Genomic insights into human complex demography:

A study of western and eastern European Romani populations

Giacomo Francesco Ena

TESI DOCTORAL UPF / 2025

DIRECTORS DE LA TESI

Dr. David Comas Martinez

Dr. Marcos Araújo Castro e Silva

DEPARTAMENT DE MEDICINA I CIÈNCIES DE LA VIDA

Universitat Pompeu Fabra
Barcelona

« The truths expressed by scientific research are not absolute and definitive truths but historical approximations, and science is a continually evolving movement. If scientific truths were indeed definitive and permanently established on an absolute and meta-historical level, science as such would have ceased to exist. Thus, science is a historical category; it provides parameters for interpreting reality that vary and will continue to vary with the changes in historical epochs. In reality, science is also a superstructure, an ideology. Therefore, science does not possess absolute validity beyond time but represents, in its history, the reflection of real power dynamics within classes and modes of production. »

Antonio Gramsci

«Sas beridadis espressas dae sa circa scientìfica no funt beridadis assolutas e definitivas, ma funt aprossimatzionis istóricas, e sa chistioni iscientìfica est unu movimentu in contìnuu desvoltu. Chi sas beridadis scientìficas fesint beridadis definitivas e crobadas in manera assoluta e metaistórica, s'iscientzia cumenti tale iat acabadu dde esistit. Duncas, s'iscientzia est una categoria istórica; oferesit parametrus de interpretatzioni de sa realidàdi chi cambiant e chi ant a cambiai cun su istramudinzu de is épocas istóricas. In realidàdi, puru sa scientzia est una superstrutura, una ideologia. Pro tantu, s'iscientzia no tenit una balidesa assoluta in pitzus de su tempu, ma rapresentat in s'istòria sua su reflussu de is relatzionis de fortzas reales a intru de is classis e is medius de produtzioni.»

Antoniu Gramsci

This thesis would not have been possible without the support of many people. I am deeply grateful to all the volunteers who participated in this research; your contributions were invaluable. I want to thank my supervisor, David, for offering me this opportunity and guiding me throughout the process, as well as my co-supervisor, Marcos, for his patience, insights, and mentorship. I also want to acknowledge the encouragement and assistance from both past and present members of the research group—it has been a privilege to learn and grow alongside you. Finally, I want to express my heartfelt gratitude to my partner and family for their unwavering moral support throughout these years. Your encouragement kept me grounded and motivated through the challenges of this experience.

ABSTRACT

The Romani people, Europe's largest transnational ethnic minority, originate from South Asia and have undergone complex demographic shifts during their diaspora. This thesis presents genome-wide and Y-DNA analyses revealing a shared ancestry among Romani groups, alongside distinct genetic structures across Europe. Mating patterns, such as sex-biased gene flow and assortative mating, have significantly shaped their genetic profile. Gene flow with non-Romani populations often offset diversity losses due to bottlenecks and founder effects. However, Spanish Roma show much higher internal genetic differentiation than non-Romani Iberians, demonstrating the presence of a substructure. The study emphasizes the need to include underrepresented populations in genetic research to capture human diversity more fully. It also highlights the potential for tailored public health strategies and promotes social understanding through the study of marginalised groups.

RESUM

El poble Romaní, la minoria ètnica transnacional més gran d'Europa, té els seus orígens al sud d'Àsia i ha experimentat complexos canvis demogràfics al llarg de la seva diàspora. Aquesta tesi presenta anàlisis de dades genòmiques i d'ADN-Y que revelen una ascendència comuna entre els grups romanís, així com estructures genètiques distintes arreu d'Europa. Els patrons d'aparellament, com el flux genètic amb biaix de sexe i l'aparellament assortatiu, han modelat de manera significativa el seu perfil genètic. El flux genètic amb poblacions no romanís ha contribuït sovint a compensar la pèrdua de diversitat causada per colls d'ampolla i efectes fundador. Tanmateix, els gitanos espanyols mostren una diferenciació genètica interna molt superior a la dels ibèrics no Romanís, la qual cosa demostra la presència d'estructura interna. L'estudi subratlla la necessitat d'incloure poblacions infrarepresentades en la recerca genètica per captar millor la diversitat humana, i destaca el potencial per a estratègies de salut pública personalitzades, alhora que promou la comprensió social mitjançant l'estudi de grups marginats.

PREFACE

This dissertation explores the genetic, demographic, and historical dimensions of the Roma people, a population that has long been under-represented in genetic and medical research. Beyond its biological relevance, the study also offers anthropological and historical insights into migration, discrimination, endogamy, sex-biased gene flow. By examining admixture patterns between Roma and other European populations, the research highlights past cultural interactions and their consequences. This multidisciplinary approach bridges genetics, anthropology, and history, addressing key gaps in our understanding of the Roma. The research involved notable challenges, including maintaining scientific rigor and ethical standards while ensuring comprehensive analysis. The computational aspects, especially data processing and coding, demanded precision and sustained effort to yield accurate results. Conversely, the sociocultural and anthropological aspects of this study were particularly rewarding, as they allowed for a broader exploration of how historical and societal factors have influenced the Roma genetic landscape.

My motivation for undertaking this research stems from a simple but enduring curiosity—the desire to investigate, discover, and contribute to scientific knowledge. This work is the product of that curiosity, combined with a commitment to advance our knowledge of human genetic diversity through the lens of an often-overlooked population.

Table of Contents

1. INTRODUCTION

1.1 The Roma people

The terminology surrounding the Roma, including the often-controversial term “Gypsy,” is addressed in section 1.6.1 to provide clarity and context.

1.1.1 The history of the Roma people

1.1.1.1 Overview of the proto-Roma, the Out-of-India and the first diaspora

The Roma are unique among historically migratory peoples in that they have never been tied to a specific homeland, either in a physical or abstract sense. While other nomadic groups, such as the Bedouin, may share a cultural or symbolic connection to particular lands, the Roma lack a tradition of an ancient or distant homeland from which their ancestors are said to have migrated. Unlike many ethnic groups, whose histories are rooted in particular lands or states, they do not claim the right to national sovereignty in any of the lands where they reside. The Roma identity has instead been shaped by a centuries-long journey across diverse regions and cultures. This mobility has played a significant role in both the formation of their cultural identity and their complex relationship with the places they have inhabited. To understand the origins of the Roma, however, it is important to consider the context and territories from which the proto-Roma emerged. The precise moment when the proto-Roma left South Asia remains uncertain, as do the specific reasons for their departure or whether they belonged to a particular caste or homogeneous social group. Although the exact timing is uncertain, the migration of the proto-Roma from South Asia is often referred to, especially in genetics, as the "Out of India" event, with subsequent movements described as the Romani diaspora. What is certain, however, is that the Roma originated in regions that, over the span of 200 years, were governed by various rulers, including the Hindu Shahi dynasty, the

Ghaznavid Empire, and later the Ghurid Empire. The proto-Roma are believed to have left during the period of Ghaznavid rule, which was a Persianate Muslim dynasty of Turkic mamluks. This entity existed from 977 to 1186 and, at its height, governed a vast territory extending from modern-day Uzbekistan to the Indus Valley, encompassing what is now North-Western India [1].

Scholars have proposed various hypotheses regarding the origins of the Roma people, with two main speculative perspectives dominating the debate. classes (for more on their social status within the Ottoman empire see section 1.1.2.2). The first hypothesis suggests that the Roma descend from an itinerant caste specialising in service-based trades, such as crafts and entertainment [2,3], gradually migrating through Asia before reaching Europe [4–6]. Some authors extend this theory further, proposing that their migration from South Asia occurred between 600 CE and 800 CE, potentially driven by economic hardship, such as famine, or socio-political pressures [7]. The second hypothesis, instead, associates the Roma with a contingent of warriors from Rajasthan who left South Asia after the year 1000 AD, to participate in the Battle of Dandanaqan against the Seljuk Empire. After their defeat, these people were reportedly taken into Seljuk territories, possibly Anatolia, as captives or displaced peoples [8].

Supporting the first hypothesis, various ethnonyms used in regions of Iran—one of the first areas the proto-Roma would have crossed upon leaving South Asia—link the "gypsies" with different arts and crafts. For instance, in Azerbaijan, the term *Aseq* refers to musicians, while in Fars, *roubaz* denotes cow traders [7]. In medieval times, they were also referred to as *Luli*, *Louri*, *Kouli*, and *Zot* (*Jat*). The latter term is reminiscent of *jāti*, a term in India used to describe endogamous social groups that form the foundation of the traditional caste system. Persian sources from the early medieval period provide some of the earliest references to these groups. Hamza-i (d. 972 CE) and Ferdowsi (d. 1033 CE) mention the arrival of "Gypsies" in Iran. According to these accounts, Bahram-i, the Sasanian king who reigned from 421–439 CE, summoned 12,000 *Luli* or "Gypsy" musicians from India [7]. Several medieval Muslim historians also recorded details about these migrations. For example, Mas'udi (d. 957 CE) described how a famine in India forced the *Zot* (or *Zatt*) to migrate in large numbers, settling in several Iranian provinces, including Kerman, Fars, and Khuzestan. Over time, they gained power but were defeated by the Abbasid Empire in 835 CE, leading to their relocation to Syria. Evidence also suggests that, as early as the late Sasanian period, groups of Hindus known as *Lat* or *Kat* (or *Zatt*) migrated from India in different waves, settling in southeastern Khuzestan in what is now modern-day Iran [7].

An argument for the second hypothesis is based on linguistic evidence regarding the Romani language. The absence of agricultural or civilian terms of Indic origin, coupled with the presence of military-related vocabulary, has been interpreted by proponents of this theory as evidence that Romani migration was linked to a military or warrior-based context [8,9]. This would support the theory that the Roma were part of a military force that left India, possibly during or after conflicts such as the Ghaznavid invasions of the Hindu Shahis [9]. Furthermore, certain linguistic features suggest that the separation of the Romani language from its Indic roots likely occurred no earlier than around AD 1000, with significant phonological changes and vocabulary reflecting influences from the period following the 9th and 10th centuries. This evidence points to the migration of the Roma from India after this date, supporting the hypothesis of an exodus during or after the early medieval period, rather than earlier [8]. In addition, some scholars propose that the Roma in Iran are descendants of eastward migrations during the Ottoman Empire in the 16th century [10], an idea further supported by linguistic similarities between the Zargari group in Iran and Roma populations in the Balkans [11,12].

However they reached West Asia, sources suggest that some of the earliest recorded mentions of the Roma can be traced back to the Byzantine Empire, which was then known as the Roman Empire [13]. Ecclesiastical sources from this period identify a group known as the Aiguptoi or "Egyptians," also referred to as the Athinganoi. One such reference comes from Patriarch Anastasios I of Constantinople, who, during his second tenure (23 June 1303 – September 1309), admonished his clergy to caution parishioners against associating with fortune-tellers, bear-leaders, and snake-charmers, specifically naming the Atsinganoi as culprits (Marsh, Adrian 2008). The Byzantine Empire is also where the term Romitoi was first recorded, which some scholars believe is linked to the origin of the term "Roma" [13]. While the term "Roma" is commonly thought to derive from the Romani word for "man", a related term, Romiti, was noted in 1384 by Lionardo di Niccolò Frescobaldi, who observed penitents outside the city walls of Modon (modern Methoni, Greece) referring to themselves by this name.

During their time in the Byzantine Empire, the Roma, as non-Christians, were likely subject to a specific head tax known as the kephalition or "capitation tax," imposed by Christian rulers. Additional levies from the Church and widespread prejudice further exacerbated their social and economic hardships [14–16]. These conditions likely contributed to their subsequent migrations across Europe. References to Roma-like groups in the Byzantine Empire become more frequent over time, though often in a negative context, variously referred to as Egyptians or Atsinganoi, as noted by Nikephoros Gregoras and the scholar Joseph Bryennius (1340–1431), who described the Atsinganoi as practitioners of magic, fortune-telling, and enchantment, perceived as detrimental to the Empire [14,17].

Although it is impossible to definitively establish whether any of these groups were the direct ancestors of the modern Roma, many of the hypotheses remains intriguing. Groups such as the Zott, Jat, Romitoi, Athinganoi, and Atsingani have all been proposed as potential precursors to the Roma, though these connections remain speculative. Some researchers argue that these groups may be more closely related to the Lom and Dom peoples, who, while sharing South Asian origins with the Roma, are considered distinct from them [18]. The Lom, primarily found in the Caucasus, and the Dom, dispersed across the Middle East and North Africa, may have migrated much earlier than the Roma. Despite being historically labelled as 'Gypsies' by outsiders, both the Lom and the Dom do not identify as Roma, and their languages bear no more similarity to Romani than other Indo-Aryan languages, such as Gujarati [19–21]. The earlier arrival of these groups may explain reports of nomadic peoples of South Asian origins in the Byzantine period, predating the arrival of the Roma. However, this provides little insight into the specific identity of the proto-Roma. One view is that within the Byzantine Empire—where the Roma are believed to have lived for at least 200 years—a distinct Romani identity have begun to emerge [14,22].

1.1.1.2 Arrival in Europe and second diaspora

The Roma, who had likely been present in the Byzantine Empire for several centuries by the 1400s, presumably arrived through Anatolia, with some groups crossing the Bosphorus Strait into the European territories of the Empire or travelling to other Greek regions by boat. While spontaneous movements of Roma into Byzantine Greek regions occurred as early as the 12th century [14], it is plausible that large groups entered as prisoners of war and were enslaved following conflicts with the Turks [23].

In the Ottoman Empire, they were often integrated into local communities, becoming part of the artisan and labour classes (for more on their social status within the Ottoman empire see section 1.1.2.2). They continued their nomadic or semi-nomadic lifestyle as long as possible, which facilitated their rapid spread across the continent.

Reliable accounts of the Roma's early history in Asia and Europe are non-existent, with only fragmentary evidence remaining. As a result, the interpretation of this part of their history continues to be a subject of scholarly debate, and the exact timing of their initial settlement in Europe remains uncertain. The period between 1200 and 1500 brought significant changes in this sense, as by the 14th century, contemporary evidence begins to unequivocally identify the Roma. Reliable records of their presence in Wallachia (part of modern-day Romania) appear as early as 1385. From this point onward, the Roma are increasingly mentioned in historical documents, demonstrating their growing recognition in many European cities [24].

The history of their migration within Europe is better documented than their earlier migration in Asia, with distinct waves of migrations recorded throughout modern and contemporary history. Generally travelling in groups of 50 to over 300 people, they would customarily approach the rulers of each city they visited, presenting safe conduct documents (similar to modern passports) [25–27] and recommendation letters. They often presented themselves as dukes or pilgrims in order to receive food, money, and goods [28], privileges granted to all pilgrims at the time [29]. When denied entry, they would temporarily camp outside the city walls, sleeping in tents in the open countryside.

The first wave of Roma migration outside the territories of the declining Byzantine Empire and the rising Ottoman Empire began between 1417 and 1421, marking what could be described as the "Out of Balkans" migration—a significant new phase in their diaspora [30–32]. This movement is often seen as the Roma’s departure from what some regard as their "second home"[33] in the Balkans, signalling the start of their expansion across Europe. Declaring their journey a pilgrimage and presenting letters of safe conduct, the Roma ensured a welcoming reception upon their first appearance in Central and

Western Europe. By 1540, when they were first reported in Åbo (modern Turku, Finland), they had reached nearly every extremity of the continent [24].

This dispersal across Europe continued into the 17th century, as Roma moved from the Balkans through Hungary and Poland into Russia, often in search of economic opportunities or refuge from local persecution. During the thirty years' War (1618-1648), famine, epidemics, and flight affected many areas of central Europe. This also led to increased movements of Roma towards Bohemia and Moravia, which in turn strengthened measures specifically targeting the newly arrived, fueling growing resentment towards the Roma [17,34]. At the same time, in the Holy Roman Empire (a decentralized federation of territories in central Europe), Roma faced increasing restrictions and expulsions, with numerous decrees calling for their removal or forced settlement. From 1697 onwards, the Holy Roman Emperor issued increasingly severe edicts, declaring that "gypsies" were to be "vogelfrei" (outlaws), treated "with all possible severity both as regards body and property", and hanged without trial [34]. As a result of these harsh measures, many Roma were forced into near-constant movement.

During the 18th century, Roma communities were scattered throughout much of Europe, but their lifestyle and movements became more constrained. The early part of this century saw the rise of European colonial powers and the expansion of empires, which also impacted Roma migration patterns [35]. In the Habsburg lands, Roma were systematically persecuted, forced into settled communities, or expelled, as authorities sought to control nomadic populations [36]. However, many Roma in the Russian Empire continued to thrive in their semi-nomadic way of life, especially in southern Russia, where they were sometimes allowed to retain their traditional practices.

The 19th century marked a critical period of transformation for many peoples in Europe, including the Roma. The century began with a series of major conflicts, the Napoleonic wars, which caused widespread casualties and displacement across large parts of continental Europe [37]. While Roma were not explicitly mentioned in casualties' records, it is reasonable to assume they were also affected by these events. The consequent dissolution of the Holy Roman Empire and the establishment of the Austrian Empire brought about a shift in policies towards the Roma, with an increasing focus on settling them in forced communities rather than expelling them [38,39]. As industrialisation advanced and the borders of nation-states were established, many Roma were pushed into permanent settlements under state policies of forced assimilation. However, at the same time, the Industrial Revolution led to unstable labour relations, which significantly increased the migration of Roma workers in search of job opportunities [40]. In the German states, the British Isles, and parts of Scandinavia, laws were enacted to curtail the Roma's nomadic lifestyle and integrate them into settled communities. Many Roma were subjected to a process of sedentarisation, where they were made to abandon their traditional occupations and lifestyles in favour of agricultural or labour-based work, often under coercion.

The abolition of slavery in the Principalities of Wallachia and Moldova triggered a significant migration of Roma people. This process unfolded through a series of laws between 1831 and 1856, culminating in the *Act for the Emancipation of All Gypsies in the Romanian Principality* [41]. This decree officially ended slavery and granted freedom to all Roma in this category, who were subsequently registered as taxpayers to the state. Following these events, a large portion of the Roma population left Romania in search of new opportunities abroad. Although precise figures are unavailable, estimates suggest that between 250,000 and 300,000 Roma were affected. This wave of migration is believed to have been particularly significant and substantial, with tens of thousands relocating to different countries [42]. From the second half of the century, just like many other Europeans, Roma started a wave of migration to the Americas [28,43,44].

By the early 20th century, many Roma in Europe had long been forced to abandon their nomadic ways, although a small number in rural areas were able to maintain a semi-nomadic existence, particularly during the warmer months. This era witnessed increased ethnic categorisation, as nationalist ideologies and the rise of racial theories in the late 19th and early 20th centuries began to impact the Roma, exacerbating their marginalised position in society. By the 1930s and 1940s, with the rise of fascism and Nazi ideology in Europe, Roma communities became targets of the racial policies of the Third Reich, the Axis powers, and their vassals, including Romania and the Scandinavian countries [45–50]. This culminated in the Romani Holocaust during World War II, where hundreds of thousands of Roma were murdered in concentration camps, further reinforcing the tragic consequences of centuries of persecution [45].

1.1.1.3 Roma mobility in contemporary history

In the aftermath of the genocide perpetrated by the Axis powers, Roma communities across several European regions—including those directly under Nazi control, as well as in Croatia and Czechoslovakia—were decimated [51–53]. Between 25% and 50% of the total European Roma population was killed, amounting to between 1.5 and 2 million individuals [45]. In addition, the Second World War was a catastrophic event that displaced millions of people [54], including an unknown and immeasurable number of Roma. This not only prompted new migrations within Europe, but also led to migration to other continents, such as the Americas and Oceania [55].

However, the end of World War II did not mark the end of Roma suffering. Most Roma Holocaust survivors in the Federal Republic of Germany (FRG, also known as West Germany) and Austria continued to face systemic discrimination, being denied official recognition and compensation for their persecution [56–59]. Local authorities and welfare administrations frequently refused to acknowledge their ethnic victimisation, often imposing bureaucratic obstacles such as requiring proof of permanent residence and employment [56–58]. In many cases, officials in the FRG who had been involved in Nazi-era policies continued to influence post-war decisions, further marginalising Roma survivors. In Austria, Roma who had lost their homes and property were unable to reclaim them due to legal technicalities.

In the German Democratic Republic (GDR, or East Germany) and Eastern European countries, despite the complete dismantling of Nazi legal apparatus, Roma were either not recognised as victims or had to demonstrate ideological loyalty to gain recognition and compensation [60]. This difficulty, even within the socialist system that granted Roma citizens equal rights to other ethnic groups, can be attributed to the fact that outside Germany, their persecution did not receive the same level of attention as the Jewish Holocaust, and awareness of Roma suffering remained minimal. Even among Roma activists and intellectuals, significant discussions on the topic only began in the 1990s [61,62]. It took decades for Roma Holocaust survivors to receive fair compensation in Western countries, with notable progress occurring only in recent years.

In Eastern European countries such as Bulgaria, Hungary and the Soviet Union, state policies generally aimed to recognise all ethnicities as equal. While scholars and activists debate the extent of discrimination against Roma in these countries, critiques—particularly from scholars based in Western academic traditions [63–65]—highlight various shortcomings in the response to Roma issues. However, even studies that offer harsh assessments of Roma living conditions under socialism, often influenced by political bias, acknowledge that the planned economy significantly improved their access to housing, education, healthcare, and stable employment [64–67]. A central point of contention is the sedentarisation of nomadic Roma during the 20th century, which critics describe as a repressive policy imposed by socialist governments, particularly in Euro-American human rights and academic discourse [68–70]. However, historical evidence indicates that many Roma had already been living in settled communities for centuries, and sedentarisation was a process that had begun long before the socialist era [68,71]. Socialist governments implemented policies that facilitated this transition, providing housing and promoting integration into urban areas, particularly between the 1950s and 1970s. Certain policies, such as property confiscation in Czechoslovakia and Poland, have been criticised as coercive, but the broader process of sedentarisation was influenced by multiple factors, including economic changes that made nomadic life increasingly difficult during the mid-20th century.

A case in point is Filákovo, in southeastern Slovakia. In the 1970s, Roma huts were demolished, and residents were relocated to newly constructed apartment buildings as part of broader urban development efforts. At the time, the local steelworks employed around 8,000 people, including hundreds of Roma. However, following the economic transition after 1989, the factory was privatised and became part of a multinational corporation. In the early 2000s, only 800 workers remained, and there were no Roma among them [24]. This reflects wider socio-economic shifts that have had a significant impact on Roma communities in the post-socialist era. The transition from socialism to capitalism was accompanied by a resurgence of nationalist and reactionary movements, during which mob violence against Roma communities occurred in several countries, including Romania [72], Ukraine [73], and Poland [74] . The fragmentation of Yugoslavia also led to ethnic conflicts, which contributed to the displacement of Balkan Roma towards the West and the occurrence of acts of violence against them [75]. Therefore this 'third migration' of Roma from Eastern to Western Europe in the second half of the 20th century was driven not only by war and political unrest but also by systemic racism, discrimination, violent pogroms, and economic hardships. These pogroms, sometimes incited by political figures, have continued into the 21st century, as evidenced by incidents in Central [76,77] and Eastern Europe [78,79], such as the 2016 anti-Roma riots in Loshchynivka, Ukraine [80]. In this case, local authorities, including the police and the mayor, failed to intervene during the attacks and, in some instances, supported the violent expulsion of Roma families. The legal system also proved complicit, as lawsuits seeking justice for the victims were dismissed, with the court ruling that the forced evictions were an expression of "direct democracy." These incidents of state-sanctioned violence resulted in further migrations of Roma to Western Europe, particularly to countries with established Roma communities and greater tolerance, such as Spain and the Netherlands [81]. Beyond these various migration waves in the late 20th century, which were primarily east-to-west within Europe, detailed information about modern Roma movements outside the continent remains limited. However, outside of Eurasia, significant Roma populations are known to exist in countries such as Brazil, Argentina, and the United States [43,82,83].

1.1.2 Roma society and social status

1.1.2.1 Historical overview of the status and evolution of Roma society

The arrival of the Roma in Europe marked the beginning of a complex and often challenging relationship with their new surroundings. Initially met with hospitality—similar to that typically extended to all pilgrims—their presence eventually led to growing hostility, largely due to societal intolerance of their cultural and lifestyle differences. As early as the second half of the 15th century, the settled population began pushing them away from the urban areas. The 1482 edict by Elector Albrecht Achilles of Brandenburg, which prohibited the "Gypsies" from staying in his territory, and the 1497 declaration during the Reichstag (Parliament session) in Lindau, which declared the "Gypsies" outlaws, marked some of the first significant steps towards the large-scale persecution of the Roma. In France, by 1504, Louis XII ordered the banishment of Roma, and by 1510, defying this ban was punishable by death by hanging [84].

Initial frictions may have arisen due to the Roma's nomadic or semi-nomadic way of life, which limited their interaction with broader European societies and, in many cases, contributed to their marginalisation. As a result, they were often assigned a lower social status and regarded as semi-outsiders. Some of their occupations and distinct social practices (for definitions see section 1.4.3 Social norms and practices), such as nomadism, palmistry and fortune-telling, were frequently viewed with suspicion by non-Roma populations, whether in Christian or Muslim societies. These social and cultural differences contributed to various forms of discrimination, marginalisation, persecution, and, at times, enslavement.

Traditionally, Roma worked as metalworkers, woodworkers, animal traders, entertainers and beggars [8,28,85]. The Christian Church viewed their practices, particularly the so-called "unchristian" medical remedies, palmistry, and other activities labelled as "sorcery," as deeply problematic [86,87]. Local workers and guilds often saw the Roma as unwelcome competition, perceiving a threat to their income and monopolies, and sought to exclude them through various means [88]. As the Roma's reputation declined and the perceived costs associated with their presence grew, cities became less willing to tolerate them.

In the modern era, rulers often viewed the Roma as a group of unproductive vagabonds who resisted integration into the established social order and lacked loyalty to any authority. Consequently, many rulers introduced anti-Roma laws aimed at controlling or excluding them entirely. However, the treatment of the Roma varied significantly across different regions of Europe. The Ottoman Empire was relatively more "tolerant" in comparison to regions such as the Romanian Principalities, the Holy Roman Empire, and later the Austrian Empire, which were among the least tolerant.

1.1.2.2 The Roma in the Ottoman Empire

The Roma became part of Ottoman society following the empire's expansion into formerly Greek Byzantine territories, where they had already lived for centuries. Most remained in these regions, largely unaffected by the political transitions that followed. This was because the Ottoman authorities, who were relatively tolerant for the era, had no major issue with the Roma—except for their nomadism. The concern was not nomadism itself but rather the challenge of tracking their tax payments [33]. Although several regulations were issued by Suleiman I (1551) and Murad III (1574) to restrict nomadism or encourage settlement, these measures were largely ineffective. Nomadism and irregular tax payments were never considered major offences, and taxation records often registered nomadic Roma as settled. In practice, formal attempts to forcibly settle the Roma were rarely enforced, allowing many to continue their preferred way of life. Under the Ottoman rule all inhabitants were classed hierarchically into various categories, the main two being the "true believers" (the Muslims) and the "infidels" (the subjected Christian and Jewish populations). These categories were treated differently, the members of each having a different status and varying obligations towards the central state. All subjects of the Ottoman Empire, including Roma, were obliged to pay different taxes and were described in numerous tax registers. In these registers, Roma citizens were described by age, occupation and family status and, according to these and other factors, were grouped into tax units, each with its respective community leader.

The earliest Ottoman tax record mentioning "Gypsies" dates to 1430, in a register detailing a grant of lands for the administrative division (sanjak) of Nicopolis (modern Nikopol, Bulgaria), mentioning that 3.5% of the approximately 12,000 households were "Gypsies" [33]. The majority of Roma in the Ottoman Empire were settled, identified primarily by their ethnicity rather than their lifestyle, reflecting an early recognition of their distinct identity. From the 15th to the 16th century, settled Roma predominantly lived in segregated quarters in towns and villages across the Balkans and Anatolia, with records confirming a significant settled population.

Sultan Suleiman the Magnificent issued a 1530 law concerning Roma in the Province of Rumelia (which included most of the Balkans), highlighting their unique administrative and legal status within the Empire. Both Christian and Muslim Roma paid taxes, with Christians taxed slightly higher, though these distinctions were inconsistent. Exceptions were granted to Roma in the "Gypsy Sanjak", who provided essential military services such as blacksmithing, fortress maintenance, and music [89]. In this context, the sanjak was a special administrative category, not a territorial unit, and covered all Roma living in Rumelia and Istanbul [90].

To ensure proper tax collection, Roma leaders were made responsible for their communities' payments, a system reinforced by later decrees, such as Sultan Selim II's 1574 order for mine workers to organise into groups for accountability. Additionally, some Roma, particularly Muslims, were directly conscripted into the Ottoman army, with an estimated 15,000–20,000 serving during the 16th and 17th centuries. Beyond their military involvement, Roma were also prominent in guilds for professions such as bear breeding, horse trading, and music, as documented by traveller Evliya Çelebi in Istanbul [91].

The 17th and 18th centuries marked a decline in central authority and the rise of the *ayân*—influential local notables or provincial elites who wielded significant political and economic power within the Ottoman Empire. With the empire engaged in near-constant warfare and experiencing frequent territorial changes, records concerning the Roma became increasingly fragmented and uncertain [13,33]. Despite this, the Roma occupied a unique position within the empire's social and administrative framework. From the outset, they were considered "citizens," granted a distinct status that set them apart from other groups and recognised for their own specific legal and social standing.

Despite this, many historical sources highlight the evident contempt directed towards the Roma by both the Ottoman authorities and local populations, who often viewed them as inferior and undeserving of any respect or attention [33,92]. However, unlike in Western Europe, the Roma in the Ottoman Empire were not subjected to systemic persecution. Instead, they were largely tolerated, with their distinct social position allowing them to live relatively unhindered, though still marginalised, within the broader societal framework [33,92].

1.1.2.3 The Roma in Wallachia

In contrast to the more favourable treatment in the Ottoman Empire, Wallachia—*de iure* a territory under Ottoman suzerainty—stands out as one of the first European regions where the Roma settled and also as a place where they endured some of the harshest treatment. In this kingdom, their status was often perceived as subordinate or entirely outside the bounds of "culture" and "civilisation" [23]. By the late 15th century, slavery had become a central feature of the regional economy in Wallachia, with the Roma playing an integral role in this system. Despite the harsh conditions, both free and enslaved Roma coexisted in Wallachia [23,93], but were consistently identified by their occupation, social status, type of residence, and familial or social relations rather than by ethnicity [23]. The pervasive dehumanisation of Roma as a class became the dominant perspective among Romanians and others, following a troubling archetype: prisoners of war were labelled as Roma, Roma were enslaved, and slaves were deemed degenerate or flawed. Once this view of the Roma as less than human became entrenched, recognising them as anything else—particularly as contributing members of society—

was wholly dismissed. This perception fostered one of the most enduring contradictions in how dominant societies have understood the Roma.

Although the Roma faced marginalisation, they historically played vital roles in pre-capitalist societies, contributing as entertainers, craftspeople, and even state functionaries. However, these contributions were often overlooked, dismissed, or undervalued, reflecting the broader social devaluation of the Roma community. This disregard persists, despite the clear utility and enjoyment their work has provided to others.

Even in such society, not all Roma were passive or marginalised; in fact, they dominated certain sectors of the economy. In medieval Wallachia and Moldavia, they were heavily involved in crafts, metallurgy, and alcohol production, with Roma kettle smiths (*caldarari*) in 17th century Iași even forming their own guild. Renowned for their music as well, the Roma were an integral part of the feudal economy [23,42].

During this period, slavery reinforced the notion that even free peasants could lose their status, as non-Roma peasants were subjected to enserfment [23]. However, as feudalism waned and capitalism emerged, the Roma became increasingly defined in ethnic and racial terms. By the 19th century, Romanians began to identify themselves in opposition to this marginalised group, whose low social status was perceived as antithetical to national progress. This process significantly influenced the development of Romanian national identity and the establishment of the Romanian nation-state [23]. Capitalist modernisation and industrialisation in Romania pushed the Roma to the margins, as mass-produced goods replaced their traditional crafts in rural economies. Many Roma turned to day labour, contributing to their poverty and social exclusion. Their involvement in theft and petty crimes further entrenched their stigmatisation. As urbanisation progressed, more Roma, including former slaves, moved to cities, securing unskilled jobs in construction and factories. Many of them settled in overcrowded, low-income districts alongside other poor communities, reinforcing their continued marginalisation in early Romanian society [94].

1.1.2.4 The Roma in modern and contemporary times

In several Central European countries, including the Holy Roman Empire and later the Austrian Empire, the Roma faced systemic marginalisation and harsh legal restrictions. The Roma were often perceived as vagrants and outsiders, accused of lawlessness and a lack of allegiance to the state [95]. During the sixteenth century, the treatment of Roma in German territories became increasingly tied to laws governing vagabonds and paupers. In contrast, neighbouring Hungary exhibited greater tolerance—partly due to the skills and trades the Roma were known for, but primarily because Hungary was under Ottoman rule, which promoted a more lenient policy towards minority groups. Meanwhile, early German legislation sought to restrict Roma entry and later escalated to violent measures.

By 1650, Duke Eberhard III issued a decree calling for the violent expulsion of the Roma, accusing them of criminal activity and social harm. This rhetoric intensified by 1667, with official documents using the term "eradication" to describe the intended policy [96]. Hostility peaked with a 1706 edict under Duke Eberhard Ludwig, which authorised the killing of Roma on sight if they refused to leave [96]. Although these extreme laws were not fully enforced, they underscore the severe persecution Roma faced during this period, mirroring similar developments in England.

Throughout the 16th and 17th centuries, other imperial decrees criminalized Roma presence, ordering their expulsion or execution if caught [28,36]. At the time, the social status of the Roma across Europe was largely defined by exclusion, marginalisation, and forced assimilation. They were often perceived as outsiders, living on the fringes of society, and subjected to restrictive laws that limited their movement, economic activities, and cultural expression. In many regions, authorities attempted to forcibly settle them, banning the nomadic lifestyle. While some Roma integrated into local economies as artisans, traders, or entertainers, they were still socially stigmatized and often relegated to the lowest socio-economic classes. Cultural representations of the Roma in literature and art fluctuated between romanticised depictions of freedom and mystery and deeply negative stereotypes associating them with criminality and deceit.

This can be seen, for example, in Kant's writings on "Gypsies" [97,98]. Kant's works reflect and reinforce racialised Enlightenment thought, portraying the Roma as an inherently different and uncivilisable group. While some scholars debate whether Kant's race theory is central to his philosophy [99], he clearly engaged with contemporary racial theories, particularly those of Heinrich Grellmann, who popularised the idea of "Gypsies" as a biologically distinct people of Indian origin [100]. Kant's few but significant mentions of Roma suggest he saw them as proof of immutable racial traits, incapable of adapting to European civilisation. His arguments aligned with broader Enlightenment ideas of progress but simultaneously justified exclusionary policies by linking race to social immobility. This racialised framework contributed to long-lasting stereotypes and shaped later exclusionary policies against Roma populations in Europe [96].

In the 1800s, attitudes towards the Roma began to shift, albeit slightly. The Habsburg rulers attempted forced assimilation policies, particularly under Maria Theresa and Joseph II, banning their language, traditional clothing, and nomadic lifestyle in an effort to integrate them into the peasantry [95,101]. Despite these measures, Roma communities remained largely excluded from mainstream society, facing persistent discrimination, social segregation, and economic hardship. Their legal status varied by region, but in many areas, they were denied land ownership, access to guilds, and basic rights, reinforcing their position at the margins of society.

By the mid-1800s, an abolitionist movement advocating for the end of slavery emerged among intellectuals in the Wallachia and Moldova Principalities. During this period, the figure of the "Gypsy" became a common subject in newspaper articles, poetry, literature, and plays. The Romani people also became prominent in British literature as a symbol of the displaced "Other," representing moral, social, and racial standards that were seen as contrary to the national ideal. In some cases, such as in Conan Doyle's *Sherlock Holmes* series, the "Gypsies" maintain their anonymity as a semi-independent, mobile, and distinct culture within Britain's borders. However, in these stories, they are never depicted as responsible for criminal activity. Instead, it is often suspicion and prejudice that lead to their displacement, as they are forced to move to avoid causing local unrest [102]. The narrative thus suggests that the Roma do not pose a threat to domestic spaces, despite their non-conformity.

Generally, in nineteenth-century literature, the portrayal of the Roma reflects a constructed image rather than a realistic representation [103]. Unable to speak for themselves, the Roma were depicted according to the stereotypes and biases of the writer or speaker—at times romanticised and alluring, at other times criminal and filthy, shifting between these extremes without consistency [104,105]. This pattern aligns with what Said, a century later, would define as Orientalism: 'the written statement is a presence to the reader by virtue of its having excluded, displaced, made supererogatory any such real thing as "the Orient" [106].

Although never explicitly referring to the Roma, Karl Marx offers an opposing view of nomadic people in *Das Capital.* Marx explores vagabondage as a result of socioeconomic forces rather than racial or cultural traits [107]. He ties vagrancy to the shift from feudalism to capitalism, where displaced rural populations were forced into cities, becoming impoverished and often labelled as criminals. Marx analyses the English legislation on poverty and vagabondage, where the latter was de facto aimed at Roma [108], and shows how itinerants were punished for refusing to work in conditions that no longer existed. Marx argues that in late capitalism, unemployment and poverty are deliberate tools to suppress wages and sustain the system. In this context, Marx views vagabondage not as a moral or racial issue, but as a structural consequence of capitalist exploitation, where the ruling class imposes these conditions to protect their wealth.

Comparing to Kant, Marx offers a different view to understand the treatment of "gypsies" in early European society, focusing on socioeconomic structures, explaining how vagabondage and criminalisation were tied to the rise of capitalism, with the ruling class reinforcing exclusion to maximise profit. Kant, on the other hand, introduces a racialised view, linking the inability of "gypsies" to adapt to their supposed racial inferiority, which solidified their exclusion [96,107]. Modern debates on Roma, and their perceived status as "anti-social," still echo these early racial and social constructs, obscuring the socioeconomic factors at play [96].

By the late 19th and early 20th centuries, as the Roma began to integrate into the broader society, questions about their inclusion in social and economic life started to arise. Nationalist movements and racial theories further entrenched their marginalisation, with some governments implementing policies aimed at erasing Roma identity through forced assimilation, sterilisation programmes, and restrictions on their language and customs [42,101,109,110].

The period leading up to 1945 saw a dramatic worsening of their social status, culminating in the genocide of Roma people during the Holocaust, when they were systematically persecuted and murdered by Nazi Germany and its allies. For centuries, the Roma have faced systemic discrimination, exclusion, and criminalisation in Europe, with antiziganism still deeply rooted in society today. State policies have historically reinforced their marginalisation, portraying them as threats to public security while subjecting them to surveillance and persecution. The shift from viewing "Gypsies" as a social category to a racial one in police work played a key role in their persecution, particularly under Nazism [46].

Following the Second World War, the Roma continued to be broadly 'orientalised' across large parts of Europe, as this type of portrayal persisted in literature and popular narratives [111,112].

In contemporary times, the Roma continue to face different levels of marginalisation, discrimination, and socio-economic exclusion across Europe. While legal persecution and forced assimilation policies have largely been abandoned, Roma communities still experience widespread poverty, limited access to education, and barriers to employment [113–115]. In many countries, such as Italy, France, Serbia, and Slovakia, they remain segregated in informal settlements with inadequate infrastructure [116–119]. Institutional discrimination persists, particularly in law enforcement, housing, and healthcare, reinforcing cycles of social exclusion. Efforts to improve Roma integration have been made, especially through policies and advocacy initiatives, yet progress remains slow, and anti-Roma sentiment continues to be an issue in many European societies [120,121].

Outside Europe, India occasionally recognised Roma heritage as part of its diasporic connections, though these acknowledgments were symbolic rather than involving substantial cultural programs. One significant moment came in 2001, when the International Romani Union (IRU, organization active for the rights of the Romani people) declared the Roma to be a part of the Indian diaspora, and the Indian government supported this declaration [122]. The recognition underscored the Roma's historical ties to India, tracing their origins back to South Asia. In 2016, during the International Roma Conference held in New Delhi, the Indian Ministry of External Affairs reiterated this connection, with the Indian Minister Sushma Swaraj expressing a sense of cultural affinity with Roma [123].

Roma activism and cultural revival movements have gained momentum, promoting Roma identity and rights. However, prejudice and social stigma still shape public perceptions, and hate crimes against Roma individuals are reported in various regions. Despite legal protections, full social inclusion remains an ongoing challenge in the modern era.

1.1.3 The evolution and realities of Roma culture

Culture is not something fixed, inherited, or unchallengeable. On the contrary, it is constantly evolving, allowing for the self-expression of our self-realisation, re-inventing as well as representing and reproducing our ethnic identities. This is as true for the Roma as it is for any other group [124]. Culture, therefore, is not merely a collection of traditions, customs, and artistic expressions. Rather, it can be understood as a historically determined, class-based phenomenon, shaped by economic structures and political power [125]. Additionally, culture is often seen as the attainment of one's identity, a higher level of awareness that enables individuals to understand their historical significance, their role in life, and their rights and responsibilities [126].

To fully understand Roma culture, or the culture of any other community, an analytical approach that is both historical and materialist is essential. This framework moves beyond deterministic or essentialist views of culture by examining the role of class position, economic forces, social structures, and ideological influences in shaping cultural traditions. Roma have historically been marginalised, excluded from dominant economic structures, and subjected to systemic discrimination. These conditions have strongly influenced Roma cultural patterns, yet their culture is not merely a passive product of oppression. Rather, it reflects a dialectical process of adaptation, resistance, and internal continuity.

Practices such as nomadism, oral storytelling, and music are long-standing aspects of Roma culture. While their evolution may have been influenced by historical conditions of exclusion, economic instability, and restrictions on land ownership and formal employment, they are also expressions of Roma identity, creativity, and mobility. Likewise, the persistence of extended kinship networks and strong communal ties can be understood not only as strategies for navigating socio-economic precarity but also as integral aspects of Roma social organisation and cultural continuity. At the same time, some practices often attributed to Roma culture—such as begging—are not inherent cultural traits but rather socio-economic strategies that have emerged under conditions of structural exclusion. The institutional framing of such practices, as seen in the Italian state [127], reflects how dominant societies interpret and reinforce stereotypes rather than addressing the material causes of economic marginalisation.

Roma culture is shaped by both external pressures—such as historical oppression and economic exclusion—and internal agency, as illustrated by the following points:

- Migrations, adaptation, discrimination, and resistance to assimilation have played a central role in shaping Roma culture.
- Oral traditions, craftsmanship, and communal living reflect both internal values and historical adaptations to exclusion.
- Music, dance, and storytelling frequently express themes of struggle, exile, and survival, directly linked to their socio-economic position.

From this perspective, Roma culture is not fundamentally different or isolated from broader socio-economic developments. Rather, dominant ideologies shape how Roma communities are perceived. In both feudal and capitalist societies, Roma people have often been stigmatised as outsiders, reinforcing stereotypes that justify their exclusion from economic participation. Ruling classes have historically portrayed them as "criminal," "lazy," or "exotic," which served to maintain their marginalisation and keep them in precarious conditions. Meanwhile, Roma culture was often commodified—seen in romanticised portrayals in art—while Roma people remained structurally disadvantaged.

Despite centuries of coexistence with various populations, Roma communities have maintained a distinct cultural identity, not through rigid isolation but through adaptation and strategic continuity. Their determination to preserve their heritage, rather than simply resisting assimilation, has been integral to their survival as a separate population [128], reflecting both resilience and historical necessity rather than an inherent resistance to change.

1.1.3.1 Traditional Roma culture

Traditional Roma culture has been shaped by a nomadic lifestyle, strong communal bonds, oral traditions, and distinct economic roles. Mobility was a defining feature of Roma life, with communities travelling in family-based groups not only out of cultural choice but also due to legal restrictions and persecution. This constant movement allowed Roma communities to remain flexible and adapt to changing circumstances, but it also imposed practical constraints on their social and economic structures.

The family and community were central to Roma society. While each community had a chief holding authority, decision-making was often collective, reflecting a communal approach to governance and social order [17,129]. Gender roles were traditionally defined, with men typically engaged in external economic activities, while women managed the domestic sphere, including child-rearing and specific trades such as fortune-telling and healing

[130–132]. These roles were shaped by necessity, helping to maintain social cohesion within the community.

Language and oral traditions played a crucial role in maintaining Roma identity, particularly in contexts where formal education and written records were limited. The predominantly spoken Romani language, at least within the community, and the practice of passing down myths, proverbs, and historical narratives through storytelling ensured that cultural knowledge was preserved and transmitted across generations. Religion and spirituality among the Roma have also been influenced by the dominant faiths of the regions they inhabit, whether Christianity, Islam, or other traditions. However, Roma communities have integrated their own folk beliefs—such as reliance on luck, protective charms, and healing rituals—into these religious practices, reinforcing community cohesion and offering an alternative framework for understanding their social reality. Cultural expression through music and dance always was one of the most recognisable aspects of Roma life. Roma musicians have left a significant mark on genres such as Flamenco, Balkan brass, and Jazz Manouche, dance and song have long served as vital means of communicating social experience and historical memory.

Economic survival for many Roma communities depended on specialised trades. The reliance on adaptable forms of work—ranging from metalworking and animal trading to providing entertainment services, was something the Roma were recognised for [28,88].

Roma communities adhered to their own system of customs and unwritten laws, which governed social behaviour, honour codes, and conflict resolution [128,133]. Loyalty to family and respect for elders were central values, and traditional justice was often administered within the community rather than through external legal systems. Clothing also carried cultural significance, with certain items, such as long coats and one-piece dresses, being considered shameful due to their denial of the body's divisions [134].

1.1.3.2 Modern Roma culture

Modern Roma communities, particularly in Europe where they are predominantly settled, have developed a culture that blends traditional Roma elements with the influences of the dominant societies they now live in. While Roma cultural aspects vary significantly by country and community, several common traits remain central to the community's way of life. Family and community are foundational, with extended kinship networks playing a crucial role in shaping daily life. Intergenerational ties are strong, and respect for elders continues to be deeply ingrained, reinforcing the collectivist values that prioritise family over individual decisions [133]. In many communities, traditional gender roles persist, reflecting long-established social norms (for definitions, see section 1.4.3 Social norms and practices) [132,135]. Reports indicate that Roma girls in Western Balkan communities (Bosnia, Serbia, Albania) are less likely to attend school, marry at a young age, refrain from working, and have low aspirations, with little evidence of change over the past decade [135]. However, in Spanish Roma communities, significant positive changes have been reported over the last decades, particularly with more Roma women actively seeking employment [136,137].

At the same time, language and identity remain important markers of Roma culture. Typically, Roma speak Romani only with other Roma they have some kind of relation to, while using the other language for any other interaction [6]. Therefore, many continue to speak Romani alongside the local language, even as younger generations are increasingly influenced by the dominant linguistic and cultural frameworks of their societies. This dual-language use reflects the ongoing negotiation between maintaining cultural identity and integrating into wider society. In “fringe” communities, so Iberian, British, Scandinavian (except Finland), the local Romani dialects are extinct and only exist in sparse loanwords and expressions [138–140]. Religion, too, plays a significant role in the lives of many Roma. This particular aspect has not changed over the centuries, with most Roma still adopting the dominant religion of their country and retaining elements of folk beliefs and taboos that are integral to their spiritual identity [133]. Cultural expressions such as music and dance remain deeply rooted in Roma identity, providing a means of both personal expression and communal solidarity. Similarly, oral traditions and storytelling continue to play an important role in cultural transmission, though their prominence is waning due to increasing literacy and the growth of digital media.

Economically, Roma have adapted to modern realities, with many working across a range of professions from manual labour to entrepreneurship and academia. However, historical exclusion from mainstream job markets has often led to a reliance on self-employment, which remains common within the community [141]. Despite progress, challenges remain. Roma communities, especially in urban areas, still face segregation and high dropout rates in education [142–144], as urbanisation has contributed to the growth of segregated Roma neighbourhoods [116]. However, there is also a growing Roma presence in more central urban areas, and school segregation is decreasing, indicating a shift in social and geographical dynamics as a result of policies specifically designed to reduce segregation [143,145–147].

Modern studies suggest that much of the common knowledge about Roma culture is outdated, with older research often based on interviews with individuals from more conservative Roma communities, highlighting the importance of recognising the diverse and evolving nature of Roma identity [133]. Certain core traditions and beliefs persist, including the increased status and respect associated with older age [83], the involvement of elders for conflict resolution [148], and the determination not to assimilate into dominant society [148]. In daily life, women are primarily responsible for domestic duties and generally do not work [149], while non-parental figures often participate in child-rearing [124]. However, these traditions coexist along with ongoing change and adaptation [133].

1.1.4 Roma language and literature

Romani is an Indo-Aryan language estimated to be spoken by at least 4 million people, with official recognition in at least 17 countries, primarily in Eastern Europe. It is subclassified into para-Romani languages, which are further divided into seven distinct groups [138]. Within this classification, each group encompasses several subgroups of Romani dialects, corresponding to different Roma communities. The seven major groups are: Balkan, Vlax, Central, Northeastern, Northwestern, British, and Iberian. Despite their diversity, all of these dialects share a common origin, shaped by linguistic influences encountered along the Roma diaspora route—most notably, the Greek language during the Byzantine period [5]. While the exact numbers of speakers remain uncertain, some groups, such as Vlax and Balkan Romani, retain a strong knowledge and usage of the language [5,150–152]. In contrast, dialects from regions like the Iberian Peninsula have largely disappeared due to centuries of assimilation, surviving mostly through a few expressions [140]. Linguistic research has played a crucial role in tracing the history of the Roma. Linguistic research has long supported the theory that the original homeland of the Roma people was Northern India, as the Romani language is classified within the Indo-Aryan branch of the Indo-European language family [21,153]. Genetic studies further corroborate this, showing that contemporary Punjabi populations exhibit the closest genetic affinity to the Roma [154,155].

Therefore, the study of the Romani language is significant not only for its preservation and perpetuation but also for the socio-historical context it provides. To this end, in recent history, there have been several attempts to safeguard Romani languages and document Roma history and culture in written form, moving beyond the traditional reliance on oral transmission.

The first initiatives to publish in Romani emerged in the Union of Soviet Socialist Republics (USSR) and the Socialist Federal Republic of Yugoslavia (SFRY) [156,157]. As part of a broader policy of promoting minority languages, the first textbooks, newspapers, and literary works in Romani were published in the USSR in the 1920s. The journal *Nevo Drom* ('New Path') became a significant platform for Romani literature and cultural expression in the 1930s [158]. In Yugoslavia, Romani publications gained traction within the framework of socialist multiculturalism, which aimed at supporting minority cultures as part of a collective national identity. From the late 1960s, the term Roma was officially adopted, and in the 1970s, writers such as Slobodan Berberski and Rajko Đurić emerged as some of the most prominent Romani authors, contributing to the development of a distinct Romani literary tradition [159]. Following the pioneering efforts in the USSR and Yugoslavia, several decades passed before other countries began supporting Romani language and culture.

Starting in the late 20th century, nations like Sweden and Finland recognised Romani as a minority language, launching initiatives to preserve it through education, publications, and media [160]. These efforts included the publication of books and the production of radio programs in Romani. Similarly, Spain saw a growing interest in Romani cultural contributions, particularly through literature and music [161], as part of broader social inclusion policies. In the United Kingdom, organizations of citizens and academic institutions began promoting Romani studies and media, providing platforms for Romani writers and artists [162]. These developments marked a shift from state-led initiatives in Eastern European countries to a mix of governmental and community-driven efforts in Western Europe, reflecting an evolving recognition of Romani identity and heritage on an international scale.

Today, Romani language is used in magazines, websites, and other media—sometimes in bilingual formats or in a lingua franca such as English, primarily targeting Roma communities.

In the academic sphere, Romani studies have gained recognition through dedicated publications:

- The Journal of the Gypsy Lore Society, first established in 1888, continued until 1982. It resumed publication in 1991 and was renamed Romani Studies in 2000. It now serves as the official journal of the Gypsy Lore Society, an international academic organization dedicated to Roma research and cultural promotion.
- The International Journal of Roma Studies (IJRS) was launched in 2019, focusing on interdisciplinary research that examines discrimination, exclusion, and the empowerment of ethnic minorities, with a particular emphasis on Roma communities.

The impact of globalisation has significantly affected the preservation of Romani languages and dialects, particularly in Western Europe. As Roma communities increasingly adopt dominant state languages for reasons of education, employment, and social integration, traditional Romani dialects are gradually fading. Younger generations are less likely to speak Romani, leading to a decline in linguistic diversity within the Roma community. For this reason, Romani is classified as Vulnerable by the UNESCO [163]. While efforts to preserve and revitalize the language continue, the widespread use of state languages and the influence of global media and technology pose ongoing challenges to the survival of Romani dialects. This trend is part of a broader global phenomenon in which minority languages struggle to survive in an increasingly interconnected world [164,165].

1.1.5 The Roma groups

The Roma are divided into various groups, each with some unique traditions, dialects, and social structures [138,166–168]. These internal divisions likely emerged due to a combination of endogamy, social pressures, and migration patterns, all of which contributed to the genetic diversity observed among Roma populations. Traditionally, Roma groups were organised along occupational lines. For example, the Kalderash derive their name from the Romanian *căldăraş* (coppersmith), itself from the Latin *caldāria*, referring to their trade as metalworkers or tinkers [128]. Similarly, the Boyash (also known as Rudari) are thought to originate from the term *baie* in Old Slavonic, referencing their historical association with mining [169]. Many of these groups are further subdivided by profession or region of origin. For instance, the Ursari (bear trainers) are associated with Bulgaria, while the Volokhy refer to a group of Roma in Ukraine originally from Wallachia. While most scholars—including geneticists—use socio-linguistic classifications to define Roma groups by country (e.g. Spanish Roma, Greek Roma), macro-region (e.g. Balkan Roma, Carpathian Roma, Iberian Roma), or linguistic subgroup (e.g. Romungro, Vlax, Rudari, and Calé), some researchers have proposed an alternative classification into five major groups: Roma, Sinti, Kale (or Calé), Manouches, and Romanichals **[170]**. Each of these groups comprises numerous subgroups and communities, often shaped by historical migration routes and local cultural influences. The Roma are primarily concentrated in Central and Eastern Europe, especially the Balkans. The Sinti, whose name reflects their origin in the Sindh region of South Asia, are found mainly in Northern Italy, France, and other parts of Western Europe, and speak a German-influenced dialect of Romani. The Kale, deriving their name from the Hindi word for "black," are present in Wales, Finland, Spain, and Portugal. The Manouches, mostly in France and Northern Italy, derive their name from the Sanskrit *manus*, meaning "human." Lastly, the Romanichals, primarily based in the UK, whose name is derived from Romani words meaning "Roma children" or "young ones." [170]

Over the centuries, Roma communities developed divergent customs and dialects depending on their regional settlements and the degree of assimilation into broader societies. In some cases, this assimilation led to the loss of the Romani language (as in the Iberian Peninsula and in England), although a distinct Roma identity was retained—albeit with significant differences from other Roma groups. Self-identification also varies: some Roma identify solely with the nation in which they reside (e.g., Romanian, Spanish), others as Roma, and some as both [166,168,169]. These layers of classification—social, linguistic, regional, and historical—reflect the diversity within Roma groups and point to the importance of considering each group within its specific context.

1.1.5.1 The Iberian Roma

The Iberian Roma, also known as Calé, are a transnational Roma community residing in modern Spain and Portugal. The earliest studies on the Iberian Roma date back to the first half of the 18th century, though they primarily focused on language and general culture [171,172]. Historical records provide solid evidence regarding their presence in the Iberian Peninsula. At the time of their arrival, the peninsula was divided into several realms: the Kingdom of Aragon controlled the northeast; the Kingdom of Navarre ruled a small part of the north; Castile governed the central territories from north to south; Granada was under the Caliphate; and Portugal already existed within its modern borders. In this fragmented political landscape, the Roma had to navigate multiple kingdoms, seeking permission from various rulers to cross their lands.

The earliest documented arrival of the Roma in the Iberian Peninsula dates to 1425 in Zaragoza, where the King of Aragon granted them permission to stay [86]. They were later documented in Navarre in 1435, allegedly on a pilgrimage to Santiago de Compostela [173]; in Jaén in 1462 [174]; in Burgos in 1475, when another group on pilgrimage to Compostela was recorded in what was then the capital of the Kingdom of Castile [175]; and in Medina del Campo in 1480, again with pilgrimage cited as their purpose [175]. In Portugal, the first literary reference to the Roma appears in Gil Vicente's *Farsa das*

Ciganas, performed in Évora in 1521 for Dom João III. Notably, there are no records of the Roma entering the Emirate of Granada, the last independent Muslim state in Western Europe [176].

There is speculation that some Roma may have arrived via an alternative route through the Arabian Peninsula, North Africa, and the Strait of Gibraltar [8,177], although no solid evidence supports this hypothesis. Their arrival in Iberia coincided with a period of significant political turmoil, as Christian kingdoms overthrew Islamic rule. In the late 15th and early 16th centuries, these kingdoms persecuted, expelled, and forcibly converted Muslims and Jews [178–181]. Before this period, Christians, Jews, and Muslims had coexisted for nearly a century. During these turbulent times, there may have been instances of contacts between the Roma and these populations, who were eventually expelled, while the Roma were forcibly settled and assimilated [86,109,182].

Institutional persecution of the Roma in Castile began with the *Pragmática* of the Catholic Monarchs in 1449, marking the start of ongoing efforts to suppress them. In Portugal, Dom João III issued a law in 1526 prohibiting the entry of Roma and ordering the expulsion of those already present [183]. However, this decree was likely not strictly enforced, as similar laws continued to be passed over the following years while Roma communities persisted [184]. In Navarre, persecution was documented as early as 1549, when the Tudela *Cortes* decreed that any Roma found within the region were to be whipped and expelled [185]. Such measures reflect the widespread discrimination and systemic exclusion the Roma faced in Iberian societies.

By the 16th century, the Spanish and the Portuguese Empires had enacted numerous anti-Roma laws [182,186–188], severely restricting the freedoms of Roma. During this period, with the first recorded presence of Roma in Brazil dating back to 1574 [189], the Portuguese Empire began forcibly deporting Roma to its colonies in the Americas, while some also migrated voluntarily in search of opportunities [44,190]. The relationship between the Roma and local authorities was consistently strained, as the Roma maintained a distinct cultural identity and were often seen as disruptive to society. They were frequently stereotyped as "dirty," "cheats," and "thieves," with little recognition of their individuality. Instead, the Roma were often viewed as a collective group, defined by negative stereotypes of barbarism, dishonesty, and immorality [44]. This atmosphere of persecution in Spain culminated in the *Gran Redada* (Great Round-up) on 30 July 1749, during which thousands of Roma were arrested and imprisoned [109].

The 19th and 20th centuries saw increasingly intense assimilation policies, which also led to the gradual disappearance of the Iberian Roma dialect [109]. During all this time, significant internal migrations within the Iberian Peninsula caused the displacement of many people, likely including the Roma [191,192]. In this period, Spain kept maintaining the same policies of strict control towards the Roma [177]. In 19th-century Portugal, following the end of absolutism and the rise of liberalism, Roma people were no longer targeted for extinction but were instead systematically marginalised. Authorities reinforced stereotypes of Roma as "nomads"—a condition they themselves perpetuated by preventing settlement and seizing property—then used this label to justify further discrimination. These policies aimed to keep Roma impoverished and culturally erased. Despite these efforts, Roma communities maintained high birth rates, cultural resilience, and continued attempts to settle and integrate into urban and rural life [193].

In the 1930s, under Francisco Franco's fascist regime in Spain, state policies towards the Roma population were shaped both by earlier imperial-era decrees and by the influence of Nazi Germany's concept of "racial hygiene." [194] For decades, Roma were widely regarded as inferior by many academics, medical professionals, and legal experts within Francoist Spain. This perception was reflected in the legal framework of the time, which included several explicitly anti-Roma measures—most notably, the obligation to speak only Spanish; the drafting of the "Social Danger Law"; and the use of the Guardia Civil as an enforcement arm of the dictatorship's policies towards the Roma people [177].

Similarly, the fascist regime of Salazar in Portugal employed comparable identity-based mechanisms of exclusion. The construction of a collective 'bad object'—whether external, such as the Spanish, or internal, such as the so-called 'bad Portuguese'—was central to the identity politics of the Estado Novo, which ruled Portugal from 1928 to 1974 [195]. Portuguese Roma were often conflated with these negatively perceived groups, with scientists of the regime using terms like 'nomadic' and 'primitive' to characterise not only Roma, but also other marginalised Portuguese populations. This homogenising discourse served to reinforce a hegemonic sense of superiority among the majority [195].

1.1.5.2 The Iberian Roma today

The Roma can be found in most parts of the world, including Europe, the Americas, Asia, Oceania, and North Africa. Despite their long presence in Europe, the Roma in Iberia continue to face varying degrees of social challenges, even into the 20th and 21st centuries, although there have been significant improvements in their conditions. In Europe, particularly in Western Europe, the largest Roma community resides in the Iberian Peninsula, comprising Spanish and Portuguese Roma. This community numbers between 1 and 2 million people, a minority of which (around 60,000 as first language) speaks the Caló, a para-Romani language that evolved in this region and is now nearly extinct [140,196]. The Iberian Roma are predominantly concentrated in the Andalusian region [197,198], with a significant presence in Catalonia and a smaller presence in Portugal. The Iberian Roma nowadays are involved in every aspect of society but still retain a certain level of closure. This can also be attributed to the fact that Romani culture has historically been relatively closed and inaccessible to outsiders [20,197,198]. Despite being much more integrated than in the past, they continue to suffer from socioeconomic inequities compared to non-Roma individuals [199–201]. This is evident in school issues, but also in health and economic disparities. In fact, it appears that in some regions Roma have lower access to education or other public services, or they are discriminated against in these sectors [200,202]. Due to higher difficulties in finding employment, many Iberian Roma are self-employed [203], with some researchers suggesting the presence of an ethnic income gap in Spain [204].

1.1.5.3 The Czech Roma

The Czech Roma have been the focus of numerous medical studies, yet remain understudied in the fields of genetics, history, and sociology, with most existing research concentrating on their situation in the 20th century. The Roma reached the regions of Moravia and Bohemia in the early 15th century [24], establishing a presence in what is now the Czech Republic for over five centuries. A large part of the Roma in this region call themselves Servika or Serviko [205], as many Czech Roma trace their origins to Slovakia following World War II, due to the extermination of the Roma population in Czechia by the Nazis [206,207]. Other groups in the Czech Republic include the Vlax and Romungro, the latter being part of the broader Carpathian Roma. As the Roma populations in both the Czech Republic and Slovakia are closely related, particularly given that the two countries were united until the 1990s, it is crucial to consider the history of the Slovak Roma to better understand the context of the Czech Roma.

The Roma began settling in present-day Slovakia as early as the 13th century, with the earliest documented mention of Roma in Slovakia dating back to 1322 in the Spiš region [208]. Their migration into Central Europe is tied to the Ottoman expansion, which pushed populations out of the Balkans [208]. While a charter issued by the Holy Roman Emperor Sigismund of Luxembourg in 1423 granted Roma certain rights, the treatment of Roma in Silesia, Bohemia, and Moravia (now part of the Czech Republic and Slovakia), was markedly different from that in Slovakia. The Holy Roman Empire developed a history of discriminatory and violent policies toward the Roma, whereas Slovakia, part of Hungary (which was under Ottoman influence), exhibited more flexibility in its treatment of Roma [209]. This was especially evident in Slovakia's guild system, which facilitated the integration of Roma into crafts during the 15th–17th centuries [208]. As a result, most Roma in Slovakia became settled, while in Czech lands, they remained nomadic, often facing precarious economic conditions [209]. This lasted until the 18th century, when the Habsburg policied towards the Roma shifted from persecution to forced assimilation [209]. By the 18th century, Roma communities had become widespread in Slovakia, particularly in southern regions. Administrative inquiries by Maria Theresa and Joseph II documented significant concentrations of Roma in areas like Bratislava, Nitra, Tekov, and Gemer [208]. In the 19th century, Roma began migrating from Wallachia and Moldavia in response to legal reforms, part of which established in Slovakia. As noted earlier, most of the Czech Roma—living in areas of Czechia that came under direct control of the Third Reich following the annexation of the Sudetenland—were completely exterminated, with only a handful of families surviving [209].

The post-WWII period saw a marked increase in the Roma population, with significant concentrations in cities like Košice, Levoča, Prešov, and Bratislava [208]. In 1945, with the establishment of Czechoslovakia, the Roma were able to move more freely between the territories, which facilitated the spread of Slovak Roma into Czech regions. A 1955 survey identified over a thousand segregated Roma settlements in Slovakia, where a majority of the population lived. In the late 1950s, the Czechoslovakian state began efforts to end nomadism, which impacted thousands of people, pushing for greater integration into settled communities [208]. By the 1960s, Slovakia had numerous Roma settlements spread across the eastern, western, and central regions, moving in particular to the areas previously occupied by ethnic Germans. In 1980, there were about 300,000 Roma in Czechoslovakia (2% of the total population) [210]. Over the 20th century, the population continued to grow and according to estimates, in 2012, over 353,000 Roma lived in Slovakia, constituting 6.5% of the total population [208].

In the Czech Republic, official census data identifies only 4,458 individuals who exclusively identify as Roma, with an additional 17,233 identifying as both Roma and Czech [211]. However, unofficial estimates from the 1990s suggested that Roma made up at least 2–3% of the country's population of 10 million [84]. Many Roma choose not to disclose their ethnicity due to fears of discrimination [212], which means the actual population size remains uncertain.

Roma communities in the Czech Republic continue to face a range of structural challenges, particularly in the areas of education, health, and social inclusion. A longstanding and issue is the segregation of Roma children in education. Many are placed in special schools or segregated "Roma classes" within mainstream institutions, significantly limiting their opportunities for academic achievement and social mobility [213]. Although racially motivated violence is relatively rare and generally addressed by law enforcement, systemic educational discrimination remains a significant concern[213]. In the realm of healthcare, persistent disparities between Roma and the majority population remain evident. Over the past two decades, recent health strategies increasingly emphasize personal responsibility, portraying Roma as primarily accountable for their own health outcomes. This individualised framing often overlooks structural determinants such as discrimination, poor housing conditions, and limited access to healthcare services [214].

1.2 Genetics of the Roma

The Roma have been subject of scientific interest for the past decades. Genetic studies began in the 1960s with early research based on blood groups, and in recent years, there has been a growing focus on studying this population. The genetics of the Roma have been relatively well studied, although research remains incomplete, with several aspects still unresolved and many Roma groups across Europe under-represented in the literature. Many studies have relied on available data rather than structured methodologies, partly due to the complexities of human genetic research, which requires extensive ethical approvals and lengthy procedures, as well as challenges in engaging with the interested communities. As a result, research on the Roma spans multiple countries but varies in technology, methodology, and sampling strategies. This lack of consistency has made it challenging to compare findings, with data often being sparse and sample sizes relatively small. Despite this, researchers have made significant breakthroughs in understanding the genetic history, population structure, and health-related genetics of the Roma.

1.2.1 The development of genetic studies on the Roma

Genetics is a biological discipline that is relatively old, with its foundations rooted in the 19th-century studies of Gregor Mendel, who is considered the father of the field. Since his discovery of the basic laws of inheritance — now known as Mendelian inheritance — around 160 years ago [215], numerous groundbreaking discoveries have shaped the field. A major turning point came in 1901 with the discovery of the ABO blood group system by Karl Landsteiner, the first reliable genetic markers in humans, which will be used in population genetic studies until the 1990s. In the early 20th century, the concept of the "gene" was formalised by Wilhelm Johannsen (1909), and Thomas Hunt Morgan's experiments proved that genes are located on chromosomes (1910s). Blood group typing quickly became a valuable tool in studying genetic variation across populations.

Among the early applications of this approach were studies on the Roma people, using blood group data to investigate aspects of their origins and genetic affinity to non-Roma. Although the very first scientific mentions of the Roma in this context are unclear, they were briefly referenced in a study from 1928 [216] and again in a 1933 forensic study [217]. However, the first proper genetic studies appeared in the 1940s, describing the frequency of blood groups of Roma communities in Hungary [218–220]. It was observed that Hungarian Roma retained gene frequencies and blood group distributions similar to those of the Rajput people in Northern India, and that they were "strikingly different from the Hungarian people generally" [218]. The Roma of the Balkans, in comparison to the more admixed English Roma, were defined as genetically "pure" [218], a term then used to denote lack of admixture. A 1948 study similarly referred to the Roma as an "unmixed group." [221] While these early studies were framed by the racialised perspectives common in Western Europe at the time, and though limited in scope, they correctly identified genetic similarities between the Roma and Northern Indian populations.

More advanced studies began to appear during the 1960s and 1970s. Alongside ABO blood group frequencies, researchers started examining additional blood group systems, protein markers (such as Haptoglobin), and enzymatic markers (like Phosphoglucomutase), supported by more structured statistical analyses. For instance, studies examined blood group distributions in Swedish Roma [222], and blood and protein markers in Roma from Slovakia [223,224] and Slovenia [225]. Notably, these early studies often included relatively large sample sizes, ranging from 160 [222] to nearly 400 individuals [223], as it was easier at the time to conduct large-scale genetic studies using inexpensive, simpler methods like blood group typing and protein electrophoresis, which produced rapid results compared to today's resource-intensive genomic sequencing. These studies generally found that the Roma were genetically distinct from other European populations, and that certain Roma groups (e.g. the Swedish and Welsh Roma) were slightly different from others, likely due to genetic drift and founder effects [222,224]. Evidence of genetic substructure within individual countries, such as Slovakia, was also observed [223].

In the early 1990s, researchers hypothesised a significant genetic distinction between Eastern and Western European Roma: the western groups were more genetically homogeneous and admixed with local European populations, while the eastern groups retained stronger affinities with Indian populations—possibly due to greater isolation, genetic drift, and internal migration among subgroups [226].

Therefore, up until the 1990s, early Roma genetic research largely focused on three questions: their South Asian origins, their genetic differentiation from non-Roma Europeans, and the differences among Roma groups themselves. The technical limitations of the time likely contributed to this narrow scope. A shift began in the 1990s, notably with the discovery of the Y-sarcoglycan gene, which had particular relevance to Roma genetic studies. The Y-sarcoglycan gene, part of the sarcoglycan complex, is essential for maintaining muscle cell membrane integrity. A founder mutation was identified in Roma groups, linked to limb-girdle muscular dystrophy (LGMD), a condition disproportionately affecting the Roma [227,228]. The mutation was estimated to have originated between 60 and 200 generations ago, possibly even before the Roma migrated out of India [227]. One medical genetics study also reported that consanguinity rates among Spanish Roma were roughly 20 times higher than in non-Roma populations—an early genetic indication of endogamy and inbreeding [229]. These early studies, however, often underestimated admixture in Roma populations—likely due to both technical constraints and cultural biases of the time.

In the beginning of the 2000s, studies based on autosomal and uniparental markers, particularly microsatellites (STRs), began to emerge [230,231]. Major findings included evidence that socially distinct Roma groups shared common origins, as revealed by unique mutational events and the first identification of founder lineages. A sex-biased pattern of admixture was observed: mitochondrial DNA (mtDNA) showed more non-Asian haplogroups, while Y-chromosome (Y-DNA) lineages retained more Asian ancestry. The MSY1 Y-lineage, known for its high mutation rate, was estimated to be about 400 years old, indicating more recent differentiation [230]. These findings confirmed that the Roma were genetically heterogeneous [231].

Such studies laid the foundation for Roma genetic research over the next two decades, defining key questions: Do all Roma share a common origin? Did they experience a founder effect, and when? How internally stratified are Roma groups? What are the timelines of their migrations? Is there evidence of sex-biased admixture?

Importantly, a shift in tone and terminology also began to emerge in the literature. Terms like "Gypsies" were still used, but "Roma" increasingly appeared, along with growing awareness of the ethical dimensions of studying marginalised populations. More attention was paid to the discrimination faced by Roma, and researchers began recommending more ethical and respectful approaches to studying their genetics.

By the late 2000s, Single Nucleotide Polymorphism (SNP)-based research had begun, with the first study on Y-SNPs in Portuguese Roma published in 2008 [232], followed by the first application of genome-wide array data on European Roma in 2012 [155]. From that point onward, some of the most transformative discoveries in Roma genetics were made, fuelled by the sequencing of the human genome and the computational power of modern bioinformatics—ushering in a new era of population genomics.

1.2.2 Origins of the Roma from a genetic perspective

TThe origins of the Roma have long been traced to South Asia, particularly the Punjab and Kashmir regions. While historical and linguistic evidence had already pointed in this direction, by the 1970s it was generally accepted among biologists that the Roma originally came from South Asia. This conclusion was strongly supported by the distribution of the haptoglobin (Hp) gene, which showed strikingly similar frequencies between all the (then-studied) European Roma groups and Indian populations [233]. Additionally, other genetic markers such as the ABO, Rhesus, and Duffy blood group systems in Roma displayed patterns very similar to those observed in populations from Punjab and western Pakistan [233]. Studies on Slovakian Roma—among the most frequently researched Roma populations—found that frequencies of phosphoglucomutase and other isoenzymes aligned with the broadly accepted hypothesis of Indian ancestry for European Roma [234].

In the 1980s, this evidence was further reinforced through newer types of markers, such as allotypic variants of immunoglobulin heavy chains, in studies focused on Hungarian Roma [235], another well-studied group. In 1990, a large-scale study combined existing data with new samples from Roma communities across nine countries, along with individuals from northwestern India, reinforced the close genetic affinities between these populations [226]. Similarly, a study of Human Leukocyte Antigen (HLA) in Spanish Roma populations provided further confirmation of South Asian origins [236].

A key development following the era of blood group and protein marker analysis was the identification of founder mutations. These included a mutation in the Y-sarcoglycan gene associated with limb-girdle muscular dystrophy (LGMD), and the 1267delG mutation causing congenital myasthenic syndromes (CMS), shared by Roma and Indian/Pakistani populations [227,228,237].

More recent genomic research using autosomal and uniparental DNA has further corroborated these conclusions, identifying a substantial Punjabi ancestral component in the Roma population [154,155,238,239]. In particular, Y-chromosomal haplogroup H1a1a-M82, which is highly prevalent among South Asian populations—particularly among Dravidian and Austroasiatic speakers—has been found at high frequencies among European Roma. However, it exhibits reduced diversity in Roma, suggesting a relatively recent founder event and migration from India [240]. A significant portion of Roma paternal ancestry has been linked specifically to Indian scheduled caste and tribal populations. Shared haplotypes between Roma and South Indian groups such as the Lingayat, Vokkaliga, and Soliga indicate a moderate contribution of South India to proto-Romani ancestry [241]. On the maternal side, key mtDNA haplogroups supporting Indian ancestry include M5a1, M35, and M35b. These have been identified in numerous Roma groups across Europe and serve as markers of their shared South Asian heritage [242,243]. The overrepresentation of Indian-origin haplogroups (e.g., M5a1, M35, H1a-M82) in both mitochondrial and Y-chromosomal studies reaffirms the Roma's South Asian roots.

Concerning the date of the Roma's departure from South Asia, haplotype coalescence estimates suggest that the emergence of the proto-Roma population occurred around 800–900 years ago, consistent with historical records linking their migration to the early Islamic invasions of India during the 11th century [154,239,244]. Further genetic evidence suggests a divergence into distinct migratory branches approximately 500–600 years ago, aligning with the period of Roma enslavement and dispersal in Romania [245]. Studies of uniparental DNA—specifically mitochondrial and Y-chromosome lineages—reinforce these findings, identifying South Asian founder lineages that provide compelling evidence for the Roma's ancestral origins [238,243,246,247]. These genetic markers further support the hypothesis of a migration from India to Europe during the early second millennium CE [238,240,248,249]. The autosomal markers in genome-wide array data and whole-genome sequences support the hypothesis, as Roma groups across Europe have been found to possess at least 20% South Asian ancestral components [239,245]. Given the substantial body of genetic evidence, it is now widely accepted that the origins of the Roma trace back to the Punjab region, which spans parts of present-day India and Pakistan.

1.2.3 Roma genetic structure and regional differentiation

1.2.3.1 Autosomal genetic structure

The Roma exhibit a distinct genetic profile among European populations. Aside from recent South Asian immigrants, they are the only group in Europe with a significant proportion of South Asian ancestry—accounting for approximately one-third of their genetic makeup in most groups [155,245,250]. The remaining two-thirds typically comprise West Eurasian ancestry, though the proportions of European and West Asian vary regionally. For example, Romanian Roma show higher genetic affinity with Romanian populations, while Spanish Roma display signatures of Spanish ancestry [154,155]. All of them, generally show components that closely resemble Iranians and other West Asian populations such as Armenian, Lebanese or Turkish [251,252]. These patterns reflect the differential admixture, the complex migration routes of the Roma and their historical interactions with local populations along the way.

Contrary to early assumptions of complete genetic isolation, numerous studies have shown that the Roma are indeed admixed with surrounding European populations, as well as with groups encountered during their diaspora [155,238,239,251]. However, the nature and extent of this admixture have been shaped by historical circumstances and enduring social barriers. Like other historically marginalised communities—such as the Ashkenazi Jews—the Roma were often perceived as genetically "unmixed," "pure," and socially isolated, and thus different from other European populations.

Early genetic research, which relied on simple markers such as ABO blood groups, lacked the resolution to detect nuanced patterns of admixture. Nevertheless, these early did hint at internal population structure among Roma groups. For instance, certain isolated Roma communities, such as those in Sweden or Wales, exhibited distinct blood group frequencies, suggesting heterogeneity within the broader Roma population [226,253]. Notably, an East–West gradient was observed: Roma groups in Western Europe appeared more admixed, with blood group frequencies more closely resembling those of local European populations, while Eastern groups retained more genetic signatures of Indian ancestry [226]. More recent genetic analyses using high-resolution markers have confirmed these initial observations. While admixture with non-Roma populations has occurred to varying degrees, the Roma have largely maintained a distinct genetic identity. This persistence is attributed in large part to their strong social organisation and endogamous marriage practices [254,255].

The regional differentiation observed among Roma populations can be partially explained by the varying historical, political, and social contexts they experienced across Europe. Geographic and cultural separation between Roma subgroups has contributed to the development of genetic substructure within the broader population. Early studies proposed two possible scenarios to account for this: either the Roma originated from a genetically sub-structured ancestral population that preserved endogamous practices, or they descended from a small, genetically homogeneous group that diverged over time due to genetic drift [254,255]. Recent research supports the former explanation, with increasing evidence of substructure within the Roma population and indications of further microgeographical differentiation [239,238,256,245]. The impact of a bottleneck is evident in the reduced genetic diversity observed in Roma compared to non-Roma European populations, as well as the high degree of homozygosity—both reflecting historical inbreeding and long-term isolation—which have further contributed to regional genetic differentiation [245,257,258].

1.2.3.2 Uniparental genetic structure

The mitochondrial genome offers several key advantages for studying human origins and genetic diversity, including its high mutation rate, strict maternal inheritance, and absence of recombination. The Y chromosome provides a complementary perspective by tracing paternal lineages. As a haploid marker passed from father to son without recombination, it mirrors patrilineal descent and, in many cultures, aligns with surnames. Its diverse mutation rates allow for the reconstruction of male-specific genealogies across various time scales, from families to populations. Together, these uniparental markers are powerful tools for tracing migration patterns and understanding population structure.

Genetic studies focusing on uniparental markers in the Roma population have a long-standing history, beginning alongside early autosomal studies that employed markers such as STRs. However, early research faced limitations due to the lack of well-defined human haplotypes and incomplete knowledge of human genome origins, as the human genome had not yet been fully sequenced. These challenges were reflected in Roma studies, where certain haplogroups were identified, but their origin remained unclear [230]. Over the past 25 years, more than 30 studies have examined mitochondrial DNA (mtDNA) and Y-chromosome haplotypes among the Roma, with most research concentrated on populations in the Balkans, Central Europe, and the Iberian Peninsula. However, genetic data from Roma groups in other regions remain scarce.

Overall, analyses reveal that the Roma carry moderate to high frequencies of distinct paternal (Y-chromosome H) and maternal (mtDNA M) lineages, both tracing back to India, particularly North-Western India [238,249]. The remaining lineages typically consist of varying frequencies of West Eurasian haplogroups, such as Y-DNA J2 and R1b [230], or mtDNA U3b and J1/J2 [243,249,259].

Regional studies further highlight the genetic diversity among Roma subgroups. Research on Serbian Roma, for example, revealed significant variation in Y-chromosome haplogroup frequencies: while some subgroups carried high levels of the Indian-origin haplogroup H-M69, others showed a predominance of Southeastern European haplogroups, including E and I—indicative of substantial local admixture [241,260]. In Slovakia and the Czech Republic, mtDNA analysis confirmed the presence of Indian-origin haplogroups M5a1 and M35. Additionally, the identification of a novel mtDNA sub-haplogroup, J1a—shared between Roma and non-Roma populations—suggests either a shared founder event or historical gene flow with surrounding groups [243].

A study of Roma in northeastern Bosnia revealed a highly diverse mtDNA profile, comprising Indian (M5a1, M35b), Indian-European (H7a, X2b), and European/Middle Eastern haplogroups. This diversity points to distinct migration pathways and early divergence from other Roma groups [260]. Among Bosnian-Herzegovinian Roma, Y-chromosome analysis showed a high level of haplotype diversity, with three-quarters of individuals carrying unique Y-STR haplotypes. The dominant paternal lineage, H1a1a-M82, was once again traced back to North-Western India, although evidence of admixture with local European populations was also present [262].

In the Iberian Peninsula, studies of the Portuguese Roma revealed Y-chromosome lineages of South Asian origin (e.g., H1a, J2a1b1), alongside markers of local Iberian admixture (e.g., R1b1c), particularly during and following their migration from the Balkans [232]. Among Spanish Roma, mtDNA diversity appeared significantly reduced in comparison to the general Iberian population. While South Asian M lineages were still evident, European founder haplogroups—especially U3b1c—were predominant. Additional West Eurasian lineages, such as J1b3a and J2b1c*, were also detected, indicating some degree of limited admixture with Iberian populations [247].

The uniparental genetic structure of the Roma population has been strongly shaped by demographic processes that have left distinct signatures in both maternal and paternal lineages. Founder effects and long-term endogamy have played a significant role in reducing genetic diversity and increasing differentiation between subgroups. The Vlax Roma—a major subgroup—exemplify this trend. Likely descended from a single founder population, they display pronounced internal genetic structuring due to strict endogamy and regional isolation [30].

The combined analysis of maternal and paternal lineages highlights how distinct migration routes and early divergence shaped subgroup-specific genetic profiles. Lineages such as mtDNA U3 and Y-chromosome J2f suggest independent migration events and region-specific admixture patterns among Roma groups, as evidenced in Polish, Lithuanian, Spanish, and Bulgarian populations [259]. Despite their overall reproductive isolation, Roma are not genetically isolated. Increased gene flow—likely influenced by modern shifts in social norms—has reduced some of the historical genetic boundaries between subpopulations [261]. The differences among Roma groups suggest admixture during various stages of their migration through the Balkans into Europe [260]. Genetic drift has further amplified these effects over generations, resulting in the distinct genetic substructure observed across European Roma populations today [262].

1.2.4 Sex bias in the Roma population

Sex bias refers to asymmetries in the genetic contributions or inheritance patterns between males and females, often resulting from historical, social, and demographic factors. In the case of the Roma, this phenomenon has been repeatedly identified across genetic studies and appears to have played a significant role in shaping their genetic landscape from their South Asian origins through their diaspora into Europe.

Initial evidence for sex-biased admixture in Roma populations emerged from early STR studies. These revealed that over 70% of Roma males belonged to a single Y-chromosome lineage, indicating low heterogeneity in the paternal gene pool [230]. In contrast, approximately 50% of maternal lineages were of non-Asian origin—suggesting that female admixture had been more extensive than male [230]. A concurrent study found that South Asian Y-chromosome lineages accounted for nearly 45% of Roma males, while Asian mtDNA haplogroup M was found across all Roma populations, comprising 26.5% of female lineages [263].

Later research expanded on these findings, comparing Portuguese Roma to non-Roma populations. Results showed markedly less genetic differentiation on the autosomes than on the X chromosome—evidence consistent with a sex-biased admixture process [264]. Given that males carry only one X chromosome, a slower rate of homogenisation on the X chromosome suggests a higher proportion of male gene flow from surrounding European populations into the Roma gene pool. Further estimates for maternal admixture in Iberian Roma were around 30%, compared to 47% for paternal lineages [249].

The disparity in genetic contributions from different West Eurasian regions is also evident when comparing mtDNA and Y-chromosome data. One study found that 49% of paternal lineages retained South Asian ancestry, whereas mtDNA showed stronger gene flow from West Eurasian sources, reaching 75% [232,256]. This confirms that while the Roma have preserved a significant portion of their paternal heritage, their maternal ancestry has been more heavily influenced by surrounding non-Roma populations.

Further analyses traced founder lineages in proto-Roma populations before their migration into Europe, revealing significantly higher gene flow from host populations into Roma than vice versa. Male gene flow was especially pronounced—ranging from 17% in Romania to 46% in Hungary—while female gene flow from Roma to non-Roma remained minimal. The persistence of the Indian mitochondrial haplogroup M in up to 23% of Roma maternal lineages, but rarely among non-Roma Europeans, supports this pattern [238].

Hence, Roma experienced regionally variable gene flow and consistently incorporated more non-Roma women than men over the course of the diaspora [256]. This is supporting the presence of patrilocal social systems. European male influence is largely confined to specific episodes, with the most notable input traced to the Balkans. In contrast, maternal lineages show widespread and ongoing integration of European women [256].

1.2.5 Effects of complex demography on human populations: the Roma as a case study

The genetic effects of complex demography on isolated populations are profound, as small population sizes and unique historical trajectories shape their genetic landscapes. Reduced genetic diversity, often the result of demographic events such as population bottlenecks or expansions, amplifies the effects of genetic drift. This, in turn, can lead to the accumulation of deleterious mutations, which increases the genetic load—the reduction in a population's mean fitness relative to its theoretical maximum [265]. During a bottleneck, many genetic variants, including harmful ones, may be lost or rise in frequency purely by chance. In such contexts, purifying selection—which would normally act to remove deleterious mutations—becomes less effective. As a result, variants that would be efficiently purged in larger populations may persist in smaller, isolated ones. Founder effects and endogamy further increase homozygosity, which can unmask recessive diseases—a pattern clearly observed in several Indian populations [266].

The Roma population offers a particularly compelling example of how complex demographic events shape genetic variation. Originating from the Indian subcontinent and spreading across Europe over the past millennium, the Roma have experienced multiple severe bottlenecks and sustained reductions in effective population size [247,258]. Compounded by prolonged periods of genetic isolation and endogamy, these events have shaped a distinct genetic profile marked by reduced diversity, high levels of homozygosity, and elevated frequencies of certain pathogenic alleles. Unlike large populations, which often recover genetic diversity after bottlenecks, the Roma have maintained relatively small effective population sizes, allowing drift to continue operating strongly and limiting the efficacy of purifying selection [245].

Similar patterns are found in other founder populations. For instance, isolated Greenlandic communities and the Finnish population—both shaped by bottlenecks and founder effects—show elevated frequencies of deleterious variants compared to broader populations [267,268]. These parallels underscore how reduced diversity and population history can impact the distribution and persistence of harmful alleles. The Roma, however, offer a distinctive case due to their wide geographic dispersion, mobility, and complex admixture patterns.

Historical records and genetic data converge to suggest that the Roma descended from a small founding group, with an estimated male effective population size of around 100 individuals [155]. This estimate is consistent with

Ottoman tax records from the 16th century, similarly pointing to a modest Roma presence in the Balkans [13,33]. Upon arriving in Europe in small groups, the Roma's gene pool was rapidly shaped by founder effects and drift, further intensified by strong endogamy and social isolation. Genetic analyses reveal subgroup-specific patterns of diversity and disease, including the independent evolution of disease-associated haplotypes—strong evidence for sustained isolation and limited gene flow between communities [154,254].

The unique demographic trajectory of the Roma—now a large, widespread population known to have migrated across multiple continents in a relatively short time span—makes them an exceptional model for studying complex human demography. Their history integrates all major population-genetic processes: sharp bottlenecks, founder effects, long-term genetic drift, endogamy, and multiple, geographically varied admixture events. These factors have exerted a combined influence, though they have manifested at distinct periods and with differing intensities across subgroups, producing a mosaic of genetic signatures within the population. The Roma also provide a case of a population that has combined fragmentation and internal reproductive isolation with significant episodes of gene flow from diverse host societies.

This singular combination offers a valuable opportunity to observe how contrasting demographic forces—such as drift versus admixture, or isolation versus mobility—interact to shape genetic diversity and disease risk in real-world human populations. In this sense, the Roma stand out as a paradigmatic case for investigating the consequences of complex demography and its lasting imprint on genetic composition.

1.2.6 Genetic overview of the Iberian Roma

The Spanish Roma have been the focus of genetic research for over two decades, though much of this work has either concentrated on medical issues or examined Roma populations more generally, without specifically addressing the Spanish Roma. Genetic studies on the Iberian Roma predominantly focused on uniparental markers [232,247,249,256,269], with smaller portions dedicated to the X chromosome [264], genome-wide array data [239], and complete genomes [245]. These studies primarily addressed broad historical questions, such as the origins and differentiation of European Roma, rather than focusing specifically on Iberian Roma groups.

Despite limitations such as partial genome data and small sample sizes, these findings have revealed several important insights. Iberian Roma share a common origin with other European Roma [154,155,238,239], possess founder lineages in both uniparental DNA [232,247,269], and belong to the first wave of migration out of the Balkans, alongside Central and Northern European Roma [249,259]. This westernmost expansion of the Roma in Europe is characterised by the highest proportion of European-like ancestral components compared to other Roma populations, and it displays significant geographical substructure within the Iberian Peninsula [239]. For instance, in the Basque Country, analysis of MAPT inversion polymorphism (tau haplotypes) showed that Roma individuals carried significantly more H1 haplotypes than the surrounding populations. This pattern was consistent with South Asian ancestry and limited gene flow with Europeans due to social endogamy [261].

Uniparental DNA studies have also revealed sex-biased admixture patterns in various Roma groups, including the Spanish Roma [232,249,256,264]. More recent studies have expanded the scope by analysing both autosomal and mitochondrial DNA, providing deeper insights into the genetic diversity, relationships with other populations, population history, and the impact of socio-cultural practices on genetic variation and population structure [247,249,256]. However, these studies have not fully addressed the internal diversity within Roma groups, their heterogeneity both within Iberia and across Europe, or how socio-cultural practices have influenced gene flow and population admixture. Additionally, the role of assortative mating in shaping the distribution of genetic variation was not explored.

1.2.7 Medical studies of the Roma

The earliest genetic investigations into the Roma population emerged in parallel with medical research, as both fields often overlapped in their initial stages. These early studies were frequently conducted within university medical departments or hospital research institutes, where clinicians had direct access to patient blood samples. Consequently, many of the first insights into Roma genetics were derived from investigations into the prevalence of hereditary disorders among Roma patients.

One of the first notable findings came from a study on Greek Roma, where congenital glaucoma was observed in 1 out of 200 individuals—an extraordinarily high frequency compared to the general population, where it affects approximately 1 in 10,000 to 20,000 people. Other hereditary conditions, such as cystic fibrosis, were also identified at elevated frequencies among Roma patients early on [233].

Metabolic disorders were another area of early detection. For example, phenylketonuria (PKU) was reported at a higher frequency among Welsh Roma [270,271]. Among the most extensively studied conditions in the early period was primary congenital glaucoma, which received considerable attention in the 1980s, particularly in Slovakia, where several epidemiological and genetic studies were conducted [272,273].

By the late 1990s, more complex and novel genetic disorders began to be identified. One such condition was Congenital Cataracts Facial Dysmorphism Neuropathy Syndrome (CCFDN), a multi-systemic genetic disorder identified in Balkan Roma [274]. Around the same time, galactokinase deficiency was found to be relatively common among Roma individuals [275], and pathogenic mutations in the *Y-sarcoglycan* gene were also widely reported [276].

Genetic conditions such as limb-girdle muscular dystrophy, galactokinase deficiency, and hereditary neuropathies are also notably prevalent among the Roma [254,257]. In some groups, carrier rates for specific mutations exceed 5%, with frequencies reaching as high as 15% in certain communities [254,257]. Interestingly, many of these genetic disorders are not exclusive to European Roma. They are also found among populations in Pakistan and India, suggesting that the mutations predate the Roma diaspora out of India [228,254,258]. This shared genetic heritage further supports the South Asian origin of the Roma and highlights how ancient genetic variants have persisted through centuries of demographic change.

However, not all pathogenic variants can be attributed solely to South Asian ancestry or the effects of endogamy. Some mutations were likely introduced through multiple admixture events with surrounding non-Roma populations during the diaspora [32,277,278]. Others—particularly those associated with complex diseases—are more likely to reflect recent shifts in lifestyle, such as the transition from a nomadic to a sedentary way of life. This is especially evident in the case of cardiovascular disease, where increased prevalence has been linked to changes in diet, reduced physical activity, and other consequences of settlement [277,278]. In some instances, admixture appears to have mitigated the prevalence of certain diseases within specific Roma subgroups. For example, a study of the Spanish Roma population suggests that genetic factors, particularly HLA haplotypes, may influence the prevalence of multiple sclerosis (MS). The researchers reported an MS prevalence of 52.9 per 100,000—comparable to that of the non-Roma Spanish population, yet notably lower than the rates observed in other European Roma groups [279].

1.2.7.1 Rare variants and pathogenic mutations

Rare variants, generally defined as alleles occurring at low frequencies within a population, are an essential aspect of genetic research, particularly in the study of complex diseases. These variants, which can arise from *de novo* mutations (a genetic alteration that is present for the first time in one family member) or specific ancestral lineage events, are of particular interest because they can contribute significantly to disease risk [280,281]. In many cases, these mutations are population-specific, influenced by the unique genetic history and demographic events of specific populations.

In populations that experienced genetic isolation, such as the Roma, the elevated prevalence of disorders is often linked to factors such as long-term genetic isolation, endogamy, and consanguineous marriage, which can increase the frequency of recessive deleterious mutations within a population [282]. Founder effects occur when a small subset of the population carries a limited genetic pool, and this genetic makeup is passed down through successive generations, leading to higher rates of certain variants. Similarly, genetic bottlenecks—periods of reduced genetic diversity due to population decline or migration—can accentuate these effects.

In addition, recent findings suggest that generally the Roma carry a higher load of deletions in intergenic and intronic regions—particularly in genes intolerant to loss-of-function mutations—indicating a slight relaxation of purifying selection in their population [283]. This pattern may reflect the effects of founder events and partial reproductive isolation, as well as admixture, which introduced additional variation while selection continued to act against strongly deleterious exonic mutations [283].

The persistence of rare variants at higher frequencies increases the likelihood that pathogenic mutations will be passed down through generations, resulting in a higher incidence of monogenic diseases. For instance, mutations causing galactokinase deficiency (due to mutations in the GALK1 gene) and Primary Congenital Glaucoma (associated with pathogenic variants in the CYP1B1 or LTBP2 genes) are more prevalent in such populations [254,284]. These diseases, which follow Mendelian inheritance patterns, can manifest with varying degrees of severity. Because many rare mutations are deleterious, they are typically subject to purifying selection; however, in isolated populations where reduced genetic diversity weakens selection, such mutations may persist. In the context of Mendelian diseases, understanding how rare variants segregate within families—whether via autosomal recessive, autosomal dominant, or even pseudodominant patterns—is crucial for elucidating disease inheritance. Additionally, X-linked and mitochondrial inheritance can further influence the expression of these conditions.

