Int J Legal Med (1995) 108:8–13

ORIGINAL ARTICLE

James A. Lambert · John K. Scranage · Ian W. Evett

Large scale database experiments to assess the significance of matching DNA profiles

Received: 4 October 1994 / Received in revised form: 16 March 1995

Abstract Over 5,700 three-probe VNTR DNA profiles collected by several United Kingdom (UK) laboratories have been compared to examine the probability of randomly matching 2 samples from different individuals. In over 16 million comparisons, using a matching rule corresponding to the matching guideline employed by the UK Forensic Science Service, no profiles were found to match at the 3 loci D1S7 (MS1), D7S21(MS31) and D12S11 (MS43a). The frequency of occurrence of a set of Caucasian profiles have been estimated with 6 reference databases. The results show that there were greater differences in the frequency estimates when using a database of Afro-Caribbean or Asian profiles, rather than a different Caucasian database. The results further demonstrate the power and robustness of the VNTR DNA profiling technique for forensic casework.

Key words DNA · VNTR Databases · Between person comparisons · Match probabilities

Introduction

Conventional assumptions in the interpretation of single locus probe VNTR DNA profiles are those of between and within-probe independence. If population substructure were a serious problem, leading to significant departures from independence, certain VNTR genotype patterns might occur more often than independence would predict, and hence the probability that 2 unrelated individuals have a matching DNA profile could be higher than reported. A number of experiments have been performed to investigate the robustness of the assumption of independence. Risch and Devlin [1] studied data on over 6,000 profiles collected by the Federal Bureau of Investigation (FBI) and Lifecodes. The 2 organisations used different restriction enzymes to cut the DNA; the FBI used Hae III, which results in smaller fragments than those analysed by Lifecodes, which used Pst I. All the data were analysed by using a 2.4% matching rule. For the Lifecodes database the probability of a three-probe match ranged from about 1 in 6,000 for Caucasians to about 1 in 120,000 for Afro-Caribbeans. When considering all possible combinations of 3 probes out of 5 used in the FBI database, only one match was observed out of more than 7.6 million comparisons.

Krane et al. [2] compared match probabilities obtained from a large mixed American database with those obtained from small databases of Italians and Finns. Using a ± 2.5% match window they found that the probability of a three-probe match varied from about 1 in 300,000 for the mixed database, to about 1 in 1,200 for the Italians. The Italian database was anonymous and may have contained samples from close relatives. Conclusions relating to data collected at the HRAS1 locus have been the subject of critical comment by Devlin et al. [3].

A further large scale study was done by Herrin [4] on more than 3,700 DNA profiles obtained from 8 laboratories in the south east of the United States. Four different match criteria were used for the comparisons, ranging from 2.5% to 10% for DNA fragments with a calculated size of up to 10 kb, and from 5% to 20% for fragments of greater size. No four-probe matches were found in several million comparisons between different individuals when match criteria equivalent to those of the participating laboratories were used.

It is difficult to compare results with respect to the above studies for a number of reasons:

1. Variations in laboratory protocol, particularly with respect to choice of restriction enzyme.
2. Different loci are probed, resulting in different degrees of discrimination.
3. Different methods are used for assessing a match between 2 profiles.

J. A. Lambert (✉) · I. W. Evett
Service Development, Birmingham Laboratory,
The Forensic Science Service, Priory House,
Gooch Street North, Birmingham B5 6QQ, UK

J. K. Scranage
16 Gander Drive, Basingstoke, Hants, UK

4. The presence or absence of close relatives in the databases.
5. Different ethnic composition of databases.

For these reasons it is a useful exercise for any organisation which is involved in DNA profiling to carry out similar experiments on its own databases, in order to assess the assumption of independence.

In the UK extensive experiments have been performed to test the robustness of data used by the Forensic Science Service (FSS) and the Metropolitan Police Forensic Science Laboratory (MPFSL). Gill et al. [5] described the basic match/binning methodology used by the FSS. This involves a 2.8% match guideline and a corresponding 5.6% sliding window to provide a conservative estimate of database frequencies.

Evett and Gill [6] used a database containing duplicate analyses of 218 individuals to study the effects of sample size and population stratification. They performed within-person comparisons on the duplicate results and between-person comparisons on the results from different individuals. They concluded that the method described by Gill et al. [5] was sufficiently robust for use in forensic casework.

