



Statistical Interpretation in Making DNA-based Identification of Mass Victims

Kyoung-Jin Shin¹, Hwan Young Lee¹, Woo Ick Yang¹, Eunho Ha²

¹Dept. of Forensic Medicine, Yonsei University College of Medicine

²Dept. of Information and Statistics, Yonsei University



DNA-based Identification of Mass Victims

- Direct Matching
 - assesses the probability or likelihood that a DNA profile from a victim's remains and a profile developed from a personal item known to belong to a missing individual would share—by chance—the same DNA profile

- Indirect Matching
 - uses methods of formal genetic kinship

<http://www.massfatality.dna.gov/Chapter12/>



Consideration for DNA-based Identification of Mass Victims

- Calculation of likelihood ratio (LR)
 - Various kinship
 - Numerous families
- Allele drop-in and drop-out
- Estimation of cold hit



Evaluation of DNA Analysis

Odds Form of Bayes' Theorem

$$\frac{P(H_p|E)}{P(H_d|E)} = \frac{P(E|H_p)}{P(E|H_d)} \times \frac{P(H_p)}{P(H_d)}$$

Posterior Odds = Likelihood Ratio × Prior Odds



Kinship probability

$$P(H_p | E) = \frac{LR \times P(H_p)}{LR \times P(H_p) + [1 - P(H_p)]}$$

if prior probability $P(H_p) = 0.5$,

$$P(H_p | E) = \frac{LR}{LR + 1}$$



1. G_p and G_q be the genotypes of two individuals P and Q

2. H_p : P and Q are related

(for example, parent - child, full - sibs, cousins, etc.)

H_d : P and Q are unrelated

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



Use of the IBD Probabilities

According to Balding and Nichols (Forensic Sci Int.1994;64 : 125 - 40)

Z_0 : 0 alleles are identical -by- descent with probability P_0

Z_1 : 1 allele is identical -by- descent with probability P_1

Z_2 : 2 alleles are identical -by- descent with probability P_2

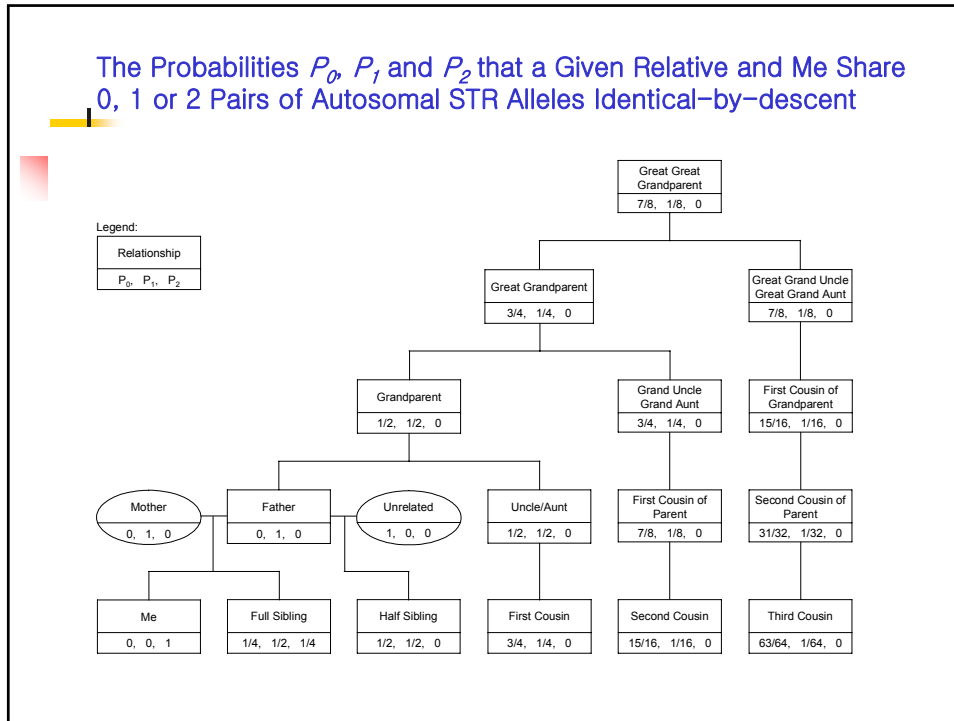
$$\begin{aligned}
 P(G_Q|G_P, H_p) &= P(G_Q|G_P, Z_0, H_p)P_0 \\
 &\quad + P(G_Q|G_P, Z_1, H_p)P_1 \\
 &\quad + P(G_P|G_P, Z_2, H_p)P_2
 \end{aligned}$$

