

Lineage Marker를 이용한 개인식별에서의 통계적 방법

이효정¹ · 이송덕² · 이승환³
박수정³ · 정수진⁴ · 이재원⁴

¹동아ST 개발본부

²서울대학교 의과대학 법의학교실

³대검찰청 DNA수사담당관실

⁴고려대학교 통계학과

접 수 : 2014년 5월 1일

수 정 : 2014년 5월 13일

게재승인 : 2014년 5월 13일

본 연구는 2012년도 DNA 감식의 국산화 및 선진화 연구사업(1333-304-260)의 지원을 받아 수행되었습니다.

책임저자 : 이재원

(136-701) 서울특별시 성북구 안암로

145, 고려대학교 통계학과

전화 : +82-2-3290-2237

FAX : +82-2-924-9895

E-mail : jael@korea.ac.kr

Statistical Evaluation of Lineage Markers in Individual Identification

Hyo Jung Lee¹, Soong Deok Lee², Seung Hwan Lee³, Su Jeong Park³,
Su Jin Jeong⁴, Jae Won Lee⁴

¹Clinical Development Department, Dong-A ST, Seoul, Korea

²Department of Forensic Medicine, Seoul National University College of Medicine, Seoul, Korea

³Supreme Prosecutor's Office, Seoul, Korea

⁴Department of Statistics, Korea University, Seoul, Korea

Mitochondrial DNA (mt DNA) and the non-recombining region of the Y chromosome are passed down, unaltered, from generation to generation, matrilineally and patrilineally, respectively. Therefore, the Y-chromosome DNA and mtDNA are known as lineage markers, and they play important roles in studies based on human migration and evolutionary history. Y-chromosome DNA is used in forensic analysis to identify individuals involved in cases of sexual assault. In this paper, we review the methods of statistical evaluation of lineage markers used in forensic identification. We also review the combined approach of autosomal and lineage marker evidence.

Key Words : Lineage marker, Mitochondrial DNA, Y chromosome, Individual identification, Match probability

서론

인간 유전자의 특정 염기서열 반복(microsatellite repeat)은 한 유전자좌(locus)에 대한 개인별 대립유전자(allele)의 다형성을 확인할 수 있기 때문에 개인식별이나 개체 간 혈연관계를 밝히는 데 사용되고 있다.^{1,2)} 이 중에서도 다중중합효소 연쇄반응(multiplex polymerase chain reaction)을 이용한 상염색체의 짧은 연쇄 반복(STR, short tandem repeats)은 법의학에서 널리 사용되고 있으며, 가계 추적이나 인구집단의 이동, 유전적 관련성 연구, 계통발생학적 연구의 필요성에 따라 부계 유전되는 Y 염색체와 모계 유전되는 미토콘드리아 DNA(Mitochondrial DNA, 이하 mtDNA)의 분석이 다양한 분야에

서 사용되고 있다.^{3,4)}

22쌍의 상염색체는 감수 분열시 재조합이 일어나지만, 비 재조합 영역의 Y 염색체(non-recombining region Y: NRY)와 mtDNA는 재조합이 일어나지 않기 때문에, 돌연변이가 일어나지 않는다면 세대와 무관하게 동일한 형태로 유전된다. 따라서 부계로 유전되는 Y 염색체와 모계로 유전되는 mtDNA는 'lineage marker'로 불린다.

mtDNA는 정자와 난자의 수정과정에서 정자의 mtDNA가 난자로 들어가지 못해 자연적인 돌연변이를 제외하고는 모계에서 미토콘드리아 일배체형(mitochondrial haplotype)을 공유하게 되며, 세포 당 다수가 존재하기 때문에 가계 추적뿐만 아니라 소량이나 부패된 샘플, 모발, 유해 등의 신원확인에도 사용될 수 있다. 80여 개의 유전자로 구성된 Y 염색체는 인류의

진화 연구나 민족 간 이동 경로 연구 등에 사용될 뿐 아니라, 성범죄와 같은 남녀조식의 혼합시료에서 가해자인 남성을 밝히는 데 매우 유용하며, 실종자 찾거나 대량재해 발생 시 희생자 신원확인, 추정부 확인 등 부계 가족 확인에 사용되고 있다. 또한, 남성에게만 존재하는 유전자를 이용하기 때문에, 소량의 혼합 시료에서도 식별할 수 있어 법의학에서 매우 중요한 역할을 하고 있다.

이처럼 'lineage marker'가 개인식별에서 유용한 도구로 사용되고 있으나, 'lineage marker'를 이용한 개인식별에서 적용될 수 있는 통계적 방법론에 대해 잘 알려지지 않은 것이 국내 실정이라 할 수 있다. 따라서 본 연구의 목적은 Y 염색체나 mtDNA의 유전적 변이를 이용한 개인식별에서 사용될 수 있는 통계적 방법을 조사하고 통계적 해석에 대한 지침을 제공하여, 'lineage marker'를 이용한 유전자 증거의 평가 과정에서 기초자료로 활용할 수 있도록 하는 것이다.

통계적 방법 및 평가

친자확인소송, 형사소송과 같은 법정 소송에서 유전자 감식 결과가 개인식별을 위한 과학적 증거로 제시될 때, 유전자 증거를 평가하기 위한 대표적인 통계적 기법 중 한 가지가 우도비(likelihood ratio, LR)에 의한 평가이다.

형사사건에서 유전자 감식 결과를 이용하는 경우 범인(criminal)이 남긴 것으로 추정되는 표본과 용의자(suspect)의 유전자형을 비교하는 개인식별 검사를 실시하게 된다. 이때 범인이 남긴 것으로 추정되는 표본과 용의자의 표본이 일치하는 경우 용의자가 범인일 가능성을 배제할 수 없게 되며, 서로 다른 두 사람이 검사한 항목에서 우연히 같은 유전자형을 가질 가능성이 어느 정도인가를 평가하게 된다.