1.2.7.2 Mendelian and complex diseases

The Roma population is known to be affected by a variety of Mendelian and complex diseases, which are largely shaped by a combination of endogamy, inbreeding (for definition, see section 1.4.2 Inbreeding and endogamy), and their complex demographic history. Mendelian diseases are genetic disorders caused by mutations in a single gene and following Mendelian inheritance patterns [285,286]. In contrast, complex diseases are disorders influenced by multiple genetic and environmental factors, rather than a single gene mutation, and do not follow simple Mendelian inheritance pattern. These conditions often run in families due to shared genetic factors and lifestyle influences [287].

In clinical genetics, consanguineous marriage is defined as a union between individuals related as second cousins or closer [288], while endogamy refers to marriage within one's own group as dictated by custom or law [289]. For cultural reasons, some populations view consanguineous marriage favourably, and endogamy remains a customary practice. Historically, this has been the case among the Roma, although it is less common today [290,291]. Endogamy and inbreeding are also observed in other groups such as North African and Arab populations [292,293], and is prevalent in physically isolated populations, such as those on islands or remote and scarcely populated territories [294–296], where it influences both Mendelian and complex diseases [297,298].

Genetic screening in Roma populations has revealed exceptionally high carrier rates for various genetic disorders, with about one in eight individuals carrying a pathogenic mutation [257]. Specific mutations have frequencies exceeding 5% in certain groups, reflecting strong founder effects similar to those observed in Finnish and Ashkenazi Jewish populations [257]. Among Iberian Roma, at least 90 distinct autosomal recessive disorders have been identified, manifesting as 91 distinct phenotypes and 111 pathogenic disease variants [299]. These findings indicate a significant health burden due to recessive Mendelian disorders, suggesting that targeted genetic screening programs could be highly beneficial for early diagnosis and intervention.

Beyond Mendelian disorders, studies have indicated a higher prevalence of type 2 diabetes, metabolic syndrome, cardiovascular disease, obesity, and hypertension within the Roma population [257,258,300]. These conditions likely result from a combination of genetic predispositions and socio-economic factors.

1.3 Population Genetics of the European Roma—A Review

Giacomo Francesco Ena, Julen Aizpurua-Iraola, Neus Font-Porterias, Francesc Calafell and David Comas.

Genes. 2022.
doi: 10.3390/genes13112068

Ena, G. F., Aizpurua-Iraola, J., Font-Porterias, N., Calafell, F., & Comas, D. (2022). Population Genetics of the European Roma—A Review. Genes, 13(11), 2068. doi: 10.3390/genes13112068

This section presents a review paper written as part of my PhD thesis, aiming to compile and summarise the current state of knowledge in the population genetics of the European Roma up to the time of publication. My contributions included writing the majority of the sections (excluding the one on sex bias), creating the figures, and refining and finalising the manuscript.

genes MDPI

Review

Population Genetics of the European Roma—A Review

Giacomo Francesco Ena [1], Julen Aizpurua-Iraola [1], Neus Font-Porterias [2], Francesc Calafell [1] and David Comas [1,*]

[1] Department of Medicine and Life Sciences (MELIS), Institute of Evolutionary Biology (CSIC-UPF), Universitat Pompeu Fabra, 08002 Barcelona, Spain
[2] Department of Biomedical Informatics, Anschutz Medical Campus, University of Colorado, Aurora, CO 80309, USA
* Correspondence: david.comas@upf.edu

Abstract: The Roma are a group of populations with a common origin that share the Romani identity and cultural heritage. Their genetic history has been inferred through multiple studies based on uniparental and autosomal markers, and current genomic data have provided novel insights into their genetic background. This review was prompted by two factors: (i) new developments to estimate the genetic structure of the Roma at a fine-scale resolution have precisely identified the ancestral components and traced migrations that were previously documented only in historical sources, clarifying and solving debates on the origins and the diaspora of the Roma; (ii) while there has been an effort to review the health determinants of the Roma, the increasing literature on their population genetics has not been subjected to a dedicated review in the last two decades. We believe that a summary on the state of the art will benefit both the public and scholars that are approaching the subject.

Keywords: population genetics; genomics; Romani; genetic structure; bottleneck; founder event

check for updates

Citation: Ena, G.F.; Aizpurua-Iraola, J.; Font-Porterias, N.; Calafell, F.; Comas, D. Population Genetics of the European Roma—A Review. *Genes* **2022**, *13*, 2068. https://doi.org/10.3390/genes13112068

Academic Editor: Luisa Azevedo

Received: 6 October 2022
Accepted: 3 November 2022
Published: 8 November 2022

1. Introduction

The Romani population constitute the largest transnational ethnic minority across the European continent, and while they are commonly known to be traditionally itinerant, most of them have been settled for centuries. Despite their long presence in numerous countries and the assimilation into part of the local culture to different extents, Romani groups share a common identity and cultural traditions [1,2]. For centuries, most of the information regarding Romani history originated from linguistics, written historical records from non-Romani populations and socio-anthropological comparisons, although recent insights have come from the characterization of their genetic background [3–7]. According to historical findings [2,8], this nomadic state of living, including cultural and social practices (for instance, endogamy) that are uncommon in the dominant European populations, has led to a cultural friction which very often resulted in social marginalization and persecution. Due to this, a significant number of the Roma might have decided not to register their ethnicity in official censuses [9]. In 2019, Romani people in Europe were estimated to number 10 million [10]; however, some Romani and international human rights organizations declare that they exceed 14 million, meaning that they constitute 1.5–1.8% of the total European population [10]. Most Romani populations are established in South-Eastern Europe, while other major groups are found in Central and Western European countries. Other Romani live outside of the European continent, in particular in the Middle East and in the Americas [2,8]. The Roma are a complex, structured population and considering just the national origin of its members usually fails to properly characterize this diversity. Understanding Romani structure requires the concerted efforts of many disciplines, as well as the involvement of the Romani public. For instance, the main linguistic units within the Roma are based on seven main dialect groups: Balkan, Vlax, Central, North-Eastern, North-Western, British and Iberian [11,12].

Linguistic and historical research has provided evidence that the Roma originated in the northern regions of South Asia, in particular the Punjab and Kashmiri regions of northern India [5,11]. Anthropological records have shown resemblances between the cultures of different Indian groups and Romani people [13], who have a similar social structure to the jatis of India where groups are usually defined by profession and the endogamous group constitutes the primary unit [2,14]. The origins of the Roma have been a subject of debate for a long time and hypotheses on their Indian origins were already made in the 19th century [15–18]. However, it was as a result of studies based on autosomal [5,7] and uniparental markers [19–21] that their genetic South Asian origin was corroborated, which was also supported by previous evidence of Mendelian mutations that are shared by the Roma and individuals of South Asian ancestry [22,23]. Because their history of migrations, bottlenecks and endogamy practices have led to the presence of a number of founder mutations [23] and sharing of unique Mendelian disorders [3,22], as well as founder mitochondrial DNA (mtDNA) and Y-chromosome (Y-DNA) lineages [6], the Roma are described as a group of founder populations with a relatively complex demography and history [21,24].

Historical records show that the Roma were initially welcomed with mixed feelings in most of the European kingdoms and became known as “Egyptians” and locally derived terms, such as ‘astigani’, ‘tsigane’ and ‘gitanos’, leading to misconceptions regarding the origins of the population. The first descriptions of the Roma included a variety of differently skilled workers, such as farmworkers, blacksmiths, craftsmen and mercenary soldiers, as well as musicians and entertainers [2,8,25]. Due to changes in the political climate, between the 15th and 16th century, the Roma were engulfed by a wave of intolerance, often forced to settle or to leave the countries in which they were travelling [26–28]. When Romani individuals did not accept these new imposed laws, they were arrested, enslaved, deported or killed [29,30]. Since then, and until the second part of the 19th century, they lived in slavery in many parts of western, central and eastern Europe [8,31]; the persecutions culminated into the genocide that occurred during the Second World War, in which it is estimated that the Nazi regime and its allies exterminated around half of the Romani population of Europe [32–37].

Romani people nowadays are full-right citizens in the European Union, but despite improvements in their living conditions, there is still an anti-Romani sentiment which has proven to be a difficult prejudice to overcome [38]. The Roma still suffer many inequalities, with high rates of unemployment and poverty, low participation in education and a poor health status [39,40], and governments are still slow in taking action against their discrimination [38,40].

2. Origins of the Roma from Genetic Evidence

The first genetic evidence of the Indian origins of the Roma came from studies based on blood group genetic markers [41–43] and founder mutations in the γ-*Sarcoglycan* gene [44,45]. The presence of several disease-causing Mendelian mutations [22], which can be explained by founder events, such as the congenital myasthenia caused by the *1267delG* mutation in the *CHRNE* gene, also supported the Indian origin of the Roma, as this form of inherited disorder was otherwise described only in patients of Indian and Pakistani genetic ancestry [23,46]. Other mutations, such as the *W24X* mutation present in the *GJB2* (connexin 26) gene related to non-syndromic hearing loss or the mutations present on the *LTBP2* gene, were found in the Roma [47–49] and are known to be present at a high frequency in a number of Indian populations [50].

However, most of the genetic evidence regarding the South Asian origin of the Roma has been provided by the analyses of uniparental lineages (mtDNA and Y-DNA), as the presence of South Asian lineages in Romani groups pointed to an Indian origin of the proto-Roma [3]. Regarding mtDNA, M haplogroups (such as M5a1, M18 and M35b), which originated in South Asia, are present in Romani groups [21]. One of the most distinct maternal lineages found in the Roma is the M5a1b1a1 haplogroup that, as expected from a founder lineage, has been observed at a high frequency but low diversity [51], and whose diversity in the Roma has been dated around 1.5 kya [51]. Other strong evidence on the Indian origin comes from Y-DNA data due to the high frequency of the Y haplogroup H1a-M82 [19,52], which is originally from North-Western India [53]. While studies based on uniparental DNA agree on the estimating dates of the origins [51,53], some have contradicted each other about the exact location within the Indian subcontinent in which the proto-Roma originated [20,21]. A study suggesting that southern India could have also contributed to the gene pool of the proto-Roma, based on the exact matches of Y-DNA H haplotypes [20], contradicted previous reports based on mtDNA and sparked an intense debate. However, as some authors noted [4], the Indian sampling of the study in question did not include groups from the north-western regions, biasing the results. More recent findings have proven that Dravidian-speaking populations from South India are also involved in the South Asian source of the Romani individuals [7], appearing to solve the contradiction regarding the identification of uniparental Roma lineages with a North-Western Indian origin and the high level of Y-STR haplotype sharing among the Roma and South Indian populations.

Based on genome-wide SNP arrays and whole-genome sequences, it has been determined that the Romani people carry approximately 20–35% South Asian ancestry [4,7], and North-West India constitutes the major source of this component [4,7,54]. The South Asian genetic ancestry is complex and can be decoupled into two components: Ancestral North Indian (ANI) and Ancestral South Indian (ASI) [55]. The former is related to the admixture events that occurred in South Asia around 2–4 kya from West Eurasian migrations [56] and, as a result, the Roma may have already carried an Ancestral West Eurasian (AWE) component derived from the South Asian ANI prior to their arrival into Europe [4,7]. The general consensus on the origins was eventually reached with the results coming from analyses of uniparental markers and whole-genome sequences, which pinpointed the origins in north/northwestern India 1.5 kya [5,6,57], in agreement with the dates obtained from uniparental markers [21,52]. Additionally, these results proved that all the analyzed Romani people descended from a common single founder population [5–7], as previously suggested by Kalaydjieva et al. [24].

3. The Romani Diaspora out of India

Historical and genetic sources agree on the fact that after the proto-Roma left North-Western India, they migrated through the Middle East, mostly through Persia (in the Iranian Plateau) and the Caucasus, Armenia in particular [4,8,13,21]. Such evidence suggests the occurrence of a southern Black Sea migration through Anatolia into the Balkans [11,58], which might account for the near absence of Arabic and the existence of a few lexical items of Georgian derivation in the Romani language [11]. At the same time, a more northerly route through the Caucasus and Crimea into Eastern Europe has been deemed as unlikely by historical sources, due to the fact that the first evidence of the Roma in Russia appears rather late, already into the 16th century [13]. Nonetheless, there is no agreement on the amount of gene flow that the Roma received during their diaspora until reaching Europe. Some genome-wide studies have suggested a long stay of Romani groups in the Middle East with substantial gene flow from local populations [59], whereas others have proposed a moderate Middle East and Caucasus gene flow during the migration across these territories [5,7]. The migration through the Middle East and Caucasus is supported by the presence of uniparental haplogroups, which most likely have a Middle Eastern and Caucasus origin, having their highest frequency in these present-day regions, such as the mtDNA U3 haplogroup in Iran and Lebanon [60] or the Y-DNA J2 haplogroup in Ingushetia and Chechnya [61]. Despite this, Middle Eastern admixture is probably overestimated due to its genetic similarity with European-related ancestry, and the presence of high-frequency uniparental haplogroups from this region in the Roma could be explained by the occurrence of high levels of genetic drift.

It has been historically suggested that the Roma reached the Anatolian peninsula between the 11th and 12th centuries [2,8,27] and the Balkans as early as the 13th century [8,31]; the dating of these events is also supported by genetic studies [4,5,7,57] (Figure 1). The Balkan genetic footprint in the Roma is well documented [4,57,59] and there is evidence of an ancestry gradient that correlates with the distance to the Balkans. In fact, the Balkan genetic component varies from 45% in Bulgarian, Greek and Serbian Roma to 25% in Baltic and Iberian Roma, which is further evidence that the dispersion into Europe took place via the Balkans [5,7]. After subsequent migrations and dispersal across Europe, Romani groups eventually reached Northern Europe and the Iberian Peninsula, the western-most part of the continent, in the 15th century [5,8]. We can speculate that one of the earlier migration waves of the Roma that started from the Balkans is the one that generated the Polish, Lithuanian and Iberian Roma [21], because only these three subpopulations show a high frequency of mtDNA U3 and Y-DNA J2 haplogroups, present in over a third of the individuals [3,51,62,63].

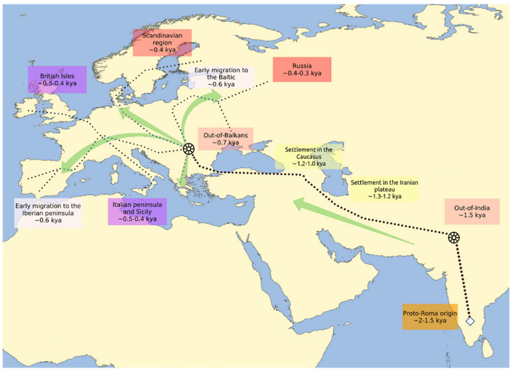


Figure 1. Early routes of the Romani diaspora. The figure is based on data from Kenrick 2007, Mendizabal et al., 2012, Moorjani et al., 2013, Font-Porterias et al., 2019 and Bianco et al., 2020, and reports the estimated date of arrival and settlement of the Romani populations in different regions of West Eurasia. The wheel symbols indicate the two main reported bottlenecks experienced by the Roma, the arrows indicate the main flow of the migrations and the dotted lines indicate the possible routes of the diaspora.

In general, Romani people carry approximately 65–80% West Eurasian (European, Middle Eastern and Caucasian) ancestry, estimated to have been acquired by extensive gene flow between the 13th and 16th centuries [4,7]. Among the West Eurasian sources, Eastern Europe constitutes the major source of European ancestry [4,6,7,54]. The different estimates of West Eurasian ancestry depend on the inclusion, or exclusion, of the South Asian AWE component in the Roma (derived from the ANI in Indian populations): estimates of the recent West Eurasian component increase from 65% [7] to 80% [4] when AWE is included. In addition, the influence of West Eurasian gene flow in the Roma is attested by the presence in the moderate to high frequencies of West Eurasian mtDNA (i.e., H, U, X, J, T) and Y-DNA lineages (i.e., J and R) [3,19,24,52,64].

4. The Genetic Differentiation and Heterogeneity among Different European Romani Subpopulations

Despite their common origin, Romani groups have experienced several population splits, founder effects, small population sizes and differential gene flow from non-Romani groups that have caused strong drift effects and rapid genetic differentiation and heterogeneity among different Romani subpopulations. The effect of this genetic heterogeneity is shown in the diverse frequency patterns of uniparental haplogroups [65–68], as well as in the frequency of disease-causing mutations [22–24] and genome-wide structures [4,54,57,69,70]. This has resulted in a marked genetic substructure with different levels of complexity [57,71].

In addition to differential bottlenecks and the loss of genetic variation experienced by each Romani group, part of the heterogeneity might be explained by different exogamy patterns expressed as gene flow within non-Romani groups, which might have been different in sources and amount [7]. As part of the recent European admixture, every Romani exhibits a Balkan common element (with different percentages in each group) and some local gene flow that are specific to each group, i.e., non-Roma Spanish into the Spanish Roma or non-Roma Polish into the Polish Roma [4,7]. The Western European Roma generally differ by exhibiting higher amounts of non-Roma European ancestry and lower levels of South Asian ancestry [4,7,72] (Figure 2).

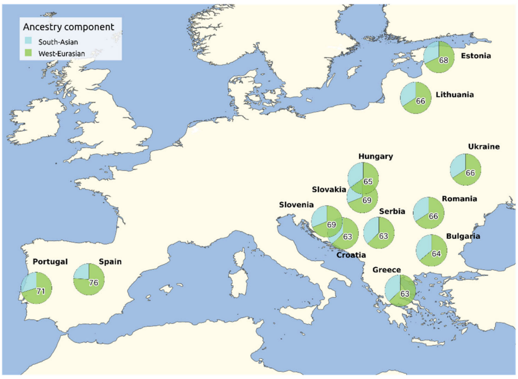


Figure 2. South Asian and West Eurasian ancestry components in Romani groups based on the results from Font-Porterias et al., 2019. The weighted means for the individuals obtained in the original analysis by genetic clusters are regrouped by country of origin. For sake of clarity, only West Eurasian ancestry percentages are shown.

Some differences between Romani groups can be spotted from the heterogeneous distribution of the mtDNA haplogroups and from the low frequency of M5 in the Polish, Iberian and Lithuanian Roma [3] to the presence of some exclusive branches, such as X2e, in the same Roma populations [73]. As other instances of differentiation, the Polish Roma frequently show haplotypes from W and K haplogroups that appear to be absent in other Romani populations [65], while the Romungro from Ukraine have the highest frequency of the J1 haplogroup and show a complete absence of the M5 haplogroup [6].

Differences in the time of admixture between the Roma and non-Romani groups have also been explored through the use of haplotype-based methods of dating, as these results show that the older dates of admixture for the Balkan Roma genetic clusters (a grouping method based on haplotype affinity) are around 800 ya, whereas the Iberian Roma genetic clusters are estimated to be around 600 ya [7]. The presence of South-Western European elements in the Iberian Roma also supports this hypothesis, as these elements are lacking in the Balkan groups. Nonetheless, Balkan genetic clusters, such as the Croatian Roma, show a higher level of more recent admixture, with an estimated date of admixture around 400 ya [7].

Local ancestry analyses have been used to infer more details about the time of these admixture events by measuring the length of ancestral chromosomal segments. As suggested by Mendizabal et al. [5], recent gene flow coming from non-Romani groups is supposed to generate individuals with long chromosomal segments of non-Roma European ancestry and individuals without traces of these segments, whereas cumulative recombination tends to shorten and spread the non-Roma European chromosomal tracts across the Roma population [74]. This is the case for several Romani groups from Central Europe and the Balkans, where a few individuals have shown very long ancestral chromosomal segments of non-Roma European origin; meanwhile, the populations overall have low mean values of admixture, which is possibly reflective of modern changes in social rules and endogamy practices, in particular [5]. The Roma from Lithuania and the Iberian Peninsula instead show higher values of non-Roma European admixture but in shorter segments, which might be proof of a more recent genetic isolation [5].

5. The Genetic Diversity of the Roma: Current Insights in the Bottleneck and Founder Events

The analysis of the demographic history of the Romani people presents two main levels of complexity. The first level of complexity comes from the gene flow with non-Romani groups that occurred all over their history, as shown in the previous sections. The second level is represented by the series of bottlenecks and splits experienced during their diaspora [5,57,75] and the different amounts of endogamy that have left traces of low intragroup diversity and high intergroup heterogeneity [6,52,63–65,76–78]. In this sense, two major bottlenecks have impacted the history of the Roma: the out-of-India and the out-of-Balkans migrations (Figure 2), events that are discernible through variations in the effective population size (*N*e), which fluctuated moderately for generations and drastically reduced in these particular instances [5,57].

According to genetic evidence, initially, the *N*e of the Roma overlapped with the *N*e of Northern Indian individuals until ~125 generations ago (ga), in agreement with the Roma origins in this region of the subcontinent [5]. The proto-Roma started to differentiate soon after the out-of-India event, proven by the occurrence of a bottleneck in which the proto-Roma *N*e became half of the parental Indian population [5,57]. It is thought that the proto-Roma generated in this circumstance left their place of origin in a single migration event [6,7]. In the period of time that goes from 125 ga to 50 ga, the Roma *N*e became lower than that of Northern Indians, and around 50 ga, it became even lower than the Dravidic groups of southern India [79,80].

The out-of-Balkans bottleneck is estimated to have happened around 900 ya, which coincides with the divergence between the Eastern and Western European Roma; the Western European Roma appear to have undergone an additional bottleneck, which reduced their *N*e to a third of that of the Eastern European Roma [5]. In the following period, both the Western and Eastern European Roma admixed with non-Roma European populations, and in a few generations, extensive gene flow from non-Roma Europeans increased the Roma *N*e, thus, compensating for the previous loss of genetic diversity [5,57]. Historically, the cessation of the *N*e reduction coincides with the settlement of the Roma in Europe and the beginning of more intense assimilation policies during the 17th century [2,57,81]. The increasing genetic distance from the Balkans and the decaying *N*e in the Western Roma suggest that cumulative drift events within Europe are one of the main catalysts of genetic differentiation within the subpopulations [5,82,83]. For this reason, Roma diversity decreased from Eastern towards Western and Northern groups due to the accumulation of drift effects during successive population splitting and migrations along the dispersion within Europe [5]. The signature of the founder effects and subsequent bottlenecks, amplified by higher levels of isolation and rates of endogamy, is also evident in the presence of a high number and total length of runs of homozygosity (RoH), in comparison to other European and Northern Indian populations [57]. In addition, a lower number of RoH was found in the Roma from the Balkans [57], possibly explicable by the fact that this group did not experience the out-of-Balkans bottleneck.

6. Gene Flow Sex Bias in the Roma Population

The Roma, as many human societies [84], traditionally have patrilocal residence patterns where the sociocultural group affiliation is patrilinearly inherited [85,86].

The genetic evidence for sexual asymmetries in gene flow has been provided by the data coming from studies on maternally inherited mtDNA and paternally inherited Y-DNA, due to their lack of recombination and well-established phylogeography [87]. Besides uniparental lineages, comparing differences in ancestry proportions between the X chromosome and the autosomes also provides information about sex-specific population history as two-thirds of the X chromosomes in a population are carried by women; therefore, the influence of female ancestry is higher than that of male ancestry [88]. In the case of the Roma, sex-biased patterns have been found in their genetic variation. The South Asian ancestral component in the Romani groups has been found in significantly different proportions in male and female lineages [3,6,89]. In this sense, South Asian lineages (mainly represented by the H1a1a4b2 haplogroup) have been found at high frequencies in the Y-chromosome genetic pool (approximately 50%), whereas their presence in the maternal counterpart (represented by the M haplogroups) is less frequent (less than 25%). This points to a higher male South Asian gene pool in the Romani groups [7]. Interestingly, comparing X chromosome ancestry proportions to those in the autosomes, South Asian ancestry had a higher result in the X chromosome than in the autosome [7].

Both uniparental South Asian lineages present clear evidence of founder effects [6], although the maternal lineages show higher levels of diversity, and their origin does not seem to be restricted to a single geographic location. In contrast, the Y-DNA lineages are more homogeneous, and their origin seems to be restricted to the region of Pakistan, reflecting a patrilocal pattern in the formation of the proto-Roma population [89].

Differences between males and females can also be observed in the West Eurasian lineages incorporated by the Roma during their diaspora. More than half of the mtDNA lineages in the Roma have a West Eurasian origin, belonging to different haplogroups (i.e., H7, J1b3 and J1c1) [6,21]. Some of these lineages present evidence of founder effects [6], and their origin seems to be explained by the independent assimilation of non-Roma females all along the Romani diaspora [6,21]. In contrast, the West Asian paternal lineage composition is dominated by Middle Eastern lineages with very low diversity (i.e., J2a1b). In a similar way, most European paternal lineages belong to a single haplogroup of Balkan origin (I1a-Z140) [89].

In agreement with the uniparental markers' data, higher Middle Eastern and Caucasian ancestry proportions were observed in the autosomes when compared to the X chromosomes, suggesting a higher influx of non-Roma males in the Middle Eastern and Caucasian regions through the Romani diaspora [7].

The differences between mtDNA and Y-DNA suggest that the inclusion of non-Romani people into Roma communities has been different according to sex: non-Roma females seem to have been adopted more easily, while the low diversity levels in male lineages suggest a sporadic inclusion of non-Roma males due to changes in the socio-historical context [2,90].

7. Biomedical Implications of the Complex Romani Population Demography

Due to the characteristic history of the Romani population, some genetic conditions have been reported to be more frequent than in non-Romani groups. Others have been reported to be less frequent or absent, while new founder mutations have been detected, highlighting a distinct genetic background in terms of disease occurrence [22,23,91,92]. The possible impact of bottlenecks on the health of the Roma is discernible from the reduced genetic diversity (compared to other Europeans and South Asians), the depletion of rare alleles and the increased high frequency of slightly deleterious genomic variants [69,92]. The high level of homozygosity is another factor that can generate a higher frequency of harmful recessive mutations [91,92]. These demographic factors might have increased the mutational load in the Roma; however, this increase might have been counterbalanced by the extensive gene flow coming from the admixture with other populations. As a consequence, the excess of deleterious mutations is limited, as most did not have time to fixate due to rapid admixture events occurring shortly after founder events [69].

Regarding selection, several pathways related to immunity, metabolism, histone modification, cardiovascular traits, and xenobiotic response have been described to be subject to selective pressures in the Romani groups [69,93]. However, these signals of positive selection are also present in European or South Asian populations, pointing that the signals found in the Roma might be explained by selection processes that occurred before the admixture between these two groups, and that the signal is maintained in the Roma due to drift or weaker positive selection after the admixture [69].

Aside from genetic factors, previous reports suggest that the health status of the Roma is typically characterized by higher mortality, higher morbidity and shorter life expectancy (LE) than the non-Romani populations of Europe [94–96], where the most common causes of mortality are cardiovascular and respiratory diseases [96]. The causes of these are still unidentified and under research; however, they could likely be caused by a combination of inherited metabolic risk factors, socio-cultural and economic factors, discrimination or cultural constraints that prevent Romani people from accessing health care which has consequences, including limited access to preventive medicine [97].

As in other human populations that have experienced founder effects and bottlenecks linked to geographical, linguistic or cultural isolation, such as Finns [98], Sardinians [99] and French Canadians [100], among others, a number of private Romani mutations associated with specific single-gene disorders (Mendelian disorders) have been reported. For instance, among these private conditions, we found hereditary motor and sensory neuropathy Lom type and Russe type, congenital-cataracts facial-dysmorphism neuropathy, congenital myasthenic syndromes, limb-girdle muscular dystrophy type 2C and galactokinase deficiency to be the most common [22,23]. On average, one out of eight Romani people is a carrier of one of the five more common Mendelian mutations [22], with carrier rates for specific mutations often exceeding 5%, and sometimes 15%, depending on the analyzed group [22,23,101]. Many of the private mutations detected in the Roma have been described as ancient mutations present in the population prior to the out-of-India diaspora and are, therefore, shared by Romani patients throughout the European continent, Pakistan and India [22].

Regarding complex diseases, multiple studies have reported a higher prevalence of type 2 diabetes (T2DM) and prediabetes (PreDM) [97,102], metabolic syndrome, cardiovascular disease, a higher occurrence of obesity, overweight and hypertension [96,103], which all may contribute to their higher mortality [104]. All the risk factors were reported to be present more regularly in the Romani population compared with the majority populations in some of the studies, including smoking, sedentary lifestyle, unhealthy dietary habits, low socio-economic status and alcohol consumption, which was also associated with a higher incidence of obesity in many of them [97]. While studies have reported that the cardiovascular risk load in the Romani population differs from the non-Romani population of the same countries, the genetic risk for cardiovascular diseases appears to be in line with other European populations [105], if not lower [106]. Obesity, hypertension and smoking are also associated with metabolic syndrome, insulin resistance, high blood pressure and elevated serum lipids that can lead to the development of metabolic disorders, such as T2DM and cardiovascular disease [96]. Considering the fact that no firm conclusions can be made with the current knowledge, a metabolic hypothesis has been proposed in an effort to identify some of the causes of obesity in the Roma. The hypothesis suggests that the Roma may have experienced inadequate maternal nutrition throughout their diaspora, which would have led to fetal undernutrition. As a result, genes that would alter glucose–insulin metabolism and help the body store more energy reserves might have been selected. After settling in Europe, they acquired better nutrition and substantially lowered their energy expenditure (due to the end of their nomadic lifestyle), which may have contributed to obesity, T2DM and metabolic syndrome, all of which negatively affect LE [93,107].

For some cardiovascular/metabolic conditions, the Roma show a higher average risk allele frequency (RAF) compared to other Europeans, but the difference in RAF has so far only been found to be statistically significant only for insulin response [92]. A similar but weaker tendency was observed when comparing the Roma to Indian groups [91,92]. Generally, it is necessary to keep in consideration that Asians are genetically more susceptible to insulin resistance and diabetes than Europeans [97], and although some of the existing studies suggest a substantial prevalence of diabetes among Romani populations, and even a higher risk of developing diabetes in Romani individuals compared to non-Roma individuals, the amount of published literature on the topic remains scarce and insufficient to draw any conclusions.

In light of these aspects, it is likely that public interventions that improve general and preventive healthcare access for the Roma would have a large impact on their health conditions and life expectancy.

8. Future Perspectives

Despite the efforts performed in recent times to provide a complete panorama of the genetic diversity of our species and to overcome the lack of genetic data of some human populations, ethnic minorities are still underrepresented in large genomic projects [108,109]. In this sense, one important challenge in studying the population genetics of the Roma is a lack of data, the sampling biases and the underrepresentation of some of the Romani groups analyzed so far. Most population genetic studies have only considered the countries of origin of the Roma individuals, regardless of their socio-cultural and ethnic characteristics. This bias might mask part of the genetic diversity of Romani groups. Genetic data are absent for some of them, such as the scanty data from Central European Sinti groups, not to mention the complete lack of knowledge of the Middle Eastern and American Roma gene pools. In relation to social issues, it is not known to what extent socio-cultural customs are a factor in limiting gene flow and population admixture in the Roma, and sociocultural norms may promote assortative mating [110]. So far, no studies on the assortative mating of the Roma have been undertaken, and this is another matter that needs to be addressed.

Minority communities regularly experience discriminatory practices in healthcare and receive inferior medical care [111]. Therefore, as in other minoritarian populations, in contrast to the one-fits-all model, a right-treatment-to-the-right-patient approach should be sought and broader access to personalized medicine should be built. It is particularly important to evaluate how to assess the frequencies of functional genetic variants in key drug responses and metabolism genes of the Roma, as these significantly influence drug response differences in different populations [112]. To conclude, for the interpretation of ethnic inequalities, further research is required to determine the impact of how genetic, inherited, environmental effects and other environmental factors interact. As ethnicity is not only influenced by genetic ancestry but also by socio-cultural factors, a multidisciplinary approach is needed to unravel the population history of Romani groups.

Future genetic studies including Romani groups should follow the general ethical guidelines required for genetic studies of human populations, including those where minority populations participate, ensuring appropriate informed consent is obtained and data sharing policies, among other ethnic regulations, are followed [113,114]. However, some recent articles have casted a shadow over the entire field by questioning the ethical review and consenting process in studies of the Roma [115]. This patronizing attitude assumes that only members of human minorities are unable to exercise their free will when consenting to participation in a genetic project, which can be considered a form of racism.

As in any other underrepresented group, the engagement of the Roma community in future genomic approaches is of pivotal importance to maximize benefits, minimize potential harms and protect the rights and interests of the participants [116,117].

Author Contributions: Conceptualization, G.F.E. and D.C.; writing—review and editing, G.F.E., J.A.-I., N.F.-P., D.C. and F.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Spanish Ministry of Science, Innovation and Universities (MCIU) and the Agencia Estatal de Investigación (AEI) [PID2019-106485GB-I00/AEI/10.13039/501100011033]; and "Unidad de Excelencia María de Maeztu" [AEI, CEX2018-000792-M].

Acknowledgments: We would like to thank the Roma FAGiC association (Federació d'Associacions Gitanes de Catalunya) and all its members and associates for its continuous support and collaboration.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hancock, I. On Romani Origins and Identity. The Romani Archives and Documentation Center, University of Texas at Austin, USA. 2006. Available online: https://www.oocities.org/romani_life_society/indexpics/HistoryOrigins.pdf (accessed on 1 June 2022).
2. Fraser, A.M. *The Gypsies (The Peoples of Europe)*; Blackwell Pub.: Hoboken, NJ, USA, 1992; ISBN 978-0-631-19605-1.
3. Gresham, D.; Morar, B.; Underhill, P.A.; Passarino, G.; Lin, A.A.; Wise, C.; Angelicheva, D.; Calafell, F.; Oefner, P.J.; Shen, P.; et al. Origins and Divergence of the Roma (Gypsies). *Am. Soc. Hum. Genet.* **2001**, *69*, 1314–1331. [CrossRef]
4. Moorjani, P.; Patterson, N.; Loh, P.-R.; Lipson, M.; Kisfali, P.; Melegh, B.I.; Bonin, M.; Kádaši, L.; Rieß, O.; Berger, B.; et al. Reconstructing Roma History from Genome-Wide Data. *PLoS ONE* **2013**, *8*, e58633. [CrossRef]
5. Mendizabal, I.; Lao, O.; Marigorta, U.M.; Wollstein, A.; Gusmão, L.; Ferak, V.; Ioana, M.; Jordanova, A.; Kaneva, R.; Kouvatsi, A.; et al. Report Reconstructing the Population History of European Romani from Genome-Wide Data. *Curr. Biol.* **2012**, *22*, 2342–2349. [CrossRef]
6. Martínez-Cruz, B.; Mendizabal, I.; Harmant, C.; Pablo, R.D.; Ioana, M.; Angelicheva, D.; Kouvatsi, A.; Makukh, H.; Netea, M.G.; Pamjav, H.; et al. Origins, Admixture and Founder Lineages in European Roma. *Eur. J. Hum. Genet.* **2016**, *24*, 937–943. [CrossRef] [PubMed]
7. Font-Porterias, N.; Arauna, L.R.; Poveda, A.; Bianco, E.; Rebato, E.; Prata, M.J.; Calafell, F.; Comas, D. European Roma Groups Show Complex West Eurasian Admixture Footprints and a Common South Asian Genetic Origin. *PLOS Genet.* **2019**, *15*, e1008417. [CrossRef] [PubMed]
8. Kenrick, D. *Historical Dictionary of the Gypsies (Romanies)*; Scarecrow Press: Lanham, MD, USA, 2007; ISBN 978-0-8108-6440-5.
9. Covrig, A. Why Roma Do Not Declare Their Identity—Careful Decision or Unpremeditated Refusal? *J. Study Relig. Ideol.* **2010**, *3*, 90–101.
10. Roma Equality, Inclusion and Participation in the EU. Available online: https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combatting-discrimination/roma-eu/roma-equality-inclusion-and-participation-eu_en (accessed on 8 June 2022).
11. Matras, Y. *Romani A Linguistic Introduction*; UK Cambridge University Press: Cambridge, UK, 2002; ISBN 0-511-03580-2.
12. Zatreanu, M.; Halwachs, D.W. Romani in Europe. Ministry of Education and Research–Bucharest Romania, Department of Linguistics at the University of Graz, Austria. 2003. Available online: https://www.coe.int/t/dg4/education/roma/Source/RomaniEurope_EN.pdf (accessed on 8 June 2022).
13. Iovita, R.P.; Schurr, T.G. Reconstructing the origins and migrations of diasporic populations: The case of the European Gypsies. *Am. Anthropol.* **2004**, *106*, 267–281.
14. Marushiakova, E.; Popov, V. *Gypsies (Roma) in Bulgaria (Volume 18)*; Peter Lang Pub Incorporated: New York, NY, USA, 1997.
15. Grierson, G.A. Doms, Jâts, and the Origin of the Gypsies. *J. Gypsy Lore Soc.* **1888**, *1*, 71.
16. Frossard, E.B.D. The Gypsies in France. *Saturday Mag.* **1835**, *6*, 39–40.
17. Crawfurd, J. On the Origin of the Gypsies. *Trans. Ethnol. Soc. Lond.* **1865**, *3*, 25–36. [CrossRef]
18. Colocci, A. *The Gypsies: History of an Errant People*; Arnaldo Forni Editore: Bologna, Italy, 1889.
19. Pamjav, H.; Zalán, A.; Béres, J.; Nagy, M.; Chang, Y.M. Genetic Structure of the Paternal Lineage of the Roma People. *Am. J. Phys. Anthropol.* **2011**, *145*, 21–29. [CrossRef] [PubMed]
20. Regueiro, M.; Rivera, L.; Chennakrishnaiah, S.; Popovic, B.; Andjus, S.; Milasin, J.; Herrera, R.J. Ancestral Modal Y-STR Haplotype Shared among Romani and South Indian Populations. *Gene* **2012**, *504*, 296–302. [CrossRef]
21. Mendizabal, I.; Valente, C.; Gusmão, A.; Alves, C.; Gomes, V.; Goios, A.; Parson, W.; Calafell, F.; Alvarez, L.; Amorim, A.; et al. Reconstructing the Indian Origin and Dispersal of the European Roma: A Maternal Genetic Perspective. *PLoS ONE* **2011**, *6*, e15988. [CrossRef]
22. Kalaydjieva, L.; Gresham, D.; Calafell, F. Genetic Studies of the Roma (Gypsies): A Review. *BMC Med. Genet.* **2001**, *2*, 5. [CrossRef]

23. Morar, B.; Gresham, D.; Angelicheva, D.; Tournev, I.; Gooding, R.; Guergueltcheva, V.; Schmidt, C.; Abicht, A.; Lochmüller, H.; Tordai, A.; et al. Mutation History of the Roma/Gypsies. *Am. J. Hum. Genet.* **2004**, *75*, 596–609. [CrossRef] [PubMed]
24. Kalaydjieva, L.; Morar, B.; Chaix, R.; Tang, H. A Newly Discovered Founder Population: The Roma/Gypsies. *BioEssays* **2005**, *27*, 1084–1094. [CrossRef]
25. Bloch, J.; David, H.; Foletier, F.D.V. *The Gypsies*; FeniXX: Singapore, 1969; ISBN 978-2-7059-1051-8.
26. Brearley, M. The Persecution of Gypsies in Europe. *Am. Behav. Sci.* **2001**, *45*, 588–599. [CrossRef]
27. Taylor, B. *Another Darkness, Another Dawn: A History of Gypsies, Roma and Travellers*; Reaktion Books: London, UK, 2014; ISBN 978-1-78023-297-3.
28. Barany, Z. The East European Gypsies in the Imperial Age. *Ethn. Racial Stud.* **2001**, *24*, 50–63. [CrossRef]
29. Crowe, D.; Kolsti, J.; Hancock, I. *The Gypsies of Eastern Europe*; Routledge: London, UK, 2016; ISBN 978-1-315-49024-3.
30. Kamusella, T.; Jaskułowski, K. *Nationalisms Today*; Peter Lang: New York, NY, USA, 2009; ISBN 978-3-03911-883-0.
31. Marushiakova, E.; Popov, V.Z.; Popov, V. *Gypsies in the Ottoman Empire: A Contribution to the History of the Balkans*; University of Hertfordshire Press: Hatfield, UK, 2001; Volume 22, ISBN 978-1-902806-02-0.
32. Lewy, G. *The Nazi Persecution of the Gypsies*; Oxford University Press: Oxford, UK, 2000; ISBN 978-0-19-802904-5.
33. Ioanid, R. *The Holocaust in Romania: The Destruction of Jews and Gypsies Under the Antonescu Regime, 1940–1944*; Ivan R. Dee: Chicago, IL, USA, 2008; ISBN 978-1-4616-9490-8.
34. Kenrick, D.; Puxon, G. *Gypsies Under the Swastika*; University of Hertfordshire Press: Hatfield, UK, 2009; ISBN 978-1-902806-80-8.
35. Trevisan, P. 'Gypsies' in Fascist Italy: From Expelled Foreigners to Dangerous Italians. *Soc. Hist.* **2017**, *42*, 342–364. [CrossRef]
36. Kotljarchuk, A. World War II and the Registration of Roma in Sweden: The Role of Experts and Census-Takers. *Holocaust Genocide Stud.* **2017**, *31*, 457–479. [CrossRef]
37. Palmer, M.D.; Burgess, S.M. *The Wiley-Blackwell Companion to Religion and Social Justice*; John Wiley & Sons: Hoboken, NJ, USA, 2020; ISBN 978-1-119-57210-7.
38. A Persisting Concern: Anti-Gypsyism as a Barrier to Roma Inclusion. Available online: https://fra.europa.eu/en/publication/2018/persisting-concern-anti-gypsyism-barrier-roma-inclusion (accessed on 5 September 2022).
39. Council of Europe Strategic Action Plan for Roma and Traveller Inclusion (2020–2025). Available online: https://edoc.coe.int/en/roma-and-travellers/8508-council-of-europe-strategic-action-plan-for-roma-and-traveller-inclusion-2020-2025.html (accessed on 14 September 2022).
40. ECRI General Policy Recommendation No. 13. Available online: https://www.coe.int/en/web/european-commission-against-racism-and-intolerance/recommendation-no.13 (accessed on 5 September 2022).
41. Bernasovský, I.; Suchý, J.; Bernasovská, K.; Vargová, T. Blood Groups of Roms (Gypsies) in Czechoslovakia. *Am. J. Phys. Anthropol.* **1976**, *45*, 277–279. [CrossRef] [PubMed]
42. Bernasovský, I.; Jurícková, J.; Ferák, V. Population Genetic Study in Gypsies (Roms) from Slovakia: Distribution of Blood Group Genetic Markers. *Anthropol. Sci.* **1994**, *102*, 409–419. [CrossRef]
43. Gáliková, J.; Vilímová, M.; Ferák, V.; Mayerová, A. Haptoglobin Types in Gypsies from Slovakia (Czechoslovakia). *Hum. Hered.* **1969**, *19*, 480–485. [CrossRef]
44. Piccolo, F.; Jeanpierre, M.; Leturcq, F.; Dodé, C.; Azibi, K.; Toutain, A.; Merlini, L.; Jarre, L.; Navarro, C.; Krishnamoorthy, R.; et al. A Founder Mutation in the γ-Sarcoglycan Gene of Gypsies Possibly Predating Their Migration Out of India. *Hum. Mol. Genet.* **1996**, *5*, 2019–2022. [CrossRef] [PubMed]
45. Lasa, A.; Piccolo, F.; de Diego, C.; Jeanpierre, M.; Colomer, J.; Rodríguez, M.J.; Urtizberea, J.A.; Baiget, M.; Kaplan, J.C.; Gallano, P. Severe Limb Girdle Muscular Dystrophy in Spanish Gypsies: Further Evidence for a Founder Mutation in the γ-Sarcoglycan Gene. *Eur. J. Hum. Genet.* **1998**, *6*, 396–399. [CrossRef]
46. Abicht, A.; Stucka, R.; Karcagi, V.; Herczegfalvi, A.; Horváth, R.; Mortier, W.; Schara, U.; Ramaekers, V.; Jost, W.; Brunner, J.; et al. A Common Mutation (E1267delG) in Congenital Myasthenic Patients of Gypsy Ethnic Origin. *Neurology* **1999**, *53*, 1564. [CrossRef] [PubMed]
47. Álvarez, A.; del Castillo, I.; Villamar, M.; Aguirre, L.A.; González-Neira, A.; López-Nevot, A.; Moreno-Pelayo, M.A.; Moreno, F. High Prevalence of the $W_{24}X$ Mutation in the Gene Encoding Connexin-26 (GJB_2) in Spanish Romani (Gypsies) with Autosomal Recessive Non-Syndromic Hearing Loss. *Am. J. Med. Genet.* **2005**, *137A*, 255–258. [CrossRef]
48. Bouwer, S.; Angelicheva, D.; Chandler, D.; Seeman, P.; Tournev, I.; Kalaydjieva, L. Carrier Rates of the Ancestral Indian $W_{24}X$ Mutation in GJB_2 in the General Gypsy Population and Individual Subisolates. *Genet. Test.* **2007**, *11*, 455–458. [CrossRef]
49. Azmanov, D.N.; Dimitrova, S.; Florez, L.; Cherninkova, S.; Draganov, D.; Morar, B.; Saat, R.; Juan, M.; Arostegui, J.I.; Ganguly, S.; et al. $LTBP_2$ and CYP_1B_1 Mutations and Associated Ocular Phenotypes in the Roma/Gypsy Founder Population. *Eur. J. Hum. Genet.* **2011**, *19*, 326–333. [CrossRef]
50. Maheshwari, M.; Vijaya, R.; Ghosh, M.; Shastri, S.; Kabra, M.; Menon, P.S.N. Screening of Families with Autosomal Recessive Non-Syndromic Hearing Impairment (ARNSHI) for Mutations in GJB_2 Gene: Indian Scenario. *Am. J. Med. Genet.* **2003**, *120A*, 180–184. [CrossRef]
51. Gómez-Carballa, A.; Pardo-seco, J.; Fachal, L.; Vega, A.; Cebey, M.; Salas, A. Indian Signatures in the Westernmost Edge of the European Romani Diaspora: New Insight from Mitogenomes. *PLoS ONE* **2013**, *8*, e75397. [CrossRef]
52. Zalán, A.; Béres, J.; Pamjav, H. Paternal Genetic History of the Vlax Roma. *Forensic Sci. Int. Genet.* **2011**, *5*, 109–113. [CrossRef] [PubMed]

53. Rai, N.; Chaubey, G.; Tamang, R.; Pathak, A.K.; Singh, V.K.; Karmin, M.; Singh, M.; Rani, D.S.; Anugula, S.; Yadav, B.K.; et al. The Phylogeography of Y-Chromosome Haplogroup H1a1a-M82 Reveals the Likely Indian Origin of the European Romani Populations. *PLoS ONE* **2012**, *7*, e48477. [CrossRef] [PubMed]
54. Melegh, B.I.; Banfai, Z.; Hadzsiev, K.; Miseta, A.; Melegh, B. Refining the South Asian Origin of the Romani People. *BMC Genet.* **2017**, *18*, 82. [CrossRef]
55. Reich, D.; Thangaraj, K.; Patterson, N.; Price, A.L.; Singh, L. Reconstructing Indian Population History. *Nature* **2009**, *461*, 489–494. [CrossRef]
56. Narasimhan, V.M.; Patterson, N.; Moorjani, P.; Rohland, N.; Bernardos, R.; Mallick, S.; Lazaridis, I.; Nakatsuka, N.; Olalde, I.; Lipson, M. The Formation of Human Populations in South and Central Asia. *Science* **2019**, *365*, eaat7487. [CrossRef]
57. Bianco, E.; Laval, G.; Font-Porterias, N.; García-Fernández, C.; Dobon, B.; Sabido-Vera, R.; Stefanovska, E.S.; Kučinskas, V.; Makukh, H.; Pamjav, H.; et al. Recent Common Origin, Reduced Population Size, and Marked Admixture Have Shaped European Roma Genomes. *Mol. Biol. Evol.* **2020**, *37*, 3175–3187. [CrossRef]
58. Matras, Y.; Bakker, P.; Kyuchukov, H. *The Typology and Dialectology of Romani*; John Benjamins B.V.: Amsterdam, The Netherlands, 1997.
59. Bánfai, Z.; Ádám, V.; Pöstyéni, E.; Büki, G.; Czakó, M.; Miseta, A.; Melegh, B. Revealing the Impact of the Caucasus Region on the Genetic Legacy of Romani People from Genome-Wide Data. *PLoS ONE* **2018**, *13*, e0202890. [CrossRef]
60. Derenko, M.; Malyarchuk, B.; Bahmanimehr, A.; Denisova, G.; Perkova, M.; Farjadian, S.; Yepiskoposyan, L. Complete Mitochondrial DNA Diversity in Iranians. *PLoS ONE* **2013**, *8*, e80673. [CrossRef]
61. Tarkhnishvili, D.; Gavashelishvili, A.; Murtskhvaladze, M.; Gabelaia, M.; Tevzadze, G. Human Paternal Lineages, Languages, and Environment in the Caucasus. *Hum. Biol.* **2014**, *86*, 113–130. [CrossRef]
62. Malyarchuk, B.A.; Perkova, M.A.; Derenko, M.V.; Vanecek, T.; Lazur, J.; Gomolcak, P. Mitochondrial DNA Variability in Slovaks, with Application to the Roma Origin. *Ann. Hum. Genet.* **2008**, *72*, 228–240. [CrossRef] [PubMed]
63. Gusmão, A.; Gusmão, L.; Gomes, V.; Alves, C.; Calafell, F.; Amorim, A.; Prata, M.J. A Perspective on the History of the Iberian Gypsies Provided by Phylogeographic Analysis of Y-Chromosome Lineages. *Ann. Hum. Genet.* **2008**, *72*, 215–227. [CrossRef] [PubMed]
64. Klarić, I.M.; Salihović, M.P.; Lauc, L.B.; Zhivotovsky, L.A.; Rootsi, S.; Janićijević, B. Dissecting the Molecular Architecture and Origin of Bayash Romani Patrilineages: Genetic Influences from South-Asia and the Balkans. *Am. J. Phys. Anthropol.* **2009**, *138*, 333–342. [CrossRef] [PubMed]
65. Malyarchuk, B.A.; Grzybowski, T.; Derenko, M.V.; Czarny, J.; Miścicka-Śliwka, D. Mitochondrial DNA Diversity in the Polish Roma. *Ann. Hum. Genet.* **2006**, *70*, 195–206. [CrossRef]
66. Kalaydjieva, L.; Calafell, F.; Jobling, M.A.; Angelicheva, D.; Knijff, P.D.; Rosser, Z.H.; Hurles, M.E.; Underhill, P.; Tournev, I.; Marushiakova, E.; et al. Patterns of Inter- and Intra-Group Genetic Diversity in the Vlax Roma as Revealed by Y Chromosome and Mitochondrial DNA Lineages. *Eur. J. Hum. Genet.* **2001**, *9*, 97–104. [CrossRef]
67. Pokupčić, K.; Cukrov, S.; Klarić, I.M.; Salihović, M.P.; Lauc, L.B.; Blažanović, A.; Janićijević, B. Y-STR Genetic Diversity of Croatian (Bayash) Roma. *Forensic Sci. Int. Genet.* **2008**, *2*, e11–e13. [CrossRef]
68. Ahmic, A.; Kalamujic, B.; Ismailovic, A.; Lasic, L.; Halilovic, E.; Mujkic, I.; Pojskic, N. Mitochondrial DNA Diversity of the Roma from Northeastern Bosnia, Bosnia and Herzegovina. *HOMO* **2018**, *69*, 347–356. [CrossRef]
69. Font-Porterias, N.; Caro-Consuegra, R.; Lucas-Sánchez, M.; Lopez, M.; Giménez, A.; Carballo-Mesa, A.; Bosch, E.; Calafell, F.; Quintana-Murci, L.; Comas, D. The Counteracting Effects of Demography on Functional Genomic Variation: The Roma Paradigm. *Mol. Biol. Evol.* **2021**, *38*, 2804–2817. [CrossRef]
70. Bánfai, Z.; Melegh, B.I.; Sümegi, K.; Hadzsiev, K.; Miseta, A.; Kásler, M.; Melegh, B. Revealing the Genetic Impact of the Ottoman Occupation on Ethnic Groups of East-Central Europe and on the Roma Population of the Area. *Front. Genet.* **2019**, *10*, 558. [CrossRef]
71. Morar, B.; Azmanov, D.N.; Kalaydjieva, L. Roma (Gypsies): Genetic Studies. In *eLS*; John Wiley & Sons: Hoboken, NJ, USA, 2013; ISBN 978-0-470-01590-2.
72. Pereira, V.; Gusmão, L.; Valente, C.; Pereira, R.; Carneiro, J.; Gomes, I.; Morling, N.; Amorim, A.; João Prata, M. Refining the Genetic Portrait of Portuguese Roma through X-Chromosomal Markers. *Am. J. Phys. Anthropol.* **2012**, *148*, 389–394. [CrossRef]
73. Reidla, M.; Kivisild, T.; Metspalu, E.; Kaldma, K.; Tambets, K.; Tolk, H.-V.; Parik, J.; Loogväli, E.-L.; Derenko, M.; Malyarchuk, B.; et al. Origin and Diffusion of MtDNA Haplogroup X. *Am. J. Hum. Genet.* **2003**, *73*, 1178–1190. [CrossRef] [PubMed]
74. Chapman, N.H.; Thompson, E.A. The Effect of Population History on the Lengths of Ancestral Chromosome Segments. *Genetics* **2002**, *162*, 449–458. [CrossRef]
75. Chaix, R.; Austerlitz, F.; Morar, B.; Kalaydjieva, L.; Heyer, E. Vlax Roma History: What Do Coalescent-Based Methods Tell Us? *Eur. J. Hum. Genet.* **2004**, *12*, 285–292. [CrossRef]
76. Irwin, J.; Egyed, B.; Saunier, J.; Szamosi, G.; O'Callaghan, J.; Padar, Z.; Parsons, T.J. Hungarian MtDNA Population Databases from Budapest and the Baranya County Roma. *Int. J. Legal Med.* **2007**, *121*, 377–383. [CrossRef]
77. Salihović, M.P.; Barešić, A.; Klarić, I.M.; Cukrov, S.; Lauc, L.B.; Janićijević, B. The Role of the Vlax Roma in Shaping the European Romani Maternal Genetic History. *Am. J. Phys. Anthropol.* **2011**, *146*, 262–270. [CrossRef] [PubMed]
78. Peričić, M.; Klarić, I.M.; Lauc, L.B.; Janićijević, B.; Đorđević, D.; Efremovska, L.; Rudan, P. Population Genetics of 8 Y Chromosome STR Loci in Macedonians and Macedonian Romani (Gypsy). *Forensic Sci. Int.* **2005**, *154*, 257–261. [CrossRef] [PubMed]

79. Juyal, G.; Mondal, M.; Luisi, P.; Laayouni, H.; Sood, A.; Midha, V.; Heutink, P.; Bertranpetit, J.; Thelma, B.K.; Casals, F. Population and Genomic Lessons from Genetic Analysis of Two Indian Populations. *Hum. Genet.* **2014**, *133*, 1273–1287. [CrossRef]
80. Mondal, M.; Bergström, A.; Xue, Y.; Calafell, F.; Laayouni, H.; Casals, F.; Majumder, P.P.; Tyler-Smith, C.; Bertranpetit, J. Y-Chromosomal Sequences of Diverse Indian Populations and the Ancestry of the Andamanese. *Hum. Genet.* **2017**, *136*, 499–510. [CrossRef]
81. Maria Theresia and Joseph II Policies of Assimilation in the Age of Enlightened Absolutism [Rombase]. Available online: http://rombase.uni-graz.at/cgi-bin/art.cgi?src=data/hist/modern/maria.en.xml (accessed on 9 June 2022).
82. Gusmão, A.; Valente, C.; Gomes, V.; Alves, C.; Amorim, A.; Prata, M.J.; Gusmão, L. A Genetic Historical Sketch of European Gypsies: The Perspective from Autosomal Markers. *Am. J. Phys. Anthropol.* **2010**, *141*, 507–514. [CrossRef]
83. Inotai, D.; Szilvasi, A.; Benko, S.; Boros-Major, A.; Illes, Z.; Bors, A.; Kiss, K.P.; Rajczy, K.; Gelle-Hossó, A.; Buhler, S.; et al. HLA Genetic Diversity in Hungarians and Hungarian Gypsies: Complementary Differentiation Patterns and Demographic Signals Revealed by HLA-A, -B and -DRB1 in Central Europe. *Tissue Antigens* **2015**, *86*, 115–121. [CrossRef]
84. Burton, M.L.; Moore, C.C.; Whiting, J.W.M.; Romney, A.K.; Aberle, D.F.; Barcelo, J.A.; Dow, M.M.; Guyer, J.I.; Kronenfeld, D.B.; Levy, J.E.; et al. Regions Based on Social Structure. *Curr. Anthropol.* **1996**, *37*, 87–123.
85. Gamella, J.F. Marriage, Gender and Transnational Migrations in Fertility Transitions of Romanian Roma Women. *Intersect. East Eur. J. Soc. Polit.* **2018**, *4*, 57–85. [CrossRef]
86. Weyrauch, W.O. *Gypsy Law: Romani Legal Traditions and Culture*; University of California Press: Berkeley, CA, USA, 2001.
87. Heyer, E.; Chaix, R.; Pavard, S.; Austerlitz, F. Sex-Specific Demographic Behaviours That Shape Human Genomic Variation. *Mol. Ecol.* **2012**, *21*, 597–612. [CrossRef] [PubMed]
88. Goldberg, A.; Rosenberg, N.A. Beyond 2/3 and 1/3: The Complex Signatures of Sex-Biased Admixture on the X Chromosome. *Genetics* **2015**, *201*, 263. [CrossRef]
89. García-Fernández, C.; Font-Porterias, N.; Kučinskas, V.; Sukarova-Stefanovska, E.; Pamjav, H.; Makukh, H.; Dobon, B.; Bertranpetit, J.; Netea, M.G.; Calafell, F.; et al. Sex-Biased Patterns Shaped the Genetic History of Roma. *Sci. Rep.* **2020**, *10*, 14464. [CrossRef]
90. Hancock, I.F. *We Are the Romani People*; University of Hertfordshire Press: Hatfield, UK, 2002; ISBN 978-1-902806-19-8.
91. Mendizabal, I.; Lao, O.; Marigorta, U.M.; Kayser, M.; Comas, D. Implications of Population History of European Romani on Genetic Susceptibility to Disease. *Hum. Hered.* **2013**, *76*, 194–200. [CrossRef]
92. Font-Porterias, N.; Giménez, A.; Carballo-Mesa, A.; Calafell, F.; Comas, D. Admixture Has Shaped Romani Genetic Diversity in Clinically Relevant Variants. *Front. Genet.* **2021**, *12*, 683880. [CrossRef]
93. Dobon, B.; Ter Horst, R.; Laayouni, H.; Mondal, M.; Bianco, E.; Comas, D.; Ioana, M.; Bosch, E.; Bertranpetit, J.; Netea, M.G. The shaping of immunological responses through natural selection after the Roma Diaspora. *Sci. Rep.* **2020**, *10*, 1–12. [CrossRef]
94. Simko, V.; Ginter, E. Short Life Expectancy and Metabolic Syndrome in Romanies (Gypsies) in Slovakia. *Cent. Eur. J. Public Health* **2010**, *18*, 16. [CrossRef]
95. Parekh, N.; Rose, T. Health Inequalities of the Roma in Europe: A Literature Review. *Cent. Eur. J. Public Health* **2011**, *19*, 139–142.
96. Dobranici, M.; Buzea, A.; Popescu, R. The Cardiovascular Risk Factors of the Roma (Gypsies) People in Central-Eastern Europe: A Review of the Published Literature. *J. Med. Life* **2012**, *5*, 382–389. [PubMed]
97. Nunes, M.A.; Kučerová, K.; Lukáč, O.; Kvapil, M.; Brož, J. Prevalence of Diabetes Mellitus among Roma Populations—A Systematic Review. *Int. J. Environ. Res. Public. Health* **2018**, *15*, 2607. [CrossRef] [PubMed]
98. Sarantaus, L.; Huusko, P.; Eerola, H.; Launonen, V.; Vehmanen, P.; Rapakko, K.; Gillanders, E.; Syrjäkoski, K.; Kainu, T.; Vahteristo, P.; et al. Multiple Founder Effects and Geographical Clustering of BRCA1 and BRCA2 Families in Finland. *Eur. J. Hum. Genet.* **2000**, *8*, 757–763. [CrossRef] [PubMed]
99. Loudianos, G.; Dessi, V.; Lovicu, M.; Angius, A.; Figus, A.; Lilliu, F.; de Virgiliis, S.; Nurchi, A.M.; Deplano, A.; Moi, P.; et al. Molecular Characterization of Wilson Disease in the Sardinian Population—Evidence of a Founder Effect. *Hum. Mutat.* **1999**, *14*, 294–303. [PubMed]
100. Gosselin, I.; Thiffault, I.; Tétreault, M.; Chau, V.; Dicaire, M.-J.; Loisel, L.; Emond, M.; Senderek, J.; Mathieu, J.; Dupré, N.; et al. Founder SH3TC2 Mutations Are Responsible for a CMT4C French-Canadians Cluster. *Neuromuscul. Disord.* **2008**, *18*, 483–492. [CrossRef] [PubMed]
101. Martínez-Frías, M.L.; Bermejo, E. Prevalence of Congenital Anomaly Syndromes in a Spanish Gypsy Population. *J. Med. Genet.* **1992**, *29*, 483–486.
102. Werissa, N.A.; Piko, P.; Fiatal, S.; Kosa, Z.; Sandor, J.; Adany, R. SNP-Based Genetic Risk Score Modeling Suggests No Increased Genetic Susceptibility of the Roma Population to Type 2 Diabetes Mellitus. *Genes* **2019**, *10*, E942. [CrossRef]
103. Zeljko, H.; Škarić-Jurić, T.; Smolej Narančić, N.; Peričić Salihović, M.; Martinović Klarić, I.; Barbalić, M.; Starčević, B.; Barać Lauc, L.; Janićijević, B. Traditional CVD Risk Factors and Socio-Economic Deprivation in Roma Minority Population of Croatia. *Coll. Antropol.* **2008**, *32*, 667–676.
104. De Courten, B.V.; de Courten, M.; Hanson, R.L.; Zahorakova, A.; Egyenes, H.P.; Tataranni, P.A.; Bennett, P.H.; Vozar, J. Higher Prevalence of Type 2 Diabetes, Metabolic Syndrome and Cardiovascular Diseases in Gypsies than in Non-Gypsies in Slovakia. *Diabetes Res. Clin. Pract.* **2003**, *62*, 95–103.
105. Škarić-Jurić, T.; Zeljko, H.M.; Tomas, Ž.; Peričić Salihović, M.; Smolej Narančić, N.; Janićijević, B. Genetic Risk Factors for Cardiovascular Diseases in the Roma Minority Population of Croatia. *Eur. J. Hum. Genet.* **2009**, *17* (Suppl. S2), 262.

106. Soltesz, B.; Piko, P.; Sandor, J.; Kosa, Z.; Adany, R.; Fiatal, S. The Genetic Risk for Hypertension Is Lower among the Hungarian Roma Population Compared to the General Population. *PLoS ONE* **2020**, *15*, e0234547. [CrossRef]
107. Ginter, E.; Simko, V.; Wsolova, L. Fall of the Iron Curtain: Male Life Expectancy in Slovakia, in the Czech Republic and in Europe. *Cent. Eur. J. Public Health* **2009**, *17*, 171. [PubMed]
108. Popejoy, A.B.; Fullerton, S.M. Genomics Is Failing on Diversity. *Nature* **2016**, *538*, 161–164. [PubMed]
109. Sirugo, G.; Williams, S.M.; Tishkoff, S.A. The Missing Diversity in Human Genetic Studies. *Cell* **2019**, *177*, 26–31. [CrossRef] [PubMed]
110. Alfonso-Sánchez, M.A.; Espinosa, I.; Gómez-Pérez, L.; Poveda, A.; Rebato, E.; Peña, J.A. Tau haplotypes support the Asian ancestry of the Roma population settled in the Basque Country. *Heredity* **2018**, *120*, 91–99. [PubMed]
111. Patel, K.; Patel, M.; Mukhi, H. Despite Advances in Precision Medicine, Disparities Persist. *Target. Ther. Oncol.* **2020**, *9*, 16–112.
112. Ingelman-Sundberg, M.; Mkrtchian, S.; Zhou, Y.; Lauschke, V.M. Integrating Rare Genetic Variants into Pharmacogenetic Drug Response Predictions. *Hum. Genomics* **2018**, *12*, 26. [CrossRef]
113. Knoppers, B.M. Framework for Responsible Sharing of Genomic and Health-Related Data. *HUGO J.* **2014**, *8*, 1–6. [CrossRef]
114. Claw, K.G.; Anderson, M.Z.; Begay, R.L.; Tsosie, K.S.; Fox, K.; Garrison, N. A Framework for Enhancing Ethical Genomic Research with Indigenous Communities. *Nat. Commun.* **2018**, *9*, 1–7. [CrossRef]
115. Lipphardt, V.; Rappold, G.A.; Surdu, M. Representing Vulnerable Populations in Genetic Studies: The Case of the Roma. *Sci. Context* **2021**, *34*, 69–100. [CrossRef]
116. Hudson, M.; Garrison, N.; Sterling, R.; Caron, N.R.; Fox, K.; Yracheta, J.; Anderson, J.; Wilcox, P.; Arbour, L.; Brown, A. Rights, Interests and Expectations: Indigenous Perspectives on Unrestricted Access to Genomic Data. *Nat. Rev. Genet.* **2020**, *21*, 377–384. [CrossRef] [PubMed]
117. Lemke, A.A.; Esplin, E.D.; Goldenberg, A.J.; Gonzaga-Jauregui, C.; Hanchard, N.A.; Harris-Wai, J.; Ideozu, J.E.; Isasi, R.; Landstrom, A.P.; Prince, A.E. Addressing Underrepresentation in Genomics Research through Community Engagement. *Am. J. Hum. Genet.* **2022**, *109*, 1563–1571. [CrossRef] [PubMed]

1.4 Overview of genetic markers utilised in this study

Over the past 20 to 30 years, genetic research has seen significant advancements in various genotyping techniques, particularly in the use of Single Nucleotide Polymorphism (SNP) arrays and Y-chromosomal Short Tandem Repeat (Y-STR) analysis. SNP arrays have evolved to allow the detection of thousands to millions of genetic variants across the genome, enabling high-resolution genome-wide association studies (GWAS) and advancing our understanding of genetic diversity, disease susceptibility, and population genetics. Similarly, Y-STR analysis has advanced through improvements in marker selection and analytical methods, and being less computationally intensive, offers a simpler and more efficient tool for studying paternal lineage and human migration patterns.

Recent advances in population genomics have been driven by high-throughput sequencing technologies, with whole-genome sequencing (WGS) providing a comprehensive view of genetic variation, capturing both common and rare genetic variations across all regions of the genome. This method allows researchers to identify structural variants, copy number variations, and mutations that could be missed with other techniques like SNP arrays. WGS is transforming our understanding of complex traits, disease mechanisms, and human evolution by providing unprecedented resolution of genetic diversity across populations. The analysis of ancient DNA has revolutionised our understanding of human migrations and past admixture events, shedding light on long-term demographic processes. Machine learning and sophisticated statistical models now allow for more precise inference of population history, natural selection, and gene-environment interactions. Large-scale biobanks and genomic databases, such as the UK Biobank and the 1000 Genomes Project, provide extensive reference panels that improve ancestry inference and trait mapping across diverse populations. Additionally, functional genomics and multi-omics approaches are increasingly being integrated to connect genetic variation with gene expression, epigenetics, and phenotypic traits, deepening our understanding of the molecular mechanisms underlying human diversity. In this thesis, data were generated using both SNP array and Y-STR analysis methodologies, chosen due to the trade-off between cost and the level of resolution required for the analysis, as whole-genome sequencing is not typically necessary for most demographic analyses. Additionally, some data from previous studies were repurposed for this research, having utilised the best available genotyping techniques at the time.

1.5.1 Genome-wide SNP arrays

This thesis utilises genome-wide SNP array data. SNP arrays target specific, pre-identified genomic loci, providing a broad but representative picture of genetic variation across the genome. While they lack the comprehensive coverage of whole-genome sequencing, SNP arrays are well-suited for a wide range of population genetics analyses, such as those relying on identity-by-descent fragments or leveraging the abundance of non-coding variants. One of the key advantages of SNP arrays is their relatively low computational and financial burden, making them more accessible for large-scale studies with thousands of individuals. Some of the limitations of SNP arrays are the susceptibility to ascertainment bias, arising from the fact that the included variants are often discovered in specific, well-characterized populations, leading to over representation of certain groups while underrepresenting polymorphisms specific to other populations. SNP arrays are also generally ineffective for studying functional variations, as they mainly capture non-coding variants and exclude large portions of coding regions, therefore in case of studies on pathogenic mutations, only the SNPs surrounding the target regions will be generally available.

Different SNP array designs address distinct research needs, incorporating markers linked to traits like disease susceptibility, immune responses, or population-specific adaptations. In this study we used *Affymetrix Human Origins* 1 *Array*, a genotyping panel specifically designed to capture global genetic diversity and ancestry-informative markers for high-resolution demographic studies. This array minimizes ascertainment bias by incorporating markers representing diverse modern populations as well as archaic hominids, offering a valuable resource for demographic and evolutionary studies.

1.5.2 Y-chromosomal Short Tandem Repeats

This thesis employs Y-chromosomal Short Tandem Repeats (Y-STRs), which are short DNA sequences repeated in tandem and located outside of pseudo-autosomal regions of the human Y chromosome. Due to their high mutation rate, Y-STRs exhibit considerable variability, are regarded as particularly valuable markers for evolutionary investigations, in particular for studying paternal lineage and population dynamics. These markers are passed down from fathers to the male offspring in a non-recombinant form, allowing for the tracing of direct male ancestry over many generations. Furthermore, Y-STRs are widely used in forensic genetics and genealogical studies due to their ability to differentiate between closely related male lineages. However, since the Y chromosome is non-recombining, the genetic diversity observed in Y-STRs reflects mutations rather than the combined effects of recombination and mutation seen in autosomal markers. For this reason, they provide a complementary, lineage-specific perspective on human population structure and evolutionary history.

For this study, both the Yfiler™ kit and the Yfiler™ Plus kit were utilised. The Yfiler™ kit targets 17 Y-STR loci, while the Yfiler™ Plus kit expands this coverage to include 10 additional loci, offering increased resolution and discriminatory power. The additional markers included in the Yfiler™ Plus kit are particularly beneficial for distinguishing between closely related male lineages and for enhancing its applicability across diverse populations.

However, as these kits are designed primarily for forensic purposes, a potential bias exists. The loci selected for these panels are optimised for individual identification and kinship analysis, often prioritising populations of forensic interest. This focus may lead to an under-representation of markers that are more informative for global population diversity or evolutionary studies, potentially skewing results when applied in a broader genetic context. Despite these limitations, the combined use of the Yfiler™ and Yfiler™ Plus kits provides robust data for investigating paternal lineage, enabling the identification of male-mediated gene flow and offering valuable insights into population history.

1.5 The importance of studying population genetics in Roma people

For the Roma people—an under-represented group with many aspects of their history remaining undocumented or unclear—population genetics serves as a powerful tool for uncovering their demographic past and the factors that have shaped their genetic diversity over time. The Roma genetic profile has been shaped by centuries of migration, isolation, and endogamy, which have led to both shared genetic traits and regional variations within the population. By examining their genetic data, we can uncover insights into their ancestral origins, migration patterns, and potential health risks that may be specific to the community. Furthermore, this research can contribute to improving healthcare and genetic counselling for Roma individuals, addressing health disparities that may arise from their specific genetic profile. One of the practical applications is identifying genetic predispositions to specific health conditions that may be more prevalent in this community. Beyond scientific and medical implications, studying Roma population genetics also helps challenge stereotypes and misconceptions, providing a deeper, evidence-based understanding of Roma identity, heritage, and diversity.

1.6 The significance of effective communication in science

1.6.1 Roma versus Gypsies

The terminology used to refer to the Roma community carries significant social and cultural weight, making it important to address the distinction between "Roma" and "Gypsies." The term "Roma" is the self-designation used by many members of the community and is derived from the Romani word *rom*, meaning “man” or “husband.” It is widely accepted in academic and official contexts, reflecting respect for the group's own identification and cultural heritage.

In contrast, the terms "Gypsy" or "Gypsies," along with their equivalents in other languages, such as *gitano* in Spanish, *cigano* in Portuguese, or *cigan* in Slavic languages, are often regarded as offensive by Roma individuals and organisations. These terms originated in a historical misunderstanding, as early European observers mistakenly believed the Roma to come from Egypt. Over time, "Gypsies" has been associated with stereotypes, discrimination, and negative connotations that have contributed to the marginalisation of Roma communities. Its usage is frequently tied to outdated and pejorative ideas, which continue to perpetuate harm.

In this thesis, the term "Roma" is intentionally used to respect the community's self-identification and to align with contemporary ethical standards in academic research. When historical sources or contexts necessitate the use of the term "Gypsies," it is presented in quotation marks and accompanied by a clear acknowledgment of its controversial nature.

1.6.2 Inbreeding and endogamy

The terms “inbreeding” and “endogamy” are used in this research to describe specific genetic phenomena, and their scientific meanings warrant careful clarification to avoid misinterpretation. In population genetics, these terms refer to measurable patterns of genetic inheritance observed in populations with particular social or demographic structures. However, outside of scientific contexts, they can carry cultural, social, or moral connotations that may lead to misunderstanding.

“Inbreeding” in its scientific use refers broadly to the increased likelihood of inheriting identical genetic material from both parents, a pattern that can arise in populations with restricted gene flow or smaller effective population sizes [301]. On the other hand, “endogamy” typically describes the social practice of marriage within a defined group, which can similarly result in higher genetic similarity among group members. While both terms may describe overlapping phenomena, endogamy is culturally specific and associated with social practices, whereas inbreeding is a broader term used to capture genetic effects, irrespective of cultural context [301].

In this thesis, the term “inbreeding” is used primarily as a neutral, scientific descriptor of genetic patterns, with “endogamy” applied where relevant to acknowledge cultural or social factors influencing genetic structure. The intention is not to ascribe value judgments to these terms but to describe observed genetic phenomena accurately. Recognising the potential for misinterpretation, this thesis makes a deliberate effort to contextualise these terms, ensuring clarity and sensitivity when discussing their implications.

Finally, we define “ancestry,” which in population genetics refers to the origins of an individual’s genome, based on the populations from which their biological ancestors originated. However, in the absence of detailed pedigree data, ancestry is often used to describe genetic similarity between individuals or populations rather than direct genealogical links. It is important to distinguish ancestry from ethnicity and race, especially since the latter is considered outdated, unscientific, and crude [302]. Ethnicity, while being scientifically valid, is not often used in population genetics studies because the term is not only related to genetics, but also to cultural factors—such as language, culture, religion, history, and socioeconomic status—that may not always align with genetic origins [303,304].

1.6.3 Social norms and practices

This dissertation employs sociological terms to elucidate the social and cultural factors influencing genetic variation within the Roma community. All of these concepts are social constructs, and for this reason they are dynamic. To ensure clarity and prevent misinterpretation, each term will be defined in accordance with established research in the field:

- **Social norms**:
  Social norms are the shared expectations that govern behaviour within a reference group, and they are often rooted in traditions, beliefs, and customs. They are not explicitly designed but emerge from repeated interactions and are often applicable to specific social contexts. Social norms can shape behaviour in various settings—such as workplaces, schools, or public spaces—and often regulate actions like politeness, dress codes, or etiquette, as well as broader social structures like marriage and gender roles. These norms are maintained by social influence and can be resistant to change due to the interdependence of people's expectations and behaviours. [305–307]
- **Social practices**:
  Social practices refer to the actual behaviours, activities and rituals that individuals or groups routinely engage in as part of social life. These can include a wide range of habitual actions, such as greeting others, attending meetings, or participating in community events, as well as a religious ceremony, or traditional dance. Social practices can

> depend on group identification and reflect behaviours shaped by social norms, but they are more about the actions themselves rather than the underlying norms or expectations that guide them. [305–307]

Thus, social norms can be specific to a culture or community (e.g., ethnic, religious, or geographic group), and are about general behavioural expectations that apply across a wide range of contexts. Norms can also be broader, emerging from everyday interactions across diverse groups and settings. Social practices reflect how these norms are effectively applied within societies.

1.6.4 Ethnicity and genetic disorders: Social challenges and perceptions

The development of genetic research has significantly contributed to the detection of genetic diseases and deleterious mutations. While this is undeniably useful in a practical sense and aids carriers of such diseases, in some cases, the identification of rare or exclusive mutations has led to negative social implications. In the worst-case scenarios, these include increased stigma from outsiders towards a population and, to some extent, the disruption of the positive image that these groups have of themselves. The social consequences of identifying a genetic disease can reflect a return to an essentialist view of identity, which assumes that genetic traits are fixed, inherent characteristics that define an individual's or group's identity. This perspective can reinforce stereotypes, deepen social divisions, and contribute to discrimination, particularly when certain genetic conditions are disproportionately associated with specific populations.

An example of this is the case of the Kashubian population, where the LCHAD enzyme deficiency, relatively common in this group, was referred to in the media as the “Kashubian gene”, leading to negative social and cultural consequences. It has been argued that a genetic disease can become a form of stigma, resulting in the perception of a "spoiled identity" within the affected community [310].

In the case of the Roma, the identification of population-specific genetic mutations or diseases could inadvertently fuel misconceptions about biological determinism, overlooking the complex interplay of genetics, culture, and environment in shaping health outcomes. For this reason, researchers in the field must carefully describe their findings and avoid potential misinterpretations of their data, clearly articulating the meaning and implications of their topics and results.

1.6.5 Engaging public and communities

For centuries, research has often reinforced the social exclusion of Roma people by legitimising stereotypes rather than challenging them. As a result, many Roma individuals have been hesitant to participate in genetic studies [311]. For this reason, over the past two decades, institutions such as the European Union have introduced programmes and frameworks aimed at supporting Roma communities within the EU [312]. Among these initiatives is a greater emphasis on involvement in research and education, encouraging more representative studies [313].

Engaging Roma, or any other community, in research is crucial as it facilitates access and helps build trust between researchers and participants. However, significant challenges remain, including low literacy rates, social stigma, and the fluid nature of identity, which can complicate participant recruitment and definitions of community membership. Researchers must therefore adopt a flexible approach, adapting to shifting community dynamics, funding changes, and other external factors that may impact engagement. Effective strategies include offering incentives, ensuring culturally sensitive communication, and tailoring research methods to meet the specific needs of vulnerable populations [314].

Given these considerations, meaningful engagement with the Roma community has been central to this research. Meetings with Roma individuals and associations have been organised to share findings, promote transparency, and ensure that the research aligns with their concerns and interests. Collaboration with Roma organisations, alongside non-biologist researchers specialising in Roma studies, further enriches the project by incorporating diverse perspectives and expertise.

This approach is not only ethically significant but also enhances the quality of the research itself. Involving the communities fosters trust and ensures that their voices are actively included in the interpretation of findings, reducing the risk of misrepresentation or the reinforcement of harmful stereotypes [314]. Moreover, it acknowledges their agency and promotes a sense of shared ownership over the knowledge produced. By working closely with both Roma associations and interdisciplinary scholars, the research benefits from broader insights and culturally informed perspectives, ultimately making it more inclusive, respectful, and impactful for the communities it seeks to understand.

2. OBJECTIVES

The general objective of this PhD thesis is to investigate the demographic history and genetic structure of the Iberian Roma population, focusing on the interplay between geography, migration, and social practices in shaping both autosomal and paternal genetic diversity. By analysing genome-wide array data from various European Roma groups and Y-STR markers from Iberia, Central, and Southwestern European Roma, this study seeks to bridge existing knowledge gaps through allele-frequency and haplotype-based methods, focusing on the following specific objectives:

a) Analyse admixture events before and after the Roma's arrival to the Iberian Peninsula, examining the impact of historical demographic events and gene flow on their genetic profile. (*Results section 3.1*)

b) Assess the genetic relatedness between Iberian Roma and other Roma and non-Roma populations across Europe, including potential historical influences from North African and Jewish groups. (*Results section 3.1*)

c) Evaluate the role of endogamy, assortative mating, and founder effects in shaping genetic diversity within Iberian Roma communities, while exploring how these factors compare to broader European Roma populations. (*Results section 3.1*)

d) Construct a comprehensive database of Y-STR haplotypes for the Spanish Roma to provide a detailed assessment of patrilineal diversity and population substructure. (*Results section 3.2*)

e) Investigate the paternal genetic structure of the Spanish Roma and its correlation with micro geographical variation to assess regional substructure. (*Results section 3.2*)

f) Infer the historical migration and demographic dynamics of Roma patrilineal lineages across Europe and the Iberian Peninsula, reconstructing the factors that contributed to their present-day genetic structure. (*Results section 3.2*)

3. RESULTS

3.1 The genetic footprint of the European Roma diaspora: Evidence from the Balkans to the Iberian Peninsula

Giacomo Francesco Ena, Aaron Giménez, Annabel Carballo-Mesa, Petra Lišková, Marcos Araújo Castro e Silva & David Comas.

Human Genetics. 2025.
doi: 10.3390/genes13112068

Ena, G. F., Giménez, A., Carballo-Mesa, A., Lišková, P., Araújo Castro e. Silva, M., & Comas, D. (2025). The genetic footprint of the European Roma diaspora: evidence from the Balkans to the Iberian Peninsula. Hum. Genet., 1–17. doi: 10.1007/s00439-025-02735-z

Human Genetics (2025) 144:463–479
https://doi.org/10.1007/s00439-025-02735-z

ORIGINAL INVESTIGATION

The genetic footprint of the European Roma diaspora: evidence from the Balkans to the Iberian Peninsula

Giacomo Francesco Ena[1] · **Aaron Giménez**[2] · **Annabel Carballo-Mesa**[3] · **Petra Lišková**[4,5] ·
Marcos Araújo Castro e Silva[1] · **David Comas**[1]

Received: 18 December 2024 / Accepted: 4 March 2025 / Published online: 17 March 2025

Abstract

The Roma people have a complex demographic history shaped by their recent dispersal from a South Asian origin into Europe, accompanied by continuous population bottlenecks and gene flow. After settling in the Balkans around 1,000 years ago, the Roma gradually dispersed across Europe, and approximately 500 years ago, they established in the Iberian Peninsula what is now one of the largest Roma populations in Western Europe. Focusing specifically on the Iberian Roma, we conducted the most comprehensive genome-wide analysis of European Roma populations to date. Using allele frequency and haplotype-based methods, we analysed 181 individuals to investigate their genetic diversity, social dynamics, and migration histories at both continental and local scales. Our findings demonstrate significant gene flow from populations encountered during the Roma's dispersal and confirm their South Asian origins. We show that, between the 14th and 19th centuries, the Roma spread westward from the Balkans in various waves, with multiple admixture events. Furthermore, our findings refute previous hypotheses of a North African dispersal route into Iberia and genetic connections to Jewish populations. The Iberian Roma exhibit ten times greater genetic differentiation compared to non-Roma Iberians, indicating significant regional substructure. Additionally, we provide the first genetic evidence of assortative mating within Roma groups, highlighting distinct mating patterns and suggesting a gradual shift towards increased integration with non-Roma individuals. This study significantly enhances our understanding of how demographic history and complex genetic structure have shaped the genetic diversity of Roma populations, while also highlighting the influence of their evolving social dynamics.

Introduction

The Romani population, often inaccurately referred to as "Gypsies", is recognised as Europe's largest transnational ethnic minority (O'Nions 2016). Conservative estimates place the Romani population at approximately 10 million, though, without a formal census, this number is likely an underestimation (Bernát and Messing 2016). While Roma groups share a common identity and cultural traditions, their linguistic diversity is notable (Matras 2002). Our understanding of the Roma's historical origins has been derived from a combination of linguistic studies, historical records, and more recent genetic research. Previous investigations trace their roots to South Asia, particularly the Punjab and Kashmir regions, as supported by both historical sources (Iovita and Schurr 2004; Kenrick 2007) and genetic findings (Kalaydjieva et al. 2001; Mendizabal et al. 2012; Moorjani et al. 2013; Martínez-Cruz et al. 2016; Font-Porterias et al. 2019; Ena et al. 2022).

✉ David Comas
david.comas@upf.edu

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

[2] Facultat de Sociologia, Universitat Autònoma de Barcelona, Barcelona, Spain

[3] Facultat de Geografia i Història, Universitat de Barcelona, Barcelona, Spain

[4] Department of Paediatrics and Inherited Metabolic Disorders, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic

[5] Department of Ophthalmology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic

Embarking on their diaspora from South Asia, the Roma travelled through what is now Afghanistan and various regions of Western Asia and the Caucasus, such as Iran, Armenia, and Anatolia, before entering Europe through the Balkans during the Middle Ages (Fraser 1992; Hancock 2006; Kenrick 2007). Upon their arrival, they faced a mixed reception that escalated into persecution and slavery, conditions that persisted in several European kingdoms until the 19th century (Brearley 2001). This tragic history culminated in the Nazi genocide during the Second World War (Lewy 2000; Kenrick and Puxon 2009). Although they are now recognised as full-fledged citizens within the European Union, the Roma community continues to grapple with discrimination, high unemployment rates, pervasive poverty, and significant health disparities (O'Nions 2011; Parekh and Rose 2011; Ivanov and Kagin 2014; Kajanova and Kmecova 2018).

Today, the Roma community is found throughout Europe, with the Iberian Peninsula hosting one of the largest Roma groups. Known by the endonym *Calé*, this group represents the westernmost edge of the Roma Diaspora within the continent. Historical accounts suggest that the Roma arrived on the Iberian Peninsula in the 15th century, with the earliest documentation in Zaragoza in 1425, having travelled from south-eastern Europe via a northern route (Pym 2007; Kenrick 2007; Sánchez 2022). There is speculation, based on oral traditions within some Roma communities, as well as historical interpretations of religious texts and the earlier belief that the Roma originated from Egypt, that some travellers may have reached the Iberian Peninsula through an alternative route via the Arabian Peninsula, North Africa, and the Strait of Gibraltar (Aparicio Gervás 2006; Hancock 2006; Pohoryles 2018), although no solid evidence supports this hypothesis.

The arrival of the Roma into Iberia coincided with a period of significant population and political turmoil, during which the Islamic rule of seven centuries was overthrown by Christian kingdoms. In the late 15th and early 16th centuries, these kingdoms persecuted, expelled, and forced religious conversions of Muslims and Jewish (Soyer 2007; Tartakoff 2012; Kimmel 2015; Carr 2017). This era was marked by considerable upheaval as Christian Spaniards, Jews, and Muslims coexisted for nearly a century. During this turbulent period, there may have been instances of genetic admixture between the Roma and these groups, who were eventually expelled in the 16th century, while the Roma were compelled to settle (Pym 2007; Abreu 2007; Sánchez 2022). Although initially tolerated, by the 16th century the first laws against the Roma were enacted in Spain and Portugal (Leblon 1985; Ortega 1994; Martínez Dhier 2007; Abreu 2007), restricting their freedom. This culminated in Spain with the "Gran Redada" (Great Round-up) on July 30, 1749, during which thousands of Roma individuals were arrested and imprisoned (Sánchez 2022). The 19th and 20th centuries saw significant internal movements within Spain and the broader peninsula, contributing to the displacement of many people, likely including the Roma (Bover and Velilla 1999; Silvestre 2005). Today, the Iberian Roma constitute the largest Roma population in Western Europe, estimated at nearly a million individuals, predominantly concentrated in the Andalusian region (Laparra 2007; Laparra et al. 2007). Despite this, they continue to suffer from socioeconomic inequities compared to non-Roma individuals (Tarnovschi et al. 2012; La Parra-Casado et al. 2018; Mendes and Magano 2022).

Previous genetic studies on the Roma population of the Iberian Peninsula have primarily focused on uniparental markers (Gusmão et al. 2008; Mendizabal et al. 2011; Gómez-Carballa et al. 2013; García-Fernández et al. 2020; Aizpurua-Iraola et al. 2022) with a smaller portion dedicated to the X chromosome, genome-wide array-data (Pereira et al. 2012; Font-Porterias et al. 2019), and complete genomes (Bianco et al. 2020). These studies have generally centred on broad historical inquiries, origins, and differentiation of European Roma on a large scale rather than specifically targeting Iberian Roma groups. Despite limitations due to partial genome data and small sample sizes, findings have shown that Iberian Roma share a common origin with other European Roma groups (Mendizabal et al. 2012; Moorjani et al. 2013; Martínez-Cruz et al. 2016; Font-Porterias et al. 2019), possess unique uniparental haplotypes (Gusmão et al. 2008; Gómez-Carballa et al. 2013; Aizpurua-Iraola et al. 2022), and appear to be part of the first out-of-Balkan migration along with Central and Northern European Roma (Mendizabal et al. 2011). This westernmost expansion of the Roma in Europe exhibits the highest proportion of European-like ancestral components compared to other Roma populations and appears to display geographical substructure within the Iberian Peninsula (Font-Porterias et al. 2019). However, these analyses have not fully explored the internal diversity of Roma groups, their heterogeneity (both within Iberia and across Europe), or how socio-cultural customs have shaped gene flow and population admixture within the Roma community. Additionally, the role of assortative mating in shaping the distribution of genetic variation among these populations has yet to be investigated.

To address these previous limitations, and as a part of a Roma community-driven initiative in collaboration with the FAGiC (Federation of Roma Associations of Catalonia), we have analysed genome-wide data from 105 Iberian Roma volunteers, comparing them with other Roma and non-Roma groups. Our genomic approach focuses on: (i) the internal genetic structure of Roma and its correlation with geography; (ii) the relationships between Iberian Roma and other non-Roma populations to explore possible contacts during the period of population upheaval following the Roma's arrival on the Iberian Peninsula; (iii) estimation of the admixture events in the Iberian Roma before and after their arrival to the Iberian Peninsula; and (iv) the assessment of endogamy patterns (understood as marriage within the population or community) and assortative mating of the European Roma. Our study provides a detailed analysis of the intricate population history of the Iberian Roma, offering insights from both micro-geographical and large-scale perspectives on the population history of European Roma.

Results

Genetic distinctiveness of the Roma within the European context

The genetic relationships between Roma individuals and other populations were assessed through Principal Component Analysis (PCA) and ADMIXTURE analyses. The plot of the first two PC components (Fig. 1b) shows that Iberian Roma, along with other European Roma groups, cluster between European and Indian populations, with minimal overlap with other groups (the first four principal components are shown in Fig. 1b and Supplementary Fig. 1). This pattern is consistent with previous genetic studies (Mendizabal et al. 2012; Ceballos et al. 2018; Font-Porterias et al. 2019; Bianco et al. 2020). The ADMIXTURE analysis at K = 4 reveals populations clustering by continental origin, with the Roma forming a distinct group, showing with two main components linked to European and South Asian populations (Supplementary Fig. 2). At the lowest cross-validation error, K = 6 (Supplementary Fig. 3), a Roma-specific component is highlighted, ranging from approximately 30–80% (Fig. 1c and S2).

To investigate gene flow patterns from populations encountered during the Roma diaspora, we analysed allele sharing using the outgroup *f3*-statistics and *f4*-statistics. In the *f3* outgroup test, the non-Roma European populations showed the highest *f3* values (Supplementary Fig. 4a-c), followed by West Asian and South Asian groups, with African populations showing the least affinity, irrespective of the Roma subgroup (Supplementary Fig. 4a-c). The *f4* tests showed that Roma are genetically closer to European groups than to South Asian ones (Supplementary Fig. 5). Among European groups, Iberian Roma were more closely related to the Basques (Supplementary Fig. 6a), while Czech Roma were genetically closer to Central Europeans (CEU) (Supplementary Fig. 6a). When examining proximity to South Asian groups, both Iberian and Czech Roma showed divergence from southern and western Indian populations (STU,

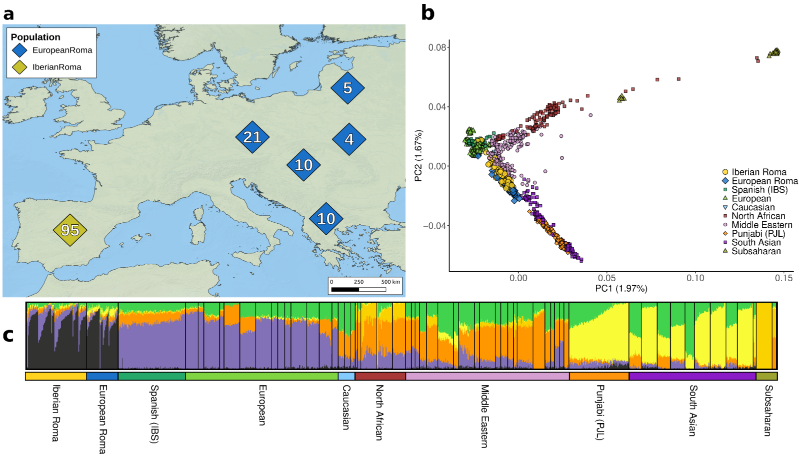


Fig. 1 Genetic population structure of the European Roma groups. (**a**) Distribution and sample sizes of Roma populations included in the present study. (**b**) Principal component analysis (PCA) of all samples analysed in this study. IBS and PJL stand for Iberian Spanish and Punjabi from the 1000 Genomes Project, respectively. (**c**) ADMIXTURE results showing the lowest cross-validation error (K = 6)

BEB, ITU) and closer genetic affinity with Pakistani and northwestern Indian groups (Supplementary Fig. 6b), supporting the hypothesis of a Punjabi origin of the Roma diaspora (Mendizabal et al. 2011; Pamjav et al. 2011; Rai et al. 2012; Martínez-Cruz et al. 2016).

We further explored the genetic relatedness among populations using FineStructure analysis, which examines haplotype similarities. This analysis revealed that most Roma individuals cluster within a single macrobranch, with the exception of two Iberian individuals clustering with non-Roma Iberians (IBS) and one Hungarian Vlax individual clustering with the general Hungarian population. These Roma individuals were not included in further dating analyses to avoid estimation biases (Supplementary Fig. 7a-c). Within the Roma macrobranch, a geographical structure emerged, with distinct clusters for Iberian, Czech, and Macedonian Roma, while other European Roma clustered together, likely due to their small sample sizes (Supplementary Fig. 7a-c). For subsequent analyses, we grouped all Iberian Roma samples into one Recipient cluster (*IberianRoma*) and all other European Roma samples into another Recipient cluster (*EuropeanRoma*).

Ancestry profiles derived from NNLS analysis showed that Roma individuals primarily shared haplotypes with European clusters, particularly *Balkan* and *CentralEurope*, followed by West and South Asian clusters, though the order varied (Supplementary Fig. 8). Minor components included other populations with negligible contributions. To simplify the interpretation of the NNLS results, we grouped the Donor clusters (Supplementary Table 1a) into 14 geographical macro-regions. Both Roma clusters show similar Donor compositions, although Iberian Roma showed a higher non-Roma Iberian ancestry and traces of southern European and North African ancestry, which were absent in other European Roma groups (Fig. 2a). We further divided the Iberian Roma in five subclusters and the European Roma in three subclusters (Supplementary Table 1b) for finer-scale analysis. Among the Iberian Roma subclusters, the European component ranged from 42 to 70%, while the South Asian component varied from 13 to 27%. Notably, the *IberianRomaSouthEast* subcluster exhibited the highest European and lowest South Asian percentages, with the other subclusters displaying more similar percentages (Supplementary Fig. 9a-b).

