Int J Legal Med (1996) 109:173–177

ORIGINAL ARTICLE

I. W. Evett · J. A. Lambert · J. S. Buckleton · B. S. Weir

Statistical analysis of a large file of data from STR profiles of British Caucasians to support forensic casework

Received: 19 April 1996 / Received in revised form: 8 October 1996

Abstract Data from nearly 2500 British Caucasians, profiled using an STR quadruplex, have been analysed. The data came from several laboratories and represent samples from different geographical distributions. Analysis of the combined files shows that previous reports of failed independence tests were the results of sampling effects. A further convincing proof is given of the robustness of the statistical methods used to estimate evidential value in casework. Comparisons between different samples show that regional effects between Scotland and the South of England have no importance from the forensic viewpoint.

Key words DNA profiling · PCR · STR · Statistics · Caucasian · Likelihood ratio

Introduction

A four locus STR quadruplex incorporating HUMVWA, HUMTHO1, HUMF13A1 and HUMFES has been developed by Kimpton et al. [1] and validated for forensic use as described by Kimpton et al. [2] and Lygo et al. [3]. Analyses of data from British Caucasians profiled using the quadruplex have been described by Gill and Evett [4], Evett et al. [5] and Lee et al. [6]. A Bayesian method of analysing the same databases is described by Foreman et al. [7]. A recent paper by Evett and Buckleton [8] has questioned the relevance of large scale independence testing and described analysis of data from 1660 British Caucasians. The purpose of the present paper is to describe the analysis of the largest set of data so far. To our knowledge, no other papers have been published on STR databases as large as the one described here: 2481 full profiles of Caucasians from throughout the British Isles.

The statistical analysis of data from restriction fragment length polymorphism (RFLP) profiling has always been problematic because of the uncertainty of differentiating between genuine and apparent homozygotes and also because of the need to take account of fragment measurement error through binning or other methods. The advent of STR profiling has largely removed such complications and so it is possible to gain a clearer impression of the scale of real or imagined population genetic effects caused by inbreeding and substructuring. However, a number of instances have been observed where significance tests on sample data have failed. In particular, as Evett et al. reported [5], a file of 257 UK Caucasians assembled at the Metropolitan Police Forensic Science Laboratory (MPFSL) was found to demonstrate very high homozygosity for F13A1 ($p = 0.00$); the same authors found that composite exact tests for within and between locus independence in a file of 1400 UK Caucasians failed for the VWA/THO1 combination ($p = 0.03$) and for the F13A1/FES combination ($p = 0.00$); and Drozd et al. [9] reported a failure of a test of Hardy-Weinberg proportions ($p < 0.01$) in a file of 200 UK Caucasians profiled at the VWA locus.

It has been something of a problem to determine whether or not these tests have any kind of population genetic rationale – such as population substructuring, for example – or whether they are nothing more than sampling effects. The analysis results presented in this paper supports the second explanation. We claim to demonstrate that the technique is sufficiently robust to allow multiplication of allele frequencies both within and between loci without using the rather arbitrary kind of correction described in Evett et al. [5], originally adopted as an interim expedient in anticipation of further data.

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

J. S. Buckleton
Institute of Environmental Science and Research Limited: Forensic, Mt. Albert, Private Bag 92-021, Auckland, New Zealand

B. S. Weir
Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA

Table 1 A summary of the sizes of the databases used in the study

Origin	Filename	VWA	THO1	F13A1	FES	Full profiles
FSS donors	1fssdon	423	423	423	423	423
Derbyshire police	1fssder	582	582	582	582	582
MPFSL donors	1mpldon	256	257	257	256	256
SPFSL casework	1fssscl2	400	400	400	400	400
MPFSL casework	1mp2	423	423	421	423	421
Tayside casework	1dund	400	400	400	400	400
Combined file	1al17_10	2484	2485	2483	2484	2481

Data

The data file sizes are shown in Table 1. The FSS and MPFSL donor files and the Derbyshire police files are as specified in Evett et al. [5]; the 140 Strathclyde samples from the Strathclyde Police Forensic Science Laboratory (SPFSL) referred to in that study were later absorbed into the file of 400 used here and analysed by Lee et al. [6]. There are two new files which have not previously been described: one of 423 individuals from casework at the MPFSL and one of 400 individuals from casework at the Tayside Police Forensic Science Laboratory, Dundee, Scotland.

