Int J Legal Med (2006) 120: 191–200
DOI 10.1007/s00414-005-0026-1

REVIEW ARTICLE

L. Gusmão · J. M. Butler · A. Carracedo · P. Gill · M. Kayser · W. R. Mayr · N. Morling · M. Prinz · L. Roewer · C. Tyler-Smith · P. M. Schneider

DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis

Received: 14 April 2005 / Accepted: 17 June 2005 / Published online: 26 August 2005

Abstract The DNA Commission of the International Society of Forensic Genetics (ISFG) regularly publishes guidelines and recommendations concerning the application of DNA polymorphisms to the problems of human identification. A previous recommendation published in 2001 has already addressed Y-chromosome polymorphisms, with particular emphasis on short tandem repeats (STRs). Since then, the use of Y-STRs has become very popular, and numerous new loci have been introduced. The current recommendations address important aspects to clarify problems regarding the nomenclature, the definition of loci and alleles, population genetics and reporting methods.

Keywords Y chromosome · Short tandem repeat · DNA commission · ISFG · Mutation

Introduction

Y-chromosome-specific short tandem repeat (STR) analysis is an important tool in the majority of laboratories working in forensic genetics. In the same way as mtDNA, Y-STR haplotypes represent the information from a non-recombining lineage that may be shared by many individuals and, therefore, do not allow individualization to the degree that autosomal markers do. Nevertheless, during the last decade, the usefulness of Y-specific information has been recognized in deficiency paternity cases with male offspring and in forensic genetic cases where the analysis of autosomal STRs failed to give clear conclusions. For example, in a large proportion of mixed male/female stains, the male profile can only be detected through analyses of Y-chromosome markers such as Y-STRs.

L. Gusmão
IPATIMUP,
Porto, Portugal

J. M. Butler
National Institute of Standards and Technology,
Gaithersburg, MD, USA

A. Carracedo
Institute of Legal Medicine, Faculty of Medicine,
Santiago de Compostela, Spain

P. Gill
Forensic Science Service,
Birmingham, UK

M. Kayser
Department of Forensic Molecular Biology,
Erasmus University,
Rotterdam, The Netherlands

W. R. Mayr
Department of Blood Group Serology and Transfusion Medicine, University of Vienna,
Vienna, Austria

N. Morling
Department of Forensic Genetics, Institute of Forensic Medicine, University of Copenhagen,
Copenhagen, Denmark

M. Prinz
Office of Chief Medical Examiner,
Department of Forensic Biology,
New York, NY, USA

L. Roewer
Institute of Legal Medicine, Humboldt University,
Berlin, Germany

C. Tyler-Smith
Department of Biochemistry, University of Oxford,
Oxford, UK

P. M. Schneider (✉)
Institute of Legal Medicine, University of Cologne,
Melatenguertel 60-62,
50823 Cologne, Germany
e-mail: peter.schneider@uk-koeln.de
Tel.: +49-221-47887506
Fax: +49-221-4783496

The use of a common nomenclature is crucial in the forensic and population genetic fields to allow communication and data comparison. Changes to established nomenclatures and the use of different nomenclatures for the same STR markers have created difficulties in inter-laboratory data exchanges and comparisons in proficiency testing trials, especially in those including new markers [1].

After the publication of the first recommendations on forensic analysis using Y-chromosomal STRs [2], the DNA Commission of the International Society for Forensic Genetics (ISFG) is now releasing additional recommendations in order to clarify some confusion that still exists in the field, mainly as a consequence of the large number of new markers that has been introduced in recent years.

Nomenclature

Although STR locus nomenclature is straightforward and Y-STRs do not require special consideration, different repeat-based nomenclatures have been published for the same alleles [3–7]. Therefore, the main aim of the present recommendations is to provide guidelines for Y-STR allele nomenclature in order to avoid future accumulation of different nomenclatures.

Locus nomenclature

Recommendations on locus nomenclature, sequence designation and structure of STRs were previously detailed [2, 8, 9]. The main issue related to Y-STR locus nomenclature that still persists arises from the amplification of more than one STR locus (region of the Y chromosome) by the same primer pair. This can occur due to the presence of multiple primer annealing sites (in most cases as a result of locus multiplication) or due to the presence of two separate Y-STR loci lying between a pair of primers. The first is observed more often at Y-STRs than at autosomal STRs due to the highly repetitive nature of the human Y chromosome [10].

1. There are situations where more than one Y-specific locus is amplified by a single primer pair and each PCR product cannot be unambiguously assigned to a specific locus (Fig. 1). DYS385 is an example of this, where, although the two amplified fragments are sometimes named DYS385a and DYS385b, it is not correct to designate them “a” and “b” if the PCR is performed in the conventional way [11, 12] because neither fragment can be assigned unequivocally to a defined locus. Therefore, the term “DYS385 loci” should be applied to this marker, with the observed fragments treated as genotypes and the alleles separated by a hyphen, e.g. “DYS385*11-14”. The same holds true in the case of other multi-copy STRs, e.g. DYS459 and DYS464, where distinction between different amplification products is not possible. However, if specific genetic analysis assures separate identification of the different Y-STRs, e.g. as is possible now for DYS385 [13], they should be designated as DYS#a*# and DYS#b*#, e.g. DYS385a*11 and DYS385b*14.
2. There are many reports of duplications of Y-STRs that are usually single-copy, with a mutation changing the number of repeat units in one of the copies, for example DYS19, DYS385, DYS389, DYS390, DYS391, DYS393, DYS437 and DYS439 [e.g. 11, 14–17]. In this situation, the observed fragments should also be treated as genotypes, with the two alleles separated by a hyphen. It is worth mentioning the importance of reporting the frequencies of such duplications for the correct interpretation of the observation of two or more DNA fragments because such results can be misinterpreted as mixed DNA profiles.
3. In some cases, two distinct Y-STRs can be present in a single amplicon sufficiently far apart from each other to allow separate typing by locus-specific primers (Fig. 2a). If new primers are designed in order to discriminate between the two Y-STRs or to reduce the amplicon size by excluding one of the variable repeat-blocks, the 5′ STR should be designated DYS#.1 and the second one DYS#.2. Note that to define the 5′ STR in accordance with the ISFG guidelines [9], the DNA strand that was originally described in the literature or the first public database entry, preferably GenBank, is used. As examples, we have the nomenclature proposed by Gusmão et al. [7] for GATA H4. Other examples are DYS448, DYS449 and DYS552 that also include two Y-STR regions. If the loci are amplified separately, they should be called DYS448.1