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

The Probabilities P_0 , P_1 and P_2 that a Given Relative and Me Share 0, 1 or 2 Pairs of Autosomal STR Alleles Identical-by-descent



General Formula for LR Calculation

According to Brenner and Weir (Theor Popul Biol. 2003;63 : 173 -8)

1. a) If two genotypes with suspected relationships are ab and cd , define the four mating combinations 1: ac ; 2: ad ; 3: bc ; 4: bd .
- b) If the two alleles in the i th pair are the same type, then u_i is the reciprocal of the frequency of that allele. Otherwise, $u_i = 0$.

$$2. a) U = \frac{u_1 + u_2 + u_3 + u_4}{4}$$

$$b) W = \frac{u_1 u_4 + u_2 u_3}{2}$$

$$3. LR = P_0 + UP_1 + WP_2$$



Common IBD Probabilities and LR

Relationship	P_0	P_1	P_2	$LR = P_0 + UP_1 + WP_2$
Not related	1	0	0	$LR = 1$
Same individual	0	0	1	$LR = W$ (Discrimination index)
Parent-child	0	1	0	$LR = U$ (Paternity Index)
Grandparent-grandchild	$\frac{1}{2}$	$\frac{1}{2}$	0	$LR = \frac{1}{2} + \frac{1}{2} U$
Full sibs	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$LR = \frac{1}{4} + \frac{1}{2} U + \frac{1}{4} W$
Half sibs	$\frac{1}{2}$	$\frac{1}{2}$	0	$LR = \frac{1}{2} + \frac{1}{2} U$
Uncle-nephew	$\frac{1}{2}$	$\frac{1}{2}$	0	$LR = \frac{1}{2} + \frac{1}{2} U$
First cousin	$\frac{3}{4}$	$\frac{1}{4}$	0	$LR = \frac{3}{4} + \frac{1}{4} U$

AutoSTR Manager

File About

Sample	Relation	AvL	XY	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S
13070030027	유해	15	XY	13	30 31	10 11	10 12	16 18	6 9	11 14	9 11	20 21	14
13070030028	유해	15	XY	13	29 30.2	11 12	13	16	6	11 12	9 10	17 23	13 15
13070030029	유해	2	.	.	28.2
13070030030	유해	15	XY	13 14	29 30	11	11 13	15 16	6 7	8	9 12	19 20	13 21
13070030031	유해	13	XY	15	29 30 30.2	8 12	10 11	15 17	.	11 13	9	19 25	13 17
13070040036	유해	14	XY	10	29 30	10 11	10 12	12 15	9	8 12	9 12	18 23	11 15
13070100078	유해	14	XY	11 15	30	11 12	10 12	15 16 17	9	9 10	12 13	22	13 21
13070100079	유해	14	XY	14 16	29 31 2	8 11	10 12	15	7 9	8 12	9	18 20	14
13070100080	유해	12	XY	13 15	29	10 12	11	16	9	9 10	10	17 19	14 14
13070100081	유해	15	XY	11 12	29 30	8 12	10 12	15	9	8 11	9	18 19	14
13070100082	유해	15	XY	12	28 2 31	11 13	10	15 16	7	8 9	9 10	17 19	12 14
13070100083	유해	14	XY	12	29 30 30.2	8 11	12	15	7 9	10 11	9 11	20 24	13 14
13070100084	유해	15	XY	14	29 31	10 11	10 12	15 16	7	8 10	11	16 17	13

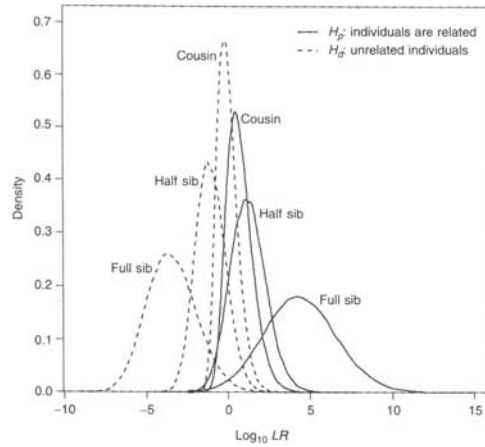
Add Group Delete Group Sample: 13070100078 Full Sibling 1 / 509 Shared A: 8 Log(LR) 2