Gill et al. [7] performed pairwise comparison tests on FSS data consisting of results for 4 probes. The tests were repeated for each of the 3 major race codes, using a 2.8% matching rule. The observed number of two-probe matches agreed closely with the expected number under the assumption of independence. No three-probe or four-probe matches were observed between profiles from different individuals. Evett and Pinchin [8] utilised the same data sets as in [7] to study regional and racial variation and to assess the effect of such variation on casework conclusions. They also used a contingency table approach to test for within-probe independence. In 2 out of 12 analyses the results were statistically significant at the 5% level, although the consequences were shown to have negligible practical significance. The results provided further confirmation of the robustness of the FSS procedures.

An efficient one-stage continuous alternative to the discrete match/binning approach is described by Evett et al. [9, 10]. The calculation takes account of the closeness of the 2 sets of measurements, together with their rarity or commonness in the relevant population, and also allows for band shift, and for the variation in precision with molecular weight. The result of the calculation is expressed as a likelihood ratio (LR), a continuous function which is derived by considering the probability of the observed measurements under 2 competing hypotheses; either the 2 samples have the same source, or they have 2 different sources. The robustness of the method is demonstrated in a series of large scale experiments performed with data collected at the MPFSL. The second paper also includes a comparison of the continuous method with match/binning and with a modified ceiling frequency approach proposed by the National Research Council Committee on DNA Technology in Forensic Science [11].

The purpose of the present study is two-fold; firstly to examine the number of random matches between different individuals in a large database (experiment 1), and secondly to study the variation in the estimates of LRs obtained when using different UK databases for reference (experiment 2).

In this study data from several UK sources have been amalgamated to form a single large database. All possible between-person comparisons have been carried out to examine the number of chance matches occurring between profiles of different individuals. Most of the previous experiments described above have been designed to test for within-probe independence and for pairwise statistical independence between 2 probes. As matches involving even 2 probes are rare, large quantities of data are required to look for three-probe or four-probe matches.

Materials and methods

The data used for experiment 1 are shown in Table 1. Included in these data are profiles from individuals of Caucasian, Afro-Caribbean and Asian origin. The restriction enzyme HinfI was used on all samples. The three-probe profiles consist of data for D1S7 (MS1), D7S21 (MS31) and D12S11 (MS43a), the four-probe profiles also include D2S44 (YNH24) data. Thus for example, 52 of the 143 profiles from the Strathclyde Police Forensic Science Laboratory include data for the probe YNH24.

The number of profiles in the reference databases used for experiment 2 are shown in Table 2. These databases are a subset of the databases used for experiment 1. The FSS Caucasian database included 202 profiles with results for all 4 probes. This database was also used in experiment 1 to estimate the frequency of occurrence of any matching profiles. A randomly selected subset of 200 four-probe profiles from the MPFSL Caucasian data was used to give a database of comparable size to the FSS Caucasian database.

Results and discussion

Experiment 1 – Probability of a random match between profiles obtained from different individuals

A total of 5724 three-probe profiles were compared against each other. This procedure resulted in 16,379,226 comparisons between different samples (the number of comparisons is given by the expression $n(n-1)/2$, where n is the number of profiles in the database). The experiment

Table 1 Numbers of DNA profiles used in experiment 1. The 3-probe profiles are for the loci D1S7 (MS1), D7S21 (MS31) and D12S11 (MS43a). The 4-probe profiles are a subset of the 3-probe profiles, and also include D2S44 (YNH24)

Organisation	3 probe profiles	4 probe profiles
Forensic Science Service (FSS)	468	456
Metropolitan Police Forensic Science Laboratory (MPFSL)	4612	2890
Strathclyde Police Forensic Science Laboratory	143	52
Cellmark Diagnostics	501	501
Totals	5724	3899

Table 2 Breakdown of the data used in experiment 2. The figures in the body of the table are numbers of profiles. The 200 MPFSL profiles have been selected at random from the full MPFSL Caucasian database

