Int J Legal Med (2004) 119 : 1–9
DOI 10.1007/s00414-004-0437-4

ORIGINAL ARTICLE

Harald Niederstätter · Burkhard Berger
Herbert Oberacher · Anita Brandstätter
Christian G. Huber · Walther Parson

Separate analysis of DYS385a and b versus conventional DYS385 typing: is there forensic relevance?

Received: 22 October 2003 / Accepted: 3 March 2004 / Published online: 8 April 2004

Abstract In order to determine to what extent the separate analysis of both copies of DYS385 improves Y-chromosomal short tandem repeat (Y-STR) haplotyping, we followed a recently published protocol for the separate amplification of DYS385a and DYS385b with modifications and compared the results with those obtained by conventional analysis in a population sample comprising 133 unrelated Caucasian males from Austria. Additionally, we typed all markers of the minimal haplotype (minHT) and a set of Y-chromosomal single nucleotide polymorphisms (Y-SNPs) in order to interpret the STR data depending on the Y-SNP haplogroup structure. The separate amplification of DYS385a and b improved the power of discrimination of this marker when compared to the results obtained with the conventional non-locus-discriminating amplification strategy. However, the degree of this improvement varied greatly between different haplogroups and was found to be highest in clade K. In the forensically relevant context of the minHT, the separate analysis of the DYS385 alleles had no effect on the differentiation of paternal lineages in our study. Furthermore, the amplicon lengths of 700–780 base pairs obtained in the course of the locus-discriminating approach restrict the applicability of this amplification strategy to high quality DNA samples.

Keywords Y chromosome · DYS385 · Y-STR · Y-SNP haplogroups · Population study

H. Niederstätter · B. Berger · H. Oberacher · A. Brandstätter
W. Parson (✉)
Institute of Legal Medicine, Medical University of Innsbruck,
Müllerstrasse 44, 6020 Innsbruck, Austria
Tel.: +43-512-5073303, Fax: +43-512-5072764,
e-mail: walther.parson@uibk.ac.at

C. G. Huber
Instrumental Analysis and Bioanalysis,
Saarland University Building 9.2, 66123 Saarbrücken, Germany

Introduction

The major limitation (and also advantage) associated with the use of Y-chromosomal markers in forensic science is that they can discriminate between different paternal lineages but not between males that belong to the same patriline. The reason for this is that the male-specific region of the Y-chromosome (MSY, Skaletsky et al. 2003) does not recombine during meiosis. Hence, the MSY is transmitted along paternal lineages and all markers in it are fully linked. Consequently, all members of a particular lineage show the same allele combination for the set of Y-chromosomal markers used, which has to be considered as a haplotype.

A combination of eight Y-chromosomal short tandem repeat systems (Y-STRs) forms the core set of the markers included in the Y-STR haplotype reference database (YHRD; Roewer et al. 2001; http://www.ystr.org). This allele combination is generally referred to as the "minimal haplotype" (minHT).

In order to improve Y-chromosomal haplotyping, a number of novel Y-STRs have been characterized and studied in different populations (e.g. Beleza et al. 2003; Berger et al. 2003; Butler et al. 2002; Redd et al. 2002; Zarrabeitia et al. 2003). On the other hand, the analysis of the Y-STR marker with the highest gene diversity in the minHT, DYS385, does not utilize the full information content of this duplicated marker when conventional PCR strategies are used, because both copies are amplified simultaneously. This makes an unambiguous assignment of the alleles to the loci DYS385a and DYS385b impossible.

DYS385a and b lie on the long arm of the Y-chromosome in palindrome P4 close (~450 bp) to the proximal and distal inner borders that separate the duplicated arms (190 kb long, 99.97% sequence identity) from the 40 kb interspersed unique spacer-sequence (Kittler et al. 2003; Seo et al. 2003; Skaletsky et al. 2003). This enables the design of locus-specific PCR primers that hybridize to the spacer-sequence. Both DYS385 alleles can be amplified separately when these primers are used in combination with a non-discriminating primer that lies in the duplicated

fragment on the opposite side of the STR (Kittler et al. 2003; Seo et al. 2003). This facilitates the correct assignment of the two alleles to their loci as long as no inversion of the spacer-sequence or the entire Y-chromosomal segment occurs.

The lengths of the PCR products that are obtained with this strategy range between 700 bp and 780 bp. For accurate downstream fragment length analysis, it is therefore necessary to reduce the size of the amplicons. This is achieved by subjecting the PCR fragments to a second round of amplification where both primers are located close to the repeat motif (Kittler et al. 2003; Seo et al. 2003).

The aims of this study were to specify to what extent the separate analysis of both DYS385 copies improves Y-STR haplotyping and to assess to what degree it is influenced by the Y-chromosomal single nucleotide polymorphism (Y-SNP) haplogroup composition of our population sample.

Materials and methods

DNA samples

Blood samples were obtained from 133 unrelated Caucasian males from Tirol (Austria). DNA was extracted and quantified as described previously (Berger et al. 2003).

PCR amplification and electrophoretic separation

All previously unpublished primers that were used in this study were designed with the computer program Primer Express version 1.5 (AB, Applied Biosystems, Foster City, Calif., USA). The sequences and final concentrations in the PCR cocktail are listed in Table 1. The PCR conditions and sequences of the primers used for the amplification of the eight Y-STR systems (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 and DYS385) that represent the core set of the Y-STR haplotype reference database were taken from Parson et al. (2001).

The locus-discriminating primers dys385a and dys385b were designed on the basis of the nucleotide sequences of the clones RP11-143C1 (GenBank entry AC007379) and RP11-569J3 (GenBank entry AC022486), respectively. These two clones belong to the RPCI-11 human BAC library (Osoegawa et al. 1998) which was constructed from blood of a single male donor that belongs to the Y-SNP haplogroup R* in the YCC2003 Tree (Jobling and Tyler-Smith 2003; http://ycc.biosci.arizona.edu). The aligned sequence of the euchromatic regions of the Y-chromosome of this particular man (Skaletsky et al. 2003) reveals 11 and 14 repeats of the GAAA motif for the proximal (5') and distal (3') copies of DYS385, respectively.

In order to prevent confusion we chose to follow the convention of Kittler et al. (2003) for the assignment of alleles to the two DYS385 loci, although it defines the proximal copy as DYS385b and the distal copy as DYS385a. Hence, the DYS385 haplotype of the blood donor for the construction of the RPCI-11 clone library is denoted as 14-11.

