

## Statistical Interpretation in Making DNA-based Identifications of Mass Victims

Kyoung-Jin Shin, Hwan Young Lee, Woo Ick Yang, Eunho Ha<sup>2</sup>

*Department of Forensic Medicine, Yonsei University College of Medicine, Seoul, Korea*

*<sup>2</sup>Department of Information and Statistics, Yonsei University at Wonju-si, Gangwon-do, Korea*

DNA profiles have been increasingly used as the most reliable means to identify remains from war or mass disaster. To establish the identity with such a large set of victims, special care should be taken to correlate remains with correct family references while avoiding coincidental match between non-relatives. Therefore we address here relevant statistical and combinatorial issues in the DNA identification of mass victims. A simple and general formula for the likelihood ratio governing any potential kinship between two DNA profiles was presented, and for that purpose, the probabilities that a given relative and an individual share autosomal identical-by-descent alleles were calculated. In addition, a method dealing with the allele drop-out in kinship analysis and the estimation of a cold hit were discussed.

**Key words** : mass victims, kinship, relatives, probability, likelihood, allele drop-out, cold hit

2  
2  
: 2008 4 28  
: 2008 5 14  
( : M10740030002 -  
07N4003 - 00210)  
:  
(220 - 710)  
234  
: (033) 760 - 2268  
FAX : (033) 763 - 4323  
E - mail: statha@yonsei.ac.kr

가  
가 가  
가  
가 , 가  
(likelihood  
(http://www.massfatality.dna.gov/  
Chapter12/). (direct matching) ratio; LR), (kinship index)  
가 , 가  
(match probability) 가  
(reference sample) (allele drop - out)<sup>1)</sup>,  
가 , 가  
(indirect matching) 가 가  
가 ,  
가 ,

cold hit

1.

$$\begin{aligned}
 LR &= \frac{P(G_P, G_Q | H_p)}{P(G_P, G_Q | H_d)} \\
 &= \frac{P(G_Q | G_P, H_p)P(G_P | H_p)}{P(G_Q | G_P, H_d)P(G_P | H_d)} \\
 &= \frac{P(G_Q, G_P | H_p)}{P(G_Q, G_P | H_d)} \\
 &= \frac{P(G_Q | G_P, H_p)}{P(G_Q | H_d)}
 \end{aligned}$$

가 가 , 가 가 가 가

identical - by - descent

(IBD)

(paternity index)

Z<sub>0</sub>: 0 alleles are identical - by - descent, with probability P<sub>0</sub>

Z<sub>1</sub>: 1 allele are identical - by - descent, with probability P<sub>1</sub>

Z<sub>2</sub>: 2 alleles are identical - by - descent, with probability P<sub>2</sub>

, 6.25

3 가

가

, Balding Nichols

2, 3)

screening

P Q

(genotype) G<sub>p</sub> G<sub>o</sub> H<sub>d</sub>(P Q가  
가 ) H<sub>p</sub>( 가 )

$$\begin{aligned}
 P(G_Q | G_P, H_p) &= P(G_Q | G_P, Z_0, H_p)P_0 + P(G_Q | G_P, Z_1, H_p)P_1 \\
 &\quad + P(G_Q | G_P, Z_2, H_p)P_2
 \end{aligned}$$

