Int J Legal Med (2003) 117 : 39–45
DOI 10.1007/s00414-002-0321-z

ORIGINAL ARTICLE

Yue-Qing Hu · Wing K. Fung

Interpreting DNA mixtures with the presence of relatives

Received: 13 February 2002 / Accepted: 6 June 2002 / Published online: 26 November 2002

Abstract The assessment of DNA mixtures with the presence of relatives is discussed in this paper. The kinship coefficients are incorporated into the evaluation of the likelihood ratio and we first derive a unified expression of joint genotypic probabilities. A general formula and seven types of detailed expressions for calculating likelihood ratios are then developed for the case that a relative of the tested suspect is an unknown contributor to the mixed stain. These results can also be applied to the case of a non-tested suspect with one tested relative. Moreover, the formula for calculating the likelihood ratio when there are two related unknown contributors is given. Data for a real situation are given for illustration, and the effect of kinship on the likelihood ratio is shown therein. Some interesting findings are obtained.

Keywords Forensic science · Kinship coefficient · Likelihood ratio · Mixed stain · Relative

Introduction

Consider a crime case in which the stain is collected from the scene and the reference sample is gathered from the suspect, through a profiling system. The suspect cannot be excluded as a contributor of the stain if the reference sample matches the crime stain. If that is not the case, a suggestion may be made that one close relative of the suspect is a probable assailant when the suspect and crime stain share very rare alleles (Sjerps and Kloosterman 1999). Usually, a series of hypotheses will be raised to explain who the contributors were, and the likelihood ratio (*LR*) is an effective tool to assess the strength of the evidence. The problem of how to assign the weight of the DNA evidence when one suspect's relative is involved in the pool of possible perpetrators was discussed by several authors over the past years. For example, Evett (1992) established a formula for the likelihood ratio in a case where the defense is "It was my brother"; Brookfield (1994) evaluated the effect on the likelihood ratio of the possibility that the suspect and the source of the crime scene DNA are relatives; Donnelly (1995) quantified the effect of close relatives on the match probability; Belin et al. (1997) described a new methodology that summarizes DNA evidence by addressing the possibility that a relative of the accused individual is the source of a crime sample; Sjerps and Kloosterman (1999) discussed the assessment of DNA profiles for close relatives of an excluded suspect; Lee et al. (1999) described a method for inference in a case where the true father may be a relative of the alleged father. These authors limited the effect of kinship on the evaluation of match probability and likelihood ratio relating to a single source DNA sample. Recently, Ayres (2000) presented adjusted *LR* formulae for various two-person relationships, incorporating the coancestry coefficient F_{ST}; Fung et al. (2002) discussed the probability of exclusion when the alleged father is a relative of the true father in paternity testing.

It is a common case that the DNA material from the crime scene was contributed by more than one person, e.g. the victim and the perpetrator. Weir et al. (1997) considered the interpretation of DNA mixtures and derived a general formula for the evaluation of the *LR*; Curran et al. (1999) and Fung and Hu (2000a) extended the results to the situation where the relatedness between persons is described by the formula given by Balding and Nichols (1994). The model of Balding and Nichols (1994) is quite general, and the formulae are simple to employ (National Research Council 1996). Harbison and Buckleton (1998) applied the Balding-Nichols formula to a simple mixed sample case. Expressions of likelihood ratios for six com-

Y.-Q. Hu
Department of Applied Mathematics, Southeast University, Nanjing 210018, People's Republic of China

W.K. Fung (✉)
Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, People's Republic of China
e-mail: wingfung@hku.hk,
Tel.: +852-28591988, Fax: +852-28589041

mon cases are reported in Fung and Hu (2002a) when the contributors to a DNA mixture belonged to different ethnic groups. Fung and Hu (2000b, 2002b) discussed the evaluation of match probability in single and multiple racial groups under the NRC-II recommendation 4.1 (National Research Council 1996). Recently, Fukshansky and Bär (2000) constructed a formula for the evaluation of *LR* when the suspect is not tested but relatives are. They considered three different kinds of relationship namely, child-parent, siblings and half-siblings. In this paper, we consider the evaluation o f *LR* incorporating the kinship coefficients (Cotterman 1941; Wenk et al. 1996), when relatives of the suspect are involved in the pool of possible perpetrators. General formulae for calculating the *LR* are given for the following two cases: the case that a relative of the tested suspect was an unknown contributor and the case that two related unknowns were contributors. It is noted that in the first case the relative is an unknown contributor since in some situations the relative concerned may refuse to cooperate or may not be approachable for various reasons, including death.

In this paper, we study the evaluation of DNA mixtures when the relatives are involved as the source contributors. We present a general formula for the evaluation of *LR* which can meet most needs. Two particular results with the presence of relatives are provided. A single unified expression for joint genotypic probabilities is also presented. The implementation of the given formulae and the effect of kinship on the *LR* can be seen from the analysis of one real example.

