Int J Legal Med (2001) 115 : 128–134

ORIGINAL ARTICLE

J. A. Thomson · K. L. Ayres · V. Pilotti · M. N. Barrett
J. I. H. Walker · P. G. Debenham

Analysis of disputed single-parent/child and sibling relationships using 16 STR loci

Received: 16 June 2000 / Accepted: 7 January 2001

Abstract This study describes the validation of short tandem repeat (STR) systems for the resolution of cases of disputed parentage where only a single parent is available for testing or where the claimed relationship of both parents is in doubt and also cases where sibship must be tested. Three separate multiplex systems the Second Generation Multiplex, Powerplex 1.2 and FFFL have been employed, giving a total of 16 STR loci. Both empirical and theoretical approaches to the validation have been adopted. Appropriate equations have been derived to calculate likelihood ratios for different relationships, incorporating a correction for subpopulation effects. An F_{ST} point estimate of 1% has been applied throughout. Empirically, 101 cases of alleged father, alleged mother and child where analysed using six SLP systems and also using the three multiplex STR systems. Of the 202 relationships tested, 197 were independently resolved by both systems, providing either clear evidence of non-parentage or strong support for the relationship.

Keywords STR profile · DNA profile · Parentage · Paternity · Likelihood ratio · F_{ST}

Introduction

The use of short tandem repeat (STR) profiling in forensic science is widespread. More recently it has also been adopted for routine use in paternity disputes (e.g. [1]). As with all PCR-based systems, STR profiling offers significant benefits in terms of sensitivity, speed and amenity to automation. Of particular benefit in parentage testing is the ability to carry out STR analysis on more convenient sample types, such as blood spots from finger or heel pricks taken onto paper stain-cards, or buccal swabs. This is particularly pertinent in cases where samples may have to be taken overseas and subsequently transported to the testing laboratory, as is often the case in the investigation of claimed family relationships of individuals wishing to emigrate to the UK.

However, in comparison with existing single locus probe (SLP) systems, individual STR loci generally have lower discriminatory power, necessitating the use of many loci in order to resolve disputed relationships satisfactorily. We have previously shown [1] that 12 STR loci can be used to satisfactorily resolve cases where paternity alone is disputed. However, this combination of loci will be less efficient when applied to cases where a single parent only is available for testing or the claimed relationship of both parents is in doubt and also when the real parent of the tested child may be closely related to the tested alleged parent. Similar problems are encountered in cases where other relationships, such as sibship (full or half), must be tested.

We have taken an empirical and a theoretical approach in order to validate 16 loci in 3 separate STR multiplexes, for resolving cases such as those described. A brief outline and preliminary results for parentage testing were given in Thomson et al. [2]. In this paper we detail our analyses more fully and in addition address the other relationships described above.

The three multiplex systems used in the analysis were the Second Generation Multiplex (SGM) system, developed by the Forensic Science Service [3], Powerplex 1.2 (Promega, Madison, Wis.) [4] and FFFL [4] (also Promega). The SGM system consists of the loci HUMAMGX/Y, HUMTH01, HUMVWFA31/A, D8S1179, D21S11, HUMFIBRA and D18S51. The Powerplex 1.2 system consists of the loci HUMAMGX/Y, D5S818, D13S317, D7S820, D16S539, HUMVWFA31/A, HUMTH01,

J. A. Thomson (✉) · V. Pilotti · P. G. Debenham
Life Sciences Group, LGC, Queens Road, Teddington, Middlesex, TW11 0LY, UK
e-mail: jat@lgc.co.uk,
Tel.: +44-20-89437604, Fax: +44-20-89437659

K. L. Ayres
Department of Applied Statistics, University of Reading, Reading, UK

M. N. Barrett · J. I. H. Walker · P. G. Debenham
University Diagnostics Ltd., Queens Road, Teddington, Middlesex, UK

HUMTPOX and HUMCSF1PO. The FFFL system consists of the loci F13A1, F13B, FES/FPS and LPL. This gives a total of 16 separate STR systems, plus the AMG sex test.

Samples from 101 cases from the Indian subcontinent involving alleged father, alleged mother and child were analysed using the established six SLP tests currently in use and also using the three multiplex STR systems. The results from the SLP and STR loci were compared in each case. In addition, we investigated the ability of the systems to resolve different relationship scenarios by simulation and in this way we were able to extend the scope of the study to cases other than straightforward parentage testing. For these simulations, cases were drawn from hypothetical populations derived from allele proportions in our Asian (Indian subcontinent) frequency database (e.g. [1]).