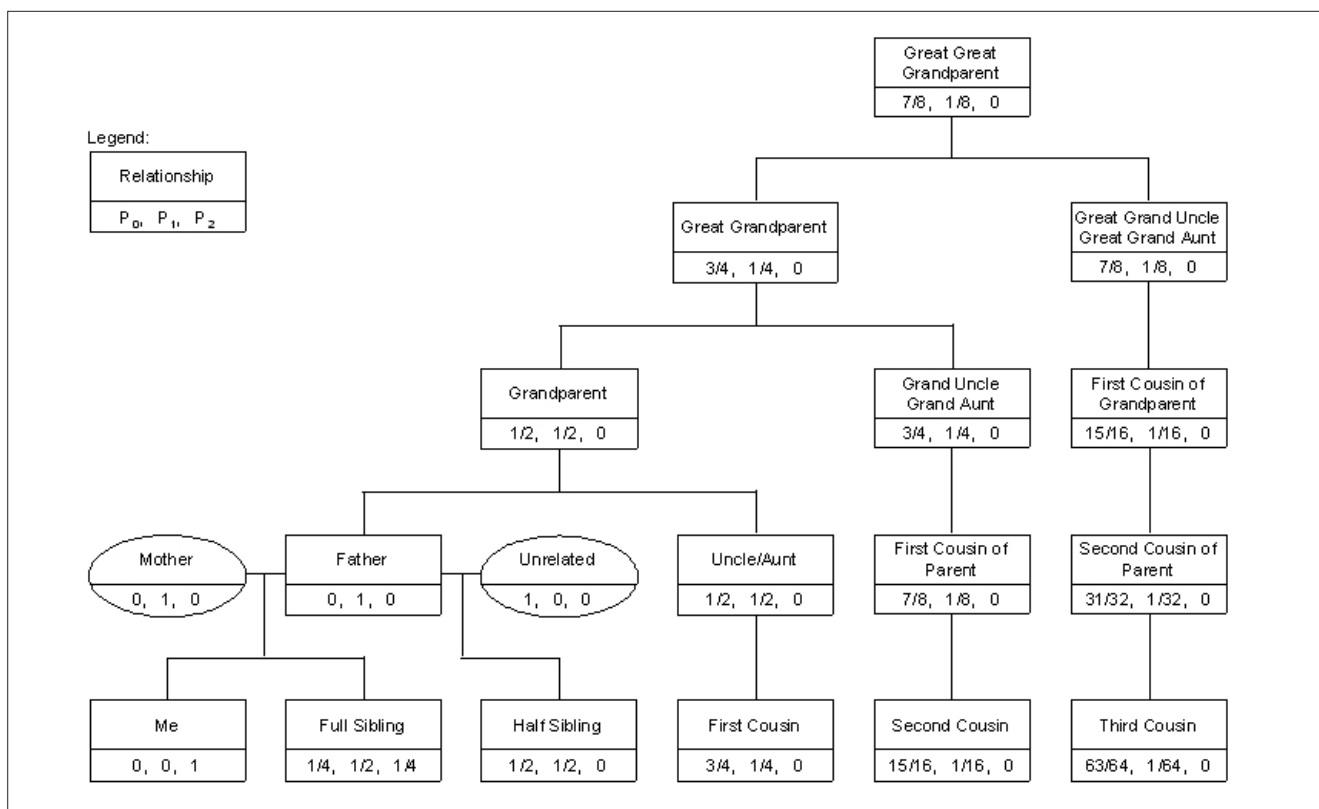


Fig. 1. The probabilities, P<sub>0</sub>, P<sub>1</sub> and P<sub>2</sub> that a given relative and me share 0, 1, or 2 pairs of autosomal STR alleles identical-by-descent

가

IBD

8

가

0, 1, 2

IBD

가

$P_0 = \frac{1}{4}, P_1 = \frac{1}{2}, P_2 = \frac{1}{4}$

, Bieber

DNA

, Brenner Weir

1

4, 5)

ab, cd

가

5)

1: ac, 2: ad, 3: bc, 4: bd

가

가

가

(false - positive)