Likelihood ratio

Suppose that the mixed stain recovered from the scene of the crime and some persons, e.g. the victim and the suspect, are tested with the aim of identifying the true perpetrator(s). The likelihood ratio,

$$LR = \frac{P\left(\text{Evidence}\,\middle|H_p\right)}{P\left(\text{Evidence}\,|H_d\right)},$$

is usually used to evaluate the weight of the evidence regarding whether the suspect has contributed to the mixed sample, where H_p and H_d are the prosecution and defense propositions, and the evidence is the genetic information obtained from the mixed stain and the typed person(s). Following Fung and Hu (2000a, 2001), let K denote the collection of genotypes (not necessarily distinct) of the typed person(s) and M denote the (distinct) genetic profile of the mixed stain. Expressing the evidence as (M,K), we have:

$$LR = \frac{P\left(M,K\,|H_p\right)}{P\left(M,K\,|H_d\right)} = \frac{P\left(M\,|K,H_p\right)}{P\left(M\,|K,H_d\right)}\frac{P\left(K\,|H_p\right)}{P\left(K\,|H_d\right)} = \frac{P\left(M\,|K,H_p\right)}{P\left(M\,|K,H_d\right)}$$

from the third law of probability. The latter equality holds because both hypotheses H_p and H_d contain no different assumptions about the relationship (and origin) of the persons with known genotypes (K), and so $P(K|H_p)$=$P(K|H_d)$. Thus the evaluation of *LR* is induced to the evaluation of the conditional probability $P(M|K,H)$ for some hypothesis H. Under H, let x be the number of unknown contributor(s) and X be their genetic profile. Since the mixture M is contributed by the known and unknown contributors, we have $U\subset X\subset M$, where set U comprises the alleles present in mixture M but not in the genetic profile of the known contributor(s) declared by H, i.e., the alleles in set U have to be contributed by the x unknown contributor(s). For the rest of the paper, we use the notation $P_x(U,M|K)$, instead of $P(M|K,H)$, to express the conditional probability $P(U\subset X\subset M|K)$, and the evaluation of *LR* becomes the evaluation of $P_x(U,M|K)$.

Without loss of generality, assume M=$\{1,2,...,m\}$ and the corresponding allele frequencies be p_1,p_2,...,p_m. By the principle of inclusion and exclusion, we have:

$$\begin{aligned} P_x\left(U,M\,|K\right) &= P\left(U\subset X\subset M\,|K\right) \\ &= P\left(X\subset M\,|K\right) - P\left(\cup_{i\in U}\left(X\subset M\setminus\{i\}\right)|K\right) \\ &= \sum_{M\setminus U\subset C\subset M}(-1)^{|M\setminus C|}\,W\left(C\right) = W\left(M\right) - \sum_{i\in U}W\left(M\setminus\{i\}\right) \\ &\quad + \sum_{i,j\in U}W\left(M\setminus\{i,j\}\right) - \cdots + (-1)^{|U|}\,W\left(M\setminus U\right), \end{aligned} \tag{1}$$

where

$$W(C) = P(X\subset C\,|K)$$

is defined for arbitrary subset C of M satisfying $M\setminus U\subset C\subset M$, $|U|$ is the cardinality of set U with $|U|\le 2x$. It is now clear that the kernel of the evaluation of *LR* is converted into the evaluation of $W(C)$.

Under the Hardy-Weinberg (HW) law, it is obvious that:

$$W(C) = \left(\sum_{i\in C}p_i\right)^{2x}$$

and this leads to the formula reported in Weir et al. (1997) and Fukshansky and Bär (1998):

$$P_x(U,M) = s^{2x} - \sum_{i\in U}(s-p_i)^{2x} + \sum_{i,j\in U}\left(s-p_i-p_j\right)^{2x} - \cdots, \tag{2}$$

where s=p_1+p_2+···+p_m. For the other forms of $W(C)$ regarding dependence and ethnicity, see Fukshansky and Bär (1999, 2000), Fung and Hu (2000a, 2000b, 2001, 2002b).

