Searching for a Genetic Relationship Between the Mitochondrial DNA and Y-chromosomal Polymorphisms

Paulina Wolańska-Nowak

TestDNA Laboratorium, Katowice, Poland

Agnieszka Parys-Proszek

Institute of Forensic Research, Section of Forensic Genetics, Krakow, Poland

Witold Pepiński

Department of Forensic Medicine, Medical University of Bialystok, Poland

Ireneusz Sołtyszewski

Department of Criminalistics and Forensic Medicine, University of Warmia and Mazury, Olsztyn, Poland
Corresponding author: e-mail: ireneusz.soltyszewski@uwm.edu.pl

Abstract: The present paper addresses the question of linkage disequilibrium (LD) between Y-chromosomal and mtDNA HVR1 distributions based on a population sample of southern Poland. Genetic independence of both polymorphisms would validate the use of the product rule in the assessment of matching probability, as is accepted for autosomal STRs. For this purpose, a total of 330 typed unrelated male samples were assigned to 17 Y-haplogroups, 32 mtDNA haplogroups, and 91 combined haplotypes (including 50 singletons). The average gene diversity for the two polymorphisms was 0.71045. The expected heterozygosity values were 0.69491 and 0.72598 for Y-chromosomal and mtDNA haplogroups, respectively. The exact test for pairwise LD indicated that there was no association between the analysed polymorphisms (median p-value=0.30781). The normalized entropy difference (NED) was 0.02. These results prove lack of association between mitochondrial and Y-chromosomal markers and provide acceptable arguments for multiplying together likelihood ratios obtained from both genetic markers without any further assumptions.

Introduction

Forensic experts are required to quantify their findings and to assess the range of uncertainty associated with the inferences that may be drawn from different genetic markers (Aitken, & Taroni, 2004). In the routine genetic identification it is not uncommon to combine likelihood ratios (LRs) obtained from autosomal, and uniparentally inherited Y-chromosomal, and/or mitochondrial markers (Wolańska-Nowak, Branicki, Parys-Proszek, & Kupiec, 2008). Both lineage polymorphisms are physically unlinked, thus a question arises as to the possible use of the combined LR values obtained from Y-STR and mtDNA analyses, considering different inheritance modes. Walsh et al. (Walsh, Redd, & Hammer, 2008) found that it is suitable to calculate combined match probabilities by multiplying Y-haplotype frequencies and appropriately corrected autosomal frequencies. Even if the population

substructure is present based on autosomal markers (coancestry coefficient $\vartheta > 0.01$), the additional effect due to the autosomal conditioning in a Y-chromosome match is low ($\vartheta \approx 0.02\text{-}0.04$) and the decrease in the combined LR is negligible. Population data demonstrated that the lineage markers occasionally present lower homogeneity in the geographic dispersion when compared to that of autosomal markers (Bardo et al., 2013; Brion et al., 2004). Hence, the levels of possible Y-chromosomal and mtDNA disequilibria (or lack there-of) should be empirically quantified, rather than assumed in the population of interest (Zaykin, Pudovkin, & Weir, 2008). The autosomal DNA estimate is based on the product rule. On the other hand, mtDNA or Y-haplotype estimates are based on frequency counts in reference databases, thus their product may provide the primary assessment of the combined match probability. This approach is problematic in terms of the information inherent both in the Y-haplotype and the mitotype that may contain strong bias to the population of origin, hence combining different types of genetic markers should be based on appropriate local databases (Buckleton, Krawczak, & Weir, 2011; Cockerton, McManus, & Buckleton, 2012). The value of DNA evidence may be updated by developing the nuclear DNA estimate using the subpopulation formula of Balding and Nichols, and an appropriate ϑ estimate for lineage markers, respectively (Balding & Nichols, 1994; Buckleton, Triggs, & Walsh, 2005; Roewer et al., 2000).

$$\theta_{mt} + (1 - \theta_{mt}) \frac{(count_{mitotype} + 2)}{N_{mt\,database} + 3}$$

and

$$\theta_{Y-STR} + (1 - \theta_{Y-STR}) \frac{(count_{Y\,haplotype} + 2)}{N_{Y\,database} + 3}$$

where ϑ_{mt} – the coancestry coefficient for the mtDNA estimate; $\vartheta_{Y\text{-}STR}$ – the coancestry coefficient for the Y-STR estimate.

