

Y-STR 일배체형의 빈도와 부계친족지수 산출을 위한 웹기반 데이터베이스

양인석¹ · 이환영^{1,2} · 박명진¹
양우익¹ · 신경진^{1,2}

¹연세대학교 의과대학 법의학과
²연세대학교 개인식별연구소

접 수 : 2012년 4월 17일
게재승인 : 2012년 5월 10일

이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단 바이오·의료기술개발사업의 지원을 받아 수행된 연구임(No. 2011-0027729).

책임저자 : 신경진
(120-752) 서울시 서대문구 연세로 50
연세대학교 의과대학 법의학과
전화 : (02) 2228-2481
FAX : (02) 362-0860
E-mail: kjshin@yuhs.ac

Web-based Y-STR Database for Haplotype Frequency Estimation and Kinship Index Calculation

In Seok Yang¹, Hwan Young Lee^{1,2}, Myung Jin Park¹, Woo Ick Yang¹,
Kyoung-Jin Shin^{1,2}

¹Department of Forensic Medicine, Yonsei University College of Medicine, Seoul, Korea

²Human Identification Research Center, Yonsei University, Seoul Korea

We developed a web-based Y chromosomal short tandem repeat (Y-STR) database (ySTRmanager, <http://ystrmanager.yonsei.ac.kr>) to facilitate calculation of Y-STR haplotype frequency estimates for random matches and kinship indices for various relationship levels. The ySTRmanager database provides 3 functions: (i) Y-STR haplotype search, (ii) kinship index calculation, and (iii) user database configuration. The Y-STR haplotype search function allows researchers to retrieve Y-STR haplotypes that meet queried Y-STR allele, Y-haplogroup affiliation, and/or sample information from a selected population in the open database, which consists of 12-17 Y-STR loci. The number of matches in a selected population, haplotype frequency estimator, and detailed results for matched and neighbor haplotypes are displayed as a set of search results. The kinship index calculation function provides kinship indices of 2 input Y-STR haplotypes for the relationship represented by the number of meioses, with consideration of target population and mutation rate of each Y-STR. In addition, ySTRmanager allows registered users to configure their own database to store and analyze Y-STR haplotype and/or mutation rate data. The stored Y-STR data can be used in the search function and in the analysis to obtain forensic statistical values. The ySTRmanager will be a useful system to analyze and manage Y-STR data in the practice of forensic genetics.

Key words : Y-chromosome, STR, database, frequency estimator, kinship index, mutation rate

서 론

Y 염색체의 짧은연쇄반복(short tandem repeat; STR) 유전좌는 친족관계에 있는 남성간의 혈연관계를 확인하거나, 부계 혈통의 추적을 위한 계통 연구에서 아주 유용한 도구로서 이용될 수 있다.¹⁻⁴⁾ Y-STR 유전좌들은 Y 염색체 상에서 재조합이 일어나지 않는 부위에 위치하여 돌연변이의 가능성을 제외하면 동일한 유전형이 아버지에게서 아들로 유전되기 때문에⁵⁾ 이들의 유전형은 각 유전좌의 대립유전자 값으로 이루어진 일

배체형(haplotype)으로 표시된다. 따라서 시료로부터 얻어진 Y-STR 일배체형이 다른 시료의 것이나 데이터베이스에 있는 것과 일치하는 경우에 “일치 확률(match probability)”을 제시하기 위해서는 상염색체 STR과 달리 해당 인구 집단 내에서 특정 일배체형이 얼마나 관찰되는지 그 빈도를 조사하는 것이 중요하다.⁶⁾ 이러한 목적을 위해 각 집단에서의 Y-STR 일배체형 정보를 담은 데이터베이스가 필요하게 된다.

현재 웹에서 접근 가능한 대표적인 데이터베이스로는 Y-STR Haplotype Reference Database (YHRD; www.yhrd.org)가 있으며, 전 세계의 다양한 민족으로부터 9개의 유전좌

연구 방법

1. Y-STR 일배체형 자료의 수집

Y-STR 일배체형 자료는 2000년대 중반 이후에 보고된 논문을 바탕으로 다음과 같은 내용을 포함하여 수집하여 데이터베이스화 저장하였으며, 이들 자료들을 대상으로 ySTRmanager의 검색이 수행되도록 하였다.

1) Y-STR 일배체형 자료는 되도록 17개의 유전자(DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385ab, DYS438, DYS439, DYS437, DYS448, DYS456, DYS458, DYS635, YGATA H4)로 구성된 것을 수집하였다.

2) 수집된 Y-STR 일배체형 자료는 민족 혹은 인구 집단에 따라 아프리카인, 서유라시아인, 동아시아인, 오세아니아인, 혼합집단 등의 메타집단(metapopulation)으로 구분할 수 있도록 하였다.

3) 경우에 따라 Y-haplogroup 및 시료가 채취된 지역과 참고문헌과 같은 정보도 함께 저장하였다.

2. 데이터베이스 설계 및 웹페이지의 개발

ySTRmanager는 MySQL 관계형 데이터베이스 프로그램(<http://www.mysql.com>)을 사용하여 공개 데이터베이스와 사용자 데이터베이스로 구성되도록 설정하였다. 각 데이터베이스에는 각 집단의 정보와 이에 속하는 시료의 정보가 서로 구분되어 입력될 수 있도록 그룹 테이블과 시료 테이블로 구성되어 있다. 로그인 없이도 공개 데이터베이스를 이용한 검색 기능과 동일부계의 관계 확인을 위한 부계친족지수 계산 기능을 이용할 수 있으며, 로그인 후에는 사용자 데이터베이스 기능을 부가적으로 이용할 수 있도록 하였다.

ySTRmanager의 웹페이지는 PHP (Personal Hypertext Preprocessor) 및 AJAX (Asynchronous JavaScript and XML)를 이용하여 직관적인 사용자 인터페이스의 구축에 초점을 맞추었다. 특히 특정 기능을 구현함에 있어서 AJAX를 활용하여 입력 값과 결과 값을 연동하여 보여줄 수 있도록 하였다.

3. ySTRmanager에서의 자료 검색

ySTRmanager에서는 Y-STR 일배체형 자료뿐만 아니라 Y-haplogroup이나 시료의 기타 정보를 이용하여 검색을 수행할 수 있도록 하였다. Y-STR 일배체형 검색은 일배체형의 범위, microvariant 대립유전자 및 대립유전자의 중복 또는 결실 그리고 off-ladder 대립유전자 등을 모두 고려하여 수행될 수

로 구성된 minimal haplotype 99,258개, 17개의 유전자로 구성된 Yfiler haplotype 40,987개의 정보를 저장하고 있다.⁷⁾ 그 외에도 US Y-STR Database (www.usystrdatabase.org)가 있는데, 여기에는 11개의 유전자로 구성된 extended haplotype부터 17개의 유전자로 구성된 Yfiler haplotype까지 총 18,719개의 정보를 담고 있다.⁸⁾ 이와 같이 YHRD와 US Y-STR database는 다량의 일배체형 자료를 바탕으로 검색 기능을 제공하고, 이로부터 특정 일배체형의 빈도를 추정할 수 있게 해준다. 하지만, YHRD에서는 1일 검색 횟수를 제한하고 있을 뿐만 아니라 검색 결과에서 일치하는 Y-STR 일배체형이 많은 경우에는 빈도가 높은 일부 자료만을 보여준다. 반면 US Y-STR database는 검색 과정에서 YHRD와 같은 제한은 없지만, 대부분 미국 현지인을 대상으로 자료를 얻었기 때문에, 여기에서 얻게 되는 Y-STR 일배체형의 빈도 자료는 미국인에 국한될 수밖에 없는 한계가 있다.

