Ena *et al. BMC Genomics* (2025) 26:1012
https://doi.org/10.1186/s12864-025-12210-8

BMC Genomics

RESEARCH Open Access

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

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

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

*Correspondence:
David Comas
david.comas@upf.edu
[1]Institut de Biologia Evolutiva (CSIC-UPF), Departament de Medicina i Ciències de la Vida, Universitat Pompeu Fabra, Barcelona, Spain
[2]Facultat de Sociologia, Universitat Autònoma de Barcelona, Barcelona, Spain
[3]Facultat de Geografia i Història, Universitat de Barcelona, Barcelona, Spain

Introduction

The Romani are the largest transnational ethnic minority in Europe. Although traditionally nomadic or seminomadic, 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 [5, 13].

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 [14, 15]. Historical records indicate that the Roma arrived in the various kingdoms of what is now Spain in 1425 [16] and in Portugal likely earlier than 1521 [15], the year of their first documented mention, having migrated from Eastern and Central Europe [16, 17]. 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, 18], or share close genetic connections with Jewish groups [19, 20], previous studies have found no genetic evidence to support these claims [21, 22]. 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 [23]. 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 [17, 23, 24]. In the 19th and 20th centuries, internal migrations further displaced Roma populations within Spain, potentially increasing admixture among groups [23]. 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 [24–26].

Over the past 25 years, genetic research on Spanish Roma has often focused on medical issues [27–30]. However, several studies on uniparental DNA have included Spanish Roma individuals, with two specifically examining Spanish mtDNA [31, 32], collectively revealing evidence of bottlenecks, identifying founder lineages, and uncovering sex-biased admixture patterns [33–36]. In addition, two studies focusing on Roma autosomal data have also included Spanish Roma samples [7, 37], while the most recent and comprehensive research to date consists of an in-depth whole-genome array study on Iberian Roma [21]. 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 [22], and another including a limited sample size of Spanish Roma that did not specifically focus on this group [33], 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 [38], in collaboration with the Roma FAGiC association (*Federació d'Associacions Gitanes de Catalunya*). Participants were selected randomly and 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 [39]. 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) [40].

After genotyping, 133 Spanish Roma samples were included in the dataset, while the inclusion of 40 Spanish Roma samples from Martinez-Cruz et al. [33] 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. [32] and Ena et al. [21] (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, along with detailed citations for each reference dataset obtained from supplementary materials of previously published studies, are listed in Supplementary Table S1. These datasets do not have centralised accession numbers but are publicly accessible via the cited sources.

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 [41]; 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 [42]. 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 [43] and calculated with R software [44]. 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 [45]. 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}) [46] were calculated using Arlequin version 3.5.1.2 [47], 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 [48], and plots were 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 [49, 50]. 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 [51]. 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 [52], 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. [53] and implemented by Pamjav et al. [54]. 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 [55], 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 [51, 56, 57].

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. [51]: (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 (Fig. 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).

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%) [58, 59]. 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) [60–63] 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; Fig. 2) is the tight clustering of all H haplotypes within short branches, indicating minimal diversity. 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

