Int J Legal Med (2007) 121:377–383
DOI 10.1007/s00414-006-0128-4

ORIGINAL ARTICLE

Hungarian mtDNA population databases from Budapest and the Baranya county Roma

Jodi Irwin · Balazs Egyed · Jessica Saunier · Gabriella Szamosi · Jennifer O'Callaghan · Zsolt Padar · Thomas J. Parsons

Received: 4 April 2006 / Accepted: 8 September 2006 / Published online: 22 December 2006

Abstract To facilitate forensic mtDNA testing in Hungary, we have generated control region databases for two Hungarian populations: 211 individuals were sampled from the urban Budapest population and 208 individuals were sampled from a Romani ("gypsy") population in Baranya county. Sequences were generated using a highly redundant approach to minimize potential database errors. The Budapest population had high sequence diversity with 180 lineages, 183 polymorphic positions, and a random match probability of 1%. In contrast, the Romani population exhibited low sequence diversity, with only 56 lineages, 109 segregating sites, and a random match probability of 8.8%. The mtDNA haplogroup compositions of the two populations were also distinct, with the large proportion of haplogroup M samples (35%) in the Roma the most obvious difference between the two populations. These factors highlight the importance of considering population structure when generating reference databases for forensic testing purposes. Comparisons between our Romani population sample and other published data indicate the need for heightened caution when sampling and using mtDNA databases of small endogamous populations. The Romani populations that we compared showed significant departures from genetic uniformity.

Keywords mtDNA · Baranya county · Romani population

Electronic supplementary material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00414-006-0128-4 and is accessible for authorized users.

J. Irwin (✉) · J. Saunier · J. O'Callaghan · T. J. Parsons
Armed Forces DNA Identification Laboratory,
Armed Forces Institute of Pathology Building 101,
1413 Reseach Boulevard,
Rockville, MD 20850, USA
e-mail: jodi.irwin@afip.osd.mil

B. Egyed · G. Szamosi · Z. Padar
Department of HaemoGenetics, Institute for Forensic Sciences,
1903 Budapest POB. 314/4, Hungary

Present address:
T. J. Parsons
International Commission on Missing Persons,
Alipasina 45 A,
71000 Sarajevo, Bosnia and Herzogovina

Introduction

MtDNA testing in the forensic context requires appropriate population databases for estimating the rarity of the questioned mtDNA profile. If the mtDNA haplotypes of the evidence and the suspect are consistent, the frequency of the haplotype, as determined from an appropriate reference database, generally provides a conservative estimate of the likelihood that the evidence haplotype originated from the suspect rather than a random individual in the population.

Because of well-established differences in mtDNA haplotype distributions between diverse population groups, the appropriate reference database for a particular questioned sample will vary depending on the circumstances surrounding the case. The conceptual ideal for the appropriate reference database would be one that reflects the mtDNA haplotype distribution of the potential pool of sample contributors. In urban populations, the ideal database might reflect diverse population groups and mitochondrial haplotypes in proportions similar to those found in the actual population. Unfortunately, very few mtDNA population databases can boast such thorough sampling and completeness, and the generation of such databases is generally impractical. Therefore, multiple

databases representing various relevant source are generally used to report frequency estimates. Especially in criminalistics, the most conservative estimate can be given the most emphasis [1–3]. This provides the suspect with the maximum benefit of the doubt. The most conservative estimate tends to be from the source population of the suspect or reference sample. However, in the absence of information about the population origin of the questioned sample (such information is usually circumstantial, but could be genetic), the population origin of the suspect or reference sample has no direct logical relevance apart from providing a conservative estimate.

We report here two new forensic mtDNA population databases from Hungary: a sample of individuals from urban Budapest and a sample of Roma from Baranya county. These two databases reflect the two largest population groups in Hungary. The population of Hungary is comprised largely (~95%) of autochthonous "Europeans." However, many ethnic minorities also reside in Hungary, with European Roma forming the largest minority (the latest national census lists self-identified Roma at 2% of the population, although the actual value is considered to be substantially higher). In Baranya county, Roma represent approximately 10% of the 440,000 or so inhabitants, and therefore constitute a significant proportion of the residents of this region.

