Int J Legal Med (1997) 110 : 244–250

ORIGINAL ARTICLE

L. A. Foreman · A. F. M. Smith · I. W. Evett

A Bayesian approach to validating STR multiplex databases for use in forensic casework

Received: 18 November 1996 / Received in revised form: 21 April 1997

Abstract A previous paper in this journal has described the conventional statistical analysis of three databases (Caucasian, Afro-Caribbean and Asians from the Indian subcontinent) where individuals are typed at six short tandem repeat (STR) loci. This paper presents a Bayesian analysis of the same data and the approach is centred on the concept of estimating coancestry coefficients from mixed databases. Posterior distributions for the three databases are presented and discussed and the consequences of implementing bootstrap estimation procedures are also shown.

Key words DNA profiling · Forensic identification · Bayesian inference · Likelihood ratio · Coancestry coefficient · Probability · Statistics · PCR · STR

Introduction

Evett et al. (1997) described the conventional statistical analysis of data from 602 Caucasians, 190 Afro-Caribbeans and 257 Asians of Indo/Pakistani descent using samples analysed with the new six-locus multiplex short tandem repeat (STR) profiling system described by Kimpton et al. (1994). The authors of the former paper expressed reservations with regard to the usefulness of conventional significance testing within the context of forensic science practice. Those reservations have been discussed in greater detail elsewhere by Evett (1996) and Evett and Buckleton (1996). In particular, the exact test (Zaykin et al. 1995) applied to data from six loci leads to 63 significance tests. Even if the idealised null hypotheses of perfect within- and between-locus independence were true – which, of course, they cannot be – a proportion of significance tests are expected to fail by chance alone. Attempts to rectify this by means of Bonferroni corrections (for multiple comparisons) merely serve to reduce the power of the testing regime and provide no useful insight as to whether effects, real or artefactual, have any meaningful impact on forensic procedures.

Departures from equilibrium conditions will be caused by population substructure and so any successful method should be applied directly to estimation of the magnitude of such phenomena, in particular through the estimation of θ, the coancestry coefficient. Classical methods of doing this, as described for example by Weir and Cockerham (1984), require that data be collected from the relevant identified subpopulations. In forensic science, however, such data are not in general available. Instead, a *convenience sample* will have been drawn from the full heterogeneous population and there is no meaningful way of assigning individuals in the sample to discrete component subpopulations. Foreman et al. (1997) showed how, in spite of this difficulty, suitable probability distributions could be derived for STR quadruplex data by developing and extending the work of Roeder et al. (1997) and Balding and Nichols (1995).

In summary, the Bayesian approach starts with a *prior* probability distribution for θ, $p(\theta)$, which is modified by a likelihood function calculated from the data, $p(DATA \mid \theta)$, yielding a *posterior* distribution, $p(\theta \mid DATA)$. The prior distribution represents our beliefs about the value of θ prior to observing any data. A flat distribution essentially corresponds to "ignorance" while a highly peaked distribution concentrated on some interval [a, b] indicates very strong prior beliefs that $\theta \in$ [a,b]. The likelihood simply describes the behaviour of the data (in our case, observed allele pairs) in terms of θ. It tends to highlight those values of θ which are supported most by the data. The process of updating our prior beliefs in the presence of further information contained in the data is then achieved via *Bayes' theorem*:

L. A. Foreman · A. F. M. Smith
Department of Mathematics,
Imperial College of Science, Technology and Medicine,
Huxley Building, 180 Queen's Gate, London SW7 2BZ, UK

I. W. Evett (✉)
Forensic Science Service, Metropolitan Laboratory,
109 Lambeth Road, London SE1 7LP, UK
FAX: +44 (171) 230 6338

$$p(\theta \mid DATA) \propto p(\theta) \times p(DATA \mid \theta) \quad (1)$$

The resulting posterior distribution represents a weighted combination of the prior information and the data; in fact, the more data are available, the more the likelihood tends to dominate the prior. The posterior value of any function, $f(\theta)$, may then be obtained by performing the mathematical operation of integration with respect to the posterior distribution. For further details on the theory of Bayesian inference, see Bernardo and Smith (1994).