가

가

2

가

Y-STR

가

$$u = \frac{u_1 + u_2 + u_3 + u_4}{4}, w = \frac{u_1 u_4 + u_2 u_3}{2}$$

가

LR =  $P_0 + UP_1 + WP_2$

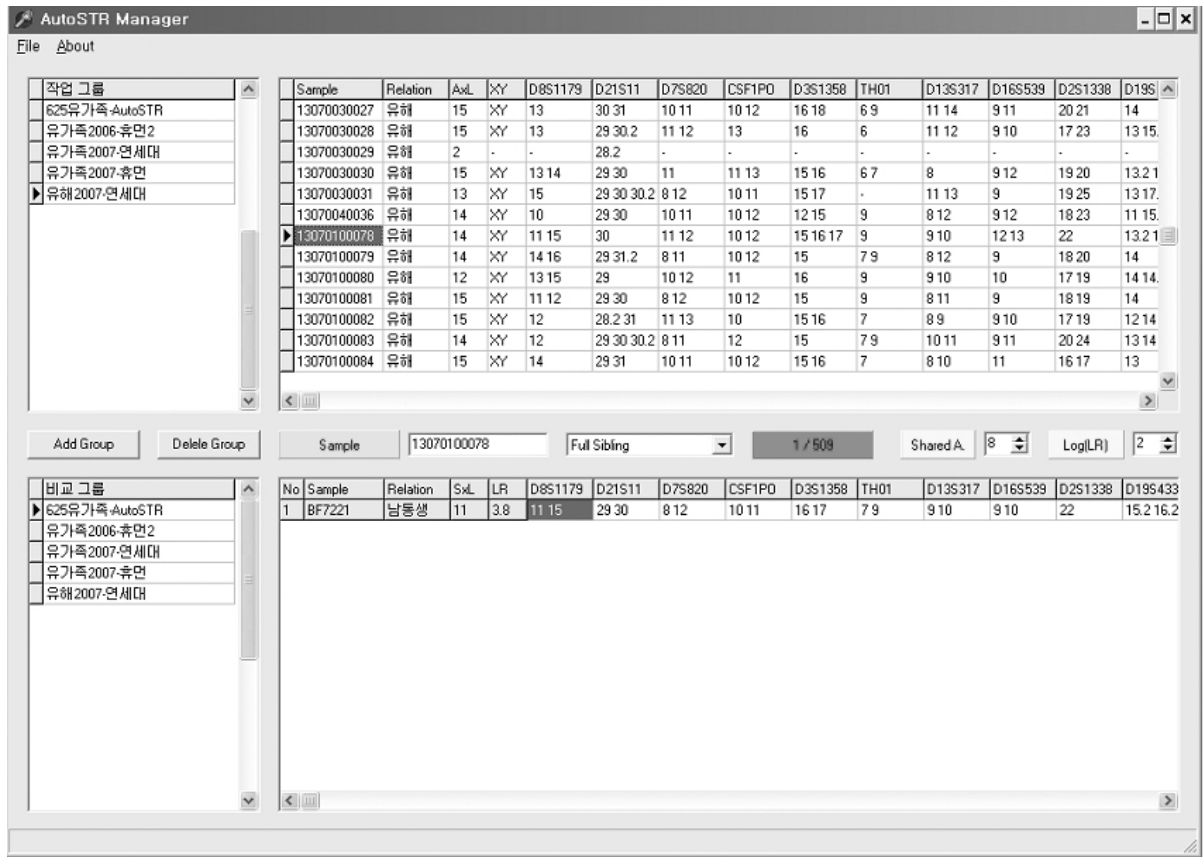


Fig. 2. A computer program for calculating likelihood ratio between two individuals under given relationship

2. 가 가

가 , 가 . DNA mini-STR PCR system F 가 , D = 1/2

6.7), 6.25 PCR system 가

STR DNA 가 STR 가

1), 3. cold hit

가 가

Buckleton Triggs 가 2 8), (closed forms)

가 3 가 (open forms)

(interpol DVI guide, <http://www.interpol.int>).

STR Y-STR 2 가 가

Buckleton Triggs 가 가 6.25

13

Table 1. The likelihood ratio for the proposition that the individuals ( $P$  and  $Q$ ) are related ( $H_p$ ) versus unrelated ( $H_d$ )

$G_P$	$G_Q$	$LR = \frac{P(G_Q   G_P, H_p)}{P(G_Q   H_d)}$
$aa$	$aa$	$P_0 + \frac{P_1}{p_a} + \frac{P_2}{p_a^2}$
	$ab$	$P_0 + \frac{P_1}{2p_a}$
	$bb$ or $bc$	$P_0$
$ab$	$aa$	$P_0 + \frac{P_1}{2p_a}$
	$ab$	$P_0 + \frac{P_1}{4} \left( \frac{1}{p_a} + \frac{1}{p_b} \right) + \frac{P_2}{2p_a p_b}$
	$ac$	$P_0 + \frac{P_1}{4p_a}$
	$cc$ or $cd$	$P_0$

$P_0, P_1$ , and  $P_2$ ; the probabilities that  $P$  and  $Q$  share 0, 1, or 2 pairs of autosomal STR alleles identical-by-descent

(http://www.withcountry.mil.kr/),

X

가

- 가

가

$$P(X=k) = \frac{\binom{M}{k} \times \binom{N-M}{n-k}}{\binom{N}{n}} \quad (k = 0, 1, \dots, n) \quad (N-M, k, n, M, n, N)$$

N

가  
가 가

M

- 가

n

cold hit

가

가

$$E(X) = n \times \frac{M}{N}$$

가

가

가

**Table 2.** The likelihood ratio (LR) for the proposition that the individuals (P and Q) are related ( $H_r$ ) versus unrelated ( $H_d$ ) where a particular allele in the profile P has dropped out ( $aN$ )

