

RomIdent Working Papers

RomIdent
Working papers

Paper No.31

Romani genetic linguistics and genetics: Results, prospects and problems

Peter Bakker

(Aarhus University)

2012

Romani Studies 5, 22: 91-111.

<http://romani.humanities.manchester.ac.uk/virtuallibrary>

The project RomIdent is financially supported by the HERA Joint Research Programme (www.heranet.info) which is co-funded by AHRC, AKA, DASTI, ETF, FNR, FWF, HAZU, IRCHSS, MHEST, NWO, RANNIS, RCN, VR and The European Community FP7 2007-2013, under the Socio-economic Sciences and Humanities programme.



EUROPEAN COMMISSION
European Research Area



Funded under Socio-economic Sciences & Humanities

Romani genetic linguistics and genetics: Results, prospects and problems

PETER BAKKER

Much progress has been made recently in the migration history of the Roms based on language data. Genetic research on the Roms, using DNA data, sheds light on the same, by studying connections between European Roms and local populations, between Roms in different locations, and between Roms and Asian groups. The genetic data confirm a migration of the ancestors of the Roms out of South Asia, through southern Persia. In most locations of settled populations in Europe, the Roms appear rather different from the local non-Roms and from other Romani groups. The groupings used by geneticists often differ from those used by linguists, and the density of data, too, differs between the two disciplines. This leads to problems of different kinds in comparing the genetic and linguistic results on migration patterns. Generally, the data point to the same historical scenario, including the dating of linguistic and genetic splits.

Keywords: dialects, genetics, geography, Gypsies, history, language, linguistics, migration, Romani, Roma

Until now, most findings on the origin of the Roms and the history and splits of Romani groupings have been established by linguistic research. An Indian origin of the Romani language was proven several centuries ago, and migration patterns were reconstructed subsequently with linguistic data. Archaeological or historical corroboration was almost completely absent. More recently, another branch of science, human genetics, has contributed additional evidence to these historical findings, especially the implications of DNA research. This article presents a brief summary of the rather extensive genetic literature on Roms, and discusses the compatibility of linguistic and genetic data. Two types of evidence are dealt with: the mutation history and the distribution of genetically transmitted diseases.

The results of linguistic and genetic research appear to be compatible: genetic research proves an Indian origin, subsequent splits and the founding of groupings by limited number of ancestors and often limited mixture with outsiders. However, a number of methodological problems are identified, especially the way of sampling by geneticists and the fact that some genetic research is not done independently from linguistic findings. If performed well, genetic and linguistic research can complement one another, and promising findings are ahead.

Peter Bakker is Associate Professor in Linguistics at Aarhus University, DK-8000 Aarhus, Denmark. Email: linpb@hum.au.dk

Romani Studies 5, Vol. 22, No. 2 (2012), 91–111

ISSN 1528–0748 (print) 1757–2274 (online)
doi: 10.3828/rs.2012.6

1. Introduction

The question where the Roms come from has long fascinated both students of the Roms and the Roms themselves. There is evidence that they arrived in Europe in the late Middle Ages and that the language originates in India. However, there are no unambiguous historical documents that mention a massive exodus from South Asia or that describe a sizable group travelling between India and Europe. Linguistic research, however, has enabled us to reconstruct a migration route between India and Europe. After speakers of an early form of Romani arrived in Europe, they spread through the continent, which, naturally, led to the development of different varieties of the language.

In this article I will try to shed new light on the history of the migration of the ancestors of the Roms from India, and their subsequent migrations, by reviewing the increasing body of literature on recent genetic research, and compare it with what linguistic research has established. In Section 2 I first sketch the historical background of linguistic research. Well aware that genetic research is controversial, even more so in the case of groups like the Roms, I will discuss the choice of these data in Section 3. In section 4 I deal with several ways in which genetic data can shed light on what we already know from linguistics, and compare changes in language with changes in genes. Section 5 deals with the history of Roms as revealed in DNA. I will focus on the Indian origin, the journey through Iranian speaking territory, evidence from genetic diseases, the comparison between Roms and neighbouring populations, and the relation between the genetic and dialectal diversity of Romani groups. In section 6 I compare dating proposals from genetic and linguistic studies. I end with conclusions (Section 7) and perspectives for further research (Section 8).

2. Background

From the moment the Roms appeared in Europe, people have wondered where the groups of people came from who dressed differently, who looked differently and who spoke a language nobody could understand. When asked that question, the Gypsies answered in a number of cases that they were from “Little Egypt”. Probably one of the first vocabularies of Romani was therefore printed more than 500 years ago in a book about Africa, but it was concluded that the language did not resemble any known African language. However, the lack of similarities with any African language, including the languages of Egypt, made scholars, already at that time, wonder where these people were from, if not from Egypt – to the extent of course that language can give a clue.

It was not until more than 250 years later that at least the riddle of the language was solved. Within one decade around 1780 a number of people claimed

that Romani was related with languages spoken in India. It was only in that time that information about the languages of the Indian subcontinent became more easily available. Jacob Bryant in England and Simon Peter Pallas, together with Catherine the Great, noticed the striking similarities between Indian words and Romani words, while in Germany Johann Rüdiger (Matras 1999a) made the proof complete by showing that not only the words, but also important parts of the Romani grammar, such as the verbal endings, were so close to Hindustani, an Indic language, that there is no doubt that the language has the same origin as Indian languages such as Hindustani, Punjabi and Bengali, all of which are historically linked to Sanskrit.

But the fact that the European language Romani has sister languages in India does not necessarily mean that the ancestors of the speakers of the Romani language ever lived in India. Just as the ancestors of speakers of African languages had shifted to the English language, once transported to North America, it is at least theoretically possible that ancestors of the Roms had picked up the language somewhere else.

For a long time we had only linguistic data that could shed light on this matter. In fact, the data from the language suggested not only an Indian origin, but also a migration route between India and Europe (early accounts in Pischel and Miklosich 1872–1880). Words for body parts, lower numerals, kinship terms, common actions and other terms used on a daily basis are clearly Indic in Romani (Boretzky 1992, Matras 2002: 20–30); they are the words least subject to borrowing from other languages. Apart from these, there are common words shared by most of the Romani dialect groupings that Romani has in common with languages like Persian, Kurdish, Armenian and Greek. If one plots the locations on a map, one can draw a migration route from India, via northern India, through Iran and Kurdistan, through Armenia and then through Greek-speaking territory (see e.g. Campbell 1998, Tcherenkov and Laederich 2004) – which may have been parts of present-day Turkey, where Greek used to be dominant. As the set of words taken from these languages is basically the same in all Romani groups, we can conclude that the ancestors of the Roms constituted one community in their migration from India to Europe. All this can be conjectured on the basis of the language data only.