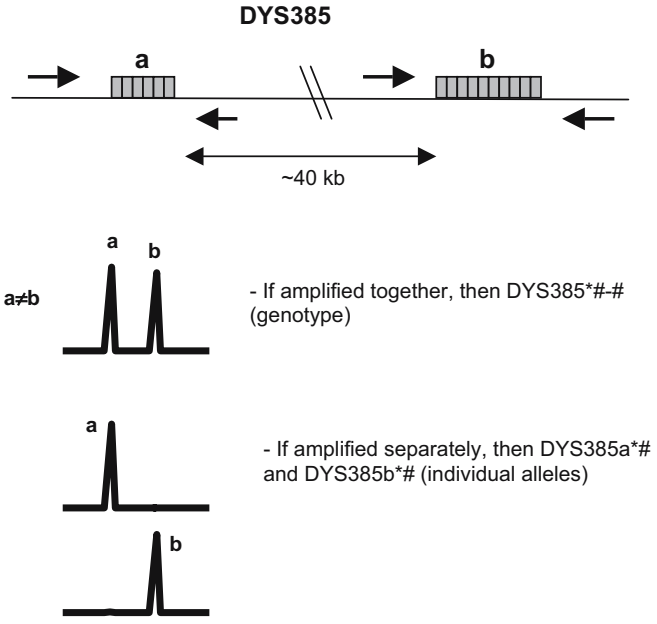

Fig. 1 Illustration of the multi-copy marker DYS385, which occurs in two inverted regions of the Y chromosome separated by about 40 kb. These regions are typically amplified together because PCR primers anneal to both regions simultaneously due to the presence of identical sequences immediately surrounding the two DYS385 copies, although a separate locus-specific amplification is also possible using a nested PCR approach [13]

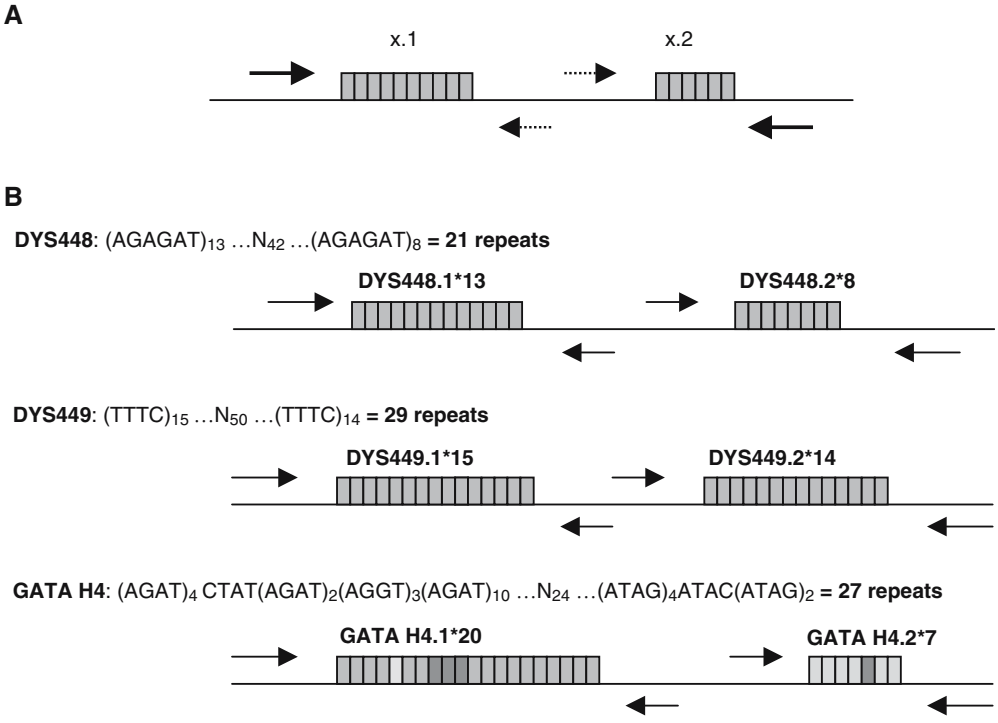

Fig. 2 **A** Two closely spaced STR repeat regions that were originally assigned to the same locus may later be subdivided. If a new PCR primer is developed that can hybridize between the two regions, then the regions should be designated .1 and .2 (e.g. DYS448.1 and DYS448.2). **B** Examples of where the original PCR primers target two blocks of STR repeats that are separated by a number of nucleotides (in these cases, 42, 50 or 24)

and DYS448.2, DYS449.1 and DYS449.2, and DYS552.1 and DYS552.2, respectively (Fig. 2b).

Allele designation of Y-STRs

Y-chromosomal STRs show the same sequence structure and mutational mechanism as autosomal STRs [14]. Therefore, the same rules apply, and allele nomenclature follows the principles previously described for autosomal STRs [9] and later emphasized for Y-chromosomal STRs [2].

