Genetica **96:** 3–12, 1995.

A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity

David J. Balding & Richard A. Nichols
School of Mathematical Sciences and School of Biological Sciences, Queen Mary & Westfield College, University of London, Mile End Road, London E1 4NS, UK

Received 16 May 1994 Accepted 26 July 1994

Key words: DNA profiles, paternity, Wright's F_{ST}, coancestry, forensic science

Abstract

A method is proposed for allowing for the effects of population differentiation, and other factors, in forensic inference based on DNA profiles. Much current forensic practice ignores, for example, the effects of coancestry and inappropriate databases and is consequently systematically biased against defendants. Problems with the 'product rule' for forensic identification have been highlighted by several authors, but important aspects of the problems are not widely appreciated. This arises in part because the match probability has often been confused with the relative frequency of the profile. Further, the analogous problems in paternity cases have received little attention. The proposed method is derived under general assumptions about the underlying population genetic processes. Probabilities relevant to forensic inference are expressed in terms of a single parameter whose values can be chosen to reflect the specific circumstances. The method is currently used in some UK courts and has important advantages over the 'Ceiling Principle' method, which has been criticized on a number of grounds.

1. Introduction

The genetic composition of human populations varies because of, among other factors, their differing evolutionary histories and patterns of dispersal and interbreeding. The magnitude of the effect of this genetic differentiation on the forensic evaluation of DNA profile evidence is controversial. It is the practice of many forensic scientists to ignore coancestry except, possibly, in cases where genetically isolated populations or close relatives are clearly involved. Some authors argue, however, that uncertainty about possible levels of differentiation may invalidate such an approach (Lewontin & Hartl, 1991; Krane *et al.*, 1992). Others take the view that typical levels of differentiation are sufficiently small that they may routinely be neglected (Chakraborty & Kidd, 1991; Roeder, 1994).

We argue for an intermediate position: even small levels of genetic differentiation can be important and the effect should not be ignored. To do so would unfairly overstate the strength of the evidence against the defendant and the error could be crucial in some cases, such as those involving partial profiles or large numbers of possible culprits, many of whom share the defendant's ethnic background. However, the forensic use of DNA profiles need not be invalidated as a consequence. One approach to allowing for population differentiation, the 'Ceiling Principle', has been proposed by the US National Research Council (NRC) (*DNA Technology in Forensic Science*, Natl. Acad. Press, Washington D.C., 1992). The principle has been widely criticized (Robertson & Vignaux, 1992; Devlin, Risch & Roeder, 1993; Morton, 1993a; Weir, 1993a). In particular, the principle is inflexible and cannot be adjusted to the circumstances of a particular case, in part because it incorporates the view that the defendant's ethnicity is irrelevant to inference. We propose a method for quantifying the effect of genetic differentiation in terms of a single parameter, which can often be interpreted in terms of coancestry. Debates about the effect of population heterogeneity in particular cases can thus be simplified to a discussion of values for the parameter appropriate to the circumstances. Our proposed method has previously been described

(Balding & Nichols, 1994) and is currently used in some UK courts. Here, we develop the justification for the method and extend its application to paternity testing. The use of DNA profile evidence when incest is alleged in paternity cases is becoming increasingly common and the proposed method is particularly appropriate in such cases.

2. Key issues in forensic inference

Although the literature on forensic identification using DNA profile evidence is now extensive, many fundamental statistical issues are still not widely appreciated. Balding and Donnelly (1995) consider the forensic identification inference problem in a general setting and their analysis clarifies several issues. In particular, they show that the weight of evidence against the defendant depends on, for each possible perpetrator other than the defendant, the ratio of the likelihood of the DNA profile data if he were the culprit, to its likelihood if the defendant were the culprit. These likelihood ratios should then be summed by the jury, weighted by their probability, based on the non-DNA evidence, that each possible culprit is the true culprit.

To facilitate the discussion, it is common to make four simplifying assumptions:

1. that the crime sample DNA is that of the culprit;
2. that matches are unequivocal;
3. that if the defendant were the culprit then the defendant and crime sample DNA profiles would be certain to match; and
4. the fact that the defendant's DNA profile was investigated is not, in itself, informative about his/her profile.

These assumptions are not valid in general, but they allow us to focus on other important issues and deviations from them can be addressed within the framework discussed here. See Balding and Donnelly (1995) for further discussion.

Under these four assumptions, each likelihood ratio is simply the conditional probability that the possible culprit has the profile given that the defendant has it, that is, the 'match probability'. Note that the match probability may also be formulated in terms of the probability that the defendant has the profile conditional on the event that the alternative culprit has it, but we find the former definition to be more convenient.

Many authors ignore the conditioning on the observed profile and take the match probability to be equivalent to the relative frequency of the defendant's profile in some population. This use of profile frequencies in place of the match probability is inappropriate for several reasons. The concept of 'match' clearly involves two profiles, not one, and there seems no logical framework for linking profile frequencies with the issue of the defendant's guilt or innocence, which is the crucial issue in court. In particular, it is unclear how to allow coherently for the possibility that the culprit is related to the defendant, or shares ancestry through common origin in a subpopulation. Perhaps most importantly, there seems no logical framework for combining the DNA evidence, quantified by a profile frequency, with the non-DNA evidence.

Correct definition of the match probability clarifies much of the current debate. A general discussion of 'reference populations' can be avoided and neither is it necessary to consider hypothetical 'random' selections of suspects. Crucially, a coherent framework becomes available for incorporating the effects of shared ancestry, on both recent and evolutionary timescales. Since match probabilities are *conditional* probabilities, they cannot be estimated directly from database relative frequencies. Correlations in profile possession must be explicitly modelled in terms of population genetic theory in addition to the available data. Consequently, the ethnicities of both defendant and possible culprits are relevant to inference. Some authorities ignore correlations in profile possession and, instead, use 'conservative' estimates of relative frequencies. The Ceiling Principle, for example, is based on this approach. However unless the correlations are specifically taken into account, it is impossible to assess what level of 'conservativeness' is appropriate.