Y 염색체와 mtDNA를 이용하는 경우에도 상염색체의 STR을 이용하는 경우와 유전자 증거에 대한 평가의 원리는 동일하게 된다. 즉, 'lineage marker'를 이용한 개인식별에서도 용의자를 범인에서 배제하는 방법은 상염색체의 STR 방법과 동일하며, 서로 다른 두 사람이 우연히 동일한 일배체형을 가질 가능성을 평가하게 되는 것이다. 단, 'lineage marker'의 경우 재조합이 일어나지 않는 특성으로 인해 유전자 간 독립성을 가정할 수 없으며, 이로 인해 상염색체의 유전자 프로파일 빈도의 추정을 위해 사용되는 곱의 법칙('product rule')을 적용할 수 없게 되는 것이다. 따라서 'lineage marker'의 경우에는 검사된 유전자에서 일배체형이 구성되고, 각 일배체형은 한 개의 대립유전자(allele)처럼 처리되어 유전자 검사체계에서 구성된 모든 가능한 일배체형은 한 개의 유전자에서 나타날 수 있는 모든 가능한 대립유전자처럼 간주되는 것이다. 따라서 상염색체의 유전자 프로파일 빈도 추정에 적용되는 곱의 법칙('product rule')은 적용할 수 없게 되므로, 일배체형이 데이터

베이스에서 관측된 수('count')를 기준이 되는 데이터베이스의 크기로 나눈 값을 이용하여 Y STR 일배체형에 대한 유전적 평가를 실시하게 된다. 이러한 방법을 집계 방법('counting method')이라고 한다.

이처럼 'lineage marker'를 이용한 유전자 검사결과를 법정에서 증거로 이용하고자 하는 경우 유전자 검사 대상자들의 유전적 특성을 가장 잘 반영하는 기준 집단(reference population)에 대한 조사가 선행되어 있어야 하며, 집단에서 나타날 수 있는 일배체형 빈도(haplotype frequency)나 유전자좌에 대한 식별력 등이 검토되어 있어야 한다. 따라서 본 논문의 통계적 방법 및 평가에서는 유전자 검사의 기초가 되는 집단 유전학적 조사를 위한 통계적 방법을 소개하고, 이를 개인식별에 적용하기 위한 방법 및 해석방법에 대해 소개하였다.

1. 집단 유전학적 조사

(1) 일배체형 빈도 추정(haplotype frequency estimation)

유전자 검사 대상자들의 유전적 특성을 가장 잘 반영하는 기준 집단(reference population)에서의 일배체형 빈도(haplotype frequency, 법의학에서 일배체형 빈도는 일배체형이 관측되는 확률을 의미한다)를 추정하기 위해서는 이를 대표할 수 있는 표본(sample)을 추출하여 표본으로부터 계산된 일배체형 빈도를 이용하게 된다. 일배체형 빈도를 추정하는 방법에는 한 개의 값으로 추정하는 점 추정(point estimation) 방법과 구간의 형태로 추정하는 구간 추정(interval estimation) 방법을 사용할 수 있다.

기준 집단으로부터 추출된 n 명의 표본에서 일배체형 A가 x 번 발생한 경우, 일배체형 A 빈도의 점 추정치는 $\hat{p} = \frac{x}{n}$ 가 되며("counting method"), 점 추정치의 표준편차(standard deviation)인 표준오차(standard error)의 추정치는

$$se(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

로 표현될 수 있고, 이는 좋은 추정량을 선택하는 기준의 하나로 사용될 수 있다. 구간 추정은 일배체형 빈도가 속할 것으로 기대되는 범위를 추정하는 것으로, 점 추정치가 기준 집단에서의 일배체형 빈도에 얼마나 가까운 값인가에 대한 정보를 포함하게 된다. 구간 추정 방법은 구간이 기준 집단에서의 일배체형 빈도를 포함할 확률을 일정한 수준(신뢰수준, $100(1-\alpha)\%$)으로 정해 놓고, 추정치의 분포를 이용하여 구간을 계산하게 되며, 다음과 같이 계산된 구간을 $100(1-\alpha)\%$ 신뢰구간(confidence interval)이라고 한다.

$$\left[\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \quad \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

단, $z_{\alpha/2}$ 는 표준정규분포(standard normal distribution)의 상위 $\alpha/2$ 에 해당하는 임계치이며(95% 신뢰구간 계산 시 임계값은 1.96임), 이의 신뢰구간은 표본의 정규분포를 가정하여 계산된 결과이므로 표본 크기가 클 때 사용할 수 있는 방법이다. 일반적으로 np 와 $n(1-p)$ 가 모두 5보다 큰 값을 가질 때 대표본 근사에 의한 신뢰구간이 정확하다고 알려져 있다. 만일 일배체형 빈도가 0.05인 경우에는 표본의 크기가 100보다 큰 경우에 이 조건이 만족이 되지만, 일배체형 빈도가 0.01인 경우에는 표본의 크기가 500보다 커야 이 조건이 만족되는 것이다. 즉, 일배체형 빈도가 0이나 1에 가까운 값을 가지게 되면 정규분포로 근사하기 위해서는 많은 수의 표본이 필요하게 되는 것이다. 그러나 mtDNA나 Y 염색체(non-recombining region Y: NRY)의 경우 상염색체와는 달리 일배체형 빈도가 낮은 경우가 많이 발생하게 되며, 이의 추정을 위해서는 많은 수의 표본을 필요로 하게 된다. 만일 많은 수의 표본을 얻지 못하여 정규분포의 근사에 의한 추정이 적합하지 않은 경우에는 이항분포(binomial distribution)를 사용하여 정확한 신뢰구간을 추정할 수 있다.⁵⁾ 이항분포를 이용한 $100(1-\alpha)\%$ 신뢰구간(confidence interval) (p_L, p_U)는 다음 식에 의해 각각 계산될 수 있다.

$$\sum_{k=x}^n \binom{n}{k} p^k (1-p)^{n-k} = \alpha/2, \quad \sum_{k=0}^x \binom{n}{k} p^k (1-p)^{n-k} = \alpha/2$$

여기서, x 는 확률변수 X 의 관측값이 되며, X 는 n 개의 표본에서 일배체형 A 가 발생한 횟수로 이항분포(n, p)를 따르게 된다. 만일 표본에서 일배체형 A 가 한 번도 관측되지 않은 경우에는 $100(1-\alpha)\%$ 단측 신뢰구간(one-sided confidence interval)이 계산될 수 있으며, 일배체형 빈도 p 의 신뢰구간은 $[0, 1-\alpha^{1/n}]$ 이고, $1-\alpha^{1/n}$ 은 n 이 100보다 큰 값인 경우 $3/n$ 에 근

사하게 된다(의학에서 “rule of three”로 알려져 있다).

