Published in final edited form as:
Forensic Sci Int Genet. 2011 March ; 5(2): 78–83. doi:10.1016/j.fsigen.2011.01.010.

The Interpretation of Lineage Markers in Forensic DNA Testing

J.S. Buckleton,
ESR Ltd, Private Bag 92021, Auckland, NEW ZEALAND

M. Krawczak, and
Institute of Medical Informatics and Statistics, Christian-Albrechts University, 24105 Kiel, GERMANY

B.S. Weir
Department of Biostatistics, University of Washington, Box 357232, Seattle, WA 98195-7232, USA

Abstract

Mitochondrial DNA (mtDNA) and the non-recombining portion of the *Y* chromosome are inherited matrilinealy and patrilinealy, respectively, and without recombination. Collectively they are termed ‘lineage markers’. Lineage markers may be used in forensic testing of an item, such as a hair from a crime scene, against a hypothesised source, or in relationship testing.

An estimate of the evidential weight of a match is usually provided by a count of the occurrence in some database of the mtDNA or Y-STR haplotype under consideration. When the factual statement of a count in the database is applied to a case, issues of relevance of the database and sampling uncertainty may arise. In this paper, we re-examine the issues of sampling uncertainty, the relevance of the database, and the combination of autosomal and lineage marker evidence. We also review the recent developments by C.H. Brenner.

Keywords

Haplotype; mtDNA; Y-haplotype; Likelihood ratio; Match probability; Theta correction; Population genetics

Introduction

Lineage markers have particular forensic utility in cases of degraded samples, in remains identification and in sexual assault cases. Attaching numerical strength to matching profiles for mitochondrial or Y-chromosome haplotypes (broadly) follows the same principles as for autosomal markers, but the lack of recombination within such haplotypes suggests some modifications to avoid seriously understating the evidential power of a match. We first present a population genetic theory approach to put lineage marker testing in the same framework as autosomal marker testing, and then we review the recent interesting work of Brenner [1] that addresses the common situation of a particular profile not appearing in a

Corresponding Author: B.S. Weir, Department of Biostatistics, University of Washington, Box 357232, Seattle, WA 98195-7232, USA, Phone: +1 (206) 221-7947, Fax: +1 (206) 543-3286, bsweir@u.washington.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

database. Our approach is based on frequencies of particular lineage markers whereas that of Brenner assumes that the particular type contains no information when most types occur only once in a database.

Autosomal Markers

Suppose an autosomal DNA profile of interest is homozygous *AA* at some locus. Population genetic theory (e.g. [2]) distinguishes between three types of frequency of allele *A*. First, there is the 'sample frequency' $\tilde{p}_A$, i.e. the frequency of *A* in a sample of, say, *n* individuals (e.g. a database). Note that $\tilde{p}_A$ is a random variable with a value that depends upon the sample. Second, there is the frequency of *A* in the population from which this sample has been taken, p^*_A, which may be termed the 'actual frequency' of *A*. Note that p^*_A also equals the expected value of $\tilde{p}_A$ taken over samples from that population. With random sampling of individuals, and if the population in question is in Hardy-Weinberg equilibrium, the number of *A* alleles in a sample of *n* genotypes has a binomial distribution with parameters 2*n* and p^*_A, so that the variance of $\tilde{p}_A$ is

$$Var\left(\tilde{p}_A\right)=\frac{1}{2n}p^*_A\left(1-p^*_A\right)$$

However, any population genetic theory aimed at sensible match probabilities must address another level of sampling, namely that inherent to the underlying evolutionary process. The actual allele frequency p^*_A is just one of many values that are possible for a given evolutionary history, and the expected value of p^*_A taken over many different realizations of this history is written as p_A, usually referred to as 'the allele frequency' of *A*. It is the *probability* that an allele drawn at random from a population, not yet specified in terms of the actual allele frequency p^*_A, is of type *A*, i.e. "probability" in this context has an evolutionary connotation. Note that the total expectation of the sample frequency $\tilde{p}_A$, taken over samples and over populations with the same evolutionary history, also equals p_A.

