels in the dataset: (A) the Iberian Peninsula and France, (B) the Franco-Cantabrian region and (C) Catalonia. In the left, the heatmaps and dendrograms based on the pairwise F_{ST} values (above the diagonal), and related p-values (below the diagonal). In the right, multidimensional scaling analyses performed with these genetic distances.

A



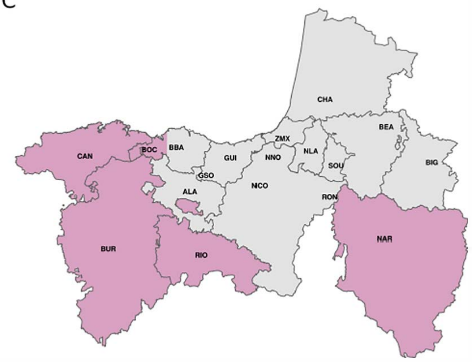
	Explained variance	P-value
Variation between groups	0.141	0.002
Variation between populations within group	0.536	0.001
Variation within populations	99.323	0.001

B



	Explained variance	P-value
Variation between groups	0.128	0.001
Variation between populations within group	0.555	0.001
Variation within populations	99.317	0.001

C

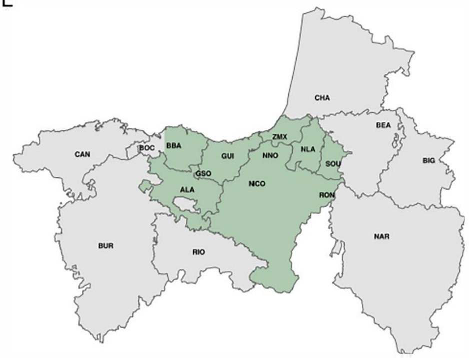


	Explainedvariance	P-value
Variation between populations	0.366	0.001
Variation within populations	99.634	-

D

	Explained variance	P-value
Variation between population	0.402	0.001
Variation within populations	99.598	-

E



	Explained variance	P-value
Variation between populations	0.641	0.001
Variation within populations	99.359	-

Figure S12. AMOVA analyses in the Franco-Cantabrian region. Different hypotheses were tested: (A) Three differentiated major groups for Spanish Peri-Basques, French Peri-Basque and Basques. Variation explained in the different levels is statistically significant, evidencing heterogeneity in the region. (B) Two major groups for Basques and non-Basques in order to see if the variation explained between groups descended. The results are also significant, with higher variation explained within groups due to the fact of considering Spanish and French Peri-Basques together. This shows evidences of heterogeneity in the Peri-Basques groups. Then, in order to see which major group explained more variation within, separately AMOVA tests were performed for: (C) Spanish Peri-Basques, (D) French Beri-Basques and (E) Basques. The higher value is observed in Basques.

	Explained variance	P-value
Variation between groups	0.012	0.268
Variation between populations within group	0.042	0.047
Variation within populations	99.946	0.008

Figure S13. AMOVA analysis in Catalonia. The analysis was performed in two strata as performed in the Franco-Cantabrian region: provinces and regions within provinces.

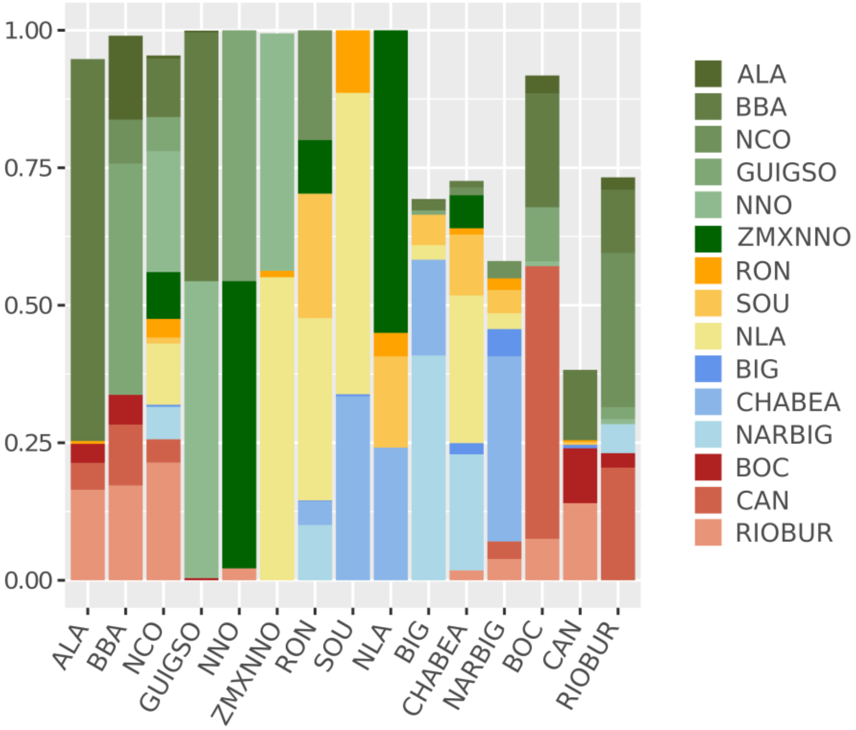


Figure S14. Inferred proportions of shared ancestry among the Franco-Cantabrian groups. This plot is in continuity with the main figure 3D. Here, we dissect the internal proportions of haplotype sharing in the Franco-Cantabrian region.

4. DISCUSSION

Population genetics’ studies including Basque samples tend to be burdened by preconceived assumptions about their differentiation, heterogeneity and origins, which have not been completely solved in previous studies, as it was reflected in chapter 1.6. Given the interest of Basques, not only in population genetics, but also in other fields as linguistics or anthropology, it is necessary to assess the main questions that have been related to them during the last decades.

The first challenge of any population genetics approach is the definition of the body of the study, the population (Kreager et al. 2015). In the present thesis I have defined our populations based on their geographical origins and I have added other variables such as the language spoken by the volunteers. The definition used in this thesis for “Basques”, as described in chapter 3.2., includes those samples coming from geographical areas where Euskara was spoken in historical times. This definition overcomes administrative or political classifications, although some overlap exists. Therefore, the term “Basque” used in this thesis should be taken within the frame of the population genetics field, and cannot be directly extrapolated to any political or administrative classification.

In this chapter, I will discuss about the different results I have achieved during my thesis to elucidate the different controversies of Basques, by improving the methodology and the sampling.

4.1. The Rhesus system in Basques from a new perspective

The analysis of the Rh system in Basques had been usually studied by classical serologic methods and the basic description of the antigens detected (Van Der Heide, Magnée, and Van Loghem 1951; Chalmers, Ikin, and Mourant 1949; Manzano et al. 2002). This procedure presents a limited accuracy, and sometimes could lead to false positive or negative results (Flegel 2011). By genotyping the variants associated with the main phenotypes, those limitations are overcome and more precise and reliable results are obtained (Westhoff 2007). Furthermore, the genetic landscape of the system in Basques had never been studied before at sequence level, that together with the subsequent demographic analyses performed, could elucidate specific patterns in the population to better understand the causes of the high frequencies of Rh-negative (Perry et al. 2012). It is complicated to study the system through the gene sequence, due to the high identity between both *RHD* and *RHCE* genes. Indeed, when considering the Rh-negative and Rh-positive phenotypes in the analysis, it becomes impossible to accurately identify the alleles since the Rh-negative comes from the homozygous deletion of the *RHD* gene (Avent and Reid 2000; Wagner and Flegel 2000; Perry et al. 2012). However, the flanking regions can mirror their genetic history and give information of the processes that could have acted along the system.

Through serologic analysis the frequencies of Rh variants have been widely studied worldwide. The Rh-negative phenotype has been reported worldwide to be lower in Asian groups, followed by the indigenous population from America and the Pacific Area, and Africans. In Europe, the reported frequencies have been detected to be higher, especially in Basques, outstanding with frequencies close to 50% (Daniels 2013; Chalmers, Ikin, and Mourant 1949). However, no remarkable singular patterns of allele, haplotype, or sequence diversity have been detected from the genotyping and sequencing analyses. The Rh-negative frequencies reported with the DNA sequences, though still higher in Basques and within the previously reported range, are not as extreme as usually shown by serologic methods (Flores-Bello et al. 2018). This result contrasts with the clear specific genetic differentiation observed in Basques with the genome-wide analysis (Flores-Bello et al., in preparation). This evokes the importance of using a proper genome-wide representation when studying demographic patterns, in order to describe the general picture of the genome and not just the history of a specific marker.

It is complicated to contextualize the origins of the higher frequency of the Rh-negative in Basques. Besides not observing specific patterns in the results, the information about the function of the *RH* genes is extremely scant, which does not allow to suggest hypotheses related to the adaptive value of the genes. The frequency of Rh variants has been suggested to be subject to a frequency-dependent equilibrium in favor of the most frequent variant (Vogel and Motulsky 1997; Flegr 2016; Cavalli-Sforza 1988). It has been proposed to have originated during the LGM due to a bottleneck and subsequent inbreeding when the Franco-Cantabrian region acted as glacial refuge (Anstee 2010). However, this is just a hypothesis with no direct evidences. In the LGM or not, signals of an old isolation are evinced based on the ROHs analyses, but farmer and steppe components are also detected in Basques (Flores-Bello et al., in preparation). Thus, if a demographic process had been the cause of the high frequency of Rh-negative, it should probably have been after admixing with Neolithic and Bronze Age incomers. Nonetheless, a plausible balancing selective pressure under the system cannot be completely discarded to explain the diverse frequencies of the Rh-negative and Rh-positive (Flatz and Brinkmann 1977).