일배체형 빈도 추정에 대한 보수적인 접근을 위해 일배체형 빈도를 포함할 확률에 대한 일정한 수준(신뢰수준) 하에 신뢰구간을 계산한 후 신뢰구간의 상한값을 제시해야 하며, 대표본 근사에 의한 신뢰구간 또는 이항분포를 사용한 정확한 신뢰구간 계산방법을 이용하여 추정할 것을 권장한다.

Table 1은 17-locus Yfiler 프로파일의 한 예제를 이용하여 Y-STR loci 수에 따라 각각 YHRD와 US Y-STR 데이터베이스를 사용하여 검색된 결과를 나타내며, 검색된 일배체형 빈도에 대한 신뢰구간의 상한치(upper bound confidence interval)가 계산된 결과이다.⁶⁾ 데이터베이스 크기는 Y-STR 유전자좌의 수에 따라 변화되며, 즉 일배체형 빈도의 추정에 사용될 분모의 값이 변화되므로 추정 결과에 영향을 받게 되는 것이다. Table 1에서 17개 검사체계를 사용하는 경우에 있어서 일배체형 빈도에 대한 신뢰구간의 상한치는 0.029%로 3,420명마다 1명씩 발생한다고 볼 수 있는 반면, 7개 검사체계를 사용하는 경우에는 3.357%로 30명마다 1명씩 발생하여 일배체형이 나타날 가능성이 약 100배 이상 높게 나타나는 것을 볼 수 있다.

집단 내 아집단(subpopulation)을 가정한 경우에 대한 일배체형 A 빈도의 점 추정치 및 분산은 다음과 같이 계산될 수 있으며, θ 가 0인 경우에는 아집단을 가정하지 않은 추정결과와 동일하게 된다.

$$\hat{p} = \frac{1}{n} \sum_v n_v \hat{p}_v, \quad \text{Var}(\hat{p}) = \frac{1}{n^2} \sum_v n_v^2 \text{Var}(\hat{p}_v) = p_A(1-p_A) \left[\frac{\sum_v n_v^2}{n^2} \theta + \frac{1-\theta}{n} \right]$$

여기서, n_v 와 \hat{p}_v 는 아집단 v 에 대한 집단 크기 및 일배체형 빈도 추정치이며, θ 는 아집단 내 대립유전자(allele)들의 유사성 정도를 아집단 간 유사성 정도에 대한 상대적인 크기로 표현한 공통조상계수(coancestry coefficient)이다. \hat{p} 는 아집단

Table 1. Example for 17-locus Yfiler Profile and Calculation of Y-STR Haplotype Frequency Estimates

17-locus Yfiler Profile											
DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385a/b	DYS393	DYS391			
17	13	24	29	18	14	11, 15	13	7			
DYS439	DYS635	DYS392	GATA-H4	DYS437	DYS438	DYS448					
13	23	13	12	15	12	19					
YHRD	<i>N</i>	<i>x</i>	\hat{p} (%)	<i>UCL</i> [†] (%)	~1 in Every	US Y-STR	<i>N</i>	<i>x</i>	\hat{p} (%)	<i>UCL</i> [†] (%)	~1 in Every
17-locus	10243	0	0	0.029	3420	17-locus	4163	0	0	0.072	1390
12-locus	13751	3	0.022	0.047	2150	12-locus	10865	7	0.064	0.112	892
11-locus	36174	18	0.050	0.073	1375	11-locus	13906	19	0.137	0.198	505
9-locus	63369	307	0.484	0.539	186	9-locus	13906	128	0.920	1.079	93
7-locus	65165	2099	3.221	3.357	30	7-locus	13906	846	6.084	6.481	15

*17-locus Yfiler, 12-locus Powerplex Y, 11-locus SWGDAM recommended, 9-locus european minimal haplotype, and 7-locus minimal haplotype minus DYS385a/b

[†]Upper bound confidence interval

에 대한 크기에 의해 가중화된 평균이 되며, $Var(\hat{p})$ 는 아집단 간 변이와 아집단 내 변이의 합으로 구성된다.

(2) 유전자 다양성 (gene diversity or power of discrimination)

유전자좌에 대한 식별력은 유전자 다양성 (gene diversity)으로 평가할 수 있으며, 이는 다음과 같이 계산될 수 있다. 단, 유전자 다양성은 개체식별력 (power of discrimination)으로 정의되기도 한다.

$$PD = 1 - \sum p_i^2$$

여기서, p_i 는 유전자좌에서 발생하는 각 일배체형에 대한 빈도를 나타내며, PD 는 유전자좌에서 나타난 어떠한 일배체형에 대해서도 서로 다른 두 사람이 동일한 일배체형을 나타내지 않을 확률을 나타낸다. 즉, 이 확률값이 1에 가까울수록 개인식별 결과의 신뢰성을 높일 수 있는 것이다.

유전자 다양성의 추정량에 대한 편의를 보정 (bias correction)한 개체식별력은 다음과 같이 계산된다.

$$PD_{Adj.} = \frac{n}{n-1} \times \{1 - \sum p_i^2\} = \frac{n}{n-1} \times PD$$

단, 모든 일배체형이 각각 한 번씩만 관측되는 경우 보정된 개체식별력은 1의 값을 가지게 되므로 PD 를 사용하여 식별력을 평가하게 된다.

2. 개인식별검사

(1) 짝확률 (match probability) 및 우도비 (likelihood ratio) 계산

개인식별 검사에서 유전자 증거에 대한 해석은 짝확률 (match probability) 또는 우도비 (likelihood ratio, LR)에 기초하게 되며, 이는 정황증거에 의해 지목된 용의자 (suspect)의 유전자형과 범인 (criminal)의 것으로 추정되는 흔적의 유전자형 검사결과를 비교하여 일치하는 경우에 계산하게 된다. 짝확률은 범죄현장에서 발견된 유전자형에 대해 용의자가 범인이 아님에도 불구하고 우연히 일치했을 확률로, mtDNA와 Y 염색체를 이용하여 짝확률을 계산하는 경우에는 일배체형 빈도를 사용하게 된다. 일배체형 A에 대한 짝확률은 다음과 같이 계산할 수 있으며, 이는 일배체형 A의 프로파일 빈도와 일치하게 된다.⁷⁾

$$GMP = P(A|A) = \frac{P(A, A)}{P(A)} = \frac{p_A^2}{p_A} = p_A$$

우도비 (likelihood ratio, LR)는 짝확률의 역수인 $1/GMP$ 로 정

의되며, 이를 개인식별지수 (identity index)라고 한다. 이 값은 “용의자가 범인이어서 일배체형이 일치했을 확률은, 용의자가 범인이 아님에도 불구하고 우연히 일치했을 확률과 비교할 때 LR배이다”라고 해석할 수 있다. 또한, 용의자와 범인이 같은 아집단에 속해 있을 가능성을 고려하는 경우에 대한 짝확률은

$$GMP = P(A|A) = \theta + (1 - \theta)p_A$$

이며, 이를 정확하게 평가하기 위해서는 공통조상계수 θ 에 대한 정확한 추정이 중요하다고 하겠다.