Some of the current debate concerning population differentiation focusses on statistical tests of hypotheses of independence in forensic databases (Geisser & Johnson, 1993; Weir, 1993b). The tests are complicated by the experimental difficulties involving apparent homozygotes. It is, in any case, difficult in principle to draw conclusions relevant to forensic inference from the outcomes of such tests. Population differentiation indubitably exists, the question of interest concerns the magnitude of its effect on match probabilities. Failure to reject a null hypothesis of no differentiation reflects some combination of insufficient, or inappropriate, data, low power against the alternatives of interest and small magnitude of effect. Such tests are thus not directly helpful in forensic inference. We propose parameter estimation, both point and interval, as an alternative to hypothesis testing.

3. Likelihood ratios for identification and paternity

3.1 Identification

We consider single-locus DNA profiles, one taken from a crime sample and one from a defendant, and make the four assumptions listed in Section 2. The match probability then depends on a number of factors. In particular, it is affected by the possibility that the individuals have matching DNA profile bands through shared inheritance from a common ancestor. For some possible culprits, the amount of ancestry shared with the defendant is largely known. This can occur, for example, when the defendant's close relatives are possible culprits. (Note that 'possible culprits' is taken to include all individuals not excluded by the non-DNA evidence, not merely those on whom suspicion falls for good reason (Lempert, 1991).) More generally, the amount of shared ancestry between defendant and possible culprit will be unknown. Frequently, however, many possible culprits will have features in common with the defendant (Lempert, 1991), such as similar physical description or location of residence, and hence defendant and possible culprit may plausibly have a large level of shared ancestry compared with two 'random' individuals.

In addition to shared ancestry, the match probability is also affected by uncertainty about relative frequencies of bands. Such uncertainty occurs because forensic databases are rarely exactly appropriate for the possible culprits in a specific crime. They typically are unplanned samples from large, heterogeneous racial groups which are subject to sampling and other sources of error.

Balding and Nichols (1994) proposed the following formulae for $\Pr(AA|AA)$ and $\Pr(AB|AB)$, the single-locus match probabilities for possible culprits not known to be close relatives of the defendant in, respectively, the homozygote and heterozygote cases:

$$\Pr(AA|AA) = \frac{(2F + (1-F)p_A)(3F + (1-F)p_A)}{(1+F)(1+2F)} \quad (1)$$

$$\Pr(AB|AB) = 2\frac{(F + (1-F)p_A)(F + (1-F)p_B)}{(1+F)(1+2F)}, \quad (2)$$

in which p_A and p_B denote the relative frequencies of alleles A and B in the population from which the database is drawn, in principle that most appropriate for the possible culprit under consideration. In practice, the homozygote case is complicated by the fact that, because of experimental difficulties, some heterozygotes may be incorrectly classified as homozygotes. Balding and Nichols (1994) give match probabilities which take this difficulty into account, as well as extensions of (1) and (2) to the case that the possible culprit under consideration is known to be a close relative of the defendant.

Two distinct justifications for (1) and (2) are given in Sections 4.1 and 4.2. Equation (2) differs slightly from that originally proposed by the authors (Nichols & Balding 1991), which employed an approximation ignoring certain higher order correlations described in Section 4.

The parameter F in (1) and (2) may be interpreted as measuring the degree of uncertainty about p_A as an estimate of the match probability for a single A allele. The case F=0 corresponds to certainty so that the single-locus match probabilities are exactly p_A^2 and $2p_Ap_B$, which, with p_A and p_B replaced by sample relative frequencies, are the values used in the so-called 'product rule'. Absolute certainty is unrealistic in practice and thus the product rule consistently overstates the strength of the evidence against the defendant (unless $p_A+p_B \geq 2/3$, which never arises for most typing systems). For realistic values of F, the effect can be important (Balding & Nichols, 1994).

In many cases, shared ancestry between defendant and possible culprit on an evolutionary timescale may be considered the most important source of uncertainty, in which case F may be approximately the same as Wright's F_{ST}. Estimates of F_{ST} are often based on populations which are geographically closely spaced. In the forensic context, however, it is of interest to compare broad racial groups with subpopulations at varying levels of stratification. At traditional loci, collations of allele frequency estimates for disparate human populations are available. A recent survey (Cavalli-Sforza & Piazza, 1993) reports F_{ST} estimates among Europeans with a median of 0.8% while the 90th percentile is about 2.8%. The corresponding values are 2.7% and 14% among Africans and 4.3% and 12% among Asians. These results may not be directly relevant to forensic inference, since such meta-studies encompass differing methodologies, sampling may concentrate on unusual populations and the loci surveyed may be subject to geographically-varying selection. In addition, mutation rates at the these loci are typically much low-

er than at the VNTR loci currently used in forensic work.

More directly relevant, in view of moves to introduce short tandem repeat (STR) loci for forensic work, is the differentiation reported at two of the three loci examined in a sample of ethnic groups classified as Greek Cypriot, Gujarati, Northern European and Pakistani (Wall *et al.*, 1993). Some loci show little differentiation within the broad racial groups (European or Asian), others show dramatic differences. There are a variety of plausible explanations for the differences between loci. M. Greenhalgh (pers. comm.) has implemented accurate automatic sequencer technology to overcome technical problems with the F13A1 locus and reports substantial differentiation between Gujarati and Pakistani populations (Fig. 1). The data of Wall *et al.* (1993) from other loci show more marked differentiation within both major ethnic groups. This greater differentiation could be a consequence of a variety of processes. Geographically-varying selection on the gene containing the STR-bearing intron, or a linked locus, could cause allele frequencies to diverge. Mutation can produce either greater differentiation or, conversely, convergence in allele frequencies depending on the mechanism. It thus seems plausible that the variation within broad racial groups varies from locus to locus, but that the more variable STR loci show at least as much differentiation as traditional loci.