4.2. The genome-wide analysis of Basques

Basques have been scarcely studied from a genome-wide perspective, usually with a weak representation of the Basque area, considering the population as a whole, and without a proper ethnolinguistic background of the samples.

Taking into account the exhaustive sampling of the region presented in this thesis, a fine-scale result has been achieved showing a complete context of the historical Basque-related area. A clear genetic differentiation of all the Basque groups is confirmed, and the so-called Peri-Basques can be explained as peri-isolated intermediates, between Basques and the external Spanish and French samples. In addition, the controversy about internal heterogeneity has been solved, detecting for the first time a very clear pattern of differentiation among the groups within the Franco-Cantabrian region correlated with the geography (Flores-Bello et al., in preparation). This correlation is based on the higher gene flow among the closest groups, but it does not seem to be related to physical barriers, not even the Pyrenees range. This reflects a shared history of the whole region, and together with the historical distribution of the Euskara. In fact, the areas where the Euskara has been considerably reduced are the ones where some proportions of gene flow with external non-Franco-Cantabrian groups are detected. This result reflects the parallel evolution of languages and genes as first proposed by Cavalli-Sforza (Cavalli-Sforza 1997). In view of the results, it can be suggested the important role of the Euskara as a cultural barrier in the genetic differentiation of Basque people, in spite of the contact with different incomers such as the Romans. Nevertheless, it is difficult to suggest if the Euskara dialects are the cause or the consequence of the substructure related to the contact between the closest geographic groups. Although five dialects are usually differentiated, their classification and distribution are solid and still debated among the scholars (Abaitua 2018). However, they agree in a remarkable linguistic difference between the west and the east, but not along the central area. Nonetheless, genetic heterogeneity is also observed among Basque groups in this area. If dialects and genetics are correlated, the high internal genetic substructure found in the present results seems not to be compatible with a recent origin, i.e. during Middle Ages, of the dialects, since this would have implied an extreme and rapid differentiation of the

Basque groups in the last few centuries. In fact, a pre-Roman genetic structure have been already suggested based on uniparental markers (Martínez-Cruz et al. 2012). Therefore, a plausible scenario could be suggested, where the geographic distribution is the cause of the genetic and linguistic distances among groups along the history.

It is noteworthy to remark the strong specific genetic profile, associated with geography, observed in the Franco-Cantabrian region, and more specifically in Basques. Haplotype-based results are more accurate to detect fine internal patterns of structure within the populations than methods based on allele frequencies. However, the results have been very clear even detecting the internal substructure in the allele frequency analyses. Furthermore, haplotype-based methods have allowed to precisely reconstruct, without pre-established geographic labels, the same groups classified according to the geographic region they belong. Haplotype-based methods have also allowed to measure the proportions of shared ancestry among the reconstructed groups, to give a genetic support to the observed substructure in the region, beyond the limitations of the ADMIXTURE analyses that are more dependent on the number of SNPs and samples.

Including ancient samples in the analyses enables to describe modern populations in a historical context. Basques have been recently suggested to present a Bronze Age / Iron Age genetic profile within an Iberian ancient context (Olalde et al. 2019). In this thesis, similar results have been achieved, but framing the analysis in a wide Western European scenario. Basques have been placed among the general European proportions when formally modeling with the three ancient components par excellence: pre-Neolithic hunter-gatherer, Neolithic farmer, and Bronze Age steppe herder groups. Furthermore, shared ancestry has not been detected with modern proxies that could represent the post-Bronze Age admixture events that affected the external populations. These results suggest that Basques could be actually explained as the result of admixture between those three ancient components. This fact puts an end to the previous assumptions of the origin of Basques as a pre-Neolithic isolated group followed by low contact with the incoming migrants in the Iberian Peninsula in the following periods. Thus, it can be suggested that the process of genetic differentiation of Basques from the external populations may have taken place at least from the Bronze Age onwards (Flores-Bello et al., in preparation) .

4.3. Future studies

In spite of overcoming most of the limitations burdened in the previous works to finally elucidate the controversies of Basques, some caveats that have been faced during the development of the thesis should be taken into account.

Given the scant information about the function of the *RH* genes, specially the ones related to the Rhesus system, further genetic and functional studies are crucial to clarify their role in the organism.

Subsequently, this would enable to better understand their evolutive processes and the distribution of their variants in the populations.

Haplotypes combine sets of SNPs that are in linkage disequilibrium allowing to obtain a higher resolution than with independent SNPs. Moreover, they are less affected by the fixation of derived alleles through drift. Then, the haplotype-based methods used in this thesis (Hellenthal et al. 2014; Lawson et al. 2012) are more reliable to study populations that have gone through bottlenecks or founder effects. The characterization of the ancestry genetic profiles is often performed on admixed populations that can be properly explained from external sources. Nonetheless, analyzing isolated populations, as is the case of Basques, turns delicate and sensitive to artifacts if an appropriate representation of the population is not considered. This caveat is mainly due to the lack of external proxies that share haplotypes with them.

Isolates are often used as representatives of a region or a historical period in the analyses. In the same way that Sardinians have been used as a Neolithic proxy when studying modern populations, Basques have been usually considered as a pre-Neolithic group. However, this is not supported by the present results in this thesis, that connect Basques to the Bronze Age’s genetic profile in Southwestern Europe. Furthermore, they have been widely used without taking into consideration their internal heterogeneity. The observed substructure within Basques highlights the importance of considering the geographic area where the samples come from.

The public genetic databases are increasingly abundant, encompassing a large number of samples and geographic areas worldwide for different genetic data, especially genome-wide array and whole-genome data. However, most of them present the limitation of offering scant geolocation information, providing the coordinates of the samples as an average value for populations instead of individually. This is important for those methods whose resolution relies on a refined geographic distribution of the samples, such as isolation by distance analyses. Moreover, most databases do not include ethnolinguistic information of the populations analysed. There are many geographical regions where several different languages coexist. Given the relevance of this information to contextualize the genetic results, as in the case of the Basques, it would be highly informative to have this information available.

Population datasets usually define wide-scale groups to depict a general demographic context. In these conditions, the intrinsic genetic traits at population level get masked. The high resolution at a fine-scale scenario that it is achieved in my thesis bring to light the importance of considering a suitable representation of the population context studied. I encourage the increase of genetics studies considering a higer density of samples at micro-geographical level in order to enable the detection of new genetic populations outliers.

Ancient DNA is an essential resource to explore the demographic history of modern populations. Providing more ancient data from the Basque area would help to deeper elucidate their demographic history, that is characterized by their genetic differentiation and internal heterogeneity. However, it is fundamental to be aware about its limitations. Due to the fragile conservation of the ancient remains, it is more complicated to obtain good quality DNA material from this type of samples. Thus, the number of samples and the coverage become a limiting factor when using ancient DNA data. As mentioned above, haplotype-based methods allow to obtain finer results with high resolution. Therefore, it would be convenient to include ancient samples in these haplotype-based analyses. Nevertheless, a phasing process is mandatory to start the analyses. Many of the ancient samples do not present enough quality to achieve a good coverage to be reliably phased. Furthermore, when running some of these haplotype-based methods such as ChromoPainter, it is recommended to include a considerable number of samples in order to represent properly the group they belong. It is also crucial to have a balanced sample size of the groups to avoid over/underrepresentation of groups. Obtaining enough samples to represent a whole ancient group is usually complicated, making difficult to apply this kind of methods with this data. Moreover, human populations become more homogeneous as they get close to the present. The recurrent admixture events during history, and the increasing connections between them, make the specificity of the genetic components be diluted. Thus, modeling modern populations with more recent ancient samples than the Iron Age, is more complicated since they are more admixed and share more genetic components.

History, linguistics, archaeology and medicine have been essential to give support and contextualize the genetic results of my thesis. Thus, I would like conclude highlighting the importance of improving the multidisciplinary communication when studying the evolution and demographic history of populations.

5. REFERENCES

Abaitua, Joseba. 2018. “Patrones geolingüísticos, áreas dialectales y cronologías absolutas del EHHA.” *Fontes linguae vasconum: Studia et documenta* 50 (126): 283–322. https://dialnet.unirioja.es/servlet/articulo?codigo=6834722.

Abasolo, Carlos Cid. 2002. “Las Fronteras de La Lengua Vasca a Lo Largo de La Historia.” *Revista de Filología Románica* 19.

Achilli, Alessandro, Chiara Rengo, Vincenza Battaglia, Maria Pala, Anna Olivieri, Simona Fornarino, Chiara Magri, et al. 2005. “Saami and Berbers—An Unexpected Mitochondrial DNA Link.” *The American Journal of Human Genetics* 76 (5): 883–86. https://doi.org/10.1086/430073.

Achilli, Alessandro, Chiara Rengo, Chiara Magri, Vincenza Battaglia, Anna Olivieri, Rosaria Scozzari, Fulvio Cruciani, 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.” *The American Journal of Human Genetics* 75 (5): 910–18. https://doi.org/10.1086/425590.

Aguirre, A, A Vicario, L I Mazón, A Estomba, M Martínez de Pancorbo, V Arrieta Picó, F Perez Elortondo, and C M Lostao. 1991. “Are the Basques a Single and a Unique Population?” *American Journal of Human Genetics* 49 (2): 450–58. http://files/430/Aguirre et al. - 1991 - Are the Basques a single and a unique population.pdf.

