A newly discovered founder population: the Roma/Gypsies

Luba Kalaydjieva, Bharti Morar, Raphaelle Chaix, and Hua Tang

Summary

The Gypsies (a misnomer, derived from an early legend about Egyptian origins) defy the conventional definition of a population: they have no nation-state, speak different languages, belong to many religions and comprise a mosaic of socially and culturally divergent groups separated by strict rules of endogamy. Referred to as “the invisible minority”, the Gypsies have for centuries been ignored by Western medicine, and their genetic heritage has only recently attracted attention. Common origins from a small group of ancestors characterise the 8–10 million European Gypsies as an unusual trans-national founder population, whose exodus from India played the role of a profound demographic bottleneck. Social and economic pressures within Europe led to gradual fragmentation, generating multiple genetically differentiated subisolates. The string of population bottlenecks and founder effects have shaped a unique genetic profile, whose potential for genetic research can be met only by study designs that acknowledge cultural tradition and self-identity. BioEssays 27:1084–1094, 2005. © 2005 Wiley Periodicals, Inc.

Introduction

Isolated founder populations have been a precious resource for Mendelian genetics, contributing knowledge on novel disorders, genes and mutations, as well as new experimental and statistical approaches designed to exploit the special characteristics of such populations. As the term implies, founder populations derive from a small number of ancestors, with subsequent isolation (geographic, cultural or religious etc.) leading to limited immigration and demographic growth mainly from within. Reduced genetic diversity and founder effect, resulting in a more homogeneous basis of inherited disorders and predispositions, make it possible for genetic studies to treat the whole population as one large family, where individuals affected by a specific condition are likely to share the same ancestral disease-causing DNA variant(s).

Unlike other founder populations, whose history, genealogy and genetic epidemiology have been extensively documented, the geographically dispersed and socially marginalised Gypsies have been ignored by European medicine for hundreds of years, and their unique genetic heritage is only now becoming a focus of interest for geneticists and medical practitioners. Private (meaning confined to this population) disease-causing mutations and evidence of founder effect were published for the first time in 1996 in two independent studies—of the novel Hereditary Motor and Sensory Neuropathy type Lom (HMSNL) and of Limb-Girdle Muscular Dystrophy 2C. Further medical genetic research, adding to the list of private mutations and aiming to understand their molecular epidemiology, 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. This epidemiological pattern suggests a complex population structure, whose understanding is essential for further genetic research, as well as for patient care and public health interventions. The issue has not been addressed in earlier population genetic investigations (summarised in ref. 13), relying on the random sampling of “pure-blood Gypsies”, classified by country of residence, with disregard for social structure and self-identity. Recent population studies, initiated and designed to answer the questions raised by medical genetics, have made a big step forward and characterised the Gypsies as one of the most interesting founder populations and a valuable part of the European genetic landscape.

The gypsies through the eyes of social scientists

The potential of founder populations to contribute to understanding the genetic basis of human disease is largely determined by historical demography: size and diversity of the founding population, subsequent demographic growth, extent of genetic isolation, and internal genetic differentiation

1Western Australian Institute for Medical Research and Centre for Medical Research, The University of Western Australia, Perth, Australia.
2Équipe de Génétique des Populations, Unité d’Éco-Athropologie, Musée de l’Homme, Paris, France.
3Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.

Funding agencies: The Welcome Trust, the Australian Research Council and the National Health and Medical Research Council.

*Correspondence to: Luba Kalaydjieva, Laboratory of Molecular Genetics, WAIMR, B Block, QEII Medical Centre, Nedlands, WA6009, Perth, Australia. E-mail: luba@cyllene.uwa.edu.au

DOI 10.1002/bies.20287

Published online in Wiley InterScience (www.interscience.wiley.com).
The Gypsies have no written chronicles, church records and historical research of their own, and the puzzle of their history has been assembled from the accounts of outsiders and from inferences of linguistic and cultural anthropology studies. This is what the social sciences tell us.