The locus-specific first round PCR was performed in 0.2 ml MicroAmp reaction tubes (AB) in a total volume of 20 µl containing 1× PCR buffer (AB), 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 300 nM locus discriminating primer dys385a or dys385b, 300 nM DYS385.2B (Schneider et al. 1998), 10 ng genomic DNA and 2.5 units AmpliTaq Gold DNA polymerase (AB). The amplification reactions were conducted on a GeneAmp PCR System 9700 (AB) by an initial denaturation step at 95°C for 10 min, 30 cycles comprising 95°C for 15 s, 54°C for 1 min and 72°C for 2 min, followed by a final extension step at 72°C for 60 min.

The unpurified first-round amplification products were subjected to separate semi-nested second round PCR amplification in order to reduce the fragment length to approximately 250–300 bp for downstream analysis steps. The reaction set-up for the reamplification of both first round amplicons comprised 1×PCR buffer (AB), 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 200 nM FAM-labelled primer DYS385A (Kayser et al. 1997), 200 nM primer DYS385.2B, 1 µl undiluted first-round PCR product and 1 unit AmpliTaq Gold DNA polymerase (AB) in a total volume of 20 µl. Thermal cycling was performed for 7 cycles comprising 95°C for 15 s, 58°C for 1 min and 72°C for 1 min after an initial denaturation at 95°C for 10 min. The final extension step was conducted at 72°C for 60 min.

Capillary gel electrophoretic separation of the PCR products and data analysis were performed as described by Berger et al. (2003).

Y-chromosomal single nucleotide polymorphism analysis

A set of nine Y-SNPs was analyzed in all samples of our study. The SNPs M9, SRY-1532 (=SRY-$10831_{a/b}$), SRY-4064 (=M40=SRY-8299), 92R7 and Tat (=M46) were analyzed by digestion of fluorochrome-labelled PCR fragments with restriction endonucleases (Hurles et al. 1998, 1999; Santos et al. 1999; Zerjal et al. 1997) followed by laser-induced fluorescence capillary electrophoresis.

The Y-SNPs M17, M35, M213 and P25 were amplified by PCR and subsequently analyzed by mass spectrometry. M213 was amplified using singleplex PCR, the other three Y-SNP markers were amplified together in one multiplex reaction. The PCR was carried out in 0.2 ml MicroAmp reaction tubes (AB) in a total volume of 20 µl containing 1× PCR buffer (AB), 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 5 µg bovine serum albumin, 10 ng genomic DNA and 5 units AmpliTaq Gold DNA polymerase (AB). Thermal cycling was carried out on a GeneAmp PCR system 9700 (AB) for 40 cycles comprising 95°C for 15 s, 64°C for 1 min (multiplex PCR) or 40°C for 1 min (singleplex PCR) and 72°C for 1 min after an initial

Table 1 Sequences of previously unpublished oligonucleotide primers used in amplification reactions and their final concentrations in the PCR cocktail (Conc)

Primer	Conc.	Sequence
dys385a	0.3 µM	5' CATGTGGCTATTGAGCACTTG 3'
dys385b	0.3 µM	5' GTCAGAGACTAGGAATGCAATTTC 3'
M17 F	0.5 µM	5' TCTGTCTACTCACCAGAGTTTGTGGTTGCTG 3'
M17 R	0.5 µM	5' GGCCACTTAACAAACCCCAAAATTCACTTAA 3'
M35 F	0.9 µM	5' AGTGCCCCAATTTTCCTTTGGGACACTG 3'
M35 R	0.3 µM	5' GCAGACTTTCGGAGTCTCTGCCTGTGTC 3'
P25 F	1.0 µM	5' GTAAAGTGAATTATCTGCCTGAAACCTGCCTG 3'
P25 R	1.0 µM	5' GATGGACCGAGATACGAGCACAATTCTATTT 3'
M213 F	0.9 µM	5' TATATCTATTACTATTACATATATTAATAAGAAGTCA 3'
M213 R	0.3 µM	5' TAAAATATTCAGAACTTAAAACATCTCGTTAC 3'

