ELSEVIER

Legal Medicine 5 (2003) S160–S163

LEGAL MEDICINE

www.elsevier.com/locate/legalmed

# Asian online Y-STR Haplotype Reference Database

Ruediger Lessig[a], Sascha Willuweit[b], Michael Krawczak[c], Fang-Chin Wu[d], Chang-En Pu[d], Wook Kim[e], Lotte Henke[f], Juergen Henke[f], Jasmin Miranda[g], Monika Hidding[g], Mark Benecke[g], Cornelia Schmitt[g], Michelle Magno[h], Gayvelline Calacal[h], Frederick C. Delfin[h], Maria Corazon A. de Ungria[h], Sahar Elias[i], Christa Augustin[j], Zaw Tun[k], Katsuja Honda[k], Manfred Kayser[l], Leonor Gusmao[m], António Amorim[m], Cintia Alves[m], Yiping Hou[n], Christine Keyser[o], Bertrand Ludes[o], Michael Klintschar[p], Uta D. Immel[p], Barbara Reichenpfader[q], Boriana Zaharova[r], Lutz Roewer[b,*]

[a]*Institute of Legal Medicine, University of Leipzig, Johannisallee 28, D-04103 Leipzig, Germany*
[b]*Institute of Legal Medicine, Humboldt-University Berlin, Hannoversche Strasse 6, D-10115 Berlin, Germany*
[c]*Institute of Medical Informatics and Statistics, University of Kiel, Kiel, Germany*
[d]*Scientific and Technical Research Center, Ministry of Justice, Hsin-Tien City, Taipei County, Taiwan, ROC*
[e]*Department of Biology, Dankook University, Cheonan, South Korea*
[f]*Institute of Bloodgroup Research, D-50672 Köln, Germany*
[g]*Institute of Legal Medicine, University of Cologne, Cologne, Germany*
[h]*Natural Sciences Research Institute, University of the Philippines, Quezon City, Philippines*
[i]*Institute of Legal Medicine, University of Damascus, Damascus, Syria*
[j]*Institute of Legal Medicine, University of Hamburg, Hamburg, Germany*
[k]*Department of Legal Medicine, University of Tsukuba, Tsukuba City, Ibaraki, Japan*
[l]*Max-Planck-Institute of Evolutionary Anthropology, Leipzig, Germany*
[m]*Institute of Pathology and Immunology, University of Porto (IPATIMUP), Porto, Portugal*
[n]*Institute of Forensic Medicine, West China University of Medical Sciences, Chengdu, PR China*
[o]*Institute of Legal Medicine, 67085 Strasbourg, France*
[p]*Institute of Legal Medicine, University Halle-Wittenberg, D-06112 Halle, Germany*
[q]*Institute of Legal Medicine, Karl Franzens University Graz, A-8010 Graz, Austria*
[r]*University Hospital of Obstetrics and Gynaecology, 1431 Sofia, Bulgaria*

**Abstract**

For several years Y-chromosomal microsatellites (short tandem repeats, STRs) have been well established in forensic practice. In this context, the genetic characteristics of the Y chromosome (i.e. its paternal inheritance and lack of recombination) render STRs particularly powerful. However, genetic differences between male populations appear to be larger for Y-STRs than for autosomal STRs, a fact that is most likely due to the higher sensitivity of Y-chromosomal lineages to genetic drift (Forensic Sci Int 118 (2001) 153). The assessment of probabilities for matches between haplotyped male persons or traces/persons requires the typing of a large number of haplotypes in the appropriate reference populations. The haplotype data of a large number of European as well as South and North American populations have been collected and are continuously published online (Y-STR Haplotype Reference Database – YHRD; http://www.ystr.org). The most recent multicentric effort has led to the

* Corresponding author. Tel.: +49-30-450525012; fax: +49-30-450525912.
*E-mail address:* roewer@charite.de (L. Roewer).

