Int J Legal Med (1997) 110 : 125–133

ORIGINAL ARTICLE

M. Kayser · A. Caglià · D. Corach · N. Fretwell · C. Gehrig · G. Graziosi · F. Heidorn · S. Herrmann · B. Herzog · M. Hidding · K. Honda · M. Jobling · M. Krawczak · K. Leim · S. Meuser · E. Meyer · W. Oesterreich · A. Pandya · W. Parson · G. Penacino · A. Perez-Lezaun · A. Piccinini · M. Prinz · C. Schmitt · P. M. Schneider · R. Szibor · J. Teifel-Greding · G. Weichhold · P. de Knijff · L. Roewer

Evaluation of Y-chromosomal STRs: a multicenter study

Received: 30 December 1996 / Received in revised form: 26 February 1997

M. Kayser · L. Roewer (✉)
Institut für Gerichtliche Medizin, Hannoversche Strasse 6,
D-10115 Berlin, Germany
FAX: +49 (30) 2093 7240

A. Caglià
Instituto di Medicina Legale,
Università Cattolica del S. Cuore, Largo Francesco Vito 1,
I-00168 Roma, Italy

D. Corach · G. Penacino
Servicio de Huellas Digitales Geneticas,
Facultad de Farmacia y Bioquimica UBA, Junin 956,
1113 Buenos Aires, Argentina

N. Fretwell · M. Jobling
Department of Genetics, University of Leicester,
University Road, Leicester LEI 7RH, UK

C. Gehrig
Institut für Rechtsmedizin, Buehlstrasse 20,
CH-3012 Bern, Switzerland

G. Graziosi
Department of Biology, University of Trieste,
Via Giorgeri 5, I-34127 Trieste, Italy

F. Heidorn
Institut für Rechtsmedizin, Hittorfstrasse 18, D-14195 Berlin,
Germany

S. Herrmann
Landeskriminalamt Berlin, Gothaer Strasse 19, D-10823 Berlin,
Germany

B. Herzog
Institut für Rechtsmedizin, Fürstengraben 2,
D-07743 Jena, Germany

M. Hidding · C. Schmitt
Institut für Rechtsmedizin, Melatengürtel 60–62,
D-50823 Köln, Germany

K. Honda
Department of Legal Medicine,
Osaka University Medical School, 2–2 Yamadaoka, Suita,
Osaka 565, Japan

M. Krawczak
Institute of Medical Genetics, University of Wales, Heath Park,
Cardiff CF4 4XN, UK

K. Leim
Institut für Rechtsmedizin, Vosstrasse 2,
69115 Heidelberg, Germany

S. Meuser · P. M. Schneider
Institut für Rechtsmedizin, Am Pulverturm 3,
D-55131 Mainz, Germany

E. Meyer
Institut für Rechtsmedizin, von-Esmarch-Strasse 86,
D-48149 Münster, Germany

W. Oesterreich
Landeskriminalamt Brandenburg, Prenzlauer Strasse 66–70,
D-16352 Basdorf, Germany

A. Pandya
CRC Chromosome Molecular Biology Group,
Department of Biochemistry, University of Oxford,
South Parks Road, Oxford, OX1 3QU, UK

W. Parson
Institut für Gerichtliche Medizin, Müllerstrasse 44,
A-6020 Innsbruck, Austria

A. Perez-Lezaun
Department Anthropologia, Facultat Biologia,
Universitat de Barcelona, Diagonal 645, Barcelona 08028, Spain

A. Piccinini
Instituto di Medicina Legale, Università Milano,
Via Mangiagalli 37, I-20133 Milano, Italy

M. Prinz
Department of Forensic Biology, OCME, 520 First Avenue,
NY 10016 New York, USA

R. Szibor
Institut für Rechtsmedizin, Leipziger Strasse 44,
D-39112 Magdeburg, Germany

J. Teifel-Greding
Bayerisches Landeskriminalamt, Maillingerstrasse 15,
D-80636 München, Germany

G. Weichhold
Institut für Rechtsmedizin, Frauenlobstrasse 7a,
D-80337 München, Germany

P. de Knijff
MGC Department of Human Genetics, Leiden University,
P.O.Box 9503, 2300 RA Leiden, The Netherlands

Abstract A multicenter study has been carried out to characterize 13 polymorphic short tandem repeat (STR) systems located on the male specific part of the human Y chromosome (DYS19, DYS288, DYS385, DYS388, DYS389I/II, DYS390, DYS391, DYS392, DYS393, YCAI, YCAII, YCAIII, DXYS156Y). Amplification parameters and electrophoresis protocols including multiplex approaches were compiled. The typing of non-recombining Y loci with uniparental inheritance requires special attention to population substructuring due to prevalent male lineages. To assess the extent of these subheterogeneities up to 3825 unrelated males were typed in up to 48 population samples for the respective loci. A consistent repeat based nomenclature for most of the loci has been introduced. Moreover we have estimated the average mutation rate for DYS19 in 626 confirmed father-son pairs as 3.2×10^{-3} (95% confidence interval limits of 0.00041–0.00677), a value which can also be expected for other Y-STR loci with similar repeat structure. Recommendations are given for the forensic application of a basic set of 7 STRs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393) for standard Y-haplotyping in forensic and paternity casework. We recommend further the inclusion of the highly polymorphic bilocal Y-STRs DYS385, YCAII, YCAIII for a nearly complete individualisation of almost any given unrelated male individual. Together, these results suggest that Y-STR loci are useful markers to identify males and male lineages in forensic practice.