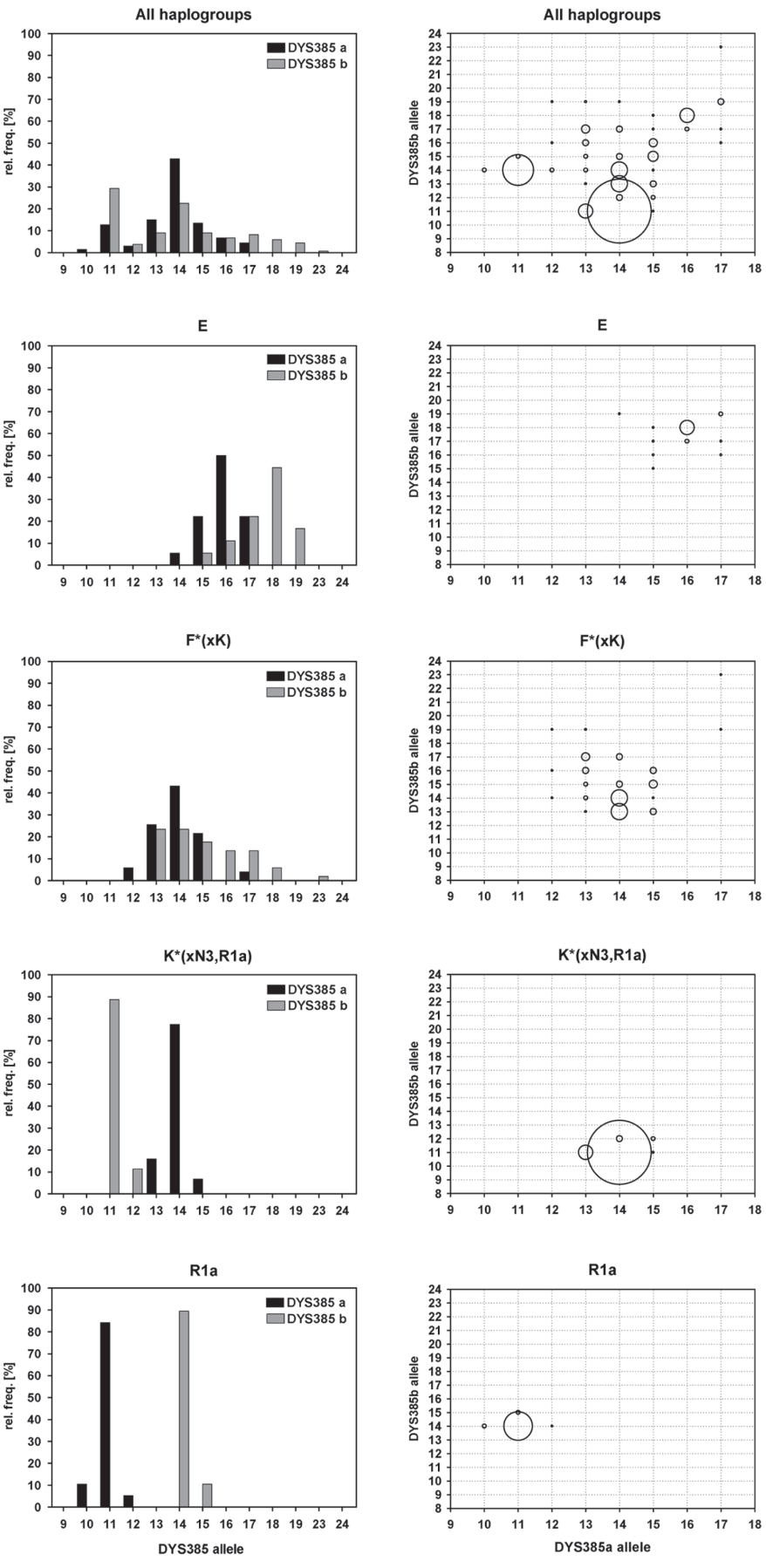

Fig. 1 Relative frequencies of the DYS385 alleles a and b (*bar graphs*) in the entire population sample and several Y-SNP haplogroups. In the bubble plots the centres of the circles depict the different DYS385 haplotypes. The diameters of the circles correspond to the population frequencies of the DYS385 haplotypes

denaturation step at 95°C for 10 min. The final extension step was set to 72°C for 30 min. The amplified fragments were analyzed directly by ion-pair reversed-phase liquid chromatography-electrospray ionization mass spectrometry (ICEMS; Berger et al. 2002; Oberacher et al. 2001; Premstaller et al. 2000).

Nomenclature and statistical analysis

The nomenclature of the Y-STR systems and the Y-SNP haplogroups followed the recommendations of the International Society for Forensic Genetics (ISFG) for the forensic use of Y-STRs (Gill et al. 2001) and the YCC2003 Tree (Jobling and Tyler-Smith 2003; Y Chromosome Consortium 2002), respectively. The combinations of the two alleles DYS385a and DYS385b were denoted as "haplotypes" (e.g. DYS385a-DYS385b: 14-11) when the locus-discriminating protocol was used for amplification. When DYS385a and b were analyzed simultaneously, the DYS385 allele combinations were referred to as "phenotypes". In this particular case, the two alleles were denoted in ascending order of their lengths, separated by a hyphen and labelled with the subscript "a" (e.g. 11-14_a) indicating that this designation does not make allele to locus assignments, as they would be ambiguous. DYS385 phenotypes or haplotypes that differed in length by a single repeat unit in one or both alleles were called "direct neighbours".

Non-"homozygous" DYS385 haplotypes that were indistinguishable from each other by means of the conventional non-locus-discriminating analysis (e.g. 11-14 and 14-11) were referred to as "complementary" (Kittler et al. 2003) throughout the text.

Estimates for the gene diversity (H) values and the sampling variances were computed with the Arlequin software (Schneider et al. 2000).

Results

The conventional amplification strategy for DYS385 allowed the discrimination of 27 different phenotypes (H=0.862±0.026) in the entire population sample. With the separate analysis of the two DYS385 copies this number increased to 34 haplotypes (H=0.919±0.015). For DYS385a the alleles 10–17 and for DYS385b the alleles 11–19 and 23 were found in our population sample (Fig. 1). The most informative marker was DYS385b with a gene diversity of 0.835 followed by DYS385a (H=0.757) and DYS390 (H=0.737). The estimated gene diversities and sampling variances of the other Y-STR markers in the different Y-SNP haplogroups found in our study are listed in Table 2.

In accordance with Kittler et al. (2003), all 15 samples (Table 3) that yielded only a single fluorescent peak when the conventional amplification strategy was used, produced two bands of identical lengths when DYS385a and DYS385b were analyzed separately.

The determination of the Y-SNP haplogroup structure of our population sample revealed that the state of the Y-SNP M9 sorted our samples in two groups of almost equal size, one with a high and the other one with a low variability for the two DYS385 loci (Figs. 1, 2).

In the concatenated Y-SNP haplogroups E and F*(xK), that share the ancestral state of M9, DYS385 was the marker with the highest information content. The diversity of DYS385 haplotypes was only marginally higher than that obtained for the DYS385 phenotypes (Fig. 2). With the conventional PCR protocol we could distinguish

Table 2 Estimated gene diversities (± sampling variances) of the Y-STR markers of the minimal haplotype in the different Y-SNP haplogroups found in this study and in combinations of them