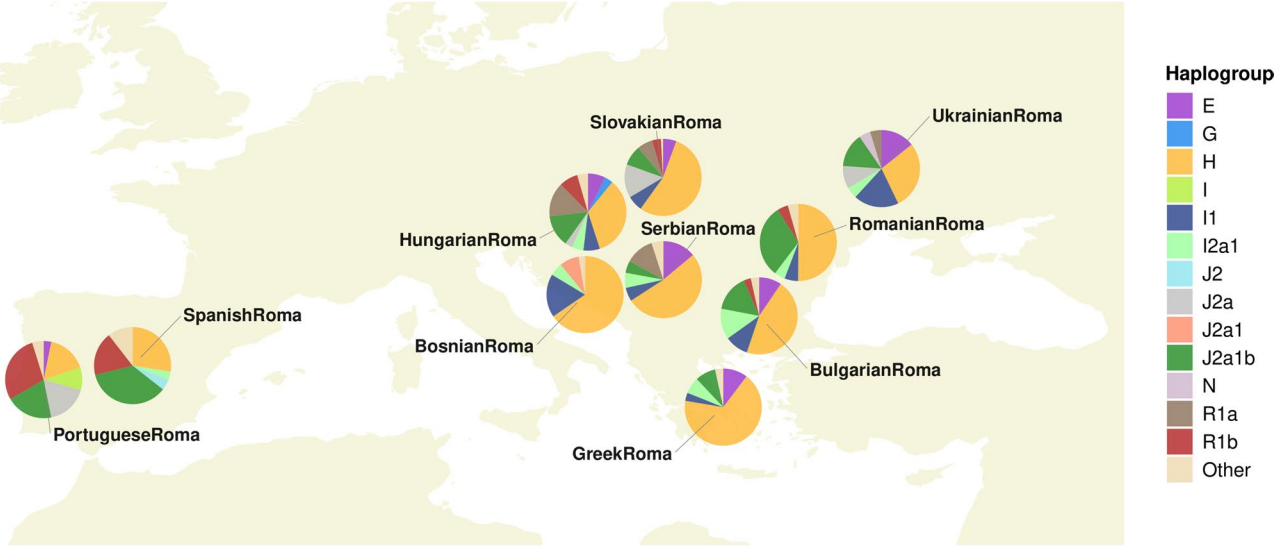

Fig. 1 Geographic distribution of Y-chromosome haplogroups in European Roma populations. Pie charts show the frequencies of the inferred haplogroups in each population. Haplogroups with frequencies below 3% are labelled as "Other" in the legend

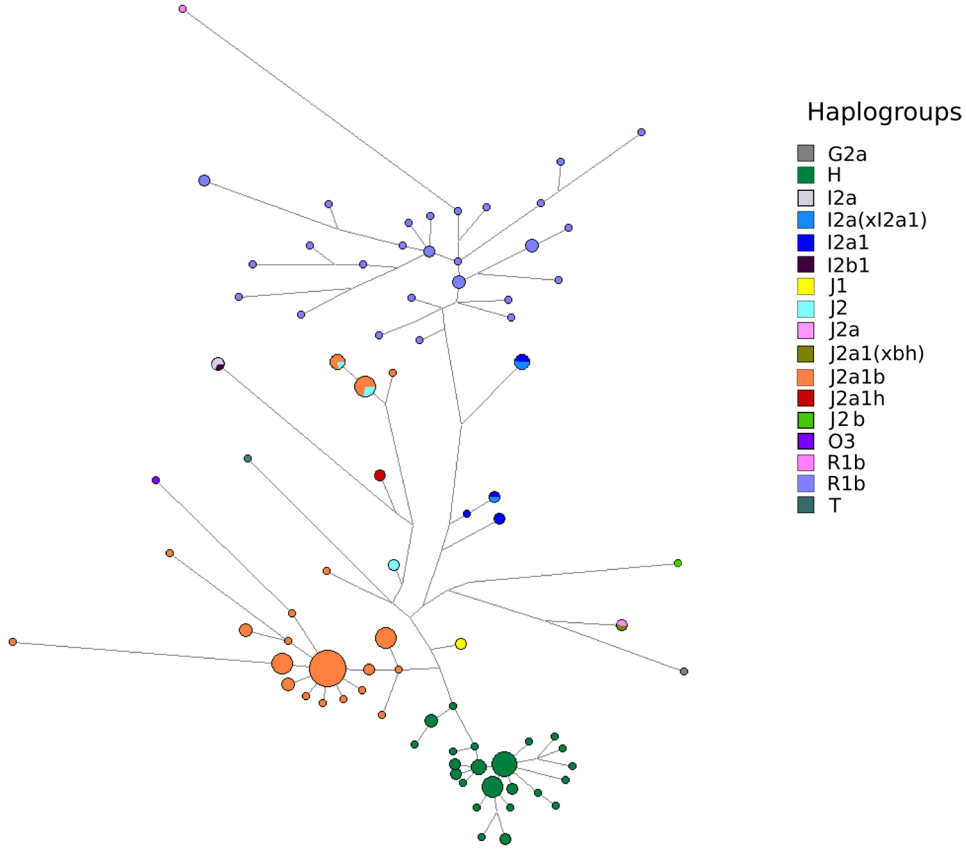

Fig. 2 Median-joining network of 173 Spanish Roma individuals based on 17 Y-chromosome STR markers. Colours represent the inferred haplogroups derived from the STR profiles

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 (Fig. 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. 3c-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.