Alday Ruiz, Alfonso. 2009. “Late Mesolithic and Early Neolithic in the Iberian Peninsula : Chronology and Phases” 60: 157–74.

Alexander, David H., John Novembre, and Kenneth Lange. 2009.

“Fast Model-Based Estimation of Ancestry in Unrelated Individuals.” *Genome Research* 19 (9): 1655–64. https://doi.org/10.1101/gr.094052.109.

Alfonso-Sánchez, M. A., S. Cardoso, C. Martínez-Bouzas, J. A. Peña, R. J. Herrera, A. Castro, I. Fernández-Fernández, and M. M. De Pancorbo. 2008. “Mitochondrial DNA Haplogroup Diversity in Basques: A Reassessment Based on HVI and HVII Polymorphisms.” *American Journal of Human Biology* 20 (2): 154–64. https://doi.org/10.1002/ajhb.20706.

Alonso, Santos, and J. A. Armour. 1998. “MS205 Minisatellite Diversity in Basques: Evidence for a Pre-Neolithic Component.” *Genome Research* 8 (12): 1289–98. https://doi.org/10.1101/gr.8.12.1289.

Alonso, Santos, Carlos Flores, Vicente Cabrera, Antonio Alonso, Pablo Martín, Cristina Albarrán, Neskuts Izagirre, Concepción de la Rúa, and Oscar García. 2005. “The Place of the Basques in the European Y-Chromosome Diversity Landscape.” *European Journal of Human Genetics* 13 (12): 1293–1302. https://doi.org/10.1038/sj.ejhg.5201482.

Alzualde, A., N. Izagirre, S. Alonso, A. Alonso, and C. de la Rúa. 2005. “Temporal Mitochondrial DNA Variation in the Basque Country: Influence of Post-Neolithic Events.” *Annals of Human Genetics* 69 (6): 665–79. https://doi.org/10.1046/j.1529-8817.2005.00170.x.

Amela Valverde, Luis. 2000. “Las Ciudades Fundadas Por Pompeyo Magno En Occidente ‘Pompaelo, Lugdunum Convenarum y Gerunda.’” *POLIS: Revista de Ideas y Formas*

Políticas de La Antigüedad Clásica 12.

Anstee, David J. 2010. “The Relationship between Blood Groups and Disease.” *Blood* 115 (23): 4635–43. https://doi.org/10.1182/blood-2010-01-261859.

Arnaiz-Villena, A., Jorge Martínez-Laso, Eduardo Gómez-Casado, Nieves Díaz-Campos, Paulo Santos, Antonio Martinho, and Henriqueta Breda-Coimbra. 1997. “Relatedness among Basques, Portuguese, Spaniards, and Algerians Studied by HLA Allelic Frequencies and Haplotypes.” *Immunogenetics* 47 (1): 37–43. https://doi.org/10.1007/s002510050324.

Arnaiz-Villena, A, D Benmamar, M Alvarez, N Diaz-Campos, P Varela, E Gomez-Casado, and J Martinez-Laso. 1995. “HLA Allele and Haplotype Frequencies in Algerians. Relatedness to Spaniards and Basques.” *Human Immunology* 43 (4): 259–68. https://doi.org/10.1016/0198-8859(95)00024-x.

Avent, Neil D, and Marion E Reid. 2000. “The Rh Blood Group System: A Review.” *Blood* 95 (2): 375–87. https://doi.org/10.1182/blood-2006-08-043232.

Bandelt, H J, P Forster, and A. Rohl. 1999. “Median-Joining Networks for Inferring Intraspecific Phylogenies.” *Molecular Biology and Evolution* 16 (1): 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036.

Barbujani, G, and G Bertorelle. 2001. “Genetics and the Population History of Europe.” *Proceedings of the National Academy of Sciences of the United States of America* 98 (1): 22–25. https://doi.org/10.1073/pnas.98.1.22.

Barral-Arca, Ruth, Sara Pischedda, Alberto Gómez-Carballa, Ana

Pastoriza, Ana Mosquera-Miguel, Manuel López-Soto, Federico Martinón-Torres, Vanesa Álvarez-Iglesias, and Antonio Salas. 2016. “Meta-Analysis of Mitochondrial DNA Variation in the Iberian Peninsula.” Edited by Alessandro Achilli. *PLOS ONE* 11 (7): e0159735. https://doi.org/10.1371/journal.pone.0159735.

Bauduer, Frédéric, Josué Feingold, and Didier Lacombe. 2005. “The Basques: Review of Population Genetics and Mendelian Disorders.” *Human Biology; an International Record of Research* 77 (5): 619–37. https://doi.org/10.1353/hub.2006.0001.

Bedialauneta Marcos, Estíbaliz. 1992. “Oihenart y Su Aportación a La Disuasión Sobre Los Límites Geográficos Del Vasco Antiguo En La ‘Notitia Utrisque Vasconiae.’” In *Oihenarten Laugarren Mendeurrena: Euskaltzaindiaren XII*, 189–201. Bilbao (Spain): Euskaltzaindia.

Behar, Doron M., Christine Harmant, Jeremy Manry, Mannis Van Oven, Wolfgang Haak, Begoña Martinez-Cruz, Jasone Salaberria, et al. 2012. “The Basque Paradigm: Genetic Evidence of a Maternal Continuity in the Franco-Cantabrian Region since Pre-Neolithic Times.” *American Journal of Human Genetics* 90 (3): 486–93. https://doi.org/10.1016/j.ajhg.2012.01.002.

Bertranpetit, J., J. Sala, F. Calafell, P. A. Underhill, P. Moral, and D. Comas. 1995. “Human Mitochondrial DNA Variation and the Origin of Basques.” *Annals of Human Genetics* 59 (Pt 1): 63–81. https://doi.org/10.1111/j.1469-1809.1995.tb01606.x.

Bertranpetit, J, and L L Cavalli-Sforza. 1991. “A Genetic Reconstruction of the History of the Population of the Iberian Peninsula.” *Annals of Human Genetics* 55 (1): 51–67. http://www.ncbi.nlm.nih.gov/pubmed/2042936.

Bodmer, Walter. 2015. “Genetic Characterization of Human Populations: From ABO to a Genetic Map of the British People.” *Genetics* 199 (2): 267–79. https://doi.org/10.1534/genetics.114.173062.

Bonaparte, Louis Lucien. 1869. “Le Verbe Basque En Tableaux, Accompagné de Notes Grammaticales, Selon Les Huit Dialectes de l’euskara : Le Guipuscoan, Le Biscaïen... / Par Le Prince Louis-Lucien Bonaparte.” http://www.liburuklik.euskadi.eus/jspui/handle/10771/25405.

Borrás, Henrike Knörr. 2008. “El Euskera En Tierras Del Romance: Rioja Alavesa, La Rioja, Burgos, Encartaciones.” *Huarte de San Juan . Filología y Didáctica de La Lengua* 10.

Bosch, E, F Calafell, D Comas, P J Oefner, P A Underhill, and J Bertranpetit. 2001. “High-Resolution Analysis of Human Y-Chromosome Variation Shows a Sharp Discontinuity and Limited Gene Flow between Northwestern Africa and the Iberian Peninsula.” *American Journal of Human Genetics* 68 (4): 1019–29. https://doi.org/10.1086/319521.

Bradburd, Gideon S., Peter L. Ralph, and Graham M. Coop. 2016. “A Spatial Framework for Understanding Population Structure and Admixture.” Edited by Montgomery Slatkin. *PLOS Genetics* 12 (1): e1005703. https://doi.org/10.1371/journal.pgen.1005703.

Bryc, K., A. Auton, M. R. Nelson, J. R. Oksenberg, S. L. Hauser, S. Williams, A. Froment, et al. 2010. “Genome-Wide Patterns of Population Structure and Admixture in West Africans and African Americans.” *Proceedings of the National Academy of Sciences* 107 (2): 786–91. https://doi.org/10.1073/pnas.0909559107.

Burke, Wylie, Steven S. Coughlin, Nancy C. Lee, Douglas L. Weed, and Muin J. Khoury. 2001. “Application of Population Screening Principles to Genetic Screening for Adult-Onset Conditions.” *Genetic Testing* 5 (3): 201–11. https://doi.org/10.1089/10906570152742245.

Bycroft, Clare, Ceres Fernandez-Rozadilla, Clara Ruiz-Ponte, Inés Quintela, Ángel Carracedo, Peter Donnelly, and Simon Myers. 2019. “Patterns of Genetic Differentiation and the Footprints of Historical Migrations in the Iberian Peninsula.” *Nature Communications* 10 (1): 551. https://doi.org/10.1038/s41467-018-08272-w.

Calafell, F., and J. Bertranpetit. 1994a. “Principal Component Analysis of Gene Frequencies and the Origin of Basques.” *American Journal of Physical Anthropology* 93 (2): 201–15. https://doi.org/10.1002/ajpa.1330930205.

———. 1994b. “Mountains and Genes: Population History of the Pyrenees.” *Human Biology* 66 (5): 823–42. http://www.ncbi.nlm.nih.gov/pubmed/8001912.

Calderón, R, C Vidales, J A Peña, A Perez-Miranda, and J M Dugoujon. 1998a. “Immunoglobulin Allotypes (GM and KM) in Basques from Spain: Approach to the Origin of the Basque

Population.” *Human Biology* 70 (4): 667–98. http://www.ncbi.nlm.nih.gov/pubmed/9686480.

———. 1998b. “Immunoglobulin Allotypes (GM and KM) in Basques from Spain: Approach to the Origin of the Basque Population.” *Human Biology* 70 (4): 667–98.