Y-SNP haplogroup	DYS 19	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385a	DYS 385b	n
E*(xE3b)	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	5
E3b	0.385±0.132	0.154±0.126	0.564±0.112	0.500±0.136	0.462±0.110	0.154±0.126	0.154±0.126	0.769±0.061	0.833±0.060	13
F*(xK)	0.633±0.051	0.471±0.064	0.722±0.035	0.700±0.031	0.303±0.073	0.184±0.070	0.700±0.023	0.711±0.037	0.833±0.019	51
K*(xN3,P)	–	–	–	–	–	–	–	–	–	1
N3	–	–	–	–	–	–	–	–	–	1
P*(xR1a,R1b)	0.324±0.081	0.400±0.085	0.617±0.064	0.628±0.041	0.354±0.079	0.180±0.077	0.410±0.090	0.396±0.083	0.215±0.077	42
R1b	–	–	–	–	–	–	–	–	–	1
R1a*	0.000±0.000	0.333±0.215	0.333±0.215	0.333±0.215	0.333±0.215	0.000±0.000	0.000±0.000	0.000±0.000	0.000±0.000	6
R1a1	0.539±0.060	0.385±0.132	0.628±0.143	0.692±0.075	0.539±0.060	0.154±0.126	0.000±0.000	0.410±0.154	0.282±0.142	13
all samples	0.711±0.021	0.548±0.031	0.729±0.017	0.737±0.020	0.508±0.017	0.527±0.033	0.522±0.044	0.757±0.028	0.835±0.018	133
E	0.294±0.119	0.111±0.096	0.451±0.117	0.386±0.128	0.366±0.112	0.111±0.096	0.111±0.096	0.686±0.080	0.752±0.075	18
F	0.657±0.023	0.583±0.028	0.723±0.021	0.760±0.017	0.519±0.015	0.556±0.030	0.568±0.043	0.700±0.034	0.794±0.023	115
K	0.604±0.046	0.389±0.068	0.675±0.035	0.709±0.027	0.446±0.049	0.507±0.050	0.333±0.073	0.642±0.047	0.551±0.051	64
EF*(xK)	0.723±0.026	0.546±0.027	0.758±0.018	0.706±0.022	0.315±0.062	0.165±0.059	0.635±0.039	0.791±0.024	0.867±0.011	69
P	0.603±0.047	0.377±0.069	0.679±0.036	0.700±0.029	0.454±0.048	0.516±0.049	0.318±0.074	0.644±0.049	0.551±0.053	62
P*(xR1a)	0.318±0.080	0.392±0.084	0.608±0.065	0.624±0.042	0.348±0.078	0.176±0.075	0.434±0.088	0.389±0.082	0.210±0.075	43
K*(xN3,R1a)	0.312±0.079	0.415±0.082	0.616±0.062	0.641±0.043	0.341±0.078	0.172±0.074	0.426±0.087	0.382±0.082	0.206±0.074	44
R1a	0.456±0.085	0.351±0.011	0.532±0.130	0.602±0.088	0.515±0.052	0.105±0.092	0 ±0.000	0.292±0.127	0.199±0.112	19

The number of samples found per haplogroup (*n*) is given in the right-most column of the table.

Table 3 Absolute population frequencies (*n*) of samples with different minimal haplotypes (*from left to right* DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385 phenotype) plus the DYS385a and DYS385b alleles and haplogroup (Y-SNP hg) affiliations

Sample	Minimal haplotype	DYS385a	DYS385b	Y-SNP hg	n
GMI_008	13-13-30-24-10-11-13-16-18	16	18	E*(xE3b)	5
GMI_009	13-13-30-24-10-11-13-16-18	16	18	E*(xE3b)	
GMI_010	13-13-30-24-10-11-13-16-18	16	18	E*(xE3b)	
GMI_011	13-13-30-24-10-11-13-16-18	16	18	E*(xE3b)	
GMI_012	13-13-30-24-10-11-13-16-18	16	18	E*(xE3b)	
GMI_038	14-13-29-23-11-13-10-11-14	14	11	P*(xR1a,R1b)	4
GMI_039	14-13-29-23-11-13-10-11-14	14	11	P*(xR1a,R1b)	
GMI_040	14-13-29-23-11-13-10-11-14	14	11	P*(xR1a,R1b)	
GMI_041	14-13-29-23-11-13-10-11-14	14	11	P*(xR1a,R1b)	
GMI_044	14-13-29-23-11-13-13-11-14	14	11	P*(xR1a,R1b)	4
GMI_045	14-13-29-23-11-13-13-11-14	14	11	P*(xR1a,R1b)	
GMI_046	14-13-29-23-11-13-13-11-14	14	11	P*(xR1a,R1b)	
GMI_047	14-13-29-23-11-13-13-11-14	14	11	P*(xR1a,R1b)	
GMI_051	14-13-29-24-11-13-13-11-14	14	11	P*(xR1a,R1b)	4
GMI_052	14-13-29-24-11-13-13-11-14	14	11	P*(xR1a,R1b)	
GMI_053	14-13-29-24-11-13-13-11-14	14	11	P*(xR1a,R1b)	
GMI_054	14-13-29-24-11-13-13-11-14	14	11	P*(xR1a,R1b)	
GMI_125	16-13-30-25-10-11-13-11-14	11	14	R1a*	4
GMI_126	16-13-30-25-10-11-13-11-14	11	14	R1a*	
GMI_127	16-13-30-25-10-11-13-11-14	11	14	R1a*	
GMI_128	16-13-30-25-10-11-13-11-14	11	14	R1a*	
GMI_084	15-12-29-22-10-11-13-14-14	14	14	F*(xK)	3
GMI_085	15-12-29-22-10-11-13-14-14	14	14	F*(xK)	
GMI_086	15-12-29-22-10-11-13-14-14	14	14	F*(xK)	
GMI_022	14-12-28-22-10-11-13-13-15	15	13	F*(xK)	2
GMI_023	14-12-28-22-10-11-13-13-15	15	13	F*(xK)	
GMI_024	14-12-28-22-10-11-14-13-14	14	13	F*(xK)	2
GMI_025	14-12-28-22-10-11-14-13-14	14	13	F*(xK)	
GMI_077	15-12-28-22-10-11-14-15-16	15	16	F*(xK)	2
GMI_078	15-12-28-22-10-11-14-15-16	15	16	F*(xK)	
GMI_042	14-13-29-23-11-13-13-11-13	13	11	P*(xR1a,R1b)	2
GMI_043	14-13-29-23-11-13-13-11-13	13	11	P*(xR1a,R1b)	
GMI_064	14-13-30-24-10-13-13-11-14	14	11	P*(xR1a,R1b)	2
GMI_065	14-13-30-24-10-13-13-11-14	14	11	P*(xR1a,R1b)	
GMI_113	15-14-30-24-11-13-13-12-15	15	12	P*(xR1a,R1b)	2
GMI_114	15-14-30-24-11-13-13-12-15	15	12	P*(xR1a,R1b)	
GMI_003	13-12-29-23-10-11-13-16-17	17	16	E3b	1
GMI_005	13-13-30-23-11-11-13-16-17	16	17	E3b	1
GMI_006	13-13-30-24-10-11-13-15-18	15	18	E3b	1
GMI_007	13-13-30-24-10-11-13-16-17	16	17	E3b	1
GMI_013	13-13-30-24-10-11-13-17-17	17	17	E3b	1
GMI_014	13-13-30-24-11-11-13-17-19	17	19	E3b	1
GMI_015	13-13-30-25-10-11-13-15-17	15	17	E3b	1
GMI_016	13-13-31-23-10-11-12-14-19	14	19	E3b	1
GMI_017	13-13-31-24-10-11-13-16-18	16	18	E3b	1
GMI_019	13-13-31-24-11-11-13-15-16	15	16	E3b	1
GMI_062	14-13-30-24-10-10-13-17-19	17	19	E3b	1
GMI_063	14-13-30-24-10-11-13-16-18	16	18	E3b	1
GMI_068	14-13-31-24-11-11-13-15-15	15	15	E3b	1
GMI_004	13-13-30-22-10-11-12-13-19	13	19	F*(xK)	1
GMI_018	13-13-31-24-10-11-13-17-19	17	19	F*(xK)	1
GMI_020	14-11-27-23-10-11-13-13-15	15	13	F*(xK)	1
GMI_021	14-12-28-22-10-11-12-12-19	12	19	F*(xK)	1
GMI_026	14-12-28-23-10-11-13-13-13	13	13	F*(xK)	1
GMI_027	14-12-28-23-10-11-13-13-14	14	13	F*(xK)	1
GMI_029	14-12-28-23-09-11-12-13-14	14	13	F*(xK)	1
GMI_031	14-12-29-22-10-11-13-13-14	14	13	F*(xK)	1
GMI_032	14-12-30-22-10-11-13-13-14	14	13	F*(xK)	1
GMI_036	14-13-29-22-10-11-14-14-15	15	14	F*(xK)	1
GMI_037	14-13-29-23-10-11-12-13-17	13	17	F*(xK)	1
GMI_061	14-13-30-23-11-11-12-17-23	17	23	F*(xK)	1
GMI_067	14-13-31-23-10-11-12-13-16	13	16	F*(xK)	1
GMI_074	15-12-28-21-10-11-13-14-14	14	14	F*(xK)	1
GMI_075	15-12-28-22-10-11-14-12-16	12	16	F*(xK)	1
GMI_076	15-12-28-22-10-11-14-13-14	13	14	F*(xK)	1
GMI_079	15-12-28-24-10-11-12-14-17	14	17	F*(xK)	1