To further examine ancestry proportions, we estimated population-level averages from the NNLS results rather than using cluster averages. These findings revealed moderate differences, suggesting lower heterogeneity among Iberian populations compared to the clusters, and also among European populations. The primary distinction between Iberian Roma and other European Roma lies in the presence of Iberian and North African components, along with smaller South Asian and Southeastern European components (Supplementary Fig. 10a-b).

In summary, the Roma populations analysed exhibit genetic profiles with varying proportions of West Eurasia and South Asia ancestry, shaped by geographic factors. Pearson's correlation tests on the NNLS ancestry components by macro-region revealed significant correlations between longitude and several ancestral components (Supplementary Table 2a-b; Supplementary Fig. 11). Specifically, there is an east-to-west and south-to-north decrease

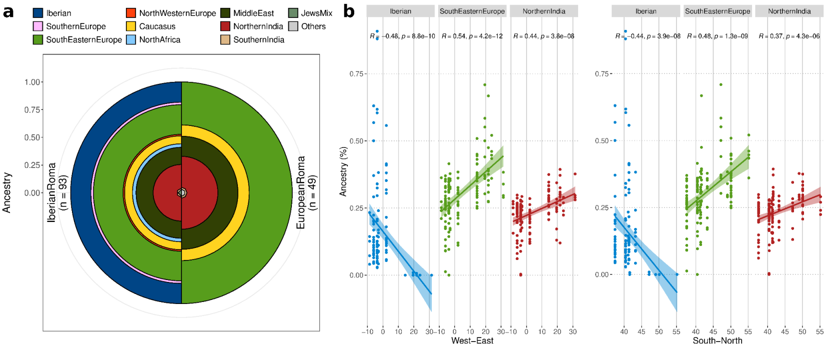


Fig. 2 Ancestry proportions in Roma and geographic correlations. (**a**) Inferred proportions of shared ancestry among Iberian (N=93) and European Roma (N=49) clusters using the NNLS method. Roma individuals that clustered outside the Roma branch in the FineStructure dendrogram were not included in this analysis. (**b**) Pearson's correlation tests for the three main ancestral components (Iberian, Southeastern European, and Indian) derived from the NNLS analysis, plotted against longitude and latitude for the Roma

in Southeastern European and Northern Indian components among European Roma, reflecting greater gene flow with non-Roma groups outside the Balkans. Conversely, an east-to-west increase in Iberian and North African components was observed among Iberian Roma, indicating higher gene flow with non-Roma Iberians (Fig. 2b).

Multiple events of admixture in European Roma groups

To estimate the timing of the major admixture events in the Roma populations, we used two methods based on an admixture pulse model. To explore admixture dynamics across various geographical scales, we conducted fastGLOBETROTTER analyses on the *IberianRoma* and *EuropeanRoma* clusters (Supplementary Tables 3 and Supplementary Fig. 12a-b), followed by analyses of the Iberian and European Roma subclusters (Supplementary Table 3a and Fig. 12c-d). The fastGLOBETROTTER analysis identified multiple waves of admixture (Fig. 3a), while MALDER detected only a single event for each significant test. For the *EuropeanRoma* cluster, two significant admixture events were identified (Fig. 3, Supplementary Table 3a-b). The first occurred approximately 30 ± 0.49 generations ago (GA) (1218 to 1243 CE, assuming 25 years per generation), involving two nearly equal ancestry sources: a major source with over 50% South Asian ancestry and a minor source predominantly consisting of Southeastern European, Caucasus, and Middle Eastern/West Eurasian ancestries. The second event, dated about 10.06 ± 1.06 GA (1722 to 1775 CE), had a major source contributing 73%, characterised by a blend of the earlier sources, evenly distributed among Southeastern European, Caucasus-Middle Eastern, and South-Asian groups. The minor source in this recent event was primarily European, with smaller contributions from West and South Asian ancestries. For the Iberian Roma cluster (Fig. 3, Supplementary Table 3a), two admixture events were also identified. The first occurred approximately 25 ± 0.21 GA (1363 to 1374 CE), with two sources: one comprising 64% European (mainly Iberian) and Caucasus-Middle East components, and the other an equal mix of South Asian and European/West Asian ancestries, including a small 2% North African component. The second event, dated to about 5 ± 0.29 GA (1864 to 1878 CE), had a major source showing a balanced contribution from European, Caucasus-Middle Eastern, and South Asian ancestries, while the minor source was predominantly West Eurasian, with a small South Asian component.

The key differences between the Iberian Roma and European Roma clusters were the presence of Iberian, North African, and Southern European components in the Iberian Roma, compared to a much larger Central and Southeastern European components in the European Roma. Additionally, the minor source in the recent admixture event of Iberian Roma included a JewsMix ancestry component, which comprises Jewish individuals from Poland, Turkey, and Morocco. This component, absent in European Roma, may indicate gene flow from North African or Jewish groups.

MALDER's estimates, which identified single admixture events for each reference pair, also revealed several statistically significant results (Supplementary Table 4). For the *EuropeanRoma*, using *Punjabi1* and *Balkan* clusters as references, the estimated date was 35.67 ± 2.18 (1001 to 1215 CE). For the *IberianRoma*, using *Punjabi1* and *IBS* as references, the estimated date was 26.41 ± 1.78 GA (1253 to 1427 CE), while using *Punjabi1* and *Balkan*

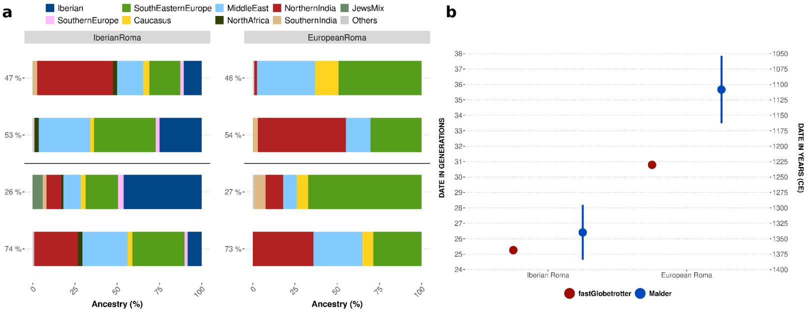


Fig. 3 Admixture events in Roma history. (**a**) Admixture events inferred by fastGLOBETROTTER, showing relative ancestry proportions grouped into macro-regions for each source of the two inferred admixture events. The plot displays recent admixture events at the bottom and older events at the top, separately for Iberian Roma (n = 93) and European Roma (n = 49). (**b**) Comparison of inferred admixture events between the fastGLOBETROTTER (representing older events) and MALDER dating methods. CE = Common Era

clusters produced a date of 28.81 ± 2.22 GA (1171 to 1389 CE). These *IberianRoma* estimates overlap with those from fastGLOBETROTTER but have larger standard errors (Fig. 3b). To further trace the earliest Roma migrations and their dispersion routes from South Asia, we used MALDER to infer admixture dates between incoming Roma and local populations. Using *Punjabi1* as a proxy for proto-Roma ancestry, we tested whether the Iberian Roma could have been formed through admixture events with various populations along the proposed dispersion path. As expected from a westward dispersal originating in South Asia, the earliest date (36.49 ± 2.22 GA) was obtained for an admixture event between the *Punjabi1* and *Iranian* clusters (Supplementary Fig. 13). The dates inferred for other groups along the way were consistent with this hypothesis, showing more recent admixture dates along an east-to-west axis, with the latest date (25.23 ± 1.81 GA) corresponding to an admixture event between *Punjabi1* and *Basque* clusters (Supplementary Fig. 13). Both methods used to infer admixture dates identified sources consistent with the NNLS analysis, primarily composed of South Asian, Southeastern European, and Middle Eastern sources, with the addition of an Iberian source specifically in the Iberian Roma.

Evolution of admixture dynamics over time

We investigated how admixture dynamics with non-Roma European populations have varied over time among different Roma groups by analysing the distribution patterns of European local ancestry segment sizes and the shared IBD segments within Roma groups and between Roma and non-Roma Europeans. We categorised these European segments based on their length, using time approximations from previous studies (Baharian et al. 2016; Harris et al. 2018; Castro e Silva et al. 2022) to infer when the haplotypes were formed (see Supplementary Note 1 and Supplementary Note 2). On average, the length of non-Roma European ancestral segments was similar across different Roma groups, ranging from 15.7 to 22.3 Mb (Supplementary Fig. 14a). When analysing the number of non-Roma European segments by length category (Supplementary Fig. 14b), Iberian Roma, along with the Romungro from Hungary and Lithuanian Roma, exhibited an increased number of segments in the longer categories (representing more recent admixture events). In the shorter categories, there were fewer differences, or none, compared to other Roma groups. The average number of segments in most categories for Iberian Roma differed significantly compared to Czech Roma and the Romungro from Ukraine (Supplementary Table 5b). This suggests that the timing of admixture with non-Roma Europeans has varied across Roma groups, with a more recent increase observed in Iberian Roma over the last 200 years. This period corresponds to the end of the Roma slavery and the contemporary era (Greenberg 2010; Marushiakova and Popov 2010).

To further support the idea of changing levels of admixture over time, we analysed the distribution of shared IBD segments within Roma populations and between Roma and non-Roma populations. Iberian Roma consistently showed significantly lower levels of within-population shared IBD compared to other Roma groups (Supplementary Fig. 15, Supplementary Table 6a-b). Conversely, shared IBD between Roma and non-Roma groups did not significantly differ in most cases when comparing Iberian Roma to other Roma groups (Supplementary Table 7). This suggests that while Iberian Roma have experienced lower levels of endogamy, they have maintained similar levels of gene flow with local non-Roma populations compared to other Roma groups.

Roma in the Iberian Peninsula exhibit high internal genetic structure

In addition to exploring the continental-scale genetic structure of European Roma, we aimed to investigate the genetic structure of Roma at a regional level. To this end, we assessed the genetic structure of Iberian samples (both Roma and non-Roma) using PCA and ADMIXTURE analyses, classifying the samples based on their geographic locations within the Iberian Peninsula (see Material and Methods) (Fig. 4a). The PCA plot reveals a clear distinction between Roma and non-Roma individuals (Fig. 4b). While non-Roma form a cohesive and distinct cluster, the Roma are more scattered across the plot, indicating regional substructure, as evidenced by the non-overlapping averages of PC1 and PC2 estimated for each Roma geographical group (Supplementary Fig. 16). The ADMIXTURE plot at the lowest cross-validation error K = 2, further differentiates Roma and non-Roma groups: Iberian non-Roma (IBS) show a single ancestral component, whereas Iberian Roma exhibit two components at varying frequencies across individuals (Fig. 4c). In contrast, the IBS display uniformity in their ancestral components even at K = 4, while the Roma show four components that are unevenly distributed (Supplementary Fig. 17; cross-validation plot in Supplementary Fig. 18).

To detect differences in ancestry profiles across Roma from different Iberian regions, we conducted an NNLS analysis, which showed that proportions are highly similar across all regions (Supplementary Fig. 19).

We formally tested the genetic structure of the Iberian samples through a series of AMOVA analyses (Supplementary Table 8). The analyses were conducted in four steps, including all individuals from the Iberian dataset except

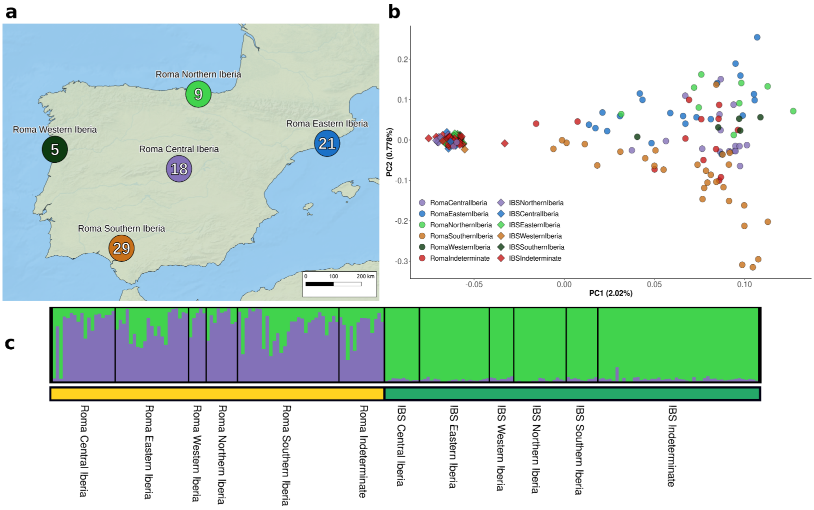


Fig. 4 Population stratification of the Iberian Peninsula. (**a**) Distribution and size of Iberian Roma populations included in this study. Iberian Indeterminate population is not represented on the map due to insufficient geographic information. (**b**) Principal Component Analysis (PCA) using the Iberian dataset. (**c**) ADMIXTURE results showing the lowest cross-validation error (K = 2)

those classified as "Indeterminate". When Iberian Roma and IBS were considered as a single group, significant genetic heterogeneity was detected between individuals (2.79%, $p = 0.00001$). When these two groups were analysed separately, 1.43% ($p < 0.0001$) of genetic heterogeneity was observed. To further explore internal heterogeneity within each group, we found that genetic heterogeneity within the Iberian Roma (1.07%, $p = 0.0001$) was more than ten times higher than that among Iberian non-Roma (0.08%, $p = 0.0021$). This finding was reinforced when Roma samples were grouped by region, showing significant heterogeneity (0.55%; $p = 0.0001$) between regions, while the regional heterogeneity in IBS is ten times lower (0.04%, $p = 0.0019$).

Finally, we estimated gene flow between Iberian Roma and IBS. A *f4* test was conducted to determine whether Iberian Roma individuals from different geographical regions share more alleles with non-Roma from the same regions. The results indicate that Roma groups share nearly equal numbers of alleles with all non-Roma groups, except for the IBS from Western Iberia (Supplementary Fig. 20a). Eastern Iberian Roma and *Indeterminate* Roma groups tend to show slightly greater genetic similarity with IBS, while comparisons involving IBS and Western and Central Iberian Roma often fall outside the significant threshold (Supplementary Fig. 20a).

To investigate interactions with Jewish and North African populations following the arrival of the Roma in the Iberian Peninsula in the 15th century, *f4* tests were conducted to analyse gene flow between these groups. The results indicate that Jews and North Africans share more alleles with the Iberian non-Roma than with the Iberian Roma, except for the Cochin Jews from India (Supplementary Fig. 20b-c). This may be due to the absence of direct gene flow between Roma and these populations, while the genetic proximity to Cochin Jews likely reflects the Indian origin of the Roma.

Mating patterns in the Roma population: endogamy and ancestry-assortative mating

We analysed shared segments of IBD to understand demographic history and mating patterns within the Roma population. Compared to other global populations (Supplementary Fig. 21), Roma exhibit a high degree of internal IBD sharing, both in terms of segment size and number (Supplementary Tables 9 and 10). This level of sharing surpasses that observed in traditionally endogamous populations, such as the Cochin Jews, Brahui, Balochi, and Mozabite (Reich et al. 2009; Henn et al. 2012; Waldman et al. 2016), as well as Jews from Libya and Tunisia (Campbell et al. 2012). Iberian Roma share more IBD segments with Basque, Punjabi, Greek, and Iberian (IBS) populations; while European Roma share the most with Greek, Punjabi, Hungarian, and Czech populations. In general, Roma groups share more IBD segments with Greek than Punjabi, although Iberian Roma show lower levels of sharing with both (Supplementary Tables 9 and 10).

The inferred historical effective population size (N_e) estimates for most Roma groups reveal a common pattern (Supplementary Fig. 22). N_e has remained low, gradually decreasing from around 50 to 8 generations ago (800–1800 CE), with the lowest N_e occurring between 7 and 29 generations ago (Supplementary Table 11). In contrast, reference populations such as Iberian, Czech, and Punjabi exhibit larger N_e values (Supplementary Fig. 22) and do not show a similar decline in N_e over time.

Genetic bottlenecks, like those experienced by the Roma, result in a higher number of shared ancestors over time and an increase in long runs of homozygosity (ROHs) (Ceballos et al. 2018). In populations where mating pairs share recent ancestors, such as the Roma, a higher number of long ROHs is expected, as they indicate recent endogamy. Conversely, many short ROHs can signal older bottlenecks followed by genetic admixture with different communities. Analysing the distribution of ROH lengths provides insight into the historical patterns of isolation and endogamy within the Roma population. To explore these patterns, we examined the length and number of ROH in our cohort (Supplementary Table 12) and found that Roma groups generally have more and longer ROHs compared to non-Roma reference populations (Supplementary Fig. 23). When analysing the average number of ROHs by length category, Roma groups exhibit fewer short ROHs but more medium-length ROHs compared to most populations (Supplementary Fig. 24a-b), with numbers comparable to South Asian groups such as Kalash and West Asian groups like the Bedouins in most categories (Supplementary Fig. 24a-d). These findings reinforce the notion that the Roma have higher levels of endogamy compared to non-Roma Europeans, primarily due to more recent isolation.

Lastly, to investigate the presence of ancestry-assortative mating within Roma populations, we applied the ANCESTOR method to the previously inferred local ancestry tracts, focusing on the proportions of European and South Asian ancestries. The results revealed that the inferred ancestries of mating pairs over the last generation showed significantly positive correlations in most Roma groups, deviating from the expectation under random mating (Supplementary Fig. 25a-b). This suggests that European and South Asian ancestry traits have influenced the mating patterns within the Roma population.

Discussion

Our comprehensive genome-wide analysis of several European Roma groups reveals that the Roma population has a genetic profile composed of approximately two-thirds West Eurasian ancestry and one-third South Asian ancestry. This finding aligns with previous genetic studies, which suggested a proto-Roma origin in the Indian subcontinent, followed by extensive gene flow during their diaspora, arrival in the Balkans, and subsequent dispersion across Europe. However, this genetic profile varies among Roma groups, showing a strong correlation between genetic ancestries and geographic location. Our results show that Western Roma (i.e. Iberian Roma) exhibit a significantly higher southwestern European component and lower southeastern European and South Asian components compared to Central and Eastern Roma groups. This patterns aligns with the long history of assimilation experienced by Roma in the Iberian Peninsula, where centuries of enforced settlement policies and assimilation laws likely led to greater incorporation of Iberian ancestry. These policies began with the 1499 *Pragmática* (a royal decree), which ordered Roma to abandon their nomadic lifestyle, acquire a trade, and serve a local lord, and culminated in the 1783 *Pragmática*, which granted citizenship but aimed at complete assimilation by prohibiting Roma culture and traditions (Martínez Dhier 2007; Sánchez 2022).

In contrast, the fragmented socio-political landscape of Central and Southeastern Europe contributed to greater endogamy and genetic isolation. Under the Ottoman Empire, Roma faced marginalisation but were generally tolerated, and settlement was not actively enforced (Marushiakova et al. 2001; Çelik 2004; Marsch 2008). However, extreme marginalisation persisted in the Ottoman suzerainties of Wallachia and Moldavia, where many Roma were segregated or enslaved (Gheorghe 1983; Achim 2004; Crowe 2016). In the Habsburg Empire, various assimilation laws were enacted, including four issued by Empress Maria Theresa (1758–1773) and a decree by Joseph II in 1783 (Crowe 2006; Zahra 2017). Despite strict sanctions, these measures were only partially effective, achieving lasting settlement in a few areas where Roma communities remain to this day, such as Burgenland in Austria (Klamper 1993; Halwachs 2005).

These differing historical circumstances likely shaped the genetic landscape of Roma populations, with greater isolation and endogamy in Eastern Roma groups leading to the preservation of a higher proportion of South Asian ancestry. This supports the notion that as Roma groups migrated westward, gene flow with local populations gradually increased. This differential gene flow may explain the distinct genetic profiles observed among European Roma groups, as demonstrated by our fineSTRUCTURE analysis, which shows that Iberian Roma form a separate cluster, while the rest of the European Roma exhibit a geographical structure. Specifically, the Czech and Macedonian Roma each form distinct clusters, while the remaining Roma samples are grouped together in a third cluster. The lack of substructure in some European Roma groups may be explained by the limited sample size.

The arrival of the Roma in Europe has been documented through historical and genetic data (Fraser 1992; Martínez-Cruz et al. 2016; Font-Porterias et al. 2019; Sánchez 2022). However, more recent migrations, such as mass movements that could be described as a second out-of-Balkans, have not been detected in previous genetic studies. Our admixture estimates performed with fastGLOBETROTTER and MALDER, suggest an earlier arrival of the proto-Roma to Eastern Europe (i.e., the European Roma groups analysed in our study) approximately 30–35 generations ago, compared to Western Europe (i.e., Iberian Roma) with estimates around 25–26 generations ago. In both cases, a significant northern Indian component is present in the admixture profiles, consistent with a North-western Indian origin of the proto-Roma (Fraser 1992; Hancock 2006; Gómez-Carballa et al. 2013; Martínez-Cruz et al. 2016; Bánfai et al. 2019; Font-Porterias et al. 2019), while other more specific ancestry components are evident in eastern and western Roma. These inferred dates slightly predate the historical records and may indicate that the earliest reports were recorded shortly after the actual arrival of the Roma to Europe. A key finding of our analysis is the evidence for previously undetected recent admixture in the Roma. Our fastGLOBETROTTER results reveal recent admixture events in both eastern and western Roma, occurring approximately 10 and 5 generations ago, respectively. The similarity of the ancestry components involved suggests ongoing migrations and gene flow during the 17th and 19th centuries among the Roma across Europe.

We assessed the effects of evolving social norms that have historically influenced relationships between Roma and non-Roma, and how these dynamics have been reflected in the genome. These changes in social norms were evaluated through the use of local ancestry tracts of European ancestry and shared IBD segments, as the presence of segments of different lengths point out to changes in admixture and endogamy over time (Baharian et al. 2016; Harris et al. 2018; Castro e Silva et al. 2022). We analysed the size and number of European ancestry tracts within Roma groups, as well as within- and between-population IBD segments, to evaluate variations in endogamy (within population) and gene flow (between populations). Finally, we categorised the segments by length and linked them to key events in Roma history. Observing the distribution of European local ancestry tracts, Iberian Roma displayed higher levels of endogamy in the past, particularly around the time of the arrival in Europe and the Iberian Peninsula. Over time, they experienced increasing gene flow with non-Roma Europeans, especially after the end of slavery and during the second out-of-Balkans migration event (Marushiakova and Popov 2010; Crowe et al. 2016), when many Roma migrated westward in search of opportunities (Marushiakova-Popova and Popov 2018). This contrasts with Southeastern European Roma groups, where gene flow with non-Roma Europeans decreased over time. Moreover, the distribution of shared IBD segments revealed an increase in endogamy levels among Southeastern European Roma and a decrease among Iberian Roma, suggesting a shift in mating patterns where, in recent times, Iberian Roma faced higher gene flow with the non-Roma, while Southeastern European Roma grew more isolated. This can also be attributed to differences in social structures between Iberian and Southeastern European Roma. These findings diverge from previous results (Mendizabal et al. 2012), where Roma from Spain, Portugal and Lithuania — all originating from a first out-of-Balkans event — showed a high number of short non-Roma European local ancestry tracts. Conversely, Southeastern European Roma groups had more long tracts. These discrepancies can be attributed to the significantly higher resolution of the current study, which included over six times more SNPs and a larger sample size for groups such as Spanish and Czech Roma, allowing a more accurate estimation of the ancestry segments.

To better understand the Roma community and their social history, we examined demographic dynamics, finding that all Roma groups have lower N_e values than non-Roma reference groups, with Iberian Roma showing a higher N_e than the other Roma groups. This difference likely stems from higher admixture with non-Roma Europeans and lower levels of isolation, which is consistent with the presence of the many long European local ancestry tracts detected in previous analyses. Interestingly, the N_e curve shows consistently low values across all Roma groups over the past 50 generations. After hitting its lowest point between 7 and 29 generations ago, the average N_e of Roma groups began to rise, suggesting a reduction in endogamy due to migration and increased gene flow with other populations. This trend is supported by the inferred admixture events (arrival into Europe and arrival into the Iberian Peninsula), aligning with earlier research (Bianco et al. 2020).

Continuing our focus on social and cultural dynamics, we conducted a detailed analysis of endogamy patterns within Roma populations by examining the distribution of ROH and IBD segments. Consanguineous unions increase the likelihood of identical genomic segments being paired within individuals, leading to the formation of ROHs (Severson et al. 2019). The presence of both short and long ROHs within Roma indicates historically high levels of inbreeding (understood as mating between relatives, (Ceballos et al. 2018), consistent with previous research (Mendizabal et al. 2012; Bianco et al. 2020; Font-Porterias et al. 2021). This is further supported by examining IBD segments, where the abundance of shared segments of any length between Roma individuals confirms the presence of a historical pattern of isolation. Specifically, the high number of segments under 8 cM of length suggests that endogamy likely peaked in a period that goes from ~ 1000 to ~ 500 years ago, corresponding to the lowest historical N_e and the estimated older dates of admixture inferred by fastGLOBETROTTER.

Assortative mating is a form of sexual selection that shapes the characteristics and reproduction of communities. Genetic evidence now shows for the first time that Roma are subject to this process, as evidenced by ANCESTOR results showing that Roma tend to choose partners with similar ancestry proportions. With their distinctive ancestry profiles, this means that Roma partners are often chosen from within the same community, which contributes to higher levels of endogamy. This mating pattern is reflected in the positive correlation between European and South-Asian ancestries across all groups. These findings align with cultural studies (Weyrauch 2001; Drummond 2011; Gamella and Álvarez-Roldán 2023) and highlight the significant role of non-random mating on the genetic profile of Roma communities.

The availability of extensive data from Iberian individuals enables a detailed analysis of the Roma's microgeographical genomic structure. Our findings reveal a distinct ancestry profile and genetic substructure between Iberian Roma and non-Roma individuals. The AMOVA indicates that Iberian Roma groups exhibit a level of heterogeneity and geographical structure ten times greater than that of non-Roma groups. Despite this genetic substructure, the allele-sharing between Iberian Roma and non-Roma groups does not demonstrate significant within-region genetic similarity. For example, Roma from Eastern Iberia do not exhibit more allele-sharing with non-Roma from the same region than with other non-Roma Iberians, suggesting that inter-group mating has occurred randomly without a clear regional pattern.

We also explored the presence of alternative ancestries with the Iberian Roma, considering their cohabitation with Christians, Jews, and Muslim North Africans from their arrival on the Iberian Peninsula in the 15th century until the expulsion of Jews and Muslim North Africans (Pérez and Hochroth 2007; Carr 2017). F-statistics indicate that non-Roma groups are genetically closer to Jews and North Africans than to the Roma, with the sole exception of Cochin Jews (from southern India), who share closer ties with the Roma due to their Indian ancestry. A small fraction of North African ancestry is present in both Iberian Roma (2–3%) and non-Roma (11%) but is absent in other Roma groups, suggesting that this North African component in Iberian Roma likely results from admixture with Iberian non-Roma after their arrival on the Peninsula. Similarly, the Jewish ancestry found in Iberian Roma (1%) likely entered their gene pool through admixture with Iberian populations (being 2% in IBS; Supplementary Fig. 26a-b) after their arrival. These results indicate minimal historical interaction between Roma, Muslims, and Jews in the Iberian Peninsula, challenging hypotheses of Jewish descent or North African migration in Roma origins (Aparicio Gervás 2006; Hancock 2006; Pohoryles 2018). Our findings support a South Asian origin for the Roma, followed by a diaspora across western Asia before reaching the Balkans and eventually dispersing across Europe until their arrival on the Iberian Peninsula, consistent with previous studies (Mendizabal et al. 2011; Martínez-Cruz et al. 2016; Bianco et al. 2020).

This study is the first to provide a genetic characterisation of the Roma, confirming their Indian origins, Central European migration routes, and revealing a genetic substructure marked by varying levels of non-Roma admixture, shaped by complex historical events and evolving social norms. Despite challenges such as uneven sample sizes and the absence of key reference groups, our findings suggest that the Roma are a highly admixed and evolving community, shaped by centuries of migrations and interactions with other populations, and highlight the need for further research, including whole-genome sequencing and studies of underrepresented Roma communities.

Materials and methods

Sampling, dataset assembly, and quality control

The present study is based on genome-wide data of 105 self-reported Spanish Roma, 5 Portuguese Roma (Font-Porterias et al. 2019, 2021), and 42 Roma individuals sampled in the Czech Republic (see Data Availability), all genotyped using the Affymetrix Axiom Genome-Wide Human Origins 1 array, which includes 630k genome-wide SNPs. Additionally, 52 Czech non-Roma controls were genotyped with the same array for this study. Beyond the newly generated array data, we included previously published sequence data from 29 Roma individuals from other European regions (North

Macedonia, Lithuania, Hungary, Ukraine) from Bianco et al. 2020. As the Roma populations exhibit appreciable population substructure, the inclusion of Czech and Eastern European Roma alongside Iberian Roma aimed to enhance the representation of Roma genetic diversity across Europe, allowing for a comparative analysis of genetic differentiation and admixture patterns between Western and Central European Roma groups.

For the genotyped Czech samples and the complete genomes in the Bianco et al. (Bianco et al. 2020) dataset, the variant calling using GATK (McKenna et al. 2010) was performed following the GATK Best Practices pipeline (Auwera et al. 2013). Sequencing reads were aligned to the GRCh37 human reference genome using BWA-MEM v 0.7.12 (Li 2013), removing PCR duplicates with Picard Tools v 2.18.6 (Picard Toolkit 2019), and performing base quality recalibration using GATK's BaseRecalibrator (BQSR). We performed joint genotype calling (HaplotypeCaller and GenotypeGVCFs), and we used GATK DepthOfCoverage program (--minBaseQuality 20 --minMappingQuality 20) to estimate sequencing depth. The final call set consists of 613,535 autosomal SNPs in the Czech dataset and 28,678,694 autosomal SNPs in the Bianco dataset.

Besides the Roma samples described above, reference samples from previously published whole genome sequences and genome-wide data from Europe, Caucasus, Central Africa, Central and Southern Asia (Lazaridis et al. 2014; Auton et al. 2015), as well as North Africa (Patterson et al. 2012; Lucas-Sánchez et al. 2023), were included in the analyses. Following the approach used in Font-Porterias et al. (2019), populations from the reference dataset were standardised to include 25 individuals to mitigate potential biases caused by highly imbalanced sample sizes while maintaining optimal representation. However, the non-Roma Spanish population (IBS) and Punjabi from Lahore (PJL) from the 1000 Genomes Project were excluded from this standardisation due to their pivotal relevance in the analysis of Roma. Using PLINK 2 (Chang et al. 2015), SNPs with a missingness rate higher than 5% and with a minor allele frequency below 0.05 were removed, and individuals with more than 10% of missing calls were removed. We used vcftools (Danecek et al. 2011) to extract biallelic SNPs and to check for related individuals using the included relatedness2 method; we then removed individuals related above the third degree of relatedness (values of kinship coefficient between 0.0442 and 0.5). The final dataset includes 335,367 autosomal SNPs in 1,189 individuals after filtering for QC (Supplementary Table 13), including 95 Iberian Roma and 21 Czech Roma unrelated individuals.

In our analyses, we refer to the combined Spanish and Portuguese Roma samples as the Iberian Roma. For some specific analyses we selected a subset of individuals containing exclusively Iberian Roma and IBS populations clustered by geographic regions (henceforth called Iberian dataset). The Iberian Roma were classified into five geographical regions (Centre, North, South, West, and East; Supplementary Fig. 27) based on their parents and grandparents' birthplaces, similar to the approach used in Aizpurua-Iraola et al. (Aizpurua-Iraola et al. 2022). In the cases where no information about the geographical origin of the volunteer and volunteer ancestors was available, or the cases where ancestors originated in different geographical areas, they were classified as *'Indeterminate'*. For the *IBS* population from the 1000GP dataset our classification strategy consisted in assigning each individual to the respective region of their sampling location. Due to the lack of detailed information about birthplace or family origins for the Roma from Central and Eastern Europe, they were assigned to their respective countries based solely on the sampling location.

(distribution of the Roma samples in Fig. 1a).

Principal component analysis (PCA) and population structure

Data was pruned for linkage disequilibrium using PLINK v2, removing SNPs with an $r^2 > 0.5$ in a sliding window of 200 SNPs, and steps of 25 SNPs. The pruned dataset contains 156,153 SNPs, which were then used to perform a PCA with SmartPCA from the EIG v6.0.1 software package (Price et al. 2006; Patterson et al. 2006) on the complete and the Iberian datasets.

To infer the global ancestry components in the Roma populations, ADMIXTURE v1.3.0 (Alexander and Lange 2011) was used with 10 iterations of cross-validation error calculation and they were plotted using PONG v1.4.9.

To estimate within-population and between-population genetic variation, an analysis of the molecular variance (AMOVA) was conducted, and diversity indices were calculated using the Poppr R package (Kamvar et al. 2014; Behr et al. 2016) on the Iberian dataset.

Patterns of allele sharing

AdmixTools2 (Maier et al. 2023) was used to estimate allele-sharing between the Roma populations and putative parental source populations from Europe, Asia, and Africa. To assess evidence of admixture in Roma populations, we conducted *f3* tests in the form of *f3*(Roma; Source1, Source2). To investigate the shared drift between the Roma and other populations from a common outgroup, we employed the three-population test (*f3*) using the entire dataset. This was done through the outgroup *f3*-statistic: *f3* (YRI; Roma, X), where YRI represents Yoruba individuals from the 1000GP

(Auton et al. 2015), Roma represents any Roma population in the dataset, and X denotes any other population.

To infer more specific patterns of allele-sharing, *f4* statistics were computed in the following way: (i) *f4*(Yoruba; Roma, Source1; Source2). The source populations used in the tests were selected as follows: (i) pairs of European reference populations; (ii) pairs of Asian reference populations; and (iii) pairs of North African and Jewish reference populations. In addition, we estimated *f4*-statistics using the Iberian dataset, to infer if patterns of allele-sharing were geographically-related between Roma and non-Roma from the Iberian Peninsula, in the way *f4*(Yoruba; Roma, IBS; IBS). To avoid biases, in the *f3* outgroup test on the complete dataset and the *f4* tests with European references only the 25 IBS individuals with the lowest North African ancestry component were used, measured according to a Local Ancestry Inference (LAI) (Supplementary Table 14).

Phasing

The haplotype phase was inferred for each chromosome using default settings with SHAPEIT v2 (Delaneau et al. 2013), using the HapMap GRCh37 genetic map (Gibbs et al. 2003) and 1000 Genomes (Phase 3) as a reference panel (Auton et al. 2015). In a pre-phasing step carried out (--check) to remove any SNPs that did not align correctly to the reference no SNP was removed. Specifically, for the identity-by-descent (IBD) analyses and IBD-based estimation of effective population size, we phased autosomal data using Beagle v 5.3 (Browning et al. 2021), following the pipeline proposed by Browning et al. (Browning et al. 2018).

Haplotype-sharing analysis

To detect fine-scale population structure, we performed ChromoPainter/fineSTRUCTURE analysis (Lawson et al. 2012; Hellenthal et al. 2014) on the complete dataset, first with the Romani populations as target, then with the *IBS* population as target, to provide a baseline of comparison for the results. We used the ChromoPainter v2 (Lawson et al. 2012) expectation-maximisation (EM) algorithm to obtain the global mutation probability (M) and the switch rate (n) parameters by setting the -in -iM switches. This analysis was performed on chromosomes 1, 7, 14, and 20, with 15 iterations of the EM algorithm, and the -a switch was used to parallelise the analysis for each subset of 20 individuals, from a total of 1,189. The resulting parameters were averaged across chromosomes, weighted by the number of SNPs per chromosome, to obtain the final values of $n = 252.38$ and $M = 0.0006$. Next, we used ChromoPainter with the previously estimated parameters, to infer number and length of shared haplotypes between every pair of individuals in the dataset, producing the coancestry matrix. ChromoCombine was then used to combine the results over all the chromosomes.

FineSTRUCTURE v4.1.1 (Hellenthal et al. 2014) was used on the coancestry matrix to group the data and determine genetic clusters based on patterns of haplotype sharing. The analysis was conducted for 2 million iterations using Markov Chain Monte Carlo, with 1 million iterations designated as “burn-in,” and sampling of values was performed every 10,000 iterations. For the complete dataset analysis, FineSTRUCTURE was run with the normalisation parameter “c” estimated as 0.262. FineSTRUCTURE dendrograms were built with the default parameter -m T.

The analysis was performed for three different seeds in the chunkcounts and chunklengths sharing coancestry matrices from ChromoPainter, and the consistency of the dendrograms between different seeds was manually evaluated. The three chunkcounts dendrograms were used as references to assign individuals to genetic clusters.

Following that, a third ChromoPainter analysis was run (Donors vs. Target), setting the -f switch to state which individuals belong to each cluster and which genetic clusters would be donor groups. In the population file, all the clusters were set both as donors (D) and recipients (R), except for the Roma clusters, which were only set as recipients. A modified version of the NNLS method implemented in the *nnls* v1.4 R package (Katharine M. Mullen and Ivo H. M. van Stokkum 2012) was used to infer the ancestry profiles of the clusters using the ChromoPainter coancestry matrix. The Donor clusters were combined into macroregions defined according to geographical criteria (i.e. European, Middle Eastern) for ease of representation.

Using the NNLS ancestry component percentages, Pearson’s correlation tests were performed to explore the potential presence of a geographic cline of variation in the ancestral components of Roma individuals (see Supplementary Note 3).

Dating admixture events

The fastGLOBETROTTER method (Wangkumhang et al. 2022) was used to infer and date admixture events in our sample sets. For our analysis, we ran fastGLOBETROTTER to detect admixture events in Europe and in the Iberian Peninsula. All the donor clusters were used as surrogate populations to represent the admixing sources in the analyses. We applied ChromoPainter v2 to paint Roma individuals using the surrogate population only as donors and inferred copying vectors for the target individuals.

fastGLOBETROTTER was run for 5 mixing iterations using the prop.ind:1 and null.ind:1 settings. After reviewing the results and evaluating clusters with multiple or single-date admixture, we conducted an additional run with fastGLOBETROTTER using 100 bootstrap resampling iterations to estimate the admixture dates, following the developer's approach for measuring 95% confidence intervals (CI) around inferred dates (Wangkumhang et al. 2022). The mean admixture date was calculated as the average across bootstrap estimates for each event, with CIs and Standard Errors derived from the bootstraps obtained using the standard model (null.ind: 1). A generation time of 25 years was assumed, as done in previous comparable studies (Martínez-Cruz et al. 2016; Font-Porterias et al. 2019; Vilà-Valls et al. 2023). The sources of each admixture event were combined into the previously defined macro-regions for ease of visualisation.

As an alternative estimate of the admixture events, MALDER (Pickrell et al. 2014) was run on the non-pruned and unphased dataset containing individuals from the Iberian-Roma and EuropeanRoma clusters. We used a set of reference clusters that reflected the diaspora route, with the addition of Basque for the IberianRoma test. The dataset was first converted from plink to Eigenstrat format using the *convertf* function from EIG v 6.0.1 (Patterson et al. 2006). Then, MALDER was run with a minimum distance of 0.005 cM and jackknife resampling. Similar to the fastGLOBETROTTER results, we converted time from admixture in generations to a date by assuming a generation time of 25 years.

Homozygosity and identity-by-descent fragment Estimation

Runs of homozygosity (ROH) were identified using PLINK v2, considering runs with at least 50 SNPs, a minimum length of 500 kb, and a maximum gap of 100 kb between two consecutive SNPs. The ROH lengths were categorised into four length categories: very short (<1 Mb), short (1 to 2.5 Mb), medium (2.5 to 5 Mb), and long (>5 Mb).

Using the dataset phased with Beagle, we identified IBD segments between pairs of individuals. To call IBD blocks, we used Hap-Ibd (Zhou et al. 2020) on the main dataset, utilising the HapMap GRCh37 genetic map to convert from base pairs to genetic positions in centiMorgans (cM). We excluded any segment under 3 cM in length to limit overestimations. After the estimation, we merged the segments using the provided tool (merge-ibd-segments.17Jan20.102.jar) from the Browning et al. (Browning et al. 2018) pipeline. To construct an IBD heatmap, we summed the IBD pairwise lengths between individuals following Han et al. (Han et al. 2017) and used the heatmaply package to plot the results in R (Galili et al. 2018). To explore variations in endogamy and gene flow between populations over time, we assessed the within- and between-population IBD sharing patterns in Roma and non-Roma European populations (see Supplementary Note 1).

Effective population size (*Ne*) Estimation

We further explored the demographic history of the Roma population using the measured IBD segments to estimate historical changes in effective population size (Ne) using IBDNe (Browning and Browning 2015). The analysis utilised the HapMap GRCh37 genetic map as a reference. We set a minimum threshold to exclude segments of IBD shorter than 3 cM. To generate a 95% CI for the results, we employed a bootstrap approach with 100 simulations.

Initially, the analysis was performed on all Roma groups and then on the IBS, Czech, and PJL reference populations. As this method is most reliable for recent periods when applied to genome-wide array data (Browning and Browning 2015), we filtered the data to retain only information from the present up to 50 generations ago. The 50-generation cutoff was chosen based on Browning and Browning (2015), which demonstrated that SNP array data provides reliable estimates of effective population size within this timeframe. Beyond 50 generations, increased uncertainty in IBD-segment detection leads to underestimation of effective population size, making the method less accurate.

Tests for ancestry-assortative mating

We conducted tests for ancestry-assortative mating using ANCESTOR. ANCESTOR is an algorithm that leverages phased local ancestry tracts to estimate the ancestral proportions of both parents for each individual. A RFMix 2 (Maples et al. 2013; Zou et al. 2015a, b) analysis was performed to generate the LAI used by ANCESTOR, using a two-group reference set consisting of European and South-Asian populations. To convert RFMix output to the required inputs we used the code from Korunes et al. (Korunes et al. 2022). The European reference group comprised 95 individuals from IBS and Balkan populations (Croatian, Greek, Hungarian, Romanian), while the South-Asian group was formed by 95 Punjabi (PJL) individuals. Using the inferred parental ancestries, we tested for assortative mating, which is indicated by a positive correlation in ancestry between the inferred mating pairs.

Local ancestry tracts size distribution

We aimed to infer the ancestry tract size distribution to compare the observed patterns for various Roma populations, with the goal of determining which groups exhibit early and later admixture with non-Roma European populations (or, at least, when this admixture was most intense). To do this, first the local ancestry inference was performed using RFMix v2 software, using two reference populations, European and South-Asian (see Supplementary Note 4). Then we identified the ancestry tracts and their length, using an in-house developed tool (details in Supplementary Note 2). Lastly, the length tract distributions of different Roma populations were compared, pooling all tracts together and then splitting them in length size categories.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00439-025-02735-z.

Acknowledgements We would like to thank the Roma FAGiC association (Federació d'Associacions Gitanes de Catalunya) and all its members and associates for its continuous support and collaboration. We also thank the Spanish National Genotyping Center (CeGen) for their help with the genotyping.

Author contributions DC contributed to the design and conception of the study. GF-E performed and implemented the data analysis. Data collection were performed by AG, AC-M, PL. The first draft of the manuscript was written by GF-E and all authors commented on previous versions of the manuscript. All authors approved the submitted version.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.
This work was supported by the Spanish Ministry of Science and Innovation and by "ERF A way of making Europe" (grant numbers CGL2019-106485GB-I00 and PID2022-138755NB-I00) funded by the MCIN and the AEI https://doi.org/10.13039/501100011033. PL was supported by Iceland, Liechtenstein and Norway governments through the EEA Grants ZD-ZDOVA2-001. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability The genomic data of Czech Roma and non-Roma individuals analysed during the current study is available in the EGA repository, under accession number EGAD50000001103.
The scripts generated during this study can be downloaded from https://github.com/gfena/AncestryLength

Declarations

Ethical approval The study has been approved by the CEIm-PSMAR IRB in Barcelona (2019/8900/I, 2022/10542/I, and 2023/10830 protocols) and the Ethics Committee of General University Hospital in Prague (reference no. 65/16 and 34/19); and it has been developed under the umbrella of the project "*El Camí del Poble Gitano: una història de diversitat*" in collaboration with the FAGiC (Giménez et al. 2019). Preliminary results were presented to the local Roma community in a meeting at the UPF in Barcelona on May 26th 2023 and retransmitted in streaming (Universitat Pompeu Fabra 2023). All the procedures were carried out in adherence to the standard guidelines and regulations outlined in the Declaration of Helsinki (General Assembly of the World Medical Association 2014).

Consent to participate For the novel data, DNA was collected from volunteers who self-identified as Romani and Czech, with all participants providing written informed consent.

Competing interests The authors declare no competing interests.

References

Abreu L (2007) Beggars, vagrants and romanies: repression and persecution in Portuguese society (14th–18th Centuries). Hygiea 6:41–66. https://doi.org/10.3384/hygiea.1403-8668.076141

Achim V (2004) The Roma in Romanian history. Central European University

Aizpurua-Iraola J, Giménez A, Carballo-Mesa A et al (2022) Founder lineages in the Iberian Roma mitogenomes recapitulate the Roma diaspora and show the effects of demographic bottlenecks. Sci Rep 12:18720. https://doi.org/10.1038/s41598-022-23349-9

Alexander DH, Lange K (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry Estimation. BMC Bioinformatics 12:246. https://doi.org/10.1186/1471-2105-12-246

Aparicio Gervás JM (2006) Breve recopilación sobre La historia Del Pueblo Gitano: desde Su Salida Del Punjab, Hasta La constitución Española de 1978. Veinte Hitos sobre La Otra historia de España. RIFOP: revista interuniversitaria de Formación Del Profesorado: continuación de La Antigua. Revista De Escuelas Normales 141–162

Auton A, Abecasis GR, Altshuler DM et al (2015) A global reference for human genetic variation. Nat 2015 526:7571. https://doi.org/10.1038/nature15393

Baharian S, Barakatt M, Gignoux CR et al (2016) The great migration and African-American genomic diversity. PLoS Genet 12:e1006059. https://doi.org/10.1371/journal.pgen.1006059

Bánfai Z, Melegh BI, Sümegi K et al (2019) Revealing the genetic impact of the Ottoman occupation on ethnic groups of East-Central Europe and on the Roma population of the area. Front Genet 10:558

Behr AA, Liu KZ, Liu-Fang G et al (2016) Pong: fast analysis and visualization of latent clusters in population genetic data. Bioinformatics 32:2817–2823. https://doi.org/10.1093/bioinformatics/btw327

Bernát A, Messing V (2016) Methodological and data infrastructure report on Roma population in the EU. In:GRID

Bianco E, Laval G, Font-Porterias N et al (2020) Recent Common Origin, Reduced Population Size, and Marked Admixture Have Shaped European Roma Genomes. Molecular Biology and Evolution. https://doi.org/10.1093/molbev/msaa156

Bover O, Velilla P (1999) Migration in Spain. Historical Background and Current Trends

Brearley M (2001) The persecution of Gypsies in Europe. Am Behav Sci 45:588–599. https://doi.org/10.1177/00027640121957367

Browning SR, Browning BL (2015) Accurate Non-parametric Estimation of recent effective population size from segments of identity by descent. Am J Hum Genet 97:404–418. https://doi.org/10.1016/j.ajhg.2015.07.012

Browning SR, Browning BL, Daviglus ML et al (2018) Ancestry-specific recent effective population size in the Americas. PLoS Genet 14:e1007385. https://doi.org/10.1371/journal.pgen.1007385

Browning BL, Tian X, Zhou Y, Browning SR (2021) Fast two-stage phasing of large-scale sequence data. Am J Hum Genet 108:1880–1890. https://doi.org/10.1016/j.ajhg.2021.08.005

Campbell CL, Palamara PF, Dubrovsky M et al (2012) North African Jewish and non-Jewish populations form distinctive, orthogonal clusters. Proceedings of the National Academy of Sciences 109:13865–13870. https://doi.org/10.1073/pnas.1204840109

Carr M (2017) Blood and faith: the purging of Muslim Spain. Oxford University Press, pp 1492–1614

Castro e Silva MA, Ferraz T, Couto-Silva CM et al (2022) Population histories and genomic diversity of South American natives. Mol Biol Evol 39:msab339. https://doi.org/10.1093/molbev/msab339

Ceballos FC, Joshi PK, Clark DW et al (2018) Runs of homozygosity: windows into population history and trait architecture. Nat Rev Genet 19:220–234. https://doi.org/10.1038/nrg.2017.109

Çelik F (2004) Exploring marginality in the Ottoman Empire: Gypsies or People of Malice (Ehl-i Fesad) as viewed by the Ottomans. EUI WORKING PAPERS:RSCAS No. 2004/39

Chang CC, Chow CC, Tellier LC et al (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4:s13742-015-0047–8. https://doi.org/10.1186/s13742-015-0047-8

Crowe DM (2006) From persecution to pragmatism: the Habsburg Roma in the eighteenth century. Austrian History Yearbook 37:99–120

Crowe D (2016) The Gypsy historical experience in Romania. The Gypsies of Eastern Europe. Routledge, pp 61–79

Crowe D, Kolsti J, Hancock I (2016) The Gypsies of Eastern Europe. Routledge

Danecek P, Auton A, Abecasis G et al (2011) The variant call format and vcftools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Delaneau O, Zagury J-F, Marchini J (2013) Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods 10:5–6. https://doi.org/10.1038/nmeth.2307

der Auwera GAV, Carneiro MO, Hartl C et al (2013) From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protocols Bioinf 43. https://doi.org/10.1002/0471250953.bi1110s43

Drummond S (2011) Mapping marriage law in Spanish Gitano communities. UBC

Ena GF, Aizpurua-Iraola J, Font-Porterias N et al (2022) Population genetics of the European Roma—A. Rev Genes 13:2068. https://doi.org/10.3390/genes13112068

Font-Porterias N, Arauna LR, Poveda A et al (2019) European Roma groups show complex West Eurasian admixture footprints and a common South Asian genetic origin. PLoS Genet 15:e1008417. https://doi.org/10.1371/journal.pgen.1008417

Font-Porterias N, Giménez A, Carballo-Mesa A et al (2021) Admixture has shaped Romani genetic diversity in clinically relevant variants. Front Genet 12:683880. https://doi.org/10.3389/fgene.2021.683880

Fraser AM (1992) The Gypsies (The peoples of Europe). Blackwell Pub

Galili T, O'Callaghan A, Sidi J, Sievert C (2018) Heatmaply: an R package for creating interactive cluster heatmaps for online publishing. Bioinformatics 34:1600–1602. https://doi.org/10.1093/bioinformatics/btx657

Gamella JF, Álvarez-Roldán A (2023) Breaking secular endogamy. The growth of intermarriage among the Gitanos/Calé of Spain (1900–2006). History Family 28:457–483

García-Fernández C, Font-Porterias N, Kučinskas V et al (2020) Sex-biased patterns shaped the genetic history of Roma. Sci Rep 10:14464. https://doi.org/10.1038/s41598-020-71066-y

General Assembly of the World Medical Association (2014) World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. J Am Coll Dent 81:14–18

Gheorghe N (1983) The origin of Roma's slavery in the Romanian principalities. Roma 7:12–27

Gibbs RA, Belmont JW, Hardenbol P et al (2003) The international HapMap project. Nature. https://doi.org/10.1038/nature02168

Giménez A, Comas D, Carballo A (2019) Origen e identidad del Pueblo Gitano. Int J Roma Stud 1:159–184 https://doi.org/10.17583/ijrs.2019.4561

Gómez-Carballa A, Pardo-Seco J, Fachal L et al (2013) Indian signatures in the Westernmost edge of the European Romani diaspora: new insight from mitogenomes. PLoS ONE 8:e75397. https://doi.org/10.1371/journal.pone.0075397

Greenberg J (2010) Report on Roma education today: from slavery to segregation and beyond. Colum L Rev 110:919–1001

Gusmão A, Gusmão L, Gomes V et al (2008) A perspective on the history of the Iberian Gypsies provided by phylogeographic analysis of Y-Chromosome lineages. Ann Hum Genet 72:215–227. https://doi.org/10.1111/j.1469-1809.2007.00421.x

Halwachs D (2005) Roma and Romani in Austria. Romani Stud 15:145–173

Han E, Carbonetto P, Curtis RE et al (2017) Clustering of 770,000 genomes reveals post-colonial population structure of North America. Nat Commun 8:14238

Hancock I (2006) On Romani origins and identity. The Romani Archives and Documentation Center, Aralık

Harris DN, Song W, Shetty AC et al (2018) Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. Proceedings of the National Academy of Sciences 115:E6526–E6535. https://doi.org/10.1073/pnas.1720798115

Hellenthal G, Busby GBJ, Band G et al (2014) A genetic atlas of human admixture history. Science 343:747–751. https://doi.org/10.1126/science.1243518

Henn BM, Botigué LR, Gravel S et al (2012) Genomic ancestry of North Africans supports Back-to-Africa migrations. PLoS Genet 8:e1002397. https://doi.org/10.1371/journal.pgen.1002397

Iovita RP, Schurr TG (2004) Reconstructing the origins and migrations of diasporic populations: the case of the European Gypsies. Am Anthropol 106(2):267–281

Ivanov A, Kagin J (2014) Roma poverty from a human development perspective. Istanbul, Turkey, UNDP, p 96

Kajanova A, Kmecova I (2018) Development of unemployment of the Roma minority in selected European countries. Cswhi 9:59–64. https://doi.org/10.22359/Cswhi_9_4_08

Kalaydjieva L, Gresham D, Calafell F (2001) Genetic studies of the Roma (Gypsies): a review. BMC Med Genet 2:5. https://doi.org/10.1186/1471-2350-2-5

Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281. https://doi.org/10.7717/peerj.281

Katharine M, Mullen, Ivo HM, van Stokkum (2012) The Lawson-Hanson algorithm for non-negative least squares (NNLS). https://citeseerx.ist.psu.edu/document?repid=rep1%26type=pdf%26doi=b84fcc0feb78f074edad012273847ce8fa5361ce. Accessed 1 Jul 2024

Kenrick D (2007) Historical dictionary of the Gypsies (Romanies). Scarecrow

Kenrick D, Puxon G (2009) Gypsies under the swastika. University of Hertfordshire

Kimmel S (2015) Parables of coercion: conversion and knowledge at the end of Islamic Spain. University of Chicago Press

Klamper E (1993) Persecution and annihilation of Roma and Sinti in Austria, 1938–1945. Romani Stud 3:55

Korunes K, KL, Soares-Souza GB, Bobrek et al (2022) Sex-biased admixture and assortative mating shape genetic variation and influence demographic inference in admixed Cabo Verdeans. G3 Genes|Genomes|Genetics 12:jkac183. https://doi.org/10.1093/g3journal/jkac183

La Parra-Casado D, Mosquera PA, Vives-Cases C, San Sebastian M (2018) Socioeconomic inequalities in the use of healthcare services: comparison between the Roma and general populations in Spain. Int J Environ Res Public Health 15:121

Laparra M (2007) Informe sobre La situación social y tendencias de Cambio En La Población gitana. Ministerio de Trabajo y Asuntos Sociales, Una primera aproximación Madrid

Laparra M, Arza J, Fernández A (2007) Diagnóstico social de La Comunidad gitana En España, vol 2011. Un análisis contrastado de la Encuesta del CIS a Hogares de Población Gitana

Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. PLoS Genet 8:e1002453. https://doi.org/10.1371/journal.pgen.1002453

Lazaridis I, Patterson N, Mittnik A et al (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513:409–413. https://doi.org/10.1038/nature13673

Leblon B (1985) Les Gitans d'Espagne: Le prix de la différence. FeniXX

Lewy G (2000) The Nazi persecution of the Gypsies. Oxford University Press

Li H (2013) Aligning sequence Reads, clone. sequences and assembly contigs with BWA-MEM

Lucas-Sánchez M, Fadhlaoui-Zid K, Comas D (2023) The genomic analysis of current-day North African populations reveals the existence of trans-Saharan migrations with different origins and dates. Hum Genet 142:305–320. https://doi.org/10.1007/s00439-022-02503-3

Maier R, Flegontov P, Flegontova O et al (2023) On the limits of fitting complex models of population history to f-statistics. eLife 12:e85492. https://doi.org/10.7554/eLife.85492

Maples BK, Gravel S, Kenny EE, Bustamante CD (2013) RFMix: A discriminative modeling approach for rapid and robust Local-Ancestry inference. Am J Hum Genet 93:278–288. https://doi.org/10.1016/j.ajhg.2013.06.020

Martínez Dhier A (2007) La condición social y jurídica de los gitanos en la legislación histórica española.(A partir de la pragmática de los Reyes Católicos de 1499). Doctoral thesis, Universidad de Granada

Martínez-Cruz B, Mendizabal I, Harmant C et al (2016) Origins, admixture and founder lineages in European Roma. Eur J Hum Genet 24:937–943. https://doi.org/10.1038/ejhg.2015.201

Marushiakova E, Popov V (2001) Gypsies in the Ottoman empire: A contribution to the history of the Balkans, vol 22. University of Hertfordshire

Marushiakova E, Popov V (2010) Gypsy/Roma European migrations from 15th century till nowadays. In: Proceedings of International Conference Romani Mobilities in Europe: Multidisciplinary Perspectives. pp 126–139

Marushiakova-Popova EA, Popov V (2018) Migration vs. inclusion: Roma mobilities from East to West. Baltic Worlds

Matras Y (2002) Romani A linguistic introduction. Cambridge University Press

McKenna A, Hanna M, Banks E et al (2010) The genome analysis toolkit: A mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297. https://doi.org/10.1101/GR.107524.110

Mendes MM, Magano O (2022) Roma/Ciganos and the condition of internal strange in Portuguese society: the construction of otherness. In: Simmel and beyond. Routledge

Mendizabal I, Valente C, Gusmão A et al (2011) Reconstructing the Indian origin and dispersal of the European Roma: A maternal genetic perspective. PLoS ONE 6:1–10. https://doi.org/10.1371/journal.pone.0015988

Mendizabal I, Lao O, Marigorta UM et al (2012) Report reconstructing the population history of European Romani from Genome-wide data. Curr Biol 22:2342–2349. https://doi.org/10.1016/j.cub.2012.10.039

Moorjani P, Patterson N, Loh P-R et al (2013) Reconstructing Roma history from Genome-Wide data. https://doi.org/10.1371/journal.pone.0058633. PLOS One 8:

O'Nions H (2011) Roma expulsions and discrimination: the elephant in Brussels. Eur J Migration Law 13:361–388. https://doi.org/10.1163/157181611X605864

O'Nions H (2016) Minority rights protection in international law: the Roma of Europe. Routledge, London

Ortega MHS (1994) Los Gitanos Españoles desde Su Salida de La India Hasta Los Primeros conflictos En La Península. Espacio Tiempo y Forma Serie IV, Historia Moderna

Pamjav H, Zalán A, Béres J et al (2011) Genetic structure of the paternal lineage of the Roma people. Am J Phys Anthropol 145:21–29. https://doi.org/10.1002/ajpa.21454

Parekh N, Rose T (2011) Health inequalities of the Roma in Europe: a literature review. Cent Eur J Public Health 19:139–142. https://doi.org/10.21101/cejph.a3661

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2:e190. https://doi.org/10.1371/journal.pgen.0020190

Patterson N, Moorjani P, Luo Y et al (2012) Ancient admixture in human history. Genetics 192(3):1065–1093. https://doi.org/10.1534/genetics.112.145037

Pereira V, Gusmão L, Valente C et al (2012) Refining the genetic portrait of Portuguese Roma through X-chromosomal markers. Am J Phys Anthropol 148:389–394. https://doi.org/10.1002/ajpa.22061

Pérez J, Hochroth L (2007) History of a tragedy: the expulsion of the Jews from Spain. University of Illinois Press

Picard toolkit (2019) Picard toolkit. Broad Institute, GitHub repository

Pickrell JK, Patterson N, Loh P-R et al (2014) Ancient West Eurasian ancestry in Southern and Eastern Africa. Proc Natl Acad Sci 111:2632–2637. https://doi.org/10.1073/pnas.1313787111

Pohoryles Y (2018) The Jewish-Romani connection: Are Gypsies descendants of tribe of Simeon? Ynetnews. Retrieved from https://www.ynetnews.com/articles/0,7340,L-5248902,00.html

Price AL, Patterson NJ, Plenge RM et al (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909. https://doi.org/10.1038/ng1847

Pym R (2007) The Gypsies of early modern Spain. Springer

Rai N, Chaubey G, Tamang R et al (2012) The phylogeography of Y-Chromosome haplogroup H1a1a-M82 reveals the likely Indian origin of the European Romani populations. PLoS ONE 7:e48477. https://doi.org/10.1371/journal.pone.0048477

Reich D, Thangaraj K, Patterson N et al (2009) Reconstructing Indian population history. Nature 461:489–494. https://doi.org/10.1038/nature08365

Sánchez DM (2022) Historia Del Pueblo Gitano En España. Los Libros De La Catarata

Severson AL, Carmi S, Rosenberg NA (2019) The effect of consanguinity on Between-Individual Identity-by-Descent sharing. Genetics 212:305–316. https://doi.org/10.1534/genetics.119.302136

Silvestre J (2005) Internal migrations in Spain, 1877–1930. Eur Rev Econ Hist 9:233–265. https://doi.org/10.1017/S1361491605001462

Soyer F (2007) The Persecution of the Jews and Muslims of Portugal: King Manuel I and the End of Religious Tolerance (1496-7). BRILL

Tarnovschi D, Preoteasa AM, Pamporov A et al (2012) Roma from Romania, Bulgaria, Italy and Spain between social inclusion and migration. Soros Foundation Romania, Bucharest

Tartakoff P (2012) Between Christian and Jew: conversion and inquisition in the crown of Aragon. University of Pennsylvania, pp 1250–1391

Universitat Pompeu Fabra (2023) Presentació del projecte El camí del Poble Gitano: una història de diversitat. Available: https://www.youtube.com/watch?v=JRP9So3bnwc

Vilà-Valls L, Aizpurua-Iraola J, Casinge S et al (2023) Genomic insights into the population history of the Resande or Swedish travelers. Genome Biol Evol 15:evad006. https://doi.org/10.1093/gbe/evad006

Waldman YY, Biddanda A, Dubrovsky M et al (2016) The genetic history of Cochin Jews from India. Hum Genet 135:1127–1143. https://doi.org/10.1007/s00439-016-1698-y

Wangkumhang P, Greenfield M, Hellenthal G (2022) An efficient method to identify, date, and describe admixture events using haplotype information. Genome Res 32:1553–1564. https://doi.org/10.1101/gr.275994.121

Weyrauch WO (2001) Gypsy law: Romani legal traditions and culture. University of California Press

Zahra T (2017) Condemned to rootlessness and unable to Budge: Roma, migration panics, and internment in the Habsburg empire. Am Hist Rev 122:702–726

Zhou Y, Browning SR, Browning BL (2020) A fast and simple method for detecting Identity-by-Descent segments in Large-Scale data. Am J Hum Genet 106:426–437. https://doi.org/10.1016/j.ajhg.2020.02.010

Zou JY, Halperin E, Burchard E, Sankararaman S (2015a) Inferring parental genomic ancestries using pooled semi-Markov processes. Bioinformatics 31:i190–i196. https://doi.org/10.1093/bioinformatics/btv239

Zou JY, Park DS, Burchard EG et al (2015b) Genetic and socioeconomic study of mate choice in Latinos reveals novel assortment patterns. Proceedings of the National Academy of Sciences 112:13621–13626. https://doi.org/10.1073/pnas.1501741112

3.2 Demographic insights into paternal genetic diversity and regional substructure in the Spanish Roma

Giacomo Francesco Ena, Aaron Giménez, Annabel Carballo-Mesa, Marcos Araújo Castro e Silva & David Comas.

Under Revision in BMC Genomics.

Ena, G. F., Giménez, A., Carballo-Mesa, A., Araújo Castro e. Silva, M., & Comas, D. (2025). Demographic insights into paternal genetic diversity and regional substructure in the Spanish Roma. Under Revision in BMC Genomics.

Demographic insights into paternal genetic diversity and regional substructure in the Spanish Roma

Giacomo F. Ena[1], Aaron Giménez[2], Annabel Carballo-Mesa[3], Marcos Araújo Castro e Silva[1], David Comas[1,*]

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

[2] Facultat de Sociologia, Universitat Autònoma de Barcelona, Barcelona, Spain

[3] Facultat de Geografia i Història, Universitat de Barcelona, Barcelona, Spain

[*]Corresponding author

david.comas@upf.edu

Abstract

Background: The Iberian (Calé) Roma constitute one of the largest Roma communities in Europe, yet their internal genetic structure and connections to other Roma groups remain understudied. This study explores the microgeographical structure of the Iberian Roma and their relationships with other Roma groups by analysing paternal lineages using 17 Y-chromosome short tandem repeat markers in a geographically stratified sample of 173 Spanish Roma individuals.

Results: The haplogroup distribution patterns indicate that the paternal genetic profile of the Spanish Roma is shaped by founder effects, population bottlenecks, and multiple admixture events with non-Roma groups. Haplogroups H and J2a1b dominate the genetic landscape, reflecting their South Asian origin and subsequent dispersal patterns through West Asia into Europe. A distinctive feature of the Spanish Roma is the high frequency of haplogroup R1b, indicating significant gene flow from non-Roma Iberian populations. The absence of North African or Jewish genetic influences rules out the possibility of a North African migration route for the Calé Roma into the Iberian Peninsula. Microgeographical analyses (AMOVA) reveal substantial genetic substructure among Calé Roma across Spanish regions, consistent with historical isolation and localised gene flow. Additionally, a striking sex-biased admixture is observed when comparing the current results with previous mitochondrial DNA (mtDNA) data, with paternal South Asian ancestry being twice as high as maternal contributions, suggesting that Roma communities have historically been more inclined to integrate non-Roma women.

Conclusions: The genetic landscape of the Iberian Roma is shaped by a complex history of founder effects, admixture, and isolation. The observed genetic substructure and sex-biased admixture reflect historical social dynamics. These results contribute to the broader understanding of Roma genetic diversity and demography in Spain and underscore the importance of integrating Y chromosome, autosomal, and mtDNA data in future studies.

Keywords

Iberian Roma, Population genetics, Migration, Demography, Y-chromosome

Introduction

The Romani are the largest transnational ethnic minority in Europe. Although traditionally nomadic or semi-nomadic, most of them have been settled for centuries and generally share a common identity and cultural traditions [1–3]. The history of Roma has been reconstructed through linguistics, historical records, and anthropological studies, which combined with genetic research, traced their origins in the northwestern region of India and southeastern Pakistan [4–8]. The Roma have a complex social structure that transcends the concept of national identities, characterised by a tradition of living in relatively closed social groups [1, 3], and using distinct varieties of the Romani language alongside local languages of the countries in which they reside [5, 9]. Romani language is classified within the Indo-Aryan branch of the Indo-European language family, and its dialects—despite regional variation—share a common origin [9, 10]. Linguistic influences acquired during the diaspora—such as Farsi loanwords from early stages of migration and Greek loanwords from the Byzantine period—reflect key historical contact zones [11, 12]. As such, linguistic research has provided an independent line of evidence for tracing Roma migrations, supporting their origins in Northern India [13, 14].

The Roma community is widespread across Europe, with one of the largest groups, the Calé Romani in Iberia (Spain and Portugal), where some members still speak the endangered Caló language [15, 16]. Historical records indicate that the Roma arrived in the various kingdoms of what is now Spain in 1425 [17] and in Portugal likely earlier than 1521 [16], the year of their first documented mention, having migrated from Eastern and

Central Europe [17, 18]. This marks six centuries of Roma presence in Spain, a history that has contributed to the country's cultural and social development. While popular myths and some scholars have hypothesised that the Iberian Roma may have migrated from North Africa [4, 19], or share close genetic connections with Jewish groups [20, 21], previous studies have found no genetic evidence to support these claims [22, 23]. In the fragmented political landscape of the time, they sought safe conducts by presenting themselves as pilgrims traveling to Santiago de Compostela, which most rulers initially granted [24]. The arrival of the Roma in present-day Spain coincided with a period of social and political turmoil, characterised by the expulsion of Jews and Muslims by the Christian kingdoms. Although initially tolerated, the Roma soon encountered increasing restrictions and persecution, with policies aimed at their forced settlement or expulsion [18, 24, 25]. In the 19th and 20th centuries, internal migrations further displaced Roma populations within Spain, potentially increasing admixture among groups [24]. Despite their centuries-long presence in Europe and some recent improvements in their conditions, the Roma have historically faced varying degrees of social marginalisation, both from the broader population and through systemic discrimination [25–27].

Over the past 25 years, genetic research on Spanish Roma has often focused on medical issues [28–31]. However, several studies on uniparental DNA have included Spanish Roma individuals, with two specifically examining Spanish mtDNA [32, 33], collectively revealing evidence of bottlenecks, identifying founder lineages, and uncovering sex-biased admixture patterns [34–37]. In addition, two studies focusing on Roma autosomal data have also included Spanish Roma samples [7, 38], while the most recent and comprehensive research to date consists of an in-depth whole-genome array study on Iberian Roma [22]. These studies provide extensive insights into their genetic diversity, relationships with other populations, population history, and the impact of socio-cultural practices on genetic variation and population structure. However, Y-chromosome genetic diversity remains largely understudied, with only one study examining paternal lineages in Portugal [23], and another including a limited sample size of Spanish Roma that did not specifically focus on this group [34], and none dedicated solely to the Spanish Roma. This leaves a critical gap in understanding the paternal genetic history of the Spanish Roma, which is essential for a more comprehensive view of their genetic diversity, admixture patterns, and historical migration. Y-DNA studies, in particular, offer unique insights by revealing sex-biased gene flow and the presence of founding lineages—elements not captured by autosomal DNA analysis. These features are crucial for understanding male-mediated gene flow and its impact on the Roma's genetic profile and demographic history.

To address these gaps, this community-driven initiative conducted in collaboration with FAGiC (Federation of Roma Associations of Catalonia), analysed up to 27 Y-DNA Short Tandem Repeats (STRs) data from 173 Spanish Roma volunteers, including 133 newly genotyped individuals. The primary aims of this study were to: i) evaluate Spanish Roma Y-chromosome diversity and whether their patrilineal lineages are geographically substructured; ii) evaluate the paternal genetic relatedness with other Roma and non-Roma populations from Europe, investigating potential influences from North Africa and Jewish groups; iii) infer the population history of the Roma patrilineal lineages through Europe and the Iberian Peninsula. This comprehensive study provides a detailed picture of the patrilineal genetic diversity of the Iberian Roma, offering insights from both micro-geographical and broader perspectives on European Roma history and demography.

Materials and Methods

Samples

We genotyped Y-STRs from 133 Spanish Roma volunteers using saliva samples. The collection of the samples was conducted under the umbrella of the 'El Camí del Poble Gitano: una història de diversitat' project [39], in collaboration with the Roma FAGiC association (*Federació d'Associacions Gitanes de Catalunya*). Participants were selected based on their self-identified Roma ancestry, with recruitment facilitated by FAGiC, which helped identify volunteers from the Spanish Roma community. To maximise sample inclusion, we used two different Y-STR kits: the Yfiler® Plus PCR Amplification Kit (27 markers) and the AmpFlSTR® Yfiler® PCR amplification kit (17 markers), based on the availability of samples and kits at the time of genotyping.

Y-STR genotyping

A total of 64 Spanish Roma samples were typed for the 27 Y-STR loci included in the Yfiler® Plus PCR Amplification Kit (Applied Biosystems/Thermo Fisher Scientific) and additional 69 Spanish Roma samples were typed for the 17 Y-STR loci included in the AmpF*l*STR® Yfiler® PCR amplification kit (Applied Biosystems, Inc.). Resulting amplicons were separated on an ABI 3730 XL Genetic Analyzer using ABI GeneScan 600 LIZ as an internal size standard, and fragment lengths were estimated by GeneMapper v4.1 [40]. Y-STR alleles were assigned by comparison with an allelic ladder provided by the manufacturer. Allelic nomenclature follows the recommendations of the International Society for Forensic Genetics (ISFG) [41].

After genotyping, 133 Spanish Roma samples were included in the dataset, while the inclusion of 40 Spanish Roma samples from Martinez-Cruz et al. [34] brings the final number to 173. For regional-scale analyses, the Spanish

Roma and non-Roma populations (references in Table S1) were grouped into five Iberian geographical regions (Centre, North, South, West, and East) based on their sampling location, following the approach used in Aizpurua-Iraola et al. [33] and Ena et al. [22] (sample distribution in Figure S1).

A total of 50 reference populations (10,307 individuals; Table S1) with Y-STR frequency data were included for comparisons with the Spanish Roma population, 11 of which are Roma groups from different European countries. All collections, biogeographical origins, reference publications, and total number of individuals analysed are listed in Supplementary Table 1.

To ensure compatibility with a broader range of reference populations, analyses were conducted using only the 17 Y-STR markers included in the AmpF*l*STR® Yfiler® PCR amplification kit. Y-chromosomal haplogroup prediction was primarily conducted using Whit Athey's Haplogroup Predictor v5 (27-Haplogroups version) based on allele frequencies from 17 Y-STR loci. However, where additional Y-STR loci were available, predictions incorporated these additional markers to enhance accuracy. Whit Athey's Haplogroup Predictor is based on the Bayesian-allele-frequency algorithm [42]; for our analysis we set the fitness score to 0, the Bayesian probability to 85%, and applied equal priors. In the case of intermediate alleles, repeat numbers were rounded to the nearest integer; missing alleles were coded as '99' in input files and considered as missing data, as performed in previous studies [43]. For the prediction, DYS389II is represented as the sum of the two parts of this marker.

Statistical Analysis

Haplotype diversity (HD) of the Spanish Roma population samples was assessed using Nei's HD formula [44] and calculated with R software [45]. Haplotype frequencies were determined by direct counting. We then calculated the confidence interval using the bootstrap method with 10,000 iterations, implemented via the 'boot' package in R [46]. The genotypic data for Y-STR in 133 novel individuals are presented in Supplementary Table S2. A permutation test was performed to compare the number of distinct haplogroups in Spanish Roma with those in other Roma populations, based on 10,000 permutations.

Population pairwise genetic distances (Slatkin R_{ST}) [47] were calculated using Arlequin version 3.5.1.2 [48], after converting raw data to the arp format via a custom script. The analysis used 17-locus haplotypes from individuals belonging to haplogroups H, J2a1b, R1b, and I2a(x) (defined as the grouping of I2a, I2a(xI2a1), and I2a1). The statistical significance of the R_{ST} values generated, based on a stepwise mutation model, was ascertained through permutation tests (10,000 iterations). The migration rate (M) matrix was computed for all the Spanish individuals using Arlequin. A multidimensional scaling (MDS) analysis was performed using the metaMDS function from the vegan package [49], and a plot was created using 'ggplot2' to visualise the genetic distances among the populations examined, based on the R_{ST} pairwise matrix. The patterns of genetic differentiation were further assessed through an analysis of molecular variance (AMOVA) conducted in Arlequin.

Median-joining networks

Y-STR haplotypes of individuals belonging to the J2a1b, H, R1b and I2a(x) haplogroups were used to generate Median-Joining networks using NETWORK 10.2.0.0 (www.fluxus-engineering.com). The networks were generated using the median-joining algorithm, with the weight of each STR locus assigned a value from 1 to 10, inversely proportional to the STR variance, following references [50, 51]. The Maximum Parsimony (MP) option was employed to infer the simplest topology with a good fit to the data. In the case of intermediate alleles, repeat numbers were rounded to the nearest integer, following the approach of previous studies [52]. For calculating networks, we excluded the constitutively duplicated loci (385a/b), as indicated by the Network User Guide, while we retained DYS389I/II after subtracting the number of repetitions in DYS389I from DYS389II. Any missing data or deleted alleles were replaced with the standard code '99' in the input files. To enhance interpretability of the analysis and address computational challenges caused by the large number of samples, we applied random sampling to limit the reference sample size to 20 in H, R1b and I2a(x) networks.

Time estimates

Y-STR haplotypes were used to estimate the time to the most recent common ancestor (TMRCA) of the H haplogroup, along with the J2a1b, R1b, and I2a(x) sub-haplogroups prevalent among Spanish Roma. To achieve this, the rho statistic (ρ) and weighted rho (ρW) were computed using a modified version of the weighted rho method [53], adjusting the mutation rates to fit our input data. We first applied the pedigree mutation rates for each STR obtained from the Y-Chromosome STR Haplotype Database (YHRD, www.yhrd.org). In addition, we applied the rho method using a median pedigree-based mutation rate of 2.5×10^{-3}, as described by Goedbloed et al. [54] and implemented by Pamjav et al. [55]. The DYS385 marker was excluded from the calculations. The statistical significance of differences in time estimates was evaluated by comparing the standard deviation of the dates.

Migration rates in the Roma population

We used MIGRATE version 5.0.6 [56], a software based on coalescent theory that applies Bayesian inference to jointly estimate all parameters of a demographic model, to infer migration patterns in the Roma population from Spain and other relevant Roma populations across Europe. The following parameters were found to provide the models with the highest likelihood after a series of exploratory analysis: one single long chain was run in three independent replicates with a sampling increment of 500 and 2,000 recorded steps, while the number of discarded trees per chain (burn-in) was set to 2,500. Based on the increment value and the number of discarded trees, each sample was visited 3,000,000 times. All models were inferred with uniform priors, using two distinct prior settings: one for effective population size and migration (Min: 0.0, Max: 50.0, Delta: 2.0), and another for divergence and divergence standard deviation (Min: 0.0, Max: 50.0, Delta: 5.0).

Metropolis-Coupled MCMC ("MCMCMC") or "heating" was applied for auxiliary searches with more permissive acceptance criteria. The search was executed with four chains at different temperatures (1.0, 1.5, 3.0, and 10,000) with an adaptive heating scheme that manipulated the temperatures according to their swapping success, as described in previous studies [52, 57, 58].

Gene flow was explored at two geographic scales: investigating the dispersal patterns within the Iberian Peninsula, and examining long-distance movements from the Balkans to Western Europe. First, we inferred the migration rate between Iberian regions, considering the Spanish Roma divided by geographic regions, also including the reference Portuguese Roma. We employed five distinct population history models to investigate migration patterns, following the approach of Almohammed et al. [52]: (i) the first model assumed all populations belonged to the same panmictic population; (ii) the second model entailed unidirectional gene flow from one population to another (East to West); (iii) the third model accounted for divergence from a common ancestral population; (iv) while the fourth model incorporated both divergence from the ancestral population and ongoing immigration; (v) the fifth model included both divergence from the ancestral population and ongoing immigration in two directions (East to West and *viceversa)* (Figure S2). Subsequently, we conducted the analysis on a continental scale, where pairwise comparisons were made between Spanish

Roma and Roma from three other European countries: Greece, Romania and Slovakia. For this analysis, we used the same previously described migration models.

To assess the relative strengths of the model fits, log marginal likelihoods were used to calculate Bayes factors using the script provided by the developer. The magnitudes of the Bayes factors provided evidence for the degree of dissimilarity between the models, which informed us about the relative fit of each model to the data.

Results

Y-chromosome haplogroup composition and diversity of Spanish Roma

The largest proportion of Y-chromosome lineages in Spanish Roma (35.8%) are assigned to haplogroup J2a1b, followed by H (27.7%), R1b (18.5%), J2 (3.5%), and I2a1 (3.5%), with smaller percentages (<3%) for the remaining haplogroups (Table S3). Differences are observed between Spanish Roma and other Roma groups (Table S4), with the exception of Portuguese Roma, who display a haplogroup frequency profile similar to that of Spanish Roma. Spanish Roma exhibit notably lower frequencies of the South Asian haplogroup H and higher frequencies of haplogroups R1b and J2a1b, both associated with West Eurasian regions (i.e., Europe and West Asia), compared to most Central and Eastern European Roma populations (Figure 1). These findings suggest higher levels of gene flow between Iberian Roma (Spanish and Portuguese Roma) and non-Roma Western European populations, in contrast to other European Roma groups. Among the haplogroups identifiable using Athey's method, no significant differences in the number of distinct haplogroups were observed between the Spanish Roma and other Roma populations (permutation test p-value = 0.0985).

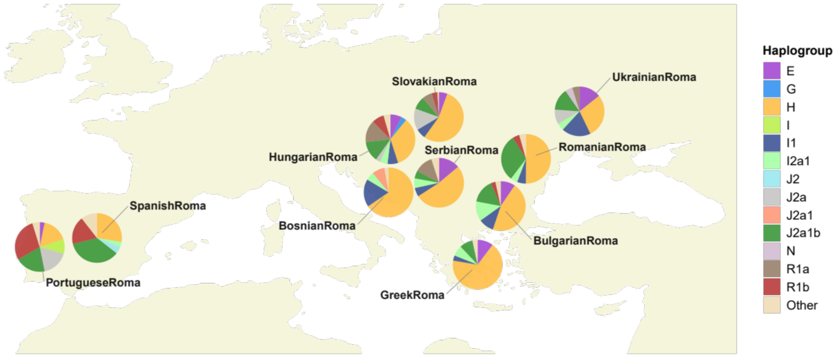


Figure 1. Geographic distribution of Y-chromosome haplogroups in European Roma populations. Pie charts show the frequencies of the inferred haplogroups in each population. Haplogroups with frequencies below 3% are labelled as "Other" in the legend.

To investigate potential admixture from North African and Jewish populations, which may have occurred during the period of cohabitation in the Iberian Peninsula in the 15th century, we examined the presence of Y-DNA haplogroups linked to these ancestries in the Spanish Roma. Our analysis revealed no significant North African influence on the Y-DNA composition of the Spanish Roma, as evidenced by the complete absence of native North African E haplogroups, such as E-M81 (which represents around 40-46% in North Africa) and the internal clade E-V65 of E-M78 (found in around 3%) [59, 60]. Similarly, we found no genetic connection between Spanish Roma and Jewish populations, as markers typically associated with Jewish ancestry (such as J1-P58, J2-M172, E-M34, and R1a-M582) [61–64] were absent in Spanish Roma. Having ruled out significant admixture and shared ancestry from North African and Jewish populations, we next turned our attention to the internal diversity of haplogroups within the Spanish Roma.

To explore this further, we constructed median-joining networks based on the Y-STR haplotypes (15 loci) for all Spanish Roma individuals, as well as independent networks for each of the four most frequent haplogroups (H, J2a1b, I2a(x), and R1b). Haplogroups were assigned within the network based on predicted clusters and, where available, Y-SNP data from reference populations. The resulting network analysis revealed distinct patterns of haplogroup diversity, highlighting the dominant features of each lineage. The dominant feature of the general Spanish Roma network (including all haplogroups; Figure 2) is the tight clustering of all H haplotypes within short branches, indicating minimal diversity.

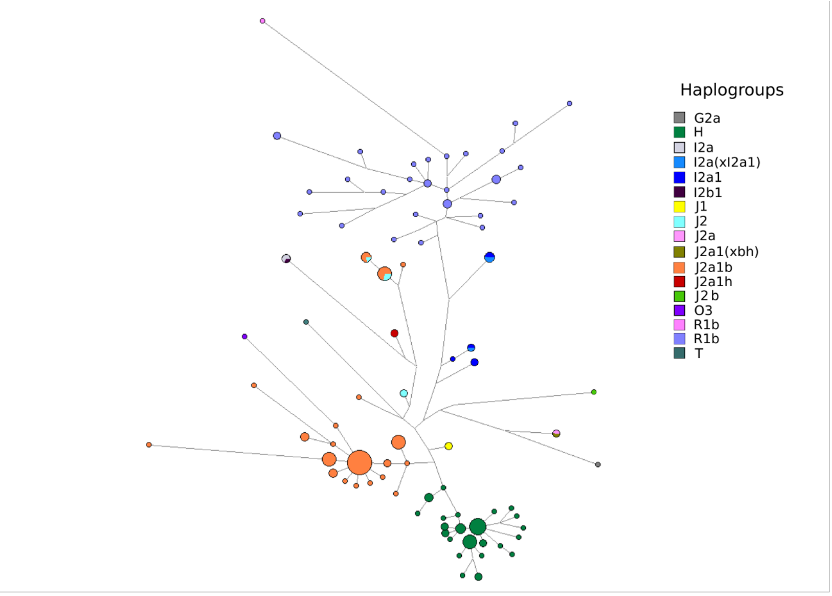


Figure 2. Median-joining network of 173 Spanish Roma individuals based on 17 Y-chromosome STR markers. Colours represent the inferred haplogroups derived from the STR profiles.

Similarly, J2a1b exhibits low diversity, although with some haplotypes more distantly located from the main cluster. R1b individuals are more dispersed across the network, reflecting greater diversity, while the other haplogroups are distributed between the main clusters without distinct patterns, except for a small subgroup of I2a(x). This network structure suggests that the H and J2a1b haplogroups in the Spanish Roma population may have experienced a founder effect, where a small initial group of ancestors contributed to the genetic pool, resulting in reduced diversity within these haplogroups compared to others.

Building upon the general network analysis of Spanish Roma, we next examined the diversity of individual haplogroups across all populations in our study. Y-STR median-joining networks for the founder haplogroups H and J2a1b show no clear clustering or differentiation within the Spanish Roma and exhibit limited diversity (Figures S3, S4, S5), suggesting an early and common divergence from other Roma groups. In contrast, the R1b and I2a(x) networks display considerable genetic diversity and a lack of clear structure among the Roma, indicating that the diversity in these haplogroups likely arose from independent admixture with local European populations during the Roma's migration through Europe (Figure S6 and Figure S7). Overall, these findings suggest that the Spanish Roma carried relatively low diversity for the founder haplogroups H and J2a1b upon their arrival in Spain, but incorporated a broader range of R1b and I2a(x) sub-haplogroups through admixture with local populations during their diaspora to the Iberian Peninsula.

To further investigate genetic differentiation within these haplogroups, we performed an MDS analysis based on pairwise genetic distances (Slatkin R_{st}). The MDS results for the H and J2a1b haplogroups (Figure 3a-b) show Roma populations tightly clustered, with a higher frequency of these haplogroups compared to non-Roma groups, highlighting the effects of founder events and genetic drift after their arrival in Europe. In contrast, for R1b and I2a(x) (Fig 3-c-d), show a more dispersed distribution among Roma populations, with no distinctive clustering and lower frequency relative to non-Roma groups, suggesting greater diversity from multiple introgressions and a more recent introduction into the Roma gene pool after their arrival in Europe.

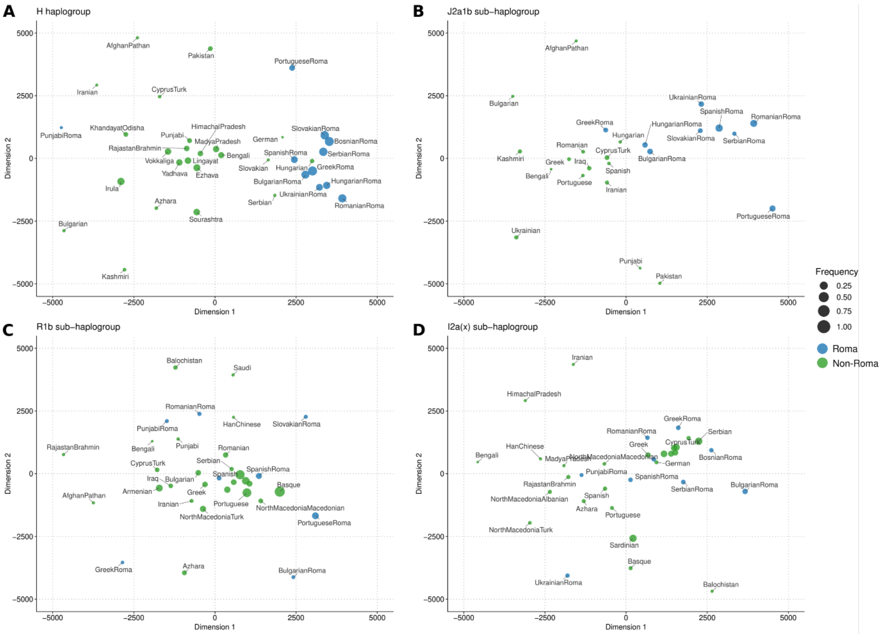


Figure 3. Multidimensional scaling (MDS) plot of the four major haplogroups found in Spanish Roma, based on RST distances. Roma populations are shown in blue and non-Roma populations in green. The size of the circles represents the population frequency of the haplogroups.

To explore the timing of these genetic events, we estimated the time to the most recent common ancestor (TMRCA) for the four major haplogroups across three contexts: the entire dataset, all Roma populations, and Spanish

Roma (Figure S8 and Table S5). In all cases, the TMRCA dates were more recent for the Spanish Roma, reflecting their more recent differentiation. The inferred dates predate the suggested out-of-India diaspora, which occurred over a thousand years ago [2, 4, 6, 7], hinting that the internal haplogroup diversity in the Roma gene pool predates their migration into Europe. Furthermore, this implies that the diversity observed in the R1b, J2a1b, and I2a(x) haplogroups was already present at the time of introgression, with no significant diversification detected subsequently within the Spanish Roma. These findings indicate that the Y-chromosome diversity in Roma groups reflects a series of distinct ancestry sources: (1) early genetic diversity from their South Asian ancestors, particularly for haplogroup H, which was present in South India; (2) the introduction of haplogroups such as J2a1b during their migration out of South Asia, but before reaching Europe; and (3) further admixture events in Europe involving haplogroups such as I2a(x) and R1b, which became prevalent once they arrived in Europe.

Geographic Y-chromosome sub-structure within the Iberian Peninsula

To examine the Spanish Roma at a finer geographic scale and compare genetic variation across Spanish regions, we measured and compared haplogroup composition and haplotype diversity within regional groups of Spanish Roma and non-Roma individuals (Table S6 and S7). The Spanish Roma groups exhibit lower haplotype diversity than the Spanish non-Roma (Table S8), likely due to smaller sample sizes and reduced genetic variability, which may result from serial population bottlenecks within the Roma population. Comparisons of haplogroup distributions between Spanish Roma and non-Roma populations reveal substantial differences, with no clear regional pattern linking Roma groups to geographically close non-Roma populations. Notable differences include the exclusive presence of haplogroup H and a higher prevalence of J2a1b in Roma groups, and the dominance of R1b in all Spanish non-Roma groups. The haplogroup distributions among Roma groups across different regions of Spain also show regional variation, which may reflect the impacts of differential admixture, isolation, and genetic drift (Table S9, Figure 4). Haplogroup H, associated with South Asian ancestry, is most frequent in Roma from Southern and Western Spain, with its frequency decreasing in Central and Northern Spain (Figure 4, Table S9). In contrast, J2a1b, linked to West Asian ancestry, is more common in Northern and Eastern Spain (Figure 4, Table S9). R1b, the dominant haplogroup in non-Roma Europeans, shows regional variability, with higher frequencies in Northern and Central Spain, suggesting increased gene flow with non-Roma populations in these areas (Figure 4, Table S9).

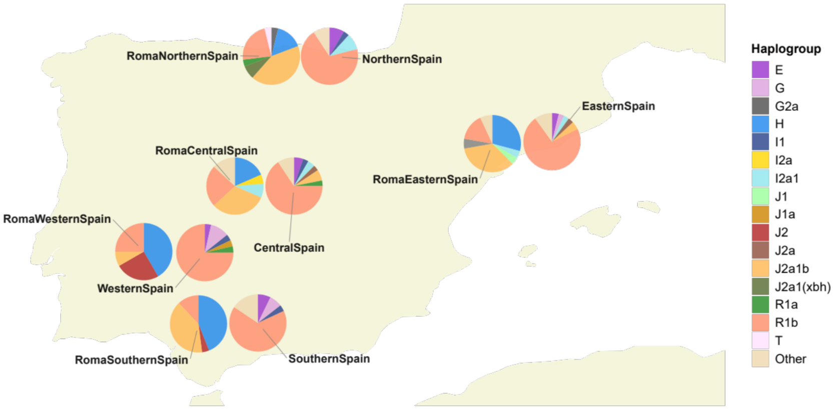


Figure 4. Geographic distribution of Y-chromosome haplogroups in Spanish Roma and non-Roma populations. Pie charts show the frequencies of inferred haplogroups in each population. Haplogroups with frequencies below 3% are labelled as "Other" in the legend.

Additionally, less frequent haplogroups, such as I2 and its sub-haplogroups, exhibit regional patterns, appearing exclusively in Central and Eastern Spain (Figure 4, Table S9). Together, these findings indicate diverse genetic influences and highlight regional substructure within Roma populations across Spain.

We next assessed the extent of genetic differentiation within the Spanish Roma using AMOVA to better understand internal heterogeneity and potential regional substructure (Table S10). Comparisons between the Y-chromosome composition of Spanish Roma and non-Roma populations showed significant differentiation, with 11.28% ($p<0.05$) of the variation observed between these groups. When analysing the Spanish Roma and Spanish non-Roma populations by region, we found that 19.77% of the genetic variation could be attributed to differences between these populations ($p<0.05$) with no significant differences within groups (0.9%, $p>0.05$). Further exploration of genetic variation within Spanish Roma and Spanish non-Roma populations revealed that regional differentiation was over six times higher among Spanish Roma (3%, $p<0.05$) compared to Spanish non-Roma (0.48%, $p<0.05$), highlighting greater internal heterogeneity within the Roma group.

Estimation of migration rates in the Roma population

To better understand regional substructure within Spain and quantify shared gene flow between groups, we used Slatkin's RST to calculate the migration rate (M) between Spanish regions (Table S11). Our analysis reveals that in all pairwise comparisons, the M values exceed 1, indicating the presence of gene flow. Notably, M values among Spanish non-Roma regions are exceptionally high ($M > 40$), suggesting an absence of population substructure. In contrast, M values between Roma regions are generally around 1, indicating more limited levels of gene flow and supporting the presence of regional substructure, consistent with our AMOVA results. Furthermore, moderate gene flow (M > 1 and < 2) was observed between Roma and Spanish non-Roma regions, with no distinctive pattern among regions. However, this demonstrates the existence of at least some level of gene flow between the Roma and Spanish non-Roma across all regions.

To explore migration patterns within and into the Iberian Peninsula, we analysed five distinct migration models, focusing on the Iberian Roma (Spanish Roma divided by geographic regions and Portuguese Roma) and their genetic structure (Table S12 A). The model with the highest likelihood indicates that genetic differentiation among regions within the Iberian Roma is best explained by divergence from a common ancestor, likely driven by historical isolation and genetic drift, with minimal recent gene flow between regions (Figure S9). The second most likely model assumes divergence with ongoing east-to-west migration, suggesting that the Spanish Roma populations originated from eastern ancestors, with continuous gene flow between eastern and western Spanish Roma groups in a west-to-east cline (Table S12 A).

To trace the movements of the Roma from southwestern Europe into the Iberian Peninsula, we broadened the scope of our analysis, testing migration models across Roma populations from southern and central Europe to the Iberian Peninsula (Table S12 B). The highest likelihood model, which assumes migration from East to West—spanning Greece, Romania, Slovakia, and Spain—indicates that the genetic structure of the Spanish Roma aligns with a stepwise westward migration process through these regions (Figure S10). This pattern suggests sequential movement with genetic differentiation likely occurring at each stage, consistent with historical routes of Roma dispersal across Europe, and limited gene flow between populations in different regions during this westward migration. The second most likely model assumes divergence from a common ancestor along a west-to-east cline, indicating that all European Roma trace their origins to eastern ancestors, with continuous gene flow from eastern to western Roma groups.