India, Persia, Armenia and Greece have traditions of writing that go back thousands of years. From documents in the local languages one can deduct changes in the local languages and try to date the changes that took place, especially when the documents can be dated. For instance Old Armenian /l/ shifted to a fricative /ɣ/, but the Armenian loans in Romani have /l/ and not, for instance, /x/ in these words. Then we can conclude that the ancestors of the Roms left the Armenian speaking areas before that change took place in Armenian, that is, before 900 CE (Tcherenkov and Laederich 2004). These

linguistic techniques allow linguists not only to reconstruct a migration route, but also to put tentative dates to the migrations of the ancestors of the Roms. Linguistic researchers tend to agree on the route (see Campbell 1998 based on Terrence Kaufman's research; Matras 2002) but not on the timing. Some date the exodus from central or north India at around 1000 CE (e.g. Hancock 2000), others 1,000 years or more earlier (Metzger 2004 on the basis of sound changes in early Indic inscriptions; Campbell 1998, citing Kaufman, based on facts like grammatical influences, parallel sound changes and the absence of Arabic loans). Recent genetic research, however, sheds light on this question.

3. Genetics

Humans, and in fact all living organisms, contain rather detailed data about their ancestors in their body. This information is stored in our genes, in our DNA. There are three relevant types of DNA which contains our genes. One inherits most DNA from both parents, and from both parents exactly half, in an arbitrary way. Other parts are inherited always from the mother, or always from the father. This knowledge can be used, for instance, in tests for fatherhood: if one compares the DNA of a child with the DNA of a father, one can establish that a certain person is or is not the father of a particular child. And if one can do it for one generation, one can also do it for two, three and many generations. Of course, the further one goes back in time, the more uncertainties one is confronted with.

If one can do this at a time depth of one generation and one can do this over hundreds, perhaps thousands of generations, one can also do this for the history of the last 1,000 or 2,000 years. If we have data on the DNA of different Romani groups and individuals, we can try and investigate the origin and history of Romani people rather than of the Romani language. One should distinguish between linguistic, ethnic and genetic groupings. Speakers of a common language may not coincide with genetic groups, and neither of them equate with ethnic groups. There is no space here to elaborate on the problems associated with this. Here we deal with links between linguistic groupings (Romani speakers) and the ethnic overarching group of the Gypsies. One aspect of Gypsy ethnicity is a common descent group, and a common descent group implies some common genetic features. Medical researchers and geneticists have investigated the question of common genetic features and other matters relating to the genetics of the Roms.

Genetic research has not been able to establish that one type of people is in some way better than another. Actually, the visible differences between humans appear to be very superficial from a genetic point of view, and in fact research points to the unity of mankind. This contrasts with the more biased

and ideological biological opinion of the first half of the twentieth century. In this article only more modern genetic research is consulted to the extent that this research can shed light on questions of the origin and migrations of the Roms.

4. Genetics and linguistics and the study of change

Both (historical) linguistics and genetics deal with changes. Once a change takes place in a speech community it will be transmitted to the next generation. Thus, if a change takes place in community A and not in community B, and A splits into C and D, researchers can conclude that A, C and D share part of their more recent history. Other communities may not have undergone the same change, so this technique makes it possible to reconstruct demographic developments and population movements and contacts.

Geneticists can reconstruct the history of individuals, families, groups and populations by comparing the DNA and isolate changes, and perhaps date them. Here too, genetic changes will be transmitted only to direct descendants, thus enabling researchers to track population movements and intermarriage patterns.

In the study of changes, genetics and linguistics share common grounds. Both research fields shed light on historical relations between groups: specific features between DNA or language varieties must be inherited or taken over from other groups, proving a historical connection of inheritance or contact. Both genetic and linguistic changes allow for relative dating and absolute dating. Relative dating (first A changed to B, then B to C) means that one can determine in which order changes have taken place, by using certain techniques.

There are also differences between genetics and linguistics: only genetics can distinguish men's from women's history, as some of the DNA is transmitted only via mothers and other DNA is transmitted through fathers. This makes it possible, for instance, to draw conclusions about historical asymmetries in marriage patterns: if only female genetic lines from one particular group are found in another group, (some of) the women apparently found their grooms elsewhere. One only needs to resort to these techniques for the historical period since interviews and fieldwork in the communities reveal current and recent marriage patterns of a Romani group (see e.g. Jakoubek and Budilová 2006).

Second, languages are transmitted socially, genes are transmitted biologically. Any individual and any group can shift to another language without leaving any trace of the original language, but one's genes cannot be changed. In that sense, genetics is a more reliable indicator of historical developments than

linguistics. In the strict biological sense, only vertical spread, not horizontal spread, is possible in genetics, whereas people can learn other languages, and add them to their repertoire, or replace them with others – a social process.

The interplay of genetics and linguistics could shed some light on linguistic and dialectological questions. The traditional view stresses the divergence of dialects, minimalizing the influence on each other after separation. Boretzky and Iglá (2004) offer such a view on Romani dialects. Matras (2002, 2005) holds a different, more areal view of dialect classification. Elšík (2006), too, has criticized the divergence view, pointing out clear areal spread in some cases, and the fact that some dialects are transitional, fitting the geographical location, but not fitting the branching metaphor of splits into distinct dialect groupings. Even the existence of some isogloss bundles would not define genetic groupings, since many other isoglosses deviate from those bundles as well, crosscutting groupings. Romani shows clear geographical patterns in its dialect distribution. These differences are more often caused by diffusion from one or more centers to neighboring communities (see Matras 1999b). Even though Matras does not completely abandon branching trees, he emphasizes the geographical patterns. This reflects the traditional difference between the tree model and the wave model in historical linguistics. Linguistic features can spread from one dialect to another only when there is contact between speakers of the two dialects. Perhaps they meet each other for business, or they have closer contacts and intermarry. Perhaps both men and women marry people from the other group, perhaps only women from one group marry men from the other group.

As geneticists can establish the gene flow between such groupings – at least to some extent – their research can help conclude whether dialect features have spread because of casual contact or because of more intensive contact, for instance in the case of mixed marriages. Romani groups as a rule are endogamous. Endogamy often means that the genetic features of a group are quite distinct from those of other Romani groups, in a process called “drift” by geneticists (e.g. Wells 2002: 19): subgroups become genetically homogenous, and contrast increasingly with other subgroups.