Established allele nomenclatures

To avoid further confusion due to nomenclature changes, the nomenclature of widely used Y-STRs should not be altered, even if the present guidelines are not followed. This is applied to the Y-STRs DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438 and DYS439, which are already included in well-known databases and widely used commercial kits in the forensic field. Using an established nomenclature (Table 1; see also the National Institute of Standards and Technology [NIST] STRBase website for details [18]), these markers are the core set of the Y Chromosome Haplotype Reference Database (YHRD) [17] and selected by the Scientific Working Group on DNA Analysis Methods (SWGDAM) for forensic DNA analysis in the USA [19]. For the same reason, no nomenclature changes are recommended for the Y-STR markers for which sequence information is available, and a nomenclature based on the recommendations of the DNA Commission of the ISFG has already been published (see Table 2). In situations where two or more nomenclatures already exist, priority should be given to the nomenclature that most closely follows the present guidelines (some

Table 1 DYS19, DYS385, DYS389 I and II, DYS390, DYS391, DYS392, DYS393, DYS438 and DYS439 repeat structure and nomenclature

GDB locus name	STR reference	Repeat structure	Nomenclature reference
DYS19/DYS394	[32]	$(TAGA)_3$**tagg**$(TAGA)_n$	[11, 51]
DYS385	[11, 51]	**(aagg)$_{6-7}$**$(GAAA)_n$	[11, 51]
DYS389 I	[11, 51]	$(TCTG)_3(TCTA)_n$	[17]
DYS389 II	[11, 51]	$(TCTG)_n(TCTA)_n$***N*$_{28}$**$(TCTG)_3(TCTA)_n$	[17]
DYS390	[11, 51]	**(tcta)$_2$**$(TCTG)_n(TCTA)_n(TCTG)_n$ $(TCTA)_n$**tca(tcta)$_2$**	[11, 51]
DYS391	[11, 51]	**(tctg)$_3$**$(TCTA)_n$	[11, 51]
DYS392	[11, 51]	$(TAT)_n$	[11, 51]
DYS393/DYS395	[11, 51]	$(AGAT)_n$	[11, 51]
DYS438	[4]	$(TTTTC)_1(TTTTA)_{0-1}(TTTTC)_n$	[7]
DYS439 (GATA A4)	[4]	$(GATA)_n$	[7]

Segments that are not included in the allele nomenclature are in bold small letters

Table 2 Y-STRs repeat structure and nomenclature

GDB locus name	STR reference	Repeat structure	Nomenclature reference
YCAII[MC]	[33]	$(CA)_n$	[53]
YCAIII[MC]	[33]	$(CA)_n$	[11, 51]
DYS388	[11, 51]	$(ATT)_n$	[28]
DYS426	[35]	$(GTT)_n$	[28]
DYS434	[4]	$(TAAT)_{1–2}(CTAT)_n$	[7]
DYS435	[4]	$(TGGA)_n$	[7]
DYS436	[4]	$(GTT)_n$	[7]
DYS437	[4]	$(TCTA)_n(TCTG)_{1–3}(TCTA)_4$	[7]
DYS441	[36]	$(TTCC)_n$	[a]
DYS442	[36]	$(TATC)_2(TGTC)_3(TATC)_n$	[a]
DYS443	[37]	$(TTCC)_n$	[37]
DYS444	[37]	$(ATAG)_n$	[a]
DYS445	[37]	$(TTTA)_n$	[37]
DYS446	[38]	$(TCTCT)_n$	[38]
DYS447	[38]	$(TAATA)_n(TAAAA)_1(TAATA)_n(TAAAA)_1$ $(TAATA)_n$	[38]
DYS448	[38]	$(AGAGAT)_n\mathbf{N_{42}}(AGAGAT)_n$	[38]
DYS449	[38]	$(TTTC)_n\mathbf{N_{50}}(TTTC)_n$	[38]
DYS450	[38]	$(TTTTA)_n$	[38]
DYS452	[38]	$(TATAC)_2(TGTAC)_2(TATAC)_n(CATAC)_1$ $(TATAC)_1(CATAC)_1(TATAC)_{3–4}$ $(CATAC)_{0–2}(TATAC)_{0–3}(CATAC)_1(TATAC)_3$	[38]
DYS453	[38]	$(AAAT)_n$	[38]
DYS454	[38]	$(AAAT)_n$	[38]
DYS455	[38]	$(AAAT)_n$	[38]
DYS456	[38]	$(AGAT)_n$	[38]
DYS458	[38]	$(GAAA)_n$	[38]
DYS459[MC]	[38]	$(TAAA)_n$	[38]
DYS460 (formerly GATA A7.1)	[3]	$(ATAG)_n$	[7]
DYS461 (formerly GATA A7.2)	[3]	$(TAGA)_n(CAGA)$	[7]
DYS462 (formerly G09411)	[39]	$(TATG)_n$	[39]
DYS463	[38]	$(AAAGG)_n(AAGGG)_n(AAGGA)_2$	[38]
DYS464[MC]	[38]	$(CCTT)_n$	[38]
DYS485	[24]	$(TTA)_n$	[54]
DYS490	[24]	$(TTA)_n$	[54]
DYS495	[24]	$(AAT)_n$	[54]
DYS504	[24]	$(TCCT)_n$	[54]
DYS505	[24]	$(TCCT)_n$	[54]
DYS508	[24]	$(TATC)_n$	[54]
DYS510	[24]	$(TAGA)_3(TACA)(TAGA)(TACA)(TAGA)_n$	[52]
DYS513	[24]	$(TATC)_n$	[52]
DYS520	[24]	$(ATAG)_n(ATAC)_n$	[54]
DYS522	[24]	$(GATA)_n$	[54]
DYS525	[24]	$(TAGA)_n$	[54]
DYS532	[24]	$(CTTT)_n$	[54]
DYS533	[24]	$(ATCT)_n$	[54]
DYS534	[24]	$(CTTT)_n$	[54]
DYS540	[24]	$(TTAT)_n$	[54]
DYS542	[24]	$(ATAG)_2ATAA(ATAG)_n$	[54]
DYS544	[24]	$(GATA)_3GATG(GATA)_n$	[52]
DYS552	[24]	$(TCTA)_3TCTG\ (TCTA)_n\mathbf{N_{40}}(TCTA)_n$	[52]

Table 2 (continued)