1344-6223/03/$ - see front matter 
doi:10.1016/S1344-6223(02)00100-1

establishment of an Asian YHRD (http://www.ystr.org/asia) which has been available since January 2002. All databases are maintained and curated at the Institute of Legal Medicine, Humboldt-University, Berlin and will soon be fused to a global repository including populations from all continents.


*Keywords*: Y chromosome; Short tandem repeat; Y-Haplotype Reference Database; Population genetics

## 1. Introduction

Most of the now widely forensically used Y-chromosomal microsatellites (short tandem repeats, STRs) were first published in 1997 [1,2]. Because male profiling is done via haplotyping of a number of linked STRs located at the human Y chromosome a set of loci has been chosen and evaluated which guarantees high discrimination, robust performance for minute amounts of DNA and swift integration in automated typing technologies. Thus, the so called 'minimal core set' of Y-STRs has been developed which has been validated for forensic applications, which includes nine loci – DYS19, DYS390, DYS391, DYS392, DYS393, DYS389I/II and DYS385I/II. These markers represent the most commonly used systems in forensic practice worldwide. The so-called extended core set additionally includes the highly polymorphic biallelic STR YCAII representing a dinucleotide polymorphism. In the last years a number of new polymorphic STRs such as DYS435, DYS439, DYS447, DYS448 and others were described [3,4]. Routine forensic cases show the usefulness of these markers. Especially in paternity cases lacking the putative father or other relatives as well as in rape or identification cases (namely mass disasters) the potential of the Y-STRs is evident. The alleles on the Y chromosome are fully linked building Y-chromosomal haplotypes of an immense variability. Thus, large haplotype reference databases comprising the relevant populations are necessary to get realistic match probabilities. For several years the haplotype data of a large number of European as well as South and North American populations have been collected and published online (Y-STR Haplotype

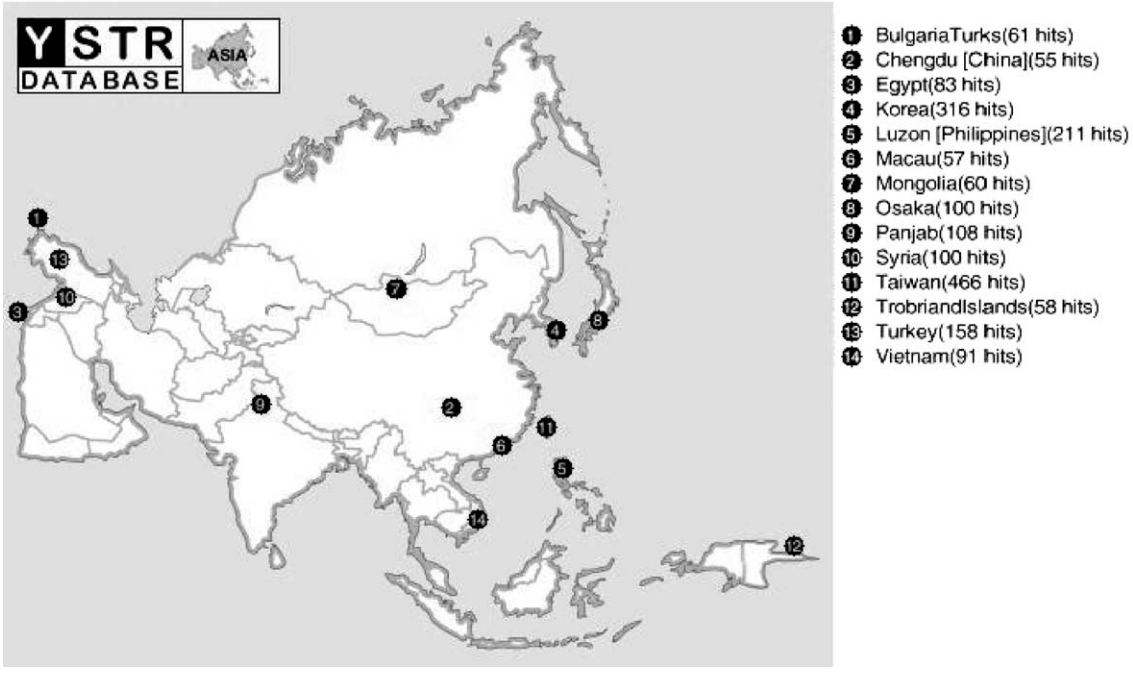


Fig. 1. Map representing the Asian populations logged to the YHRD until September 2002.