No	Sample	Relation	SxL	LR	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433
1	BF7221	남동생	11	3.8	11 15	29 30	8 12	10 11	16 17	7 9	9 10	9 10	22	15 21 6.2

비교 그룹: 625유가족-AutoSTR, 유가족2006-유연2, 유가족2007-연세대, 유가족2007-유연, 유해2007-연세대



Distribution of the LR



Buckleton et al., Forensic DNA Evidence Interpretation



Genotyping of Highly Degraded DNA

- Efficient DNA extraction
- Mini-STR
 - Identifiler + Minifiler (ABI)
 - Yfiler + Y-miniplex (Park et al, 2007)
- ❖ LR calculation should be modified in light of allele drop-out

The Likelihood Ratio (LR) for the Proposition that the Individuals (P and Q) are Related (H_p) versus Unrelated (H_o) where a Particular Allele in the Profile P has Dropped Out (aN)

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

D : the probability that a particular allele in profile P has dropped out
 P_0, P_1, P_2 : the probabilities that P and Q share 0, 1 or 2 pair of autosomal STR alleles identical-by-descent
 LR equation from Buckleton *et al.* of which some notations are modified

The screenshot displays the mtDNAmanager web application. The 'Working Sample' section lists various samples with their IDs and expected/estimated haplotypes. The 'Target Group' section lists family and human samples. The 'Matching Result' table shows the following data:

No. of Matched Samples	No. of Target Samples	Match Probability
1	15	0.1765

The 'Matching Result' table also includes a detailed view of the match between sample BF7221 and target sample np 18024-16569, showing their respective haplotypes and allele frequencies.



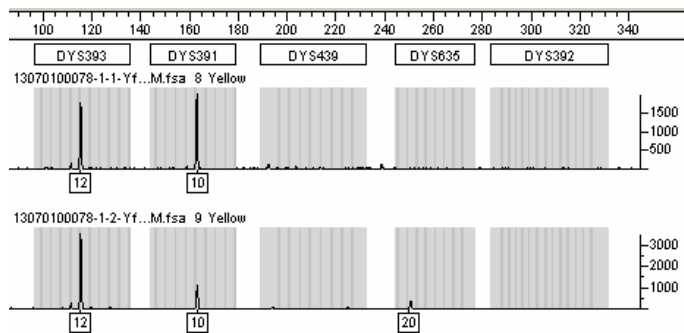
Mitochondrial DNA Control Region Sequence of a Victim and the Alleged Brother

Sample	HV1	HV2	HV3	Haplo group
Victim	16223T-16362C	73G-194T-263G-309.1C-315.1C	489C-523d-524d	D4b2b
Brother	16223T-16362C	73G-194T-263G-309.1C-315.1C	489C-523d-524d	D4b2b