Specifically, simulated families were constructed by first generating all necessary unrelated individuals in each family (e.g. mother and father) with joint genotype proportions given by the recursive formula:

$$P_{m+1}(i) = \frac{m_i F_{ST} + (1 - F_{ST}) p_i}{1 + (m-1) F_{ST}}$$

which is the probability that the $(m + 1)$th gene sampled is of allele type A_i, given that m_i of the m genes already sampled are A_i (e.g. Balding and Nichols [5]): here p_i is the population relative frequency of allele A_i, and F_{ST} is the coancestry coefficient (an F_{ST} value of 1% was implemented in each simulation). Children were then generated by randomly choosing one gene at each locus from each parent.

Throughout this study, F_{ST} has been used to model recent shared ancestry of individuals in the population, following the recommendations of Balding and Nichols [5]. We have used the formulae outlined by Ayres [6] and have implemented an F_{ST} value of 1% in our likelihood ratio calculations. Foreman and Lambert [7] have recently showed that estimates of F_{ST} for UK Asian and Afro-Caribbean subpopulations are typically less than 1% when a relevant population database is considered.

Materials and methods

Samples were obtained from 101 cases of disputed parentage submitted to the University Diagnostics (UDL) laboratory. In each case, the relationship of the child to both alleged parents was in doubt prior to any DNA analysis. All parties had consented to the use of their samples for further research.

All of these cases had been previously analysed (as part of the routine casework) using the six SLP systems YNH24 [8], TBQ7 [9], EFD52 [10] (Promega), LH1 (Life Technologies, Paisley, UK), MR24/1 (formerly Amersham UK, now UDL) and 3'alpha HVR [11]. Analysis followed the manufacturers' directions, as previously described [1].

SGM and Powerplex 1.2 amplification and electrophoresis on a PE-ABI 377 DNA sequencer was also performed as previously described [1]. FFFL amplification was performed according to the manufacturer's instructions (under conditions very similar to those used for Powerplex 1.2) and electrophoresis was carried out under the same conditions as for SGM.

Results were obtained in the form of likelihood ratios (i.e. the ratio of the probability of observing the DNA profiles under two different relationship scenarios), implementing the coancestry coefficient F_{ST} in each calculation. Likelihood ratios (LRs) were calculated for all 202 alleged single-parent/child relationships tested. In addition to the usual LR for parentage (i.e. paternity or maternity as appropriate) versus unrelatedness (henceforth denoted PI), the LR in favour of a first-degree relative of the alleged parent being the true parent (as opposed to the alleged parent) was calculated. Moreover, a simulation study was performed using theoretical populations generated from observed allele proportions and LR distributions for relationships such as sibship were approximated via this method.

Results

Parentage testing

Comparison of SLP and STR results

SLP and STR results for the 202 alleged parent/child relationships tested were compared. Full STR profiles for the 16 loci and 6 SLP results were obtained for every sample. For each relationship, the result is classed as one of three types; (1) no mismatches between alleged parent and child, (2) single SLP mismatch or 1–2 STR mismatches, a result of either non-parentage or of mutation, and (3) two or more mismatches for SLPs, or three or more mismatches for STRs (evidence in favour of non-parentage). These results are summarised in Table 1.

Of the 202 single-parent/child alleged relationships tested, 181 resulted in a combined (SLP + STR) PI > 1,000, providing evidence in favour of the claimed relationship. Of these, 178 were supported (PI > 1,000, no mismatches) by both the SLP and STR systems independently. On the basis of allele sharing alone, two relationships were only supported by STRs (SLPs resulted in a single mismatch possibly attributable to a mutation event), and one was supported by SLPs only (for this case the STR locus D13S317 showed a mismatch, possibly at-

Table 1 Summary of results of 202 single-parent/child relationships tested with SLPs and STRs (number of cases in each indicated category)

SLP result	STR result				
	All inclusion	Single mismatch	Two mismatch	> Two mismatches	Total
All inclusion	178	1	0	0	179
Single mismatch	2	0	1	0	3
> One mismatch	0	1	0	19	20
Total	180	2	1	19	202

tributable to a null allele or primer binding site mismatch) [12]. Of the putative relationships, 19 resulted in a number of mismatches (inconsistent with parentage) at both SLP and STR loci independently (two or more mismatches for SLP; three or more mismatches for STR). One claimed relationship exhibiting five mismatches with SLP resulted in only one mismatch at the STR loci. In this case, the PI based on the 15 non-excluding STR loci was low (429) and incorporating a simple mutation model for the mismatch locus (see Discussion) resulted in a PI of 0.2, which weakly supports the hypothesis that an unrelated person is the true parent relative to the hypothesis that the tested individual is the true parent.