For a very wide class of evolutionary models, the variance of the actual allele frequency p^*_A over populations is

$$Var\left(p^*_A\right)=P_A\left(1-p_A\right)\theta$$

which introduces parameter θ as a measure of normalized variance of allele frequencies over populations. Note that θ is synonymous with Wright's fixation index F_{ST} [3] which is usually interpreted as a measure of similarity of alleles within populations, relative to the similarity of alleles between populations. In fact, under most evolutionary scenarios, as similarity increases within populations, so does the dissimilarity between populations. Two special cases are noteworthy: $\theta = 1$, when all alleles within a population are the same because of a shared ancestry, but are different from alleles in other populations, and $\theta = 0$, when there is no excess similarity of alleles within populations due to shared ancestry.

If there is Hardy-Weinberg equilibrium within the population of interest, the probability that a randomly drawn profile is *AA* equals $P^*_{AA}=\left(p^*_A\right)^2$ and this has an expected value over populations of

$$E\left(P_{AA}^{*}\right)=E\left(p_{A}^{*}\right)^{2}=p_{A}^{2}+p_{A}\,(1-p_{A})\,\theta \qquad (1)$$

Still assuming Hardy-Weinberg equilibrium, the probability that two randomly and independently drawn individuals in the same population of interest are both *AA* is $\left(p_{A}^{*}\right)^{4}$, so the probability that an unknown person has *AA* given that one instance of *AA* has been seen in the population of interest is Pr(*AA*|*AA*) = Pr(*AA*,*AA*)/Pr(*AA*), and this equals $\left(p_{A}^{*}\right)^{2}$. Thus, the match probability is the same as the profile probability in this case.

This simple result breaks down, however, if there is population structure and if the relevant subpopulation is either not known or has not been sampled. Thus, let us suppose that p_{A}^{*} is the actual allele frequency in the relevant subpopulation, but that p_{A}^{*} has been estimated by a sample allele frequency $\tilde{p}_{A}$ that is from the whole population. The quantity θ now refers to the variation in actual allele frequencies over subpopulations. Note that there is an analogous quantity for the variation among populations but we will ignore this variation here. In other words we ignore dependencies in terms of p_{A}^{*} between subpopulations.

Finding the expected value of $\left(p_{A}^{*}\right)^{4}$ for structured populations is simplified under additional assumptions about the evolutionary process, namely drift-mutation equilibrium and equal mutation rates between alleles. One situation where the second assumption holds is the ‘infinite alleles’ model under which every mutation gives rise to a new allele, and this may be a reasonable approximation for long mtDNA sequences or haplotypes comprising many Y-STR markers. Under the above assumptions, and adding selective neutrality, Wright [3] showed that the p_{A}^{*} values follow a Beta distribution over subpopulations, i.e.

$$p_{A}^{*}\sim Be\left(\frac{(1-\theta)\,p_{A}}{\theta},\frac{(1-\theta)\,(1-p_{A})}{\theta}\right) \qquad (2)$$

This distribution has expected value p_A and variance $p_A(1-p_A)\theta$. Moreover, Balding and Nichols [4] showed that, in a structured population where p_A refers to the whole population, the Beta distribution assumption yields

$$E\left(p_{A}^{*}\right)^{4}=\frac{p_{A}\,[\theta+(1-\theta)\,p_{A}]\,[2\theta+(1-\theta)\,p_{A}]\,[3\theta+(1-\theta)\,p_{A}]}{(1+\theta)\,(1+2\theta)}$$

This leads to the match probability result given by the National Research Council [5]:

$$Pr\,(AA|AA)=\frac{[2\theta+(1-\theta)\,p_{A}]\,[3\theta+(1-\theta)\,p_{A}]}{(1+\theta)\,(1+2\theta)} \qquad (3)$$

The quantity in Equation 3 is greater than both the profile probability $Pr\,(AA)=p_{A}^{2}+p_{A}\,(1-p_{A})\,\theta$ and the “product rule” value p_{A}^{2}. In other words, variation of allele frequencies over subpopulations implies that the population frequency p_A, estimated by a population-wide $\tilde{p}_{A}$, cannot be used directly to calculate match probabilities for a particular subpopulation.

Since its introduction by Balding and Nichols in 1994, the match probability in Equation 3 has been extended to allow for inbreeding [6], mixtures [7] and relatedness [8]. It has been of substantial benefit for the interpretation of matching autosomal profiles, and it was endorsed by the US National Research Council [5]. For a single-contributor stain, for example, when both the stain and the suspect have profile *AA*, two hypotheses of interest may be

H_p: The suspect is the source of the stain.