이처럼 개인식별검사에서 유전자 증거에 대한 해석은 짝확률 또는 개인식별지수에 근거하게 되지만, 이에 대한 해석 시에는 주의를 필요로 한다. 유전자 증거는 다른 정황증거들과 함께 재판부의 판결을 돕는 판단근거의 한 부분이 되는 것으로 유전자 증거 자체가 용의자가 범인일 확률 또는 용의자가 범인이 아닐 확률을 나타내는 것은 아니기 때문이다.

(2) 상염색체와 Y 염색체의 통합적 접근법

대량 재해나 실종자 확인 등과 같은 개인식별에서 상염색체 유전자와 Y 염색체 유전자가 동시에 조사되는 경우가 발생된다. 이때의 관심은 식별력을 높이기 위해 상염색체와 Y 염색체에서 결과한 유전자 정보를 통합해서 평가할 수 있는 가에 있다. 하지만 이러한 통합적 접근법에 대해 현재 많은 연구가 진행되지 못했으며, 이로 인해 통합적 접근법을 사용하는 경우는 드물다고 할 수 있다.

통합적 접근법에 대해 최근 진행된 연구 결과들을 살펴보면, Sinha⁸⁾은 상염색체 STR의 대립유전자와 Y STR 대립유전자 간의 연관성을 조사하여, 유전자 간 연관성이 대체로 존재하지 않는 것을 확인하였다. 이를 근거로 상염색체 STR의 프로파일 빈도와 Y-STR의 일배체형 빈도의 곱으로 평가되는 통합적 접근법은 사용될 수 있으며, 이러한 방법은 보수적이라는 판단을 내렸다. Walsh⁹⁾의 연구에서도 이러한 통합적 접근법을 지지하는 내용의 논문을 발간하였다. Sinha⁸⁾의 방법과 동일하게 상염색체 STR과 Y STR 간의 연관성을 조사하였으나, 대립유전자 간 연관성이 아닌 상염색체 STR의 유전자형과 Y STR의 일배체형 간의 연관성을 조사하였다. 연관성 검정에 대한 접근 방법은 Sinha⁸⁾와 달랐지만, 연관성 결과는 동일하게 낮은 결과를 보였다. 또한, Walsh⁹⁾는 NRC II에서 권장되는 공통조상계수 θ 의 0.01~0.03에 대해 Y STR 결과가 서로 일치한다는 조건하에서의 θ 값을 재계산하였다. 그 결과 θ 값이 0.01에 대해서는 0.013~0.022, 0.03에 대해서는 0.032~0.040의 값을 보였으며, Y STR이 일치하는 경우에도 θ 는 약간 증가한다는 것을 확인하였다. 이러한 결과들을 근거로 상염색 STR에서 계산된 짝확률과 Y STR에서 계산된 짝확률의 곱에 의해 계산된 결합 확률은 보수적인 추정값이 될 수 있다고 하였으며, Y STR

이 일치하는 경우에 대해 상염색체 STR의 정확률 계산 시 θ 값은 가장 보수적인 값인 0.04를 사용할 것을 권장하였다.

이처럼 Sinha⁸⁾와 Walsh⁹⁾는 통합적 접근법에 대해 지지하는 내용의 논문을 발간한 반면, Amorim¹⁰⁾은 두 염색체의 통합적 접근법은 사용하지 않아야 한다는 결론을 내렸는데, 이에 대한 사유로는 lineage marker가 일치하는 경우 같은 혈통의 사람의 표본이 될 수 있기 때문에 상염색체 STR에서 사용되는 통계적 가설과는 다르게 적용되어야 한다는 것이다. 또한, 상염색체와 Y 염색체는 통계적 접근 방법이 이질적이며, 통합적 접근은 편의된 결과를 얻기 쉽다는 점을 이유로 제시하였다. 통합적 접근법은 반대하였지만, 두 염색체에 대한 유전자 증거의 분석 결과를 각각 보고하는 방식을 제안하였으며, 혈연관계가 없다는 가정을 명백하게 받아들일 수 있는 경우에 한해서 유전자 증거의 결합은 가능할 것이라는 의견을 제시하였다.

이와 같이 통합적 접근법의 사용에 대해 연구자마다 다른 주장들이 제기되고 있으므로, 상염색체 유전자와 Y 염색체 유전자를 통합하여 한 개의 측도로 유전자 증거를 평가하기 위해서는 통합적 접근 방법론에 대한 추가적인 연구가 필요할 것으로 판단된다.

(3) 유전적 증거의 귀인(source attribution)

개인식별 검사에서 유전자 증거에 대한 해석은 유전자 증거를 통해 알 수 있는 유일한 수치인 정확률 또는 우도비(likelihood ratio, LR)에 기초하게 되지만, 이러한 값이 용의자가 범인일 확률 또는 용의자가 범인이 아닐 확률을 나타내는 것은 아니다. 따라서 사건 현장에서 발견된 유전자 증거의 흔적과 일치하는 유전자를 가진 용의자가 발견되었을 때, 이 용의자가 그 흔적을 남긴 사람이라고 판단할 수 있는 기준이 필요하게 된다. 어떤 기준에 의해 그 흔적을 남긴 사람이 용의자의 것으로 추정된다고 선언되는 데 이러한 선언을 ‘유전적 증거의 귀인(source attribution, 이하 SA)’이라고 한다.