Key words Short tandem repeats · Y chromosome · Population analysis · Mutation rate · Y haplotype analysis

Introduction

Currently, the analysis of the male-specific portion of total human DNA is limited to the gender identification of a given DNA sample. Various techniques have been employed for this purpose but most of the routine sex determination tests are based on the rapid and sensitive PCR amplification of the X-Y homologous gene amelogenin [1]. This and other forensic tests do not make use of the discriminatory potential of the male-specific part of the Y-chromosome which contributes to the overall variability of the human genome. The major part of the human Y chromosome (i.e. the long arm) consists of polymorphic sequences which are organized into large interspersed tandemly repeated arrays [2]. RFLP analysis of these and other hypervariable sequences that are several thousands of basepairs long has been used to characterize individual human Y chromosomes [3, 4]. In addition a number of point mutations occurring in low copy number sequences have been employed for male differentiation [5, 6]. Due to the lack of sensitivity, technical and statistical problems none of these methods for male identification have been introduced in forensic practice. Forensic identification techniques in the 1990s are dominated by the development and rapid establishment of PCR-amplifiable short tandem repeats (STRs) or microsatellites. Whereas a number of autosomal STRs have been rapidly evaluated for forensic use and are successfully used in routine casework [7, 8], the development of Y-chromosomal-linked STRs – the systems of choice for male identification – remained stagnant. One hypothetical reason could be that STR loci appear to occur less frequently on the Y chromosome compared with autosomes: following the study of Spurdle and Jenkins the ratio is 1:4 [9]. Only three dimeric (YCAI, YCAII, YCAIII) [10], one tetrameric (27H39 or DYS19) [11] and one pentameric Y-STR [12] have been described in more detail in the first half of the decade. From these loci only the system DYS19 has been tested for forensic applicability [13] and introduced to a number of laboratories for routine forensic applications [14, 15]. Extensive worldwide population studies for DYS19 [16–18] revealed marked differences in allele distributions for major ethnic groups (i.e. Amerindians, Asians, Europeans, Africans). Distinct DYS19 alleles predominated in particular populations, whereas similarities in frequency patterns were observed in geographically neighbouring populations with non-restricted gene flow. This observation reflects an inherent property of uniparentally inherited DNA sequences, such as the non-pseudoautosomal proportion of the Y chromosome. These sequences do not recombine during meiosis, with the result that Y chromosomes are transferred unchanged from generation to generation establishing paternal lineages. Populations and subgroups can be dominated by male founder lineages over a long time-span which – as long as they exist – can only be modified by mutational events. This mode of inheritance of Y-linked alleles has to be taken into account when statistical interpretation of Y-STR data has to be done. The establishment of well defined local databases is therefore an essential for using Y-STRs in practice. As has been shown previously for DYS19 [11], other characteristics of Y-STRs are expected to be similar to autosomal systems: allele numbers correlate with the number of repeats [19] and the allele distribution is essentially unimodal as expected for regular repeats evolving according to the single step mutation model (SMM) [20]. In preliminary studies no slippage mutation was found for DYS19 in 129 meioses [11, 21].

The ability to detect and discriminate male DNA makes Y chromosomal STR systems an advantageous addition to well-established autosomal PCR-based systems. Y-STRs will increase the success rate for identifying the male component in male/female cell mixtures in body fluids where differential lysis was unsuccessful or too risky, for example for highly degraded samples or very low sperm counts. Other possible applications include the detection of male epithelial cells in ejaculates of vasectomized individuals, or the determination of the number of semen contributors in multiple rape cases. By total omission of the time-consuming and sometimes ineffective differential lysis step, it can also be used as a rapid screening tool for excluding suspects. Only cases with a non-exclusion will have to be processed further.

In addition, Y-specific STRs can be applied for parentage testing to identify patrilineages in deficiency cases with deceased putative fathers.

Recently, primers for a series of highly polymorphic Y-specific STRs have been developed and tested [22]. Based on worldwide population data which have been collected on the occasion of the 1st Forensic Y-User Workshop held in Berlin 1996 we report here population and family data for most of the known informative Y-STRs. The objective was to establish locus-specific databases for routine forensic application. For this purpose a consistent allele nomenclature following the recommendations of the ISFH was introduced for most of the systems. We have further collected family data to give a solid estimate for Y-STR mutation rates. A list of PCR single locus and multiplex protocols will be given and Y-STR standard and extended haplotype formats for practical use in stain and paternity analysis will be recommended.

Materials and methods

Population samples

For the population survey 3825 male DNA samples from 48 different subpopulation groups from Europe, America, Asia, Africa, and Oceania were analysed for one or several loci (Appendix, Tables 1, 2). Because of patrilinear inheritance of Y-markers samples from similar geographical regions have not been pooled. DNA was extracted according to published procedures. Males were analysed at the following tetranucleotide STR loci: DYS19, DYS385, DYS389I/II, DYS390, DYS391, DYS393, at the trinucleotide loci DYS388, DYS392, at the dinucleotide loci DYS288, YCAI, YCAII, YCAIII and at the pentanucleotide locus DXYS156Y. Different amplification, cycling and electrophoresis protocols have been used to analyse the Y-STRs. Silver staining procedures were done according to published protocols [21] or to the German DNA Profiling Group (GEDNAP) standard protocols, respectively. Protocols for analyses with frequently used ALF or ALF express (Pharmacia) and 373A automated sequencers (Applied Biosystems Division/Perkin Elmer) are given in the following:

Amplification conditions

For DYS19, DYS389 I/II, DYS390, DY391, DYS392, DY393, DYS385, DYS288, DYS388, YCAI, YCAII, YCAIII, DXYS156Y 1 U Taq, 0.1–0.8 µM each primer, 200 µM dNTPs, 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM TrisHCl pH 8.3 were used in a total volume of 25 µl. The forward primers were labeled with the appropriate dyes recommended by the suppliers of the automated sequencers used in this study.

Cycling conditions (PTC-100, MJ Research, Inc.)