Y-STR 일배체형 자료는 아버지와 아들의 관계뿐만 아니라 동일한 부계혈통의 남자들 간의 관계를 확인하기 위한 유전자 검사 결과의 해석에 이용되는데,^{2,5)} 여기서 얻어지는 부계친족지수(kinship index, KI)는 이들의 관계를 설명하는데 중요한 통계지표가 된다. 부계친족지수는 유전자검사서 얻어진 두 개의 Y-STR 일배체형이 서로 완벽하게 일치한 경우에는 집단에서 그 일배체형이 나타난 빈도로부터 계산할 수 있고, 돌연변이가 일어나 Y-STR 유전자의 대립유전자 반복수(repeat number)가 한 개 증가하거나 감소한 경우(+1 또는 -1 repeat)에는 Rolf 등⁹⁾이 제시한 방법에 따라 산출할 수 있다. 이 방법은 현재까지 보고된 Y-STR 유전자에서의 돌연변이가 대부분 대립유전자 반복수가 한 개 증가하거나 감소하는 것으로 나타난다는 점에서¹⁰⁻¹³⁾ 법의 실무에 쉽게 적용될 수 있다. 하지만 이 방법은 특정 유전자의 돌연변이율이 아닌 평균 돌연변이율을 사용한다는 점에서 특정 유전자의 돌연변이에 의한 영향을 제대로 반영한다고 할 수 없다. 따라서 보다 정확한 부계친족지수를 계산하기 위해서는 각각의 유전자의 돌연변이율을 반영할 필요성이 있고, 더불어 많은 시료를 분석하는 법의 실무에서는 다양한 조건에서 부계친족지수를 자동으로 산출하여 제시하는 프로그램의 개발이 요구되고 있다.

본 연구에서는 한국인의 Y-STR 일배체형 자료를 포함하여 2000년대 중반 이후에 논문을 통해 보고된 여러 민족 또는 인구집단에 속하는 13,700여개의 Y-STR 일배체형 자료 수집하여 데이터베이스화하고 이를 관리할 수 있는 웹 프로그램인 ySTRmanager를 개발하였다. ySTRmanager는 수집된 자료를 대상으로 Y-STR 일배체형을 검색할 수 있는 기능뿐만 아니라 특정 Y-STR 유전자에서 돌연변이가 나타났을 경우 그 유전자의 돌연변이율을 적용하여 다양한 조건에서 자동으로 부계친족지수를 계산하는 기능과 사용자의 Y-STR 자료를 저장하고 관리할 수 있는 기능을 제공할 수 있도록 설계되었다.

있도록 설계하였고, 나아가 각 항목별로 독립적으로도 검색할 수 있는 기능도 준비하였다. 또한 로그인 후에는 사용자 데이터베이스에 저장된 Y-STR 일배체형 각각에 대한 검색 또는 이를 대상 집단으로 수행하는 검색 그리고 두 개 집단 내의 모든 Y-STR 일배체형에 대한 교차비교도 가능하도록 하였고, 이 경우에도 Y-STR 일배체형의 범위나 대립유전자의 상태를 고려한 검색이 수행될 수 있도록 관련 옵션을 두었다.

4. Y-STR 일배체형의 빈도 추정

검색된 특정 Y-STR 일배체형의 빈도를 추정하기 위해서 ySTRmanager에서는 Clopper와 Pearson이 제시한 방법¹⁴⁾을 바탕으로 95%의 단측신뢰구간(0-95% 범위)의 상단 값(p_0)을 계산하여¹⁵⁾ 추정된 빈도 값을 제공한다.

$$\sum_{k=0}^x \binom{n}{k} p_0^k (1-p_0)^{n-k} = 0.05 \quad \text{Eq. (1)}$$

여기서 x 는 집단 내에서 특정 Y-STR 일배체형과 일치하는 수이며, n 은 집단의 크기를 나타낸다. 또한 이 방법은 일치하는 Y-STR 일배체형이 없을 때에도 빈도 값의 추정이 가능한데, 이러한 경우에는 Eq. (1)이 $p_0 = 1 - 0.05^{1/n}$ 으로 되며, $n \geq 100$ 인 집단에서 $3/n$ 으로 간단히 p_0 사용할 수 있다.

5. 돌연변이율에 기반한 부계친족지수 산출

ySTRmanager에서는 두 개의 Y-STR 일배체형이 일치하는 경우뿐만 아니라 그렇지 않은 경우에서도 돌연변이를 고려하여 부계친족지수를 산출하여 제시한다. 보다 정확한 부계친족지수 값을 제시하기 위해서 모든 유전좌에서 나타나는 돌연변이율(mutation rate)의 평균값이 아닌, 각 유전좌에서의 돌연변이율 값을 사용하는 Buckleton 등의 방법¹⁶⁾을 프로그램으로 구현하였으며, 이에 필요한 Y-STR 돌연변이율은 STRbase (<http://www.cstl.nist.gov/biotech/strbase/>)에서 제공하는 자료¹⁷⁾를 이용하였다.

1) 두 개의 Y-STR 일배체형이 완벽히 일치하는 경우

$$KI = \frac{\prod_{l=1}^N (1 - \mu_l)^m}{f} \quad \text{Eq. (2)}$$

2) 돌연변이로 인하여 두 개의 Y-STR 일배체형이 일치하지 않는 경우

$$KI \approx \frac{\prod_{l=1, l \neq k}^N (1 - \mu_l)^m \mu_k^{x-y} (1 - \mu_k)^{m-1}}{f} \approx \frac{\prod_{l=1, l \neq k}^N (1 - \mu_l)^m \mu_k (1 - \mu_k)^{m-1}}{2f} \quad \text{Eq. (3)}$$

여기서 N 은 Y-STR 유전좌의 총 수, l 은 돌연변이가 없는 유전좌, k 는 돌연변이가 일어난 유전좌, μ_l 및 μ_k 는 각 유전좌의 돌연변이율, m 은 감수분열의 수, f 는 Y-STR 일배체형의 빈도를 나타낸다. 또한 μ_k^{x-y} 는 k 유전좌에서 대립유전자가 $x \rightarrow y$ 로 되는 돌연변이율이다. 여기서 $x \rightarrow y$ 가 +1 또는 -1 단계로 나타나는 돌연변이이고, 이것이 k 유전좌에서 관찰되는 모든 돌연변이의 대략 1/2을 차지한다는 가정에 의해서 μ_k^{x-y} 는 $\mu_k/2$ 로 대체될 수 있다.

부계친족지수의 산출을 위해서는 두 개의 Y-STR 일배체형 자료와 함께 감수분열의 수, 일배체형 빈도의 추정을 위한 대상 집단, 그리고 각 유전좌에 대한 돌연변이율 자료를 선택하도록 하였다. 감수분열의 수는 1에서 10까지 선택이 가능하며, 1은 아버지와 아들, 2는 할아버지와 손자 또는 형제간의 관계를 나타낸다. 각 유전좌에서의 돌연변이율은 기본적으로 제공되는 자료를 사용할 수도 있지만, 사용자가 자신의 자료를 데이터베이스에 업로드한 후에 그 자료를 이용할 수도 있다. DYS385ab처럼 유전좌의 중복으로 multi-copy로 존재하는 유전좌의 경우에 Y-STR 일배체형을 표기할 때에는 하나의 유전좌처럼 다루겠지만, 부계친족지수의 계산 과정에서 이들은 각각의 유전좌로 구분하여 처리된다.