Appropriate surveys are not yet available at the VNTR loci which predominate in current forensic work. Direct estimation of F_{ST} from forensic databases is hampered by, among other factors, sensitivity to assumptions about apparent homozygotes and the ill-defined sampling frame of the databases. In addition, the ethnic origin of individuals in databases is often not known in sufficient detail to permit the investigation of population differentiation at the finer levels of stratification which may be appropriate for forensic inference. Reliable estimates will require substantial surveys of individuals of known ethnicity at varying levels of stratification. The preliminary evidence which is available suggests that F_{ST} at VNTR loci may typically be smaller than at traditional loci, which is plausible in view of the higher mutation rates, and that values may differ substantially from locus to locus. Morton (1993b) gave average point estimates of around 0.1% for US Caucasians and 1% for US Blacks. Because of the difficulties discussed above, and below, the values appropriate for forensic inference are likely to be substantially larger.

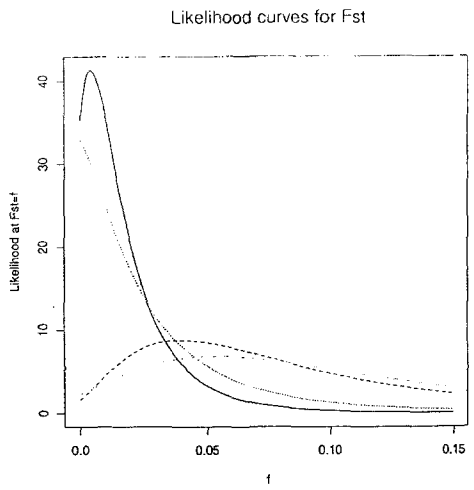

Fig. 1. Likelihood curves for F_{ST} measuring the differentiation at the STR loci CD4 and F13A1. For each locus the differentiation of a Cypriot sample from a North European database (C-N) and a Gujarati sample from a Pakistani (G-P) database is shown. At the origin ($f = 0$) the highest pair of curves relate to the F13A1 locus, and the lowest to CD4. Within each pair the G-P curve is higher (at $f = 0$). The CD4 data are from Wall *et al.*, (1993) and the F13A1 data from M. Greenhalgh (unpublished). Sample sizes for the Pakistani, North European, Gujarati and Greek Cypriot populations were 186, 58, 66, 50 (F13A1) and 50, 88, 44, 80 (CD4). The y-axis is scaled so that the curves can be directly interpreted as posterior densities with respect to a uniform prior for F_{ST}.

Morton (1992) proposed the use of formulae due to Yasuda (1968) which agree with (1) and (2) up to terms in F. Ignoring terms of order F^2 is reasonable when F is small compared with p_A and p_B, such as occurs with most traditional loci, but is often not appropriate for VNTR data.

Equations (1) and (2) deal with the single-locus case. In principle, it is not reasonable to assume independence across loci because of a sequential effect similar to that described by Donnelly (1995). If, as is usually the case, there is some uncertainty about the amount of shared ancestry between defendant and possible culprit, each successive single-locus match makes a higher level of shared ancestry more plausible, and hence a subsequent match is somewhat less surprising than the first. This feature of forensic inference differs from the usual use of F_{ST} in population genetics. The effect can in principle be accounted for by regarding F in equations (1) and (2) as having a distribution of possible values. Final single-locus match probabilities should then be obtained by integration with respect to this distribution. Four-locus match probabilities based on (1) and (2) involve powers of F up to order eight and the upper tail of the distribution will thus have a