The subject of theta estimates for the Polish populations was discussed elsewhere (Pepinski et al., 2004a; Pepinski et al., 2004b; Wolańska-Nowak, 2000; Wolańska-Nowak, 2007; Wolańska-Nowak, Branicki, Parys-Proszek, & Kupiec, 2009; Wolańska-Nowak, Parys-Proszek, & Sołtyszewski, 2014). Similarly to autosomal markers, the higher the ϑ estimate for Y-chromosomal and mtDNA haplotypes, the lower the value of DNA evidence. Thus, if disequilibrium is expected between these markers, LR should not be calculated by a product rule. The combined likelihood ratio (LR_{comb}) on the hypothesis that the suspect's DNA is in the evidence may be computed as below:

$$LR_{comb} = \frac{\Pr(E_A, E_{Y-STR}, E_{mtDNA} | H_p)}{\Pr(E_A, E_{Y-STR}, E_{mtDNA} | H_d)}$$

$$= \frac{1}{\Pr(E_A | E_{Y-STR}, E_{mtDNA}, H_d) \Pr(E_{Y-STR} | E_A, E_{mtDNA}, H_d) \Pr(E_{mtDNA} | E_A, E_{Y-STR}, H_d)}$$

whe ... mosome haplotypes; H_p – hypothesis that the evidence comes from the suspect's relative; H_d –hypothesis that the evidence comes from a random, unrelated person.

Considering a possible substructure in the Polish population (Wolańska-Nowak, 2000; Wolańska-Nowak, 2007) the objective of the study was to verify empirically the presence of pairwise LD between mtDNA haplogroup, Y-haplotypes, and Y-haplogroup distributions that would give rationale for using the following formula:

$$LR_{comb} = \frac{\Pr(E_A, E_Y, E_M | H_p)}{\Pr(E_A, E_Y, E_M | H_d)} \approx \frac{\Pr(E_A | H_p)}{\Pr(E_A | H_d)} \frac{\Pr(E_Y | H_p)}{\Pr(E_Y | H_d)} \frac{\Pr(E_M | H_p)}{\Pr(E_M | H_d)}$$

2. Material and methods

2.1. Haplogrouping the samples

Mitochondrial HVR1, and Y-haplotypes were analysed in 330 unrelated male samples from southern Poland. Surname repetition was avoided and used as a criterion for non-relatedness. DNA was extracted with Qiagen magnetic beads by Biorobot M48 (Qiagen) and quantified using Quantifiler kit (Applied Biosystems) in 7500 Real-Time PCR System. Y-haplotypes were obtained using Yfiler™ PCR Amplification Kit and 3130xl Genetic Analyzer with GeneMapper ID-X v.1.0 software (Applied Biosystems). HVR1 mitotypes were obtained using BigDye Terminator Cycle Sequencing Ready Reaction kit v.1.1 and 3100 Genetic Analyser (Applied Biosystems). Mutations were defined by aligning and comparing with the revised Cambridge Reference Sequence (rCRS) using SeqScape software. In the analysed sample of 330 male individuals more than 50 polymorphic sites were found in the HVR1, hence the data were haplogrouped to decrease the number of polymorphisms and to strengthen the power of the statistical tests. The haplogroups were predicted using a Web-based tool mtDNAmanager (*mtmanager.yonsei.ac.kr*) which allows automatic estimate of the most-probable mtDNA haplogroups based on their control region sequences (Lee et al., 2008). Y-haplogrouping was performed using Athey's Haplogroup Predictor (*hprg.com/hapest5/*) (Athey, 2006).

2.2. Statistical analysis

The pairwise analysis of possible LD between the mtDNA and Y-haplogroups was performed by extension of Fisher's exact test for R×C contingency tables using Arlequin v.3.5 software (*cmpg.unibe.ch/software/arlequin35/*) (Raymond & Rousset, 1995). A total of 100,000,000 steps were performed in the Markov chain following 10,000 dememorization steps (the initial burn-in for the Markov sampler). Data for the input files were set in two formats: I. mtDNA haplotypes vs. Y-haplogroups; II. mtDNA haplogroups vs. Y-haplotypes. As an additional measure of the association between haplogroup variants the concept of Shannon entropy was introduced (Siegert, Roewer, & Nothnagel, 2015). It is defined as an expected value of the information content of a discrete random variable. Assuming that a haplogroup is a discrete random variable taking a finite number of possible variants with probabilities $p(x_{1, 2 \ldots x_n})$, respectively, a number that indicates the amount of uncertainty can be determined, thus function H may be interpreted as the uncertainty associated with the event that a haplogroup = x_i.

$$H = -\sum_{i}^{n} p(x_i) \times \ln p(x_i)$$

where: H − entropy; $p(x_i)$ − the relative frequency of n haplogroup(s) variants.