5. The history of the Roms as revealed in DNA

Even though linguistic data point unambiguously to an Indian origin of the Romani language, there are no archaeological traces of the journey of the ancestors of the Roms before they arrived in Europe. And even though the reconstructed route must have led them through regions where people were literate long before people were so in Europe, there are no unambiguous written references to a considerable group of people of Indian origin who certainly were ancestral to

the Roms. No historical documents in Persia, Armenia, or Byzantium have been unearthed that shed light on the exodus from India to Europe, a journey one would presume must have taken at least several centuries. There are documents relating to the immigration of Indian populations into Persia, but it cannot be established that these people were the ancestors of the Roms.

DNA can shed light on this question. It can reveal where an individual has his or her roots. Genetic research on the Roms has started several decades ago. Some 300 publications have been published in specialist medical and biological journals. In this article I will summarize a number of recent studies from that realm.

The medical-genetic literature on the Roms deals with at least three areas. First, the state of health of the Roms. Researchers express their concern that Roms generally are much less healthy than the rest of the population in most countries. This, however, even though it is very important, does not reveal anything about the origin of the Roma, and will not be dealt with here. Overviews have been made by Hajioff and McKee (2000) and Zeman et al. (2003).

Second, the DNA of Roms or specific Romani groups is being compared with other groups, most typically the surrounding (non-Romani) populations, populations in India and/or Roms in other countries.

A third branch of research focuses on hereditary diseases that are caused by a genetic change. Every small change in a person's DNA will have a 50 percent change to be inherited by his or her children. This research has an important practical applicability. If such a DNA change can cause a defect or an illness, medical researchers can try to find a cure for the disease. Alternatively, these results can be used as a source for the history of Romani groups. The spread through the Romani groups and populations may reveal the spread of population groups and intermarriage patterns of the past, and thus shed a light on the migration history of the Roms.

In this section, we will use some of the studies of the second and third type to investigate the migration history of the Roms. We will try to compare these with the rather detailed findings from linguists about possible migrations. In both types of research, there are contradictions and uncertainties with regards to the results. There are both discrepancies and similarities between the results from the two disciplines.

5.1. Indian origin

Genetic studies establish beyond any doubt that the bulk of the ancestors of the Roma, in both male and female line, have their origins in India. This is especially clear in the high frequency of Haplogroup M in the female lineage, and Haplogroup H in the male lineage. Haplogroups can be said to be genetic descent groups, analogous to dialect groups in linguistics, but here based on

genetic features. The members of haplogroups M and H all descend from a single ancestor in whose genes an irreversible change has taken place that has been inherited by all of his or her descendants.

Kalaydjieva et al. (2005) report that 30 percent of the (Bulgarian) Roms they studied have Haplogroup M ancestry in their female ancestry. This haplogroup is very common in India (Thangaraj et al. 2006), independent of the affiliation of the languages spoken by the populations. In the southern part of Pakistan and northwest India, 50 percent of the population belong to this haplogroup. This haplogroup is also present in Europe in non-Romani populations (Gresham et al. 2001: 1324). One particular haplotype within the M haplogroup “is shared by nearly 30% of Gypsy men, an astonishing degree of preservation of a highly differentiated lineage” (Kalaydjieva et al. 2005: 1086), suggesting a limited number of ancestors as this number refers to all Romani men studied. Haplogroup H is highly frequent in India (though not in Pakistan), but rare elsewhere in the world, except among the Roms, where it is found among 35.6 percent of people studied, from 13 (not among southern Kalajci) of 14 subgroups (Gresham et al. 2001: 1321–3).

Similar figures are available for the male lineage. Haplogroup VI-68 was found in all 14 sampled Romani populations in Gresham et al. (2001: 1318–19), and in low frequencies in India, suggesting a profound bottleneck effect. A bottleneck effect is identified when only a limited sample of a larger population can be found at a certain location, suggesting that a smaller subgroup is ancestral to a population, for instance, much Indian genetic variation and limited Roma variation. Haplogroup VI-68 has been found in only one non-Romani person in Europe. Haplogroup VI-56 has likewise been found in one non-Romani individual in Europe (in Sardinia), but it was found in 12.7 percent of all Romani males. It has also been found in Central Asia, Pakistan, and the Middle East (Gresham et al. 2001: 1320). Two subvariants of haplogroup H common among Roms have not been detected in the Near East or Europe (Gresham et al. 2001: 1325).

In addition, there is a genetically transmitted disease (see 5.3) that is prevalent among Gypsies. One such disease (CMS) is also found in South Asia more than elsewhere, suggesting an “old age and wide geographical dispersal in South Asia” (Morar et al. 2004: 602). According to the authors this would constitute the “strongest evidence to date for the Indian origin” of the Roms (Morar et al. 2004: 604).

Both the M and the H haplogroups are almost absent among non-Romani European populations, proving a genetic inheritance from South Asia in both male and female lines. The existence of genetically transmitted diseases common to Indians and European Roms provides supportive evidence.

A few studies found that Roms differ genetically from specific groups in India. Nagy et al. (2006), for example, compared Slovak Roms with two endog-

amous *jatis* (castes based on professions, also found among many groups of Roms) in India, and concluded that Roms differ genetically significantly from both of them. Clearly, cultural similarities in subsistence and marriage patterns are not sufficient to conclude that there is a common origin. Other studies, however, do link Roms with certain other Indian groups. In the broad sample used by Wells et al. (2001: 10247–8), comprising 49 Eurasian population groups, the Sinte Romanies appeared closest to the Hunza of northern Pakistan and Bartangi of Pamir (Tajikistan). No other Romani groups were included in their sample.

The genetic evidence thus points unambiguously to a south Asian origin of the Romani population.

5.2. *Travel through Iran?*

There is linguistic evidence of influence of Persian on Romani (see Matras 2002 and sources quoted therein). Genetic research on the populations of Iran is limited. Some of the scarce data may indicate limited contact between the ancestors of the Romani and Iranian populations. Haplogroup H-M82 is typical for Roma, but it is also found among 2 percent of the population of Iran, and only in the south (Regueiro et al. 2006). This may be an indication of limited male gene flow in Iran, and as this is only found in south Iran, it may be indicative of a southern migration route through Iran. More research is needed here.

5.3. *Rare diseases among Roma*

There are a number of genetically transmitted diseases that are more typical for Roms than for non-Roms. Kalaydjieva et al. (2005: 1091) list 16 of them, all of them autosomal, that is, they are transmitted such that both sons and daughters can inherit the disease. Some of these diseases are encountered (virtually) only among Roms, not among other Europeans, and others have a higher incidence among Roma, and some are only found in certain Romani subgroups, not in others. The diseases came about by a genetic change in one person generations ago, and the descendants of this individual, but no others, are carriers. Although these carriers need not develop the disease and show the symptoms, their descendants may, but no individuals who do not descend from the first carrier.

I will give a few examples from the cases that are best studied. I will not specify the nature and symptoms of these diseases.

