Genomic Insights into the Population History of the *Resande* or Swedish Travelers

Laura Vilà-Valls[1], Julen Aizpurua-Iraola[1], Sebastian Casinge[2], Karin Bojs[2], André Flores-Bello[1,†], Neus Font-Porterias[1,†,#], and David Comas [1,*,†]

[1]Institute of Evolutionary Biology (UPF-CSIC), Department of Medicine and Life Sciences, Universitat Pompeu Fabra, Barcelona, Spain
[2]Independent scholar, Stockholm, Sweden

[†]These authors contributed equally to this work.

[#]Present Affiliation: Department of Biomedical Informatics, Anschutz Medical Campus, University of Colorado, Aurora, CO 80309, USA.

*Corresponding author: E-mail: david.comas@upf.edu.

Accepted: 09 January 2023

Abstract

The *Resande* are a minority ethnic group in Sweden, who were characterized by an itinerant way of life, and they have been suggested to originate from the mixture between Swedish and Romani populations. Because the population history of the *Resande* has been scarcely studied, we analyzed genome-wide genotype array data from unrelated *Resande* individuals in order to shed light on their origins and demographic history for the first time from a genetic perspective. Our results confirm the Romani-related ancestry of this population and suggest an admixture event between a Romani-like population and a general Swedish-like population that occurred approximately between the mid-18th and mid-19th centuries, two centuries after the arrival of the first historically reported Romani families in Sweden. This inferred date suggests that the Romani group involved in the admixture is related to the pre-18th-century arrivals of Romani in Scandinavia. In addition, a reduction in the population size is detected previous to the admixture event, suggesting a subtle signal of isolation. The present work constitutes a step forward toward a better representation of ethnic minorities and underrepresented groups in population genetic analyses. In order to know in more detail the complete history of human populations, it is time to focus on studying populations that have not been previously considered for a general scenario and that can provide valuable information to fill in the gaps that still remain uncovered.

Key words: Resande population, admixture, genome-wide autosomal data, Sweden, population genetics, Romani population.

Introduction

The *Resande* or *Resandefolket*, translated into English as Swedish Travelers, are a minority ethnic group in Sweden currently estimated to number ~20,000 individuals (Government of Sweden: Ministry of Culture 1999).

The *Resande* were historically considered as an itinerant group, as the ethnonym *Resande* (translated from Swedish as "traveler") suggests. Hence, their traditional occupations have been closely associated with an itinerant lifestyle, such as horse dealers, horse traders, artisans, farmers, animal healers, and glass/porcelain dealers, among others (Heymowski 1969; Tervonen 2010; Wiedner n.d.). This lifestyle started to change with the societal changes, industrialization, and economic development in the 20th century, although there is still a small proportion of the *Resande* that maintain a traditional lifestyle (Wiedner n.d.).

The presence of Romani (or Roma) groups in Sweden was first described in the year 1512, apparently coming from Scotland and England (Fraser 1992), with a continuous influx in the 16th century and especially the 17th century. Not least, Sweden's time as a major military power in Europe during the 17th century and its need to recruit

Significance

The *Resande* are a minority ethnic group in Sweden for whom no genetic studies have yet been done, and whose Romani ancestry has long been questioned by academics. The present study fills this gap by using genome-wide genotype data to study their origins and demographic history. It demonstrates that the *Resande* have a Romani-related ancestry and that the Romani and the general Swedish populations admixed between the mid-18th and mid-19th centuries, increasing the genetic diversity of the *Resande* population.

soldiers played an important role in this process (Machiels 2002). Although the origins of the *Resande* have been related to the first migration (Fraser 1992; Tervonen 2010), the genealogical evidence suggests that contemporary *Resande* share Romani ancestors who immigrated from Continental Europe (Germany, France, and the Low Countries) in the late 1600s (Lindwall 2014). The European Romani originated in South Asia around 1,500 years ago, from a group closely related to the Punjabi from Northwestern India (Fraser 1992; Matras 1995; Hancock 2002; Mendizabal et al. 2012; Font-Porterias et al. 2019), although their South Asian–related ancestry is more complex and also involved populations with high Ancestral South Indian (ASI) ancestry (Font-Porterias et al. 2019). They are known to have spread throughout Europe from the Balkans (Fraser 1992), admixing with local European populations, resulting in different proportions of West Eurasian components among European Romani groups (Font-Porterias et al. 2019).

Romani groups and other Traveler communities across Europe have been subject to discrimination for many years (Machiels 2002; Tervonen 2010). In Sweden, many laws were imposed against Romani during the 16th and 17th centuries, which authorized their persecution and deportation among other repressive measures (Machiels 2002). Particularly for the *Resande*, previous studies on pedigrees suggest that the social exclusion suffered during centuries resulted in a high incidence of endogamy practices (Heymowski 1969), understood as marriages within the population or community (Ceballos et al. 2018). During the 20th century, the *Resande* suffered harsh assimilation measures that were intended to destroy their culture and lifestyle to forcibly integrate them into the general Swedish society (Government of Sweden: Ministry of Culture 2014; Wiedner n.d.). In the late 20th century, the *Resande*, together with other Romani groups in Sweden, were ensured some acceptance, legal protection, and the status of national minority (Wiedner n.d.; Government of Sweden: Ministry of Culture 1999).

The Romani-related ancestry of the *Resande* has long been questioned by academics. During the 20th century, two different hypotheses were discussed: the *Resande* as a mixed population of Romani and Swedes, and the *Resande* as individuals who were socially excluded from the general Swedish population (Heymowski 1955, 1969). The latter hypothesis became the predominant one in Swedish academia (Lindholm & Svanberg 1988; Svensson 1993). More recently, genealogists (Lindgren & Lindwall 1992) and historians (Minken 2009) have indicated a Romani-related genetic component in the *Resande* ancestry in the 17th and 18th centuries, also confirmed by recent DNA genealogy (Bojs & Sjölund 2016). Moreover, language samples of inflected Romani are found among *Resande* ancestors as late as the 18th century, and nowadays, the Romani language is preserved in a para-Romani variety (Bakker 2020).

Although the population genetics and demography of many groups are well-known and have been studied for decades, some groups and ethnic minorities still remain underrepresented in genetic studies (Popejoy & Fullerton 2016). In this case, there are no previous genetic studies on the *Resande* despite the fact that the general Swedish population has been extensively described (Humphreys et al. 2011). Therefore, the present work aims to fill this gap by using genome-wide genotype data with both allele frequency- and haplotype-based approaches, providing a robust fine-scale characterization.

