Int J Legal Med (2011) 125:519–525
DOI 10.1007/s00414-010-0514-9

ORIGINAL ARTICLE

Pedigree likelihood ratio for lineage markers

Jianye Ge · Arthur Eisenberg · Jiangwei Yan · Ranajit Chakraborty · Bruce Budowle

Received: 12 July 2010 / Accepted: 8 September 2010 / Published online: 21 September 2010

Abstract Lineage-based haplotype markers (e.g., Y chromosome STRs and mitochondrial DNA sequences) are important adjunct tools to the autosomal markers for kinship analysis and for specialized kinship applications such as database searching. Traditionally, the prosecution or kinship hypothesis considers the haplotypes in the same lineage and the probability of genotype data given the lineage hypothesis is simply set at 1 if the number of mismatched loci or nucleotides between the questioned person and the references is less than a predefined threshold. In this study, a kinship hypothesis based on a fixed relationship of the questioned person in the reference family is introduced. A graphical model is proposed to calculate the probability of the genotype data given the kinship hypothesis, which is the product of haplotype frequency of the founder in the pedigree and the transmission probability from the founder to all descendants. Proper mutation models are suggested for Y chromosome STRs and mitochondrial DNA sequence variants (i.e., SNPs) to calculate the transmission probability. The methods to infer the genotypes of the untyped individuals in the pedigree and the computational complexity of handling these untyped individuals are also addressed. Lastly, numerical examples of the applications are given to demonstrate the kinship hypothesis and the algorithms.

Keywords Pedigree likelihood ratio · Lineage markers · Y chromosome · Mitochondrial DNA · Mutations

J. Ge · A. Eisenberg · R. Chakraborty · B. Budowle
Department of Forensic and Investigative Genetics,
University of North Texas Health Science Center,
Ft Worth, TX 76107, USA

J. Ge (✉) · A. Eisenberg · R. Chakraborty · B. Budowle
Institute of Investigative Genetics, University of North Texas Health Science Center,
Ft Worth, TX 76107, USA
e-mail: Jianye.Ge@unthsc.edu

J. Yan
Beijing Institute of Genomics,
China Academy of Science,
Beijing, China

Introduction

The Y chromosome and the mitochondrial DNA (mtDNA) genome are inherited through the paternal and a maternal lineage, respectively. The genetic markers carried on these lineage chromosomes can be useful for kinship analysis to provide additional support or to exclude relationships particularly for complex kinships, for more distant relationships, and for database searches (with very limited family member candidates) [1–14]. The discrimination power of current multiplex Y-STR systems used for identification, although dependent on the size of the population sample, can reach 0.999 with 10 Y-STRs and 0.9999 with 16 Y-STRs [15, 16]. Haplotypes of the hypervariable regions I and II (HVI and HVII, respectively) of the mtDNA genome yield a similar discrimination power which is dependent on the length of sequence used and also on the size of the reference population database [13]. Generally, a DNA profile comprised of both mtDNA and Y haplotypes will be shared once in about every 1,000,000 individuals using current typing strategies. Although biologically independent of autosomal markers, Walsh et al. [17] and Budowle et al. [15] statistically tested for dependence and found that for the forensic markers on the Y chromosome and mtDNA haplotypes there were no detectable departures from independence with the forensically selected autosomal STR loci. Therefore, the likelihood ratios calculated for autosomal markers, Y-STR haplotypes, and mtDNA sequences can be directly multiplied together.