Frequencies

In this, as in the previous papers we have cited, THO1 9.3 is classified as allele 10 and F13A1 3.2 is classified as allele 3.

Many papers have now shown frequency distributions for these for loci (see, for example, Gill and Evett [4]) and, for illustration we here show only those for the databases which have not previously been reported: the Tayside and London casework collections. This is also interesting because these two represent the geographical extremes of the data we have analysed. The two frequency distributions are shown in Figs. 1–4.

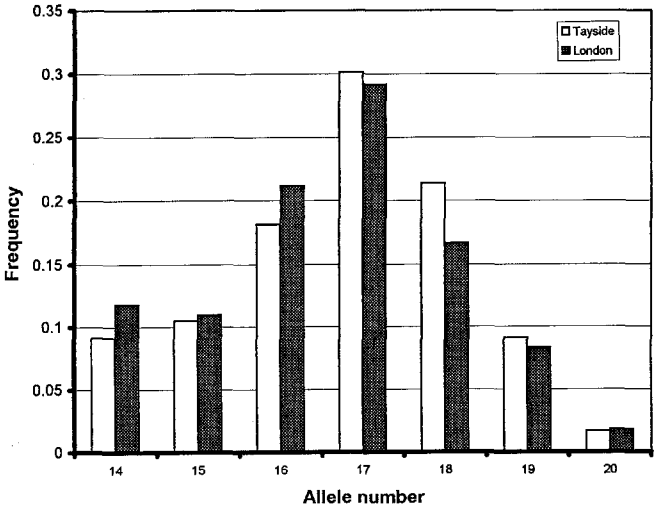

Fig. 1 Allele frequency distributions for Tayside and MPFSL casework: VWA

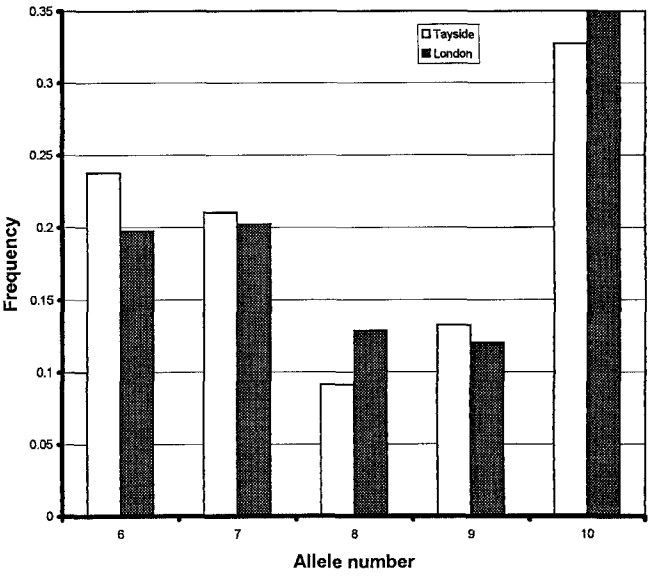

Fig. 2 Allele frequency distributions for Tayside and MPFSL casework: THO1

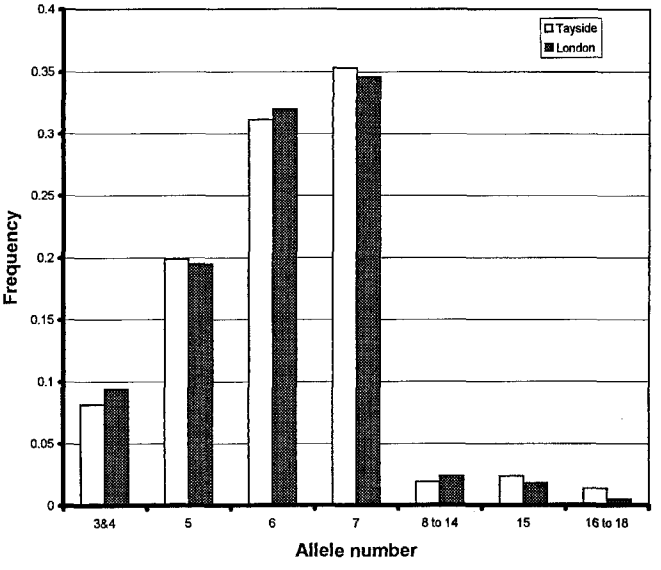

Fig. 3 Allele frequency distributions for Tayside and MPFSL casework: F13A1

Conventional tests on the combined database

The conventional goodness-of-fit test of Hardy-Weinberg proportions was carried out for each locus. The p-values were determined empirically by comparing each test statistic with those calculated from 2000 shuffled databases, following Guo and Thompson [10]. The p-values were 0.07, 0.80, 0.42 and 0.41 respectively.