미연방수사국(Federal bureau of Investigation, FBI)의 경우 상염색체에 대한 특정 유전자 프로파일의 나타날 가능성이 미국 인구 크기의 1000배가 되는 집단에서 나타날 가능성보다 낮은 경우를 SA의 판단 기준으로 채택하기도 하였다. 또한, 미국의 많은 실험실에서는 유전자 프로파일의 빈도가 사전에 정의된 정확률의 한계치(match probability threshold)를 초과하는 경우에 SA에 대한 선언문을 사용한다.¹¹⁾ 정확률의 한계치는 n 개의 표본에서 특정 유전자 프로파일이 $100(1-a)\%$ 의 신뢰도로 한 번도 발생하지 않을 확률에 근거하게 된다. n 개의 표본에서 유전자 프로파일이 한 번도 발생하지 않을 확률은 $(1-p)^n$ 이 되고, 이 값이 $1-a$ 이상인 경우, 즉 $p \leq 1-(1-a)^{1/n}$ 인 경우에 유전자 증거가 되는 프로파일은 집단에서 유일하다고 $100(1-a)\%$ 의 신뢰도로 평가하게 되는 것이다. Table 2는 다양한 표본(집단) 크기에서 다양한 신뢰도에 따른 SA를 위한 정확률의 한계치(match probability threshold)를 나타낸다.⁶⁾

대한민국을 예로 들면, 인구 크기의 1,000배 기준을 적용하는 경우 대한민국의 인구수는 약 5천만 명이므로 유전자 프로파일의 확률이 5백억 분의 1 (5.00×10^{-10})보다 작은 경우 이러한 유전자 프로파일은 대한민국에서 유일한 것으로 결론 내릴 수 있을 것이다. 또한, 99% 신뢰도에 의한 정확률 한계치의 평가기준을 적용하는 경우에는 2백억 분의 1 (2.01×10^{-10})보다 작은 경우에 유일성을 판단 내릴 수 있을 것이다.

다음에 기술된 내용은 유전자 증거에 대한 평가 보고서에 작성될 수 있는 SA의 해석에 대한 한 예로, 99% 신뢰도에 의한 정확률 한계치의 평가기준을 적용한 경우이다.⁶⁾

“일란성 쌍생아(identical twins) 또는 가까운 친척 관계(close relatives)의 가능성을 배제하는 경우 $100(1-a)\%$ 의 신뢰도(reasonable scientific certainty)로 Q (questioned source, Q sample)와 K (known reference sample, K sample)의 유전자는 동일한 사람에서 기인한 것으로 결론 내릴 수 있다.” 또는

Table 2. Match Probability Threshold for Source Attribution at Various Population Size and Confidence Levels

	Sample Size (n)	Confidence levels			
		90%	95%	99%	99.9%
	10	1.05×10^{-2}	5.12×10^{-3}	1.00×10^{-3}	1.00×10^{-4}
	50	2.10×10^{-3}	1.03×10^{-3}	2.01×10^{-4}	2.00×10^{-5}
	100	1.05×10^{-3}	5.13×10^{-4}	1.00×10^{-4}	1.00×10^{-5}
	1,000	1.05×10^{-4}	5.13×10^{-5}	1.01×10^{-5}	1.00×10^{-6}
	100,000	1.05×10^{-6}	5.13×10^{-7}	1.01×10^{-7}	1.00×10^{-8}
	1,000,000	1.05×10^{-7}	5.13×10^{-8}	1.01×10^{-8}	1.00×10^{-9}
	10,000,000	1.05×10^{-8}	5.13×10^{-9}	1.01×10^{-9}	1.00×10^{-10}
Korea (2011)	50,000,000	2.11×10^{-9}	1.03×10^{-9}	2.01×10^{-10}	2.00×10^{-11}
U.S. (1999)	260,000,000	4.05×10^{-10}	1.97×10^{-10}	3.87×10^{-11}	3.85×10^{-12}
U.S. (2005)	300,000,000	3.51×10^{-10}	1.71×10^{-10}	3.35×10^{-11}	3.33×10^{-12}
	1,000,000,000	1.05×10^{-10}	5.13×10^{-11}	1.01×10^{-11}	1.00×10^{-12}
World pop.	6,000,000,000	1.76×10^{-11}	8.55×10^{-12}	1.68×10^{-12}	1.67×10^{-13}

“유전적으로 무관한 사람들의 표본에서 이러한 유전자 증거의 프로파일은 $100(1-\alpha)\%$ 의 신뢰도(reasonable scientific certainty)로 발견될 수 없다.”

이처럼 용의자가 범죄현장의 흔적을 남긴 사람이라고 판단할 수 있는 기준은 상염색체 유전자를 이용한 개인식별에는 적용 가능하지만, 일배체형 빈도로 계산되는 mtDNA와 Y 염색체에서는 사용이 어렵게 된다. 따라서 혼합 표본 등에서 상염색체 유전자에 대한 증거력이 감소하여 SA를 평가하기 어려운 경우에서 mtDNA와 Y 염색체에 대한 유전적 평가 결과가 이용될 수 있을 것으로 판단된다. 단, 상염색체와 lineage marker에 대한 유전자 검사 결과를 함께 평가하기 위해서는 유전자 증거력에 대한 많은 경험을 통한 신중한 판단이 필요할 것이다.

3. 공통조상계수의 추정

전체 집단(population)은 여러 개의 아집단(subpopulation)으로 구성될 수 있다. 예를 들어 미국의 경우 인구집단을 African American, Caucasian, SE Hispanic, SW Hispanic 등으로 나누며, 영국의 경우 Caucasian, Afro-Caribbean, Asian 등으로 구분한다. 이처럼 전체 집단에 이질적인 아집단(heterogeneous population)이 있으며, 아집단에서의 일배체형 빈도가 전체 집단과 다른 경우 개인식별에 잘못된 판단을 내리게 되는 것이다. 따라서 유전자 증거를 개인식별에 이용하기 위해서는 전체집단과 아집단의 관계에 대한 연구가 필요하게 되는 것이다.

전체 집단에서 아집단으로의 세분을 집단유전학 측면에서 정량화시킨 척도는 집단에서의 동계교배계수(inbreeding coefficient)이며, 동계교배계수는 집단 내에서의 동계교배계수(within-population inbreeding coefficient, f), 전체 동계교배계수(total inbreeding coefficient, F), 공통조상계수(coancestry coefficient, θ)로 구분될 수 있다. 본 논문에서는 개인식별에서 중요한 아집단 내 유사성 척도인 공통조상계수를 추정하는 방법들에 대해 살펴보도록 한다.