GDB locus name	STR reference	Repeat structure	Nomenclature reference
DYS556	[24]	$(AATA)_n$	[54]
DYS557	[24]	$(TTTC)_n$	[54]
DYS561	[24]	$(GATA)_n(GACA)_4$	[52]
DYS570	[24]	$(TTTC)_n$	[54]
DYS575	[24]	$(AAAT)_n$	[54]
DYS576	[24]	$(AAAG)_n$	[54]
DYS587	[24]	$(ATACA)_n[(GTACA)(ATACA)]_3$	[52]
DYS593	[24]	$(AAAAC)_2AAAAT(AAAAC)_4(AAAAT)_n$	[52]
DYS594	[24]	$(TAAAA)_n$	[54]
DYS632	[24]	$(CATT)_n$	[54]
DYS635 (formerly GATA-C4)	[3]	$(TCTA)_4(TGTA)_2(TCTA)_2(TGTA)_2(TCTA)_2(TGTA)_{0,2}(TCTA)_n$	[7]
DYS641	[24]	$(TAAA)_n$	[54]
DYS643	[24]	$(CTTTT)_n$	[54]
GATA-A10	[3]	$(TCCA)_2(TATC)_n$	[7]
GATA-H4	[3]	$(AGAT)_4CTAT(AGAT)_2(AGGT)_3(AGAT)_n\mathbf{N_{24}}(ATAG)_4(ATAC)_1(ATAG)_2$	[7]

MC Multi-copy Y-STR
[a]Modified in order to observe the ISFG recommendations

examples are DYS435, DYS437, DYS460, DYS635, GATA A10 and GATA H4).

Nomenclature guidelines

Ideally, alleles should be designated according to the total number of repeats included in a simple or complex sequence structure that varies among individuals. Due to the impracticability of sequencing all samples, the only way to identify the main sources of variation is by sequence analyses of individuals sampled from a wide range of haplotypes. Since mutation rates of Y-STRs are about 100,000 times higher than those of Y-SNPs [14, 20, 21], the choice of samples from different Y-SNP-defined haplogroups, rather than different-sized alleles from a single population, will increase the genetic distance between sequences and, consequently, maximise the chance of identifying locus sequence heterogeneity. A high proportion of the polymorphic Y-STRs described in humans is also present in chimpanzees and can be amplified using the same primers [22, 23]. Chimpanzee sequence information may also be used to identify regions that are likely to vary.

1. It is recommended that alleles are named according to the total number of contiguous variant and non-variant repeats determined from sequence data. Single interruptions within repetitive blocks should be considered as part of the locus (e.g. DYS452, where the single CATAC sequence interrupting the other repeats in several places should be included in the total number of repeats). In a complex STR, single repeat units located adjacent to the main array and consisting of the same sequence as the main variable repeat should be considered as part of the locus structure since the entire structure could have evolved from a single array. Therefore, these single units are included in the allele nomenclature. For example, a hypothetical STR allele with the sequence ...$(GATA)_n(GACA)_2(GATA)$... is considered to have n+2+1 repeats.
2. The inclusion of non-variant repeats dispersed throughout the amplified region can be a disadvantage in nomenclature standardization since, in forensic genetics, new primers may be designed in order to amplify smaller fragments that may exclude the non-variant repeats. For this reason, repetitive motifs that are not adjacent to the variable stretch, have three or less units and show no size variation within humans or between humans and chimpanzees should not be included in allele nomenclature. For example, alleles at a hypothetical STR with the sequence ...$(GATA)_n(GACA)_2N_8(GATA)_3$... is called n+2, excluding the non-adjacent $(GATA)_3$ repetitive stretch from the allele nomenclature.
   A distinction has to be made between the number of nucleotides that constitute an interruption within a single locus and the number of nucleotides that form a boundary between two separate loci. In accordance with current usage [24], we recommend that the distinction be based on the number of nucleotides in the interrupting section compared to the number of nucleotides in the Y-STR repeats. If the number of interrupting nucleotides is similar to or less than the number of nucleotides in the repeats, the region is considered one unit, with a length corresponding to the total number of nucleotides. Thus, ...$(GATA)_n(GACA)_2N_4(GATA)_3$... is considered as one complex locus with n+6 units, while ...$(GATA)_n(GACA)_2N_5(GATA)_3$... is considered to be two loci with n+2 and 3 units, respectively, of which n+2 would be included in the allele nomenclature.
3. Sometimes, allele length variation indicating the presence of intermediate alleles can be detected in

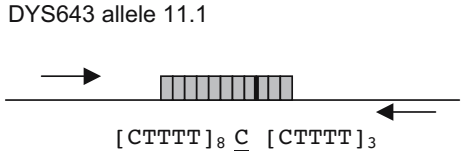

Fig. 3 Schematic illustration of an insertion occurring within the STR repeat region of the locus DYS643 that results in an allele which is one nucleotide longer than the more common allele containing only 11 CTTTT repeats

addition to variation in integral numbers of repeats. Such alleles have been created by insertion/deletion events and fall into two classes. A partial repeat can be found inside the locus and, in this case, it is recommended that the allele be designated according to Olaisen et al. [9] and Gill et al. [2], e.g. Fig. 3. Intermediate alleles of this type have been detected at the following loci: DYS19, DYS385, DYS389 I, DYS390, DYS392, DYS393 and DYS438 (YHRD [17], Reliagene [25] and Promega [26] databases).

4. Intermediate-sized alleles can also arise from mutations, usually insertions/deletions of 1 bp, in the flanking sequences which alter the length of the PCR product. A common example is the deletion of a T in the flanking region of the DYS385 locus [27]. This variant is only detected when using a reverse primer located downstream of the deletion site, e.g. the first primers described by Kayser et al. [11] and the ones included in the Y-Plex 6 kit (Reliagene), but not with those published by Schneider et al. [12] or the ones included in the PowerPlex Y System kit (Promega Corporation) (Fig. 4). It is, therefore, expected that allele designation discrepancies will be observed in some DYS385 alleles when different primer pairs (or commercial kits) are used, e.g. [28, 29]. In order to have a standard nomenclature that allows comparisons to be made when different primers are used for the amplification of alleles carrying this type of flanking polymorphism, we recommend that these variants not be included as part of the repeat size. Instead, they should be considered as additional information, indicated after the number of complete repeat units. For example, an allele with 11 repeats and a T insertion at base 40 upstream from the repeat is not named 11.1 but 11 (U40Tins), where 11 stands for the number of complete repeats, U40 indicates that the polymorphism is located 40 bp upstream of the repeat and Tins indicates that a T has been inserted. In the case of DYS385, a T deletion can occur inside a homopolymeric T tract located between nucleotide positions 74 and 80 downstream of the repeat (Fig. 4, box 3). Since the exact position of the deletion is unknown, we recommend, for nomenclature purposes, that the deletion be assigned to the highest numbered end of the homopolymeric region (the same strategy is used in the nomenclature of mtDNA variations residing within homopolymeric stretches; [30]). Thus, a DYS385 allele with 18 repeats and a T deletion at the homopolymeric T tract located downstream the repeat is named 18 (D80Tdel) and not, e.g. 17.3. For database purposes or QA schemes where data are compiled from laboratories using different primer sets, alleles with the same number of repeats are considered to lie in the same class, and differences in the flanking sequences are noted separately.

