ELSEVIER

Forensic Science International 118 (2001) 116–121

Forensic Science International

www.elsevier.com/locate/forsciint

Mutations at Y-STR loci: implications for paternity testing and forensic analysis

Manfred Kayser[a,*], Antti Sajantila[b,1]

[a]*Max Planck Institute for Evolutionary Anthropology, Inselstrasse 22, D-04103 Leipzig, Germany*
[b]*Department of Forensic Medicine, Laboratory for Forensic Biology, P.O. Box 40, University of Helsinki, Helsinki 00014, Finland*

Received 14 August 2000; accepted 1 November 2000

Abstract

Knowledge about mutation rates and the mutational process of Y-chromosomal short-tandem-repeat (STR) or microsatellite loci used in paternity testing and forensic analysis is crucial for the correct interpretation of resulting genetic profiles. Therefore, we recently analysed a total of 4999 male germline transmissions from father/son pairs of confirmed paternity (probability $\geq 99.9\%$) at 15 Y-STR loci which are commonly applied to forensics. We identified 14 mutations. Locus specific mutation rate estimates varied between 0 and 8.58×10^{-3}, and the overall average mutation rate estimate was 2.80×10^{-3} (95% CIL 1.72×10^{-3}–4.27×10^{-3}). In two confirmed father/son pairs mutation at two Y-STRs were observed. The probability of two mutations occurring within the same single germline transmission was estimated to be statistically not unexpected. Additional alleles caused by insertion polymorphisms have been found at a number of Y-STRs and a frequency of 0.12% was estimated for DYS19. The observed mutational features for Y-STRs have important consequences for forensic applications such as the definition of criterions for exclusions in paternity testing and the interpretation of genetic profiles in stain analysis. In order to further enrich our knowledge of Y-STR mutations we suggest the establishment of a Y-STR mutation database and ask the forensic community for data contribution.

Keywords: Paternity testing; Y chromosome; Short-tandem-repeat loci; Y-STR; Microsatellites; Mutations

1. Introduction

DNA evidence in forensic analysis and paternity testing is based on the interpretation of similarities or differences at genetic marker loci. In paternity testing, differences at genetic marker loci between the putative father and the offspring are attributed to non-biological paternity, and thus, lead to the exclusion of biological paternity. However, spontaneous mutations in the germline of the putative father at any genetic marker locus used in the analysis can lead to an erroneous exclusion since such mutations result in differences between the parent and the offspring. Since new alleles occur due to mutation events, there is a natural correlation between the degree of polymorphism and the underlined mutation rate of a given locus: the higher the mutation rate, the more variable the locus is. For forensic purposes highly polymorphic loci are usually attractive due to their high power of discrimination.

During the last decade, short-tandem-repeat (STR) loci or microsatellites evolved to be the markers of choice for forensic case work because of their high power of discrimination and easy way of analysis. For special applications, such as deficiency paternity testing of male offspring with a deceased alleged father, STR loci from the Y chromosome have been shown to be useful [1–3], since they are transmitted without recombination from fathers to sons, and therefore, are able to characterise a paternal lineage. For investigation of sexual assault cases Y-STRs are especially suitable, since they provide a male-specific DNA profile which avoids problems of mixed stain interpretation [1,3–5].

* Corresponding author. Tel.: +49-341-9952506; fax: +49-341-9952555.
E-mail addresses: kayser@eva.mpg.de (M. Kayser), antti.sajantila@helsinki.fi (A. Sajantila).
[1] Tel.: +358-9-19127472; fax: +358-9-19127518.

0379-0738/01/$ – see front matter
PII: S0379-0738(00)00480-1

However, since highly polymorphic STR loci applied to forensics constantly evolve through mutational processes, the interpretation of the genetic profiles requires knowledge about the mutation rates at each locus used.

