Review

Assessing the Forensic Value of DNA Evidence from Y Chromosomes and Mitogenomes

Mikkel M. Andersen [1,2,]* and David J. Balding [3,4]

1. Department of Mathematical Sciences, Aalborg University, 9220 Aalborg, Denmark
2. Section of Forensic Genetics, Department of Forensic Medicine, University of Copenhagen, 1165 Copenhagen, Denmark
3. Melbourne Integrative Genomics, University of Melbourne, Melbourne 3010, Australia; dbalding@unimelb.edu.au
4. Genetics Institute, University College London, London WC1E 6BT, UK
* Correspondence: mikl@math.aau.dk

Abstract: Y chromosome and mitochondrial DNA profiles have been used as evidence in courts for decades, yet the problem of evaluating the weight of evidence has not been adequately resolved. Both are lineage markers (inherited from just one parent), which presents different interpretation challenges compared with standard autosomal DNA profiles (inherited from both parents). We review approaches to the evaluation of lineage marker profiles for forensic identification, focussing on the key roles of profile mutation rate and relatedness (extending beyond known relatives). Higher mutation rates imply fewer individuals matching the profile of an alleged contributor, but they will be more closely related. This makes it challenging to evaluate the possibility that one of these matching individuals could be the true source, because relatives may be plausible alternative contributors, and may not be well mixed in the population. These issues reduce the usefulness of profile databases drawn from a broad population: larger populations can have a lower profile relative frequency because of lower relatedness with the alleged contributor. Many evaluation methods do not adequately take account of distant relatedness, but its effects have become more pronounced with the latest generation of high-mutation-rate Y profiles.

1. Weight of Evidence for Lineage Marker Profiles

Standard DNA profiles use autosomal DNA, inherited from both parents. We focus here on DNA profiles obtained from the Y chromosome and the mitochondrial genome (mitogenome/mtDNA), which are inherited only from the father and from the mother, respectively. Because of this uniparental inheritance over generations, these DNA profiles are called lineage markers. We outline the forensic value of lineage markers in general, give a brief, historical review and critique of evaluation methods and make recommendations for improved practice. A key message is that the high mutation rates of the latest generation of Y chromosome short tandem repeat (STR) profiles have effects that exaggerate the deficiencies of previous methods of analysis, but understanding these effects highlights ways forward. The mutation rate of even the whole mitogenome is lower, but insights from the high-mutation-rate setting are also informative for evaluating and communicating the weight of evidence.

We focus on the simplest scenario in which there is a good-quality, single-contributor DNA profile (either Y or mitogenome) obtained from an evidence sample, and a matching

reference profile from a known individual Q who is alleged to be the source of the evidence sample (sometimes Q is called the person of interest, PoI). Hypotheses of interest are:

$$H_Q : \text{the evidence profile came from } Q,$$
$$H_X : \text{the evidence profile came from } X,$$

where X is an alternative to Q as the source of the DNA whose profile is not available. The likelihood ratio (LR) comparing the strength of the DNA profile evidence for H_Q relative to H_X is then

$$\text{LR}(X,Q) = \frac{\text{P(profile evidence} \mid H_Q)}{\text{P(profile evidence} \mid H_X)} = \frac{1}{\text{P}(X \text{ has profile } q)}, \tag{1}$$

where q is the profile of Q. We omit background information and other evidence from the notation; see [1] for a discussion. The denominator of (1) is a probability for the unknown profile of X, given the observed q and possibly a database of profiles.

1.1. The Effect of Relatedness and Mutation Rate on the LR

For lineage markers, the relatedness of two individuals is fully captured by a single number, G, of generations (or germline transfers, or meioses) that separate X and Q, following either female-only or male-only ancestors. Degree-1 ($G = 1$) relative pairs are parent/offspring, $G = 2$ for siblings and grandparent/grandchild, while $G = 3$ for avuncular relationships such as aunt/niece as well as great-grandparent/great-grandchild. As G increases, the relatedness becomes less likely to be known, but there is a female-lineage G and a male-lineage G for all pairs of individuals, "unrelated" only means that G is large and/or unknown.

Given $G = g$ for X and Q, we can approximate (1) by

$$\text{LR}(X,Q) = \frac{1}{(1-\mu)^g} \approx e^{g\mu} \tag{2}$$

where μ is the profile mutation rate, which is the probability that parent and child have non-matching profiles. For Y-STR loci with mutation rates $\mu_l, l = 1, \ldots, L$, and assuming independence of mutation events across loci, we have

$$\mu = 1 - \prod_{l=1}^{L}(1-\mu_l) < \sum_{l=1}^{L}\mu_l . \tag{3}$$

For the mitogenome, the lower mutation rate, larger number of sites, and possible germline selection events as part of the mitochondrial bottleneck [2] make it impractical to use (3) at individual sites, but it can be employed over genome regions [3,4].

The LR Formula (2) is not exact for at least two reasons. Firstly, it assumes that μ is constant over generations and independent of the current profile state. In fact, accurate estimates are hard to obtain but the STR mutation rate is likely to depend on the allele sequence (including its length and presence of a partial repeat) [5]. However, these effects are relatively unimportant for the total mutation rate over many loci. Secondly, it is based on assuming no mutations in the lineage path connecting X and Q, whereas a match can also arise following an even number of mutations between X with Q such that the effect of each mutation is reversed by another mutation. However, profiles consisting of many loci and with multiple possible mutation events at each locus are very unlikely to match if there is any mutation between X and Q [6,7].