In kinship analysis cases, lineage markers often are used as an adjunct tool to the autosomal STR markers. Methods to calculate the likelihood ratio (LR) for the lineage markers have been discussed and have been applied in real cases [4, 7, 9–12, 14]; these approaches mainly focus on the manner to calculate haplotype frequency. The prosecution or kinship hypothesis usually considers the haplotypes in the same lineage instead of a fixed relationship of the questioned person in the reference family. Yet, little has been described on the probability of genotype data given the lineage or kinship hypothesis. In most cases, this probability is simply set at 1 if the number of mismatched loci or nucleotides between the questioned person and the references is less than a predefined threshold (e.g., less than two mismatched nucleotides infers that mtDNA sequence cannot be excluded as possibly being from the same lineage [18]). However, mutations and inconsistent haplotypes among the questioned person and the reference(s) should be taken into account in the likelihood calculation. Rolf et al. [6] suggested a method for calculating the probability of segregating the haplotypes in multiple generations. However, they did not include a mutation model to deal with multiple step mutations and complex mutation events. For example, two single locus Y-STR haplotypes {11} and {10} are more likely to be in the same lineage explained by mutation than haplotypes {11} and {8.2}, although both scenarios vary from each other at one mismatched locus. In addtion, the details of the pedigree structure were not considered in [6]. Instead, only the number of transmission events or generations between individuals was considered.

Ge et al. [19] described the methods to calculate the lineage-based LR with multiple references, in which the same mutation model for the autosomal STRs was suggested for Y-STRs, only the closest available relatives were considered and the haplotype was directly compared to the haplotype of the questioned person. However, given the hypothesis that the questioned person has a fixed relationship with the references, the LRs of the lineage markers should be calculated in a pedigree in a similar manner to that used for autosomal markers [19–22]. In this study, we provide more sophisticated algorithms to formalize the likelihood calculations of both Y-STR and mtDNA haplotypes, in which the likelihood is calculated in a pedigree fashion and the kinship hypothesis is based on a fixed relationship rather than solely on lineage. The computational complexity is analyzed and methods to reduce the complexity are discussed. Preliminary mutation models for the mtDNA are introduced to calculate the transmission probability between mtDNA haplotypes. Lastly, numerical examples of the applications are given to explain the algorithms.

Methods

The general principle of the LR calculations of lineage markers is the same as that for autosomal markers, which compares the probabilities of genotype data, including both the questioned person and the reference family member(s), given the hypothesis of kinship H_k: the questioned person (Q) is a specific member of the family (F), or the hypothesis of non-kinship or unrelated H_{nk}: Q is unrelated to F. The LR is represented by the following expression:

$$LR = \frac{\Pr(Q, F|H_k)}{\Pr(Q, F|H_{nk})} = \frac{\Pr(Q, F|H_k)}{\Pr(Q)^* \Pr(F)} \quad (1)$$

$\Pr(Q)$ is the haplotype frequency of Q, namely, the Y-STR or mtDNA haplotype frequency in a database with selected loci or ranges. $\Pr(F)$ is the pedigree probability of F. $\Pr(Q, F\backslash H_k)$ is the pedigree probability of both Q and F given H_k. Note that H_k is different from the hypothesis "Q belongs to the lineage of F", in which the position of Q in F is not determined.

Although the mutation mechanisms of Y-STRs and mtDNA haplotypes are different, the principles of the pedigree likelihood are the same for Y-STR and mtDNA haplotypes. In the following, the details are introduced of the algorithms to obtain the pedigree likelihoods using Y-STR haplotypes as a template. Pedigree likelihood of the mtDNA can be calculated by the same algorithm but with a different mutation model. Main estimates of the lineage haplotype frequency will be presented.

Pedigree likelihood for Y-STR

The pedigree likelihood based on the Y-STR haplotypes is actually the product of the haplotype frequency of the founder (i.e., the person without antecedent relatives in the pedigree) and the transmission probability from the founder to all descendants, which is dependent on the mutation steps between each father–son transmission event. The whole pedigree is regarded as a directed acycling graph [23], in which vertices are the family members and the directed edges represent the genetic transmission from fathers to sons (see Fig. 1 as an example).