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

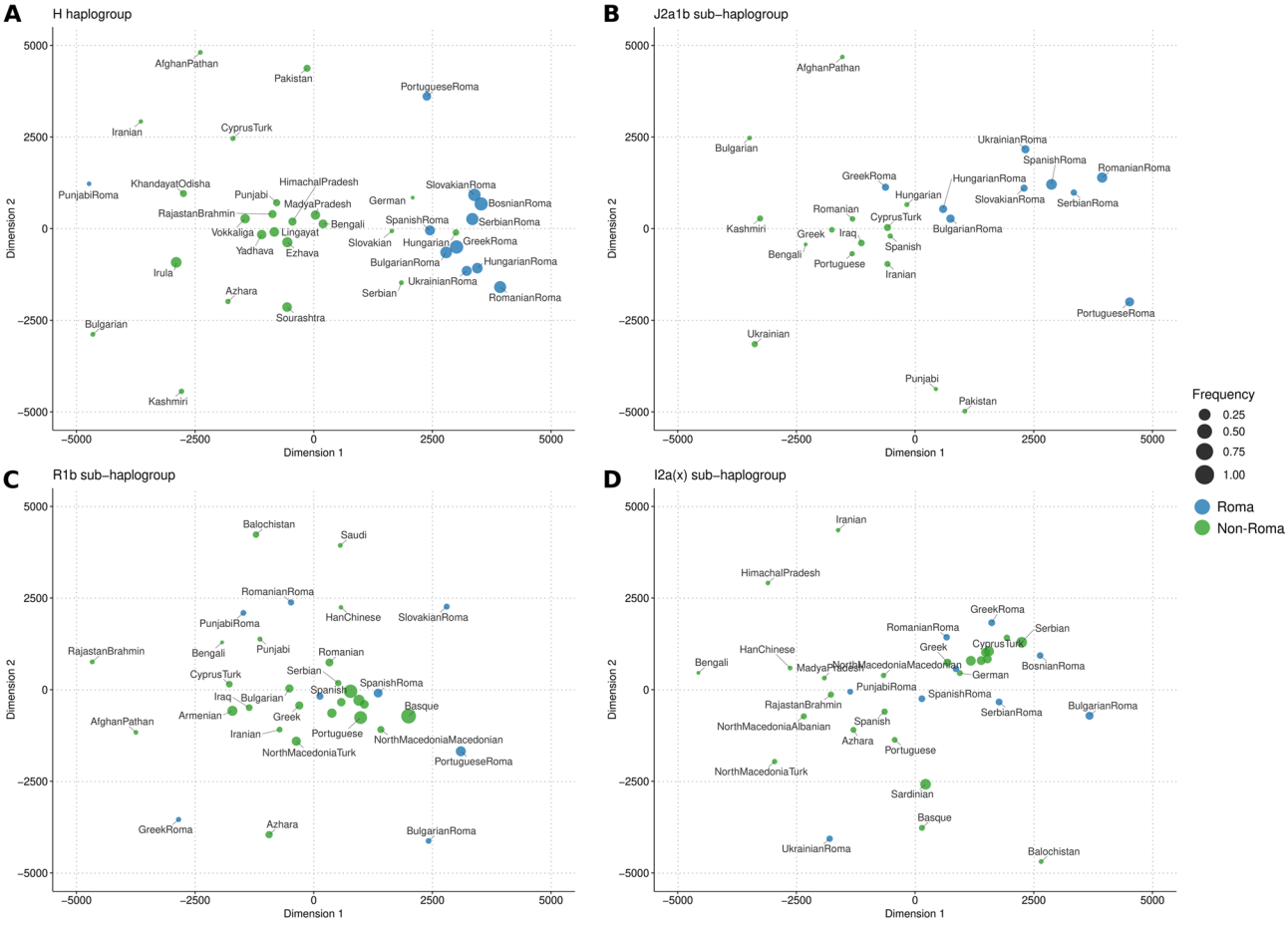

Fig. 3 Multidimensional scaling (MDS) plots of the four major haplogroups found in Spanish Roma, based on Rst distances: (**a**) Haplogroup H; (**b**) Haplogroup J2a1b; (**c**) Haplogroup R1b; (**d**) Haplogroup I2a(x). Roma populations are shown in blue and non-Roma populations in green. The size of the circles represents the population frequency of the haplogroups

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, Fig. 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 (Fig. 4, Table S9). In contrast, J2a1b, linked to West Asian ancestry, is more common in Northern and Eastern Spain (Fig. 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 (Fig. 4, Table S9). Additionally, less frequent haplogroups, such as I2 and its sub-haplogroups, exhibit regional patterns, appearing exclusively in Central and Eastern Spain (Fig. 4, Table S9). Together, these findings indicate diverse genetic influences and highlight regional substructure within Roma populations across Spain.