Based on the form of the right side of Eq. 2, the subset U of M, and the sum s for the allele frequencies in M, we define a function:

$$Q(r,U,s) = s^r - \sum_{i\in U}(s-p_i)^r + \sum_{i,j\in U}\left(s-p_i-p_j\right)^r - \cdots$$

for integer-valued r, and introduce (Fukshansky and Bär 2000):

$$\begin{aligned} L(r,u,s) &\equiv L_\phi(r,u,s) = Q\left(r,\{1,2,\ldots,u\},s\right), \\ L_i(r,u-1,s) &\equiv L_{\{i\}}(r,u-1,s) = Q\left(r,\{1,2,\ldots,u\}\setminus\{i\},s\right), \\ L_{ij}(r,u-2,s) &\equiv L_{\{i,j\}}(r,u-2,s) = Q\left(r,\{1,2,\ldots,u\}\setminus\{i,j\},s\right), \end{aligned}$$

for any distinct $1\le i, j\le u$. It is noted that the calculation of $Q(r,U,s)$ by a computer program is straightforward and so are the calculations of $L(r,u,s)$, $L_i(r,u–1,s)$, and $L_{ij}(r,u–2,s)$.

Table 1 Values of kinship coefficients for commonly encountered relationships

Relationship	k_0	k_1	k_2
Parent-child	0	1/2	0
Siblings	1/4	1/4	1/4
Half-siblings	1/2	1/4	0
Grandparent-child	1/2	1/4	0
Uncle-niece	1/2	1/4	0
Cousins	3/4	1/8	0
Second cousins	15/16	1/32	0
Unrelated	1	0	0

Table 2 Expressions of joint genotype probabilities for seven possible pairs

$$P(R = ii, S = ii) = k_0 p_i^4 + 2k_1 p_i^3 + k_2 p_i^2$$

$$P(R = ii, S = jj) = k_0 p_i^2 p_j^2$$

$$P(R = ii, S = ij) = 2k_0 p_i^3 p_j + 2k_1 p_i^2 p_j$$

$$P(R = ii, S = jk) = 2k_0 p_i^2 p_j p_k$$

$$P(R = ij, S = ij) = 4k_0 p_i^2 p_j^2 + 2k_1 p_i p_j (p_i + p_j) + 2k_2 p_i p_j$$

$$P(R = ij, S = ik) = 4k_0 p_i^2 p_j p_k + 2k_1 p_i p_j p_k$$

$$P(R = ij, S = kl) = 4k_0 p_i p_j p_k p_l$$

Evaluation of *LR* with the inclusion of relatives

It is not uncommon in practice that the relative(s) is(are) involved in the pool of possible contributors and Eq. 2 should be adjusted to meet this need. Let k_0, $2k_1$ and k_2 be the kinship coefficients (Cotterman 1941; Wenk et al. 1996), or equivalently the probabilities that two persons will share 0, 1 or 2 alleles identical by descent (ibd). See Table 1 for values of kinship coefficients for eight commonly encountered relationships. In order to find the conditional probability $W(C)=P(X\subset C|K$ for any $M\backslash U\subset C\subset M$ used in Eq. 1, we derive the following formula for calculating the probability of joint genotypes for two related non-inbreeding individuals $R=r_1r_2$ and $S=s_1s_2$,

$$\begin{aligned} &P(R = r_1r_2, S = s_1s_2) \\ &= k_0 P(S) P(R) + k_1 \left(2 - \delta_{r_1r_2}\right) \left[I_S(r_1) p_{r_2} + I_S(r_2) p_{r_1}\right] p_{s_1} p_{s_2} + k_2 P(S) \delta_{SR} \\ &= k_0 P(S) P(R) + k_1 \left(2 - \delta_{s_1s_2}\right) \left[I_R(s_1) p_{s_2} + I_R(s_2) p_{s_1}\right] p_{r_1} p_{r_2} + k_2 P(S) \delta_{SR}, \end{aligned} \tag{3}$$

where $\mathcal{S}(\mathcal{R})$ is the genetic profile of $S(R)$, for example, $\mathcal{S} = \{s_1, s_2\}$ for a heterozygous $S=s_1s_2$ and $\mathcal{S} = \{s_1\}$ for a homozygous $S=s_1s_1$, I is the indicator function, $\delta_{SR}=1$ if $R=S$ and $\delta_{SR}=0$ if $R\neq S$. Note the second equality in Eq. 3 follows the symmetry of R and S. The proof of Eq. 3 is given in the Appendix. As an application of Eq. 3, we list all seven joint genotype probabilities in Table 2, which coincide with Eqs. 4.13–4.18 of Evett and Weir (1998).

From Eq. 3, we can conclude the following two equations and their derivations are given in the Appendix:

$$P(\mathcal{R} \subset C|S) = k_0 \left(\sum_{i\in C} p_i\right)^2 + k_1 [I_C(s_1) + I_C(s_2)] \sum_{i\in C} p_i + k_2 I_C(s_1) I_C(s_2), \tag{4}$$

$$P(\mathcal{S} \subset C, \mathcal{R} \subset C) = \left(\sum_{i\in C} p_i\right)^2 \left[k_0 \left(\sum_{i\in C} p_i\right)^2 + 2k_1 \sum_{i\in C} p_i + k_2\right]. \tag{5}$$

In the remainder of this section, we also denote

$$A = \{1, \ldots, u\}, \quad B = M\backslash A, \quad Z = \{m+1, m+2, \ldots\},$$

where i and j represent any distinct alleles from A, b_1 and b_2 represent any alleles from B, and z_1 and z_2 represent any alleles from Z.