Match Probability = 1/144

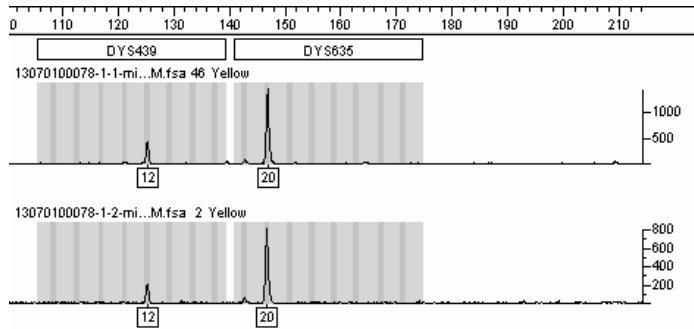


Genotyping of Y-STRs using Yfiler kit from a Victim





Genotyping of Y-STRs using Y-miniplex from a Victim



Y-STR Genotypes of a Victim and the Alleged Brother

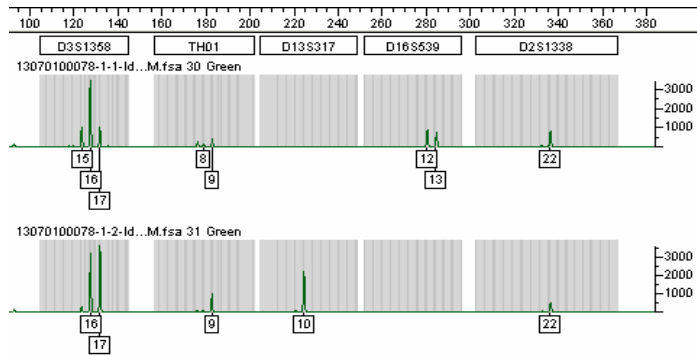


STR locus	Allele		Shared Allele
	Victim	Brother	
DYS19	-	17	-
DYS389I	12	12	12
DYS389II	-	28	-
DYS390	25	25	25
DYS391	10	10	10
DYS392	14	14	14
DYS393	12	12	12
DYS385a/b	12	12-20	12
DYS437	15	15	15
DYS438	11	11	11
DYS439	12	12	12
DYS448	-	20	-
DYS456	15	15	15
DYS458	15	15	15
DYS635	20	20	20
GATA H4.1	21	21	21

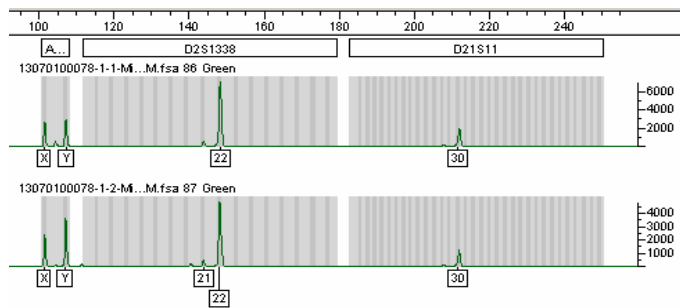
Match Probability = 1/133



Genotyping of Autosomal STRs using Identifiler kit from a Victim



Genotyping of Autosomal STRs using Minifiler kit from a Victim



An Example of Likelihood Ratio (LR) Calculation for Full Sibling 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
<i>D21S11</i>	<i>30</i>	29-30	<i>0.924</i>	<i>1.084</i>	<i>1.135</i>
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
<i>TH01</i>	<i>9</i>	7-9	<i>0.733</i>	<i>0.873</i>	<i>0.931</i>
D13S317	9-10	9-10	8.854	8.854	8.854
D16S539	12-13	9-10	0.250	0.250	0.250
<i>D2S1338</i>	<i>22</i>	22	<i>336.985</i>	<i>31.200</i>	<i>22.756</i>
D19S433	13.2-14	15.2-16.2	0.250	0.250	0.250
<i>vWA</i>	<i>19</i>	15-19	<i>3.124</i>	<i>3.345</i>	<i>3.363</i>
TPOX	-	8-11	-	-	-
D18S51	11-15	13-15	0.922	0.922	0.922
<i>D5S818</i>	<i>11</i>	9-10	<i>0.250</i>	<i>0.420</i>	<i>0.466</i>
<i>FGA</i>	<i>23</i>	23-25	<i>1.435</i>	<i>1.625</i>	<i>1.661</i>
Cumulative LR			2,996.3	791.3	734.3

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

Probability for Kinship Determination

N : No. of victims

M : No. of genotyped skeletal remains

n : No. of genotyped families

$$P(X = k) = \frac{\binom{M}{k} \times \binom{N-M}{n-k}}{\binom{N}{n}}$$

(단, $n-k \leq N-M, k \leq n, k \leq M, n \leq N$)



Estimation of Cold Hit

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

❖ For example, identification of Korean War victims

$$200 \times \frac{1,000}{100,000} = 2$$



Summary

- A simple and general formula for likelihood ratio calculation
- Evaluation of kinship index in light of allele drop-out
- Estimation of cold hit in DNA-based identification on open forms of mass disaster



Consideration for Use of LR

- LR threshold for determination of kinship
- Rate of allele drop-out
 - Degree of DNA degradation
 - STR locus used in PCR system
- Strategy to increase LR
 - Use of mtDNA, Y-STR haplotypes
 - Combined calculation of two more relatives' genotypes



Thank you for your attention!

kishin@yuhs.ac

<http://forensic.yonsei.ac.kr>