6. 사용자 데이터베이스 관리

(1) Y-STR 일배체형 자료 관리와 업로드

사용자 자신의 Y-STR 일배체형 자료를 관리하기 위하여 새로운 그룹을 생성하거나 기존 그룹의 정보를 수정 혹은 삭제할 수 있는 기능을 준비하였다. 사용자 데이터베이스에 Y-STR 일배체형 자료를 업로드하기 위해서는 탭으로 구분된 형식의 텍스트 파일을 작성해야 한다. 여기에는 시료명, 17개 유전좌의 대립유전자 값, Y-haplogroup, 시료의 기타 정보가 포함되는데, 웹페이지의 링크를 통해 예제 파일을 다운로드할 수 있도록 하였다. 각 Y-STR의 대립유전자 값의 다양한 형태를 고려하여 자료를 데이터베이스에 업로드하는 과정에서 자료의 정확성 검사를 수행하도록 하였다. 또한 업로드된 Y-STR 일배체형 및 대립유전자에 대한 일배체형의 빈도, 일배체형의 다양성, 유전좌의 다양성 등과 같은 기본적인 통계량을 분석할 수 있고, 더불어 업로드된 Y-STR 일배체형 자료의 정보를 수정하거나 특정 자료를 삭제할 수 있는 기능도 설계하였다.

(2) Y-STR 유전좌의 돌연변이율 자료 관리

17개의 Y-STR 유전좌에 대해서 각각 돌연변이율 값을 탭으로 구분된 형식의 텍스트 파일로 작성하여 입력할 수 있도록 하였다. 돌연변이율 값은 5번째 자리에서 반올림하여 소수점 4 자리까지 입력하고, Y-STR 일배체형 자료와 마찬가지로 예제 파일을 다운로드하여 편리하게 사용자 자신의 돌연변이율 자료를 구축할 수 있도록 하였다.

결 과

ySTRmanager는 <http://ystrmanager.yonsei.ac.kr>로 접속하여 사용이 가능하며, 특정 Y-STR 일배체형의 빈도를 추정하기 위한 검색 기능, 동일부계의 관계를 확인하기 위한 부계 친족지수 계산 기능, 그리고 사용자 자료의 업로드 및 통계량 분석을 위한 기능을 제공하고 있다.

1. Y-STR 일배체형 데이터베이스 및 검색 기능

ySTRmanager에는 동아시아인, 서유라시아인, 혼합집단으로 구분되는 3개 메타집단에 속하는 총 35개 민족 또는 인구 집단으로부터 얻어진 13,790개 Y-STR 일배체형 자료가 저장되어 있다(Table 1). 각 집단의 Y-STR 일배체형 자료들은 대부분 12 또는 17개의 유전자좌로 구성되어 있으며, 이들은 특정 Y-STR 일배체형의 빈도를 추정하기 위한 검색에서 공개 데이터베이스로 제공한다.

특정 Y-STR 일배체형의 검색은 그 일배체형의 범위(haplotype range)를 선택하고 각 유전자좌에서의 대립유전자 값을 입력한 후에 대상 집단을 선택함으로써 이루어진다. 일배체형의 범위는 minimal haplotype (9 loci), extended

haplotype (11 loci), PowerPlex Y haplotype (12 loci), Yfiler haplotype (17 loci) 중에서 하나를 선택함으로써 검색에 포함되는 유전자좌들을 지정할 수 있다. 각 유전자좌에서의 대립유전자 값은 목록 상자에서 선택하여 입력하는데, 만약 대립유전자 값이 없다면 “Others”를 선택한 후에 원하는 값을 텍스트 상자에 직접 입력해야 한다. 특정 대립유전자 값(예, 14 또는 14.1)을 입력한 경우에는 그와 일치하는 결과만 얻을 수 있는 반면에, 와일드카드(*)를 사용하여 대립유전자 값(예, 14.*)을 입력하게 되면 일반적인 대립유전자(예, 14) 및 microvariant 대립유전자(예, 14.1)를 모두 결과로 얻을 수 있다. 와일드카드는 또한 multi-copy 유전자좌의 검색에도 사용할 수 있다. 예를 들어 “14.*,15”라고 입력한 경우라면 “14,15”, “14.1,15”, “14.2,15”가 모두 검색될 것이다. 대립유전자 값을 직접 입력하는 경우에는 그 값이 잘못 입력되는 것을 막기 위해 아래의 규칙을 따라야 한다. 유전자의 중복으로 인하여 두 개 이상의 대립유전자 값이 입력되는 경우에는 쉼표로 구분하여 입력한다(예, DYS19: 14,15 또는 DYS385ab: 16,20). 또한 그 값이 작은 순서대로 입력해야 한다(예, 20,16 → 16,20). DYS385ab 유전자좌에서 두 개의 대립유전자 값이 동일한 경우라도 이들을 모두 표기해야 한다(예, DYS385ab: 16 → 16,16). 특정 유전자좌에서 유전자의 결실이 관찰된 경우(null allele) 대립유전자 값은 숫자 “0”으로 표기한다. 하지만, 그 외

Fig. 1. Search setting in ySTRmanager. In order to search a particular Y-STR haplotype (Y-haplotype) from a selected population, user should provide Y-haplotype information (circled “1”) and then select a target population (circled “2”). Y-STR allele values can be provided by selecting a value in list box for each Y-STR loci interested or by direct input of allele values into text box after selecting “Others” in list box. Y-haplogroup and/or sample information can be combined to Y-haplotype or used independently to search haplotypes in a selected population.

의 경우에서 대립유전자의 값을 결정할 수 없을 때에는 빈칸으로 두면 된다. Off-ladder 대립유전자(예, <10 또는 >19)도 입력이 가능하며, “>19”로 입력했다면 “20” 이상의 일치하는 값이 있을 때에 검색 결과로 출력된다. YGATA H4 유전좌는 8–15(H4)의 ladder 범위로 대립유전자 값을 입력해야 한다. 만약 17–24(H4.1) 또는 24–31(H4.1+H4.2)의 ladder 범위에서 대립유전자 값이 정해졌다면, 통일된 명명법의 사용을 위해 이들의 대립유전자 값에서 각각 9와 16을 빼줌으로써 교정해야 한다.

Y-STR 일배체형 자료와 함께 Y-haplogroup 또는 시료의 기타 정보를 종합하여 검색할 수도 있고, Y-haplogroup 또는 시료의 기타 정보만을 이용한 검색을 통하여 입력된 정보와 관련된 Y-STR 일배체형의 존재 여부도 알아 볼 수 있다. 이를 위해 Y-haplogroup 검색에 대해서는 두 가지 옵션을 지원한다: i) 입력한 값과 완벽하게 일치하는 경우(Exact match), ii) 입력한 단어가 포함되는 경우(Include the keyword). 한편 시료의 기타 정보 검색에서는 입력된 검색어를 포함하는 모든 결과를 출력한다.