Discussion

The paternal genetic diversity observed within the Spanish Roma can be explained by demographic events including bottlenecks, founder effects, and multiple episodes of admixture with populations encountered throughout their diaspora, as well as following their settlement in the Iberian Peninsula. The Spanish Roma exhibit a distinctive paternal genetic profile with lower frequencies of South Asian haplogroups and a higher representation of Western European lineages, in contrast to Roma groups outside the Iberian Peninsula. This profile is characterised by a predominance of haplogroups J2a1b (34.1%) and H (27.7%), associated with West Asian and South Asian origins, respectively [65, 66]. For comparison, the frequency of haplogroup H in other European Roma ranges from a minimum of 28.6% in Ukrainian Roma to 67.2% in Greek Roma, possibly showing lower admixture with non-Roma. In the case of J2a1b, apart from Romanian Roma (30.7%), frequencies in other groups range from 19.8% in Portuguese Roma to 5.1% in Serbian Roma, suggesting that the increased frequency in Spanish Roma may result from genetic drift and endogamous practices, as observed with other haplogroups in previous studies on other Roma groups [67–69]. It is important to note that endogamy, combined with the practice of patrilocal residence observed both historically and in modern times among Spanish Roma [70, 71]—where women traditionally moved to their husband's community—has likely influenced haplogroup frequencies in both paternal and maternal lineages. Overall, the frequency of these haplogroups in the Roma is consistent with their South Asian origin and subsequent migration through West Asia during their diaspora. Haplogroup R1b, commonly associated with Central and Western European populations [72], is found at a relatively high frequency in Spanish Roma (18.5%) compared to non-Iberian

Roma groups in our dataset, where it remains below 8%. This haplogroup is particularly prevalent in Spanish (68%) and Basque (88%) populations, as estimated from data combined from earlier research [34, 73–75]. Taken together, this suggests substantial gene flow from Iberian populations into the Roma following their arrival in the region. This pattern of admixture with non-Roma populations in Iberia aligns with findings from autosomal DNA studies [22, 38] and likely reflects the history of forced assimilation occurred within the peninsula over several centuries [17, 24, 76].

Therefore, Roma populations likely arrived in Europe with existing diversity within haplogroups H and J2a1b, later incorporating additional lineages such as R1b and I2a(x), which are also prevalent in Europe [77], through admixture with local European populations. This process is reflected in the median-joining networks and MDS analyses, where haplogroups H and J2a1b show tight clustering and higher frequencies with minimal differentiation from other European Roma. In contrast, the R1b and I2a(x) haplogroups exhibit greater diversity and dispersion but occur at lower frequency in Roma compared to non-Roma populations, likely indicating multiple admixture events with European populations during their migration through Europe. Collectively these results align with historical accounts of the Roma's migration routes [2, 3] and are consistent with previous genetic studies on mtDNA [33], Y-chromosome [34, 69], and whole-genome array data [22] in Spanish Roma.

To explore alternative routes of the diaspora into Iberia, we examined genetic relationships with North African and Jewish groups. Our analyses reveal no significant North African influence on the Spanish Roma paternal gene pool, despite its presence in the non-Roma Spanish population [78], where haplogroup E-M81, associated with Amazigh ancestry [79–81], accounts for around 1%. This contrasts with the Portuguese Roma, where 3.2% carry E haplogroups of African origin [23]. In comparison, predictions for the Portuguese non-Roma, based on combined data from multiple studies [82, 83], indicate that 12% carry the E1b1b haplogroup. However, due to the lack of Y-SNP data, we cannot confirm how many of these haplotypes are specifically of North African origin. The presence of E haplogroups in the Portuguese Roma, but not in the Spanish Roma, suggests that North African haplogroups entered the Roma gene pool via non-Roma Iberian populations—who had previously admixed with North African groups—rather than through direct contact between Roma and North Africans before their arrival in Iberia. Besides, as 2-3% North African ancestry was observed in the autosomal DNA of Iberian Roma [22], the absence of E haplogroups in our Spanish Roma sample may reflect random variation and the haplogroup's low frequency in the population. These findings corroborate mtDNA studies [33], which also found no evidence of North African gene flow into the Spanish Roma. Additionally, we found no genetic link between Spanish Roma and Jewish populations, as paternal lineages commonly associated with Jewish ancestry were absent in the Spanish Roma and rarely present in other Roma groups. For example, J1-P58, which is predominant in several Jewish groups [84], was observed only twice in the Spanish non-Roma and in Bulgarian Roma, while E-M34 was detected in Serbian Roma at a frequency of 3.8%. These results suggest no evidence of admixture between Spanish Roma and Jewish populations, consistent with prior autosomal DNA findings [22].

Increased genetic drift, resulting from bottlenecks, founder events, and periods of isolation, has contributed to the preservation at high frequencies of specific founder lineages—defined by their presence in Roma and absence in non-Roma European populations [34, 37]— within the Roma population [85]. In this context, previous research has identified both maternal and paternal founder lineages within the Roma population [33, 34, 37]. These lineages include South Asian haplogroups (such as H-52 and H-

M82) and Western European haplogroups (I-P259, J-M92, and J-M67), which has been explained as a result of a single Roma origin from North-Western India, with admixture and bottlenecks during their diaspora through Middle East and Europe. Despite the limited resolution of our STR-based data, we identified the presence of some paternal founder lineages such as H1a1 and J2a1b*. Beyond that, evidence from the median-joining network analysis shows that some newly-genotyped samples, particularly within haplogroups such as H and J2a1b, cluster with reference individuals carrying known founder lineages.

While mtDNA studies reported an 86% West-Eurasian and 14% South Asian maternal heritage in Iberian Roma [33], our results show a significantly higher South Asian contribution (30%) on the paternal side in the Spanish Roma. This discrepancy aligns with observations of sex-biased admixture in European Roma when analysing complete mtDNA and Y-chromosome sequences [36]. The migration of the proto-Roma appears to have been mostly male-driven, with South Asian paternal lineages preserved and maintained at higher frequencies in the population, while a large amount of West Eurasian maternal lineages was incorporated during their diaspora. Cultural customs, where non-Roma women are more likely to join the community through marriage with Roma men [86–88], likely explain, at least in part, the observed sex-biased ancestry.

We provide evidence of regional population structure within the Spanish Roma, as AMOVA analysis shows that genetic differentiation among geographic regions is six times greater than in the Spanish non-Roma. This is consistent with autosomal DNA findings [22] but contrasts with maternal DNA studies [33], which showed less pronounced inter-regional differentiation. This suggests that the preferential integration of non-Roma women has likely homogenised mtDNA across regions, while more restricted and region-specific paternal gene flow has led to the higher differentiation observed in Y-DNA and autosomal markers. Although Roma groups have experienced historical gene flow between regions, prolonged isolation during later periods—potentially linked to the transition from a nomadic to a settled lifestyle—may have contributed to the observed genetic substructure. This is supported by the migration rate analysis, which reveals significant but low M values between Roma populations across different Spanish regions—substantially lower than those observed between non-Roma Spanish from different regions. This may depend on the patrilocal residence patterns [70, 71], which could help explain the reduced paternal gene flow between regions.

Finally, we contribute to understanding Roma migrations in the Iberian Peninsula and Europe by applying coalescent methods and Bayesian demographic model inference. At the European scale, the model with the highest likelihood suggests a westward migration pattern, consistent with previous autosomal DNA findings [22, 38], which identified multiple migration waves from the Balkans and Southwestern Europe towards Iberia, as well as with historical sources [2, 3]. At the Iberian scale, the model with the highest likelihood indicates divergence from ancestral populations without ongoing migration. These findings suggest that the Roma in the Iberian Peninsula diverged early from their ancestral populations and subsequently experienced limited gene flow. This supports the presence of a genetic structure shaped by geographic distribution and highlights distinctive genetic patterns within Iberian Roma compared to broader European Roma populations.

Conclusions

This study sheds light on the paternal genetic profile of Spanish Roma, highlighting the influence of different populations, demographic processes and cultural practices on their genetic structure. These findings contribute to the broader understanding of Roma migration patterns in Europe, offering valuable insights into the complex demographic history of Spanish Roma. While Y-STR analysis provides valuable insight into the paternal genetic structure of the Spanish Roma, its resolution is limited compared to whole Y-chromosome sequencing. Future studies should prioritise generating high-resolution paternal sequence data from a diverse set of European Roma populations to detect deep genealogical branches, identify new founder lineages, and capture a broader range of genetic diversity. Integrating this with autosomal, mtDNA and X-chromosome analyses will provide a more comprehensive view of Roma's demographic history.

Declarations

Ethics approval and consent to participate

This study was approved by the local IRB (Comitè d'Ètica de la Investigació, Parc de Salut Mar, references 2016/6723/I, on 7th June 2016; 2019/8900/I, on 15th January 2020; 2022/10542/I on 3rd November 2022; and 2023/10830 on 5th April 2023). All methods in this study were performed following standard guidelines and regulations.

All participants self-identified as Spanish Roma, and appropriate written consent was obtained from all donors.

Availability of data and materials

The newly generated raw Y-STR data are available in Supplementary Table S2.

Code Availability

The scripts generated during this study can be downloaded from https://github.com/gfena/Y-STR-tools. Please note that these scripts are provided as-is and are not supported.

Competing Interests

The authors have declared that no competing interests exist.

Funding

This work was supported by the Spanish Ministry of Science and Innovation and by "ERF A way of making Europe" (grant numbers CGL2019-106485GB-I00 and PID2022-138755NB-I00) funded by the MCIN and the AEI (DOI:10.13039/501100011033). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

DC contributed to the design and conception of the study. GF-E performed and implemented the data analysis. Data collection were performed by AG, AC-M. MACS contributed to the discussion and contextualization of the results. The first draft of the manuscript was written by GF-E and all authors commented on previous versions of the manuscript. All authors approved the submitted version.

Acknowledgments

We would like to thank the Roma FAGiC association (Federació d'Associacions Gitanes de Catalunya) and all its members and associates for its continuous support and collaboration.

This work was supported by the Spanish Ministry of Science and Innovation and by "ERF A way of making Europe" (grant numbers CGL2019-106485GB-I00 and PID2022-138755NB-I00) funded by the MCIN and the AEI (DOI:10.13039/501100011033).

References

1. Hancock IF. We are the Romani people. Univ of Hertfordshire Press; 2002.

2. Kenrick D. Historical Dictionary of the Gypsies (Romanies). Scarecrow Press; 2007.

3. Fraser AM. The Gypsies (The Peoples of Europe). Blackwell Pub.; 1992.

4. Hancock I. On Romani Origins and Identity. The Romani Archives and Documentation Center The University of Texas at Austin; 2006.

5. Beníšek M. The historical origins of Romani. Palgrave Handb Romani Lang Linguist. 2020;:13–47.

6. Moorjani P, Patterson N, Loh P-R, Lipson M, Kisfali P, Melegh BI, et al. Reconstructing Roma History from Genome-Wide Data. PLOS One. 2013;8.

7. Mendizabal I, Lao O, Marigorta UM, Wollstein A, Gusmão L, Ferak V, et al. Report Reconstructing the Population History of European Romani from Genome-wide Data. Curr Biol. 2012;22:2342–9.

8. Ena GF, Aizpurua-Iraola J, Font-Porterias N, Calafell F, Comas D. Population Genetics of the European Roma—A Review. Genes. 2022;13:2068.

9. Matras Y. Romani A Linguistic Introduction. 2002.

10. Bakker P, Ki͡uchukov K, Ki͡uchukov K. What is the Romani language? Univ of Hertfordshire Press; 2000.

11. Matras Y, Adamou E. Romani and contact linguistics. Palgrave Handb Romani Lang Linguist. 2020;:329–52.

12. Elsík V. Loanwords in Selice Romani, an Indo-Aryan language of Slovakia. Loanwords Worlds Lang Comp Handb. 2009;:260–303.

13. Beníšek M. The historical origins of Romani. Palgrave Handb Romani Lang Linguist. 2020;:13–47.

14. Bakker P. Romani genetic linguistics and genetics: Results, prospects and problems. Romani Stud. 2012;22:91–111.

15. Gamella JF, Fernández C, Adiego I-X. The long agony of Hispanoromani. The remains of Caló in the speech of Spanish Gitanos. Romani Stud. 2015;25:53–93.

16. Bastos JGP, Bastos SP. Gypsies (Ciganos) in Portugal, Today. Stud Eur. 2000;8:99.

17. Pym R. The Gypsies of Early Modern Spain. Springer; 2007.

18. Ortega MHS. Los gitanos españoles desde su salida de la India hasta los primeros conflictos en la península. Espac Tiempo Forma Ser IV Hist Mod. 1994.

19. Aparicio Gervás JM. Breve recopilación sobre la historia del pueblo gitano: desde su salida del Punjab, hasta la Constitución Española de 1978. Veinte hitos sobre la "otra" historia de España. RIFOP Rev Interuniv Form Profr Contin Antig Rev Esc Norm. 2006;:141–62.

20. Pohoryles Y. The Jewish-Romani connection: Are Gypsies descendants of tribe of Simeon? Ynetnews. 2018.

21. El Origen judío de los gitanos. http://www.hashavuabogota.com/articulos/704. Accessed 13 Dec 2024.

22. Ena GF, Giménez A, Carballo-Mesa A, Lišková P, Araújo Castro e Silva M, Comas D. The genetic footprint of the European Roma diaspora: evidence from the Balkans to the Iberian Peninsula. Hum Genet. 2025;144:463–79.

23. Gusmão A, Gusmão L, Gomes V, Alves C, Calafell F, Amorim A, et al. A Perspective on the History of the Iberian Gypsies Provided by

Phylogeographic Analysis of Y-Chromosome Lineages. Ann Hum Genet. 2008;72:215–27.

24. Sánchez DM. Historia del pueblo gitano en España. Los Libros De La Catarata; 2022.

25. Martínez Dhier A. La condición social y jurídica de los gitanos en la legislación histórica española.(A partir de la pragmática de los Reyes Católicos de 1499). 2007.

26. Halwachs DW. The Changing Status of Romani in Europe. In: Hogan-Brun G, Wolff S, editors. Minority Languages in Europe: Frameworks, Status, Prospects. London: Palgrave Macmillan UK; 2003. p. 192–207.

27. La Parra Casado D, Gil González D, de la Torre Esteve M. The social class gradient in health in Spain and the health status of the Spanish Roma. Ethn Health. 2016;21:468–79.

28. Font-Porterias N, Giménez A, Carballo-Mesa A, Calafell F, Comas D. Admixture Has Shaped Romani Genetic Diversity in Clinically Relevant Variants. Front Genet. 2021;12:683880.

29. Casals F, Anglada R, Bonet N, Rasal R, van der Gaag KJ, Hoogenboom J, et al. Length and repeat-sequence variation in 58 STRs and 94 SNPs in two Spanish populations. Forensic Sci Int Genet. 2017;30:66–70.

30. Mavillard F, Pérez-Florido J, Ortuño FM, Valladares A, Álvarez-Villegas ML, Roldán G, et al. The Iberian Roma genetic variant server; population structure, susceptibility to disease and adaptive traits. 2023;:2023.08.25.23294490.

31. Tenorio J, Navas P, Barrios E, Fernández L, Nevado J, Quezada CA, et al. A founder EIF2AK4 mutation causes an aggressive form of pulmonary arterial hypertension in Iberian Gypsies. Clin Genet. 2015;88:579–83.

32. Gómez-Carballa A, Pardo-Seco J, Fachal L, Vega A, Cebey M, Martinón-Torres N, et al. Indian Signatures in the Westernmost Edge of the European

Romani Diaspora: New Insight from Mitogenomes. PLOS ONE. 2013;8:e75397.

33. Aizpurua-Iraola J, Giménez A, Carballo-Mesa A, Calafell F, Comas D. Founder lineages in the Iberian Roma mitogenomes recapitulate the Roma diaspora and show the effects of demographic bottlenecks. Sci Rep. 2022;12:18720.

34. Martínez-Cruz B, Mendizabal I, Harmant C, de Pablo R, Ioana M, Angelicheva D, et al. Origins, admixture and founder lineages in European Roma. Eur J Hum Genet. 2016;24:937–43.

35. Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, Wise C, et al. Origins and Divergence of the Roma (Gypsies). Am J Hum Genet. 2001;69:1314–31.

36. García-Fernández C, Font-Porterias N, Kučinskas V, Sukarova-Stefanovska E, Pamjav H, Makukh H, et al. Sex-biased patterns shaped the genetic history of Roma. Sci Rep. 2020;10:14464.

37. Mendizabal I, Valente C, Gusmão A, Alves C, Gomes V, Goios A, et al. Reconstructing the Indian Origin and Dispersal of the European Roma : A Maternal Genetic Perspective. PLOS One. 2011;6:1–10.

38. Font-Porterias N, Arauna LR, Poveda A, Bianco E, Rebato E, Prata MJ, et al. European Roma groups show complex West Eurasian admixture footprints and a common South Asian genetic origin. PLOS Genet. 2019;15:e1008417.

39. Giménez A, Comas D, Carballo A. Origen e identidad del Pueblo Gitano. Int J Roma Stud. 2019;1:159–84.

40. Currie-Fraser E, Shah P, True S. Data analysis using GeneMapper® v4. 1: comparing the newest generation of GeneMapper software to legacy Genescan® and Genotyper® Software. J Biomol Tech JBT. 2010;21 3 Suppl:S31.

41. ISFG - isfg.org. https://www.isfg.org/. Accessed 31 Jul 2024.

42. Athey TW. Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. J Genet Geneal. 2006;2:34–9.

43. Khubrani YM, Wetton JH, Jobling MA. Extensive geographical and social structure in the paternal lineages of Saudi Arabia revealed by analysis of 27 Y-STRs. Forensic Sci Int Genet. 2018;33:98–105.

44. Nei M. Molecular evolutionary genetics. Columbia university press; 1987.

45. R Core Team. R: A language and environment for statistical computing. 2024.

46. Canty AJ. Resampling methods in R: the boot package. Newsl R Proj Vol. 2002;2:2–7.

47. Slatkin M. A measure of population subdivision based on microsatellite allele frequencies. Genetics. 1995;139:457–62.

48. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010;10:564–7.

49. Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, et al. vegan: Community Ecology Package. R package version 2.6-4. 2022. Github; 2023.

50. Bosch E, Calafell F, González-Neira A, Flaiz C, Mateu E, Scheil H-G, et al. Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. Ann Hum Genet. 2006;70:459–87.

51. Barbieri C, Butthof A, Bostoen K, Pakendorf B. Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia. Eur J Hum Genet. 2013;21:430–6.

52. Almohammed EK, Hadi A, Al-Asmakh M, Lazim H. The Qatari population's genetic structure and gene flow as revealed by the Y chromosome. PLOS ONE. 2023;18:e0290844.

53. Solé-Morata N, Villaescusa P, García-Fernández C, Font-Porterias N, Illescas MJ, Valverde L, et al. Analysis of the R1b-DF27 haplogroup shows that a large fraction of Iberian Y-chromosome lineages originated recently in situ. Sci Rep. 2017;7:7341.

54. Goedbloed M, Vermeulen M, Fang RN, Lembring M, Wollstein A, Ballantyne K, et al. Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFlSTR® Yfiler® PCR amplification kit. Int J Legal Med. 2009;123:471–82.

55. Pamjav H, Fóthi Á, Fehér T, Fóthi E. A study of the Bodrogköz population in north-eastern Hungary by Y chromosomal haplotypes and haplogroups. Mol Genet Genomics. 2017;292:883–94.

56. Beerli P, Mashayekhi S, Sadeghi M, Khodaei M, Shaw K. Population genetic inference with MIGRATE. Curr Protoc Bioinforma. 2019;68:e87.

57. Lazim H, Almohammed EK, Hadi S, Smith J. Population genetic diversity in an Iraqi population and gene flow across the Arabian Peninsula. Sci Rep. 2020;10:15289.

58. Adnan A, Rakha A, Lazim H, Nazir S, Al-Qahtani WS, Abdullah Alwaili M, et al. Are Roma People Descended from the Punjab Region of Pakistan: A Y-Chromosomal Perspective. Genes. 2022;13:532.

59. D'Atanasio E, Iacovacci G, Pistillo R, Bonito M, Dugoujon J-M, Moral P, et al. Rapidly mutating Y-STRs in rapidly expanding populations: Discrimination power of the Yfiler Plus multiplex in northern Africa. Forensic Sci Int Genet. 2019;38:185–94.

60. Bekada A, Fregel R, Cabrera VM, Larruga JM, Pestano J, Benhamamouch S, et al. Introducing the Algerian Mitochondrial DNA and

Y-Chromosome Profiles into the North African Landscape. PLOS ONE. 2013;8:e56775.

61. Nogueiro I, Manco L, Gomes V, Amorim A, Gusmão L. Phylogeographic analysis of paternal lineages in NE Portuguese Jewish communities. Am J Phys Anthropol. 2010;141:373–81.

62. Manco L, Albuquerque J, Sousa MF, Martiniano R, de Oliveira RC, Marques S, et al. The Eastern side of the Westernmost Europeans: Insights from subclades within Y-chromosome haplogroup J-M304. Am J Hum Biol. 2018;30:e23082.

63. Behar DM, Saag L, Karmin M, Gover MG, Wexler JD, Sanchez LF, et al. The genetic variation in the R1a clade among the Ashkenazi Levites' Y chromosome. Sci Rep. 2017;7:14969.

64. Rootsi S, Behar DM, Järve M, Lin AA, Myres NM, Passarelli B, et al. Phylogenetic applications of whole Y-chromosome sequences and the Near Eastern origin of Ashkenazi Levites. Nat Commun. 2013;4:2928.

65. Di Giacomo F, Luca F, Popa LO, Akar N, Anagnou N, Banyko J, et al. Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. Hum Genet. 2004;115:357–71.

66. Mahal DG, Matsoukas IG. The geographic origins of ethnic groups in the Indian subcontinent: exploring ancient footprints with Y-DNA haplogroups. Front Genet. 2018;9:4.

67. Regueiro M, Rivera L, Chennakrishnaiah S, Popovic B, Andjus S, Milasin J, et al. Ancestral modal Y-STR haplotype shared among Romani and South Indian populations. Gene. 2012;504:296–302.

68. Salihović MP, Barešić A, Klarić IM, Cukrov S, Lauc LB, Janićijević B. The role of the Vlax Roma in shaping the European Romani maternal genetic history. Am J Phys Anthropol. 2011;146:262–70.

69. Regueiro M, Stanojevic A, Chennakrishnaiah S, Rivera L, Varljen T, Alempijevic D, et al. Divergent patrilineal signals in three Roma populations. Am J Phys Anthropol. 2011;144:80–91.

70. Martin E, Gamella JF. Marriage practices and ethnic differentiation: The case of Spanish Gypsies (1870–2000). Hist Fam. 2005;10:45–63.

71. Gamella JF, Muntean VM. Marriage and the reproductive regime of a digitally connected Roma diaspora. J Contemp Cent East Eur. 2023;31:533–59.

72. Myres NM, Rootsi S, Lin AA, Järve M, King RJ, Kutuev I, et al. A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. Eur J Hum Genet. 2011;19:95–101.

73. García O, Yurrebaso I, Mancisidor ID, López S, Alonso S, Gusmão L. Data for 27 Y-chromosome STR loci in the Basque Country autochthonous population. Forensic Sci Int Genet. 2016;20:e10–2.

74. Gaibar M, Esteban E, Moral P, Gómez-Gallego F, Santiago C, Bandrés F, et al. STR genetic diversity in a Mediterranean population from the south of the Iberian Peninsula. Ann Hum Biol. 2010;37:254–67.

75. Solé-Morata N, Bertranpetit J, Comas D, Calafell F. Y-chromosome diversity in Catalan surname samples: insights into surname origin and frequency. Eur J Hum Genet. 2015;23:1549–57.

76. Dhier AM. Expulsión o asimilación, esa es la cuestión : los gitanos en Castilla durante el gobierno de la Monarquía Absoluta. Expulsión O Asimilación Esa Es Cuestión Los Gitanos En Castilla Durante El Gob Monarquía Absol. 2011;:173–230.

77. Navarro-López B, Granizo-Rodríguez E, Palencia-Madrid L, Raffone C, Baeta M, de Pancorbo MM. Phylogeographic review of Y chromosome haplogroups in Europe. Int J Legal Med. 2021;135:1675–84.

78. Adams SM, Bosch E, Balaresque PL, Ballereau SJ, Lee AC, Arroyo E, et al. The Genetic Legacy of Religious Diversity and Intolerance: Paternal Lineages of Christians, Jews, and Muslims in the Iberian Peninsula. Am J Hum Genet. 2008;83:725–36.

79. Solé-Morata N, García-Fernández C, Urasin V, Bekada A, Fadhlaoui-Zid K, Zalloua P, et al. Whole Y-chromosome sequences reveal an extremely recent origin of the most common North African paternal lineage E-M183 (M81). Sci Rep. 2017;7:15941.

80. Fadhlaoui-Zid K, Martinez-Cruz B, Khodjet-el-khil H, Mendizabal I, Benammar-Elgaaied A, Comas D. Genetic structure of Tunisian ethnic groups revealed by paternal lineages. Am J Phys Anthropol. 2011;146:271–80.

81. Reguig A, Harich N, Barakat A, Rouba H. Phylogeography of E1b1b1b-M81 Haplogroup and Analysis of Its Subclades in Morocco. Hum Biol. 2014;86:105–12.

82. Alves C, Gomes V, Prata MJ, Amorim A, Gusmão L. Population data for Y-chromosome haplotypes defined by 17 STRs (AmpFlSTR YFiler) in Portugal. Forensic Sci Int. 2007;171:250–5.

83. Pontes ML, Cainé L, Abrantes D, Lima G, Pinheiro M de F. Allele frequencies and population data for 17 Y-STR loci (AmpFℓSTR® Y-filerTM) in a Northern Portuguese population sample. Forensic Sci Int. 2007;170:62–7.

84. Hammer MF, Behar DM, Karafet TM, Mendez FL, Hallmark B, Erez T, et al. Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish priesthood. Hum Genet. 2009;126:707–17.

85. Kalaydjieva L, Morar B, Chaix R, Tang H. A newly discovered founder population: the Roma/Gypsies. BioEssays. 2005;27:1084–94.

86. Drummond S. Mapping Marriage Law in Spanish Gitano Communities. UBC Press; 2011.

87. Matras Y. The Romani Gypsies. Harvard University Press; 2015.

88. Gamella JF, Álvarez-Roldán A. Breaking secular endogamy. The growth of intermarriage among the Gitanos/Calé of Spain (1900–2006). Hist Fam. 2023;28:457–83.

4. DISCUSSION

This chapter follows a top-down structure, beginning with macro-scale analyses of Roma origins, broad patterns of gene flow, and continent-wide genetic variability. It then narrows in scope to examine microgeographical patterns, including regional differentiation within the Iberian Roma and the ancestry of specific pathogenic variants in Czech Roma. This organisation allows for a comprehensive understanding of Roma population history—from its deep ancestral roots to the contemporary regional dynamics observed today.

4.1 New evidence on Roma. Insights and contributions of this thesis

Genetic data on Roma is fragmentary, with some groups more represented than others and varying levels of resolution. Many studies have a narrow focus—some are limited to medical research describing specific mutations, while others provide only basic population data with a low number of SNPs. Even whole-genome studies often rely on very small sample sizes, sometimes using only two or three individuals per group. Moreover, data collection continues to be challenging due to low participation, which is partly driven by historical and ongoing fears of discrimination and persecution. This issue is further compounded by difficulties in obtaining samples from certain communities and countries, as well as the strict ethical requirements that apply to all human populations. Despite these limitations, this thesis—utilising the largest Roma array-data dataset to date—aims to provide a comprehensive study focused on Iberia, while also considering Central and Southwestern European Roma.

In addition to expanding the available genetic data, this thesis provides an extensive genetic characterisation of Roma populations, focusing on demographic studies using array data, defining the Y-chromosome profile, and assessing the genetic background of pathogenic variants related to rare ocular diseases. More specifically, this work includes the first studies on the demographic history of Roma populations in the Czech Republic, as well as contributions to the understanding of genetic ancestry and the distribution of pathogenic variants within this group. Additionally, this research integrates genetic findings with socio-cultural factors, offering a deeper understanding of the genetic landscape of Central, Southwestern, and Iberian Roma. It also represents the first population genetics study to take into consideration mating patterns, having inferred assortative mating in Roma populations. Besides, this thesis provides new, more detailed evidence regarding the genetic influence of populations across the Roma diaspora, addressing several key questions that remained unanswered prior to our work.

4.1.1 Origins, dispersal, and patterns of gene flow in Roma

A central question regarding the Roma concerns their origins, the shared nature of these origins across groups, and the patterns, routes, and scale of their migrations, along with the resulting gene flow. Previous studies have traced the origins of the Roma population to South Asia, specifically North-Western India, where the proto-Roma emerged approximately 1,000 years ago [154,155,239,244], a conclusion supported by both historical records and clear South Asian ancestry in modern Roma genomes [231,237,239,240]. The arrival of the proto-Roma in Europe thus marks a pivotal moment in their genetic history, setting the stage for centuries of interaction with local populations.

Building on these foundations, our analyses suggests that the proto-Roma first entered Europe around 30-35 generations ago, carrying a distinct north-western Indian genetic component. Early admixture events, particularly among eastern Roma populations, reflects interactions with local groups as the proto-Roma moved through the Middle East and into the Balkans. These events align with earlier genetic studies confirming the Roma's South Asian origin [238,269]. Our Y-DNA analysis further supports the persistence of South Asian lineages, such as H1a1, even after the Roma's long diaspora across Europe. These lineages, common in the Indian subcontinent, further corroborate the Roma's South Asian origin, consistent with previous studies of both maternal and paternal lineages [232,238,269].

Turning to the question of how Roma reached the Iberian Peninsula, we tested hypotheses of a North African entry route. While autosomal data reveal a modest North African signal in the genomes of Iberian Roma, Y-DNA analysis detects none in Spanish Roma, very little in Portuguese Roma, and none elsewhere. The overall level of North African admixture is lower than that seen in non-Roma Iberians, and Y-chromosome haplogroups of North African origin are virtually absent, leading us to reject a North African entry into Iberia, in line with previous studies [232]. Likewise, although autosomal DNA shows trace Jewish ancestry, its frequency is lower than in non-Roma Iberians, implying regional admixture rather than direct gene flow from Jewish communities.

Our findings also indicate two major waves of European admixture. Roma who settled in the Balkans experienced a pulse around 31 generations ago, whereas those in Western Europe—especially Iberia—underwent a later wave approximately 25–26 generations ago. The genetic profiles of Iberian Roma, with their pronounced southwest-European component, suggest that westward migration was accompanied by increasing integration into local populations, perhaps driven by imperial-era assimilation policies [86].

By contrast, Southeastern European Roma—such as the Czech and Macedonian groups—have remained relatively isolated. Their stronger retention of South Asian signatures reflects prolonged marginalisation, historic practices of endogamy and fewer opportunities for intermarriage. These patterns were likely shaped by more severe policies of persecution within the Holy Roman Empire and later under the Habsburgs [95], as well as the institution of slavery in the Romanian principalities [93]. As a result, these groups display lower levels of West Eurasian admixture and a more distinct genetic profile, when compared to Iberian Roma.

This regional contrast highlights how multiple migration waves and variable social contexts have shaped Roma genetic diversity. Moreover, analysis of more recent admixture events reveals continued gene flow in both eastern and western Roma around 10 and 5 generations ago, respectively. These late pulses demonstrate that, despite historic barriers to mixing, the Roma have remained genetically dynamic and responsive to new external influences throughout their European diaspora.

Ultimately, the varying proportions of South Asian and West Eurasian ancestry across Roma groups attest to the complex interplay of ancient migrations, regional interactions and evolving social norms. This evidence reinforces the view of the Roma as an admixed and dynamic population, shaped by centuries of migration, isolation, and integration across the continent.

4.1.2 Sex bias gene flow in Spanish Roma

Our analyses confirm previous findings of sex-biased gene flow in the Roma population. As seen in other Roma communities [238], Spanish Roma display markedly lower diversity in the Y-chromosome compared to mitochondrial DNA, a pattern shaped by sociocultural dynamics throughout their diaspora [256]. This asymmetry in diversity points to a male-driven migration out of South Asia, with South Asian paternal haplogroups—such as H—preserved at higher frequencies, while maternal lineages are overwhelmingly of West Eurasian origin.

Research on mtDNA indicate that over 85% of maternal lineages in Iberian Roma derive from West Eurasian sources [247,249], whereas West Eurasian Y-DNA lineages account for approximately 70% of paternal ancestry in our results. This disparity is best explained by higher integration of non-Roma women in the communities, leading to the incorporation of West Eurasian mitochondrial lineages into the gene pool, while South Asian Y-chromosomes were maintained across generations. Limited integration of non-Roma men, in contrast, is reflected in the lower frequency of South Asian mtDNA lineages, suggesting that Roma women had fewer opportunities to enter non-Roma communities.

Nonetheless, the presence of considerable West Asian and European Y-haplogroups among Spanish Roma shows that gene flow from non-Roma males did occur—likely at various stages during their migration and settlement in Europe. These instances may reflect the inclusion of men from outside the Roma community who perhaps shared a similar itinerant or marginalised lifestyle. The relatively high differentiation and lower frequency of European Y-lineages compared to non-Roma populations suggests sporadic introgression, whereas the high frequency and low diversity of South and West Asian haplogroups indicate founder effects and population bottlenecks early in the Roma diaspora.

These patterns are further influenced by likely patrilocal social structures, wherein women traditionally move to live with the husband's family after marriage. This would promote genetic continuity through the paternal line, reinforcing the observed sex bias [256]. Altogether, the combined effects of social exclusion, endogamy, patrilocality, and asymmetric integration of non-Roma individuals have shaped the distinctive patterns of maternal and paternal ancestry observed in the Iberian Roma today.

4.1.3 Variability across Roma groups: geographic and genetic substructure

The Roma population exhibits significant spatial genetic variability, with distinct substructures emerging across different Roma groups in Europe [155,239,245]. Our genetic analysis reveals that the composition of Roma ancestry is not uniform, but rather exhibits a strong correlation with geographic distribution, reflecting both historical migration routes and regional interactions. Western Roma, particularly Iberian Roma, show a higher proportion of Western European ancestry, while Central and Eastern Roma groups retain a greater share of South Asian and Southeastern European ancestry. This genetic-geographic pattern is further supported by fineSTRUCTURE analysis, which clusters Iberian Roma separately from other European Roma groups, highlighting the genetic differentiation between Western and Eastern Roma populations.

Most notably, through statistical testing of these ancestry components in relation to geographic location, particularly longitude and latitude, we have formally demonstrated a correlation. Specifically, as one moves westward, there is a significant increase in West Eurasian, particularly European, ancestry. Conversely, moving eastward, we observe a statistically significant increase in South Asian ancestry. Additionally, we note a slightly significant cline, with North African components being more pronounced in the Iberian Roma compared to other groups, reflecting the introgression of North African genetic material via the Iberian non-Roma population. This genetic substructure within the Roma is therefore again reflecting the complex and differential admixture patterns.

Additionally, regional substructure within groups like the Spanish Roma, as revealed by AMOVA and migration rate analysis, further emphasizes the role of local factors in shaping genetic diversity. Iberian Roma show a higher degree of genetic differentiation between regions compared to non-Roma Spanish, reflecting the historical patterns of settlement and isolation within the Iberian Peninsula. This variability underscores the dynamic nature of Roma genetic history, shaped by both ancient migrations and more recent events, such as forced assimilation, regional isolation, and varying levels of gene flow. The differing patterns of genetic substructure across Roma groups reinforce the need for region-specific studies to better understand the complex history and genetic diversity within the Roma population.

4.1.4 Endogamy and evolving social norms within Roma

Endogamy has been a significant factor in shaping the genetic structure of Roma populations over time. Our study aligns with the hypothesis that endogamy levels within Roma groups have evolved in response to both historical and social factors [155]. In particular, the Iberian Roma demonstrate higher levels of endogamy during the initial periods of their arrival in Europe and in the Iberian Peninsula. This increased isolation in the past likely served to preserve their distinct genetic identity, a practice that would have been crucial during periods of forced settlement and marginalisation.

As the Roma integrated into European societies and faced changing social norms, the dynamics of endogamy began to shift. Over time, particularly after the end of slavery and during subsequent waves of migration, Iberian Roma experienced increasing gene flow with non-Roma Europeans, which contributed to the gradual reduction of endogamy levels. This pattern contrasts with that of Southeastern European Roma, who, despite similar historical challenges, maintained a higher degree of isolation, likely due to differing social structures and external pressures. The decline in gene flow between Southeastern European Roma and non-Roma populations has led to a greater degree of genetic similarity within these groups, as evidenced by the longer shared IBD segments and a higher degree of inbreeding observed in these populations.

The analysis of European local ancestry tracts and shared IBD segments highlights the historical shifts in social norms regulating mating patterns within Roma communities. The Iberian Roma exhibit evidence of increasing gene flow from surrounding European populations, particularly during and following the second migration out of the Balkans in the 19th century—a migration event closely associated with the abolition of slavery in the Romanian principalities. In contrast, Southeastern European Roma exhibit evidence of more persistent endogamy, as suggested by the greater length of IBD segments and the higher levels of genetic isolation. These findings suggest that endogamy in Roma communities is not a static practice, but rather one shaped by complex historical events, including migration, forced assimilation, and changing social norms across generations.

The shift in endogamy patterns also reflects broader societal changes, where social and economic pressures likely influenced the extent to which Roma communities interacted with non-Roma populations. As Roma began to encounter more opportunities for integration and mobility, particularly in the wake of socio-political changes in Europe, the practice of endogamy likely weakened, resulting in the increasing genetic admixture seen in contemporary Iberian Roma populations. However, the persistence of endogamy in certain groups, such as Southeastern European Roma, illustrates how social norms can evolve at different rates depending on the region and local conditions.

Regional variation in endogamy patterns and gene flow is a key aspect of understanding the genetic structure of Roma populations across Europe. Endogamy has historically been a prominent feature of Roma communities [315]. However, these practices have varied considerably depending on geographical location, historical context, and the level of integration with surrounding populations.

4.1.5 Assortative mating and the impact on Roma's genetic profile

Assortative mating, the non-random mating pattern where individuals with similar traits are more likely to mate with each other, plays a critical role in shaping the genetic profile of populations [316]. For the Roma, assortative mating is an important factor that has contributed to the genetic structure of their communities. This phenomenon has been particularly significant in Roma populations due to their unique social, cultural, and historical contexts.

In Roma communities, mating patterns have been influenced by social stratification, cultural preferences, and endogamy [290,291,317]. These practices have reinforced the genetic cohesion within specific subgroups of Roma, limiting the gene flow between different Roma groups as well as between Roma and non-Roma populations [290]. For example, within the Roma, subgroups with distinct linguistic or regional backgrounds tend to marry within their own groups, perpetuating genetic similarities [290]. This has led to the formation of distinct genetic clusters within the broader Roma population. In particular, linguistic, cultural, and geographical boundaries have reinforced assortative mating, maintaining a degree of genetic isolation even as the Roma moved across different regions of Europe [317].

The social norms within Roma communities often emphasize the importance of maintaining ethnic and cultural identity through marriage [17]. This cultural preference for endogamy is typically associated with the idea of preserving Roma traditions, customs, and kinship networks [133], which has further contributed to assortative mating. For example, Roma communities may prefer to marry individuals with similar social status, familial ties, or shared social practices [315], thus reducing genetic diversity within those groups.

The effects of assortative mating in Roma populations can be observed in several ways. First, it can lead to the persistence of certain genetic traits or disorders that are more likely to be passed down through generations within tightly-knit subgroups [318]. This type of mating pattern can have both beneficial and detrimental consequences to health. On the one hand, assortative mating may help preserve beneficial traits that enhance survival and adaptation within a particular environment [319]. On the other hand, it may increase the risk of recessive genetic disorders, as certain genetic variations become more concentrated within small populations [318].

Moreover, assortative mating in Roma groups may limit the introduction of new genetic material from outside populations, especially during periods when Roma were geographically isolated or subjected to discrimination. This restriction of gene flow can result in a higher degree of genetic homogeneity within Roma subgroups, which is reflected in the genetic profiles of contemporary Roma populations. For example, when compared to non-Roma European populations, Roma tend to exhibit more distinct genetic markers due to centuries of endogamy and assortative mating.

However, in more recent generations, as Roma have faced social and political integration into broader European societies, assortative mating patterns have likely shifted. Increasing social mobility and changes in cultural attitudes have made it more common for Roma individuals to marry outside of their immediate social or regional group, leading to more genetic admixture and the gradual reduction of genetic isolation. This shift in mating patterns has contributed to greater genetic diversity within Roma communities, particularly in regions where Roma populations have been exposed to higher levels of gene flow from surrounding populations [155,256,317].

In conclusion, assortative mating has played a fundamental role in shaping the genetic profile of Roma populations. The combination of social norms and practices, as well as historical experiences have reinforced genetic cohesion within subgroups, leading to distinct genetic profiles across different Roma communities. While assortative mating has contributed to genetic isolation in the past, changing social dynamics and increasing gene flow from non-Roma populations are reshaping the genetic landscape of the Roma today.

4.2 Microgeographic structure and medical implications

4.2.1 Microgeographical structure in Roma

Genetic analyses of Roma populations across Europe have revealed notable microgeographic variation, reflecting differences in historical admixture levels, population size, and sociocultural practices such as endogamy. Studies have consistently shown that Western European Roma groups—such as those in Spain, Portugal, and Wales—exhibit higher levels of genetic admixture with surrounding non-Roma populations than their Eastern European counterparts, who often display more genetic continuity with the original Roma gene pool [239,245,270]. MtDNA and Y-chromosome data further underscore the regional diversity in Roma genetic structure. Demographic events and varying degrees of endogamy have been key drivers of this genetic variation. Although endogamy has been a longstanding feature of Roma history, its intensity and consistency differ notably across regions. Recognizing these regional patterns underscores the need to consider local social and historical contexts when examining the genetic diversity of marginalised populations.

4.2.2 Genetic substructure within the Iberian Roma

The genetic substructure within Iberian Roma reveals significant microgeographical variation. A notable feature is the absence of strong geographic patterns in allele sharing between Roma and non-Roma populations. Our results show that Roma groups do share alleles with local non-Roma populations, but this affinity does not follow a clear geographic pattern. For example, Roma groups from northern regions are not necessarily closer to local non-Roma than to Roma from other regions. This suggests that historical and social factors, rather than simple geographic proximity, have played a more important role in shaping genetic differentiation.

Haplotype-sharing analyses, including fineSTRUCTURE and NNLS, reveal that regional Roma clusters display moderately varying ancestry proportions, further emphasizing the non-uniform nature of Iberian Roma. This is also evident in the dendrogram, where Roma from different regions form distinct subclusters, though these do not strictly correspond to administrative or geographic boundaries. In the paternal lineage, we observe low to moderate differences in haplogroup frequencies across regions—for example, southern Spanish Roma show higher proportions of South Asian ancestry, while eastern groups appear more admixed with local non-Roma populations.

These regional genetic distinctions reflect the complex interplay of migration history, social practices, and endogamy within the Iberian Roma community.

4.3 Limitations and Potential Biases in the Study

4.3.1 Challenges with sample size and representation

The study of Roma genetics faces significant challenges related to sample size and representation. Obtaining large, diverse, and representative samples is difficult due to social barriers, recruitment challenges, and historical exclusion of Roma communities from research. Many Roma populations are dispersed across various regions, and obtaining participation from these communities is hindered by concerns about privacy, stigma, and marginalisation. Furthermore, Roma groups are unevenly distributed across Europe, and regional substructures—shaped by historical migration patterns and varying levels of integration with non-Roma populations—complicate efforts to capture the full genetic diversity of the population. Without proper geographic representation, studies may fail to reflect the genetic variation present across different Roma groups, leading to biased conclusions. The combination of small sample sizes and uneven geographic coverage can hinder the ability to accurately understand Roma genetic structure and history, making it essential to address these issues in future research.

4.3.2 Methodological constraints

The study of Roma genetics, as most other human populations, is also limited by several methodological constraints that can impact the accuracy and depth of the findings. One of the primary challenges is the resolution of genetic data, particularly when using lower-resolution techniques like SNP arrays or Y-STR analysis. These methods, while informative, miss rare variants and deeper genealogical branches, potentially underestimating the complexity of Roma genetic history. Furthermore, the use of historical records to create models to infer migration patterns and gene flow is prone to oversimplification, and the accuracy of these inferences depends heavily on the assumptions made during analysis. The lack of well-defined and comprehensive reference groups—especially from regions where Roma populations are under-represented—can also lead to inaccuracies in admixture and ancestry estimations. Additionally, methodological choices, such as the algorithms used for admixture analysis or the clustering methods for population substructure, may introduce biases depending on how they are applied to the available genetic data. These constraints emphasize the need for the application of high-resolution sequencing technologies and the inclusion of a broader range of reference populations to better capture the intricate genetic history of the Roma.

4.3.3 The need for improved reference databases

A significant limitation of the study is the lack of comprehensive and diverse reference databases, which can impact the interpretation of genetic data, particularly in terms of admixture estimates and population comparisons. The Roma, with their complex history of migration and admixture with various populations across Europe and Asia, present a challenge in terms of accurately characterising their genetic profiles. The absence of well-represented reference groups from regions with under-represented Roma populations—especially those in more remote or less-studied areas—limits the resolution and accuracy of ancestry analysis. Inadequate reference data can lead to errors in estimating the proportion of genetic contributions from different populations, skewing the interpretation of historical migration patterns and gene flow. The creation of more inclusive, global reference databases, particularly from regions with large and diverse Roma communities, would allow for more precise comparisons and a better understanding of Roma genetic structure. Such databases could also help capture the genetic diversity across the entire Roma diaspora, offering more accurate models of their demographic history and enabling future studies to draw from a broader and more robust set of genetic data.

4.4 Conclusions and future directions

4.4.1 Concluding Remarks

The study of Roma population genetics has provided significant insights into the complex history, migration patterns, and social dynamics of this unique ethnic group. Key findings indicate that the Roma's genetic structure reflects a complex interplay of historical events, geographic factors, and social norms such as endogamy. The initial migration of the Roma from northern India to Europe, followed by centuries of settlement across different European regions, has generated unique genetic diversity in various Roma groups. This genetic diversity is further shaped by the effects of isolation, assortative mating, and the influence of local populations through gene flow, while maintaining a clear separation from non-Roma populations in certain areas, such as among Iberian Roma. The importance of integrating diverse genetic data, including mitochondrial and Y-chromosome studies, has highlighted how these lineages have contributed to the Roma's distinct genetic profile.

Furthermore, the limitations of current research, particularly in terms of sample size, geographic representation, and reference databases, underscore the need for more comprehensive and diverse data collection. This study not only enriches our understanding of Roma population genetics but also opens the door for future research that could enhance our grasp of the Roma's unique place in the genetic landscape of Europe.

4.4.2 Pathways for further research

4.4.2.1 The need for Whole-Genome Sequencing

The advancement of whole-genome sequencing (WGS) represents a crucial step forward in the study of Roma population genetics. Until now genetic studies of the Roma populations have primarily relied on targeted sequencing methods or smaller genomic regions, which has limited the resolution of genetic variation observed. Whole-genome sequencing would enable a more comprehensive analysis of the genetic structure and variation across the entire genome, providing deeper insights into both rare and common genetic variants that contribute to the unique genetic landscape of the Roma. This approach will improve the accuracy of admixture analyses and clarify the relationships between Roma subgroups and their ancestral populations, offering more precise models for studying migration and population structure. This technology would also facilitate the identification of previously unknown genetic diseases and help trace the evolution of genetic traits over time. It is crucial that future studies incorporate whole-genome sequencing to enhance our understanding of Roma's genetic history and health, allowing for a more accurate and inclusive representation of the population's genetic diversity.

4.4.2.2 Importance of targeted sequencing

Targeted sequencing is crucial for studying the genetic background of pathogenic variants, particularly in rare diseases among the Roma population.

Given the high prevalence of founder effects and population-specific mutations, this approach enables precise identification of disease-associated variants at the individual level. By focusing on relevant genomic regions, targeted sequencing enhances sensitivity in detecting rare pathogenic mutations while reducing the complexity and cost of whole-genome analysis. However, a key challenge in studying these diseases is the absence of many pathogenic variants in reference databases, making standard approaches less effective. Targeted sequencing on a broader scale overcomes this limitation by allowing in-depth analysis of specific mutation regions, not only in affected individuals but also in non-affected individuals from the same population. This comparative approach is essential for distinguishing true pathogenic variants from benign polymorphisms, particularly within the genetic profile of the Roma community. By sequencing both groups, researchers can more accurately assess variant frequency, inheritance patterns, and potential disease mechanisms, ultimately improving genetic diagnosis, risk assessment, and personalised treatment strategies for rare conditions.

4.4.2.3 Expanding sample coverage and regional diversity

Expanding sample coverage and regional diversity is essential for gaining a comprehensive understanding of Roma population genetics. While current studies have provided valuable insights, the sampling methods often focus on specific subgroups or geographic regions, which may not capture the full extent of the genetic diversity within the Roma population. By increasing the geographic scope of future studies to include under-represented areas, particularly in Eastern and Southeastern Europe, researchers can uncover regional variations in genetic structure that may have been overlooked. Additionally, incorporating a wider range of Roma subgroups—each with potentially distinct histories of migration, endogamy, and genetic influence —will provide a more complete picture of the population's genetic diversity. Ensuring that both urban and rural populations are included in genetic studies is also crucial, as different environmental, social, and cultural factors may contribute to regional variation in genetic profiles. Expanding the sample diversity will also help improve the accuracy of admixture models and provide deeper insights into the complex patterns of gene flow and migration. This approach will not only refine our understanding of Roma population history but also enhance the inclusivity and representativeness of genetic research in the Roma community.

4.4.2.4 Exploring the influence of social norms and practices on genetic structure

Exploring the role of social norms and practices in shaping genetic structure is key to understanding the ways in which the Roma population’s genetic profile has evolved over time. Social practices, such as endogamy and assortative mating, along with regional migration patterns, play a significant role in shaping the genetic makeup of Roma groups. The norms that regulate these practices tend to limit gene flow between different Roma subgroups and surrounding non-Roma populations, helping to preserve distinct genetic markers within Roma communities. By investigating how these social practices influence the genetic diversity, researchers can gain insights into the mechanisms behind genetic similarity within specific Roma subgroups, as well as variation across regions. Furthermore, understanding the influence of social practices such as marriage customs, social cohesion, and identity on genetic structure could provide valuable context for interpreting historical and contemporary patterns of gene flow. This research could also reveal how cultural adaptation and social networks have influenced genetic drift and selection over generations, offering a more nuanced understanding of the dynamic relationship between genetics, culture, and social structure within human populations.

4.4.2.5 Integrating maternal and paternal lineage studies

Integrating maternal and paternal lineage studies is crucial for gaining a more comprehensive understanding of the genetic structure and history of the Roma population. MtDNA and Y-chromosome analysis provide valuable insights into the maternal and paternal inheritance patterns, respectively, and can reveal distinct genetic trajectories within Roma communities. MtDNA is passed down through the maternal line, while Y-chromosome analysis focuses on the paternal line, allowing researchers to track two separate, yet complementary, genetic histories. By examining both lineages, researchers can uncover migration patterns, genetic bottlenecks, and the extent of gene flow between Roma and non-Roma populations over time. For example, differences in the maternal and paternal genetic profiles could reveal asymmetries in migration or social practices, such as preferential mate selection or gender-specific patterns of settlement. Additionally, studies of the X chromosome—due to its unique mode of inheritance—can further illuminate sex-biased demographic events and complement uniparental and autosomal findings. Integrating these studies would enable a clearer understanding of how Roma groups have evolved genetically, considering both the direct genetic contributions from the maternal and paternal lineages and the social dynamics influencing them.

5. BIBLIOGRAPHY

1. Lee, R. Roma in Europe: "Gypsy" Myth and Romani Reality—New Evidence for Romani History. in *"Gypsies" in European Literature and Culture: Studies in* European *Culture and History* 1–28 (Springer, 2008).

2. Sampson, J. On the Origin and Early Migrations of the Gypsies. *Journal of the Gypsy Lore Society* **2**, 156–169 (1923).

3. Gmelch, S. B. Groups that don't want in: Gypsies and other artisan, trader, and entertainer minorities. *Annual review of anthropology* **15**, 307–330 (1986).

4. Matras, Y. A conflict of paradigms. *Romani Studies* **14**, 193–209 (2004).

5. Matras, Y. *Romani A Linguistic Introduction*. (2002).

6. Matras, Y. *The Romani Gypsies*. (Harvard University Press, 2015).

7. Amanolahi, S. The Gypsies of Iran (a brief introduction). *Iran & the Caucasus* 109–118 (1999).

8. Hancock, I. *On Romani Origins and Identity*. (2006).

9. Hancock, I., Kenrick, D. & Acton, T. The emergence of Romani as a koine outside of India. *Scholarship and the Gypsy struggle* 1–13 (2000).

10. Marushiakova, E. & Popov, V. Migrations West to East in the Times of the Ottoman Empire: The Example of a Gypsy/Roma Group in Modern Iran. *Anthropology of the Middle East* **5**, 93–99 (2010).

11. Baghbidi, H. R. The Zargari language: An endangered European Romani in Iran. *Romani Studies* **13**, 123–148 (2003).

12. Pstrusińska, J. *Secret Languages of Afghanistan and Their Speakers*. (Cambridge Scholars Publishing, 2014).

13. Marsch, A. A brief history of gypsies in Turkey. *European Roma Rights Center Country Reports Series* 5–20 (2008).

14. Soulis, G. C. The Gypsies in the Byzantine Empire and the Balkans in the Late Middle Ages. *Dumbarton Oaks Papers* **15**, 141 (1961).

15. Winstedt, E. O. The Gypsies of Modon and the'Wyne of Romeney'. *Romani Studies* **3**, 57 (1909).

16. Frescobaldi, L. di N. *Viaggio in Egitto e in Terra Santa (1384) Con Un Discorso Dell'editore Sopra Il Commercio Degl'Italiani Nel Secolo XIV*. (Rome, 1818).

17. Fraser, A. M. *The Gypsies (The Peoples of Europe)*. (Blackwell Pub., 1992).

18. Marushiakova-Popova, E. A. & Popov, V. The 'Gypsies'(Dom–Lom–Rom) in southern Caucasus (Georgia, Armenia, Azerbaijan). *Preserving the Roma Memories* (2020).

19. Hancock, I. On the migration and affiliation of the Dōmba: Iranian words in Rom, Lom and Dom Gypsy. *Matras* **1995**, 25–51 (1995).

20. Hancock, I. F. *We Are the Romani People*. vol. 28 (Univ of Hertfordshire Press, 2002).

21. Bakker, P., Ki͡uchukov, K. & Ki͡uchukov, K. *What Is the Romani Language?* vol. 21 (Univ of Hertfordshire Press, 2000).

22. Carmona, S. Nuevas perspectivas sobre la génesis de la historia gitana. *Quaderns de la Mediterrània= Cuadernos del Mediterráneo* 321–327 (2013).

23. Beck, S. The origins of Gypsy slavery in Romania. *Dialectical Anthropology* 53–61 (1989).

24. Tcherenkov, L. & Laederich, S. *The Roma. History, Language, and Groups*. (Basel: Schwabe Verlag Basel, 2004).

25. Scholz, L. *Borders and Freedom of Movement in the Holy Roman Empire*. (Oxford University Press, 2020).

26. Burns, R. I. The Guidaticum Safe-Conduct in Medieval Arago-Catalonia: a Mini-Institution for Muslims, Christians and Jews. *Medieval encounters* **1**, 51–113 (1995).

27. Williams, M. K. Diplomatic Safe Conducts across Sixteenth-Century Habsburg-Ottoman Borders. in *Borders and Mobility Control in and between Empires and Nation-States* 33–65 (Brill, 2022).

28. Taylor, B. *Another Darkness, Another Dawn: A History of Gypsies, Roma and Travellers.* (Reaktion books, 2014).

29. Webb, D. Pilgrims and pilgrimage in the medieval West. (2001).

30. Salihović, M. P. *et al.* The role of the Vlax Roma in shaping the European Romani maternal genetic history. *American journal of physical anthropology* **146**, 262–270 (2011).

31. Ena, G. F., Aizpurua-Iraola, J., Font-Porterias, N., Calafell, F. & Comas, D. Population Genetics of the European Roma—A Review. *Genes* **13**, 2068 (2022).

32. Font-Porterias, N. *et al.* The Counteracting Effects of Demography on Functional Genomic Variation: The Roma Paradigm. *Molecular Biology and Evolution* **38**, 2804–2817 (2021).

33. Marushiakova, E., Popov, V. Z., Popov, V. & Descartes), C. de recherches tsiganes (Université R. *Gypsies in the Ottoman Empire: A Contribution to the History of the Balkans.* (Univ of Hertfordshire Press, 2001).

34. Crowe, D. *A History of the Gypsies of Eastern Europe and Russia*. (Springer, 2016).

35. Crowe, D. M. The international and historical dimensions of Romani migration in Central and Eastern Europe. *Nationalities Papers* **31**, 81–94 (2003).

36. Barany, Z. The East European Gypsies in the imperial age. *Ethnic and Racial Studies* **24**, 50–63 (2001).

37. Aaslestad, K. Napoleonic empire and migration. *The Encyclopedia of Global Human Migration* (2013).

38. Janoski, T. The difference that empire makes: Institutions and politics of citizenship in Germany and Austria. *Citizenship Studies* **13**, 381–411 (2009).

39. Panayi, P. *Ethnic Minorities in 19th and 20th Century Germany: Jews, Gypsies, Poles, Turks and Others*. (Routledge, 2014).

40. Jezernik, B. The Gypsies as the original sin of modernity. in *Miasto mozaika. Opis kulturowy* 289–306 (Wydawnictwo Uniwersytetu Łódzkiego).

41. Achim, V. *The Gypsies in the Romanian Principalities-the Emancipation Laws, 1831-1856*. (Editura Academiei Române, 2004).

42. Achim, V. *The Roma in Romanian History*. (Central European University Press, 2004).

43. Bernal, J. F. ¿ Y los gitanos de Argentina? *O Tchatchipen: lil ada trin tchona rodipen romani= revista trimestral de investigación gitana* 15–17 (2002).

44. Teixeira, R. C. História dos ciganos no Brasil. *Recife: Núcleo de Estudos Ciganos* 15–21 (2008).

45. Hancock, I. Romanies and the Holocaust: A Re-evaluation and Overview. in *The historiography of the Holocaust* 383–396 (Springer, 2004).

46. Lewy, G. *The Nazi Persecution of the Gypsies*. (Oxford University Press, 2000).

47. Trevisan, P. ‘Gypsies’ in Fascist Italy: from expelled foreigners to dangerous Italians. *Social History* **42**, 342–364 (2017).

48. Kotljarchuk, A. World War II and the Registration of Roma in Sweden: The Role of Experts and Census-Takers. *Holocaust and Genocide Studies* **31**, 457–479 (2017).

49. Ioanid, R. *The Holocaust in Romania: The Destruction of Jews and Gypsies Under the Antonescu Regime, 1940-1944*. (Ivan R. Dee, 2008).

50. Lutz, B. D. Gypsies as Victims of the Holocaust. *Holocaust and Genocide Studies* **9**, 346–359 (1995).

51. Reinhartz, D. Unmarked graves: the destruction of the Yugoslav Roma in the Balkan Holocaust, 1941–1945. *Journal of Genocide Research* **1**, 81–89 (1999).

52. Korb, A. Ustaša mass violence against Gypsies in Croatia, 1941-1942. *The Nazi Genocide of the Roma: Reassessment and Commemoration* 72–95 (2013).

53. Kenrick, D. The Nazis and the Gypsies: A fresh look. *Jewish Quarterly* **41**, 46–47 (1994).

54. Barkhof, S. & Smith, A. K. *War and Displacement in the Twentieth Century: Global Conflicts*. (Routledge, 2014).

55. Kirk, D. & Huyck, E. Overseas migration from Europe since world war II. *American Sociological Review* **19**, 447–456 (1954).

56. Margalit, G. *Germany and Its Gypsies: A Post-Auschwitz Ordeal*. (University of Wisconsin Pres, 2002).

57. Milton, S. Persecuting the survivors: The continuity of ‘Anti-Gypsyism’in postwar Germany and Austria. *Sinti and Roma: Gypsies in German-Speaking Society and Literature* 35–47 (1998).

58. Perkins, J. Continuity in modern German history? The treatment of gypsies. *Immigrants & minorities* **18**, 62–82 (1999).

59. Von dem Knesebeck, J. *The Roma Struggle for Compensation in Post-War Germany.* (University of Hertfordshire Press, 2011).

60. Donert, C. *The Rights of the Roma.* (Cambridge University Press, 2017).

61. Martins-Heub, K. 'Genocide in the 20th Century' REFLECTIONS ON THE COLLECTIVE IDENTITY OF GERMAN ROMA AND SINTI (GYPSIES) AFTER NATIONAL SOCIALISM. *Holocaust and Genocide Studies* **4**, 193–211 (1989).

62. Milton, S. Gypsies and the Holocaust. *The History Teacher* **24**, 375–387 (1991).

63. Marushiakova, E. & Popov, V. Politics of multilingualism in Roma education in early Soviet Union and its current projections. *Social Inclusion* **5**, 48–59 (2017).

64. Barany, Z. Politics and the Roma in state-socialist Eastern Europe. *Communist and Post-Communist Studies* **33**, 421–437 (2000).

65. Silverman, C. Bulgarian Gypsies: Adaptation in a Socialist Context. *Nomadic Peoples* 51–62 (1986).

66. Stewart, M. Gypsies, the work ethic, and Hungarian socialism. in *Socialism* (Routledge, 1993).

67. Kostelancik, D. J. The Gypsies of Czechoslovakia: Political and ideological considerations in the development of policy. *Studies in Comparative Communism* **22**, 307–321 (1989).

68. Marushiakova, E. & Popov, V. 'Letter to Stalin': Roma activism vs. Gypsy nomadism in Central, South-Eastern and Eastern Europe before WWII. *Social Inclusion* **8**, 265–276 (2020).

69. Abakunova, H. Roma in Soviet Ukraine: Ways of Life and Forced Sedentarisation Before and After the Second World War. *Multiethnica: Journal of the Hugo Valentin Centre* **42**, 63–80 (2022).

70. Petrova, D. The Roma: Between a myth and the future. *Social Research: An International Quarterly* **70**, 111–161 (2003).

71. Bartash, V. The sedentarisation of Roma in the Soviet Union after 1956: A case study from the former Belarussian Soviet Socialist Republic1. *Romani Studies* **25**, 23–51 (2015).

72. Haller, I. The Mendacious Government: Implementation of the Romanian Pogrom Judgments. *Roma Rights Quarterly* 23–28 (2010).

73. Laederich, S. Roma in the Ukraine: From the origins to present days, the war and refugees. *Zurich: Rroma Foundation* (2022).

74. Belavusau, U. Anti-Roma Hate Speech in the Czech Republic, Hungary and Poland. (2013).

75. Marushiakova, E. & Popov, V. Gypsy/Roma European migrations from 15th century till nowadays. in *Proceedings of International Conference "Romani Mobilities in Europe: Multidisciplinary Perspectives* 126–139 (2010).

76. Ryšavý, Z. Czech Republic: Hundreds of rampaging neo-Nazis attempt pogrom against Romani people in Ostrava. *Romea.cz - Everything about Roma in one place* https://romea.cz/en/czech-republic/czech-republic-hundreds-of-rampaging-neo-nazis-attempt-pogrom-against-romani-people-in-ostrava/ (2013).

77. Ten years after the Roma killings in Hungary: there's nothing 'so-called' about antigypsyism. *European Roma Rights Centre* http://www.errc.org/news/ten-years-after-the-roma-killings-in-hungary-theres-nothing-so-called-about-antigypsyism.

78. Kolchyna, G. E., Viktoryia. Attacked and abandoned: Ukraine's forgotten Roma. *Al Jazeera* https://www.aljazeera.com/features/2018/11/23/attacked-and-abandoned-ukraines-forgotten-roma.

79. Anti-Roma pogroms in Ukraine: on C14 and tolerating terror. *European Roma Rights Centre* http://www.errc.org/news/anti-roma-pogroms-in-ukraine-on-c14-and-tolerating-terror.

80. Chillingly retrograde Ukrainian court ruling justifies anti-Roma pogrom as 'direct democracy'. *Human Rights in Ukraine* https://khpg.org//en/1608809962.

81. El nuevo Eurobarómetro sobre discriminación muestra el antigitanismo como la discriminación más extendida. *Fundación Secretariado Gitano* https://www.gitanos.org/actualidad/el_nuevo_eurobarometro_sobre_discriminacion_muestra_el_antigitanismo_como_la_discriminacion_mas_extendida/index.html.es.

82. Vasconcelos, M., Ribeiro, J. & Costa, E. Dados oficiais sobre os povos romani (ciganos) no Brasil–2013. *AMSK Brasil, Brasília* (2013).

83. Sutherland, A. *Gypsies: The Hidden Americans.* (Waveland Press, 1986).

84. Liegeois, J.-P. & Gheorghe, N. *Roma/Gypsies: A European Minority.* vol. 94 (Minority Rights Group, 1995).

85. Kemény, I. History of Roma in hungary. *East European Monographs* **702**, 1 (2005).

86. Pym, R. *The Gypsies of Early Modern Spain.* (Springer, 2007).

87. Brearley, M. The persecution of Gypsies in Europe. *American Behavioral Scientist* **45**, 588–599 (2001).

88. Lucassen, L., Willems, W., Cottaar, A. & Lucassen, L. The clink of the hammer was heard from daybreak till dawn: Gypsy occupations in Western Europe (nineteenth-twentieth centuries). *Gypsies and other itinerant groups: A socio-historical approach* 153–173 (1998).

89. Altinöz, İ. Osmanlı toplumunda çingeneler. (2013).

90. Dingeç, E. XVI. Yüzyılda Osmanlı Ordusunda Çingeneler. *Süleyman Demirel Üniversitesi Fen-Edebiyat Fakültesi Sosyal Bilimler Dergisi* **2009**, 33–46 (2009).

91. Çelebi, E. *Narrative of Travels in Europe, Asia, and Africa in the Seventeenth Century.* vol. 1 (Cambridge University Press, 2012).

92. Çelik, F. Exploring marginality in the Ottoman Empire: Gypsies or People of Malice (Ehl-i Fesad) as viewed by the Ottomans. (2004).

93. Gheorghe, N. The origin of Roma's slavery in the Romanian principalities. *Roma* **7**, 12–27 (1983).

94. Marin, M. On the fringes of mainstream: Assessing the extreme poverty of the Roma people in Romania—A historical perspective. in *The Political Economy of Extreme Poverty in Eastern Europe* 17–37 (Routledge, 2024).

95. Zahra, T. "Condemned to Rootlessness and Unable to Budge": Roma, Migration Panics, and Internment in the Habsburg Empire. *The American Historical Review* **122**, 702–726 (2017).

96. Tittel, L. S. Racial and social dimensions of antiziganism: the representation of Gypsies in the political theory. (2021).

97. Kant, I. *Zerstreute Aufsätze.* vol. 1 (na, 1793).

98. Kant, I. *Von Den Verschiedenen Rassen Der Menschen.* (BoD–Books on Demand, 2018).

99. Kleingeld, P. Kant's second thoughts on race. *The Philosophical Quarterly* **57**, 573–592 (2007).

100. Grellmann, H. M. G. *Historischer Versuch Über Die Zigeuner: Betreffend Die Lebensart Und Verfassung, Sitten Und Schicksale Dieses Volks Seit Seiner Erscheinung in Europa Und Dessen Ursprung.* (Bey Johann Christian Dieterich, 1787).

101. Crowe, D. M. From persecution to pragmatism: the Habsburg Roma in the eighteenth century. *Austrian History Yearbook* **37**, 99–120 (2006).

102. Goodson, A. The "Gypsies" as displaced others in Doyle's Sherlock Holmes stories. (2018).

103. Bardi, A. The Gypsy as trope in Victorian and modern British literature. *Romani Studies* **16**, 31–42 (2006).

104. Saul, N. *Gypsies and Orientalism in German Literature and Anthropology of the Long Nineteenth Century.* (MHRA, 2007).

105. Houghton-Walker, S. *Representations of the Gypsy in the Romantic Period.* (OUP Oxford, 2014).

106. Said, E. W. Orientalism. *The Georgia Review* **31**, 162–206 (1977).

107. Egger, L. Karl Marx und die Kritik des Antiziganismus. *Freie Assoziation* **26**, 63–67 (2024).

108. Dragomir, C. I. Nomads,"Gypsies," and Criminals in England and India from the Seventeenth to the Nineteenth Century. *Critical Romani Studies* **2**, 62–81 (2019).

109. Sánchez, D. M. *Historia del pueblo gitano en España.* (Los Libros De La Catarata, 2022).

110. Braun, K., Herrmann, S. L. & Brekke, O. A. Sterilization policies, moral rehabilitation and the politics of amends. *Critical Policy Studies* **8**, 203–226 (2014).

111. Lee, K. Orientalism and gypsylorism. *Social Analysis: The International Journal of Social and Cultural Practice* **44**, 129–156 (2000).

112. Marushiakova-Popova, E. A. & Popov, V. Orientalism in Romani studies: the case of Eastern Europe. *Languages of Resistance* (2017).

113. Ringold, D., Orenstein, M. A. & Wilkens, E. *Roma in an Expanding Europe: Breaking the Poverty Cycle.* (World Bank Publications, 2005).

114. Frazer, H. & Marlier, E. Promoting the social inclusion of Roma. (2011).

115. Ciaian, P. & Kancs, d'Artis. *Causes of the Social and Economic Marginalisation: The Role of Social Mobility Barriers for Roma*. (2016).

116. Picker, G. *Racial Cities: Governance and the Segregation of Romani People in Urban Europe.* (Routledge, 2017).

117. Filčák, R. & Škobla, D. Sanitation infrastructure at the systemic edge: Segregated Roma settlements and multiple health risks in Slovakia. *International Journal of Environmental Research and Public Health* **18**, 6079 (2021).

118. Vuksanović-Macura, Z. The mapping and enumeration of informal Roma settlements in Serbia. *Environment and Urbanization* **24**, 685–705 (2012).

119. Chaudhuri, I. N. Community mobilization for slum upgrading through sanitation in Roma informal settlements in the Paris region. *Frontiers in public health* **5**, 213 (2017).

120. Loveland, M. T. & Popescu, D. The Gypsy threat narrative: Explaining anti-Roma attitudes in the European Union. *Humanity & Society* **40**, 329–352 (2016).

121. Townsend, E. Hate speech or genocidal discourse? An examination of anti-Roma sentiment in contemporary Europe. *PORTAL: Journal of Multidisciplinary International Studies* **11**, 1–23 (2014).

122. 'India should declare Romas as national minority of Indian origin'. *Hindustan Times* https://www.hindustantimes.com/india/india-should-declare-romas-as-national-minority-of-indian-origin/story-xlQJRtOLtSTUqId652QOnJ.html (2016).

123. Standard, B. Romas are India's children: Sushma Swaraj. https://www.business-standard.com/article/news-ians/romas-are-india-s-children-sushma-swaraj-116021201051_1.html (2016).

124. Acton, T. A. & Mundy, G. *Romani Culture and Gypsy Identity.* (Univ of Hertfordshire Press, 1997).

125. Lenin, V. I. *On Culture and Cultural Revolution.* (Wildside Press LLC, 2008).

126. Gramsci, A. Socialism and culture. *Pre-Prison Writings* 9–10 (1916).

127. RUGGIU, I. Is begging a Roma cultural practice? Answers from the Italian legal system and anthropology. *Romani Studies* **26**, 31–62 (2016).

128. Weyrauch, W. O. *Gypsy Law: Romani Legal Traditions and Culture.* (Univ of California Press, 2001).

129. Puxon, G. *Roma: Europe's Gypsies.* vol. 209 (Minority Rights Group, 1992).

130. Cohn, W. *The Gypsies.* (Addison Wesley Publishing Company, 1973).

131. Leland, C. G. *Gypsy Sorcery and Fortune Telling.* (T. Fisher Unwin, 1891).

132. Afanasieva, L., Bukrieieva, I., Glyns' ka, L., Hlebova, N. I. & Semikin, M. Family roles of Roma and ways to overcome its socio-cultural isolation. *Journal of History Culture and Art Research* **9**, 288–296 (2020).

133. Edden, N., Hughes, K., McCormack, E. & Prendergast, C. *A Study Investigating the Cultural Traditions and Customs of the Romani Community in Gorton, Manchester.* (Manchester: The University of Manchester, 2011).

134. Stewart, M. *The Time of the Gypsies.* (Routledge, 2019).

135. Browne, E. Gender norms in the Western Balkans. (2017).

136. Herrero-Arias, R., Parra-Casado, D. L., Ferrández-Ferrer, A., Sanchís-Ramón, M.-J. & Ortiz-Barreda, G. Experiences and constructions of womanhood and motherhood among Spanish Roma women. *European Journal of Women's Studies* **30**, 164–180 (2023).

137. Preoteasa, A. M. Roma women and precarious work: Evidence from Romania, Bulgaria, Italy and Spain. *Revista de cercetare şi Intervenţie Socială* 155–168 (2013).

138. Matras, Y. *The Classification of Romani Dialects: A Geographic-Historical Perspective.* (na, 2005).

139. Matras, Y. *Romani in Britain.* (Edinburgh University Press, 2010).

140. Bakker, P. Notes on the genesis of Caló and other Iberian Para-Romani varieties. *romani in Contact. the History, Structure and Sociology of a Language* 125–150 (1995).

141. O'Higgins, N. & Ivanov, A. Education and employment opportunities for the Roma. *Comparative Economic Studies* **48**, 6–19 (2006).

142. Carrasco, S., Narciso, L. & Bertran-Tarrés, M. Neglected aspirations: Academic trajectories and the risk of early school leaving amongst immigrant and Roma youth in Spain. in *Comparative perspectives on early school leaving in the European Union* 164–182 (Routledge, 2018).

143. Rostas, I. & Kostka, J. Structural dimensions of Roma school desegregation policies in Central and Eastern Europe. *European Educational Research Journal* **13**, 268–281 (2014).

144. Surdu, L., Vincze, E. & Wamsiedel, M. Roma school participation, non-attendance and discrimination in Romania. *Bucureşti: UNICEF* (2011).

145. Hojsík, M. Examples of successful housing desegregation as a precondition of Roma integration. (2024).

146. Méreiné-Berki, B., Málovics, G. & Creţan, R. "You become one with the place": Social mixing, social capital, and the lived experience of urban desegregation in the Roma community. *Cities* **117**, 103302 (2021).

147. Rostas, I. *Ten Years after: A History of Roma School Desegregation in Central and Eastern Europe.* (Central European University Press, 2012).

148. Weyrauch, W. O. & Bell, M. A. Autonomous Lawmaking: The Case of the" Gypsies". (1993).

149. Matras, Y. Notes on Romani culture. *Computer Printout* (2008).

150. Matras, Y. & Adamou, E. Romani and contact linguistics. *The Palgrave Handbook of Romani Language and Linguistics* 329–352 (2020).

151. Friedman, V. A. The Romani language in the Republic of Macedonia: Status, usage, and sociolinguistic perspectives. *Acta Linguistica Hungarica* **46**, 317–339 (1999).

152. Lee, R. *Learn Romani.* (Univ of Hertfordshire Press, 2005).

153. Matras, Y. *Romani in Contact: The History, Structure and Sociology of a Language.* (John Benjamins Publishing, 1995).

154. Kamusella, T. & Jaskułowski, K. *Nationalisms Today.* (Peter Lang, 2009).

155. Marushiakova, E. & Popov, V. Beginning of Romani literature: the case of Alexander Germano. *Romani studies* **30**, 135–161 (2020).

156. Zahova, S. History of Romani literature. *časopis romistických studií | 26. ročník | 1/2019* 183 (2014).

157. Zahova, S. Romani writers and the legacies of Yugoslavia. *Baltic Worlds* **11**, 25–33 (2018).

158. Zahova, S. Romani language literature. *The Palgrave Handbook of Romani Language and Linguistics* 539–569 (2020).

159. Hertrampf, M. O. M. Romani Literature (s) as Minor Literature (s) in the Context of World Literature: A Survey of Romani Literatures in French and Spanish. *Critical Romani Studies* **3**, 42–57 (2020).

160. Toninato, P. Romani literature in Britain and Australia: An overview.

161. Moseley, C. *Atlas of the World's Languages in Danger*. (Unesco, 2010).

162. Mufwene, S. Colonisation, globalisation, and the future of languages in the twenty-first century. *International Journal on Multicultural Societies* **4**, 162–193 (2002).

163. Zhang, F. Globalisation and Minority Languages: A Critical Analysis of the Impact of Globalisation on Minority Languages. *International Journal of Frontiers in Sociology* **6**, (2024).

164. Csepeli, G. & Simon, D. Construction of Roma identity in Eastern and Central Europe: perception and self-identification. *Journal of Ethnic and Migration Studies* **30**, 129–150 (2004).

165. Talewicz-Kwiatkowska, J. One culture or multiple cultures? The diversity of Roma people in Poland. in *Colloquia Humanistica* 77–89 (Instytut Slawistyki Polskiej Akademii Nauk, 2014).

166. Kapralski, S. Memory, identity, and Roma transnational nationalism. *Transnational Memory. Circulation, Articulation, Scales, ed. by Chiara De Cesari and Ann Rigney (Berlin: De Gruyter, 2014)* 195–218 (2014).

167. Sorescu-Marinković, A., Kahl, T. & Sikimić, B. *Boyash Studies: Researching "Our People"*. (Frank & Timme GmbH, 2021).

168. Spinelli, S. *Rom Genti Libere*. (Baldini & Castoldi, 2013).

169. Borrow, G. *The Zincali, Or An Account of the Gypsies of Spain: With an Original Collection of Their Songs and Poetry, and a Copious Dictionary of Their Language*. (J. Murray, 1843).

170. Borrow, G. H. *The Bible in Spain and the Gypsies of Spain*. (R. Carter, 1848).

171. MARTIN SANCHEZ, D. El pueblo gitano en el País Vasco y Navarra (1435-1802). (Universidad del País Vasco-Euskal Herriko Unibertsitatea, 2016).

172. Lovera, C. J. Aportaciones documentales a la historia de los gitanos en Andalucía. *Boletín del instituto de estudios giennenses* 41–56 (1980).

173. Catalunya, M. V. del P. G. de & Salinas, J. Castilla y León. *Museu Virtual del Poble Gitano a Catalunya* https://www.museuvirtualgitano.cat/es/historia/comunidades-autonomas/castilla-y-leon/.

174. Martínez, M. M. Los gitanos en el sureste peninsular de los siglos XV y XVI. *Boletín del Instituto de Estudios Almerienses. Letras* 91–102 (1995).

175. Aparicio Gervás, J. M. Breve recopilación sobre la historia del pueblo gitano: desde su salida del Punjab, hasta la Constitución Española de 1978. Veinte hitos sobre la 'otra' historia de España. *RIFOP : Revista interuniversitaria de formación del profesorado: continuación de la antigua Revista de Escuelas Normales* 141–162 (2006).

176. Soyer, F. *The Persecution of the Jews and Muslims of Portugal: King Manuel I and the End of Religious Tolerance (1496-7)*. (BRILL, 2007).

177. Tartakoff, P. *Between Christian and Jew: Conversion and Inquisition in the Crown of Aragon, 1250-1391*. (University of Pennsylvania Press, 2012).

178. Kimmel, S. *Parables of Coercion: Conversion and Knowledge at the End of Islamic Spain*. (University of Chicago Press, 2015). doi:10.7208/9780226278315.

179. Carr, M. *Blood and Faith: The Purging of Muslim Spain, 1492-1614*. (Oxford University Press, 2017).

180. Abreu, L. Beggars, Vagrants and Romanies : Repression and Persecution in Portuguese Society (14th–18th Centuries). *Hygiea* **6**, 41–66 (2007).

181. Marques, A. de O. História de Portugal, vol. I. *I, pág* **126**, 40–41 (1976).

182. Serrão, J. Dicionário de história de Portugal. (1963).

183. Idoate, F. Los gitanos en Navarra. *Príncipe de Viana* **10**, 443–474 (1949).

184. Leblon, B. *Les Gitans d'Espagne: Le prix de la différence*. (FeniXX, 1985).

185. Ortega, M. H. S. Los gitanos españoles desde su salida de la India hasta los primeros conflictos en la península. *Espacio Tiempo y Forma. Serie IV, Historia Moderna* (1994).

186. Martínez Dhier, A. La condición social y jurídica de los gitanos en la legislación histórica española.(A partir de la pragmática de los Reyes Católicos de 1499). (2007).

187. Castro, D. S. Política de imigração e as minorias étnicas durante o estado novo: o caso dos ciganos. *Conversas & Controvérsias* **2**, (2011).

188. Moonen, F. A História esquecida dos Ciganos no Brasil. *Sæculum-Revista de História* **2**, 123–138 (1996).

189. Bover, O. & Velilla, P. Migration in Spain: Historical Background and Current Trends. SSRN Scholarly Paper at https://doi.org/10.2139/ssrn.219871 (1999).

190. Silvestre, J. Internal migrations in Spain, 1877–1930. *European Review of Economic History* **9**, 233–265 (2005).

191. Bastos Pereira, J. G. From negative visibility to positive invisibility–who needs to use Gypsies [Roma] as the anti-citizens of Civilized Europe? *Journal of Gypsy Studies* **2**, 3–38 (2020).

192. Arias, D. L. Anti-sociedad gitana. Aceptabilidad y esquizofrenia en el romanticismo, el franquismo y el multiculturalismo. *Imago crítica. Revista de Antropología, Comunicación y Estudios Culturales* 177–188 (2022).

193. Pereira Bastos, S. O Estado Novo e os seus Vadios. *Contribuição para o Estudo das Identidades Marginais e da Sua Repressão, Lisboa, Pub. Dom Quixote* (1997).

194. Gamella, J. F., Fernández, C., Nieto, M. & Adiego, I.-X. La agonía de una lengua. Lo que queda del caló en el habla de los gitanos. Parte I. Métodos, fuentes y resultados generales. *Gazeta de antropología* **27**, (2011).

195. Laparra, M., Arza, J. & Fernández, A. Diagnóstico social de la comunidad gitana en España. *Un análisis contrastado de la Encuesta del CIS a Hogares de Población Gitana* **2011**, (2007).

196. Laparra, M. Informe sobre la situación social y tendencias de cambio en la población gitana. *Una primera aproximación. Madrid: Ministerio de Trabajo y Asuntos Sociales* (2007).

197. Tarnovschi, D., Preoteasa, A. M., Pamporov, A., Kabakchieva, P. & Gitano, F. S. Roma from Romania, Bulgaria, Italy and Spain between Social Inclusion and Migration. *Bucharest: Soros Foundation Romania* (2012).

198. La Parra-Casado, D., Mosquera, P. A., Vives-Cases, C. & San Sebastian, M. Socioeconomic Inequalities in the Use of Healthcare Services: Comparison between the Roma and General Populations in Spain. *International Journal of Environmental Research and Public Health* **15**, 121 (2018).

199. Mendes, M. M. & Magano, O. Roma/Ciganos and the condition of internal strange in Portuguese society: The construction of otherness. in *Simmel and Beyond* (Routledge, 2022).

200. Pinto, F., Malo, M. Á. & Martínez, R. Ethnical incomes' gap in Spain. The case of Roma. *Social Science Quarterly* **105**, 1383–1395 (2024).

201. Messing, V. Patterns of Roma employment in Europe. *Neujobs Policy Brief* 4 (2014).

202. Pinto, F., Malo, M. Á. & Martínez, R. Ethnical incomes' gap in Spain. The case of Roma. *Social Science Quarterly* **105**, 1383–1395 (2024).

203. Eder-Jordan, B. Oral and Written Šukar Laviben of the Roma: The Beginning of a Romani Literary Historiography. in *Or Words to That Effect: Orality and the writing of literary history* 194–205 (John Benjamins Publishing Company, 2016).

204. Petráš, R. The Creation and Legal Status of the Slovak, Roma and Vietnamese Minorities in the Czech Republic. *Acta Humana–Emberi Jogi Közlemények* **3**, 39-46-39–46 (2015).

205. Roma in the Czech Republic Foreigners in Their Own Land. *Refworld* https://www.refworld.org/reference/countryrep/hrw/1996/en/21907.

206. Matlovičová, K., Matlovič, R., Mušinka, A. & Židová, A. The Roma population in Slovakia. *Basic characteristics of the Roma population with emphasis on the spatial aspects of its differentiation In: Penczes, J., Radics, Z.(Eds.): Roma popuation on the peripheries of the Visegrad countries. Spatial trends and social challenges. Debrecen* 77–104 (2012).

207. Kalvoda, J. The gypsies of Czechoslovakia. *Nationalities Papers* (1991) doi:10.1080/00905999108408204.

208. Kalibová, K. Gypsies in the Czech Republic and the Slovak Republic: Geographic and demographic characteristics. *GeoJournal* **30**, 255–258 (1993).

209. Ethnicity. *Census 2021* https://scitani.gov.cz/ethnicity.

210. Covrig, A. Why Roma do not Declare their Identity - Careful Decision or Unpremeditated Refusal? *Journal for the Study of Religions and Ideologies* **3**, 90–101 (2010).

211. Cviklova, L. Direct and indirect racial discrimination of Roma people in Bulgaria, the Czech Republic and the Russian Federation. *Ethnic and Racial Studies* **38**, 2140–2155 (2015).

212. Slepickova, L. & Bobakova, D. F. Conceptualization of Roma in policy documents related to social inclusion and health in the Czech Republic. *International Journal of Environmental Research and Public Health* **17**, 7739 (2020).

213. Miko, I. Gregor Mendel and the principles of inheritance. *Nature Education* **1**, 134 (2008).

214. Hirszfeld, L. *Konstitutionsserologie Und Blutgruppenforschung.* (Springer, 1928).

215. Kitchin, D. H. Blood tests: the four blood groups. *The Police Journal* **6**, 51–63 (1933).

216. Cappell, D. F. The Blood Groups of Mankind: A New Approach to Anthropology *Presidential Address to the Royal Medico-Chirurgical Society of Glasgow, on 24th October, 1947. *Glasgow Med J* **28**, 397–413 (1947).

217. Morgan, W. T. J. Occurrence and nature of human blood-group substances. *British Medical Bulletin* **2**, 165–168 (1944).

218. MacFarlane, E. W. E. Blood groups of Eurasians and Parsees in Calcutta. *The American Naturalist* **76**, 520–526 (1942).

219. Lundman, B. Geography of human blood groups (A, B, O system). *Evolution* **2**, 231–237 (1948).

220. Beckman, L. & Takman, J. On the anthropology of a Swedish Gypsy population. *Hereditas* **53**, 272–280 (1965).

221. Gáliková, J., Vilímová, M., Ferák, V. & Mayerová, A. Haptoglobin Types in Gypsies from Slovakia (Czechoslovakia). *Human Heredity* **19**, 480–485 (1969).

222. Seyedna, S. Y. *Some Serological Attributes of the European Gypsies. M. Sc. thesis* (University of Durham England, 1979).

223. Avčin, M. Gypsy isolates in Slovenia. *Journal of Biosocial Science* **1**, 221–233 (1969).

224. Mastana, S. S. & Papiha, S. S. Origin of the Romany gypsies — genetic evidence. *Zeitschrift für Morphologie und Anthropologie* **79**, 43–51 (1992).

225. Lasa, A. *et al.* Severe limb girdle muscular dystrophy in Spanish gypsies: further evidence for a founder mutation in the γ-sarcoglycan gene. *Eur J Hum Genet* **6**, 396–399 (1998).

226. Piccolo, F. *et al.* A Founder Mutation in the γ-Sarcoglycan Gene of Gypsies Possibly Predating Their Migration Out of India. *Human Molecular Genetics* **5**, 2019–2022 (1996).

227. Martínez-Frías, M. L. & Bermejo, E. Prevalence of congenital anomaly syndromes in a Spanish gypsy population. *J Med Genet* **29**, 483–486 (1992).

228. Kalaydjieva, L. *et al.* Patterns of inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages. *Eur J Hum Genet* **9**, 97–104 (2001).

229. Peričić, M. *et al.* Population genetics of 8 Y chromosome STR loci in Macedonians and Macedonian Romani (Gypsy). *Forensic Science International* **154**, 257–261 (2005).

230. Gusmão, A. *et al.* A Perspective on the History of the Iberian Gypsies Provided by Phylogeographic Analysis of Y-Chromosome Lineages. *Annals of Human Genetics* **72**, 215–227 (2008).

231. Mendizabal, I. *et al.* Reconstructing the Population History of European Romani from Genome-wide Data. *Current Biology* **22**, 2342–2349 (2012).

232. Bartsocas, C. S. *et al.* Genetic structure of the Greek gypsies. *Clinical genetics* **15**, 5–10 (1979).

233. Siváková, D. Distribution of three red-cell enzyme polymorphisms (ACP, PGM1 and AK) in gypsies from Slovakia (Czechoslovakia). *Annals of Human Biology* **10**, 449–452 (1983).

234. Van Loghem, E., Tauszik, T., Hollan, S. & Nijenhuis, L. E. Immunoglobulin Allotypes in Hungarian Gypsies. *International Journal of Immunogenetics* **12**, 131–137 (1985).

235. de Pablo, R. *et al.* Distribution of HLA antigens in Spanish Gypsies: A comparative study. *Tissue Antigens* **40**, 187–196 (1992).

236. Abicht, A. *et al.* A common mutation (ε1267delG) in congenital myasthenic patients of Gypsy ethnic origin. *Neurology* **53**, 1564–1564 (1999).

237. Moorjani, P. *et al.* Reconstructing Roma History from Genome-Wide Data. *PLOS ONE* **8**, e58633 (2013).

238. Martínez-Cruz, B. *et al.* Origins, admixture and founder lineages in European Roma. *Eur J Hum Genet* **24**, 937–943 (2016).

239. Font-Porterias, N. *et al.* European Roma groups show complex West Eurasian admixture footprints and a common South Asian genetic origin. *PLOS Genetics* **15**, e1008417 (2019).

240. Melegh, B. I., Banfai, Z., Hadzsiev, K., Miseta, A. & Melegh, B. Refining the South Asian Origin of the Romani people. *BMC Genetics* **18**, 82 (2017).

241. Bianco, E. *et al.* Recent Common Origin, Reduced Population Size, and Marked Admixture Have Shaped European Roma Genomes. *Molecular Biology and Evolution* **37**, 3175–3187 (2020).

242. Adnan, A. *et al.* Are Roma People Descended from the Punjab Region of Pakistan: A Y-Chromosomal Perspective. *Genes* **13**, 532 (2022).

243. Malyarchuk, B. A. *et al.* Mitochondrial DNA variability in Slovaks, with application to the Roma origin. *Annals of human genetics* **72**, 228–240 (2008).

244. Aizpurua-Iraola, J., Giménez, A., Carballo-Mesa, A., Calafell, F. & Comas, D. Founder lineages in the Iberian Roma mitogenomes recapitulate the Roma diaspora and show the effects of demographic bottlenecks. *Sci Rep* **12**, 18720 (2022).

245. Morar, B. *et al.* Mutation History of the Roma/Gypsies. *The American Journal of Human Genetics* **75**, 596–609 (2004).

246. Gresham, D. Genetic variation and disease in the Roma (Gypsies). *Theses: Doctorates and Masters* (2001).

247. Gusmão, A. *et al.* A genetic historical sketch of European Gypsies: The perspective from autosomal markers. *American Journal of Physical Anthropology* **141**, 507–514 (2010).

248. Bánfai, Z. *et al.* Revealing the Genetic Impact of the Ottoman Occupation on Ethnic Groups of East-Central Europe and on the Roma Population of the Area. *Frontiers in Genetics* **10**, 558 (2019).

249. Bánfai, Z. *et al.* Revealing the impact of the Caucasus region on the genetic legacy of Romani people from genome-wide data. *PLOS ONE* **13**, e0202890 (2018).

250. Kalaydjieva, L., Gresham, D. & Calafell, F. Genetic studies of the Roma (Gypsies): a review. *BMC Med Genet* **2**, 5 (2001).

251. Kalaydjieva, L., Morar, B., Chaix, R. & Tang, H. A newly discovered founder population: the Roma/Gypsies. *BioEssays* **27**, 1084–1094 (2005).

252. Lohmueller, K. E. The distribution of deleterious genetic variation in human populations. *Current Opinion in Genetics & Development* **29**, 139–146 (2014).

253. Machha, P. *et al.* Endogamy and high prevalence of deleterious mutations in India: evidence from strong founder events. *Journal of Genetics and Genomics* (2025) doi:10.1016/j.jgg.2025.02.001.

254. Pedersen, C.-E. T. *et al.* The Effect of an Extreme and Prolonged Population Bottleneck on Patterns of Deleterious Variation: Insights from the Greenlandic Inuit. *Genetics* **205**, 787–801 (2017).

255. Lim, E. T. *et al.* Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS genetics* **10**, e1004494 (2014).

256. Behar, D. M. *et al.* MtDNA evidence for a genetic bottleneck in the early history of the Ashkenazi Jewish population. *Eur J Hum Genet* **12**, 355–364 (2004).

257. Gresham, D. *et al.* Origins and Divergence of the Roma (Gypsies). *The American Journal of Human Genetics* **69**, 1314–1331 (2001).

258. Pereira, V. *et al.* Refining the genetic portrait of Portuguese Roma through X-chromosomal markers. *American Journal of Physical Anthropology* **148**, 389–394 (2012).

259. Mendizabal, I. *et al.* Reconstructing the Indian Origin and Dispersal of the European Roma: A Maternal Genetic Perspective. *PLOS ONE* **6**, e15988 (2011).

260. García-Fernández, C. *et al.* Sex-biased patterns shaped the genetic history of Roma. *Sci Rep* **10**, 14464 (2020).

261. Harper, P. S. Genetic Variation in Wales. *J R Coll Physicians Lond* **10**, 321–332 (1976).

262. Williams, E. M. & Harper, P. R. Genetic study of Welsh gypsies. *Journal of Medical Genetics* **14**, 172–176 (1977).

263. Genĉík, A. Epidemiology and genetics of primary congenital glaucoma in Slovakia. Description of a form of primary congenital glaucoma in gypsies with autosomal-recessive inheritance and complete penetrance. *Dev Ophthalmol* **16**, 76–115 (1989).

264. Genčík, A., Genčíková, A. & Gerinec, A. Genetic heterogeneity of congenital glaucoma. *Clinical Genetics* **17**, 241–248 (1980).

265. Thomas, JamesD., Doucette, MargaretM., Catanzano Thomas, D. & Stoeckle, JohnD. DISEASE, LIFESTYLE, AND CONSANGUINITY IN 58 AMERICAN GYPSIES. *The Lancet* **330**, 377–379 (1987).

266. Tournev, I. *et al.* Congenital cataracts facial dysmorphism neuropathy syndrome, a novel complex genetic disease in Balkan gypsies: Clinical and electrophysiological observations. *Annals of Neurology* **45**, 742–750 (1999).

267. Kalaydjieva, L. *et al.* A Founder Mutation in the GK1 Gene Is Responsible for Galactokinase Deficiency in Roma (Gypsies). *The American Journal of Human Genetics* **65**, 1299–1307 (1999).

268. Todorova, A., Ashikov, A., Beltcheva, O., Tournev, I. & Kremensky, I. C283Y mutation and other C-terminal nucleotide changes in the γ-

sarcoglycan gene in the Bulgarian gypsy population. *Human Mutation* **14**, 40–44 (1999).

269. Amorim, C. E. G. *et al.* The population genetics of human disease: The case of recessive, lethal mutations. *PLoS genetics* **13**, e1006915 (2017).

270. Saint Pierre, A. & Génin, E. How important are rare variants in common disease? *Briefings in Functional Genomics* **13**, 353–361 (2014).

271. Momozawa, Y. & Mizukami, K. Unique roles of rare variants in the genetics of complex diseases in humans. *Journal of human genetics* **66**, 11–23 (2021).