Reference Database – YHRD) [5]. The European database – http://www.ystr.org/europe – includes as of September 2002 11,610 minimal haplotypes from 81 populations. A total of 1705 minimal haplotypes of 30 different US populations have been collected in the reference database of the USA – http://www.ystr.org/usa [6]. The USA database has been substructured with separate entries for African-Americans (ten population samples, 599 minimal haplotypes), European-Americans (11 population samples, 628 minimal haplotypes) and Hispanics (nine population samples, 478 haplotypes). The YHRD Asia – http://www.ystr.org/asia – represents 14 populations with currently 1924 minimal and 362 extended haplotypes as of September 2002 (Figs. 1 and 2). So far 95 institutes mainly working in forensic genetics have passed the required quality control exercises and contributed haplotyped Y chromosomes. More than 80,000 users have visited the three repositories during the last 2 years.

## 2. Database

The YHRD databases are curated at the Institute of Legal Medicine, Humboldt-University of Berlin, Germany. They are continuously updated and are publicly available via the World Wide Web. All databases are based on the same core format described above for the minimal haplotype including seven highly polymorphic, forensically evaluated Y-STR markers (DYS19, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS385I/II). These 'minimal haplotypes' can be extended by typing of the duplicated locus YCAII. The haplotype diversity for the minimal haplotype format is 0.9976 in Europe ($n = 11,610$), 0.9996 in Asia ($n = 1924$) and 0.9968 in the USA ($n = 1705$).

The main feature of the databases is the search function enabling the user to receive realistic frequency estimates for single Y-STR alleles, as well as for partial, minimal or extended Y-chromosomal haplotypes. The query results in observed frequency estimates which are given for the whole database as well as for any single population. Maps provide information about the geographical distribution of matched reference haplotypes. For rare haplotypes which are not represented in the database, a frequency extrapolation algorithm ('haplotype

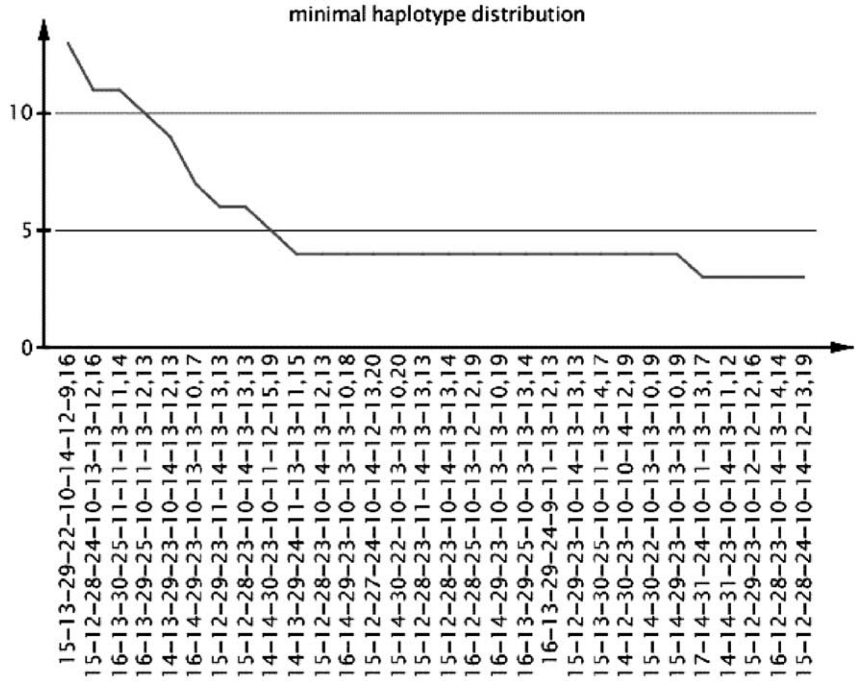


Fig. 2. The 30 most frequent minimal haplotypes in the YHRD Asia (of 1924 haplotypes in 14 populations).