One alleged relationship resulted in a single SLP mismatch and two STR mismatches. The SLP results alone resulted in a PI of 14.4 and the LR of a first-degree relative being the father as opposed to the tested parent was 8.7. For the STR systems the corresponding LR values were 0.0013 and 6,567. The combined PI over the 22 SLP and STR systems (incorporating a simple mutation model for the mismatch loci, see the Discussion) was 0.16. Hence, an unrelated man is only 6 times more likely to be the father than the tested man, despite three mismatches (assuming prior odds of paternity of 0.5). However, the LR of a first-degree relative being the father as opposed to the tested man was 57,000. It is therefore important to calculate the PI even when several mismatches are seen (and calculating the LR for a first-degree relative of the alleged parent being the true parent can help resolve such cases).

Likelihood ratio distributions

Figure 1 shows the observed and expected distributions for the PI in cases where no mismatches are seen (previously given in Thomson et al. [2]). Where possible, one non-excluded parent from each of the 101 cases was selected at random to avoid bias. PI distributions for the STR systems are shown for F_{ST} = 0% (to allow direct comparison with the SLP data), and for F_{ST} = 1%, for both the observed cases and also for 10,000 parent-child pairs randomly generated from allele frequency tables (with F_{ST} = 1%).

The STR distributions are more diffuse (i.e. longer tails) than the SLP distribution and the increased variability is due to the greater number of systems used. This variability results in a slightly increased number of cases with a PI < 1,000 (2% for F_{ST} = 0%; 4% for F_{ST} = 1%) and a corresponding rise in the number of cases with very high values for the PI.

The LR for the hypothesis that a first-degree relative of the alleged parent is the true parent versus that an unrelated individual is the parent (L1) can be calculated by the formula L1 = (PI/2) + 0.5 (see e.g. [13] for the case when the relative is a brother of the putative father). Dividing L1 by PI results in the LR in favour of a first-degree relative versus the alleged parent being the real parent. This is an important measure in the resolution of immigration casework where occasionally nephews or nieces of the alleged parent are claimed to be their children. Figure 2 shows the distribution of the measure L1/PI for the 95 non-excluded (presumed actual parent-child) cases for the SLP and STR systems (for STRs, results are given for F_{ST} = 0% and F_{ST} = 1%), and for 10,000 simulated parent-