2. How often do mutations at Y-STR loci occur?

In contrast to autosomal STRs, data of Y-STR mutation rates were limited until recently. Different strategies have been applied to estimate the mutation rates at Y-STR loci. Heyer et al. [6] analysed nine Y-STR loci in multigenerational pedigrees. With this approach, a large number of meioses can be covered by testing of a low number of descendants. They identified four mutations in 1917 germline transmissions and estimated an average mutation rate of 2.1×10^{-3} (95% confidence-interval limits (CIL) 0.6×10^{-3}–4.9×10^{-3}). The disadvantage of using multigenerational pedigrees is that paternity within the pedigree cannot be established since only remote descendants in the male lineages are analysed. Only the analysis of father/son pairs of confirmed biological paternity can reveal reliable estimates of mutation rates. Such an analysis can either be performed on DNA directly extracted from blood or using DNA extracted from immortalised cell lines. The latter approach was used by Bianchi et al. [7] for seven Y-STRs. The disadvantage of using cell lines for mutation studies is the difficulty in differentiating somatic mutations produced by the cell line propagation from true germline mutations that occurred in the donor individual. Bianchi et al. [7] observed two mutations but attributed these to somatic events and concluded that no germline mutations occurred in 1743 meioses. Our method of choice for getting Y-STR mutation rates is the direct analysis of confirmed father/son pairs since it supplies most reliable estimates. In a previous study, we used this approach for the single Y-STR locus DYS19 and estimated a locus-specific mutation rate of 3.19×10^{-3} (95% CIL 0.41×10^{-3}–6.7×10^{-3}) based on 626 meioses [1].

However, it is known from autosomal STRs and therefore expected also for Y-chromosomal STRs, that mutation rates differ strongly between loci depending on a number of sequence-specific parameters. Thus, a complete study investigating all Y-STRs used in forensics is needed to supply the forensic community with a reliable knowledge about the rate of spontaneous mutations of every Y-STR applied to case work analyses. The first attempt to supply the needed information was published this year by us together with a number of colleagues [8]. In this study, we analysed 15 Y-STRs: DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, YCAI, YCAII and DYS413, all of which are used in forensics, in a total of 4999 meioses from confirmed father/son pairs (probability of paternity >99.9%) and identified 14 mutations (Table 1). Mutations were confirmed by DNA sequence analysis and probabilities of paternity in the mutated father/son pairs were >99.99999%. The locus-specific mutation rate estimates were between 0 and 8.58×10^{-3} and the average mutation rate estimates was 2.80×10^{-3} (95% CIL 1.72×10^{-3}–4.27×10^{-3}) across all 15 Y-STRs (Table 1). In other words: depending on the Y-STR locus analysed approximately up to eight of every thousand father/son pairs show a mutation and on average a Y-STR mutation occurs in about three of every thousand father/son pairs.

Table 1
Mutation rates of Y-chromosomal STR loci as revealed by direct observation in father/son pairs of confirmed paternity

Y-STR locus	Repetitive DNA sequence	Number of mutations observed	Number of meioses analysed	Mutation rate (95% CIL)[a] $\times 10^{-3}$
DYS19	CTAT/CTAC	2	996	2.01 (0.26–6.82)
DYS385[b]	AAGG/GAAA	2	952	2.10 (0.27–7.15)
DYS389I	CTGT/CTAT	1	425	2.35 (0.06–12.5)
DYS389II[c]	CTGT/CTAT	2	425	4.71 (0.60–16.0)
DYS390	CTGT/CTAT	4	466	8.58 (2.54–20.3)
DYS391	CTGT/CTAT	2	415	4.82 (0.61–16.4)
DYS392	ATT	0	415	0 (7.15)
DYS393	GATA	0	415	0 (7.15)
YCAI[b]	CA	0	150	
YCAII[b]	CA	0	240	2.04 (0.06–10.93)[d]
DYS413[b]	CA	1	100	
Overall		14	4999	2.80 (1.72–4.27)

[a] 95% CIL: 95% Poisson confidence-interval limits of the mutation rates (for the instances in which a mutation has been observed, both the 97.5% upper CIL and 97.5% lower CIL are given, for the instances in which no mutation has been observed, only the 95% upper CIL is given).
[b] Consists of two Y-specific loci.
[c] Excludes DYS389I.
[d] Average mutation rate across all dinucleotide loci is given due to low number of meioses studied at single loci.

3. Why different Y-STR loci have different mutation rates?

Mutation rates depend on the molecular structure of the respective genetic marker locus. For autosomal STRs it is known that the mutation rate relies on the sequence of the repetitive array. Our Y-STR data revealed a higher average mutation rate for tetranucleotide repeat Y-STRs of 3.17×10^{-3} (95% CIL 1.89×10^{-3}–4.94×10^{-3}) than for dinucleotide repeat Y-STRs of 2.04 (95% CIL $0.06-10.93 \times 10^{-3}$). However, the difference is not statistically significant ($P = 0.73$), and about ten times more meioses and more loci were analysed at tetranucleotide Y-STRs than at dinucleotide Y-STRs. A higher tetra- versus dinucleotide repeat mutation rate has also been found in family analyses of autosomal STRs [9], but using population data the opposite result was obtained [10]. These findings also demonstrate the difficulties of comparing results concerning STR mutation rates obtained from various data sets, where different strategies have been used in collecting the data.