The current Gypsy population of Europe, around 10 million people, is widely accepted by social scientists, with different opinions dating the exodus to 1000–1500 years ago, and the arrival in the Byzantine Empire to the 9th–11th century AD. Settling in the Balkans has taken place in the 13th–14th century, and the early diaspora into the rest of Europe completed by the end of the 15th century. The Balkans have traditionally hosted the largest proportion of European Gypsies, referred to as Balkan and Vlax Roma, depending on their history of migrations and related dialects of Romanes (classified according to the borrowings from other languages). The Balkan Roma, speakers of “Balkan” or stratum I dialects of Romanes, descend from the early Gypsy settlers in the lands south of the Danube River, within the limits of the Ottoman Empire. The ancestors of the Vlax Roma continued the journey north of the Danube, into the Wallachian Principalities (present-day Romania), where Gypsy slavery was instituted in the early 14th century. Slavery has not only played an important role in the social and biological history of the Vlax Roma, but is also related to important later migrations. During the Austrian–Turkish wars in the late 17th–early 18th century, temporary occupation of Wallachia allowed large numbers to escape into the ethnically tolerant Ottoman Empire; these are speakers of “old Vlax” (stratum II) dialects of Romanes. By the end of the 19th–beginning of the 20th century, abolition of slavery led to mass migrations of Vlax Roma, speaking “new Vlax” (stratum III) dialects, to all parts of Europe. The latest migration wave, making the Roma a visible presence in the big cities of Western Europe, was triggered by the social and economic changes in Eastern Europe of the 1990s and the wars in former Yugoslavia.

Adding to the complexity of historical migrations is the traditional organisation of Gypsy society, which has long fascinated cultural anthropologists and has been meticulously described in many studies. The primary unit of this social organisation is the Gypsy group, whose close resemblance to the professional jatis of India has been interpreted as further support for the Indian origins of the Gypsies. Identity is defined by the group ethnonym (usually reflecting a traditional trade), customs and traditions, organs of self-rule, language, history of migrations and religion. Persisting in the midst of the nation states of Europe, Gypsy groups are separated from each other by strict rules of endogamy, yet recognise no political boundaries and may spread across many countries: members of the same group in different countries are eligible marriage partners, while intermarriage between neighbouring communities in the same small town is often proscribed. Group structure and diversity are particularly well preserved in the Balkans. In the early 20th century, a British ethnologist serving as the British consul in the Black Sea port city of Varna, described 19 different Gypsy “tribes” in the North-East of Bulgaria alone.

Drawing a convoluted picture of historical migrations and a mosaic social structure, linguists and cultural anthropologists also favour an ethnically and linguistically diverse founding population. Hypotheses on the origins of the proto-Gypsies range from the lowest strata of the Indian caste system, to a mixed society of warriors and camp followers fighting off the early Islamic incursions in the north of India. Some scenarios assume a single founding event—the exodus from India, while others allow for a slow continuous trickle of small nomadic bands. A geneticist’s summary of these data would describe the Gypsies as a conglomerate of Asian populations of largely unknown history and current size, that may fit the observed pattern of Mendelian disorders but does not satisfy the criteria for a large founder population of serious potential for research into complex disorders.

The picture emerging from recent population genetic studies

Population genetic studies of the Gypsies in the last few years have been facilitated by new technologies, growing knowledge of variation in the human genome and rapidly accumulating data on global genetic diversity (please see box for a brief description of methodology). The design of these studies, based on the social tradition of the Gypsy people and made possible by a close interaction with cultural anthropologists, has aimed at understanding the common biological history of the Gypsy population of Europe, as well as the relationship between cultural and genetic identity of a diversity of Roma groups (Fig. 1).