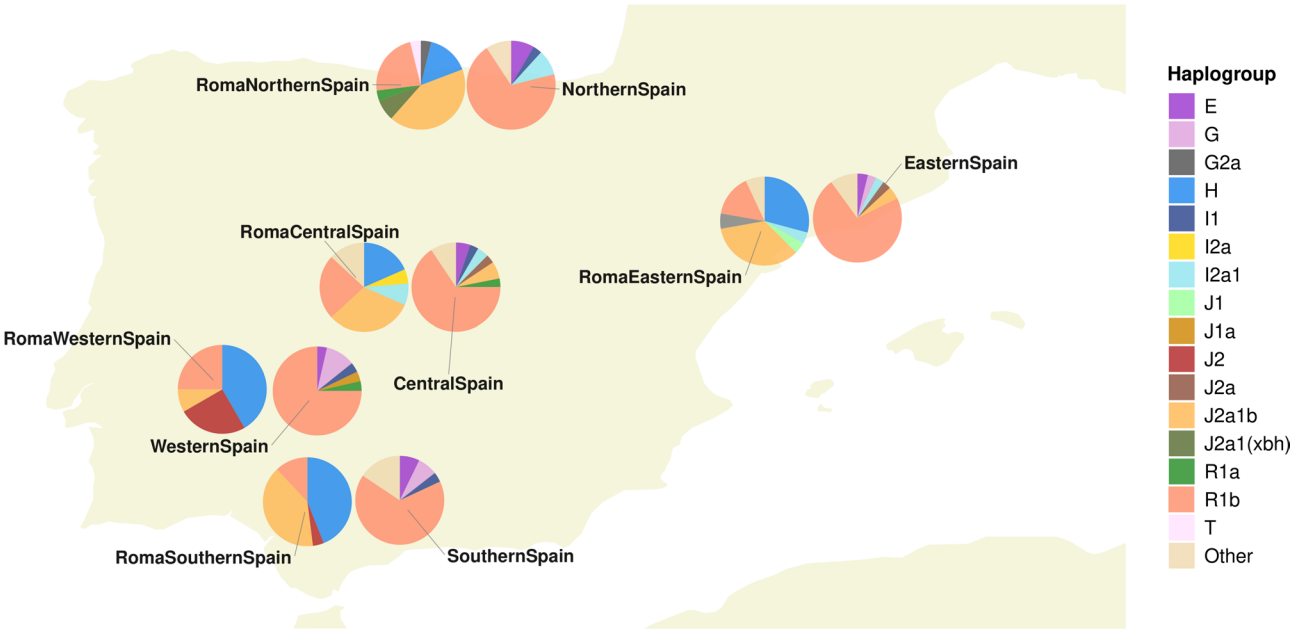

Fig. 4 Geographic distribution of Y-chromosome haplogroups in Spanish Roma and non-Roma populations. Pie charts show the frequencies of inferred haplogroups in each population. Haplogroups with frequencies below 3% are labelled as "Other" in the legend

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 [64, 65]. 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 [66–68]. It is important to note that endogamy, combined with the practice of patrilocal residence observed both historically and in modern times among Spanish Roma [69, 70]—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 [71], 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 [33, 72–74]. 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 [21, 37] and likely reflects the history of forced assimilation occurred within the peninsula over several centuries [16, 23, 75].

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 [76], 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 [32], Y-chromosome [33, 68], and whole-genome array data [21] 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 [77], where haplogroup E-M81, associated with Amazigh ancestry [78–80], accounts for around 1%. This contrasts with the Portuguese Roma, where 3.2% carry E haplogroups of African origin [22]. In comparison, predictions for the Portuguese non-Roma, based on combined data from multiple studies [81, 82], 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 [21], 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 [32], 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 [83], 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 [21].

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 [33, 36]—within the Roma population [84]. In this context, previous research has identified both maternal and paternal founder lineages within the Roma population [32, 33, 36]. 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 [32], 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 [35]. 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 [85–87], 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 [21] but contrasts with maternal DNA studies [32], 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 [69, 70], 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 [21, 37], 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.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-12210-8.