Suppose there are N father–son pairs (i.e., FS_1, FS_2, ..., FS_N) in the pedigree and the genotype of each individual in the pedigree is known, the cumulative transmission probability of the pedigree, TP, is the cumulative product of the transmission probability of each father–son pair at each locus (Eq. 2), assuming the mutation events at each locus are independent:

$$TP = \prod_{i=1}^{N} \prod_{k=0}^{K} TP(FS_{ik}) \quad (2)$$

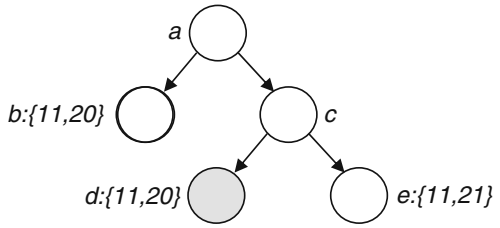

Fig. 1 A two-locus Y-STR pedigree including individuals from *a* to e. The vertices represent individuals and the directed edges are father–son transmissions. *b*, *d*, and *e* have Y-STR genotypes available; *a* and *c* are untyped. Individual *d* is the questioned person with genotype {11,20}, and the missing person d has a brother *e* and an uncle *b* as the family references

where $TP(FS_{ik})$ is the transmission probability of *i*th father–son pair at *k*th locus, which depends on the mutation rate at this locus (μ) and the mutation steps (x) between the genotypes of the father–son pair. Since the majority of Y-STR mutations are slippage mutations, the same mechanism as the autosomal STRs [24], the same mutation model for autosomal STRs, two-phase model [25–27], can be used for Y-STRs. This mutation model appears to be the most suitable for the Y and autosomal STRs [27]. The transmission probability between two alleles can be presented as Eq. (3)

$$TP(FS_{ik}) = \begin{cases} 1-\mu & x=0 \\ 1/2\mu\alpha(1-\alpha)^{x-1} & x=1,2,...,\infty \end{cases} \quad (3)$$

in which α is the probability of being a one-step mutation. Equal probabilities for gaining or losing repeats are assumed. According to [28], about 95% of mutations result in one-step differences, similar to that of the autosomal STRs, hence α is set at 0.95 (although any reasonable value can be set). The mutation rates vary according to the markers [24].

The mutation step (x) between two integer alleles or two fractional alleles with the same fraction is defined as the absolute numerical difference between the alleles. For example, the mutation step is 2 between {10} and {12} or {10.2} and {12.2}. The mechanism of mutations between integer and fractional (e.g., {10}à{10.2}) STR alleles or fractional alleles with different fractions (e.g., {10.1}à{10.2}) is different from slippage-based mutation. The probability of a partial repeat mutation is much lower than that of the average STR slippage mutation rates (i.e., 2×10^{-3}) and likely higher than that of the SNP mutation rates (i.e., 10^{-8}). Because there are no data on partial repeat mutations, we arbitrarily set the probability at 10^{-5}; further investigations are needed to establish a more accurate probability. For mutations increasing the number of alleles (e.g., {10}à{10, 11}), a duplication or insertion may explain the event. The slippage mutation and the duplication event are assumed to be independent. We also arbitrarily set the probability of insertion events at 10^{-5}. For mutations which lose some alleles (e.g., {10, 11}à{10}), we assume only slippage mutation as it is more likely.

The pedigree likelihood is the product of the haplotype frequency of the founder, Pr(founder), and the transmission probability within the pedigree (Eq. 4). The computational complexity is on the order of $N\times K$ for a pedigree with all individuals having unambiguous genotypes.

$$L = \Pr(\text{founder}) \times TP \quad (4)$$

In situations where genotypes are not available at some loci for some individuals in the pedigree, the genotypes are inferred. The likelihood of the pedigree with untyped individuals is the sum of the likelihood of all possible subpedigrees with all individual's genotypes typed or inferred. Note that different subpedigrees may have different founder profiles (see Appendix for an example).

The genotype assignment with minimum mutation steps to explain the pedigree is preferred. Assume that the genotypes of the untyped individuals can only be composed of one or more of the observed alleles in the pedigree or the alleles between these observed alleles. The number of alleles should be the same as those of the majority of the observed genotypes in the pedigree. For example, in Fig. 1, the individual *a* can only be {11} for the first locus. The chance that *a* has a genotype {12} or {10, 11} at the first locus is remote and thus can be ignored in the likelihood calculation.