H_d: The suspect is not the source of the stain.

and the likelihood ratio for H_p versus H_d might be simply

$$LR=\frac{1}{Pr\,(AA|AA)}$$

Matching with Lineage Markers

For an mtDNA or Y-chromosome haplotype of type *A*, the match and profile probabilities for a specific subpopulation are both equal to p_A^*, the profile frequency in that subpopulation. If there was a sample available from that subpopulation, it would furnish an estimate of this quantity although, as described in the following section, care is needed if the profile is not present in the sample.

With population structure, match and profile probabilities are no longer the same and the expected value of $\left(p_A^*\right)^2$ is needed. There is no need to invoke the Beta distribution here since Equation 1 already provides the match probability

$$Pr\,(A|A)=\theta+(1-\theta)\,p_A \tag{4}$$

This can be given a numerical value if reasonable estimates of p_A (e.g. $\tilde{p}_A$) and θ are available. Note that the match probability within a particular subpopulation is also greater than the haplotype frequency in the whole population since $\theta + (1 - \theta)p_A > p_A$. Two haplotypes are more likely to be the same when they are taken from the same subpopulation than when they are taken randomly from the whole population.

Estimation of Haplotype Frequencies

Statistical Sampling Only

If a sample from the relevant subpopulation is available, the sample haplotype frequency $\tilde{p}_A$ is a reasonable estimate of the actual haplotype frequency p_A^* and therefore can be used to estimate p_A, leading to the match probability. If haplotype *A* is seen *x* times in a sample of *n* haplotypes, $\tilde{p}_A=x/n$. There has been a tendency in the literature arising from [9] to regard this quantity as being normally distributed and to acknowledge sampling variation with a 100(1 – α)% confidence interval of the form

$$\tilde{p}_A \pm z_{(1-\alpha/2)}\sqrt{\frac{\tilde{p}_A\left(1-\tilde{p}\ A\right)}{n}}\,(\text{two}\quad\text{sided})$$

and

$$\tilde{p}_A \pm z_{(1-\alpha)} \sqrt{\frac{\tilde{p}_A\left(1- \tilde{p}\ A\right)}{n}}\ \text{(one sided)}$$

For 95% confidence, $\alpha = 0.05$ so that $z_{(1-\alpha/2)} = 1.96$ and $z_{(1-\alpha)} = 1.645$.

As acknowledged by the original authors [9], the actual sampling distribution is binomial

$$\begin{aligned} x & \sim B\left(n, p_A^*\right) \\ Pr(x) & = \binom{n}{x}\left(p_A^*\right)^x\left(1-p_A^*\right)^{n-x} \end{aligned}$$

and this is approximately normal only for moderately large values of np_A^*, say five or more. An exact confidence interval based on the binomial distribution was described by Clopper and Pearson [10]. An upper 100(1 – α)% one-sided confidence limit p_0 is found by solving

$$\sum_{k=0}^{x}\binom{n}{k} p_0^k (1-p_0)^{n-k} = \alpha \tag{5}$$

If the sample does not contain any copies of A, then Equation 5 gives $p_0 = 1 - \alpha^{1/n}$ and, for a 95% confidence limit, this is approximately $3/n$ once n is 100 or more. These confidence limits do not, however, address variation of p_A^* over subpopulations.

Statistical and Genetic Sampling

If there is population structure and Equation 4 is used for calculating the match probability, it is still possible to account for the effect of sampling variation on the relevant expression $\theta + (1-\theta)\tilde{p}_A$. Different samples will provide different values of $\tilde{p}_A$ and the variation among these values will decrease as the sample size increases. Use of Equation 4, however, explicitly acknowledges variation among subpopulations and this variation, in turn, inflates the variance of $\tilde{p}_A$. If a sample of n haplotypes consists of n_v haplotypes from the vth subpopulation, the sample haplotype frequency $\tilde{p}_A$ is a weighted average of the sample frequencies $\tilde{p}_{A_v}$ for each subpopulation, where the weights are the n_v's. The sample sizes n_v are usually not known but $\sum_v n_v = n$. As we are assuming the subpopulations are independent

$$\begin{aligned} \tilde{p}_A & = \frac{1}{n}\sum_v n_v \tilde{p}_{A_v} \\ Var\left(\tilde{p}_A\right) & = \frac{1}{n^2}\sum_v n_v^2 Var\left(\tilde{p}_{A_v}\right) \\ & = p_A\left(1-p_A\right)\left[\frac{\sum_v n_v^2}{n^2}\theta + \frac{1-\theta}{n}\right] \end{aligned} \tag{6}$$

The variance expression is derived in the Appendix. The term $p_A\left(1-p_A\right)\sum_v n_v^2\theta/n^2$ is the 'among subpopulation' component of variance, and $p_A(1 - p_A)(1 - \theta)/n$ is the 'within subpopulation' component.