Organisation	Locus			
	D1S7 MS1	D7S21 MS31	D12S11 MS43a	D2S44 YNH24
Forensic Science Serrvice (FSS)				
a) Caucasian	207	208	210	208
b) Afro-Caribbean	151	165	170	155
c) Asian	129	138	141	144
Metropolitan Police Forensic Science Laboratory (MPFSL)				
e) Caucasian – full database	1598	1613	1603	883
d) Caucasian – random subset	200	200	200	200
Strathclyde Police Forensic Science Laboratory				
f) Caucasian	144	145	144	52

was performed using a 2.8% matching rule; 2 bands are declared as matching if the difference between them is less than 2.8% of their mean size: 2.8% was chosen as this is the matching guideline used by the FSS and has been found to work well in casework. Wider matching guidelines are more appropriate when comparing data from different laboratories, consequently the experiment was also run with a 4% matching rule. When a comparison between 2 samples resulted in a match, the FSS Caucasian database was used to estimate a frequency of occurrence by taking bin widths of 5.6% and 8%, corresponding to the 2.8% and 4% matching rules. The main purpose of this experiment was to study the number of matches obtained, hence the FSS Caucasian database was used to estimate a frequency of occurrence, regardless of the ethnic origin of the profiles. The effect of using different databases to estimate frequency of occurrence was studied in experiment 2. The 3 probes MS1, MS31 and MS43a were used for the calculation. Within and between-probe independence was assumed and each result was expressed as a LR by taking 1/frequency of occurrence.

The experiment was also repeated by using the continuous approach. In this calculation there is no concept of a match, simply the closeness of the agreement between 2 profiles, and the LR is a continuous function in the range zero to infinity. The great majority of between-person comparisons will give LRs close to zero. A LR of 1,000 or greater is considered to provide strong support for the proposition that 2 samples have originated from the same individual.

Many of the profiles recorded in the data files have only a single band for one or more loci. This could indicate that the DNA is homozygous at that locus, or that the sample is heterozygous, but for some reason, only one band has been detected. For the purpose of comparing profiles for crime intelligence purposes it is usual to class a single-banded profile as matching a two-banded profile, provided that the single band corresponds to the upper band of the two-banded profile. This procedure has been followed for the results displayed in Table 3 which shows the number of between-person comparisons which gave a LR in excess of one. The results for 2.8% match binning show that a match was obtained in about 1 in 300,000 comparisons (52/16,379,226), and that the LR exceeded 1,000 in about 1 in 400,000 comparisons. As would be expected, more matches are found when 4% match binning is used. About 1 in 100,000 comparisons gave a LR in excess of 1,000. The continuous calculation gives a LR greater than one in about 1 in 200,000 comparisons, and a LR greater than 1,000 in about 1 in 700,000 comparisons. The single result shown in Table 3 which gave a LR in excess of a million involved a comparison of 6 bands in one profile against 5 bands in the other. These results show good agreement with earlier work [10], where similar calculations on a smaller database gave a LR greater than one in about 1 in 125,000 comparisons, and a LR greater than 1,000 in about 1 in 400,000 comparisons.

Table 3 Number of between-person comparisons giving a LR greater than 1: 5724 3-probe profiles gave a total of 16,379,266 between-person comparisons. Results are shown for 2.8% and 4% matching rules, and also for the continuous Bayesian calculation described by Evett et al. [9, 10]

LR greater than . . .	2.3%	4%	Bayes
1	52	185	84
10	52	185	56
100	52	185	39
1,000	39	147	23
10,000	30	84	8
100,000	19	24	4
1,000,000	2	2	1

It may be more appropriate to examine the number of between-person comparisons where all the bands correspond. This approach has been followed in some previous studies e.g. Herrin [4], and where for example the database is composed of data from liquid blood samples, and it is assumed that a repeat analysis of a particular sample would always give the same number of bands. Using the 2.8% matching rule with this approach gives no matching profiles for the 3 probes MS1, MS31, MS43a.

A 4% rule with 3 probes, results in 10 pairs of matching profiles, all of which give LRs greater than 100,000. When the continuous calculation was applied to these 10

results, 7 of the LRs were considerably less than one. The band shifts in these 7 cases did not correspond, and it is likely that all the matches would have been ruled out by a visual inspection of the corresponding profiles. Of the 3 remaining results 2 showed fairly pronounced conflicting shifts, and may have been classed as different; the final result gave a LR of 1,600 and would probably have been classed as matching. Thus by applying a 4% matching guideline, rather than a rule, it is unlikely that any more than 3, and possibly only one out of the 16 million comparisons would have been classed as a match. It is important to draw a distinction between the experiments described here, where the computer slavishly applies a rigid matching rule, and forensic casework, where visual inspection of the profiles is an essential stage in the comparison process.