Suppose all individuals have the same number of observed alleles (m; e.g., m=2 means that all individuals have two alleles), there are X number of possible alleles ($X\geq m$), including both observed alleles and the alleles between these observed alleles, in the whole pedigree (e.g., X=4 if three individuals in a pedigree have alleles {8,9}, {9,11}, and {8,11} at a locus, respectively) and Y number of untyped individuals at a specific locus the number of possible subpedigrees (NP) at this locus is

$$\text{NP} = \binom{X}{m}^{Y} \quad (5)$$

Using the second locus in Fig. 1 as an example, m=1 (e.g., all individuals in the pedigree have only one allele at the second locus), X=2 (e.g., all possible alleles in the pedigree at the second locus are {20} and {21}), and Y=2 (e.g., there are two untyped individuals a and c); the number of subpedigrees is 2^2=4. The computational complexity is exponential mainly depends on the Y. Fortunately, in most real cases, both X and Y are less than 3 or 4, which is acceptable and the computation can be completed in a reasonable time (i.e., less than 1 s with the current computation ability). For any pedigrees or subpedigrees with all the genotypes determined, the computational complexity is linear to the number of transmission with the pedigrees (i.e., the number of edges in the acycling graph).

Pedigree likelihood for mtDNA

The algorithm for pedigree likelihood calculation based on mtDNA haplotypes is the same as that for Y-STR haplotypes, except that the mtDNA and Y-STR loci have different mechanisms and rates of mutation. mtDNA mutations tend to be due to base substitution (i.e., transition and transversion), deletion, or insertion. The independence of mutation events between bases is assumed.

The mutation rates for the mtDNA genome are known to be higher than those of nuclear DNA [13, 29]. Many nucleotide substitution models [30] have been proposed, such as Kimura model, Tamura model, HKY model, and unrestricted model, etc. Unfortunately, there has been no systematic study for which model is the best fit for human mtDNA sequences. The Kimura model [31] was used (simply as a starting point), which assigns different substitution rates to transitions and transversions, but the substitution rates of two different transitions, as well as two different transversions, are assumed equal. The mutation rates are assumed to be even for all nucleotides, although relatively high mutation rates were observed at some hotspots [32, 33]. More sophisticated models can be implemented with further data. According to the summarized data in Howell et al. [34], 15 mutations in 1,246 transmission events were observed in HVI and HVII, which leads to an average mutation rate 2.0×10^{-5} across HVI and HVII (i.e., 15/(1,246×610)). Table 1 summarizes the mutation rates of HVI and HVII for the Kimura model [35, 36].

To accommodate heteroplasmy, the mtDNA pedigree likelihood of a base position is the sum of likelihood of subpedigrees with a nucleotide fixed for each individual. For untyped individuals at a base position, the genotype can be all four possible nucleotides (although transitions tend to be favored over transversions at many sites). Suppose there are m mtDNA typed individuals in a pedigree with T_1, T_2, ..., T_m number of nucleotides, respectively, the number of possible subpedigrees with fixed nucleotides for all typed individuals is the product of the number of nucleotides of individuals (i.e., $\prod_{k=1}^{m} T_k$). To save the computational time, the transmission probability may be directly set at 1 for bases at which all individuals share at least one nucleotide. Subpedigrees with transversion mutation may be ignored if there are other subpedigrees which only contain transition mutations or no mutation.

Table 1 Human mitochondrial DNA nucleotides mutation rates ($\times10^{-6}$) assuming Kimura model

Nucleotides	Mutations					
	A	T	C	G	Insertion	Deletion
A		0.2	0.2	3.6	0.5	0.5
T	0.2		3.8	0.2	0.5	0.5
C	0.2	3.6		0.2	0.5	0.5
G	3.6	0.2	0.2		0.5	0.5

Haplotype frequency

There are many reasonable suggestions to estimate haplotype frequency of the lineage markers. Let n be the database size and x be the number of observations of the lineage profile. The suggested haplotype frequency estimates mainly include the followings.