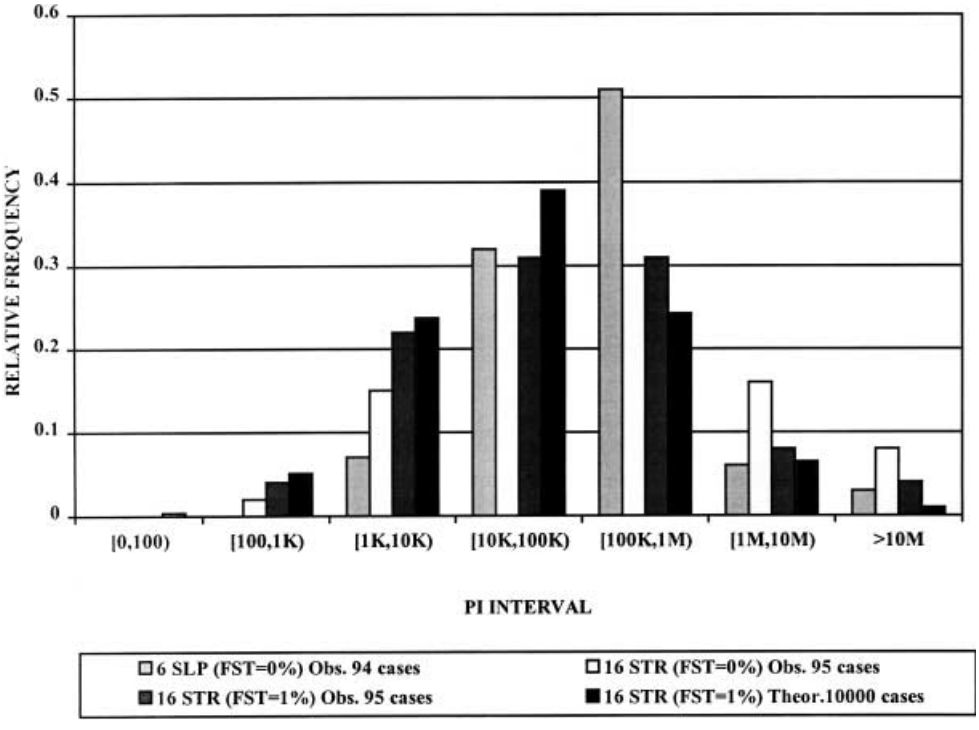

Fig 1 Distribution of PI in parent-child pairs for six SLP systems and for 16 STR systems. PIs were calculated in the 95 test cases where at least 1 alleged parent was not excluded. A theoretical distribution based on 10,000 simulated parent-child pairs was also calculated

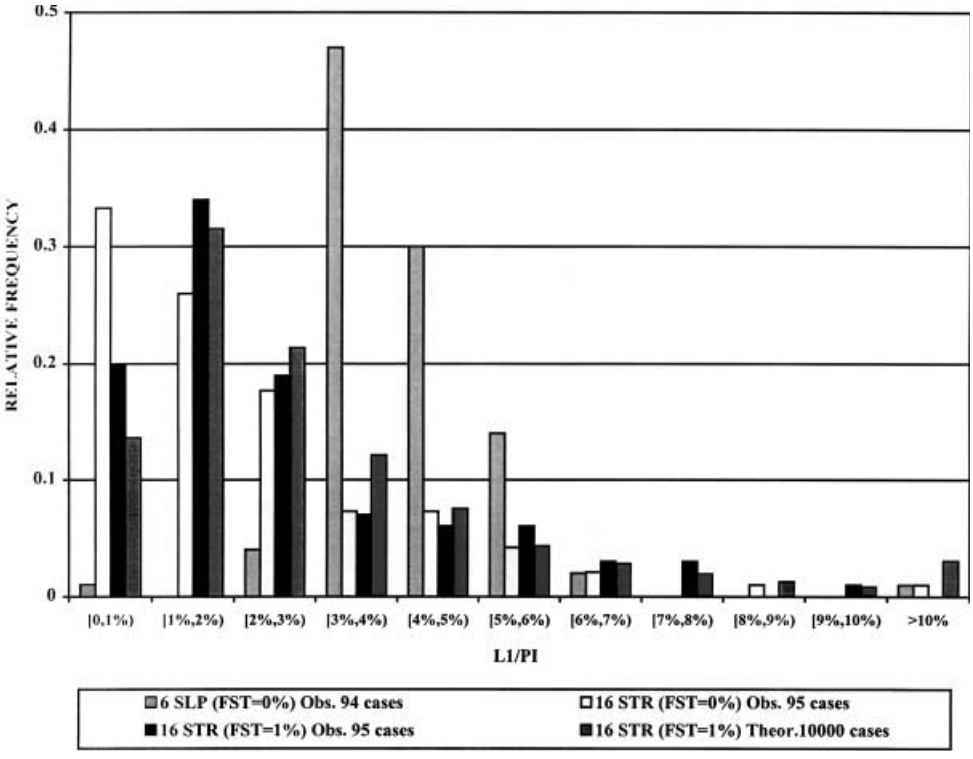

Fig. 2 Distribution of L1/PI ratios in 95 observed relationships (where parentage is assumed to hold true for the alleged parent) and in 10,000 simulated parent-child pairs. L1/PI is the relative likelihood of a first-degree relative of the alleged parent being the true parent, versus the alleged parent being the true parent