surveying method') has been proposed [7] which is already included as a feature at the European database. The calculation starts by using the button 'Calculate' and provides a posteriori frequency estimates based on an empirical a priori frequency distribution curve. This method is available for calculating haplotype frequencies in all large and sufficiently homogeneous populations. Thus, it is applicable in large parts of Europe but still not for the Asian repository which is yet neither large enough nor sufficiently homogeneous from a population genetic point of view.

The minimal haplotype used to individualize Y chromosomes and the backbone of the YHRD is especially sensitive for population substructuring. In addition to its forensic benefit the database has useful information about the phylogeography of the logged populations as well as about the ethnic origin of unknown Y chromosomes which can be searched against the reference database. The assessment of the inter-population variability of Y-STR haplotypes is based on the AMOVA (analysis of molecular variance) approach. The genetic distance values (Phist) for population pairs, the fraction of different haplotypes as well as the fraction of population-specific haplotypes can be inspected at the site 'Population analysis'.

In addition the YHRD websites provide contact information about the collaborating laboratories (International Forensic Y-User Group – IFYUG). Furthermore, high-throughput multiplex protocols with detailed information about loci, primers, PCR conditions, etc. are available at the site 'Primers & Protocols'.

## 3. Discussion

Two years after the first launch of the YHRD the usefulness of publicly available online repositories of haplotype data becomes evident. The collection is growing fast with weekly updates and reaches at least for the European database the state of a qualified forensic tool. The YHRD has already helped to solve crime, paternity and genealogical casework, and has supported molecular anthropological research.

The worldwide effort to increase the number of haplotyped Y chromosomes will lead to the release of a global database in 2003.

## References

[1] Kayser M, Caglià A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hidding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Weichhold G, de Knijff P, Roewer L. Evaluation of Y-chromosomal STRs: a multicenter study. Int J Leg Med 1997;110:125–133.

[2] de Knijff P, Kayser M, Caglià A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hidding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Weichhold G, Roewer L. Chromosome Y microsatellites: population genetic and evolutionary aspects. Int J Leg Med 1997;110:134–140.

[3] Ayub Q, Mohyuddin A, Qamar R, Mazhar K, Zerjal T, Mehdi SQ, Tyler-Smith C. Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. Nucleic Acids Res 2000;28:e8.

[4] White PS, Tatum OL, Deaven LL, Longmire JL. New, male-specific microsatellite markers from the human Y chromosome. Genomics 1999;57:433–437.

[5] Roewer L, Krawczak M, Willuweit S, Nagy M, Alves C, Amorim A, Anslinger K, Augustin C, Betz A, Bosch E, Cagliá A, Carracedo A, Corach D, Dekairelle AF, Dobosz T, Dupuy BM, Füredi S, Gehrig C, Gusmaõ L, Henke J, Henke L, Hidding M, Hohoff C, Hoste B, Jobling MA, Kärgel HJ, de Knijff P, Lessig R, Liebeherr E, Lorente M, Martínez-Jarreta B, Nievas P, Nowak M, Parson W, Pascali VL, Penacino G, Ploski R, Rolf B, Sala A, Schmidt U, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Kayser M. Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. Forensic Sci Int 2001;118:103–111.

[6] Kayser M, Brauer S, Willuweit S, Schädlich H, Batzer MA, Zawacki J, Prinz M, Roewer L, Stoneking M. Online Y-chromosomal short tandem repeat (STR) haplotype reference database for U.S. populations. J Forensic Sci 2002;47:513–519.

[7] Roewer L, Kayser M, de Knijff P, Anslinger K, Cagliá A, Corach D, Füredi S, Henke L, Hidding M, Kärgel HJ, Lessig R, Nagy M, Pascali VL, Parson W, Rolf B, Schmitt C, Szibor R, Teifel-Greding J, Krawczak M. 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.