Drozd et al. [9] have observed an excess of the 17,17 genotype and a deficiency of the 16,17 genotype at the VWA locus in a database of 200 British Caucasians. Our observed numbers, with expected numbers in brackets were: 194(193.06) of 17,17 and 266(287.71) of 16,17. If

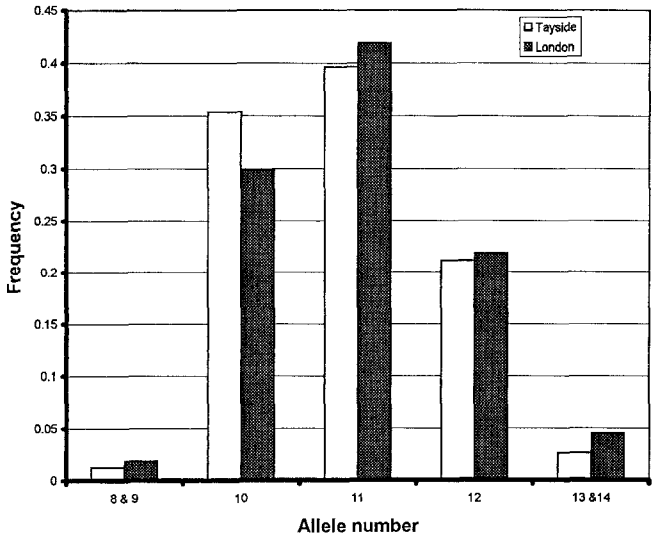

Fig. 4 Allele frequency distributions for Tayside and MPFSL casework: FES

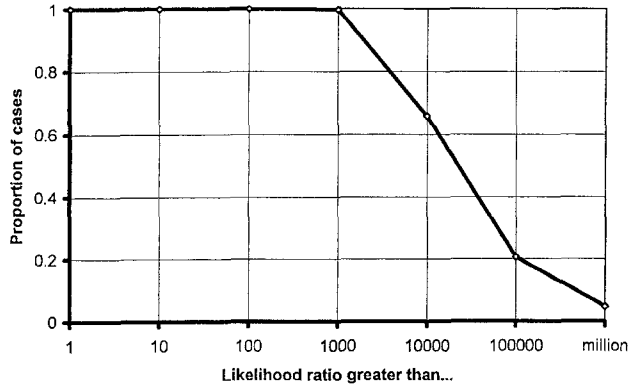

Fig. 5 Within-person experiment: simulates cases in which the suspect is truly the offender. The vertical scale represents the proportion of cases in which there is a match and the LR exceeds the value at the corresponding position on the horizontal scale

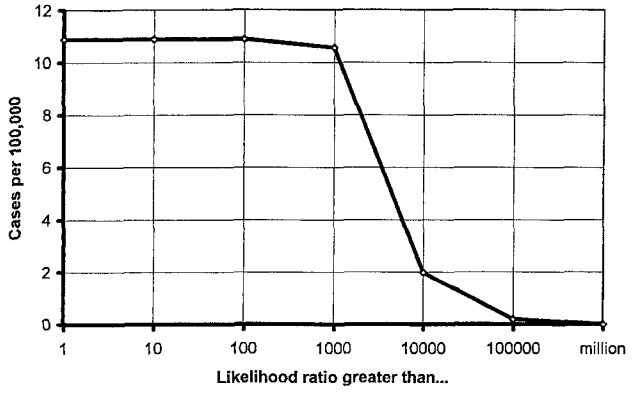

Fig. 6 Between-person experiment: simulates cases in which the suspect and the offender are two different unrelated individuals. The vertical scale represents the number of cases in 100,000 in which there is a match and the LR exceeds the value at the corresponding position on the horizontal scale

the latter were a real deficiency of the 16,17 genotype then it would have no practical significance.

The conventional test for excess homozygosity was carried out, again determining p-values from shuffled databases. The p-values were 0.15, 0.86, 0.24 and 0.46. We note that there is no sign of the excess homozygosity which was found in the first MPFSL database and conclude that this was a sampling phenomenon.