Galactokinase (P28T mutation) is a disease that is rare worldwide, but it has a high incidence among the Roms (Hunter et al. 2002). It was first discovered among Bulgarian Roms in 1999, later in “different European countries” (Hunter et al. 2002: 602).

Muscular Dystrophy_{SG} (Piccolo et al. 1996) is a genetic muscle disease found in 1996 in seven to ten Gypsy families in France, Spain, and Italy.

Infantile GM1-gangliosidosis is a disease caused by mutation R59H, which causes infant mortality (Sinigerska et al. 2006). At a global level, the disease is rare (1: 100,000–300,000), but in some regions it is much more frequent, for example, in Malta (1: 3,700) and southern Brazil (1: 17,000). The disease has been encountered and studied in eleven Bulgarian Gypsy families, one Italian (Balkan Gypsy) case, and eight patients in Brazil of unknown ethnicity. Among the Rudaris in Bulgaria (a Gypsy-like group who are speakers of Romanian, not of Romani), the carrier rate is extremely high: one in ten persons carries the mutation, but fortunately not all carriers develop symptoms of the disease. From groups other than Rudari, 413 Roms were tested too, and not a single carrier was found among them, which means that this is a Rudari-specific disease. With less confidence can we conclude that the Brazilian patients must be descendants of one or more Rudari founders who were carriers. The high incidence in the population of Malta is unclear.

In addition there are five rare single-gene diseases that are relatively common among Roma: CMS, which is found in several Romani groups and in South Asia; HMSNL, which is found among Roms individuals across Europe, and “more common among the Vlax” (Morar et al. 2004: 601); CCFDN, which is found only among Vlax Roma; LGMDC2, found among Western European Roma, and some Balkan Roma, but not elsewhere; and finally GD, found across Europe, most commonly among the Vlax Roms (Morar et al. 2004).

There are two types of pattern: some diseases are found in all main Gypsy groupings, others are found only in certain subgroups (e.g. Vlax). The fact that some of these diseases are found only in Romani groups that correspond to particular dialect groupings within Romani may mean that the disease emerged (i.e. after the first person became a carrier of the disease) after the speakers of these dialect groupings had split off from the rest of Roms, and since avoided, or had no opportunity for, marital contacts outside their group. Some genetic patterns are “observed in all Gypsy chromosomes, supporting common origins” (Morar et al. 2004: 602). The authors also found differences between Romani branches, for instance the “differences between the Balkan and Vlax [. . .] was highly significant” (Morar et al. 2004: 601), and in addition there was a “striking lack of haplotype sharing between Gypsy groups in the Vlax category (Rudari, Kalderash, Lom [meaning, here and elsewhere, Roms from the town of Lom in Bulgaria, not the Armenian Gypsy group of the same name], Kalajdzhi)”. This leads Morar et al. (2004) to a number of conclusions. First, there is an obvious founder effect in Western European Gypsies (p. 603) because of “high carrier rates” (p. 604). A founder effect is caused by the fact that there is a small number of ancestors of a specific group, leading to limited genetic diversity. Second, branches covering several groups can be established: “The patterns observed were distinctly different in the four Gypsy groups

[Kalderash, Lom, Rudari, Turgovzi], reflecting an independent history of random recombinations” (p. 603). The distinctness of subgroups suggest “profound secondary and tertiary bottleneck events” (p. 604), that is, again clans and groups that descend from a small group of ancestors. Third, even though the “time frame [. . .] is unclear”, some tentative dating can be suggested: the “founding of the Proto-Gypsy population” can be related to a “single recent founding event ca. 800–900 years ago” (p. 604) (upper limit), and there was at that time a “small group of related individuals” (p. 605). Gresham et al. (2001: 1320) reach a similar date for one of the founding lineages: 992 years ago. Morar et al. (2004), too, discuss the number of founders of Vlax subgroups: ca. 100 individuals. The formation of subbranches of Romani groupings can be dated 425–600 years back. Fourth, there has always been little intermarriage with local populations or with other branches of Romani populations. Because of the limited gene flow, we can conclude that the “history of endogamy is as old as Gypsy groups” (p. 607).

In brief, the studies on the spread of genetic diseases corroborate what was known from cultural anthropological and linguistic studies: “exodus, diaspora, and subsequent fragmentation into small, geographically dispersed and isolated communities” (p. 607) have caused the current diversity among Roms, and their clear genetic distinctness from surrounding populations, and in many respects also from other Romani groupings.

5.4. *Roms and neighboring populations*

There have been a number of studies on genetic differences between Romani groups and surrounding populations. The results are quite similar, so I will take only three studies as examples here (see also Bakker 2009).

Peričić et al. (2005) compared Macedonians and Macedonian Roms, and found that the two groups were genetically different. Macedonians fitted in with the southeastern European populations, but the Roms did not. The Macedonians showed more diversity than the Roms. Another study concluded, incidentally, that Macedonian Roms stood out genetically from other European Romani populations (Nagy et al. 2006).

Deligiannides et al. (1996) studied genetic differences between Greeks and Greek Roms. They collected 96 samples from unrelated healthy Roms in settlements throughout Greece, and compared them with the Greek population. In 11 out of the 13 items studied, the comparison between the Roms and the Greeks revealed statistically significant differences. Apparently a rather limited gene flow had taken place between the two population groups. They also compared the Roms of Greece with a Rom population from Hungary, and this revealed statistically significant differences in five out of the eight items studied. This appears to be what most studies find: Roms differ genetically from the

surrounding population, but also from other Romani groups elsewhere.

Kalaydjieva et al. (2005) analyze such cases in terms of rules of intermarriage with other Romani groups. Marital connections with people from other groupings exist, but they can lead to banishment from the group and are generally rare – though there are significant intragroup differences in this respect. Marital connections between groups can be hierarchical (only one group marries into the other group, not the other way around) and/or asymmetric (e.g. only men find wives elsewhere, women do not find men elsewhere). The authors' research also confirms the ethnographical observations that Roms tend to marry only within their group, that there are often strong barriers against marrying men or women from other Romani groups, even those living in the same town, and that liaisons between Roms and non-Roms are probably more common than liaisons with partners from other Romani groups (at least in the Bulgarian context). Here, the genetic and anthropological data point in the same direction. Better knowledge of these patterns may shed light on the reasons for the spread of linguistic features.

5.5. *Diversification*

Linguists have distinguished a number of dialect branches of Romani: Vlax (subdivided into Northern and Southern, called New and Old Vlax in some linguistic and genetic literature), Balkan (South Balkan I, II), Central (North and South), Northern, Northeastern, and a number of minor groups whose classification is not agreed upon (Wales, Iberian peninsula, Havati/Doljenski of Slovenia and Italy). See Matras (2002, 2005) and, for a slightly different classification, Boretzky and Iglá (2004).