272. Morar, B., Azmanov, D. N. & Kalaydjieva, L. Roma (Gypsies): Genetic Studies. in *eLS* (John Wiley & Sons, Ltd, 2013). doi:10.1002/9780470015902.a0006239.pub3.

273. Alliance, G. & Health, D. of C. D. of. Classic Mendelian Genetics (Patterns of Inheritance). in *Understanding Genetics: A District of Columbia Guide for Patients and Health Professionals* (Genetic Alliance, 2010).

274. Chial, H. Mendelian genetics: patterns of inheritance and single-gene disorders. *Nature Education* **1**, 63 (2008).

275. Schork, N. J. Genetics of Complex Disease. *Am J Respir Crit Care Med* **156**, S103–S109 (1997).

276. Hamamy, H. Consanguineous marriages. *J Community Genet* **3**, 185–192 (2012).

277. Lathrop, M. & Pison, G. Méthode statistique d'étude de l'endogamie. Application à l'étude du choix du conjoint chez les Peul Bandé. *Population (french edition)* 513–541 (1982).

278. Martin, E. & Gamella, J. F. Marriage practices and ethnic differentiation: The case of Spanish Gypsies (1870–2000). *The History of the family* **10**, 45–63 (2005).

279. Gamella, J. F. & Muntean, V. M. Marriage and the reproductive regime of a digitally connected Roma diaspora. *Journal of Contemporary Central and Eastern Europe* **31**, 533–559 (2023).

280. Ben Halim, N. *et al.* Consanguinity, endogamy, and genetic disorders in Tunisia. *J Community Genet* **4**, 273–284 (2013).

281. Bittles, A. H. & Hamamy, H. A. Endogamy and consanguineous marriage in Arab populations. *Genetic disorders among Arab populations* 85–108 (2010).

282. Utomo, A. J. & McDonald, P. F. Internal migration, group size, and ethnic endogamy in Indonesia. *Geographical Research* **59**, 56–77 (2021).

283. Rudan, I. *et al.* Effects of inbreeding, endogamy, genetic admixture, and outbreeding on human health: a "1001 Dalmatians" study. *Croatian medical journal* **47**, 601–610 (2006).

284. Hernandez, M., Garcia-Moro, C., Moral, P. & Gonzalez-Martin, A. Population evolution in 20th-century Easter Island: Endogamy and admixture. *Human Biology* 359–377 (2000).

285. Ben Halim, N. *et al.* Consanguinity, endogamy, and genetic disorders in Tunisia. *Journal of community genetics* **4**, 273–284 (2013).

286. Romdhane, L. *et al.* Consanguinity and inbreeding in health and disease in North African populations. *Annual review of genomics and human genetics* **20**, 155–179 (2019).

287. Quinn, S. *et al.* Catalogue of inherited autosomal recessive disorders found amongst the Roma population of Europe. *European Journal of Medical Genetics* **73**, 104989 (2025).

288. Dobranici, M., Buzea, A. & Popescu, R. The cardiovascular risk factors of the Roma (gypsies) people in Central-Eastern Europe: a review of the published literature. *Journal of medicine and life* **5**, 382 (2012).

289. Mendizabal, I., Lao, O., Marigorta, U. M., Kayser, M. & Comas, D. Implications of Population History of European Romani on Genetic Susceptibility to Disease. *HHE* **76**, 194–200 (2013).

290. Font-Porterias, N., Giménez, A., Carballo-Mesa, A., Calafell, F. & Comas, D. Admixture Has Shaped Romani Genetic Diversity in Clinically Relevant Variants. *Front Genet* **12**, 683880 (2021).

291. Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M. & Wilson, J. F. Runs of homozygosity: windows into population history and trait architecture. *Nat Rev Genet* **19**, 220–234 (2018).

292. Sussman, R. W. *The Myth of Race: The Troubling Persistence of an Unscientific Idea.* (Harvard University Press, 2014).

293. Gabbert, W. Concepts of ethnicity. *Latin American and Caribbean Ethnic Studies* **1**, 85–103 (2006).

294. Baumann, T. Defining ethnicity. *The SAA archaeological record* **4**, 12–14 (2004).

295. Mackie, G., Moneti, F., Shakya, H. & Denny, E. What are social norms? How are they measured. *University of California at San Diego-UNICEF Working Paper, San Diego* (2015).

296. Bicchieri, C., Muldoon, R. & Sontuoso, A. Social Norms. in *The Stanford Encyclopedia of Philosophy* (eds. Zalta, E. N. & Nodelman, U.) (Metaphysics Research Lab, Stanford University, 2023).

297. Legros, S. & Cislaghi, B. Mapping the Social-Norms Literature: An Overview of Reviews. *Perspect Psychol Sci* **15**, 62–80 (2020).

298. Valsiner, J. *The Oxford Handbook of Culture and Psychology.* (Oxford University Press, 2013).

299. Frese, M. Cultural practices, norms, and values. *Journal of Cross-Cultural Psychology* **46**, 1327–1330 (2015).

300. Kwaśniewska, A. The Social Consequences of Identifying a Genetic Disease with an Endogamous Ethnic Group: The Kashubian Case. *Ethnologia Europaea* **52**, (2022).

301. Munté, A., Serradell, O. & Sordé, T. From research to policy: Roma participation through communicative organization. *Qualitative Inquiry* **17**, 256–266 (2011).

302. Aiello, E., Mondejar, E. & Pulido, M. A. Communicative methodology of research and recognition of the Roma people. *International Review of Qualitative Research* **6**, 254–265 (2013).

303. 2020-2030 EU Roma Strategic Framework. *European Commission - European Commission* https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1813.

304. Condon, L. *et al.* Engaging gypsy, Roma, and traveller communities in research: Maximizing opportunities and overcoming challenges. *Qualitative Health Research* **29**, 1324–1333 (2019).

305. Gómez-Carballa, A. *et al.* Indian Signatures in the Westernmost Edge of the European Romani Diaspora: New Insight from Mitogenomes. *PLOS ONE* **8**, e75397 (2013).

6. ANNEXES

6.1 Supplementary Information: The genetic footprint of the European Roma diaspora: Evidence from the Balkans to the Iberian Peninsula

Authors

Giacomo Francesco Ena, Aaron Giménez, Annabel Carballo-Mesa, Petra Lišková, Marcos Araújo Castro e Silva & David Comas.

The genetic footprint of the European Roma diaspora: Evidence from the Balkans to the Iberian Peninsula

*Giacomo F. Ena1 (ORCID: 0000-0002-7749-5182), Aaron Giménez2, Annabel Carballo-Mesa3, Petra Lišková4,5 (ORCID: 0000-0003-3519-643X), Marcos Araújo Castro e Silva1 (ORCID: 0000-0002-9873-3717), David Comas1 (ORCID: 0000-0002-5075-0956), **

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

2 Facultat de Sociologia, Universitat Autònoma de Barcelona, Barcelona, Spain

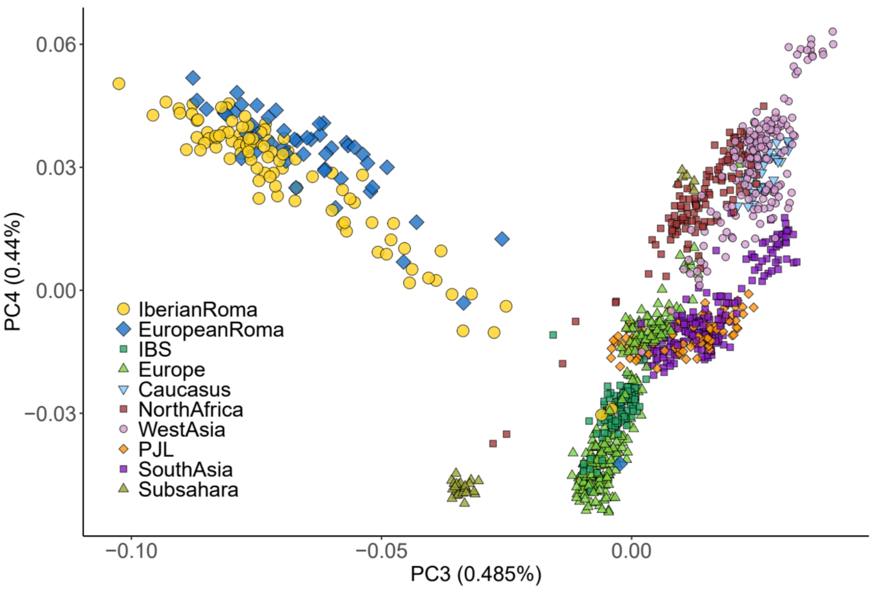
3 Facultat de Geografia i Història, Universitat de Barcelona, Barcelona, Spain

4 Department of Paediatrics and Inherited Metabolic Disorders, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic

5 Department of Ophthalmology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic

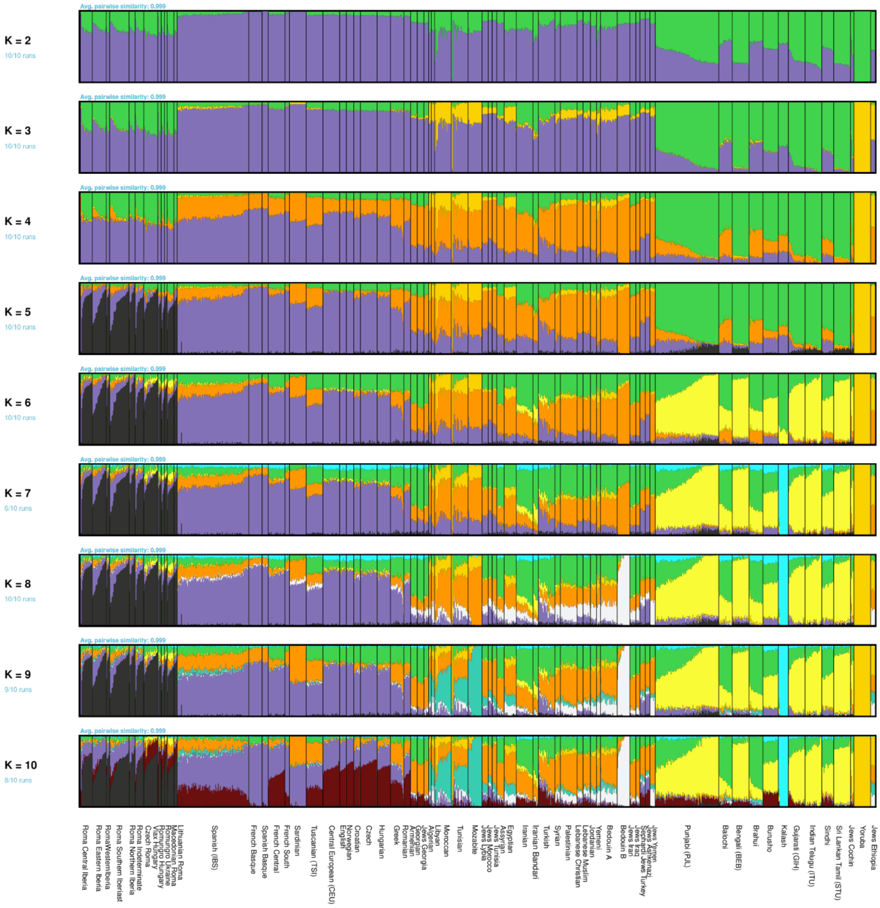
** Corresponding author*

E-mail: david.comas@upf.edu



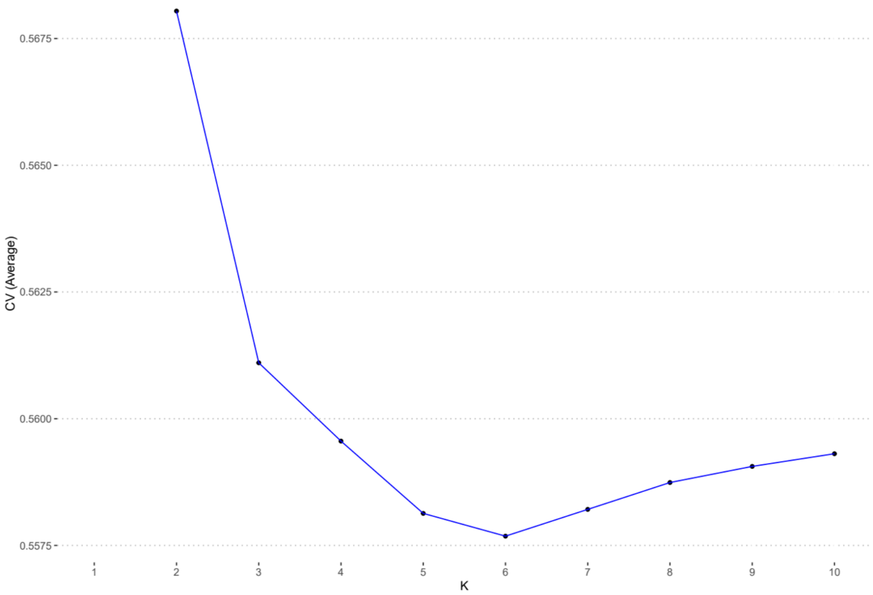
Supplementary Figure 1 Principal Component Analysis on an Afro-Eurasian scale.

PCA displaying Principal Components 3 and 4, including all Roma samples and references analysed in this study.

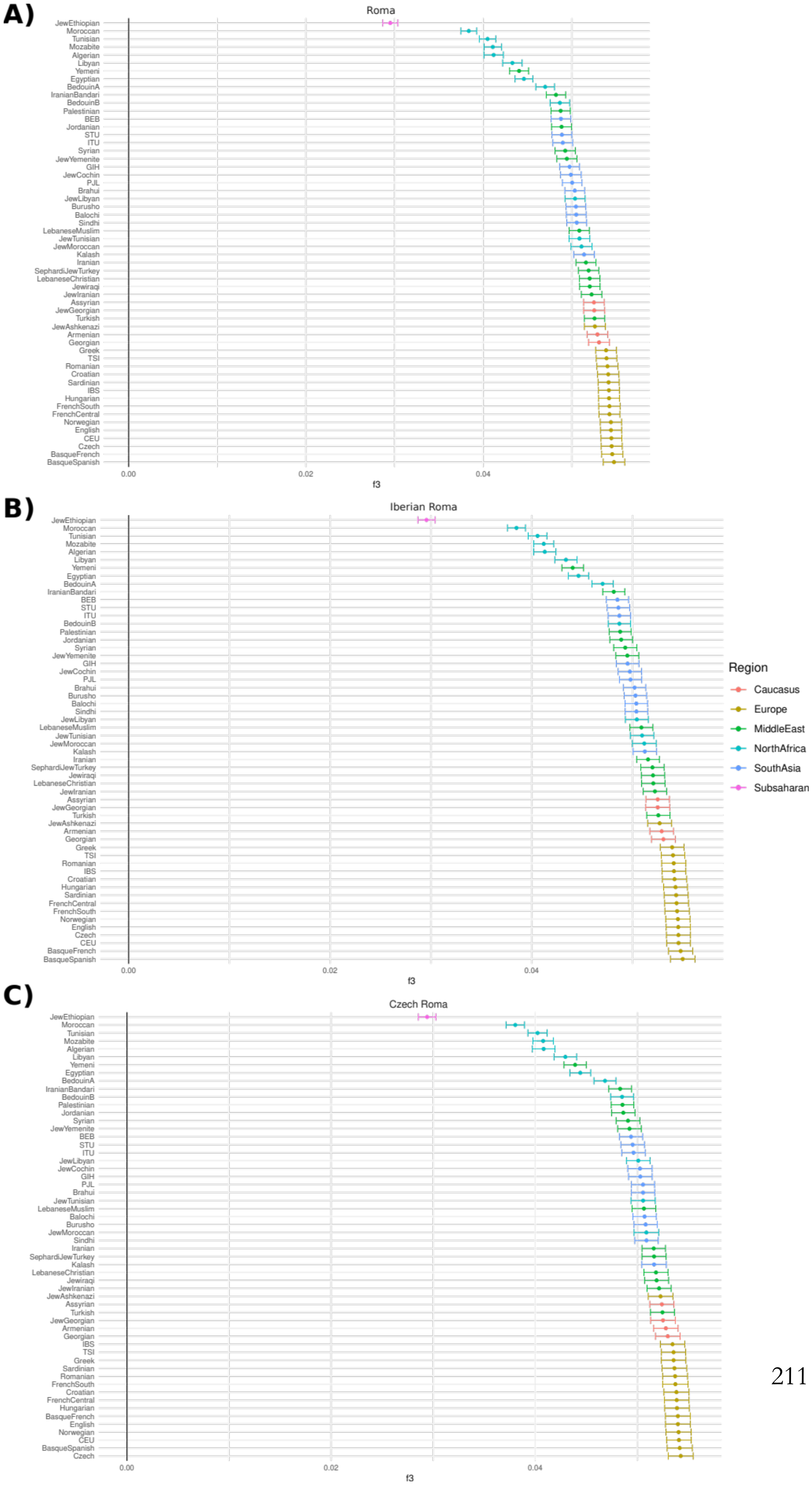


Supplementary Figure 2 ADMIXTURE results on the whole dataset.

Results from all ADMIXTURE runs performed on the main dataset, with values of K ranging from 2 to 10.

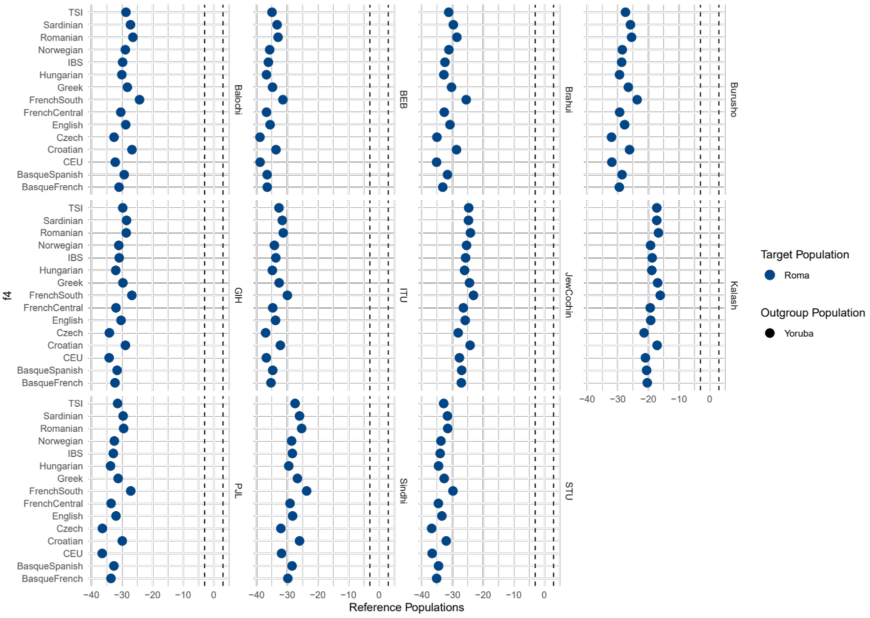


Supplementary Figure 3 Plot of cross-validation values for ADMIXTURE on the whole dataset.

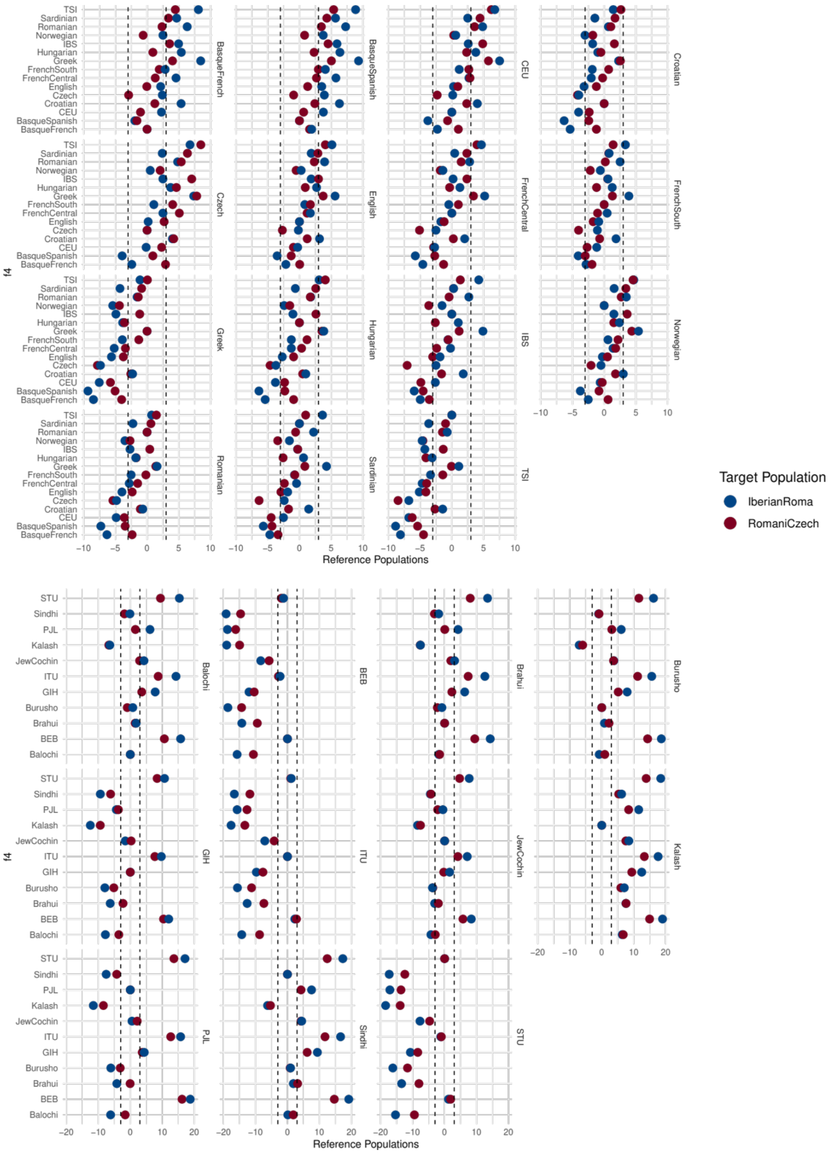

A)
Roma
JewEthiopian
Moroccan
Tunisian
Mozabite
Algerian
Libyan
Yemeni
Egyptian
BedouinA
IranianBandari
BedouinB
Palestinian
BEB
Jordanian
STU
ITU
Syrian
JewYemenite
GIH
JewCochin
PJL
Brahui
JewLibyan
Burusho
Balochi
Sindhi
LebaneseMuslim
JewTunisian
JewMoroccan
Kalash
Iranian
SephardiJewTurkey
LebaneseChristian
Jewiraqi
JewIranian
Assyrian
JewGeorgian
Turkish
JewAshkenazi
Armenian
Georgian
Greek
TSI
Romanian
Croatian
Sardinian
IBS
Hungarian
FrenchSouth
FrenchCentral
Norwegian
English
CEU
Czech
BasqueFrench
BasqueSpanish
0.00
0.02
0.04
f3
B)
Iberian Roma
JewEthiopian
Moroccan
Tunisian
Mozabite
Algerian
Libyan
Yemeni
Egyptian
BedouinA
IranianBandari
BEB
STU
ITU
BedouinB
Palestinian
Jordanian
Syrian
JewYemenite
GIH
JewCochin
PJL
Brahui
Burusho
Balochi
Sindhi
JewLibyan
LebaneseMuslim
JewTunisian
JewMoroccan
Kalash
Iranian
SephardiJewTurkey
Jewiraqi
LebaneseChristian
JewIranian
Assyrian
JewGeorgian
Turkish
JewAshkenazi
Armenian
Georgian
Greek
TSI
Romanian
IBS
Croatian
Hungarian
Sardinian
FrenchCentral
FrenchSouth
Norwegian
English
Czech
CEU
BasqueFrench
BasqueSpanish
0.00
0.02
0.04
f3
Region
Caucasus
Europe
MiddleEast
NorthAfrica
SouthAsia
Subsaharan
C)
Czech Roma
JewEthiopian
Moroccan
Tunisian
Mozabite
Algerian
Libyan
Yemeni
Egyptian
BedouinA
IranianBandari
BedouinB
Palestinian
Jordanian
Syrian
JewYemenite
BEB
STU
ITU
JewLibyan
JewCochin
GIH
PJL
Brahui
JewTunisian
LebaneseMuslim
Balochi
Burusho
JewMoroccan
Sindhi
Iranian
SephardiJewTurkey
Kalash
LebaneseChristian
Jewiraqi
JewIranian
JewAshkenazi
Assyrian
Turkish
JewGeorgian
Armenian
Georgian
IBS
TSI
Greek
Sardinian
Romanian
FrenchSouth
Croatian
FrenchCentral
Hungarian
BasqueFrench
English
Norwegian
CEU
BasqueSpanish
Czech
0.00
0.02
0.04
f3

Supplementary Figure 4 f3 outgroup statistical test for admixture.

(a) f3 outgroup test in the form f3(Yoruba, Roma, Reference), incorporating all reference populations. (b) f3 outgroup test in the form f3(Yoruba, Iberian Roma, Reference), incorporating all reference populations. (c) f3 outgroup test in the form f3(Yoruba, Czech Roma, Reference), incorporating all reference populations.

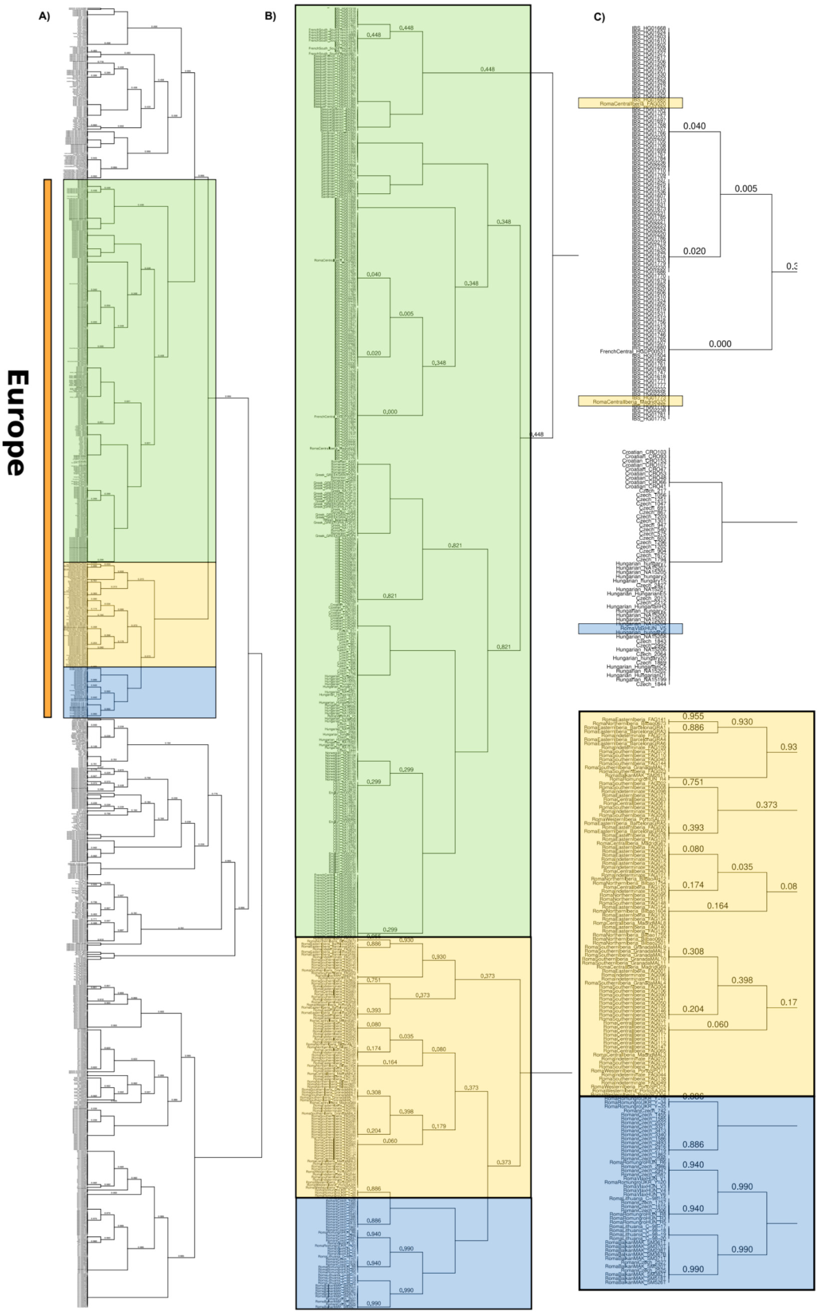


Supplementary Figure 5 f4 statistical test for shared drift. f4 test in the form f4(Yoruba, Roma, reference, reference), including only references from Europe and South Asia.



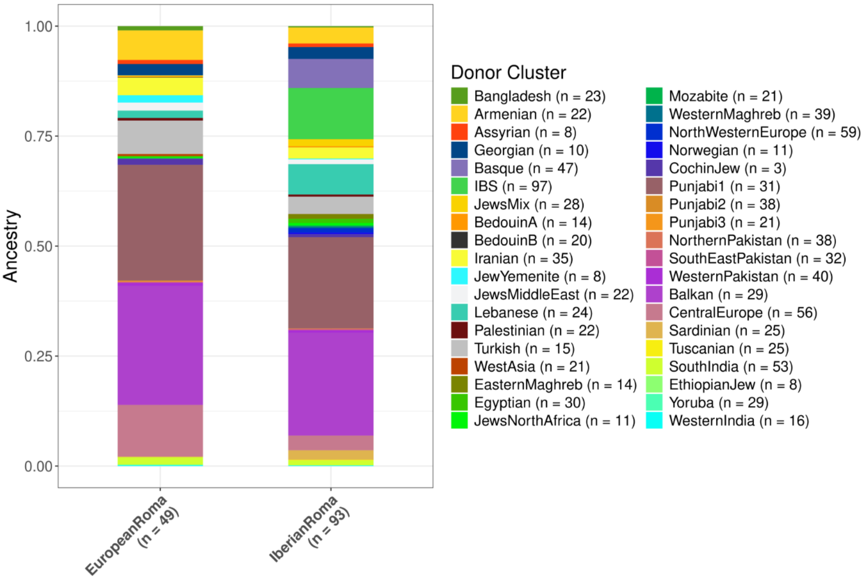
Supplementary Figure 6 f4 statistical test for shared drift.

(a) f4 test including only the European reference populations. (b) f4 test including only the Asian reference populations.

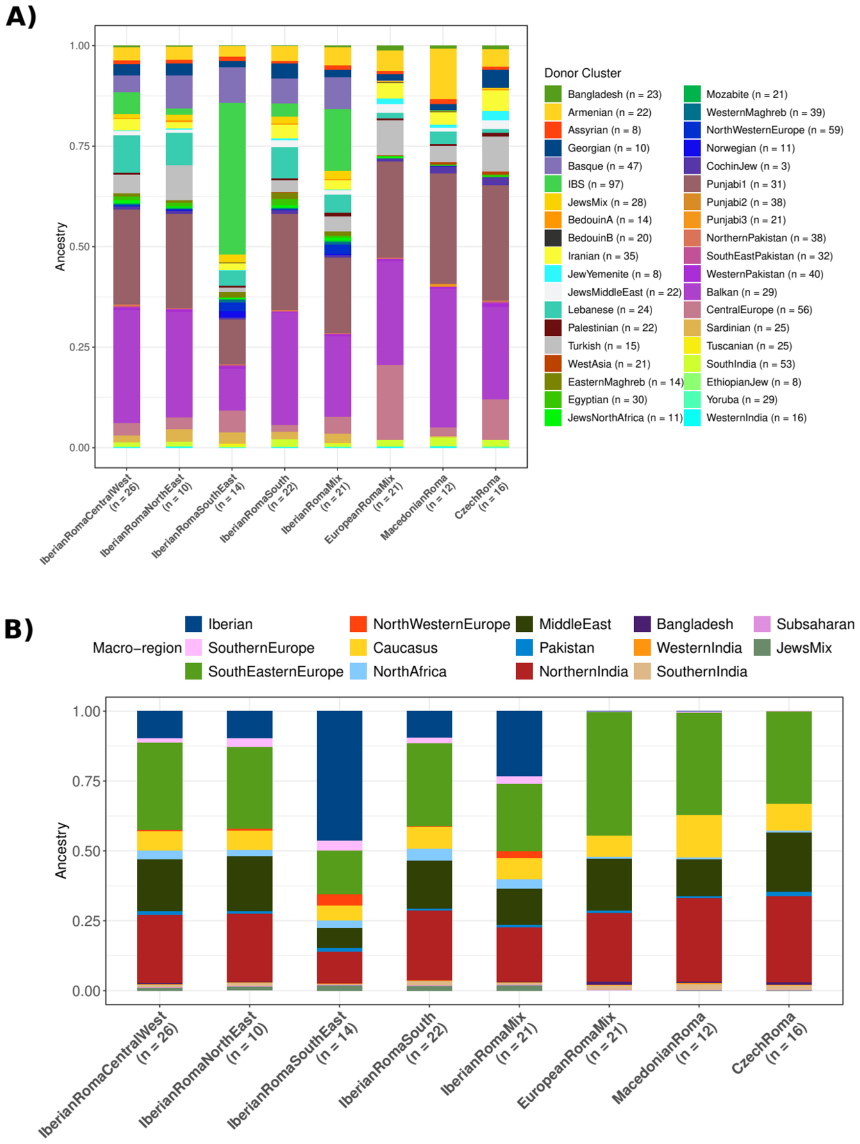

A)
Europe
B)
C)
0.040
0.005
0.020
0.000
RomaCentralIberia_FA0207
FrenchCentral_HG01607
RomaCentralIberia_Madrid022
RomaVlaxHUN_V5
0.955
0.930
0.886
0.93
0.751
0.373
0.393
0.080
0.035
0.174
0.08
0.164
0.308
0.398
0.204
0.17
0.060
0.886
0.940
0.990
0,448
0,348
0,040
0,005
0,020
0,000
0,821
0,299
0,930
0,886
0,751
0,373
0,393
0,080
0,035
0,174
0,164
0,308
0,398
0,204
0,179
0,060
0,940
0,990

Supplementary Figure 7 Composition of the fineSTRUCTURE dendrogram.

(a) Whole-tree: European populations are shown in green, the Iberian Roma branch in yellow, and the other European Roma branch in blue. (b) Detailed view of the Europe macro-branch, highlighting the three Roma individuals that cluster outside the Roma macro-branches. (c) Details of the Roma macro-branches, including the section containing the three separated Roma individuals.

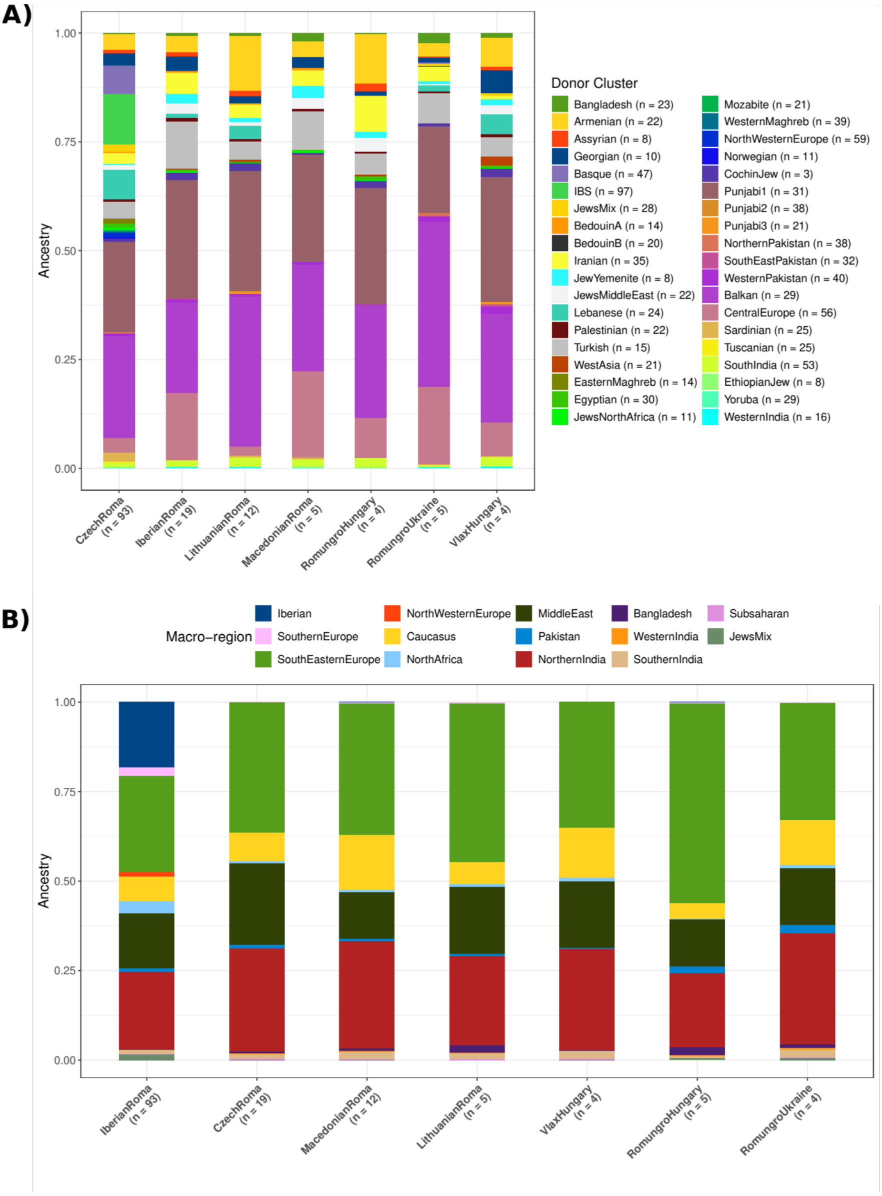


Supplementary Figure 8 Composition of the Roma genetic profile based on NNLS analysis. *NNLS results derived from the classification of Roma populations based on genetic clustering. The analysis grouped all Roma individuals into two distinct clusters.*



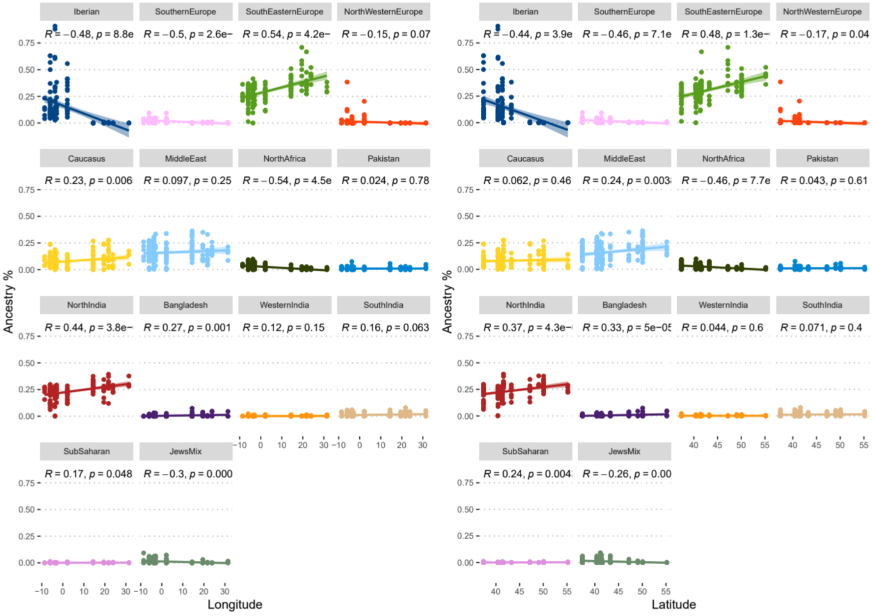
Supplementary Figure 9 Results of the NNLS analysis on the subclusters.

(a) NNLS results categorised by genetic cluster, grouped by donor cluster. (b) NNLS results categorised by genetic cluster, grouped by donor macro-region.

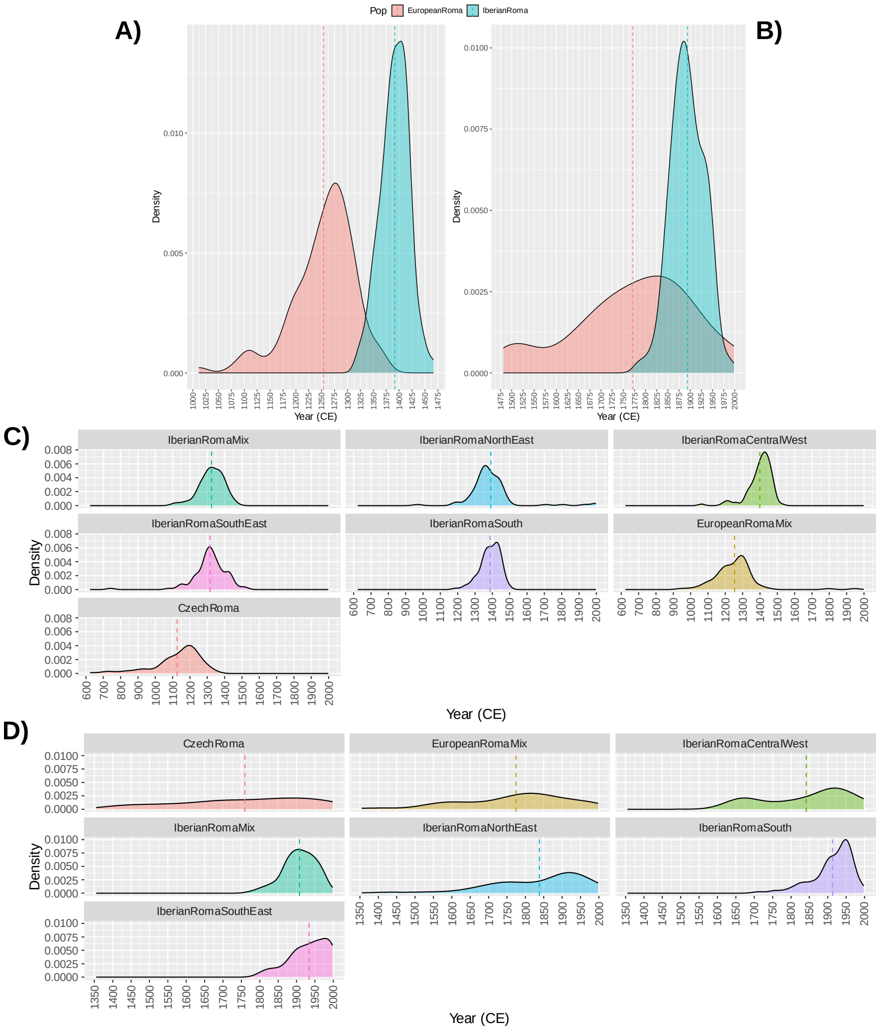


Supplementary Figure 10 Results of the NNLS analysis on geographical regions. (a) NNLS results grouped by donor clusters. (b) NNLS results grouped by donor macro-region.

Supplementary Figure 11 Linear regression plot and Pearson correlation coefficients related to the gradient in NNLS ancestry components.



Pearson correlation analysis based on longitude and latitude, conducted across all donor clusters grouped into macro-regions for each Roma individual.

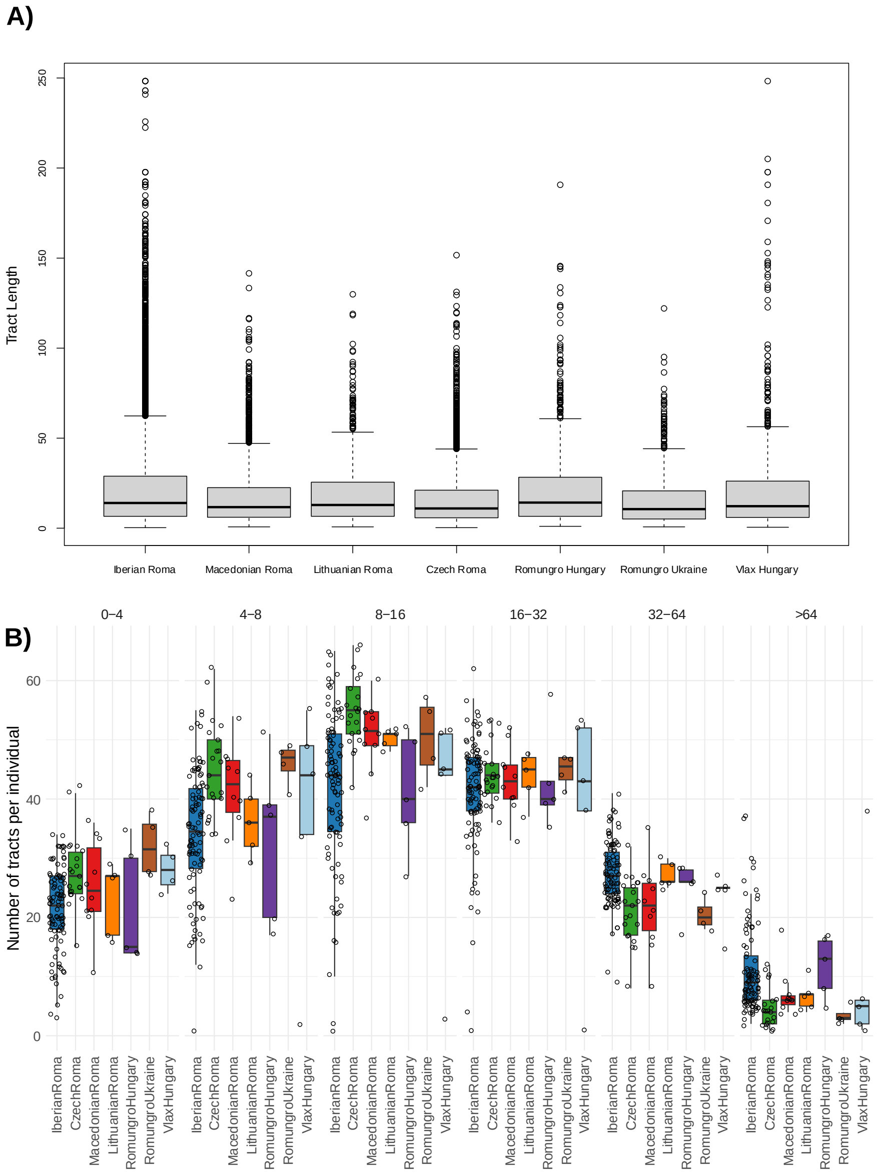


Supplementary Figure 12 Bootstrap analysis of the fastGLOBETROTTER estimated dates for recipient genetic clusters. Vertical lines represent the average admixture time for each cluster, with dates expressed in years (CE).

(a) Bootstrap analysis of Date 1 for the main Roma clusters. (b) Bootstrap analysis of Date 2 for the main Roma clusters. (c) Bootstrap analysis of Date 1 for subclusters within the main Roma clusters. (d) Bootstrap analysis of Date 2 for subclusters within the main Roma clusters.

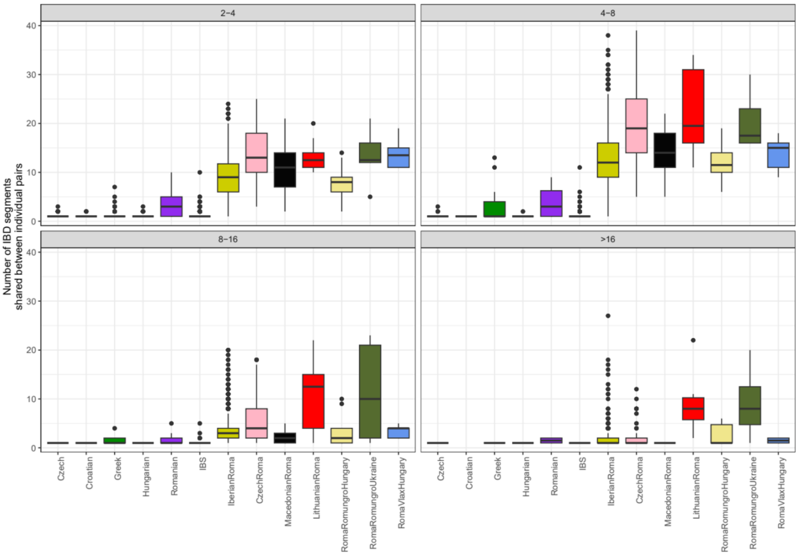
Supplementary Figure 13 Migration of the Roma diaspora.

Map illustrating the inferred admixture events identified by MALDER and proposed migration routes across the Roma diaspora. The starting point is based on historical estimations.

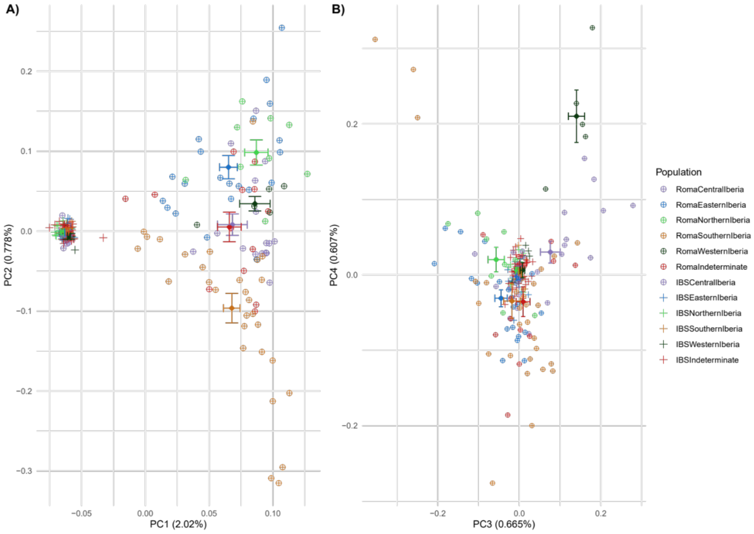


Supplementary Figure 14 Distribution of European local ancestry tracts.

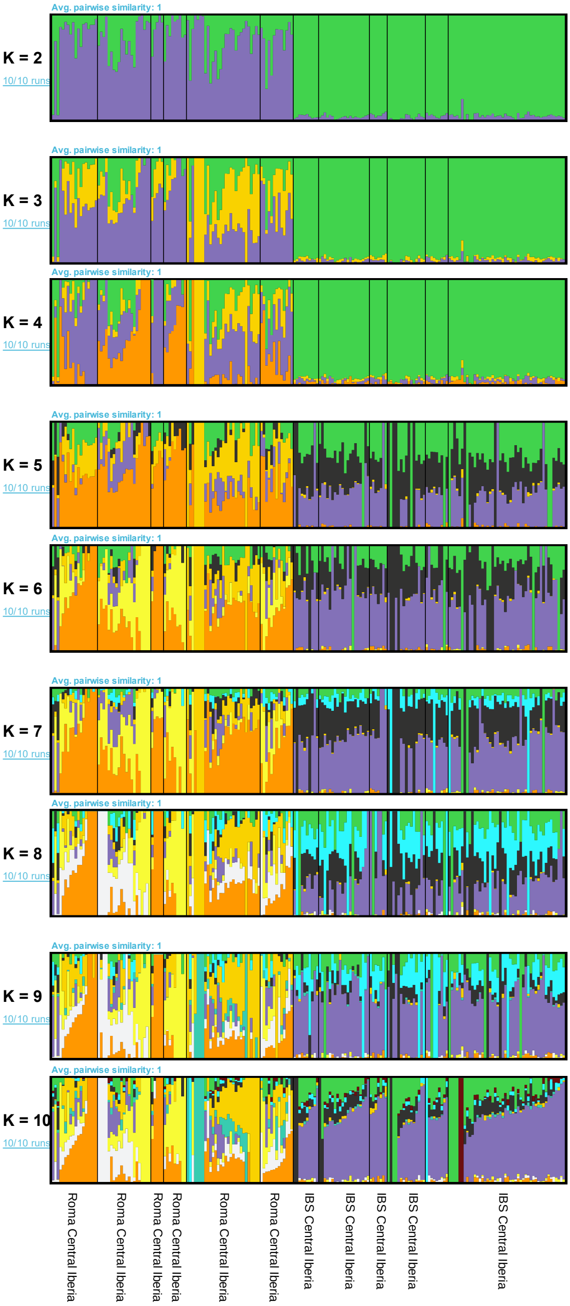
(a) Distribution of local ancestry tracts across each Roma population. (b) Distribution of local ancestry tract numbers by length category for each Roma population.



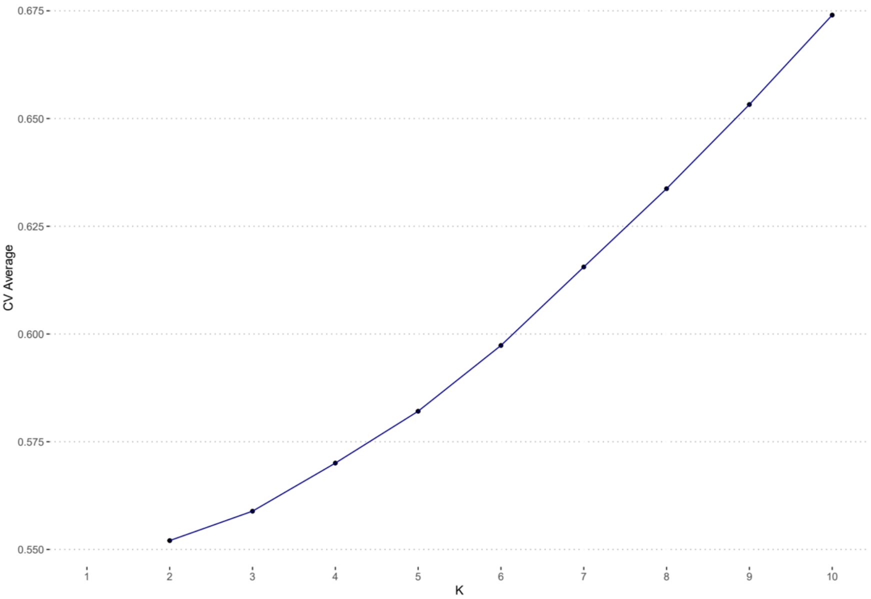
Supplementary Figure 15 Distribution of shared IBD segments within populations by length category in Roma and non-Roma reference populations.



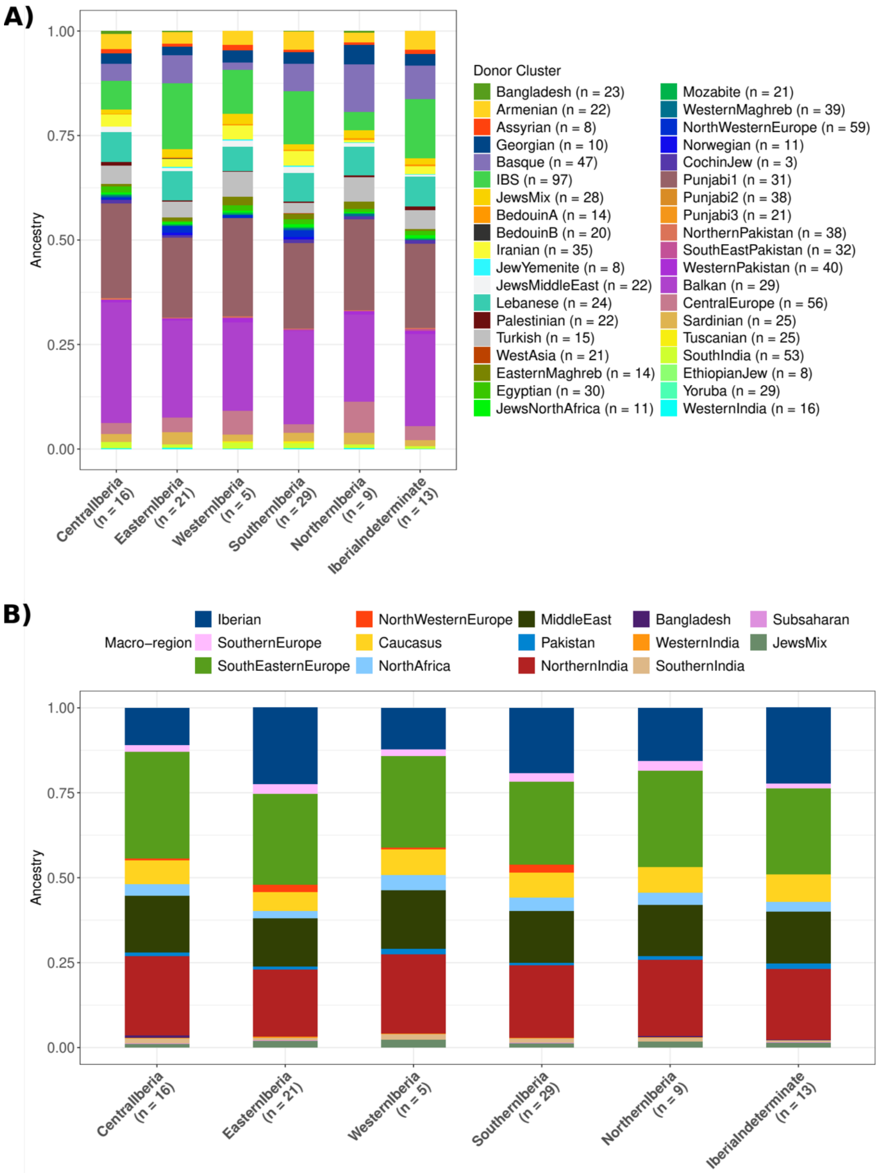
Supplementary Figure 16 Principal Component Analysis on the Iberian dataset. PCA conducted using the Iberian dataset. We calculated the confidence intervals at a significance level of 95% for each centroid, representing the average of each principal component.



Supplementary Figure 17 ADMIXTURE results on the Iberian dataset. Results from ADMIXTURE analyses using the Iberian dataset for all performed runs (K = 2 to K = 10).

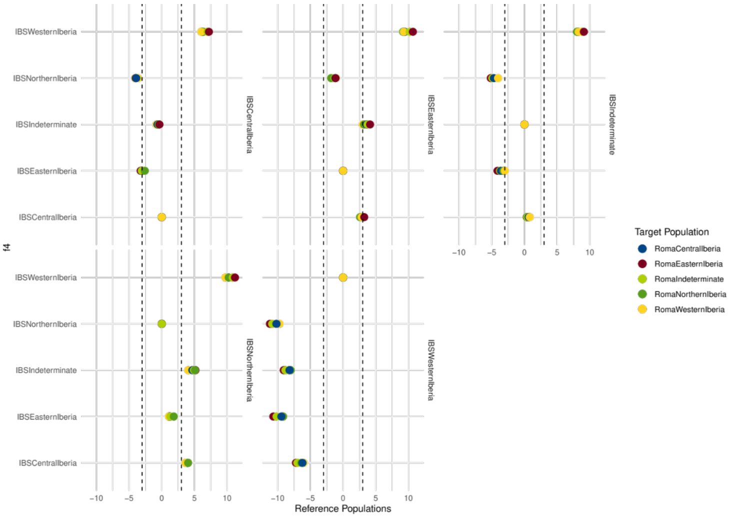


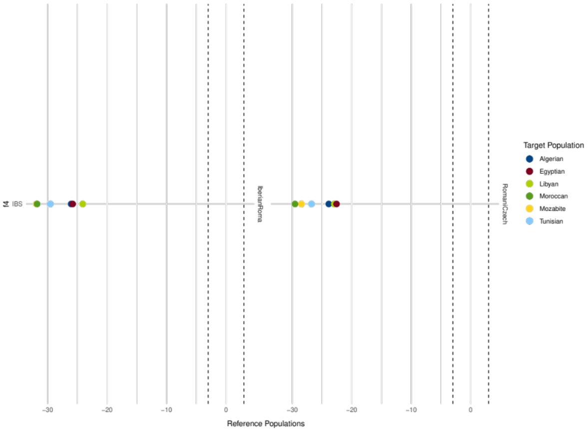
Supplementary Figure 18 Plot of cross-validation values for the ADMIXTURE on the Iberian dataset.

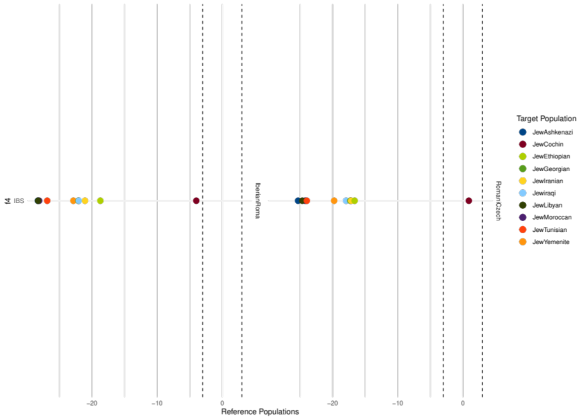


Supplementary Figure 19 Results of the NNLS analysis on the Iberian dataset.

(a) NNLS performed by genetic cluster, with results grouped by donor macro-region. (b) NNLS performed by genetic cluster, with results grouped by donor cluster.

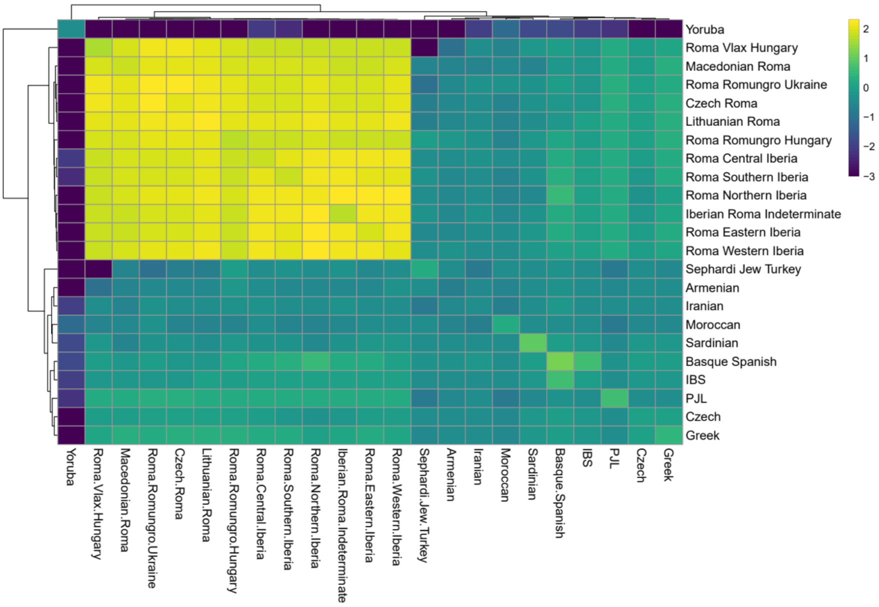

f4
IBSWesternIberia
IBSNorthernIberia
IBSIndeterminate
IBSEasternIberia
IBSCentralIberia
IBSCentralIberia
IBSEasternIberia
IBSIndeterminate
IBSNorthernIberia
IBSWesternIberia
−10
−5
0
5
10
Reference Populations
Target Population
RomaCentralIberia
RomaEasternIberia
RomaIndeterminate
RomaNorthernIberia
RomaWesternIberia


f4
IBS
IberianRoma
RomaniCzech
−30
−20
−10
0
Reference Populations
Target Population
Algerian
Egyptian
Libyan
Moroccan
Mozabite
Tunisian


f4
IBS
IberianRoma
RomaniCzech
−20
−10
0
Reference Populations
Target Population
JewAshkenazi
JewCochin
JewEthiopian
JewGeorgian
JewIranian
JewIraqi
JewLibyan
JewMoroccan
JewTunisian
JewYemenite

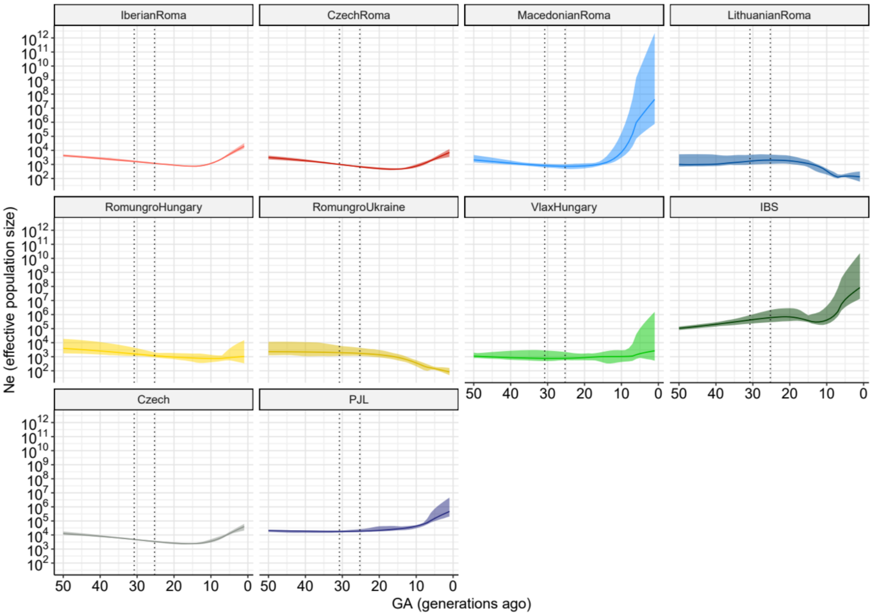
Supplementary Figure 20 f4 statistical tests for admixture.

(a) f4 test on the Iberian dataset. (b) f4 test for admixture with North African populations. (c) f4 test for admixture with Jewish populations.

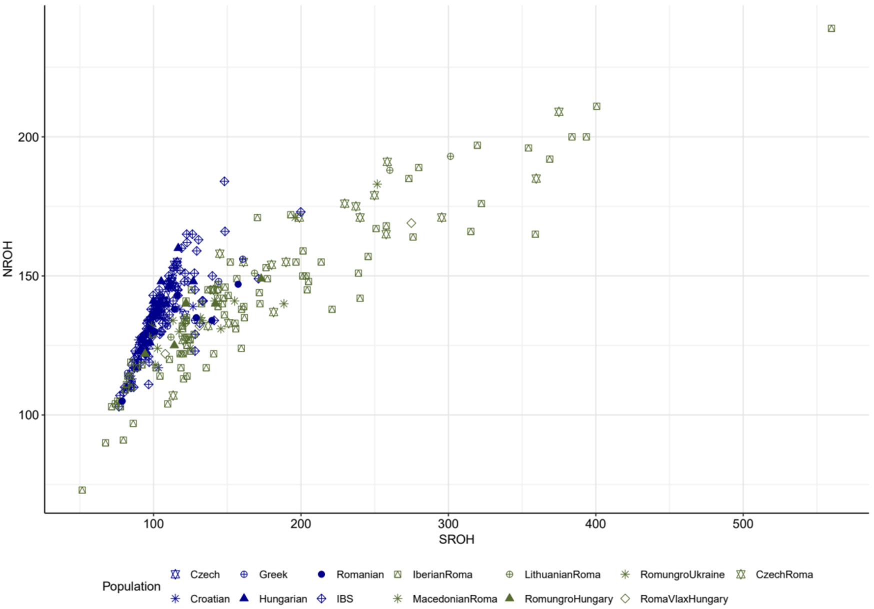


Supplementary Figure 21 Heatmap of pairwise shared IBD segments in the Roma population and reference populations.

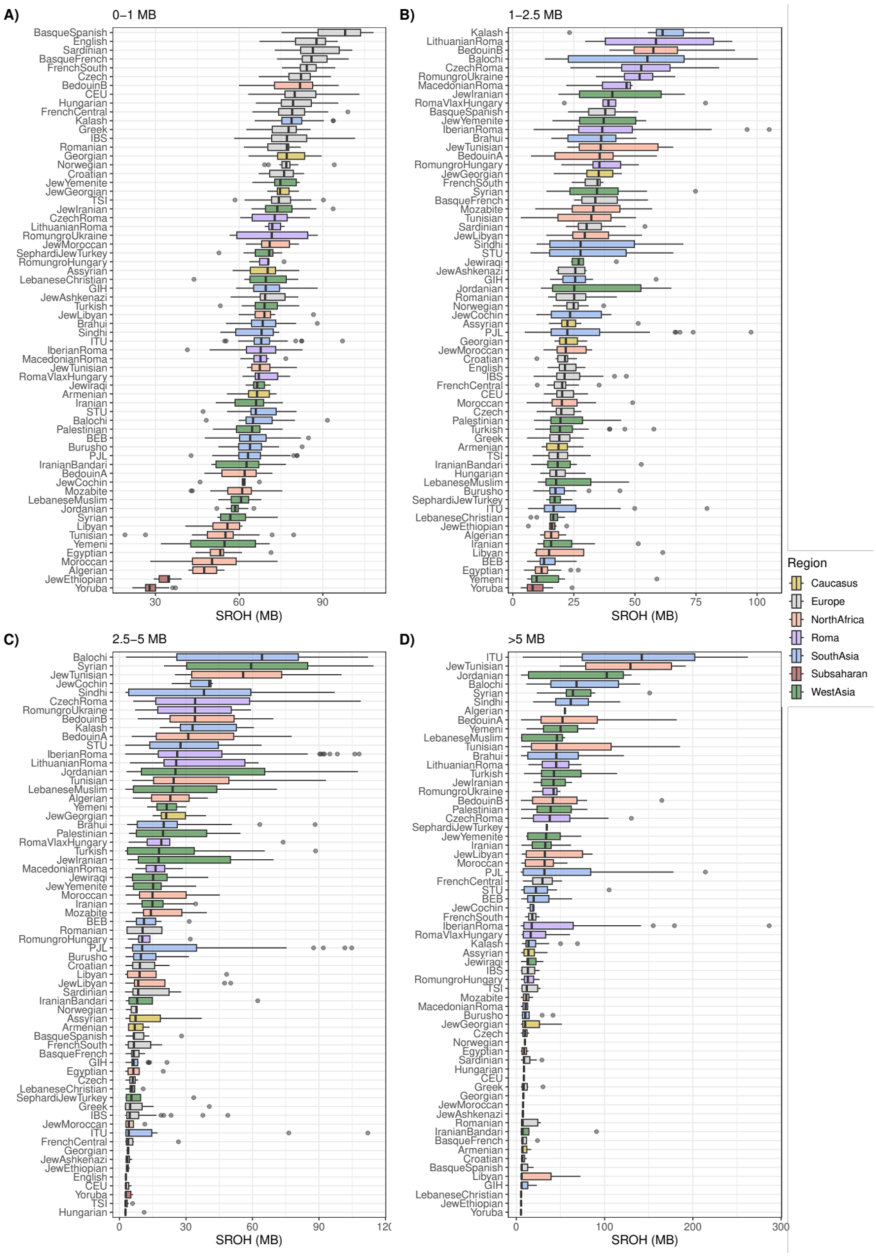
The heatmap displays IBD segment sharing between pairs of populations included in the present study. Reference populations were selected to represent each macro-region. Data is presented on a logarithmic scale.



Supplementary Figure 22 Effective population size trend for each Roma population over the past 50 generations. The vertical lines indicate the dates of the Roma's arrival in Europe (circa 31 generations ago) and in Spain (circa 25 generations ago), as inferred by fastGLOBETROTTER analysis. The horizontal lines represent the detected effective population size (Ne) values for each generation, with shaded areas indicating the confidence intervals for these values.

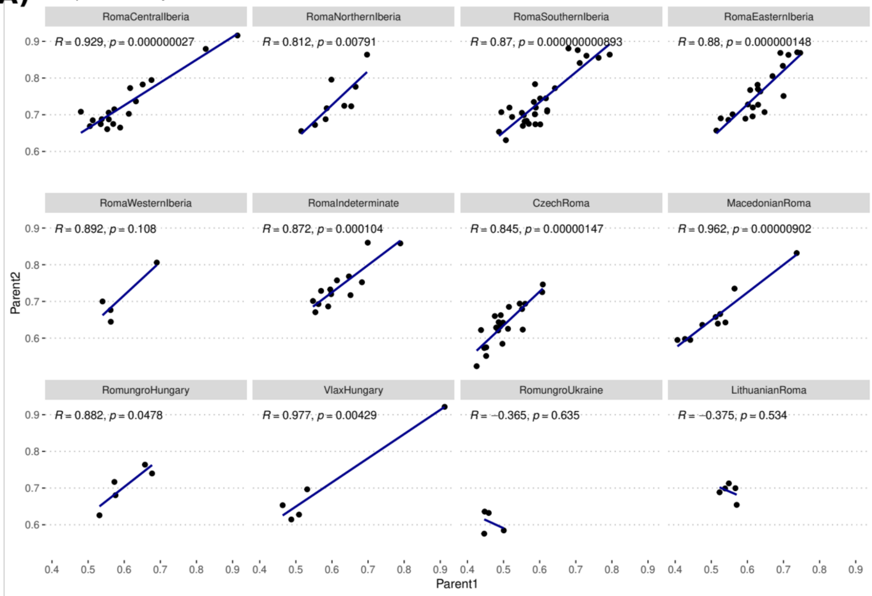


Supplementary Figure 23 Cumulative number (NROH) and size (SROH) of ROHs by individual for Roma and a limited set of references.

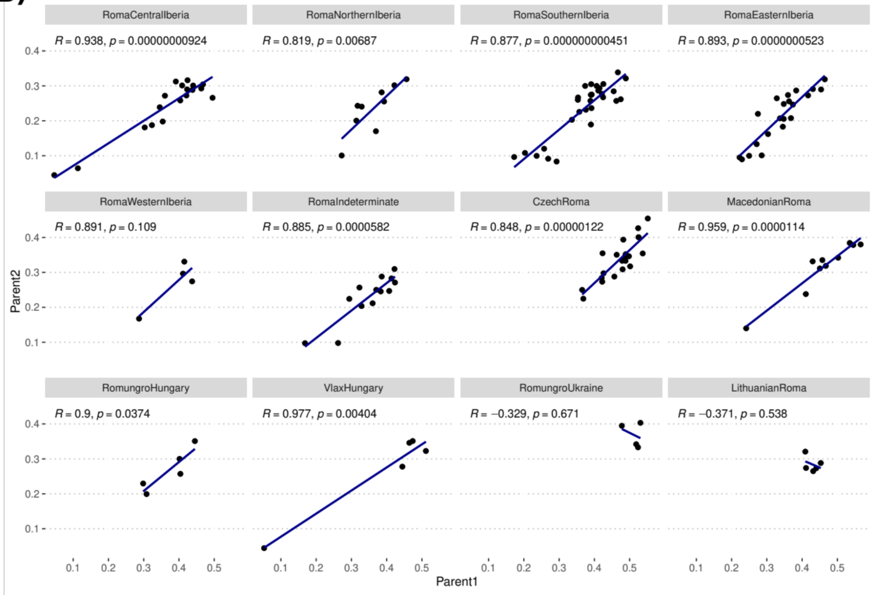


Supplementary Figure 24 (a-d) Average number of ROHs per individual by length class in each population of the dataset.

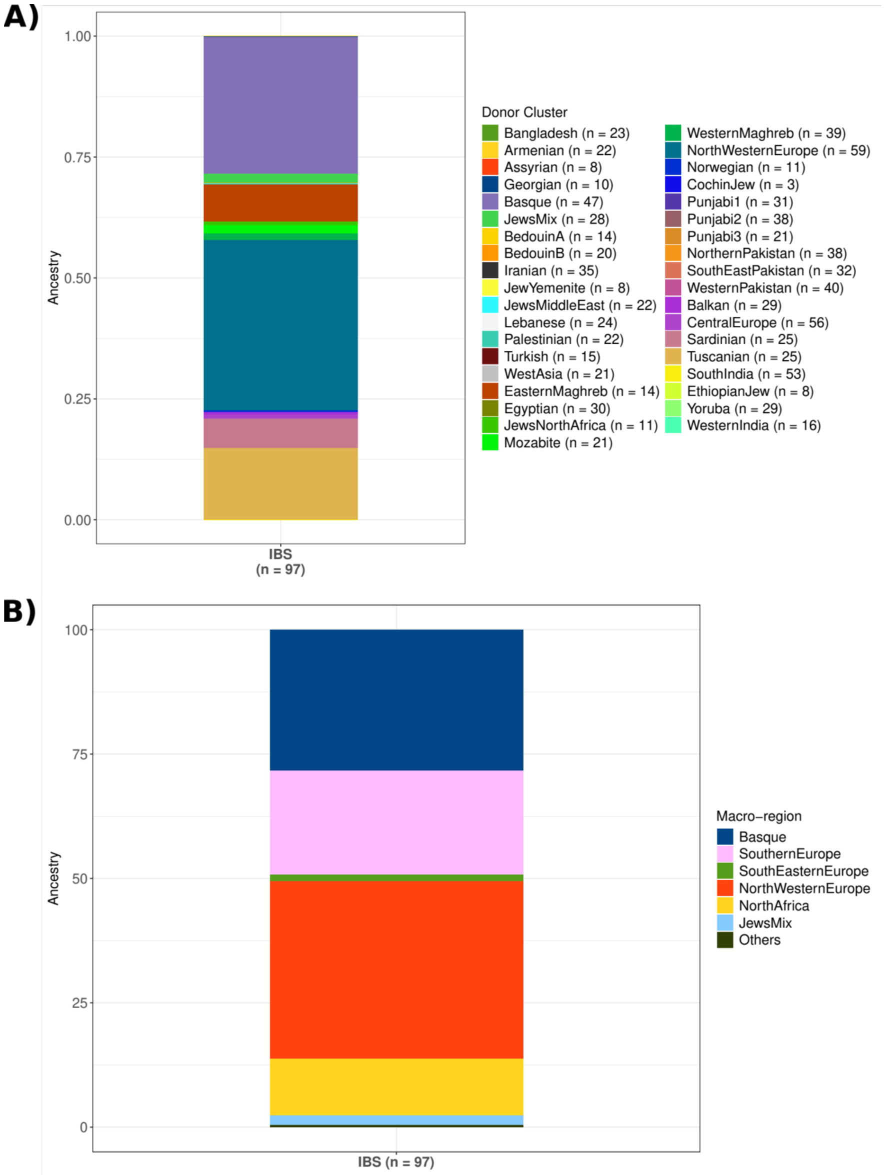
A) European Ancestry



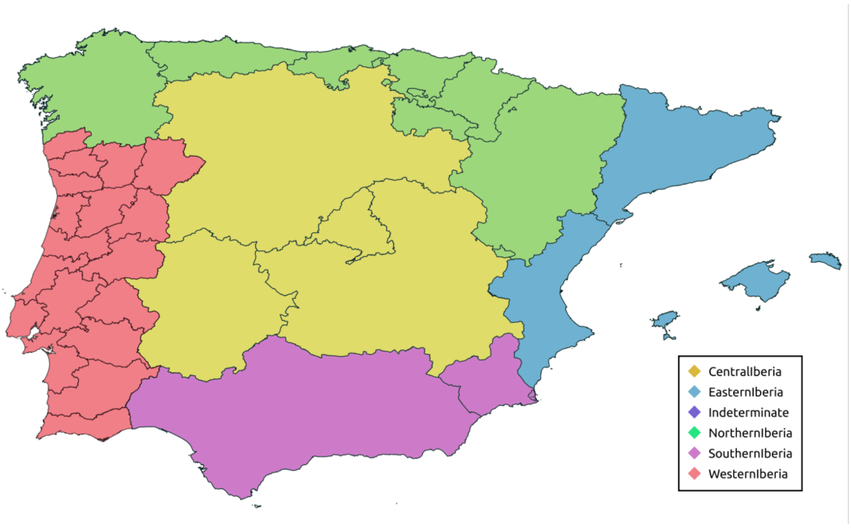
B) South Asian Ancestry



Supplementary Figure 25 Assortative mating analysis results for each Roma population. (a) Assortative mating test for the European ancestral component. (b) Assortative mating test for the South Asian ancestral component.



Supplementary Figure 26 Results of the NNLS analysis with IBS as the recipient cluster. (a) NNLS analysis by genetic cluster, with results grouped by donor cluster. (b) NNLS analysis by genetic cluster, with results grouped by donor macro-region.



Supplementary Figure 27 Distribution of the Iberian Roma samples. Distribution of the Iberian Roma individuals based on their grandparents' birthplace. The colours represent the macro-regions used for classifying the Iberian Roma within the Iberian Peninsula.

Supplementary Notes

Supplementary Note 1 IBD sharing patterns.

The length and number of shared IBD segments between populations provide insights into the timing of admixture events and help identify trends in changing levels of endogamy. In this case, the characterization of between-population shared IBD segments is used as an indicator of admixture level over time, while the IBD segments shared within the population are a proxy for the population size and inbreeding level. We expressed shared IBD segments in three categories of comparison: 1) within-population, which are the IBD shared between individuals from the same Roma group; 2) Roma-nonRoma, the IBD shared between Roma and non-Roma individuals; 3) between-Roma, the IBD shared between individuals of different Roma groups. The length categories were assigned following an approximation (Baharian et al. 2016; Harris et al. 2018; Castro e Silva et al. 2022) which can be used to infer when the IBD segments were formed based on their length:

E(generations ago) $\cong 3/(2l); l =$ *IBD segment length in morgans*

Using this approximation we selected four periods: 1) Arrival into Europe ~< 1100 CE (segments of 0-4 cM) (Fraser 1992; Kenrick 2007); 2) out-of-Balkans and arrival into Iberian peninsula ~ 1100–1570 CE (segments of 4-8 cM) (Fraser 1992; Kenrick 2007); 3) second out-of-Balkans waves ~ 1570–1800 CE (segments of 8-16 cM) (Achim 2004; Marushiakova and Popov 2010); 4) ending of slavery and contemporary period ~> 1800 (segments >16 cM) (Achim 2004; Greenberg 2010). Where 4 cM corresponds to circa 940 years ago (37.5 generations ago * 25 years per generation), and 16 cM corresponds to circa 230 years ago (9.37 generations ago * 25 years per generation). The average number of IBD segments and the average length of shared IBD were then calculated for each length category, by dividing these averages, one at a time, by the product of the sample sizes of the 2 populations being compared, as in Castro e Silva et al. (2022). Two-sided and one-sided Wilcoxon tests with Bonferroni correction were then computed to test for significance in the differences in shared IBD numbers within and between populations.

Supplementary Note 2 Distribution of European Local Ancestry tracts. The measurement of LAI tract lengths of European ancestry was performed on the RFmix output files generated for the ANCESTOR (Zou et al. 2015a, b) analysis. Measurement was performed using AncestryLength.rb, a script designed to measure the length of segments in the chromosomes. This measurement is calculated using the formula:

Ending Position of the switch ending point – Starting Position of the switch starting point

Here, the 'switch' refers to the SNP where the ancestry changes for a given haplotype. The switch points are identified as the first and last SNP of the segment within a certain random forest window. AncestryLength.rb can iterate through multiple ancestries and individuals. It was run with two settings: --hap, which omits haplotype information (needed when considering segments of the two haplotypes as part of the same individual), and --MB, which converts the tract length to MegaBases. The tracts were divided in categories following the same criteria used in the IBD test, assuming that 1 MB = 1 cM. We grouped the segments into six categories, where the longer segments are divided into three additional categories: 16-32 MB, 32-64 MB, >64 MB, which correspond to time periods going from circa 1800 CE until the last generation. The mean tract length and the number of tract length by length category were computed for each Roma group. One-sided Wilcoxon tests with Bonferroni correction were then computed to test for significance in the differences.

Supplementary Note 3 Genetic variation and geographic distribution. To assess the relationship between the inferred proportion of each ancestry component (summary of CHROMOPAINTER copying vectors obtained with NNLS) and the geographic distribution (Longitude and Latitude) of Roma individuals, Pearson's correlation tests were performed. Specifically, correlation tests were conducted for each ancestry component with longitude and latitude. The correlation coefficient (r), t-value, p-value, and 95% confidence intervals were calculated and tabulated for each test. The results were used to evaluate the strength and direction of the linear associations. For each individual we used the geographic coordinates of the principal city within their region. In the case of the "Indeterminate" group, coordinates were assigned to the geographical centre of the Iberian Peninsula. Last, for non-Iberian Roma groups we assigned the coordinates of the geographic centre of their respective countries of origin.

Supplementary Note 4 Local Ancestry Inference.

The local ancestry inference was estimated using the software RFMix v2 (Maples et al. 2013) with the -e switch set to perform 5 iterations of the expectation-maximization optimisation of the model. The analysis was conducted with two (European and Punjabi) parental populations, analysing both the Roma and 10 additional European individuals to provide a baseline of comparison. For the analysis, 58 individuals were used as a reference for each component. The European component was selected from Spanish (IBS), Czech, Croatian, Greek, Hungarian, and Romanian populations, while the Punjabi component came from Punjabi (PJL). To measure the North African ancestry component in the IBS, an analysis was conducted using two parental populations (European and North African), including all IBS individuals and 59 reference individuals for each component. The European component was represented by Basques from France, French, Central Europeans (CEU), and Tuscans (TSI), while the North African component was represented by Moroccans, Libyans, Algerians, and Tunisians.

References

Achim V (2004) The Roma in Romanian History. Central European University Press

Baharian S, Barakatt M, Gignoux CR, et al (2016) The Great Migration and African-American Genomic Diversity. PLOS Genetics 12:e1006059. https://doi.org/10.1371/journal.pgen.1006059

Castro e Silva MA, Ferraz T, Couto-Silva CM, et al (2022) Population Histories and Genomic Diversity of South American Natives. Molecular Biology and Evolution 39:msab339. https://doi.org/10.1093/molbev/msab339

Fraser AM (1992) The Gypsies (The Peoples of Europe). Blackwell Pub.

Greenberg J (2010) Report on Roma Education Today: From Slavery to Segregation and Beyond. Colum L Rev 110:919–1001

Harris DN, Song W, Shetty AC, et al (2018) Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. Proceedings of the National Academy of Sciences 115:E6526–E6535. https://doi.org/10.1073/pnas.1720798115

Kenrick D (2007) Historical Dictionary of the Gypsies (Romanies). Scarecrow Press

Maples BK, Gravel S, Kenny EE, Bustamante CD (2013) RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. The American Journal of Human Genetics 93:278–288. https://doi.org/10.1016/j.ajhg.2013.06.020

Marushiakova E, Popov V (2010) Gypsy/Roma European migrations from 15th century till nowadays. In: Proceedings of International Conference "Romani Mobilities in Europe: Multidisciplinary Perspectives. pp 126–139

Zhou Y, Browning SR, Browning BL (2020) A Fast and Simple Method for Detecting Identity-by-Descent Segments in Large-Scale Data. The American Journal of Human Genetics 106:426–437. https://doi.org/10.1016/j.ajhg.2020.02.010

Zou JY, Halperin E, Burchard E, Sankararaman S (2015a) Inferring parental genomic ancestries using pooled semi-Markov processes. Bioinformatics 31:i190–i196. https://doi.org/10.1093/bioinformatics/btv239

Zou JY, Park DS, Burchard EG, et al (2015b) Genetic and socioeconomic study of mate choice in Latinos reveals novel assortment patterns. Proceedings of the National Academy of Sciences 112:13621– 13626. https://doi.org/10.1073/pnas.1501741112

Supplementary Tables

a) Composition of the donor clusters.

Cluster	Population	N	Macro-cluster
Balkan	Greek	18	SouthEasternEurope
	Romanian	10	
	Turkish	1	
CentralEurope	Croatian	10	
	Czech	25	
	Hungarian	20	
	Roma Vlax from Hungary	1	
Basque	Basque French	20	Iberian
	Basque Spanish	9	
	French South	7	
	IBS	12	
IBS	French Central	1	
	IBS	96	
Sardinian	Sardinian	25	SouthernEurope
Tuscanian	TSI	25	
NorthWesternEurope	CEU	25	NorthWesternEurope
	English	10	
	FrenchCentral	24	
Norwegian	Norwegian	11	
Armenian	Armenian	10	Caucasus
	Assyrian	3	
	Greek	2	
	Syrian	1	
	Turkish	6	
Assyrian	Assyrian	8	
Georgian	Georgian	10	
BedouinA	BedouinA	15	MiddleEast
BedouinB	BedouinB	19	
Iranian	Iranian	25	
	Iranian Bandari	6	
	Cochin Jew	2	
	Turkish	2	
JewsMiddleEast	Jew Georgian	7	
	Jew Iranian	9	
	Jew Iraqi	6	
Lebanese	Jordanian	1	
	Lebanese Christian	9	
	Lebanese Muslim	11	
	Syrian	2	
	Turkish	1	
WestAsia	BedouinA	2	
	Jordanian	8	
	Palestinian	3	
	Syrian	5	
	Yemeni	3	
Palestinian	Palestinian	22	
Turkish	Turkish	15	
JewsMix	Ashkenazi Jews (Poland)	7	JewsMix
	Moroccan Jews	6	
	Sephardic Jews (Turkey)	8	
JewsNorthAfrica	Jew Libyan	9	NorthAfrica
	Jew Tunisian	7	
	Tunisian	2	
Mozabite	Mozabite	21	
WesternMaghreb	Algerian	3	
	Moroccan	21	
	Tunisian	14	
EasternMaghreb	Libyan	5	
	Tunisian	9	
Egyptian	BedouinA	8	
	Egyptian	18	
	Iranian Bandari	2	
	Yemeni	2	
EthiopianJew	EthiopianJew	7	Subsaharan
	Yemeni	1	
Yoruba	Moroccan	4	
	Yoruba	25	
Punjabi1	PJL	31	NorthIndia
Punjabi2	PJL	34	
	GIH	4	
Punjabi3	BEB	1	
	GIH	4	
	PJL	16	
CochinJew	Cochin Jew	3	WesternIndia
WesternIndia	GIH	16	
SouthIndia	BEB	1	SouthIndia
	ITU	25	
	STU	25	
	PJL	2	
NorthernPakistan	Burusho	23	Pakistan
	Kalash	15	
SouthEastPakistan	Sindhi	18	
	Balochi	1	
	GIH	1	
	PJL	12	
WesternPakistan	Balochi	19	
	Brahui	21	
Bangladesh	BEB	23	Bangladesh

b) Composition of the recipient clusters

Cluster	Population	N	Subcluster
IberianRoma	Eastern Iberia	5	IberianRomaSouthEast
	Indeterminate	2	
	Northern Iberia	1	
	Southern Iberia	6	
	Western Iberia	4	IberianRomaCentralWest
	Southern Iberia	1	
	Northern Iberia	1	
	Indeterminate	5	
	Eastern Iberia	4	
	Central Iberia	11	
	Southern Iberia	17	IberianRomaSouth
	Central Iberia	1	
	Eastern Iberia	1	
	Indeterminate	3	
	Central Iberia	1	IberianRomaNorthEast
	Eastern Iberia	5	
	Northern Iberia	4	
	Central Iberia	3	IberianRomaMix
	Eastern Iberia	6	
	Southern Iberia	5	
	Western Iberia	1	
	Indeterminate	3	
	Northern Iberia	3	
EuropeanRoma	Czech Roma	13	CzechRoma
	Roma Romungro from Ukraine	3	
	Czech Roma	2	MacedonianRoma
	Macedonian Roma	10	
	Czech Roma	6	EuropeanRomaMix
	Roma Romungro from Hungary	5	
	Roma Romungro from Ukraine	1	
	Lithuanian Roma	5	
	Roma Vlax from Hungary	4	

Supplementary Table 1. Composition of fineSTRUCTURE genetic clusters based on the dendrogram.