(1) Weir-Cockerham 방법

공통조상계수 θ 를 추정하는 가장 대표적인 방법은 Weir and Cockerham¹²⁾에 의해 제안된 방법으로, 집단들에서 나타나는 대립유전자(일배체형) 빈도 분포의 평균과 분산을 이용하여 θ 를 추정하는 방법이다. r 개의 각 집단에서의 대립유전자 빈도의 추정치를 이용하여 집단 내에서의 대립유전자 빈도의 변이를 나타내는 집단 내 평방향 MSW와 집단 간의 대립유전자 빈도의 변이를 나타내는 집단 간 평방향 MSA를 다음과 같이 계산한다.

$$MSA = \frac{2n}{r-1} \sum_{s=1}^r (\hat{p}_s - \bar{p})^2, \quad MSW = \frac{2n}{r(2n-1)} \sum_{s=1}^r \hat{p}_s(1-\hat{p}_s)$$

여기서, r 은 아집단의 개수, n 은 전체 표본의 크기를 나타내며, $\bar{p} = \frac{1}{r} \sum_{s=1}^r \hat{p}_s$ 는 아집단 각각에서의 대립유전자(일배체형) 빈도에 대한 평균을 나타낸다.

공통조상계수 θ 는 집단 간 평방향 MSA 및 집단 내 평방향 MSW를 이용하여 다음 공식에 의해 추정할 수 있다.

$$\hat{\theta} = \frac{\sum_K (MSA - MSW)}{\sum_K (MSA + (2n-1)MSW)}$$

여기서 K 는 대립유전자(일배체형)의 개수를 의미하며, 전체 집단에서 모든 일배체형이 유일하게 한 번씩만 발생한 경우에는 Weir-Cockerham 방법의 의한 공통조상계수 θ 의 추정값은 0의 값을 가지게 된다.

(2) Ewens 방법

Weir and Cockerham¹²⁾에 의해 제안된 방법이 집단들에서 나타나는 대립유전자(또는 일배체형) 빈도 분포의 평균과 분산을 이용하였다면, Ewens¹³⁾의 방법은 집단에서의 서로 다른 대립유전자(또는 일배체형)의 수를 이용하여 공통조상계수 θ 를 추정하였다.

Ewens¹³⁾는 집단에서 서로 다른 대립유전자(또는 일배체형)의 수 k 를 이용하여 집단에서 임의로 추출된 2개의 대립유전자가 동일할 확률인 $1/(1+\psi)$ 을 추정할 수 있음을 보였다. $1/(1+\psi)$ 는 θ 로, 돌연변이율(infinite-alleles mutation rate)이 μ 인 크기가 N 인 집단에서의 ψ 는 $2N\mu$ 로, θ 는 $1/(1+2N\mu)$ 가 된다. 따라서 돌연변이율이 증가할수록 θ 는 감소하게 되며, 일배체형의 돌연변이율은 유전자좌의 수가 많아질수록 증가하게 된다.

ψ (또는 θ)의 최대우도추정량(maximum likelihood estimate)은 k 의 관측된 값이 이의 기댓값(expected value)과 동일하다는 가정 하에서 계산되는 것으로, 수치적 방법을 이용하여 θ 를 추정하게 된다.

$$E(k) = \sum_{i=0}^{n-1} \frac{\psi}{\psi + i} = \sum_{i=1}^{n-1} \frac{1-\theta}{1+(i-1)\theta}$$

이처럼 Ewens 방법은 θ 의 추정에 있어 특정 대립유전자의 빈도를 사용하지 않게 된다. 또한, 만일 전체 집단에서 모든 일배체형이 유일하게 한번 씩만 발생한 경우, 즉 $k=n$ 인 경우에는 (무한하게) 큰 돌연변이율에 따라 θ 의 추정값은 Weir-Cockerham 방법과 마찬가지로 0의 값을 가지게 된다.

(3) 기타

상염색체에 대해 θ 는 0.01~0.05로 정의되는 것이 보통이다.¹⁴⁾ θ 를 추정하기 위해 자료를 이용한 연구들도 많이 진행되었는데, Cavalli-Sforza¹⁵⁾는 다양한 인구집단에 대한 연구에 기초하여 θ 값을 추정하였고, 추정값은 대체로 보다 낮은 결과를 나타내었다.

Balding and Nichols¹⁶⁾는 지리적으로 정의된 아집단에 대한 자료를 사용하였으며, i 번째 아집단의 j 번째 유전자좌에 대한 θ_{ij} 를 $1/(1-\alpha_i+\beta_j)$ 로 모형화한 후, α_i 와 β_j 에 대한 로그정규 사전분포(lognormal prior distribution)와 대립유전자 비율에 대한 디리클레 사전분포(Dirichlet prior distribution)의 가정 하에 θ 를 추정하는 방법을 연구하였다.

Foreman¹⁷⁾의 연구에서도 집단의 자료를 임의의 아집단으로 분할한 자료를 사용하였으며, θ 에 대한 사전분포로 베타분포를 대립유전자 비율에 대해서는 디리클레 사전분포(Dirichlet prior distribution)를 가정 하에 θ 를 추정하였다.

이외에도 공통조상계수 θ 를 추정하기 위한 방법들이 많이 제안되고 있으며, 집단(population)이 여러 개의 아집단(subpopulation)으로 구성된 경우 개인식별에서 유전자 증거의 평가값에 영향을 주게 되므로 공통조상계수 θ 의 정확한 추정을 위한 연구가 중요하다고 하겠다.

4. 유전자 데이터베이스

mtDNA 또는 Y 염색체의 데이터베이스(population database)는 일배체형의 기대빈도를 추정하는 데 중요한 역할을 한다. 전 세계적으로 다양한 집단에서 혈연관계가 없는 사람들의 mtDNA 또는 Y 염색체의 유전자 정보를 수집하기 위한 노력이 계속돼 왔다. 데이터베이스에서 추정된 일배체형 빈도는 개인식별을 위한 유전자 증거의 평가값으로 사용되기 때문에 높은 품질의 정보를 수집하는 것이 중요하다고 하겠다. 본 논문에서는 mtDNA 또는 Y 염색체의 데이터베이스 종류를 소개하고, 데이터베이스 생성 방법 등에 대한 자세한 논의는 생략하기로 한다.

mtDNA의 데이터베이스는 FBI mtDNA Database, EMPOP, mtDNAMANAGER 등이 있다. FBI mtDNA Database는 14개의 다양한 집단에 대한 4,839개의 mtDNA 프로파일에 대한 데이터베이스를 만들었으며, 이는 FBI 홈페이지에서 MitoSearch 분석도구를 이용하여 내려받기(download) 할 수 있으며 특정

프로파일에 대한 검색이 가능하다. EMPOP (European DNA Profiling Group mitochondrial DNA population database project)는 유럽의 mtDNA 데이터베이스 구축을 위한 프로젝트로, 이 프로젝트에 의해 수천 개의 mtDNA 프로파일이 수집되었다. 이는 EMPOP의 웹사이트 <http://www.empop.org>에서 접근 가능하며, 온라인에서 검색이 가능하다. mtDNAMANAGER (online mtDNA searchable population database)는 연세대학교에서 개발된 데이터베이스로, 현재 5개의 인종(African, West Eurasian, East Asian, Oceanian, and Admixed)에 대한 자료를 포함하고 있다. 이는 웹사이트 <http://mtmanager.yonsei.ac.kr>에서 접근 가능하다.