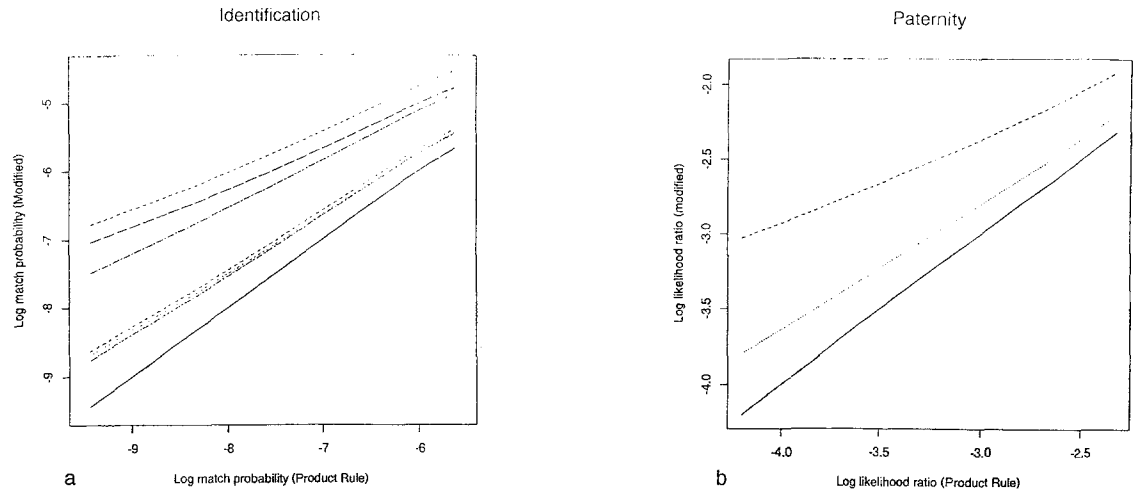

Fig. 2. (a) Match probabilities ($\log_{10}$) for four-locus, all-heterozygote profiles when the possible culprit under consideration is not known to be a relative of the defendant. The straight line $y = x$ gives the result of the unmodified product rule, which corresponds to $F = 0$. The upper set of three curves is for $F = 5\%$, the lower set is for $F = 1\%$. Within each set of three curves, the middle curve represents the modified match probabilities based on (2), the higher and lower curves represent, respectively, the approximations of Nichols and Balding (1991) and Morton (1992). The four pairs of allele frequencies are $(\alpha, 2\alpha)$, $(9\alpha, 5\alpha)$, $(6\alpha, 8\alpha)$ and $(7\alpha, 7\alpha)$ for $0.01 \leq \alpha \leq 0.03$. (b) Likelihood ratios ($\log_{10}$) for four-locus profiles when the alternative father under consideration is not known to be a relative of the alleged father and the genotypes of mother, child and alleged father at each locus are of the form AB, BC and CD. From lower to higher, the three lines represent the likelihood ratio with, respectively, $F = 0$, $F = 1\%$ and $F = 5\%$. The frequencies of the child's paternal alleles at the four loci are α, 9α, 6α, and 7α for $0.01 \leq \alpha \leq 0.03$.

large influence on the resulting match probability. This effect may be approximately accounted for by using an 'effective' value of F in the upper tail of the distribution.

As an illustration, suppose that F has probability density proportional to $(1-f)^{100}$ at $F = f$, so that the mean and standard deviation of F are both near 1% while the median is close to 0.7%. The shape of this distribution is chosen to reflect a high probability that the amount of shared ancestry between defendant and possible culprit, and hence F, is small and a small probability for F to be large. Ignoring this distribution and wrongly assuming that $F = 1\%$ leads to a three-fold error in the match probability for a four-locus, all-heterozygote profile with $p = 5\%$ for each band. When $p = 1\%$, the error is 34-fold. The appropriate 'effective' values of F are, respectively, 1.9% and 2.7%. Point estimates of the mean or median of F_{ST} are thus not directly relevant to forensic inference.

Figure 2(a) illustrates the effect of allowing for realistic levels of population differentiation. For a range of four-locus, all-heterozygote profiles, it compares match probabilities calculated using the product rule with those obtained from (1) and (2), and two approximations (Nichols & Balding, 1991; Morton, 1992), when F takes values 1% and 5%. With $F = 5\%$, the product rule can understate the appropriate match probability by two orders of magnitude which can be important, especially in cases involving little or no evidence other than the DNA profiles. Balding and Donnelly (1995) show that a small match probability does not necessarily provide convincing proof of guilt and thus a change of one or two orders of magnitude, even in a very small match probability, can be crucial in some cases.

3.2 Paternity

The principles which lead to (1) and (2) can be extended to allow for uncertainty, including that due to possible shared ancestry, in paternity testing. Suppose that we have single-locus DNA profiles for each of mother, child and alleged father. Evaluating the probability that the alleged father is the true father requires, for every other possible father, the ratio of the likelihood of the observed DNA profiles if he were the true father to their likelihood if the alleged father were the true father. These likelihood ratios each depend on the amount of shared ancestry among the mother, alleged father and alternative possible father. If the alternative father under consideration is not directly related to either the mother or the alleged father, but has a similar level of shared ancestry with both of them, then, under assumptions analogous to those which led to (1) and (2), the single-locus likelihood ratios for the pos-

Table 1. Single-locus likelihood ratios for paternity when the mother's genotype is AB. Blank entries indicate that the alleged father is excluded.

Alleged Father	Child AA	Child AB	Child AC
AA	$\frac{3F+(1-F)p_A}{1+3F}$	$\frac{4F+(1-F)(p_A+p_B)}{1+3F}$	
AB	$2\left(\frac{2F+(1-F)p_A}{1+3F}\right)$	$\frac{4F+(1-F)(p_A+p_B)}{1+3F}$	
AC	$2\left(\frac{2F+(1-F)p_A}{1+3F}\right)$	$2\left(\frac{3F+(1-F)(p_A+p_B)}{1+3F}\right)$	$2\left(\frac{F+(1-F)p_C}{1+3F}\right)$
CC			$\frac{2F+(1-F)p_C}{1+3F}$
CD			$2\left(\frac{F+(1-F)p_C}{1+3F}\right)$

Table 2. Single-locus likelihood ratios for paternity when the mother's genotype is AA. Blank entries indicate that the alleged father is excluded.

Alleged Father	Child AA	Child AB
AA	$\frac{4F+(1-F)p_A}{1+3F}$	
AB	$2\left(\frac{3F+(1-F)p_A}{1+3F}\right)$	$2\left(\frac{2F+(1-F)p_B}{1+3F}\right)$
BB		$\frac{2F+(1-F)p_B}{1+3F}$
BC		$2\left(\frac{F+(1-F)p_B}{1+3F}\right)$

sible observed genotypes are given in Tables 1 and 2. If the alternative father has substantial shared ancestry with either mother or alleged father but not both, then $1+F$ may be more appropriate in the denominator of the likelihood ratio, in place of $1+3F$. The difference will, however, usually be unimportant.

Most current practice employs the values in Tables 1 and 2 but with $F = 0$, corresponding to no shared ancestry and complete certainty about band relative frequencies. An inappropriate assumption of certainty thus leads, for realistic values of the parameters, to an overstatement of the probability that the alleged father is the true father. The magnitude of the overstatement for a four-locus profile is illustrated in Figure 2(b). With $F = 5\%$, ignoring uncertainty can lead to an order of magnitude overstatement of the likelihood ratio.