For G unknown, in place of (2), we have (see also [8]):

$$\text{LR}(X,Q) = \frac{1}{\sum_{g=1}^{\infty}(1-\mu)^g \text{P}(G=g)} \qquad \text{where} \quad \sum_{g=1}^{\infty} \text{P}(G=g) = 1. \tag{4}$$

Equation (4) requires a probability distribution for *G*, which can be informed by population genetic models and by the available information about alternative possible sources of the DNA profile. We introduced *X* as a specific alternative individual but, in practice, there are usually many alternative sources of the DNA. In either case, $P(G = g)$ can be interpreted as the probability that the unknown alternative source of the DNA is a degree-*g* relative of *Q*. If, for example, it is known that all of the degree-*g* relatives of *Q* are excluded as possible sources, then $P(G = g) = 0$ in (4), and the other values must be assigned to maintain $\sum_{g=1}^{\infty} P(G = g) = 1$.

Although we can rarely compute it accurately, (4) tells us how to evaluate a lineage marker profile match: we need to assess, given the known circumstances, the probability that an alternative source of the DNA has relatedness *g* with the alleged source *Q*, and weight this probability by $(1-\mu)^g$, which means that individuals with $g \gg 1/\mu$ contribute little to the LR.

Most presentations of lineage marker DNA profile evidence mention something like "maternally-related individuals will share the same mitochondrial DNA profile" and then proceed to assess weight of evidence assuming that *X* and *Q* are unrelated. This practice is potentially misleading, because we are all maternally related and we are also all paternally related, what is important is the degree *g* of the relatedness (or that $g \gg 1/\mu$).

1.2. The Role of Databases in Evidence Evaluation

Most methods reviewed below make some use of a database of profiles to provide information about population profile frequencies. When multiple databases are available, the one with ancestry closest to that of *Q* should usually be chosen [9,10], unless the database size is very small and there is an alternative database of larger size drawn from a population with similar ancestry.

Databases are usually not random samples [7,11], and not drawn from the population relevant to a specific case. For lineage markers, the important role of relatedness implies that the way the database is sampled can have a big impact on inferences. On the one hand, databases sampling may be biased towards including more sets of related individuals, for example because relatives of a suspected contributor may also be suspects. On the other hand, some databases implement a policy of excluding close relatives.

For Y and the mitogenome, the most important international databases, both highly respected for their data quality [12], are available online at www.YHRD.org (accessed on 1 August 2021) [13] and www.EMPOP.online (accessed on 1 August 2021) [14,15]. In July 2021, EMPOP had 48,572 mtDNA sequences (of which 4289 cover the entire mitogenome, the rest span some or all of the control region, which is the most variable part of the mitogenome [15]). YHRD had 337,449 minimal (8 STR loci) profiles, of which 97,087 were 27-locus Yfiler Plus profiles. Both databases contain samples from multiple worldwide populations—in some cases, including several subpopulations.

1.3. The Probability of Matching Profiles

Equation (4) has profound implications for the role of databases in computing the LR. For any of the lineage markers in Table 1, the space of possible profiles is so vast that it is extremely unlikely that two distinct lineages will generate the same (complete) profile by neutral mutations. Although matching Y-STR profiles have been reported between males with different single-nucleotide polymorphism (SNP) subhaplotypes, the matching males were in the same geographic region and relatedness is the most likely explanation [16,17], noting that the Y chromosome accumulates several single-nucleotide mutations per generation [18]. Unless *μ* is very low, the possibility of a match between distantly-related individuals is negligible relative to the much more likely event of a match between pairs of individuals who are related closely enough to generate the match, but distantly enough that the relatedness is not recognised. Relatedness also affects match probabilities for autosomal profiles, but recombination weakens its effects, except in the case of monozygotic twins [19].

Table 1. Estimates of μ for lineage marker profiles. The per-year mitogenome mutation rates in [3,4] have been multiplied by a generation time of 25 years.

Type	Markers	Profile Mutation Rate μ [% Per Generation]	Reference
Ychr	PowerPlex Y (12 loci)	2.5%	[7,20]
Ychr	Yfiler (17 loci)	4.4%	[7,20]
Ychr	PowerPlex Y23 (23 loci)	8.3%	[7,20]
Ychr	Yfiler Plus (27 loci)	13.5%	[7,20]
Mitogenome	16,070 sites	0.3–1.9%	[3,21]
Mitogenome	16,494 sites	0.4–2.3%	[4,21]

For the mitogenome, μ is estimated to be approximately 1 per 70 generations (Table 1). Thus, alternative sources X will almost certainly be female-line relatives of Q with G up to a few hundred [21]. This is greatly beyond the known relatives of Q, but is still closer than a random pair of individuals in a large population, who are typically separated by thousands of generations [21]. For older Y profile kits, μ is a small multiple of the mitogenome value; but for more recent Y profile kits, μ can be an order of magnitude higher, approximately 1 per 7.5 generations for Yfiler Plus profiles (Table 1). In that case, profile matches occur between pairs of individuals separated by at most a few tens of generations [7].

Therefore, the majority of profiles in broadly-defined databases such as YHRD or EMPOP are drawn from individuals that are too distantly related to Q to be relevant to a particular case. The frequency of q can depend sensitively on the choice of population, according to its average relatedness with Q. Alternative sources of the DNA may be concentrated in the same subpopulation as Q, defined by geographical origin and/or social factors such as ethnicity and religion. A relevant consideration is that if Q is in fact not the source of the DNA, the false allegation may in part be due to Q resembling the true source in some characteristics such as appearance, place of residence, or social background, which tends to increase $P(G = g)$ for small values of g.

2. Review of Evaluation Methods

Most methods have not addressed the fundamental effects of relatedness, mutation rate and database sampling frame discussed above. We will review methods assuming, as most authors have done, that the available database is appropriate, and return to these issues in the Discussion.