$G_P$	$G_Q$	$LR = \frac{P(G_Q   G_P, H_r)}{P(G_Q   H_d)}$
$aN$	$aa$	$P_0 + \frac{(D + p_a)P_1 + (1 + D)P_2}{p_a(2D + \overline{D}p_a)}$
	$ab$	$P_0 + \frac{(D + (1 + D)p_a)P_1 + 2DP_2}{2p_a(2D + \overline{D}p_a)}$
	$bb \text{ or } bc$	$P_0 + \frac{DP_1}{2D + \overline{D}p_a}$

$D$ ; the probability that a particular allele in the profile P has dropped out

$P_0, P_1, P_2$ ; the probabilities that P and Q share 0, 1, or 2 pairs of autosomal STR alleles identical-by-descent  
LR equation from Buckleton *et al.* of which some notations are modified<sup>8)</sup>

**Table 3.** An example of likelihood ratio (LR) calculation for full siblings with probabilities of dropout:  $D = 0, \frac{1}{4}$  and  $\frac{1}{2}$

STR Locus	Genotype		LR (Sibling Index)		
	Victim	Brother	$D=0$	$D=\frac{1}{4}$	$D=\frac{1}{2}$
D8S1179	11-15	11-15	11.962	11.962	11.962
<b>D21S11</b>	<b>30</b>	<b>29-30</b>	<b>0.924</b>	<b>1.084</b>	<b>1.135</b>
D7S820	11-12	8-12	0.729	0.729	0.729
CSF1PO	10-12	10-11	0.750	0.750	0.750
D3S1358	16-17	16-17	3.514	3.514	3.514
<b>TH01</b>	<b>9</b>	<b>7-9</b>	<b>0.733</b>	<b>0.873</b>	<b>0.931</b>
D13S317	9-10	9-10	8.854	8.854	8.854
D16S539	12-13	9-10	0.250	0.250	0.250
<b>D2S1338</b>	<b>22</b>	<b>22</b>	<b>336.985</b>	<b>31.200</b>	<b>22.756</b>
D19S433	13.2-14	15.2-16.2	0.250	0.250	0.250
vWA	19	15-19	3.124	3.345	3.363
TPOX	-	8-11	-	-	-
D18S51	11-15	13-15	0.922	0.922	0.922
<b>D5S818</b>	<b>11</b>	<b>9-10</b>	<b>0.250</b>	<b>0.420</b>	<b>0.466</b>
<b>FGA</b>	<b>23</b>	<b>23-25</b>	<b>1.435</b>	<b>1.625</b>	<b>1.661</b>
Cumulative LR			2,996.3	791.3	734.3

The STR loci and alleles which have possibility of allele dropout, and their LR are represented in bold italic.

LRs were calculated using allele frequencies of 15 autosomal STRs as reported in Han *et al.*<sup>9,10)</sup>

, 6.25 가 100,000 (M)  
 200 (M) 가 1,000 (n)  
 , 2

$(= \frac{200 \times 1,000}{100,000})$  , 6.25  
 가 가

, 가 가  
 , 2 Y-STR  
 가

가 screening

PCR system  
 가  
 cold hit

가 ,  
 6.25 가 가  
 가 가

1. Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J. An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forensic Sci Int* 2000;112:17-40.
2. Buckleton JS, Triggs CM, Walsh SJ. *Forensic DNA Evidence Interpretation*. Boca Raton, FL: CRC Press; 2005.
3. Balding DJ, Nichols RA. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Sci Int* 1994;64:125-40.
4. Brenner CH, Weir BS. Issues and strategies in the DNA identification of World Trade Center victims. *Theor Popul Biol* 2003;63:173-8.
5. Bieber FR, Brenner CH, Lazer D. Human genetics. Finding criminals through DNA of their relatives. *Science* 2006;312:1315-6.
6. Butler JM, Shen Y, McCord BR. The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J Forensic Sci* 2003;48:1054-64.
7. Park MJ, Lee HY, Chung U, Kang SC, Shin KJ. Y-STR analysis of degraded DNA using reduced-size amplicons. *Int J Legal Med* 2007;121:152-7.
8. Buckleton J, Triggs C. Dealing with allelic dropout when reporting the evidential value in DNA relatedness analysis. *Forensic Sci Int* 2006;160:134-9.
9. Han GR, Lee YW, Lee HL, et al. A Korean population study of the nine STR loci FGA, VWA, D3S1358, D18S51, D21S11, D8S1179, D7S820, D13S317 and D5S818. *Int J Legal Med* 2000;114:41-4.
10. Han GR, Kwon JS, Yang YS, Choi JH, Kim CY, Shin KJ. Population data of the COfiler STR loci in Koreans. *J Forensic Sci* 2004;49:176-7.