The probability value $p(x_i)$ is inversely related with the uncertainty (*planetcalc.com/2476/*). This concept also allows describing the association of variants between two different haplogroups (mtDNA and Y-chromosomal) by use of the Normalized Entropy Difference (NED) (Siegert, Roewer, & Nothnagel, 2015), which is defined as:

$$NED_{mtDNA,Y} = 2 \times \frac{(H_{mtDNA} + H_Y) - H_{mtDNA,Y}}{H_{mtDNA} + H_Y}$$

Where: H_{mtDNA} and H_Y − entropies for mtDNA and Y-haplogroups, respectively; $NED_{mtDNA,Y}$ − the combined entropy of both polymorphisms.

$H_{mtDNA} + H_Y$ equals the expected entropy under stochastic independence of the two estimates, and thus equals the maximum possible entropy of both lineage markers, hence NED equals 0.

3. Results and discussion

3.1. Frequency of the mtDNA and Y-haplogroups

To assess the relationship between mtDNA, and Y-STR data we decided to haplogroup results to reduce the number of genetic variants. For this reason only HVR1 data were considered. Budowle et al. (2009) noticed that the estimates became lower as more polymorphic markers were used and each haplotype became less frequent or even unique in the dataset. Within the analysed population sample 32 mitochondrial, 17 Y-chromosomal, and 91 combined haplogroups (including 50 singletons - all combinations with a frequency of 0.0033) were observed. The average combined gene diversity was 0.71045+/-0.48157). The expected "gene diversity" values for mtDNA, and Y-haplogroups were 0.72598 and 0.69491, respectively. The results are presented in Tables 1 and 2. The most frequent mtDNA haplogroups in the analysed population were: H, H5, and HV0 (frequencies of 0.50909, 0.06667, and 0.06364, respectively). Haplogroup H also accounts for 40-50% of mtDNA pool in most of European populations (Roostalu et al., 2007) (Table 1, see attachment).

We were aware that a precaution is required when using Athey's Haplogroup Predictor, since Muzzio et al. (2011) reported a 4.8 % error in the Y-haplogroup prediction for Nicaraguan population data. In our opinion however, this was due to excluding the most population-sensitive markers DYS385I and DYS385II from their calculations. The most frequent Y-haplogroups were: R1a, R1b, and I2a (frequencies of 0.51818, 0.15758, and 0.06667, respectively). These results are concordant with those obtained by Peričić et al. (2005). According to the survey of Y-haplogroups within minorities and subpopulations of Poland, the ϑ coefficient ranged from 0.004 among Upper Silesians and Polish Highlanders (Wolańska-Nowak, 2007) to 0.309 (Belorussians and Old Believers) (Pepinski et al., 2004a; Pepinski et al., 2004b). Within the same population, ϑ values for lineage markers appear to be higher than those for autosomal markers. This is easily understood, because the effective population sizes for lineage markers are usually ¼ of those for autosomal markers and often significantly lower due to differential reproductive success. Hence Y-STRs are more sensitive to genetic drift resulting in statistically significant

interpopulation differences. Due to diverse migration patterns ϑ values for mtDNA and Y-STRs may differ. Unfortunately no data on mtDNA haplogroup frequency distribution over Polish subpopulations are available. (Table 2, see attachment).

Frequency distributions of the combined mtDNA and Y-haplogroups reflect the most common haplogroups in the population of southern Poland: H R1a, H R1b, and HV0 R1a with frequencies of 0.26667, 0.07576, and 0.04242, respectively. Table 3 presents the combined observed haplogroup frequencies. The remaining combinations were found only once with a frequency of 0.0033 each (Table 3, see attachment).

3.2. Results of the linkage disequilibrium tests

Based on the exact test, LD between the mtDNA and Y-haplogroups was insignificant (p =0.0500) (Table 4). The median p-value was 0.3078. We assume that inclusion of HVR2 would increase haplogroup diversity, and thus decrease the chance of finding disequilibrium between the markers of interest. Practically, we do not expect it to have any significant effect on our results (Table 4, see attachment).