2.1. Adjusted Database Counts

The denominator of the LR (1), is often assumed to equal the *match probability*, π_q, the relative frequency of q in a population of alternative sources of the DNA. Many approaches aim to estimate π_q based on the count k_q of q in a database of size n. The database relative frequency k_q/n has long been recognised as an unsatisfactory estimator of π_q, because we often have $k_q = 0$ and yet Q has the profile and so $\pi_q > 0$.

Further, the requirement to avoid overstatement of evidence converts here to preferring a "conservative" over-estimate of π_q rather than an under-estimate. Excessive conservativeness wastes information, and our goal should be to evaluate the evidence as accurately as possible, while taking care to guard against being anti-conservative. However, the latter requirement is difficult to satisfy, and there is no agreement about to what extent we should seek to eliminate any risk of being anti-conservative.

Various adjustments to k_q and n have been proposed to avoid zero estimates and introduce an upward bias. We will discuss them in order of increasing conservativeness as measured by $\hat{\pi}_q$ when $k_q = 0$.

2.1.1. Adjustment Based on the Database Frequency Spectrum

Brenner's κ (kappa) estimate [22] (based on [23]) is $\hat{\pi}_q = (1-\kappa)/n$ when $k_q = 0$, where κ is the fraction of singleton profiles (observed only once in the database). Intuitively, a large κ corresponds to high profile diversity, which justifies a low estimate for π_q. The estimator can be very small if the database consists mainly of singletons, which is the case for high-mutation-rate Y profiles. In particular, if the database consists only of singletons, then $\kappa = 1$ and $\hat{\pi}_q = 0$. Cereda's estimators [24,25] are based on the numbers of both singletons and doubletons (profiles observed exactly twice). All of these approaches take no account of individuals that share profile q due to relatedness with Q, and the estimates can be strongly influenced by the way the database is sampled, as discussed above.

2.1.2. Augmenting the Database

If we compute the database relative frequency after adding q, then both k_q and n are augmented by one to obtain $\hat{\pi}_q = (k_q+1)/(n+1)$. If we add two copies of q to the database, corresponding to the profiles of both Q and X under H_X, we obtain [26]

$$\hat{\pi}_q = \frac{k_q + 2}{n + 2}. \tag{5}$$

Other methods can be modified similarly by augmenting the database with one or two copies of q.

Use of (5) is conservative in the sense that we do not know whether we have observed one or two individuals with q (that is, we do not know whether H_X is true). Both the above estimators can alternatively be derived as Bayesian posterior probabilities given a uniform prior for π_q, the first using the original database, while (5) uses the database augmented with one copy of q. A uniform prior is conservative in that it assigns weight to unrealistically high values of π_q.

2.1.3. Upper Confidence Limit (UCL)

An alternative to (5) is an upper confidence limit (usually 95%) for π_q [27,28]. This is the smallest binomial "success" probability π such that the probability of observing up to k_q successes in n trials is ≤ 0.05. That is, the UCL is the largest π such that

$$\sum_{x=0}^{k_q} \binom{n}{x} \pi^x (1-\pi)^{n-x} \leq 0.05 \tag{6}$$

which is sometimes called the Clopper–Pearson formula [29]. The UCL represents a standard scientific approach to controlling the risk of overstating the evidence: it provides an answer to the question of how big the unknown π_q could reasonably be, given observations k_q and n. The 95% UCL is larger (more conservative) than (5). For example, when $k_q = 0$, from (6) and $\ln(0.05) \approx -3$, the UCL is just under $3/n$, whereas (5) is just under $2/n$.

2.2. The Discrete Laplace Method

The Discrete Laplace method [30,31] models π_q for all possible q, making it also useful for model-based clustering and mixture analyses [32,33]. Intuitively, based on the database profiles, clusters are identified corresponding to the descendants of a recent common ancestor, and probabilities are computed for the observed profiles to belong to each cluster. The ancestors are treated as unrelated and so their profiles are independent. Profile probabilities are then computed by assuming that profiles descend independently from each ancestor according to a Discrete Laplace (double geometric) distribution.

2.3. Coancestry Adjustment for Population Substructure

The population genetic parameter F_{ST}, often referred to as θ in forensic DNA profiling, has been widely used to correct match probabilities for autosomal DNA profiles [34]. An analogous adjustment for lineage marker profiles [1,35] is

$$\text{LR} = \frac{1}{\theta + (1-\theta)\pi_q} \tag{7}$$

There are several interpretations of θ—one is that it represents the average level of relatedness of individuals within a subpopulation, relative to a larger population. In forensic settings, the larger population can be interpreted as the population from which the available profile database was drawn, while the subpopulation is not usually well defined but it is assumed to include Q and some or all of the alternative possible sources of the evidence sample. The denominator of (7) can be loosely interpreted as follows. Under H_X, there is probability θ that X is a relative of Q, sufficiently close that a match is very likely, while with probability $1-\theta$, X comes from the broader population which includes more distant relatives of Q such that the match probability is well estimated by the database relative frequency.

Equation (7) can be used in conjunction with the Discrete Laplace or adjusted database count methods to estimate π_q. However, θ cannot be directly estimated because the value will depend on the available case-specific background information as well as the population that the reference database was drawn from. For example, if the database is drawn from a broad population that extends greatly beyond the population relevant to the case, then a larger θ would be needed than if the database came from a smaller more homogeneous region that includes most alternative donors of the DNA in a particular case.