In this paper, we adopt the generic notation **x** to represent the STR profile of an individual typed at six loci labelled as 1. D18, 2. D21, 3. THO1, 4. D8, 5. FGA and 6. VWA. The current datasets are each composed of n profiles, $D = \{\mathbf{x}_1^d, \mathbf{x}_2^d, \ldots, \mathbf{x}_n^d\}$, where n is 602 (Caucasian), 190 (Afro-Caribbean) and 257 (Asian). Based on these data, empirical estimates for the set of allele distributions, γ, exhibited in each racial group may be obtained using the appropriate database allele proportions. These are presented in the form of simple "look-up" tables in Evett et al. (1997).

We denote by $\mathbf{x}^0$ the STR profile of the offender (O) which is recovered from the crime scene and by $\mathbf{x}^s$ the STR profile of the suspect accused of the crime (S). Then, if the profiles match and **x** denotes the shared profile, the likelihood ratio (LR) representing the strength of evidence in support of the hypothesis that the suspect and offender are the same person is given by

$$LR(\mathrm{x}) = \frac{1}{p\left(\mathbf{x}^0 = \mathbf{x} \middle| \mathbf{x}^s = \mathbf{x}, S \neq O\right)}.$$

The denominator is referred to as the conditional match probability and represents how likely it is that an individual exhibits the profile **x**, given that another individual does. Thus, the match probability will vary according to how the suspect and the offender are genetically related when they are not the same person. Foreman et al. (1997) address the two main cases which are again considered here for the new STR data:

(a) actual offender belongs to the same subpopulation of the suspect's racial group,
(b) actual offender belongs to another subpopulation in the same/different racial group as the suspect.

The match probability of **x** under scenario (a) may be computed using a formula derived for a single-locus genotype by Balding and Nichols (1994, 1995):

$$\begin{cases} \dfrac{[2\theta + (1-\theta)\gamma_i][3\theta + (1-\theta)\gamma_i]}{(1+\theta)(1+2\theta)}, & \text{for the homozygote } (A_i, A_i) \\ \dfrac{2[\theta + (1-\theta)\gamma_i][\theta + (1-\theta)\gamma_j]}{(1+\theta)(1+2\theta)}, & \text{for the heterozygote } (A_i, A_j) \end{cases}$$

This is expressed in terms of the subpopulation coancestry coefficient, θ, and allele proportions, γ_i and γ_j, exhibited in the suspect's racial group. The conditional match probability for the full profile **x** is then obtained by multiplying terms such as (2) across all six loci. Under scenario (b), where the offender and suspect are effectively unrelated, the conditional match probability reduces to the proportion of the profile **x** in the racial group of interest and may be computed using the product rule; i.e. by effectively multiplying together component allele proportions. This is equivalent to substituting $\theta = 0$ in (2) for all loci.

One issue which may be important when presenting DNA evidence in court relates to the idea of "conservativeness"; i.e. do the methods of LR calculation tend to yield figures which err in favour of the suspect? It can be shown that, for all practical realisations of the allele distributions γ, the match probability expression in (2) is an increasing function of θ given γ, resulting in more conservative values for larger θ. Therefore, assuming the suspect and offender to belong to the *same* subpopulation of a racial group P (where $\theta > 0$ and (2) is used) is more conservative than assuming they belong to a *different* subpopulation in P (where $\theta = 0$ and the product rule is implemented), since the former will yield larger match probabilities and, hence, lower LR values. Thus, excluding the special case where the suspect and offender are close blood relatives, which will usually be dealt with separately, it is common practice at the Forensic Science Service to calculate the LR under scenario (a) for the offender's racial group when known; otherwise, the minimum value across all racial groups is reported.