Cann, H. M., Claudia de Toma, Lucien Cazes, Marie-Fernande Legrand, Valerie Morel, Laurence Piouffre, Julia Bodmer, et al. 2002. “A Human Genome Diversity Cell Line Panel.” *Science* 296 (5566): 261b – 262. https://doi.org/10.1126/science.296.5566.261b.

Cann, Rebecca L. 2006. “DNA Markers and Human Evolution.” In *Encyclopedia of Life Sciences*. Chichester: John Wiley & Sons, Ltd. https://doi.org/10.1038/npg.els.0005078.

Cardoso, Sergio, Laura Valverde, Miguel A. Alfonso-Sánchez, Leire Palencia-Madrid, Xabier Elcoroaristizabal, Jaime Algorta, Susana Catarino, et al. 2013. “The Expanded MtDNA Phylogeny of the Franco-Cantabrian Region Upholds the Pre-Neolithic Genetic Substrate of Basques.” Edited by Alessandro Achilli. *PLoS ONE* 8 (7): e67835. https://doi.org/10.1371/journal.pone.0067835.

Caro Baroja, Julio. 1973. *Los Pueblos Del Norte de La Península Ibérica*. 2nd ed. San Sebastián (Spain): Txertoa.

Carretero Díaz, José Miguel, Ana Isabel Ortega, Laura Juez Aparicio, Alfredo Pérez González, Juan Luis Arsuaga, Raquel Pérez Martínez, and Maria Cruz Ortega. 2008. “A Late Pleistocene-Early Holocene Archaeological Sequence of Portalón de Cueva Mayor.” *Munibe Antropologia* -

Arkeologia, ISSN 1132-2217, Nº 59, 2008, Págs. 67-80, no. 59: 67–80.

Cava, Ana. 1990. “The Neolithic in the Basque Country.” *Munibe* 42: 97–106. http://www.aranzadi.eus/fileadmin/docs/Munibe/1990081089AA.pdf.

Cavalli-Sforza, L. L. 1988. “The Basque Population and Ancient Migrations in Europe.” *MUNIBE (Antropología y Arqueología)* 6 (Suppl): 129–37. http://www.aranzadi-zientziak.org/fileadmin/docs/Munibe/1988129137.pdf.

———. 1997. “Genes, Peoples, and Languages.” *Proceedings of the National Academy of Sciences of the United States of America* 94 (15): 7719–24. https://doi.org/10.1073/pnas.94.15.7719.

Cavalli-Sforza, L. L., and W. F. Bodmer. 1971. *The Genetics of Human Populations*. Dover Publications.

Cavalli-Sforza, L. L., and W. F. Edwards. 1963. “Analysis of Human Evolution.” In *Genetics Today: Proceedings of the 11th International Congress of Genetics*, 923–33. The Hague: Oxford: Pergamon Press.

Cavalli-Sforza, L. L., Paolo Menozzi, and Alberto Piazza. 1994. *The History and Geography of Human Genes*. Princeton (New Jersey): Princeton University Press.

Ceberio Rodríguez, Manu. 2009. “Nuevas Aportaciones Al Estudio de La Transición de La Edad Del Hierro a Época Romana En Gipuzkoa : El Caso de Santiagomendi (Astigarraga).” *Munibe Antropologia-Arkeologia* 60 (1): 219–41.

Chadarevian, Soraya de. 2010. “Genetic Evidence and Interpretation in History.” *BioSocieties* 5 (3): 301–5. https://doi.org/10.1057/biosoc.2010.15.

Chalmers, J. N. M., E. W. Ikin, and A. E. Mourant. 1948. “Basque Blood Groups.” *Nature* 162 (4105): 27. http://www.ncbi.nlm.nih.gov/pubmed/18869685.

Chalmers, J. N. M., Elizabeth W. Ikin, and A. E. Mourant. 1949. “The ABO, MN and Rh Blood Groups of the Basque People.” *American Journal of Physical Anthropology* 7 (4): 529–44. https://doi.org/10.1002/ajpa.1330070405.

Charlesworth, B, and D Charlesworth. 2017. “Population Genetics from 1966 to 2016.” *Heredity* 118 (1): 2–9. https://doi.org/10.1038/hdy.2016.55.

Chikhi, L, G Destro-Bisol, G Bertorelle, V Pascali, and G Barbujani. 1998. “Clines of Nuclear DNA Markers Suggest a Largely Neolithic Ancestry of the European Gene Pool.” *Proceedings of the National Academy of Sciences of the United States of America* 95 (15): 9053–58. https://doi.org/10.1073/pnas.95.15.9053.

Collins, Roger. 1990. *The Basques*. Basil Blackwell. https://books.google.es/books/about/The_Basques.html?id=ageHQgAACAAJ&redir_esc=y.

Comas, D., F. Calafell, E. Mateu, A. Pérez-Lezaun, and J. Bertranpetit. 1998. “HLA Evidence for the Lack of Genetic Heterogeneity in Basques.” *Annals of Human Genetics* 62 (Pt 2): 123–32. https://doi.org/10.1046/j.1469-1809.1998.6220123.x.

Comas, D., F. Calafell, E. Mateu, A. Pérez-Lezaun, E. Bosch, and J. Bertranpetit. 1997. “Mitochondrial DNA Variation and the Origin of the Europeans.” *Human Genetics* 99 (4): 443–49. https://doi.org/10.1007/s004390050386.

Comrie, Bernard, Maria Polinsky, and Stephen Matthews. 2003. *The Atlas of Languages. The Origin and Development of Languages throughout the World*. New York: Facts on File.

Côrte-Real, H B, V A Macaulay, M B Richards, G Hariti, M S Issad, A Cambon-Thomsen, S Papiha, J Bertranpetit, and B C Sykes. 1996. “Genetic Diversity in the Iberian Peninsula Determined from Mitochondrial Sequence Analysis.” *Annals of Human Genetics* 60 (4): 331–50. http://www.ncbi.nlm.nih.gov/pubmed/8865993.

Daniels, G. 2013. *Human Blood Groups: 3rd Edition*. *Human Blood Groups: 3rd Edition*. https://doi.org/10.1002/9781118493595.

Domene Verdú, José Fernando. 2011. *Lengua Vasca: Originalidad y Riqueza de Una Lengua Diferente*. Edited by Club Universitario. Alicante.

Edwards, A. W., and L. L. Cavalli-Sforza. 1965. “A Method for Cluster Analysis.” *Biometrics* 21 (June): 362–75. http://www.ncbi.nlm.nih.gov/pubmed/14338671.

Estornés Lasa, Bernardo. 1983. “El Dialecto Roncalés En La Muga de Aragon y Bearne.” In *Piarres Laffiteri Omenaldia, Homenaje a Pierre Lafitte*, 215–29. Euskaltzaindia. https://dialnet.unirioja.es/servlet/articulo?codigo=3865065.

Etcheverry, M. A. 1945. “El Factor Rhesus: Su Genética y Su Importancia Clínica.” *El Día Médico* 17: 1237–59.

———. 1947. “El Factor Rh En Personas de Ascendencia Iberica e Italica Residentes En La Argentina.” *La Semana Medica* 2: 1–8.

Eusko Jaurlaritza. 2016. “Sixth Sociolinguistic Survey 2006.” Donostia. http://www.euskadi.eus/contenidos/informacion/ikerketa_soziolinguistikoak/es_def/adjuntos/2016 VI INK SOZLG - Euskal Herria gaz.pdf.

F. Ruiz, Teofilo. 2018. *The Western Mediterranean and the World: 400 CE to the Present*. 2nd ed. Malden: Wiley-Blackwell.

Fisher, R. A. 1947. “The Rhesus Factor: A Study in Scientific Method.” *American Scientist* 35 (1): 95–102. http://www.ncbi.nlm.nih.gov/pubmed/20282984.

Fisher, R. A., and G. L. Taylor. 1940. “Scandinavian Influence in Scottish Ethnology.” *Nature* 145 (3676): 590–590. https://doi.org/10.1038/145590a0.

Flatz, G., and H.-B. Brinkmann. 1977. “A Selection Model for the Rhesus Blood-Group System.” *Journal of Human Evolution* 6 (2): 135–40. https://doi.org/10.1016/S0047-2484(77)80118-8.

Flegel, Willy A. 2011. “Molecular Genetics and Clinical Applications for RH.” *Transfusion and Apheresis Science : Official Journal of the World Apheresis Association : Official Journal of the European Society for Haemapheresis* 44 (1): 81–91. https://doi.org/10.1016/j.transci.2010.12.013.

Flegr, Jaroslav. 2016. “Heterozygote Advantage Probably Maintains Rhesus Factor Blood Group Polymorphism: Ecological Regression Study.” Edited by Calogero Caruso.

PLOS ONE 11 (1): e0147955. https://doi.org/10.1371/journal.pone.0147955.

Flores-Bello, André, David Mas-Ponte, Miruna E. M.E. Rosu, Elena Bosch, Francesc Calafell, and David Comas. 2018. “Sequence Diversity of the Rh Blood Group System in Basques.” *European Journal of Human Genetics* 26 (12): 1859–66. https://doi.org/10.1038/s41431-018-0232-1.

Flores, Carlos, Nicole Maca-Meyer, Ana M González, Peter J Oefner, Peidong Shen, Jose A Pérez, Antonio Rojas, Jose M Larruga, and Peter A Underhill. 2004. “Reduced Genetic Structure of the Iberian Peninsula Revealed by Y-Chromosome Analysis: Implications for Population Demography.” *European Journal of Human Genetics* 12 (10): 855–63. https://doi.org/10.1038/sj.ejhg.5201225.