Population-genetic estimates of θ have been reported for many human subpopulation-population pairs [36,37] and from simulation scenarios [9]. Most available estimates of θ are for autosomal loci. For lineage markers, the smaller effective population size (there are 4 copies of each autosome for every Y chromosome) tends to increase the value of θ, but the higher mutation rate of the Y tends to act in the opposite direction. Many population genetic studies do not estimate θ relative to a forensic database, but instead use a hypothetical ancestral reference population which leads to smaller estimates. Further, few studies are available at the fine geographical scale relevant for many cases.

In general, every alternative X can have a different value of θ reflecting their level of ancestry shared with Q relative to the database population. However, rather than try to choose a distribution of θ values appropriate for the alternative contributors in a case, usual practice is to take a single value from the upper tail of that distribution. Ref. [36] argued for $\theta = 0.03$ as a conservative, default value for autosomal profiles, finding through simulations based on real data that it remains appropriate even if X in fact comes from a different continent than the database population.

2.4. Estimating the Number of Matches in the Population

If we knew that K_q individuals have a profile matching Q, and these are well-mixed in a population of alternative sources of the DNA of size N, then the LR would equal the inverse of the match probability, which is

$$\text{LR} = \frac{N}{K_q}. \tag{8}$$

We have noted above that this is not directly useful in practice, in contrast with the central role of the LR in the interpretation of autosomal profiles, because the choice of population and hence the value of N is problematic for lineage marker profiles. As we consider larger suspect populations, K_q tends to increase at a slower rate than N, because the matching individuals are relatives of Q and they are expected to form a smaller proportion of the population when it is more broadly defined.

An alternative to reporting the LR is to report estimates of K_q. A precise estimate is not usually possible, but probability distributions for K_q can be obtained using simulation under different population genetic models [7,21,38], which can include various mutation rates and mechanisms, and demographic factors such as population growth and structure, as well as variance in offspring number. The simulations can also take into account database frequencies, provided the database sampling scheme is known, which is only feasible in practice if the database can be assumed to be a random sample from the population. Often conditioning on k_q and n from a database does not greatly alter the distribution of K_q, since it merely confirms that the profile is rare, as expected from the mutation rate [7].

A further advantage of the simulation approach is that it can easily incorporate available information about the numbers of close relatives of Q, and about their profiles if available [38].

Using the malan (MAle Lineage ANalysis) software [20], the distribution of K_q was found in high-mutation-rate settings to be insensitive to the modelling assumptions [7,38]. In the case of Yfiler Plus, the number of matching males is typically <10 and rarely more than a few tens. This approach has been extended to mitogenomes [21], but due to their lower mutation rate, the distribution of K_q was spread over a wider range, and was more sensitive to the population genetic model. While there remains merit in reporting to jurors an estimate of K_q, the arguments for this are less compelling than in high-mutation-rate settings, and conversely problems with the match probability and LR are lessened.

Given the estimates of K_q, a juror can assess how likely it is that one of these matching individuals was the source of the evidence profile, rather than Q. The population size N may play some role in these considerations, but is relatively unimportant. Since K_q is a count, it is likely to be more interpretable to a juror than an LR or a match probability [39]: it can be presented using phrases such as "the number of individuals with profile q is unlikely to be more than…".

2.5. Methods Not Widely Used

2.5.1. Population Genetic Modelling Using the Coalescent

The first method for computing match probabilities using an explicit genetic model was based on a genealogical tree with $n + 2$ leaves, representing the database augmented with one copy of q plus a leaf node representing an unobserved profile [40]. Given a mutation model and a demographic model describing the population size, growth rates and structure, a Markov chain Monte Carlo algorithm updates the tree structure and branch lengths, as well as the profiles at the internal nodes of the tree and the unobserved leaf node. The distribution of profiles in the population is estimated from equilibrium frequencies at the unobserved leaf. The method was found to perform well in comparison to methods available at the time [41], but it is computationally demanding, particularly for large databases. This is because the whole profile space is explored, whereas only the probability of q is needed for forensic identification.

2.5.2. Frequency Surveying

The frequency surveying method [42–44] is based on pairwise distances, measured in mutational steps, between the Y profiles from individual i and j. An exponential regression is then based on these distances and used to establish a Beta prior distribution in a Bayesian model with a binomial likelihood. One disadvantage of this approach is that the differences in mutation rates across Y-STR loci are ignored, and only counting the number of mutational steps discards information. This method is still available at www.YHRD.org (accessed on 1 August 2021) [13], but is no longer recommended [11].

2.5.3. Graphical Statistical Models

These Y profile models [45,46] are computationally fast and allow intermediate alleles, which the population-genetic models above cannot accommodate. However, they do not exploit much genetic information and alleles are merely assumed to be different categories.

3. Recommendations from Forensic Authorities

For Y profiles, the Discrete Laplace method (Section 2.2) is currently recommended in the Philippines [47] and in Germany [48], where it was first used in court in a case from 2015 [49]. In that case, profiles were available for some male-line relatives of the suspected contributor, which can be taken into account in assessing the strength of the evidence by the malan simulation approach [38] (Section 2.4). The Polish Speaking Working Group of the International Society for Forensic Genetics (ISFG) [50] currently recommends using the κ method (Section 2.1.1). The UK Forensic Regulator has recently commended use of (5), the adjusted database frequency [10].

Below we briefly summarise and comment on recommendations from two other forensic authorities.

3.1. Scientific Working Group on DNA Analysis Methods (USA, Canada)

3.1.1. Y Profiles

The SWGDAM 2014 guidelines [51] note that "the profile probability is not the same as the match probability", but they do not distinguish between low- and high-mutation-rate profiles. They recommend that the profile probability is estimated by the unadjusted database relative frequency k_q/n, or with a UCL (Section 2.1.3), and that this value be used within (7) to adjust for population structure (Section 2.3).