DYS19: 94° C – 3 min soak, 94° C – 30 s, 51° C – 30 s, 72° C – 90 s, 30 cycles
DYS389 I/II, DYS390, DY391, DYS392: 94° C – 3 min soak, 94° C – 15 s, 58° C – 20 s, 72° C – 20 s, 5 cycles; 94° C – 15 s, 54° C – 20 s, 72° C – 20 s, 30 cycles
DYS393, DYS288, DXYS156Y: 94° C – 3 min soak, 94° C – 15 s, 58° C – 20 s, 72° C – 20 s, 30 cycles
DYS388, DYS385, YCAI, YCAIII: 94° C – 3 min soak, 94° C – 60 s, 55° C – 60 s, 72° C – 90 s, 30 cycles
YCAII: 94° C- 3 min soak, 94° C – 60 s, 45° C – 60 s, 72° C – 90 s, 30 cycles

Electrophoresis conditions (ALF automated sequencer)

Of the amplified PCR products 1 µl (up to 1:10 diluted) was loaded mixed with 4 µl 1% dextrane blue/formamide loading dye and appropriate amounts of both internal length standards on denaturing Hydrolink Long Ranger 6% polyacrylamide/bisacrylamide gels (AT Biochem) and run in 1 × TBE (0.09 M Tris, 0.09 M Boric acid, 0.001 M EDTA) as gel and electrophoresis buffer with 1000 V, 40 mA, 40 W. PCR fragments were automatically analysed by the Fragment Manager Version 1.2 software (Pharmacia).

Recommended multiplex reactions
for the ALF automated sequencer (non-overlapping systems)

(1) Triplex I (ALF): DYS19
DYS389 I
DYS389 II

(2) Triplex II (ALF): DYS390
DYS391
DYS393

Amplification conditions (Triplex I and II, ALF)

See single locus PCR systems.

Cycling conditions (Triplex I and II, ALF)

94° C – 3 min soak, 94° C – 15 s, 58° C – 20 s, 72° C – 20 s, 5 cycles; 94° C – 15 s, 54° C – 20 s, 72° C – 20 s, 30 cycles (PTC-100, MJ Research, Inc.).

Electrophoresis conditions (Triplex I and II, ALF)

See single locus electrophoresis systems.

Recommended multiplex reactions for ABI 373A
automated sequencer (partially overlapping systems) [23]

(1) Triplex I (373A):	DYS391:	Primer 1 – FAM labeled
	DYS392:	Primer 1 – JOE labeled
	DYS393:	Primer 1 – FAM labeled
(2) Quadruplex I (373A):	DYS19:	Primer 1 – JOE labeled
	DYS390:	Primer 1 – FAM labeled
	DY389 I/II:	Primer 1 – JOE labeled

Amplification conditions (Triplex I, 373A)

2 U Amplitaq Gold (Perkin Elmer), 0.3 µM each primer DY391, DYS392, 0.06 µM each primer DYS393, 200 µM dNTPs, 2 mM $MgCl_2$, 50 mM KCl, 10 mM TrisHCl pH8.3, 0.01% gelatin

Cycling conditions (Triplex I, 373A)

95° C – 10 min soak, 94° C – 1 min, 55° C – 1 min, 72° C – 2 min, 29 cycles (Thermalcycler 480, Perkin Elmer).

Electrophoresis conditions (Triplex I, 373A)

6% polyacrylamide/bisacrylamide (19:1), 8 M urea, 1 × TBE buffer (0.09 M Tris, 0.09 M boric acid, 0.001 M EDTA) as gel and electrophoresis buffer, separation distance 24 cm (well-to-read), 2500V, 40 mA, 30 W, laser power 40 mW, on a 373A automated sequencer (Applied Biosystems Division/Perkin Elmer). Ampli-

fied samples (2 µl) were loaded mixed with 3 µl formamide, 1.5 µl loading buffer and 0.5 µl internal standard GS500.

Amplification conditions (Quadruplex I, 373A)

2 U AmpliTaqGold (Perkin Elmer), 0.24 µM each primer DYS19, DYS390, 0.12 µM each primer DYS389, 200 µM dNTPs, 2 mM $MgCl_2$, 50 mM KCl, 10 mM TrisHCl, pH 8.3, 0.01% gelatin.

Cycling conditions (Quadruplex I, 373A)

95° C – 10 min soak, 94° C – 1 min, 55° C – 1 min, 72° C – 2 min, 29 cycles (Thermalcycler 480 Perkin Elmer).

Electrophoresis conditions (Quadruplex I, 373A)

6% polyacrylamide/bisacrylamide (19:1), 7 M urea, 1 × TBE buffer (0.09 M Tris, 0.09 M boric acid, 0.001 M EDTA) as gel and electrophoresis buffer, separation distance 24 cm (well-to-read), 2500 V, 40 mA, 30 W, laser power 20 mW, on a 373 A automated sequencer (Applied Biosystems Division/Perkin Elmer). Amplified samples (2 µl) were loaded mixed with 3 µl formamide, 1.5 µl loading buffer and 0.5 µl internal standard GS500.

The use of Amplitaq Gold (Perkin Elmer) instead of regular Taq polymerase was crucial for successful ABI 373A multiplex analysis. Presently Triplex I (373A) does not display any overlapping alleles, but since the distance between the longest known DYS392 allele and the shortest known DYS391 allele is only 12 bp, the occurrence of additional alleles might cause an identification problem. Therefore both loci were labeled using different colours. The primers for locus DYS391 produce one or two constant fragments of 260 bp and 264 bp length for high amounts (> 10 ng) of female DNA, which does not interfere with the interpretation of the results.