1 $p=\frac{x}{n}$
2 $p=\frac{x+2}{n+2}$ [37, 38]
3 $p=1-\alpha^{1/n}$, where α is (1−confidence interval), usually 0.05 [38, 39]
4 The "surveying" methods [40, 41]
5 $p=\frac{x}{n}+\theta\left(1-\frac{x}{n}\right)$ [15, 16], where θ is the measure of population substructure
6 Upper bound of 95% binomial confidence interval [16]; the Clopper and Pearson's binomial confidence interval is suggested

Discussion

This study provides formalized methods for calculating the pedigree likelihood based on lineage markers, in which the hypothesis of kinship is defined by a fixed relationship with the reference family. The traditional direct comparison between the haplotypes has been the hypothesis that the reference(s) and the questioned person are in the same lineage. These two hypotheses are conceptually distinguishable, although the likelihoods given them may not be significantly far apart. The pedigree likelihood usually gives more conservative results compared with the direct comparison method.

There are several ways to implement the process and reduce computation time. For simple cases with all haplotypes "matching," the transmission probability within a pedigree may be set at 1. For example, suppose the haplotypes contain 17 Y-STR loci with mutation rates of 2×10^{-3} across all loci, the transmission probability is about 0.968 from father to son, or 0.938 from grandfather to grandson (i.e., two transmission events). Similar values can be obtained for mtDNA. However, when the possibility of mutations present in the pedigree, the formalized algorithms need to accommodate various mutation types and mutation steps. Another way to reduce computational time is to search those subpedigrees which have the fewest mutation steps in all subpedigrees and only calculate the likelihoods of these minimum subpedi-

grees. This shortcut will not substantially change the likelihood of the pedigree since for STRs one more step mutation is much less likely. For mtDNA, the subpedigrees with minimum number of the transversions, deletions, and insertions are the main contributors to the total pedigree likelihood.

A recent study on mtDNA confirmed [42] what has been an assumed basis for forensic interpretations that heteroplasmy is ubiquitous in human cells. This observation supports that individual humans have multiple haplotypes instead of one. Assuming the mutation at each base is independent from other bases, for an mtDNA sequence range with 610 nucleotides, the transmission probability between two identical mtDNA sequences is about 0.9879. Starting from a single zygotic cell with 1,000 homogeneous mitochondrial genomes, after 20 cell divisions any sampled cell (e.g., white blood cell) has about a 99.99% chance of having at least two mutations on at least one mitochondrial DNA sequence. The interpretation rule of two or more mismatched nucleotides for excluding two samples (based on HVI and HVII sequences) as originating from the same source was appropriate for the sequencing technology applied [18]. Although, there is a small probability of a false exclusion, two differences residing on a single molecule in a pool of mitochondrial genomes would not be detectable. However, with a high resolution technology (e.g., the technology used in [42]), the same rule [18] would not strictly apply. At present, the mtDNA mutation model assumes uniform mutation rates across all bases of the mtDNA sequence. However, mutation hotspots do exist on mtDNA [32, 33]. Further studies are required to develop more sophisticated mutation models for mtDNA, which may allow base-specific mutation rates.

These algorithms will be implemented into the software program MPKin [19].

Acknowledgements This work was partially supported by USA National Institute of Justice (2009-DN-BX-K188) and National Natural Science Foundation of China (No. 81072511).

Appendix—numerical examples

Case 1: Y-STR

Figure 1 is used as an example to explain the Y-STR haplotype-based likelihood ratio calculation. Let H_k be the hypothesis that d is the brother of e and nephew of b, and H_{nk} be the hypothesis that d is unrelated to the family of b and e. The likelihood of the family given H_k is the sum of the following four likelihoods, each of which is the likelihood of a subpedigree with untyped data determined.