Table 3 (continued)

Sample	Minimal haplotype	DYS385a	DYS385b	Y-SNP hg	n
GMI_080	15-12-28-24-11-11-12-13-17	13	17	F*(xK)	1
GMI_081	15-12-28-24-11-11-12-14-17	14	17	F*(xK)	1
GMI_083	15-12-28-25-10-11-12-13-17	13	17	F*(xK)	1
GMI_087	15-12-29-22-10-11-14-13-14	13	14	F*(xK)	1
GMI_088	15-12-29-23-10-11-13-14-14	14	14	F*(xK)	1
GMI_089	15-12-29-23-10-11-13-15-15	15	15	F*(xK)	1
GMI_090	15-12-29-23-10-11-14-14-14	14	14	F*(xK)	1
GMI_091	15-12-29-23-10-12-15-14-15	14	15	F*(xK)	1
GMI_092	15-12-29-24-10-11-14-14-14	14	14	F*(xK)	1
GMI_093	15-12-30-21-10-11-14-14-15	14	15	F*(xK)	1
GMI_094	15-12-30-22-10-11-14-14-14	14	14	F*(xK)	1
GMI_096	15-13-29-22-10-11-12-13-15	13	15	F*(xK)	1
GMI_097	15-13-29-22-10-11-13-13-14	14	13	F*(xK)	1
GMI_098	15-13-29-23-11-11-12-13-16	13	16	F*(xK)	1
GMI_108	15-13-31-23-10-12-14-15-15	15	15	F*(xK)	1
GMI_110	15-13-31-24-11-11-13-15-15	15	15	F*(xK)	1
GMI_112	15-14-30-24-10-12-15-15-16	15	16	F*(xK)	1
GMI_115	15-14-32-23-10-11-13-13-16	13	16	F*(xK)	1
GMI_116	16-12-28-24-10-11-12-13-17	13	17	F*(xK)	1
GMI_117	16-12-28-24-11-11-12-14-17	14	17	F*(xK)	1
GMI_118	16-12-29-22-10-10-14-15-15	15	15	F*(xK)	1
GMI_119	16-12-29-23-10-12-14-13-14	14	13	F*(xK)	1
GMI_124	16-13-30-24-11-11-13-14-15	14	15	F*(xK)	1
GMI_134	17-12-29-24-10-11-14-12-14	12	14	F*(xK)	1
GMI_135	17-13-31-24-11-11-13-13-15	13	15	F*(xK)	1
GMI_033	14-12-30-26-11-13-13-11-14	14	11	K*(xN3,P)	1
GMI_100	15-13-29-23-11-13-14-11-14	11	14	N3	1
GMI_001	12-12-28-25-11-13-12-11-14	14	11	P*(xR1a,R1b)	1
GMI_028	14-12-28-23-11-13-13-11-14	14	11	P*(xR1a,R1b)	1
GMI_030	14-12-28-24-10-13-13-11-14	14	11	P*(xR1a,R1b)	1
GMI_034	14-13-28-23-11-11-13-11-13	13	11	P*(xR1a,R1b)	1
GMI_035	14-13-28-24-11-13-13-12-14	14	12	P*(xR1a,R1b)	1
GMI_048	14-13-29-24-10-13-13-11-14	14	11	P*(xR1a,R1b)	1
GMI_049	14-13-29-24-10-14-13-11-14	14	11	P*(xR1a,R1b)	1
GMI_055	14-13-29-24-11-13-13-11-15	15	11	P*(xR1a,R1b)	1
GMI_056	14-13-29-24-11-13-14-11-13	13	11	P*(xR1a,R1b)	1
GMI_057	14-13-29-24-11-14-13-11-13	13	11	P*(xR1a,R1b)	1
GMI_058	14-13-29-25-10-13-13-11-14	14	11	P*(xR1a,R1b)	1
GMI_059	14-13-29-25-11-13-13-11-14	14	11	P*(xR1a,R1b)	1
GMI_060	14-13-30-22-11-13-13-11-13	13	11	P*(xR1a,R1b)	1
GMI_066	14-13-30-25-10-14-13-11-14	14	11	P*(xR1a,R1b)	1
GMI_069	14-13-31-24-11-13-13-11-14	14	11	P*(xR1a,R1b)	1
GMI_070	14-14-30-23-11-13-13-11-13	13	11	P*(xR1a,R1b)	1
GMI_071	14-14-30-23-12-13-13-11-14	14	11	P*(xR1a,R1b)	1
GMI_072	14-14-30-25-11-13-14-11-14	14	11	P*(xR1a,R1b)	1
GMI_073	14-14-31-23-11-13-13-11-14	14	11	P*(xR1a,R1b)	1
GMI_082	15-12-28-24-11-13-13-12-14	14	12	P*(xR1a,R1b)	1
GMI_099	15-13-29-23-11-13-13-11-14	14	11	P*(xR1a,R1b)	1
GMI_101	15-13-29-24-11-13-12-11-14	14	11	P*(xR1a,R1b)	1
GMI_102	15-13-29-24-11-13-14-11-14	14	11	P*(xR1a,R1b)	1
GMI_109	15-13-31-24-10-13-12-12-14	14	12	P*(xR1a,R1b)	1
GMI_121	16-13-30-24-10-11-13-11-14	11	14	R1a*	1
GMI_133	16-14-31-25-11-11-13-11-14	11	14	R1a*	1
GMI_095	15-13-27-25-10-11-13-11-14	11	14	R1a1	1
GMI_103	15-13-29-25-10-11-13-11-15	11	15	R1a1	1
GMI_104	15-13-30-24-10-11-13-12-14	12	14	R1a1	1
GMI_105	15-13-30-26-10-10-13-11-14	11	14	R1a1	1
GMI_106	15-13-30-26-11-11-13-11-14	11	14	R1a1	1
GMI_111	15-13-31-25-10-11-13-11-14	11	14	R1a1	1
GMI_120	16-13-29-25-10-11-13-11-14	11	14	R1a1	1
GMI_122	16-13-30-24-11-11-13-10-14	10	14	R1a1	1
GMI_123	16-13-30-24-11-11-13-11-15	11	15	R1a1	1
GMI_129	16-13-32-25-11-11-13-11-14	11	14	R1a1	1
GMI_130	16-14-30-24-11-11-13-11-14	11	14	R1a1	1
GMI_131	16-14-30-25-11-11-13-11-14	11	14	R1a1	1
GMI_132	16-14-30-26-11-11-13-10-14	10	14	R1a1	1
GMI_050	14-13-29-24-11-13-12-11-14	14	11	R1b	1
Total					133