A young founder population of common Indian descent

Unambiguous proof of the Indian ancestry of the Gypsies comes from three genetic marker systems: Y chromosome haplogroup H-M82, mtDNA haplogroup M (Fig. 2), and the pathogenic 1267delG mutation in CHRNA12 (causing autosomal recessive congenital myasthenia), found on the same ancestral chromosomal background in Gypsy, Indian and Pakistani subjects. While confirming the centuries-old

Challenges

BioEssays 27.10 1085
Challenges

**Box: Brief description of methodology**

Genetic variation in the Gypsy population has been studied in nearly 2000 subjects, representing different parts of Europe and divergent Gypsy groups. The analyses included polymorphisms on the paternally inherited Y chromosome, the maternally transmitted mitochondrial (mt) DNA and biparentally inherited autosomal markers (neutral variants and disease-causing mutations). Slowly evolving or unique event polymorphisms, which define Y chromosome and mtDNA haplogroups and reflect ancient human history, were used in comparison to global diversity data to trace the continental origins of Gypsy lineages. Markers of higher mutability, such as Y chromosomal and autosomal microsatellites (with haplotypes defined as the combination of marker alleles carried on the same chromosome), the Y chromosomal minisatellite MSY1, and sequence variation in the mitochondrial control region were used to assess current diversity and infer recent population history. The population genetics analysis tool Arlequin was used to examine genetic diversity (using parameters such as the observed number of haplotypes, number of polymorphic sites, and gene diversity), test demographic models based on mtDNA sequence variation, and assess genetic distances between populations using the number of pairwise differences as the molecular distance. The dating of historical events was based on coalescence time analysis of Y chromosome and autosomal haplotypes. Historical demography parameters were inferred from genetic data using the coalescence-based approach incorporated in the Batwing software.

Population structure was examined as described.

The Gypsy group was born in Europe

During the long journey from India to Europe, the Gypsies added new words to their language and newcomers to their society (Fig. 2). There is no documented history of their early social organisation; however, old European chronicles already mention small groups of 100–300 people with horses and dogs, headed by “barons, kings or princes” and provide some indirect evidence of tribal endogamy and hostilities between groups. The splits leading to group formation could have occurred at any time since the founding, while the description of group organisation as a “fluid mosaic” implies multiple divisions and mergers and changing rules of endogamy.

Genetic dating of the population fissions is again based on microsatellite variation in Y chromosomes and in the autosomal regions surrounding four private disease-causing mutations. All marker systems suggest that the earliest splits occurred 20–24 generations ago, i.e. from the late 13th century onwards, when Gypsy settlement in Europe has been documented beyond doubt. The genetic data point to a gradual process of consecutive separations, which has continued until as recently as 6–8 generations ago.

The size of the founding population of individual groups is estimated at several hundreds, with less than 100 for the effective male population size of some groups. These figures may be inflated, as they are based on the overall current diversity, including lineages that are likely to have been

linguistic theory of the Indian origins is no great triumph for modern genetic research, the major, unexpected and most-significant result of these studies is the strong evidence of the common descent of all Gypsies regardless of declared group identity, country of residence and rules of endogamy. A staggering 47.3% of men carry H-M82 Y chromosomes, mtDNA haplogroup M accounts for almost 30% of Gypsies subjects, and congenital myasthenia is one of the most common Mendelian disorders of this population with 4% average carrier rates of the 1267delG mutation.

The Y chromosome H-M82 and the mtDNA M haplogroups are characterised by limited internal haplotype and sequence diversity, which can easily be explained with the accumulation of mutations within the Gypsy population subsequent to its founding, rather than variation among the founders. Within the H-M82 haplogroup, an identical 8-microsatellite Y chromo-

some haplotype is shared by nearly 30% of Gypsy men, an astonishing degree of preservation of a highly differentiated lineage, previously described only in Jewish priests. Analysis of the highly mutable minisatellite MSY1 in these Y chromosomes provides further evidence of limited variation, with diversity values lower than in Finns, Basques and Cook Islanders. Similarly, nearly all mtDNA haplogroup M sequences fall into a single sub-haplogroup, M5. These closely related lineages, which form minute tight clusters within the overall diversity of Asia, signal genetic homogeneity among the founders—possibly a small group of related individuals separating from a single, ethnically defined population.