Finally, we consider the case that the alternative father under consideration is known to be a close relative of the alleged father (but the DNA profile of the former is not available). Let r denote the probability that an allele drawn from the alternative father matches one of the alleged father's alleles at that locus through inheritance from the known ancestors, so that $r = 1/2$ when they are brothers and $r = 1/4$ for either uncle-nephew or half-brothers. The selection of an allele from the alternative father is exactly equivalent to selecting with probability r an allele from the alleged father and with probability $1-r$ an allele from an apparently unrelated person, typically in the same subpopulation. The single-locus likelihood ratio is thus

$$r + (1-r)LR, \tag{3}$$

where LR denotes the appropriate value from Tables 1 or 2.

4. Derivation of likelihood ratios

If two individuals are drawn from a randomly-mating subpopulation then, when the subpopulation frequencies are known, the probability of observing any four specified alleles can be expressed in terms of the product of the corresponding frequencies. For example, two AA homozygotes are observed with probability $\tilde{p}_A^4$, where $\tilde{p}_A$ is the subpopulation frequency of A alleles, while two AB heterozygotes are observed with probability $4\tilde{p}_A^2\tilde{p}_B^2$ (the constant 4 occurs because of the two possible orderings of each of the two AB pairs). When the subpopulation frequencies are unknown, the probability is given by the expectation of the product. The

probabilities given at (1) and (2) can thus be expressed in the form

$$\Pr(AA|AA) = \frac{\mathrm{E}(\tilde{p}_A^4)}{\mathrm{E}(\tilde{p}_A^2)} \quad (4)$$

$$\Pr(AB|AB) = 2\frac{\mathrm{E}(\tilde{p}_A^2\tilde{p}_B^2)}{\mathrm{E}(\tilde{p}_A\tilde{p}_B)}. \quad (5)$$

The expectations in (4) and (5) must be based on a model for the evolution of the population at each locus. The evolution of VNTR loci is complicated and traditional population genetic models may not accurately describe their behaviour (Harding, 1992; Jeffreys *et al.*, 1994). Here, we formulate expressions for moments such as (4) and (5) which are valid under a range of evolutionary models. We develop justifications for these expressions using two approaches, the first based directly on a specific evolutionary model and the second using more general statistical arguments.

4.1 Genetical derivation

To specify fully the expected frequencies of pairs of diploid genotypes would require nine parameters (Cockerham, 1971). It is clearly not feasible to estimate all these parameters for each of the possible culprits relevant to a particular case. Here we specify a genetic model under which each of the nine parameters, and hence the expected frequencies, can be expressed in terms of a single parameter. The model is reasonably general and the simplification to a single, readily interpreted parameter is very helpful in a court environment, in which the use of more complicated, multi-parameter models may be inappropriate.

We consider a randomly-mating subpopulation, partly isolated from a large population, in which migration and mutation events occur independently and at constant rates. We write $\theta/(2N)$ for the sum of the two rates, where N denotes the subpopulation size (number of alleles). The probability F that two alleles are identical by descent (ibd) through an ancestor in the same subpopulation is simply the probability that, in tracing back the two lineages, an immigration or mutation event does not occur prior to the lineages coalescing in a common ancestor. Coalescences occur independently of migrations and mutations at rate $1/N$ while the total rate at which mutations or migrations occur on the two lineages is θ/N and thus we have

$$F = \frac{1/N}{1/N + \theta/N} = \frac{1}{1+\theta}.$$

The probability that two alleles drawn from the subpopulation are both type A is given by the familiar formula (Crow & Kimura, 1970)

$$\Pr(AA) = \mathrm{E}(\tilde{p}_A^2) = F\pi_A + (1-F)\pi_A^2, \quad (6)$$

in which we introduce π_A for the probability that a migration or mutation event produces an allele of type A. The first term in (6) is the probability that the two alleles are ibd and the most recent common ancestor was of type A, while the second term gives the probability that the two alleles are not ibd and are, in effect, the results of independent draws from a mechanism which generates A alleles with probability π_A. If the subpopulations are in equilibrium, the value of π_A is naturally estimated by p_A, the population relative frequency of A alleles, and henceforth we replace π_A with p_A. The probability that the two alleles are distinct, of types A and B say, is

$$\Pr(AB) = 2\mathrm{E}(\tilde{p}_A\tilde{p}_B) = 2(1-F)p_Ap_B.$$

Consider next $\mathrm{E}(\tilde{p}_A^3)$, the probability that three alleles chosen randomly from the subpopulation are all of type A. Tracing the three lineages backward in time, the rate at which any two coalesce is $3/N$, while the total rate at which mutations or migrations occur on the three lineages is $3\theta/(2N)$. The probability that the first event is a coalescence is then

$$\frac{3/N}{3/N + 3\theta/(2N)} = \frac{2}{2+\theta} = \frac{2F}{1+F}.$$

Continuing backwards in time, the two remaining lineages coalesce prior to a mutation or migration event with probability F. The probability that all three alleles are ibd is thus $2F^2/(1+F)$. Similarly, it can be seen that the probabilities that precisely one and zero pairs of alleles are ibd are, respectively, $3F(1-F)/(1+F)$ and $(1-F)^2/(1+F)$. The probability that all three alleles are of type A is given by the above terms multiplied by the probability that a type A allele is generated at each mutation or migration event, and hence

$$\begin{aligned} E(\tilde{p}_A^3) &= p_A\frac{2F^2}{1+F} + p_A^2\frac{3F(1-F)}{1+F} + p_A^3\frac{(1-F)^2}{1+F} \\ &= \frac{p_A}{1+F}(F + p_A(1-F))(2F + p_A(1-F)). \end{aligned} \quad (7)$$