Gamble, Clive, William Davies, Paul Pettitt, Lee Hazelwood, and Martin Richards. 2006. “The Late Glacial Ancestry of Europeans: Combining Genetic and Archaeological Evidence.” *Documenta Praehistorica* 33 (December): 1–10. https://doi.org/10.4312/dp.33.1.

Gangal, Kavita, Graeme R. Sarson, and Anvar Shukurov. 2014. “The Near-Eastern Roots of the Neolithic in South Asia.” Edited by David Caramelli. *PLoS ONE* 9 (5): e95714. https://doi.org/10.1371/journal.pone.0095714.

Garagnani, Paolo, Hafid Laayouni, Anna González-Neira, Martin Sikora, Donata Luiselli, Jaume Bertranpetit, and Francesc Calafell. 2009. “Isolated Populations as Treasure Troves in Genetic Epidemiology: The Case of the Basques.” *European*

Journal of Human Genetics 17 (11): 1490–94. https://doi.org/10.1038/ejhg.2009.69.

García, O, R Fregel, J M Larruga, V Álvarez, I Yurrebaso, V M Cabrera, and A M González. 2011. “Using Mitochondrial DNA to Test the Hypothesis of a European Post-Glacial Human Recolonization from the Franco-Cantabrian Refuge.” *Heredity* 106 (1): 37–45. https://doi.org/10.1038/hdy.2010.47.

Gardiner, R M. 2002. “The Human Genome Project: The next Decade.” *Archives of Disease in Childhood* 86 (6): 389–91. https://doi.org/10.1136/adc.86.6.389.

Gattepaille, Lucie M, and Mattias Jakobsson. 2012. “Combining Markers into Haplotypes Can Improve Population Structure Inference.” *Genetics* 190 (1): 159–74. https://doi.org/10.1534/genetics.111.131136.

Goldsworthy, Adrian. 2009. *How Roma Fell: Death of a Superpower*. Yale University Press.

González, Ana M, Oscar García, José M Larruga, and Vicente M Cabrera. 2006. “The Mitochondrial Lineage U8a Reveals a Paleolithic Settlement in the Basque Country.” *BMC Genomics* 7 (1): 124. https://doi.org/10.1186/1471-2164-7-124.

Gorrochategui Churruca, Joaquín. 1984. *Onomástica Indígena de Aquitania*. Bilbao (Spain): Euskal Herriko Unibertsitatea - Universidad del País Vasco.

Gorrochategui, Javier, and María Josè Yarritu. 1990. “The Upper Neolithic - Bronze Age in the Northern Basque Country.” *Munibe* 42: 107–23.

Green, Richard E, Johannes Krause, Adrian W Briggs, Tomislav

Maricic, Udo Stenzel, Martin Kircher, Nick Patterson, et al. 2010. “A Draft Sequence of the Neandertal Genome.” *Science (New York, N.Y.)* 328 (5979): 710–22. https://doi.org/10.1126/science.1188021.

Günther, Torsten, Cristina Valdiosera, Helena Malmström, Irene Ureña, Ricardo Rodriguez-Varela, Óddny Osk Sverrisdóttir, Evangelia a. Daskalaki, et al. 2015. “Ancient Genomes Link Early Farmers from Atapuerca in Spain to Modern-Day Basques.” *Proceedings of the National Academy of Sciences* 112 (38): 11917–22. https://doi.org/10.1073/pnas.1509851112.

Haak, Wolfgang, Iosif Lazaridis, Nick Patterson, Nadin Rohland, Swapan Mallick, Bastien Llamas, Guido Brandt, et al. 2015. “Massive Migration from the Steppe Was a Source for Indo-European Languages in Europe.” *Nature* 522 (7555): 207–11. https://doi.org/10.1038/nature14317.

Heide, H. M. Van Der, Willy Magnée, and J.J. J Van Loghem. 1951. “Blood Group Distribution in Basques.” *American Journal of Human Genetics* 3 (4): 356–61. http://files/432/Van Der Heide, Magnée, Van Loghem - 1951 - Blood group distribution in Basques.pdf.

Helgason, Agnar, Sigrún Sigurðardóttir, Jeffrey R. Gulcher, Ryk Ward, and Kári Stefánsson. 2000. “MtDNA and the Origin of the Icelanders: Deciphering Signals of Recent Population History.” *The American Journal of Human Genetics* 66 (3): 999–1016. https://doi.org/10.1086/302816.

Hellenthal, Garrett, George B J Busby, Gavin Band, James F Wilson, Cristian Capelli, Daniel Falush, and Simon Myers.

2014. “A Genetic Atlas of Human Admixture History.” *Science (New York, N.Y.)* 343 (6172): 747–51. https://doi.org/10.1126/science.1243518.

Hernández, Candela L., Jean M. Dugoujon, Andrea Novelletto, Juan N. Rodríguez, Pedro Cuesta, and Rosario Calderón. 2017. “The Distribution of Mitochondrial DNA Haplogroup H in Southern Iberia Indicates Ancient Human Genetic Exchanges along the Western Edge of the Mediterranean.” *BMC Genetics* 18 (1): 46. https://doi.org/10.1186/s12863-017-0514-6.

Hervella, Montserrat, Neskuts Izagirre, Santos Alonso, Rosa Fregel, Antonio Alonso, Vicente M. Cabrera, and Concepción de la Rúa. 2012. “Ancient DNA from Hunter-Gatherer and Farmer Groups from Northern Spain Supports a Random Dispersion Model for the Neolithic Expansion into Europe.” Edited by Michael Hofreiter. *PLoS ONE* 7 (4): e34417. https://doi.org/10.1371/journal.pone.0034417.

Hirschfeld, Ludwik, and Hanka Hirschfeld. 1919. “Serological Differences between the Blood of Different Races. The Result of Researches on the Macedonian Front.” *The Lancet* 194 (5016): 675–79. https://doi.org/10.1016/S0140-6736(01)48686-7.

Holsinger, Kent E., and Bruce S. Weir. 2009. “Genetics in Geographically Structured Populations: Defining, Estimating and Interpreting FST.” *Nature Reviews Genetics* 10 (9): 639–50. https://doi.org/10.1038/nrg2611.

Hualde, José Ignacio, Joseba A. Lakarra, and R.L. Trask, eds. 1996. *Towards a History of the Basque Language*. Vol. 131. Current

Issues in Linguistic Theory. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/cilt.131.

Hudson, R. R., M. Slatkin, and W. P. Maddison. 1992. “Estimation of Levels of Gene Flow from DNA Sequence Data.” *Genetics* 132 (2): 583–89. http://www.ncbi.nlm.nih.gov/pubmed/1427045.

Igartua, Iván, and Zabier Zabaltza. 2016. *A Brief History of the Basque Language*. Donostia: Etxepare Euskal Institutua.

Iriondo, Mikel, M. C. Barbero, and C. Manzano. 2003. “DNA Polymorphisms Detect Ancient Barriers to Gene Flow in Basques.” *American Journal of Physical Anthropology* 122 (1): 73–84. https://doi.org/10.1002/ajpa.10212.

Izagirre, N., and C. de la Rúa. 1999. “An MtDNA Analysis in Ancient Basque Populations: Implications for Haplogroup V as a Marker for a Major Paleolithic Expansion from Southwestern Europe.” *American Journal of Human Genetics* 65 (1): 199–207. https://doi.org/10.1086/302442.

Jobling, Mark, Edward Hollox, Matthew Hurles, Toomas Kivisild, and Chris Tyler-smith. 2013. *Human Evolutionary Genetics*. New York (USA): Garland Science.

Jones, Horace Leonard. 1917. *The Geography of Strabo*. London: Heinemann.

Jorde, L.B., W.S. Watkins, M.J. Bamshad, M.E. Dixon, C.E. Ricker, M.T. Seielstad, and M.A. Batzer. 2000. “The Distribution of Human Genetic Diversity: A Comparison of Mitochondrial, Autosomal, and Y-Chromosome Data.” *The American Journal of Human Genetics* 66 (3): 979–88.

https://doi.org/10.1086/302825.

Kreager, Philip, Bruce Winney, Stanley Ulijaszek, and Cristian Capelli, eds. 2015. *Population in the Human Sciences: Concepts, Models, Evidence*. New York (USA): Oxford University Press.

Kurlansky, Mark. 1999. *The Basque History of the World*. Penguin Books.

la Rúa, C. de. 1992. “Craniofacial Factors in the Basque Skull. A Comparative Study.” *Homo* 43: 135–61.

Laayouni, H., F. Calafell, and J. Bertranpetit. 2010. “A Genome-Wide Survey Does Not Show the Genetic Distinctiveness of Basques.” *Human Genetics* 127 (4): 455–58. https://doi.org/10.1007/s00439-010-0798-3.

LaFramboise, Thomas. 2009. “Single Nucleotide Polymorphism Arrays: A Decade of Biological, Computational and Technological Advances.” *Nucleic Acids Research* 37 (13): 4181–93. https://doi.org/10.1093/nar/gkp552.

Lander, E S, L M Linton, B Birren, C Nusbaum, M C Zody, J Baldwin, K Devon, et al. 2001. “Initial Sequencing and Analysis of the Human Genome.” *Nature* 409 (6822): 860–921. https://doi.org/10.1038/35057062.