Based on the accumulated diversity, the most-recent common ancestor of Gypsy men carrying H-M82 Y chromosomes is estimated to have lived around 39 generation ago. Estimates in the same range, 32–36 generations ago, are obtained for the mean coalescence times of the microsatellite haplotypes surrounding the two most-common and widespread founder mutations, 1267delG (congenital myasthenia) and R148X (Hereditary Motor and Sensory Neuropathy Lom), used as an indicator of the time when these two mutations were introduced into the overall Gypsy population. Assuming a generation time of 30 years, these dates translate to 960–1170 years ago, supporting the exodus from India as the single recent founding event.
introduced by admixture subsequent to the founding. Such small sizes are compatible with the early historical descriptions of Gypsy groups, as well as with the data provided by the Ottoman tax registries, for example listing a total of ~5,700 Gypsies in the territory of present-day Bulgaria in the early 16th century. The demographic growth of Gypsy groups has been slow, with growth rate estimates based on Y chromosome variation in the range of 1.0017–1.0084, substantially lower than the 1.016–1.027 proposed for other European populations.

Given the common origins and lack of pre-existing differentiation of the early Gypsies, the formation of small strictly demarcated jati-like groups is surprising in itself. Even more surprising is the fact that the restitution of the ancient Indian tradition has taken place several centuries after the exodus, in the midst of European culture and social organisation. Social anthropologists attribute a major role in this process to the economy of “symbiotic” nomadism. The Gypsies have always made a living by providing services to the surrounding population. Such dependence on the macro-society would have favoured disintegration as a result of the economic pressure resulting from limited needs for specific services. The low capacity of small regions and communities to sustain blacksmiths, basket makers, and musicians etc. could lead to a continual budding off of branches initially maintaining contact with the old clan and eventually evolving into new, geographically dispersed endogamous groups. The hostility of the surrounding society has also exerted a profound influence on this process. (1) Slavery in Romania involved the separation of groups based on their owner and trade, (2) murderous persecution of the Gypsies in Western Europe in the early centuries after their arrival has made the small mobile group a means of survival, and (3) even the ethnically diverse and tolerant Ottoman Empire has encouraged the
separations by banning communication between Muslim and Christian Gypsies and imposing higher taxes on Christians.\textsuperscript{(21)}

The evolution of Gypsy groups into genetic subisolates

Interruption between groups is generally proscribed or subject to strict rules, which can be complicated, hierarchical and asymmetrical for the two sexes. Failure to comply leads to the expulsion of the couple or the assimilation of the member of the higher-ranking into the lower-ranking group.\textsuperscript{(21)} Interestingly, the “blood purity” laws are applied much more rigorously between Roma groups than to the surrounding population, and gadje outsiders, especially women, are often readily assimilated. In some cases, mixed marriages can result in the formation of a new Gypsy group, for example the Zhutane Roma in Bulgaria (literally Jewish Roma) born in the common deportation camps during World War II.\textsuperscript{(21)}

The strongest genetic indication of severely limited intergroup migrations is provided by a study of the chromosomal haplotypes surrounding the R148X founder mutation in three Vlax Roma groups, where R148X occurs at very high frequencies, with carrier rates ranging between 10 and 16\%.\textsuperscript{(29)} While the common origin of the mutation is clearly documented by a shared small conserved region around the mutated site, extended haplotypes spanning larger genetic distances have evolved independently, generating unique, group-specific haplotype profiles.\textsuperscript{(29)} The striking lack of haplotype sharing is evidence of strict inter-group endogamy throughout the history of Roma groups.

At the same time, genetic data support the declared tolerance towards outsiders, which appears to have always been part of Gypsy culture. Non-Indian Y chromosome and mtDNA lineages contribute to the genetic profile of the Gypsies (Fig. 2), with high internal diversity suggesting multiple admixture events.\textsuperscript{(25)} A higher proportion of admixed lineages and a stronger signal of demographic growth are observed in the Balkan than in the Vlax Roma (Fig. 3). The majority of Balkan groups included in the genetic studies have settled centuries ago, in contrast to a long history of nomadism persisting among the Vlax Roma until their forced sedentarisation in the 1950s. Nomadic groups are known to be more conservative and to adhere strictly to social and cultural traditions.\textsuperscript{(20,21)} The social reasons behind the different demographic regimes and proportion of admixed lineages may be related to internal differences in the tolerance to intermarriage and/or to the historically closer relationships with the surrounding populations. The harsh conditions of nomadic life may also have played a role, affecting life expectancy, population growth and genetic diversity.