Historical records indicate that the Romani populations of Europe originally arrived in the Balkans in the eleventh to twelfth centuries with additional migrations following into the fifteenth century [4, 5] and later dates as well. Multiple historical and linguistic data support the theory that these migrations originated in India [5, 6], although no written records exist to confirm this hypothesis. Genetic data (classical markers, Y-STR, and mtDNA) from various European Romani populations have supported the Indian origins of the European Roma [6–8]. However, these studies have also demonstrated that despite their common origin, culturally distinct Romani subpopulations possess significantly different mtDNA structures. Romani populations in Europe have been subject to severe discrimination and persecution over the centuries, including mass executions in a Roma holocaust in World War II and ethnic cleansing in the recent Balkan conflicts ([9]; and see European Roma Rights Center, http://www.errc.org). Establishment of appropriate mtDNA population databases for the Roma could be considered particularly desirable to assist in identifying missing persons or individuals unearthed from mass graves, for which mtDNA has a well-established role. It is also important that the significance of mtDNA matching, should it occur in criminalistic casework involving Romani individuals, is not overestimated through comparison to inappropriate databases. For such small, culturally isolated populations, it is also important to assess if there is genetic uniformity between different localities, again to ensure that comparisons are not made to inappropriate databases.

Materials and methods

DNA samples and extraction

Blood samples were obtained from 211 unrelated individuals from urban Budapest, Hungary and 205 unrelated, self-declared Roma from Baranya county, Hungary. DNA was extracted as in Egyed et al. [10]. To the best of our knowledge, these samples represent individuals who are unrelated by at least two generations. The individual samples in this study were essentially the same as those in a report on nuclear short tandem repeats (STR) [10], with a few individuals more (Roma) or less (Budapest). Additional STR loci from a subset of the samples in this paper were published in Füredi et al. [11] and Egyed et al. [12].

Amplification and sequencing of mtDNA

Polymerase chain reaction (PCR) amplification and amplicon sequencing were performed as in Brandstätter et al. [13] with minor modifications. Automated PCR setup was performed on a Corbett CAS-1200 robotic workstation and sequencing was performed on a Tecan Genesis 200. Primers used for sequencing, in addition to the amplification primers, were: F16190 (CCCCATGCTTACAAGCAAGT), F16450 (GCT CCG GGC CCA TAA CAC TTG), F15 (CAC CCT ATT AAC CAC TCA CG), F314 (CCG CTT CTG GCC ACA GCA CT), R285 (GTT ATG ATG TCT GTG TGG AA), and R16400 (GTC AAG GGA CCC CTA TCT GA).

Sequences were aligned using Sequencher software (GeneCodes, Ann Arbor, MI). The nucleotide positions considered for analysis were 16024–16569 and 1–576, as defined by the revised Cambridge reference sequence (rCRS, [14, 15]). Sequences are denoted in this paper as a list of polymorphisms that differ from the rCRS. Observed point heteroplasmies were denoted by the appropriate International Union of Biochemistry code, and observed length heteroplasmies were treated consistently, with the predominant molecule described.

To ensure data quality, a redundant multilaboratory approach to data generation and analysis was used. Details of the approach can be found in Brandstätter et al. [13]. Sequence data in electronic form are available from the authors upon request or may be downloaded directly from GenBank (Baranya Roma, GenBank accession numbers DQ359273–DQ359477; Budapest, GenBank accession numbers DQ359478–DQ359688).

Analysis of population data and comparison to other populations

Pairwise comparisons within and between the Romani and Budapest populations were performed using custom software (LISA, Future Technologies, Fairfax, VA), with C-stretch insertions at 16193, 309, and 573 ignored. Genetic diversity indices and random match probabilities were calculated by hand [(1−sum of squares of haplotype frequency)×(n/n−1); see [16]]. Random match probabilities were generated both empirically and as the sum of squares of the haplotype frequencies. For these comparisons, entire control region data were used.

The Baranya county Romani population was compared to six other European Romani populations described in Gresham et al. [7]. Only populations for which 20 or more individuals were characterized were used in our analyses. Thus, the 205 Hungarian Roma were compared to five Bulgarian Romani populations (Monteni, n=42; Lom, n=43, Kalderash, n=23; Kalaidjii North, n=25; Turgovzi, n=25) and one Spanish Romani population (n=25) [7]. Analysis of molecular variance (AMOVA) were conducted on HV1 data spanning positions 16023 to 16384, using 1,000 permutation replicates, an alpha of 0.3, and the Kimura two-parameter method for calculating distance [17].