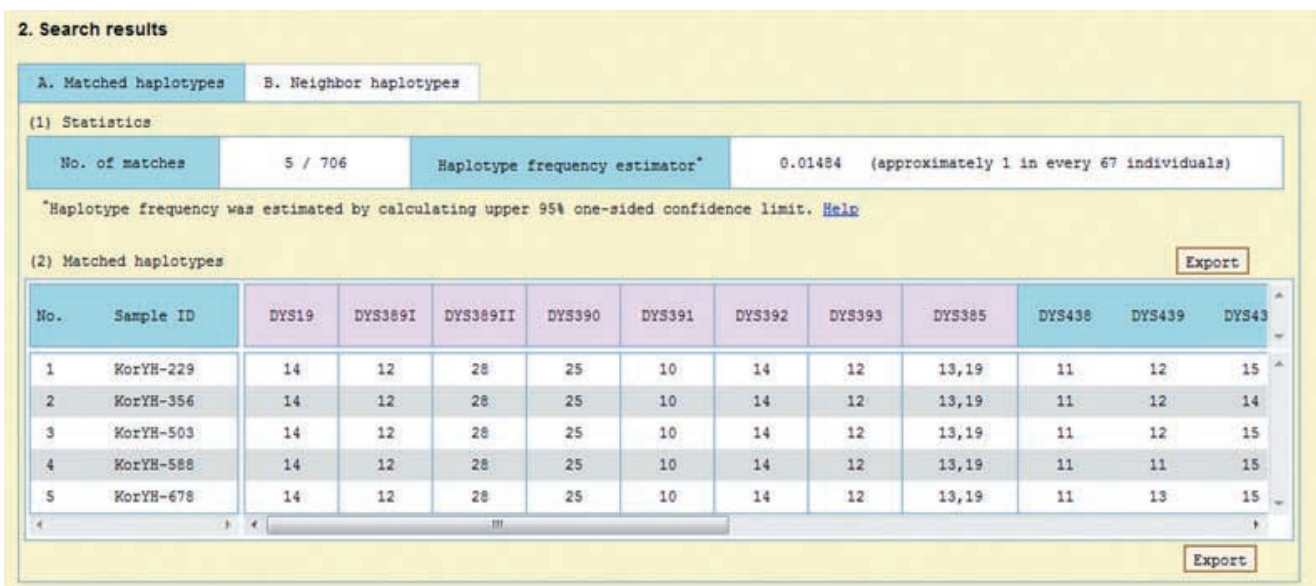
검색결과는 선택된 집단 내에서 일치하는 Y-STR 일배체형(matched haplotype; Fig. 2A)과 이웃하는 일배체형(neighbor haplotype; Fig. 2B)을 탭 메뉴로 구성하여 보여준다. 일치하는 일배체형에 대해서는 선택된 집단 내에서 얼마나 관찰되는지를 보여주고, 이들의 빈도는 Clopper와 Pearson의 방법¹⁴⁾에 따라 95%의 단측신뢰구간의 상단 값¹⁵⁾으로 추정 값을 제시하였다. 이들 일배체형을 갖는 시료에 대해서 각 유전좌에서의 대립유전자 값과 함께 시료의 기타 정보 및 Y-haplogroup 정보도 결과에서 확인할 수 있으며, “Export” 버튼을 이용하여 이들 자료를 탭으로 구분된 텍스트 파일로 저장할 수 있다. 이웃하는 일배체형에 대해서는 입력한 일배체형을 기준으로 각 일배체형 당 하나의 유전좌에서 +1 또는 -1 단계의 돌연변이를 포함하고 있으며, 이들이 집단 내에서 관찰되는 수가 작은 순서대로 정렬하여 보여준다. 이러한 돌연변이는 동일 부계의 관계를 확인하는 검사에서 나타날 수 있기 때문에, 이웃하는 일배체형을 이용하여 계산된 부계친족지수 결과도 함께 보여주었다. 이웃하는 일배체형의 결과에서 감수분열의 수 및 돌연변이율 자료를 선택함으로써 다양한 조건으로 부계친족

Table 1. Information of Y-STR Haplotype Data Collected in ySTRmanager

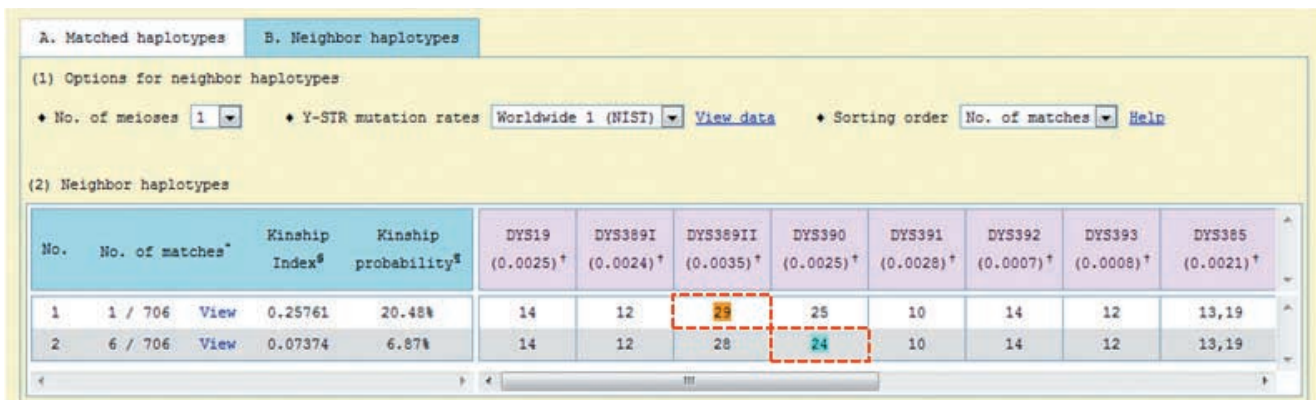
Metapopulation	Population ^a	No. of haplotypes	No. of Y-STR loci	Reference
East Asian	Korean (3)	2253	17	18 – 20
	Chinese Han (7)	1104	11, 12, or 17	21 – 27
	Chinese 8 minor populations ^b	1337	11, 12, or 17	28 – 35
	Japanese (2)	2245	17	36, 37
	Taiwanese Han	200	17	38
	Taiwanese Paiwan	208	17	39
	Malay (2)	520	12 or 17	40, 41
	West Eurasian	Austrian	135	17
Danish		185	12	43
German		279	11	44
Hungarian		215	12	45
Italian		155	17	46
Polish		255	17	47
Portuguese (2)		425	17	48, 49
Resident Basques		197	17	50
Russian		545	17	51
Serbian		185	17	52
Spanish (2)		395	14 or 17	53, 54
Swiss		150	12	55
UK Caucasian		250	12	56
US Caucasian		100	17	57
Admixed		Argentine	224	12
	Brazilian	500	17	59
	Colombian (2)	950	9 or 13	60, 61
	Ecuadorian	120	13	62
	Honduran	128	12	63
	Mexican-Mestizo	357	12	64
	Venezuelan	173	12	65
	Total	35 populations	13,790 haplotypes	

^a If the same population data was collected from two or more different sources, the number of sources was indicated in parentheses.

^b Chinese 8 minor populations indicate Hui, Korean, Manchu, Naxi, Salar, Tibetan, Tuija, and Yi populations living in China.



a



b

Fig. 2. Search results of ySTRmanager. The number of matches for the particular haplotype and its haplotype frequency estimator in a selected population are displayed in search result (the top table of a). Individual detailed information including sample name, Y-haplogroup affiliation, and sample information for matched haplotypes are shown in a table, which can be downloaded by clicking “Export” button (the bottom table of a). Neighbor haplotypes are displayed in a table only when there is at least one matched haplotype in a selected population (b). Allele values at Y-STR loci with single-step mutations (+1 repeat gain or -1 repeat loss) in neighbor haplotypes are represented by different backgrounds (dashed box in b). Detailed sample information having the neighbor haplotypes can also be appeared in a new window by clicking “View” link.