5. Mutations in the flanking regions of a Y-STR other than insertion/deletions do not interfere with the allele size estimation to any significant degree and, consequently, do not affect the allele nomenclature. However, a point mutation in the primer binding region may result in the lack of sufficient binding of a primer and thereby cause a lack of a detectable amount of PCR product, resulting in a null ('silent') allele. Point mutations have been described in the flanking regions of DYS391, DYS437, DYS438 [31] and DYS392 (J.M. Butler, personal communication). At DYS391, a C→G substitution can be present at base 87 downstream from the repeat (D87C→G), and DYS437 U3C→T, DYS438 D7A→C and DYS392 (U180C→G) are found in some individuals. In order to optimize Y-STR typing, we recommend that point mutations verified by sequence analysis be published or reported to the YHRD [17] using the nomenclature described above.

6. Kayser et al. [24] described 166 Y-chromosome-specific STR polymorphisms that are potentially useful in population and forensic genetic analyses. In the nomenclature of these Y-STRs, if no additional sequence variation is found, we recommend that the authors' locus delimitation criteria be taken into account and the present recommendations be followed.

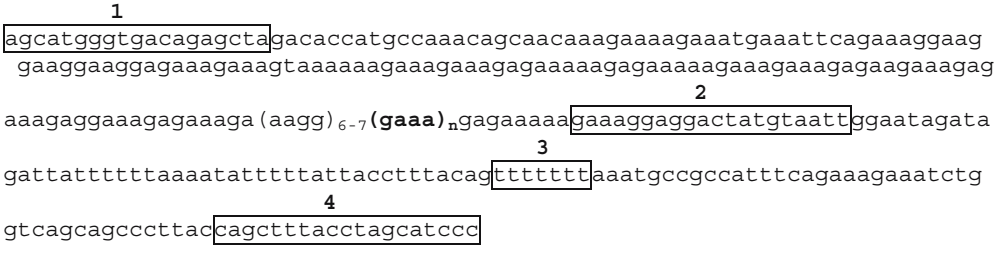

Fig. 4 The DYS385 sequence. *Boxes 1* and *4* contain the annealing sequences for the forward and reverse primers described by Kayser et al. [11]. In *box 2* is the annealing position for the alternative reverse primer described by Schneider et al. [12]. *Box 3* shows the position of the T deletion detected by Füredi et al. [27]

7. It is important that journal editors, reviewers and organisers of QA schemes focus on the use of standardized nomenclatures in order to obtain uniformity and avoid the spread of confusing nomenclatures.
8. It is also important that commercial Y-STR kits follow the nomenclature recommendations so that direct comparisons between results obtained with different kits are possible.

Locus selection for forensic applications

At present, about 220 different male-specific STRs that are potentially useful for forensic genetics have been identified on the human Y chromosome [3, 4, 11, 24, 32–40]. For most of them, relevant data on sequence variation and discriminatory capacity are still scarce, and it is therefore premature to recommend any of the novel Y-STRs for forensic purposes. Nevertheless, due to the fact that a large number of markers are now available, some criteria for the selection of new Y-STRs for forensic genetic investigations will be suggested.

1. In forensic investigations of small amounts of DNA, the availability of large multiplexes allowing fast typing of many markers is very important. Therefore, it is recommended that the potential for multiplexing be taken into consideration when Y-STRs are selected.
2. In forensic analyses, Y-STRs are often used to determine the number of individuals contributing to a mixture of DNA in a stain. For this purpose, single-copy loci are ideal since it may be difficult to draw definite conclusions from multi-copy loci.
3. If there is a choice between equally polymorphic simple and complex Y-STRs, preference should be given to simple Y-STRs since they are favoured by population geneticists and their use will facilitate database sharing between the fields.
4. If a ‘new’ Y-STR is considered for addition to an existing set of Y-STRs, the additional information the extra Y-STR will add to the information obtained by the original set of Y-STRs needs to be investigated. Due to the lack of recombination between Y-specific loci, the whole haplotype is transmitted as a single marker, and haplotype diversity defined by a set of STRs must be established by frequency estimates of the whole haplotype. The haplotype diversity cannot be predicted by combining the average diversity at each single locus. The two main factors that contribute to the single-locus diversity within a population are the presence of distinct lineages differing in their modal Y-STR alleles (where the combination of lineages may be population-specific) and the variation accumulated within each lineage by mutation. Only the second of these will contribute to the decrease in the association between alleles of different loci and therefore be reflected in the Y-STR diversity at the haplotype level [41]. Therefore, it is recommended that Y-STR diversities be studied in Y-SNP-defined haplogroups rather than in specific populations in order to choose the best markers to increase Y-STR haplotype discrimination capacity in forensic genetics.

Mutation

With the large number of Y-STR polymorphisms being described, as well as the development of new multiplex kits incorporating an increasing number of these markers, it is expected that in the near future, forensic laboratories will be able to use highly discriminating sets of Y-STRs.