Lao, Oscar, Timothy T. Lu, Michael Nothnagel, Olaf Junge, Sandra Freitag-Wolf, Amke Caliebe, Miroslava Balascakova, et al. 2008. “Correlation between Genetic and Geographic Structure in Europe.” *Current Biology* 18 (16): 1241–48. https://doi.org/10.1016/j.cub.2008.07.049.

las Heras, C. de, J. A. Lasheras, Álvaro Arrizabalaga, and Marco de la Rosilla, eds. 2012. *Rethinking the Gravettian: New Approaches for the Cantabrian Region in Its Peninsular and Pyrenean Contexts*. Secretaría General Técnica. Centro de Publicaciones. Ministerio de Educación, Cultura y Deporte.

Laska, Magdalena Janina, Kari Konstantin Karl Kari K Nissen, Bjørn Andersen Nexø, Andy Purvis, Thomas Granier, Sophie Beley, Jacques Chiaroni, et al. 2009. "A composite estimate of primate phylogeny." Edited by Gregory P. Copenhaver. *Bioinformatics* 8 (1): 11. https://doi.org/10.1098/rstb.1995.0078.

Lawson, Daniel John, Garrett Hellenthal, Simon Myers, and Daniel Falush. 2012. "Inference of Population Structure Using Dense Haplotype Data." Edited by Gregory P. Copenhaver. *PLoS Genetics* 8 (1): e1002453. https://doi.org/10.1371/journal.pgen.1002453.

Lazaridis, Iosif, Dani Nadel, Gary Rollefson, Deborah C. Merrett, Nadin Rohland, Swapan Mallick, Daniel Fernandes, et al. 2016. "Genomic Insights into the Origin of Farming in the Ancient Near East." *Nature* 536 (7617): 419–24. https://doi.org/10.1038/nature19310.

Lazaridis, Iosif, Nick Patterson, Alissa Mittnik, Gabriel Renaud, Swapan Mallick, Karola Kirsanow, Peter H. Sudmant, et al. 2014. "Ancient Human Genomes Suggest Three Ancestral Populations for Present-Day Europeans." *Nature* 513 (7518): 409–13. https://doi.org/10.1038/nature13673.

Levine, P., N. J. Newark, and R .E. Stetson. 1939. "An Unusual Case of Intragroup Agglutination." *JAMA* 113: 126–27.

Li, Jun Z., Devin M Absher, Hua Tang, Audrey M Southwick, Amanda M Casto, Sohini Ramachandran, Howard M Cann, et al. 2008. “Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation.” *Science (New York, N.Y.)* 319 (5866): 1100–1104. https://doi.org/10.1126/science.1153717.

López-Mugartza Iriarte, Juan Karlos. 2008. “Euskera, Language of the Pyrenees: The Case of Ansó (Aragon) and Roncal (Navarre).” *Anuario Del Seminario de Filología Vasca Julio de Urquijo: International Journal of Basque Linguistics and Philology* 42 (2): 303–36. https://dialnet.unirioja.es/servlet/articulo?codigo=4018182.

Lopez-Parra, A M, L Gusmão, L Tavares, C Baeza, A Amorim, M S Mesa, M J Prata, and E Arroyo-Pardo. 2009. “In Search of the Pre- and Post-Neolithic Genetic Substrates in Iberia: Evidence from Y-Chromosome in Pyrenean Populations.” *Annals of Human Genetics* 73 (1): 42–53. https://doi.org/10.1111/j.1469-1809.2008.00478.x.

López, Saioa, and Santos Alonso. 2013. “Genetics and the History of the Basque People.” https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470015902.a0020803.pub2.

Luchaire, Achille. 1877. *Les Origines Linguistiques de l’Aquitaine*. Pau (France): Bulletin de la Société des Sciences, Lettres et Arts.

Manzano, C, A I Aguirre, M Iriondo, M Martín, L Osaba, and C de la Rúa. 1996. “Genetic Polymorphisms of the Basques from

Gipuzkoa: Genetic Heterogeneity of the Basque Population." *Annals of Human Biology* 23 (4): 285–96. http://www.ncbi.nlm.nih.gov/pubmed/8830916.

Manzano, C, la Rúa C De, M Iriondo, L I Mazón, A Vicario, and A Aguirre. 2002. "Structuring the Genetic Heterogeneity of the Basque Population: A View from Classical Polymorphisms." *Human Biology; an International Record of Research* 74 (1): 51–74. https://doi.org/10.1353/hub.2002.0008.

Manzano, C, J M Orue, and C de la Rúa. 1996. "The 'Basqueness' of the Basques of Alava: A Reappraisal from a Multidisciplinary Perspective." *American Journal of Physical Anthropology* 99 (2): 249–58. http://www.ncbi.nlm.nih.gov/pubmed/8967326.

Mardulyn, Patrick. 2012. "Trees and/or Networks to Display Intraspecific DNA Sequence Variation?" *Molecular Ecology* 21 (14): 3385–90. https://doi.org/10.1111/j.1365-294X.2012.05622.x.

Martínez-Cruz, Begoña, Christine Harmant, Daniel E. Platt, Wolfgang Haak, Jeremy Manry, Eva Ramos-Luis, David F. Soria-Hernanz, et al. 2012. "Evidence of Pre-Roman Tribal Genetic Structure in Basques from Uniparentally Inherited Markers." *Molecular Biology and Evolution* 29 (9): 2211–22. https://doi.org/10.1093/molbev/mss091.

McVean, Gil. 2009. "A Genealogical Interpretation of Principal Components Analysis." Edited by Molly Przeworski. *PLoS Genetics* 5 (10): e1000686. https://doi.org/10.1371/journal.pgen.1000686.

Mithen, Steven J. 2004. *After the Ice : A Global Human History, 20,000-5000 BC*. Harvard University Press. https://www.hup.harvard.edu/catalog.php?isbn=9780674019997.

Mittnik, Alissa, Chuan-Chao Wang, Saskia Pfrengle, Mantas Daubaras, Gunita Zariņa, Fredrik Hallgren, Raili Allmäe, et al. 2018. “The Genetic Prehistory of the Baltic Sea Region.” *Nature Communications* 9 (1): 442. https://doi.org/10.1038/s41467-018-02825-9.

Mitxelena Elissalt, Luis. 1976. *Lenguas Indígenas y Lengua Clásica En Hispania*. Paris (France): Les Belles Lettres.

———. 1981. “Lengua Común y Dialectos Vascos.” *International Journal of Basque Linguistics and Philology* 15 (1): 289–313.

Montero, Manuel. 2008. *Historia General Del País Vasco*. Donostia: Txertoa.

Mourant, A. E. 1947. “The Blood Groups of the Basques.” *Nature* 160 (4067): 505–6. https://doi.org/10.1038/160505b0.

———. 1954. *The Distribution of the Human Blood Groups*. Oxford, UK: Blackwell Scientific Publications.

Mourant, A. E., A. C. Kopeć, and K. Domaniewska-Sobczak. 1976. *The Distribution of the Human Blood Groups and Other Polymorphisms*. 2nd ed. Oxford University Press.

Mujika Zufiría, Serapio. 1914. “El Obispado de Bayona Con Relación a Los Pueblos de Guipúzcoa Adscritos a Dicha Diócesis.” *International Journal on Basque Dialects* 8 (2): 185–229.

Nei, M. 1973. “Analysis of Gene Diversity in Subdivided

Populations.” *Proceedings of the National Academy of Sciences of the United States of America* 70 (12): 3321–23. https://doi.org/10.1073/pnas.70.12.3321.

Nelis, Mari, Tõnu Esko, Reedik Mägi, Fritz Zimprich, Alexander Zimprich, Draga Toncheva, Sena Karachanak, et al. 2009. “Genetic Structure of Europeans: A View from the North-East.” *PloS One* 4 (5): e5472. https://doi.org/10.1371/journal.pone.0005472.

Novembre, John, Toby Johnson, Katarzyna Bryc, Zoltán Kutalik, Adam R Boyko, Adam Auton, Amit Indap, et al. 2008. “Genes Mirror Geography within Europe.” *Nature* 456 (7218): 98–101. https://doi.org/10.1038/nature07331.

Novembre, John, and Matthew Stephens. 2008. “Interpreting Principal Component Analyses of Spatial Population Genetic Variation.” *Nature Genetics* 40 (5): 646–49. https://doi.org/10.1038/ng.139.

Nuñez Astrain, Luis. 2004. *El Euskera Arcaico : Extensión y Parentescos*. Txalaparta. https://dialnet.unirioja.es/servlet/libro?codigo=624365.

Olalde, Iñigo, Swapan Mallick, Nick Patterson, Nadin Rohland, Vanessa Villalba-Mouco, Marina Silva, Katharina Dulias, et al. 2019. “The Genomic History of the Iberian Peninsula over the Past 8000 Years.” *Science* 363 (6432): 1230–34. https://doi.org/10.1126/science.aav4040.

Owen, R. 2000. “Karl Landsteiner and the First Human Marker Locus.” *Genetics* 155 (3): 995–98. http://www.ncbi.nlm.nih.gov/pubmed/10880463.

Pala, Maria, Anna Olivieri, Alessandro Achilli, Matteo Accetturo, Ene Metspalu, Maere Reidla, Erika Tamm, et al. 2012. “Mitochondrial DNA Signals of Late Glacial Recolonization of Europe from near Eastern Refugia.” *American Journal of Human Genetics* 90 (5): 915–24. https://doi.org/10.1016/j.ajhg.2012.04.003.