In the following, we list two cases which are commonly encountered in practice in the interpretation of mixed DNA samples and the corresponding formula for evaluating $P_x(U,M|K)$ are given for $U=\{1,...,u\}$ therein.

Tested suspect with an unknown relative and unknown suspect with a tested relative

Assume that one suspect S was typed in a crime and the proposition about the source contributors of the DNA mixture is:

H: One relative, R, of the suspect and other x–1 unknowns were contributors, where R is not typed. Write K as (S, K_0) and $X = \mathcal{R} \cup X_0$, we have from Eq. 4 for any $M\backslash A\subset C\subset M$

$$\begin{aligned} P(X \subset C|K) &= P(\mathcal{R} \subset C|S)\, P(X_0 \subset C) \\ &= \left(\sum_{i\in C} p_i\right)^{2(x-1)} \left\{k_0 \left(\sum_{i\in C} p_i\right)^2 + k_1 [I_C(s_1) + I_C(s_2)] \sum_{i\in C} p_i \right. \\ &\qquad \left. + k_2 I_C(s_1) I_C(s_2)\right\} \end{aligned} \tag{6}$$

where K_0 is the collection of genotypes of the typed person(s) except S, X_0 is the genetic profile of the unknown contributor(s) except R. Substituting Eq. 6 into Eq. 1 and using the notations L, L_i, L_{ij} introduced previously, we have, for a given hypothesis H:

$$\begin{aligned} P_x(U, M|K) &= k_0 L(2x, u, s) \\ &+ k_1 \left[I_A(s_1) L_{s_1}(2x-1, u-1, s) + I_B(s_1) L(2x-1, u, s)\right] \\ &+ k_1 \left[I_A(s_2) L_{s_2}(2x-1, u-1, s) + I_B(s_2) L(2x-1, u, s)\right] \\ &+ k_2 I_B(s_1) I_B(s_2) L(2x-2, u, s) \\ &+ k_2 [I_M(s_1) I_M(s_2) - I_B(s_1) I_B(s_2)] L_{S\cap A}(2x-2, u-|\mathcal{S}\cap A|, s) \end{aligned} \tag{7}$$

after simplification using the fact that $I_C(s_1)$ is always 0 for any $s_1\in Z$, $I_C(s_1)$ is always 1 for any $s_1\in B$, and $I_C(s_1)$ may take the value 0 or 1 if $s_1\in A$.

Detailed expressions of $P_x(U,M|K)$ are given in Table 3 when suspect S can have seven different kinds of genotypes. Table 3 also shows that the computation of $P_x(U,M|K)$ is relatively simple for the given kinship coefficients k_0, $2k_1$, k_2. In order to find the *LR* using Eq. 7 or Table 3, it is necessary to have a precise specification of the allele i out of the set A. For alleles in sets B and Z, solely the fact of being part of the set, not the precise specification of alleles, is of importance for the calculation of $P_x(U,M|K)$.

Table 3 Expressions for the conditional probability $P_x(U,M|K)$ for a tested suspect S with an unknown relative, or a tested relative R with a non-tested suspect

Case	S/R	Conditional probability
1	ii	$k_0L(2x,u,s)+2k_1L_i(2x-1,u-1,s)+k_2L_i(2x-2,u-1,s)$
2	ij	$k_0L(2x,u,s)+k_1[L_i(2x-1,u-1,s)+L_j(2x-1,u-1,s)]+k_2L_{ij}(2x-2,u-2,s)$
3	ib_1	$k_0L(2x,u,s)+k_1[L(2x-1,u,s)+L_i(2x-1,u-1,s)]+k_2L_i(2x-2,u-1,s)$
4	iz_1	$k_0L(2x,u,s)+k_1L_i(2x-1,u-1,s)$
5	b_1b_2	$k_0L(2x,u,s)+2k_1L(2x-1,u,s)+k_2L(2x-2,u,s)$
6	b_1z_1	$k_0L(2x,u,s)+k_1L(2x-1,u,s)$
7	z_1z_2	$k_0L(2x,u,s)$

Note: $i\neq j\in A$, $b_1,b_2\in B$, and $z_1,z_2\in Z$

Consider a case where the suspect is not tested for some reason, e.g. death, and the suspect's relative is tested instead. The proposition can be formulated as:

H: the suspect (not typed) and other x–1 unknowns were contributors.