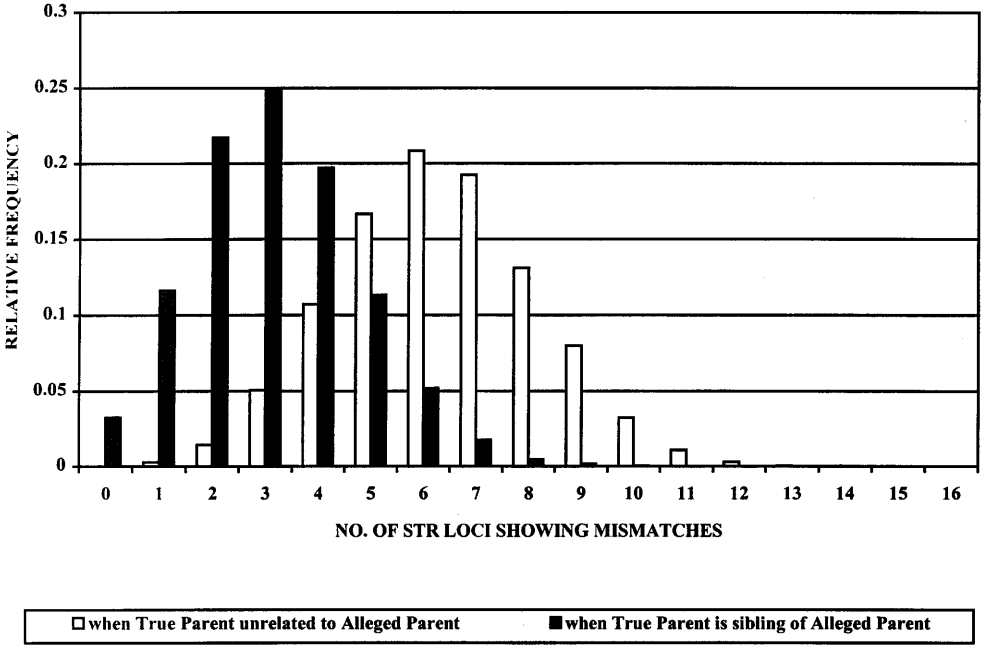

Fig. 3 Distribution of the number of STR loci showing mismatches (exclusions) in alleged single-parent/child pairs when the claimed parent is not the parent but (1) unrelated, or (2) a sibling of the true parent (based on 10,000 simulated cases with F_{ST} = 1%)

child pairs (F_{ST} = 1%). These distributions show that when the alleged parent is the real parent, 16 STRs result in lower L1/PI values than do the 6 SLPs and therefore provide stronger evidence to support the true relationship.

Exclusion power

When the alleged parent is not the true parent we have investigated the ability of this STR panel to exclude the claimed relationship. A total of 10,000 unrelated pairs of individuals were simulated from population allele proportions (with F_{ST} = 1%) and a further 10,000 uncle-nephew pairs were simulated. The number of excluding loci (mismatches) in each pair was summed. As shown in Fig. 3, when the true parent is actually unrelated to the alleged parent, 3 or more mismatches will be expected in 98.3% of cases (63.4% when the true parent is actually a sibling of the alleged parent), and 2 or more in 99.7% (85.1%) of cases. Hence, in only 0.3% of cases will one or zero mis-

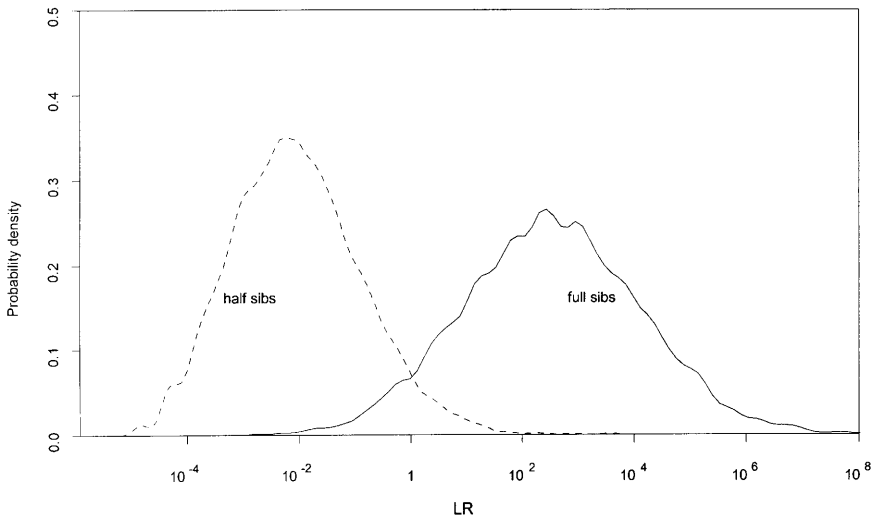

Fig. 4 Distribution of the LR in favour of two children who have the same tested mother having the same (untested) father, versus having different unrelated (untested) fathers, when they are actually full siblings (——) and actually half siblings (-----), based on 10,000 simulated cases with F_{ST} = 1% (*the x-axis is shown on a* log_{10} scale)