When $\theta = 0$, Equation 6 reduces to the binomial variance $p_A(1 - p_A)/n$. The same result is obtained if every haplotype is systematically sampled from a different subpopulation so that the haplotypes are evolutionarily independent. There is no among subpopulation contribution to the variance. At the opposite extreme, if all haplotypes were sampled from the same subpopulation, they would all be dependent on each other and Equation 6 gives $Var\left(\tilde{p}_A\right)=p_A\left(1-p_A\right)\left[\theta+\left(1-\theta\right)/n\right]$, showing that there is a non-zero variance even for very large sample sizes n. There are still among- and between-subpopulation contributions to the variance. Only when $\theta = 1$, and all alleles within a subpopulation are completely dependent, does the within-subpopulation component of variance disappear.

The individual sample sizes n_v are not known but they may reflect the composition of the population. For a population in which there are s equally sized subpopulations, the variance would be

$$Var\left(\tilde{p}_A\right)=p_A\left(1-p_A\right)\left[\frac{\theta}{s}+\frac{1-\theta}{n}\right]$$

If s is large enough so that θ/s can be ignored relative to $(1 - \theta)/n$, then

$$Var\left[\theta+\left(1-\theta\right)\tilde{p}_A\right]\approx p_A\left(1-p_A\right)\frac{\left(1-\theta\right)^3}{n}$$

We believe it helpful to accompany match probabilities or likelihood ratios by measures of sampling variation because the latter reflect the effects of basing the former upon estimates from samples of profiles. Two likelihood ratios of 1,000 may be viewed quite differently if one had an estimated standard deviation of 1 and the other had an estimated standard deviation of 100.

Bayesian Approach

An alternative framework is to suppose that the actual subpopulation frequency p_A^* is a random variable with a prior distribution that is conveniently taken to be Beta. Although any Beta distribution could be assumed here, including the uniform distribution as a special case, there is population-genetic appeal in taking the distribution in Equation 2. This is combined with the binomial sampling distribution to give a Beta posterior distribution when a sample of size n contains x copies of A:

$$p_A^*\sim Be\left(\frac{\left(1-\theta\right)p_A}{\theta}+x,\ \frac{\left(1-\theta\right)\left(1-p_A\right)}{\theta}+\left(n-x\right)\right)$$

The expected value of this distribution is $\left[\left(1-\theta\right)p_A/n+\theta\tilde{p}_A\right]/\left[\left(1-\theta\right)/n+\theta\right]$, a weighted mean of p_A and $\tilde{p}_A$ with decreasing weight on the prior as the sample size increases. The Bayesian analog of a confidence interval is the credible interval based on $100(1 - \alpha)$% of the posterior distribution as opposed to $100(1 - \alpha)$% of the sampling distribution of an estimate. Numerical values for the boundaries of these credible intervals can be calculated with a program obtainable from the senior author of this paper.

Estimating *θ*

For autosomal markers, it is usual [5] to assign a value in the range 0.01 to 0.05 to θ. Because θ describes the normalized variance of allele frequencies over subpopulations, it is necessary in a non-Bayesian framework to have data from at least two subpopulations in order to estimate θ. Recall that if a sample was available from the relevant subpopulation, the sample haplotype frequency from those data could be used directly for calculating match probabilities without having to invoke θ. Otherwise, any estimation of θ would have to use data from a set of samples from subpopulations that were considered relevant. The average haplotype frequency over these samples could be used as an estimate of p_A, along with θ estimated from the between-subpopulation variation of the $\tilde{p}_A$ values.

A classical estimator for θ was given by Weir and Cockerham [11] and this estimator will return a value of zero if every haplotype appears only once in all samples combined. There would be no evidence then of more haplotype similarity within than among samples, as is required for θ to be greater than zero. Indeed, Budowle et al. [12,13] noticed that estimates became smaller as more Y-STR markers were used and each haplotype became less frequent in the data, although they did not encounter data where every haplotype was unique.