Of the 84 comparisons which gave a LR greater than one with the continuous calculation, only 2 involved a three-probe match. One of these gave a LR of 1,600 and has been described above. The remaining comparison gave a LR of 7,200; the band shifts agree quite well, although the difference between one pair of bands exceeds 4%.

The standard FSS protocol involves the routine use of 4 probes in casework comparisons, with a fifth probe available if required. When the 3899 four-probe profiles were compared, generating over 7.5 million between-person comparisons, *no* four-probe matches were found, even allowing for missing bands.

Experiment 2 – Variation in LRs obtained by using different databases to assess the significance of matching profiles

The databases used are shown in Table 2. As an example of the content of the databases, smoothed fragment weight distributions for the locus D1S7 (MS1) are shown in Fig. 1. The distributions for the Afro-Caribbeans and Asians are different from each other, and from the other 4 data sets, which all represent Caucasians. The 4 Caucasian data sets show some similarities, although the effect of sample size can be clearly seen; the distribution of the smallest data set, from Strathclyde, is much "spikier" than that of the full MPFSL data.

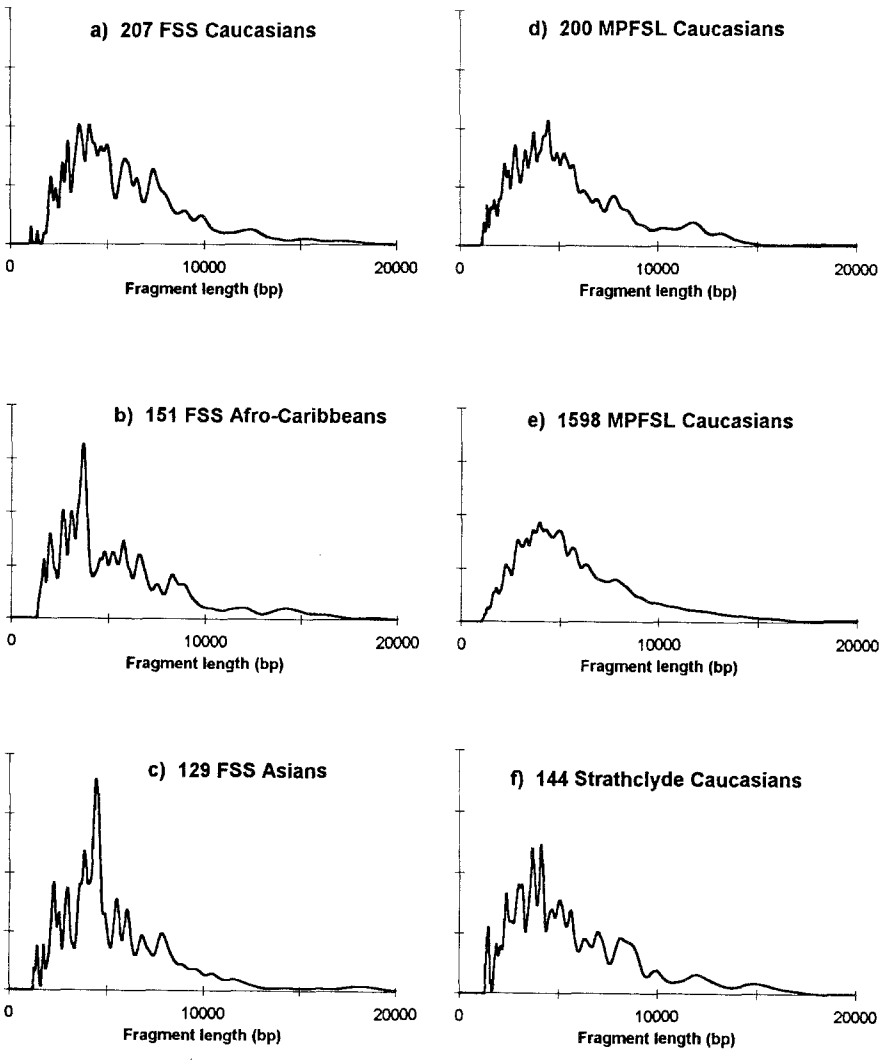

Fig. 1 Smoothed fragment length distributions for the locus D1S7 (MS1). The number of profiles included in each database is also listed in Table 2