Cluster	Test	t value	p-value	Lower 95% CI	Upper 95% CI	r
Iberian	Longitude	-6,572	0,000	-0,599	-0,347	-0,483
SouthEasternEurope	Longitude	7,574	0,000	0,409	0,644	0,536
NorthWesternEurope	Longitude	-1,788	0,076	-0,305	0,016	-0,148
SouthernEurope	Longitude	-6,809	0,000	-0,610	-0,362	-0,496
Caucasus	Longitude	2,765	0,006	0,065	0,376	0,226
NorthAfrica	Longitude	-7,561	0,000	-0,643	-0,408	-0,536
MiddleEast	Longitude	1,161	0,248	-0,068	0,256	0,097
Pakistan	Longitude	0,282	0,778	-0,140	0,187	0,024
NorthIndia	Longitude	5,815	0,000	0,296	0,562	0,439
Bangladesh	Longitude	3,348	0,001	0,112	0,416	0,270
WesternIndia	Longitude	1,445	0,151	-0,044	0,278	0,120
SouthIndia	Longitude	1,875	0,063	-0,008	0,311	0,155
SubSaharan	Longitude	1,996	0,048	0,002	0,320	0,165
JewsMix	Longitude	-3,774	0,000	-0,444	-0,146	-0,302
Iberian	Latitude	-5,812	0,000	-0,562	-0,296	-0,438
SouthEasternEurope	Latitude	6,492	0,000	0,342	0,595	0,478
NorthWesternEurope	Latitude	-2,071	0,040	-0,326	-0,008	-0,171
SouthernEurope	Latitude	-6,161	0,000	-0,579	-0,320	-0,459
Caucasus	Latitude	0,736	0,463	-0,103	0,223	0,062
NorthAfrica	Latitude	-6,143	0,000	-0,578	-0,318	-0,458
MiddleEast	Latitude	2,940	0,004	0,079	0,388	0,240
Pakistan	Latitude	0,516	0,607	-0,121	0,205	0,043
NorthIndia	Latitude	4,780	0,000	0,222	0,505	0,372
Bangladesh	Latitude	4,184	0,000	0,177	0,469	0,331
WesternIndia	Latitude	0,520	0,604	-0,121	0,206	0,044
SouthIndia	Latitude	0,849	0,397	-0,094	0,232	0,071
SubSaharan	Latitude	2,905	0,004	0,076	0,385	0,237
JewsMix	Latitude	-3,221	0,002	-0,407	-0,102	-0,261

Supplementary Table 2. Pearson's correlation tests for the geographic cline across all donor clusters grouped by macro-regions.

a) Composition of fastGLOBETROTTER estimated dates and sources for each recipient genetic cluster

IberianRoma **INFERRED SOURCES AND DATES** **('best guess' conclusion: multiple dates)**

2 DATE FIT EVIDENCE DATE ESTIMATES SINGLE BEST FITTING DONORS

Generations Date1	Generations Date2	MaxScore 2 events	Proportion Date1 Source1	Bestmatch Date1 Source1	Bestmatch Date1 Source2	Proportion Date2 Source1	Bestmatch Date2 Source1	Bestmatch Date2 Source2
4,77	24,55	0,89	0,26	IBS	Turkish	0,47	Punjabi1	Tuscanian

2 DATE FIT SOURCES DATE1 PC1:

proportion	JewsMiddleEast	WesternMaghreb	CochinJew	SouthEastIndia	Georgian	Sardinian	Lebanese	JewsMix	Iranian	Punjabi1	CentralEurope	Balkan	Basque	IBS
0,26	0,00	0,01	0,02	0,02	0,03	0,03	0,04	0,06	0,06	0,07	0,09	0,10	0,12	0,34
proportion	WesternMaghreb	JewsNorthAfrica	Assyrian	SouthEastIndia	Sardinian	Georgian	EasternMaghreb	IBS	Iranian	Basque	Turkish	Lebanese	Punjabi1	Balkan
0,74	0,00	0,00	0,00	0,01	0,02	0,02	0,02	0,03	0,05	0,05	0,10	0,12	0,26	0,31

2 DATE FIT SOURCES DATE2 PC1:

proportion	JewsNorthAfrica	CochinJew	EasternMaghreb	Sardinian	JewsMix	SouthEastIndia	Turkish	Georgian	CentralEurope	IBS	Basque	Lebanese	Balkan	Punjabi1		
0,47	0,00	0,01	0,02	0,02	0,03	0,03	0,03	0,03	0,04	0,05	0,06	0,10	0,15	0,44		
proportion	JewsMiddleEast	CochinJew	Punjabi1	Assyrian	JewsMix	WesternMaghreb	CentralEurope	EasternMaghreb	Georgian	Sardinian	Basque	Iranian	Lebanese	Turkish	IBS	Balkan
0,53	0,00	0,00	0,00	0,00	0,01	0,01	0,01	0,02	0,02	0,02	0,08	0,10	0,10	0,11	0,17	0,36

EuropeanRoma **INFERRED SOURCES AND DATES** **('best guess' conclusion: multiple dates)**

2 DATE FIT EVIDENCE DATE ESTIMATES SINGLE BEST FITTING DONORS

Generations Date1	Generations Date2	MaxScore 2 events	Proportion Date1 Source1	Bestmatch Date1 Source1	Bestmatch Date1 Source2	Proportion Date2 Source1	Bestmatch Date2 Source1	Bestmatch Date2 Source2
9,29	30,70	0,59	0,27	CentralEurope	Punjabi1	0,46	Balkan	Punjabi1

2 DATE FIT SOURCES DATE1 PC1:

proportion	EthiopianJew	Turkish	JewYemenite	Punjabi1	CochinJew	Iranian	SouthEastIndia	Georgian	CentralEurope	Balkan		
0,27	0,01	0,01	0,01	0,05	0,06	0,06	0,07	0,07	0,32	0,35		
proportion	SouthEastIndia	Georgian	Assyrian	CochinJew	WestAsia	JewYemenite	JewsNorthernMiddleEast	Iranian	Armenian	Turkish	Balkan	Punjabi1
0,73	0,00	0,00	0,00	0,01	0,02	0,02	0,02	0,05	0,06	0,17	0,29	0,35

2 DATE FIT SOURCES DATE2 PC1:

proportion	EthiopianJew	SouthEastIndia	Assyrian	CochinJew	JewYemenite	JewsMix	WestAsia	Georgian	Armenian	Iranian	Turkish	CentralEurope	Balkan
0,46	0,00	0,00	0,01	0,02	0,03	0,04	0,04	0,04	0,09	0,12	0,13	0,19	0,31
proportion	JewYemenite	SouthEastIndia	CochinJew	Turkish	Balkan	Punjabi1							
0,54	0,01	0,03	0,03	0,13	0,30	0,49							

IberianRomaCentralWest **INFERRED SOURCES AND DATES** **('best guess' conclusion: multiple dates)**

2 DATE FIT EVIDENCE DATE ESTIMATES SINGLE BEST FITTING DONORS

Generations Date1	Generations Date2	MaxScore 2 events	Proportion Date1 Source1	Bestmatch Date1 Source1	Bestmatch Date1 Source2	Proportion Date2 Source1	Bestmatch Date2 Source1	Bestmatch Date2 Source2
4,05	23,15	0,60	0,15	IBS	Turkish	0,45	Punjabi1	Balkan

2 DATE FIT SOURCES DATE1 PC1:

proportion	JewsMix	JewsMiddleEast	CochinJew	JewsNorthAfrica	CentralEurope	Sardinian	WesternPakistan	Punjabi1	Basque	IBS					
0,15	0,00	0,01	0,02	0,02	0,02	0,04	0,06	0,19	0,29	0,36					
proportion	JewsNorthAfrica	JewsMiddleEast	Assyrian	Egyptian	EasternMaghreb	CochinJew	Sardinian	WesternMaghreb	SouthEastIndia	Iranian	Georgian	Turkish	Lebanese	Punjabi1	Balkan
0,85	0,00	0,00	0,00	0,00	0,01	0,01	0,01	0,02	0,02	0,02	0,05	0,12	0,14	0,25	0,37

2 DATE FIT SOURCES DATE2 PC1:

proportion	JewsMiddleEast	Assyrian	Sardinian	Turkish	SouthEastIndia	Iranian	CochinJew	IBS	Basque	Lebanese	Balkan	Punjabi1		
0,45	0,00	0,00	0,00	0,00	0,01	0,01	0,01	0,01	0,05	0,18	0,19	0,53		
proportion	JewsNorthAfrica	WesternMaghreb	JewsMix	CentralEurope	SouthEastIndia	Sardinian	EasternMaghreb	Georgian	Basque	IBS	Lebanese	Iranian	Turkish	Balkan
0,55	0,01	0,01	0,01	0,02	0,02	0,03	0,03	0,04	0,05	0,06	0,07	0,10	0,14	0,42

IberianRomaSouth **INFERRED SOURCES AND DATES** **('best guess' conclusion: multiple dates)**

2 DATE FIT EVIDENCE DATE ESTIMATES SINGLE BEST FITTING DONORS

Generations Date1	Generations Date2	MaxScore 2 events	Proportion Date1 Source1	Bestmatch Date1 Source1	Bestmatch Date1 Source2	Proportion Date2 Source1	Bestmatch Date2 Source1	Bestmatch Date2 Source2
3,17	23,87	0,67	0,20	IBS	Turkish	0,50	Punjabi1	Tuscanian

2 DATE FIT SOURCES DATE1 PC1:

proportion	Mozabite	IBS	JewsNorthAfrica	EasternMaghreb	JewsMiddleEast	WesternMaghreb	CochinJew	Punjabi1	Sardinian	SouthEastIndia	b	Basque	Balkan
0,20	0,00	0,00	0,00	0,00	0,02	0,02	0,04	0,05	0,06	0,07	0,14	0,22	0,37
proportion	Egyptian	WesternMaghreb	SouthEastIndia	JewsMiddleEast	Sardinian	EasternMaghreb	Basque	Georgian	Iranian	Turkish	Lebanese	Punjabi1	Balkan
0,80	0,01	0,01	0,01	0,01	0,01	0,02	0,04	0,05	0,06	0,07	0,14	0,29	0,29

2 DATE FIT SOURCES DATE2 PC1:

proportion	Egyptian	Sardinian	EasternMaghreb	Georgian	SouthEastIndia	Basque	Lebanese	Balkan	Punjabi1					
0,50	0,01	0,01	0,01	0,02	0,04	0,05	0,16	0,22	0,48					
proportion	SouthEastIndia	JewsNorthAfrica	CochinJew	WesternMaghreb	JewsMiddleEast	Sardinian	EasternMaghreb	JewsMix	Georgian	Lebanese	Iranian	Basque	Turkish	Balkan
0,50	0,00	0,00	0,01	0,02	0,02	0,03	0,03	0,05	0,06	0,07	0,09	0,10	0,11	0,40

IberianRomaSouthEast **INFERRED SOURCES AND DATES** **('best guess' conclusion: multiple dates)**

2 DATE FIT EVIDENCE DATE ESTIMATES SINGLE BEST FITTING DONORS

Generations Date1	Generations Date2	MaxScore 2 events	Proportion Date1 Source1	Bestmatch Date1 Source1	Bestmatch Date1 Source2	Proportion Date2 Source1	Bestmatch Date2 Source1	Bestmatch Date2 Source2
5,88	28,62	0,67	0,34	NorthWestEurope	IBS	0,34	IBS	IBS

2 DATE FIT SOURCES DATE1 PC1:

proportion	JewsNorthAfrica	Lebanese	Turkish	JewsMix	WesternMaghreb	Georgian	Sardinian	Armenian	Iranian	Balkan	Basque	CentralEurope	Punjabi1
0,34	0,00	0,00	0,02	0,03	0,03	0,04	0,06	0,07	0,08	0,13	0,13	0,19	0,21
proportion	Sardinian	CochinJew	CentralEurope	WesternPakistan	Iranian	Basque	Turkish	Punjabi1	Balkan	Lebanese	IBS		
0,66	0,00	0,00	0,01	0,01	0,01	0,04	0,04	0,07	0,08	0,09	0,66		

2 DATE FIT SOURCES DATE2 PC1:

proportion	EasternMaghreb	JewsNorthAfrica	Georgian	Iranian	CochinJew	Turkish	Sardinian	Basque	JewsMix	Lebanese	Armenian	Balkan	Punjabi1	IBS
0,34	0,00	0,00	0,00	0,00	0,01	0,01	0,01	0,03	0,03	0,05	0,07	0,08	0,34	0,36
proportion	Punjabi1	EasternMaghreb	WesternPakistan	WesternMaghreb	Georgian	Sardinian	Turkish	Iranian	Lebanese	Basque	Balkan	CentralEurope	IBS	
0,66	0,00	0,00	0,01	0,01	0,02	0,03	0,05	0,05	0,06	0,09	0,10	0,11	0,47	

IberianRomaNorthEast **INFERRED SOURCES AND DATES** **('best guess' conclusion: multiple dates)**

2 DATE FIT EVIDENCE DATE ESTIMATES SINGLE BEST FITTING DONORS

Generations Date1	Generations Date2	MaxScore 2 events	Proportion Date1 Source1	Bestmatch Date1 Source1	Bestmatch Date1 Source2	Proportion Date2 Source1	Bestmatch Date2 Source1	Bestmatch Date2 Source2
8,34	24,68	0,73	0,20	Basque	Turkish	0,43	Punjabi1	Tuscanian

2 DATE FIT SOURCES DATE1 PC1:

proportion	Iranian	Yoruba	WesternPakistan	EthiopianJew	Sardinian	CochinJew	Georgian	SouthEastIndia	Punjabi1	Balkan	Basque							
0,20	0,00	0,00	0,01	0,01	0,01	0,02	0,03	0,05	0,19	0,23	0,44							
proportion	Basque	Egyptian	Mozabite	SouthEastIndia	JewYemenite	JewsNorthAfrica	CochinJew	EasternMaghreb	Assyrian	Georgian	JewsMix	Sardinian	Iranian	CentralEurope	Turkish	Lebanese	Punjabi1	Balkan
0,80	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,02	0,03	0,04	0,04	0,04	0,05	0,12	0,13	0,25	0,26

2 DATE FIT SOURCES DATE2 PC1:

proportion	SouthEastIndia	Sardinian	Egyptian	Georgian	Basque	Turkish	Lebanese	Balkan	Punjabi1								
0,43	0,01	0,01	0,02	0,02	0,05	0,06	0,11	0,17	0,55								
proportion	Punjabi1	Mozabite	EasternMaghreb	JewsNorthAfrica	CochinJew	SouthEastIndia	Assyrian	Egyptian	JewsMix	Georgian	Sardinian	CentralEurope	Iranian	Lebanese	Basque	Turkish	Balkan
0,57	0,00	0,00	0,00	0,00	0,01	0,01	0,01	0,02	0,04	0,04	0,05	0,05	0,07	0,08	0,12	0,13	0,35

IberianRomaMix	**INFERRED SOURCES AND DATES**	**('best guess' conclusion: multiple dates)**																	
	2 DATE FIT EVIDENCE DATE ESTIMATES SINGLE BEST FITTING DONORS																		
Generations Date1	Generations Date2	MaxScore 2 events	Proportion Date1 Source1	Bestmatch Date1 Source1	Bestmatch Date1 Source2	Proportion Date2 Source1	Bestmatch Date2 Source1	Bestmatch Date2 Source2											
4,73	26,84	0,75	0,46	IBS	Turkish	0,42	Punjabi1	Tuscanian											
2 DATE FIT SOURCES DATE1 PC1:																			
proportion	WesternPakistan	Mozabite	EasternMaghreb	JewsMiddleEast	Sardinian	WesternMaghreb	Georgian	Iranian	CentralEurope	Lebanese	Punjabi1	Basque	Balkan	IBS					
0,46	0,00	0,00	0,01	0,01	0,01	0,01	0,02	0,07	0,07	0,09	0,09	0,11	0,14	0,39					
proportion	Assyrian	JewsNorthAfrica	CochinJew	Georgian	CentralEurope	SouthEastIndia	Iranian	JewsMix	EasternMaghreb	Sardinian	Armenian	Basque	Lebanese	Turkish	Punjabi1	Balkan			
0,54	0,00	0,00	0,01	0,01	0,01	0,02	0,02	0,02	0,03	0,03	0,04	0,04	0,07	0,14	0,27	0,28			
2 DATE FIT SOURCES DATE2 PC1:			FIT	SOURCES	DATE2	PC1:													
proportion	SouthEastIndia	Sardinian	Assyrian	EasternMaghreb	WesternMaghreb	Iranian	CentralEurope	Georgian	Armenian	Turkish	Basque	Lebanese	Balkan	IBS	Punjabi1				
0,42	0,00	0,00	0,01	0,01	0,01	0,01	0,02	0,02	0,02	0,02	0,04	0,07	0,15	0,17	0,45				
proportion	Punjabi1	WesternPakistan	Mozabite	JewsNorthAfrica	JewsMiddleEast	CochinJew	Georgian	Armenian	SouthEastIndia	JewsMix	EasternMaghreb	Sardinian	CentralEurope	Iranian	Lebanese	Basque	Turkish	IBS	Balkan
0,58	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,02	0,02	0,02	0,03	0,04	0,05	0,07	0,08	0,09	0,11	0,19	0,26

CzechRoma	**INFERRED SOURCES AND DATES**	**('best guess' conclusion: multiple dates)**											
	2 DATE FIT EVIDENCE DATE ESTIMATES SINGLE BEST FITTING DONORS												
Generations Date1	Generations Date2	MaxScore 2 events	Proportion Date1 Source1	Bestmatch Date1 Source1	Bestmatch Date1 Source2	Proportion Date2 Source1	Bestmatch Date2 Source1	Bestmatch Date2 Source2					
9,56	33,24	0,39	0,27	Punjabi1	Turkish	0,49	Turkish	Punjabi1					
2 DATE FIT SOURCES DATE1 PC1:													
proportion	Assyrian	JewYemenite	WestAsia	CochinJew	Armenian	Balkan	JewsMiddleEast	Georgian	Punjabi1				
0,27	0,01	0,01	0,02	0,04	0,05	0,10	0,13	0,17	0,48				
proportion	EthiopianJew	Palestinian	WesternPakistan	CochinJew	SouthEastIndia	JewYemenite	Iranian	CentralEurope	Punjabi1	Turkish	Balkan		
0,73	0,01	0,01	0,01	0,02	0,03	0,04	0,05	0,12	0,22	0,22	0,28		
2 DATE FIT SOURCES DATE2 PC1:													
proportion	Punjabi1	WesternPakistan	Assyrian	EthiopianJew	WestAsia	JewYemenite	JewsMiddleEast	CochinJew	Georgian	Iranian	Turkish	CentralEurope	Balkan
0,49	0,00	0,00	0,00	0,01	0,02	0,03	0,05	0,05	0,11	0,13	0,15	0,17	0,28
proportion	CochinJew	WestAsia	Palestinian	JewYemenite	SouthEastIndia	Turkish	Balkan	Punjabi1					
0,51	0,00	0,01	0,01	0,03	0,04	0,13	0,22	0,57					

EuropeanRomaMix	**INFERRED SOURCES AND DATES**	**('best guess' conclusion: multiple dates)**												
	2 DATE FIT EVIDENCE DATE ESTIMATES SINGLE BEST FITTING DONORS													
Generations Date1	Generations Date2	MaxScore 2 events	Proportion Date1 Source1	Bestmatch Date1 Source1	Bestmatch Date1 Source2	Proportion Date2 Source1	Bestmatch Date2 Source1	Bestmatch Date2 Source2						
12,63	30,45	0,60	0,40	CentralEurope	Turkish	0,46	Punjabi1	Balkan						
2 DATE FIT SOURCES DATE1 PC1:														
proportion	EthiopianJew	WestAsia	SouthEastIndia	JewYemenite	Iranian	CochinJew	Turkish	Punjabi1	Balkan	CentralEurope				
0,40	0,01	0,01	0,01	0,02	0,03	0,05	0,12	0,13	0,19	0,42				
proportion	WestAsia	Georgian	Assyrian	SouthEastIndia	JewYemenite	JewsMiddleEast	Iranian	Armenian	Turkish	Balkan	Punjabi1			
0,60	0,00	0,00	0,00	0,02	0,02	0,03	0,05	0,08	0,16	0,31	0,32			
2 DATE FIT SOURCES DATE2 PC1:														
proportion	JewYemenite	CochinJew	CentralEurope	Turkish	Balkan	Punjabi1								
0,46	0,02	0,03	0,10	0,15	0,18	0,52								
proportion	Punjabi1	Assyrian	EthiopianJew	Georgian	CochinJew	WestAsia	JewYemenite	SouthEastIndia	JewsMiddleEast	Armenian	Iranian	Turkish	CentralEurope	Balkan
0,54	0,00	0,00	0,00	0,01	0,01	0,01	0,02	0,03	0,04	0,08	0,09	0,15	0,22	0,34

MacedonianRoma	**INFERRED SOURCES AND DATES**	**('best guess' conclusion: multiple dates)**						
	1 DATE FIT EVIDENCE DATE ESTIMATES SINGLE BEST FITTING DONORS							
Generations Date1	Proportion Source1	Max R2 Fit 1 Date	Fit Quality 1 Event	Bestmatch Event 1 Source 1	Bestmatch Event 1 Source 2			
28,59	0,49	0,98	1,00	Balkan	Punjabi1			
1 DATE FIT SOURCES DATE1 PC1:								
proportion	SouthEastIndia	WestAsia	CochinJew	Lebanese	Turkish	Iranian	Armenian	Balkan
0,49	0,02	0,02	0,03	0,06	0,07	0,10	0,10	0,61
proportion	Lebanese	SouthEastIndia	CochinJew	Balkan	Armenian	Punjabi1		
0,51	0,02	0,03	0,03	0,18	0,19	0,55		

b) Bootstrap results and dating statistics for all recipient clusters.

Recipient Cluster	Date 1 (Mean)	Mean admixture time 1 (CE)	Date 1 – Standard Error	Date 1 - 95% Lower CI	Date 1 - 95% Upper CI	95% Lower CI (CE)	95% Upper CI (CE)	Date 2 (Mean)	Mean admixture time 2 (CE)	Date 2 – Standard Error	Date 2 - 95% Lower CI	Date 2 – 95% Upper CI	95% Lower CI (CE)	95% Upper CI (CE)
IberianRoma	25,26	1369	0,11	25,04	25,47	1363	1374	5,16	1871	0,15	4,87	5,46	1864	1878
EuropeanRoma	30,79	1230	0,25	30,30	31,28	1218	1243	10,06	1749	0,54	8,99	11,12	1722	1775
CzechRoma	35,90	1103	0,56	34,78	37,01	1075	1131	10,54	1736	0,70	9,16	11,93	1702	1771
MacedonianRoma	29,25	1269	0,22	28,80	29,69	1258	1280	NA	NA	NA	NA	NA	NA	NA
EuropeanRomaMix	30,79	1230	0,51	29,79	31,79	1205	1255	9,94	1751	0,58	8,80	11,09	1723	1780
IberianRomaMix	27,96	1301	0,28	27,40	28,51	1287	1315	4,60	1885	0,18	4,24	4,95	1876	1894
IberianRomaNorthEast	25,27	1368	0,56	24,16	26,37	1341	1396	7,39	1815	0,51	6,39	8,40	1790	1840
IberianRomaSouth	25,45	1364	0,23	24,99	25,90	1352	1375	4,39	1890	0,22	3,95	4,82	1879	1901
IberianRomaSouthEast	28,29	1293	0,39	27,51	29,07	1273	1312	3,56	1911	0,21	3,14	3,97	1901	1921
IberianRomaCentralWest	24,97	1376	0,29	24,41	25,54	1362	1390	7,26	1818	0,46	6,34	8,18	1796	1841

Supplementary Table 3. Dating estimates results.

Target	Reference A	Reference B	P-value	Z-score	Mean admixture time (GA)	Standard Error (GA)	Mean admixture time (CE)	95% Lower CI (CE)	95% Upper CI (CE)	Mean admixture time – SE (CE)	Mean admixture time + SE (CE)
IberianRoma	Punjabi1	Iranian	0,00	13,93	36,49	2,62	1088	959	1216	1022	1153
IberianRoma	Punjabi1	Lebanese	0,00	11,31	33,09	2,93	1173	1029	1316	1100	1246
IberianRoma	Punjabi1	Turkish	0,00	12,48	32,24	2,58	1194	1068	1320	1130	1259
IberianRoma	Punjabi1	Jordanian	0,00	13,28	32,30	2,43	1193	1073	1312	1132	1253
IberianRoma	Punjabi1	Georgian	0,00	11,59	36,20	3,12	1095	942	1248	1017	1173
IberianRoma	Punjabi1	Armenian	0,00	16,15	34,09	2,11	1148	1044	1251	1095	1201
IberianRoma	Punjabi1	Balkan	0,00	13,00	28,81	2,22	1280	1171	1389	1224	1335
IberianRoma	Punjabi1	CentralEurope	0,00	13,65	27,66	2,03	1309	1209	1408	1258	1359
IberianRoma	Punjabi1	Basque	0,00	13,91	25,23	1,81	1369	1281	1458	1324	1415
IberianRoma	**Punjabi1**	**IBS**	**0,00**	**14,83**	**26,41**	**1,78**	**1340**	**1253**	**1427**	**1295**	**1384**
EuropeanRoma	Punjabi1	Iranian	0,00	12,85	40,70	3,17	983	827	1138	903	1062
EuropeanRoma	Punjabi1	Lebanese	0,00	11,50	39,91	3,47	1002	832	1172	916	1089
EuropeanRoma	Punjabi1	Turkish	0,00	13,02	37,79	2,90	1055	913	1197	983	1128
EuropeanRoma	Punjabi1	Jordanian	0,00	10,69	39,73	3,72	1007	824	1189	914	1100
EuropeanRoma	Punjabi1	Georgian	0,00	10,79	44,84	4,16	879	675	1083	775	983
EuropeanRoma	Punjabi1	Armenian	0,00	12,91	38,58	2,99	1036	889	1182	961	1110
EuropeanRoma	**Punjabi1**	**Balkan**	**0,00**	**16,03**	**35,67**	**2,18**	**1108**	**1001**	**1215**	**1054**	**1163**
EuropeanRoma	Punjabi1	CentralEurope	0,00	15,80	34,63	2,19	1134	1027	1242	1080	1189

Supplementary Table 4. Results of statistically significant MALDER tests.

Cases of detected admixture dates that were statistically significant (p-value<0.01), z-scores for each test and the average admixture time expressed in generations ago (GA) with relative Standard Error and time in C.E.

a) Average segment length across all categories.

Pairs	P-value (Alpha 0.05)
Iberian Roma and Czech Roma	0,000
Iberian Roma and Macedonian Roma	0,000
Iberian Roma and Lithuanian Roma	0,021
Iberian Roma and Romungro Hungary	0,908
Iberian Roma and Romungro Ukraine	0,000
Iberian Roma and Vlax Hungary	0,016

Category	Pair	P-value (Alpha 0.05)
0-4	Iberian Roma and Czech Roma	0,000
0-4	Iberian Roma and Macedonian Roma	0,113
0-4	Iberian Roma and Lithuanian Roma	0,593
0-4	Iberian Roma and Romungro Hungary	0,923
0-4	Iberian Roma and Romungro Ukraine	0,006
0-4	Iberian Roma and Vlax Hungary	0,048
4-8	Iberian Roma and Czech Roma	0,000
4-8	Iberian Roma and Macedonian Roma	0,025
4-8	Iberian Roma and Lithuanian Roma	0,761
4-8	Iberian Roma and Romungro Hungary	0,848
4-8	Iberian Roma and Romungro Ukraine	0,010
4-8	Iberian Roma and Vlax Hungary	0,273
8-16	Iberian Roma and Czech Roma	0,000
8-16	Iberian Roma and Macedonian Roma	0,021
8-16	Iberian Roma and Lithuanian Roma	0,085
8-16	Iberian Roma and Romungro Hungary	0,776
8-16	Iberian Roma and Romungro Ukraine	0,168
8-16	Iberian Roma and Vlax Hungary	0,862
16-32	Iberian Roma and Czech Roma	0,240
16-32	Iberian Roma and Macedonian Roma	0,731
16-32	Iberian Roma and Lithuanian Roma	0,596
16-32	Iberian Roma and Romungro Hungary	0,893
16-32	Iberian Roma and Romungro Ukraine	0,428
16-32	Iberian Roma and Vlax Hungary	0,825
32-64	Iberian Roma and Czech Roma	0,000
32-64	Iberian Roma and Macedonian Roma	0,011
32-64	Iberian Roma and Lithuanian Roma	0,949
32-64	Iberian Roma and Romungro Hungary	0,490
32-64	Iberian Roma and Romungro Ukraine	0,008
32-64	Iberian Roma and Vlax Hungary	0,165
> 64	Iberian Roma and Czech Roma	0,000
> 64	Iberian Roma and Macedonian Roma	0,036
> 64	Iberian Roma and Lithuanian Roma	0,185
> 64	Iberian Roma and Romungro Hungary	0,547
> 64	Iberian Roma and Romungro Ukraine	0,004
> 64	Iberian Roma and Vlax Hungary	0,099
0-4	Lithuanian Roma and Czech Roma	0,311
0-4	Lithuanian Roma and Macedonian Roma	0,668
0-4	Lithuanian Roma and Iberian Roma	0,593
0-4	Lithuanian Roma and Romungro Hungary	0,674
0-4	Lithuanian Roma and Romungro Ukraine	0,105
0-4	Lithuanian Roma and Vlax Hungary	0,389
4-8	Lithuanian Roma and Czech Roma	0,034
4-8	Lithuanian Roma and Macedonian Roma	0,157
4-8	Lithuanian Roma and Iberian Roma	0,761
4-8	Lithuanian Roma and Romungro Hungary	0,841
4-8	Lithuanian Roma and Romungro Ukraine	0,032
4-8	Lithuanian Roma and Vlax Hungary	0,463
8-16	Lithuanian Roma and Czech Roma	0,068
8-16	Lithuanian Roma and Macedonian Roma	0,537
8-16	Lithuanian Roma and Iberian Roma	0,085
8-16	Lithuanian Roma and Romungro Hungary	0,248
8-16	Lithuanian Roma and Romungro Ukraine	1,000
8-16	Lithuanian Roma and Vlax Hungary	0,340

16-32	Lithuanian Roma and Czech Roma	0,922
16-32	Lithuanian Roma and Macedonian Roma	0,759
16-32	Lithuanian Roma and Iberian Roma	0,596
16-32	Lithuanian Roma and Romungro Hungary	0,548
16-32	Lithuanian Roma and Romungro Ukraine	1,000
16-32	Lithuanian Roma and Vlax Hungary	1,000
32-64	Lithuanian Roma and Czech Roma	0,019
32-64	Lithuanian Roma and Macedonian Roma	0,085
32-64	Lithuanian Roma and Iberian Roma	0,949
32-64	Lithuanian Roma and Romungro Hungary	0,665
32-64	Lithuanian Roma and Romungro Ukraine	0,019
32-64	Lithuanian Roma and Vlax Hungary	0,105
> 64	Lithuanian Roma and Czech Roma	0,095
> 64	Lithuanian Roma and Macedonian Roma	0,803
> 64	Lithuanian Roma and Iberian Roma	0,185
> 64	Lithuanian Roma and Romungro Hungary	0,115
> 64	Lithuanian Roma and Romungro Ukraine	0,064
> 64	Lithuanian Roma and Vlax Hungary	0,462
0-4	Romungro Hungary and Czech Roma	0,240
0-4	Romungro Hungary and Macedonian Roma	0,500
0-4	Romungro Hungary and Iberian Roma	0,923
0-4	Romungro Hungary and Lithuanian Roma	0,674
0-4	Romungro Hungary and Romungro Ukraine	0,217
0-4	Romungro Hungary and Vlax Hungary	0,459
4-8	Romungro Hungary and Czech Roma	0,090
4-8	Romungro Hungary and Macedonian Roma	0,219
4-8	Romungro Hungary and Iberian Roma	0,848
4-8	Romungro Hungary and Lithuanian Roma	0,841
4-8	Romungro Hungary and Romungro Ukraine	0,190
4-8	Romungro Hungary and Vlax Hungary	0,690
8-16	Romungro Hungary and Czech Roma	0,008
8-16	Romungro Hungary and Macedonian Roma	0,086
8-16	Romungro Hungary and Iberian Roma	0,776
8-16	Romungro Hungary and Lithuanian Roma	0,248
8-16	Romungro Hungary and Romungro Ukraine	0,190
8-16	Romungro Hungary and Vlax Hungary	0,753
16-32	Romungro Hungary and Czech Roma	0,281
16-32	Romungro Hungary and Macedonian Roma	0,623
16-32	Romungro Hungary and Iberian Roma	0,893
16-32	Romungro Hungary and Lithuanian Roma	0,548
16-32	Romungro Hungary and Romungro Ukraine	0,268
16-32	Romungro Hungary and Vlax Hungary	1,000
32-64	Romungro Hungary and Czech Roma	0,090
32-64	Romungro Hungary and Macedonian Roma	0,218
32-64	Romungro Hungary and Iberian Roma	0,490
32-64	Romungro Hungary and Lithuanian Roma	0,665
32-64	Romungro Hungary and Romungro Ukraine	0,174
32-64	Romungro Hungary and Vlax Hungary	0,202
> 64	Romungro Hungary and Czech Roma	0,007
> 64	Romungro Hungary and Macedonian Roma	0,137
> 64	Romungro Hungary and Iberian Roma	0,547
> 64	Romungro Hungary and Lithuanian Roma	0,115
> 64	Romungro Hungary and Romungro Ukraine	0,037
> 64	Romungro Hungary and Vlax Hungary	0,249

Supplementary Table 5. Two-sided Wilcoxon rank sum test with continuity correction for European LAI tracts.

A) One-sided test with the alternative hypothesis of less.

Pair 1	Pair 2	IBD Length Category	P-value (Alpha 0.05)
IberianRoma_IberianRoma	CzechRoma_CzechRoma	2-4	0,000
IberianRoma_IberianRoma	LithuanianRoma_LithuanianRoma	2-4	0,000
IberianRoma_IberianRoma	MacedonianRoma_MacedonianRoma	2-4	0,001
IberianRoma_IberianRoma	RomaRomungroHungary_RomaRomungroHungary	2-4	0,708
IberianRoma_IberianRoma	RomaRomungroUkraine_RomaRomungroUkraine	2-4	0,014
IberianRoma_IberianRoma	RomaVlaxHungary_RomaVlaxHungary	2-4	0,002
IberianRoma_IberianRoma	CzechRoma_CzechRoma	4-8	0,000
IberianRoma_IberianRoma	LithuanianRoma_LithuanianRoma	4-8	0,000
IberianRoma_IberianRoma	MacedonianRoma_MacedonianRoma	4-8	0,025
IberianRoma_IberianRoma	RomaRomungroHungary_RomaRomungroHungary	4-8	0,643
IberianRoma_IberianRoma	RomaRomungroUkraine_RomaRomungroUkraine	4-8	0,002
IberianRoma_IberianRoma	RomaVlaxHungary_RomaVlaxHungary	4-8	0,197
IberianRoma_IberianRoma	CzechRoma_CzechRoma	8-16	0,000
IberianRoma_IberianRoma	LithuanianRoma_LithuanianRoma	8-16	0,000
IberianRoma_IberianRoma	MacedonianRoma_MacedonianRoma	8-16	1,000
IberianRoma_IberianRoma	RomaRomungroHungary_RomaRomungroHungary	8-16	0,691
IberianRoma_IberianRoma	RomaRomungroUkraine_RomaRomungroUkraine	8-16	0,090
IberianRoma_IberianRoma	RomaVlaxHungary_RomaVlaxHungary	8-16	0,323
IberianRoma_IberianRoma	CzechRoma_CzechRoma	>16	0,000
IberianRoma_IberianRoma	LithuanianRoma_LithuanianRoma	>16	0,000
IberianRoma_IberianRoma	MacedonianRoma_MacedonianRoma	>16	0,904
IberianRoma_IberianRoma	RomaRomungroHungary_RomaRomungroHungary	>16	0,282
IberianRoma_IberianRoma	RomaRomungroUkraine_RomaRomungroUkraine	>16	0,001
IberianRoma_IberianRoma	RomaVlaxHungary_RomaVlaxHungary	>16	0,273

b) One-sided test with the alternative hypothesis of greater.

Pair 1	Pair 2	IBD Length Category	P-value (Alpha 0.05)
IberianRoma_IberianRoma	CzechRoma_CzechRoma	2-4	1,000
IberianRoma_IberianRoma	LithuanianRoma_LithuanianRoma	2-4	1,000
IberianRoma_IberianRoma	MacedonianRoma_MacedonianRoma	2-4	0,999
IberianRoma_IberianRoma	RomaRomungroHungary_RomaRomungroHungary	2-4	0,293
IberianRoma_IberianRoma	RomaRomungroUkraine_RomaRomungroUkraine	2-4	0,986
IberianRoma_IberianRoma	RomaVlaxHungary_RomaVlaxHungary	2-4	0,998
IberianRoma_IberianRoma	CzechRoma_CzechRoma	4-8	1,000
IberianRoma_IberianRoma	LithuanianRoma_LithuanianRoma	4-8	1,000
IberianRoma_IberianRoma	MacedonianRoma_MacedonianRoma	4-8	0,975
IberianRoma_IberianRoma	RomaRomungroHungary_RomaRomungroHungary	4-8	0,357
IberianRoma_IberianRoma	RomaRomungroUkraine_RomaRomungroUkraine	4-8	0,998
IberianRoma_IberianRoma	RomaVlaxHungary_RomaVlaxHungary	4-8	0,803
IberianRoma_IberianRoma	CzechRoma_CzechRoma	8-16	1,000
IberianRoma_IberianRoma	LithuanianRoma_LithuanianRoma	8-16	1,000
IberianRoma_IberianRoma	MacedonianRoma_MacedonianRoma	8-16	0,000
IberianRoma_IberianRoma	RomaRomungroHungary_RomaRomungroHungary	8-16	0,309
IberianRoma_IberianRoma	RomaRomungroUkraine_RomaRomungroUkraine	8-16	0,910
IberianRoma_IberianRoma	RomaVlaxHungary_RomaVlaxHungary	8-16	0,677
IberianRoma_IberianRoma	CzechRoma_CzechRoma	>16	1,000
IberianRoma_IberianRoma	LithuanianRoma_LithuanianRoma	>16	1,000
IberianRoma_IberianRoma	MacedonianRoma_MacedonianRoma	>16	0,097
IberianRoma_IberianRoma	RomaRomungroHungary_RomaRomungroHungary	>16	0,719
IberianRoma_IberianRoma	RomaRomungroUkraine_RomaRomungroUkraine	>16	0,999
IberianRoma_IberianRoma	RomaVlaxHungary_RomaVlaxHungary	>16	0,728

Supplementary Table 6. Wilcoxon rank sum test with continuity correction for the number of shared IBD segments by length category.

POP1	POP2	IBD Length Category	Average IBD Length	Average IBD Number	Cumulative IBD Length	Cumulative IBD Number	Category
CzechRoma	CzechRoma	2-4	47,48	13,74	9971	2885	Within-population
CzechRoma	CzechRoma	4-8	107,82	19,74	22641	4146	Within-population
CzechRoma	CzechRoma	8-16	54,76	5,33	10460	1018	Within-population
CzechRoma	CzechRoma	>16	51,69	2,33	3463	156	Within-population
IberianRoma	IberianRoma	2-4	30,49	8,82	131436	38041	Within-population
IberianRoma	IberianRoma	4-8	68,62	12,65	294169	54235	Within-population
IberianRoma	IberianRoma	8-16	35,85	3,45	138577	13346	Within-population
IberianRoma	IberianRoma	>16	33,88	1,59	35097	1644	Within-population
LithuanianRoma	LithuanianRoma	2-4	45,53	13,30	455	133	Within-population
LithuanianRoma	LithuanianRoma	4-8	122,96	22,20	1230	222	Within-population
LithuanianRoma	LithuanianRoma	8-16	125,95	11,40	1260	114	Within-population
LithuanianRoma	LithuanianRoma	>16	220,46	9,14	1543	64	Within-population
MacedonianRoma	MacedonianRoma	2-4	38,01	11,02	1710	496	Within-population
MacedonianRoma	MacedonianRoma	4-8	73,89	14,00	3325	630	Within-population
MacedonianRoma	MacedonianRoma	8-16	20,78	2,08	790	79	Within-population
MacedonianRoma	MacedonianRoma	>16	18,92	1,00	95	5	Within-population
RomaRomungroHungary	RomaRomungroHungary	2-4	27,64	8,10	276	81	Within-population
RomaRomungroHungary	RomaRomungroHungary	4-8	63,98	11,90	640	119	Within-population
RomaRomungroHungary	RomaRomungroHungary	8-16	38,11	3,67	343	33	Within-population
RomaRomungroHungary	RomaRomungroHungary	>16	58,98	2,67	177	8	Within-population
RomaRomungroUkraine	RomaRomungroUkraine	2-4	45,21	13,17	271	79	Within-population
RomaRomungroUkraine	RomaRomungroUkraine	4-8	112,56	20,00	675	120	Within-population
RomaRomungroUkraine	RomaRomungroUkraine	8-16	121,45	11,17	729	67	Within-population
RomaRomungroUkraine	RomaRomungroUkraine	>16	237,41	9,25	950	37	Within-population
RomaVlaxHungary	RomaVlaxHungary	2-4	48,24	13,83	289	83	Within-population
RomaVlaxHungary	RomaVlaxHungary	4-8	72,71	14,00	436	84	Within-population
RomaVlaxHungary	RomaVlaxHungary	8-16	33,45	3,40	167	17	Within-population
RomaVlaxHungary	RomaVlaxHungary	>16	35,97	1,50	72	3	Within-population
CzechRoma	Croatian	2-4	3,40	1,00	58	17	Roma-nonRoma
CzechRoma	Croatian	4-8	4,99	1,08	30	7	Roma-nonRoma
CzechRoma	Croatian	8-16	8,09	1,00	4	1	Roma-nonRoma
CzechRoma	Czech	2-4	3,61	1,09	190	57	Roma-nonRoma
CzechRoma	Czech	4-8	4,96	1,00	50	10	Roma-nonRoma
CzechRoma	Czech	8-16	10,48	1,00	10	1	Roma-nonRoma
CzechRoma	Greek	2-4	3,98	1,17	197	58	Roma-nonRoma
CzechRoma	Greek	4-8	6,52	1,22	163	31	Roma-nonRoma
CzechRoma	Greek	8-16	10,82	1,00	49	5	Roma-nonRoma
CzechRoma	Hungarian	2-4	3,70	1,10	148	44	Roma-nonRoma
CzechRoma	Hungarian	4-8	5,21	1,06	44	9	Roma-nonRoma
CzechRoma	IBS	2-4	3,86	1,16	461	139	Roma-nonRoma
CzechRoma	IBS	4-8	6,86	1,33	230	45	Roma-nonRoma
CzechRoma	IBS	8-16	11,22	1,13	45	5	Roma-nonRoma
CzechRoma	Romanian	2-4	3,62	1,05	69	20	Roma-nonRoma
CzechRoma	Romanian	4-8	5,42	1,09	30	6	Roma-nonRoma
IberianRoma	Croatian	2-4	3,68	1,06	173	50	Roma-nonRoma
IberianRoma	Croatian	4-8	4,98	1,00	140	28	Roma-nonRoma
IberianRoma	Czech	2-4	3,53	1,07	664	202	Roma-nonRoma
IberianRoma	Czech	4-8	5,06	1,01	169	34	Roma-nonRoma
IberianRoma	Czech	8-16	11,52	1,00	12	1	Roma-nonRoma
IberianRoma	Greek	2-4	3,77	1,12	725	217	Roma-nonRoma
IberianRoma	Greek	4-8	6,26	1,19	701	134	Roma-nonRoma
IberianRoma	Greek	8-16	9,54	1,00	124	13	Roma-nonRoma
IberianRoma	Hungarian	2-4	3,59	1,08	611	183	Roma-nonRoma
IberianRoma	Hungarian	4-8	5,21	1,05	208	42	Roma-nonRoma
IberianRoma	IBS	2-4	3,90	1,16	3287	980	Roma-nonRoma
IberianRoma	IBS	4-8	7,20	1,36	1907	360	Roma-nonRoma
IberianRoma	IBS	8-16	13,31	1,29	346	34	Roma-nonRoma
IberianRoma	IBS	>16	25,03	1,00	50	2	Roma-nonRoma
IberianRoma	Romanian	2-4	3,64	1,08	269	80	Roma-nonRoma
IberianRoma	Romanian	4-8	5,04	1,03	101	21	Roma-nonRoma
LithuanianRoma	Croatian	2-4	3,24	1,00	15	5	Roma-nonRoma
LithuanianRoma	Croatian	4-8	5,60	1,00	11	2	Roma-nonRoma
LithuanianRoma	Czech	2-4	3,58	1,10	36	11	Roma-nonRoma
LithuanianRoma	Czech	4-8	4,91	1,00	25	5	Roma-nonRoma
LithuanianRoma	Greek	2-4	4,06	1,21	49	15	Roma-nonRoma
LithuanianRoma	Greek	4-8	5,79	1,09	67	13	Roma-nonRoma
LithuanianRoma	Greek	8-16	11,01	1,00	6	1	Roma-nonRoma
LithuanianRoma	Hungarian	2-4	3,33	1,00	30	9	Roma-nonRoma
LithuanianRoma	Hungarian	4-8	5,24	1,00	13	3	Roma-nonRoma
LithuanianRoma	IBS	2-4	3,98	1,20	137	42	Roma-nonRoma
LithuanianRoma	IBS	4-8	7,52	1,46	105	21	Roma-nonRoma
LithuanianRoma	IBS	8-16	9,36	1,00	5	1	Roma-nonRoma
LithuanianRoma	Romanian	2-4	3,73	1,11	17	5	Roma-nonRoma
LithuanianRoma	Romanian	4-8	4,82	1,00	5	1	Roma-nonRoma
MacedonianRoma	Croatian	2-4	3,30	1,00	17	5	Roma-nonRoma
MacedonianRoma	Croatian	4-8	4,88	1,00	15	3	Roma-nonRoma
MacedonianRoma	Croatian	8-16	8,09	1,00	4	1	Roma-nonRoma
MacedonianRoma	Czech	2-4	3,47	1,07	75	23	Roma-nonRoma
MacedonianRoma	Czech	4-8	5,52	1,00	39	7	Roma-nonRoma
MacedonianRoma	Greek	2-4	3,61	1,10	90	28	Roma-nonRoma
MacedonianRoma	Greek	4-8	7,15	1,34	114	22	Roma-nonRoma
MacedonianRoma	Greek	8-16	10,01	1,00	30	3	Roma-nonRoma
MacedonianRoma	Greek	>16	18,23	1,00	9	1	Roma-nonRoma
MacedonianRoma	Hungarian	2-4	3,35	1,00	65	20	Roma-nonRoma
MacedonianRoma	Hungarian	4-8	5,06	1,00	23	5	Roma-nonRoma
MacedonianRoma	IBS	2-4	3,79	1,15	190	58	Roma-nonRoma
MacedonianRoma	IBS	4-8	8,34	1,65	96	19	Roma-nonRoma
MacedonianRoma	IBS	8-16	13,31	1,33	20	2	Roma-nonRoma
MacedonianRoma	Romanian	2-4	3,48	1,00	31	9	Roma-nonRoma
MacedonianRoma	Romanian	4-8	5,26	1,00	29	6	Roma-nonRoma

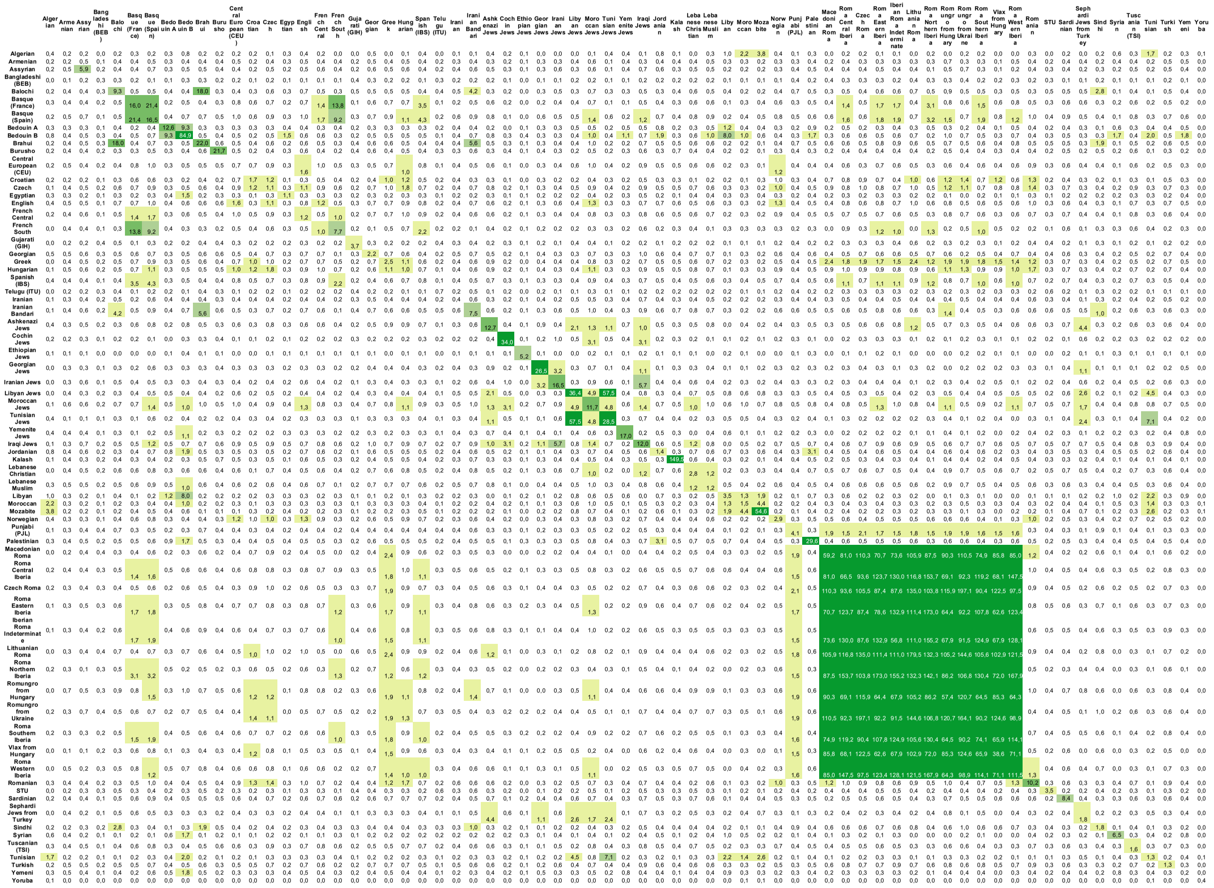
RomaRomungroHungary	Croatian	2-4	3,82	1,14	13	4	Roma-nonRoma
RomaRomungroHungary	Croatian	4-8	4,87	1,00	10	2	Roma-nonRoma
RomaRomungroHungary	Croatian	8-16	12,13	1,00	6	1	Roma-nonRoma
RomaRomungroHungary	Czech	2-4	3,89	1,18	54	17	Roma-nonRoma
RomaRomungroHungary	Czech	4-8	4,56	1,00	18	4	Roma-nonRoma
RomaRomungroHungary	Greek	2-4	3,83	1,14	56	17	Roma-nonRoma
RomaRomungroHungary	Greek	4-8	5,43	1,08	33	7	Roma-nonRoma
RomaRomungroHungary	Greek	8-16	9,17	1,00	9	1	Roma-nonRoma
RomaRomungroHungary	Hungarian	2-4	3,86	1,17	44	14	Roma-nonRoma
RomaRomungroHungary	Hungarian	4-8	4,80	1,00	10	2	Roma-nonRoma
RomaRomungroHungary	IBS	2-4	3,88	1,17	157	48	Roma-nonRoma
RomaRomungroHungary	IBS	4-8	6,63	1,19	53	10	Roma-nonRoma
RomaRomungroHungary	IBS	8-16	9,10	1,00	9	1	Roma-nonRoma
RomaRomungroHungary	Romanian	2-4	3,25	1,00	18	6	Roma-nonRoma
RomaRomungroHungary	Romanian	4-8	4,74	1,00	7	2	Roma-nonRoma
RomaRomungroUkraine	Croatian	2-4	3,81	1,10	19	6	Roma-nonRoma
RomaRomungroUkraine	Croatian	4-8	5,50	1,00	8	2	Roma-nonRoma
RomaRomungroUkraine	Czech	2-4	3,35	1,05	35	11	Roma-nonRoma
RomaRomungroUkraine	Czech	4-8	4,46	1,00	18	4	Roma-nonRoma
RomaRomungroUkraine	Greek	2-4	4,11	1,25	33	10	Roma-nonRoma
RomaRomungroUkraine	Greek	4-8	5,04	1,08	33	7	Roma-nonRoma
RomaRomungroUkraine	Greek	8-16	19,24	2,00	10	1	Roma-nonRoma
RomaRomungroUkraine	Hungarian	2-4	3,30	1,00	30	9	Roma-nonRoma
RomaRomungroUkraine	Hungarian	4-8	4,27	1,00	13	3	Roma-nonRoma
RomaRomungroUkraine	Hungarian	8-16	9,06	1,00	9	1	Roma-nonRoma
RomaRomungroUkraine	IBS	2-4	3,37	1,04	86	27	Roma-nonRoma
RomaRomungroUkraine	IBS	4-8	6,83	1,28	62	12	Roma-nonRoma
RomaRomungroUkraine	Romanian	2-4	3,23	1,00	3	1	Roma-nonRoma
RomaRomungroUkraine	Romanian	4-8	4,74	1,00	5	1	Roma-nonRoma
RomaVlaxHungary	Croatian	2-4	4,23	1,25	8	3	Roma-nonRoma
RomaVlaxHungary	Croatian	4-8	5,40	1,00	22	4	Roma-nonRoma
RomaVlaxHungary	Czech	2-4	4,14	1,21	29	9	Roma-nonRoma
RomaVlaxHungary	Czech	4-8	5,10	1,00	20	4	Roma-nonRoma
RomaVlaxHungary	Greek	2-4	3,78	1,17	23	7	Roma-nonRoma
RomaVlaxHungary	Greek	4-8	6,28	1,31	41	9	Roma-nonRoma
RomaVlaxHungary	Greek	8-16	9,53	1,00	10	1	Roma-nonRoma
RomaVlaxHungary	Hungarian	2-4	3,93	1,17	24	7	Roma-nonRoma
RomaVlaxHungary	Hungarian	4-8	5,64	1,00	20	4	Roma-nonRoma
RomaVlaxHungary	IBS	2-4	3,89	1,17	68	21	Roma-nonRoma
RomaVlaxHungary	IBS	4-8	6,84	1,25	82	15	Roma-nonRoma
RomaVlaxHungary	IBS	8-16	11,56	1,00	6	1	Roma-nonRoma
RomaVlaxHungary	Romanian	2-4	3,22	1,00	5	2	Roma-nonRoma
RomaVlaxHungary	Romanian	4-8	4,25	1,00	9	2	Roma-nonRoma
CzechRoma	IberianRoma	2-4	28,74	8,33	28163	8163	Between-Roma
CzechRoma	IberianRoma	4-8	52,04	9,82	50923	9613	Between-Roma
CzechRoma	IberianRoma	8-16	17,76	1,81	11817	1205	Between-Roma
CzechRoma	IberianRoma	>16	18,17	1,04	445	26	Between-Roma
CzechRoma	LithuanianRoma	2-4	38,54	11,14	2023	585	Between-Roma
CzechRoma	LithuanianRoma	4-8	73,20	13,88	3843	729	Between-Roma
CzechRoma	LithuanianRoma	8-16	25,76	2,53	1185	117	Between-Roma
CzechRoma	LithuanianRoma	>16	17,92	1,00	36	2	Between-Roma
CzechRoma	MacedonianRoma	2-4	36,70	10,65	3835	1113	Between-Roma
CzechRoma	MacedonianRoma	4-8	60,53	11,45	6356	1203	Between-Roma
CzechRoma	MacedonianRoma	8-16	18,00	1,83	1341	136	Between-Roma
CzechRoma	MacedonianRoma	>16	21,03	1,00	53	3	Between-Roma
CzechRoma	RomaRomungroHungary	2-4	32,84	9,53	1724	501	Between-Roma
CzechRoma	RomaRomungroHungary	4-8	62,23	11,68	3267	613	Between-Roma
CzechRoma	RomaRomungroHungary	8-16	21,54	2,11	959	94	Between-Roma
CzechRoma	RomaRomungroHungary	>16	20,57	1,08	134	7	Between-Roma
CzechRoma	RomaRomungroUkraine	2-4	46,94	13,54	1971	569	Between-Roma
CzechRoma	RomaRomungroUkraine	4-8	102,52	18,74	4306	787	Between-Roma
CzechRoma	RomaRomungroUkraine	8-16	42,61	4,21	1747	173	Between-Roma
CzechRoma	RomaRomungroUkraine	>16	23,20	1,14	255	13	Between-Roma
CzechRoma	RomaVlaxHungary	2-4	39,99	11,53	1760	508	Between-Roma
CzechRoma	RomaVlaxHungary	4-8	81,31	15,33	3496	659	Between-Roma
CzechRoma	RomaVlaxHungary	8-16	29,18	2,95	1123	114	Between-Roma
CzechRoma	RomaVlaxHungary	>16	17,40	1,00	52	3	Between-Roma
IberianRoma	CzechRoma	2-4	28,74	8,33	28163	8163	Between-Roma
IberianRoma	CzechRoma	4-8	52,04	9,82	50923	9613	Between-Roma
IberianRoma	CzechRoma	8-16	17,76	1,81	11817	1205	Between-Roma
IberianRoma	CzechRoma	>16	18,17	1,04	445	26	Between-Roma
IberianRoma	LithuanianRoma	2-4	32,34	9,36	7551	2187	Between-Roma
IberianRoma	LithuanianRoma	4-8	63,01	11,77	14681	2744	Between-Roma
IberianRoma	LithuanianRoma	8-16	23,76	2,39	4502	453	Between-Roma
IberianRoma	LithuanianRoma	>16	19,34	1,07	135	8	Between-Roma
IberianRoma	MacedonianRoma	2-4	26,29	7,63	12197	3541	Between-Roma
IberianRoma	MacedonianRoma	4-8	44,06	8,38	20467	3891	Between-Roma
IberianRoma	MacedonianRoma	8-16	14,88	1,54	3660	378	Between-Roma
IberianRoma	MacedonianRoma	>16	17,57	1,00	97	6	Between-Roma
IberianRoma	RomaRomungroHungary	2-4	22,57	6,56	5247	1526	Between-Roma
IberianRoma	RomaRomungroHungary	4-8	38,52	7,38	8878	1700	Between-Roma
IberianRoma	RomaRomungroHungary	8-16	14,73	1,50	1937	198	Between-Roma
IberianRoma	RomaRomungroHungary	>16	17,48	1,00	52	3	Between-Roma
IberianRoma	RomaRomungroUkraine	2-4	30,33	8,78	5642	1634	Between-Roma
IberianRoma	RomaRomungroUkraine	4-8	51,12	9,63	9508	1791	Between-Roma
IberianRoma	RomaRomungroUkraine	8-16	17,59	1,82	2507	259	Between-Roma
IberianRoma	RomaRomungroUkraine	>16	18,20	1,00	64	4	Between-Roma
IberianRoma	RomaVlaxHungary	2-4	29,14	8,49	5507	1605	Between-Roma
IberianRoma	RomaVlaxHungary	4-8	45,61	8,68	8597	1636	Between-Roma
IberianRoma	RomaVlaxHungary	8-16	14,64	1,54	1705	179	Between-Roma
IberianRoma	RomaVlaxHungary	>16	17,22	1,00	34	2	Between-Roma
LithuanianRoma	CzechRoma	2-4	38,54	11,14	2023	585	Between-Roma
LithuanianRoma	CzechRoma	4-8	73,20	13,88	3843	729	Between-Roma
LithuanianRoma	CzechRoma	8-16	25,76	2,53	1185	117	Between-Roma
LithuanianRoma	CzechRoma	>16	17,92	1,00	36	2	Between-Roma
LithuanianRoma	IberianRoma	2-4	32,34	9,36	7551	2187	Between-Roma
LithuanianRoma	IberianRoma	4-8	63,01	11,77	14681	2744	Between-Roma
LithuanianRoma	IberianRoma	8-16	23,76	2,39	4502	453	Between-Roma
LithuanianRoma	IberianRoma	>16	19,34	1,07	135	8	Between-Roma
LithuanianRoma	MacedonianRoma	2-4	33,44	9,64	836	241	Between-Roma
LithuanianRoma	MacedonianRoma	4-8	60,43	11,54	1511	289	Between-Roma
LithuanianRoma	MacedonianRoma	8-16	16,77	1,69	302	31	Between-Roma

LithuanianRoma	RomaRomungroHungary	2-4	32,69	9,52	409	119	Between-Roma
LithuanianRoma	RomaRomungroHungary	4-8	54,86	10,36	686	130	Between-Roma
LithuanianRoma	RomaRomungroHungary	8-16	19,65	2,00	197	20	Between-Roma
LithuanianRoma	RomaRomungroHungary	>16	16,34	1,00	25	2	Between-Roma
LithuanianRoma	RomaRomungroUkraine	2-4	43,34	12,45	433	125	Between-Roma
LithuanianRoma	RomaRomungroUkraine	4-8	72,54	13,65	725	137	Between-Roma
LithuanianRoma	RomaRomungroUkraine	8-16	29,16	2,89	277	28	Between-Roma
LithuanianRoma	RomaRomungroUkraine	>16	20,86	1,00	10	1	Between-Roma
LithuanianRoma	RomaVlaxHungary	2-4	40,27	11,65	403	117	Between-Roma
LithuanianRoma	RomaVlaxHungary	4-8	67,99	12,70	680	127	Between-Roma
LithuanianRoma	RomaVlaxHungary	8-16	24,46	2,56	196	21	Between-Roma
LithuanianRoma	RomaVlaxHungary	>16	16,52	1,00	8	1	Between-Roma
MacedonianRoma	CzechRoma	2-4	36,70	10,65	3835	1113	Between-Roma
MacedonianRoma	CzechRoma	4-8	60,53	11,45	6356	1203	Between-Roma
MacedonianRoma	CzechRoma	8-16	18,00	1,83	1341	136	Between-Roma
MacedonianRoma	CzechRoma	>16	21,03	1,00	53	3	Between-Roma
MacedonianRoma	IberianRoma	2-4	26,29	7,63	12197	3541	Between-Roma
MacedonianRoma	IberianRoma	4-8	44,06	8,38	20467	3891	Between-Roma
MacedonianRoma	IberianRoma	8-16	14,88	1,54	3660	378	Between-Roma
MacedonianRoma	IberianRoma	>16	17,57	1,00	97	6	Between-Roma
MacedonianRoma	LithuanianRoma	2-4	33,44	9,64	836	241	Between-Roma
MacedonianRoma	LithuanianRoma	4-8	60,43	11,54	1511	289	Between-Roma
MacedonianRoma	LithuanianRoma	8-16	16,77	1,69	302	31	Between-Roma
MacedonianRoma	RomaRomungroHungary	2-4	31,01	9,10	775	228	Between-Roma
MacedonianRoma	RomaRomungroHungary	4-8	49,50	9,40	1238	235	Between-Roma
MacedonianRoma	RomaRomungroHungary	8-16	14,11	1,41	226	23	Between-Roma
MacedonianRoma	RomaRomungroHungary	>16	18,04	1,00	18	1	Between-Roma
MacedonianRoma	RomaRomungroUkraine	2-4	38,61	11,23	772	225	Between-Roma
MacedonianRoma	RomaRomungroUkraine	4-8	59,66	11,45	1193	229	Between-Roma
MacedonianRoma	RomaRomungroUkraine	8-16	15,68	1,63	235	25	Between-Roma
MacedonianRoma	RomaRomungroUkraine	>16	17,39	1,00	9	1	Between-Roma
MacedonianRoma	RomaVlaxHungary	2-4	33,56	9,78	688	201	Between-Roma
MacedonianRoma	RomaVlaxHungary	4-8	59,71	11,35	1194	227	Between-Roma
MacedonianRoma	RomaVlaxHungary	8-16	18,72	1,96	262	28	Between-Roma
RomaRomungroHungary	CzechRoma	2-4	32,84	9,53	1724	501	Between-Roma
RomaRomungroHungary	CzechRoma	4-8	62,23	11,68	3267	613	Between-Roma
RomaRomungroHungary	CzechRoma	8-16	21,54	2,11	959	94	Between-Roma
RomaRomungroHungary	CzechRoma	>16	20,57	1,08	134	7	Between-Roma
RomaRomungroHungary	IberianRoma	2-4	22,57	6,56	5247	1526	Between-Roma
RomaRomungroHungary	IberianRoma	4-8	38,52	7,38	8878	1700	Between-Roma
RomaRomungroHungary	IberianRoma	8-16	14,73	1,50	1937	198	Between-Roma
RomaRomungroHungary	IberianRoma	>16	17,48	1,00	52	3	Between-Roma
RomaRomungroHungary	LithuanianRoma	2-4	32,69	9,52	409	119	Between-Roma
RomaRomungroHungary	LithuanianRoma	4-8	54,86	10,36	686	130	Between-Roma
RomaRomungroHungary	LithuanianRoma	8-16	19,65	2,00	197	20	Between-Roma
RomaRomungroHungary	LithuanianRoma	>16	16,34	1,00	25	2	Between-Roma
RomaRomungroHungary	MacedonianRoma	2-4	31,01	9,10	775	228	Between-Roma
RomaRomungroHungary	MacedonianRoma	4-8	49,50	9,40	1238	235	Between-Roma
RomaRomungroHungary	MacedonianRoma	8-16	14,11	1,41	226	23	Between-Roma
RomaRomungroHungary	MacedonianRoma	>16	18,04	1,00	18	1	Between-Roma
RomaRomungroHungary	RomaRomungroUkraine	2-4	27,55	8,10	276	81	Between-Roma
RomaRomungroHungary	RomaRomungroUkraine	4-8	66,96	12,35	670	124	Between-Roma
RomaRomungroHungary	RomaRomungroUkraine	8-16	21,33	2,20	213	22	Between-Roma
RomaRomungroHungary	RomaRomungroUkraine	>16	24,14	1,00	48	2	Between-Roma
RomaRomungroHungary	RomaVlaxHungary	2-4	30,81	8,81	323	93	Between-Roma
RomaRomungroHungary	RomaVlaxHungary	4-8	61,64	11,75	616	118	Between-Roma
RomaRomungroHungary	RomaVlaxHungary	8-16	19,68	1,83	118	11	Between-Roma
RomaRomungroHungary	RomaVlaxHungary	>16	17,07	1,00	9	1	Between-Roma
RomaRomungroUkraine	CzechRoma	2-4	46,94	13,54	1971	569	Between-Roma
RomaRomungroUkraine	CzechRoma	4-8	102,52	18,74	4306	787	Between-Roma
RomaRomungroUkraine	CzechRoma	8-16	42,61	4,21	1747	173	Between-Roma
RomaRomungroUkraine	CzechRoma	>16	23,20	1,14	255	13	Between-Roma
RomaRomungroUkraine	IberianRoma	2-4	30,33	8,78	5642	1634	Between-Roma
RomaRomungroUkraine	IberianRoma	4-8	51,12	9,63	9508	1791	Between-Roma
RomaRomungroUkraine	IberianRoma	8-16	17,59	1,82	2507	259	Between-Roma
RomaRomungroUkraine	IberianRoma	>16	18,20	1,00	64	4	Between-Roma
RomaRomungroUkraine	LithuanianRoma	2-4	43,34	12,45	433	125	Between-Roma
RomaRomungroUkraine	LithuanianRoma	4-8	72,54	13,65	725	137	Between-Roma
RomaRomungroUkraine	LithuanianRoma	8-16	29,16	2,89	277	28	Between-Roma
RomaRomungroUkraine	LithuanianRoma	>16	20,86	1,00	10	1	Between-Roma
RomaRomungroUkraine	MacedonianRoma	2-4	38,61	11,23	772	225	Between-Roma
RomaRomungroUkraine	MacedonianRoma	4-8	59,66	11,45	1193	229	Between-Roma
RomaRomungroUkraine	MacedonianRoma	8-16	15,68	1,63	235	25	Between-Roma
RomaRomungroUkraine	MacedonianRoma	>16	17,39	1,00	9	1	Between-Roma
RomaRomungroUkraine	RomaRomungroHungary	2-4	27,55	8,10	276	81	Between-Roma
RomaRomungroUkraine	RomaRomungroHungary	4-8	66,96	12,35	670	124	Between-Roma
RomaRomungroUkraine	RomaRomungroHungary	8-16	21,33	2,20	213	22	Between-Roma
RomaRomungroUkraine	RomaRomungroHungary	>16	24,14	1,00	48	2	Between-Roma
RomaRomungroUkraine	RomaVlaxHungary	2-4	46,12	13,19	369	106	Between-Roma
RomaRomungroUkraine	RomaVlaxHungary	4-8	76,79	14,25	614	114	Between-Roma
RomaRomungroUkraine	RomaVlaxHungary	8-16	31,70	3,13	254	25	Between-Roma
RomaRomungroUkraine	RomaVlaxHungary	>16	17,87	1,00	9	1	Between-Roma
RomaVlaxHungary	CzechRoma	2-4	39,99	11,53	1760	508	Between-Roma
RomaVlaxHungary	CzechRoma	4-8	81,31	15,33	3496	659	Between-Roma
RomaVlaxHungary	CzechRoma	8-16	29,18	2,95	1123	114	Between-Roma
RomaVlaxHungary	CzechRoma	>16	17,40	1,00	52	3	Between-Roma
RomaVlaxHungary	IberianRoma	2-4	29,14	8,49	5507	1605	Between-Roma
RomaVlaxHungary	IberianRoma	4-8	45,61	8,68	8597	1636	Between-Roma
RomaVlaxHungary	IberianRoma	8-16	14,64	1,54	1705	179	Between-Roma
RomaVlaxHungary	IberianRoma	>16	17,22	1,00	34	2	Between-Roma
RomaVlaxHungary	LithuanianRoma	2-4	40,27	11,65	403	117	Between-Roma
RomaVlaxHungary	LithuanianRoma	4-8	67,99	12,70	680	127	Between-Roma
RomaVlaxHungary	LithuanianRoma	8-16	24,46	2,56	196	21	Between-Roma
RomaVlaxHungary	LithuanianRoma	>16	16,52	1,00	8	1	Between-Roma
RomaVlaxHungary	MacedonianRoma	2-4	33,56	9,78	688	201	Between-Roma
RomaVlaxHungary	MacedonianRoma	4-8	59,71	11,35	1194	227	Between-Roma
RomaVlaxHungary	MacedonianRoma	8-16	18,72	1,96	262	28	Between-Roma
RomaVlaxHungary	RomaRomungroHungary	2-4	30,81	8,81	323	93	Between-Roma
RomaVlaxHungary	RomaRomungroHungary	4-8	61,64	11,75	616	118	Between-Roma
RomaVlaxHungary	RomaRomungroHungary	8-16	19,68	1,83	118	11	Between-Roma
RomaVlaxHungary	RomaRomungroHungary	>16	17,07	1,00	9	1	Between-Roma
RomaVlaxHungary	RomaRomungroUkraine	2-4	46,12	13,19	369	106	Between-Roma
RomaVlaxHungary	RomaRomungroUkraine	4-8	76,79	14,25	614	114	Between-Roma
RomaVlaxHungary	RomaRomungroUkraine	8-16	31,70	3,13	254	25	Between-Roma
RomaVlaxHungary	RomaRomungroUkraine	>16	17,87	1,00	9	1	Between-Roma

Supplementary Table 7. Distribution of shared IBD segments by category with general statistical gata.

	Source of Variation	Percentage of variation
a) Iberian Roma	Between individuals	1,07
	Within individuals	98,93
		Genetic differentiation (ΦPT): 0.0001*
b) IBS	Between individuals	0,08
	Within individuals	99,92
		Genetic differentiation (ΦPT): 0.0021*
c) Roma-IBS	Between individuals	2,79
	Within individuals	97,21
		Genetic differentiation (ΦPT): 0.0001*
d) Iberian Roma	Between regions	0,55
	Between ind. within reg.	1,47
	Within individuals	97,98
		Genetic differentiation (ΦPT): 0.0001*
e) IBS	Between regions	0,04
	Between ind. within reg.	0,38
	Within individuals	99,58
		Genetic differentiation (ΦPT): 0.0019*
f) Roma-IBS	Between populations	1,43
	Between ind. within pop.	1,24
	Within individuals	97,33
		Genetic differentiation (ΦPT): 0.0001*

Supplementary Table 8. AMOVA analysis of the Iberian dataset. Six hypotheses were tested across the Iberian peninsula: a) genetic differentiation among Iberian Roma individuals, b) genetic differentiation among IBS individuals, c) genetic differentiation between Roma and IBS individuals, d) genetic differentiation among Roma individuals grouped by region, e) genetic differentiation among IBS individuals grouped by region, and f) genetic differentiation between Roma and IBS populations.

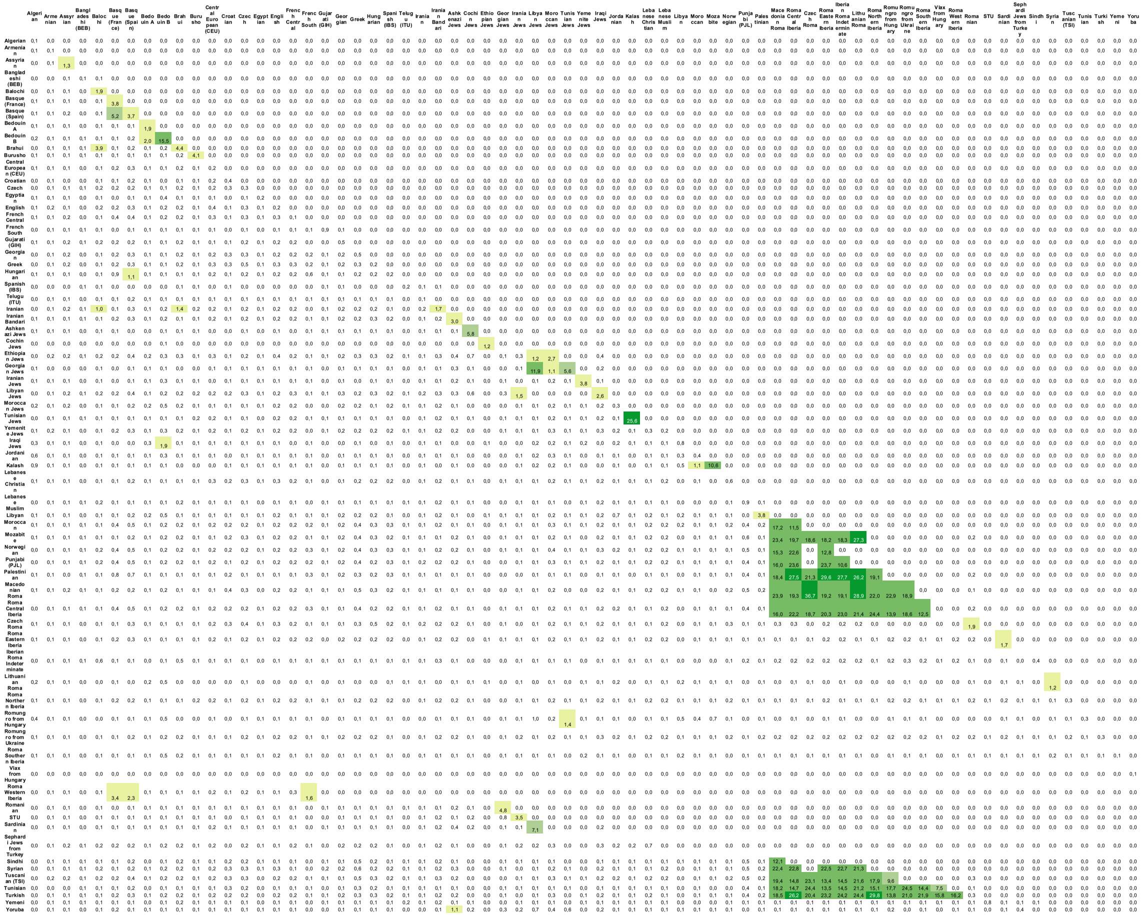


	Algerian	Armenian	Assyrian	Bangladeshi (BEB)	Balochi	Basque (France)	Basque (Spain)	Bedouin A	Bedouin B	Brahui	Burusho	Central European (CEU)	Croatian	Czech	Egyptian	English	French Central	French South	Gujarati (GIH)	Georgian	Greek	Hungarian	Spanish (IBS)
Algerian	0,4	0,2	0,2	0,0	0,2	0,3	0,2	0,3	0,8	0,2	0,2	0,2	0,2	0,1	0,3	0,4	0,2	0,0	0,0	0,5	0,0	0,1	0,4
Armenian	0,2	0,2	0,5	0,1	0,4	0,4	0,5	0,3	0,4	0,4	0,4	0,5	0,2	0,4	0,3	0,5	0,4	0,4	0,2	0,5	0,4	0,5	0,4
Assyrian	0,2	0,5	5,9	0,2	0,4	0,4	0,7	0,3	0,5	0,5	0,4	0,4	0,2	0,5	0,2	0,5	0,6	0,4	0,2	0,6	0,5	0,6	0,4
Bangladeshi (BEB)	0,0	0,1	0,2	0,3	0,3	0,2	0,1	0,1	0,3	0,3	0,2	0,2	0,1	0,2	0,1	0,1	0,1	0,1	0,4	0,3	0,2	0,2	0,1
Balochi	0,2	0,4	0,4	0,3	9,3	0,5	0,5	0,4	0,4	18,0	0,3	0,4	0,3	0,6	0,3	0,7	0,5	0,3	0,5	0,5	0,5	0,5	0,4
Basque (France)	0,3	0,4	0,4	0,2	0,5	16,0	21,4	0,2	0,5	0,4	0,3	0,8	0,6	0,7	0,2	0,7	1,4	13,8	0,1	0,6	0,7	0,7	3,5
Basque (Spain)	0,2	0,5	0,7	0,1	0,5	21,4	16,5	0,4	0,7	0,7	0,5	1,0	0,6	0,9	0,3	1,0	1,7	9,2	0,3	0,7	0,9	1,1	4,3
Bedouin A	0,3	0,3	0,3	0,1	0,4	0,2	0,4	12,6	9,3	0,3	0,3	0,3	0,3	0,3	0,4	0,4	0,3	0,4	0,2	0,3	0,3	0,3	0,3
Bedouin B	0,8	0,4	0,5	0,3	0,4	0,5	0,7	9,3	84,9	0,5	0,4	0,5	0,2	0,5	1,5	0,6	0,6	0,3	0,2	0,5	0,5	0,5	0,5
Brahui	0,2	0,4	0,5	0,3	18,0	0,4	0,7	0,3	0,5	22,0	0,6	0,5	0,4	0,6	0,2	0,6	0,5	0,3	0,4	0,6	0,6	0,5	0,4
Burusho	0,2	0,4	0,4	0,2	0,3	0,3	0,5	0,3	0,4	0,6	21,7	0,5	0,2	0,4	0,3	0,6	0,4	0,2	0,4	0,6	0,4	0,5	0,4
Central European (CEU)	0,2	0,5	0,4	0,2	0,4	0,8	1,0	0,3	0,5	0,5	0,5	0,7	0,7	0,9	0,3	1,6	1,0	0,5	0,3	0,5	0,7	1,0	0,8
Croatian	0,2	0,2	0,2	0,1	0,3	0,6	0,6	0,3	0,2	0,4	0,2	0,7	1,7	1,2	0,1	0,3	0,5	0,4	0,2	0,4	1,0	1,2	0,5
Czech	0,1	0,4	0,5	0,2	0,6	0,7	0,9	0,3	0,5	0,6	0,4	0,9	1,2	1,1	0,3	1,1	0,9	0,6	0,2	0,7	1,0	1,8	0,7
Egyptian	0,3	0,3	0,2	0,1	0,3	0,2	0,3	0,4	1,5	0,2	0,3	0,3	0,1	0,3	1,1	0,3	0,3	0,3	0,2	0,3	0,2	0,3	0,3
English	0,4	0,5	0,5	0,1	0,7	0,7	1,0	0,4	0,6	0,6	0,6	1,6	0,3	1,1	0,3	0,8	1,2	0,5	0,3	0,7	0,7	0,9	0,8
French Central	0,2	0,4	0,6	0,1	0,5	1,4	1,7	0,3	0,6	0,5	0,4	1,0	0,5	0,9	0,3	1,2	0,5	1,0	0,2	0,7	0,7	1,0	0,9
French South	0,0	0,4	0,4	0,1	0,3	13,8	9,2	0,4	0,3	0,3	0,2	0,5	0,4	0,6	0,3	0,5	1,0	7,7	0,2	0,1	0,5	0,7	2,2
Gujarati (GIH)	0,0	0,2	0,2	0,4	0,5	0,1	0,3	0,2	0,2	0,4	0,4	0,3	0,2	0,2	0,2	0,3	0,2	0,2	3,7	0,3	0,2	0,2	0,2
Georgian	0,5	0,5	0,6	0,3	0,5	0,6	0,7	0,3	0,5	0,6	0,6	0,5	0,4	0,7	0,3	0,7	0,7	0,1	0,3	2,2	0,7	0,6	0,4
Greek	0,0	0,4	0,5	0,2	0,5	0,7	0,9	0,3	0,5	0,6	0,4	0,7	1,0	1,0	0,2	0,7	0,7	0,5	0,2	0,7	2,5	1,1	0,6
Hungarian	0,1	0,5	0,6	0,2	0,5	0,7	1,1	0,3	0,5	0,5	0,5	1,0	1,2	1,8	0,3	0,9	1,0	0,7	0,2	0,6	1,1	1,0	0,7
Spanish (IBS)	0,4	0,4	0,4	0,1	0,4	3,5	4,3	0,3	0,5	0,4	0,4	0,8	0,5	0,7	0,3	0,8	0,9	2,2	0,2	0,4	0,6	0,7	0,8
Telugu (ITU)	0,0	0,2	0,2	0,3	0,4	0,1	0,2	0,2	0,1	0,4	0,3	0,1	0,2	0,2	0,1	0,2	0,2	0,2	0,4	0,2	0,2	0,1	0,1
Iranian	0,1	0,3	0,4	0,2	0,5	0,2	0,6	0,4	0,4	0,4	0,3	0,4	0,4	0,4	0,2	0,4	0,4	0,2	0,3	0,5	0,4	0,4	0,4
Iranian Bandari	0,1	0,4	0,6	0,2	4,2	0,5	0,9	0,4	0,7	5,6	0,6	0,5	0,3	0,7	0,2	0,7	0,6	0,1	0,4	0,7	0,6	0,6	0,5
Ashkenazi Jews	0,4	0,3	0,5	0,2	0,3	0,6	0,8	0,2	0,8	0,5	0,3	0,6	0,3	0,8	0,3	0,6	0,6	0,4	0,2	0,5	0,9	0,9	0,7
Cochin Jews	0,2	0,2	0,3	0,1	0,2	0,2	0,1	0,0	0,3	0,3	0,4	0,1	0,1	0,2	0,1	0,2	0,2	0,2	0,3	0,2	0,2	0,2	0,3
Ethiopian Jews	0,1	0,1	0,1	0,0	0,0	0,0	0,0	0,1	0,4	0,1	0,1	0,0	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,1
Georgian Jews	0,0	0,3	0,3	0,1	0,3	0,2	0,6	0,3	0,3	0,3	0,4	0,3	0,3	0,3	0,2	0,3	0,3	0,2	0,2	0,4	0,4	0,4	0,3
Iranian Jews	0,0	0,3	0,6	0,1	0,5	0,4	0,5	0,3	0,3	0,4	0,3	0,4	0,2	0,4	0,2	0,6	0,4	0,2	0,1	0,3	0,4	0,3	0,3
Libyan Jews	0,0	0,4	0,3	0,2	0,3	0,4	0,5	0,5	0,4	0,4	0,2	0,6	0,2	0,5	0,2	0,4	0,4	0,2	0,4	0,3	0,4	0,4	0,4
Moroccan Jews	0,1	0,6	0,6	0,2	0,7	0,7	1,4	0,5	1,0	1,0	0,5	1,0	0,4	0,9	0,4	1,3	0,8	0,3	0,4	0,7	0,8	1,1	0,9
Tunisian Jews	0,4	0,1	0,1	0,1	0,3	0,1	0,6	0,2	0,4	0,2	0,2	0,4	0,3	0,4	0,2	0,3	0,3	0,3	0,1	0,3	0,3	0,3	0,4
Yemenite Jews	0,4	0,4	0,4	0,1	0,3	0,1	0,2	0,5	1,1	0,2	0,3	0,2	0,2	0,2	0,3	0,3	0,3	0,2	0,2	0,3	0,2	0,3	0,3
Iraqi Jews	0,1	0,3	0,7	0,2	0,5	0,5	1,2	0,5	0,7	0,7	0,6	0,9	0,5	0,9	0,5	0,7	0,8	0,6	0,2	1,0	0,7	0,9	0,7
Jordanian	0,8	0,4	0,6	0,2	0,3	0,4	0,7	0,8	1,9	0,5	0,3	0,5	0,3	0,5	0,2	0,7	0,6	0,2	0,1	0,6	0,5	0,5	0,5
Kalash	0,1	0,4	0,3	0,2	0,4	0,3	0,4	0,2	0,3	0,5	0,7	0,5	0,3	0,5	0,1	0,6	0,5	0,3	0,2	0,4	0,5	0,5	0,4
Lebanese Christian	0,0	0,4	0,5	0,2	0,6	0,6	0,8	0,3	0,6	0,6	0,4	0,6	0,1	0,7	0,4	0,5	0,6	0,4	0,2	0,7	0,6	0,8	0,5
Lebanese Muslim	0,3	0,5	0,5	0,2	0,5	0,6	0,9	0,5	1,0	0,6	0,4	0,6	0,2	0,6	0,4	0,6	0,7	0,1	0,2	0,7	0,7	0,7	0,6
Libyan	1,0	0,3	0,2	0,1	0,4	0,1	0,2	1,2	8,0	0,2	0,2	0,2	0,2	0,3	0,3	0,3	0,2	0,3	0,1	0,1	0,2	0,1	0,3
Moroccan	2,2	0,3	0,2	0,1	0,2	0,3	0,4	0,4	1,0	0,2	0,2	0,3	0,1	0,3	0,3	0,3	0,3	0,3	0,1	0,3	0,3	0,3	0,4
Mozabite	3,8	0,2	0,2	0,1	0,2	0,4	0,5	0,4	0,6	0,1	0,1	0,1	0,3	0,2	0,4	0,2	0,2	0,3	0,1	0,2	0,2	0,3	0,5
Norwegian	0,4	0,3	0,3	0,1	0,4	0,6	0,8	0,3	0,4	0,4	0,3	1,2	1,0	1,0	0,3	1,3	0,9	0,3	0,2	0,6	0,5	0,9	0,7
Punjabi (PJL)	0,1	0,3	0,4	0,4	0,7	0,3	0,5	0,2	0,3	0,7	0,4	0,4	0,3	0,4	0,2	0,4	0,4	0,2	0,4	0,4	0,4	0,4	0,4
Palestinian	0,3	0,4	0,5	0,2	0,5	0,5	0,6	0,9	1,7	0,5	0,3	0,4	0,4	0,5	0,4	0,5	0,5	0,3	0,2	0,6	0,5	0,5	0,5
Macedonian Roma	0,0	0,2	0,4	0,4	0,3	0,4	0,6	0,2	0,3	0,6	0,2	0,4	0,7	0,9	0,2	0,4	0,5	0,0	0,2	0,4	2,4	0,9	0,6
Roma Central Iberia	0,2	0,4	0,4	0,3	0,5	1,4	1,6	0,5	0,6	0,6	0,4	0,6	0,8	0,8	0,3	0,8	0,7	0,9	0,3	0,5	1,8	1,0	1,1
Czech Roma	0,2	0,3	0,4	0,3	0,4	0,5	0,6	0,2	0,6	0,5	0,4	0,3	0,9	1,0	0,2	0,6	0,5	0,3	0,3	0,7	1,9	0,9	0,7
Roma Eastern Iberia	0,1	0,3	0,5	0,3	0,6	1,7	1,8	0,3	0,5	0,8	0,4	0,7	0,7	0,8	0,3	0,8	0,7	1,2	0,3	0,6	1,7	0,9	1,1
Iberian Roma Indeterminate	0,1	0,3	0,4	0,2	0,6	1,7	1,9	0,4	0,6	0,9	0,4	0,6	0,4	0,7	0,3	0,7	0,6	1,0	0,4	0,6	1,5	0,8	1,1
Lithuanian Roma	0,0	0,3	0,4	0,4	0,7	0,4	0,7	0,3	0,7	0,6	0,4	0,5	1,0	1,0	0,2	1,0	0,5	0,0	0,6	0,5	2,4	0,9	0,9
Roma Northern Iberia	0,2	0,3	0,1	0,3	0,5	3,1	3,2	0,5	0,2	0,5	0,2	0,3	0,6	0,5	0,2	0,6	0,3	1,3	0,2	0,7	1,2	0,6	1,2
Romungro from Hungary	0,0	0,7	0,5	0,3	0,9	0,8	1,5	0,3	1,0	0,7	0,5	0,6	1,2	1,2	0,1	0,8	0,8	0,2	0,3	0,6	1,9	1,1	0,8
Romungro from Ukraine	0,2	0,3	0,7	0,4	0,5	0,6	0,7	0,2	0,7	0,5	0,4	0,4	1,4	1,1	0,0	0,5	0,7	0,5	0,2	0,2	1,9	1,3	0,7
Roma Southern Iberia	0,2	0,3	0,3	0,2	0,5	1,5	1,9	0,4	0,5	0,6	0,4	0,6	0,7	0,7	0,2	0,6	0,6	1,0	0,3	0,5	1,8	0,9	1,0
Vlax from Hungary	0,0	0,1	0,1	0,2	0,3	0,6	0,8	0,1	0,3	0,4	0,3	0,4	1,2	0,8	0,1	0,5	0,3	0,5	0,1	0,4	1,5	0,9	0,6
Roma Western Iberia	0,0	0,5	0,2	0,3	0,5	0,8	1,2	0,5	0,7	0,8	0,4	0,6	0,6	0,8	0,1	0,6	0,6	0,2	0,2	0,7	1,4	1,0	1,0
Romanian	0,3	0,4	0,4	0,2	0,3	0,5	1,0	0,4	0,4	0,5	0,4	0,9	1,3	1,4	0,3	1,0	0,7	0,2	0,4	0,4	1,2	1,7	0,7
STU	0,0	0,2	0,3	0,3	0,5	0,3	0,4	0,2	0,3	0,5	0,2	0,3	0,2	0,3	0,1	0,3	0,3	0,1	0,4	0,4	0,3	0,3	0,2
Sardinian	0,2	0,4	0,4	0,1	0,5	0,6	0,9	0,4	0,5	0,5	0,5	0,6	0,4	0,7	0,2	0,6	0,7	0,6	0,2	0,6	0,7	0,7	0,7
Sephardi Jews from Turkey	0,0	0,4	0,2	0,1	0,5	0,6	0,5	0,3	0,3	0,3	0,2	0,4	0,3	0,3	0,2	0,4	0,4	0,1	0,1	0,3	0,3	0,3	0,4
Sindhi	0,2	0,2	0,3	0,2	2,8	0,3	0,4	0,1	0,3	1,9	0,5	0,4	0,2	0,4	0,2	0,3	0,4	0,3	0,3	0,4	0,4	0,4	0,3
Syrian	0,6	0,4	0,2	0,1	0,1	0,2	0,1	0,6	1,7	0,1	0,2	0,1	0,1	0,2	0,2	0,1	0,1	0,3	0,1	0,2	0,2	0,2	0,2
Tuscanian (TSI)	0,3	0,4	0,5	0,1	0,4	0,6	0,8	0,3	0,4	0,5	0,6	0,6	0,3	0,9	0,3	0,7	0,8	0,6	0,2	0,5	0,5	0,7	0,7
Tunisian	1,7	0,2	0,2	0,1	0,1	0,2	0,3	0,4	2,0	0,2	0,1	0,2	0,1	0,3	0,3	0,3	0,3	0,4	0,1	0,2	0,2	0,2	0,3
Turkish	0,2	0,5	0,5	0,2	0,5	0,5	0,7	0,4	0,5	0,6	0,3	0,5	0,5	0,7	0,2	0,7	0,6	0,2	0,2	0,7	0,7	0,7	0,5
Yemeni	0,3	0,5	0,4	0,1	0,4	0,2	0,6	0,5	1,8	0,5	0,2	0,3	0,3	0,3	0,4	0,5	0,4	0,3	0,2	0,4	0,5	0,4	0,4
Yoruba	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0