From various studies of autosomal STRs in human, fruit flies and yeast a dependency of the mutation rate to the length of the repetitive array is known with higher rates for longer repetitive arrays [11–14]. However, for Y-STRs various statistical tests could not show a significant correlation of the length of the repetitive array and the mutation rate [8] which may be attributed to the relatively small number of mutations observed. Nevertheless, an effect of allele length is indicated also for Y-STRs by the observation that all of the 14 Y-STR mutations we identified occurred at uninterrupted repetitive arrays of $\geq$11 repeats of identical sequence (homogeneous repeats). Similarly, the four Y-STR mutation reported by Heyer et al. [6] occur at uninterrupted arrays of $\geq$10 homogeneous repeats. This is in agreement with mutations at autosomal STRs, where in a study by Brinkmann et al. [11] all of the 23 mutation have been identified in uninterrupted arrays of $\geq$10 homogeneous repeats. Altogether, these data indicate that the probability of human STRs to mutate increases when the length of the array of uninterrupted homogeneous repeats accumulates, possibly having a threshold of approximately 10 or 11 repeats. Moreover, our data indicates that not only the number of homogeneous repeats but also the number of total uninterrupted repeats as long as they belong to the same size type (e.g. tetranucleotide repeats of different sequence) might influence the mutational process. Evidence comes from the Y-STR locus DYS390, which has the highest mutation rate (8.58×10^{-3} versus $0-4.82 \times 10^{-3}$ for all tetranucleotide Y-STRs analysed) and the highest number of total tetranucleotide repeats (25.8 versus 11.7–19.2 for all tetranucleotide Y-STRs analysed) but a similar number of homogeneous repeats (longest array of 10.8 versus 10.0–13.1 for all tetranucleotide Y-STRs analysed). These data show that locus-specific knowledge of the mutation rate is crucial since generalisation is not appropriate due to locus-specific molecular differences resulting in locus-specific mutation rate differences.

Our observation of mutations only in alleles of a particular number of repeats and repeat structure, may suggest that within each locus, there are also allele-specific mutation rates. Certainly, more data are needed to further investigate allele-specific mutation rates and to establish the resulting forensic implications.

4. Forensic implications of mutations at Y-STR loci

The average Y-STR mutation rate of 2.8×10^{-3} as estimated from our data is not significantly different from average mutation rates of autosomal STR loci commonly used in forensics as recently obtained from family analyses: 2.1×10^{-3} [11], 2.7×10^{-3} [15], and 0.6×10^{-3} [16]. Does this indicate that similar recommendations concerning mutations should be implied for autosomal and Y-chromosomal STR loci? The common practice in paternity testing is that a difference at one or two out of 6–15 STR loci commonly analysed is attributed to mutation rather than non-paternity, whereas differences at more than two loci are interpreted as non-biological paternity. In order to investigate the relevance of this, we analysed 415 father/son pairs of confirmed paternity at nine Y-STR loci [8]. Interestingly, we observed one father/son pair with mutations at two Y-STR loci and a second pair with two mutations was found when expanding the number of loci to 15 Y-STRs in a subset of 50 meioses. Given the number of loci and meioses analysed and the specific average mutation rate of the loci used we calculated that mutations at two loci within a single germlines transmission are statistically not unexpected [8]. Considering the underlying mutation rate of 2.88×10^{-3}, which is an average of the specific loci analysed and is in the range of rates estimated for other Y-chromosomal and autosomal STRs, this observation has consequences for the general practice of distinguishing between exclusions and mutations in STR-based paternity testing. Rare observations of multiple STR mutations within the same germline transmission in the literature [17] might be due to the approach applied to mutation studies. Only the analysis of a large number of STR loci in the same families, as performed in our Y-STR study, is able to reveal multiple mutations within single germline transmissions. However, most of the STR mutation studies are based on pooled data of the same loci from different families [11,15,16], which will clearly be unable to detect multiple mutations.