Small population size, strong drift effects, limited intergroup gene flow and differential admixture have led to rapid genetic differentiation between Gypsy groups. The extent of

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{The origins of Gypsy Y chromosome and mtDNA haplogroups. Linguistic analysis has charted the Gypsy migration from India to Europe through the Hindu Kush, along Persia and the southern shoreline of the Caspian Sea, through the southern Caucasus (Armenia) and westwards to Anatolia and Byzantium. Genetic evidence of the geographical origins of paternal and maternal lineages in the Gypsies is based on the study of unique event polymorphisms of the Y chromosome (left panel) and mtDNA (right panel), compared to published data on other populations.\textsuperscript{(51–62)} Y chromosome haplogroup H-M82 and mtDNA haplogroup M can be unambiguously assigned to Asia and the Indian subcontinent. The wide geographical distribution of the remaining lineages precludes the tracing of their origins in the Gypsy population. The internal diversity of the latter lineages points to multiple admixture events during the journey or after the arrival of the Gypsies in Europe.}
\end{figure}
divergence becomes evident in comparison to global populations (Fig. 4). The most striking feature of the multidimensional scaling (MDS) plot, based on mtDNA sequences, is the dispersal and large genetic distances separating Gypsy groups. A similar and even more pronounced pattern is presented by the tribal groups of India, in sharp contrast to other (especially European) populations, which cluster closely together. Significant internal substructure in the Gypsy population is also supported by other genetic systems, namely Y chromosomes, neutral autosomal polymorphisms, and disease mutations and their linked haplotypes (refs (25,29) and Table 1). On a condensed time scale, the history of the Gypsies (and, to a large extent, the social phenomena behind it) reproduces the more widely known history of the Jews, with the exodus, diaspora and fragmentation into small geographically dispersed communities. Possibly smaller historical population size, stronger drift effects and persisting group separation, maintained by social tradition and marriage customs, account for the marked internal differentiation of the Roma, making the group the basic unit of genetic research, as it is of social organisation.

Genetic correlates of group identity
Analysis of the correspondence between genetic affinities and the cultural anthropology classification of Roma groups shows that most criteria defining group identity do not reflect genetic similarities. Neither ethnonyms nor religion shared between Roma groups predict common genetic profiles and, contrary to other Europeans, (34) genetic and geographical distances show no correlation. (25) Higher order classification into metagroups, divided by particularly rigid rules of endogamy, separates groups which are genetically very close. (26) The only cultural anthropological criterion reflecting genetic relatedness is the history of migrations, hence dialects of Romanes. (25,29) The Balkan and Vlax Roma, and the Gypsies of Western Europe have been separated since the early migrations of the 14th century. Within each migrational category, the earliest internal splits have taken place simultaneously with the migrational separations, but the gradual and prolonged process of group formation has maintained intermarriage within each category for generations, with differential admixture from surrounding European populations contributing to the distinctive genetic profiles of migrational categories. (25,29)

The impact on genetic disease
A list of the private Gypsy mutations, known to be associated with single-gene disorders, is presented in Table 2. The list is not exhaustive and reflects the interests of individual researchers rather than comprehensive studies of the genetic disease heritage of the population. The major lesson learned from Mendelian genetics so far is that the identification of a
A small number of affected Gypsy families sharing the same private mutation usually signals a widespread problem, involving large numbers of patients in many countries. Currently available frequency data\(^{(29,37)}\) show that an average of 1 in 8 subjects in the general Gypsy population is a carrier of one of the five mutations tested. Within individual Gypsy groups, carrier rates for specific mutations often exceed 5\% and can be as high as 16\%. Mendelian disorders are thus a considerable health burden, making community-based carrier testing programs a highly beneficial public health initiative. Such programs would be facilitated by the allelic homogeneity, which allows simple testing procedures and high detection rates.