Extending this argument to arbitrary numbers of alleles leads to recursive formulae of the following form:

$$\begin{aligned} E(\tilde{p}_A^{r+1}\tilde{p}_B^s\tilde{p}_C^t\tilde{p}_D^u) &= \mathrm{E}(\tilde{p}_A^r\tilde{p}_B^s\tilde{p}_C^t\tilde{p}_D^u) \\ &\times \left(\frac{rF + p_A(1-F)}{1 + (r+s+t+u-1)F}\right), \end{aligned} \quad (8)$$

for all integers $r, s, t, u \geq 0$. Similar formulae apply for more than four distinct alleles. Note that equation (6) is valid for F corresponding to shared ancestry either through known ancestors, such as parents and grandparents, or on an evolutionary timescale. In general, however, the derivation of (8) is only valid when F has the latter interpretation. Shared ancestry through known relatives is discussed by Balding and Nichols (1994).

Equations (1) and (2) follow from (4), (5) and (8). The formulae of Tables 1 and 2 also follow from (8). For example, consider the case that mother, child, and alleged father's genotypes are, respectively, AB, AC, and CD. If the alleged father were the true father then the conditional probability of the child's genotype, given the parent's genotypes, would be simply 1/4. Further, the parent's genotypes represent outcomes A, B, C and D in four draws from the subpopulation and thus have likelihood $\mathrm{E}(\tilde{p}_A\tilde{p}_B\tilde{p}_C\tilde{p}_D)$. The the joint likelihood is therefore $\mathrm{E}(\tilde{p}_A\tilde{p}_B\tilde{p}_C\tilde{p}_D)/4$. If the alternative father were the true father then the likelihood of the child's maternal allele, given the mother's genotype, is 1/2 and we have, under this hypothesis, observed two C alleles, one from the alleged father and one from the true father. The joint likelihood in this case is thus $\mathrm{E}(\tilde{p}_A\tilde{p}_B\tilde{p}_C^2\tilde{p}_D)/2$. Substituting from (8), the ratio of these joint likelihoods gives

$$2\left(\frac{F+(1-F)p_C}{1+3F}\right),$$

as given in Table 1.

Note that the moments (8) are exactly those which follow from assuming that $(\tilde{p}_A, \tilde{p}_B, \tilde{p}_C, \tilde{p}_D, 1-\tilde{p}_A-\tilde{p}_B-\tilde{p}_C-\tilde{p}_D)$ is jointly Dirichlet distributed with parameter vector

$$(\theta p_A-1, \theta p_B-1, \theta p_C-1, \theta p_D-1, \\ \theta(1-p_A-p_B-p_C-p_D)-1), \qquad (9)$$

and $\theta = (1-F)/F$. Although our derivation does not start from the Dirichlet assumption, it would in any case be a natural family of distributions to consider for modelling uncertainty about relative frequencies.

4.2 Statistical derivation

Equation (6) was interpreted in terms of a specific evolutionary model which may not be accurate for VNTR loci. However, it follows from the results of Lindley (1990) that equation (6) is more general than the above derivation suggests. Sufficient conditions for (6) are that $\tilde{p}_A = p_A$ whenever p_A is either zero or one and that both the expectation and variance of $\tilde{p}_A$ given $p_A, p_B, p_C, \ldots$, are twice differentiable functions of p_A and do not depend on $p_B, p_C, \ldots$. In addition to (6), it follows immediately from these assumptions that

$$\mathrm{E}(\tilde{p}_A|p_A, p_B, p_C, \ldots) = \mathrm{E}(\tilde{p}_A|p_A) = p_A.$$

We henceforth suppress the explicit conditioning in the moments and write, for example, $\mathrm{E}(\tilde{p}_A)$ in place of $\mathrm{E}(\tilde{p}_A|p_A)$.

We now extend Lindley's argument to derive (8) for $r+s+t+u \leq 4$. In addition to the assumptions in the previous paragraph, we require that the third, fourth and fifth moments of $\tilde{p}_A$ are each sufficiently differentiable functions of p_A. Further, we assume that joint moments of $\tilde{p}_A$ and $\tilde{p}_B$ are differentiable functions of p_A and p_B, and similarly for more complicated joint moments up to order five. Finally, we assume that the even and odd central moments of $\tilde{p}_A$ are functions of p_A which are, respectively, symmetric and anti-symmetric about $p_A = 1/2$. This assumption is natural because of the arbitrary labelling of the alleles.

Let $h(p_A) = \mathrm{E}((\tilde{p}_A-p_A)^3)$. Then

$$\begin{aligned} &h(p_A+p_B+p_C+p_D) = \\ &h(p_A+p_B+p_C) + h(p_A+p_B+p_D) \\ &+h(p_A+p_C+p_D) + h(p_B+p_C+p_D) \\ &-h(p_A+p_B) - h(p_A+p_C) - h(p_A+p_D) \\ &-h(p_B+p_C) - h(p_B+p_D) - h(p_C+p_D) \\ &+h(p_A) + h(p_B) + h(p_C) + h(p_D). \end{aligned} \qquad (10)$$