Nomenclature

According to the recommendations of the International Society of Forensic Haemogenetics [24] the number of variable repeats has been used to designate the various Y-STR alleles. Consistent allele designation for all loci was ensured by the distribution of sequenced allelic ladders or typed standard DNA to all laboratories involved in this study. The allelic ladders for DYS19, DYS389I/II, DYS390, DYS391, DYS392, DYS393 were produced and kindly distributed by P. de Knijff, Leiden University (manuscript in preparation). The sequence information for DYS385 alleles was provided by S. Meuser and P.M. Schneider, Mainz (manuscript in preparation), and for YCAIII by C. Tyler-Smith, Oxford. For those Y-STRs (YCAI, YCA II, DXYS156Y, DYS288, DYS388) where alleles have not been fully sequenced either the allele lengths were used to define alleles or a simple numerical nomenclature was used with the shortest observed allele defined as allele 1. Typed standard DNA for all loci are available from the Institut für Gerichtliche Medizin, Humboldt-Universität Berlin.

Statistical calculations

Gene diversity

As a measure for the Y-STR locus informativeness the gene diversity [25] given as

$$P(Y) = \Sigma (P_i)^2$$

was used. For Y-linked systems this value is identical with frequently used measures as individualisation potential PI or discrimination index DI [26]. The mean exclusion chance or power of exclusion [27], a value which is used in paternity calculations, is for Y-inherited polymorphic systems identical with the gene diversity.

Mutation rate

For a given mutation rate µ, the number of mutations X observed among M independent meioses follows a binomial distribution of parameter µ and M given as

$$P(X = x) = \binom{M}{x} \mu^x (1-\mu)^{M-x}$$

The best estimator of µ is the value that maximises P: = X/M. The 95% interval is given by the two values of µ so that P = 2.5%. These two values have been estimated by simulation.

Results and discussion

To establish an efficient male identification system, STR loci located on the Y chromosome have been characterized in terms of molecular, population and family genetics and tested for forensic application. For further investigations 13 polymorphic systems located in the non-pseudoautosomal region of the Y chromosome were selected. The repeat sequences of DYS384, DYS394 and DYS395 deposited in the Genome Data Bank (GDB) have been excluded from this survey for the following reasons: DYS384 (GDB-ID: G00-316-229) is not Y-specific, PCR-primers amplify both in males and in females, DYS395 (GDB-ID: G00-456-772) is synonymous with DYS393, whereas DYS394 (GDB-ID: G00-456-738) describes a short largely invariant TA repeat within the DYS19 sequence. Other sequences published in the GDB or elsewhere [10–12, 22] which have been found to be Y-specific and male-specific are DYS19, DYS288, DYS385, DYS388, DYS389I/II, DYS390, DYS391, DYS392, DYS393, YCAI, YCAII, YCAIII, DXYS156Y. Characteristics of these systems are shown in the Appendix, Table 1. The locus DXYS156 is not strictly male-specific since it is located both on the Y and on the X chromosome. Because of unambiguous expression of X-specific (X1-X5) and Y-specific length variants (Y6-Y8) the male-specific sublocus DXYS156Y was chosen for further studies. The loci DYS385, YCAI, YCAII and YCAIII show two male-specific PCR products after amplification. Most probably the repeated sequences are duplicated on the Y chromosome with identical flanking sites allowing duplex amplification of length variable alleles from two independent loci. Because of overlapping sizes of the length variants the origin of the alleles from either of the two loci cannot be unambiguously determined which contrasts previous attempts to separate a locus A and B for YCAI-III [10]. Instead, we have designated the allelic pairs at the respective loci as classes of alleles. Thus, for bilocal Y-STRs the frequencies in a population refer to allele classes and not to single alleles (Appendix, Tables 1, 2i-l). Another feature has been observed at the locus DYS389. As reported elsewhere [28, 29] two products of different sizes can be amplified by the primers described in the GDB. Sequence analysis shows that the priming site of the forward primer has been duplicated. Thus, the larger (DYS389II) product includes 3 CTGT/CTAT repeat stretches, whilst the smaller (DYS389I) product includes just two. The difference of the length ranges of the alleles

generated at locus DYS389 is around 100 bp allowing an unambiguous assignment of the alleles to either of the two systems DYS389I and DYS389II. In contradiction to other reports [29] a correlation of repeat lengths between the DYS389I and DYS389II alleles has not been observed in all individuals. Thus, both systems have been included and studied in this survey.

All other loci are monolocal and share common features with autosomal STRs. The allele distribution in the population is unimodal and symmetric with a local maximum for alleles of medium size. An exception is DYS392 which was found to be bimodal, showing two maxima in several population samples. The highly polymorphic bilocal systems show complex allele distributions with a high number of rare alleles.

Alleles were found in the short size range below 150 bp (DYS288, DYS388, DYS393, YCAI), in a medium size range between 150 and 260 bp (DYS19, DYS389I, DYS390, DYS392, YCAII, YCAIII, DXYS156Y) and in the higher size range above 260 bp (DYS389II, DYS385, DYS391).