Results

Resande Differentiation and Genetic Structure in a European Context

The genetic affinity of *Resande* individuals (N = 9) with other populations was analyzed within a large population panel that included a wide range of Indian, Middle Eastern, Caucasian, and European populations and a Romani reference population from the Iberian Peninsula (see Materials and Methods). Focusing on the European context, the PCA shows that *Resande* samples are separated from the general Swedish and other Northern European samples and fall closer to the Romani cline (fig. 1*A*). This singular pattern of the *Resande* individuals is also shown in the PCA of the whole data set, where Romani samples fall between the non-Romani European and Indian populations (supplementary fig. S1*A*, Supplementary Material online), in accordance with previous genetic studies (Mendizabal et al. 2012; Font-Porterias et al. 2019) and their historical demographic records (Fraser 1992).

Individual ancestries were estimated with ADMIXTURE. The results for the European context (fig. 1*B* and supplementary fig. S2*A*, Supplementary Material online)

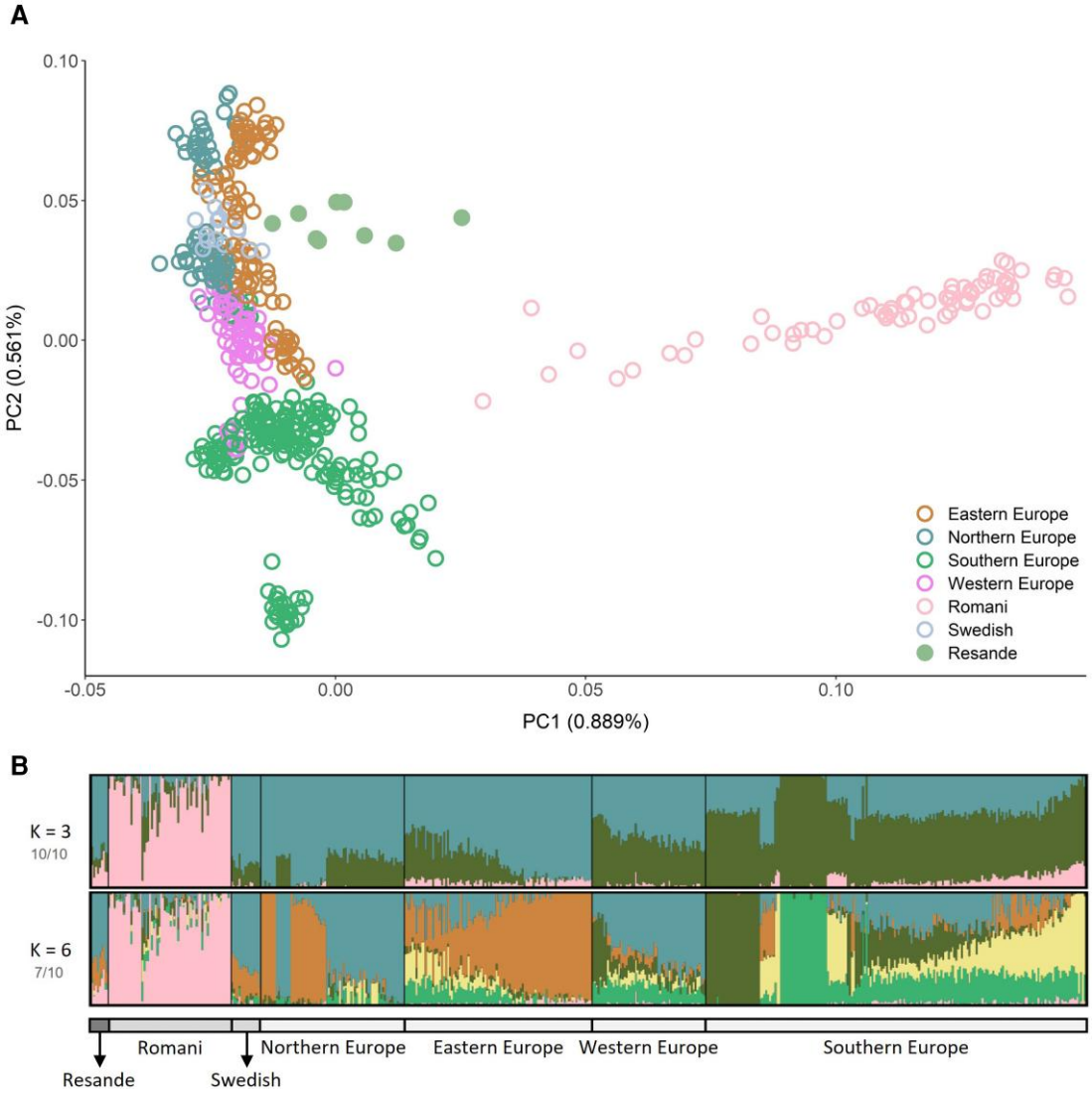

Fig. 1.—*Resande* genetic structure within the European context. (*A*) Principal component analysis including the European references and the *Resande*, general Swedish, and Romani populations. Grouped by geographical region. (*B*) ADMIXTURE results for K = 3 (lowest cross-validation error) and K = 6. The plots correspond to the representatives of the major modes of each K. The number of runs out of 10 classified in the major mode at each value of K are indicated in the plot. Within the large "Southern Europe" group, the two outlier populations in K = 6 are Basque (olive green) and Sardinian (sea green).

show at K = 3, with the lowest cross-validation error (supplementary fig. S2*B*, Supplementary Material online), a major component (pink) in Romani, present in all *Resande* samples at intermediate frequencies (17.7% ± 7.2), refining the PCA results. In addition, the *Resande* show a major component (blue) (69.3% ± 6.2) found at high frequencies in the general Swedish, Northern European, and Eastern European populations. Thus, *Resande* samples have a similar genetic pattern to Swedish samples across all K's, except for the presence of a Romani-like component (fig. 1*B* and supplementary fig. S2*A*, Supplementary Material online). With a larger dataset, the *Resande* show a pattern of ancestral components similar to the general Swedish population, but the component maximized in South Asia (rosy brown), which has a high presence in Romani and also has a higher proportion in the *Resande* at K = 3 (10.6% ± 2.3) compared with the European reference populations (4.9% ± 1.4 for the Swedish population) (supplementary fig. S1*B*, Supplementary Material online). These results point to a genetic connection between the *Resande* and the Romani populations that was confirmed with the admixture f3 statistics test, which was used to test admixture between two source populations, showing a significant (z-score = -4.078) negative score of −0.001012 in the form f3(*Resande*; Swedish, Romani), providing evidence of

admixture in the *Resande* between Swedish and Romani populations.