The urban Budapest population was compared to the SWGDAM "Caucasian" database. To include the maximum number of samples in the analysis, only positions falling within HVI (16024–16365) and HVII (73–340) were considered for these comparisons.

mtDNA haplogroup affiliation

Haplogroups were assigned to sequences based on motifs of control region polymorphisms that are either specific to, or associated with, recognized haplogroups [7, 8, 18–33]. In cases where there were one or several differences from reported "defining" polymorphisms, we evaluated the significance with relation to the relative evolutionary rate of the sites in question (TJ Parsons, unpublished data, but similar to relative rates in, e.g., [3]). In two cases, we were unable to assign the sample to a particular haplogroup due to a lack of defining polymorphisms.

Results and discussion

Population comparisons

Romani individuals from Baranya county, Hungary (n=205), and native Hungarian individuals from Budapest (n=211) were sequenced for the mtDNA control region. These two population samples differed dramatically at the mtDNA level. To begin with, the genetic diversity of mtDNA within the Romani sample was much lower than in the Budapest sample (0.915 vs 0.995), with an intrapopulation pairwise match probability in the Roma ~13-fold higher than in the Budapest sample (Table 1). In the Budapest sample, 183 variable sites were observed, while only 109 sites were observed in the Roma. Despite the similar sample size, the Budapest sample had more than three times the number of distinct haplotypes than the Roma and a fivefold greater proportion of the number of haplotypes that were unique in the population samples. The five most common haplotypes in the Romani population account for ~44% of the sample, with the most common type occurring in ~16% of the population. A close relative of the most common mtDNA type, differing only by the presence of an AC repeat indel at position 524, occurred at ~8% of the population.

Pairwise sequence comparisons were performed in both populations. The mismatch distribution for the Budapest samples was smoothly unimodal, centered at nine differences (mean=9.7). The ragged, bimodal histogram of the Romani sample showed two peaks: 1 at zero differences, which reflects the high empirical random match probability of 7.12%, and another at 12 differences. The peak at 12 differences likely represents the multiple pairwise comparisons between divergent mtDNA types (such as the common M haplogroup and H haplogroup samples). Unimodal mismatch distributions tend to reflect populations that have expanded from an initial pool, while ragged distributions tend to result from a more constant population size over time and/or heterogeneous founder events [34].

For both population samples, the control region sequences were assigned to probable evolutionary haplogroups by identification of haplogroup-diagnostic or haplogroup-associated sequence motifs. The haplogroup representation within the Budapest and Romani samples was also strikingly different. The Budapest sample was very similar in haplogroup representation to that of the pooled "Caucasian" (US and European) samples in the FBI/SWGDAM database.

Table 1 Diversity measures for urban Budapest and Baranya County Roma populations

Population statistics	Urban Budapest (n=211)	Baranya county Roma (n=205)
Random match prob.		
Empirical	0.53%	7.12%
Sum of squares	1.00%	8.97%
Haplotypes	180 (167 unique)	57 (33 unique)
Polymorphic sites	183	109
Mean pairwise differences	9.5	9.8
Genetic diversity	0.995	0.915

Indices were generated from the entire control region data (16024–16576), with C insertions at 16193, 16309, and 16573 ignored.

Of the Budapest samples, 94% fell into various western European haplogroups, with 40% of all samples belonging to haplogroup H. The most common haplotype, 16519C 263G 315.1C, was seen in 6.2% of the samples (HVI/HVII only: 263G 315.1C, seen in 7.5% of samples). In the SWGDAM database, this HVI/HVII haplotype was observed in 6.7% of the samples. The high frequency of this type and of haplogroup H is distinctive of western European populations (e.g., [13, 18, 35]).