Population analysis for 13 Y-STRs

Allele frequencies were calculated for 13 Y-chromosomal STR loci in up to 48 populations from Europe, America, Asia, Africa and Oceania in up to 3825 male individuals which are known to be unrelated. Assignement of males to a population sample means belonging to a defined geographical region, otherwise the sample is designated as "pooled". Several alleles have been found for each of the loci and the gene diversities are low (DYS288, DYS388, DXYS156Y, YCAI), moderate (DYS391, DYS392, DYS393), high (DYS19, DYS390, DYS389I/II, YCAII) or very high (DYS385, YCAIII) (Appendix, Table 1). The most common alleles differ between major population groups, with a generally observed shift of the allele lengths to higher sizes in Asia, whereas more shorter alleles have been observed in African and American individuals. Substructuring within apparently homogeneous populations (e.g. Germans, Italians) which is to be expected for non-recombining Y-linked markers has been observed for different regional samples (Appendix, Tables 2 and de Knijff et al. this issue). In indigenous populations a striking restriction of the Y-chromosome gene pool can be observed (Appendix, Tables 2). This is most probably caused by isolation and traditional mating patterns [30, and de Knijff et al. this issue]. Different types of aberrant alleles have been seen at the Y-STR loci surveyed including sequence variants (DYS389), intermediate sized alleles (DYS19), "null alleles" (DYS19), duplication and triplication of alleles (DYS19, DYS390, DYS385). Very rarely observed alleles with intermediate sizes (2 bp longer or shorter than canonical alleles) arise most probably by variation in the largely invariant TA stretch (DYS394) located within the amplified DYS19 sequence. The more frequent occurrence of both "null alleles" and multiplicated alleles at Y-chromosomal STRs [see also 17] compared with autosomal loci is probably caused by the particular Y chromosome structure. Cytogenetic analyses show that the heterochromatic region on the long arm of the Y chromosome displays an increased length fluctuation which can be explained by the increased frequency of sister chromatid exchange events for repetitive DNA and the lack of chromosome pairing during meiosis [31]. This could also affect the deletion and multiplication of STR loci. The loci YCAI-III and DYS385 consists of two copies of the STRs including the flanking region and DYS389 has a duplicated 5′-primer site. Santos et al. [17] reported a father/son pair with three DYS19 alleles. In this study we have seen two individuals with biallelic patterns for DYS19, one biallelic individual for DYS390 and one triplication of alleles for DYS385.

Because of the restricted individualisation potential the Y-STRs DYS288, DYS388, YCAI and DXYS156Y will not be recommended for forensic routine application.

Highly variable bilocal Y-STRs DYS385, YCAII, YCAIII

Four bilocal Y-linked STRs have been investigated, the dinucleotide systems YCAI, YCAII, YCAIII [10] and and the tetranucleotide system DYS385. Of these YCAI was found to be largely invariant (Appendix, Tables 1, 2) and will therefore not be recommended for forensic application. However, systems DYS385, YCAIII and YCAII show very high discriminatory potentials (Appendix, Table 1). These STRs consist most probably of two homologous loci on the Y chromosome with different repeat lengths enhancing the variability significantly. In contrast to DYS389 the alleles can not be assigned to either of the two loci because of overlapping size ranges. Nevertheless, a designation as allelic pairs or allele classes seems to be appropriate for practical use. For the still incompletely sequenced YCAI and YCAII alleles a simple preliminary nomenclature was used, with the shortest allele found for each locus designated allele 1.

The size range for the YCAII alleles is between 147–165 bp allowing efficient forensic analysis of degraded DNA samples. The system DYS385 is expressed in the higher molecular weight range (360–412 bp) which makes analysis of low quality DNA samples sometimes more problematic. Advantageous is the tetrameric structure of DYS385 which allows unambiguous classification of the alleles. For YCAIII and YCAI a number of cross-amplified non-Y specific bands have been observed which render an exact interpretation sometimes more difficult.

Mutation rates

For all 13 STRs, with the exception of DYS288, DYS388 and DXYS156Y, family studies have been carried out to check holandric transmission of the alleles (Table 1). For DYS19 446 confirmed father/son pairs of Caucasian ancestry have been analysed in this study and 180 published meioses [14, 21] have been added to calculate a mutation

Table 1 Average mutation rates for Y-STRs

	DYS19	DYS390	DYS391	DYS392	DYS393	DYS385	DYS389 I	DYS389 II	YCA II	YCA I	YCA III
Sample	*n* = 626	*n* = 94	*n* = 41	*n* = 42	*n* = 42	*n* = 104	*n* = 55	*n* = 53	*n* = 113	*n* = 72	*n* = 42
Mutations	2	1	0	0	0	1	0	1	0	0	0
Mutation rate	0.0032	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.	n.c.

n.c. not calculated because of the low sample size

Table 2 Y-STR haplotype analysis

	Sample	No. of different haplotypes	Discrimination capacity (%)
Yh 1	70 Germans (Berlin 1)	63	90.0
Yh 2		68	97.1
Yh 3		69	98.6
Yh 4		70	100.0
Yh 1	54 Swiss (Bern)	51	79.7
Yh 1	100 Italians (Roma)	79	79.0
Yh 1	63 Trobriands	40	63.5
Yh 2, Yh 3		49	77.8
Yh 1	88 Dutch (Leiden)	65	73.9
Yh 1	30 Basques	17	56.7
Yh 1	36 Chinese	34	94.4
Yh 1	40 Mongolians	29	72.5
Yh 1	62 Inuit	22	35.5
Yh 1	54 Surinams	47	87.0
Yh 1	31 Pygmees	18	58.1
Yh 1	100 Buenos Aires, Caucasians	77	77.0
Yh2		91	91.0
Yh3		94	94.0
Yh1, Yh2	16 Mapuches	15	93.7
Yh3		16	100.0
Yh1, Yh2, Yh3	12 Tehuelches	11	91.7
Yh1, Yh2, Yh3	6 Wichis	4	66.7
Yh5	81 New York African Americans	48	59.2
Yh5	88 New York Hispanics	50	56.8
Yh5	25 New York Asians	17	68.0

Yh 1: DYS19 + DYS389 I/II + DYS390 + DX391 + DYS392 + DYS393
Yh 2: Yh1 + DYS385
Yh 3: Yh2 + YCA II
Yh 4: Yh2 + YCA III
Yh 5: DYS19 + DYS389 I/II, DYS390
Haplotypes have been analysed in the laboratories listed in the legend to Appendix, Tables 2

rate for DYS19 on the basis of 626 meioses. All alleles were found to be inherited regularly with the exception of two slippage mutation events for DYS19, one slippage mutation for DYS385 and two slippage mutations occurring both at DYS389II and DYS390 in the same father/son pair whose paternity is beyond any reasonable doubt after extensive testing. We have calculated an average mutation rate for DYS19 of 0.0032, with 95% confidence interval limits of 0.00041–0.0067. Our mutation rate calculation for DYS19 is close to the estimate of Weber & Wong of 0.0021 for a set of chromosome 19 tetranucleotide STRs [32]. For the other Y-STR loci included in this study the number of meioses counted is too low to calculate solid mutation rates. Nevertheless, from our data, 41–113 meioses for 10 loci, as well as from the 248 meioses surveyed by Heyer et al. [33], the mutation rates for the tetra- and trimeric Y-STRs DYS19, DYS389I/II, DYS390, DYS391, DYS392 and DYS393 will be expected to lie in the range found for DYS19. Aberrant allele patterns such as duplicated alleles at DYS19, DYS390 and triplicated alleles at DYS385 have been found to be inherited from father to son and were not regarded as mutations.