Inference for the coancestry coefficient, θ

We concentrate on evaluating LRs in case (a) where the suspect and offender, if not the same person, are assumed to belong to the same subpopulation of the racial group P. Since individuals are classified solely by broad racial group and information at subpopulation level is unavailable, it is necessary to implement Equation (2) to evaluate the match probability in this case. The set of allele distributions, γ, exhibited in P will typically be estimated by $\hat{\gamma}$; i.e. the allele proportion estimates as observed in the augmented dataset D', where the suspect's profile is added twice to the observed database D (see Section 3 of Foreman et al. (1997) for the Bayesian explanation). Thus, we are reduced to drawing inference about $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_6)$, the set of coancestry coefficients, one for each locus, corresponding to individuals in the subpopulation of interest with respect to $\hat{\gamma}$ in P.

Following Foreman et al. (1997), our analyses for $\boldsymbol{\theta}$ are conducted "off-line" without knowledge of the particular matching profile **x** and, in this case, inference is based only on the observed dataset D. Under a Bayesian approach, we make inference about the unknown quantity $\boldsymbol{\theta}$ via the posterior distribution $p(\boldsymbol{\theta} \mid D, \hat{\gamma})$. The likelihood adopted here is termed the profile-product likelihood in our earlier paper and the quantity $\boldsymbol{\theta}$ in this case may be interpreted either as a combined coancestry measure across all subpopulations at the finest level of subdivision of P (Foreman et al. 1997) or as a measure of excess homozygosity in the racial group (Roeder et al. 1997). Further-

Table 1 Prior and posterior distribution summaries for $\boldsymbol{\theta}$ under Beta(1.5,50) priors when the θ_j's are modelled independently

Locus	5th percentile	Median	Mean	75th percentile	95th percentile
Prior					
All loci	0.0035	0.0233	0.0291	0.0401	0.0748
Caucasian					
D18	0.0016	0.0100	0.0116	0.0161	0.0270
D21	0.0015	0.0087	0.0103	0.0142	0.0245
THO1	0.0019	0.0114	0.0135	0.0187	0.0326
D8	0.0014	0.0087	0.0104	0.0144	0.0253
FGA	0.0024	0.0125	0.0141	0.0192	0.0306
VWA	0.0047	0.0204	0.0219	0.0295	0.0444
Average			0.0136		
Afro-Caribbean					
D18	0.0032	0.0182	0.0210	0.0291	0.0484
D21	0.0028	0.0167	0.0194	0.0268	0.0455
THO1	0.0018	0.0111	0.0136	0.0188	0.0342
D8	0.0016	0.0108	0.0132	0.0183	0.0330
FGA	0.0044	0.0232	0.0259	0.0355	0.0566
VWA	0.0048	0.0271	0.0305	0.0421	0.0677
Average			0.0206		
Asian					
D18	0.0024	0.0152	0.0180	0.0249	0.0430
D21	0.0020	0.0122	0.0146	0.0203	0.0353
THO1	0.0026	0.0159	0.0189	0.0263	0.0457
D8	0.0035	0.0203	0.0229	0.0315	0.0515
FGA	0.0041	0.0201	0.0224	0.0305	0.0482
VWA	0.0036	0.0206	0.0234	0.0324	0.0529
Average			0.0200		

more, independent Beta(1.5,50) priors are adopted for θ_j at each locus. This corresponds to vague prior beliefs that each θ_j is likely to be less than 0.1 since values higher than this tend to be associated with close blood relations; e.g. $\theta = 0.125$ corresponds to coancestry between uncle and niece. The Beta(1.5,50) distribution is centred about a median value of 0.023 and a mean of 0.029 so our prior beliefs are that θ is much higher than suggested by other analyses of STR data (e.g. Gill and Evett 1995). This may therefore be considered a "cautious" choice of prior since it is biased in such a way as to be non-prejudicial to the suspect. The form of the posterior distribution, $p(\boldsymbol{\theta} \mid D,\hat{\gamma})$, resulting from the application of Bayes' theorem in (1), is too complicated to allow evaluation of match probabilities directly by integration. However, the Gibbs sampler is a powerful stochastic simulation technique, belonging to the general family of Markov chain Monte Carlo methods (Smith and Roberts 1993), which may be employed to generate iteratively an approximate sample of $\boldsymbol{\theta}$ values from $p(\boldsymbol{\theta} \mid D,\hat{\gamma})$. The value of any posterior quantity of interest (in particular, the match probability of **x**) may then be estimated by finding an average over this sample. The Gibbs sampling strategy which applies here is outlined in the Appendix of Foreman et al. (1997).