The additional test performed on LD between the mtDNA and Y-haplotypes also proved the lack of significant association among these markers. The median of the p-values was 0.4727. The results are presented in Table 5 (see attachment).

When the concept of Shannon entropy as well as all haplogroup variants were considered, the entropy values were as follows: H_Y=3.99, H_{mtDNA}=3.08, and $H_{mtDNA,Y}$=5.5. Hence, the normalized entropy difference NED=0.444. After confining the analysis to variants of frequencies>0.009 (to avoid biases associated with LD between haplogroups with rare variants) the corresponding values were as follows: H_Y=2.22, H_{mtDNA}=2.56, and $H_{mtDNA,Y}$=4.79, with NED=0.02. These results provide additional information about the lack of association between the mitochondrial and Y-chromosomal markers. The STR loci on Y chromosome are physically linked together, so the lack of disequilibrium of mtDNA haplogroups with particular alleles is a further evidence of population independence on frequencies of these markers. On the other hand, at the locus DYS456 of the lowest p value (although still >0.05) alleles that reveal certain degree of coupling with rare mtDNA haplogroups are found, which may be explained by their common genealogy (Table 6, see attachment).

In our opinion, both sensitivity of the test for LD used and the sample size are reasonable to address the issues concerning relationship between the lineage markers. Another approach (Buckleton, Triggs, & Walsh, 2005) raised the question however, as to whether the LRs calculated for lineage and autosomal markers may be multiplied together in order to obtain a single value of measure for a genetic evidence. Amorim (2008) challenged the practice of LR multiplication, mainly on the grounds that lineage markers are not individual-specific but are instead shared by the suspect's lineage. He argued that, instead of standard definitions of prosecution and defence hypotheses, H_p and H_d:

H_p: The suspect is the source of the evidence

H_d: The suspect is not the source of the evidence

the prosecution and defence hypotheses should rather be stated:

H'_p: The suspect or a lineage relative is the source of the evidence

H'_d: Neither the suspect nor a lineage relative is the source of the evidence,

when lineage markers are used for genetic analysis. Due to these reasons, Amorim concluded that "the combination of likelihood ratios from the two sources of data should be avoided." Our previous experience and the results presented in this manuscript are contrary to Amorim's suggestion. Unless the circumstances of an individual case suggest differently, there seems to be neither a logical nor a legal basis for changing the prosecution hypothesis from H_p to H'_p, i.e. for incriminating individuals with no prior evidence of being the sample donor. Moreover, if the possibility of mutation is neglected, then the likelihood values of the two hypotheses H_p and H'_p would be identical for lineage markers. In both instances, the match probability refers to a hypothetical case when two unrelated individuals share a DNA profile. This means that both defence hypotheses, H_p and H'_p are also formally the same. With this in mind, the LR is also the same for all members of the suspect's lineage (Balding & Nichols, 1994; Cockerton, McManus, & Buckleton, 2012). The main reason for legal disregarding the question of lineage markers in the familial search is apparently the use of autosomal markers that offer high expected power of exclusion of also close relatives from being the source of the evidence. According to Walsh et al. (2008) for two individuals with identical Y haplotypes, the mean time span back to their common ancestor is about ten generations. Consequently, a distant relationship may be assumed given that the increase in shared autosomal alleles is very low. The similar relationship is apparent for mtDNA. Previous analyses of autosomal STR loci revealed that ϑ value for the population of southern Poland was negligible (Wolańska-Nowak, 2000; Wolańska-Nowak, 2007). Nowadays, its updated substructure based on distributions of autosomal, mitochondrial, and Y-chromosomal markers demonstrates high degree of differentiation without major non-random population processes, including drift. The advantage of combining both marker systems was shown by Ayadi et al. (2007) In five out of seven disputable paternity cases Y-STRs were more informative for the inclusion when compared to autosomal STRs, but were similarly or less informative when maternal data were considered. These authors multiplied LRs obtained for Y STR and autosomal STRs assuming *a priori* independence. This approach is particularly useful in analyses of deficiency or mutational cases.