Haplotype estimation was performed to provide a robust fine-scale characterization of ancestry. ChromoPainter and fineSTRUCTURE were used to describe in detail the *Resande* genetic substructure and identify fine-scale patterns. Notable levels of differentiation in the *Resande* were detected with fineSTRUCTURE. *Resande* samples clustered together within the European branch but formed a specific cluster (supplementary fig. S3*A*, Supplementary Material online), close to some Northern and Eastern Europeans (supplementary fig. S3*B*, Supplementary Material online), except for three *Resande* individuals who clustered with most of the Swedish individuals, in the *Swedish1* branch, for which one specific cluster was defined: *Swedish1_RES* (supplementary Table S1, Supplementary Material online). This result suggests a *Resande*-specific genetic profile with high haplotype sharing between them and the rest of the European groups, especially with Northern Europeans. In contrast, Romani individuals from the Iberian Peninsula clustered together within the Middle East/Caucasus reference populations, forming a unique branch divided in two general branches (supplementary fig. S3*A*, Supplementary Material online), which were defined as two different Romani clusters: *Romani1* and *Romani2* (supplementary Table S1, Supplementary Material online). Consistent with this result, ancestry profiles obtained with the non-negative least-square (NNLS) analysis showed that all *Resande* individuals shared most of their haplotypes with Northern European (mostly *Swedish1* and *Swedish2*) and Romani (*Romani2*) clusters, with average proportions of 79.3% and 16.8%, respectively (fig. 2 and supplementary Table S2, Supplementary Material online). This result provides evidence of Romani-related ancestry in *Resande* individuals and shows a very similar pattern among the *Resande* volunteers analyzed.

The Romani cluster that shares haplotypes with the *Resande* (*Romani2*) presents higher proportions of haplotype sharing with clusters related to the Romani-related ancestry (Mendizabal et al. 2012; Bánfai et al. 2018; Font-Porterias et al. 2019) and lower proportions of clusters related to the Iberian Peninsula origin of the Romani samples (supplementary fig. S4, Supplementary Material online and supplementary Table S3, Supplementary Material online). The three individual samples that clustered outside the *Resande*-specific branch are the ones with the highest proportion of haplotype sharing with the Northern European genetic clusters (fig. 2 and supplementary Table S2, Supplementary Material online).

Admixture Events and Demographic Patterns in the *Resande*

Admixture events in the *Resande* were formally inferred after the evidence shown in the previous analyses, using GLOBETROTTER, which enabled the exploration and dating of admixture events as well as the estimation of ancestry profiles. A single admixture event was detected between a Northern European–like major source (Source 1) and a Romani-like minor source (Source 2), with a proportion of 79% and 21%, respectively (fig. 3*A*). Source 1 was mainly formed by surrogate populations from Northern Europe, potentially representing the Swedish population when the admixture event took place. In contrast, Source 2 was formed by surrogate populations associated with the Romani diaspora and their West Eurasian ancestry (Mendizabal et al. 2012; Bánfai et al. 2018), including Southeastern surrogate populations, which was equivalent to the Balkan populations that contributed the most to the West Eurasian ancestry of Romani (fig. 3*A*) (Font-Porterias et al. 2019).

The date inferred for this single-detected admixture event is between 5.6 and 10.1 (95% CI, M = 7.9) generations ago (GA). Considering 28 years per generation, the event is placed in a confidence interval between 1,739 and 1,865 AD (fig. 3*B*).

The effective population size (Ne) estimations through time in the *Resande*, obtained from IBD segments, show a remarkable Ne decrease between 1,500 and 1,800 AD (8–18 GA) with a minimum at 8 generations ago (Ne of 1,060 in a 95% CI of 797–1840). However, this Ne trend increases after the admixture date inferred with GLOBETROTTER (supplementary fig. S5, Supplementary Material online). Similarly, Romani clusters show a long uninterrupted Ne reduction trend, followed by a population size increase with a minimum of around 13–14 generations ago. In contrast, the Swedish population Ne trend does not show a decline in population size, as seen in the *Resande* and Romani.

The genetic relationship among the *Resande*, Romani, and Swedish populations was investigated from the pairwise sharing of IBD segments. The *Resande* internally share more IBD segments than the Swedes (Wilcoxon test *P*-value = 4.19E-21), but less than the Iberian Romani (Wilcoxon test *P*-value = 6.29E-26), in addition to sharing significantly more IBDs with the Romani than with the Swedes (Wilcoxon test *P*-value = 2.53E-71) (supplementary fig. S6, Supplementary Material online).

Runs of homozygosity (ROH) were studied to describe the potential genetic isolation and the levels of inbreeding, that is "mating of individuals or organisms that are closely related through common ancestry" (supplementary File 6, Supplementary Material online), in the *Resande*. The mean total number (NROH) and total length (SROH) of ROHs in the *Resande* in most of the length categories is lower than the values of the Romani reference population (figs. 4*A* and 4*B*), which is known to present higher levels of inbreeding than non-Romani European populations (Font-Porterias et al. 2019). In the intermediate length categories, 1–2 Mb and 2–4 Mb, *Resande* values are slightly but not significantly higher than those of the European reference populations, as shown in figures 4*A* and 4*B*.