Ewens' Sampling Theory

Ewens' [14] approach, and that of Brenner [1] discussed below, do not focus on particular lineage types, A, as has been the case so far. Ewens gave a treatment of allele frequency distributions that required mutation-drift equilibrium and allelic exchangeability. These are the same assumptions made earlier by Wright [3] but they go beyond those made by Weir and Cockerham [11] who did not restrict the mutation model and did not assume equilibrium. Weir and Cockerham worked only with the mean and variance of the distribution of allele frequencies over populations whereas Wright's and Ewens' assumptions lead to the whole distribution. Ewens' conclusions apply, for example, to the infinite alleles model where drift and mutation are in balance. He showed that the number, k, of distinct alleles (or haplotypes for that matter) in a sample is sufficient for estimating parameter ψ, where $1/(1 + \psi)$ is the probability that two alleles (haplotypes) drawn randomly from a population are the same. The latter probability is also a definition of θ, and for a haploid population of size N with an infinite-alleles mutation rate of μ, $\psi = 2N\mu$ so that $\theta = 1/(1 + 2N\mu)$. Note that θ decreases as the mutation rate increases, and that haplotype mutation rates increase as the number of constituent markers increase.

A maximum likelihood estimate of ψ (or θ) is found by setting the observed value of k equal to its expected value

$$E(k) = \sum_{i=0}^{n-1} \frac{\psi}{\psi+i} = \sum_{i=0}^{n-1} \frac{1-\theta}{1+(i-1)\theta}$$

Numerical methods are needed to solve this equation for θ. Note that the frequencies of particular alleles are not used here. If every haplotype in a sample is unique, then $k = n$, which provides an estimate of zero for θ corresponding to (infinitely) large mutation rates.

Brenner's Approach

Brenner [1] addressed the issues by an approach that is reminiscent of, and refers to, Ewens' work, although he does not specify an explicit genetic model. Even though Brenner has a purely statistical approach, we discuss it here because of the parallels to the work of Ewens. Instead of employing the sufficient statistic k, Brenner used k, the proportion of haplotypes seen only once in a sample of size (n – 1) augmented by the crime-scene haplotype. He approximated the match probability between an innocent suspect and a previously unseen trace haplotype as $(1 - k)/n$. In most cases, this will lead to a likelihood ratio of $n/(1 - k)$ rather than the n that would result if the match probability was set to $1/n$, and for this reason Brenner refers to $1/(1 - k)$ as the "inflation factor".

If the trace haplotype had been seen (x–1) times in a database of size (n–1), Brenner modifies the likelihood ratio to $n/[x(1 - k)]$. Note that, if all haplotypes are seen only once in a database, then $k = 1$ and the match probability is zero. Note also that the same quantification of evidential strength is applied to all types of a lineage marker.

Ewens has shown, and Brenner repeats, that the expected value of k is

$$E(k)=\frac{\psi}{n+\psi-1}\approx\frac{\psi}{n+\psi}=\frac{1-\theta}{1+(n-1)\theta}$$

although Ewens did not suggest using this equation to estimate ψ or θ.

Searches of Frequency Databases

Using Equation 4, the problem of deriving match probabilities essentially reduces to one of estimating haplotype frequencies (in addition to θ). It is therefore helpful to compare some of the "counting estimates" of p_A that can be obtained from a database of (n–1) haplotypes $H_1, H_2, \ldots, H_{n-1}$ when information is also available on the haplotype H_A of a suspect and H_T of some trace evidence. A generic estimator $\widehat{p}^g_A$ of the population frequency of H_A is the sample frequency $\tilde{p}_A$. An adapted estimator $\widehat{p}^a_A$ adds the suspect's type to the database, so that the count of A is increased by one, and divides by n. Brenner added the trace haplotype, rather than the suspect's haplotype, to the database but the numerical results will be the same when suspect and trace types match. If $1_{\{X\}}$ takes the value of one when X is true and the value zero otherwise, the generic and adapted estimates of the population frequency of A can be written as

$$\begin{aligned}\widehat{p}^g_A &= \frac{1}{n-1}\sum_{i=1}^{n-1} 1_{\{H_i=A\}}\\ \widehat{p}^a_A &= \frac{1}{n}\sum_{i=1}^{n} 1_{\{H_i=A\}}\end{aligned}$$

(The suspect's type is the nth type in the augmented database.) The expected value of $\widehat{p}^g_A$ is just p_A from standard statistical theory. Here "expected" means the expectation over all possible databases. If attention is restricted to those databases that do not contain haplotypes of type A, then the conditional expectation of $\widehat{p}^g_A$ is zero, which is anti-conservative with respect to the match probability. The adapted estimate $\widehat{p}^a_g$, in contrast, has an expected value of $[(n - 1)p_A/n + 1/n]$ over all samples and so is biased but conservative (on average). In particular, when conditioning on databases that do not contain type A, $\widehat{p}^a_g$ has an expected value of $1/n$.