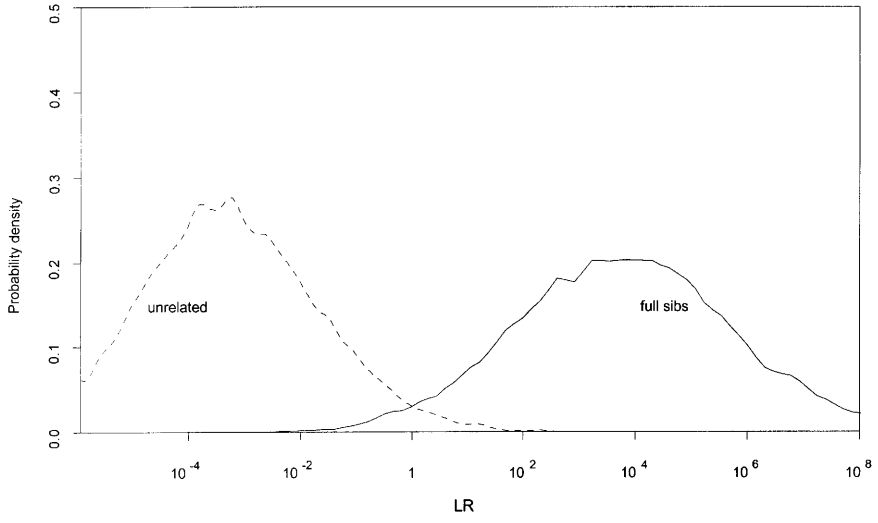

Fig. 5 Distribution of the LR in favour of two individuals being full siblings versus being unrelated when they are actually full sibs (——), and when actually unrelated (-----), based on 10,000 simulated cases with F_{ST} = 1%

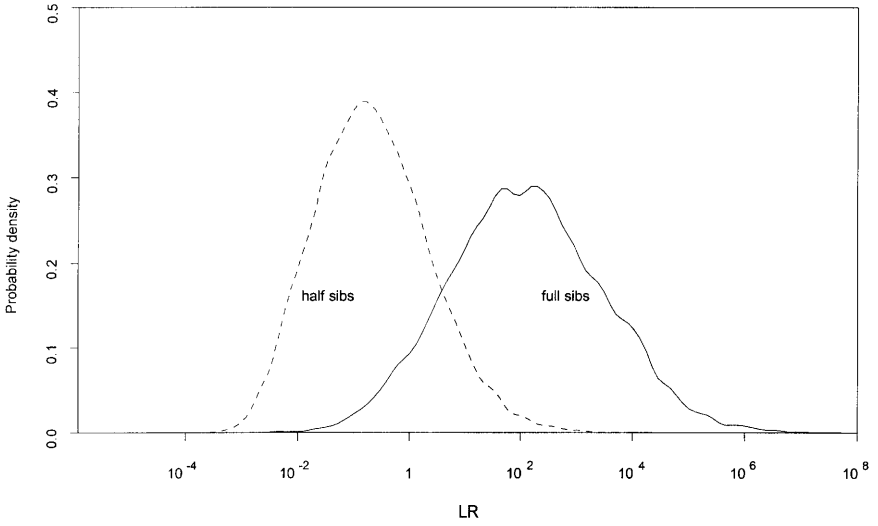

Fig. 6 Distribution of the LR in favour of two individuals being full siblings versus being half siblings (no parents tested) when they are actually full sibs (——) and when actually half sibs (-----), based on 10,000 simulated cases with F_{ST} = 1%

matches be observed (14.9% for uncle-nephew; 11.6% showing one mismatch, 3.3% showing zero mismatches). When the true parent is actually a sibling of the alleged parent and no mismatches are found, our results (not shown) indicate that the LR for the alleged parent will be < 100 in about 98% of cases and < 1,000 in virtually every case. Low LR values such as these would merit further investigation and consequently very few, if any, uncle-nephew relationships would be wrongly interpreted as parent-child.