Table 1 summarises the prior and posterior distributions for $\boldsymbol{\theta}$ in each racial group based on a Gibbs sampler run of $T = 30{,}000$ iterations. Figures 1–3 illustrate further by comparing prior and posterior densities for θ at locus D21. It can be seen that the posterior distribution in each case is more peaked and concentrated about lower values than originally believed and specified in the prior. Furthermore, the larger the database, the more peaked the posterior distribution which clearly illustrates the effect of

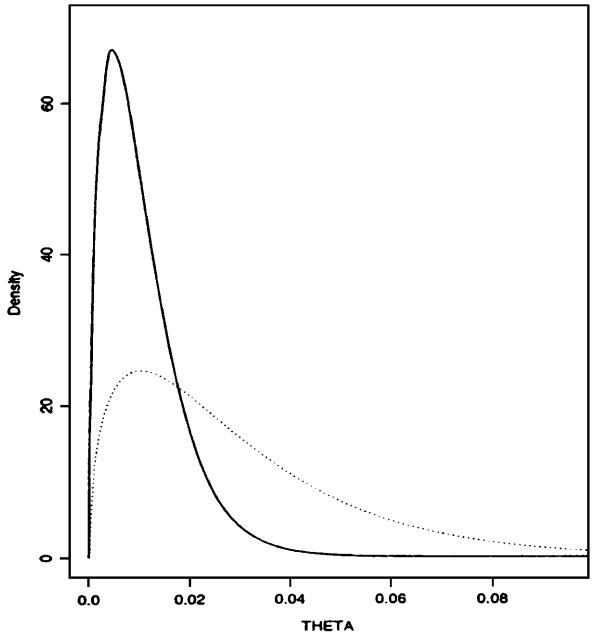

Fig. 1 Caucasian: comparison of the prior (· · ·) and posterior (—) density for θ at locus D21

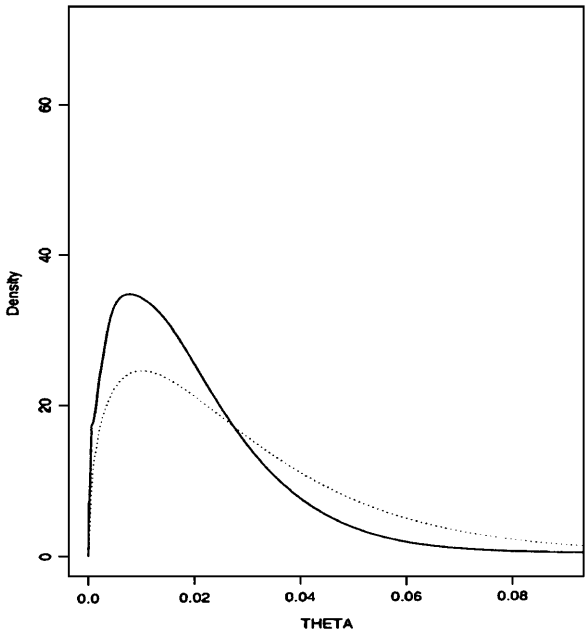

Fig. 2 Afro-Caribbean: comparison of the prior (· · ·) and posterior (—) density for θ at locus D21

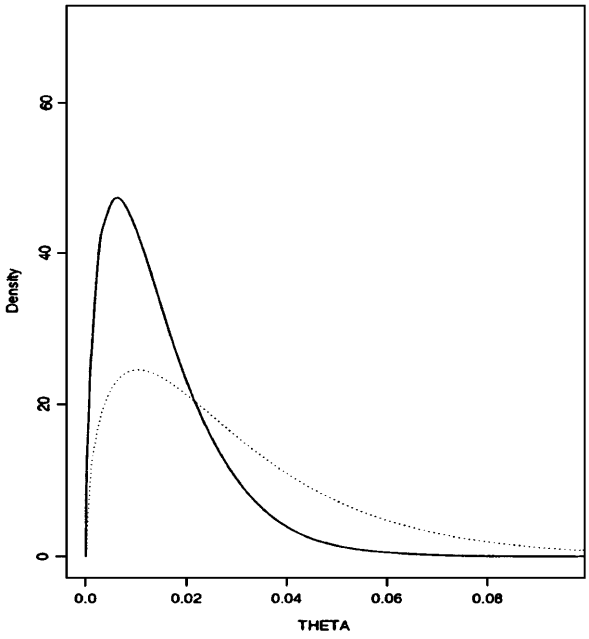

Fig. 3 Asian: comparison of the prior (· · ·) and posterior (—) density for θ at locus D21