As a consequence of the observations of two Y-STR mutations within the same father/son pair and the observed mutation rates the criterion for exclusion in paternity testing should be defined, so that an exclusion from paternity is based on exclusion constellations at the minimum of three Y-STR loci, requiring the analysis of a sufficient number of Y-STR loci (e.g. nine Y-STRs), as we already recommended earlier [2]. Furthermore, since similar mutation rates and

features have been observed for autosomal and Y-chromosomal STRs, the same criterion should be used for paternity testing based on autosomal STRs. So far, none of the recommendations of the DNA commission of the International Society for Forensic Genetics (ISFG) regarding the use of STR systems in paternity testing and forensic analysis includes clear guidelines for the consideration of mutations [18–20]. In the early ISFG guidelines on VNTR loci it is noted, that mutations have to be adequately addressed, but no details are given how to do so [21]. However, the German Association of Expert Witnesses for Paternity Testing already recommends that the exclusion from paternity based on DNA evidence should be confirmed by differences at the minimum of three loci (Peter M. Schneider, personal communication), which is in agreement with our data and conclusions. Naturally, when the number of analysed loci increases, the potential number of observed mutations increases. Since there is a complicated relationship of allele length and internal structure, it is not possible to give a straight forward number of loci/number of mutations relationship for paternity exclusion purposes. Thus, a careful analysis of locus and allele specific characteristics would be desirable before using a particular Y-STR locus in paternity testing.

To avoid the practical problem of STR mutations especially in paternity testing, STR loci can be combined with or replaced by single-nucleotide polymorphisms (SNPs), which have an essentially lower mutation rate. However, since the Y-SNP data is currently being collected from various populations it would be premature to estimate the number of Y-SNPs needed to reach the power of exclusion obtained using Y-STRs. Furthermore, since Y-SNPs seem to have a high population specificity (Jobling, this issue), it is indeed of utmost importance to evaluate the diversity and allele frequencies of the SNPs in order to find suitable ones for a particular population.

Another mutational feature especially but not exclusively of STRs from the Y chromosome is the occurrence of additional alleles due to insertion polymorphisms of larger chromosomal region including the STR locus, followed by a mutational change in the number of repeats within the STR locus. Due to the lack of recombination, the Y chromosome is predestined to accumulate various types of repetitive sequences [22,23], and thus, additional alleles seem to be more frequent at Y-chromosomal than autosomal STRs. They have been observed at various Y-STR loci [1,8,24–26], e.g. at DYS19 and DYS385 (Fig. 1). For DYS19 a