Our investigations suggest that mutation rates for Y chromosomal STRs are in the range of those for autoso-

mal STRs. Even the rates for the highly polymorphic systems DYS385 and YCAII are not significantly increased (this study and [33]). This finding again confirms the idea that slippage creates mutations at STR sites and argues against any recombination-related mechanism.

Y-STR haplotype analysis

Because non-pseudoautosomal Y loci are genetically linked, Y-STR single locus typing can be expanded to a much more discriminative Y haplotype analysis. Different sets of STRs have been used to construct highly informative Y chromosome haplotypes [22, 29]. We now demonstrate that a set of seven monolocal Y-chromosomal STRs called Y haplotype 1 (Yh1) consisting of the Y-STRs DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 can discriminate between most of the male individuals in a population (discrimination capacity between 74% and 90% in local European populations) and 63% even in a native population on the isolated Trobriand islands in the Western Pacific (Table 2). As demonstrated in this study two multiplex reactions (Triplex I, Quadruplex I run on a ABI 373A automated sequencer) can be used to perform a 7-locus Y-haplotyping. Three PCRs are necessary for an ALF-based haplotype analysis, two triplexes and a single locus DYS392 analysis. For forensic purposes, it would be preferable if even more information could be obtained from a Y-chromosomal haplotype. Nearly complete individualisation for a male or patrilineage can be obtained by further inclusion of the systems DYS385, YCAII and YCAIII (Table 2). This high discrimination capacity could potentially allow us to assign one haplotype to one male (and to all his patrilinear relatives) and would thus circumvent the need for a calculation based on the product rule. Any statistical calculation of a non-exclusion is problematic because of the prevalence of patrilineages in a population which means that some haplotypes are consequently overrepresented. Ideally the forensic Y haplotype analysis must therefore include as many polymorphic loci as possible to raise the exclusion chance to nearly 100% leaving only patrilinear male relatives undiscriminated.

Forensic application – stain analysis

Polymorphic Y-STR systems have the advantage that they provide additional information besides the proof of the presence of male DNA in a forensic stain. In particular for rape and sexual assault cases the Y-STR analysis offers a new approach. Since contrary to autosomal systems, even high amounts of female DNA do not inhibit the amplification of the Y-linked alleles [23], Y-STRs allow the generation of a specific DNA profile of the male offender even after omission of the time-consuming differential lysis step. This approach is not only faster, but also minimizes the loss of sperm DNA, reduces the risk of contamination, and allows a more unambiguous determination of the number of semen contributors. With the inclusion and multiplexing of several loci (7-locus haplotyping + 3 bilocal systems) the degree of discrimination will easily be sufficient to replace the autosomal STRs for sexual offender databases. Nevertheless, single locus Y-STR analysis is advisable for sensitive stain analysis. A combination of a Y-STR with an autosomal STR could be necessary to exclude false negative (female) results. Different systems have been proposed to be multiplexed with DYS19, eg. D12S66, D12S67 [11, 21].

To investigate the applicability of Y-STRs in stain analysis, 48 vaginal swabs which were collected from rape cases between 14 and 32 years ago and had been stored dry at room temperature, have been analysed without separating sperm from female cells in 5 Y-STR systems (DYS19, DYS390, YCAII, DYS389I/II). DYS390 was the most sensitive and alleles were determined in almost all (97.9%) of the samples. DYS19 and YCAII loci were succesfully typed in 95.8% and 87.5% of the samples, respectively. However, amplification of DYS389I/II was the most difficult and successful only in 54.1% of the samples [K. Honda, personal communication]. Stain analysis results show that at least for some Y-STR systems an optimization of primer sequences and PCR conditions will be necessary.

A series of experiments was carried out for mixtures of DNA from two males and from male and female individuals for DYS19 alone and for the Y-STR quadruplex I (373A) consisting of DYS19, DYS390, DYS389I/II. Results were compared to the autosomal STR system VWA. While for the male/male mixtures and the autosomal system an unambiguous detection of the minor component was only possible up to ratios of 1:10 and 1:50 respectively, in male/female DNA mixtures the Y-STR alleles could be identified for the highest ratio tested, 1:2000, or 400 pg male DNA in 800 ng female DNA [23].

Forensic application – paternity analysis

Due to their inheritance along male lineages, Y-STR polymorphisms have a considerable potential for paternity testing, especially in deficiency cases [34]. For Y-STRs the average probability of exclusion in paternity testing and the probability of discrimination in criminal identification are identical. Thus, Y-linked polymorphic traits have a much higher exclusion power than equally variable autosomal loci (compare gene diversity values, Appendix, Table 1). Y-STRs have been included for paternity testing by different laboratories. At the Institut für Gerichtliche Medizin, Berlin in 39 exclusion cases with male offspring, 31 putative fathers were found to be excluded with DYS19, a rate of 79%. Similar numbers have been reported from Munich with 28 out of 35 exclusions confirmed with DYS19 (80%) [35] and from Verona with 19 out of 25 exclusions found with DYS19 (76%) [14]. These rates are in the expected range (Appendix, Table 1).