Palencia-Madrid, Leire, Sergio Cardoso, Christine Keyser, Juan Carlos López-Quintana, Amagoia Guenaga-Lizasu, and Marian M de Pancorbo. 2017. “Ancient Mitochondrial Lineages Support the Prehistoric Maternal Root of Basques in Northern Iberian Peninsula.” *European Journal of Human Genetics* 25 (5): 631–36. https://doi.org/10.1038/ejhg.2017.24.

Patterson, Nick, Priya Moorjani, Yontao Luo, Swapan Mallick, Nadin Rohland, Yiping Zhan, Teri Genschoreck, Teresa Webster, and David Reich. 2012. “Ancient Admixture in Human History.” *Genetics* 192 (3): 1065–93. https://doi.org/10.1534/genetics.112.145037.

Pembleton, Luke W, Noel O I Cogan, and John W Forster. 2013. “StAMPP: An R Package for Calculation of Genetic Differentiation and Structure of Mixed-Ploidy Level Populations.” *Molecular Ecology Resources* 13 (5): 946–52. https://doi.org/10.1111/1755-0998.12129.

Pereira, Luísa, Martin Richards, Ana Goios, Antonio Alonso, Cristina Albarrán, Oscar Garcia, Doron M Behar, et al. 2005. “High-Resolution MtDNA Evidence for the Late-Glacial Resettlement of Europe from an Iberian Refugium.” *Genome Research* 15 (1): 19–24. https://doi.org/10.1101/gr.3182305.

Pérez-Lezaun, Anna, Francesc Calafell, E. Mateu, David Comas, Rosalía Ruiz-Pacheco, and J. Bertranpetit. 1996. “Microsatellite Variation and the Differentiation of Modern Humans.” *Human Genetics* 99 (1): 1–7. https://doi.org/10.1007/s004390050299.

Perez-Miranda, A.M., M.A. Alfonso-Sanchez, M.C. Vidales, R. Calderon, and J.A. Pena. 2004. “Genetic Polymorphism and Linkage Disequilibrium of the HLA-DP Region in Basques from Navarre (Spain).” *Tissue Antigens* 64 (3): 264–75. https://doi.org/10.1111/j.1399-0039.2004.00284.x.

Pérez-Miranda, Ana M., Miguel A. Alfonso-Sánchez, Arif Kalantar, Susana García-Obregón, Marian M. Pancorbo, José A. Peña, and Rene J. Herrera. 2005. “Microsatellite Data Support Subpopulation Structuring among Basques.” *Journal of Human Genetics* 50 (8): 403–14. https://doi.org/10.1007/s10038-005-0268-2.

Perry, George H., Yali Xue, Richard S. Smith, Wynn K. Meyer, Minal ÇalIşkan, Omar Yanez-Cuna, Arthur S. Lee, et al. 2012. “Evolutionary Genetics of the Human Rh Blood Group System.” *Human Genetics* 131 (7): 1205–16. https://doi.org/10.1007/s00439-012-1147-5.

Peter, Benjamin M. 2016. “Admixture, Population Structure, and F-Statistics.” *Genetics* 202 (4): 1485–1501. https://doi.org/10.1534/genetics.115.183913.

Petkova, Desislava, John Novembre, and Matthew Stephens. 2016. “Visualizing Spatial Population Structure with Estimated Effective Migration Surfaces.” *Nature Genetics* 48 (1): 94–

100. https://doi.org/10.1038/ng.3464.

Piazza, A., N. Cappello, E. Olivetti, and S. Rendine. 1988. “The Basques in Europe: A Genetic Analysis.” *MUNIBE (Antropología y Arqueología)1* 6: 167–77.

Pickrell, Joseph K., and David Reich. 2014. “Toward a New History and Geography of Human Genes Informed by Ancient DNA.” *Trends in Genetics* 30 (9): 377–89. https://doi.org/10.1016/J.TIG.2014.07.007.

Posth, Cosimo, Nathan Nakatsuka, Iosif Lazaridis, Pontus Skoglund, Swapan Mallick, Thiseas C Lamnidis, Nadin Rohland, et al. 2018. “Reconstructing the Deep Population History of Central and South America.” *Cell* 175 (5): 1185-1197.e22. https://doi.org/10.1016/j.cell.2018.10.027.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. “Inference of Population Structure Using Multilocus Genotype Data.” *Genetics* 155 (2): 945–59. http://www.ncbi.nlm.nih.gov/pubmed/10835412.

Rasmussen, Morten, Yingrui Li, Stinus Lindgreen, Jakob Skou Pedersen, Anders Albrechtsen, Ida Moltke, Mait Metspalu, et al. 2010. “Ancient Human Genome Sequence of an Extinct Palaeo-Eskimo.” *Nature* 463 (7282): 757–62. https://doi.org/10.1038/nature08835.

Reich, David, Richard E. Green, Martin Kircher, Johannes Krause, Nick Patterson, Eric Y. Durand, Bence Viola, et al. 2010. “Genetic History of an Archaic Hominin Group from Denisova Cave in Siberia.” *Nature* 468 (7327): 1053–60. https://doi.org/10.1038/nature09710.

Renfrew, Colin. 2010. “Archaeogenetics - Towards a ‘New Synthesis’?” *Current Biology* 20 (4): R162–65. https://doi.org/10.1016/j.cub.2009.11.056.

Richards, M, H Côrte-Real, P Forster, V Macaulay, H Wilkinson-Herbots, A Demaine, S Papiha, R Hedges, H J Bandelt, and B Sykes. 1996. “Paleolithic and Neolithic Lineages in the European Mitochondrial Gene Pool.” *American Journal of Human Genetics* 59 (1): 185–203. http://www.ncbi.nlm.nih.gov/pubmed/8659525.

Rodríguez-Ezpeleta, Naiara, Jon Álvarez-Busto, Liher Imaz, María Regueiro, María Nerea Azcárate, Roberto Bilbao, Mikel Iriondo, Ana Gil, Andone Estonba, and Ana María Aransay. 2010. “High-Density SNP Genotyping Detects Homogeneity of Spanish and French Basques, and Confirms Their Genomic Distinctiveness from Other European Populations.” *Human Genetics* 128 (1): 113–17. https://doi.org/10.1007/s00439-010-0833-4.

Rohlfs, Gerhard. 1935. *Le Gascon: Études de Philologie Pyrénéenne*. Halle (Germany): Max Niemeyer.

Rootsi, Siiri. 2006. “Y-Chromosome Haplogroup I Prehistoric Gene Flow in Europe.” *Documenta Praehistorica* 33 (December): 17–20. https://doi.org/10.4312/dp.33.3.

Rosario, Calderón, and Esther Rebato. 1997. “Historia de La Antropología Biológica En El País Vasco (A History of Biological Anthropology in the Basque Country).” *Revista Internacional de Los Estudios Vascos* 42 (1): 45–62.

Ruiz, Teofilo F. 2018. *The Western Mediterranean and the World :*

400 CE to the Present. John Wiley & Sons, Ltd.

Sajantila, A, P Lahermo, T Anttinen, M Lukka, P Sistonen, M L Savontaus, P Aula, et al. 1995. “Genes and Languages in Europe: An Analysis of Mitochondrial Lineages.” *Genome Research* 5 (1): 42–52. https://doi.org/10.1101/gr.5.1.42.

Samida, Stefanie, and Jörg Feuchter. 2016. “Why Archaeologists, Historians and Geneticists Should Work Together – and How.” *Medieval Worlds* medieval worlds (Volume 2016.4): 5–21. https://doi.org/10.1553/medievalworlds_no4_2016s5.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. “DNA Sequencing with Chain-Terminating Inhibitors.” *Proceedings of the National Academy of Sciences* 74 (12): 5463–67. https://doi.org/10.1073/pnas.74.12.5463.

Sazzini, Marco, Stefania Sarno, and Donata Luiselli. 2014. “The Mediterranean Human Population: An Anthropological Genetics Perspective.” In *The Mediterranean Sea,* 529–51. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-6704-1_31.

Schaefer, Nathan K., and Beth Shapiro. 2019. “New Middle Chapter in the Story of Human Evolution.” *Science* 365 (6457): 981–82. https://doi.org/10.1126/science.aay3550.

Schaefer, Nathan K., Beth Shapiro, and Richard E. Green. 2016. “Detecting Hybridization Using Ancient DNA.” *Molecular Ecology* 25 (11): 2398–2412. https://doi.org/10.1111/mec.13556.

Schraiber, Joshua G., and Joshua M. Akey. 2015. “Methods and Models for Unravelling Human Evolutionary History.” *Nature*

Reviews Genetics 16 (12): 727–40. https://doi.org/10.1038/nrg4005.

Semino, O., G Passarino, P J Oefner, A A Lin, S Arbuzova, L E Beckman, G De Benedictis, et al. 2000. “The Genetic Legacy of Paleolithic Homo Sapiens Sapiens in Extant Europeans: A Y Chromosome Perspective.” *Science* 290 (5494): 1155–59. https://doi.org/10.1126/science.290.5494.1155.

Shinde, Vasant, Vagheesh M Narasimhan, Nadin Rohland, Swapan Mallick, Matthew Mah, Mark Lipson, Nathan Nakatsuka, et al. 2019. “An Ancient Harappan Genome Lacks Ancestry from Steppe Pastoralists or Iranian Farmers.” *Cell* 0 (0). https://doi.org/10.1016/j.cell.2019.08.048.

Slatkin, M. 1991. “Inbreeding Coefficients and Coalescence Times.” *Genetical Research* 58 (2): 167–75. http://www.ncbi.nlm.nih.gov/pubmed/1765264.