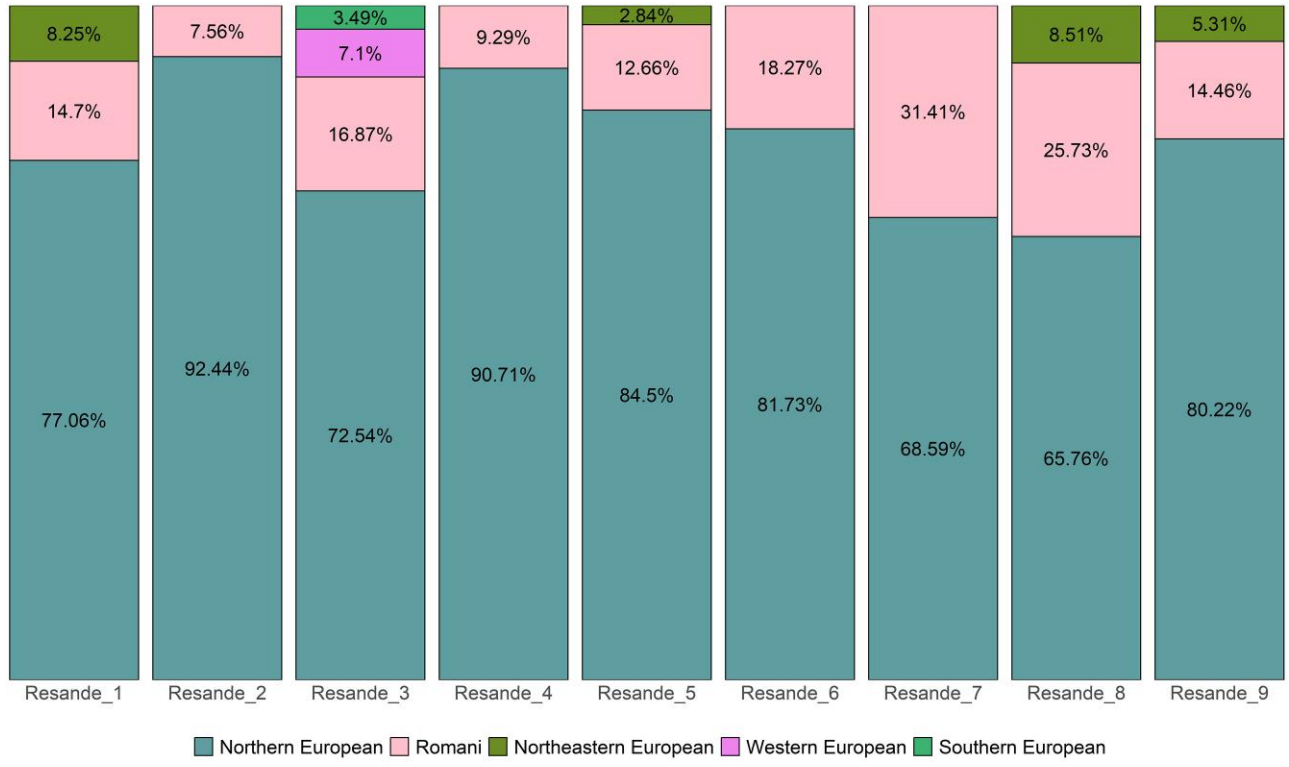

Fig. 2.—Haplotype sharing proportions in *Resande* individuals inferred with the NNLS analysis, considering the Romani genetic clusters as donor groups. Values lower than 2.5% are excluded and the rest are normalized. Major geographic areas are used to group the genetic clusters. The specific proportions for each cluster are shown in supplementary Table S2, Supplementary Material online. Note that the Northeastern European group contains samples from both Northern and Eastern European populations (supplementary Table S1, Supplementary Material online).

Nevertheless, Romani values in these two categories are significantly higher than those of the *Resande* (supplementary Table S4*A* and S4*B*, Supplementary Material online). This result suggests that the levels of inbreeding in the *Resande* population are lower than in the general Romani and more closely resemble those of non-Romani Europeans. Nonetheless, given the small sample size of the *Resande* individuals analyzed, this result should be taken with caution. Furthermore, the small number of individuals (N = 3) represented in the last category (fig. 4*C*) prevents us to make any inferences on recent inbreeding due to long ROHs.

Paternal and Maternal Lineages in the *Resande*

Uniparental lineages are used to assess the genetic diversity and population history of the target populations. Most of the uniparental lineages identified in the *Resande* and Swedish individuals are frequently found in other Northern European populations (Simoni et al. 2000; Jobling & Tyler-Smith 2003). The H haplogroup is the most frequent maternal lineage in both populations, with H1 being the most common and in agreement with that in the literature (Lappalainen et al. 2008). With regard to paternal lineages, the most common haplogroups are I1 and J2a1b in the *Resande* and I1 and R1a in the Swedes (supplementary fig. S7, Supplementary Material online and supplementary Table S5, Supplementary Material online). The haplogroup I1, frequent in Sweden, has its highest frequencies in Northern Europe (Rootsi et al. 2004; Lappalainen et al. 2008). The Y-chromosome haplogroup J2a1b is present in one of three *Resande* samples but absent in Swedish samples. This haplogroup, which originated in West Asia (Singh et al. 2016), has previously been described in Romani in association with their diaspora through the Middle East, Caucasus, and Europe and their subsequent settlement. No *Resande* individual in our dataset has shown South Asian uniparental lineages, not even the maternal and paternal lineages M5a1b and H1a1a4b2, respectively; which have been claimed to reflect the Romani diaspora out of India (Martínez-Cruz et al. 2016; García-Fernández et al. 2020). Nevertheless, the absence of these South Asian lineages in the *Resande* sample might be explained by the limited sample size analyzed in the present study.

Discussion

The genetic study of the *Resande* confirms the Romani-related ancestry of this population, connected to an