In this case the formula for the evaluation of $P_x(U,M|K)$ is the same as Eq. 7 with the interchange of R and S. Thus, detailed expressions about the seven possible genotypes of R can also be referred to Table 3. It is found that the expression of $p(n,k)$ in Table 2 of Fukshansky and Bär (2000) when $R=A_iB$, $E=E_2$ is not complete with the absence of $L_i(2n-1,k-1,s)/4$.

Two related persons were unknown contributors

Here we consider the case where two related persons are declared to have contributed to the mixed stain. The proposition can be written as:

H: two related persons R_1 and R_2, and x–2 unknowns were contributors.

Under this situation, it can be shown from Eqs. 1 and 5 that $P_x(U,M|K)$ has a simple form which is given as:

$$P_x(U,M\,|K) = k_0L(2x,u,s) + 2k_1L(2x-1,u,s) + k_2L(2x-2,u,s). \quad (8)$$

In this case, we do not need a table such as Table 3 of the previous section for expressions on various possible combinations of genotypes.

Application

In this section, we apply Eqs. 7 and 8 to a rape case that happened in Hong Kong (Fung and Hu 2000b). The Profiler PCR-STR system was employed, and the results of the first three loci were selected, because it happened by chance that the combinations of victim and suspect genotypes for these three systems were both heterozygous, both homozygous, and one heterozygous and one homozygous, respectively, thereby giving a range of examples (Fung and Hu 2000b). The details can be referred to in Table 4. The following two explanations to the evidence are first considered:

- H_p: contributors were the victim and the suspect
- H_{d1}: contributors were the victim and one relative of the suspect.

Table 4 Alleles detected in a rape case in Hong Kong

Locus	Mixture (M)	Victim (V)	Suspect (S)	Frequency
D3S1358	14		14	0.033
	15	15		0.331
	17		17	0.239
	18	18		0.056
vWA	16		16	0.155
	18	18		0.158
FGA	20	20		0.042
	24	24		0.166
	25		25	0.106

Here, the victim, the suspect and the unknown are assumed to come from the same ethnic population. The relationship between the relative and the suspect is described by the kinship coefficients k_0, $2k_1$, k_2. Table 5 lists the likelihood ratios for six commonly encountered relationships, including the unrelated case. As we can see from Table 5, the effect of kinship on *LR* is substantial. For example in locus D3S1358, the maximum *LR* value (63.40) is 20 times the minimum one (3.11).

If the evidence was collected from somewhere other than the victim's body (Fung and Hu 2000b), another set of explanations should be used, which is:

- H_p: contributors were the victim and the suspect
- H_{d2}: contributors were one relative of the suspect and one unknown.

The *LRs* are also reported in Table 5, which are larger than those given earlier. However, the effect of kinship on *LR* is not as large as before. As in the previous case, the relationship of kinship has the effect of giving a smaller *LR* (compared with the unrelated situation), with the smallest *LR* for the siblings relationship.

Finally, we consider the following explanations about who the source contributors of the mixed stain were:

- H_p: contributors were the victim and the suspect
- H_{d3}: contributors were two related persons (relatives).

Equation 8 can be applied to evaluate the *LR* for various relationships (Table 5). Unlike the other two previous cases, under the current set of hypotheses, the *LR* at locus D3S1358 for the siblings relationship is the highest (*LR*=1,140), while that for the unrelated situation is the

Table 5 The effect of kinship relationship on the likelihood ratios in a rape case of Hong Kong (see Table 4), where the prosecution proposition is H_p: contributors were the victim and the suspect, and the defense proposition takes three different forms

Defense proposition	Relationship	Likelihood ratios			Overall
		D3S1358	vWA	FGA	
H_{d1}	Parent-child	7.35	3.19	3.18	74.56
	Siblings	3.11	2.35	2.38	17.39
	Half-siblings[a]	13.18	5.18	5.42	370.04
	Cousins	21.82	7.52	8.33	1,366.84
	Second cousins	42.94	11.36	13.99	6,824.30
	Unrelated	63.40	13.70	18.07	15,695.24
H_{d2}	Parent-child	66.11	37.12	113.83	279,339
	Siblings	56.47	29.36	116.62	193,351
	Half-siblings[a]	107.33	56.60	172.87	1,050,164
	Cousins	155.94	76.73	233.39	2,792,575
	Second cousins	236.14	104.64	316.49	7,820,369
	Unrelated	285.01	119.08	359.11	12,187,830
H_{d3}	Parent-child	[b]	43	226	9,718
	Siblings	1140	39	343	15,249,780
	Half-siblings[a]	570	64	277	10,104,960
	Cousins	380	83	313	9,872,020
	Second cousins	304	107	346	11,254,688
	Unrelated	285	119	359	12,175,485

H_{d1}: contributors were the victim and one relative of the suspect.
H_{d2}: contributors were one relative of the suspect and one unknown.
H_{d3}: contributors were two related persons (relatives).
[a]The same as the grandparent-child and the uncle-niece relationship.
[b]The parent-child relationship is impossible for a mixture of four distinct alleles.

lowest (*LR*=285). However, the lowest *LRs* at vWA and FGA go to the siblings and the parent-child relationships, respectively. The effect of kinship is mixed under this particular set of hypotheses.