A number of features have spread from one branch to another, which has led some linguists to cast doubt on the branching metaphor, stressing the areal patterns. Perhaps genetics can shed light on these questions. In areas where linguists have detected the spread of linguistic features from one dialect group to another, or one dialect branch to another, can geneticists establish gene flow between these groups? In other words, did these linguistic features spread through communication or through intermarriage? And, if there was intermarriage, were there specific patterns differing for males and females? As yet, there are no clear answers, but some suggestions have been made.

Kalaydjieva et al. (2001) compared three groups they subsume under the label "Vlax". However, they use this term for all Gypsy groups with historical Romanian connections, whereas Romani specialists use the term only for Romani dialects with historical Romanian influence. Thus, the Rudaris are speakers of Romanian, not Romani. The authors studied genetic features of the males and females among Bulgarian Lom (Southern Vlax), Kalderash (Northern Vlax), and Rudari (Romanian speaking Gypsies). They concluded

that the three groups had been part of the same ancestral population for most of their history, and the authors calculated the first split at around 350 or 360 years ago, whereby the Lom split off, and a second split between 100 and 150 years ago. Genetically, the Rudari are halfway between the Lom and Kalderash (Chaix et al. 2004: 287).

Based on the genetic diversity, geneticists have also tried to calculate the size of founding populations. The number of female founders were between 300 and 1,000 for both Lom and Rudari, and between 2,500 and 8,000 for the Kalderash, whereas the number of males is estimated as under 100 for all three groups. Chaix et al. (2004: 290) point to the 1893 census of Transsylvania, where the group size for Romani clans is about 2,000. The geneticists' estimates thus seem to correlate with historical data.

6. Dating migrations and splits

Linguists and geneticists have a number of techniques to date language changes and genetic mutations, respectively. On the basis of rates of change in the lexicon (see Fraser 1992a) and by comparing sound changes and grammatical changes in Romani with parallel changes in neighboring languages, and the absence and presence of loanwords from coterritorial languages, linguists have made estimates of the dates of emigration from South Asia and the route of migration in present-day Romani. Fraser calculated the breakup of Romani, based on the choice of three divergent dialects, viz. Welsh Romani (Northern or isolate), Paspatian Romani of Turkey (Balkan) and Kalderash Romani (Vlax branch), by using glottochronology. Despite the shortcomings of that technique – which the author is aware of – the dates coincide with the approximate dates based on what is known on migration within Europe: the breakup would have taken place between 1040 CE and 1200 CE. The split of Romani from Central Indic was calculated as dating from 390 BCE or earlier. At the time of publication of Fraser's study, most specialists assumed a movement out of India, and hence a split between languages of India and Romani, at around 1000 CE. In more recent research, however, older splits have been suggested see, for example, Metzger (2004) and Tcherenkov and Laederich (2004).

When comparing population groups, geneticists can calculate a most recent common ancestor, the one person from whom all descend partially. The most recent common ancestor of all Roms would have lived 1500 years ago, in other words, around 500 CE, but later according to others, between 960 and 1170 years ago (Kalaydjieva et al. 2005: 1086).

Geneticists also make estimates of dates, on the basis of assumptions of a certain rhythm in mutations. Geneticists have estimated dates for mutations for some of the single-gene diseases among Roms. In their comparative study

of the transmission of genetically transmitted diseases, Morar et al. (2004) and Gresham et al. (2001) both point to a single founding of the Proto-Gypsy population “by a small group of related individuals” which would go back to a “single recent founding event”, 800 to 900 years ago (Morar et al. 2004: 604–5) (that would be the upper limit). The subgroups had been founded between 425 and 600 years ago, and Vlax 500 years ago. On the basis of these studies of gene flow, the authors conclude that endogamy is not a recent phenomenon, but that the “history of endogamy is as old as Gypsy groups” (607). For instance, Galactokinase (P28T mutation) was estimated to go back 750 years in history, perhaps even 1000 years (Hunter et al. 2002: 605). Dystropy is estimated to be at least 1200 years old:

If a generation is 20 years, this would indicate that the C283Y mutation in the [gamma]-SG gene is at least 1200 years old. If the genetic clock had been reset by a bottleneck around the time of wandering (fixation of one haplotype [a set of connected genetic changes] in a small population), the age of the mutation would be even older. We therefore assume that the C283Y mutation predates the commonly accepted date of migration of Gypsies out of Northern India. (Piccolo et al. 1996: 2020)

If a disease was brought from India, one would expect the existence of the disease not only in all major Romani groupings, but also in India. But no Indian cases have been reported.

Morar et al. (2004) try to relate these diseases with particular Romani groupings as well, providing at the same time an estimate of a split. More concretely, HMSNL is found across Europe, or “more common among the Vlax” (Morar et al. 2004: 604) (estimated age 850 years, or between 700 and 1075 years), and CCFDN is found only among Vlax Roms (estimated age: 500 years: 400–650). LGMDC2 is found among Western European Roma, some in the Balkan, but not elsewhere, and not among the Vlax (estimated age: 600 years: 525–775).

Dating the emergence of diseases seems to follow splits of the major Romani groupings that have been established on the basis of linguistic and historical evidence. Diseases limited to specific groups (e.g. Rudari, Vlax) have shallower estimated time depths than those found in all groups. Among the latter, an origin in India is regularly suggested.

7. Conclusions

The genetic studies summarized here make a number of things clear. Roms are genetically closer to Indians than to European populations. Roms are heterogeneous between countries, and even between dialectal group affiliations in the same region. The internal diversity of the Roms shows that they are genetically far more heterogeneous than autochthonous European populations (the Macedonian case mentioned above (Peričić et al. 2005) seems to be an

exception). This is due to the founder effect: each dialect group descends from a limited group of ancestors, thus strengthening genetic homogeneity within the group. Social and economic pressures within Europe led to gradual fragmentation of Romani groups, generating multiple genetically differentiated subisolates (Kalaydjieva et al. 2005: 1084).

Chaix et al. (2004: 285) mentioned research results of a genetic study of Vlax Roms indicating that “genetic affinities reflect [-] to a large extent the linguistic classification in Balkan, Vlax and Western European.” Speakers of Northern and Central dialects presumably left south-eastern Europe for the West, whereas the Vlax remained isolated in Romania between the 1500s and the late 1700s (Southern Vlax) and the mid-1800s (Northern Vlax), and “each of these large migrational/linguistic categories comprises numerous socially and culturally defined endogamous groups” (Chaix et al. 2004: 285). Vlax Roms display “limited [genetic] diversity, with a single ancestral Y-chromosome shared by 73% of the males.” The historical and genetic dates thus correlate to a significant extent.