The potential to distinguish between relatives belonging to the same paternal lineage will be increased due to the accumulation of Y-STR mutations from generation to generation. In paternity and identity testing including male relatives, it is necessary to take the mutation rates into account. Studies of Y-STR mutation rates are few and have so far considered a restricted number of markers (data concerning Y-STR mutations and respective references are compiled at the YHRD). Based on an average mutation rate of 2.8×10^{-3} [14], haplotypes including nine Y-STRs (e.g. the YHRD minimal haplotype) are expected to show at least one allele mismatch between father and son in about one out of 40 pairs analysed (see Table 3). This value will increase to one out of 20 pairs for males two generations apart from each other and in father/son pairs when 18 to 19 STRs are typed. As expected from the mutation rate estimates, verified father/son pairs with mutations at more than one Y-STR have been reported [14, 21].

STR mutation rates, including Y-STRs, show not only inter- but also intra-locus variation depending on the locus structure and the allele repeat lengths (e.g. [14, 21, 42]). A large amount of data is necessary to estimate reliable mutation rates, which are crucial for the interpretation of the genetic results in certain situations. Therefore, in addition to the efforts that are being made in publishing population data and in population databasing, the publication of mutation data from father/son pairs with confirmed paternity is encouraged. Selective publication of studies in which mutations are found would lead to upwardly biased estimates of mutation rates, so all such studies should be published, irrespective of outcome, for example by the submission to the YHRD [17].

1. In order to make the compilation of data published by different groups possible, the inclusion of the following information is recommended:
   - The sequences of the alleles involved in the mutations
   - Allele distribution in the fathers’ population allowing estimation of allele-specific mutation rates
   - When available, the father’s age at the birth of the son (in both cases with and without mutations)
2. Estimates of mutation rates must be based on the number of observed mutations and the total number of mutations possible from the transmissions of alleles.

Table 3 Probability of finding no mutations or at least one mutation between two Y-STR haplotypes one and two generations apart

Number of STRs (n)	One generation		Two generations	
	Probability of no mutation $[(1-\mu)^n]$	Probability of at least one mutation $[1-(1-\mu)^n]$	Probability of no mutation $[(1-\mu)^{2n}]$	Probability of at least one mutation $[1-(1-\mu)^{2n}]$
1	0.99720000	0.00280000	0.99440784	0.00559216
2	0.99440784	0.00559216	0.988846952	0.011153048
3	0.99162350	0.00837650	0.983317162	0.016682838
4	0.98884695	0.01115305	0.977818295	0.022181705
5	0.98607818	0.01392182	0.972350179	0.027649821
6	0.98331716	0.01668284	0.966912641	0.033087359
7	0.98056387	0.01943613	0.961505511	0.038494489
8	0.97781829	0.02218171	0.956128618	0.043871382
9	0.97508040	0.02491960	0.950781794	0.049218206
10	0.97235018	0.02764982	0.94546487	0.05453513
11	0.96962760	0.03037240	0.940177679	0.059822321
12	0.96691264	0.03308736	0.934920055	0.065079945
13	0.96420529	0.03579471	0.929691832	0.070308168
14	0.96150551	0.03849449	0.924492847	0.075507153
15	0.95881330	0.04118670	0.919322935	0.080677065
16	0.95612862	0.04387138	0.914181934	0.085818066
17	0.95345146	0.04654854	0.909069683	0.090930317
18	0.95078179	0.04921821	0.903986019	0.096013981
19	0.94811960	0.05188040	0.898930785	0.101069215
20	0.94546487	0.05453513	0.89390382	0.10609618
21	0.94281757	0.05718243	0.888904967	0.111095033
22	0.94017768	0.05982232	0.883934068	0.116065932
23	0.93754518	0.06245482	0.878990967	0.121009033
24	0.93492006	0.06507994	0.874075509	0.125924491
25	0.93230228	0.06769772	0.869187539	0.130812461
26	0.92969183	0.07030817	0.864326903	0.135673097
27	0.92708870	0.07291130	0.859493449	0.140506551
28	0.92449285	0.07550715	0.854687024	0.145312976
29	0.92190427	0.07809573	0.849907478	0.150092522
30	0.91932294	0.08067706	0.845154659	0.154845341
31	0.91674883	0.08325117	0.840428419	0.159571581
32	0.91418193	0.08581807	0.835728609	0.164271391
33	0.91162222	0.08837778	0.831055081	0.168944919
34	0.90906968	0.09093032	0.826407688	0.173592312
35	0.90652429	0.09347571	0.821786284	0.178213716
36	0.90398602	0.09601398	0.817190723	0.182809277
37	0.90145486	0.09854514	0.812620862	0.187379138
38	0.89893078	0.10106922	0.808076556	0.191923444
39	0.89641378	0.10358622	0.803557663	0.196442337
40	0.89390382	0.10609618	0.79906404	0.20093596

These values were estimated for haplotypes including 1–40 STRs and using the Y-STR average mutation rate value calculated by Kayser et al. [14] (μ=2.8×10^{-3})

Some differences in DYS385 mutation rate estimates can be attributed to different methodologies. Some authors have reported the number of mutations for both DYS385 loci, taking into account only the number of meioses analysed [21], while others have counted each locus separately, considering the number of allele transmissions, which for a duplicated Y-STR such as DYS385 equals two times the number of meioses [14]. Therefore, it is recommended that for multi-copy loci (e.g. DYS385, DYS464), mutation rates should be estimated by considering the number of mutations observed in the total number of allele transmissions.

Y x-STR haplotype frequency estimation

Y-chromosomal STRs constitute a single haplotype, and the frequency of a Y-STR haplotype is assessed in the relevant population. It is not valid to multiply together individual allele frequencies. When a match is established using Y-STR haplotype analysis, the frequency of the Y-STR haplotype in a population is needed for the calculation of a match probability. A number of strategies have been proposed to determine this (e.g. [43, 44]), and they are currently the subject of scientific evaluation. Individual laboratories must establish relevant, regional Y-STR hap-

lotype databases. Also, multi-regional Y-STR databases are available (YHRD [17]; Reliagene [25]; Promega [26]; Applied Biosystems [45]). Most of the databases provide haplotype frequency estimates for larger regions, e.g. for the major population groups in the USA or for geographically or linguistically derived meta-populations. However, pooling of different regions is only valid if there is no population substructure, i.e. no statistically significant difference between the Y-STR haplotype distributions in different regions. Population substructure has been shown in a number of regional groups within the same (but not between different) major US populations [46, 47] and also in some European groups [48, 49]. However, such statistical analyses—and subsequent conclusions—are highly dependent on the amount of data available. Recently, it was shown that with the increased size of the YHRD [17], clusters of regional groups could be identified in Europe that show non-significant differences within the cluster but significant differences between clusters, indicating Y-STR-haplotype-based population substructure [50]. These effects thus need to be considered as well when haplotype frequencies are estimated.