4. Conclusions

In 330 samples from southern Poland a total of 32 different mitochondrial, 17 Y-haplogroups, and 91 combined haplogroups were observed. We proved that there was no association between the Y-STR and mtDNA polymorphisms based on the population of southern Poland. In our opinion, the problem of a possible relationship between the lineage markers should be explored empirically for each population of interests. As long as the frequency estimation is aided by large databases (YHRD, EMPOP), the theta correction appears unnecessary. If autosomal data are available, low match probability values allow exclusion of the suspect's parents, siblings and offspring from being the source of the evidence. This is not the case when considering the

lineage markers separately. The obtained results provide acceptable arguments for the use of product rule for LR values derived from different types of genetic polymorphisms, without any further assumptions.

References

Aitken, C.G.G., Taroni, F. (2004) Statistics and the Evaluation of Evidence for Forensic Scientists. Statistics in Practice. *John Wiley & Sons*

Amorim, A.A. (2008) Cautionary note on the evaluation of genetic evidence from uniparentally transmitted markers. *Forensic Science International Genetics, 2*(4), 376-378. doi: 10.1016/j.fsigen.2008.04.001

Athey, T.W. (2006) Haplougroup prediction from Y-STR values using a Bayesian-allele-frequency approach. *Journal of Genetic Genealogy, 2*, 34-39

Ayadi, I., Mahfoudh-Lahiani, N., Makni, H., Ammar-Keskes, L., Rebai, A. (2007) Combining autosomal and Y-chromosomal short tandem repeat data in paternity testing with male child: methods and application. *Journal of Forensic Sciences, 52*(5), 1068-1072. doi: 10.1111/j.1556-4029.2007.00513.x

Badro, D.A., Douaihy, B., Haber, M., Youhanna, S.C., Salloum, A., Ghassibe-Sabbagh, M., Johnsrud, B., Khazen, G., Matisoo-Smith, E., Soria-Hernanz, D., Spencer Wells, R., Tyler-Smith, C., Platt, D.E., Zalloua, P.A. (2013) Y-Chromosome and mtDNA Genetics Reveal Significant Contrasts in Affinities of Modern Middle Eastern Populations with European and African Populations. *PLOS One, 8*(1), e54616. doi: 10.1371/journal.pone.0054616

Balding, D.J., Nichols, R.A. (1994) DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Science International, 64*(2-3), 125-140. doi: org/10.1016/0379-0738(94)90222-4

Brion, M.B., Quintans, M., Zarrabeitia, A., Gonzalez-Neira, A., Salas, V., Lareu, C., Tyler-Smith, C., Carracedo, A. (2004) Micro-geographical differentiation in Northern Iberia revealed by Y-chromosomal DNA analysis. *Gene 329*, 17-25. doi: 10.1016/j.gene.2003.12.035

Buckleton, J., Triggs, C.M., Walsh, S.J. (eds) (2005) Forensic DNA evidence interpretation. *CRC Press*

Buckleton, J.S., Krawczak, M., Weir, B.S. (2011) The interpretation of lineage markers in forensic DNA testing. *Forensic Science International Genetics, 5*(2), 78-83. doi: 10.1016/j.fsigen.2011.01.010

Budowle, B., Ge, J., Aranda, X., Planz, J., Eisenberg, A., Chakraborty, R. (2009) Texas population substructure and its impact of estimating the rarity of Y haplotypes from DNA evidence. *Journal of Forensic Sciences, 54*(5), 1016-1021. doi: 10.1111/j.1556-4029.2009.01105.x

Cockerton, S., McManus, K., Buckleton, J.S. (2012) Interpreting lineage markers in view of subpopulation effects. *Forensic Science International Genetics, 6*(3), 393-397. doi: 10.1016/j.fsigen.2011.04.020

Lee, H.Y., Song, I., Ha, E., Cho, S-B., Yang, W.I., Shin, K.J. (2008) mtDNAmanager: a Web-based tool for the management and quality analysis of mitochondrial DNA control-region sequences. *BMC Bioinformatics 9*, 483-498. doi: 10.1186/1471-2105-9-483

Muzzio. M., Ramallo, V., Motti, J.M.B., Santos, M.R., López Camelo, J.S., Bailliet, G. (2011) Software for Y-haplogroup predictions: a word of caution. *International Journal of Legal Medicine, 125*(1), 143-147. doi: 10.1007/s00414-009-0404-1