Since $(1 - k)$ is the proportion of haplotypes in a database that occur more than once, Brenner's [1] estimator $\widehat{p}^k_A$ of p_A may be written as

$$\widehat{p}_{A}^{k} = \frac{1-k}{n} = \frac{1}{n^2}\sum_{i=1}^{n} 1_{\{\exists_{j \neq i}, H_j = H_i\}}$$

Taking expectations over all augmented databases that contain only one copy of *A* gives the following conditional expectation of $\widehat{p}_{A}^{k}$:

$$E\left(\widehat{p}_{A}^{k} | H_1 \neq H_A, \ldots, H_{n-1} \neq H_A\right) = \frac{n-1}{n^2}\left[1 - \sum_{i=1}^{a} \frac{p_i(1-p_A-p_i)^{n-2}}{(1-p_A)^{n-1}}\right]$$

where p_i is the population frequency of the *i*th type different from *A*, and *a* is the number of such types in the sample. Without an evolutionary model such as that described by Wright [3] and Ewens [14], it is difficult to determine just what is being estimated by $\widehat{p}_{A}^{k}$. Moreover, the conditional expectation of $\widehat{p}_{A}^{k}$ does not depend only on p_A itself but is a complex function of all other sample haplotype frequencies. Therefore, it is difficult to judge for a particular database if, and to what extent, $\widehat{p}_{A}^{k}$ is either conservative or anti-conservative. We emphasize that both scenarios are possible, as Brenner [1] demonstrated himself by providing an example for each of them. Finally, Brenner's estimator implies that

$$np_A \approx (n-1)\, p_A \approx 1 - (1-p_A)^{n-1}$$

is reasonably estimated by 1 – *k*. In other words, the probability that a singleton (i.e. *A*) from the original database would occur more than once in a second realization of the same database (i.e. $[1 - (1 - p_A)^{n-1}]$) is assumed to be equal to the proportion of non-singletons in the original database (i.e. 1 – *k*). Why, and under which genetic model, such a link should exist is not detailed by Brenner [1].

Combination of Lineage and Autosomal Markers

Recently, Amorim [15] raised the issue whether likelihood ratios calculated from lineage and autosomal markers may be multiplied in order to obtain a single measure of the genetic evidence in a given case. Amorim 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 whole lineage. He argued that instead of

H_p: The suspect is the source of the stain.

H_d: The suspect is not the source of the stain.

the prosecution and defense hypotheses should read

H_p^*: The suspect or somebody from their lineage is the source of the stain.

H_d^*: Neither the suspect nor anybody from their lineage is the source of the stain.

when lineage markers are employed for genetic analysis. Due to these imponderabilities, Amorim [15] concluded that "the combination of likelihood ratios from the two sources of data should be avoided."

Unless the circumstances of an individual case suggest differently, however, 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 for which there is no prior evidence of being the stain donor. Moreover, if the possibility of mutation is neglected, then the likelihoods of the two hypotheses H_p and H_p^* would be identical for lineage markers anyway. Furthermore, the way match probabilities are usually translated into likelihoods of either H_d or H_d^* implies that the formulations given above are slightly incorrect. In both instances, the match probability refers to a situation where an "unknown person" has a particular genotype, given that one instance of this genotype has been seen in the population of interest. This means that the two defense hypotheses, H_d and H_d^*, are also formally the same. Note that, under this view, the likelihood ratio is also the same for all members of the suspect's lineage.

Amorim's criticism was partly motivated by his perception of a common practice of sampling lineages rather than sampling at random. Under such regimes, people with the same surname would be deliberately avoided when constructing a frequency database which, in turn, would lead to biased haplotype frequency estimates. However, while lineage-based sampling may have been exercised in the early days of some genetic databases, to our knowledge, this is not (or no longer) the case for major projects such as, for example, YHRD (www.yhrd.org) or EMPOP (empop.org). This notwithstanding, it may be worth reiterating that databases underlying match probabilities need to be representative of the population(s) of interest. Moreover, such databases should also be large enough to reveal whether previously proposed adjustments of LR products for dependencies between, and different levels of coancestry of, lineage and autosomal markers [16] are valid.