The exact test (Zaykin et al. [11]) was carried out for all 15 combinations: 4 single locus, 6 two locus, 4 three locus and one four locus. The smallest p-value was 0.12 for the composite test for VWA/THO1. We note, in particular the p-value for the composite test for F13A1/FES was 0.34, which is a remarkable change from the value 0.00 observed for the earlier sample of 1400.

Robustness investigations

The limitations of conventional independence testing based on classical statistical significance testing have been discussed elsewhere: Evett et al. [5] and Evett and Buckleton [8]. It is our view that the robustness of a particular forensic science comparative technique, whether DNA-based or otherwise, is best investigated by simulating two kinds of cases. First, cases in which the suspect is truly the offender: in the present situation this can be done by calculating a likelihood (LR) ratio as the inverse of the four locus genotype frequency for every member of the database. Second, cases in which the suspect and the offender are different unrelated people: in the present situation this can be done by carrying out all of the between-person comparisons possible, in this case approximately 3.1 million. Most comparisons result in a non-match, but in those situations in which a match does occur a LR is calculated in the same way as was done for the within-person comparisons. We repeat that the between-person comparison experiment is in exactly the same spirit as that carried out by Tippett et al. [12] when investigating the power of house paint comparison in forensic casework and later by Gaudette and Keeping [13]who considered the power of forensic hair comparison. Lambert et al. [14] have described large scale experiments based on DNA RFLP data in which up to 16 million comparisons were made.

The results of these experiments can be compared with those in Evett et al. [5], but with an important difference. In that previous work, as we have explained, we were not entirely sure of the reasons for the small p-values found for the FES/F13A1 and VWA/THO1 composite tests. We therefore proposed a rather ad hoc correction procedure to be applied to allele frequencies in casework based on that first proposed by Nichols and Balding [15] and carried out the experiments using various degrees of conservative correction factors. We now see that the effects we observed were, as we had originally suspected, no more than artefacts and we see no further need for using the corrected frequencies, at least for British Caucasians.

Therefore when we carried out the within and between-person comparisons for this study we made no cor-

rections to the allele frequencies. The four locus genotype frequencies were calculated by multiplying the uncorrected allele frequencies within and between loci. The within-person experiment is summarised in Fig. 5. This shows that in about 65% of cases in which the suspect is truly the offender the LR will exceed 10,000 and in about 20% of such cases it will exceed 100,000.

The results of the between-person comparison experiment are shown in Fig. 6. Note here that the vertical scale is in terms of cases per 100,000. Thus, when the suspect and offender are two different unrelated people, the chance that their profiles would match with a LR in excess of 10,000 is 1 in 50,000. The chance of a match with a LR in excess of 100,000 is about 1 in half a million.

We would point out that casework calculations would be more conservative than the figures illustrate because of procedures for guarding against unreasonably small allele frequencies by adding the suspect's profile either once or twice to the database. This would be in the spirit of the procedure used by Berry et al. [16] and the size bias correction proposed by Balding and Nichols [17].

Regional effects

It follows from the previous sections that there are good arguments for using the combined database for casework throughout Great Britain. Balding and Nichols [17] have devised a formula which is relevant when the suspect and offender might credibly considered to be members of the same subpopulation. That correction involves the use of F_{ST} and as a guide to appropriate magnitudes of that factor we show values of F_{ST} and F_{IT} calculated by the method of Weir and Cockerham [18] from pairwise comparisons between the Tayside database and each of the other databases in turn: again we have selected these comparisons because Tayside represents the most northerly sample we have analysed. The order of the comparisons roughly reflects geographical separation, bearing in mind that the FSS and first MPFSL samples include people from throughout the UK, though their content would be predominantly South of the England/Scottish border (Table 2).

We stress how very small the F_{ST} values are. They are small for all loci and they confirm values previously reported for other large populations – see, for example, Weir [19]. The consistency of estimates from many loci and populations argues against long tails in the distribution of F_{ST} values for the loci being used in forensic science and against the routine use of values as large as 0.03–0.05 for the UK or US Caucasian population.

We note also that the F_{IT} values are also almost all very small. There is a clear exception to this for the Tayside/MPFSL1 comparison at the F13A1 locus. However, this is obviously a manifestation of the excess of homozygosity observed at this locus in the latter database and previously reported by Evett et al. [5]. The evidence from all of the other Caucasian databases strongly supports the explanation that this excess is a sampling phenomenon.