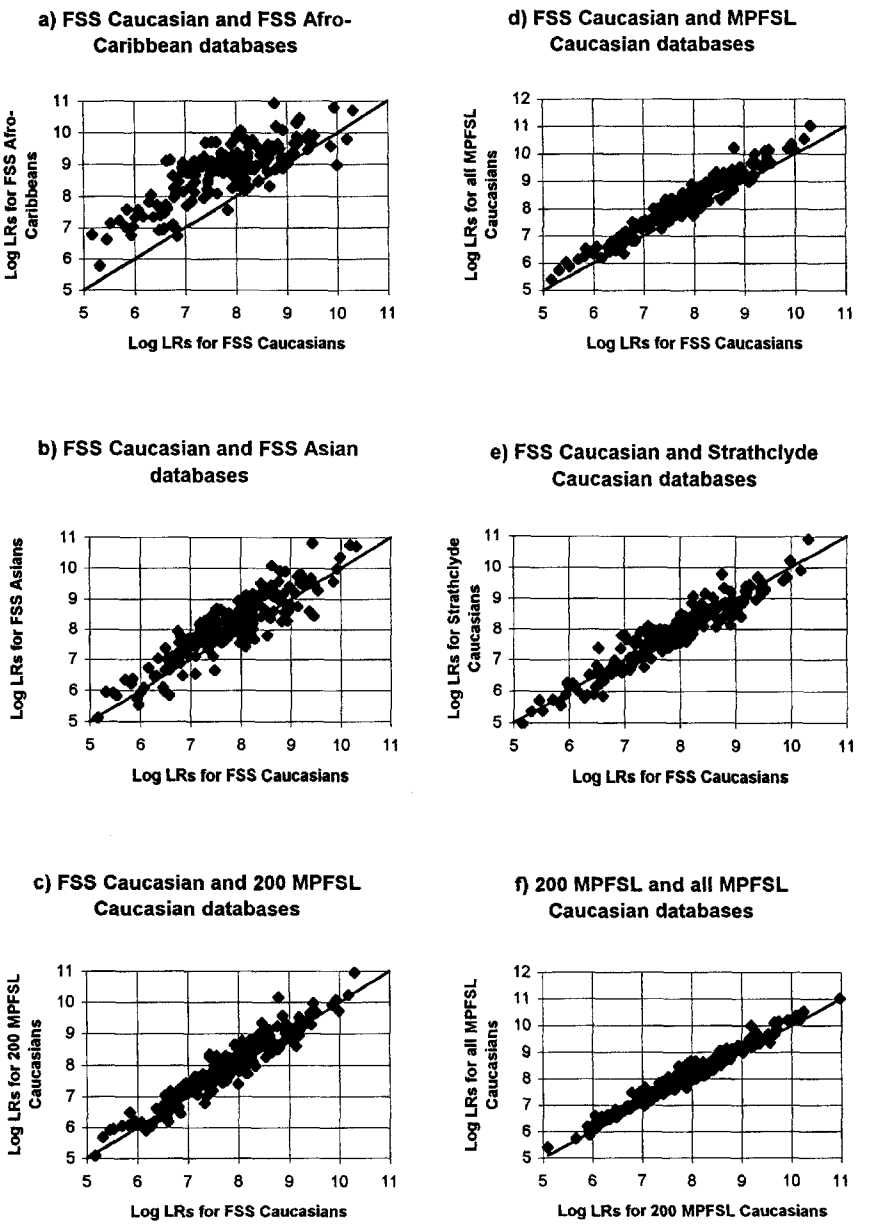

Fig. 2 Comparison of LRs for FSS Caucasian samples with FSS Caucasian using 6 different reference databases. The 2.8% sliding window method was used to calculate frequencies, which were then converted to log LRs. a–e compare the FSS Caucasian database against the other 5 databases. f compares a subset of the MPFSL Caucasian database against the full MPFSL Caucasian database