Adding to the list of private mutations and aiming to understand their molecular epidemiology, further medical genetic research has “revealed a peculiar combination of genetic homogeneity and mutation sharing by affected subjects across Europe [. . .] and, at the same time, an internal mosaic of striking differences in the prevalence of genetic disorders and mutations between neighbouring Gypsy communities in the same country” (Kalaydjieva et al. 2005: 1084).

Gresham et al. (2001: 1328) conclude that country of residence, or geography, has no relevance to the genetic structure of Romani groups. Neither have group labels referring to trade in the Balkans (cf. Marushiakova and Popov 2008). However, migration patterns and hence the major dialect groupings are “by far the most significant” factor responsible for differences between Rom groups (Gresham et al.: 1328).

An unambiguous proof of the Indian ancestry of the Gypsies comes from three genetic marker systems: Y chromosome haplogroup H-M82, mtDNA haplogroup M(25), and the pathogenic 1267delG mutation in CHRNE(12) (causing autosomal recessive congenital myasthenia), found on the same ancestral chromosomal background in Gypsy, Indian, and Pakistani subjects (Kalaydjieva et al. 2005: 1085).

8. Perspectives

The genetic studies surveyed here have basically confirmed what linguists and anthropologists already had found out: the Indian origin of the Roms (Gresham et al. 2001: 1328 mention the “strong evidence of Asian origins”),

their migration route to Europe (to a limited extent), their dispersion into groups after their arrival in Europe, and a history of endogamy within extended clans. This provides rather devastating evidence against the claims of historians Leo Lucassen (1996) and Wim Willems (1997) and social anthropologist Judith Okely (1983) that the Roms are Europeans and emerged through social stigmatization.

Unfortunately, geneticists show some linguistic naivety (especially in earlier publications) concerning their sampling among the Roms. There are a number of ambiguous wordings referring to Gypsy groups in some of their writings that are crucial distinctions for linguistics. First, names of locations that geneticists use are often ambiguous. This is especially the case when they mention states (Spain, France, Bulgaria, Hungary, etc.). Linguists, anthropologists, and historians who work with Roms know that there is a huge dialectal diversity among Roms in each country, distinguishing groups that are also significant in social patterns among Roms, for instance in marriage patterns. And if it is not specified which dialect they speak, country information is much too vague to be useful for linguists. The same is true for names of cities, such as Lom in Bulgaria, where at least Balkan, Northern Vlax, and Southern Vlax dialects are represented. Were all three groups part of their sample or just one? The geneticists do not specify whether their samples included people from all groupings, or one, and in both cases the results are of little value for linguistics, if not specified. Similarly, in Vitoria (Basque Country, Spain), the unspecified Gypsies consulted could be Basque Gypsies, Gitanos (Spanish Gypsies), or Vlax Gypsies (Kalderash), all of whom are present in the area (Cárdaba et al. 2001).

A second type of problem in compatibility is that linguists and geneticists sometimes use different classifications – “Old Vlax” versus “New Vlax”, for example – where linguists tend to use Southern Vlax and Northern Vlax. Old Vlax can be equated with Southern Vlax, and New Vlax with Northern Vlax, but still the terminology can be confusing when Southern Vlax have migrated to towns where other groups were already settled.

Third, Romanian-speaking Rudari are usually listed as Vlax by geneticists, even though they do not belong linguistically in the Vlax group of Romani as they are not Romani speakers. They may be Gypsies, but not speakers of Vlax Romani.

A fourth problem is that geneticists sometimes use group names not used in linguistic literature, for example, Lom, Musicians, Turgovzi, Feredjelli (however, see Marushiakova and Popov 1997 for some of them; Elšík and Matras deal linguistically with the Sliven Muzikanta dialect of Bulgaria, but they are not in the index; they are discussed here: 93–4, 100, 110, 113, 133, 135, 151, 158, 160–1, 165, 168, 172, 180, 183, 191, 202, 207, 209–10, 214, 237, 244, 254–5, 258,

263–4, 266, 269, 274, 276–7, 289, 292, 298–9, 304, 307, 309, 415–16, 419, 421, 435, 437). These names do not provide enough information about their linguistic affiliation. Geneticists' linguistic groupings are often unclear to linguists, and sometimes wrong from a linguistic point of view. Sometimes a name of a subgroup is associated with a country where linguists were unaware of their existence, or for whom no linguistic data are available, for example Sinte (Northern/Northeastern) in Central Asia, more specifically Uzbekistan (Wells et al. 2001).

A more serious problem is the risk of circularity. Linguists have proposed group classifications that are based on linguistic differences but which also reflect ethnic or sub-ethnic groupings. They have also proposed tentative dates of splits, in some cases in conjunction with early documents and in line with documented historical events. They have suggested contacts between some neighboring groups that have led to the blurring of some linguistic group classifications. When one reads that the genetic studies in many areas corroborate what is known from linguistics, anthropology, and history, this can be seen as a success of earlier studies, but it also raises suspicion: could it be that the geneticists took the linguistic data as a point of departure, and sought to confirm these? Most of the geneticists do refer to linguistic literature and linguistic groupings. At any rate, in some cases one suspects that their data are not completely independent of non-genetic data. This assumption may be confirmed when other genetic techniques are employed using only genetic data, without reference to pre-existing groupings based on dialectal differences. In some cases computers automatically produce structures, for instance so-called neighbor-joining trees (introduced recently in linguistics by McMahon and McMahon 2005 and others). Here, sometimes quite different groupings of Romani groups were produced. The result appears quite incompatible with what is known from linguistics about the differentiation of Romani speakers into groups.

An example is shown in Figure 1. Here the British Roms and Welsh Roms, believed to have split off two centuries ago on historical grounds (Sampson 1968), appear quite different from one another genetically. Similarly, there is no linguistic or historical reason to assume that Hungarian Roms are any closer to Indian groups such as Punjabis and Rajput than other Romani groups. "Swedish Roma" are probably Kalderash, a Vlax group, and should be expected to be close to "Wallachian Roma" (Vlax) of Slovakia, but they are not. Also, the location of "Slovenian Roma" between the Welsh Roms and Swedish Roms is quite unexpected. A similarly unexpected tree can be found in Morar et al. (2004: 602). In short, in such cases the linguistic/historical and genetic data seem incompatible, even beyond the sampling problems discussed above.

This does not mean that linking genetic and linguistic research is hopeless. If researchers in both fields become more aware of the knowledge, methods, possibilities, problems, and limits of both fields, fruitful research would lie ahead.