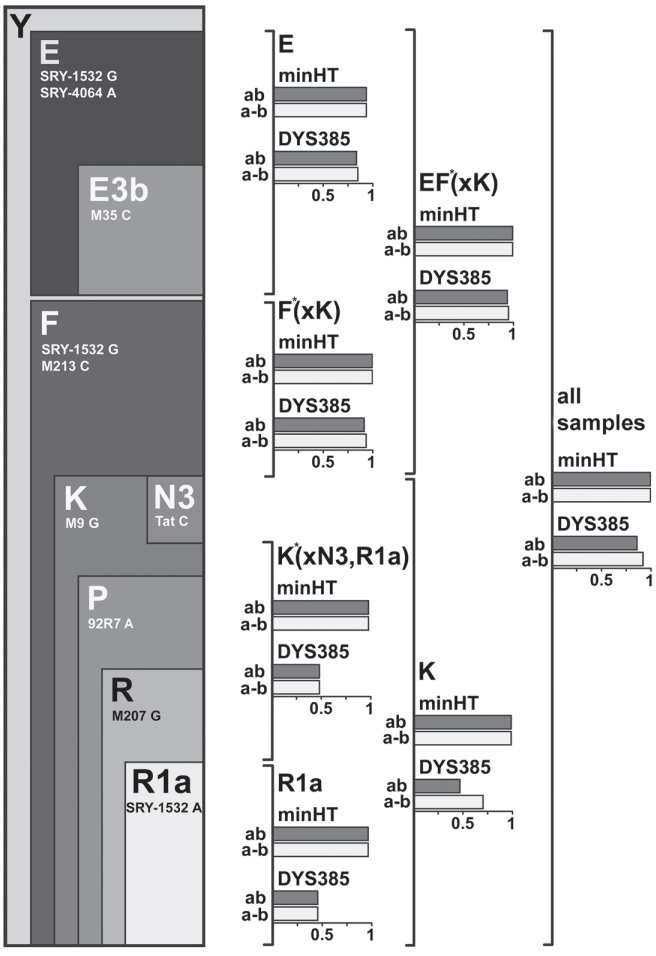

Fig. 2 Schematic representation of the Y-SNP haplogroup topology, nucleotide states of haplogroup-defining markers and bar graphs depicting the gene diversities of DYS385 alone (DYS385) or the minimal haplotype including DYS385 (minHT) in the different clades. The DYS385 alleles were either amplified simultaneously (conventional PCR protocol; ab) or separately (locus-specific PCR protocol; a-b). Please note that the Y-SNP M207 was not typed in this study

between 22 different phenotypes (H=0.942±0.012) whereas the separate analysis of the two alleles allowed the differentiation of 26 haplotypes (H=0.950±0.010). In haplogroups E and F*(xK), the four phenotypes 13-14$_a$, 13-15$_a$, 14-15$_a$ and 16-17$_a$ that comprised complementary haplotypes (e.g. 13-14 and 14-13) accounted for 31.9% of all samples belonging to these two clades. The STRs with the highest gene diversities in EF*(xK) were DYS385b (0.867) and DYS385a (0.791) followed by DYS389II (0.758).

Haplogroup K, which is characterized by the derived state of M9, contained almost 50% of all samples in our study but only 6 out of the 27 different DYS385 phenotypes found. All samples had either the phenotype 11-14$_a$ (population frequency 34.6%) or one of its direct neighbours (Fig. 1). The gene diversity of the DYS385 phenotypes found in haplogroup K was only 0.471±0.073. The occurrence of phenotype 11-14$_a$ and its direct neighbours was restricted to clade K with the exception of a single sample assigned to haplogroup F*(xK) that exhibited the DYS385 phenotype/haplotype 12-14 (Fig. 1, Table 3).