Concluding remarks

A formula is derived for calculating the match probability when one relative of the suspect was the contributor of the mixed stain. We assume that the other unknown contributors are unrelated to the suspect and the population satisfies the Hardy-Weinberg law and linkage equilibrium. The other case is also studied and it also involves the relationship of two persons (relatives).

If we want to discuss the case where two or more relatives of the suspect are involved in the pool of perpetrators, we first have to develop the theory of kinship among three or more persons. This is a much more complicated task and we are not aware of any general theory in the literature. However, cases with two or more relatives of the suspect involved are not so common in practice.

The independence assumption of alleles may be relaxed to allow for the possible existence of population substructure. We are working in this direction and hope to report our findings in the near future.

Acknowledgements The authors are very grateful to two referees for valuable comments which improved the presentation of the paper. This work was supported in part by the SEU Science Foundation (9207011146) and the Hong Kong RGC Competitive Earmarked Research Grant.

Appendix

A1 Proof of Eq. 3

As in Evett and Weir (1998), suppose individual R has alleles a and b and S has alleles c and d at some autosomal locus, where alleles a and c are of paternal and alleles b and d are of maternal origin. It is noted that alleles a and b take a unique value for a homozygous R and two values for a heterozygous R, and also for S . Then the kinship coefficients k_0, $2k_1$, k_2 can be expressed as k_0=P(no ibd allele), $2k_1$=$P(a{\equiv}c)$+$P(a{\equiv}d)$+$P(b{\equiv}c)$+$P(b{\equiv}d)$, and k_2=$P(a{\equiv}c,b{\equiv}d)$+$P(a{\equiv}d,b{\equiv}c)$, where the equivalence sign ($\equiv$) is used to indicate an ibd (identical by descent) relationship.

Let *IBDA* be the ibd alleles between the two individuals R and S, then all the possibilities for *IBDA* are: *IBDA*=none, *IBDA*=r_1, *IBDA*=r_2 (if $r_2{\neq}r_1$), and *IBDA*=r_1,r_2. It is obvious that:

$$P\,(R = r_1r_2, S = s_1s_2, IBDA = \text{none}) = k_0P(R)P(S). \tag{9}$$

For two homozygous R and S where R=S=ii, it is concluded that:

$$\begin{aligned} P(R = ii, S = ii, IBDA = i, i) &= P(a = i, b = i; c = i, d = i, IBDA = i, i) \\ &= P(a = i, b = i; c = i, d = i, a \equiv c, b \equiv d) \\ &\quad + P(a = i, b = i; c = i, d = i, a \equiv d, b \equiv c) \\ &= [P(a \equiv c, b \equiv d) + P(a \equiv d, b \equiv c)]\, p_i^2. \end{aligned}$$

For two heterozygous R and S where R=S=ij, it is concluded that:

$$\begin{aligned} P(R = ij,\ S = ij,\ IBDA = i, j) &= P(a = i,\ b = j;\ c = i,\ d = j,\ IBDA = i, j) \\ &\quad + P(a = j,\ b = i;\ c = i,\ d = j,\ IBDA = i, j) \\ &\quad + P(a = i,\ b = j;\ c = j,\ d = i,\ IBDA = i, j) \\ &\quad + P(a = j,\ b = i;\ c = j,\ d = i,\ IBDA = i, j) \\ &= P(a = i,\ b = j;\ c = i,\ d = j,\ a \equiv c,\ b \equiv d) \\ &\quad + P(a = j,\ b = i;\ c = i,\ d = j,\ a \equiv d,\ b \equiv c) \\ &\quad + P(a = i,\ b = j;\ c = j,\ d = i,\ a \equiv d,\ b \equiv c) \\ &\quad + P(a = j,\ b = i;\ c = j,\ d = i,\ a \equiv c,\ b \equiv d) \\ &= [P(a \equiv c,\ b \equiv d) + P(a \equiv d,\ b \equiv c)] \cdot 2p_i p_j. \end{aligned}$$

Thus, we obtain

$$P\left(R = r_1r_2,\ S = s_1s_2,\ IBDA = r_1, r_2\right) = k_2P(S)\delta_{RS}. \tag{10}$$