Supplementary Material 1.

Supplementary Material 2.

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.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).

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.

Authors' 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.

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:https://doi.org/10.13039/501100011033). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability

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

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 informed consent was obtained from all donors participating in the present study.

Competing interests

The authors declare no competing interests.

Received: 24 March 2025 / Accepted: 9 October 2025

Published online: 07 November 2025

References

1. Hancock IF. We are the Romani people. Univ of Hertfordshire; 2002.
2. Kenrick D. Historical dictionary of the Gypsies (Romanies). Scarecrow; 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. Benišek 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. https://doi.org/10.1371/journal.pone.0058633.
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. Popul Genet Eur Roma—A Rev Genes. 2022;13:2068.
9. Matras Y. Romani A Linguistic Introduction. 2002.
10. Bakker P, Kṭ̄uchukov K, Ki[illegible]chukov K. What is the Romani language? Univ of Hertfordshire; 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. Bakker P. Romani genetic linguistics and genetics: results, prospects and problems. Romani Stud. 2012;22:91–111.
14. 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.
15. Bastos JGP, Bastos SP. Gypsies (Ciganos) in Portugal, today. Stud Eur. 2000;8:99.
16. Pym R. The Gypsies of early modern Spain. Springer; 2007.
17. 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.
18. 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.
19. Pohoryles Y. The Jewish-Romani connection: Are Gypsies descendants of tribe of Simeon? Ynetnews. 2018.
20. El Origen judío de los gitanos. http://www.hashavuabogota.com/articulos/704. Accessed 13 Dec 2024.
21. 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.
22. 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(2):215–27.
23. Sánchez DM. Historia Del Pueblo Gitano En España. Los Libros De La Catarata; 2022.
24. 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.
25. 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. pp. 192–207.
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.
31. 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.
32. 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.
33. 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.

34. 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.
35. 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.
36. 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.
37. 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.
38. Giménez A, Comas D, Carballo A. Origen e identidad Del Pueblo Gitano. Int J Roma Stud. 2019;1:159–84.
39. 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.
40. ISFG - isfg.org. https://www.isfg.org/. Accessed 31 Jul 2024.
41. Athey TW. Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. J Genet Geneal. 2006;2:34–9.
42. 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.
43. Nei M. Molecular evolutionary genetics. Columbia university; 1987.
44. R Core Team. R: A language and environment for statistical computing. 2024.
45. Canty AJ. Resampling methods in R: the boot package. Newsl R Proj. 2002;2:2–7.
46. Slatkin M. A measure of population subdivision based on microsatellite allele frequencies. Genetics. 1995;139:457–62.
47. 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.
48. 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.
49. 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.
50. 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.
51. 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.
52. 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(1):7341.
53. 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.
54. 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.
55. Beerli P, Mashayekhi S, Sadeghi M, Khodaei M, Shaw K. Population genetic inference with MIGRATE. Curr Protoc Bioinformatics. 2019;68:e87.
56. 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.
57. 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.
58. 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.
59. 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.
60. 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.
61. 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.
62. 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.
63. 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.
64. 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.
65. 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.
66. 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.
67. 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.
68. 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.
69. Martin E, Gamella JF. Marriage practices and ethnic differentiation: the case of Spanish Gypsies (1870–2000). The History of the Family. 2005;10:45–63.
70. Gamella JF, Muntean VM. Marriage and the reproductive regime of a digitally connected Roma diaspora. J Contemp Cent East Eur. 2023;31:533–59.
71. 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.
72. 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.
73. 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.
74. 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.
75. 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.
76. 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.
77. 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.
78. 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.
79. 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.
80. 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.
81. 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.
82. Pontes ML, Cainé L, Abrantes D, Lima G, de Pinheiro M. 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.

83. 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.
84. Kalaydjieva L, Morar B, Chaix R, Tang H. A newly discovered founder population: the Roma/Gypsies. BioEssays. 2005;27:1084–94.
85. Drummond S. Mapping marriage law in Spanish Gitano communities. UBC; 2011.
86. Matras Y. The Romani Gypsies. Harvard University Press; 2015.
87. Gamella JF, Álvarez-Roldán A. Breaking secular endogamy. The growth of intermarriage among the Gitanos/Calé of Spain (1900–2006). The History of the Family. 2023;28:457–83.

Publisher's note