Y 염색체의 데이터베이스는 가장 크고 전 세계적으로 가장 많이 사용되고 있는 YHRD (Y-STR Haplotype Reference Database, <http://www.YHRD.org>)가 있으며, SWGDAM에서 권장하는 11개 유전자좌를 이용하여 5개의 다른 U.S. 그룹에 대한 일배체형 빈도에 대한 추정이 가능하도록 개발된 US Y-STR 데이터베이스(U.S.-population-specific Y-STR Database, <http://www.usystrdatabase.org>) 등이 있다.

유전자 증거 평가 예제

본 연구에서는 상염색체 및 Y 염색체에 대한 분석 예제를 이용하여 통계적 분석과정 및 해석방법에 대해 살펴보기로 한다. 첫 번째 예제에서는 Y 염색체를 이용하여 유전자 증거를 평가하고 해석하는 방법을 설명하였으며, 두 번째 예제에서는 상염색체와 Y 염색체를 이용하여 유전자 증거를 평가하는 방법을 설명하였다. 본 연구의 분석 예제에서 사용된 모든 유전자 자료 및 결과는 분석과정을 설명하기 위한 가상의 자료로 구성한 것이다.

1. Y 염색체를 이용한 유전자 증거 평가 예제 및 해석

Table 3의 유전자 감정 자료는 Yfiler STR (Applied Biosystems)을 이용한 검사 결과에 대한 가상의 예제이다. 강간사건의 피해자에서 정액을 분리하여 범인의 것으로 추정되는 남자에 대한 Y 염색체의 유전자형을 조사한 결과로, 본 사건에서 용의자는 없는 것으로 가정하였다.

유전자 증거에 대한 평가를 하기에 앞서 유전자 증거에 사용된 Y 염색체의 유전자 검사체계에 대한 개체식별력을 평가한

Table 3. Example for 17-locus Yfiler Profile

DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385a/b	DYS393	DYS391
15	14	23	29	17	16	10, 18	13	10
DYS439	DYS635	DYS392	GATA-H4	DYS437	DYS438	DYS448		
12	21	13	11	14	13	18		

결과 99.95%의 식별력을 보였다고 가정하였다. 이 때 개체식별력 99.95%가 의미하는 바는 유전자 증거가 한 사람의 것이고, 집단 및 혈족관계가 배제되는 임의의 두 사람이 동일한 일배체형을 나타내지 않을 확률을 의미하게 되는 것이다.

본 예제의 상황에서 용의자는 없는 것으로 가정하였으므로, 범인의 것으로 추정되는 남자에 대한 Y 염색체의 일배체형에 대해 유전자 데이터베이스를 사용하여 검색하였다고 가정하자. 만일 유전자 데이터베이스에 포함된 표본 수가 전체 2,000명이라고 가정하고, 2,000명의 표본 중 10명에서 동일한 서열을 보였다고 가정했을 때, 일배체형 빈도 및 95% 신뢰구간의 추정값을 계산하면 0.5% [0.2%, 0.8%]로 추정할 수 있다. 추정된 신뢰구간의 상한값인 0.8%를 기준으로 할 때 이러한 일배체형은 124명당 1명씩 발생하는 것으로 해석할 수 있겠다.

2. 상염색체와 Y 염색체를 이용한 유전자 증거 평가 예제 및 해석

Table 4의 유전자 검사 결과는 상염색체 6개 유전자좌 및 Yfiler (Applied Biosystems)를 이용한 Y 염색체 검사 결과에 대한 가상의 예제이다. 범인의 것으로 추정되는 유전자 증거 표본과 용의자의 유전자형이 일치하여 용의자가 범인일 가능성을 배제할 수 없는 것으로 가정하였다.

유전자 증거에 대한 평가를 하기에 앞서 유전자 증거에 사용된 상염색체의 유전자 검사체계에 대한 개체식별력을 평가한 결과 99.9999893%의 식별력을 보였다고 가정하자. 이때 개체식별력 99.9999893%가 의미하는 바는 검사에서 이용된 6개의 유전자좌를 ‘정황증거에 의해 한 명의 용의자가 지목되고, 범죄현장의 증거 역시 한 사람의 것인 경우’에 이용할 때,

유전자형 비교를 통해 범인이 아닌 사람을 99.9999893% 구별할 수 있다는 것을 의미하는 것이다.

유전자 증거에 대한 평가 결과는 다양한 가정 하에 계산되었다. 첫째 용의자와 범인이 친척관계나 집단관계의 가능성이 없다고 가정하였으며, 둘째 용의자와 범인이 동일한 아집단에 속해 있는 것으로 가정하였다. 아집단을 고려한 우도비 계산 시 공통조상계수는 0.03을 사용하였다. 셋째 용의자와 범인이 동일한 아집단에 속해 있으며 Y 염색체 STR 결과가 서로 일치한다고 가정하였다. Y 염색체 STR 결과가 일치하는 조건 하에서 아집단에 대한 고려가 필요한 경우, 공통조상계수는 0.03 대신 0.04를 적용하여 가장 보수적으로 평가하였다. Table 4는 각 가정 하에서 유전자 증거에 대한 평가 결과로, 다음과 같이 해석할 수 있다.

- (a) 이 정액이 용의자와 유전적으로 무관한 다른 사람에게서 나왔을 때 이와 같은 유전자 검사결과를 얻게 될 확률은 7.6269×10^{-7} 이다.
- (b) 이 정액이 용의자와 동일한 아집단에 속하고 혈연관계가 없는 다른 사람에게서 나왔을 때 이와 같은 유전자 검사결과를 얻게 될 확률은 1.9647×10^{-6} 이다.
- (c) 이 정액이 용의자와 동일한 아집단에 속하고 혈연관계가 없는 다른 사람에서 나왔으며 Y-STR 검사 결과가 일치했을 때 이와 같은 상염색체 유전자 검사결과를 얻게 될 확률은 2.5909×10^{-6} 이다.