Similarly, we have

$$P\left(R = r_1r_2,\ S = s_1s_2,\ IBDA = r_1\right) = 2k_1I_S(r_1)p_{r_2}p_{s_1}p_{s_2}. \tag{11}$$

Based on Eqs. 9–11, we have

$$\begin{aligned} P\left(R = r_1r_2,\ S = s_1s_2\right) &= P\left(R = r_1r_2,\ S = s_1s_2,\ IBDA = \text{none}\right) \\ &\quad + P\left(R = r_1r_2,\ S = s_1s_2,\ IBDA = r_1\right) \\ &\quad + \left(1 - \delta_{r_1r_2}\right)P\left(R = r_1r_2,\ S = s_1s_2,\ IBDA = r_2\right) \\ &\quad + P\left(R = r_1r_2,\ S = s_1s_2,\ IBDA = r_1, r_2\right) \\ &= k_0P(R)P(S) + k_1\left[2I_S(r_1)\,p_{r_2} + 2\left(1 - \delta_{r_1r_2}\right)I_S(r_2)\,p_{r_1}\right] \\ &\quad p_{s_1}p_{s_2} + k_2P(S)\delta_{RS}. \end{aligned}$$

Thus, Eq. 3 holds.

A2 Proof of Eqs. 4 and 5

We first give the following results:

$$\sum_{\mathcal{R}\subset C} P(R) = \left(\sum_{i\in C} p_i\right)^2, \tag{12}$$

$$\sum_{\mathcal{R}\subset C} I_{\mathcal{R}}(s_1)\,p_{r_1}p_{r_2} = I_C(s_1)\,p_{s_1}\sum_{i\in C} p_i, \tag{13}$$

$$\sum_{\mathcal{R}\subset C} \delta_{SR} = I_C(s_1)\,I_C(s_2). \tag{14}$$

Note that Eqs. 12 and 14 are straightforward. For $\mathcal{R}\subset C$, the genotype R may be homozygous or heterozygous, viz. $R{=}r_1r_1$ or $R{=}r_1r_2$, where $r_1{\neq}r_2$ and $r_1,r_2{\in}C$. In order to guarantee $I_{\mathcal{R}}(s_1) = 1$, we can assume $r_1{=}s_1$ without loss of generality. So Eq. 13 follows from:

$$\begin{aligned} \sum_{\mathcal{R}\subset C} I_{\mathcal{R}}(s_1)\,p_{r_1}p_{r_2} &= I_C(s_1)\left(p_{s_1}^2 + \sum_{i\neq s_1, i\in C} p_{s_1}p_i\right) \\ &= I_C(s_1)\,p_{s_1}\sum_{i\in C} p_i. \end{aligned}$$

According to Eq. 3, we like to find the summation of joint genotype probability $P(R,S)$ over all $\mathcal{R}\subset C$ for any given set C of M. It is observed from Eq. 3 that the summation comes down to find the summations corresponding to the coefficients of k_0, k_1 and k_2 therein over all $\mathcal{R}\subset C$, designated as T_0, T_1, T_2, respectively. From Eqs. 12–14, it is obvious that:

$$\begin{aligned} T_0 &= P(S)\sum_{\mathcal{R}\subset C} P(R) = P(S)\left(\sum_{i\in C} p_i\right)^2, \\ T_1 &= \sum_{\mathcal{R}\subset C}\left(2 - \delta_{r_1r_2}\right)\left[I_{\mathcal{S}}(r_1)\,p_{r_2} + I_{\mathcal{S}}(r_2)\,p_{r_1}\right]p_{s_1}p_{s_2} \\ &= \sum_{\mathcal{R}\subset C}\left(2 - \delta_{s_1s_2}\right)\left[I_{\mathcal{R}}(s_1)\,p_{s_2} + I_{\mathcal{R}}(s_2)\,p_{s_1}\right]p_{r_1}p_{r_2} \\ &= \left(2 - \delta_{s_1s_2}\right)p_{s_1}p_{s_2}\left[I_C(s_1) + I_C(s_2)\right]\sum_{i\in C} p_i \\ &= P(S)\left[I_C(s_1) + I_C(s_2)\right]\sum_{i\in C} p_i, \\ T_2 &= P(S)\sum_{r_1,r_2\in C}\delta_{SR} = P(S)I_C(s_1)\,I_C(s_2). \end{aligned}$$