지수의 값을 계산할 수 있다. 또한 일치하는 일배체형의 수, 부계친족지수, 돌연변이율의 3가지 옵션으로 이웃하는 일배체형의 결과를 정렬시킬 수 있다. 이웃하는 일배체형도 “View” 링크를 통해서 선택된 집단 내에서의 이들 일배체형을 갖는 시료의 정보를 확인할 수 있다. 위의 검색결과에서처럼 특정한 하나의 집단 내에서 뿐만 아니라, ySTRmanager의 공개 데이터베이스에 저장된 여러 집단에 대해서 입력한 Y-STR 일배체형이 어떠한 빈도로 나타나는지도 알아 볼 수 있다(Worldwide frequency).

ySTRmanager는 로그인한 사용자가 자신의 데이터베이스를 구축한 경우 좀 더 다양한 검색기능을 제공하고 있다. 사용

자 데이터베이스에 포함된 Y-STR 일배체형이 특정 집단 내에서 이들 각각이 어떠한 빈도를 나타나는지 알아 볼 수 있다. 이 경우에 공개 데이터베이스 또는 사용자 데이터베이스의 Y-STR 일배체형 자료 모두를 대상 집단으로 지정할 수 있다. 또한 각 사용자 Y-STR 일배체형 자료도 특정 일배체형의 빈도를 추정하기 위한 대상 집단으로도 이용될 수 있다. 이러한 기능과 더불어 선택된 두 개 집단에 포함된 모든 Y-STR 일배체형 자료에 대한 교차비교 기능을 제공한다(Cross comparison). 이 경우에 비교하고자 하는 집단은 사용자 데이터베이스에서만 선택 가능하지만, 비교의 대상이 되는 집단은 공개 및 사용자 데이터베이스 모두에서 선택할 수 있다. 한편

사용자 데이터베이스에 포함된 Y-STR 일배체형에 대한 검색을 수행할 경우에는 대립유전자형에 와일드카드를 사용할 수 없기 때문에, 이를 보완하기 위해서 microvariant 대립유전자를 무시하고 검색할 것인지 여부(Ignore microvariant allele)를 묻는 옵션을 두었다. 만약에 사용자가 microvariant 대립유전자를 무시하는 옵션을 선택하였다면 검색 대상이 되는 Y-STR 일배체형의 모든 유전좌에 대해 적용된다. 예를 들어 검색하고자 하는 일배체형이 “14.2-12-28-22-13,16”으로 주어졌다고 하면, 이것은 “14.*-12.*-28.*-22.*-13.*,16.*”로 일배체형을 직접 입력한 경우와 동일한 결과가 나오게 된다.

2. 다양한 조건에서의 부계친족지수 산출

ySTRmanager에서는 Y-STR 일배체형 자료를 이용한 동일부계의 관계의 확인을 위한 통계지표로 사용되는 부계친족지수를 계산해주는 기능을 제공한다. Table 2에서는 아버지와 두 아들간의 친자검사를 예로서 보여주고 있다. 이 경우 감수분열의 수는 아버지와 아들의 관계이므로 1이 된다. 두 아들의 일배체형의 빈도는 706명의 한국인 집단에서 나타나는 빈도를 이용하여 추정하였다. 이들 일배체형 자료를 비교하였을 때, 한 명의 아들은 아버지와 완벽하게 일치하는 반면에, 다른 한 명의 아들은 하나의 유전좌에서 차이(DYS389II: 28 → 27)가 확인된다(Table 2A). 그리고 이 두 가지 경우에 대해서 각각의 부계친족지수를 계산하였다(Table 2B). 먼저 두 개의 일배체형이 일치하는 경우와 그렇지 않은 경우에서 얻어진 부계친족지수를 비교하였을 때에 크게 차이를 보임을 확인할 수 있다(146.38209 vs. 0.25707). 다음으로 각 유전좌마다의 돌연변이율을 사용한 경우와 이들의 평균값을 사용한 경우를 비

교하였을 때, 두 개의 일배체형이 완벽히 일치하는 경우에는 얻어진 부계친족지수간의 차이가 크게 나타나지 않지만(146.38209 vs. 146.32388), 그렇지 않은 경우에는 상대적으로 그 차이가 커짐을 알 수 있다(0.25707 vs. 0.15856). Minimal haplotype에서 돌연변이율의 평균값은 0.0022이며, DYS389II 유전좌의 돌연변이율은 0.0035이다. 여기에서 짐작할 수 있듯이 돌연변이가 관찰된 유전좌의 돌연변이율이 높을수록 계산되는 부계친족지수도 커짐을 알 수 있다. 따라서 각 유전좌마다의 돌연변이율 값을 사용함으로써 Y-STR 일배체형 자료를 이용한 동일부계의 확인검사에서도 보다 정확한 부계친족지수를 얻을 수 있다.

3. 사용자 자료의 업로드 및 통계량 분석

사용자는 웹사이트의 “Y-STR data” 및 “Mutation data” 메뉴를 통해서 특정 집단에서 얻어진 Y-STR 일배체형, Y-haplogroup, 시료의 기타 정보 및 각 Y-STR 유전좌에서 발생할 수 있는 돌연변이에 대한 빈도 자료를 사용자 데이터베이스에 업로드할 수 있고, 이러한 자료들의 관리와 업데이트를 수행할 수 있다. 한편 사용자 자료를 업로드하는 과정에서 자료의 정확성 검사를 수행하는 기능을 돕으로써 데이터베이스에 저장되기 전에 가능한 오류(예, 불필요한 문자, 대립유전자 또는 돌연변이율 값이 비정상적인 범위에 있는 경우, microvariant 대립유전자 값을 형식에 맞지 않게 표기한 경우)를 줄일 수 있다. 또한 업로드된 Y-STR에 대한 기본적인 통계량(일배체형의 빈도, 일배체형의 다양성, 유전자의 다양성 등)도 확인할 수 있는데, 이러한 정보는 수집된 Y-STR 자료의 특성을 파악하는데 도움이 될 것이다(자료 제시 없음).

Table 2. An Example of Kinship Test Among Alleged Father and Two Sons

A. Comparison of Y-STR Haplotypes Among Alleged Father and Two Sons

Loci	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385
Mutation rates ^a	0.0025	0.0024	0.0035	0.0025	0.0028	0.0007	0.0008	0.0021
Alleged father	14	12	28	23	10	14	12	13,20
Son 1	14	12	28	23	10	14	12	13,20
Son 2	14	12	27	23	10	14	12	13,20

^aMutation rates for all Y-STR loci were obtained from STRbase.¹⁷⁾

B. Kinship Index Calculation for the Above Two Cases

	Alleged father and son 1	Alleged father and son 2
Matched count for son's haplotype in a population (M / N) ^a	1 / 706	1 / 706
Frequency estimate for son's haplotype ^b	0.00670	0.00670
Kinship index	146.38209 (146.32388) ^c	0.25707 (0.15856) ^c
Kinship probability (prior probability: 0.5)	98.32% (98.32%) ^c	20.45% (13.69%) ^c

^aM and N indicate matched samples and population size, respectively.

^bFrequency estimate was obtained by calculating upper 95% one-sided confidence interval as shown in Eq. (1).

^cThe values in parentheses were calculated using the average value (0.0022) of mutation rates for the minimal haplotype.