The Gypsies have already made a contribution to Mendelian genetics through four novel private conditions: Hereditary Motor and Sensory Neuropathy types Lom and Russe—two severe forms of autosomal recessive Charcot-Marie-Tooth disease;\(^{(9,31,38)}\) Congenital Cataracts Facial Dysmorphism Neuropathy syndrome, a developmental disorder affecting the basal transcription machinery;\(^{(35)}\) and a new form of immune deficiency, characterised by absence of CD8\(^{+}\) cells.\(^{(45)}\) Furthermore, private mutations in the Gypsies, causing known disorders of wide ethnic distribution, facilitate genotype–phenotype correlations in large genetically homogeneous groups of affected individuals carrying the same molecular defect.\(^{(11,12,50)}\) Similar to other founder populations, gene

**Figure 4.** Genetic distances between Gypsy groups and other Eurasian populations. Fst between Gypsy groups and other Eurasian populations\(^{(63–75)}\) were computed for the mtDNA hypervariable segment 1, using the number of differences between sequences as the molecular distance. Multidimensional scaling analysis was performed to represent the distance matrix in 2 dimensions. Stress = 14.9\%. Colour codes: Gypsy black, European red, Middle Eastern green, Central Asian light blue, Pakistani blue, and Indian orange. Abbreviations: Gypsy populations include the Roma groups from Bulgaria (shown in Fig. 1), as well as Gypsies from Spain (spR); Lithuania (liR), Serbia (seR), Romania (roR) and Hungary (huR). Indian tribal populations: Koragas (kor), Bettakurumba (bet), Mullukurunan (muln), Mullukurumba (mulm), Paniya (pan), Apatani (apa), Adi (adi), Nishi (nis), Solligas (sol), Andh (anh), Naga (nog), Kuruchian (kru), Jenkurumba (jen), Thoti (tho), Pardhi (pad), Yerava (yer), Kattunaiken (kat).
mapping is based on the assumption of a single founder mutation inherited from a common ancestor on a chromosomal segment identical by descent (IBD). The unique advantage of the Gypsies as a founder population is related to the internal structure and the genetic differentiation between Gypsy groups. Independent haplotype diversification makes the informed sampling of affected individuals from different Gypsy groups a particularly efficient approach, where the critical gene interval is defined as the minimum region shared IBD between groups. Incorporating population history in gene-cloning strategies has prompted the novel “Not-Quite Identical By Descent” (NQIBD) approach, which could be applicable to other young isolates as well. In this approach, the identification of the mutated gene was greatly facilitated by tracing the origin of the mutation to a recent event occurring on a common haplotype background, which allowed the detection

**Table 1. Genetic differentiation of Roma groups in Bulgaria**

<table>
<thead>
<tr>
<th>Genetic cluster</th>
<th>BLA N ¼ 54</th>
<th>DAR N ¼ 46</th>
<th>KAL N ¼ 33</th>
<th>KAN N ¼ 69</th>
<th>KAS N ¼ 39</th>
<th>LOM N ¼ 71</th>
<th>MUS N ¼ 51</th>
<th>RUD N ¼ 51</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.1</td>
<td>37</td>
<td>2</td>
<td>1</td>
<td>10</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>C.2</td>
<td>7</td>
<td>27</td>
<td>1</td>
<td>12</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>C.3</td>
<td>1</td>
<td>2</td>
<td>21</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>C.4</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>21</td>
<td>6</td>
<td>8</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C.5</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>C.6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td>32</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>C.7</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td>C.8</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>45</td>
</tr>
</tbody>
</table>

Analysis of 28 unlinked autosomal microsatellites selected from the World Diversity Panel (49) among the top 100 markers providing the best differentiation between European and Asian populations (the set overlaps partly with the markers used in a similar analysis of Jewish populations (33)). Stratification was assessed using the software tool Structure (vs 2.0) (81) under a non-admixture model, assuming 8 underlying sub-populations. The observed genetic clusters follow the boundaries of self-declared Roma group identity, with most clusters formed mainly (in some cases exclusively) by individuals from a specific group. Significant genetic differentiation is also revealed by analysis of the pairwise differences between mtDNA sequences and Y chromosome haplotypes in different Roma groups (25), where Fst = 0.05, P > 0 is obtained for mtDNA sequences and Fst = 0.14 P > 0 for Y chromosome haplotypes.