the data in the learning process; i.e. in strengthening posterior beliefs. We note that locus VWA for the Afro-Caribbean group is the only case where the posterior distribution is located about higher values than the prior, although the difference is minimal. This observation seems to support our claim that the choice of Beta(1.5,50) priors is a cautious one in the sense that the data, through the likelihood, are indicating smaller $\boldsymbol{\theta}$ values. The magnitude and variation of the θ_j's across loci may offer an explanation for significant test results obtained in Evett et al. (1997). For example, we observed relatively large θ values for loci FGA and VWA in the Afro-Caribbean group (Table 1), the same loci which were responsible for the failure of homozygosity and exact tests in Evett et al. (1997). This confirms that the presence of a high level of coancestry within subpopulations is the main factor governing departures from allele independence assumptions. As in our earlier paper, we use mean values to summarise the posterior distribution of θ_j at each locus j and, from Table 1, these are approximately 0.015 in the Caucasian group and 0.02 in the Afro-Caribbean and Asian groups, averaging across loci. By definition, we would expect the most likely values of $\boldsymbol{\theta}$ to be higher than F_{ST} values estimated using classical procedures and commonly appearing in the literature; e.g. Gill and Evett (1995). However, this is confounded by the effects of database size; i.e. for the relatively small datasets analysed here, the likelihood does not have the power, in terms of volume of data, to overwhelm the cautious prior information. This may account, in part, for the fact that posterior values in the Caucasian group tend to be lower than for the other two groups. However, there are grounds for expecting higher coancestry levels within the Asian group in particular, due to cultural and social factors governing marriages and, thus, imposing a pattern of preferred mating within certain subgroups.

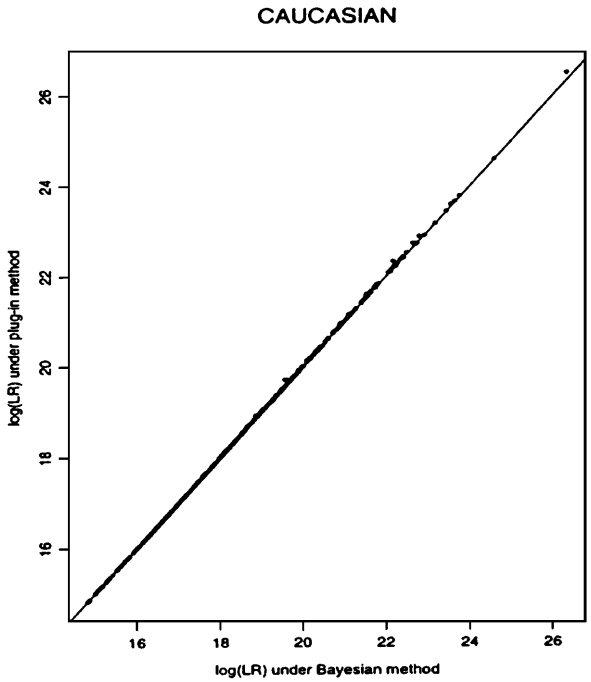

Fig. 4 Caucasian: scatterplots comparing $log(LR)$'s under the Bayesian and plug-in methods

The output sample of $\boldsymbol{\theta}$ values from the Gibbs sampler may be used to evaluate match probabilities. The fully Bayesian method involves integration of the match probability expression, using Equation (2), with respect to the posterior distribution $p(\boldsymbol{\theta} \mid D,\hat{\gamma})$. This may be estimated by

Table 2 Comparison of the average θ_j value, $\bar{\theta}$, for each racial group under various priors