Discussion

We have considered the interpretation of both autosomal and lineage markers from a statistical genetic viewpoint. We have been careful to distinguish between three frequencies for a genetic profile: the value in a sample, the value in a subpopulation of relevance and the value over all subpopulations. We have adopted the established scientific meaning of "probability" as being a theoretical construct obeying the Laws of Probability. The probability of an allele drawn randomly from a population being of a certain type can be estimated by the frequency of that type in a sample and we do not understand Brenner's discussion of frequencies and probabilities [1], especially when he states that "Probability, by contrast, depends by definition only on the available data."

We believe there is merit in focusing on the type, or name [1], of a lineage marker although we acknowledge that there is certain equivalence among all types that are not seen in a database. It appears useful to us to attach different evidential weights to lineage marker types that occur different numbers of times in a database. We also believe it to be useful to compare sample frequencies in databases from different populations or ethnicities. Given our consideration of sample profile frequencies, it is natural for us to treat the variation inherent in these sample values and we see merit both in confidence intervals and in Bayesian credible intervals.

Brenner's match probability, like those emerging from the counting estimates of p_A discussed above, does not refer to a specific profile type. Instead, it treats types as "arbitrary names" [1]. Particularly in the case of short tandem repeat (STR) markers, however, such ignorance of type implies a substantial loss of information because the evolutionary relationship between STR haplotypes is encrypted in their constituent repeat lengths. A model to exploit type information in Bayesian fashion has been introduced in [17], and has recently been validated and improved [18]. Whether or not type-based frequency estimation warrants further investigation is a matter of debate, but it is difficult to perceive any useful alternative to the counting method that ignores the likely evolutionary relationship of a haplotype of interest and the remainder of a database. Brenner's approach [1], although intuitively appealing, seems not to be a solution to this dilemma. Moreover, as we have shown, Brenner's approach [1] also suffers from potential anti-conservativeness in the way it inherently estimates haplotype frequencies. It is therefore at odds with the requirement that forensic match probabilities should not lead to over-statements of the strength of the evidence.

Brenner's match probabilities and likelihood ratios seem most applicable to profiles not seen in a database. It would be interesting to compare the utility of his use of *k*, the proportion of singletons in a database, with Ewens' use of *k*, the number of different types in a database. The other comparison that may be fruitful is the upper credible interval based on the use of a Beta prior for profile probability and Brenner's "inflated" probability ($k - 1$)/n. In any case, we would not apply the inflation factor to the upper confidence limit for a profile not seen in a database because that seems to confound approaches.

In our view, the problem addressed by Amorim [15] on combining autosomal and lineage markers is mainly one of communication, not calculation. As a (usually unrequested) by-product, a small match probability for autosomal markers excludes the suspect's parents, siblings and offspring from being the source of the stain as well. Courts need to be informed that this is not necessarily the case with lineage markers and, therefore, with likelihood ratio products combining match probabilities for autosomal and lineage markers.

Acknowledgments

This work was supported in part by grant GM 075091 from the US National Institutes of Health.

Appendix

The derivation of the variance of sample allele frequencies from a structured population can be based on indicator variables x_{vi} for the *i*th allele sampled from the *v*th subpopulation. Each $x = 1$ if the corresponding allele is of type A and $x = 0$ if the allele is not of type A. Then, under the model in this paper, taking expectations over samples and subpopulations:

$$\begin{aligned} E(x_{vi}) &= p_A \\ E\left(x_{vi}^2\right) &= p_A \\ E\left(x_{vi}x_{vi'}\right) &= p_A^2 + p_A(1 - p_A)\theta, i \neq i' \\ E\left(x_{vi}x_{v'i'}\right) &= p_A^2, v \neq v' \end{aligned}$$