The 202 profiles in the FSS Caucasian database with results for all 4 probes were used as test samples. Frequencies were calculated by assuming that each profile in the test database is a perfect match with itself, then applying a 2.8% sliding window to enter the reference database. A sliding window frequency is obtained by counting the number of bands falling within a 5.6% window, centred on the band in question, and dividing by the number of bands in the database. An additional step involves sliding the window 2.8% in each direction and recording the 5.6% window which contains the greatest number of bands. The frequency estimates for the 202 profiles were calculated by assuming within and between probe independence, then expressed as log LRs by taking the log of 1/frequency. This process was repeated for each of the 6 reference databases.

The log LRs for individual profiles are compared in Fig. 5 where 5a shows that using an Afro-Caribbean database for assessing a profile from a Caucasian gives a less conservative figure in the great majority of cases. Figure 5b shows that there is a similar, although less pronounced effect with an Asian reference database: 5c–e show that there are less differences in the LRs when a Caucasian database collected from a different region of the British Isles is used. This agrees with the generally accepted principle that greater genetic variation occurs between race codes than between sub-populations of the same race code.

The effect of database size can be seen in Fig. 5f, which compares the LRs obtained from the full and random subset of the MPFSL Caucasian data. The sliding window method tends to give more conservative figures when a smaller database is used, as the corresponding frequency distribution of fragment lengths is more spiky, and the sliding window is therefore more likely to find one of

the spikes. This result provides further support for the robustness of the FSS methods and the adequate size of our databases.

Conclusions

DNA VNTR profiling is an extremely powerful technique and the results of these experiments show that concerns with respect to lack of within and between-probe independence have been greatly exaggerated.

The notion of "conservativeness" only has value when the suspect and the perpetrator are not the same person. In this work, by comparing a large number of profiles which have originated from different individuals, we have confirmed the inherent conservativeness of VNTR DNA profiling with UK populations.

The results of these experiments raise the issue of the uniqueness of VNTR profiles when 4 or more probes are used. The uniqueness of a human fingerprint is accepted and its value as an identification tool is undisputed. Yet this belief is based on common sense, rather than experimentation. We are not aware of any reported experiments of comparable size to ours on fingerprints.

Acknowledgements The authors wish to gratefully thank the MPFSL, Cellmark Diagnostics and Strathclyde Police Forensic Science Laboratory, for allowing us to use their databases.

References

1. Risch NJ, Devlin B (1992) On the probability of matching DNA fingerprints. Science 255: 717–720
2. Krane DE, Allen RW, Sawyer SA, Petrov DA, Hartl D (1992) Genetic differences at four DNA typing loci in Finnish, Italian and mixed Caucasian populations. Proc Natl Acad Sci USA 89: 10583–10587
3. Devlin B, Krontiris T, Risch N (1993) Population genetics of the HRAS1 minisatellite locus. Am J Hum Genet 53: 1298–1305
4. Herrin G (1993) Probability of matching RFLP patterns from unrelated individuals. Am J Hum Genet 52: 491–497
5. Gill P, Evett IW, Woodroffe S, Lygo JE, Millican ES, Webster M (1991) Databases, quality control and interpretation of DNA profiling in the Home Office Forensic Science Service. Electrophoresis 12: 204–209
6. Evett IW, Gill P (1991) A discussion of the robustness of methods for assessing the evidential value of DNA single locus profiles in crime investigations. Electrophoresis 12: 226–230
7. Gill P, Woodroffe S, Lygo JE, Millican ES (1991) Population genetics of four hypervariable loci. Int J Legal Med 104: 221–227
8. Evett IW, Pinchin R (1991) DNA single locus profiles: tests for the robustness of statistical procedures within the context of forensic science. Int J Legal Med 104: 267–272
9. Evett IW, Scranage JK, Pinchin R (1992) An efficient statistical procedure for interpreting DNA single locus profiling data in crime cases. J Forensic Sci Soc 32: 307–326
10. Evett IW, Scranage JK, Pinchin R (1993) An illustration of the advantages of efficient statistical methods for RFLP analysis in forensic science. Am J Hum Genet 52: 498–505
11. National Research Council Committee on DNA Technology in Forensic Science 1992. DNA technology in forensic science. National Academy Press, Washington, DC