Table 3 Deficiency case analysed by 10 Y-STRs

	Son? of the deceased alleged father	Nephew? of the deceased alleged father
DYS19	14	14
DYS385	11–14	11–14
DYS393	12	12
YCAII	2–6	2–6
YCAIII	23–23	23–23
DYS389II	16 (26)*	16 (25)*
DYS389I	10	9
DYS390	25	24
DYS391	11	10
DYS392	13	15

Exclusion constellations are underlined
* It has to be taken into account, that when both DYS389I and DYS389II differ by a repeat unit change in the same direction, this has to be counted as a difference only in DYS389I. Because the DYS389I product is included in the DYS389II product a substraction is necessary which avoids the incorrect double counting of a single exclusion

The use of Y-STRs is of course limited to the 50% of cases in which the child being tested is male. There is, however one situation in which Y-linked polymorphisms may be uniquely suited for paternity testing, namely in deficiency cases, when the alleged father is deceased and relatives have to be tested in his place. Any male relative in patrilineage with the deceased alleged father will provide complete information on his Y-chromosomal molecular constitution, permitting a stringent test for paternity for male children. The application for deficiency cases has already been reported [34–36]. The necessity to use a large panel of Y-STR systems to achieve a correct result is demonstrated by a deficiency case with only a nephew of the deceased father available for testing (Table 3). In six Y-STRs both the nephew and the possible child of the deceased putative father carry the same allele or allele pair, in four other systems they show exclusion constellations. The relative similarity of genotypes with identity in six systems and mostly one-step repeat differences in four other systems suggests a (historical) relatedness of the alleged father's and the true father's patrilineage.

Tracing back patrilineages into the past depends on the mutation rate of the applied Y markers. With an assumed mutation rate given in this study a single Y-STR allele will survive unmodified for about 1000 generations with a 95% probability. This timespan is long enough to allow meaningful patrilineage comparisons in a historical context [37].

In conclusion the main forensic applications for Y-STRs are (1) the identification of male DNA in rape cases with male/female stain mixtures and (2) paternity analysis especially for deficiency cases with a male offspring. For these applications the proposed Y-STR haploytyping set of seven systems (Yh1) plus at least three other highly polymorphic systems (DYS385, YCAII, YCAIII) provides a reproducible and sensitive method which should be added to the analytic repertoire of forensic and paternity laboratories. For the interpretation of Y-STR typing results the specific inheritance mode of the Y chromosome has to be considered. Human Y-linked polymorphisms (in the non-pseudoautosomal portion of the chromosome) are haploid, do not undergo recombination, and thus establish patrilineages. In addition, the number of Y chromosomes in any population is one-quarter the number of autosomes, rendering Y-linked sequences more prone to genetic drift. Moreover, the population dynamics of the male carriers of the Y-chromosome is different from females, increasing the predisposition to genetic drift [38]. Thus, special requirements for the establishment of population databases have to be formulated. Only extended regional population data should be used to calculate allele frequencies. If possible all nine loci mentioned above should be tested to reach an exclusion constellation. Non-exclusions remain unclear since all (actual and historical) patrilinear relatives carry the same Y markers. A collaborative study is currently under way to collect more Y haplotype data (Yh1-type). Those haplotype frequencies should reflect the distribution of male lineages in a population much more realistically than single Y-STR frequency data.

Acknowledgements We are grateful to all donors for providing their blood and E. Hagelberg, C. Tyler-Smith, W. Schiefenhövel and J. B. Clegg for providing us with DNA samples. M.K. and L.R. are supported by the DFG (Ro 1040/2–1 and Ro 1040/2–2), M. Krawczak by the DFG (Kr 1093/5–1), A.P.-L. by a DGCYT grant (Spain) PB95–0267-C02-01 to J. Bertrantpetit, A.P. by the BBSRC, N.F. by an MRC-HGMP studentship and M.A.J. by a Wellcome Trust Career Development Fellowship. The authors wish to thank Prof. Dr. G. Geserick and the staff of the Institut für Gerichtliche Medizin, Humboldt-Universität Berlin for supporting the 1st Forensic Y-User Workshop from April 19–20, 1996 in Berlin. We thank Chris Tyler-Smith for helpful comments on the manuscript.

References

1. Mannucci A, Sullivan KM, Ivanov PL, Gill P (1994) Forensic application of a rapid and quantitative DNA sex test by amplification of the X-Y homologous gene amelogenin. Int J Legal Med 106:190–193
2. Cooke H (1976) Repeated sequences specific to human males. Nature 262:182–186
3. Tyler-Smith C (1987) Structure of repeated sequences in the centromeric region of the human Y chromosome. Dev Suppl 101:93–100
4. Tyler-Smith C, Brown WRA (1987) Structure of the major block of alphoid satellite DNA on the human Y chromosome. J Mol Biol 195:457–470
5. Ngo KY, Vergnaud G, Johnsson C, Lucotte G, Weissenbach J (1986) A DNA probe detecting multiple haplotypes of the human Y chromosome. Am J Hum Genet 38:407–418
6. Jakubiczka S, Arnemann J, Cooke HJ, Krawczak M, Schmidtke J (1989) A search for restriction fragment length polymorphisms on the human Y chromosome. Hum Genet 84:86–88
7. Urquhart A, Kimpton CP, Downes TJ, Gill P (1994) Variation in short tandem repeat sequences – a survey of twelve microsatellte loci for the use as forensic identification markers. Int J Legal Med 107:13–20

8. Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet 55:175–189
9. Spurdle AB, Jenkins T (1992) The Y chromosome as a tool for studying human evolution. Curr Opin Genet Dev 2:487–491
10. Mathias N, Bayes M, Tyler-Smith C (1994) Highly informative compound haplotypes for the human Y chromosome. Hum Mol Genet 3:115–123
11. Roewer L, Arnemann J, Spurr NK, Grzeschik K-H, Epplen JT (1992) Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. Hum Genet 89:389–394
12. Chen H, Lowther W, Avramopoulos D, Antonarakis SE (1994) Homologous loci DXYS156X and DXYS156Y contain a polymorphic pentanucleotide repeat (TAAAA)n and map to human X and Y chromosomes. Hum Mutat 4:208–211
13. Roewer L, Epplen JT (1992) Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in case work. Forensic Sci Int 53:163–171
14. Trabetti E, Casartelli A, Pignatti PF (1994) Application of a Y-linked tetranucleotide repeat polymorphism to paternity cases. In: Bär W, Fiori A, Rossi U (eds) Advances in forensic haemogenetics 5. Springer, Berlin Heidelberg New York, pp 187–189
15. Gehrig C, Hochmeister M, Borer UV, Dirnhofer R (1995) Allele frequency of 3 chromosome Y-specific STR loci and their application to forensic casework. Proceedings of the 6th international symposium on human identification, Scottsdale, USA
16. Gomolka M, Hundrieser J, Nürnberg P, Roewer L, Epplen JT, Epplen C (1994) Selected di- and tetranucleotide microsatellites from chromosomes 7, 12, 14 and Y in various Eurasian populations. Hum Genet 93:592–596
17. Santos FR, Gerelsaikhan T, Munkhtuja B, Oyunsuren T, Epplen J, Pena SDJ (1996) Geographic differences in the allele frequencies of the human Y-linked tetranucleotide polymorphism DYS19. Hum Genet 97:309–313
18. Müller S, Gomolka M, Walter H (1994) The Y-specific SSLP of the locus DYS19 in four European populations. Hum Hered 44:298–300
19. Weber JL (1990) Informativeness of human (dC-dA)n.(dg-dT)n polymorphisms. Genomics 7:524–530
20. Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. Proc Natl Acad Sci USA 75:2868–2872
21. Santos FR, Pena SDJ, Epplen JT (1993) Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. Hum Genet 90: 655–656
22. Roewer L, Kayser M, Dieltjes P, Nagy M, Bakker E, Krawczak M, de Knijff P (1996) Analysis of molecular variance (AMOVA) of Y-chromosome specific microsatellites in two closely related human populations. Hum Mol Genet 5:1029–1033
23. Prinz M, Boll K, Baum H, Shaler B (1997) Multiplexing of Y chromosome specific STRs and performance for mixed samples. Forensic Sci Int (in press)
24. DNA Commission of the ISFH (1994) Report concerning further recommendations of the DNA Commission of the ISFH regarding PCR-based polymorphisms in STR (short tandem repeats). Int J Legal Med 107:159–160
25. Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323
26. Sensabaugh GF (1982) Biochemical markers of individuality. In: Saferstein R (ed) Forensic science handbook. Prentice-Hall, Englewood Cliffs, pp 340-415
27. Chakravarti A, Li CC (1983) The effect of linkage on paternity calculations. In: Walker RH (ed) Inclusion probabilities in parentage testing. American Association of Blood Banks, Arlington, VA, pp 411–420
28. Jobling MA, Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. Trends Genet 11:449–456
29. Cooper G, Amos W, Hoffman D, Rubinsztein DC (1996) Network analysis of human Y microsatellite haplotypes. Hum Mol Genet 5:1759–1766
30. Roewer L, Nagy M, Schmidt P, Epplen JT, Herzog-Schröder G (1993) Microsatellite and HLA class II oligonucleotide typing in a population of Yanomami Indians. In: Pena SDJ, Chakraborty R, Epplen JT, Jeffreys A (eds) DNA fingerprinting: the state of the science. Birkhäuser, Basel Boston Berlin, pp 221–230
31. Vogel F, Motulsky AG (1986) Human genetics. Springer, Berlin Heidelberg New York
32. Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123–1128
33. Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P (1997) Estimating Y-chromosome-specific microsatellite mutation frequencies using deep rooting pedigrees. Hum Mol Genet (in press)
34. Santos FR, Epplen JT, Pena SDJ (1993) Testing deficiency paternity cases with a Y-linked tetranucleotide repeat polymorphism. In: Pena SDJ, Chakraborty R, Epplen JT, Jeffreys A (eds) DNA fingerprinting: The state of the science. Birkhäuser, Basel Boston Berlin, pp 221–230
35. Weichhold GM, Keil W, Bayer B (1996) The Y-linked locus Y27H39 (DYS19), frequency distribution in South Bavarian and application to paternity testing. In: Carracedo A, Brinkmann B, Bär W (eds) Advances in forensic haemogenetics 6. Springer, Berlin Heidelberg New York, pp 641–643
36. Kishida T, Wang W, Fukuda M, Tamaki Y (1996) Two deficiency cases of disputed paternity: sibling or halfsibling? Nippon Hoigaku Zasshi 50:78–81
37. Skorecki K, Selig S, Blazer S, Bradman R, Bradman N, Waburton PJ, Ismajlowicz M, Hammer MF (1997) Y chromosomes of Jewish priests. Nature 385:32
38. Santos FR, Bianchi NO, Pena SDJ (1996) Worldwide distribution of human Y-chromosome haplotypes. Genome Res 6:601–6113
39. Hammer MF, Horai S (1995) Y chromosomal variation and the peopling of Japan. Am J Hum Genet 56:951–962
40. Yamamoto Y, Inoue S, Okamoto O, Murakami H, Miyaishi S, Ishizu H (1996) A short tandem repeat (STR) polymorphism of microsatellite Y-27H39 (DYS19) locus on the human Y chromosome. Proceedings of the ISFH Hakone Symposium on DNA polymorphisms, Hakone, Japan
41. Wang W, Kishida T, Fukuda M, Tamaki Y (1996) The Y-27H39 polymorphism in a Japanese population. Int J Legal Med 109:157–158