The guidelines also discuss the importance of identifying the relevant population(s) and the difficulty in choosing an appropriate θ value, and they suggest default values for three kits. For PowerPlex Y23, the most discriminatory (highest profile mutation rate) of the three, they suggest $\theta = 2 \times 10^{-5}$ for African Americans, Asians, Caucasians and Hispanics; and $\theta = 3 \times 10^{-4}$ when Native Americans are considered.

3.1.2. mtDNA

The SWGDAM 2019 guidelines for mtDNA [52] are similar to those for Y profiles [51], except for the θ adjustment. They write: "*It is recognized that population substructure exists for mtDNA haplotypes. However, determination of an appropriate theta (θ) value is complicated by the variety of primer sets, covering different portions of HV1 and/or HV2, which may be applied to forensic casework. SWGDAM has not yet reached consensus on the appropriate statistical approach to estimating θ for mtDNA comparisons*".

3.2. International Society for Forensic Genetics (ISFG)

3.2.1. Y Profiles

The ISFG 2020 guidelines [11] state that "Information on the degree of paternal relatedness in the suspect population as well as on the familial network is, however, needed to interpret Y chromosome results in the best possible way" but did not delineate how to achieve this. The guidelines focus on estimating a profile relative frequency from a database, for which they recommend the Discrete Laplace method (Section 2.2), which we support in low-mutation-rate settings, but the guidelines do not distinguish low- and high-mutation-rate settings. They note arguments for avoiding estimates of the match probability π_q, referencing [7], but they write "the appropriate wording for statements not relying on population databases need to be validated in the context of the national guidelines".

3.2.2. mtDNA

The ISFG 2014 guidelines on mtDNA [15] focus on estimates of π_q. They do not recommend a particular estimator, but they mention (5) (Section 2.1.2) and the UCL (Section 2.1.3).

4. Some Further Issues

4.1. Combination with Autosomal Evidence

In many respects, lineage marker profiles resemble a single locus of an autosomal profile, with a higher mutation rate and hence greater allelic diversity than for typical autosomal loci but only one allele rather than two. The problem of the strong role for

relatedness that we have highlighted in this review also arises at a single autosomal locus, although, due to diploidy, there are four lineages connecting Q and X at an autosomal locus, one for each pairing of an allele from Q with one from X.

It is possible to compute a combined LR for autosomal and lineage marker profiles that both match Q. This is rarely done in practice, perhaps because an autosomal profile match is typically so informative that the relatively small additional evidential strength of the lineage marker match is outweighed by the additional interpretation issues. One possible approach would be to multiply a conservative estimate of K_q by the autosomal match probability, to obtain an estimate of the expected number of individuals matching Q at both lineage marker and autosomal profiles. Alternatively, if the lineage marker profile mutation rate is low, it may be acceptable to multiply autosomal and lineage marker LRs, each obtained using a suitable value of θ.

4.2. Locus Order for Duplicated Y Loci

The STR loci that form a Y profile have known locations on the Y chromosome, except for the ordering of some pairs of duplicated loci including DYS385a and DYS385b. Two males with the same pair of alleles DYS385a/b may or may not match because of the unknown order. When the two alleles from Q are the same as those from the evidence profile, the evaluation problem can be overcome by omitting the duplicated loci, which tends to underestimate evidential strength, or by assuming a match at both loci, which tends to overstate evidential strength, by at most a factor of two and arguably much less.

4.3. Consistency

A consistency principle has been proposed requiring that the strength of evidence for a Y profile cannot be less than that for any of the subprofiles obtained by omitting one or more loci [53]. This reasonable requirement is not enforced by most of the methods discussed here, because a Y profile is treated as a single allele and subprofiles are not considered. The problem can be important when there is a good sample size for a particular DNA profiling kit, but a reduced sample size for a more detailed profile that includes additional loci [10,51]. The method of Section 2.4 based on the distribution of the number K_q of matching individuals does respect the principle, because adding additional loci to a profile increases the mutation rate, and hence stochastically reduces K_q.

4.4. Partial Profiles

Often the DNA from a contributor of interest is at a very low level and/or degraded, so that a partial profile can arise if no allele is observed at some Y-STR loci, due to allele drop-out and/or masking by the alleles of a known contributor. The principles of interpretation are unchanged, using only the loci at which an allele is observed. The fewer the loci observed, the lower the profile mutation rate, which increases the number of matching individuals and decreases their average relatedness with Q. These quantities can be assessed using simulation including only the observed loci [20,54].

4.5. Mixtures

Many evidence profiles come from multiple unknown contributors. When there is a large discrepancy in the amounts of DNA, it may be possible to deconvolve the mixture (assign alleles to distinct, unknown contributors) based on peak heights. Mixture examples with peak height information are included in [55].

Surprisingly, for high-mutation-rate Y profiles, a two-male mixture profile in which q is fully represented is almost as powerful as observing q as a single-source evidence profile [38]. Although many pairs of possible Y profiles could give rise to the observed mixture, the great majority of these possible profiles do not exist in the worldwide human population, which is minuscule compared with the vast number of possible profiles. Since we know that q does exist in the population, the pair of profiles that includes q can be

much more likely than all the other profile pairs combined, unless one or more of the other profiles has been observed in a database [38].

For low-mutation-rate Y profiles, the Discrete Laplace method (Section 2.2) can be used to deconvolve mixtures using estimated population frequencies [33].

A "profile-centred" (HC) method [55] is based on (4) and focusses on the number of generations G in the lineage linking Q with alternative contributor X. Beyond some threshold on G, the method assumes that the match probability is low and uses a population frequency estimator similar to the Discrete Laplace method. The HC method assumes a constant-size, random mating population of size N to assign weights for the mixture donors (using formulas for the probability that two random persons are g generations apart). The HC method uses the κ method (Section 2.1.1) for calibration, so that these methods are the same in the special case of good-quality single-source profiles.