The majority of the Romani samples were assigned to known haplogroups. Thirty-five percent of the samples derived from haplogroup M (including the most common haplotype), but these were represented by only six haplotypes. Nearly all of the remaining samples fell into various western European haplogroups (H, X, T, J, U, V, I, W, K). Only three sequences (1.5%) were attributable to other Eurasian or African haplogroups (two individuals from Eurasian haplogroup N1b1, one from African haplogroup L3f). The mtDNA haplogroup distribution of the Baranya county Roma was much more similar to that seen in a study of 14 European Roma HVI sequences by Gresham et al. [7]. Gresham et al. [7] observed a large proportion of haplogroup M samples (26.5%), with the remaining mtDNA lineages comprised largely of western European haplogroups. The single most common haplotype seen in the Baranya county Roma was also the most common haplotype observed overall in Gresham et al. [7]. However, the overall representation of this haplotype in the Baranya county Romani population was much greater than in the composite European Romani population. This M5a haplotype (73G 263G 315.1C 489C 524.1A 524.2C 16129A 16223T 16291T 16298C 16519C) was seen 51 times in our Hungarian Romani database, representing 24.5% of all sequences and 73% of the observed M lineages. In the composite European Romani populations, this haplotype represented 10% of all sequences and 40% of the observed M lineages. The second most common haplotype seen in the Gresham et al. [7] study, a haplogroup M sequence, which represented 8.4% of the total database, was not observed at all in the Baranya county Roma.

Despite the general similarity of the Hungarian to other European Romani populations based on haplogroup representation, especially compared to autochthonous European populations, there were clearly substantial haplotype frequency differences among the Romani populations we compared. Previous genetic studies and knowledge of population history and social structure (e.g., [8]) also indicated the potential for significant differences among Romani populations that should be evaluated carefully with regard to potential forensic applications. We found no published complete control region sequences for the Roma, so we further explored Roma interpopulation variation by comparing the Baranya county Roma data to HV1 sequences from six other European Romani populations [7]. Diversity measures from the seven populations are shown in Table 2. All Romani population samples displayed significantly lower genetic diversities than the population sample from urban Budapest (H=0.995). Relative to the other Romani populations, the Hungarian Roma displayed one of the lowest genetic diversities with only 43 haplotypes in 205 individuals. The Hungarian Roma did, however, have a somewhat higher average pairwise difference between samples and were polymorphic at more than twice as many sites as many of the other Romani populations (the latter is certainly a result of a much larger sample size).

AMOVA analyses conducted on the seven European Romani populations indicated that over 5.15% of the total variation in the pooled population sample could be attributed to interpopulation differences between the Roma, with the remainder attributable to variation within populations. Pairwise F_{st} values for the Hungarian Baranya county Roma and each of the other European Romani populations ranged from 0.014 to as much as 0.100, and all pairwise F_{st}s were significant at the 0.05 level (Table 3). Likewise, pairwise F_{st} values between the non-Hungarian Romani populations ranged between −0.001 and 0.095. Of the 21 total pairwise comparisons of all Romani populations, only four were not significant at the 0.05 level. To get a feel for the magnitude of this interpopulation differentiation, we can compare to a similar analysis combining our Budapest sample with four European-derived populations represented in the FBI/SWGDAM database (US "Caucasian," Austria, France, Greece; [36, 37]). In this case, only 0.02% of the

Table 2 Diversity measures for various Roma populations based on HVI data (this study, [7, 38])

	Lom (Bulgarian Vlax Roma)	Rudari (Bulgarian Vlax Roma)	Kalderas (Bulgarian Vlax Roma)	Baranya County, Hungary Roma	Spanish Roma	Turgovzi Roma
Sample size	43	42	23	205	25	25
Genetic diversity	0.948	0.913	0.949	0.914	0.9733	0.977
Alleles	18	15	15	43	11	19
Polymorphic sites	23	25	22	54	23	31
Mean pairwise differences	4.28	4.72	4.44	5.17	4.07	6.24

HVI defined by positions 16024 and 16384, in this case.