The sample allele frequency can be expressed as the average of the *x*'s for all alleles in the sample:

$$\tilde{p}_A = \frac{1}{n}\sum_v \sum_i x_{vi}$$

and, taking expectations,

$$\begin{aligned}
E\left(\tilde{p}_A\right) &= \tfrac{1}{n}\sum_v\sum_i p_A = p_A \\
E\left(\widehat{p}_A^2\right) &= \tfrac{1}{n^2}\left(\sum_v\sum_i E\left(x_{vi}^2\right) + \sum_v\sum_{i\neq i'} E\left(x_{vi}x_{vi'}\right) + \sum_{v\neq v'}\sum_{i,i'} E\left(x_{vi}x_{v'i'}\right)\right) \\
&= \tfrac{1}{n^2}\left(np_A + \sum_v n_v\left(n_v - 1\right)\left[p_A^2 + p_A\left(1 - p_A\right)\theta\right] + \sum_{v\neq v'} n_v n_{v'} p_A^2\right) \\
&= \tfrac{1}{n^2}\left(np_A + \left(\sum_v n_v^2 - n\right)\left[p_A^2 + p_A\left(1 - p_A\right)\theta\right] + \left(n^2 - \sum_v n_v^2\right)p_A^2\right) \\
Var\left(\tilde{p}_A\right) &= p_A\left(1 - p_A\right)\left(\frac{\sum_v n_v^2\theta}{n^2} + \frac{1-\theta}{n}\right)
\end{aligned}$$

References

[1]. Brenner CH. Fundamental problem of forensic mathematics - The evidential value of a rare haplotype. Forensic Science International: Genetics 2010;4:281–291. [PubMed: 20457055]

[2]. Cockerham CC. Variance of gene frequencies. Evolution 1969;23:72–84.

[3]. Wright S. Evolution in Mendelian populations. Genetics 1931;16:97–159. [PubMed: 17246615]

[4]. Balding DJ, Nichols RA. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Science International 1994;64:125–140. [PubMed: 8175083]

[5]. National Research Council. The Evaluation of Forensic DNA Evidence. National Academy Press; Washington, D.C.: 1996.

[6]. Ayres KL, Overall ADJ. Allowing for within-subpopulation inbreeding in forensic match probabilities. Forensic Science International 1999;103:207–216.

[7]. Curran J, Triggs CM, Buckleton J, Weir BS. Interpreting DNA mixtures in structured populations. Journal of Forensic Sciences 1999;44:987–995. [PubMed: 10486951]

[8]. Weir BS. The rarity of DNA profiles. Annals of Applied Statistics 2007;1:358–370. [PubMed: 19030117]

[9]. Holland MM, Parsons TJ. Mitochondrial DNA sequence analysis Validation and use for forensic casework. Forensic Science Review 1999;11:21–50.

[10]. Clopper CJ, Pearson ES. The use of confidence or fiducial intervals illustrated in the case of the binomial. Biometrika 1934;26:404–413.

[11]. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution 1984;38:1358–1370.

[12]. Budowle B, Ge J, Low J, Lai C, Yee WH, Law G, Tan WF, Chang YM, Perumal R, Keat PY, Mizuno N, Kasai K, Sekiguchi K, Chakraborty R. The effects of Asian population substructure on Y STR forensic analyses. Legal Medicine 2009;11:64–69. [PubMed: 19038565]

[13]. Budowle B, Ge J, Aranda X, Planz J, Eisenberg A, Chakraborty R. Texas population substructure and its impact on estimating the rarity of Y STR haplotypes from DNA evidence. Journal of Forensic Sciences 2009;454:1016–1021. [PubMed: 19627418]

[14]. Ewens WJ. The sampling theory of selectively neutral alleles. Theoretical Population Biology 1972;3:87–112. [PubMed: 4667078]

[15]. Amorim A. A cautionary note on the evaluation of genetic evidence from uniparentally transmitted markers. Forensic Science International: Genetics 2008;2:376–378. [PubMed: 19083851]

[16]. Walsh B, Redd AJ, Hammer MF. Joint match probabilities for Y chromosomal and autosomal markers. Forensic Science International 2008;174:234–238. [PubMed: 17449208]

[17]. Roewer L, Kayser M, de Knijff P, Anslinger K, Caglia A, Corach D, F́'uredi S, Geserick G, Henke L, Hidding M, Ḱ'argel HJ, Lessig R, Nagy M, Pascali WL, Parson W, Rolf B, Schmitt C, Szibor R, Teifel-Greding J, Krawczak M. 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 2000;114:31–43. [PubMed: 10924848]

[18]. Willuweit S, Caliebe A, Andersen MM, Roewer L. Y-STR frequency surveying method: A critical reappraisal. Forensic Science International: Genetics. 2011 in press.