Therefore we have:

$$\begin{aligned} \sum_{\mathcal{R}\subset C} P(R,S) &= k_0P(S)\left(\sum_{i\in C} p_i\right)^2 + k_1P(S)\left[I_C(s_1) + I_C(s_2)\right]\sum_{i\in C} p_i \\ &\quad + k_2P(S)I_C(s_1)\,I_C(s_2), \end{aligned}$$

and

$$\begin{aligned} P(\mathcal{R}\subset C, \mathcal{S}\subset C) &= \sum_{\mathcal{S}\subset C}\sum_{\mathcal{R}\subset C} P(R,S) = k_0\left(\sum_{i\in C} p_i\right)^2\sum_{\mathcal{S}\subset C} P(S) \\ &\quad + k_1\sum_{i\in C} p_i\sum_{\mathcal{S}\subset C} P(S)\left[I_C(s_1) + I_C(s_2)\right] \\ &\quad + k_2\sum_{\mathcal{S}\subset C} P(S)I_C(s_1)\,I_C(s_2) \\ &= k_0\left(\sum_{i\in C} p_i\right)^2\left(\sum_{i\in C} p_i\right)^2 + 2k_1\sum_{i\in C} p_i\left(\sum_{i\in C} p_i\right)^2 \\ &\quad + k_2\left(\sum_{i\in C} p_i\right)^2. \end{aligned}$$

Thus, Eq. 5 holds.

Furthermore,

$$\begin{aligned} P(\mathcal{R}\subset C\,|\mathcal{S}) &= \sum_{\mathcal{R}\subset C} P(R,S)/P(S) \\ &= k_0\left(\sum_{i\in C} p_i\right)^2 + k_1\left[I_C(s_1) + I_C(s_2)\right]\sum_{i\in C} p_i \\ &\quad + k_2I_C(s_1)\,I_C(s_2). \end{aligned}$$

This is just Eq. 4.

References

Ayres KL (2000) Relatedness testing in subdivided populations. Forensic Sci Int 114:107–115

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

Belin TR, Gjertson DW, Hu MY (1997) Summarizing DNA evidence when relatives are possible suspects. J Am Stat Assoc 92: 706–716

Brookfield JFY (1994) The effect of relatives on the likelihood ratio associated with DNA profile evidence in criminal cases. J Forensic Sci Soc 34:193–197

Cotterman CW (1941) Relatives and human genetic analysis. Sci Month 53:227–234

Curran JM, Triggs CM, Buckleton J, Weir BS (1999) Interpreting DNA mixtures in structured populations. J Forensic Sci 44:987–995

Donnelly P (1995) Nonindependence of matches at different loci in DNA profiles: quantifying the effect of close relatives on the match probability. Heredity 75:26–34

Evett IW (1992) Evaluating DNA profiles in case where the defense is "It is my brother". J Forensic Sci Soc 32:5–14

Evett IW, Weir BS (1998) Interpreting DNA evidence. Sinauer Associates, Sunderland, Massachusetts

Fukshansky N, Bär W (1998) Interpreting forensic DNA evidence on the basis of hypotheses testing. Int J Legal Med 111:62–66

Fukshansky N, Bär W (1999) Biostatistical evaluation of mixed stains with contributors of different ethnic origin. Int J Legal Med 112:383–387

Fukshansky N, Bär W (2000) Biostatistics for mixed stains: the case of tested relatives of a non-tested suspect. Int J Legal Med 114:78–82

Fung WK, Hu YQ (2000a) Interpreting forensic DNA mixtures: allowing for uncertainty in population substructure and dependence. J R Statist Soc A 163:241–254

Fung WK, Hu YQ (2000b) Interpreting DNA mixtures based on the NRC-II recommendation 4.1. Forensic Sci Commun. Available at http://www.fbi.gov/hq/lab/fsc/backissu/oct2000/fung.htm

Fung WK, Hu YQ (2001) The evaluation of mixed stains from different ethnic origin: general result and common cases. Int J Legal Med 115:48–53

Fung WK, Hu YQ (2002a) The statistical evaluation of DNA mixtures with contributors from different ethnic groups. Int J Legal Med 116:79–86

Fung WK, Hu YQ (2002b) Evaluating mixed stains with contributors of different ethnic groups under the NRC-II Recommendation 4.1. Statist Med (in press)

Fung WK, Chung Y, Wong D (2002) Power of exclusion revisited: probability of excluding relatives of the true father from paternity. Int J Legal Med 116:64–67

Harbison SA, Buckleton JS (1998) Applications and extensions of subpopulation theory: a caseworkers guide. Sci Justice 38:249–254

Lee JW, Lee HS, Park M, Hwang JJ (1999) Paternity probability when a relative of the father is an alleged father. Sci Justice 39:223–230

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

Sjerps M, Kloosterman AD (1999) On the consequences of DNA profile mismatches for close relatives of an excluded suspect. Int J Legal Med 112:176–180

Weir BS, Triggs CM, Starling L, Stowell LI, Walsh KAJ, Buckleton J (1997) Interpreting DNA mixtures. J Forensic Sci 42:213–222

Wenk RE, Traver M, Chiafari FA (1996) Determination of sibship in any two persons. Transfusion 36:259–262