	Caucasian	Afro-Caribbean	Asian
Prior			
Uniform on [0,0.05]	0.0142	0.0209	0.0211
Beta(1.5,50)	0.0136	0.0206	0.0200
Beta(1.5,100)	0.0102	0.0132	0.0131

averaging the match probability expression over the sample of values for $\boldsymbol{\theta}$ generated from the posterior distribution via the Gibbs sampler. The "plug-in" method provides an efficient approximation to full integration by simply substituting an estimate, $\tilde{\boldsymbol{\theta}}$, for $\boldsymbol{\theta}$ in the match probability formula. We may compute the LR associated with a match for each profile in the database of all three racial groups under both methods. Figure 4 presents a scatterplot comparing "*log*(*LR*) under the Bayesian method" versus "*log*(*LR*) under the plug-in method with posterior mean estimates at each locus, $\tilde{\boldsymbol{\theta}}$" for each profile in the Caucasian database. To aid comparison, the "$x = y$" line is also plotted, along which exact agreement exists between the two methods. In fact, there is very close agreement for most of the observed profiles when mean θ estimates are substituted in the plug-in method. Similar plots were obtained for the other two racial groups. When the θ_j's are modelled independently, therefore, the above analyses support the case for adopting the mean value of θ_j averaged across loci, $\bar{\theta}$, as a single summary measure of the full posterior distribution for $\boldsymbol{\theta}$ representative of the general degree of substructure exhibited within a racial group. This facilitates comparisons between populations and with values obtained in other studies.

In order to illustrate the effect of adopting alternative priors, we repeated the analyses under a flat uniform prior restricted to the interval [0,0.05] and a Beta(1.5,100) prior. The former is more cautious than a Beta(1.5,50) prior, reflecting prior ignorance, and the latter is less cautious, exhibiting a mean value of 0.0148 compared with a mean of 0.0291 under the Beta(1.5,50) distribution. It was found that adopting the less cautious and tighter prior distributions tends to yield tighter posterior distributions with shorter tails and located about lower values. This is illustrated in Table 2 which presents the single summary value, $\bar{\theta}$, for each racial group under each of the three priors discussed here. It can be seen that both the constrained uniform prior and the Beta(1.5,50) prior yield similar results, suggesting that the latter prior can be considered reasonably non-informative. In the final case, where much stronger prior knowledge of small $\boldsymbol{\theta}$ is assumed, $\bar{\theta}$ values are much lower. Thus, it is clear that the choice of prior distribution has a non-negligible effect on the posterior distributions of $\boldsymbol{\theta}$ when studying the small databases currently available for this STR system.

Until now, we have assumed the θ_j's to be independent a priori. However, in the absence of disturbing forces, all the θ_j will be equal. In reality, they are likely to originate from the same value and, apart from the effects of differing mutation rates, they undergo the same evolutionary process. Therefore, it might be considered more appropriate to assume the θ_j's to be constant across loci, equal to θ, say. In this case, a similar Gibbs sampling strategy may be applied. The resulting posterior distribution summaries of θ are presented in Table 3. By comparison of Tables 1 and 3, we can see that the posterior mean of θ for each racial group is approximately half the corresponding value for $\bar{\theta}$ when the θ_j's are modelled independently. This must be, in part, due to the fact that all the data across loci are employed to draw inference about the single quantity θ associated with each racial group. Thus, 6 times more "estimating data" are available in this case to dominate the cautious prior information and yield posterior distributions closer to zero than those given in Table 1.