5. Discussion

Our review of methods for assessing the forensic value of DNA evidence from lineage markers, namely Y profiles and mitogenomes, has emphasised the key roles of the profile mutation rate and the (male-line or female-line) relatedness G, which are not adequately addressed by many methods. In particular, it is unsatisfactory to inform courts that, for example, a Y profile match is likely for male-line relatives and then to proceed as if the alleged contributor Q is unrelated to the alternative contributors, because (4) highlights that the evidential weight depends on the relatedness of Q with each alternative source of the evidence sample DNA profile.

Values of G are typically unknown in actual populations except when G is very small, but the distribution of G can be investigated via simulation in population genetic models (Section 2.4). There are also some well documented actual human populations that can be studied in more detail. To date these are relatively isolated and not typical of many cosmopolitan urban populations. A national-scale project is underway in Denmark aimed at tracing lineages over a century for almost the whole population [56]. At this time-depth, lineage paths up to about $G = 8$ for two contemporary young adults can be traced, provided that there are no migrants in the lineage.

When the profile mutation rate is low (say below 0.05 per generation, which holds for the mitogenome and older Y-STR profiling kits), most of the individuals with profiles matching Q are separated by at least several tens of generations. There are typically hundreds of matching individuals—in some cases, thousands of them [21]. That is enough individuals and at sufficient genetic distance that many of them will differ from Q in many characteristics, thus lessening the problems discussed above. In these cases, methods based on estimating the population match probability π_q, such as the Discrete Laplace method recommended by the ISFG (Section 2.2) may be acceptable, provided that the role of relatedness is also adequately explained possibly through a θ adjustment. As discussed in Section 2.3, appropriate values for θ can depend on case-specific details and the reference database.

For high-mutation-rate profiles (say above 0.1 per generation), most males matching Q are related to him within a few generations, and there seems no satisfactory alternative to summarising to a court the distribution of the number of close relatives of Q expected to match, as well as their degree of relatedness (Section 2.4). The UCL, for example, will be conservative if the alternative sources of the DNA include few close relatives of Q, but without addressing that question directly we cannot be sure. Methods that rely on a database may be affected by a non-conservative bias, because a broad database population may have lower average relatedness with Q than the alternative sources of the DNA in a particular case, and they may be adversely affected by the role of relatives in the database sampling frame.

Better models and rate estimates for mutation will improve the simulation-based approach to approximating the number of matching individuals. Some data are available but details such as the dependence of mutation rate on current profile state have been

little studied, particularly for mitogenomes. Mutation models allowing for allele-specific mutation rates have been investigated [5], but the available data were insufficient to estimate the model parameters accurately.

We believe that the relatedness, mutation and database issues that we have highlighted here can be adequately addressed, and that fair and comprehensible evaluation of lineage marker profile evidence is now within reach.

References

1. Balding, D.J.; Steele, C.D. *Weight-of-Evidence for Forensic DNA Profiles*, 2nd ed.; Wiley: Chichester, UK, 2015.
2. Johnston, I.; Burgstaller, J.; Havlicek, V.; Kolbe, T.; Rülicke, T.; Brem, G.; Poulton, J.; Jones, N. Stochastic modelling, Bayesian inference, and new in vivo measurements elucidate the debated mtDNA bottleneck mechanism. *Elife* **2015**, *4*, e07464. [CrossRef] [PubMed]
3. Rieux, A.; Eriksson, A.; Li, M.; Sobkowiak, B.; Weinert, L.A.; Warmuth, V.; Ruiz-Linares, A.; Manica, A.; Balloux, F. Improved Calibration of the Human Mitochondrial Clock Using Ancient Genomes. *Mol. Biol. Evol.* **2014**, *31*, 2780–2792. [CrossRef] [PubMed]
4. Översti, S.; Onkamo, P.; Stoljarova, M.; Budowle, B.; Sajantila, A.; Palo, J.U. Identification and analysis of mtDNA genomes attributed to Finns reveal long-stagnant demographic trends obscured in the total diversity. *Sci. Rep.* **2017**, *7*, 6193. [CrossRef]
5. Jochens, A.; Caliebe, A.; Rösler, U.; Krawczak, M. Empirical Evaluation Reveals Best Fit of a Logistic Mutation Model for Human Y-chromosomal Microsatellites. *Genetics* **2011**, *189*, 1403–1411. [CrossRef] [PubMed]
6. Brenner, C.H. Understanding Y haplotype matching probability. *Forensic Sci. Int. Genet.* **2014**, *8*, 233–243. [CrossRef]
7. Andersen, M.M.; Balding, D.J. How convincing is a matching Y-chromosome profile? *PLoS Genet.* **2017**, *13*, e1007028. [CrossRef]
8. Caliebe, A.; Krawczak, M. Match probabilities for Y-chromosomal profiles: A paradigm shift. *Forensic Sci. Int. Genet.* **2018**, *37*, 200–203. [CrossRef]
9. Steele, C.D.; Balding, D.J. Choice of population database for forensic DNA profile analysis. *Sci. Justice* **2014**, *54*, 487–493. [CrossRef]
10. Forensic Science Regulator. Guidance for Y-STR Analysis Delivered into the Criminal Justice System in England and Wales: Forensic Science Regulator Guidance Y-STR Profiling (FSR-G-227). 2021. Available online: https://www.gov.uk/government/publications/y-str-profiling (accessed on 1 August 2021).
11. Roewer, L.; Andersen, M.M.; Ballantyne, J.; Butler, J.M.; Caliebe, A.; Corach, D.; D'Amato, M.E.; Gusmão, L.; Hou, Y.; de Knijff, P.; et al. DNA commission of the International Society of Forensic Genetics (ISFG): Recommendations on the interpretation of Y-STR results in forensic analysis. *Forensic Sci. Int. Genet.* **2020**, *48*, 102308. [CrossRef] [PubMed]
12. Zimmermann, B.; Röck, A.; Huber, G.; Krämer, T.; Schneider, P.M.; Parson, W. Application of a west Eurasian-specific filter for quasi-median network analysis: Sharpening the blade for mtDNA error detection. *Forensic Sci. Int. Genet.* **2011**, *5*, 133–137. [CrossRef] [PubMed]
13. Willuweit, S.; Roewer, L. The New Y Chromosome Haplotype Reference Database. *Forensic Sci. Int. Genet.* **2015**, *15*, 43–48. [CrossRef]
14. Parson, W.; Dür, A. EMPOP—A forensic mtDNA database. *Forensic Sci. Int. Genet.* **2007**, *1*, 88–92. [CrossRef]
15. Parson, W.; Gusmão, L.; Hares, D.; Irwin, J.; Mayr, W.; Morling, N.; Pokorak, E.; Prinz, M.; Salas, A.; Schneider, P.; et al. DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing. *Forensic Sci. Int. Genet.* **2014**, *13*, 134–142. [CrossRef]
16. Larmuseau, M.; Vanderheyden, N.; Van Geystelen, A.; van Oven, M.; de Knijff, P.; Decorte, R. Recent Radiation within Y-chromosomal Haplogroup R-M269 Resulted in High Y-STR Haplotype Resemblance. *Ann. Hum. Genet.* **2014**, *78*, 92–103. [CrossRef] [PubMed]
17. Solé-Morata, N.; Bertranpetit, J.; Comas, D.; Calafell, F. Recent Radiation of R-M269 and High Y-STR Haplotype Resemblance Confirmed. *Ann. Hum. Genet.* **2014**, *78*, 253–254. [CrossRef] [PubMed]