By means of the separate analysis of both DYS385 alleles of the Y-chromosomes attributed to haplogroup K, the modal phenotype 11-14$_a$ was split into the 2 complementary haplotypes 14-11 and 11-14 which have population frequencies of 23.3% and 11.3%, respectively. The DYS385 haplotypes 14-11 and 11-14 plus their direct neighbours accounted for 100% of all haplotypes found in clades K*(xN3,R1a) and R1a, respectively (Fig. 1). The single sample that belonged to haplogroup N3 had the DYS385 haplotype 11-14. Independent of the selected amplification strategy, the diversities of the combined DYS385 alleles in haplogroups K*(xN3,R1a) and R1a were 0.482±0.083 and 0.456±0.132, respectively (Fig. 2). However, with the locus-specific approach the gene diversity of the combined copies of DYS385 in haplogroup K increased from 0.471±0.073 to a value of 0.704±0.047 (Fig. 2). This was the largest gain in discrimination power due to the separate analysis of DYS385a and b we observed in our data set. The most informative marker in haplogroup K was DYS390 followed by DYS389II and DYS385a with gene diversities of 0.709, 0.675 and 0.642, respectively. Within the subclades K*(xN3,R1a) and R1a the gene diversities of DYS385a were 0.382 and 0.292 and for DYS385b they were 0.206 and 0.199, respectively (Table 2).

Both the conventional and the locus-specific strategy for the amplification of DYS385 in combination with the other Y-STRs of the minHT gave identical gene diversities in all Y-SNP haplogroups (Fig. 2). With both PCR approaches we were able to differentiate between 109 haplotypes (H=0.995±0.002) in our data set comprising 133 samples. Therefore, the separate analysis of DYS385a and b – in the context of the minimal haplotype – had no detectable effect on the differentiation of paternal lineages in our population sample.

Discussion

The observed variation of Y-chromosomal markers residing in the MSY is based on the number of different paternal lineages and their relative frequencies in a particular population sample. Therefore, information about the Y-SNP haplogroup structure is desirable when the results of population studies using (novel) Y-STRs are to be compared or conclusions for other populations have to be extrapolated from a given data set. Kittler et al. (2003) reported that the diversity of DYS385a and b depends strongly on the Y-SNP haplogroup structure. Furthermore, they stated that the simultaneous amplification of both DYS385 alleles leads to a population structure-dependent number of erroneous allele to locus assignments. Moreover, this would be deleterious in forensic applications and could be avoided by the separate analysis of both DYS385 alleles. In order to assess the forensic relevance of the locus-specific DYS385 PCR protocol and to compare our results with data from the literature, we typed, in addition to the

Y-STR markers of the minHT, a set of Y-SNPs to gain information about the haplogroup structure of our Austrian population sample.

The separate analysis of DYS385a and b, in order to discriminate between different paternal lineages that are characterized by non-unique "heterozygous" DYS385 phenotypes, is particularly beneficial when the proportion of complementary DYS385 haplotypes with similar frequencies in the data set is high. This becomes obvious when haplogroups K and EF*(xK) are compared. These clades can be seen as models for hypothetical populations with (a) a low DYS385 diversity and a high proportion of phenotypes comprising complementary haplotypes (haplogroup K), (b) a low diversity and a complete lack of complementary DYS385 haplotypes [clades K*(xN3,R1a) and R1a] and (c) a high diversity at the two DYS385 loci and a low to medium amount of DYS385 phenotypes that can be resolved into complementary haplotypes by means of the locus-specific analysis [EF*(xK)].

The locus-discriminating approach revealed that clade K comprised solely the complementary haplotypes 11-14 and 14-11 plus their direct neighbours, which were, in accordance with the findings of Kittler et al. (2003), perfectly associated with the subclades R1a and K*(xN3,R1a), respectively (Fig. 1). The only Y-chromosome that showed the DYS385 haplotype 11-14 outside of clade R1a was attributed to the single sample belonging to haplogroup N3. The ratio of the population frequencies of the complementary haplotypes 14-11 and 11-14 was roughly 2:1 and the separate analysis of both DYS385 alleles resulted in a pronounced increase in the observed diversity of this marker in haplogroup K (Fig. 2, Table 2). However, within its subclades K*(xN3,R1a) and R1a, the locus-discriminating approach did not improve the power of discrimination of DYS385 due to the complete lack of complementary haplotype pairs. This is worth mentioning as the geographical distributions of the haplogroups P*(xR1a) – a subclade of K*(xN3,R1a) – and R1a are strongly clinal in Europe. According to Rosser et al. (2000), haplogroup P*(xR1a) reaches its highest population frequencies in the north-western part of Europe (98.5% in the west of Ireland) and R1a accounts for approximately 50% of all Y-chromosomes found in central eastern Europe. Kittler et al. (2003) reported population frequencies of 65.3% for P*(xR1a) and 0% for R1a in Spain, whereas the respective values found in a sample set from Poland were 12.8% and 61.7%. In Germany and Austria, these two haplogroups comprised 65.7% [P*(xR1a)/R1a: ~1:1; Kittler et al. 2003] and 46.6% [P*(xR1a)/R1a: ~2:1; this study] of the Y-chromosomes analyzed. Consequently, the increase in the discrimination power of DYS385 due to the resolution of the modal phenotype 11-14_a (plus its direct neighbours) into the haplotypes 14-11 and 11-14 and their direct neighbours is expected to be rather low in eastern and western Europe and to be largest in central Europe (e.g. Germany and Austria).

The concatenated haplogroups E and F*(xK) contained complementary DYS385 haplotypes to an extent that was well below that found in haplogroup K (Fig. 1, Table 3). Hence, the separate analysis of the two DYS385 alleles resulted only in a minor increase of the gene diversity (Fig. 2).

From the forensic viewpoint the statement in Kittler et al. (2003) "... the amount of incorrect allele-locus assignment for DYS385 when analysed simultaneously following the conventional DYS385 PCR protocol strongly depends on the population investigated and their hg1:hg3 proportion." needs clarification. According to the ISFG recommendations on Y-STR analysis (Gill et al. 2001; "allele nomenclature, F: duplicated systems such as DYS385 where the observed fragments cannot be assigned unequivocally to a defined genetic locus have to be treated as genotypes and the alleles should be separated by hyphen, e.g. DYS385* 11-14") the DYS385 phenotypes obtained with the conventional PCR protocol cannot be regarded as incorrect since this convention – contrary to the statement in Kittler et al. (2003) – does not assign the shorter allele to DYS385a and the longer allele to DYS385b. Therefore, only a loss of information results from the simultaneous amplification of DYS385a and b, but no population structure-dependent incorrect allele-locus assignments (allegedly ranging from ~13% in Poland up to ~65% in Spain; Kittler et al. 2003) are made.