고 찰

ySTRmanager 데이터베이스에는 동아시아인, 서유라시아인, 혼합집단에 속하는 민족 또는 인구 집단으로부터 수집된 12 또는 17개의 Y-STR 유전자로 구성된 Y-STR 일배체형 자료가 집적되어 있으며, 이것을 대상으로 ySTRmanager는 특정 Y-STR 일배체형에 대한 검색 결과를 제공하고 있다. 이때 Y-STR 일배체형 자료뿐만 아니라 Y-haplogroup 및 시료의 기타 정보를 조합하여 검색을 수행할 수 있도록 함으로써 사용자에게 다양한 조건의 검색 기능을 제공하고 있다. 이러한 점은 사용자가 데이터베이스에 저장된 자신의 Y-STR 자료를 대상으로 검색하고자 할 때에 더 유용하게 이용될 것이다. 물론 Y-haplogroup을 이용한 검색은 YHRD에서도 지원되고 있지만, ySTRmanager에서와 달리 YHRD에서는 Y-STR 일배체형을 이용한 검색과는 별개로 작동한다. 또한 microvariant 대립유전자 검색을 위한 와일드카드(*) 및 여러 옵션을 제공함으로써 법의 실무에 유연하게 대처할 수 있도록 하였다.

검색을 통해서 얻어지는 일배체형의 빈도를 추정하는 방법에 있어서 YHRD는 베이지안 접근법으로 일배체형의 빈도를 전망하는 방법(frequency surveying method)⁶⁶⁾을 사용하는 반면에, US Y-STR database에서는 ySTRmanager와 동일하게 Clopper와 Pearson의 방법¹³⁾을 따른다. ySTRmanager가 YHRD와 US Y-STR database와 비교해서 특징적인 부분은 동일부계의 확인 검사에서 사용되는 부계친족지수의 계산 기능과 사용자의 자료를 데이터베이스에 업로드하고 이를 관리하게 해주는 기능에 있다. 이 두 가지 기능은 검색 기능 못지않게 Y-STR 자료를 다루는 연구자들의 실무에 필요한 사항이 될 것이다. 특정 일배체형의 빈도(p)는 크기가 n 인 집단에서 x 만큼 관찰되는 경우에 $p = x/n$ 로 나타낼 수 있다. 하지만, 이러한 방법은 한 집단 내에서 시료의 수집이 어떻게 이루어지느냐에 따라서 빈도가 달라질 가능성이 있다. Holland와 Parsons는 정규 근사법(normal approximation method)을 사용하여 95%의 신뢰구간 내에서 시료 수집에 따른 차이를 보정하는 것을 보여주었다($p \pm 1.96 \sqrt{p(1-p)/n}$).⁶⁷⁾ 이 방법은 이해와 계산이 쉽다는 장점을 갖고 있지만, p 가 0 또는 1과 같을 경우에는 위의 방정식으로 계산이 불가능하다. 때문에, ySTRmanager에서는 Clopper와 Pearson이 제시한 방법⁹⁾을 도입하여 이항 분포를 기반으로 정확한 신뢰구간을 계산함으로써 특정 일배체형의 빈도를 추정할 수 있도록 하였다. 실제로 위의 두 가지 방법으로 얻어진 일배체형 빈도에 대한 추정 값을 비교하였을 때, 정규 근사법 보다는 Clopper와 Pearson의 방법에서 큰 값이 얻어지는 것을 확인할 수 있었으며, 특히 수집된 시료의 수(n)가 적을수록 이들 추정 값 사이의 차이가 크게 나타남을 알 수 있었다(자료 제시 없음). 이것은 한 집단에서 수집된 시료의 수에 따라 달라질 수 있는 빈도에 대한 보정을 Clopper와

Pearson의 방법이 더 잘 해주고 있음을 보여주는 것이다. 게다가 이들 추정 값이 클수록 재판에서 피고인에게 우호적으로 작용하여 억울하게 죄를 뒤집어쓰우는 확률을 낮추게 될 것으로 본다.

Y-STR 일배체형은 부계혈통으로만 유전된다는 점을 이용하여 동일부계의 관계를 확인하는 검사에 사용된다.^{2,5)} 여기서 부계친족지수는 법의 실무에서 중요한 통계지표로 사용되는데, 모든 유전자에서 동일한 돌연변이의 빈도를 보인다는 가정을 사용하여 돌연변이율의 평균값을 사용하여 이를 계산해왔다.⁹⁾ 두 개의 Y-STR 일배체형이 완벽하게 일치하는 경우라면 계산되는 부계친족지수의 값이 충분히 크기 때문에, Y-STR 유전자의 돌연변이율 값을 적용함에 있어서 평균값을 사용하든지 아니면 유전자마다 다른 값을 사용하든지 간에 큰 차이가 없을 것이다. 하지만, 검사에 사용된 유전자 중에서 하나라도 돌연변이가 확인될 경우에는 얻어지는 부계친족지수가 크게 감소하기 때문에, Y-STR 유전자의 돌연변이율 값을 적용함에 있어 이들의 평균값을 사용하는 것은 경우에 따라 부계친족지수 계산에서 부정확성을 오히려 키울 수 있다. 이러한 점에서 ySTRmanager는 각각의 돌연변이 Y-STR 유전자에서 관찰된 돌연변이율 값을 사용하여 부계친족지수를 산출하고, 계산 과정에서 사용되는 Y-STR 일배체형 빈도 값을 Clopper와 Pearson이 제시한 방법¹⁴⁾으로 추정된 값을 사용함으로써 부계친족지수의 정확도를 더 높일 수 있었다. 따라서 ySTRmanager에서 특정 Y-STR 일배체형 빈도 추정치로부터 계산되는 부계친족지수는 법의 실무에 보다 유용하게 적용될 수 있을 것으로 여겨진다.

Stepwise mutation model (SMM)⁶⁸⁾은 STR 유전자에서 나타나는 돌연변이를 설명하기에 적합한 모델로 받아들여지고 있으며, 이 모델에 의하면 STR 유전자에서의 돌연변이는 대립유전자의 반복된 수가 하나 또는 작은 수로 증가하거나 감소하는 형태로 나타난다. 특히 돌연변이가 오로지 한 단계만 증가 또는 감소하여 나타나는 모델을 “strict SMM”이라고 한다.⁶⁹⁾ ySTRmanager에서는 각 유전자에서 일어나는 돌연변이가 strict SMM에 따른다는 가정을 바탕으로 부계친족지수를 계산하고 있다. 이러한 접근은 두 단계 이상의 돌연변이, 유전자의 중복 및 결실, 그리고 비전형적 대립유전자가 관찰된 경우에는 적용할 수 없다는 단점이 있다. 하지만, 법의 유전학 및 집단 유전학 분야에서 주로 사용되는 17개의 Y-STR 유전자는 비교적 낮은 돌연변이율($10^{-4} \sim 10^{-3}$)을 보여주고 있으며, 발견된 돌연변이의 대부분은 대립유전자의 크기가 한 단계 증가 또는 감소하는 형태로 나타난다.¹⁰⁻¹³⁾ 따라서 ySTRmanager에서 제공하는 부계친족지수의 계산 기능은 동일부계를 확인하는 검사에서 돌연변이가 관찰되는 다양한 상황에도 유연하게 적용될 수 있을 것으로 생각된다.