18. Balanovsky, O. Toward a consensus on SNP and STR mutation rates on the human Y-chromosome. *Hum. Genet.* **2017**, *136*, 575–590. [CrossRef] [PubMed]

19. Tvedebrink, T.; Morling, N. Identical twins in forensic genetics—Epidemiology and risk based estimation of weight of evidence. *Sci. Justice* **2015**, *55*, 408–414. [CrossRef] [PubMed]

20. Andersen, M.M. malan: MAle Lineage ANalysis. *J. Open Source Softw.* **2018**, *3*. [CrossRef]

21. Andersen, M.M.; Balding, D.J. How many individuals share a mitochondrial genome? *PLoS Genet.* **2018**, *14*, e1007774. [CrossRef]

22. Brenner, C.H. Fundamental problem of forensic mathematics—The evidential value of a rare haplotype. *Forensic Sci. Int. Genet.* **2010**, *4*, 281–291. [CrossRef]

23. Robbins, H.E. Estimating the Total Probability of the Unobserved Outcomes of an Experiment. *Ann. Math. Stat.* **1968**, *39*, 256–257. [CrossRef]

24. Cereda, G. Impact of model choice on LR assessment in case of rare haplotype match (frequentist approach). *Scand. J. Stat.* **2017**, *44*, 230–248. [CrossRef]

25. Cereda, G. Bayesian approach to LR assessment in case of rare type match. *Stat. Neerl.* **2017**, *71*, 141–164. [CrossRef]

26. Balding, D.J. *Weight-of-Evidence for Forensic DNA Profiles*, 1st ed.; Wiley: Chichester, UK, 2005.

27. Holland, M.; Parsons, T. Mitochondrial DNA Sequence Analysis—Validation and Use for Forensic Casework. *Forensic Sci. Rev.* **1999**, *11*, 21–50. [PubMed]

28. Budowle, B.; Ge, J.; Chakraborty, R. Basic Principles for Estimating the Rarity of Y-STR Haplotypes Derived from Forensic Evidence. In *18th International Symposium on Human Identification*; Promega: Madison, WI, USA, 2007; Available online: https://www.promega.com/products/pm/genetic-identity/ishi-conference-proceedings/18th-ishi-oral-presentations-mvc/ (accessed on 1 August 2021).

29. Clopper, C.; Pearson, E. The use of confidence or fiducial intervals illustrated in the case of the binomial. *Biometrika* **1934**, *26*, 404–413. [CrossRef]

30. Andersen, M.M.; Eriksen, P.S.; Morling, N. The discrete Laplace exponential family and estimation of Y-STR haplotype frequencies. *J. Theor. Biol.* **2013**, *329*, 39–51. [CrossRef]

31. Andersen, M.M. Discrete Laplace mixture model with applications in forensic genetics. *J. Open Source Softw.* **2018**, *3*. [CrossRef]

32. Andersen, M.M.; Eriksen, P.S.; Morling, N. Cluster analysis of European Y-chromosomal STR haplotypes using the discrete Laplace method. *Forensic Sci. Int. Genet.* **2014**, *11*, 182–194. [CrossRef] [PubMed]

33. Andersen, M.M.; Eriksen, P.S.; Mogensen, H.S.; Morling, N. Identifying the most likely contributors to a Y-STR mixture using the discrete Laplace method. *Forensic Sci. Int. Genet.* **2015**, *15*, 76–83. [CrossRef]

34. Balding, D.J.; Nichols, R.A. DNA profile match probability calculation: How to allow for population stratification, relatedness, database selection and single bands. *Forensic Sci. Int.* **1994**, *64*, 125–140. [CrossRef]