	Telugu (ITU)	Iranian	Iranian Bandari	Ashkenazi Jews	Cochin Jews	Ethiopian Jews	Georgian Jews	Iranian Jews	Libyan Jews	Moroccan Jews	Tunisian Jews	Yemenite Jews	Iraqi Jews	Jordanian	Kalash	Lebanese Christian	Lebanese Muslim	Libyan	Moroccan	Mozabite	Norwegian	Punjabi (PJL)	Palestinian
Algerian	0,0	0,1	0,1	0,4	0,2	0,1	0,0	0,0	0,0	0,1	0,4	0,4	0,1	0,8	0,1	0,0	0,3	1,0	2,2	3,8	0,4	0,1	0,3
Armenian	0,2	0,3	0,4	0,3	0,2	0,1	0,3	0,3	0,4	0,6	0,1	0,4	0,3	0,4	0,4	0,4	0,5	0,3	0,3	0,2	0,3	0,3	0,4
Assyrian	0,2	0,4	0,6	0,5	0,3	0,1	0,3	0,6	0,3	0,6	0,1	0,4	0,7	0,6	0,3	0,5	0,5	0,2	0,2	0,2	0,3	0,4	0,5
Bangladeshi (BEB)	0,3	0,2	0,2	0,2	0,1	0,0	0,1	0,1	0,2	0,2	0,1	0,1	0,2	0,2	0,2	0,2	0,2	0,1	0,1	0,1	0,1	0,4	0,2
Balochi	0,4	0,5	4,2	0,3	0,2	0,0	0,3	0,5	0,3	0,7	0,3	0,3	0,5	0,3	0,4	0,6	0,5	0,4	0,2	0,2	0,4	0,7	0,5
Basque (France)	0,1	0,2	0,5	0,6	0,2	0,0	0,2	0,4	0,4	0,7	0,1	0,1	0,5	0,4	0,3	0,6	0,6	0,1	0,3	0,4	0,6	0,3	0,5
Basque (Spain)	0,2	0,6	0,9	0,8	0,1	0,0	0,6	0,5	0,5	1,4	0,6	0,2	1,2	0,7	0,4	0,8	0,9	0,2	0,4	0,5	0,8	0,5	0,6
Bedouin A	0,2	0,4	0,4	0,2	0,0	0,1	0,3	0,3	0,5	0,5	0,2	0,5	0,5	0,8	0,2	0,3	0,5	1,2	0,4	0,4	0,3	0,2	0,9
Bedouin B	0,1	0,4	0,7	0,8	0,3	0,4	0,3	0,3	0,4	1,0	0,4	1,1	0,7	1,9	0,3	0,6	1,0	8,0	1,0	0,6	0,4	0,3	1,7
Brahui	0,4	0,4	5,6	0,5	0,3	0,1	0,3	0,4	0,4	1,0	0,2	0,2	0,7	0,5	0,5	0,6	0,6	0,2	0,2	0,1	0,4	0,7	0,5
Burusho	0,3	0,3	0,6	0,3	0,4	0,1	0,4	0,3	0,2	0,5	0,2	0,3	0,6	0,3	0,7	0,4	0,4	0,2	0,2	0,1	0,3	0,4	0,3
Central European (CEU)	0,1	0,4	0,5	0,6	0,1	0,0	0,3	0,4	0,6	1,0	0,4	0,2	0,9	0,5	0,5	0,6	0,6	0,2	0,3	0,1	1,2	0,4	0,4
Croatian	0,2	0,4	0,3	0,3	0,1	0,0	0,3	0,2	0,2	0,4	0,3	0,2	0,5	0,3	0,3	0,1	0,2	0,2	0,1	0,3	1,0	0,3	0,4
Czech	0,2	0,4	0,7	0,8	0,2	0,1	0,3	0,4	0,5	0,9	0,4	0,2	0,9	0,5	0,5	0,7	0,6	0,3	0,3	0,2	1,0	0,4	0,5
Egyptian	0,1	0,2	0,2	0,3	0,1	0,1	0,2	0,2	0,2	0,4	0,2	0,3	0,5	0,2	0,1	0,4	0,4	0,3	0,3	0,4	0,3	0,2	0,4
English	0,2	0,4	0,7	0,6	0,2	0,1	0,3	0,6	0,4	1,3	0,3	0,3	0,7	0,7	0,6	0,5	0,6	0,3	0,3	0,2	1,3	0,4	0,5
French Central	0,2	0,4	0,6	0,6	0,2	0,1	0,3	0,4	0,4	0,8	0,3	0,3	0,8	0,6	0,5	0,6	0,7	0,2	0,3	0,2	0,9	0,4	0,5
French South	0,2	0,2	0,1	0,4	0,2	0,1	0,2	0,2	0,2	0,3	0,3	0,2	0,6	0,2	0,3	0,4	0,1	0,3	0,3	0,3	0,3	0,2	0,3
Gujarati (GIH)	0,4	0,3	0,4	0,2	0,3	0,1	0,2	0,1	0,4	0,4	0,1	0,2	0,2	0,1	0,2	0,2	0,2	0,1	0,1	0,1	0,2	0,4	0,2
Georgian	0,2	0,5	0,7	0,5	0,2	0,1	0,4	0,3	0,3	0,7	0,3	0,3	1,0	0,6	0,4	0,7	0,7	0,1	0,3	0,2	0,6	0,4	0,6
Greek	0,2	0,4	0,6	0,9	0,2	0,0	0,4	0,4	0,4	0,8	0,3	0,2	0,7	0,5	0,5	0,6	0,7	0,2	0,3	0,2	0,5	0,4	0,5
Hungarian	0,1	0,4	0,6	0,9	0,2	0,1	0,4	0,3	0,4	1,1	0,3	0,3	0,9	0,5	0,5	0,8	0,7	0,1	0,3	0,3	0,9	0,4	0,5
Spanish (IBS)	0,1	0,4	0,5	0,7	0,3	0,1	0,3	0,3	0,4	0,9	0,4	0,3	0,7	0,5	0,4	0,5	0,6	0,3	0,4	0,5	0,7	0,4	0,5
Telugu (ITU)	0,7	0,2	0,5	0,1	0,2	0,0	0,2	0,1	0,1	0,3	0,1	0,1	0,2	0,2	0,2	0,2	0,2	0,2	0,1	0,1	0,2	0,4	0,2
Iranian	0,2	0,4	0,6	0,3	0,2	0,1	0,4	0,2	0,5	0,5	0,3	0,4	0,7	0,4	0,3	0,4	0,5	0,2	0,2	0,2	0,3	0,4	0,4
Iranian Bandari	0,5	0,6	7,5	0,6	0,2	0,1	0,3	0,6	0,5	1,0	0,4	0,3	0,9	0,6	0,7	0,8	0,7	0,3	0,2	0,2	0,6	0,6	0,5
Ashkenazi Jews	0,1	0,3	0,6	12,7	0,4	0,1	0,9	0,4	2,1	1,3	1,1	0,7	1,0	0,5	0,3	0,5	0,8	0,0	0,4	0,2	0,4	0,3	0,5
Cochin Jews	0,2	0,2	0,2	0,4	34,0	0,1	0,2	1,0	0,5	3,1	0,5	0,4	3,1	0,2	0,3	0,2	0,1	0,1	0,2	0,1	0,2	0,3	0,1
Ethiopian Jews	0,0	0,1	0,1	0,1	0,1	5,2	0,1	0,0	0,0	0,2	0,1	0,1	0,2	0,1	0,1	0,1	0,0	0,2	0,1	0,0	0,0	0,1	0,3
Georgian Jews	0,2	0,4	0,3	0,9	0,2	0,1	26,5	3,2	0,3	0,7	0,1	0,4	1,1	0,1	0,3	0,3	0,4	0,1	0,1	0,2	0,3	0,2	0,3
Iranian Jews	0,1	0,2	0,6	0,4	1,0	0,0	3,2	16,5	0,3	0,9	0,6	0,1	5,7	0,4	0,4	0,6	0,4	0,2	0,3	0,2	0,4	0,3	0,3
Libyan Jews	0,1	0,5	0,5	2,1	0,5	0,0	0,3	0,3	36,4	4,9	57,5	0,4	0,8	0,3	0,4	0,7	0,5	0,8	0,6	0,8	0,3	0,3	0,5
Moroccan Jews	0,3	0,5	1,0	1,3	3,1	0,2	0,7	0,9	4,9	11,7	4,8	0,6	1,4	0,7	0,5	1,0	1,0	0,6	1,0	0,7	0,8	0,5	0,8
Tunisian Jews	0,1	0,3	0,4	1,1	0,5	0,1	0,1	0,6	57,5	4,8	28,5	0,3	0,7	0,4	0,4	0,2	0,3	0,5	0,5	0,4	0,3	0,2	0,4
Yemenite Jews	0,1	0,4	0,3	0,7	0,4	0,1	0,4	0,1	0,4	0,6	0,3	17,0	0,2	0,5	0,3	0,0	0,4	0,6	0,1	0,1	0,2	0,2	0,3
Iraqi Jews	0,2	0,7	0,9	1,0	3,1	0,2	1,1	5,7	0,8	1,4	0,7	0,2	12,0	0,6	0,5	1,2	0,8	0,0	0,5	0,2	0,7	0,5	0,7
Jordanian	0,2	0,4	0,6	0,5	0,2	0,1	0,1	0,4	0,3	0,7	0,4	0,5	0,6	1,4	0,3	0,7	0,6	0,7	0,3	0,6	0,4	0,3	3,1
Kalash	0,2	0,3	0,7	0,3	0,3	0,1	0,3	0,4	0,4	0,5	0,4	0,3	0,5	0,3	149,5	0,6	0,4	0,3	0,2	0,1	0,4	0,4	0,5
Lebanese Christian	0,2	0,4	0,8	0,5	0,2	0,1	0,3	0,6	0,7	1,0	0,2	0,0	1,2	0,7	0,6	2,8	1,2	0,2	0,3	0,3	0,4	0,4	0,7
Lebanese Muslim	0,2	0,5	0,7	0,8	0,1	0,0	0,4	0,4	0,5	1,0	0,3	0,4	0,8	0,6	0,4	1,2	1,2	0,5	0,4	0,2	0,6	0,4	0,5
Libyan	0,2	0,2	0,3	0,0	0,1	0,2	0,1	0,2	0,8	0,6	0,5	0,6	0,0	0,7	0,3	0,2	0,5	3,5	1,3	1,9	0,2	0,1	0,7
Moroccan	0,1	0,2	0,2	0,4	0,2	0,1	0,1	0,3	0,6	1,0	0,5	0,1	0,5	0,3	0,2	0,3	0,4	1,3	1,5	4,4	0,2	0,2	0,4
Mozabite	0,1	0,2	0,2	0,2	0,1	0,0	0,2	0,2	0,8	0,7	0,4	0,1	0,2	0,6	0,1	0,3	0,2	1,9	4,4	54,6	0,2	0,1	0,4
Norwegian	0,2	0,3	0,6	0,4	0,2	0,0	0,3	0,4	0,3	0,8	0,3	0,2	0,7	0,4	0,4	0,4	0,6	0,2	0,2	0,2	2,9	0,3	0,5
Punjabi (PJL)	0,4	0,4	0,6	0,3	0,3	0,1	0,2	0,3	0,3	0,5	0,2	0,2	0,5	0,3	0,4	0,4	0,4	0,1	0,2	0,1	0,3	4,1	0,3
Palestinian	0,2	0,4	0,5	0,5	0,1	0,3	0,3	0,3	0,5	0,8	0,4	0,3	0,7	3,1	0,5	0,7	0,5	0,7	0,4	0,4	0,5	0,3	29,6
Macedonian Roma	0,2	0,3	0,5	0,3	0,0	0,0	0,2	0,2	0,4	0,4	0,1	0,0	0,4	0,4	0,2	0,5	0,4	0,3	0,3	0,2	0,5	1,9	0,4
Roma Central Iberia	0,3	0,5	0,6	0,5	0,2	0,1	0,3	0,3	0,5	0,8	0,3	0,2	0,6	0,5	0,5	0,7	0,6	0,6	0,4	0,4	0,6	1,5	0,6
Czech Roma	0,3	0,4	0,5	0,6	0,3	0,1	0,4	0,4	0,3	0,5	0,2	0,2	0,7	0,4	0,3	0,6	0,6	0,2	0,3	0,2	0,4	2,1	0,5
Roma Eastern Iberia	0,3	0,4	0,8	0,6	0,3	0,2	0,3	0,4	0,5	1,3	0,2	0,2	0,9	0,6	0,4	0,7	0,6	0,2	0,4	0,2	0,5	1,7	0,5
Iberian Roma Indeterminate	0,3	0,5	0,6	0,8	0,3	0,1	0,1	0,4	0,4	1,0	0,2	0,2	0,6	0,5	0,3	0,5	0,5	0,3	0,3	0,3	0,5	1,5	0,5
Lithuanian Roma	0,2	0,4	0,6	1,2	0,1	0,0	0,3	0,1	0,2	0,8	0,2	0,1	0,9	0,4	0,4	0,4	0,6	0,2	0,3	0,2	0,5	1,8	0,6
Roma Northern Iberia	0,3	0,4	0,3	0,5	0,2	0,0	0,2	0,2	0,3	0,4	0,3	0,1	0,5	0,4	0,1	0,4	0,2	0,0	0,2	0,3	0,6	1,5	0,3
Romungro from Hungary	0,2	0,4	1,4	0,7	0,1	0,0	0,0	0,3	0,5	1,1	0,4	0,4	0,6	0,6	0,4	0,9	0,9	0,1	0,3	0,3	0,6	1,9	0,6
Romungro from Ukraine	0,3	0,5	0,4	0,3	0,2	0,2	0,2	0,4	0,4	0,5	0,1	0,1	0,7	0,4	0,4	0,7	0,7	0,3	0,5	0,1	0,6	1,9	0,6
Roma Southern Iberia	0,2	0,4	0,5	0,5	0,3	0,2	0,2	0,2	0,4	0,9	0,2	0,2	0,7	0,6	0,2	0,5	0,5	0,1	0,3	0,3	0,4	1,6	0,4
Vlax from Hungary	0,3	0,4	0,3	0,4	0,1	0,1	0,2	0,5	0,5	0,2	0,4	0,0	0,6	0,2	0,4	0,6	0,4	0,0	0,3	0,0	0,1	1,5	0,3
Roma Western Iberia	0,3	0,4	0,6	0,6	0,1	0,1	0,3	0,1	0,5	1,1	0,0	0,2	0,5	0,3	0,5	0,7	0,5	0,0	0,4	0,2	0,3	1,6	0,6
Romanian	0,2	0,6	0,6	0,6	0,2	0,0	0,3	0,2	0,5	0,7	0,3	0,4	0,7	0,4	0,4	0,2	0,6	0,1	0,3	0,2	1,0	0,3	0,5
STU	0,6	0,3	0,6	0,3	0,2	0,0	0,1	0,2	0,2	0,5	0,1	0,2	0,5	0,3	0,3	0,5	0,3	0,1	0,2	0,1	0,2	0,4	0,2
Sardinian	0,2	0,4	0,5	0,7	0,1	0,2	0,4	0,4	0,6	0,7	0,4	0,3	0,7	0,5	0,3	0,7	0,6	0,1	0,4	0,4	0,5	0,3	0,5
Sephardi Jews from Turkey	0,1	0,1	0,5	4,4	0,6	0,1	1,1	0,6	2,6	1,7	2,4	0,1	0,7	0,3	0,2	0,5	0,3	0,2	0,4	0,7	0,3	0,1	0,3
Sindhi	0,3	0,3	1,0	0,3	0,2	0,1	0,1	0,2	0,2	0,4	0,3	0,2	0,5	0,4	0,4	0,4	0,4	0,2	0,1	0,1	0,4	0,9	0,4
Syrian	0,2	0,2	0,2	0,2	0,2	0,3	0,1	0,4	0,1	0,4	0,1	0,2	0,4	0,5	0,1	0,2	0,4	1,0	0,5	0,2	0,2	0,1	0,4
Tuscanian (TSI)	0,2	0,4	0,6	0,6	0,3	0,1	0,1	0,5	0,2	0,8	0,4	0,3	0,8	0,5	0,4	0,7	0,5	0,2	0,3	0,3	0,7	0,4	0,5
Tunisian	0,1	0,2	0,3	0,3	0,2	0,1	0,2	0,2	4,5	0,8	7,1	0,2	0,3	0,2	0,1	0,3	0,3	2,2	1,4	2,6	0,2	0,1	0,4
Turkish	0,1	0,5	0,6	0,6	0,2	0,1	0,2	0,6	0,4	0,7	0,4	0,4	0,9	0,6	0,4	0,6	0,6	0,3	0,3	0,2	0,5	0,3	0,5
Yemeni	0,1	0,2	0,3	0,4	0,2	0,1	0,5	0,3	0,3	0,5	0,2	0,8	0,4	0,4	0,1	0,4	0,4	0,9	0,3	0,3	0,2	0,3	0,4
Yoruba	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,1	0,0	0,0	0,0

	Macedonian Roma	Roma Central Iberia	Czech Roma	Roma Eastern Iberia	Iberian Roma Indeterminate	Lithuanian Roma	Roma Northern Iberia	Romungro from Hungary	Romungro from Ukraine	Roma Southern Iberia	Vlax from Hungary	Roma Western Iberia	Romanian	STU	Sardinian	Sephardi Jews from Turkey	Sindhi	Syrian	Tuscanian (TSI)	Tunisian	Turkish	Yemeni	Yoruba
Algerian	0,0	0,2	0,2	0,1	0,1	0,0	0,2	0,0	0,2	0,2	0,0	0,0	0,3	0,0	0,2	0,0	0,2	0,6	0,3	1,7	0,2	0,3	0,1
Armenian	0,2	0,4	0,3	0,3	0,3	0,3	0,3	0,7	0,3	0,3	0,1	0,5	0,4	0,2	0,4	0,4	0,2	0,4	0,4	0,2	0,5	0,5	0,0
Assyrian	0,4	0,4	0,4	0,5	0,4	0,4	0,1	0,5	0,7	0,3	0,1	0,2	0,4	0,3	0,4	0,2	0,3	0,2	0,5	0,2	0,5	0,4	0,0
Bangladeshi (BEB)	0,4	0,3	0,3	0,3	0,2	0,4	0,3	0,3	0,4	0,2	0,2	0,3	0,2	0,3	0,1	0,1	0,2	0,1	0,1	0,1	0,2	0,1	0,0
Balochi	0,3	0,5	0,4	0,6	0,6	0,7	0,5	0,9	0,5	0,5	0,3	0,5	0,3	0,5	0,5	0,5	2,8	0,1	0,4	0,1	0,5	0,4	0,0
Basque (France)	0,4	1,4	0,5	1,7	1,7	0,4	3,1	0,8	0,6	1,5	0,6	0,8	0,5	0,3	0,6	0,6	0,3	0,2	0,6	0,2	0,5	0,2	0,0
Basque (Spain)	0,6	1,6	0,6	1,8	1,9	0,7	3,2	1,5	0,7	1,9	0,8	1,2	1,0	0,4	0,9	0,5	0,4	0,1	0,8	0,3	0,7	0,6	0,0
Bedouin A	0,2	0,5	0,2	0,3	0,4	0,3	0,5	0,3	0,2	0,4	0,1	0,5	0,4	0,2	0,4	0,3	0,1	0,6	0,3	0,4	0,4	0,5	0,0
Bedouin B	0,3	0,6	0,6	0,5	0,6	0,7	0,2	1,0	0,7	0,5	0,3	0,7	0,4	0,3	0,5	0,3	0,3	1,7	0,4	2,0	0,5	1,8	0,0
Brahui	0,6	0,6	0,5	0,8	0,9	0,6	0,5	0,7	0,5	0,6	0,4	0,8	0,5	0,5	0,5	0,3	1,9	0,1	0,5	0,2	0,6	0,5	0,0
Burusho	0,2	0,4	0,4	0,4	0,4	0,4	0,2	0,5	0,4	0,4	0,3	0,4	0,4	0,2	0,5	0,2	0,5	0,2	0,6	0,1	0,3	0,2	0,0
Central European (CEU)	0,4	0,6	0,3	0,7	0,6	0,5	0,3	0,6	0,4	0,6	0,4	0,6	0,9	0,3	0,6	0,4	0,4	0,1	0,6	0,2	0,5	0,3	0,0
Croatian	0,7	0,8	0,9	0,7	0,4	1,0	0,6	1,2	1,4	0,7	1,2	0,6	1,3	0,2	0,4	0,3	0,2	0,1	0,3	0,1	0,5	0,3	0,0
Czech	0,9	0,8	1,0	0,8	0,7	1,0	0,5	1,2	1,1	0,7	0,8	0,8	1,4	0,3	0,7	0,3	0,4	0,2	0,9	0,3	0,7	0,3	0,0
Egyptian	0,2	0,3	0,2	0,3	0,3	0,2	0,2	0,1	0,0	0,2	0,1	0,1	0,3	0,1	0,2	0,2	0,2	0,2	0,3	0,3	0,2	0,4	0,0
English	0,4	0,8	0,6	0,8	0,7	1,0	0,6	0,8	0,5	0,6	0,5	0,6	1,0	0,3	0,6	0,4	0,3	0,1	0,7	0,3	0,7	0,5	0,0
French Central	0,5	0,7	0,5	0,7	0,6	0,5	0,3	0,8	0,7	0,6	0,3	0,6	0,7	0,3	0,7	0,4	0,4	0,1	0,8	0,3	0,6	0,4	0,0
French South	0,0	0,9	0,3	1,2	1,0	0,0	1,3	0,2	0,5	1,0	0,5	0,2	0,2	0,1	0,6	0,1	0,3	0,3	0,6	0,4	0,2	0,3	0,0
Gujarati (GIH)	0,2	0,3	0,3	0,3	0,4	0,6	0,2	0,3	0,2	0,3	0,1	0,2	0,4	0,4	0,2	0,1	0,3	0,1	0,2	0,1	0,2	0,2	0,0
Georgian	0,4	0,5	0,7	0,6	0,6	0,5	0,7	0,6	0,2	0,5	0,4	0,7	0,4	0,4	0,6	0,3	0,4	0,2	0,5	0,2	0,7	0,4	0,0
Greek	2,4	1,8	1,9	1,7	1,5	2,4	1,2	1,9	1,9	1,8	1,5	1,4	1,2	0,3	0,7	0,3	0,4	0,2	0,5	0,2	0,7	0,5	0,0
Hungarian	0,9	1,0	0,9	0,9	0,8	0,9	0,6	1,1	1,3	0,9	0,9	1,0	1,7	0,3	0,7	0,3	0,4	0,2	0,7	0,2	0,7	0,4	0,0
Spanish (IBS)	0,6	1,1	0,7	1,1	1,1	0,9	1,2	0,8	0,7	1,0	0,6	1,0	0,7	0,2	0,7	0,4	0,3	0,2	0,7	0,3	0,5	0,4	0,0
Telugu (ITU)	0,2	0,3	0,3	0,3	0,3	0,2	0,3	0,2	0,3	0,2	0,3	0,3	0,2	0,6	0,2	0,1	0,3	0,2	0,2	0,1	0,1	0,1	0,0
Iranian	0,3	0,5	0,4	0,4	0,5	0,4	0,4	0,4	0,5	0,4	0,4	0,4	0,6	0,3	0,4	0,1	0,3	0,2	0,4	0,2	0,5	0,2	0,0
Iranian Bandari	0,5	0,6	0,5	0,8	0,6	0,6	0,3	1,4	0,4	0,5	0,3	0,6	0,6	0,6	0,5	0,5	1,0	0,2	0,6	0,3	0,6	0,3	0,0
Ashkenazi Jews	0,3	0,5	0,6	0,6	0,8	1,2	0,5	0,7	0,3	0,5	0,4	0,6	0,6	0,3	0,7	4,4	0,3	0,2	0,6	0,3	0,6	0,4	0,0
Cochin Jews	0,0	0,2	0,3	0,3	0,3	0,1	0,2	0,1	0,2	0,3	0,1	0,1	0,2	0,2	0,1	0,6	0,2	0,2	0,3	0,2	0,2	0,2	0,0
Ethiopian Jews	0,0	0,1	0,1	0,2	0,1	0,0	0,0	0,0	0,2	0,2	0,1	0,1	0,0	0,0	0,2	0,1	0,1	0,3	0,1	0,1	0,1	0,1	0,0
Georgian Jews	0,2	0,3	0,4	0,3	0,1	0,3	0,2	0,0	0,2	0,2	0,2	0,3	0,3	0,1	0,4	1,1	0,1	0,1	0,1	0,2	0,2	0,5	0,0
Iranian Jews	0,2	0,3	0,4	0,4	0,4	0,1	0,2	0,3	0,4	0,2	0,5	0,1	0,2	0,2	0,4	0,6	0,2	0,4	0,5	0,2	0,6	0,3	0,0
Libyan Jews	0,4	0,5	0,3	0,5	0,4	0,2	0,3	0,5	0,4	0,4	0,5	0,5	0,5	0,2	0,6	2,6	0,2	0,1	0,2	4,5	0,4	0,3	0,0
Moroccan Jews	0,4	0,8	0,5	1,3	1,0	0,8	0,4	1,1	0,5	0,9	0,2	1,1	0,7	0,5	0,7	1,7	0,4	0,4	0,8	0,8	0,7	0,5	0,0
Tunisian Jews	0,1	0,3	0,2	0,2	0,2	0,2	0,3	0,4	0,1	0,2	0,4	0,0	0,3	0,1	0,4	2,4	0,3	0,1	0,4	7,1	0,4	0,2	0,0
Yemenite Jews	0,0	0,2	0,2	0,2	0,2	0,1	0,1	0,4	0,1	0,2	0,0	0,2	0,4	0,2	0,3	0,1	0,2	0,2	0,3	0,2	0,4	0,8	0,0
Iraqi Jews	0,4	0,6	0,7	0,9	0,6	0,9	0,5	0,6	0,7	0,7	0,6	0,5	0,7	0,5	0,7	0,7	0,5	0,4	0,8	0,3	0,9	0,4	0,0
Jordanian	0,4	0,5	0,4	0,6	0,5	0,4	0,4	0,6	0,4	0,6	0,2	0,3	0,4	0,3	0,5	0,3	0,4	0,5	0,5	0,2	0,6	0,4	0,0
Kalash	0,2	0,5	0,3	0,4	0,3	0,4	0,1	0,4	0,4	0,2	0,4	0,5	0,4	0,3	0,3	0,2	0,4	0,1	0,4	0,1	0,4	0,1	0,0
Lebanese Christian	0,5	0,7	0,6	0,7	0,5	0,4	0,4	0,9	0,7	0,5	0,6	0,7	0,2	0,5	0,7	0,5	0,4	0,2	0,7	0,3	0,6	0,4	0,0
Lebanese Muslim	0,4	0,6	0,6	0,6	0,5	0,6	0,2	0,9	0,7	0,5	0,4	0,5	0,6	0,3	0,6	0,3	0,4	0,4	0,5	0,3	0,6	0,4	0,0
Libyan	0,3	0,6	0,2	0,2	0,3	0,2	0,0	0,1	0,3	0,1	0,0	0,0	0,1	0,1	0,1	0,2	0,2	1,0	0,2	2,2	0,3	0,9	0,0
Moroccan	0,3	0,4	0,3	0,4	0,3	0,3	0,2	0,3	0,5	0,3	0,3	0,4	0,3	0,2	0,4	0,4	0,1	0,5	0,3	1,4	0,3	0,3	0,1
Mozabite	0,2	0,4	0,2	0,2	0,3	0,2	0,3	0,3	0,1	0,3	0,0	0,2	0,2	0,1	0,4	0,7	0,1	0,2	0,3	2,6	0,2	0,3	0,1
Norwegian	0,5	0,6	0,4	0,5	0,5	0,5	0,6	0,6	0,6	0,4	0,1	0,3	1,0	0,2	0,5	0,3	0,4	0,2	0,7	0,2	0,5	0,2	0,0
Punjabi (PJL)	1,9	1,5	2,1	1,7	1,5	1,8	1,5	1,9	1,9	1,6	1,5	1,6	0,3	0,4	0,3	0,1	0,9	0,1	0,4	0,1	0,3	0,3	0,0
Palestinian	0,4	0,6	0,5	0,5	0,5	0,6	0,3	0,6	0,6	0,4	0,3	0,6	0,5	0,2	0,5	0,3	0,4	0,4	0,5	0,4	0,5	0,4	0,0
Macedonian Roma	59,2	81,0	110,3	70,7	73,6	105,9	87,5	90,3	110,5	74,9	85,8	85,0	1,2	0,4	0,2	0,2	0,6	0,1	0,4	0,1	0,6	0,2	0,0
Roma Central Iberia	81,0	66,5	93,6	123,7	130,0	116,8	153,7	69,1	92,3	119,2	68,1	147,5	1,0	0,4	0,5	0,4	0,6	0,3	0,5	0,2	0,8	0,3	0,0
Czech Roma	110,3	93,6	105,5	87,4	87,6	135,0	103,8	115,9	197,1	90,4	122,5	97,5	0,9	0,4	0,5	0,2	0,6	0,2	0,5	0,2	0,7	0,3	0,0
Roma Eastern Iberia	70,7	123,7	87,4	78,6	132,9	111,4	173,0	64,4	92,2	107,8	62,6	123,4	0,8	0,5	0,6	0,3	0,7	0,1	0,6	0,3	0,7	0,3	0,0
Iberian Roma Indeterminate	73,6	130,0	87,6	132,9	56,8	111,0	155,2	67,9	91,5	124,9	67,9	128,1	0,6	0,5	0,5	0,4	0,6	0,3	0,5	0,3	0,9	0,3	0,0
Lithuanian Roma	105,9	116,8	135,0	111,4	111,0	179,5	132,3	105,2	144,6	105,6	102,9	121,5	0,9	0,4	0,4	0,2	0,8	0,0	0,3	0,1	0,7	0,2	0,0
Roma Northern Iberia	87,5	153,7	103,8	173,0	155,2	132,3	142,1	86,2	106,8	130,4	72,0	167,9	0,5	0,2	0,3	0,4	0,4	0,0	0,3	0,1	0,6	0,1	0,0
Romungro from Hungary	90,3	69,1	115,9	64,4	67,9	105,2	86,2	57,4	120,7	64,5	85,3	64,3	1,0	0,4	0,7	0,8	0,6	0,0	0,6	0,3	0,8	0,4	0,0
Romungro from Ukraine	110,5	92,3	197,1	92,2	91,5	144,6	106,8	120,7	164,1	90,2	124,6	98,9	0,4	0,4	0,5	0,1	1,0	0,2	0,6	0,3	0,6	0,4	0,0
Roma Southern Iberia	74,9	119,2	90,4	107,8	124,9	105,6	130,4	64,5	90,2	74,1	65,9	114,1	0,7	0,4	0,5	0,5	0,6	0,1	0,4	0,2	0,8	0,4	0,0
Vlax from Hungary	85,8	68,1	122,5	62,6	67,9	102,9	72,0	85,3	124,6	65,9	38,6	71,1	0,5	0,2	0,6	0,0	0,6	0,0	0,3	0,2	0,5	0,0	0,0
Roma Western Iberia	85,0	147,5	97,5	123,4	128,1	121,5	167,9	64,3	98,9	114,1	71,1	111,5	1,3	0,4	0,6	0,4	0,7	0,0	0,3	0,1	0,7	0,3	0,0
Romanian	1,2	1,0	0,9	0,8	0,6	0,9	0,5	1,0	0,4	0,7	0,5	1,3	10,2	0,3	0,6	0,3	0,3	0,4	0,7	0,1	0,9	0,4	0,0
STU	0,4	0,4	0,4	0,5	0,5	0,4	0,2	0,4	0,4	0,4	0,2	0,4	0,3	3,5	0,2	0,2	0,4	0,2	0,3	0,1	0,3	0,2	0,0
Sardinian	0,2	0,5	0,5	0,6	0,5	0,4	0,3	0,7	0,5	0,5	0,6	0,6	0,6	0,2	8,4	0,4	0,3	0,3	0,6	0,3	0,6	0,4	0,0
Sephardi Jews from Turkey	0,2	0,4	0,2	0,3	0,4	0,2	0,4	0,8	0,1	0,5	0,0	0,4	0,3	0,2	0,4	1,8	0,2	0,2	0,3	0,5	0,3	0,4	0,0
Sindhi	0,6	0,6	0,6	0,7	0,6	0,8	0,4	0,6	1,0	0,6	0,6	0,7	0,3	0,4	0,3	0,2	1,8	0,1	0,4	0,1	0,3	0,2	0,0
Syrian	0,1	0,3	0,2	0,1	0,3	0,0	0,0	0,0	0,2	0,1	0,0	0,0	0,4	0,2	0,3	0,2	0,1	6,5	0,3	0,4	0,2	0,8	0,0
Tuscanian (TSI)	0,4	0,5	0,5	0,6	0,5	0,3	0,3	0,6	0,6	0,4	0,3	0,3	0,7	0,3	0,6	0,3	0,4	0,3	1,6	0,3	0,7	0,3	0,0
Tunisian	0,1	0,2	0,2	0,3	0,3	0,1	0,1	0,3	0,3	0,2	0,2	0,1	0,1	0,1	0,3	0,5	0,1	0,4	0,3	1,3	0,2	0,4	0,1
Turkish	0,6	0,8	0,7	0,7	0,9	0,7	0,6	0,8	0,6	0,8	0,5	0,7	0,9	0,3	0,6	0,3	0,3	0,2	0,7	0,2	1,3	0,3	0,0
Yemeni	0,2	0,3	0,3	0,3	0,3	0,2	0,1	0,4	0,4	0,4	0,0	0,3	0,4	0,2	0,4	0,4	0,2	0,8	0,3	0,4	0,3	0,2	0,0
Yoruba	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,0	0,0	0,4

Supplementary Table 9. Heatmap showing the average pairwise IBD sharing length between populations in the dataset, with lengths expressed in centimorgans (cM).



	Algerian	Armenian	Assyrian	Bangladeshi (BEB)	Balochi	Basque (France)	Basque (Spain)	Bedouin A	Bedouin B	Brahui	Burusho	Central European (CEU)	Croatian	Czech	Egyptian	English	French Central	French South	Gujarati (GIH)	Georgian	Greek	Hungarian	Spanish (IBS)	Telugu (ITU)	Iranian	Iranian Bandari	Ashkenazi Jews	Cochin Jews	Ethiopian Jews	Georgian Jews	Iranian Jews	Libyan Jews	Moroccan Jews	Tunisian Jews	Yemenite Jews	Iraqi Jews	Jordanian	Kalash	Lebanese Christian	Lebanese Muslim	Libyan	Moroccan	Mozabite	Norwegian	Punjabi (PJL)	Palestinian	Macedonian Roma	Roma Central Iberia	Czech Roma	Roma Eastern Iberia	Iberian Roma Indeterminate	Lithuanian Roma	Roma Northern Iberia	Romungro from Hungary	Romungro from Ukraine	Roma Southern Iberia	Vlax from Hungary	Roma Western Iberia	Romanian	STU	Sardinian	Sephardi Jews from Turkey	Sindhi	Syrian	Tuscanian (TSI)	Tunisian	Turkish	Yemeni	Yoruba
Algerian	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Armenian	0,1	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Assyrian	0,0	0,1	1,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Bangladeshi (BEB)	0,0	0,0	0,1	0,1	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Balochi	0,0	0,1	0,1	0,0	1,9	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Basque (France)	0,1	0,1	0,1	0,0	0,1	3,8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Basque (Spain)	0,1	0,1	0,2	0,0	0,1	5,2	3,7	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Bedouin A	0,1	0,1	0,1	0,0	0,1	0,1	0,1	1,9	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Bedouin B	0,2	0,1	0,1	0,1	0,1	0,1	0,2	2,0	15,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Brahui	0,0	0,1	0,1	0,1	3,9	0,1	0,2	0,1	0,2	4,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Burusho	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,2	4,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Central European (CEU)	0,1	0,1	0,1	0,0	0,1	0,2	0,3	0,1	0,1	0,2	0,1	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Croatian	0,0	0,1	0,0	0,0	0,1	0,1	0,2	0,1	0,0	0,1	0,1	0,2	0,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Czech	0,0	0,1	0,1	0,1	0,2	0,2	0,2	0,1	0,1	0,2	0,1	0,2	0,3	0,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Egyptian	0,1	0,1	0,1	0,0	0,1	0,0	0,1	0,1	0,4	0,1	0,1	0,1	0,0	0,1	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
English	0,1	0,2	0,1	0,0	0,2	0,2	0,3	0,1	0,2	0,2	0,1	0,4	0,1	0,3	0,1	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
French Central	0,1	0,1	0,2	0,0	0,1	0,4	0,4	0,1	0,2	0,2	0,1	0,3	0,1	0,3	0,1	0,3	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
French South	0,0	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,0	0,1	0,1	0,1	0,9	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Gujarati (GIH)	0,1	0,1	0,2	0,1	0,2	0,2	0,2	0,1	0,1	0,2	0,1	0,1	0,1	0,2	0,1	0,2	0,2	0,0	0,0	0,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Georgian	0,0	0,1	0,2	0,0	0,1	0,2	0,3	0,1	0,1	0,2	0,1	0,2	0,3	0,3	0,1	0,2	0,2	0,2	0,1	0,2	0,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Greek	0,0	0,1	0,2	0,0	0,1	0,2	0,3	0,1	0,1	0,2	0,1	0,3	0,3	0,5	0,1	0,3	0,3	0,2	0,1	0,2	0,3	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Hungarian	0,1	0,1	0,1	0,0	0,1	0,9	1,1	0,1	0,1	0,1	0,1	0,2	0,1	0,2	0,1	0,2	0,2	0,6	0,1	0,1	0,2	0,2	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Spanish (IBS)	0,0	0,0	0,0	0,1	0,1	0,0	0,0	0,0	0,0	0,1	0,1	0,0	0,1	0,0	0,0	0,0	0,0	0,0	0,1	0,1	0,0	0,0	0,0	0,2	0,1	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Telugu (ITU)	0,0	0,1	0,1	0,0	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Iranian	0,0	0,1	0,2	0,1	1,0	0,1	0,3	0,1	0,2	1,4	0,2	0,2	0,1	0,2	0,1	0,2	0,2	0,0	0,1	0,2	0,2	0,2	0,1	0,0	0,2	1,7	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Iranian Bandari	0,1	0,1	0,1	0,1	0,1	0,2	0,3	0,1	0,2	0,1	0,1	0,2	0,1	0,2	0,1	0,2	0,2	0,1	0,1	0,2	0,2	0,3	0,2	0,0	0,1	0,2	3,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Ashkenazi Jews	0,1	0,1	0,1	0,0	0,1	0,1	0,0	0,0	0,1	0,1	0,1	0,0	0,0	0,1	0,0	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1	5,8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Cochin Jews	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Ethiopian Jews	0,0	0,2	0,2	0,1	0,2	0,2	0,4	0,2	0,3	0,3	0,1	0,3	0,1	0,2	0,1	0,4	0,2	0,1	0,1	0,2	0,3	0,3	0,2	0,1	0,1	0,3	0,4	0,7	0,0	0,1	0,3	1,2	2,7	0,0	0,0	0,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Georgian Jews	0,1	0,0	0,0	0,0	0,1	0,0	0,1	0,1	0,1	0,0	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,3	0,1	0,0	0,0	0,1	11,9	1,1	5,6	0,0	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Iranian Jews	0,1	0,1	0,1	0,0	0,1	0,0	0,1	0,1	0,3	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,0	0,1	0,1	0,0	0,1	0,1	0,2	0,1	0,0	0,1	0,0	0,1	0,2	0,1	3,8	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Libyan Jews	0,0	0,1	0,2	0,1	0,2	0,2	0,4	0,1	0,2	0,2	0,2	0,2	0,2	0,3	0,1	0,2	0,3	0,1	0,1	0,3	0,2	0,3	0,2	0,1	0,2	0,3	0,3	0,6	0,0	0,3	1,5	0,0	0,0	0,0	0,0	2,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Moroccan Jews	0,2	0,1	0,2	0,0	0,1	0,1	0,2	0,2	0,5	0,2	0,1	0,1	0,1	0,1	0,1	0,2	0,2	0,0	0,0	0,2	0,2	0,1	0,2	0,1	0,1	0,2	0,1	0,0	0,0	0,0	0,1	0,1	0,2	0,1	0,1	0,2	0,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Tunisian Jews	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,2	0,1	0,1	0,0	0,2	0,2	0,1	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,2	0,1	0,1	0,0	0,1	0,1	0,1	0,2	0,1	0,1	0,2	0,1	25,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Yemenite Jews	0,1	0,2	0,1	0,0	0,1	0,2	0,3	0,1	0,3	0,2	0,1	0,2	0,1	0,2	0,1	0,2	0,2	0,0	0,1	0,2	0,2	0,2	0,2	0,1	0,1	0,2	0,2	0,0	0,0	0,1	0,1	0,1	0,3	0,1	0,1	0,3	0,2	0,1	0,3	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Iraqi Jews	0,3	0,1	0,1	0,0	0,1	0,0	0,0	0,3	1,9	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,0	0,0	0,1	0,0	0,1	0,0	0,1	0,1	0,0	0,0	0,1	0,0	0,0	0,2	0,2	0,1	0,2	0,0	0,2	0,1	0,1	0,1	0,8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Jordanian	0,6	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,3	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,0	0,0	0,0	0,1	0,2	0,3	0,1	0,0	0,1	0,1	0,0	0,1	0,1	0,3	0,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Kalash	0,9	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,0	0,0	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,0	0,1	0,0	0,1	0,0	0,0	0,0	0,1	0,2	0,2	0,1	0,0	0,0	0,1	0,0	0,1	0,1	0,5	1,1	10,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Lebanese Christian	0,1	0,1	0,1	0,0	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,3	0,2	0,3	0,1	0,3	0,2	0,1	0,0	0,1	0,2	0,2	0,2	0,0	0,1	0,2	0,1	0,1	0,0	0,1	0,1	0,1	0,2	0,1	0,1	0,2	0,1	0,1	0,1	0,2	0,0	0,1	0,0	0,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Lebanese Muslim	0,0	0,1	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,1	0,0	0,0	0,1	0,1	0,1	0,0	0,1	0,2	0,1	0,1	0,1	0,1	0,0	0,0	0,0	0,1	0,9	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Libyan	0,1	0,1	0,1	0,0	0,1	0,1	0,2	0,2	0,5	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,1	0,1	0,2	0,2	0,2	0,1	0,0	0,1	0,2	0,1	0,0	0,1	0,1	0,1	0,1	0,2	0,1	0,1	0,2	0,7	0,1	0,2	0,1	0,2	0,1	0,1	0,1	0,0	3,8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Moroccan	0,1	0,1	0,1	0,1	0,1	0,4	0,5	0,1	0,2	0,2	0,1	0,2	0,2	0,2	0,1	0,2	0,2	0,2	0,1	0,2	0,4	0,3	0,3	0,1	0,1	0,2	0,1	0,1	0,0	0,1	0,1	0,1	0,3	0,1	0,1	0,2	0,1	0,1	0,2	0,2	0,2	0,1	0,1	0,2	0,4	0,2	17,2	11,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Mozabite	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,2	0,2	0,1	0,1	0,2	0,3	0,1	0,2	0,2	0,1	0,1	0,2	0,4	0,3	0,2	0,1	0,1	0,1	0,2	0,1	0,0	0,1	0,1	0,1	0,2	0,1	0,0	0,2	0,1	0,1	0,2	0,2	0,0	0,1	0,1	0,1	0,6	0,1	23,4	19,7	18,6	18,2	18,3	27,3	0,0	0,0	0,0	0,0	0,0	0,0	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Norwegian	0,0	0,1	0,1	0,1	0,2	0,4	0,5	0,1	0,2	0,2	0,1	0,2	0,2	0,2	0,1	0,2	0,2	0,3	0,1	0,2	0,4	0,3	0,3	0,1	0,1	0,2	0,2	0,1	0,0	0,1	0,1	0,1	0,4	0,1	0,1	0,3	0,2	0,1	0,2	0,2	0,1	0,1	0,1	0,1	0,5	0,1	15,3	22,6	0,0	12,8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Punjabi (PJL)	0,0	0,1	0,1	0,1	0,2	0,4	0,5	0,1	0,2	0,2	0,1	0,2	0,1	0,2	0,1	0,2	0,2	0,2	0,1	0,2	0,4	0,2	0,3	0,1	0,2	0,2	0,2	0,1	0,0	0,0	0,1	0,1	0,3	0,1	0,1	0,2	0,1	0,1	0,2	0,2	0,1	0,1	0,1	0,1	0,4	0,1	16,0	23,6	0,0	23,7	10,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Palestinian	0,1	0,1	0,0	0,1	0,1	0,8	0,7	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,3	0,1	0,2	0,3	0,2	0,3	0,1	0,1	0,1	0,1	0,1	0,0	0,0	0,0	0,1	0,1	0,1	0,0	0,1	0,1	0,0	0,1	0,1	0,0	0,1	0,1	0,1	0,4	0,1	18,4	27,5	21,3	29,6	27,7	26,2	19,1	0,0	0,0	0,0	0,0	0,0	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Macedonian Roma	0,1	0,1	0,2	0,1	0,1	0,2	0,2	0,1	0,2	0,1	0,1	0,1	0,4	0,3	0,0	0,2	0,2	0,1	0,1	0,1	0,5	0,3	0,2	0,1	0,1	0,1	0,1	0,1	0,0	0,0	0,1	0,1	0,2	0,0	0,0	0,2	0,1	0,1	0,2	0,2	0,1	0,1	0,0	0,2	0,5	0,2	23,9	19,3	36,7	19,2	19,1	28,9	22,0	22,9	18,9	0,0	0,0	0,0	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Roma Central Iberia	0,0	0,1	0,1	0,1	0,1	0,4	0,5	0,1	0,2	0,2	0,1	0,2	0,2	0,2	0,1	0,2	0,2	0,3	0,1	0,1	0,4	0,2	0,2	0,1	0,1	0,2	0,1	0,1	0,0	0,1	0,1	0,1	0,2	0,1	0,1	0,2	0,2	0,1	0,1	0,2	0,0	0,1	0,1	0,1	0,4	0,1	16,0	22,2	18,7	20,3	23,0	21,4	24,4	13,9	18,6	12,5	0,0	0,0	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Czech Roma	0,1	0,1	0,1	0,0	0,1	0,1	0,3	0,1	0,1	0,1	0,1	0,2	0,3	0,4	0,1	0,3	0,2	0,1	0,1	0,1	0,3	0,5	0,2	0,1	0,2	0,2	0,2	0,1	0,0	0,1	0,0	0,1	0,2	0,1	0,1	0,2	0,1	0,1	0,1	0,2	0,0	0,1	0,1	0,3	0,1	0,1	0,3	0,3	0,0	0,2	0,2	0,2	0,0	0,0	0,0	0,0	0,0	0,0	1,9	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Roma Eastern Iberia	0,1	0,1	0,1	0,0	0,1	0,2	0,3	0,1	0,1	0,1	0,1	0,2	0,1	0,2	0,1	0,2	0,2	0,2	0,0	0,2	0,2	0,2	0,2	0,0	0,1	0,2	0,2	0,0	0,1	0,1	0,1	0,1	0,2	0,1	0,1	0,2	0,1	0,1	0,2	0,2	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,2	0,1	0,1	0,2	0,2	0,1	0,1	0,2	0,2	0,0	1,7	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Iberian Roma Indeterminate	0,0	0,1	0,1	0,1	0,6	0,1	0,1	0,0	0,1	0,5	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,3	0,1	0,0	0,0	0,0	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,0	0,0	0,0	0,1	0,2	0,1	0,2	0,2	0,2	0,2	0,2	0,2	0,1	0,2	0,3	0,2	0,1	0,2	0,1	0,0	0,1	0,0	0,4	0,0	0,0	0,0	0,0	0,0	0,0
Lithuanian Roma	0,2	0,1	0,1	0,0	0,0	0,1	0,0	0,2	0,5	0,0	0,0	0,0	0,0	0,1	0,0	0,0	0,0	0,1	0,0	0,1	0,0	0,0	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,0	0,1	0,0	0,1	0,0	0,1	0,1	0,2	0,0	0,1	0,1	0,2	0,1	0,1	0,1	0,0	0,1	0,0	0,1	0,1	0,0	0,1	0,0	0,0	0,0	0,1	0,0	0,0	0,0	0,1	0,0	0,1	0,1	0,0	1,2	0,0	0,0	0,0	0,0	0,0
Roma Northern Iberia	0,1	0,1	0,1	0,0	0,1	0,2	0,2	0,1	0,1	0,1	0,2	0,2	0,1	0,2	0,1	0,2	0,2	0,1	0,1	0,1	0,2	0,2	0,2	0,0	0,1	0,2	0,2	0,1	0,0	0,0	0,1	0,1	0,2	0,1	0,1	0,2	0,1	0,1	0,2	0,2	0,1	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,1	0,1	0,2	0,2	0,1	0,1	0,1	0,2	0,1	0,2	0,1	0,1	0,1	0,3	0,0	0,0	0,0	0,0
Romungro from Hungary	0,4	0,1	0,1	0,0	0,0	0,1	0,1	0,1	0,5	0,0	0,0	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,0	0,0	0,1	0,1	0,1	0,0	0,1	0,1	1,0	0,2	1,4	0,1	0,1	0,1	0,0	0,1	0,1	0,5	0,4	0,7	0,1	0,0	0,1	0,0	0,1	0,0	0,1	0,1	0,0	0,0	0,1	0,1	0,1	0,0	0,0	0,0	0,0	0,1	0,1	0,0	0,1	0,1	0,3	0,0	0,0	0,0
Romungro from Ukraine	0,1	0,1	0,1	0,0	0,1	0,1	0,2	0,1	0,2	0,2	0,1	0,1	0,1	0,2	0,1	0,2	0,2	0,1	0,1	0,2	0,2	0,2	0,1	0,0	0,1	0,2	0,2	0,1	0,0	0,1	0,2	0,1	0,2	0,1	0,1	0,3	0,2	0,1	0,2	0,2	0,1	0,1	0,1	0,2	0,1	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,1	0,2	0,2	0,1	0,2	0,1	0,1	0,1	0,2	0,1	0,3	0,0	0,0
Roma Southern Iberia	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,5	0,2	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,0	0,2	0,1	0,1	0,0	0,1	0,1	0,2	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,0
Vlax from Hungary	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,1
Roma Western Iberia	0,0	0,1	0,1	0,0	0,1	3,4	2,3	0,1	0,1	0,1	0,1	0,2	0,1	0,2	0,1	0,1	0,3	1,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Romanian	0,0	0,1	0,1	0,0	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,0	4,8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
STU	0,0	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,0	0,0	0,1	0,1	0,1	0,1	0,0	0,1	0,2	0,1	0,2	0,0	0,8	3,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Sardinian	0,0	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,2	0,4	0,2	0,0	0,1	0,1	7,1	0,0	0,0	0,0	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Sephardi Jews from Turkey	0,0	0,1	0,2	0,1	0,2	0,2	0,2	0,1	0,2	0,2	0,1	0,2	0,0	0,2	0,1	0,1	0,2	0,1	0,0	0,2	0,2	0,2	0,1	0,0	0,1	0,2	0,1	0,0	0,0	0,1	0,2	0,2	0,3	0,0	0,0	0,3	0,2	0,2	0,7	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Sindhi	0,0	0,1	0,1	0,1	0,1	0,1	0,2	0,0	0,1	0,2	0,1	0,1	0,2	0,2	0,1	0,1	0,1	0,0	0,1	0,1	0,5	0,2	0,1	0,1	0,1	0,2	0,1	0,0	0,0	0,1	0,1	0,1	0,1	0,0	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,5	0,1	12,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Syrian	0,0	0,1	0,1	0,1	0,2	0,1	0,2	0,1	0,2	0,2	0,1	0,2	0,3	0,3	0,1	0,3	0,1	0,0	0,2	0,2	0,6	0,2	0,2	0,1	0,1	0,2	0,3	0,0	0,0	0,1	0,0	0,1	0,2	0,1	0,0	0,2	0,1	0,1	0,1	0,2	0,0	0,1	0,1	0,1	0,5	0,2	22,4	22,8	0,0	22,5	22,7	21,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Tuscanian (TSI)	0,0	0,2	0,2	0,1	0,2	0,2	0,4	0,1	0,2	0,2	0,1	0,2	0,3	0,3	0,0	0,2	0,2	0,1	0,1	0,1	0,5	0,3	0,2	0,1	0,1	0,3	0,2	0,0	0,0	0,0	0,1	0,1	0,3	0,1	0,1	0,2	0,2	0,1	0,2	0,2	0,0	0,1	0,1	0,2	0,5	0,2	19,4	14,8	23,1	13,4	14,5	21,6	17,9	9,6	0,0	0,0	0,0	0,0	0,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Tunisian	0,0	0,0	0,0	0,1	0,1	0,1	0,2	0,0	0,1	0,1	0,1	0,1	0,3	0,2	0,0	0,1	0,1	0,1	0,0	0,1	0,3	0,2	0,1	0,1	0,1	0,1	0,1	0,0	0,0	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,0	0,0	0,4	0,1	18,2	14,7	24,4	13,5	14,5	21,2	15,1	17,7	24,5	14,4	7,5	0,0	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Turkish	0,0	0,1	0,1	0,1	0,1	0,2	0,3	0,1	0,2	0,2	0,1	0,2	0,2	0,2	0,0	0,2	0,2	0,1	0,1	0,2	0,3	0,3	0,2	0,1	0,1	0,2	0,2	0,0	0,0	0,1	0,0	0,1	0,3	0,0	0,1	0,2	0,1	0,1	0,2	0,1	0,0	0,1	0,1	0,1	0,4	0,2	18,5	26,2	20,4	23,2	24,2	24,4	29,8	13,8	21,0	21,9	15,8	16,2	0,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Yemeni	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,2	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,2	0,1	0,2	0,1	0,1	0,0	0,0	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,0	0,0	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,8	0,1	0,1	0,1	0,0	0,0	0,0	0,0	0,0	0,0
Yoruba	0,0	0,1	0,1	0,0	0,1	0,2	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,0	0,0	0,1	0,1	0,1	0,1	0,0	0,0	0,1	1,1	0,2	0,0	0,3	0,2	0,7	0,4	0,6	0,0	0,2	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,0	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,2	0,0	0,1	0,0	0,1	0,1	0,0	0,1	0,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0

Supplementary Table 10. Heatmap showing the average pairwise number of IBD segments shared between populations in the dataset.

Population	Generation	Ne	CI 95 % (Lower)	CI 95% (Higher)
Roma Central Iberia	14	805	579	1030
Roma Eastern Iberia	13	557	463	655
Roma Northern Iberia	8	266	200	375
Roma Southern Iberia	14	671	564	759
Roma Western Iberia	9	132	113	220
Iberian Roma Indeterminate	15	535	386	746
Macedonian Roma	24	738	469	1060
Lithuanian Roma	7	126	101	199
Czech Roma	16	460	403	532
Romungro Hungary	9	744	431	1280
Romungro Ukraine	1	82,3	48,9	160
Roma Vlax Hungary	29	765	498	2630
Czech	15-16	2420	2200	2680
IBS	50	106000	79400	139000
PJL	31	17700	14900	20500

Supplementary Table 11. Lowest Ne value for each Roma population and three reference populations.

Population	Category	Avg ROH number	Avg ROH length (MB)	Population	Category	Avg ROH number	Avg ROH length (MB)
Algerian	0-1	71,5	47,9	JewIranian	0-1	110,6	75,6
Algerian	1-2.5	12,0	16,1	JewIranian	1-2.5	29,4	42,2
Algerian	2.5-5	6,5	22,9	JewIranian	2.5-5	8,3	29,7
Algerian	>5	7,0	55,1	JewIranian	>5	5,5	41,8
Armenian	0-1	97,4	65,9	JewLibyan	0-1	102,1	69,5
Armenian	1-2.5	15,0	18,8	JewLibyan	1-2.5	23,3	32,8
Armenian	2.5-5	2,3	7,8	JewLibyan	2.5-5	5,3	17,8
Armenian	>5	1,7	9,9	JewLibyan	>5	5,2	42,2
Assyrian	0-1	102,5	69,7	JewMoroccan	0-1	106,3	72,2
Assyrian	1-2.5	18,7	24,9	JewMoroccan	1-2.5	17,2	23,0
Assyrian	2.5-5	3,7	13,0	JewMoroccan	2.5-5	1,6	5,6
Assyrian	>5	2,0	16,8	JewMoroccan	>5	1,0	7,6
BEB	0-1	96,9	65,4	JewTunisian	0-1	100,7	68,9
BEB	1-2.5	11,3	14,1	JewTunisian	1-2.5	29,4	42,5
BEB	2.5-5	3,5	12,7	JewTunisian	2.5-5	16,0	57,4
BEB	>5	3,4	26,9	JewTunisian	>5	14,3	124,9
Balochi	0-1	98,9	67,3	JewYemenite	0-1	110,5	75,1
Balochi	1-2.5	33,8	50,7	JewYemenite	1-2.5	27,1	37,7
Balochi	2.5-5	15,9	57,3	JewYemenite	2.5-5	4,5	14,6
Balochi	>5	9,7	76,7	JewYemenite	>5	5,2	36,3
BasqueFrench	0-1	127,3	86,5	Jewiraqi	0-1	99,7	66,9
BasqueFrench	1-2.5	26,4	36,0	Jewiraqi	1-2.5	20,5	28,9
BasqueFrench	2.5-5	2,2	7,0	Jewiraqi	2.5-5	4,7	16,6
BasqueFrench	>5	1,3	10,9	Jewiraqi	>5	2,6	16,7
BasqueSpanish	0-1	141,0	96,1	Jordanian	0-1	86,3	58,7
BasqueSpanish	1-2.5	27,1	37,5	Jordanian	1-2.5	22,9	31,3
BasqueSpanish	2.5-5	3,3	10,3	Jordanian	2.5-5	11,3	40,8
BasqueSpanish	>5	1,7	10,4	Jordanian	>5	9,4	74,7
BedouinA	0-1	90,1	61,1	Kalash	0-1	115,3	79,5
BedouinA	1-2.5	22,3	32,2	Kalash	1-2.5	42,2	62,8
BedouinA	2.5-5	9,3	33,5	Kalash	2.5-5	11,3	38,0
BedouinA	>5	7,9	63,1	Kalash	>5	2,9	21,7
BedouinB	0-1	116,0	79,4	LebaneseChristian	0-1	100,8	68,0
BedouinB	1-2.5	40,5	59,5	LebaneseChristian	1-2.5	12,7	15,7
BedouinB	2.5-5	10,9	37,9	LebaneseChristian	2.5-5	2,0	6,1
BedouinB	>5	5,9	47,4	LebaneseChristian	>5	1,0	5,5
Brahui	0-1	101,4	69,2	LebaneseMuslim	0-1	88,9	60,0
Brahui	1-2.5	23,7	33,5	LebaneseMuslim	1-2.5	17,0	23,5
Brahui	2.5-5	6,8	24,0	LebaneseMuslim	2.5-5	8,2	28,8
Brahui	>5	6,8	50,9	LebaneseMuslim	>5	4,4	33,0
Burusho	0-1	95,5	64,4	Libyan	0-1	77,8	53,7
Burusho	1-2.5	14,1	19,1	Libyan	1-2.5	18,4	24,7
Burusho	2.5-5	3,5	12,1	Libyan	2.5-5	4,6	16,0
Burusho	>5	1,9	15,3	Libyan	>5	4,0	28,2
CEU	0-1	120,4	81,4	LithuanianRoma	0-1	104,4	72,6
CEU	1-2.5	16,6	21,2	LithuanianRoma	1-2.5	40,8	59,6
CEU	2.5-5	1,3	3,7	LithuanianRoma	2.5-5	9,6	33,9
CEU	>5	1,0	8,6	LithuanianRoma	>5	6,3	44,2
Croatian	0-1	111,8	75,2	MacedonianRoma	0-1	99,0	67,4
Croatian	1-2.5	15,8	20,6	MacedonianRoma	1-2.5	28,8	41,3
Croatian	2.5-5	3,3	11,5	MacedonianRoma	2.5-5	5,3	17,1
Croatian	>5	1,3	7,9	MacedonianRoma	>5	1,5	9,9
Czech	0-1	120,3	81,4	Moroccan	0-1	75,6	50,6
Czech	1-2.5	16,9	21,0	Moroccan	1-2.5	16,0	21,8
Czech	2.5-5	2,0	5,8	Moroccan	2.5-5	5,2	18,4
Czech	>5	1,5	9,8	Moroccan	>5	3,8	29,7
CzechRoma	0-1	105,5	72,3	Mozabite	0-1	87,3	59,6
CzechRoma	1-2.5	36,0	53,3	Mozabite	1-2.5	23,1	33,5
CzechRoma	2.5-5	11,8	41,4	Mozabite	2.5-5	5,4	18,0
CzechRoma	>5	6,1	46,4	Mozabite	>5	1,7	11,1
Egyptian	0-1	79,4	53,2	Norwegian	0-1	114,5	77,5
Egyptian	1-2.5	10,2	12,8	Norwegian	1-2.5	19,2	25,0
Egyptian	2.5-5	2,4	8,3	Norwegian	2.5-5	1,7	6,3
Egyptian	>5	1,5	9,2	Norwegian	>5	1,0	9,8
English	0-1	125,4	85,2	PJL	0-1	94,5	63,9
English	1-2.5	17,8	22,2	PJL	1-2.5	19,9	27,7
English	2.5-5	1,0	2,9	PJL	2.5-5	6,6	22,9
FrenchCentral	0-1	117,5	79,0	PJL	>5	6,7	53,1
FrenchCentral	1-2.5	16,1	20,2	Palestinian	0-1	94,2	63,7
FrenchCentral	2.5-5	2,2	7,8	Palestinian	1-2.5	16,5	22,6
FrenchCentral	>5	3,0	29,7	Palestinian	2.5-5	6,8	24,5
FrenchSouth	0-1	124,7	84,7	Palestinian	>5	5,4	41,4
FrenchSouth	1-2.5	24,3	32,4	RomaVlaxHungary	0-1	102,4	69,2
FrenchSouth	2.5-5	2,6	9,2	RomaVlaxHungary	1-2.5	30,2	43,6
FrenchSouth	>5	1,5	18,1	RomaVlaxHungary	2.5-5	7,8	26,4
GIH	0-1	104,0	70,3	RomaVlaxHungary	>5	2,8	24,9
GIH	1-2.5	19,7	26,1	Romanian	0-1	110,3	74,5
GIH	2.5-5	2,5	7,8	Romanian	1-2.5	19,0	25,2
GIH	>5	1,4	10,3	Romanian	2.5-5	3,0	11,1
Georgian	0-1	114,3	77,6	Romanian	>5	2,0	14,2
Georgian	1-2.5	18,4	22,9	RomungroHungary	0-1	103,6	69,6
Georgian	2.5-5	1,0	3,9	RomungroHungary	1-2.5	26,4	36,8
Georgian	>5	1,0	8,0	RomungroHungary	2.5-5	4,0	13,7
Greek	0-1	111,7	75,7	RomungroHungary	>5	2,0	14,9
Greek	1-2.5	14,8	18,7	RomungroUkraine	0-1	104,5	72,1
Greek	2.5-5	3,0	10,2	RomungroUkraine	1-2.5	35,0	51,1
Greek	>5	1,3	12,0	RomungroUkraine	2.5-5	9,8	33,6
Hungarian	0-1	119,0	80,3	RomungroUkraine	>5	5,0	36,7
Hungarian	1-2.5	14,9	18,5	STU	0-1	100,1	67,6
Hungarian	2.5-5	1,4	4,3	STU	1-2.5	21,0	30,3
Hungarian	>5	1,0	8,7	STU	2.5-5	8,0	28,4
IBS	0-1	115,8	78,0	STU	>5	3,9	27,8
IBS	1-2.5	17,9	22,7	Sardinian	0-1	129,2	88,0
IBS	2.5-5	2,4	8,3	Sardinian	1-2.5	24,5	32,1
IBS	>5	2,0	13,9	Sardinian	2.5-5	3,5	12,2
ITU	0-1	103,1	69,5	Sardinian	>5	1,8	12,5
ITU	1-2.5	16,4	22,4	SephardiJewTurkey	0-1	101,4	68,0
ITU	2.5-5	6,0	20,6	SephardiJewTurkey	1-2.5	14,3	17,7
ITU	>5	15,7	137,3	SephardiJewTurkey	2.5-5	3,2	10,8
IberianRoma	0-1	98,9	67,5	SephardiJewTurkey	>5	4,0	34,6
IberianRoma	1-2.5	27,6	39,6	Sindhi	0-1	97,4	65,7
IberianRoma	2.5-5	10,1	35,4	Sindhi	1-2.5	22,8	32,8
IberianRoma	>5	5,5	42,3	Sindhi	2.5-5	10,2	36,0
Iranian	0-1	95,5	64,4	Sindhi	>5	8,0	64,0
Iranian	1-2.5	14,5	19,3	Syrian	0-1	87,1	59,4
Iranian	2.5-5	4,5	16,1	Syrian	1-2.5	25,0	36,5
Iranian	>5	4,0	30,7	Syrian	2.5-5	16,7	61,3
IranianBandari	0-1	90,1	61,5	Syrian	>5	9,0	74,7
IranianBandari	1-2.5	15,8	21,7	TSI	0-1	110,3	74,7
IranianBandari	2.5-5	5,4	18,5	TSI	1-2.5	15,2	19,0
IranianBandari	>5	3,2	24,6	TSI	2.5-5	1,1	3,3
JewAshkenazi	0-1	104,3	70,8	TSI	>5	1,8	15,0
JewAshkenazi	1-2.5	17,9	24,3	Tunisian	0-1	78,2	53,4
JewAshkenazi	2.5-5	1,3	3,9	Tunisian	1-2.5	20,6	29,3
JewAshkenazi	>5	1,0	7,4	Tunisian	2.5-5	9,4	32,4
JewCochin	0-1	87,8	59,6	Tunisian	>5	7,9	68,3
JewCochin	1-2.5	17,8	25,1	Turkish	0-1	103,2	69,5
JewCochin	2.5-5	10,0	35,4	Turkish	1-2.5	17,0	22,8
JewCochin	>5	2,7	17,0	Turkish	2.5-5	6,5	23,0
JewEthiopian	0-1	49,6	33,7	Turkish	>5	6,5	50,3
JewEthiopian	1-2.5	11,9	15,6	Yemeni	0-1	77,7	53,4
JewEthiopian	2.5-5	1,0	3,7	Yemeni	1-2.5	13,3	18,9
JewEthiopian	>5	1,0	5,5	Yemeni	2.5-5	6,0	21,3
JewGeorgian	0-1	111,9	75,6	Yemeni	>5	7,0	50,2
JewGeorgian	1-2.5	24,0	34,2	Yoruba	0-1	42,0	28,6
JewGeorgian	2.5-5	7,0	24,6	Yoruba	1-2.5	7,3	9,6

Supplementary Table 12. General ROH statistics.

Europe	N	Source	North Africa	N	Source	Middle-East	N	Source
Ashkenazi Jews	7	Lazaridis et al. 2014	Algerian	4	Patterson et al. (2012)	Bedouin A	25	Lazaridis et al. 2014
Central European from Utah (CEU)	25	Auton et al. 2015	Egyptian	18	Lazaridis et al. 2014	Bedouin B	19	Lazaridis et al. 2014
Croatian	10	Lazaridis et al. 2014	Libyan	5	Lazaridis et al. 2014	Iranian	25	Lazaridis et al. 2014
Czech	25	Newly Genotyped	Libyan Jews	9	Lazaridis et al. 2014	Iranian Bandari	8	Lazaridis et al. 2014
English	10	Lazaridis et al. 2014	Moroccan	25	Lazaridis et al. 2014 Lucas-Sánchez et al. 2023	Iranian Jews	9	Lazaridis et al. 2014
Basque from France	20	Lazaridis et al. 2014	Moroccan Jews	6	Lazaridis et al. 2014	Iraqi Jews	6	Lazaridis et al. 2014
Basque from Spain	9	Lazaridis et al. 2014	Mozabite	21	Patterson et al. (2012)	Jordanian	9	Lazaridis et al. 2014
French Central	25	Lazaridis et al. 2014	Tunisian	25	Lazaridis et al. 2014 Lucas-Sánchez et al. 2023	Lebanese Christians	9	Lazaridis et al. 2014
French South	7	Lazaridis et al. 2014	Tunisian Jews	7	Lazaridis et al. 2014	Lebanese Muslims	11	Lazaridis et al. 2014
Greek	20	Lazaridis et al. 2014				Palestinian	25	Lazaridis et al. 2014
Hungarian	20	Lazaridis et al. 2014				Syrian	8	Lazaridis et al. 2014
Spanish	107	Lazaridis et al. 2014				Turkish	25	Lazaridis et al. 2014
Norwegian	11	Lazaridis et al. 2014				Turkish Sephardi Jews	8	Lazaridis et al. 2014
Roma Central Iberia	19	Font et al. 2019 Font et al. 2021				Yemenite	6	Lazaridis et al. 2014
Roma Eastern Iberia	21	Font et al. 2019 Font et al. 2021				Yemenite Jews	8	Lazaridis et al. 2014
Roma Western Iberia	5	Font et al. 2019Font et al. 2021						
Roma Northern Iberia	9	Font et al. 2019Font et al. 2021						
Roma Southern Iberia	29	Font et al. 2019 Font et al. 2021						
Roma Indeterminate	12	Font et al. 2019 Font et al. 2021						
Czech Roma	21	Newly Genotyped						
Macedonian Roma	10	Bianco et al. 2020						
Lithuanian Roma	5	Bianco et al. 2020						
Roma Vlax Hungary	5	Bianco et al. 2020						
Roma Romungro Hungary	5	Bianco et al. 2020						
Roma Romungro Ukraine	4	Bianco et al. 2020						
Romanian	10	Lazaridis et al. 2014						
Sardinian	25	Lazaridis et al. 2014						
Tuscanian (TSI)	25	Auton et al. 2015						
South Asia	**N**	**Source**	**Subsaharan Africa**	**N**	**Source**	**Caucasus**	**N**	**Source**
Balochi	20	Lazaridis et al. 2014	Ethiopian Jews	7	Lazaridis et al. 2014	Armenian	10	Lazaridis et al. 2014
Bengali from Bangladesh (BEB)	25	Auton et al. 2015	Yoruba (YRB)	25	Auton et al. 2015	Assyrian	11	Lazaridis et al. 2014
Brahui	21	Lazaridis et al. 2014				Georgian	10	Lazaridis et al. 2014
Burusho	23	Lazaridis et al. 2014				Georgian Jews	7	Lazaridis et al. 2014
Cochin Jews	5	Lazaridis et al. 2014						
Gujarati Indians from Houston (GIH)	25	Auton et al. 2015						
Indian Telugu from UK (ITU)	25	Auton et al. 2015						
Kalash	15	Lazaridis et al. 2014						
Punjabi from Lahore (PJL)	95	Auton et al. 2015						
Sindhi	18	Lazaridis et al. 2014						
Sri Lankan Telugu from UK (STU)	25	Auton et al. 2015						

Supplementary Table 13. Sample size of the dataset populations after quality control.

Supplementary Table 14. Two-way local ancestry inference with IBS designated as the target population.

6.2 Demographic insights into paternal genetic diversity and regional substructure in the Spanish Roma

Authors

Giacomo Francesco Ena, Aaron Giménez, Annabel Carballo-Mesa, Marcos Araújo Castro e Silva & David Comas.

Demographic insights into paternal genetic diversity and regional substructure in the Spanish Roma

*Giacomo F. Ena[1], Aaron Giménez[2], Annabel Carballo-Mesa[3], Marcos Araújo Castro e Silva[1], David Comas[1, *]*

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

[2] Facultat de Sociologia, Universitat Autònoma de Barcelona, Barcelona, Spain

[3] Facultat de Geografia i Història, Universitat de Barcelona, Barcelona, Spain

**Corresponding author*

david.comas@upf.edu

Supplementary Figures

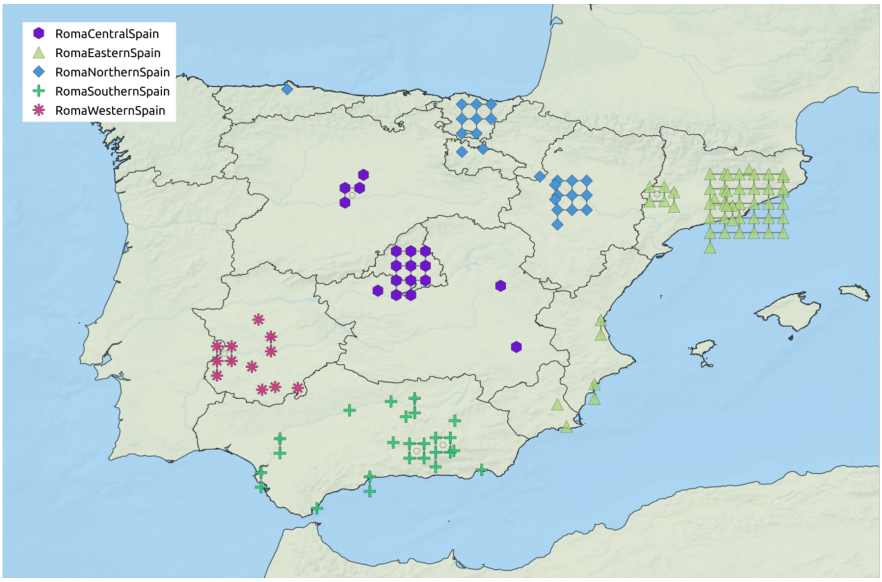


Figure S1. Geographic distribution of Spanish Roma individuals throughout Spain. Map created in QGIS software.

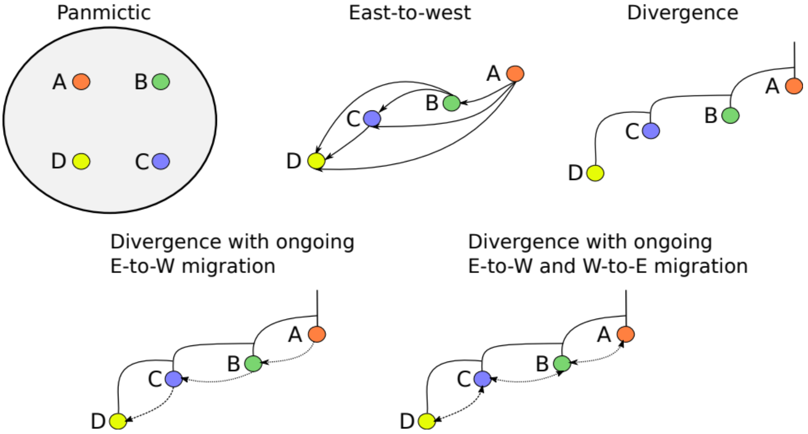


Figure S2. Five migration model routes used to investigate the movements of Roma populations.

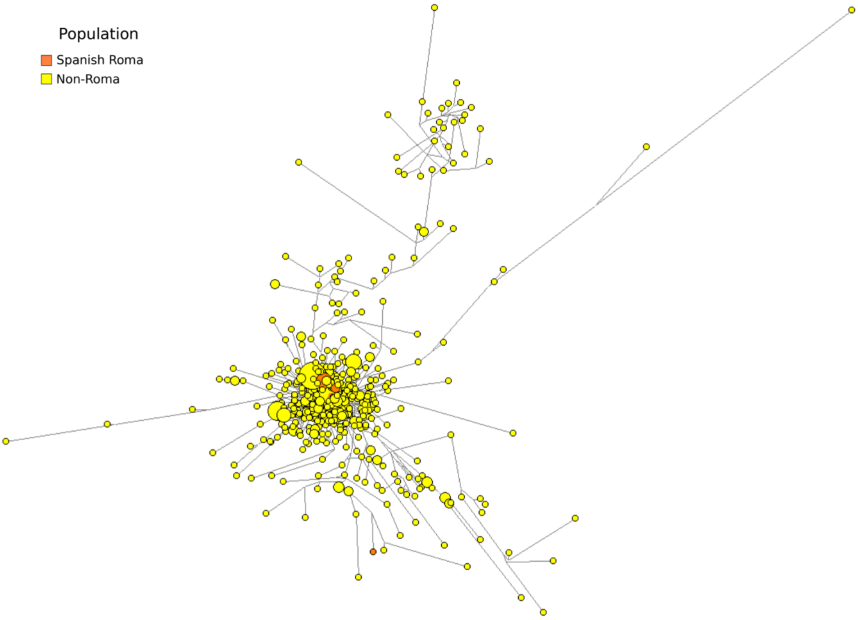


Figure S3. Median-joining network of Y-STR haplotypes showing the distribution of predicted H haplogroups across Spanish Roma and all reference populations. The network includes a random sample of up to 20 individuals per non-Roma groups.

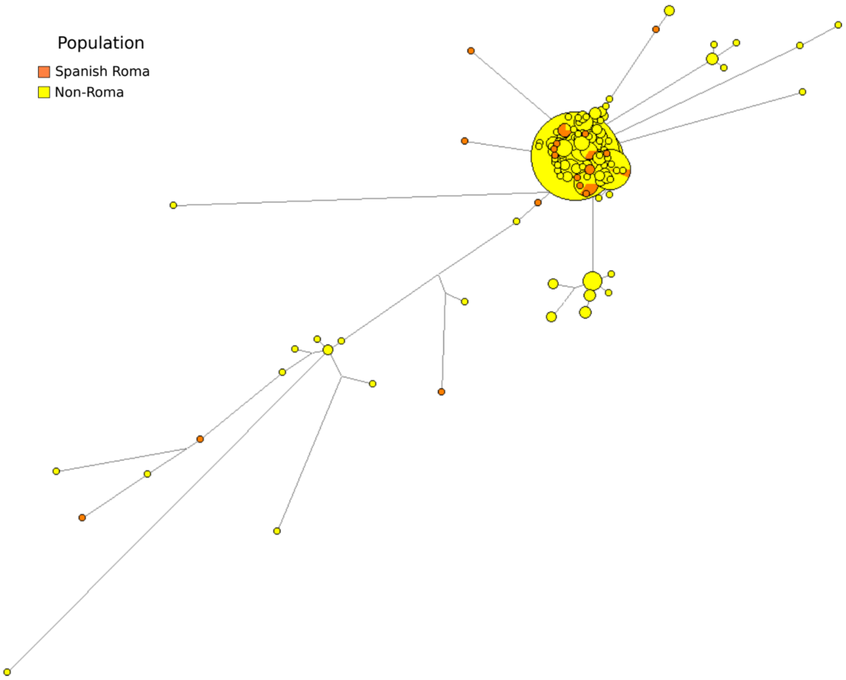


Figure S4. Median-joining network of Y-STR haplotypes showing the distribution of predicted H haplogroups across Roma groups.

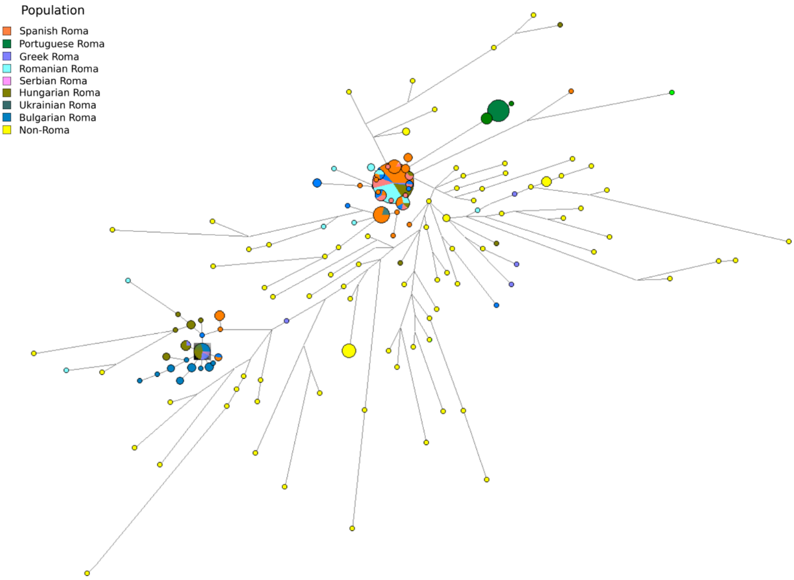


Figure S5. Median-joining network of Y-STR haplotypes showing the distribution of predicted J2a1b haplogroups across all populations.

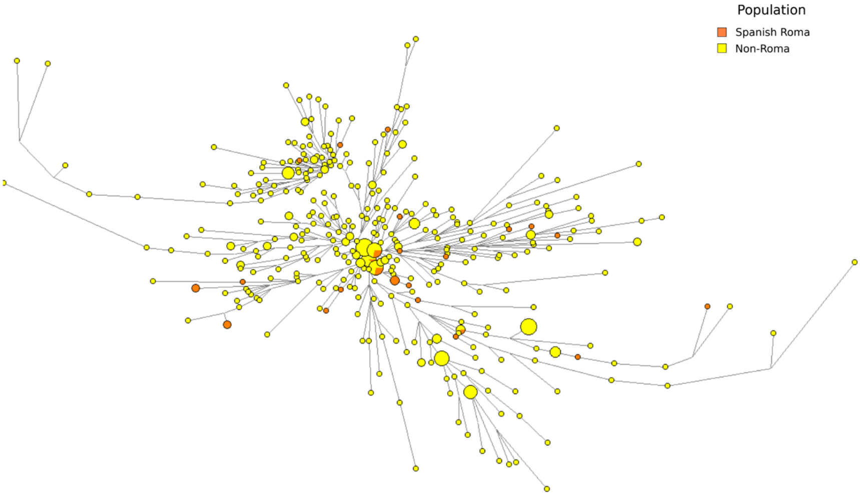


Figure S6. Median-joining network of Y-STR haplotypes illustrating the distribution of predicted R1b haplogroups across Spanish Roma and all reference populations. The network includes a random sample of up to 20 individuals per non-Roma groups.

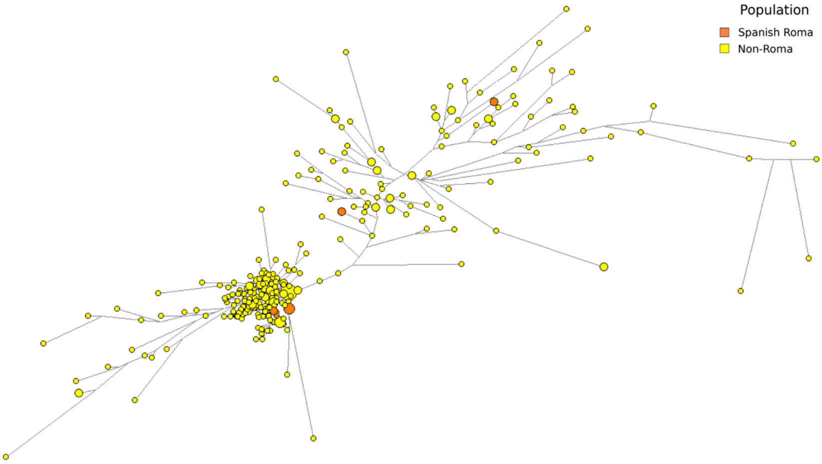


Figure S7. Median-joining network of Y-STR haplotypes illustrating the distribution of predicted I2a(x) haplogroups across Spanish Roma and all reference populations. The network includes a random sample of up to 20 individuals per non-Roma groups.

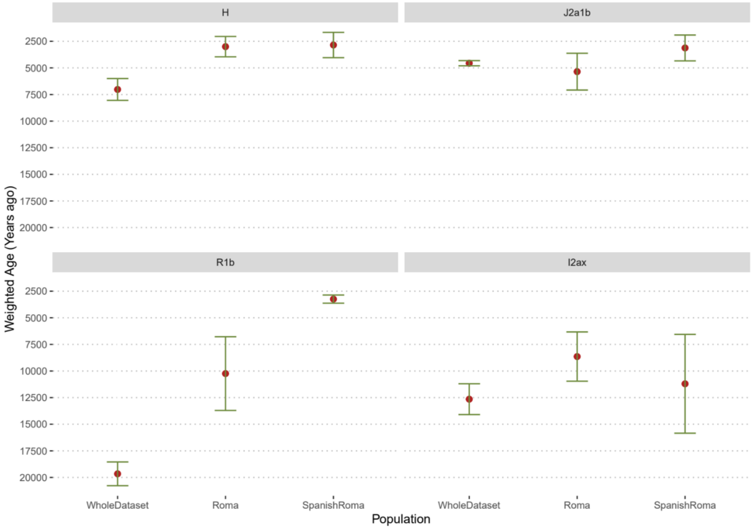


Figure S8. Time to the most recent common ancestor (TMRCA) estimates for the four major haplogroups, calculated using the rho statistics method and evaluated at three hierarchical levels.

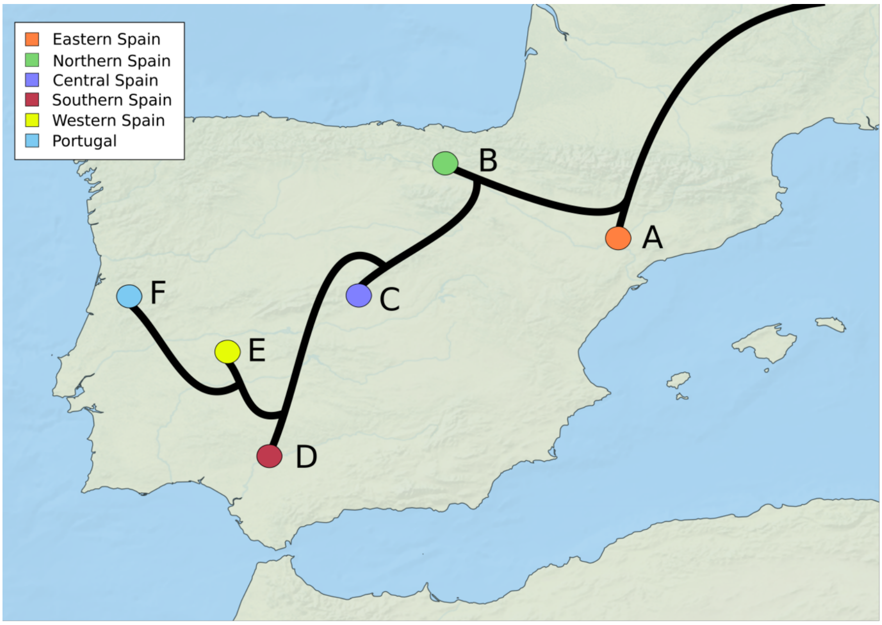


Figure S9. The most dominant migration routes of Roma individuals from the six different regions of the Iberian Peninsula. Map created in QGIS software.

Figure S10. The most dominant migration routes between Roma populations in the Iberian Peninsula, Slovakia, Romania, and Greece. Map created in QGIS software.

Supplementary Tables

Population	N	Reference	Haplogroup Prediction Availability
Spanish Roma	144	This project	//
Basque	199	García, O., Yurrebaso, I., Mancisidor, I. D., López, S., Alonso, S., & Gusmão, L. (2016). Data for 27 Y-chromosome STR loci in the Basque Country autochthonous population. Forensic Sci. Int. Genet., 20, e10–e12. doi: 10.1016/j.fsigen.2015.09.010	No
Bosnian Roma	110	Halilović, E., Ahmić, A., Kalajdžić, A., Ismailović, A., Čakar, J., Lasić, L., ...Pojskić, N. (2022). Paternal genetic structure of the Bosnian-Herzegovinian Roma: A Y-chromosomal STR study. Am. J. Hum. Biol., 34(6), e23719. doi: 10.1002/ajhb.23719	Yes
German	1718	Purps, J., Siegert, S., Willuweit, S., Nagy, M., Alves, C., Salazar, R., ...Roewer, L. (2014). A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. Forensic Sci. Int. Genet., 24854874. Retrieved from https://pubmed.ncbi.nlm.nih.gov/24854874	Yes
North Macedonia (Macedonian)	103	Jankova, R., Seidel, M., Videtič Paska, A., Willuweit, S., & Roewer, L. (2019). Y-chromosome diversity of the three major ethno-linguistic groups in the Republic of North Macedonia. Forensic Sci. Int. Genet., 42, 165–170. doi: 10.1016/j.fsigen.2019.07.007	Yes + SNP
North Macedonia (Turk)	109		
North Macedonia (Albanian)	102		
Portugal	214	Alves, C., Gomes, V., Prata, M. J., Amorim, A., & Gusmão, L. (2007). Population data for Y-chromosome haplotypes defined by 17 STRs (AmpFlSTR YFiler) in Portugal. Forensic Sci. Int., 171(2), 250–255. doi: 10.1016/j.forsciint.2006.10.022	No
Portugal North	158	Pontes, M. L., Cainé, L., Abrantes, D., Lima, G., & Pinheiro, M. F. (2007). Allele frequencies and population data for 17 Y-STR loci (AmpFℓSTR® Y-filer™) in a Northern Portuguese population sample. Forensic Sci. Int., 170(1), 62–67. doi: 10.1016/j.forsciint.2006.04.008	No
Portuguese Roma	126	Gusmão, A., Gusmão, L., Gomes, V., Alves, C., Calafell, F., Amorim, A., & Prata, M. J. (2008). A Perspective on the History of the Iberian Gypsies Provided by Phylogeographic Analysis of Y-Chromosome Lineages. Ann. Hum. Genet., 72(2), 215–227. doi: 10.1111/j.1469-1809.2007.00421.x	Yes
Sardinian	308	Lacerenza, D., Aneli, S., Di Gaetano, C., Critelli, R., Piazza, A., Matullo, G., ...Calò, C. (2017). Investigation of extended Y chromosome STR haplotypes in Sardinia. Forensic Sci. Int. Genet., 27, 172–174. doi: 10.1016/j.fsigen.2016.12.009	Yes + SNP
Serbian	195	Zgonjanin, D., Alghafri, R., Antov, M., Stojiljković, G., Petković, S., Vuković, R., & Drašković, D. (2017). Genetic characterization of 27 Y-STR loci with the Yfiler® Plus kit in the population of Serbia. Forensic Sci. Int. Genet., 31, e48–e49. doi: 10.1016/j.fsigen.2017.07.013	No
Serbian Roma	158	Regueiro, M., Rivera, L., Chennakrishnaiah, S., Popovic, B., Andjus, S., Milasin, J., & Herrera, R. J. (2012). Ancestral modal Y-STR haplotype shared among Romani and South Indian populations. Gene, 504(2), 296–302. doi: 10.1016/j.gene.2012.04.093	Yes + SNP
Spain (Catalunya)	130	Solé-Morata, N., Bertranpetit, J., Comas, D., & Calafell, F. (2015). Y-chromosome diversity in Catalan surname samples: insights into surname origin and frequency. Eur. J. Hum. Genet., 23, 1549–1557. doi: 10.1038/ejhg.2015.14	Yes + SNP
Spain (Almeria, Granada)	49	Gaibar, M., Esteban, E., Moral, P., Gómez-Gallego, F., Santiago, C., Bandrés, F., ...Fernández-Santander, A. (2010). STR genetic diversity in a Mediterranean population from the south of the Iberian Peninsula. Ann. Hum. Biol.	No
European populations (General and Roma) from Bulgaria, Greece, Hungary, Romania, Slovakia, Spain, Ukraine	1745	Martínez-Cruz, B., Mendizabal, I., Harmant, C., de Pablo, R., Ioana, M., Angelicheva, D., ...Comas, D. (2016). Origins, admixture and founder lineages in European Roma. Eur. J. Hum. Genet., 24, 937–943. doi: 10.1038/ejhg.2015.201	Yes
West and Central Asia			
Population	**N**	**Reference**	**Haplogroup Prediction Availability**
Afghan	108	Lacau, H., Bukhari, A., Gayden, T., La Salvia, J., Regueiro, M., Stojkovic, O., & Herrera, R. J. (2011). Y-STR profiling in two Afghanistan populations. Legal Med., 13(2), 103–108. doi: 10.1016/j.legalmed.2010.11.004	No
Armenian	412	Lowery, R. K., Herrera, K., Uribe, G., Reguiero, M., & Herrera, R. J. (2013). Sub-population structure evident in forensic Y-STR profiles from Armenian geographical groups. Legal Med., 15(2), 85–90. doi: 10.1016/j.legalmed.2012.10.003	Yes
Azhara (Durand Line)	258	Adnan, A., Rakha, A., Nazir, S., Alghafri, R., Hassan, Q., Wang, C.-C., & Lu, J. (2021). Forensic features and genetic legacy of the Baloch population of Pakistan and the Hazara population across Durand line revealed by Y-chromosomal STRs. Int. J. Legal Med., 135(5), 1777–1784. doi: 10.1007/s00414-021-02591-2	Yes
Balochi (Durand Line)	111	Adnan, A., Rakha, A., Nazir, S., Alghafri, R., Hassan, Q., Wang, C.-C., & Lu, J. (2021). Forensic features and genetic legacy of the Baloch population of Pakistan and the Hazara population across Durand line revealed by Y-chromosomal STRs. Int. J. Legal Med., 135(5), 1777–1784. doi: 10.1007/s00414-021-02591-2	Yes
Turkish from Cyprus	185	Teralı, K., Zorlu, T., Bulbul, O., & Gurkan, C. (2014). Population genetics of 17 Y-STR markers in Turkish Cypriots from Cyprus. Forensic Sci. Int. Genet., 10, e1–e3. doi: 10.1016/j.fsigen.2014.01.003	No
Iraqi Arab	222	Lazim, H., Almohammed, E. K., Hadi, S., & Smith, J. (2020). Population genetic diversity in an Iraqi population and gene flow across the Arabian Peninsula. Sci. Rep., 10(15289), 1–11. doi: 10.1038/s41598-020-72283-1	Yes
Iranian	215	Tabrizi, A. A., Hedjazi, A., Kerachian, M. A., Honarvar, Z., Dadgarmoghaddam, M., & Raoofian, R. (2015). Genetic profile of 17 Y-chromosome STR haplotypes in East of Iran. Forensic Sci. Int. Genet., 14, e6–e7. doi: 10.1016/j.fsigen.2014.10.010	No
Saudi Arabian	482	Khubrani, Y. M., Wetton, J. H., & Jobling, M. A. (2018). Extensive geographical and social structure in the paternal lineages of Saudi Arabia revealed by analysis of 27 Y-STRs. Forensic Sci. Int. Genet., 33, 98–105. doi: 10.1016/j.fsigen.2017.11.015	Yes
South Asia			
Population	**N**	**Reference**	**Haplogroup Prediction Availability**
Bengali	557	Hasan, M., Sufian, A., Momtaz, P., Mazumder, A. K., Khondaker, J. A., Bhattacharjee, S., ...Akhteruzzaman, S. (2018). Phylogenetic analysis and forensic evaluation among Rakhine, Marma, Hajong, and Manipuri tribes from four culturally defined regions of Bangladesh using 17 Y-chromosomal STRs. Int. J. Legal Med., 132(6), 1641–1644. doi: 10.1007/s00414-018-1915-4	No
Brahmin from Rajasthan	93	Sharma, S., Yadav, R., Sahajpal, V., Singh, M., Ranga, S., Kadian, L., ...Ahuja, P. (2022). Y-23 mediated genetic data analysis of endogamous Brahmin population of Rajasthan, India. Data in Brief, 42, 108061. doi: 10.1016/j.dib.2022.108061	No
Himachal Pradesh	166	Mohapatra, B. K., Chauhan, K., Shrivastava, P., Sharma, A., Dagar, S., & Kaitholia, K. (2019). Haplotype data for 17 Y-STR loci in the population of Himachal Pradesh, India. Int. J. Legal Med., 133(5), 1401–1402. doi: 10.1007/s00414-019-02080-7	No
Leh, Madras, Soligas	58	Regueiro, M., Rivera, L., Chennakrishnaiah, S., Popovic, B., Andjus, S., Milasin, J., & Herrera, R. J. (2012). Ancestral modal Y-STR haplotype shared among Romani and South Indian populations. Gene, 504(2), 296–302. doi: 10.1016/j.gene.2012.04.093	No
Kashmiri from Pakistan	47	Adnan, A., Rakha, A., Noor, A., van Oven, M., Ralf, A., & Kayser, M. (2018). Population data of 17 Y-STRs (Yfiler) from Punjabis and Kashmiris of Pakistan. Int. J. Legal Med., 132(1), 137–138. doi: 10.1007/s00414-017-1611-9	No
Khandayat from Odisha	90	Nayak, B. P., Khajuria, H., & Gupta, S. (2015). Y-STR haplotype diversity among the Khandayat population of Odisha, India. Egyptian Journal of Forensic Sciences, 5(2), 57–61. doi: 10.1016/j.ejfs.2014.07.003	No
Lingayat	101	Chennakrishnaiah, S., Perez, D., Gayden, T., Rivera, L., Regueiro, M., & Herrera, R. J. (2013). Indigenous and foreign Y-chromosomes characterize the Lingayat and Vokkaliga populations of Southwest India. Gene, 526(2), 96–106. doi: 10.1016/j.gene.2013.04.074	Yes
Madya Pradesh	442	Kaitholia, K., Dash, H. R., Shrivastava, P., Kumawat, R. K., Dixit, S., & Chaubey, G. (2021). Forensic characterization and genetic evaluation in the Central Indian population using 27 Y-STRs. Int. J. Legal Med., 135(3), 791–792. doi: 10.1007/s00414-020-02486-8	Yes
Pakistan general population	445	Adnan, A., Rakha, A., Lao, O., & Kayser, M. (2018). Mutation analysis at 17 Y-STR loci (Yfiler) in father-son pairs of male pedigrees from Pakistan. Forensic Sci. Int. Genet., 36, e17–e18. doi: 10.1016/j.fsigen.2018.07.001	No
Punjabi from Pakistan	252	Perveen, R., Rahman, Z., Shahzad, M. S., Israr, M., Shafique, M., Shan, M. A., ...Husnain, T. (2014). Y-STR haplotype diversity in Punjabi population of Pakistan. Forensic Sci. Int. Genet., 9, e20–e21. doi: 10.1016/j.fsigen.2013.12.004	Yes + SNP
Punjabi from Pakistan	71	Adnan, A., Rakha, A., Noor, A., van Oven, M., Ralf, A., & Kayser, M. (2018). Population data of 17 Y-STRs (Yfiler) from Punjabis and Kashmiris of Pakistan. Int. J. Legal Med., 132(1), 137–138. doi: 10.1007/s00414-017-1611-9	No
Tamil Nadu (Ezhava, Irula, Sourashra, Yadhava)	322	ArunKumar, G., Soria-Hernanz, D. F., Kavitha, V. J., Arun, V. S., Syama, A., Ashokan, K. S., ...The Genographic Consortium. (2012). Population Differentiation of Southern Indian Male Lineages Correlates with Agricultural Expansions Predating the Caste System. PLoS One, 7(11), e50269. doi: 10.1371/journal.pone.0050269	Yes
Vokkaliga	102	Chennakrishnaiah, S., Perez, D., Gayden, T., Rivera, L., Regueiro, M., & Herrera, R. J. (2013). Indigenous and foreign Y-chromosomes characterize the Lingayat and Vokkaliga populations of Southwest India. Gene, 526(2), 96–106. doi: 10.1016/j.gene.2013.04.074	Yes

Supplementary Table S1: Collection of populations used in this project, with sources and availability of predicted haplogroup information.