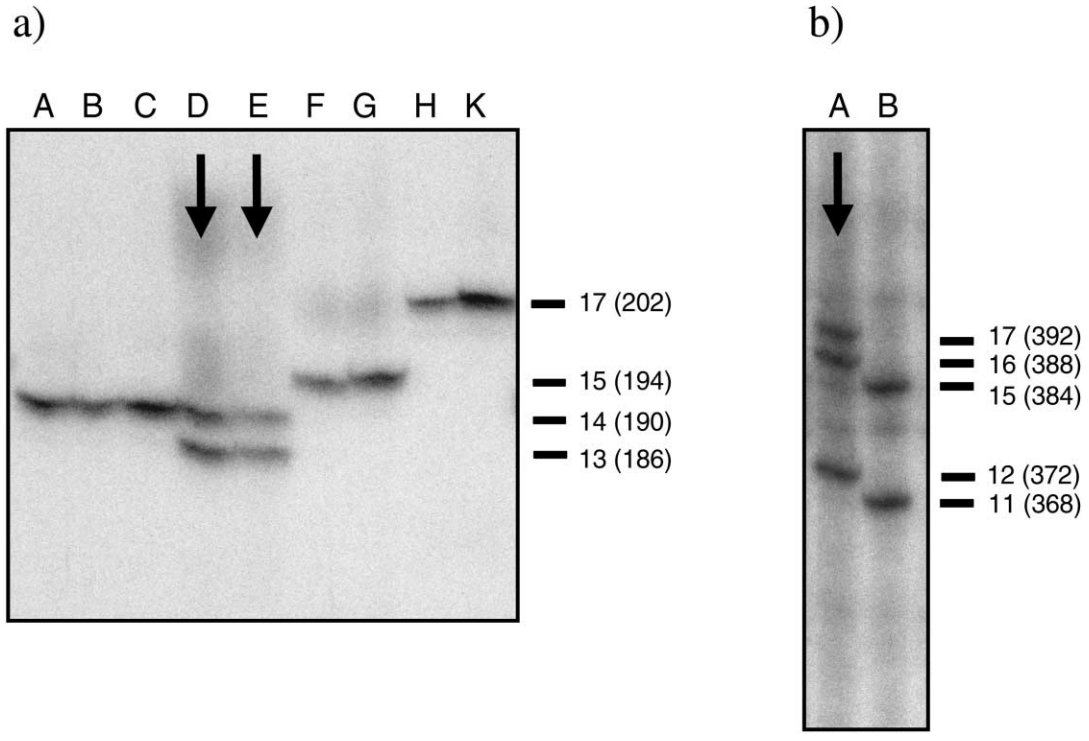

Fig. 1. Additional Y-STR alleles caused by Y-chromosomal insertion polymorphisms with subsequent mutational changes in the number of repeats within the respective Y-STR locus: (a) autoradiograph of a DYS19 locus duplication (alleles 13, 14) in a Turkish father/son pair (lane D, E), lane A control (allele 14), lane B, C: two brothers (allele 14), lane F, G (allele 15) and H, K (allele 17): two father/son pairs; (b) autoradiograph of a DYS385 locus triplication (alleles 12, 16, 17) in an individual from Kiribati Island/Micronesia (lane A), lane B control (alleles 11, 15); analysis: PAGE (4%) of ^{32}P labelled PCR-products, allele designation according to number of repeats and in parentheses fragment length in base pairs as described in Kayser et al. [1].

frequency of allele duplications of 0.12% was estimated from analysing 7772 individuals [8], which is most likely on underestimate since it only represents duplicated and diverged alleles. This has to be taken into consideration in forensic analysis where additional alleles in genetic profiles from a crime scene are usually interpreted as mixed profiles from more than one culprit. Consequently, mutations resulting in additional Y-STR alleles may lead to wrong conclusions in crime case investigations, i.e. in rape cases where the question of one or more culprits is of special importance.

In order to further enrich our knowledge about Y-STR mutation rates we suggest the establishment of a Y-STR mutation database. In this respect, we would like to ask the forensic community to contribute (1) Y-STR data from confirmed father/son pairs independent from finding a putative mutation or not, and (2) DNA samples of the complete family (including father/son(s) and mother) for those cases where deviations from Mendelian inheritance between the father and the son are observed. The authors are dedicated to further characterise the sequence structure of the Y-STR mutation and make the findings available for the forensic community. Data and samples contributions should be submitted to either of the authors.

Acknowledgements

We are grateful to L. Roewer, M. Hedman, L. Henke, J. Henke, S. Brauer, C. Krüger, M. Krawczak, M. Nagy, T. Dobosz, R. Szibor, P. de Knijff and M. Stoneking for fruitful collaboration.

References

[1] 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, A. Piccinini, A. Perez-Lezaun, 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, Int. J. Legal. Med. 110 (1997) 125–133, and 141–149.

[2] M. Kayser, C. Krüger, M. Nagy, G. Geserick, L. Roewer, Y-chromosomal DNA-analysis in paternity testing: experiences and recommendations, in: B. Olaisen, B. Brinkmann, P.J. Lincoln (Eds.), Progress in Forensic Genetics, Vol. 7, Elsevier, Amsterdam, 1998, pp. 494–496.

[3] L. Roewer, M. Kayser, P. de Knijff, K. Anslinger, D. Corach, S. Füredi, G. Geserick, L. Henke, M. Hidding, H.J. Kärgel, R. Lessig, M. Nagy, V.L. Pascali, W. Parson, B. Rolf, C. Schmitt, R. Szibor, J. Teifel-Greding, M. Krawczak, 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. 114 (2000) 31–43.

[4] K. Honda, L. Roewer, P. de Knijff, Male DNA typing from 25-year-old vaginal swabs using Y chromosomal STR polymorphisms in retrial request case, J. Forensic Sci. 44 (1999) 868–872.

[5] M. Prinz, K. Boll, H. Baum, B. Shaler, Multiplexing of Y chromosome specific STRs and performance of mixed samples, Forensic Sci. Int. 85 (1997) 209–218.

[6] E. Heyer, J. Puymirat, P. Dieltjes, E. Bakker, P. de Knijff, Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees, Hum. Mol. Genet. 6 (1997) 799–803.