35. Buckleton, J.; Krawczak, M.; Weir, B. The interpretation of lineage markers in forensic DNA testing. *Forensic Sci. Int. Genet.* **2011**, *5*, 78–83. [CrossRef]

36. Steele, C.D.; Syndercombe Court, D.; Balding, D.J. Worldwide Estimates Relative to Five Continental-Scale Populations. *Ann. Hum. Genet.* **2014**, *78*, 468–477. [CrossRef]

37. Buckleton, J.; Curran, J.; Goudet, J.; Taylor, D.; Thiery, A.; Weir, B. Population-specific FST values for forensic STR markers: A worldwide survey. *Forensic Sci. Int. Genet.* **2016**, *23*, 91–100. [CrossRef]

38. Andersen, M.M.; Balding, D.J. Y-profile evidence: Close paternal relatives and mixtures. *Forensic Sci. Int. Genet.* **2019**, *38*, 48–53. [CrossRef] [PubMed]

39. Hoffrage, U.; Lindsey, S.; Hertwig, R.; Gigerenzer, G. Communicating Statistical Information. *Science* **2000**, *290*, 2261–2262. [CrossRef] [PubMed]

40. Wilson, I.J.; Weale, M.E.; Balding, D.J. Inferences from DNA Data: Population Histories, Evolutionary Processes and Forensic Match Probabilities. *J. R. Stat. Soc. Ser. A* **2003**, *166*, 155–201. [CrossRef]

41. Andersen, M.M.; Caliebe, A.; Jochens, A.; Willuweit, S.; Krawczak, M. Estimating trace-suspect match probabilities for singleton Y-STR haplotypes using coalescent theory. *Forensic Sci. Int. Genet.* **2013**, *7*, 264–271. [CrossRef]

42. Roewer, L.; Kayser, M.; de Knijff, P.; Anslinger, K.; Betz, A.; Caglia, A.; Corach, D.; Füredi, S.; Henke, L.; Hidding, M.; et al. 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.* **2000**, *114*, 31–43. [CrossRef]

43. Krawczak, M. Forensic evaluation of Y-STR haplotype matches: A comment. *Forensic Sci. Int.* **2001**, *118*, 114–115. [CrossRef]

44. Willuweit, S.; Caliebe, A.; Andersen, M.; Roewer, L. Y-STR frequency surveying method: A critical reappraisal. *Forensic Sci. Int. Genet.* **2011**, *5*, 84–90. [CrossRef]

45. Andersen, M.M.; Curran, J.; de Zoete, J.; Taylor, D.; Buckleton, J. Modelling the dependence structure of Y-STR haplotypes using graphical models. *Forensic Sci. Int. Genet.* **2018**, *37*, 29–36. [CrossRef]

46. Andersen, M.M.; Caliebe, A.; Kirkeby, K.; Knudsen, M.; Vihrs, N.; Curran, J.M. Estimation of Y haplotype frequencies with lower order dependencies. *Forensic Sci. Int. Genet.* **2020**, *46*, 102214. [CrossRef]

47. Rodriguez, J.; Laude, R.; De Ungriaa, M. An integrated system for forensic DNA testing of sexual assault cases in the Philippines. *Forensic Sci. Int. Synerg.* **2021**, *3*, 100133. [CrossRef] [PubMed]

48. Willuweit, S.; Anslinger, K.; Bäßler, G.; Eckert, M.; Fimmers, R.; Hohoff, C.; Kraft, M.; Leuker, C.; Molsberger, G.; Pich, U.; et al. Joint recommendations of the project group "Statistical analysis of DNA" and the German Stain Commission on the statistical analysis of Y-chromosomal DNA typing results. *Rechtsmedizin* **2018**, *28*, 138–142. [CrossRef]
49. Roewer, L. Y-chromosome short tandem repeats in forensics—Sexing, profiling, and matching male DNA. *WIREs Forensic Sci.* **2019**, *1*, e1336. [CrossRef]
50. Rębała, K.; Branicki, W.; Pawłowski, R.; Spólnicka, M.; Kupiec, T.; Parys-Proszek, A.; Woźniak, M.; Grzybowski, T.; Boroń, M.; Wróbel, M.; et al. Recommendations of the Polish Speaking Working Group of the International Society for Forensic Genetics on forensic Y chromosome typing. *Arch. Forensic Med. Criminol.* **2020**, *70*, 1–18. [CrossRef]
51. SWGDAM. SWGDAM Interpretation Guidelines for Y-Chromosome STR Testing. 2014. Available online: https://www.swgdam.org/publications (accessed on 1 August 2021).
52. SWGDAM. SWGDAM MtDNA Interpretation Guidelines. 2019. Available online: https://www.swgdam.org/publications (accessed on 1 August 2021).
53. Cowell, R.G. Consistent estimation of Y STR haplotype probabilities. *Forensic Sci. Int. Genet.* **2020**, *49*, 102365. [CrossRef]
54. Andersen, M.M. mitolina: MITOchondrial LINeage Analysis. *J. Open Source Softw.* **2019**, *4*, 1266. [CrossRef]
55. Taylor, D.; Curran, J.M.; Buckleton, J. Likelihood ratio development for mixed Y-STR profiles. *Forensic Sci. Int. Genet.* **2018**, *35*, 82–96. [CrossRef]
56. Using Artificial Intelligence for Mapping the Danish Genealogy and Strengthening Research. Available online: https://novonordiskfonden.dk/en/news/kunstig-intelligens-skal-kortlaegge-danskernes-stam-trae-og-styrke-forskning/ (accessed on 17 June 2021).