**Table 2. Private mutations associated with Mendelian disorders in the Gypsies**

<table>
<thead>
<tr>
<th>Disorder</th>
<th>OMIM #</th>
<th>Inheritance</th>
<th>Locus</th>
<th>Gene</th>
<th>Mutation</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>With demonstrated founder effect</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HMSN-Lom&lt;sup&gt;1&lt;/sup&gt;</td>
<td>601455</td>
<td>AR</td>
<td>8q24</td>
<td>NDRG1</td>
<td>R148X</td>
<td>9,31</td>
</tr>
<tr>
<td>Congenital cataracts facial dysmorphism neuropathy&lt;sup&gt;1&lt;/sup&gt;</td>
<td>604168</td>
<td>AR</td>
<td>18qter</td>
<td>CTPH1</td>
<td>IVS6 + 389C &gt; T</td>
<td>35</td>
</tr>
<tr>
<td>Limb-girdle muscular dystrophy 2C&lt;sup&gt;1&lt;/sup&gt;</td>
<td>253700</td>
<td>AR</td>
<td>13q12</td>
<td>SGCG</td>
<td>C283Y</td>
<td>10</td>
</tr>
<tr>
<td>Congenital myasthenia&lt;sup&gt;1&lt;/sup&gt;</td>
<td>254210</td>
<td>AR</td>
<td>17p13</td>
<td>CHRNE</td>
<td>1267delG</td>
<td>13</td>
</tr>
<tr>
<td>Galactokinase deficiency&lt;sup&gt;1&lt;/sup&gt;</td>
<td>230200</td>
<td>AR</td>
<td>17q24</td>
<td>GKI</td>
<td>P28T</td>
<td>36,37</td>
</tr>
<tr>
<td>HMSN-Russe</td>
<td>605285</td>
<td>AR</td>
<td>10q23</td>
<td></td>
<td></td>
<td>38</td>
</tr>
<tr>
<td>Gitelman syndrome</td>
<td>263800</td>
<td>AR</td>
<td>15q13</td>
<td>SLC12A3</td>
<td>IVS9 + 1G &gt; T</td>
<td>39</td>
</tr>
<tr>
<td>Neuronal ceroid lipofuscinosis</td>
<td>606725</td>
<td>AR</td>
<td>15q21</td>
<td>CLN6</td>
<td>Unidentified, shared haplotype</td>
<td>40</td>
</tr>
<tr>
<td><strong>Reported in a small number of families, no extensive mutation frequency data yet</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Congenital glaucoma</td>
<td>231300</td>
<td>AR</td>
<td>2p21</td>
<td>CYP1B1</td>
<td>E387K</td>
<td>41</td>
</tr>
<tr>
<td>Glanzmann thrombasthenia</td>
<td>273800</td>
<td>AR</td>
<td>17q21</td>
<td>ITG2A2B</td>
<td>IVS15-1G &gt; A</td>
<td>42</td>
</tr>
<tr>
<td>Ataxia-teleangiectasia</td>
<td>208900</td>
<td>AR</td>
<td>11q22</td>
<td>ATM</td>
<td>9010_9037del28</td>
<td>43</td>
</tr>
<tr>
<td>Von Willebrand disease</td>
<td>277480</td>
<td>AR</td>
<td>12p13</td>
<td>VWF</td>
<td>Q1311X</td>
<td>44</td>
</tr>
<tr>
<td>CDB&lt;sup&gt;+&lt;/sup&gt; T-cell deficiency</td>
<td>186910</td>
<td>AR</td>
<td>2p12</td>
<td>CD8A</td>
<td>G90S</td>
<td>45</td>
</tr>
<tr>
<td>Congenital nemaline myopathy</td>
<td>102610</td>
<td>AR</td>
<td>1q42</td>
<td>ACTA1</td>
<td>R39X</td>
<td>46</td>
</tr>
<tr>
<td>Polycystic kidney disease</td>
<td>173900</td>
<td>AD</td>
<td>4q21</td>
<td>PKD2</td>
<td>R306X</td>
<td>47</td>
</tr>
<tr>
<td>Hyperparathyroidism-jaw tumour syndrome</td>
<td>145000</td>
<td>AD</td>
<td>1q31.2</td>
<td>HRPT2</td>
<td>Exon 8 del2 (TG or GT)</td>
<td>48</td>
</tr>
</tbody>
</table>