[7] N.O. Bianchi, C.I. Catanesi, G. Bailliet, V.L. Martinez-Margnac, C.M. Bravi, L.B. Vidal-Rioja, R.J. Herrera, J.S. Lopez-Camelo, Characterisation of ancestral and derived Y-chromosome haplotypes of new world native populations, Am. J. Hum. Genet. 63 (1998) 1862–1871.

[8] M. Kayser, L. Roewer, M. Hedmann, K. Henke, J. Henke, S. Brauer, C. Krüger, M. Krawczak, M. Nagy, T. Dobosz, R. Szibor, P. de Knijff, M. Stoneking, A. Sajantila, Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs, Am. J. Hum. Genet. 66 (2000) 1580–1588.

[9] J.L. Weber, C. Wong, C, Mutation of human short tandem repeats, Hum. Mol. Genet. 2 (1993) 1123–1128.

[10] R. Chakraborty, M. Kimmel, D. Stivers, L.J. Davison, R. Deka, Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci, Proc. Natl. Acad. Sci. U.S.A. 94 (1994) 1041–1046.

[11] B. Brinkmann, M. Klintschar, F. Neuhuber, J. Hühne, B. Rolf, Mutation rate in human microsatellites: influence of the structure and length of tandem repeat, Am. J. Hum. Genet. 62 (1998) 1408–1415.

[12] C. Schlötterer, R. Ritter, B. Harr, G. Brem, High mutation rate of a long microsatellite allele in Drosophila melanogaster provides evidence for allele-specific mutation rates, Mol. Biol. E 15 (1998) 1269–1274.

[13] M. Wierdl, M. Dominska, T.D. Petes, Microsatellite instability in yeast: dependence on the length of the microsatellite, Genetics 146 (1997) 769–779.

[14] H. Ellegren, Heterogeneous mutation processes in human microsatellite DNA sequences, Nat. Genet. 24 (2000) 400–402.

[15] L. Henke, J. Henke, Mutation rate in human microsatellites, Am. J. Hum. Genet. 64 (1999) 1473.

[16] A. Sajantila, M. Lukka, A.-C. Syvänen, Experimentally observed germline mutations at human micro- and minisatellite loci, Eur. J. Hum. Genet. 7 (1999) 263–266.

[17] P.R. Gunn, K. Trueman, P. Stapleton, D.B. Klarkowski, DNA analysis in disputed parentage: the occurrence of two apparently false exclusions of paternity, both at short tandem repeat (STR) loci, in one child, Electrophoresis 18 (1997) 1650–1652.

[18] Recommendations of the DNA Commission of the International Society for Forensic Haemogenetics relating to the use of PCR-based polymorphisms, Forensic Sci. Int. 55 (1992) 1–3.

[19] DNA recommendations — 1994 report concerning further recommendations of the DNA commission of the ISFH

regarding PCR based polymorphisms in STR (short tandem repeat) systems. Int. J. Legal. Med. 107 (1994) 159–160.

[20] DNA recommendations — further report of the DNA Commission of the ISFH regarding the use of tandem repeat systems. Forensic Sci. Int. 87 (1997) 179–184.

[21] 1991 Report concerning recommendations of the DNA Commission of the International Society for Forensic Haemogenetics relating to the use of DNA Polymorphisms. Forensic Sci. Int. 52 (1992) 125–130.

[22] S. Foote, D. Vollrath, A. Hilton, D.C. Page, The human Y chromosome: overlapping DNA clones spanning the euchromatic region, Science 258 (1992) 60–66.

[23] B. Charlesworth, P. Sniegowski, W. Stephan, The evolutionary dynamics of repetitive DNA in eucaryotes, Nature 371 (1994) 215–220.

[24] F.R. Santos, L. Rodriguez-Delfin, S.D.J. Pena, J. Moore, K.M. Weiss, North and South Amerindians may have the same major founder Y chromosome haplotype, Am. J. Hum. Genet. 58 (1996) 1369–1370.

[25] F.R. Santos, T. Gerelsaikhan, B. Munkhtuja, T. Oyunsuren, J.T. Epplen, S.D.J. Pena, Geographic differences in the allelic frequencies of the human Y-linked tetranucleotide polymorphism DYS19, Hum. Genet. 97 (1996) 39–313.

[26] A.J. Redd, S.L. Clifford, M. Stoneking, Multiplex DNA typing of short-tandem-repeat loci on the Y chromosome, Biol. Chem. 378 (1997) 923–927.