1. $a=\{11,20\}$ and $c=\{11,20\}$
   $L_1=\Pr(\{11,20\})\times\Pr(11\text{à}11)^4\times\Pr(20\text{à}20)^3\times\Pr(20\text{à}21)$
2. $a=\{11,20\}$ and $c=\{11,21\}$
   $L_2=\Pr(\{11,20\})\times\Pr(11\text{à}11)^4\times\Pr(20\text{à}20)\times\Pr(20\text{à}21)\times\Pr(21\text{à}20)\times\Pr(21\text{à}21)$
3. $a=\{11,21\}$ and $c=\{11,20\}$
   $L_3=\Pr(\{11,21\})\times\Pr(11\text{à}11)^4\times\Pr(20\text{à}20)\times\Pr(20\text{à}21)\times\Pr(21\text{à}20)^2$
4. $a=\{11,21\}$ and $c=\{11,21\}$
   $L_4=\Pr(\{11,21\})\times\Pr(11\text{à}11)^4\times\Pr(21\text{à}20)^2\times\Pr(21\text{à}21)^2$

The likelihood of the family given H_{nk} is the product of the haplotype frequency of d, $\Pr(\{11,20\})$, and the likelihood of the family only containing b and e, which is the sum of the following four subpedigrees with inferred genotypes for a and c.

1. $a=\{11,20\}$ and $c=\{11,20\}$
   $L_1=\Pr(\{11,20\})\times\Pr(\{11,20\})\times\Pr(11\text{à}11)^4\times\Pr(20\text{à}20)^2\times\Pr(20\text{à}21)$
2. $a=\{11,20\}$ and $c=\{11,21\}$
   $L_2=\Pr(\{11,20\})\times\Pr(\{11,20\})\times\Pr(11\text{à}11)^4\times\Pr(20\text{à}20)\times\Pr(20\text{à}21)\times\Pr(21\text{à}21)$
3. $a=\{11,21\}$ and $c=\{11,20\}$
   $L_3=\Pr(\{11,20\})\times\Pr(\{11,21\})\times\Pr(11\text{à}11)^4\times\Pr(20\text{à}21)\times\Pr(21\text{à}20)^2$
4. $a=\{11,21\}$ and c$=\{11,21\}$
   $L_4=\Pr(\{11,20\})\times\Pr(\{11,21\})\times\Pr(11\text{à}11)^4\times\Pr(21\text{à}20)\times\Pr(21\text{à}21)^2$

Let the mutation rate $\mu=0.002$, one-step mutation proportion be 0.95, haplotype frequency of {11,20} be 0.2, and haplotype frequency of {11,21} be 0.3. $\Pr(11\text{à}11)=\Pr(20\text{à}20)=\Pr(21\text{à}21)=(1-\mu)$ and $\Pr(21\text{à}20)=\Pr(20\text{à}21)=1/2\mu\alpha$ in terms of Eq. (3). The likelihood ratio is 1.43. To reduce computational time, L_2, L_3, and L_4 in H_k and L_3 in H_{nk}, which have higher number of mutation steps than others, can be ignored, and the likelihood is still about 1.43.

Case 2: mtDNA

Figure 2 is used as an example to explain the mtDNA-based likelihood ratio calculation. Let H_k be the hypothesis that c is the sister of b, and H_{nk} be the hypothesis that c is unrelated to the family of b. The likelihood of the family

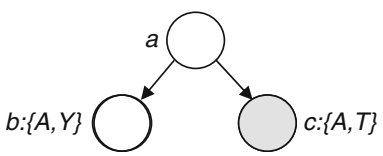

Fig. 2 A two-base mtDNA haplotype pedigree including individuals from a to c. Individual c is the questioned person with mtDNA sequence {A,T} and the missing person c has a sister b with a mtDNA sequence {A,Y}. Y presents a heteroplasmic base, which could be nucleotides C or T. The mtDNA of the mother a is not available

given H_k is the sum of the following four likelihoods, each of which is the likelihood of a subpedigree with untyped data determined.