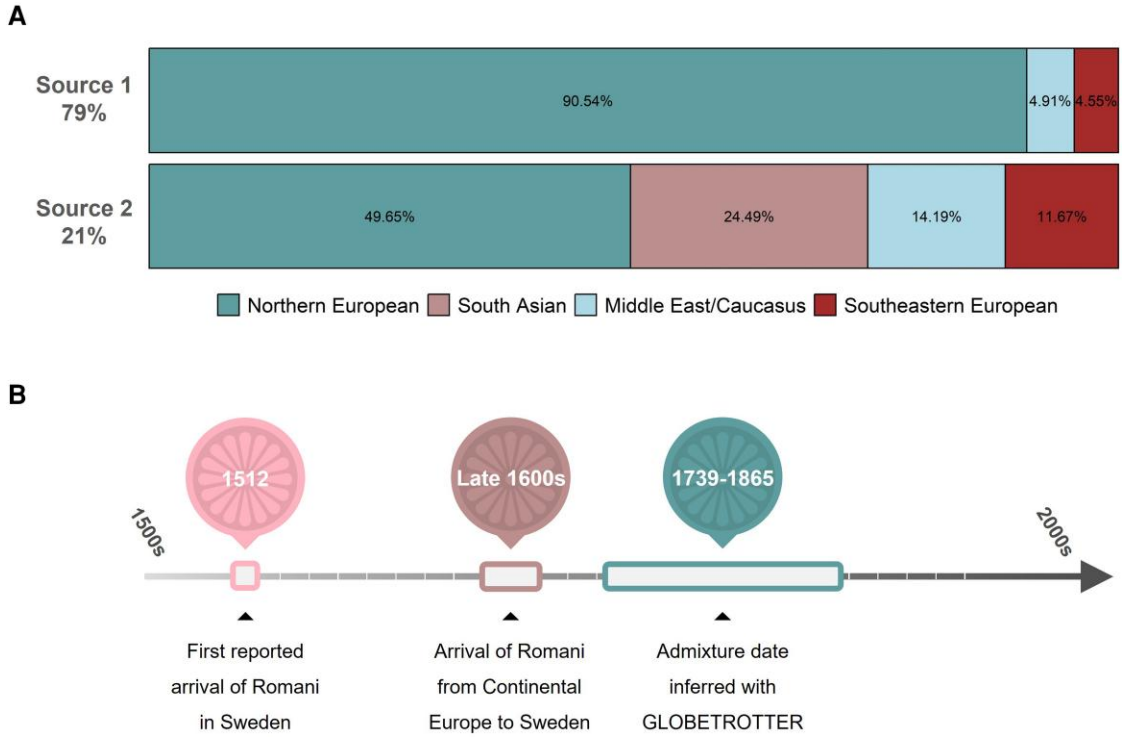

Fig. 3.—Admixture analysis inferred with GLOBETROTTER. (*A*) Estimation of the source populations involved in the admixture event (one date) and the surrogates forming these sources. The Romani genetic clusters are not included as possible surrogates. Genetic clusters are grouped by major geographical areas (supplementary Table S1, Supplementary Material online). (*B*) Timeline of the *Resande* history including the admixture date inferred with a 95% confidence interval, considering 28 years per generation ago, and two reported arrivals of Romani in the country: in 1,512 (Fraser 1992) and in the late 1600 s (Lindwall 2014).

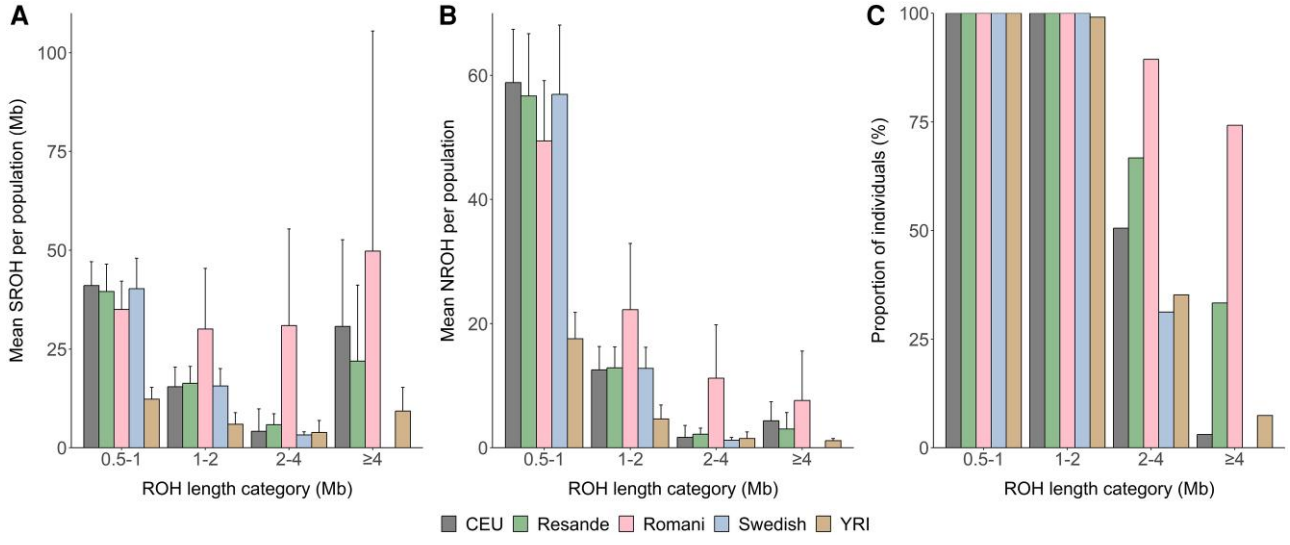

Fig. 4.—Runs of homozygosity (ROH) analysis. (*A*) Average of the total length of ROHs (SROH) per population and length category. (*B*) Average of the total number of ROHs (NROH) per population and length category. (*C*) Proportion of the total number of samples in each length category per population. Two populations with well-known demographic histories are included as references: CEU and YRI.

admixture event that occurred after the arrival of the first Romani in Sweden. Our estimates of the admixture suggest that it did not occur immediately after their first reported arrival, but just over two centuries later, between the mid-18th and mid-19th centuries. Therefore, the admixture took place in one single pulse, which could have involved the descendants of the supposed Romani group of the first migration described in 1,512 and/or the Romani groups from Continental Europe that arrived in the second half of the 17th century. Although there were later Romani migrations such as the East European *Kalderash* Romani during the late 19th century, the Finnish Romani, non-Nordic Romani, and Recent Romani (Government of Sweden: Ministry of Culture 1999), no admixture events involving these later migrations were inferred in our analyses. As a result of the admixture with a non-Romani Northern European population, we detected in the *Resande* a Northern European–related component present with a higher proportion than the Romani component.

Maternal and paternal lineage analyses in our *Resande* samples have detected a Caucasian paternal lineage (i.e., J2a1b) previously detected as founder in the Romani (Martínez-Cruz et al. 2016; Singh et al. 2016; García-Fernández et al. 2020), which might indicate their Romani ancestry, although South Asian lineages previously described in European Romani groups are absent and most of the identified uniparental lineages in the *Resande* are common in other Northern European populations. Finally, the absence of Romani founder lineages in the mtDNA suggests possible sex-biased patterns in which European gene flow is more notable in maternal lineages, previously traced in Romani due to sociocultural factors (García-Fernández et al. 2020). However, this must be taken with caution due to the limited number of samples.

Regarding the demographic estimates, the effective population size dynamics in the *Resande* changed the course with the admixture event, after an initial reduction trend, which might be the result of a bottleneck, a founder effect, or the social exclusion and endogamy practices (i.e., within-population marriages) described in the *Resande* (Heymowski 1969), although the IBD sharing within the *Resande* population is less than that within the Romani reference population and the levels of inbreeding detected with the ROH analyses are subtle and more similar to those of non-Romani European populations.