Since, by the assumptions above, no term on the RHS of (10) is a function of each of p_A, p_B, p_C and p_D, we have

$$\frac{\partial^4}{\partial p_A \partial p_B \partial p_C \partial p_D} h(p_A+p_B+p_C+p_D) = \frac{\partial^4}{\partial p_A^4} h(p_A) = 0,$$

so that $h(p_A)$ is a polynomial of degree three in p_A. The boundary conditions imply that this polynomial has roots at 0 and 1 and hence p_A and $(1-p_A)$ are both factors. By symmetry, the third factor must be $(1-2p_A)$ and hence

$$h(p_A) = \kappa p_A(1-p_A)(1-2p_A), \qquad (11)$$

for some constant κ. Expanding (11) and substituting from (6) we obtain

$$\mathrm{E}(\tilde{p}_A^3) = \kappa p_A + 3(F-\kappa)p_A^2 + (1-3F+2\kappa)p_A^3.$$

In order to assign the value of κ, we note that

$$\frac{\partial}{\partial p_A}\mathrm{E}(\tilde{p}_A^3)\,|_{p_A=0} = \kappa. \qquad (12)$$

The LHS can be interpreted as the probability that three alleles are of the same type in the limit as the number of distinct alleles increases and the relative frequency of mutation to each allele vanishes in an isolated randomly-mating population. Thus κ should agree with the value for the probability that three random alleles are of the same type given by the Ewens Sampling Formula (Ewens, 1979, equation (3.76)) for the infinite alleles model (in which every mutation is to a distinct type). This gives $\kappa = 2F^2/(1+F)$ and (7) follows.

The other third-order moments follow from (7). For example,

$$\frac{\partial^3}{\partial p_A \partial p_B \partial p_C} \mathrm{E}((\tilde{p}_A+\tilde{p}_B+\tilde{p}_C)^3) = \frac{\partial^3}{\partial p_A^3} \mathrm{E}(\tilde{p}_A^3) = 6\frac{(1-F)^2}{1+F}. \quad (13)$$

Expanding the LHS of (13), the only term which is a function of each of p_A, p_B and p_C, and hence does not vanish in the differentiation, is $6\mathrm{E}(\tilde{p}_A\tilde{p}_B\tilde{p}_C)$ and thus

$$\frac{\partial^3}{\partial p_A \partial p_B \partial p_C} E(\tilde{p}_A\tilde{p}_B\tilde{p}_C) = \frac{(1-F)^2}{1+F}. \quad (14)$$

Using (14) and the boundary conditions, it follows that

$$\mathrm{E}(\tilde{p}_A\tilde{p}_B\tilde{p}_C) = \frac{(1-F)^2}{1+F} p_A p_B p_C.$$

Turning now to fourth order moments, $\mathrm{E}((\tilde{p}_A - p_A)^4)$ can similarly be shown to be a polynomial of degree four in p_A which, from the boundary conditions and symmetry, is of the form

$$\mathrm{E}((\tilde{p}_A-p_A)^4) = p_A(1-p_A)(\kappa + \lambda p_A(1-p_A)),$$

for some constants κ and λ. Invoking again the Ewens Sampling Formula we have

$$\frac{\partial}{\partial p_A} \mathrm{E}(\tilde{p}_A^4) \mid_{p_A=0} = \kappa = \frac{6F^3}{(1+F)(1+2F)}.$$

To obtain a value for λ, we argue as at (13) and (14) that

$$\frac{1}{24}\frac{\partial^4}{\partial p_A^4} \mathrm{E}(\tilde{p}_A^4) = \frac{\partial^4}{\partial p_A \partial p_B \partial p_C \partial p_D} \mathrm{E}(\tilde{p}_A\tilde{p}_B\tilde{p}_C\tilde{p}_D) = \frac{(1-F)^3}{(1+F)(1+2F)}, \quad (15)$$

the final expression being the probability that four randomly-drawn alleles are distinct in an infinite-alleles model.

Similarly for fifth-order moments, we proceed from the observation that $\mathrm{E}((\tilde{p}_A-p_A)^5)$ is a polynomial of degree five which, from the boundary conditions and symmetry, is of the form

$$\mathrm{E}((\tilde{p}_A-p_A)^5) = p_A(1-p_A)(1-2p_A)(\kappa+\lambda p_A(1-p_A)).$$

We omit the further details.

4.3 Discussion

The substantive assumption in the derivation of Section 4.2 is that, given p_A, moments of $\tilde{p}_A$ up to order five are conditionally independent of $p_B, p_C, \ldots$, and similarly for joint moments. If nothing were known about mutation, the value of p_B, for example, might be informative about it and hence about $\tilde{p}_A$. This dependence may, however, be unimportant given partial knowledge about mutation. Similarly, p_B may be informative about the genealogy of the whole population, but this may also be unimportant for the very large population sizes of the major racial groups into which forensic databases are usually classified.

Equations (1) and (2) are appropriate for a range of genetic typing systems. The most common such system, based on VNTR loci, is problematic because of their complicated evolution. For example, mutation events at VNTR loci frequently generate new alleles of a similar length to the progenitor allele (Jeffreys *et al.*, 1988). This constraint on mutation seems to produce patterns that persist over evolutionary time scales. Waye and Eng (1994) investigated a thalassaemia deletion linked to a VNTR locus. There was an atypical set of VNTR allele lengths on haplotypes that bore the deletion. These alleles were rare on other chromosomes, and had a narrow distribution of lengths. Presumably the newly arisen deletion (or one of its early descendants) bore a rare VNTR allele, and its present day descendants have a narrow range of lengths produced by the subsequent mutations.

Even the more general assumptions of Section 4.2 may fail to encompass exactly the complicated behaviour of VNTR loci. In particular, as a consequence of the mutation process, the frequencies of alleles of similar lengths are positively correlated (Nichols & Balding, 1991) and this effect is not accounted for in (2). When more detailed knowledge of VNTR evolution becomes available, it may prove possible to improve the proposed method to include such length dependent correlations, possibly by using only one

additional parameter. Given present knowledge, however, we believe that the proposed method captures the primary effects of coancestry and other sources of uncertainty. In court, the single parameter has proved a common currency in which experts can attempt to quantify their disagreement. The calculations presented here have been used to assess the consequences for the evaluation of the DNA evidence.