Resolution of other relationship questions

The ability of this 16 STR panel to resolve other questions of familial relationships has also been addressed using a LR approach. Specific scenarios that we have so far considered are (1) likelihood of common paternity for two alleged full siblings presented with their mother (Fig. 4.) and (2) relatedness of alleged siblings presented without parents (Figs. 5 and 6).

For each scenario, theoretical distributions of LRs have been approximated by simulation both for the case where the relationship is as claimed and when not as claimed. Details are shown in the figure legends. From these distributions, it is possible to estimate the number of such cases where LR values greater than 100 or 1000 in favour of the tested relation will be observed. For scenario (1), an LR of > 100 in favour of common paternity is expected in approximately 63% of cases when common paternity is true, but only 0.2% of cases when it is not true. For LRs > 1,000 the corresponding figures are 38% when the relationship is true and 0.1% when it is not. Corresponding values for the other relationships tested are as follows: for full sibs versus unrelatedness (Fig. 5); LR > 100 in 82% of cases when full sibship is true, 0.1% when it is false; LR > 1,000 in 66% when true, 0.01% when false. For full sibs versus half-sibs (Fig. 6); LR > 100 in 52% of cases when full sibship is true, 1% when they are actually half sibs; LR > 1,000 in 25% when true, 0.1% when false.

Discussion

The usefulness of STR loci in relatedness testing has been the subject of many recent studies [e.g. 1, 14, 15, 16, 17]. These studies all find that STR multiplexes have a high power of discrimination for paternity testing.

This study has confirmed those findings and has also addressed the issues of investigating other relationships. Our results have shown that the panel of 16 STR systems implemented here form the basis of a highly informative and convenient system for investigation of parentage and other claimed relationships. In terms of the ability to exclude non-parents and to provide a high level of certainty in cases of non-exclusion, the panel performs very similarly to the six SLP systems previously in use in the UDL laboratory.

We have touched only briefly on the issue of mutation when comparing the STR and SLP results and did not include it in our simulation study. Nonetheless, the possibility remains in practice that a small number of mismatches will be seen among the 16 STR loci as a result of mutation events. For the SGM and Powerplex 1.2 systems, Thomson et al. [1] calculated that a single mutant allele (across these 12 loci) would be passed from parent to offspring approximately 1% of the time (based on their own observed mutation rates, together with those of Brinkmann et al. [18]). However, the PI can be readily calculated in the presence of a putative mutation (Ayres [6]; see also Egeland et al. [19] for the case when $F_{ST} = 0$). In the SLP/STR results discussed above, we applied a simple mutation model, in which all mutations were equally likely. Alternative models more suited to the biological mechanisms of STRs, such as stepwise mutation models [20] can also be implemented. A suitable stepwise model would assume a greater likelihood that a mutation will be single-step (e.g. TH01 8 to 9) than multi-step (e.g. TH01 7 to 9 or TH01 6 to 9). Published data [18] indicate that single-step mutations are around 20 times more frequent that multi-step mutations, so these ratios can be incorporated into the model and amended if necessary as further data become available. Hence, the relative likelihood in favour of parentage can be routinely reported even in the presence of mismatches.

Our analyses have made use of the likelihood ratio, which is a standard statistical procedure in parentage testing. Combined with a prior odds equally in favour of the putative parent and an unrelated individual, the Essen-Möller probability of parentage [21], that is PI/(1 + PI), can be calculated. For each alleged relationship studied, we have assumed that a specific alternative relationship (hence LR) is of interest. For cases where multiple alternatives need to be considered simultaneously, the LR will involve a summation over these alternatives (see e.g. [22] for the case of paternity testing, with F_{ST} assumed as zero).

Throughout our analyses we have implemented the parameter F_{ST} in order to model recent shared ancestry among the individuals tested. This is particularly important when these individuals are from isolated communities, as alleles that are not identical by descent from a close common relative are more likely (in comparison with a large randomly mating population) to be identical by descent on an ancestral (subpopulation) level. We have used an F_{ST} value of 1% throughout, which is an appropriate value for most UK Caucasian [23], UK Asian and Afro-Caribbean populations [7], although other values can be readily applied. Higher values may be required when considering individuals from genetically isolated subpopulations, or when appropriate population databases are not available (estimates of F_{ST} can be obtained from subpopulation [23] or population [24] data).