The Ne reduction trend has also been reported in isolated populations and regions in Northern Europe, such as in some counties of Norway or the Orkney Islands in the UK, although this is not a general trend (Xue et al. 2017; Mattingsdal et al. 2021). Regarding the Romani groups, a period of Ne reduction has been previously reported to reverse its course with the settlement of Romani in Europe, the beginning of more intense politics of assimilation, and the admixture events that took place later. This decrease was probably caused by multiple founder events in their expansion, derived from social and political persecution with laws concerning this ethnic minority, and/or endogamy practices (Fraser 1992; Font-Porterias et al. 2019, 2021). Therefore, we could suggest that the demographic patterns found in the *Resande* are similar to those in most European Romani groups, and these patterns have been influenced by discriminatory and repressive measures, together with social marginalization.

Given that this study has now demonstrated that the *Resande* have Romani-related ancestry, further studies could investigate the internal heterogeneity of the population by improving the sampling strategy. Many factors that have not been considered in this study, such as geographic patterns, surnames, lifestyle, and traditional professions, may have shaped their current genetic landscape, so it would be worthwhile to include these elements in the future to fully characterize the population history of the *Resande*. In addition, enlarging the sample size of the *Resande* and general Swedish populations, and including different Romani reference populations, one more similar to the supposedly Romani group that arrived in Sweden in the early 16th century from Scotland and England (e.g., *Romanichals*, also known as English Travelers) (Fraser 1992) and one group representing the Romani from Continental Europe that arrived during the 1600 s (e.g., Sinti Romani) (Lindwall 2014), would help to refine the genetic origins and ancestry of this population. Moreover, we could investigate the connection between the *Resande* and Finnish Kale Romani, whose historical and genealogical research, including DNA genealogy, suggest that they originated from the same population diverging after Finland was ceded to Russia in 1,809 (Fraser 1992; Machiels 2002).

In the broader context of human genetics studies, this work highlights the importance of studying ethnic minorities to better understand the complete history of human populations, thereby enhancing the interpretation of biomedical studies (Popejoy & Fullerton 2016).

Although the present study demonstrates the Romani-related ancestry of the *Resande* or *Resandefolket*, it aims to characterize and describe this population from a genetic perspective only, without any intention of using genetics to supplant culture or history as major *Resande* identifiers. The results obtained could be useful to support the historical records and better understand the past of this population, but they should not be construed as a denial of their singularity as a minority ethnic group. In order to address possible concerns and to present our results to a nonspecialist audience, we have prepared a glossary, a summary, and some questions and answers that are available in English and Swedish (supplementary Files 2–7, Supplementary Material online).

Materials and Methods

Dataset Description and Quality Control

Saliva samples from 13 *Resande* and 18 general Swedish volunteers were collected with appropriate written informed consent (supplementary File 1, Supplementary Material online) from unrelated volunteers self-identified as members of each group. The project was supported by the *Frantzwagner Sällskapet* Association and the corresponding IRB approvals (*Comitè Ètic d'Investigació Clínica (CEIC)-Parc de Salut Mar, Institut Hospital del Mar d'Investigacions Mèdiques*, Barcelona, 2016/6723/I and 2019/8900/I), and all methods adhered to the tenets of the Declaration of Helsinki. For the sake of clarity, note that the label "Swedish" is used to refer samples from the general non-Romani Swedish population. Additionally, 83 Romani samples previously analyzed (Flores-Bello et al. 2021; Font-Porterias et al. 2021) were used for comparison. DNA was extracted from saliva samples and genotyped with the Axiom Genome-Wide Human Origins Array (Patterson et al. 2012). Genotype calling was performed by using the software Axiom Analysis Suite 3.1.51.0. Samples that passed the genotype calling (10 *Resande*, 16 Swedish, and 75 Romani) had an average quality control call rate of 99.2%. After passing the suggested thresholds, 600,225 autosomal single-nucleotide polymorphisms (SNPs) remained in our dataset.

Samples were merged with publicly available data from Europe, Middle East, and Caucasus to represent the whole Western Eurasian context (Lazaridis et al. 2016) and PJL and ITU as proxies for the South Asian ancestries of Romani (Altshuler et al. 2012).

Quality control filtering was applied with PLINK v1.9 (Purcell et al. 2007) to filter out genotyping errors: individuals missing more than 10% genotype data, SNPs missing in more than 5% of the samples, and SNPs failing the Hardy–Weinberg exact test with a P-value $< 10^{-5}$. Ten individual samples (1 *Resande* and 9 Romani) from our dataset were removed due to high genetic relatedness to other samples (PI_HAT > 0.125), ending up with 9 *Resande*, 16 Swedish, and 66 Romani samples. Once the datasets were merged, 437,387 variants remained after removing those with a minor-allele-frequency (MAF) below 0.01 [non-linkage disequilibrium (LD)-pruned data set]. A LD pruning was performed for allele frequency methods with a window size of 200 SNPs, shifting by 25 SNPs, and a maximum pairwise LD threshold (r^2) of 0.5, resulting in 189,826 variants (LD-pruned data set). Thus, the final datasets resulted in 1,071 individuals from 54 different populations and the previous number of variants.

Allele Frequency Methods

To inspect the population structure of the *Resande,* a Principal Component Analysis (PCA) was performed using SmartPCA from EIGENSOFT v6.0.1 (Patterson et al. 2006) in two different situations: including all populations described in the previous section and including only European references, without Indian, Middle Eastern, and Caucasian samples.

To examine the patterns of the population structure, ADMIXTURE v1.3.0 (Alexander et al. 2009) was run for 10 independent iterations, with K ancestral components from 2 to 12. Then, PONG v1.4.9 (Behr et al. 2016) was used to combine the different iteration results of ADMIXTURE, in order to obtain the major modes for each K and plot them.

The qp3Pop programme from AdmixTools (Patterson et al. 2012) was used to test admixture between two source populations in the form f3(*Resande*; Swedish, Romani), with the *Resande* set as the target population.

Haplotype-Based Methods

SHAPEIT v2 (O'Connell et al. 2014) was used to phase the data, using the HapMap phase III (International HapMap Consortium 2003) genetic map and the 1000G dataset (Altshuler et al. 2012) as the reference panel. First, the data were aligned to the reference, but because our dataset already contained samples from 1000G, no SNPs were removed due to mismatches. Haplotype sharing between individual samples was estimated with ChromoPainter (Lawson et al. 2012), which reconstructs chromosomes of each target or "recipient" sample, as a mosaic of haplotypes from reference or "donor" samples. First, it was run to estimate the global mutation probability (M) and the switch rate (n) parameters, in chromosomes 1, 7, 14, and 20, for 15 iterations of the expected-maximization (EM) algorithm, with parameters -in -iM. Therefore, the parameters obtained were averaged across chromosomes weighting by the number of SNPs per chromosome, to obtain the final parameters. The resulting parameter values of the switch rate (n) and the global mutation (M) were 258.430092 and 0.000568, respectively. Then, ChromoPainter was run adjusting the estimated parameters, to obtain the final count and length-sharing coancestry matrices. ChromoCombine was used to combine the matrices of all chromosomes.