지금까지의 대다수의 Y-STR 연구는 Promega의 PowerPlex® Y kit 및 Applied Biosystems의 Yfiler® kit의 보

급이 보편화되면서 12 또는 17개의 유전좌로 구성된 일배체형으로 결과를 보고해왔다.¹⁸⁻⁶⁵⁾ 때문에, YHRD 및 US Y-STR database와 함께 ySTRmanager도 이러한 연구결과의 바탕에서 데이터베이스가 구축되었다. 한편 2009년에 Vermeulen 등⁷⁰⁾은 single-copy 유전좌에 대한 연구를 통해서 이들을 단독으로 사용하거나 기존의 12 또는 17개의 유전좌에 추가하여 사용함으로써 보다 높은 식별력을 얻을 수 있었으며, 특히 DYS481, DYS533, DYS549, DYS570, DYS576, DYS643로 구성된 6개 유전좌가 동일부계의 구분에 유용하다고 보고하였다. 이러한 결과는 최근 Promega의 PowerPlex® Y23 kit의 개발에 반영되었다. 2000년대 중반에 Applied Biosystem사의 Yfiler® kit의 사용이 보편화에 따라 17개의 유전좌에 대한 많은 연구보고가 있었던 것에 비추어 볼 때, 앞으로의 Y-STR 연구에서는 23개의 유전좌를 분석한 것이 많을 것으로 예상된다. 따라서 향후 ySTRmanager도 23개의 Y-STR 유전좌로 구성된 데이터베이스를 관리할 수 있도록 업그레이드하여 이러한 추세에 대비할 필요가 있을 것으로 보인다.

Acknowledgments

ySTRmanager의 개발 과정에서 기능 개선 및 사용자 인터페이스 향상을 위해 좋은 의견을 많이 주었던 연세대학교 법의학과 연구원들(이은영, 안자현, 최아진, 김은혜)에게 감사의 뜻을 전합니다.

참 고 문 헌

1. Roewer L, Arnemann J, Spurr NK, Grzeschik KH, Epplen JT. Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. *Hum Genet* 1992;89:389-94.
2. Jobling MA, Pandya A, Tyler-Smith C. The Y chromosome in forensic analysis and paternity testing. *Int J Legal Med* 1997;110:118-24.
3. Kayser M, Caglia A, Corach D, et al. Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med* 1997;110:125-33, appendix 141-9.
4. de Knijff P, Kayser M, Caglia A, et al. Chromosome Y microsatellites: population genetic and evolutionary aspects. *Int J Legal Med* 1997;110:134-49.
5. Jobling MA, Tyler-Smith C. Fathers and sons: the Y chromosome and human evolution. *Trends Genet* 1995;11:449-56.
6. Roewer L, Kayser M, de Knijff P, et al. A new method for the evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males. *Forensic Sci Int* 2000;114:31-43.
7. Willuweit S, Roewer L; International Forensic Y Chromosome User Group. Y chromosome haplotype reference database (YHRD): update. *Forensic Sci Int Genet* 2007;1:83-7.
8. Ge J, Budowle B, Planz JV, Eisenberg AJ, Ballantyne J, Chakraborty R. US forensic Y-chromosome short tandem repeats database. *Leg Med (Tokyo)* 2010;12:289-95.
9. Rolf B, Keil W, Brinkmann B, Roewer L, Fimmers R. Paternity testing using Y-STR haplotypes: assigning a probability for paternity in cases of mutations. *Int J Legal Med* 2001;115:12-5.
10. Kurihara R, Yamamoto T, Uchihi R, et al. Mutations in 14 Y-STR loci among Japanese father-son haplotypes. *Int J Legal Med* 2004;118:125-31.
11. Lee HY, Park MJ, Chung U, et al. Haplotypes and mutation analysis of 22 Y-chromosomal STRs in Korean father-son pairs. *Int J Legal Med* 2007;121:128-35.
12. Decker AE, Kline MC, Redman JW, Reid TM, Butler JM. Analysis of mutations in father-son pairs with 17 Y-STR loci. *Forensic Sci Int Genet* 2008;2:e31-5.
13. Goedbloed M, Vermeulen M, Fang RN, et al. Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFISTR Yfiler PCR amplification kit. *Int J Legal Med* 2009;123:471-82.
14. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 1934;26:404-13.
15. Buckleton JS, Krawczak M, Weir BS. The interpretation of lineage markers in forensic DNA testing. *Forensic Sci Int Genet* 2011;5:78-83.
16. Buckleton JS, Triggs CM, Walsh SJ. Forensic DNA evidence interpretation. 1st ed. Boca Raton: CRC press; 2005. p. 388-9.
17. Summary list of Y chromosome STR loci and available fact sheets. Available from: http://www.cstl.nist.gov/strbase/ys-tr_fact.htm.
18. Park MJ, Lee HY, Kim NY et al. Genetic characteristics of 22 Y-STR loci in Koreans. *Korean J Leg Med* 2007;31:162-70.
19. Kim SH, Kim NY, Hong SB, et al. Genetic polymorphisms of 16 Y chromosomal STR loci in Korean population. *Forensic Sci Int Genet* 2008;2:e9-10.
20. Seong KM, Yoo SY, Hwang JH, Kim SH, Chung KW, Cho NS. Population genetic polymorphisms of 17 Y-chromosomal STR loci in South Koreans. *Forensic Sci Int Genet* 2011;5:e122-3.
21. Hu SP. Polymorphism of Y-chromosomal STR haplotypes in the Chaoshan Han Chinese in South China. *Forensic Sci Int* 2006;158:80-5.
22. Hu SP. Genetic polymorphism of 12 Y-chromosomal STR loci in the Minnan Han Chinese in Southeast China. *Forensic Sci Int* 2006;159:77-82.
23. Zhu B, Deng Y, Zhang F, et al. Genetic analysis for Y chromosome short tandem repeat haplotypes of Chinese Han population residing in the Ningxia province of China. *J Forensic Sci* 2006;51:1417-20.
24. Yang B, Gu M, Wang G, Li X, Liu Y, Yang W. Population data for 11 Y-chromosome STRs in northeast China Han.