Table 2 Comparisons between the Tayside database and the other databases (a) F_{ST} and (b) F_{IT}

(a)

	SPFSL	Derbyshire	FSS	MPFSL 1	MPFSL 2
VWA	0.001	0.001	0.001	0.002	0.001
THO1	–0.001	–0.001	0.000	0.000	0.001
F13A1	0.000	–0.001	0.001	–0.001	–0.001
FES	0.002	0.003	0.001	–0.001	0.002
Average:	0.000	0.001	0.001	0.000	0.001

(b)

	SPFSL	Derbyshire	FSS	MPFSL 1	MPFSL 2
VWA	–0.025	–0.001	0.003	0.001	0.027
THO1	–0.016	–0.022	–0.003	–0.004	–0.012
F13A1	0.000	–0.005	–0.017	0.049	–0.006
FES	–0.007	0.010	–0.021	–0.019	0.026
Average:	–0.012	–0.005	0.000	0.007	–0.003

Discussion

In the introduction we referred to failures of significance tests on STR data which had previously been reported by Evett et al. [5] and Drozd et al. [9]. The results from this larger study give good support to the explanation that these are manifestations of sampling effects. The regional comparisons described in the previous section are evidence that population substructuring among British Caucasians is of negligible consequence as far as STR loci are concerned.

Finally, the robustness experiments we have described give further evidence of the capabilities of the VWA/THO1/F13A1/FES quadruplex as a powerful forensic technique.

Acknowledgements We wish to express our thanks to the members of staff of the following laboratories who freely made their data available for analysis: Metropolitan Police Forensic Science Laboratory; Strathclyde Police Forensic Science Laboratory; Tayside Police Forensic Science Laboratory; Forensic Science Service.

References

1. Kimpton C, Gill P, Walton A, Urquhart A, Millican ES, Adams M (1993) Automated DNA profiling employing 'multiplexing' amplification of short tandem repeat loci. PCR Methods Appl 3:13–22
2. Kimpton C, Fisher E, 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
3. Lygo JE, Johnson DJ, Holdaway S, Woodroffe S, Whitaker JP, Clayton TM, Kimpton CP, Gill P (1994) The validation of short tandem repeat (STR) loci for the use in forensic casework. Int J Legal Med 107:77–89

4. Gill P, Evett I (1995) Population genetics of short tandem repeat (STR) loci. Genetica 96:69–87
5. Evett IW, Gill PD, Scranage JK, Weir BS (1996) Establishing the robustness of STR statistics for forensic applications. Am J Hum Genet 58:398–407
6. Lee LD, Fairley M, Lambert JA, Evett IW (1996) Validation of a frequency database for four STR loci for use in casework in the Strathclyde Police Forensic Science Laboratory. Forensic Sci Int 79:43–48
7. Foreman LA, Smith AFM, Evett IW (1996) Bayesian analysis of DNA profiling data in forensic identification applications. J R Stat Soc (Series A) (in press)
8. 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 New York, pp 79–86
9. Drozd MA, Archard L, Lincoln PJ, Morling N, Nelleman LJ, Phillips C, Soteriou B, Syndercombe Court D (1994) An investigation of the HUMVWA31A locus in British Caucasians. Forensic Sci Int 69:161–170
10. Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportions for multiple alleles. Biometrics 48:361–372
11. Zaykin D, Zhivotovsky L, Weir BS (1995) Exact tests for association between alleles at arbitrary numbers of loci. Genetica 96:169–178
12. Tippett CF, Emerson VJ, Fereday MJ, Lawton F, Lampert SM (1968) The evidential value of the comparison of paint flakes from sources other than vehicles. J Forensic Sci Soc 8:61–65
13. Gaudette BD, Keeping ES (1974) An attempt at determining probabilities in human scalp hair comparison. J Forensic Sci 19:599–606
14. Lambert JA, Scranage JK, Evett IW (1995) Large database experiments to assess the significance of matching DNA profiles. Int J Legal Med 108:8–13
15. Nichols RA, Balding DJ (1991) Effects of population structure on DNA fingerprint analysis in forensic science. Heredity 66: 297–302
16. Berry DA, Evett IW, Pinchin RA (1992) Statistical inference in crime investigations using deoxyribonucleic acid profiling. Appl Statist 41:499–531
17. Balding DJ, Nichols RA (1994) DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int 64: 125–140
18. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370
19. Weir BS (1994) Effects of inbreeding on forensic calculations. Annu Rev Genet 28:597–621