Recommendations on the estimation of Y-STR haplotype frequencies and estimation of the weight of the evidence of Y-STR typing will be presented separately as guidelines for the interpretation of forensic genetic evidence.

References

1. Gomez J, Carracedo A (2000) The 1998–1999 collaborative exercises and proficiency testing program on DNA typing of the Spanish and Portuguese Working Group of the International Society for Forensic Genetics (GEP-ISFG). Forensic Sci Int 114:21–30
2. Gill P, Brenner C, Brinkmann B, Budowle B, Carracedo A, Jobling MA, De Knijff P, Kayser M, Krawczak M, Mayr WR, Morling N, Olaisen B, Pascali V, Prinz M, Roewer L, Schneider PM, Sajantila A, Tyler-Smith C (2001) DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs. Forensic Sci Int 124:5–10
3. White PS, Tatum OL, Deaven LL, Longmire JL (1999) New, male-specific microsatellite markers from the human Y chromosome. Genomics 57:433–437
4. Ayub Q, Mohyuddin A, Qamar R, Mazhar K, Zerjal T, Tyler-Smith C (2000) Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. Nucleic Acids Res 28(2):e8
5. González-Neira A, Elmoznino M, Lareu MV, Sánchez-Diz P, Gusmão L, Prinz M, Carracedo A (2001) Sequence structure of 12 novel Y chromosome microsatellites and PCR amplification strategies. Forensic Sci Int 122:19–26
6. Hou YP, Zhang J, Li YB, Wu J, Zhang SZ, Prinz M (2001) Allele sequences of six new Y-STR loci and haplotypes in the Chinese Han population. Forensic Sci Int 118:153–159
7. Gusmão L, Gonzalez-Neira A, Alves C, Lareu M, Costa S, Amorim A, Carracedo A (2002) Chimpanzee homologous of human Y specific STRs. A comparative study and a proposal for nomenclature. Forensic Sci Int 126:129–136
8. Gill P, Brinkmann B, d'Aloja E, Andersen J, Bar W, Carracedo A, Dupuy B, Eriksen B, Jangblad M, Johnsson V, Kloosterman AD, Lincoln P, Morling N, Rand S, Sabatier M, Scheithauer R, Schneider PM, Vide MC (1997) Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature. Forensic Sci Int 87:185–192
9. Olaisen B, Bar W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr WR, Rand S (1998) DNA recommendations 1997 of the International Society for Forensic Genetics. Vox Sang 74:61–63
10. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou SF, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T, Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlfing T, Scott K, Schultz B, Strong C, Tin-Wollam A, Yang SP, Waterston RH, Wilson RK, Rozen S, Page DC (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825–837
11. Kayser M, Caglia 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 (1997) Evaluation of Y-chromosomal STRs: a multicenter study. Int J Legal Med 110:125–133, appendix 141–149
12. Schneider PM, Meuser S, Waiyawuth W, Seo Y, Rittner C (1998) Tandem repeat structure of the duplicated Y-chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations. Forensic Sci Int 97:61–70
13. Kittler R, Erler A, Brauer S, Stoneking M, Kayser M (2003) Apparent intra-chromosomal exchange on the human Y chromosome explained by population history. Eur J Hum Genet 11:304–314
14. Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Krüger C, Krawczak M, Nagy M, Dobosz T, Szibor R, De Knijff P, Stoneking M, Sajantila A (2000) 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 66:1580–1588
15. Bosch E, Jobling MA (2003) Duplications of the *AZFa* region of the human Y chromosome are mediated by homologous recombination between HERVs and are compatible with male fertility. Hum Mol Genet 12:341–347
16. Diederiche M, Martín P, Amorim A, Corte-Real F, Gusmão L (2005) A case of double alleles at three Y-STR loci: forensic implications. Int J Legal Med 119(4):223–225
17. YHRD. Y-Chromosome Haplotype Reference Database. http://www.yhrd.org, publication date: July 15, 2005; access date: August 1, 2005
18. NIST. National Institute of Standards and Technology, STRBase website. http://www.cstl.nist.gov/div831/strbase/y_strs.htm, publication date: July 21, 2005; access date: August 1, 2005
19. Butler JM (2003) Recent developments in Y-short tandem repeat and Y-single nucleotide polymorphism analysis. Forensic Sci Rev 15:91–111
20. Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW (2000) Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. Proc Natl Acad Sci U S A 97:7360–7365
21. Dupuy BM, Stenersen M, Egeland T, Olaisen B (2004) Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci. Hum Mutat 23:117–124
22. Gusmão L, Gonzalez-Neira A, Alves C, Sanchez-Diz P, Dauber EM, Amorim A, Carracedo A (2002) Genetic diversity of Y-specific STRs in chimpanzees (Pantroglodytes). Am J Primatol 57:21–29