Table 3 Fst values for Baranya County Roma and other European Roma Populations

	Hungarian Roma (n=205)	Kalderash (Bulgaria) (n=23)	Monteni (Bulgaria) (n=42)	Lom (Bulgaria) (n=43)	Turgovzi (Bulgaria) (n=25)	Kalaidjii North (Bulgaria) (n=20)	Spanish Roma (n=25)
Hungarian Roma	–	0.021*	0.088**	0.013*	0.035*	0.038*	0.100**
Kalderash		–	0.033*	0.004	−0.001	0.034	0.085**
Monteni			–	0.058**	0.044**	0.095**	0.12**
Lom				–	0.011*	0.009	0.084**
Turgovzi					–	0.011*	0.068**
Kalaidjii North						–	0.066*
Spanish Roma							–

Data for the non-Hungarian Roma were taken from Gresham et al. [7].
*p value <0.05
**p value <0.01

total variation can be attributed to interpopulation differences. Within these autochthonous European populations, there was only one significant pairwise F_{st} value (Budapest to France, F_{st}=0.004, p=0.04), whose significance is eliminated with the application of the Bonferroni correction. Despite the distinctiveness of the Romani populations at the level of haplogroup representation, consistent with a common ancestry, there are clearly factors that restrict matrilineal gene flow among various Romani groups, even those within close geographical proximity (see also [7, 8, 38]). The Budapest sample, on the other hand, is quite similar to other autochthonous European samples and there appears to be no basis for treating the Budapest population sample separately from other combined European databases for the purposes of forensic comparisons.

Database reporting issues

Population databases are used to indicate the relative rarity of mtDNA sequences encountered in forensic casework and, in most cases, the evidentiary significance of mtDNA matching improves as database size increases. As a result, large mtDNA databases are generally desirable and pooling of genetically similar databases is something that can be done to establish larger databases—if this is warranted by an absence of population differentiation. We have established that the Budapest sample is similar to other European samples and need not be treated separately. However, given the substructure detected among Romani populations, it would likely be inappropriate to pool the various Romani population databases to create a larger database. To further address the issue, we examined the frequency of each Romani population's most common haplotype in all of the other Romani populations and in a combined "European Roma" database. In nearly all cases, the frequency of the most common haplotype in one population was underestimated (or in many cases absent) in all other populations. The only exceptions to this were among those populations that shared the same most common type. Furthermore, in all but one case, the use of the pooled database significantly underestimated (p<0.05) the frequency of each individual population's most common haplotype, suggesting that frequency estimates of particular sequences would be most conservative if populations were considered separately.

Heteroplasmy

The data were carefully scrutinized for point mutation heteroplasmy, and in each population we detected multiple instances where the sequence data from both strands strongly indicated a true heteroplasmic mixture. Point heteroplasmies were observed in 9 Romani individuals and 11 Budapest individuals at a total of 17 positions. Three individuals were heteroplasmic at 16183, two individuals were heteroplasmic at 16093, and the remaining 15 individuals were heteroplasmic at 15 unique sites, for a total incident rate of 4.8%. This rate is higher than is typically reported from blood samples (e.g., [39]), but is consistent with the rates we are observing in other databasing projects. This is likely due to increased scrutiny, very clean sequence data, and increased sensitivity of current chemistry and detection platforms.

Of course, length heteroplasmy was common in the C-stretch regions of HVI and HVII. We also observed possible/apparent length heteroplasmy in the AC dinucleotide repeat element at position 523. As reported elsewhere (e.g., [40, 41], a majority of the samples showed five AC repeats and the number of elements was highly variable in the populations. Heteroplasmy was regularly observed, albeit at an extremely low level, in samples with five repeats. However, in all cases where the predominant repeat number was six or greater, we were able to consistently detect apparent repeat length heteroplasmy, which was generally more pronounced with seven or more predomi-

nant repeat units. As the number of repeats increased, the proportion of minor component molecules and our ability to detect them also increased. This observation is consistent with our understanding of STRs: that the incidence of PCR amplification stutter increases with an increasing number of repeats [42]. However, it is also possible that the observed heteroplasmy is intrinsic. We are not aware that others have reported heteroplasmies with such frequency, and in many instances the apparent mixture is low enough to be hard to detect without very clean data. Chung et al. [43] report just three instances of AC length heteroplasmy in 500 individuals using fluorescently labeled amplicons and reported that PCR from cloned amplicons with greater than six repeats did produce some stutter. Properly designed cloning/amplification experiments could conclusively determine if these length mixtures are due to intrinsic heteroplasmy or PCR artifacts.

Conclusions

Because of the unique cultural and genetic heritage of European Romani populations, we investigated the importance of generating independent mtDNA reference databases for the Romani population of Baranya county, Hungary and the population of Budapest, Hungary. We also examined the degree of mtDNA differentiation between endogamous populations of European Roma and considered the possibility of generating a combined database for European Romani populations.