FineSTRUCTURE v2.1.0 (Lawson et al. 2012) was used to group the resulting data from ChromoPainter into homogenous genetic clusters. First, it was run for 2 million Markov Chain Monte Carlo (MCMC) iterations with 1 million "burn-in" iterations and sampling values every 10,000 iterations. FineSTRUCTURE dendrograms were built with the default parameter -m T. The analysis was performed for three different seeds in both chunkcounts and chunklengths sharing coancestry matrices from ChromoPainter. After checking the consistency of the dendrograms between different seeds, we used the

chunkcount dendrograms as references to define the genetic clusters (supplementary Table S1, Supplementary Material online). For the three *Resande* individuals who did not cluster with the others in a specific branch, one unique cluster was defined: *Swedish1_RES*.

Next, ChromoPainter was run again indicating which genetic clusters would be donor groups. In this case, we first set all clusters as donors except *Resande* clusters, and, second, all clusters as donors, except *Resande* and Romani clusters.

GLOBETROTTER (Hellenthal et al. 2014) was used to explore and date admixture events and estimate ancestry profiles. First, the NNLS method was run to infer the haplotype sharing proportions of the different genetic clusters per individual sample.

Moreover, GLOBETROTTER was run to characterize admixture events in our target genetic cluster (*Resande*), using as surrogate populations all donor clusters, in this case without considering Romani clusters as donors, because it has been shown that Iberian Romani are admixed with non-Romani European populations (Font-Porterias et al. 2019). First, the null.ind parameter was set to 1 in order to estimate *P*-values for admixture evidence. Second, the null.ind parameter was set to 0 to estimate proportions, dates, and sources of admixture. The 95% confidence interval (CI) and the median (M) for the date estimate were inferred by performing 100 bootstrap iterations.

The effective population size (Ne) through generations was estimated for the Romani genetic clusters, *Romani1* and *Romani2* (supplementary Table S1, Supplementary Material online), and the *Resande* and Swedish populations, from identity-by-descent (IBD) segments. IBDseq (Browning & Browning 2013) was used to detect IBD segments, with default parameters, from the non-LD-pruned dataset. Then, IBDNe (Browning & Browning 2015) was run using the IBD segments obtained with a minimum length of 2 cM and the HapMap GRCh37 genetic map (International HapMap Consortium 2003).

The IBD segments inferred with IBDseq were used to explore the IBD sharing between pairs of individuals, after converting base pairs to genetic positions in centiMorgans using the HapMap GRCh37 genetic map (International HapMap Consortium 2003). To construct an IBD heatmap, we excluded IBD segments shorter than 3 cM and then we summed the IBD pairwise lengths between individuals (Han et al. 2017). To test whether the IBD sharing was significantly different, we performed Wilcoxon tests.

In order to estimate ROHs with PLINK 1.9 (Purcell et al. 2007), external populations from 1000G (Altshuler et al. 2012) were included to test for inbreeding: Utah residents with Northern and Western European ancestry (CEU) and Yoruba in Ibadan, Nigeria (YRI), as reference populations with and without the "Out-of-Africa" bottleneck, respectively. The ROH analysis was performed for the LD-pruned data set considering ROHs with at least 50 SNPs, a minimum length of 500 kb, and a maximum gap between two consecutive SNPs of 100 kb.

Uniparental Markers Analyses

Mitochondrial DNA (mtDNA) was PCR-amplified with identical conditions in four segments using four primer pairs (García et al. 2021). Genetic libraries were prepared following the *Illumina mtDNA Genome guideline* (Illumina 2016) and sequencing was performed according to the *Illumina MiSeq guideline* (Illumina ® Inc. 2018). Mapping was done following the GATK best practices (Van der Auwera et al. 2013) and individual haplogroups were assigned with HaploGrep (Weissensteiner et al. 2016). To select good-quality mtDNA sequences, a minimum of 15× of coverage in all four amplified regions and a HaploGrep quality score of at least 80% were considered (supplementary Table S6, Supplementary Material online) (Seo et al. 2015). A total of 10 of 12 *Resande* and 17 of 18 Swedish samples passed the quality thresholds.

For the Y-chromosome analysis, 17 short tandem repeats (STR) present in the AmpFLSTR Yfiler PCR Amplification Kit (Thermo Fisher Scientific) were genotyped following the manufacturer's recommendations. The Whit Atheys' haplogroup predictor (http://www.hprg.com/hapest5/) (Athey 2006) was used to predict Y-chromosome haplogroups from the STRs.

Supplementary material

Supplementary data are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

We would like to thank all volunteers who contributed with their DNA samples and who made this study possible and acknowledge the constant support of the Romani associations *Frantzwagner Sällskapet* and *Federació d'Associacions Gitanes de Catalunya* (FAGiC). We specially thank Jon Pettersson and Egil Asprem for their support.

Funding

This study was supported by the Ministerio de Ciencia, Innovación y Universidades (MCIU, Spain) and the Agencia Estatal de Investigación (AEI) grant numbers CGL2016-75389-P and PID2019-106485GB481I00/AEI/10.13039/501100011033, and "Unidad de Excelencia María de Maeztu" (MCIN and AEI, DOI:10.13039/501100011033, CEX2018-000792-M).

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability

Genome-wide genotype array data and whole mtDNA sequences have been deposited at the European Genome-phenome Archive (EGA), under accession number EGAS00001006176.

Literature Cited

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

Altshuler DM, et al. 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491:56–65.

Athey TW. 2006. Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. J Genet Geneal. 2: 34–39.

Bakker P. 2020. Para-Romani varieties. In: The palgrave handbook of Romani language and linguistics. London: Palgrave Macmillan. p. 353–386. doi: 10.1007/978-3-030-28105-2

Bánfai Z, et al. 2018. Revealing the impact of the Caucasus region on the genetic legacy of Romani people from genome-wide data. PLoS One 13:e0202890.