- Forensic Sci Int 2006;164:65-71.
25. Yan J, Tang H, Liu Y, et al. Genetic polymorphisms of 17 Y-STRs haplotypes in Chinese Han population residing in Shandong province of China. *Leg Med (Tokyo)* 2007;9:196-202.
 26. Zhang H, Yun L, Li Y, et al. Haplotype of 12 Y-STR loci of the PowerPlex Y-system in Sichuan Han ethnic group in west China. *Forensic Sci Int* 2008;175:244-9.
 27. Yanmei Y, Tao G, Yubao Z, et al. Genetic polymorphism of 11 Y-chromosomal STR loci in Yunnan Han Chinese. *Forensic Sci Int Genet* 2010;4:e67-9.
 28. Guo H, Yan J, Jiao Z, et al. Genetic polymorphisms for 17 Y-chromosomal STRs haplotypes in Chinese Hui population. *Leg Med (Tokyo)* 2008;10:163-9.
 29. Zhang Y, Zhang H, Cui Y, et al. Population genetics for Y-chromosomal STRs haplotypes of Chinese Korean ethnic group in northeastern China. *Forensic Sci Int* 2007;173:197-203.
 30. Bai R, Shi M, Yu X, Zhang J, Bai L. Y-chromosomal STRs haplotypes in Chinese Manchu ethnic group. *Forensic Sci Int Genet* 2008;3:e13-5.
 31. Xin N, Chen T, Yu B, Li S. 12 Y-STRs haplotypes in Chinese Naxi ethnic minority group. *Forensic Sci Int* 2008;174:244-8.
 32. Zhu B, Shen C, Xun X, et al. Population genetic polymorphisms for 17 Y-chromosomal STRs haplotypes of Chinese Salar ethnic minority group. *Leg Med (Tokyo)* 2007;9:203-9.
 33. Yi Z, Jun W, XingBo S, XiaoJun L, Liu D, BinWu Y. Genetic analysis of 17 Y-chromosomal STRs haplotypes of Chinese Tibetan ethnic minority group. *Leg Med (Tokyo)* 2010;12:108-11.
 34. Shi M, Bai R, Wan L, Yu X, Chang L. Population genetics for Y-chromosomal STRs haplotypes of Chinese Tujia ethnic group. *Forensic Sci Int Genet* 2008;2:e65-8.
 35. Zhu B, Shen C, Qian G, et al. Genetic polymorphisms for 11 Y-STRs haplotypes of Chinese Yi ethnic minority group. *Forensic Sci Int* 2006;158:229-33.
 36. Mizuno N, Nakahara H, Sekiguchi K, Yoshida K, Nakano M, Kasai K. 16 Y chromosomal STR haplotypes in Japanese. *Forensic Sci Int* 2008;174:71-6.
 37. Hashiyada M, Umetsu K, Yuasa I, et al. Population genetics of 17 Y-chromosomal STR loci in Japanese. *Forensic Sci Int Genet* 2008;2:e69-70.
 38. Hwa HL, Tseng LH, Ko TM, et al. Seventeen Y-chromosomal short tandem repeat haplotypes in seven groups of population living in Taiwan. *Int J Legal Med* 2010;124:295-300.
 39. Wu FC, Ho CW, Pu CE, et al. Y-chromosomal STRs haplotypes in the Taiwanese Paiwan population. *Int J Legal Med* 2011;125:39-43.
 40. Chang YM, Perumal R, Keat PY, Kuehn DL. Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations (Malays, Chinese and Indians) in Malaysia. *Forensic Sci Int* 2007;167:70-6.
 41. Yong RY, Lee LK, Yap EP. Y-chromosome STR haplotype diversity in three ethnic populations in Singapore. *Forensic Sci Int* 2006;159:244-57.
 42. Berger B, Lindinger A, Niederstätter H, Grubwieser P, Parson W. Y-STR typing of an Austrian population sample using a 17-loci multiplex PCR assay. *Int J Legal Med* 2005;119:241-6.
 43. Hallenberg C, Nielsen K, Simonsen B, Sanchez J, Morling N. Y-chromosome STR haplotypes in Danes. *Forensic Sci Int* 2005;155:205-10.
 44. Rodig H, Roewer L, Gross A, et al. Evaluation of haplotype discrimination capacity of 35 Y-chromosomal short tandem repeat loci. *Forensic Sci Int* 2008;174:182-8.
 45. Völgyi A, Zalán A, Szvetnik E, Pamjav H. Hungarian population data for 11 Y-STR and 49 Y-SNP markers. *Forensic Sci Int Genet* 2009;3:e27-8.
 46. Turrina S, Atzei R, De Leo D. Y-chromosomal STR haplotypes in a Northeast Italian population sample using 17plex loci PCR assay. *Int J Legal Med* 2006;120:56-9.
 47. Soltyszewski I, Pepinski W, Spolnicka M, Kartasinska E, Konarzewska M, Janica J. Y-chromosomal haplotypes for the AmpFISTR Yfiler PCR Amplification Kit in a population sample from Central Poland. *Forensic Sci Int* 2007;168:61-7.
 48. Alves C, Gomes V, Prata MJ, Amorim A, Gusmão L. Population data for Y-chromosome haplotypes defined by 17 STRs (AmpFISTR Yfiler) in Portugal. *Forensic Sci Int* 2007;171:250-5.
 49. Pontes ML, Cainé L, Abrantes D, Lima G, Pinheiro MF. Allele frequencies and population data for 17 Y-STR loci (AmpFISTR Y-filer) in a Northern Portuguese population sample. *Forensic Sci Int* 2007;170:62-7.
 50. Valverde L, Köhneemann S, Rosique M et al. 17 Y-STR haplotype data for a population sample of Residents in the Basque Country. *Forensic Sci Int Genet* In press, 2012.
 51. Roewer L, Willuweit S, Krüger C, et al. Analysis of Y chromosome STR haplotypes in the European part of Russia reveals high diversities but non-significant genetic distances between populations. *Int J Legal Med* 2008;122:219-23.
 52. Veselinovic IS, Zgonjanin DM, Maletin MP, et al. Allele frequencies and population data for 17 Y-chromosome STR loci in a Serbian population sample from Vojvodina province. *Forensic Sci Int* 2008;176:e23-8.
 53. Martín P, García-Hirschfeld J, García O, et al. A Spanish population study of 17 Y-chromosome STR loci. *Forensic Sci Int* 2004;139:231-5.
 54. Sánchez C, Barrot C, Xifró A, et al. Haplotype frequencies of 16 Y-chromosome STR loci in the Barcelona metropolitan area population using Y-Filer kit. *Forensic Sci Int* 2007;172:211-7.
 55. Haas C, Wangenstein T, Giezendanner N, Kratzer A, Bär W. Y-chromosome STR haplotypes in a population sample from Switzerland (Zurich area). *Forensic Sci Int* 2006;158:213-8.

56. Ballard DJ, Phillips C, Thacker CR, Robson C, Revoir AP, Syndercombe Court D. Y chromosome STR haplotypes in three UK populations. *Forensic Sci Int* 2005;152:289-305.
57. Decker AE, Kline MC, Redman JW, Reid TM, Butler JM. Analysis of mutations in father-son pairs with 17 Y-STR loci. *Forensic Sci Int Genet* 2008;2:e31-5.
58. Marino M, Furfuro S. Genetic population data of 12 Y-chromosome STRs loci in Mendoza population (Argentina). *Forensic Sci Int Genet* 2010;4:e89-93.
59. Pereira RW, Monteiro EH, Hirschfeld GC, Wang AY, Grattapaglia D. Haplotype diversity of 17 Y-chromosome STRs in Brazilians. *Forensic Sci Int* 2007;171:226-36.
60. Builes JJ, Bravo ML, Gómez C, et al. Y-chromosome STRs in an Antioquian (Colombia) population sample. *Forensic Sci Int* 2006;164:79-86.
61. Builes JJ, Martínez B, Gómez A, et al. Y chromosome STR haplotypes in the Caribbean city of Cartagena (Colombia). *Forensic Sci Int* 2007;167:62-9.
62. Baeza C, Guzmán R, Tirado M, et al. Population data for 15 Y-chromosome STRs in a population sample from Quito (Ecuador). *Forensic Sci Int* 2007;173:214-9.
63. Matamoros M, Yurrebaso I, Gusmão L, García O. Population data for 12 Y-chromosome STR loci in a sample from Honduras. *Leg Med (Tokyo)* 2009;11:251-5.
64. Luna-Vázquez A, Vilchis-Dorantes G, Aguilar-Ruiz MO, et al. Haplotype frequencies of the PowerPlex Y system in a Mexican-Mestizo population sample from Mexico City. *Forensic Sci Int Genet* 2008;2:e11-3.
65. Borjas L, Bernal LP, Chiurillo MA, et al. Usefulness of 12 Y-STRs for forensic genetics evaluation in two populations from Venezuela. *Leg Med (Tokyo)* 2008;10:107-12.
66. Willuweit S, Caliebe A, Andersen MM, Roewer L. Y-STR Frequency Surveying Method: a critical reappraisal. *Forensic Sci Int Genet* 2011;5:84-90.
67. Holland MM, Parsons TJ. Mitochondrial DNA sequence analysis - validation and use for forensic case work. *Forensic Sci Rev* 1999;11:21-50.
68. Ota T, Kimura M. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet Res* 1973;22(2):201-4.
69. Fan H, Chu JY. A brief review of short tandem repeat mutation. *Genomics Proteomics Bioinformatics* 2007;5(1):7-14.
70. Vermeulen M, Wollstein A, van der Gaag K, et al. Improving global and regional resolution of male lineage differentiation by simple single-copy Y-chromosomal short tandem repeat polymorphisms. *Forensic Sci Int Genet* 2009;3(4):205-13.