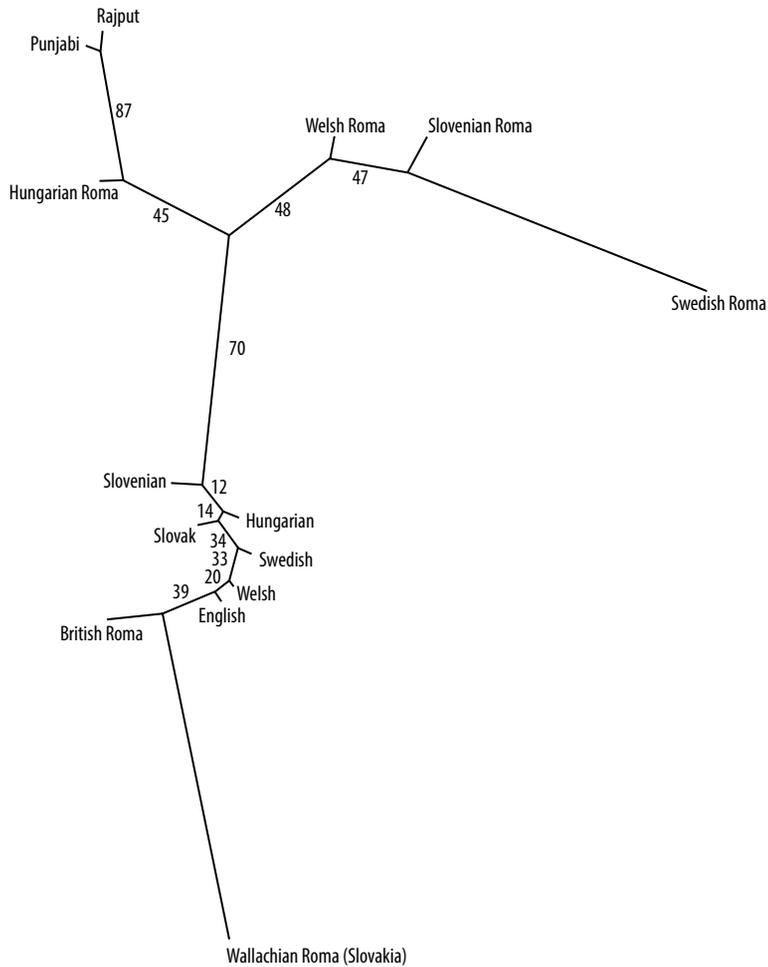


FIGURE 1. Neighbour-adjointing tree of selected population groups (from Kalaydjieva et al. 2001)

References

- Bakker, Peter. 2009. Genetic roots of the Roms. *In: Kyuchukov, Hristo, ed. New Studies in Romology. Sofia: Wini 1837. 17–45.*
- Boretzky, Norbert. 1992. Zum Erbwortschatz des Romani. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikation* 45: 227–51.
- Birgit Iгла. 2004. *Kommentierter Dialektatlas des Romani. Wiesbaden: Harrassowitz.*
- Campbell, Lyle. 1998. *Historical linguistics: An introduction. Edinburgh: Edinburgh University Press.*
- Cárdaba, Blanca, Miriam F. Moffatt, Eduardo Fernández, Aurora Jurado, Marta Rojo,

- Milagros García, Ignacio J. Ansoategui, Isabel Cortegano, Ignacio Arrieta, Miguel A. Etxenagusia, Victoria del Pozo, Javier Urraca, Esther Aceituno, Soledad Gallardo, Pilar Palomino, William Cookson, and Carlos Lahoz. 2001. Allergy to *Dermatophagoides* in a group of Spanish Gypsies: Genetic restrictions. *International Archives of Allergy and Immunology* 125(4): 297–306.
- Chaix, Raphaëlle, Frederic Austerlitz, Bharti Morar, Luba Kalaydjieva, and Evelyne Heyer. 2004. Vlach Roma history: What do coalescent-based methods tell us? *European Journal of Human Genetics* 12: 285–92.
- Deligiannidis, Panagiotis, Costas Triantaphyllidis, Dimitris Psaroulis, and Anastasia Kouvatsi. 2006. Forensic evaluation of 13 STR loci in the Roma population (Gypsies) of Greece. *Forensic Science International* 157: 198–200.
- Elišák, Viktor. 2006. Review of Boretzky and Igla 2004. *Romani Studies. Fifth Series*. 16: 105–11.
- Fraser, Angus. 1992a. Looking into the seeds of time (Romani glotto-chronology). *Tsiganologische Studien* 1: 135–66.
- 1992b. *The Gypsies*. Oxford: Blackwell.
- Gresham, David, Bharti Morar, Peter A. Underhill, Giuseppe Passarino, Alice A. Lin, Cheryl Wise, Dora Angelicheva, Francesc Calafell, Peter J. Oefner, Peidong Shen, Ivailo Tournev, Rosario de Pablo, Vaidutis Kučinskis, Anna Perez-Lezaun, Elena Marushiakova, Vesselin Popov, and Luba Kalaydjieva. 2001. Origins and divergence of the Roma (Gypsies). *American Journal of Human Genetics* 69: 1314–31.
- Hajioff, Steve, and Martin McKee. 2000. The health of the Roma people: A review of the published literature. *Journal of Epidemiology and Community Health* 54: 864–9.
- Hancock, Ian F. 2000. The emergence of a koine outside of India. *In: Thomas Acton, ed. Scholarship and the Gypsy struggle: Commitment in Romani Studies. A collection of papers and poems to celebrate Donald Kenrick's seventieth year*. Hatfield: University of Hertfordshire Press. 1–13.
- Hunter, Michael, Evelyne Heyer, Frederic Austerlitz, Dora Angelicheva, Vanya Nedkova, Paz Briones, Anna Gata, Rosario de Pablo, Aranka László, Nils Bosshard, Richard Gitzelmann, Attila Tordai, Lajos Kalmar, Csaba Szalai, Istvan Balogh, Constantin Lupu, Axinia Corches, Gabriela Popa, Anna Perez-Lezaun, and Luba Kalaydjieva. 2002. The P28T mutation in the GALK1 gene accounts for galactokinase deficiency in Roma (Gypsy) patients across Europe. *Pediatric Research* 51: 602–6.
- Jakoubek, Marek, and Lenka Budilová. 2006. Kinship, social organisation and genealogical manipulations in Gypsy osadas in eastern Slovakia. *Romani Studies. Fifth Series*. 16: 63–82.
- Kalaydjieva, Luba, David Gresham, and Francesc Calafell. 2001. Genetic studies of the Roma (Gypsies): A review. *BMC Medical Genetics* 2: 5.
- Bharti Morar, Raphaëlle Chaix and Hua Tang. 2005. A newly discovered founder population: The Roma/Gypsies. *Bio-Essays* 27: 1084–94.
- Lucassen, Leo. 1996. *Zigeuner. Die Geschichte eines polizeilichen Ordnungsbegriffes in Deutschland 1700–1945*. Köln-Wien: Böhlau-Verlag.
- McMahon, April, and Rob McMahon. 2005. *Language classification by numbers*. Oxford: Oxford University Press.
- Marushiakova, Elena, and Vesselin Popov. 1997. *Gypsies (Roma) in Bulgaria. (= Studien zur Tsiganologie und Folkloristik, 18.)* Frankfurt am Main: Peter Lang.