Behr AA, Liu KZ, Liu-Fang G, Nakka P, Ramachandran S. 2016. Pong: fast analysis and visualization of latent clusters in population genetic data. Bioinformatics 32:2817–2823.

Bojs K, Sjölund P. 2016. Svenskarna och deras fäder de senaste 11 000 åren. Stockholm: Albert Bonniers.

Browning BL, Browning SR. 2013. Detecting identity by descent and estimating genotype error rates in sequence data. Am J Hum Genet. 93:840–851.

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.

Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. 2018. Runs of homozygosity: windows into population history and trait architecture. Nat Rev Genet. 19:220–234.

Flores-Bello A, Font-Porterias N, Aizpurua-Iraola J, Duarri-Redondo S, Comas D. 2021. The genetic scenario of mercheros: an under-represented group within the Iberian Peninsula. BMC Genomics 22:1–9.

Font-Porterias N, et al. 2019. European Roma groups show complex west eurasian admixture footprints and a common south Asian genetic origin. PLOS Genet. 15:e1008417.

Font-Porterias N, et al. 2021. The counteracting effects of demography on functional genomic variation: the roma paradigm. Mol Biol Evol. 38:2804–2817.

Fraser A. 1992. The gypsies. Oxford: Wiley-Blac.

García-Fernández C, et al. 2020. Sex-biased patterns shaped the genetic history of roma. Sci Rep. 10:14464.

García A, et al. 2021. Ancient and modern mitogenomes from central Argentina: new insights into population continuity, temporal depth and migration in South America. Hum Mol Genet. 30:1200–1217.

Government of Sweden: Ministry of Culture. 1999. Nationella minoriteter i sverige proposition: legal documents. Stockholm: Government Office of Sweden.

Government of Sweden: Ministry of Culture. 2014. The dark unknown history: white paper on abuses and rights violations against roma in the 20th century. Ministry Publications Series (Ds): Stockholm.

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

Hancock IF. 2002. We are the Romani people. Hatfield: University of Hertfordshire Press.

Hellenthal G, et al. 2014. A genetic atlas of human admixture history. Science 343:747–751.

Heymowski A. 1955. Om 'tattare' och 'resande'. Uppsala: Sociologiska Institutionen.

Heymowski A. 1969. Swedish 'Travellers' and their ancestry: a social isolate or an ethnic minority? Uppsala: Almqvist & Wiksells Boktrycheri AB.

Humphreys K, et al. 2011. The genetic structure of the Swedish population. PLoS One 6:e22547.

Illumina. 2016. Human mtDNA Genome Guide (15037958). https://support.illumina.com/downloads/human_mtdna_genome_guide_15037958.html.

Illumina ® Inc. 2018. Illumina® MiSeq System Guide. https://www.illumina.com/company/legal.html.

International HapMap Consortium. 2003. The international HapMap project. Nature 426:789–796.

Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet. 4: 598–612.

Lappalainen T, et al. 2008. Migration waves to the baltic sea region. Ann Hum Genet. 72:337–348.

Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of population structure using dense haplotype data. PLoS Genet. 8: 1002453.

Lazaridis I, et al. 2016. Genomic insights into the origin of farming in the ancient near east. Nature 536:419–424.

Lindgren L, Lindwall B. 1992. Resande anor. In: Hjord B, editors. Migration: utvandrare och invandrare i gångna tider. Stockholm: Sveriges Släktforskarförbund. p. 71–104.

Lindholm A, Svanberg I. 1988. Tattare. In: Svanberg, I, Runblom, H, editors. Det mångkulturella sverige. En handbok om etniska grupper och minoriteter. Stockholm: Gidlund. p. 416–420.

Lindwall B. 2014. Anor från landsvägen. Hur jag finner mina förfäder bland resandefolket. Solna: Sveriges släktforskarförbund.

Machiels T. 2002. Keeping the distance or taking the chances. Roma and travellers in Western Europe. Brussels: European Network Against Racism (ENAR).

Martínez-Cruz B, et al. 2016. Origins, admixture and founder lineages in European roma. Eur J Hum Genet. 24:937–943.

Matras Y. 1995. Romani in contact: the history, structure and sociology of a language. Amsterdam: John Benjamins Publishing.

Mattingsdal M et al. 2021. The genetic structure of Norway. Eur J Hum Genet. 2021. 29:1710–1718.

Mendizabal I, et al. 2012. Reconstructing the population history of European Romani from genome-wide data. Curr Biol. 22:2342–2349.

Minken A. 2009. Tatere i norden før 1850. Sosio-økonomiske og etniske fortolkningsmodeller. Tromsø: University of Tromsø.

O'Connell J, et al. 2014. A general approach for haplotype phasing across the full Spectrum of relatedness. PLoS Genet. 10: 1004234.

Patterson N, et al. 2012. Ancient admixture in human history. Genetics 192:1065–1093.

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

Popejoy AB, Fullerton SM. 2016. Genomics is failing on diversity. Nature 538:161.

Purcell S, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 81: 559–575.

Rootsi S, et al. 2004. Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. Am J Hum Genet. 75:128–137.

Seo SB, et al. 2015. Underlying data for sequencing the mitochondrial genome with the massively parallel sequencing platform Ion TorrentTM PGMTM. BMC Genomics 16:24–27.
Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G. 2000. Geographic patterns of mtDNA diversity in Europe. Am J Hum Genet. 66:262–278.
Singh S, et al. 2016. Dissecting the influence of neolithic demic diffusion on Indian Y-chromosome pool through J2-M172 haplogroup. Sci Rep. 6:1–10.
Svensson B. 1993. Bortom all ära och redlighet. Tattarnas spel med rättvisan. Stockholm: Nordiska Museets Förlag.
Tervonen M. 2010. 'Gypsies', travellers' and 'peasants'. A study on ethnic boundary drawing in Finland and Sweden, c.1860-1925. Florence: European University Institute.
Van der Auwera GA, et al. 2013. From fastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr. Protoc. Bioinforma 43:11.10.1–11.10.33.
Weissensteiner H, et al. 2016. Haplogrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. 44:W58–W63.
Wiedner J. Factsheets on Roma: travellers in Norway and Sweden. Counc. Eur. http://romafacts.uni-graz.at/?c=culture&ch=3&f=4.
Xue Y, et al. 2017. Enrichment of low-frequency functional variants revealed by whole-genome sequencing of multiple isolated European populations. Nat Commun. 8:1–7.

Associate editor: Dr. Kirk Lohmueller