23. Erler A, Stoneking M, Kayser M (2004) Development of Y-chromosomal microsatellite markers for nonhuman primates. Mol Ecol 13:2921–2930
24. Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyuddin A, Mehdi SQ, Rosser Z, Stoneking M, Jobling MA, Sajantila A, Tyler-Smith C (2004) A comprehensive survey of human Y-chromosomal microsatellites. Am J Hum Genet 74:1183–1197
25. Reliagene. Y-STR Haplotype Reference Database for U.S. Populations. http://www.reliagene.com, publication date: not available; access date: August 1, 2005
26. Promega Corporation. PowerPlex Y Haplotype Database. http://www.promega.com/techserv/tools/pplexy/, publication date: February 2005; access date: August 1, 2005
27. Füredi S, Woller J, Padar Z, Angyal M (1999) Y-STR haplotyping in two Hungarian populations. Int J Legal Med 113:38–42
28. Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, Hammer MF (2002) A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. Forensic Sci Int 129:10–24
29. Schoske R, Vallone PM, Kline MC, Redman JW, Butler JM (2004) High-throughput Y-STR typing of U.S. populations with 27 regions of the Y chromosome using two multiplex PCR assays. Forensic Sci Int 139:107–121
30. Carracedo A, Bär W, Lincoln P, Mayr W, Morling N, Olaisen B, Schneider P, Budowle B, Brinkmann B, Gill P, Holland M, Tully G, Wilson M (2000) DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. Forensic Sci Int 110:79–85
31. Gusmão L, Alves C, Costa S, Amorim A, Brion M, González-Neira A, Carracedo A (2002) Point mutations in the flanking regions of the Y-chromosome specific STRs DYS391, DYS437 and DYS438. Int J Legal Med 116:322–326
32. Roewer L, Arnemann J, Spurr NK, Grzeschik KH, Epplen JT (1992) Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. Hum Genet 89:389–394
33. Mathias N, Bayes M, Tyler-Smith C (1994) Highly informative compound haplotypes for the human Y chromosome. Hum Mol Genet 3:115–123
34. 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
35. Jobling MA, Samara V, Pandya A, Fretwell N, Bernasconi B, Mitchell RJ, Gerelsaikhan T, Dashnyam B, Sajantila A, Salo PJ, Nakahori Y, Disteche CM, Thangaraj K, Singh L, Crawford MH, Tyler-Smith C (1996) Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. Hum Mol Genet 5:1767–1775
36. Iida R, Tsubota E, Matsuki T (2001) Identification and characterization of two novel human polymorphic STRs on the Y chromosome. Int J Legal Med 115:54–56
37. Iida R, Tsubota E, Sawazaki K, Masuyama M, Matsuki T, Yasuda T, Kishi K (2002) Characterization and haplotype analysis of the polymorphic Y-STRs DYS443, DYS444 and DYS445 in a Japanese population. Int J Legal Med 116:191–194
38. Redd AJ, Agellon AB, Kearney VA, Contreras VA, Karafet T, Park H, de Knijff P, Butler JM, Hammer MF (2002) Forensic value of 14 novel STRs on the human Y chromosome. Forensic Sci Int 130:97–111
39. Bosch E, Lee AC, Calafell F, Arroyo E, Henneman P, De Knijff P, Jobling MA (2002) High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions. Forensic Sci Int 125:42–51
40. Mohyuddin A, Ayub Q, Siddiqi S, Carvalho-Silva DR, Mazhar K, Rehman S, Firasat S, Dar A, Tyler-Smith C, Mehdi SQ (2004) Genetic instability in EBV-transformed lymphoblastoid cell lines. Biochim Biophys Acta 1670:81–83
41. Beleza S, Alves C, González-Neira A, Lareu M, Amorim A, Carracedo A, Gusmão L (2003) Extending STR markers in Y chromosome haplotypes. Int J Legal Med 117:27–33
42. Brinkmann B, Klintschar M, Neuhuber F, Huhne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 62:1408–1415
43. Roewer L, Kayser M, de Knijff P, Anslinger K, Betz A, Caglia A, Corach D, Füredi S, Henke L, Hidding M, Kargel HJ, Lessig R, Nagy M, Pascali VL, Parson W, Rolf B, Schmitt C, Szibor R, Teifel-Greding J, Krawczak M (2000) 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 114:31–43
44. Krawczak M (2001) Forensic evaluation of Y-STR haplotype matches: a comment. Forensic Sci Int 118:114–115
45. Applied Biosystems. Y-Filer Haplotype Database. http://www.appliedbiosystems.com/yfilerdatabase, publication date: not available; access date: August 1, 2005
46. Kayser M, Brauer S, Willuweit S, Schadlich H, Batzer MA, Zawacki J, Prinz M, Roewer L, Stoneking M (2002) Online Y-chromosomal short tandem repeat haplotype reference database (YHRD) for U.S. populations. J Forensic Sci 47:513–519
47. Kayser M, Brauer S, Schädlich H, Prinz M, Batzer MA, Zimmerman PA, Boatin BA, Stoneking M (2003) Y chromosome STR haplotypes and the genetic structure of U.S. populations of African, European, and Hispanic ancestry. Genome Res 13:624–634
48. 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, Dobosz T, Dupuy BM, Füredi S, Gehrig C, Gusmão L, Henke J, Henke L, Hidding M, Hohoff C, Hoste B, Jobling M, 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 (2001) Online reference database of Y-chromosomal short tandem repeat (STR) haplotypes. Forensic Sci Int 118:106–113
49. Ploski R, Wozniak M, Pawlowski R, Monies DM, Branicki W, Kupiec T, Kloosterman A, Dobosz T, Bosch E, Nowak M, Lessig R, Jobling MA, Roewer L, Kayser M (2002) Homogeneity and distinctiveness of Polish paternal lineages revealed by Y chromosome microsatellite haplotype analysis. Hum Genet 110:592–600
50. Roewer L, Croucher PJ, Willuweit S, Lu TT, Kayser M, Lessig R, de Knijff P, Jobling MA, Tyler-Smith C, Krawczak M (2005) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. Hum Genet 116:279–291
51. De Knijff P, Kayser M, Caglia 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 (1997) Chromosome Y microsatellites: population genetic and evolutionary aspects. Int J Legal Med 110:134–149
52. Dai HL, Wang XD, Li YB, Wu J, Zhang J, Zhang HJ, Dong JG, Hou YP (2004) Characterization and haplotype analysis of 10 novel Y-STR loci in Chinese Han population. Forensic Sci Int 145:47–55
53. Schmidt U, Meier N, Lutz S (2003) Y-chromosomal STR haplotypes in a population sample from southwest Germany (Freiburg area). Int J Legal Med 117:211–217
54. Butler JM, Decker AE, Vallone PM, Kline MC (2005) Allele frequencies for 27 Y-STR loci with U.S. Caucasian, African-American and Hispanic samples. Forensic Sci Int (in press)