The Baranya county Romani population was found to be significantly different from all other European Romani populations included in the AMOVA analysis. These results, as well as our observations on the relative rarity of each population's most common type in all other Romani and pooled populations, suggest that mtDNA frequency estimates of particular sequences would be most conservative if Romani populations were considered separately, instead of pooled together and treated as a single database. Additional sampling and larger sample sizes from Romani populations will provide additional information on the extent of mtDNA differentiation between, and haplotype representation among, various populations. Until then, mtDNA frequency estimates should be treated with much caution, due to small sample sizes and extreme subdivision. Similarly, but to an obviously larger degree, the diversity measures presented here highlight striking differences in the mtDNA compositions of the Hungarian Romani and Budapest—and other European—populations. Clearly, these databases should be maintained separately. The results of these comparisons suggest the merit of additional investigations into other distinct ethnic minority populations in Hungary and emphasize the need for particularly careful database sampling and analysis when dealing with small, isolated populations.

Acknowledgements The authors wish to thank Hans-Jürgen Bandelt (University of Hamburg, Germany) for assistance with haplogroup assignment and Miklos Angyal (Forensic Science Unit, Baranya County Police Department, Hungary) for carefully collecting the Romani population samples. We would also like to thank Rebecca Just, Katharine Strouss, Kimberly Sturk, Toni Diegoli, Carla Paintner, and Ilona Lletmanyii (AFDIL), as well as Walther Parson and Anita Brandstätter (EMPOP, University of Innsbruck, Austria) for assistance with database confirmation, and James Ross (AFDIL) and Jon Norris (Future Technologies, Fairfax, VA) for bioinformatics assistance. We thank James C. Canik, Kevin S. Carroll, Brion C. Smith (AFDIL), and Sandor Füredi, Janos Woller, Sophia Lontai-Santora (IFS, Budapest, Hungary) for logistical, administrative, and moral support. The opinions and assertions contained herein are solely those of the authors and are not to be construed as official or as views of the US Department of Defense or the US Department of the Army.

References

1. National Research Council Committee on DNA Technology in Forensic Science (1992) DNA technology in forensic science. National Academy Press, Washington DC
2. Röhl A, Brinkmann B, Forster L, Forster P (2001) An annotated mtDNA database. Int J Legal Med 115:29–39
3. Forster P, Calì F, Röhl A et al (2002) Continental and subcontinental distributions of mtDNA control region types. Int J Legal Med 116:99–108
4. Rex-Kiss B, Szabo L, Szabo S, Hartmann E (1973) ABO, MN, Rh blood groups, Hp types and Hp level, Gm(1) factor investigations on the Gypsy population of Hungary. Hum Biol 45:41–61
5. Kalaydjieva L, Gresham D, Calafell F (2001) Genetic studies of the Roma (Gypsies): a review. BMC Med Genet 2:5
6. Iovita R, Schurr T (2004) Reconstructing the origins and migrations of diasporic populations: the case of the European Gypsies. Am Anthropol 106:267–281
7. Gresham D, Morar B, Underhill PA et al (2001) Origins and divergence of the Roma (gypsies). Am J Hum Genet 69:1314–1331
8. Kalaydjieva L, Calafell F, Jobling M et al (2001) Patterns of inter and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages. Eur J Hum Genet 9:97–104
9. Human Rights Watch (1999) Abuses against Serbs and Roma in the new Kosovo. Human Rights Watch Report, vol 11, No 10
10. Egyed B, Füredi S, Angyal M, Boutrand L, Vandenberghe A, Woller J, Pádár Z (2000) Analysis of eight STR loci in two Hungarian populations. Int J Legal Med 113:272–275
11. Füredi S, Angyal M, Kozma Z, Sétáló J, Woller J, Pádár Z (1997) Semi-automatic DNA profiling in a Hungarian Romany population using the STR loci HumVWFA31, HumTH01, HumTPOX, and HumCSF1PO. Int J Legal Med 110:184–187
12. Egyed B, Füredi S, Angyal M, Balogh I, Kalmar L, Pádár Z (2005) Analysis of the population heterogeneity in Hungary using fifteen forensically informative STR markers. Forensic Sci Int 158:244–249
13. Brandstätter A, Peterson CT, Irwin JA, Mpoke S, Koech DK, Parson W, Parsons TJ (2004) Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. Int J Legal Med 118(5):294–306

14. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147
15. Anderson S, Bankier AT, Barrell BG et al (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465
16. Nei M (1987) Molecular evolutionary Genetics. Columbia University Press, New York
17. Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491
18. Achilli A, Rengo C, Magri C, Battaglia V et al (2004) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool Am J Hum Genet 75:910–918
19. Achilli A, Rengo C, Battaglia V et al (2005) Saami and Berbers—an unexpected mitochondrial DNA link. Am J Hum Genet 76:883–886
20. Allard MW, Miller K, Wilson M, Monson K, Budowle B (2002) Characterization of the Caucasian haplogroups present in the SWGDAM forensic mtDNA dataset for 1771 human control region sequences. J Forensic Sci 47:1215–1223
21. Bandelt HJ, Alves-Silva J, Guimaraes PE et al (2001) Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. Ann Hum Genet 65:549–563
22. Brehm A, Pereira L, Bandelt HJ, Prata MJ, Amorim A (2002) Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. Ann Hum Genet 66:49–60
23. Finnilä S, Lehtonen MS, Majamaa K (2001) Phylogenetic network for European DNA. Am J Hum Genet 68:1315–1320
24. Lutz S, Weisser HJ, Heizmann J, Pollak S (1998) Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. Int J Legal Med 111:67–77
25. Mishmar D, Ruiz-Pesini E, Golik P et al (2003) Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci USA 100:171–176
26. Palanichamy MG, Sun C, Agrawal S et al (2004) Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet 75:966–978
27. Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. Nat Genet 23:437–441
28. Reidla M, Kivisild T, Metspalu E et al (2003) Origin and diffusion of mtDNA haplogroup X. Am J Hum Genet 73:1178–1190
29. Richards M, Macaulay V, Hickey E et al (2000)Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67:1251–1276, Nov
30. Salas A, Richards M, De la Fe T et al (2002) The making of the African mtDNA landscape. Am J Hum Genet 71:1082–1111
31. Torroni A, Rengo C, Guida V et al (2001) Do the four clades of the mtDNA haplogroup L2 evolve at different rates? Am J Hum Genet 69:1348–1356
32. Watson E, Forster P, Richards M, Bandelt HJ (1997) Mitochondrial footprints of human expansions in Africa. Am J Hum Genet 61:691–704
33. Wilson FH, Hariri A, Farhi A et al (2004) A cluster of metabolic defects caused by mutation in a mitochondrial tRNA. Science 306:1190–1194
34. Rogers AR, Harpending HC (1992) Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9:552–569
35. Loogväli EL, Roostalu U, Malyarchuk BA et al (2004) Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. Mol Biol Evol 21:2012–2021
36. Monson KL, Miller KWP, Wilson MR, DiZinno JA, Budowle B (2003) The mtDNA population database: an integrated software and database resource for forensic comparison. Forensic Science Communications Online. Available at http://www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm
37. Miller KW, Budowle B (2001) A compendium of human mitochondrial DNA control region: development of an international standard forensic database. Croat Med J 42:315–327
38. Chaix R, Austerlitz F, Morar B, Kalaydjieva L, Heyer E (2004) Vlax Roma history: what do coalescent-based methods tell us? Eur J Hum Genet 12:285–292
39. Melton T (2004) Mitochondrial DNA heteroplasmy. Forensic Sci Rev 16:1–20
40. Lutz S, Wittig H, Weissner HJ et al (2000) Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions? Forensic Sci Int 113:97–101
41. Bini C, Ceccardi S, Luiselli D, Ferri G, Pelotti S, Colalongo C, Falconi M, Pappalardo G (2003) Different informativeness of the three hypervariable mitochondrial DNA regions in the population of Bologna (Italy). Forensic Sci Int 135:48–52
42. Butler JM (2005) Forensic DNA typing: biology, technology, and genetics of STR markers, 2nd edn. Elsevier, New York
43. Chung U, Lee HY, Yoo J, Park MJ, Shin K (2005) Mitochondrial DNA CA dinucleotide repeats in Koreans: the presence of length heteroplasmy. Int J Legal Med 119:50–53