Pepinski, W., Niemcunowicz-Janica, A., Skawronska, M., Koc-Zorawska, E., Janica, J., Soltyszewski, I. (2004a) Y-chromosome STR haplotypes in a population sample of the Byelorussian minority living in the northeastern Poland. *Forensic Science International, 140*(1), 117-121. doi: 10.1016/j.forsciint.2003.11.003

Pepinski, W., Niemcunowicz-Janica, A., Skawronska, M., Koc-Zorawska, E., Janica, J., Soltyszewski, I. (2004b) Y-chromosome STR haplotypes and alleles in the population sample of Old Believers residing in the Northeastern Poland. *Forensic Science International,* 143(1), 65-68. doi: 10.1016/j.forsciint.2004.02.009

Peričić, M.L., Lauc, B., Martinović Klarić, I., Rootsi, S., Janićijević, B., Rudan, I., Terzić, R., Čolak, I., Kvesić, A., Popović, D., Šijački, A., Behluli, D., Dorđević, L., Efremovska, D.D., Bajec, B.D., Stefanović, R., Villems, R., Rudan, P. (2005) High-Resolution Phylogenetic Analysis of Southeastern Europe Traces Major Episodes of Paternal Gene Flow Among Slavic Populations. *Molecular Biology and Evolution, 22*(10), 1964-1975. doi: org/10.1093/molbev/msi185

Raymond, M., Rousset, F. (1995) An Exact test For Population Differentiation. *Evolution, 49*(6), 1280-1283. doi: 10.1111/j.1558-5646.1995.tb04456.x

Roewer, L., Kayser, M., de Knijff, P., Anslinger, K., Caglia, A., Corach, D., Ƒuredi, S., Geserick, G., Henke, L., Hidding, M., Kargel, H.J., Lessig, R., Nagy, M., Pascali, W.L., Parson, W., Rolf, B., Schmitt, C., Szibor, R., Teifel-Greding, J., Krawczak, M. (2000) A new method for the evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males. *Forensic Science International, 114*(1), 31-43

Roostalu, U., Kutuev, I., Loogväli, E-L., Metspalu, E., Tambets, K., Reidla, M., Khusnutdinova, E.K., Usanga, E., Kivisild, T., Villems, R. (2007) Origin and Expansion of Haplogroup H, the Dominant Human Mitochondrial DNA Lineage in West Eurasia: The Near Eastern and Caucasian Perspective. *Molecular Biology and Evolution, 24*(1), 436-448. doi: 10.1093/molbev/msl173

Siegert, S., Roewer, L., Nothnagel, M. (2015) Shannon's equivocation for forensic Y-STR marker selection. *Forensic Science International Genetics, 16,* 216-225. doi: 10.1016/j.fsigen.2015.02.001

Walsh, B., Redd, A.J., Hammer, M.F. (2008) Joint match probabilities for Y chromosomal and autosomal markers. *Forensic Science International*, *174*(2-3), 234-238. doi: 10.1016/j.forsciint.2007.03.014

Wolańska-Nowak, P. (2000) Application of subpopulation theory to evaluation of DNA evidence. *Forensic Science International, 113*(1-3), 63-69

Wolańska-Nowak, P. (2007) On the subtle Southern Polish population subdivision. *Problems of Forensic Sciences, 69,* 265-279

Wolańska-Nowak, P., Branicki, W., Parys-Proszek, A., Kupiec, T. (2008) Examples of combining genetic evidence - Bayesian network approach. *Forensic Science International Genetics Supplement Series, 1,* 669-670

Wolańska-Nowak, P., Branicki, W., Parys-Proszek, A., Kupiec, T. (2009) A population data for 17 Y-chromosome STR loci in South Poland population sample – some DYS458.2 variants uncovered and sequenced. *Forensic Science International Genetics, 4*(1), 43-44. doi: org/10.1016/j.fsigen.2009.04.009

Wolańska-Nowak, P., Parys-Proszek, A., Sołtyszewski, I. (2014) Y chromosome haplogroups distribution in Poland and deduction on subpopulation origin of unknown male suspect – Bayesian Network approach. *Law and Forensic Science, 8*(2), 123-146

Zaykin, D.V., Pudovkin, A., Weir, B.S. (2008) Correlation-Based Inference for Linkage Disequilibrium with Multiple Alleles. *Genetics, 180*(1), 533-545. doi: 10.1534/genetics.108.089409