References

1. Thomson JA, Pilotti V, Stevens P, Ayres KL, Debenham PG (1999) Validation of short tandem repeat analysis for the investigation of cases of disputed paternity. Forensic Sci Int 100: 1–16

2. Thomson JA, Ayres KL, Pilotti V, Barrett MN, Walker JIH, Debenham PG (2000) Validation of multiplex STR systems for the investigation of parentage and other familial relationships. In: Sensabaugh GF, Lincoln PJ, Olaisen B (eds) Progress in forensic genetics 8. Elsevier Science, Amsterdam, pp 374–376
3. Oldroyd NJ, Urquhart A, Kimpton CP, Downes TJ, Millican ES, Watson SK, Gill P (1995) A highly discriminating octoplex short tandem repeat polymerase chain reaction system suitable for human individual identification. Electrophoresis 16:334–337
4. Lins AB, Micka KA, Sprecher CJ, Taylor JA, Bacher JW, Rabbach DR, Bever RA, Creacy SD, Schumm JW (1998) Development and population study of an eight-locus short tandem repeat (STR) multiplex system. J Forensic Sci 43:1168–1180
5. 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
6. Ayres KL (2000) Relatedness testing in subdivided populations. Forensic Sci Int 114:107–115
7. Foreman LA, Lambert JA (2000) Genetic differentiation within and between four UK ethnic groups. Forensic Sci Int 114:7–20
8. Nakamura Y, Gillilan S, O'Connell P, Leppert M, Lathrop GM, Lalouel GM, White R (1987) Isolation and mapping of a polymorphic DNA sequence pYNH24 on chromosome 2 (D2S44). Nucleic Acids Res 15:23
9. Nakamura Y, Fujimoto E, O'Connell P, Leppert M, Lathrop GM, Lalouel GM, White R (1987) Isolation and mapping of a polymorphic DNA cEFD52 on chromosome 17q (D17S26). Nucleic Acids Res 16:786
10. Bragg T, Nakamura Y, Jones C, White R (1987) Isolation and mapping of a polymorphic DNA sequence (cTBQ7) on chromosome 10 (D10S28). Nucleic Acids Res 16:11395
11. Jarman A, Nichols RD, Wetherall DJ, Clegg JB, Higgs DR (1986) Molecular characterisation of a hypervariable region downstream of the human α-globin gene cluster. EMBO J 5: 1857–1863
12. Budowle B (2000) STR allele concordance between different primer sets – a brief summary. Profiles DNA 3:10–11
13. Evett IW, Weir BS (1998) Interpreting DNA evidence. Sinauer, Sunderland, Mass., p 180
14. Henke L, Fimmers R, Josephi E, Cleef S, Dulmer M, Henke J (1999) Usefulness of conventional blood groups, DNA-minisatellites, and short tandem repeat polymorphisms in paternity testing: a comparison. Forensic Sci Int 103:133–142
15. Brinkmann B, Pfeiffer H, Schürenkamp M, Hohoff C (2001) The evidential value of STRs. An analysis of exclusion cases. Int J Legal Med 114:173–177
16. Calafell F (2000) The probability distribution of a number of loci indicating exclusion in a core set of STR markers. Int J Legal Med 114:61–65
17. Zupanic Pajnic I, Sterlinko H, Balazic J, Komel R (2001) Parentage testing with 14 STR loci and population data of 5 STRs in the Slovevian population. Int J Legal Med 114: 178–180
18. Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 62:1408–1415
19. Egeland T, Mostad PF, Olaisen B (1997) A computerised method for calculating the probability of pedigrees from genetic data. Sci Justice 37:269–274
20. Ohta T, Kimura M (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet Res 22:201–204
21. Essen-Möller E (1938) Die Beweiskraft der Aehnlichkeit im Vaterschaftsnachweis; theoretische Grundlagen. Mitteilungen der Anthropologischen Gesellschaft (Wien) 68:9–53
22. Lee JW, Lee H-S, Park M, Hwang J-J (1999) Paternity probability when a relative of the father is an alleged father. Sci Justice 39:223–230
23. Balding DJ, Nichols RA (1997) Significant genetic correlations among Caucasians at forensic DNA loci. Heredity 78:583–589
24. Foreman LA, Evett IW, Smith AFM (1997) Bayesian analysis of deoxyribonucleic acid profiling data in forensic identification applications (with discussion). J R Statist Soc A 160:429–469