1. a={A, T} and b={A,T}
   $L_1 = \Pr(\{A,T\}) \times \Pr(A \grave{a} A)^2 \times \Pr(T \grave{a} T)^2$
2. a={A, T} and b={A,C}
   $L_2 = \Pr(\{A,T\}) \times \Pr(A \grave{a} A)^2 \times \Pr(T \grave{a} T) \times \Pr(T \grave{a} C)$
3. a={A, C} and b={A,T}
   $L_3 = \Pr(\{A,C\}) \times \Pr(A \grave{a} A)^2 \times \Pr(C \grave{a} T)^2$
4. a={A, C} and b={A,C}
   $L_4 = \Pr(\{A,C\}) \times \Pr(A \grave{a} A)^2 \times \Pr(C \grave{a} T) \times \Pr(C \grave{a} C)$

The likelihood of the family given H_{nk} is the product of the haplotype frequency of c, Pr({A,T}), and the likelihood of the family only containing b, which is the sum of frequencies of two haplotypes, Pr({A,T}) and Pr({A,C}). Let the frequency of {A,T} and {A,C} be 0.2 and 0.3, respectively. Using the mutation rates in Table 1, the likelihood ratio is about 2.0.

References

1. Chakraborty R (1985) Paternity testing with genetic markers: are Y-linked genes more efficient than autosomal ones? Am J Med Genet 21:297–305
2. Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. Int J Legal Med 110:118–124
3. Corach D, Risso LF, Marino F, Penacino G, Sala A (2001) Routine Y-STR typing in forensic casework. Forensic Sci Int 118:131–135
4. Roewer L, Kayser M, de Knijff P, Anslinger K, Betz A, Caglia A et al (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 Sci Int 114:31–43
5. Kayser M, Sajantila A (2001) Mutations at Y-STR loci: implications for paternity testing and forensic analysis. Forensic Sci Int 118:116–121
6. Rolf B, Keil W, Brinkmann B, Roewer L, Fimmers R (2001) Paternity testing using Y-STR haplotypes: assigning a probability for paternity in cases of mutations. Int J Legal Med 115:12–15
7. Gill P, Brenner C, Brinkmann B, Budowle B, Carracedo A, Jobling MA et al (2001) DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs. Forensic Sci Int 124:5–10
8. Budowle B, Sinha SK, Lee HS, Chakraborty R (2003) Utility of Y-chromosome STR haplotypes in forensic applications. Forensic Sci Rev 15:153–164
9. Gusmao L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, Morling N, Prinz M, Roewer L, Tyler-Smith C, Schneider PM (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. Forensic Sci Int 157:187–197
10. Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ (1996) Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. Nature Genet 12:417–420
11. Carracedo A, Bär W, Lincoln P, Mayr W, Morling N, Olaisen B, Schneider P, Budowle B, Brinkmann B, Gill P, Holland M, Tully G, Wilson M (2000) DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. Forensic Sci Int 110:79–85
12. Tully GW, Bär BB, Carracedo A, Gill P, Morling N, Parson W, Schneider P (2001) Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. Forensic Sci Int 124:83–91
13. Budowle B, Allard MW, Wilson MR, Chakraborty R (2003) Forensic and mitochondrial DNA: application, debates, and foundations. Annu Rev Genom Hum Genet 4:119–141
14. Just RS, Loreille OM, Molto JE, Merriwether DA, Woodward SR, Matheson C, Creed J, McGrath SE, Sturk-Andreaggi K, Coble MD, Irwin JA, Ruffman A, Parr RL. Titanic's unknown child: The critical role of the mitochondrial DNA coding region in a re-identification effort, Forensic Science International: Genetics, In Press, Corrected Proof, Available online 2 April 2010, ISSN 1872–4973. doi: 10.1016/j.fsigen.2010.01.012
15. Budowle B, Ge J, Aranda X, Planz J, Eisenberg A, Chakraborty R (2009) Texas population substructure and its impact on estimating the rarity of Y STR haplotypes from DNA evidence. J Forensic Sci 54(5):1016–1021
16. Ge J, Budowle B, Planz JV, Eisenberg AJ, Ballantyne J, Chakraborty R. US forensic Y Chromosome Short Tandem Repeats Database, Legal Medicine, In Press, Corrected Proof, Available online 3 September 2010, ISSN 1344–6223. doi:10.1016/j.legalmed.2010.07.006
17. Walsh B, Redd A, Hammer M (2008) Joint match probabilities for Y chromosomal and autosomal markers. Forensic Sci Int 174 (2):234–238
18. Budowle B, DiZinno J, Wilson M (1999) Interpretation guidelines for mitochondrial DNA sequencing. 10th International Symposium of Human Identification
19. Ge J, Budowle B, Chakraborty R (2010) DNA identification by pedigree likelihood ratio accommodating population substructure and mutations, Investigative Genetics, In press
20. Brenner CH (1997) Symbolic kinship program. Genetics 145:535–542
21. Hepler AB, Weir BS (2008) Object-oriented Bayesian networks for paternity cases with allelic dependencies. Forensic Sci Int Genet 2:166–175
22. Egeland T, Mostad PF, Mevag B, Stenersen M (2000) Beyond traditional paternity and identification cases: selecting the most probable pedigree. Forensic Sci Int 110:47–59
23. Christofides N (1975) Graph theory: an algorithmic approach. Academic Press, New York, pp 170–174
24. Ge J, Budowle B, Aranda XG, Planz JV, Eisenberg AJ, Chakraborty R (2009) Mutation rates at Y chromosome short tandem repeats in Texas populations. Forensic Sci Int Genet 3:179–184
25. Chakraborty R, Stivers DN, Zhong Y (1996) Estimation of mutation rates from parentage exclusion data: applications to STR and VNTR loci. Mut Res 354:41–48
26. DiRienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci U S A 91:3166–3170
27. Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetic analysis. Mol Ecol 11:1591–1604
28. AABB annual report 2008 http://www.aabb.org/sa/facilities/Documents/rtannrpt08.pdf
29. Sigurðardóttir S, Helgason A, Gulcher JR, Stefansson K, Donnelly P (2000) The mutation rate in the human mtDNA control region. Am J Hum Genet 66(5):1599–1609

30. Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, Oxford, p 35
31. Kimura K (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16(2):111–120
32. Galtier N, Enard D, Radondy Y, Bazin E, Belkhir K (2006) Mutation hot spots in mammalian mitochondrial DNA. Genome Res 16:215–222
33. Stoneking M (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. Am J Hum Genet 67:1029–1032
34. Howell N, Christy Bogolin Smejkal DA, Mackey PF Chinnery, Turnbull DM, Herrnstadt C (2003) The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. Am J Hum Genet 72:659–670
35. Meyer S, Weiss G, von Haeseler A (1999) Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics 152:1103–1110
36. Budowle B, Wilson MR, DiZinno JA, Stauffer C, Fasano MA, Holland MM, Monson KL (1999) Mitochondrial DNA regions HVI and HVII population data. Forensic Sci Int 103(1):23–35
37. Balding DJ, Nichols RA (1994) DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int 64(2–3):125–140
38. Tully G, Bär W, Brinkmann B, Carracedo A, Gill P, Morling N, Parson W, Schneider P (2001) Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. Forensic Sci Int 124(1):83–91
39. Holland MM, Parsons TJ (1999) Mitochondrial DNA sequence analysis—validation and use for forensic casework. Forensic Sci Rev 11:21–50
40. Roewer L, Kayser M, de Knijff P, Anslinger K, Betz A, Caglia A, Corach D, Füredi S, Henke L, Hidding M, Kärgel HJ, Lessig R, Nagy M, Pascali VL, 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 Sci Int 114:31–43
41. Krawczak M (2001) Forensic evaluation of Y-STR haplotype matches: a comment. Forensic Sci Int 118:114–115
42. He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, Diaz LA Jr, Kinzler KW, Vogelstein B, Papadopoulos N (2010) Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. Nature 464:610–614