Acknowledgements

We are grateful to Wilson Wall and Matthew Greenhalgh for providing access to the data for Figure 1 and Peter Donnelly for valuable discussions. Work supported in part by the UK Science and Engineering Research Council under grants GR/F 98727 (DJB) and GR/G 11101 (RAN).

References

Balding, D.J. & P. Donnelly, 1995. Inference in forensic identification. To appear J. Roy. Statist. Soc. 158.

Balding, D.J. & R.A. Nichols, 1994. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci. Inter. 64: 125-140.

Cavalli-Sforza, L.L. & A. Piazza, 1993. Human genomic diversity in Europe: a summary of recent research and prospects for the future. Eur. J. Hum. Genet. 1: 3-18.

Chakraborty, R. & K.K. Kidd, 1991. The utility of DNA typing in forensic work. Science 254: 1735-1739.

Cockerham, C.C., 1971. Higher order probability functions of identity of alleles by descent. Genetics 69: 235-246.

Crow, J.F. & M. Kimura, 1970. An Introduction to Population Genetics Theory. New York: Harper and Row.

Devlin, B., N. Risch & K. Roeder, 1993. Statistical evaluation of DNA fingerprinting: a critique of the NRC's report. Science 259: 748, 749, 837.

Donnelly, P., 1995. The non-independence of matches at different loci in DNA profiles: quantifying the effect of close relatives on the match probability. (to appear) Heredity.

Ewens, W.J., 1979. Mathematical Population Genetics. Berlin: Springer-Verlag.

Geisser, S. & W. Johnson, 1993. Testing independence of fragment lengths within VNTR loci. Am. J. Hum. Genet. 53: 1103-1106.

Harding, R.M., 1992. VNTRs in review. Evol. Anthrop. 1: 62-71.

Jeffreys, A.J., N.J. Royale, V. Wilson & Z. Wong, 1988. Spontaneous mutation rates to new alleles at tandem repetitive hypervariable loci in human DNA. Nature 332: 278-281.

Jeffreys, A.J., K. Tamaki, A. MacLeod, D.G. Monckton, D.L. Neil & J.A.L. Armour, 1994. Complex gene conversion events in germline mutation at human minisatellites. Nature Genetics 6: 136-145.

Krane, D.E., R.W. Allen, S.A. Sawyer, D.A. Petrov & D.L. Hartl, 1992. Genetic differences at four DNA typing loci in Finnish, Italian and mixed Caucasian populations. Proc. Nat. Acad. Sci. USA 89: 10583-10587.

Lempert, R., 1991. Some caveats concerning DNA as criminal identification evidence: with thanks to the Reverend Bayes. Cardozo Law Rev. 13: 303-341.

Lewontin, R.C. & D.L. Hartl, 1991. Population genetics in forensic DNA typing. Science 254: 1745-1750.

Lindley, D.V., 1990. The present position in Bayesian statistics. Statist. Sci. 5: 44-89.

Morton, N.E., 1992. Genetic structure of forensic populations. Proc. Nat. Acad. Sci. USA 89: 2556-2560.

Morton, N.E., 1993a. DNA in court. Eur. J. Hum. Genet. 1: 172-178.

Morton, N.E., 1993b. Kinship bioassay on hypervariable loci in blacks and caucasians. Proc. Nat. Acad. Sci. USA 90: 1892-1896.

Nichols, R.A. & D.J. Balding, 1991. Effects of population structure on DNA fingerprint analysis in forensic science. Heredity 66: 297-302.

Robertson, B. & T. Vignaux, 1992. Why the NRC report on DNA is wrong. New Law J.: 1619-1621.

Roeder, K., 1994. DNA fingerprinting: a review of the controversy. Statist. Sci. 9: 222-278.

Wall, W.J., R. Williamson, M. Petrou, D. Papaioannou & B.H. Parkin, 1993. Variation of short tandem repeats within and between populations. Hum. Molec. Genet. 2: 1023-1029.

Waye, J.S. & B. Eng, 1994. Allelic stability of a VNTR locus 3'αHVR: Linkage disequilibrium with the common α-thalassaemia-1 deletion of South-East Asia (–SEA/). Hum. Heredity 44: 61-67.

Weir, B.S., 1993a. Forensic Population Genetics and the National Research Council (NRC). Am. J. Hum. Genet. 52: 437-440.

Weir, B.S., 1993b. Independence tests for VNTR alleles defined as quantile bins. Am. J. Hum. Genet. 53: 1107-1113.

Yasuda, N., 1968. An extension of Wahlund's principle to evaluate mating type frequency. Am. J. Hum. Genet. 20: 1-23.

Note added in proof

Subsequent work (M. Greenhalgh, pers. comm.) suggests that the data from the CD4 locus (Fig. 1) may be affected by laboratory error. Recent data continue to indicate different values of F_{ST} at different loci, in some cases showing as much variation as at traditional loci.

Editor's comments

The authors' work offers a sound approach to accommodating the effects of population structure, based on use of Wright's F_{ST}. Their equations 1 and 2 are very convenient, and are good approximations to the exact results given by Weir (1994). As they point out, good estimates of F_{ST} are needed. The comments about the 'generally mixed' results of independence tests may be met, in part, by the paper of Maiste and Weir in this volume. The authors cite Krane *et al.* (1992) but had not seen the subsequent rebuttal by Budowle *et al.* (1994). The work of Wall *et al.* (1993) contained errors, as noted in Greenhalgh *et al.* (1994).