Dealing with the effect of small data samples on calculations

There is no simple answer to the question "how large should my database be?". In practice, the size will emerge from compromise, with the availability of samples a major factor. In the previous paper, we dealt with larger databases than are covered by the present study and one issue to be addressed is the sensitivity of the posterior distribution to sampling variation. Of course, results will vary from one observed dataset to another, to a greater extent when the sample sizes are smaller. The effect of small-sample variation on γ estimates will be minimal once the suspect's profile, $\mathbf{x}^s = \mathbf{x}$, has been added twice to the database, since this ensures a conservative correction (often termed the *size-bias correction*). In the case of the co-ancestry coefficients, it has already been demonstrated that the Beta(1.5,50) priors we are adopting are quite cautious so we would expect to obtain larger $\boldsymbol{\theta}$ values (and, thus, smaller *LR* values) than would be observed under larger data samples. However, one option for taking account of sample variation is to adopt $\boldsymbol{\theta}$ estimates representing some "upper bound" on posterior mean values as

Table 3 Prior and posterior distribution summaries for θ under Beta(1.5,50) priors when the θ_j's are assumed constant across loci, equal to θ

	5th percentile	Median	Mean	75th percentile	95th percentile
Prior	0.0035	0.0233	0.0291	0.0401	0.0748
Caucasian	0.0011	0.0059	0.0065	0.0090	0.0141
Afro-Caribbean	0.0019	0.0098	0.0110	0.0150	0.0242
Asian	0.0025	0.0118	0.0127	0.0173	0.0262

Table 4 Summary of the "bootstrap" distribution of posterior means for $\boldsymbol{\theta}$ under Beta(1.5,50) priors when the θ_j's are modelled independently

Locus	5th percentile	Median	Mean	75th percentile	95th percentile
Caucasian					
D18	0.0071	0.0111	0.0121	0.0140	0.0202
D21	0.0066	0.0102	0.0109	0.0127	0.0174
THO1	0.0086	0.0132	0.0138	0.0158	0.0212
D8	0.0068	0.0101	0.0107	0.0121	0.0164
FGA	0.0081	0.0137	0.0147	0.0180	0.0243
VWA	0.0111	0.0213	0.0229	0.0277	0.0413
Average					0.0235
Afro-Caribbean					
D18	0.0130	0.0202	0.0221	0.0252	0.0380
D21	0.0121	0.0191	0.0205	0.0238	0.0337
THO1	0.0105	0.0135	0.0142	0.0157	0.0199
D8	0.0101	0.0130	0.0133	0.0144	0.0179
FGA	0.0151	0.0252	0.0272	0.0321	0.0462
VWA	0.0176	0.0293	0.0306	0.0359	0.0479
Average					0.0339
Asian					
D18	0.0117	0.0173	0.0188	0.0219	0.0307
D21	0.0095	0.0139	0.0149	0.0169	0.0232
THO1	0.0123	0.0179	0.0192	0.0221	0.0302
D8	0.0131	0.0215	0.0233	0.0272	0.0405
FGA	0.0123	0.0212	0.0225	0.0264	0.0371
VWA	0.0140	0.0220	0.0242	0.0283	0.0395
Average					0.0335

they vary across different small datasets of size n. This is in an attempt to avoid the suggestion in court that, had we observed a different dataset, the results would be very different, possibly to the advantage of the suspect.

Given only the single observed dataset, $D = \{\mathbf{x}_1^d, \mathbf{x}_2^d \ldots, \mathbf{x}_n^d\}$, from the racial group of interest, one procedure which is often adopted to deal with the problems of small-sample variation is the re-sampling method known as *bootstrapping* (Young 1994). This allows us to approximate the underlying variation in mean $\boldsymbol{\theta}$ values which would be observed in different datasets originating from the racial group. To explain briefly, any random sample of n profiles can be considered representative of the entire distribution of profiles underlying the racial group. An empirical estimate of this profile distribution simply assigns a probability of – to each of the n observed profiles. Thus, by uniform re-sampling (with replacement) of profiles from the original observed database, D, we may simulate a new dataset of n profiles, D^{new}, which is approximately drawn from the full racial group. A suitable re-sampling strategy may, thus, be described as follows:

(1) A sample of n profiles is drawn uniformly and with replacement from the original observed database, D, to yield a new dataset, $D^{new} = \{\mathbf{x}_1^{new}, \mathbf{x}_2^{new}, \ldots, \mathbf{x}_n^{new}\}$.
(2) Based on this new dataset, D^{new}, the associated set of allele distributions, $\hat{\gamma}^{new}$, is estimated and the Gibbs sampler is run for T iterations, under adoption of Beta(1.5,50) priors for the θ_j's. The posterior mean, $\bar{\theta}_j$, of each θ_j is recorded in $\boldsymbol{\theta}$.
(3) Steps (1) & (2) are repeated R times to obtain a sample of posterior means for $\boldsymbol{\theta}$ across different datasets, $\{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(R)}\}$.