Soares, Pedro, Alessandro Achilli, Ornella Semino, William Davies, Vincent Macaulay, Hans-Jürgen Bandelt, Antonio Torroni, and Martin B. Richards. 2010. “The Archaeogenetics of Europe.” *Current Biology* 20 (4): R174–83. https://doi.org/10.1016/J.CUB.2009.11.054.

Solé-Morata, Neus, Patricia Villaescusa, Carla García-Fernández, Neus Font-Porterias, María José Illescas, Laura Valverde, Francesca Tassi, et al. 2017. “Analysis of the R1b-DF27 Haplogroup Shows That a Large Fraction of Iberian Y-Chromosome Lineages Originated Recently in Situ.” *Scientific Reports* 7 (1): 1–13. https://doi.org/10.1038/s41598-017-07710-x.

Solomon, E., and W. F. Bodmer. 1979. “Evolution of Sickle Variant Gene.” *Lancet (London, England)* 1 (8122): 923. https://doi.org/10.1016/s0140-6736(79)91398-9.

Stone, W. H., J. J. Ely, G. S. Manis, and J. L. VandeBerg. 1993. “Classical Genetic Markers and DNA Markers: A Commensal Marriage.” *Primates* 34 (3): 365–76. https://doi.org/10.1007/BF02382632.

Stoneking, Mark, and Johannes Krause. 2011. “Learning about Human Population History from Ancient and Modern Genomes.” *Nature Reviews Genetics* 12 (9): 603–14. https://doi.org/10.1038/nrg3029.

Straus, Lawrence Guy. 1990. “Human Occupation of Euskalerria during the Last Glacial Maximum: The Basque Solutrean.” *MUNIBE (Antropología y Arqueología)* 42: 33–40.

Szécsényi-Nagy, Anna, Christina Roth, Guido Brandt, Cristina Rihuete-Herrada, Cristina Tejedor-Rodríguez, Petra Held, Íñigo García-Martínez-de-Lagrán, et al. 2017. “The Maternal Genetic Make-up of the Iberian Peninsula between the Neolithic and the Early Bronze Age.” *Scientific Reports* 7 (1): 15644. https://doi.org/10.1038/s41598-017-15480-9.

Tang, Hua, Jie Peng, Pei Wang, and Neil J. Risch. 2005. “Estimation of Individual Admixture: Analytical and Study Design Considerations.” *Genetic Epidemiology* 28 (4): 289–301. https://doi.org/10.1002/gepi.20064.

Templeton, Alan R. 2019. *Human Population Genetics and Genomics*. Academic Press.

The International HapMap Consortium. 2003. “The International

HapMap Project.” *Nature* 426 (6968): 789–96. https://doi.org/10.1038/nature02168.

———. 2005. “A Haplotype Map of the Human Genome.” *Nature* 437 (7063): 1299–1320. https://doi.org/10.1038/nature04226.

Thierry, Truffaut. 2010. “Contribution à La Connaissance de La Permanence Ursine Dans Les Diverses Manifestations Culturelles, Cultuelles et Festives Dans Le Périmètre de l’ancienne Province Romain de Novempopulanie.” *Munibe Antropologia - Arkeologia, ISSN 1132-2217, Nº 61, 2010, Págs. 407-434*, no. 61: 407–34.

Thomas, Hugh. 2005. *Rivers of Gold: The Rise of the Spanish Empire, from Columbus to Magellan*. Random House.

Torroni, A, H J Bandelt, L D’Urbano, P Lahermo, P Moral, D Sellitto, C Rengo, et al. 1998. “MtDNA Analysis Reveals a Major Late Paleolithic Population Expansion from Southwestern to Northeastern Europe.” *American Journal of Human Genetics* 62 (5): 1137–52. https://doi.org/10.1086/301822.

Trask, R. L. 1997. *The History of Basque*. Routledge.

Urteaga Artigas, María. 1988. “Los Bronces Romanos de Higer. Hondarribia. Guipúzcoa.” *Munibe Antropologia - Arkeologia,* no. 40: 111–22.

Valverde, Laura, Maria José Illescas, Patricia Villaescusa, Amparo M Gotor, Ainara García, Sergio Cardoso, Jaime Algorta, et al. 2016. “New Clues to the Evolutionary History of the Main European Paternal Lineage M269: Dissection of the Y-SNP S116 in Atlantic Europe and Iberia.” *European Journal of*

Human Genetics 24 (3): 437–41. https://doi.org/10.1038/ejhg.2015.114.

Venter, J. Craig, Mark D. Adams, Eugene W. Myers, Peter W. Li, Richard J. Mural, Granger G. Sutton, Hamilton O. Smith, et al. 2001. “The Sequence of the Human Genome.” *Science* 291 (5507): 1304–51. https://doi.org/10.1126/science.1058040.

Villaescusa, Patricia, María José Illescas, Laura Valverde, Miriam Baeta, Carolina Nuñez, Begoña Martínez-Jarreta, Maria Teresa Zarrabeitia, Francesc Calafell, and Marian M. de Pancorbo. 2017. “Characterization of the Iberian Y Chromosome Haplogroup R-DF27 in Northern Spain.” *Forensic Science International: Genetics* 27 (March): 142–48. https://doi.org/10.1016/J.FSIGEN.2016.12.013.

Villar, Francisco, and Blanca María Prósper. 2005. *Vascos, Celtas e Indoeuroeos: Genes y Lenguas*. Salamanca (Spain): Ediciones Universidad de Salamanca.

Villasante, Luis. 1979. *Historia de La Literatura Vasca*. 2nd ed. Burgos: Aranzazu.

Vogel, Friedrich, and Arno G. Motulsky. 1997. *Human Genetics*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-03356-2.

Wagner, F F, and W a Flegel. 2000. “RHD Gene Deletion Occurred in the Rhesus Box.” *Blood* 95 (12): 3662–68. https://doi.org/10.1016/S0887-7963(01)80058-4.

Waldman, Carl, and Catherine Mason. 2006. *Encyclopedia of European Peoples*. Edited by Facts On File. New York (USA).

Wangkumhang, Pongsakorn, and Garrett Hellenthal. 2018.

“Statistical Methods for Detecting Admixture.” *Current Opinion in Genetics & Development* 53 (December): 121–27. https://doi.org/10.1016/J.GDE.2018.08.002.

Weber, J L, and P E May. 1989. “Abundant Class of Human DNA Polymorphisms Which Can Be Typed Using the Polymerase Chain Reaction.” *American Journal of Human Genetics* 44 (3): 388–96. http://www.ncbi.nlm.nih.gov/pubmed/2916582.

Weir, B. S., and C. Clark Cockerham. 1984. “Estimating F-Statistics for the Analysis of Population Structure.” *Evolution* 38 (6): 1358. https://doi.org/10.2307/2408641.

Weir, B. S., and W. G. Hill. 2002. “Estimating F-Statistics.” *Annual Review of Genetics* 36 (1): 721–50. https://doi.org/10.1146/annurev.genet.36.050802.093940.

Wells, H. G. 1920. *The Outline Of History: Being a Plain History of Lifeand Mankind*. London: Cassell & Company LTD.

Westhoff, Connie M. 2007. “Rh Complexities: Serology and DNA Genotyping.” *Transfusion* 47 (s1): 17S-22S. https://doi.org/10.1111/j.1537-2995.2007.01305.x.

Wolinsky, Howard. 2008. “Our History, Our Genes. Population Genetics Lets Researchers Look Back in Time at Human Migrations.” *EMBO Reports* 9 (2): 127–29. https://doi.org/10.1038/sj.embor.7401164.

Wren, Colin D., and Ariane Burke. 2019. “Habitat Suitability and the Genetic Structure of Human Populations during the Last Glacial Maximum (LGM) in Western Europe.” Edited by Michael D. Petraglia. *PLOS ONE* 14 (6): e0217996. https://doi.org/10.1371/journal.pone.0217996.

Wright, S. 1951. “The Genetical Structure of Populations.” *Annals of Eugenics* 15 (4): 323–54. http://www.ncbi.nlm.nih.gov/pubmed/24540312.

Young, Kristin L, Guangyun Sun, Ranjan Deka, and Michael H Crawford. 2011. “Autosomal Short Tandem Repeat Genetic Variation of the Basques in Spain.” *Croatian Medical Journal* 52 (3): 372–83. https://doi.org/10.3325/cmj.2011.52.372.

Zlojutro, M., A. G. Apraiz, R. Roy, J. Palikij, and Michael H. Crawford. 2006. “Autosomal STR Variation in a Basque Population: Vizcaya Province.” *Human Biology* 78 (5): 599–618. https://doi.org/10.1353/hub.2007.0007.

Zuazo, Koldo. 2010. *El Euskera y Sus Dialectos*. Irún: Alberdania S. L.

Zubillaga Gil, Eliseo. 1990. “La Romanización En Alava: Valoración Arqueológica.” *Munibe Antropologia - Arkeologia, ISSN 1132-2217, Nº 42, 1990 (Ejemplar Dedicado a: Homenaje a D. José Miguel de Barandiarán), Págs. 327-336*, no. 42: 327–36.

6.APPENDIX

6.1. Patterns of genetic structure and adaptive positive selection in the Lithuanian population from high-density SNP data

Urnikyte A, Flores-Bello A, Mondal M, Molyte A, Comas D, Calafell F, et al. Patterns of genetic structure and adaptive positive selection in the Lithuanian population from high-density SNP data. Sci Rep. 2019 Dec 1;9(1). DOI: 10.1038/s41598-019-45746-3