Region	DYS393	DYS390	DYS19	DYS391	DYS385a	DYS385b	DYS439	DYS389I	DYS392	DYS389II	DYS389II (subtracted)	DYS458	DYS437	DYS448	DYS449	DYS460	GATA-H4	DYS456	DYS576	DYS570	DYS438	DYS481	DYS635
RomaCentralSpain	12	23	14	10	13	16	12	14	11	31	17	17	14	20	32	9	10	15	18	16	9	23	21
RomaCentralSpain	12	23	14	10	13	16	11	14	11	31	17	17	14	20	99	99	10	15	99	99	9	99	21
RomaCentralSpain	12	22	15	10	15	16	11	14	99	30	16	18	14	19	99	99	12	15	99	99	9	99	20
RomaCentralSpain	12	23	14	10	13	16	11	15	11	32	17	17	14	20	99	99	10	15	99	99	9	99	21
RomaCentralSpain	13	23	14	11	11	14	12	13	13	29	16	17	15	19	99	99	11	15	99	99	12	99	23
RomaCentralSpain	13	25	14,1	11	11	15	12	15	13	30	15	17,1	16	19	29	10	12	15	19	16	12	22	24
RomaCentralSpain	13	23	16	10	14	16	13	13	11	30	17	17	15	20	99	99	11	15	99	99	10	99	22
RomaCentralSpain	13	25	13	11	11	14	12	14	13	30	16	18	15	19	30	10	11	16	20	17	12	20	23
RomaCentralSpain	12	23	15	11	14	19	11	12	11	30	18	17	15	20	99	99	12	14	99	99	9	99	23
RomaCentralSpain	12	22	15	10	15	17	11	14	11	30	16	18	14	19	31	11	12	15	18	16	9	23	21
RomaCentralSpain	14	23	15	9	13	16	13	14	12	32	18	15	14	20	25	11	11	13	18	18	10	27	21
RomaCentralSpain	12	22	14	10	13	16	11	14	11	31	17	17	14	20	31	9	10	15	18	16	9	23	21
RomaCentralSpain	12	24	15	10	13	17	12	12	11	30	18	16	16	19	30	11	11	13	16	17	9	23	22
RomaCentralSpain	12	23	14	10	13	16	11	14	11	31	17	17	14	20	32	9	10	15	18	16	9	23	21
RomaCentralSpain	12	22	15	11	12	14	12	14	11	29	15	18	14	19	29	11	12	15	18	16	9	22	20
RomaCentralSpain	12	22	15	10	15	17	11	14	11	30	16	18	14	19	31	11	12	15	19	16	9	23	20
RomaCentralSpain	12	22	15	10	15	17	12	14	11	30	16	17	14	19	32	11	11	15	20	17	9	23	20
RomaCentralSpain	13	24	14	10	11	14	13	13	13	29	16	18	15	19	28	11	11	16	17	17	12	19	23
RomaEasternSpain	12	23	14	12	13	19	11	13	11	30	17	18	14	20	99	99	11	14	99	99	10	99	21
RomaEasternSpain	12	23	14	10	14	17	12	13	11	29	16	17	14	20	99	99	11	15	99	99	10	99	20
RomaEasternSpain	13	23	14	12	11	14	12	14	13	30	16	16	14	19	99	99	12	16	99	99	12	99	23
RomaEasternSpain	13	24	14	11	11	14	12	13	13	30	17	17	15	19	99	99	12	15	99	99	12	99	23
RomaEasternSpain	12	23	14	10	13	16	11	14	11	32	18	17	14	20	32	9	10	15	18	17	9	23	21
RomaEasternSpain	13	23	15	11	11	14	12	13	13	29	16	17	14	18	99	99	10	15	99	99	12	99	23
RomaEasternSpain	12	22	15	10	15	17	11	14	11	30	16	18	14	19	31	11	12	15	18	16	9	23	20
RomaEasternSpain	12	22	15	10	15	17	11	14	11	30	16	18	14	19	99	99	12	15	99	99	9	99	20
RomaEasternSpain	12	23	14	10	13	16	11	14	11	32	18	17	14	20	99	99	10	15	99	99	9	99	21
RomaEasternSpain	12	22	15	10	15	17	11	14	11	30	16	17	14	19	99	99	12	15	99	99	9	99	20
RomaEasternSpain	12	23	14	10	13	16	11	14	11	32	18	17	14	20	99	99	10	15	99	99	9	99	21
RomaEasternSpain	12	23	14	10	13	16	12	14	11	31	17	17	14	20	99	99	10	15	99	99	9	99	21
RomaEasternSpain	13	23	17	10	12	99	12	14	11	30	16	17	15	20	99	99	12	14	99	99	10	99	21
RomaEasternSpain	12,1	23	13	10	10	13	11	13	11,1	29	16	16	14	18,1	28	10	11	14	18	18	11	21	23
RomaEasternSpain	12	23	14	10	13	16	11	14	11	31	17	17	14	20	99	99	10	15	99	99	9	99	23
RomaEasternSpain	12	22	17	10	13	15	12	13	11	32	19	15	14	21	99	99	12	15	99	99	9	99	21
RomaEasternSpain	12	23	14	10	13	16	11	15	11	32	17	17	14	20	99	99	10	15	99	99	9	99	21
RomaEasternSpain	12	22	15	10	15	17	11	15	11	31	16	17	14	19	99	99	12	15	99	99	9	99	20
RomaEasternSpain	12	22	15	10	15	17	11	15	11	31	16	17	14	19	99	99	12	15	99	99	9	99	20
RomaEasternSpain	12	23	14	10	14	17	12	13	11	29	16	17,2	14	20	25	11	11	15	19	19	10	25	20
RomaEasternSpain	12	23	14	10	13	16	11	14	11	31	17	17	14	20	99	99	10	15	99	99	9	99	21
RomaEasternSpain	13	25	13	11	11	14	12	14	13	30	16	18	15	19	30	10	11	16	20	17	12	20	23
RomaEasternSpain	12	22	15	10	15	17	11	14	11	30	16	18	15	19	99	99	12	15	99	99	9	99	20
RomaEasternSpain	13	24	14	10	11	14	12	13	13	29	16	18	15	19	99	99	12	15	99	99	13	99	23
RomaEasternSpain	12	23	14	10	13	16	11	14	11	30	16	17	14	20	99	99	10	15	99	99	9	99	21
RomaEasternSpain	14	24	15	11	11	14	11	14	12	30	16	17	14	18	28	11	11	16	18	17	12	22	24
RomaEasternSpain	12	22	16	10	13	16	11	13	11	32	19	15	14	21	31	10	12	15	17	20	9	23	21
RomaEasternSpain	12	22	16	10	13	15	11	13	11	32	19	15	14	21	31	10	12	15	17	19	9	23	21
RomaEasternSpain	13	24	17	10	14	15	12	12	11	30	18	18	15	19	29	10	12	15	19	18	10	29	23
RomaEasternSpain	12	23	14	10	13	16	11	14	11	31	17	17	14	20	33	9	10	15	18	15	9	23	21
RomaEasternSpain	12	23	14	10	13	16	11	14	11	31	17	17	14	20	32	9	10	15	18	16	9	23	21
RomaEasternSpain	12	23	15	10	15	17	11	14	11	30	16	18	14	19	31	11	12	15	20	16	9	22	20
RomaEasternSpain	12	23	15	9	13	16	12	13	11	29	16	14	14	21	31	10	11	15	15	16	9	23	22
RomaEasternSpain	13	24	14	11	12	14	12	13	13	29	16	17	15	19	29	11	12	15	19	18	12	22	23
RomaEasternSpain	12	23	15	10	15	17	11	14	11	30	16	18	14	19	31	11	12	15	19	16	9	22	20
RomaEasternSpain	12	22	16	10	13	15	11	13	11	32	19	15	14	21	31	10	12	15	17	17	9	23	21
RomaEasternSpain	12	22	16	10	15	18	12	14	11	30	16	17	14	19	32	11	11	15	18	16	99	22	20
RomaEasternSpain	12	23	14	10	13	16	11	14	11	31	17	17	14	20	32	9	10	15	18	16	9	23	21
RomaEasternSpain	13	22	15	10	15	16	11	14	11	30	16	17	14	19	32	11	12	15	19	17	9	23	20
RomaEasternSpain	12	23	14	10	13	16	12	14	11	31	17	17	14	20	32	9	10	15	18	16	9	23	21
RomaEasternSpain	12	22	15	10	15	16	11	14	11	30	16	17	14	19	32	11	12	15	19	18	9	23	20
RomaEasternSpain	13	24	15	10	10	14	12	13	13	29	16	19	15	19	29	11	12	15	18	19	12	22	23
RomaEasternSpain	12	23	14	10	13	16	11	14	11	31	17	17	14	20	33	9	10	15	18	16	9	23	21
RomaEasternSpain	13	24	14	10	9	14	12	14	13	29	15	17	15	19	30	11	12	15	18	17	12	22	24
RomaEasternSpain	12	22	15	10	15	17	11	14	11	30	16	18	14	19	31	11	12	15	19	16	9	23	20
RomaEasternSpain	12	22	15	10	15	17	11	14	11	30	16	17	14	18	31	10	12	15	18	17	9	23	20
RomaEasternSpain	13	24	17	10	14	15	12	12	11	30	18	18	15	19	29	10	12	15	18	18	10	29	23
RomaEasternSpain	12	23	15	10	15	17	11	14	11	30	16	18	14	19	31	11	12	15	19	16	9	22	21
RomaEasternSpain	12	23	14	10	13	16	11	14	11	31	17	17	14	20	32	9	10	15	19	16	9	23	21
RomaEasternSpain	12	23	14	10	13	16	11	14	11	32	18	17	14	20	99	99	10	15	99	99	9	99	21
RomaEasternSpain	10	23	14	10	13	16	12	14	11	30	16	17	14	20	99	99	10	15	99	99	9	99	21
RomaEasternSpain	12	23	15	10	15	99	11	14	11	30	16	17	14	19	99	99	13	15	99	99	9	99	20
RomaNorthernSpain	12	23	14	10	13	16	11	14	11	32	18	17	14	20	32	9	10	15	18	16	9	23	21
RomaNorthernSpain	13	24	14	13	14	99	13	14	13	30	16	19	15	18	99	99	12	15	99	99	12	99	23
RomaNorthernSpain	12	23	14	10	13	16	11	15	11	32	17	17	14	20	99	99	10	15	99	99	9	99	21
RomaNorthernSpain	12	23	14	10	16	99	12	13	11	29	16	16	15	21	99	99	12	15	99	99	9	99	21
RomaNorthernSpain	12	23	14	10	13	16	11	14	11	31	17	17	14	20	99	99	10	15	99	99	9	99	21
RomaNorthernSpain	12	23	14	10	16	99	12	13	11	29	16	16	15	21	99	99	12	15	99	99	9	99	21
RomaNorthernSpain	12	23	14	10	13	16	11	15	11	32	17	18	14	20	99	99	10	15	99	99	9	99	21
RomaNorthernSpain	13	23	13	10	13	16	11	14	13	30	16	14	15	19	99	99	11	15	99	99	9	99	21
RomaNorthernSpain	13	24	14	11	11	14	12	14	13	30	16	17	15	19	29	10	12	15	17	19	12	22	24
RomaNorthernSpain	12	22	15	10	15	17	11	14	11	30	16	18	14	19	31	11	12	16	19	16	7	23	20
RomaNorthernSpain	13	24	14	10	11	14	11	13	14	29	16	18	15	19	99	99	12	16	99	99	12	99	23
RomaNorthernSpain	12	23	14	10	13	16	11	14	11	31	17	17	14	20	99	99	10	15	99	99	9	99	21
RomaNorthernSpain	13	24	14	11	12	14	11	13	13	29	16	17	15	19	99	99	12	15	99	99	12	99	23
RomaNorthernSpain	13	25	15	10	11	99	10	13	11	32	19	16	14	20	99	99	12	16	99	99	11	99	23
RomaNorthernSpain	12	23	14	10	13	16	11	14	11	32	18	17	14	20	32	9	10	15	18	17	9	23	21
RomaNorthernSpain	12	23	14	10	13	16	11	14	11	32	18	17	14	20	32	9	10	15	18	17	9	23	21
RomaNorthernSpain	12	23	14	10	13	16	11	99	11	32	99	17	14	20	32	9	10	15	18	17	9	23	21
RomaNorthernSpain	12	23	14	10	13	16	11	15	11	32	17	18	14	20	32	9	10	15	18	16	9	23	21
RomaNorthernSpain	13	25	13	10	11	14	12	14	13	30	16	18	14	19	30	10	11	16	20	17	12	20	23
RomaNorthernSpain	14	22	15	11	99	14	10	12	11	28	16	18	15	20	30	11	12	14	18	16	10	21	21
RomaNorthernSpain	12	22	15	10	15	17	11	15	11	31	16	18	14	19	31	11	12	16	19	16	9	23	20
RomaNorthernSpain	12	23	14	10	13	16	11	15	11	32	17	18	14	20	32	9	10	15	18	16	9	23	21
RomaNorthernSpain	12	22	15	10	16	17	11	14	11	30	16	17	14	20	32	11	12	15	19	17	9	23	20
RomaNorthernSpain	13	24	14	11	11	15	12	13	13	29	16	18	15	19	30	11	12	16	17	19	12	22	23
RomaNorthernSpain	12	23	14	10	13	16	11	14	11	31	17	17	14	20	32	9	10	15	18	16	9	23	21
RomaNorthernSpain	12	23	15	10	15	17	11	14	11	30	16	19	14	19	31	11	13	15	20	17	9	22	20
RomaSouthernSpain	12	22	15	10	15	16	11	15	11	31	16	20	14	19	99	99	12	15	99	99	9	99	20
RomaSouthernSpain	12	23	15	10	15	17	11	14	11	30	16	18	14	19	99	99	12	15	99	99	9	99	20
RomaSouthernSpain	12	23	14	10	13	16	11	14	11	31	17	17	14	20	99	99	10	15	99	99	9	99	21
RomaSouthernSpain	12	22	14	10	13	16	12	13	11	30	17	16	15	20	99	99	11	15	99	99	9	99	21
RomaSouthernSpain	12	23	14	10	13	16	11	14	11	31	17	17	14	20	99	99	10	15	99	99	9	99	21
RomaSouthernSpain	12	23	15	10	15	17	11	14	11	30	16	18	14	19	99	99	12	15	99	99	9	99	20
RomaSouthernSpain	13	25	14	11	11	14	11	13	13	29	16	17	15	19	99	99	12	16	99	99	12	99	23
RomaSouthernSpain	12	22	15	10	15	17	11	14	11	30	16	18	14	19	99	99	12	15	99	99	9	99	20
RomaSouthernSpain	12	23	14	10	13	16	12	14	11	30	16	17	14	20	99	99	10	15	99	99	9	99	21
RomaSouthernSpain	13	23	15	10	15	17	11	14	11	30	16	18	14	19	99	99	12	15	99	99	9	99	20
RomaSouthernSpain	12	23	14	10	13	16	11	14	11	31	17	17	14	20	99	99	10	15	99	99	9	99	20
RomaSouthernSpain	13	24	14	11	11	14	12	14	13	31	17	20	14	19	99	99	12	16	99	99	12	99	23
RomaSouthernSpain	12	23	14	10	13	16	12	14	11	31	17	17	14	20	99	99	10	15	99	99	9	99	21
RomaSouthernSpain	12	23	14	10	13	16	12	14	11	31	17	17	14	20	99	99	10	15	99	99	9	99	21
RomaSouthernSpain	13	25	13	11	11	14	12	14	13	30	16	18	15	19	30	10	11	16	20	17	12	20	23
RomaSouthernSpain	12	22	16	10	13	15	11	13	11	32	19	15	14	21	31	10	12	15	17	19	9	23	21
RomaSouthernSpain	12	22	15	10	15	17	11	14	11	30	16	20	14	19	32	11	12	15	17	16	9	23	20
RomaSouthernSpain	12	23	14	10	13	16	11	14	11	31	17	17	14	20	33	9	10	15	17	15	9	23	21
RomaSouthernSpain	12	22	15	10	15	17	11	14	11	30	16	18	14	19	31	11	12	15	19	16	9	22	20
RomaSouthernSpain	12	23	14	10	12	16	12	14	11	31	17	17	14	20	99	99	10	15	99	99	9	99	21
RomaSouthernSpain	12	23	14	10	13	16	12	14	11	31	17	17	14	20	99	99	10	15	99	99	9	99	21
RomaSouthernSpain	12	22	15	10	15	16	11	14	11	30	16	18	14	19	99	99	12	15	99	99	9	99	20
RomaSouthernSpain	12	22	15	10	15	17	11	14	11	30	16	18	14	19	99	99	12	15	99	99	9	99	20
RomaSouthernSpain	12	23	15	10	15	17	11	14	11	30	16	17	14	19	99	99	13	15	99	99	9	99	20
RomaSouthernSpain	12	23	15	10	15	17	11	14	11	30	16	18	14	19	99	99	12	15	99	99	9	99	20
RomaWesternSpain	12	22	16	10	13	15	11	13	11	31	18	15	14	21	99	99	12	15	99	99	9	99	21
RomaWesternSpain	12	22	15	10	15	17	12	14	11	30	16	17	14	20	99	99	11	15	99	99	9	99	20
RomaWesternSpain	12	22	16	10	13	15	11	13	11	31	18	15	14	21	99	99	12	15	99	99	9	99	21
RomaWesternSpain	13	23	14	11	11	15	11	13	13	30	17	17	15	19	99	99	12	16	99	99	12	99	23
RomaWesternSpain	13	23	14	12	11	14	12	13	13	31	18	16	15	19	99	99	13	17	99	99	12	99	23
RomaWesternSpain	12	22	15	10	15	17	12	14	11	30	16	17	14	19	99	99	11	15	99	99	9	99	20
RomaWesternSpain	12	22	15	10	16	99	11	14	11	30	16	17	14	19	99	99	12	16	99	99	9	99	20
RomaWesternSpain	12	23	14	10	13	16	11	14	11	31	17	17	14	20	99	99	10	15	99	99	9	99	21
RomaWesternSpain	12	22	15	10	15	17	11	14	11	30	16	18	14	20	99	99	12	15	99	99	9	99	20
RomaWesternSpain	12	22	16	10	13	15	11	13	11	31	18	15	14	21	31	9	12	15	17	17	9	23	21
RomaWesternSpain	13	23	14	11	11	14	12	13	13	30	17	16	15	19	29	11	13	17	17	17	12	23	23
RomaWesternSpain	12	22	15	10	16	17	11	14	11	30	16	17	14	19	31	11	12	15	20	17	9	23	20

Supplementary Table S2: Raw data for the 133 newly genotyped Spanish Roma individuals.

	Probability (%)																											
Region	**C3**	**E1a**	**E1b1a**	**E1b1b**	**G1**	**G2a**	**G2c**	**H**	**I1**	**I2a (xI2a1)**	**I2a1**	**I2b (xI2b1)**	**I2b1**	**J1**	**J2a1b**	**J2a1h**	**J2a1 (x bh)**	**J2b**	**L**	**N**	**O2**	**O3**	**Q**	**R1a**	**R1b**	**R2**	**T**	**Predicted haplogroup (>85%)**
RomaCentralSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,9	92,2	0	2,9	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaCentralSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaCentralSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaCentralSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	3,1	95,7	0	1,2	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaCentralSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaCentralSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaCentralSpain	0	0	0	0	0	0,1	0	0	0	99,9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	I2a (xI2a1)
RomaCentralSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaCentralSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,1	0	99,6	0	0	0	0	0,3	0	0	0	0	0	J2a1 (x bh)
RomaCentralSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaCentralSpain	0	0	0	7,4	0	0,1	0	0	0	0	0	0	92	0,4	0	0	0	0	0	0	0	0,1	0	0	0	0	0	I2b1
RomaCentralSpain	0	0	0	0	0	0	0	0,4	0	0	0	0	0	0,7	97,7	0	1,2	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaCentralSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	J2b
RomaCentralSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaCentralSpain	0	0	0	0	0	0	0	99,5	0	0	0	0	0	0	0,3	0	0,2	0	0	0	0	0	0	0	0	0	0	H
RomaCentralSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaCentralSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaCentralSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	J1
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	99,8	0,1	0	0	0	0	0	0	0	0	0	0	0	0	J1
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaEasternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaEasternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaEasternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,9	92,2	0	2,9	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	I2a1
RomaEasternSpain	0	0	0	0,1	0	0	0	0	0	0	0	0	0	0,5	0	0	1,2	0	0	2,4	0	75,8	3,5	16,1	0,2	0,3	0	
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99,9	0	0,1	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	95,4	0	4,5	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaEasternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	35	62,3	2,7	0	0	0	0	0	0	0	0	0	0	
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	3,1	95,7	0	1,2	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaEasternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaEasternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	99,8	0,1	0	0	0	0	0	0	0	0	0	0	0	0	J1
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaEasternSpain	0	0	0	0	0	0	0	99,9	0	0	0	0	0	0	0,1	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,4	0,1	0	0	0,4	99	0,1	0	R1b
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	85,8	1,2	0	0	0	0	0	0	0	0	0	0	J2a1h
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	46	26,9	27,1	0	0	0	0	0	0	0	0	0	0	
RomaEasternSpain	0	0	0	0	0	0	0	0	0	96,8	0	0,1	0	0	0	0	0	0	0	0	0	3,2	0	0	0	0	0	I2a(xI2a1)
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaEasternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	J2a1h
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaEasternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	46	26,9	27,1	0	0	0	0	0	0	0	0	0	0	
RomaEasternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaEasternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,9	92,2	0	2,9	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaEasternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaEasternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaEasternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaEasternSpain	0	0	0	0	0	0	0	0	0	96,8	0	0,1	0	0	0	0	0	0	0	0	0	3,2	0	0	0	0	0	I2a(xI2a1)
RomaEasternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaEasternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	3,1	95,7	0	1,2	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	1,2	38,2	0,7	59,8	0	0	0	0	0	0	0	0	0	0	
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	1,2	38,2	0,7	59,8	0	0	0	0	0	0	0	0	0	0	
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	1	98,6	0	0,4	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	T
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaNorthernSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	R1a
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	7	91,1	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	1	98,6	0	0,4	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,6	0	99	0,4	0	R1b
RomaNorthernSpain	0	0	0	0	0	99,6	0	0	0,4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	G2a
RomaNorthernSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	1	98,6	0	0,4	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaNorthernSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaNorthernSpain	0	0	0	0	0	0	0	99,8	0	0	0	0	0	0	0,2	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaNorthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaSouthernSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaSouthernSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaSouthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaSouthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98,9	0	1,1	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaSouthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaSouthernSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaSouthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaSouthernSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaSouthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,9	92,2	0	2,9	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaSouthernSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaSouthernSpain	0	0	0	0	0	0	0	0,1	0	0	0	0	0	40	59,6	0	0,2	0	0	0	0	0	0	0	0	0	0	
RomaSouthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaSouthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,9	92,2	0	2,9	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaSouthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,9	92,2	0	2,9	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaSouthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaSouthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	46	26,9	27,1	0	0	0	0	0	0	0	0	0	0	
RomaSouthernSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaSouthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaSouthernSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaSouthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	5,5	83,1	0	11,3	0	0	0	0	0	0	0	0	0	0	
RomaSouthernSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,9	92,2	0	2,9	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaSouthernSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaSouthernSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaSouthernSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaSouthernSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaWesternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	46	26,9	27,1	0	0	0	0	0	0	0	0	0	0	
RomaWesternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaWesternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	46	26,9	27,1	0	0	0	0	0	0	0	0	0	0	
RomaWesternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaWesternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaWesternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaWesternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaWesternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	4,7	93,5	0	1,8	0	0	0	0	0	0	0	0	0	0	J2a1b
RomaWesternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H
RomaWesternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	46	26,9	27,1	0	0	0	0	0	0	0	0	0	0	
RomaWesternSpain	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	R1b
RomaWesternSpain	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	H

Supplementary Table S3: Predicted Y-haplogroups for the 133 newly genotyped Spanish Roma individuals.

Haplogroup	Bosnian Roma (N)	Bosnian Roma (%)	Bulgarian Roma (N)	Bulgarian Roma (%)	Greek Roma (N)	Greek Roma (%)	Hungarian Roma (N)	Hungarian Roma (%)	Portuguese Roma (N)	Portuguese Roma (%)	Romanian Roma (N)	Romanian Roma (%)	Serbian Roma (N)	Serbian Roma (%)	Slovakian Roma (N)	Slovakian Roma (%)	Spanish Roma (N)	Spanish Roma (%)	Ukrainian Roma (N)	Ukrainian Roma (%)	Basque (N)	Basque (%)	Spanish (N)	Spanish (%)	Portuguese (N)	Portuguese (%)
E	1	0,9	22	9,5	6	10,3	15	7,4	4	3,2	1	1,1	13,9	22	5,6	6	0	0	14,3	3	9	4,5	107	7,4	46	12,4
G	1	0,9	1	0,4	0	0	7	3,5	1	0,8	0	0	0	0	0	0	0	0	0	0	5	2,5	57	3,9	24	6,5
H	72	65,5	106	45,7	39	67,2	69	34,2	21	16,7	44	50	51,9	82	54,2	58	27,7	48	28,6	6	0	0	0	0	0	0
I1	20	18,2	23	9,9	2	3,4	14	6,9	1	0,8	5	5,7	5,7	9	6,5	7	0	0	19	4	0	0	33	2,3	16	4,3
I2a1	6	5,5	30	12,9	4	6,9	10	5	0	0	4	4,5	6,3	10	0	0	3,5	6	4,8	1	7	3,5	47	3,2	8	2,2
J2a1	9	8,2	0	0,0	0	0	0	0	1	0,8	0	0	0	0	0	0	0	0	0	0	0,0	0	7	0,5	5	1,3
J2b	1	0,9	0	0	0	0	0	0	1	0,8	1	1,1	0,6	1	0	0	0,6	1	0	0	1,0	0,5	21	1,4	4	1,1
J1a	0	0	2	0,9	0	0	0	0	0	0	0	0	0	0	0,9	1	0	0	0	0	0,0	0	1	0,1	0	0
J2a	0	0	3	1,3	0	0	6	3	22	17,5	0	0	2,5	4	14	15	0,6	1	9,5	2	0,0	0	51	3,5	5	1,3
J2a1b	0	0	36	15,5	5	8,6	27	13,4	25	19,8	27	30,7	5,1	8	8,4	9	34,1	59	14,3	3	0,0	0	30	2,1	8	2,2
R1a	0	0	2	0,9	0	0	29	14,4	0	0	1	1,1	12	19	6,5	7	0,6	1	4,8	1	0,0	0	17	1,2	2	0,5
R1b	0	0	7	3	1	1,7	16	7,9	36	28,6	4	4,5	1,3	2	3,7	4	18,5	32	0	0	175,0	87,9	987	67,9	223	59,9
J2	0	0	0	0	1	1,7	0	0	0	0	0	0	0	0	0	0	3,5	6	0	0	0,0	0	0	0	0	0
C	0	0	0	0	0	0	1	0,5	0	0	0	0	0	0	0	0	0	0	0	0	0,0	0	1	0,1	0	0
I2a	0	0	0	0	0	0	2	1	0	0	1	1,1	0,6	1	0	0	1,2	2	0	0	0,0	0	22	1,5	0	0
N	0	0	0	0	0	0	1	0,5	0	0	0	0	0	0	0	0	0	0	4,8	1	0,0	0	2	0,1	1	0,3
O	0	0	0	0	0	0	1	0,5	0	0	0	0	0	0	0	0	0	0	0	0	0,0	0	0	0	1	0,3
Q	0	0	0	0	0	0	2	1	1	0,8	0	0	0	0	0	0	0	0	0	0	0,0	0	0	0	1	0,3
R2a	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0,0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	12	9,5	0	0	0	0	0	0	0	0	0	0	0,0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	1	0,8	0	0	0	0	0	0	0	0	0	0	0,0	0	1	0,1	0	0
G2a	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,6	1	0	0	0,0	0	0	0	0	0
I2a(xI2a1)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1,7	3	0	0	0,0	0	1	0,1	3	0,8
I2b1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,6	1	0	0	1,0	0,5	0	0	6	1,6
J1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1,7	3	0	0	1,0	0,5	36	2,5	12	3,2
J2a1(xbh)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1,7	3	0	0	0,0	0	0	0	0	0
J2a1h	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2,31	4	0	0	0,0	0	0	0	0	0
O3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,6	1	0	0	0,0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,6	1	0	0	0,0	0	21	1,4	7	1,9
I2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,0	0	8	0,6	0	0
L	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,0	0	1	0,1	0	0
R2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,0	0	2	0,1	0	0

Supplementary Table S4: Distribution of haplogroups in all Roma populations across Europe and non-Roma Iberian populations.

			Pedigree Mutation Rate by STR				*Median Pedigree Mutation Rate (2.5×10^{-3})*			
Haplogroup	**Population**	**N**	**Weighted age**	**SD**	**Lower Date**	**Upper Date**	**Weighted age**	**SD**	**Lower Date**	**Upper Date**
H	Whole Dataset	1090	7026	1023	8049	6003	4250	295	4545	3955
H	Roma	545	3003	955	3958	2048	2334	371	2705	1963
H	Spanish Roma	48	2850	1189	4039	1661	1458	244	1702	1214
R1b	Whole Dataset	2548	19658	1115	20773	18543	12734	445	13179	12289
R1b	Roma	101	10243	3463	13706	6780	6640	1030	7670	5610
R1b	Spanish Roma	31	3245	385	3630	2860	2925	269	3194	2656
J2a1b	Whole Dataset	294	4560	242	4802	4318	4712	221	4933	4491
J2a1b	Roma	200	5352	1728	7080	3624	4410	666	5076	3744
J2a1b	Spanish Roma	60	3126	1216	4342	1910	4222	1651	5873	2571
I2ax	Whole Dataset	600	12647	1448	14095	11199	9918	672	10590	9246
I2ax	Roma	79	8643	2313	10956	6330	8211	1885	10096	6326
I2ax	Spanish Roma	10	11202	4641	15843	6561	16000	8178	24178	7822

Supplementary Table S5: Time to the most recent common ancestor (TMRCA) estimates for the four major haplogroups, calculated using the rho statistic method and assessed at three hierarchical levels.

Haplotype	STR Sequence	Frequency	Frequency (%)
SpanishRoma_ht33	12-23-14-10-13-16-11-14-11-17-17-14-20-10-15-9-21	21	12,14
SpanishRoma_ht10	12-22-15-10-15-17-11-14-11-16-18-14-19-12-15-9-20	8	4,62
SpanishRoma_ht38	12-23-14-10-13-16-11-14-11-18-17-14-20-10-15-9-21	7	4,05
SpanishRoma_ht43	12-23-14-10-13-16-12-14-11-17-17-14-20-10-15-9-21	7	4,05
SpanishRoma_ht50	12-23-15-10-15-17-11-14-11-16-18-14-19-12-15-9-20	7	4,05
SpanishRoma_ht25	12-22-16-10-13-15-11-13-11-19-15-14-21-12-15-9-21	5	2,89
SpanishRoma_ht24	12-22-16-10-13-15-11-13-11-18-15-14-21-12-15-9-21	4	2,31
SpanishRoma_ht18	12-22-15-10-15-17-12-14-11-16-17-14-19-11-15-9-20	3	1,73
SpanishRoma_ht39	12-23-14-10-13-16-11-15-11-17-17-14-20-10-15-9-21	3	1,73
SpanishRoma_ht40	12-23-14-10-13-16-11-15-11-17-18-14-20-10-15-9-21	3	1,73
SpanishRoma_ht88	13-24-17-10-14-15-12-12-11-18-18-15-19-12-15-10-23	3	1,73
SpanishRoma_ht91	13-25-13-11-11-14-12-14-13-16-18-15-19-11-16-12-23	3	1,73
SpanishRoma_ht96	14-23-15-9-13-16-13-14-12-18-15-14-20-11-13-10-21	3	1,73
SpanishRoma_ht4	12-22-15-10-15-16-11-14-11-16-17-14-19-12-15-9-20	2	1,16
SpanishRoma_ht8	12-22-15-10-15-17-11-14-11-16-17-14-18-12-15-9-20	2	1,16
SpanishRoma_ht11	12-22-15-10-15-17-11-14-11-16-18-14-19-12-15-9-21	2	1,16
SpanishRoma_ht16	12-22-15-10-15-17-11-15-11-16-17-14-19-12-15-9-20	2	1,16
SpanishRoma_ht26	12-22-16-10-13-16-11-13-11-19-15-14-21-12-15-9-21	2	1,16
SpanishRoma_ht31	12-23-14-10-13-16-11-14-11-16-17-14-20-10-15-9-21	2	1,16
SpanishRoma_ht47	12-23-14-10-16-99-12-13-11-16-16-15-21-12-15-9-21	2	1,16
SpanishRoma_ht54	12-23-15-11-14-19-11-12-11-18-17-15-20-12-14-9-23	2	1,16
SpanishRoma_ht55	12-23-15-9-13-16-12-13-11-16-14-14-21-11-15-9-22	2	1,16
SpanishRoma_ht81	13-24-14-11-11-15-12-13-13-16-18-15-19-12-16-12-23	2	1,16
SpanishRoma_ht84	13-24-14-11-12-14-12-13-13-16-17-15-19-12-15-12-23	2	1,16
SpanishRoma_ht97	14-24-15-11-11-14-11-14-12-16-17-14-18-11-16-12-24	2	1,16
SpanishRoma_ht1	10-23-14-10-13-16-12-14-11-16-17-14-20-10-15-9-21	1	0,58
SpanishRoma_ht2	12-22-14-10-13-16-11-14-11-17-17-14-20-10-15-9-21	1	0,58
SpanishRoma_ht3	12-22-14-10-13-16-12-13-11-17-16-15-20-11-15-9-21	1	0,58
SpanishRoma_ht5	12-22-15-10-15-16-11-14-11-16-18-14-19-12-15-9-20	1	0,58
SpanishRoma_ht6	12-22-15-10-15-16-11-14-99-16-18-14-19-12-15-9-20	1	0,58
SpanishRoma_ht7	12-22-15-10-15-16-11-15-11-16-20-14-19-12-15-9-20	1	0,58
SpanishRoma_ht9	12-22-15-10-15-17-11-14-11-16-17-14-19-12-15-9-20	1	0,58
SpanishRoma_ht12	12-22-15-10-15-17-11-14-11-16-18-14-19-12-16-7-20	1	0,58
SpanishRoma_ht13	12-22-15-10-15-17-11-14-11-16-18-14-20-12-15-9-20	1	0,58
SpanishRoma_ht14	12-22-15-10-15-17-11-14-11-16-18-15-19-12-15-9-20	1	0,58
SpanishRoma_ht15	12-22-15-10-15-17-11-14-11-16-20-14-19-12-15-9-20	1	0,58
SpanishRoma_ht17	12-22-15-10-15-17-11-15-11-16-18-14-19-12-16-9-20	1	0,58
SpanishRoma_ht19	12-22-15-10-15-17-12-14-11-16-17-14-20-11-15-9-20	1	0,58
SpanishRoma_ht20	12-22-15-10-16-17-11-14-11-16-17-14-19-12-15-9-20	1	0,58
SpanishRoma_ht21	12-22-15-10-16-17-11-14-11-16-17-14-20-12-15-9-20	1	0,58
SpanishRoma_ht22	12-22-15-10-16-99-11-14-11-16-17-14-19-12-16-9-20	1	0,58
SpanishRoma_ht23	12-22-15-11-12-14-12-14-11-15-18-14-19-12-15-9-20	1	0,58
SpanishRoma_ht27	12-22-16-10-15-18-12-14-11-16-17-14-19-11-15-99-20	1	0,58
SpanishRoma_ht28	12-22-17-10-13-15-12-13-11-19-15-14-21-12-15-9-21	1	0,58
SpanishRoma_ht29	12-23-14-10-12-16-12-14-11-17-17-14-20-10-15-9-21	1	0,58
SpanishRoma_ht30	12-23-14-10-13-16-11-14-11--14-17-14-20-10-15-9-21	1	0,58
SpanishRoma_ht32	12-23-14-10-13-16-11-14-11-17-17-14-20-10-15-9-20	1	0,58
SpanishRoma_ht34	12-23-14-10-13-16-11-14-11-17-17-14-20-10-15-9-23	1	0,58
SpanishRoma_ht35	12-23-14-10-13-16-11-14-11-17-17-14-20-10-16-9-21	1	0,58
SpanishRoma_ht36	12-23-14-10-13-16-11-14-11-17-17-14-20-9-15-9-21	1	0,58
SpanishRoma_ht37	12-23-14-10-13-16-11-14-11-17-18-14-20-10-15-9-21	1	0,58
SpanishRoma_ht41	12-23-14-10-13-16-11-99-11-32-17-14-20-10-15-9-21	1	0,58
SpanishRoma_ht42	12-23-14-10-13-16-12-14-11-16-17-14-20-10-15-9-21	1	0,58
SpanishRoma_ht44	12-23-14-10-13-16-99-14-11-17-17-14-20-10-15-9-21	1	0,58
SpanishRoma_ht45	12-23-14-10-14-17-12-13-11-16-17-14-20-11-15-10-20	1	0,58
SpanishRoma_ht46	12-23-14-10-14-17-12-13-11-16-17.2-14-20-11-15-10-20	1	0,58
SpanishRoma_ht48	12-23-14-12-13-19-11-13-11-17-18-14-20-11-14-10-21	1	0,58
SpanishRoma_ht49	12-23-15-10-15-17-11-14-11-16-17-14-19-13-15-9-20	1	0,58
SpanishRoma_ht51	12-23-15-10-15-17-11-14-11-16-18-14-19-12-15-9-21	1	0,58
SpanishRoma_ht52	12-23-15-10-15-17-11-14-11-16-19-14-19-13-15-9-20	1	0,58
SpanishRoma_ht53	12-23-15-10-15-99-11-14-11-16-17-14-19-13-15-9-20	1	0,58
SpanishRoma_ht56	12-24-15-10-13-17-12-12-11-18-16-16-19-11-13-9-22	1	0,58
SpanishRoma_ht57	12.1-23-13-10-10-13-11-13-11.1-16-16-14-18.1-11-14-11-23	1	0,58
SpanishRoma_ht58	13-22-15-10-15-16-11-14-11-16-17-14-19-12-15-9-20	1	0,58
SpanishRoma_ht59	13-23-13-10-13-16-11-14-13-16-14-15-19-11-15-9-21	1	0,58
SpanishRoma_ht60	13-23-14-11-11-14-12-13-13-16-17-15-19-11-15-12-23	1	0,58
SpanishRoma_ht61	13-23-14-11-11-14-12-13-13-17-16-15-19-13-17-12-23	1	0,58
SpanishRoma_ht62	13-23-14-11-11-15-11-13-13-17-17-15-19-12-16-12-23	1	0,58
SpanishRoma_ht63	13-23-14-12-11-14-12-13-13-18-16-15-19-13-17-12-23	1	0,58
SpanishRoma_ht64	13-23-14-12-11-14-12-14-13-16-16-14-19-12-16-12-23	1	0,58
SpanishRoma_ht65	13-23-15-10-15-17-11-14-11-16-18-14-19-12-15-9-20	1	0,58
SpanishRoma_ht66	13-23-15-11-11-14-12-13-13-16-17-14-18-10-15-12-23	1	0,58
SpanishRoma_ht67	13-23-16-10-14-16-13-13-11-16-17-15-20-11-15-10-22	1	0,58
SpanishRoma_ht68	13-23-16-10-14-16-13-13-11-17-17-15-20-11-15-10-22	1	0,58
SpanishRoma_ht69	13-23-16-10-14-16-99-13-11-17-17-15-20-11-15-10-22	1	0,58
SpanishRoma_ht70	13-23-17-10-12-12-12-14-11-16-17-15-20-12-14-10-21	1	0,58
SpanishRoma_ht71	13-23-17-10-12-99-12-14-11-16-17-15-20-12-14-10-21	1	0,58
SpanishRoma_ht72	13-24-14-10-11-14-11-13-14-16-18-15-19-12-16-12-23	1	0,58
SpanishRoma_ht73	13-24-14-10-11-14-12-13-13-16-18-15-19-12-15-13-23	1	0,58
SpanishRoma_ht74	13-24-14-10-11-14-12-13-14-15-17-15-19-12-15-12-23	1	0,58
SpanishRoma_ht75	13-24-14-10-11-14-13-13-13-16-18-15-19-11-16-12-23	1	0,58
SpanishRoma_ht76	13-24-14-10-9-14-12-14-13-15-17-15-19-12-15-12-24	1	0,58
SpanishRoma_ht77	13-24-14-11-11-14-12-13-13-17-17-15-19-12-15-12-23	1	0,58
SpanishRoma_ht78	13-24-14-11-11-14-12-14-13-16-17-15-19-12-15-12-24	1	0,58
SpanishRoma_ht79	13-24-14-11-11-14-12-14-13-17-20-14-19-12-16-12-23	1	0,58
SpanishRoma_ht80	13-24-14-11-11-14-14-14-13-16-18-15-20-12-15-12-23	1	0,58
SpanishRoma_ht82	13-24-14-11-11-16-12-13-13-16-18-15-19-12-16-12-23	1	0,58
SpanishRoma_ht83	13-24-14-11-12-14-11-13-13-16-17-15-19-12-15-12-23	1	0,58
SpanishRoma_ht85	13-24-14-11-12-14-12-13-13-16-17-15-19-12-16-12-23	1	0,58
SpanishRoma_ht86	13-24-14-13-14-99-13-14-13-16-19-15-18-12-15-12-23	1	0,58
SpanishRoma_ht87	13-24-15-10-10-14-12-13-13-16-19-15-19-12-15-12-23	1	0,58
SpanishRoma_ht89	13-24-17-10-14-15-99-12-11-18-18-15-19-12-15-10-23	1	0,58
SpanishRoma_ht90	13-25-13-10-11-14-12-14-13-16-18-14-19-11-16-12-23	1	0,58
SpanishRoma_ht92	13-25-14-11-11-14-11-13-13-16-17-15-19-12-16-12-23	1	0,58
SpanishRoma_ht93	13-25-14.1-11-11-15-12-15-13-15-17.1-16-19-12-15-12-24	1	0,58
SpanishRoma_ht94	13-25-15-10-11-99-10-13-11-19-16-14-20-12-16-11-23	1	0,58
SpanishRoma_ht95	14-22-15-11-99-14-10-12-11-16-18-15-20-12-14-10-21	1	0,58

Supplementary Table S6: Y-STR haplotypes for 173 Spanish Roma individuals, including both newly genotyped and previously genotyped individuals.

Haplotype	STR Sequence	Frequency	Frequency (%)
spanish_ht190	13-24-14-10-11-14-12-13-13-16-17-14-18-12-16-12-23	3	0,66
spanish_ht267	13-24-14-11-11-14-12-13-13-16-17-15-19-12-15-12-23	3	0,66
spanish_ht321	13-24-14-11-12-14-11-13-13-16-18-15-19-12-15-12-23	3	0,66
spanish_ht100	13-23-14-10-13-14-12-12-11-16-16-16-21-11-14-10-22	2	0,44
spanish_ht192	13-24-14-10-11-14-12-13-13-16-17-15-19-12-15-12-23	2	0,44
spanish_ht231	13-24-14-11-11-13-13-14-13-16-17-15-19-12-15-12-23	2	0,44
spanish_ht258	13-24-14-11-11-14-12-13-13-16-16-15-18-11-17-12-24	2	0,44
spanish_ht261	13-24-14-11-11-14-12-13-13-16-16-15-19-12-16-12-23	2	0,44
spanish_ht273	13-24-14-11-11-14-12-13-13-16-18-15-19-12-16-12-23	2	0,44
spanish_ht277	13-24-14-11-11-14-12-13-13-17-16-15-19-12-16-12-23	2	0,44
spanish_ht307	13-24-14-11-11-14-14-14-13-16-17-15-19-12-15-12-23	2	0,44
spanish_ht322	13-24-14-11-12-14-11-13-13-17-18-15-99-12-15-12-23	2	0,44
spanish_ht324	13-24-14-11-12-14-12-13-13-15-16-15-19-12-15-12-23	2	0,44
spanish_ht326	13-24-14-11-12-14-12-13-13-16-17-15-19-12-15-12-23	2	0,44
spanish_ht327	13-24-14-11-12-14-12-13-13-17-18-14-18-12-16-12-23	2	0,44
spanish_ht339	13-24-14-12-11-14-11-13-13-16-16-15-19-12-15-12-23	2	0,44
spanish_ht388	13-25-14-11-11-14-13-14-13-16-17-15-19-12-16-12-23	2	0,44
spanish_ht1	10-24-14-11-11-14-12-13-13-16-18-15-20-12-17-12-23	1	0,22
spanish_ht2	10-30-16-12-17-14-21-22-11-99-15-10-16-15-14-20-12	1	0,22
spanish_ht3	12-22-14-10-11-13-14-13-13-16-15-15-19-12-14-12-23	1	0,22
spanish_ht4	12-22-14-10-14-16-11-13-11-17-15-14-20-12-15-10-23	1	0,22
spanish_ht5	12-22-15-10-12-15-11-13-11-16-16-14-22-11-16-9-21	1	0,22
spanish_ht6	12-22-15-10-12-16-11-13-11-16-16-14-22-11-16-9-22	1	0,22
spanish_ht7	12-22-15-10-14-16-12-12-11-16-17-16-22-11-16-10-21	1	0,22
spanish_ht8	12-23-14-10-12-13-12-13-11-17-18-14-20-11-15-10-20	1	0,22
spanish_ht9	12-23-14-10-12-17-11-13-11-16-182-14-21-11-15-10-21	1	0,22
spanish_ht10	12-23-14-10-13-16-10-12-11-16-15-14-20-11-16-9-21	1	0,22
spanish_ht11	12-23-14-10-13-16-10-13-11-16-15-14-20-12-18-9-21	1	0,22
spanish_ht12	12-23-14-10-13-16-12-14-11-17-17-14-20-10-15-9-21	1	0,22
spanish_ht13	12-23-14-10-13-17-12-12-11-18-16-15-20-11-14-9-22	1	0,22
spanish_ht14	12-23-14-10-14-15-12-13-11-17-17-16-20-13-17-9-22	1	0,22
spanish_ht15	12-23-14-10-14-17-12-13-11-16-16-15-21-12-16-9-23	1	0,22
spanish_ht16	12-23-14-10-14-18-11-13-11-16-15-16-19-10-15-9-20	1	0,22
spanish_ht17	12-23-14-10-16-20-11-13-14-16-18-16-19-11-15-10-23	1	0,22
spanish_ht18	12-23-14-11-14-17-11-13-11-17-18-15-20-12-15-9-24	1	0,22
spanish_ht19	12-23-14-11-14-18-12-12-11-16-15-15-20-11-13-9-21	1	0,22
spanish_ht20	12-23-15-10-12-17-11-13-11-17-15-16-21-12-16-9-24	1	0,22
spanish_ht21	12-23-15-10-13-16-11-13-11-16-14-14-21-12-15-9-21	1	0,22
spanish_ht22	12-23-15-10-13-18-11-14-11-18-99-14-20-10-15-10-20	1	0,22
spanish_ht23	12-23-15-11-14-16-12-12-11-16-18-16-20-11-13-9-21	1	0,22
spanish_ht24	12-23-15-9-13-16-12-14-11-17-14-14-21-12-15-9-21	1	0,22
spanish_ht25	12-23-15-9-13-16-12-14-11-17-15-14-21-12-16-9-21	1	0,22
spanish_ht26	12-23-16-10-15-17-12-12-11-16-19-16-19-12-13-9-21	1	0,22
spanish_ht27	12-24-13-10-17-18-11-12-11-17-14-14-20-11-15-10-22	1	0,22
spanish_ht28	12-24-13-10-17-18-11-12-11-17-15-14-20-11-15-10-22	1	0,22
spanish_ht29	12-24-13-11-11-11-12-13-13-16-16-15-19-13-15-13-23	1	0,22
spanish_ht30	12-24-14-10-11-13-13-13-13-16-17-15-19-12-15-12-23	1	0,22
spanish_ht31	12-24-14-10-11-14-11-13-13-16-18-15-19-11-16-12-23	1	0,22
spanish_ht32	12-24-14-10-11-14-12-13-13-16-16-15-19-13-15-12-23	1	0,22
spanish_ht33	12-24-14-10-11-14-12-13-13-16-17-15-19-11-16-12-23	1	0,22
spanish_ht34	12-24-14-10-11-14-12-13-13-16-17-15-19-12-15-12-23	1	0,22
spanish_ht35	12-24-14-10-11-14-12-13-13-16-18-14-19-11-15-12-23	1	0,22
spanish_ht36	12-24-14-10-12-12-11-14-11-16-19-14-20-11-15-10-21	1	0,22
spanish_ht37	12-24-14-10-12-16-11-13-11-17-15-14-21-11-16-9-24	1	0,22
spanish_ht38	12-24-14-10-13-16-11-13-11-17-16-15-20-11-15-9-22	1	0,22
spanish_ht39	12-24-14-10-13-17-11-12-11-17-14-15-20-11-15-9-23	1	0,22
spanish_ht40	12-24-14-10-13-18-13-13-11-17-182-14-20-11-15-10-22	1	0,22
spanish_ht41	12-24-14-10-14-16-12-14-11-17-18-15-21-12-17-9-23	1	0,22
spanish_ht42	12-24-14-10-14-17-11-12-11-17-14-15-20-11-15-10-24	1	0,22
spanish_ht43	12-24-14-11-11-14-11-14-13-17-17-15-19-12-15-12-23	1	0,22
spanish_ht44	12-24-14-11-11-14-12-13-11-14-18-15-19-12-15-12-23	1	0,22
spanish_ht45	12-24-14-11-11-14-12-13-13-16-16-15-19-12-16-12-23	1	0,22
spanish_ht46	12-24-14-11-11-14-12-13-13-16-16-15-20-13-15-12-23	1	0,22
spanish_ht47	12-24-14-11-11-14-12-13-13-16-17-15-19-12-16-12-23	1	0,22
spanish_ht48	12-24-14-11-11-14-12-13-13-16-17-15-20-13-15-12-23	1	0,22
spanish_ht49	12-24-14-11-11-14-13-13-13-16-18-15-19-12-16-12-23	1	0,22
spanish_ht50	12-24-14-11-11-15-13-13-13-16-17-15-19-13-16-12-23	1	0,22
spanish_ht51	12-24-14-11-11-16-12-13-15-15-19-15-19-12-15-12-23	1	0,22
spanish_ht52	12-24-14-11-13-14-12-13-13-16-17-14-18-11-15-12-23	1	0,22
spanish_ht53	12-24-14-13-11-15-12-13-13-14-16-15-19-12-16-12-23	1	0,22
spanish_ht54	12-24-14-9-11-16-13-14-11-18-17-14-18-12-14-10-22	1	0,22
spanish_ht55	12-24-15-10-13-16-11-12-11-17-16-15-21-12-14-9-21	1	0,22
spanish_ht56	12-24-15-10-13-17-12-12-11-16-16-16-19-11-13-9-22	1	0,22
spanish_ht57	12-24-15-10-14-16-13-12-11-16-16-16-19-11-13-9-21	1	0,22
spanish_ht58	12-24-15-10-15-18-12-12-11-16-15-16-19-12-13-9-21	1	0,22
spanish_ht59	12-24-15-9-13-16-13-14-11-16-14-14-21-12-16-9-22	1	0,22
spanish_ht60	12-25-13-10-11-13-13-13-13-16-17-15-19-12-15-12-23	1	0,22
spanish_ht61	12-25-14-10-11-13-12-13-13-16-17-15-19-12-15-12-23	1	0,22
spanish_ht62	12-25-14-11-14-15-12-14-11-16-19-16-20-11-13-9-21	1	0,22
spanish_ht63	12-25-15-10-14-16-11-13-11-17-18-15-19-11-14-9-20	1	0,22
spanish_ht64	12-25-15-10-14-19-11-12-11-16-16-14-19-11-13-9-22	1	0,22
spanish_ht65	12-25-16-11-11-13-12-13-13-16-16-15-19-12-16-12-23	1	0,22
spanish_ht66	13-20-14-10-11-14-11-14-13-16-17-15-19-12-15-12-23	1	0,22
spanish_ht67	13-21-15-9-12-14-12-12-11-17-18-16-20-11-15-10-22	1	0,22
spanish_ht68	13-22-14-10-13-14-11-12-11-16-16-16-20-11-13-10-21	1	0,22
spanish_ht69	13-22-14-10-13-14-12-13-11-16-14-15-20-11-14-10-21	1	0,22
spanish_ht70	13-22-14-10-14-14-11-12-11-16-15-16-20-11-14-10-23	1	0,22
spanish_ht71	13-22-14-10-14-15-11-12-11-16-15-16-20-11-14-10-24	1	0,22
spanish_ht72	13-22-14-10-15-15-11-12-11-17-15-16-20-11-14-10-21	1	0,22
spanish_ht73	13-22-14-11-11-14-11-13-13-16-16-15-99-12-16-12-23	1	0,22
spanish_ht74	13-22-14-11-11-15-12-13-13-16-18-15-20-12-16-12-23	1	0,22
spanish_ht75	13-22-14-11-12-15-12-12-13-16-17-15-19-12-16-12-22	1	0,22
spanish_ht76	13-22-15-10-12-14-11-12-11-17-18-16-21-11-15-10-21	1	0,22
spanish_ht77	13-22-15-10-13-14-11-12-11-16-15-16-20-11-15-8-20	1	0,22
spanish_ht78	13-22-15-10-14-15-11-12-11-17-16-16-22-11-15-10-21	1	0,22
spanish_ht79	13-22-15-11-12-14-12-13-11-16-18-16-20-11-14-10-22	1	0,22
spanish_ht80	13-23-13-10-11-11-13-14-13-16-18-15-19-12-17-12-23	1	0,22
spanish_ht81	13-23-13-10-13-15-12-13-13-15-17-14-19-11-16-9-21	1	0,22
spanish_ht82	13-23-13-10-15-16-12-13-11-17-16-14-20-11-18-10-24	1	0,22
spanish_ht83	13-23-13-9-14-15-10-14-11-16-19-14-20-12-15-10-21	1	0,22
spanish_ht84	13-23-14-10-11-11-13-15-13-16-19-15-19-12-17-12-23	1	0,22
spanish_ht85	13-23-14-10-11-12-12-14-13-16-18-15-19-12-17-12-23	1	0,22
spanish_ht86	13-23-14-10-11-13-13-14-13-15-18-15-19-12-17-12-23	1	0,22
spanish_ht87	13-23-14-10-11-14-11-13-13-16-18-14-19-12-16-12-23	1	0,22
spanish_ht88	13-23-14-10-11-14-11-13-13-16-18-15-19-12-16-12-23	1	0,22
spanish_ht89	13-23-14-10-11-14-11-13-13-17-18-11-18-11-17-12-24	1	0,22
spanish_ht90	13-23-14-10-11-14-12-13-13-15-15-15-18-12-16-12-23	1	0,22
spanish_ht91	13-23-14-10-11-14-12-13-13-16-16-14-18-12-16-12-23	1	0,22
spanish_ht92	13-23-14-10-11-14-12-13-13-16-18-15-19-12-15-12-23	1	0,22
spanish_ht93	13-23-14-10-11-14-13-13-13-16-17-15-18-12-16-12-23	1	0,22
spanish_ht94	13-23-14-10-11-15-12-13-13-16-17-14-18-11-15-12-23	1	0,22
spanish_ht95	13-23-14-10-11-15-13-13-13-16-16-15-19-12-15-12-24	1	0,22
spanish_ht96	13-23-14-10-11-15-13-13-13-17-17-15-18-12-16-12-23	1	0,22
spanish_ht97	13-23-14-10-12-14-12-12-11-16-16-16-21-11-14-10-21	1	0,22
spanish_ht98	13-23-14-10-13-14-11-12-11-16-16-16-20-11-14-10-21	1	0,22
spanish_ht99	13-23-14-10-13-14-12-12-11-16-15-16-21-11-14-10-22	1	0,22

spanish_ht101	13-23-14-10-13-14-12-13-11-16-15-14-20-11-15-10-24	1	0,22
spanish_ht102	13-23-14-10-14-14-12-12-11-17-15-16-20-11-13-10-22	1	0,22
spanish_ht103	13-23-14-11-11-13-12-14-13-16-16-15-19-12-16-12-23	1	0,22
spanish_ht104	13-23-14-11-11-14-11-13-13-16-16-15-19-12-16-12-23	1	0,22
spanish_ht105	13-23-14-11-11-14-11-13-13-16-18-15-19-12-16-12-23	1	0,22
spanish_ht106	13-23-14-11-11-14-11-13-13-17-17-14-18-12-16-12-23	1	0,22
spanish_ht107	13-23-14-11-11-14-11-13-14-16-16-14-18-11-15-12-23	1	0,22
spanish_ht108	13-23-14-11-11-14-11-14-13-15-16-15-19-12-15-12-23	1	0,22
spanish_ht109	13-23-14-11-11-14-12-13-13-13-17-14-18-11-17-13-23	1	0,22
spanish_ht110	13-23-14-11-11-14-12-13-13-16-17-14-18-11-15-12-23	1	0,22
spanish_ht111	13-23-14-11-11-14-12-14-13-16-18-15-19-12-16-12-25	1	0,22
spanish_ht112	13-23-14-11-11-14-13-13-13-16-18-16-19-11-14-12-25	1	0,22
spanish_ht113	13-23-14-11-11-14-13-13-13-16-19-15-18-12-15-12-23	1	0,22
spanish_ht114	13-23-14-11-11-14-13-13-13-17-17-15-19-12-17-12-23	1	0,22
spanish_ht115	13-23-14-11-11-15-12-13-13-16-17-15-20-12-17-13-23	1	0,22
spanish_ht116	13-23-14-11-11-15-13-14-14-17-17-16-20-12-16-12-23	1	0,22
spanish_ht117	13-23-14-11-11-15-14-13-13-16-17-15-19-12-17-12-23	1	0,22
spanish_ht118	13-23-14-11-12-13-12-13-13-17-18-14-18-10-15-12-23	1	0,22
spanish_ht119	13-23-14-11-12-14-10-13-13-16-17-15-19-12-16-12-23	1	0,22
spanish_ht120	13-23-14-11-13-14-11-12-11-16-15-16-19-11-14-10-22	1	0,22
spanish_ht121	13-23-14-12-11-13-12-13-13-16-18-15-19-12-16-12-23	1	0,22
spanish_ht122	13-23-14-12-11-14-11-13-13-16-17-15-19-12-15-12-23	1	0,22
spanish_ht123	13-23-14-12-11-14-12-13-13-15-16-15-19-12-15-12-23	1	0,22
spanish_ht124	13-23-14-9-12-17-12-14-11-16-15-14-21-12-14-10-21	1	0,22
spanish_ht125	13-23-15-10-10-14-12-13-13-17-18-15-18-12-16-12-25	1	0,22
spanish_ht126	13-23-15-10-11-14-12-13-13-17-18-15-18-12-16-12-25	1	0,22
spanish_ht127	13-23-15-10-11-14-12-14-13-16-16-15-19-12-16-12-23	1	0,22
spanish_ht128	13-23-15-10-12-18-11-12-11-18-18-16-20-11-15-10-22	1	0,22
spanish_ht129	13-23-15-10-12-18-11-13-11-16-17-14-20-10-15-10-22	1	0,22
spanish_ht130	13-23-15-10-14-14-11-12-11-17-18-16-21-11-14-10-21	1	0,22
spanish_ht131	13-23-15-10-14-16-11-13-13-15-16-14-19-11-15-9-21	1	0,22
spanish_ht132	13-23-15-10-14-16-11-13-13-16-18-14-19-11-15-9-21	1	0,22
spanish_ht133	13-23-15-10-14-16-11-13-13-17-17-14-19-11-16-9-21	1	0,22
spanish_ht134	13-23-15-10-14-16-11-13-13-17-18-14-19-11-15-9-21	1	0,22
spanish_ht135	13-23-15-11-11-12-13-12-14-17-17-14-18-11-17-12-23	1	0,22
spanish_ht136	13-23-15-11-11-13-11-14-11-16-18-16-20-11-15-10-21	1	0,22
spanish_ht137	13-23-15-11-11-14-12-14-13-17-17-15-18-12-15-12-24	1	0,22
spanish_ht138	13-23-15-11-11-14-14-13-13-16-17-14-18-11-15-12-23	1	0,22
spanish_ht139	13-23-15-11-11-15-13-13-13-16-16-14-18-11-15-12-23	1	0,22
spanish_ht140	13-23-15-11-12-12-12-14-12-15-17-15-21-13-15-10-24	1	0,22
spanish_ht141	13-23-15-11-16-17-10-13-15-17-15-14-19-11-15-9-21	1	0,22
spanish_ht142	13-23-16-10-12-12-11-13-11-15-16-15-21-12-14-10-22	1	0,22
spanish_ht143	13-23-16-10-12-13-11-13-11-15-18-15-21-13-14-10-22	1	0,22
spanish_ht144	13-23-16-10-14-15-11-13-12-17-16-16-22-11-15-10-21	1	0,22
spanish_ht145	13-23-16-11-11-14-12-13-13-16-19-15-19-12-16-12-22	1	0,22
spanish_ht146	13-23-17-10-12-12-11-13-11-15-16-15-21-12-14-10-22	1	0,22
spanish_ht147	13-23-17-10-12-12-11-13-11-15-17-15-21-12-14-10-21	1	0,22
spanish_ht148	13-23-17-10-12-12-11-13-11-15-17-15-21-12-14-10-22	1	0,22
spanish_ht149	13-23-17-10-12-12-11-13-11-15-17-15-22-12-14-10-22	1	0,22
spanish_ht150	13-23-17-10-12-12-11-13-11-15-18-15-21-11-14-10-22	1	0,22
spanish_ht151	13-23-17-11-12-12-11-13-11-15-18-15-21-12-14-10-22	1	0,22
spanish_ht152	13-23-17-9-12-12-13-13-11-15-16-15-21-11-14-10-21	1	0,22
spanish_ht153	13-24-12-11-10-12-11-13-13-16-18-15-19-12-16-12-23	1	0,22
spanish_ht154	13-24-13-10-11-14-12-13-13-16-17-16-19-12-15-12-23	1	0,22
spanish_ht155	13-24-13-10-14-14-10-14-11-16-18-14-20-12-15-10-21	1	0,22
spanish_ht156	13-24-13-10-14-19-13-14-11-18-15-14-19-12-17-9-23	1	0,22
spanish_ht157	13-24-13-10-15-18-12-13-11-17-17-14-20-12-17-10-22	1	0,22
spanish_ht158	13-24-13-10-16-16-12-12-10-18-18-14-20-9-15-10-22	1	0,22
spanish_ht159	13-24-13-10-17-17-10-12-11-17-14-14-20-11-15-10-21	1	0,22
spanish_ht160	13-24-13-10-17-18-13-12-11-19-17-14-20-11-17-10-24	1	0,22
spanish_ht161	13-24-13-11-11-14-11-13-13-16-17-15-19-12-16-12-23	1	0,22
spanish_ht162	13-24-13-11-11-14-12-13-13-17-18-15-19-12-15-12-23	1	0,22
spanish_ht163	13-24-13-11-11-14-12-14-13-16-17-14-18-11-17-12-23	1	0,22
spanish_ht164	13-24-13-11-11-15-12-13-13-16-17-15-19-12-16-12-23	1	0,22
spanish_ht165	13-24-13-11-12-14-12-13-13-16-17-15-19-11-15-12-23	1	0,22
spanish_ht166	13-24-13-11-12-14-12-13-13-16-17-15-19-12-15-12-23	1	0,22
spanish_ht167	13-24-13-11-12-14-13-13-13-16-18-15-19-11-16-12-23	1	0,22
spanish_ht168	13-24-13-11-16-18-11-13-11-17-15-14-20-12-15-10-23	1	0,22
spanish_ht169	13-24-13-9-13-14-10-13-11-16-18-14-20-12-15-10-21	1	0,22
spanish_ht170	13-24-13-9-13-14-10-14-11-16-18-14-20-11-15-10-21	1	0,22
spanish_ht171	13-24-13-9-13-14-10-14-11-16-18-14-20-12-15-10-21	1	0,22
spanish_ht172	13-24-13-9-13-15-10-14-11-16-17-14-20-11-16-10-22	1	0,22
spanish_ht173	13-24-13-9-13-16-10-14-12-16-16-14-20-12-16-10-21	1	0,22
spanish_ht174	13-24-14-10-10-15-11-13-13-16-17-15-19-12-15-12-23	1	0,22
spanish_ht175	13-24-14-10-11-11-12-13-13-16-17-15-20-12-16-12-24	1	0,22
spanish_ht176	13-24-14-10-11-13-11-14-13-16-17-15-19-12-15-12-23	1	0,22
spanish_ht177	13-24-14-10-11-13-12-13-13-16-16-15-19-12-16-12-23	1	0,22
spanish_ht178	13-24-14-10-11-13-12-14-13-16-17-14-19-12-15-12-23	1	0,22
spanish_ht179	13-24-14-10-11-13-13-13-13-16-17-15-20-12-15-12-23	1	0,22
spanish_ht180	13-24-14-10-11-14-11-12-13-16-16-15-99-12-15-12-24	1	0,22
spanish_ht181	13-24-14-10-11-14-11-13-13-16-17-15-19-11-15-12-23	1	0,22
spanish_ht182	13-24-14-10-11-14-11-13-13-16-17-15-20-12-15-12-23	1	0,22
spanish_ht183	13-24-14-10-11-14-11-13-13-17-16-15-19-12-17-12-23	1	0,22
spanish_ht184	13-24-14-10-11-14-12-12-13-16-19-99-99-12-15-12-23	1	0,22
spanish_ht185	13-24-14-10-11-14-12-13-13-15-99-15-19-12-16-12-23	1	0,22
spanish_ht186	13-24-14-10-11-14-12-13-13-16-14-15-19-12-15-12-22	1	0,22
spanish_ht187	13-24-14-10-11-14-12-13-13-16-16-15-19-11-15-12-23	1	0,22
spanish_ht188	13-24-14-10-11-14-12-13-13-16-16-15-19-12-15-12-24	1	0,22
spanish_ht189	13-24-14-10-11-14-12-13-13-16-17-14-18-12-15-12-23	1	0,22
spanish_ht191	13-24-14-10-11-14-12-13-13-16-17-15-19-11-15-12-23	1	0,22
spanish_ht193	13-24-14-10-11-14-12-13-13-16-18-15-19-12-15-12-23	1	0,22
spanish_ht194	13-24-14-10-11-14-12-13-13-16-18-15-19-12-16-12-23	1	0,22
spanish_ht195	13-24-14-10-11-14-12-13-13-17-16-15-18-12-15-12-23	1	0,22
spanish_ht196	13-24-14-10-11-14-12-13-13-17-18-14-18-11-16-12-23	1	0,22
spanish_ht197	13-24-14-10-11-14-12-13-14-16-16-15-19-12-16-12-23	1	0,22
spanish_ht198	13-24-14-10-11-14-12-14-13-16-17-15-19-11-16-12-23	1	0,22
spanish_ht199	13-24-14-10-11-14-12-14-13-16-18-14-18-11-15-12-23	1	0,22
spanish_ht200	13-24-14-10-11-14-12-14-13-16-18-14-18-11-15-12-24	1	0,22
spanish_ht201	13-24-14-10-11-14-12-14-13-17-17-14-18-11-15-12-24	1	0,22
spanish_ht202	13-24-14-10-11-14-12-14-13-17-17-14-18-11-16-12-23	1	0,22
spanish_ht203	13-24-14-10-11-14-13-13-13-15-17-14-18-11-16-12-23	1	0,22

spanish_ht206	13-24-14-10-11-14-13-14-13-17-18-14-18-11-16-12-23	1	0,22
spanish_ht207	13-24-14-10-11-15-12-13-13-16-18-15-19-12-15-12-23	1	0,22
spanish_ht208	13-24-14-10-11-15-13-13-13-16-16-15-20-13-15-13-23	1	0,22
spanish_ht209	13-24-14-10-12-13-11-13-11-16-17-15-19-12-15-12-23	1	0,22
spanish_ht210	13-24-14-10-12-14-11-13-13-16-17-15-19-11-16-12-23	1	0,22
spanish_ht211	13-24-14-10-12-14-11-13-13-16-18-15-19-12-16-12-24	1	0,22
spanish_ht212	13-24-14-10-12-14-11-14-13-15-17-15-19-12-15-12-23	1	0,22
spanish_ht213	13-24-14-10-12-14-12-13-13-16-17-14-18-12-15-12-23	1	0,22
spanish_ht214	13-24-14-10-12-14-12-13-13-16-17-14-18-12-16-12-23	1	0,22
spanish_ht215	13-24-14-10-13-13-11-13-13-18-17-15-19-12-16-12-24	1	0,22
spanish_ht216	13-24-14-10-14-15-12-14-13-16-16-14-19-11-16-9-23	1	0,22
spanish_ht217	13-24-14-10-16-18-12-13-11-17-15-14-19-12-17-11-21	1	0,22
spanish_ht218	13-24-14-10-17-18-12-13-11-18-16-14-20-12-15-10-21	1	0,22
spanish_ht219	13-24-14-10-18-21-12-13-11-19-18-14-19-13-16-10-20	1	0,22
spanish_ht220	13-24-14-11-10-14-12-13-12-16-15-15-19-12-16-12-23	1	0,22
spanish_ht221	13-24-14-11-10-14-12-13-13-16-17-16-19-12-16-11-23	1	0,22
spanish_ht222	13-24-14-11-10-14-13-13-13-16-16-15-19-12-17-12-23	1	0,22
spanish_ht223	13-24-14-11-10-14-14-13-13-16-18-15-19-12-16-12-24	1	0,22
spanish_ht224	13-24-14-11-11-13-11-13-13-16-17-14-18-11-15-11-23	1	0,22
spanish_ht225	13-24-14-11-11-13-11-13-13-16-17-14-18-11-16-12-23	1	0,22
spanish_ht226	13-24-14-11-11-13-11-13-13-16-17-15-19-11-15-12-23	1	0,22
spanish_ht227	13-24-14-11-11-13-12-12-13-16-16-15-19-13-16-12-23	1	0,22
spanish_ht228	13-24-14-11-11-13-12-13-13-16-19-14-18-12-15-12-23	1	0,22
spanish_ht229	13-24-14-11-11-13-13-13-13-16-17-15-19-12-15-12-23	1	0,22
spanish_ht230	13-24-14-11-11-13-13-13-13-17-18-15-18-12-17-13-23	1	0,22
spanish_ht232	13-24-14-11-11-13-13-14-13-17-17-14-18-11-15-12-23	1	0,22
spanish_ht233	13-24-14-11-11-13-14-13-13-16-17-15-19-12-17-13-24	1	0,22
spanish_ht234	13-24-14-11-11-14-11-12-13-16-17-14-18-11-16-12-23	1	0,22
spanish_ht235	13-24-14-11-11-14-11-12-13-16-18-15-19-11-17-12-23	1	0,22
spanish_ht236	13-24-14-11-11-14-11-13-13-16-16-15-19-12-16-12-23	1	0,22
spanish_ht237	13-24-14-11-11-14-11-13-13-16-17-14-18-11-16-12-24	1	0,22
spanish_ht238	13-24-14-11-11-14-11-13-13-16-17-14-18-12-15-12-23	1	0,22
spanish_ht239	13-24-14-11-11-14-11-13-13-16-17-15-19-13-15-12-23	1	0,22
spanish_ht240	13-24-14-11-11-14-11-13-13-16-18-15-19-12-15-12-23	1	0,22
spanish_ht241	13-24-14-11-11-14-11-13-13-16-18-15-20-12-15-12-23	1	0,22
spanish_ht242	13-24-14-11-11-14-11-13-14-16-17-15-19-12-15-12-24	1	0,22
spanish_ht243	13-24-14-11-11-14-11-14-13-16-16-15-19-12-15-12-24	1	0,22
spanish_ht244	13-24-14-11-11-14-11-14-13-16-17-14-18-11-16-12-23	1	0,22
spanish_ht245	13-24-14-11-11-14-11-14-13-16-17-15-17-12-16-12-23	1	0,22
spanish_ht246	13-24-14-11-11-14-11-14-13-16-18-15-19-12-15-12-23	1	0,22
spanish_ht247	13-24-14-11-11-14-11-14-13-16-19-14-19-11-15-12-23	1	0,22
spanish_ht248	13-24-14-11-11-14-11-14-13-16-19-15-19-11-15-12-23	1	0,22
spanish_ht249	13-24-14-11-11-14-11-14-13-17-17-14-18-11-15-12-23	1	0,22
spanish_ht250	13-24-14-11-11-14-11-14-13-17-19-15-19-12-17-12-23	1	0,22
spanish_ht251	13-24-14-11-11-14-11-14-14-16-18-15-19-12-15-12-25	1	0,22
spanish_ht252	13-24-14-11-11-14-11-15-13-16-15-14-18-12-16-12-23	1	0,22
spanish_ht253	13-24-14-11-11-14-11-15-13-16-17-15-19-11-15-12-23	1	0,22
spanish_ht254	13-24-14-11-11-14-12-12-13-16-17-15-19-12-16-12-23	1	0,22
spanish_ht255	13-24-14-11-11-14-12-13-12-16-16-15-19-12-15-12-23	1	0,22
spanish_ht256	13-24-14-11-11-14-12-13-12-16-18-15-19-11-16-12-23	1	0,22
spanish_ht257	13-24-14-11-11-14-12-13-13-16-15-15-18-12-16-12-23	1	0,22
spanish_ht259	13-24-14-11-11-14-12-13-13-16-16-15-19-11-15-12-23	1	0,22
spanish_ht260	13-24-14-11-11-14-12-13-13-16-16-15-19-12-15-12-23	1	0,22
spanish_ht262	13-24-14-11-11-14-12-13-13-16-17-15-18-12-16-12-23	1	0,22
spanish_ht263	13-24-14-11-11-14-12-13-13-16-17-15-19-11-15-12-23	1	0,22
spanish_ht264	13-24-14-11-11-14-12-13-13-16-17-15-19-11-16-12-23	1	0,22
spanish_ht265	13-24-14-11-11-14-12-13-13-16-17-15-19-11-16-12-24	1	0,22
spanish_ht266	13-24-14-11-11-14-12-13-13-16-17-15-19-11-17-12-23	1	0,22
spanish_ht268	13-24-14-11-11-14-12-13-13-16-17-15-19-12-16-12-23	1	0,22
spanish_ht269	13-24-14-11-11-14-12-13-13-16-17-15-19-13-15-12-23	1	0,22
spanish_ht270	13-24-14-11-11-14-12-13-13-16-17-15-19-13-16-12-23	1	0,22
spanish_ht271	13-24-14-11-11-14-12-13-13-16-18-15-18-12-15-12-23	1	0,22
spanish_ht272	13-24-14-11-11-14-12-13-13-16-18-15-19-12-15-12-23	1	0,22
spanish_ht274	13-24-14-11-11-14-12-13-13-16-18-15-19-13-15-12-23	1	0,22
spanish_ht275	13-24-14-11-11-14-12-13-13-16-19-15-19-12-16-12-23	1	0,22
spanish_ht276	13-24-14-11-11-14-12-13-13-17-15-15-19-11-15-12-24	1	0,22
spanish_ht278	13-24-14-11-11-14-12-13-13-17-17-14-18-12-16-12-23	1	0,22
spanish_ht279	13-24-14-11-11-14-12-14-12-17-16-14-18-11-16-12-23	1	0,22
spanish_ht280	13-24-14-11-11-14-12-14-13-13-18-15-19-12-15-12-23	1	0,22
spanish_ht281	13-24-14-11-11-14-12-14-13-15-16-14-18-11-16-12-23	1	0,22
spanish_ht282	13-24-14-11-11-14-12-14-13-15-17-15-19-12-15-12-23	1	0,22
spanish_ht283	13-24-14-11-11-14-12-14-13-16-17-14-18-10-15-13-23	1	0,22
spanish_ht284	13-24-14-11-11-14-12-14-13-16-17-14-18-11-15-12-23	1	0,22
spanish_ht285	13-24-14-11-11-14-12-14-13-16-17-14-18-11-15-13-23	1	0,22
spanish_ht286	13-24-14-11-11-14-12-14-13-16-17-15-19-12-15-12-23	1	0,22
spanish_ht287	13-24-14-11-11-14-12-14-13-16-18-14-18-11-16-12-23	1	0,22
spanish_ht288	13-24-14-11-11-14-12-14-13-16-18-14-19-12-15-12-23	1	0,22
spanish_ht289	13-24-14-11-11-14-12-14-13-17-16-14-18-11-15-12-23	1	0,22
spanish_ht290	13-24-14-11-11-14-12-14-13-17-17-14-18-12-16-12-23	1	0,22
spanish_ht291	13-24-14-11-11-14-12-14-14-16-16-15-19-12-16-12-23	1	0,22
spanish_ht292	13-24-14-11-11-14-12-14-14-16-17-14-18-11-16-12-23	1	0,22
spanish_ht293	13-24-14-11-11-14-13-13-13-16-16-14-18-11-15-12-23	1	0,22
spanish_ht294	13-24-14-11-11-14-13-13-13-16-17-14-18-11-15-12-23	1	0,22
spanish_ht295	13-24-14-11-11-14-13-13-13-16-17-15-19-12-16-12-23	1	0,22
spanish_ht296	13-24-14-11-11-14-13-13-13-16-17-15-20-12-16-12-24	1	0,22
spanish_ht297	13-24-14-11-11-14-13-13-13-16-19-16-19-12-15-13-23	1	0,22
spanish_ht298	13-24-14-11-11-14-13-13-13-16-21-15-19-12-14-12-24	1	0,22
spanish_ht299	13-24-14-11-11-14-13-13-13-17-16-15-19-12-17-12-23	1	0,22
spanish_ht300	13-24-14-11-11-14-13-13-13-17-18-14-18-11-14-12-23	1	0,22
spanish_ht301	13-24-14-11-11-14-13-13-13-17-19-15-19-11-15-13-23	1	0,22
spanish_ht302	13-24-14-11-11-14-13-14-13-16-16-14-18-11-17-12-23	1	0,22
spanish_ht303	13-24-14-11-11-14-13-14-13-16-16-15-19-12-15-12-23	1	0,22
spanish_ht304	13-24-14-11-11-14-13-14-13-16-17-15-19-12-15-12-23	1	0,22
spanish_ht305	13-24-14-11-11-14-13-14-13-17-16-15-19-12-15-12-23	1	0,22
spanish_ht306	13-24-14-11-11-14-14-13-13-16-18-15-19-12-17-12-23	1	0,22
spanish_ht308	13-24-14-11-11-15-11-13-13-16-17-15-19-11-15-12-23	1	0,22
spanish_ht309	13-24-14-11-11-15-11-14-13-16-18-15-19-12-15-12-24	1	0,22
spanish_ht310	13-24-14-11-11-15-12-13-13-16-17-14-18-12-16-12-23	1	0,22
spanish_ht311	13-24-14-11-11-15-12-13-13-16-17-15-19-12-17-12-23	1	0,22
spanish_ht312	13-24-14-11-11-15-12-13-13-16-18-15-19-12-16-12-24	1	0,22
spanish_ht313	13-24-14-11-11-15-12-13-13-17-18-15-19-12-13-12-23	1	0,22
spanish_ht314	13-24-14-11-11-15-12-14-13-16-17-14-18-11-15-12-23	1	0,22
spanish_ht315	13-24-14-11-11-15-12-14-13-16-18-15-20-12-15-12-23	1	0,22
spanish_ht316	13-24-14-11-11-15-13-12-13-16-16-15-19-12-15-12-23	1	0,22
spanish_ht317	13-24-14-11-11-15-13-13-13-16-18-15-18-12-17-12-23	1	0,22
spanish_ht318	13-24-14-11-11-16-13-14-13-16-16-15-19-12-15-12-23	1	0,22

spanish_ht323	13-24-14-11-12-14-12-13-12-16-17-14-18-12-16-12-24	1	0,22
spanish_ht325	13-24-14-11-12-14-12-13-13-16-17-14-18-12-16-12-23	1	0,22
spanish_ht328	13-24-14-11-12-14-12-14-13-16-17-15-19-13-15-12-23	1	0,22
spanish_ht329	13-24-14-11-12-14-13-13-13-16-16-15-19-12-16-12-23	1	0,22
spanish_ht330	13-24-14-11-12-14-13-13-13-16-17-14-18-11-17-12-23	1	0,22
spanish_ht331	13-24-14-11-12-14-13-13-13-16-17-15-18-12-16-12-24	1	0,22
spanish_ht332	13-24-14-11-12-15-11-13-13-16-17-15-18-10-17-12-23	1	0,22
spanish_ht333	13-24-14-11-12-15-11-14-13-16-18-15-19-12-15-12-23	1	0,22
spanish_ht334	13-24-14-11-12-15-13-13-13-16-17-15-18-11-17-12-25	1	0,22
spanish_ht335	13-24-14-11-12-15-9-12-13-17-17-14-19-12-15-12-23	1	0,22
spanish_ht336	13-24-14-11-12-16-12-13-13-16-17-15-19-11-15-12-23	1	0,22
spanish_ht337	13-24-14-11-13-14-11-13-11-17-16-15-20-12-15-9-22	1	0,22
spanish_ht338	13-24-14-12-10-14-13-13-12-17-17-16-19-12-15-12-24	1	0,22
spanish_ht340	13-24-14-12-11-14-12-13-13-17-16-15-19-12-16-12-23	1	0,22
spanish_ht341	13-24-14-12-11-14-14-13-13-16-17-14-20-12-15-13-23	1	0,22
spanish_ht342	13-24-14-13-11-14-12-13-13-16-19-15-19-12-16-12-23	1	0,22
spanish_ht343	13-24-14-9-11-15-12-13-13-16-17-14-19-12-16-13-24	1	0,22
spanish_ht344	13-24-15-10-11-13-12-13-13-16-16-15-19-12-16-12-23	1	0,22
spanish_ht345	13-24-15-10-11-14-12-13-13-16-16-14-19-13-15-12-23	1	0,22
spanish_ht346	13-24-15-10-11-14-12-13-13-17-17-15-19-12-15-12-23	1	0,22
spanish_ht347	13-24-15-10-12-14-12-13-13-17-17-14-19-12-15-12-23	1	0,22
spanish_ht348	13-24-15-10-13-17-11-12-11-16-17-15-19-11-13-9-21	1	0,22
spanish_ht349	13-24-15-11-11-12-12-14-13-16-17-15-19-12-16-12-23	1	0,22
spanish_ht350	13-24-15-11-11-13-11-13-13-16-17-14-18-11-16-12-23	1	0,22
spanish_ht351	13-24-15-11-11-13-12-13-13-17-16-15-19-11-15-12-24	1	0,22
spanish_ht352	13-24-15-11-11-14-11-13-13-16-17-15-19-13-17-12-23	1	0,22
spanish_ht353	13-24-15-11-11-14-11-13-13-17-17-15-19-11-16-11-23	1	0,22
spanish_ht354	13-24-15-11-11-14-12-13-13-16-18-14-18-11-15-12-23	1	0,22
spanish_ht355	13-24-15-11-11-14-12-13-13-17-15-14-19-13-17-12-23	1	0,22
spanish_ht356	13-24-15-11-11-14-12-13-13-17-18-15-19-12-15-12-24	1	0,22
spanish_ht357	13-24-15-11-11-14-13-13-13-16-17-14-18-11-16-12-23	1	0,22
spanish_ht358	13-24-15-11-11-14-13-13-13-16-17-14-18-11-17-10-23	1	0,22
spanish_ht359	13-24-15-11-11-14-15-13-13-16-18-15-20-11-16-11-24	1	0,22
spanish_ht360	13-24-15-11-11-15-12-14-13-16-18-15-19-12-16-12-23	1	0,22
spanish_ht361	13-24-15-11-11-15-12-15-13-16-17-15-19-12-14-12-23	1	0,22
spanish_ht362	13-24-15-11-14-14-11-13-13-16-16-15-19-12-15-13-23	1	0,22
spanish_ht363	13-24-16-10-15-16-12-14-13-16-16-15-19-11-14-9-21	1	0,22
spanish_ht364	13-24-16-10-15-17-11-14-13-16-18-14-19-11-17-9-22	1	0,22
spanish_ht365	13-24-16-11-14-15-13-13-11-18-17-14-20-11-15-10-23	1	0,22
spanish_ht366	13-24-17-10-12-13-11-13-11-15-17-15-22-12-14-10-22	1	0,22
spanish_ht367	13-24-17-9-11-12-12-14-11-15-17-15-21-11-14-10-21	1	0,22
spanish_ht368	13-24-17-9-12-12-13-13-11-15-16-14-21-11-13-10-20	1	0,22
spanish_ht369	13-24-99-11-17-20-11-13-11-17-17-15-20-12-17-10-21	1	0,22
spanish_ht370	13-25-13-10-17-18-12-13-11-18-14-14-20-12-15-10-21	1	0,22
spanish_ht371	13-25-14-10-11-14-12-13-13-16-17-15-99-12-16-12-23	1	0,22
spanish_ht372	13-25-14-10-11-14-12-13-13-17-17-14-18-12-16-12-23	1	0,22
spanish_ht373	13-25-14-10-11-14-12-14-13-17-18-14-18-11-15-12-23	1	0,22
spanish_ht374	13-25-14-11-10-14-11-14-13-16-18-14-18-11-15-12-23	1	0,22
spanish_ht375	13-25-14-11-11-12-12-13-13-16-16-15-20-12-16-12-23	1	0,22
spanish_ht376	13-25-14-11-11-13-12-13-13-16-17-14-19-12-15-12-23	1	0,22
spanish_ht377	13-25-14-11-11-14-11-12-13-16-17-14-18-11-17-12-25	1	0,22
spanish_ht378	13-25-14-11-11-14-11-13-13-17-17-15-19-12-16-12-23	1	0,22
spanish_ht379	13-25-14-11-11-14-12-13-12-17-17-14-18-11-16-12-23	1	0,22
spanish_ht380	13-25-14-11-11-14-12-13-13-15-17-15-20-12-15-12-23	1	0,22
spanish_ht381	13-25-14-11-11-14-12-13-13-16-17-14-18-12-16-12-23	1	0,22
spanish_ht382	13-25-14-11-11-14-12-13-13-16-18-15-19-12-16-12-23	1	0,22
spanish_ht383	13-25-14-11-11-14-12-14-13-16-17-14-18-11-16-12-23	1	0,22
spanish_ht384	13-25-14-11-11-14-12-14-13-16-17-15-19-13-15-12-23	1	0,22
spanish_ht385	13-25-14-11-11-14-13-13-13-16-16-15-19-12-15-12-23	1	0,22
spanish_ht386	13-25-14-11-11-14-13-13-13-16-16-15-19-12-16-13-23	1	0,22
spanish_ht387	13-25-14-11-11-14-13-13-13-16-18-14-18-11-16-12-25	1	0,22
spanish_ht389	13-25-14-11-12-13-11-13-13-16-18-15-19-12-15-12-23	1	0,22
spanish_ht390	13-25-15-10-11-14-10-13-11-17-15-14-20-12-15-11-23	1	0,22
spanish_ht391	13-25-15-10-17-19-12-13-11-17-15-14-21-11-15-10-21	1	0,22
spanish_ht392	13-25-15-11-11-12-11-12-13-16-17-15-19-12-15-12-23	1	0,22
spanish_ht393	13-25-15-11-11-14-12-13-13-16-18-14-18-11-16-12-23	1	0,22
spanish_ht394	13-25-15-11-12-12-12-13-13-18-16-16-99-11-15-11-24	1	0,22
spanish_ht395	13-25-15-11-12-13-12-14-13-18-16-16-99-12-15-11-24	1	0,22
spanish_ht396	13-25-15-11-12-15-12-13-13-99-17-16-20-12-15-12-23	1	0,22
spanish_ht397	13-25-16-11-11-14-11-13-11-17-15-13-19-13-16-11-23	1	0,22
spanish_ht398	13-25-17-9-12-12-12-13-11-15-17-15-21-11-15-10-22	1	0,22
spanish_ht399	13-25-17-9-12-12-12-14-11-15-17-15-21-11-14-10-21	1	0,22
spanish_ht400	13-26-14-11-11-14-12-12-13-17-17-14-19-12-15-12-23	1	0,22
spanish_ht401	13-27-14-11-11-15-11-13-13-16-18-15-19-12-16-12-23	1	0,22
spanish_ht402	14-21-16-10-13-16-11-12-11-17-17-16-23-11-15-10-23	1	0,22
spanish_ht403	14-21-16-10-16-16-11-12-11-18-17-14-21-11-15-11-22	1	0,22
spanish_ht404	14-22-15-10-14-14-11-13-11-18-16-16-22-12-15-10-20	1	0,22
spanish_ht405	14-22-15-10-14-15-11-14-11-16-20-16-21-12-14-10-22	1	0,22
spanish_ht406	14-22-15-7-14-15-11-12-11-17-16-15-20-12-15-10-20	1	0,22
spanish_ht407	14-22-16-10-14-15-11-12-11-16-16-16-21-13-15-11-20	1	0,22
spanish_ht408	14-23-14-11-11-14-12-13-13-16-18-14-18-11-15-12-23	1	0,22
spanish_ht409	14-23-14-11-11-14-13-12-13-16-17-14-18-11-16-12-23	1	0,22
spanish_ht410	14-23-14-11-11-15-13-12-13-16-17-14-18-11-16-12-23	1	0,22
spanish_ht411	14-23-14-11-12-13-10-14-14-16-17-14-19-11-15-10-22	1	0,22
spanish_ht412	14-23-14-11-12-13-10-14-14-16-17-14-19-11-15-10-23	1	0,22
spanish_ht413	14-23-15-10-15-15-11-12-11-17-15-16-21-13-16-10-21	1	0,22
spanish_ht414	14-23-15-10-15-16-11-13-12-16-15-15-20-11-16-10-20	1	0,22
spanish_ht415	14-23-15-10-16-16-12-14-12-17-17-15-21-12-13-10-21	1	0,22
spanish_ht416	14-23-17-10-12-12-11-13-11-15-17-15-21-12-14-10-22	1	0,22
spanish_ht417	14-24-14-10-11-14-11-13-12-16-20-15-20-11-15-12-23	1	0,22
spanish_ht418	14-24-14-10-11-14-11-13-13-16-17-15-20-12-16-12-23	1	0,22
spanish_ht419	14-24-14-10-11-14-12-13-13-16-17-14-18-11-15-12-25	1	0,22
spanish_ht420	14-24-14-10-11-14-12-13-13-17-17-15-19-11-15-12-23	1	0,22
spanish_ht421	14-24-14-10-11-15-11-13-13-16-17-15-19-13-15-12-23	1	0,22
spanish_ht422	14-24-14-11-11-14-11-13-13-16-17-14-18-11-15-12-23	1	0,22
spanish_ht423	14-24-14-11-11-14-11-13-13-16-18-15-19-12-17-12-24	1	0,22
spanish_ht424	14-24-14-11-11-14-12-14-13-17-17-14-18-11-16-12-23	1	0,22
spanish_ht425	14-24-14-11-12-14-12-13-13-17-16-15-19-12-16-12-23	1	0,22
spanish_ht426	14-24-14-9-18-20-12-13-11-17-13-14-22-12-15-10-20	1	0,22
spanish_ht427	14-25-13-10-17-18-11-12-11-17-14-14-20-11-16-10-22	1	0,22
spanish_ht428	14-25-13-9-16-17-12-14-11-17-15-14-20-11-15-10-22	1	0,22
spanish_ht429	14-25-14-10-12-15-11-13-13-16-17-15-19-12-16-12-23	1	0,22
spanish_ht430	14-25-16-9-12-12-12-14-11-15-16-15-21-11-14-10-22	1	0,22
spanish_ht431	14-26-17-10-12-15-12-13-12-17-18-15-20-10-16-10-21	1	0,22
spanish_ht432	15-21-15-10-13-16-13-12-11-17-17-16-23-11-16-10-21	1	0,22
spanish_ht433	15-22-15-9-12-14-12-12-11-20-17-16-21-11-14-10-21	1	0,22
spanish_ht434	15-23-15-10-15-15-11-13-12-17-15-14-20-11-15-10-21	1	0,22

Supplementary Table S7: Y-STR haplotypes for 456 Spanish non-Roma individuals.

Population	HD	CI lower	CI upper
RomaCentralSpain	0,979	0,917	0,973
RomaEasternSpain	0,971	0,936	0,972
RomaNorthernSpain	0,969	0,886	0,960
RomaSouthernSpain	0,960	0,870	0,953
RomaWesternSpain	0,955	0,727	0,955
CentralSpain	1,000	0,986	0,992
EasternSpain	1,000	0,990	0,994
NorthernSpain	0,999	0,987	0,993
SouthernSpain	1,000	0,984	0,991
WesternSpain	1,000	0,942	0,979

Supplementary Table S8: Haplotype diversity (HD) across Spanish regions in Roma and Spanish non-Roma populations.

Haplogroup	Central Spain (N)	Central Spain (%)	Eastern Spain (N)	Eastern Spain (%)	NorthernSpain (N)	NorthernSpain (%)	SouthernSpain (N)	SouthernSpain (%)	Western Spain (N)	Western Spain (%)	RomaCentralSpain (N)	RomaCentralSpain (%)	RomaEasternSpain (N)	RomaEasternSpain (%)	RomaNorthern Spain (N)	RomaNorthern Spain (%)	RomaSouthern Spain (N)	RomaSouthern Spain (%)	RomaWesternSpain (N)	RomaWesternSpain (%)
E	5	5,21	5	3,85	10	8,4	6	7,23	1	3,57	0	0	0	0	0	0	0	0	0	0
G	1	1,04	4	3,08	1	0,84	6	7,23	3	10,71	0	0	0	0	0	0	0	0	0	0
I1	3	3,13	1	0,77	4	3,36	3	3,61	1	3,57	0	0	0	0	0	0	0	0	0	0
I2a1	4	4,17	4	3,08	11	9,24	0	0	0	0	3	7,89	3	4,17	0	0	0	0	0	0
J1	2	2,08	2	1,54	0	0	1	1,2	0	0	0	0	3	4,17	0	0	0	0	0	0
J2a	3	3,13	4	3,08	0	0	2	2,41	0	0	1	2,63	0	0	0	0	0	0	0	0
J2a1b	6	6,25	6	4,62	3	2,52	0	0	0	0	12	31,58	25	34,72	11	42,31	10	40	1	8,33
J2b	2	2,08	2	1,54	1	0,84	2	2,41	0	0	1	2,63	0	0	0	0	0	0	0	0
N	2	2,08	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R1a	3	3,13	0	0	0	0	1	1,2	1	3,57	0	0	0	0	1	3,85	0	0	0	0
R1b	63	65,63	94	72,31	83	69,75	55	66,27	21	75	9	23,68	11	15,28	6	23,08	3	12	3	25
T	2	2,08	2	1,54	3	2,52	2	2,41	0	0	0	0	0	0	1	3,85	0	0	0	0
E1b1b	0	0	1	0,77	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G2a	0	0	1	0,77	0	0	0	0	0	0	0	0	0	0	1	3,85	0	0	0	0
I2	0	0	1	0,77	1	0,84	1	1,2	0	0	0	0	0	0	0	0	0	0	0	0
I2a	0	0	2	1,54	2	1,68	1	1,2	0	0	2	5,26	0	0	0	0	0	0	0	0
J2a1	0	0	1	0,77	0	0	1	1,2	0	0	0	0	1	1,39	0	0	0	0	0	0
I2a(xI2a1)	0	0	0	0	0	0	1	1,2	0	0	0	0	2	2,78	0	0	0	0	0	0
L	0	0	0	0	0	0	1	1,2	0	0	0	0	0	0	0	0	0	0	0	0
J1a	0	0	0	0	0	0	0	0	1	3,57	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	0	0	0	7	18,42	21	29,17	4	15,38	11	44	5	41,67
I2a (xI2a1)	0	0	0	0	0	0	0	0	0	0	1	2,63	0	0	0	0	0	0	0	0
I2b1	0	0	0	0	0	0	0	0	0	0	1	2,63	0	0	0	0	0	0	0	0
J2a1(xbh)	0	0	0	0	0	0	0	0	0	0	1	2,63	0	0	2	7,69	0	0	0	0
J2	0	0	0	0	0	0	0	0	0	0	0	0	1	1,39	0	0	1	4	3	25
J2a1h	0	0	0	0	0	0	0	0	0	0	0	0	4	5,56	0	0	0	0	0	0
O3	0	0	0	0	0	0	0	0	0	0	0	0	1	1,39	0	0	0	0	0	0

Supplementary Table S9: Distribution of haplogroups by region for Roma populations and Spanish non-Roma populations.

Test	Source of Variation	Percentage of variation
Roma and Spanish non-Roma by region as a single population	Between regions	11.28 *
	Within population	88.72 *
	Total	100
Roma and Spanish non-Roma by region as two populations	Between populations	19.77 *
	Among populations within regions	0,9
	Within populations	79.33 *
	Total	100
Spanish Roma by region	Between regions	3 *
	Within regions	97 *
	Total	100
Spanish non-Roma by region	Between regions	0,48
	Within regions	99.52 *
	Total	100

*** p-value < 0.05**

Supplementary Table S10: AMOVA tests performed on Spanish Roma and Spanish non-Roma populations.

Population	WesternSpain	SouthernSpain	NorthernSpain	EasternSpain	CentralSpain	RomaCentralSpain	RomaEasternSpain	RomaNorthernSpain	RomaSouthernSpain	RomaWesternSpain
WesternSpain										
SouthernSpain	368,58									
NorthernSpain	224,43	75,74								
EasternSpain	152,46	89,62	93,92							
CentralSpain	46,02	49,39	63,51	87,13						
RomaCentralSpain	2,97	2,85	2,43	2,98	2,24					
RomaEasternSpain	1,82	1,78	1,60	1,86	1,38	17,25				
RomaNorthernSpain	2,42	2,34	2,05	2,43	1,76	43,01	21,05			
RomaSouthernSpain	1,39	1,37	1,23	1,40	1,00	8,20	44,28	10,80		
RomaWesternSpain	2,25	2,20	1,89	2,30	1,60	8,91	15,32	7,84	7,54	

Supplementary Table S11: Migration rate (M) matrix between Spanish regions in Roma and Spanish non-Roma populations.

A

Routes	Model	Log(mL)	LBF	Model-probability
Panmictic	1	-17630,54	-5653,45	0
East-West	2	-15454,36	-3477,27	0
Divergence with migration in two directions	5	-14205,64	-2228,55	0
Divergence with migration	4	-14017,46	-2040,37	0
Divergence	3	-11977,09	0	1

B

Routes	Model	Log(mL)	LBF	Model-probability
Panmictic	1	-15567,13	-2550,27	0
Divergence with migration in two directions	5	-15329,45	-2312,59	0
Divergence	3	-13518,98	-502,12	0
Divergence with migration	4	-13178,29	-161,43	0
East-West	2	-13016,86	0	1

Supplementary Table S12: Comparison of likelihood estimates in Migrate tests. A) Iberian models; B) European models.

Log(ml) = logarithm of the maximum likelihood, LBF = Log Bayes Factor.

This thesis is a result of the I+D+I project with reference AEI-PID2019-106485GB-I00/AEI/10.13039/501100011033, funded by Ministerio de Ciencia, Innovación y Universidades (MCIU), and by Agencia Estatal de Investigación (AEI).