- Marushiakova, Elena, and Vesselin Popov. 2008. Ethnonyms and professionyms. Gypsy groups in the Balkans. *Etnonimiy i profesjonimiy. Grupy Cyganow na Balkanach. Studia Romologica* 1: 177–98.
- Matras, Yaron. 1999a. Johann Rüdiger and the study of Romani in 18th-century Germany. *Journal of the Gypsy Lore Society. Fifth Series.* 9: 89–116.
- Matras, Yaron. 1999b. S/h alternation in Romani: An historical and functional interpretation. *Grazer Linguistische Studien* 51: 99–129.
- 2002. *Romani: A linguistic introduction.* Cambridge: Cambridge University Press.
- 2005. The classification of Romani dialects: A geographic-historical perspective. *In: Dieter Halwachs, Barbara Schrammel and Gerd Ambrosch, eds. General and applied Romani linguistics.* Munich: Lincom Europa. 7–26.
- Metzger, Matthias. 2004. Romani and Central New Indo-Aryan. Paper presented at the International Colloquial on Romani linguistics, Graz, November 2004.
- Miklosich, Franz. 1872–1880. *Über die Mundarten und Wanderungen der Zigeuner Europas X–XII.* Vienna: Karl Gerold's Sohn.
- Morar, Bharti, David Gresham, Dora Angelicheva, Ivailo Tournev, Rebecca Gooding, Velina Guerguelcheva, Carolin Schmidt, Angela Abicht, Hanns Lochmuller, Attila Tordai, Lajos Kalmar, Melinda Nagy, Veronika Karcagi, Marc Jeanpierre, Agnes Herczegfalvi, David Beeson, Viswanathan Venkataraman, Kim Warwick Carter, Jeff Reeve, Rosario de Pablo, Vaidutis Kučinskas, and Luba Kalaydjieva. 2004. Mutation history of the Roma/Gypsies. *American Journal of Human Genetics* 75: 596–609.
- Nagy, Melinda, Lotte Henke, Jürgen Henke, Prasanta K. Chatthopadhyay, Antónia Völgyi, Andrea Zalán, Orsolya Peterman, Jarmila Bernasovska, and Horolma Pamjav. 2006. Searching for the origin of Romanies: Slovakian Romani, Jats of Haryana and Jat Sikhs Y-STR data in comparison with different Romani populations. *Forensic Science International* 169: 19–26.
- Okely, Judith. 1983. *The Traveller-Gypsies.* Cambridge: Cambridge University Press.
- Peričić, Marijana, Irena Martinović Klarić, Lovorka Barać Lauc, Branka Janičijević, Dobrovoje Đorđević, Ljudmila Efremovska, and Pavao Rudan. 2005. Population genetics of 8 Y chromosome STR loci in Macedonians and Macedonian Romani (Gypsy). *Forensic Science International* 154: 257–61.
- Piccolo, F., Marc Jeanpierre, F. Leturcq, C. Dodé, K. Azibi, A. Toutain, L. Merlini, L. Jarre, C. Navarro, R. Krishnamoorthy, F. M. S. Tomé, J. A. Urtizbera, J. S. Beckmann, K. P. Campbell, and J.-C. Kaplan. 1996. A founder mutation in the gamma-sarcoglycan gene of Gypsies possibly predating their migration out of India. *Human Molecular Genetics* 5: 2019–2022.
- Pischel, Richard. 1883. *Die Heimath der Zigeuner.* *Deutsche Rundschau* 36: 353–75.
- Regueiro, M. A. M. Cadenas, T. Gayden, P. A. Underhill, R. J. Herrera. 2006. Iran: Tricontinental nexus for Y-chromosome driven migration. *Human Heredity* 61: 132–43.
- Sampson, John. 1968. *The dialect of the Gypsies of Wales (1st publ 1926).* Oxford: Clarendon Press.
- Sinigerska Ivanka, David Chandler, Vijesh Vaghjiani, Irfet Hassanova, Rebecca Gooding, Amelia Morrone, Ivo Kremensky, and Luba Kalaydjieva. 2006. Founder mutation causing infantile GM1-gangliosidosis in the Gypsy population. *Molecular Genetics and Metabolism* 88: 93–5.

- Tcherenkov, Lev, and Laederich, Stéphane. 2004. *The Rroma. Otherwise known as Gypsies, Gitanos, Tsiganes, Tigani, Cingene, Zigeuner, Bohemiens, Travellers, Fahrende, etc.* Vol. 1: History, Language, and Groups, Vol. 2: Traditions and Texts. Basel: Schwabe Verlag.
- Thangaraj, Kumarasamy, Gyaneshwer Chaubey, Vijay Kumar Singh, Ayyasamy Vanniarajan, Ismail Thanseem, Alla G. Reddy, and Lalji Singh. 2006. In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup 'M' in India. *BMC Genomics* 7: 151.
- Wells, Spencer. 2002. *The journey of man: A genetic Odyssey.* Princeton: Princeton University Press.
- Nadira Yuldasheva, Ruslan Ruzibakiev, Peter A. Underhill, Irina Evseeva, Jason Blue-Smith, Li Jin, Bing Su, Ramasamy Pitchappan, Sadagopal Shanmugalakshmi, Karupiah Balakrishnan, Mark Read, Nathaniel M. Pearson, Tatiana Zerjal, Matthew T. Webster, Irakli Zholoshvili, Elena Jamarjashvili, Spartak Gambarov, Behrouz Nikbin, Ashur Dostiev, Ogonazar Aknazarov, Pierre Zalloua, Igor Tsoy, Mikhail Kitaev, Mirsaid Mirrakhimov, Ashir Chariev, and Walter F. Bodmer. 2001. The Eurasian heartland: A continental perspective on Y-chromosome diversity. *Proceedings of the National Academy of Sciences of the United States of America* 98(18): 10244–9.
- Willems, Wim. 1997. *In search of the true Gypsy: From enlightenment to final solution.* London: Frank Cass Publishers.
- Zeman, Catherine L., Diane E. Depken, and David S. Senchina. 2003. Roma health issues: A review of the literature and discussion. *Ethnicity and Health* 8: 223–49.