결론 및 고찰

본 연구에서는 Y 염색체나 mtDNA의 유전적 변이를 이용한

Table 4. Example for Combined Approach of Autosomal and Lineage Marker Evidence

Locus	Crime sample	Suspect	Not excluded	Likelihood ratio		
				[a]*	[b] [†]	[c] [‡]
D8S1179	14-15	14-15	O	18.48	15.19	14.33
D21S11	30-31	30-31	O	13.74	11.26	10.65
D7S820	11-11	11-11	O	8.40	6.51	6.05
D3S1358	15-17	15-17	O	6.51	6.05	5.92
D5S818	10-11	10-11	O	7.74	7.11	6.92
FGA	22-23	22-23	O	12.20	10.63	10.20
Cumulative Identity Index				1.3111×10^6	508994	385970
Cumulative Match Probability				7.6269×10^{-7}	1.9647×10^{-6}	2.5909×10^{-6}
Yfiler	15,14,23,28, 18,16,10-20,	15,14,23,28, 18,16,10-20,	O	MP = 1/2000		
STR	12,10,14,20,	12,10,14,20,				
haplotype	14,12,14,12, 19	14,12,14,12, 19				

*Suspect and offender are not related.

[†]Suspect and offender are the members of the same subpopulation ($\theta = 0.03$).

[‡]Suspect and offender are the members of the same subpopulation and suspect and offender have same Y-STR profile ($\theta = 0.04$).

개인식별에서 사용될 수 있는 통계적 방법을 조사하였다. 유전자 검사 대상이 되는 사람들의 유전적 특성을 잘 반영하는 집단에서 나타날 수 있는 일배체형 빈도의 추정방법과 검사체계의 식별력 평가방법을 정리하였다. 또한 'lineage marker'를 이용한 개인 식별 검사에서 유전자 증거에 대한 해석의 기초가 되는 정확률 및 우도비의 계산 방법 및 상염색체와 Y 염색체를 이용한 개인식별의 통합적 접근 방법에 대한 연구 결과들을 종합적으로 평가하였다. 또한, 전체 집단이 여러 개의 아집단으로 구성된 경우에 있어 공통조상계수의 추정방법을 정리하였으며 가상의 유전자 자료를 분석하여 통계적 분석 방법 및 해석에 대한 지침을 제공하였다.

'lineage marker'가 법의학적인 개인 식별에서 유용한 도구로 사용되고 있지만, 상염색체와 비교할 때 식별력이 크게 감소하는 제한점을 지닌다. 이러한 'lineage marker'의 제한점으로 상염색체와 Y 염색체의 통합적 접근 방법에 의해 식별력을 높이고자 하는 연구들이 진행되고 있으나, 상염색체와 'lineage marker'의 유전자 증거에 따른 통합적 접근법은 혈연관계가 없다는 가정을 명백하게 받아들일 수 있는 경우에 한해서 고려될 수 있을 것으로 판단된다. 따라서 두 염색체에 대한 유전자 증거의 평가 결과는 각각 보고하는 방식을 권장하며, 유전자 증거의 통합적 사용에는 주의가 필요하다고 하겠다. 또한, 유전자 증거에 대한 확률적 평가 결과에 대한 판단은 법정에서 이루어져야 하며, 많은 경험을 통한 신중한 판단이 필요할 것이다.

본 연구에서 소개된 'lineage marker'를 이용한 개인 식별에서의 통계적 방법은 'lineage marker'를 이용한 유전자 증거의 평가 과정에서 기초자료로 활용될 수 있을 것으로 판단된다.

참 고 문 헌

1. Jeffreys AJ, Wilson V, Thein SL. Individual-specific 'fingerprints' of human DNA. *Nature* 1985;316:76-9.
2. Di Rienzo A, Peterson AC, Garza JC, et al. Mutational processes of simple-sequence repeat loci in human populations. *Proc Natl Acad Sci USA* 1994;91:3166-70.
3. Brown WM, George M Jr, Wilson AC. Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci USA* 1979;76:1967-71.
4. Ward RH, Frazier BL, Dew-Jager K, et al. Extensive mitochondrial diversity within a single Amerindian tribe. *Proc Natl Acad Sci USA* 1991;88:8720-4.
5. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 1934;26:404-13.
6. Butler JM. *Fundamentals of Forensic DNA Typing*. Elsevier Inc.; 2010. p. 253-396.
7. Buckleton JS, Krawczak M, Weir BS. The interpretation of lineage markers in forensic DNA testing. *Forensic Sci Int Genet* 2011;5:78-83.
8. Sinha SK, Budowle B, Chakraborty R, et al. Utility of the Y-STR typing systems Y-PLEXTM 6 and Y-PLEXTM 5 in forensic casework and Y-STR haplotype database for three major population groups in the United States. *J Forensic Sci* 2004;49:691-700.
9. Walsh B, Redd AJ, Hammer MF. Joint match probabilities for Y chromosomal and autosomal markers. *Forensic Sci Int* 2008;174:234-8.
10. Amorim A. A cautionary note on the evaluation of genetic evidence from uniparentally transmitted markers. *Forensic Sci Int Genet* 2008;2:376-8.
11. Budowle B, Chakraborty R, Carmody G, et al. Source Attribution of a Forensic DNA Profile. *Forensic Science Communications* 2000;2. Available at <http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/jan2001/july2000/source.htm>
12. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution* 1984;38:1358-70.
13. Ewens WJ. The sampling theory of selectively neutral alleles. *Theor Popul Biol* 1972;3:87-112.
14. National Research Council Committee on DNA Forensic Science. *The evaluation of forensic DNA evidence*. Washington, DC: National Academy Press; 1996.
15. Cavalli-Sforza LL, Menozzi P, Piazza A. *The History and Geography of Human Genes*. Princeton University Press; 1994. p. 302-42.
16. Balding DJ, Nichols RA. Significant genetic correlation among Caucasians at forensic DNA loci. *Heredity* 1997;78:583-9.
17. Foreman LA, Smith AFM, Evett IW. Bayesian analysis of deoxyribonucleic acid profiling data in forensic identification applications. *J R Statist Soc* 1997;160:429-69.