<sup>1</sup>Used in the genetic analyses of Gypsy population history (ref. 29 and 37). AR autosomal recessive, AD autosomal dominant.
of the mutation by comparing the sequence of a small part of the critical gene region between the two chromosomes of a single carrier parent.

The potential of the Gypsies to help research into common, genetically complex disorders is still in need of solid proof. The specific advantage in this case is again related to the combination of a primary founder effect—a primary founder effect of the mutation by comparing the sequence of a small part of the critical gene region between the two chromosomes of a single carrier parent.

The potential of the Gypsies to help research into common, genetically complex disorders is still in need of solid proof. The specific advantage in this case is again related to the combination of a primary founder effect—founder effects 6–24 generations ago giving rise to divergent subisolates. The strong primary founder effect raises expectations of a relatively homogeneous genetic basis of complex disorders across subisolates and, similar to the general approach used in Mendelian disorders, their unique, group-specific, haplotype profiles could provide a powerful tool for positional cloning. Based on the current data, the best approach to the identification of susceptibility genes would involve initial crude mapping in extended multiplex families from a single Gypsy group, with replication, fine mapping and association studies based on the informed cross-sampling of genetically related groups. At the same time, presumed high frequencies of diverse susceptibility alleles in all populations warrant caution and, in combination with the contribution of differential admixture to the genetic divergence of Gypsy groups, question the concept of the Gypsies as a single isolated founder population suitable for the study of genetically complex disorders. The internal stratification of the Gypsy population warrants caution in association studies, with an informed selection of cases and controls based on knowledge of the cultural anthropology and genetic affinities of the individual Gypsy groups involved.

Table 3 shows the estimated probabilities of sharing the same susceptibility allele between Gypsy groups, given a frequency of 0.01 of that allele in an individual group. These estimates, based on neutral autosomal polymorphisms, are in agreement with the observed distributions of private mutations causing Mendelian disorders, where the highest frequency attained by a given mutation in a particular Gypsy group correlated with the overall number of groups where the mutation was detected. The real situation is likely to differ between complex disorders and a true assessment can only be based on empirical studies, some (e.g. an investigation of bipolar affective disorder) currently in progress.

### Conclusion

Many challenges have already been met by recent research into the genetics of the Gypsies. Breaking the divide between cultural and genetic anthropology has been crucial to unraveling biological history and inferring the parameters of historical demography. Involvement in these studies has been beneficial for all sides—researchers have come to appreciate this unusual population and its multiple social and medical problems, and affected families and communities have seen improvements in health care. Yet the big challenge still lies ahead. The story that we have told is the genetic image of xenophobia, the fate of an Asian people fractured and dispersed among Europeans and responding to hostility with a labyrinth of walls of endogamy. It may well be applicable to the numerous Asian, Middle Eastern and African immigrant communities now forming in the midst of prosperous united Europe. Without true integration and access to education and health care, genetic studies of the Gypsies will stall and what has already been achieved will come to be regarded as a curious and brief episode outside of mainstream genetics.

### Acknowledgments

We wish to thank the numerous Gypsy families and communities for participation in these studies, Dr. Ivalo Tournevi and other colleagues from Bulgaria and other countries for sample collection and for providing data on other populations, and Elena Marushiakova and Vesselin Popov for expert advice on the cultural anthropology of the Roma and guidance throughout these studies.
Challenges

75. Schneider S, Roessli D, Excoffier L. 1999. Arlequin ver. 2.0: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
77. Schneider S, Roessli D, Excoffier L. 2000. Arlequin ver. 2.0: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.