In forensic applications, DYS385 is rather analyzed in combination with other Y-STRs than alone. Together with seven other markers, DYS385 forms the core set of Y-STR systems included in the YHRD, which comprises with more than 13,000 European and approximately 20,300 haplotypes worldwide, the largest collection of Y-chromosomal STR profiles to date (December 2003). Despite the marked increase in the observed DYS385 diversity in a major haplogroup comprising almost 50% of our samples when the locus-specific approach was used instead of the conventional protocol, the separate analysis of DYS385a and b showed no effect on the differentiation of paternal lineages across all haplogroups when DYS385 was analyzed in combination with the other core loci. Thus, the additional information obtained with the locus-specific amplification of the two DYS385 copies becomes redundant in the context of the minimal haplotype. Therefore, to increase the power of discrimination of the minHT the addition of novel Y-STR markers is clearly superior to the separate analysis of the two DYS385 copies. On the other hand, this demonstrates the insensitivity of the minHT to the loss of information that is associated with the conventional analysis of DYS385, even when large population structure-related differences in the observed DYS385 diversities between the two amplification strategies exist. This is of particular importance as the YHRD contains DYS385 data only on the phenotype level.

The major limitation for the general use of the separate amplification of DYS385 alleles in forensic casework results from the large fragment lengths (700–780 bp) obtained in the first round of PCR amplification. Samples containing degraded DNA, such as the major portion of casework samples, are likely to yield an insufficient amount of PCR product for downstream analysis. The conventional protocol requires only template molecules with

fragment lengths of about 250–300 bp and is therefore considered superior to the separate amplification of the two DYS385 alleles in terms of sensitivity. Furthermore, the nested PCR protocol that has to be followed for the locus-discriminating approach is not suitable for PCR-multiplexing and results in an increased risk of carry-over contamination and sample mix-up due to the extra manipulation steps that are associated with the re-amplification step.

Taking together all the points mentioned above, the conventional PCR protocol for the amplification of DYS385 remains the better choice for routine applications in forensics.

Acknowledgements We thank Alexandra Lindinger for excellent technical assistance and two anonymous reviewers for valuable comments and suggestions regarding the designation of DYS385 phenotypes.

References

Beleza S, Alves C, Gonzalez-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

Berger B, Hölzl G, Oberacher H, Niederstätter H, Huber CG, Parson W (2002) Single nucleotide polymorphism genotyping by on-line liquid chromatography-mass spectrometry in forensic science of the Y-chromosomal locus M9. J Chromatogr B 782: 89–97

Berger B, Niederstätter H, Brandstätter A, Parson W (2003) Molecular characterization and Austrian Caucasian population data of the multi-copy Y-chromosomal STR DYS464. Forensic Sci Int 137:221–230

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

Gill P, Brenner C, Brinkmann B et al. (2001) DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs. Int J Legal Med 114:305–309, Forensic Sci Int 124:5–10

Hurles ME, Irven C, Nicholson J et al. (1998) European Y-chromosomal lineages in Polynesians: a contrast to the population structure revealed by mtDNA. Am J Hum Genet 63:1793–1806

Hurles ME, Veitia R, Arroyo E et al. (1999) Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. Am J Hum Genet 65:1437–1448

Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4:598–612

Kayser M, Caglià A, Corach D et al. (1997) Evaluation of Y-chromosomal STRs: a multicenter study. Int J Legal Med 110: 125–129

Kittler R, Erler A, Brauer S, Stoneking M, Kayser M (2003) Apparent intrachromosomal exchange on the human Y chromosome explained by population history. Eur J Hum Genet 11: 304–314

Oberacher H, Parson W, Mühlmann R, Huber CG (2001) Analysis of polymerase chain reaction products by on-line liquid chromatography-mass spectrometry for genotyping of polymorphic short tandem repeat loci. Anal Chem 73:5109–5115

Osoegawa K, Woon PY, Zhao B, Frengen E, Tateno M, Catanese JJ, Jong PJ de (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 52:1–8

Parson W, Niederstätter H, Köchl S, Steinlechner M, Berger B (2001) When autosomal short tandem repeats fail: optimized primer and reaction design for Y-chromosome short tandem repeat analysis in forensic casework. Croat Med J 42:285–287

Premstaller A, Oberacher H, Huber CG (2000) High-performance liquid chromatography-electrospray ionization mass spectrometry of single- and double stranded nucleic acids using monolithic capillary columns. Anal Chem 72:4386–4393

Redd AJ, Agellon AB, Kearney VA et al. (2002) Forensic value of 14 novel STRs on the human Y chromosome. Forensic Sci Int 130:97–111

Roewer L, Krawczak M, Willuweit S et al. (2001) Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. Forensic Sci Int 118:106–113

Rosser ZH, Zerjal T, Hurles ME et al. (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am J Hum Genet 67:1526–1543

Santos FR, Carvalho-Silva DR, Pena SDJ (1999) PCR-based DNA profiling of human Y chromosomes. In: Epplen JT, Lubjuhn T (eds) Methods and tools in biosciences and medicine. Birkhäuser, Basel, pp 133–152

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

Schneider S, Roessli D, Excoffier L (2000) Arlequin ver. 2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland

Seo Y, Takami Y, Takahama K, Yoshizawa M, Nakayama T, Yukawa N (2003) A method for genotyping Y chromosome-linked DYS385a and DYS385b loci. Leg Med 5:228–232

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ et al. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825–837

Y Chromosome Consortium (2003) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res 12:339–348

Zarrabeitia MT, Riancho JA, Gusmão L, Lareu MV, Sanudo C, Amorim A, Carracedo A (2003) Spanish population data and forensic usefulness of a novel Y-STR set (DYS437, DYS438, DYS439, DYS460, DYS461, GATA A10, GATA C4, GATA H4). Int J Legal Med 117:306–311

Zerjal T, Dashnyam B, Pandya A et al. (1997) Genetic relationships of Asians and northern Europeans, revealed by Y-chromosomal DNA analysis. Am J Hum Genet 60:1174–1183