The resulting distribution of posterior means, given by the sample recorded in Step (3) above, should approximately mimic the variation we would expect to see in $\boldsymbol{\theta}$ given different datasets of size n drawn from the entire racial group.

We implemented the above bootstrap re-sampling procedure for each racial group by re-sampling $R = 500$ new datasets of size n from the original observed dataset. For each of these, the Gibbs sampler was run for $T = 30{,}000$ iterations and posterior mean values of θ were recorded. Table 4 presents a summary of the distribution of posterior means across different bootstrap datasets. We can see that there exists a reasonable amount of variation within bootstrap distributions; e.g. by inspection of differences between 5th and 95th percentile values. Until further collection of data is possible and in an attempt to be cautious, the upper 95th percentile value may be adopted as a suitable estimate of $\boldsymbol{\theta}$ for substitution in the plug-in method. This corresponds to the adoption of θ values of approximately 0.02 (Caucasian), 0.03 (Afro-Caribbean) and 0.03 (Asian), on average. Adopting values of this magnitude is in agreement with recommendations made for PCR-based systems in the most recent report by the National Research Council (1996). When the variation in posterior means of $\boldsymbol{\theta}$ across different datasets falls below some preset "acceptable" level, we may then be justified in abandoning the bootstrap procedure.

We note that an alternative to bootstrapping is simply to adopt some upper percentile of the posterior distribution for $\boldsymbol{\theta}$ based on the single observed dataset D. For example, the upper posterior 95th percentile values for the θ_j's given in Table 1 are all higher than the corresponding 95th percentile values for the posterior means given in Table 4. If these are deemed too high, then a lower percentile may be considered more appropriate. In any case, under a Bayesian approach, all the uncertainty due to sampling variation is contained in the posterior distribution and, hence, we may avoid the need for lengthy bootstrap calculations in each case.

References

Balding DJ, Nichols RA (1994) DNA profile match probability calculations: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int 64: 125–140

Balding DJ, Nichols RA (1995) A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. Genetica 96: 3–12

Bernardo JM, Smith AFM (1994) Bayesian theory. John Wiley & Sons, Chichester

Evett IW (1996) Expert evidence and forensic misconceptions of the nature of exact science. Sci Justice 36: 118–122

Evett IW, Buckleton JS (1996) Statistical analysis of STR data. In: Carracedo A, Brinkmann B, Bär W (eds) Advances in forensic haemogenetics 6. Springer, Berlin Heidelberg New York, pp 79–86

Evett IW, Gill PD, Lambert JA, Oldroyd N, Frazier R, Watson S, Panchal S, Connolly A, Kimpton C (1997) Statistical analysis of data for three British ethnic groups from a new STR multiplex. Int J Legal Med 110: 5–9

Foreman LA, Smith AFM, Evett IW (1997) Bayesian analysis of DNA profiling data in forensic identification applications. J R Stat Soc A 160. In press

Gill P, Evett IW (1995) Population genetics of short tandem repeat (STR) loci. Genetica 96: 69–87

Kimpton C, Fisher D, Watson S, Adams M, Urquhart A, Lygo JE, Gill P (1994) Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. Int J Legal Med 106: 302–311

National Research Council (1996) The evaluation of forensic DNA evidence. National Academy Press, Washington DC

Roeder K, Escobar M, Kadane JB, Balazs I (1997) Measuring heterogeneity in forensic databases using hierarchical Bayes models. Biometrika. Submitted

Smith AFM, Roberts GO (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. J R Stat Soc B 55: 3–23

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. Evolution 38: 1358–1370

Young GA (1994) Bootstrap: more than a stab in the dark? Stat Sci 9: 382–415

Zaykin D, Zhivotovsky L, Weir BS (1995) Exact test for association between alleles at arbitrary numbers of loci. Genetica 96: 169–178