ELSEVIER

Forensic Science International 130 (2002) 97–111

Forensic Science International

www.elsevier.com/locate/forsciint

Forensic value of 14 novel STRs on the human Y chromosome

Alan J. Redd[a,*], Al B. Agellon[a], Veronica A. Kearney[a], Veronica A. Contreras[a], Tatiana Karafet[a], Hwayong Park[a,c], Peter de Knijff[b], John M. Butler[d], Michael F. Hammer[a]

[a]Division of Biotechnology, University of Arizona, Biosciences West room 239, Tucson, AZ 85721, USA
[b]Forensic Laboratory for DNA Research, MGC-Department of Human and Clinical Genetics, Leiden University Medical Center, RA Leiden 2300, The Netherlands
[c]Genome Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejon 305-333, South Korea
[d]National Institute of Standards and Technology, Biotechnology Division, Gaithersburg, MD 20899, USA

Received 11 February 2002; accepted 10 September 2002

Abstract

We identified and characterized 14 novel short-tandem-repeats (STRs) on the Y chromosome and typed them in two samples, a globally diverse panel of 73 cell lines, and 148 individuals from a European–American population. These Y-STRs include eight tetranucleotide repeats (DYS449, DYS453, DYS454, DYS455, DYS456, DYS458, DYS459, and DYS464), five pentanucleotide repeats (DYS446, DYS447, DYS450, DYS452, and DYS463), and one hexanucleotide repeat (DYS448). Sequence data were obtained to designate a repeat number nomenclature. The gene diversities of an additional 22 Y-STRs, including the most commonly used in forensic databases, were directly compared in the cell line DNAs. Six of the 10 most polymorphic markers include the newly identified Y-STRs. Furthermore, these novel Y-STRs greatly improved the resolution of paternal lineages, above the level obtained with commonly used Y-STRs, in the European–American population.
© 2002 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Humans; Y-chromosome; Short-Tandem-Repeat (STR); Microsatellite; Y-STR; Gene diversity; DYS446; DYS447; DYS448; DYS449; DYS450; DYS452; DYS453; DYS454; DYS455; DYS456; DYS458; DYS459; DYS463; DYS464

1. Introduction

Short-tandem-repeats (STRs) markers on the Y chromosome are valuable tools in sexual assault cases. Sex crime evidence is often made up of sample mixtures of semen from the assailant and cells from the victim. Conventional techniques can separate sperm from the female components, however, complete separation is not always achieved. In addition, sample mixtures that derive from vasectomized or azoospermic males preclude sperm-based separations. Moreover, sample mixtures that are analyzed with autosomal markers via PCR can suffer from competition between a relatively small male DNA component and a large female DNA component. Targeting male-specific polymorphisms on the non-recombining portion of the Y-chromosome (NRY) does not require the separation of sperm from female cells, and thus it improves the likelihood of obtaining male-specific DNA profiles in mixed samples. Y-STR typing will be especially useful in sample mixtures involving: (1) one or more male semen donors, (2) vasectomized or azoospermic men, and (3) the presence of other body–fluid mixtures (e.g. saliva–skin, skin–sweat) from victims and suspects of different sex [1–4]. The primary limitation of Y-STRs in forensic applications is the lack of independence of these markers on the NRY, that is, the absence of recombination. Y-STRs commonly differentiate unrelated Y chromosomes (i.e. *paternal Y lineages*), while autosomal STRs can differentiate any two *individuals* with high statistical confidence. Nonetheless, Y-STRs provide a valuable addition to the forensic scientist's tool kit. As more variable Y-STRs are discovered the potential to distinguish paternal lineages increases.

* Corresponding author. Tel.: +1-520-626-8003; fax: +1-520-626-8060.
E-mail address: ajr@u.arizona.edu (A.J. Redd).

0379-0738/02/$ – see front matter © 2002 Elsevier Science Ireland Ltd. All rights reserved.
PII: S0379-0738(02)00347-X

Approximately 35 Y-STRs have been described to date [1,5–15]. In total, these Y-STRs include: four dinucleotide (YCAI, YCAII, YCAII, and DYS288) six trinucleotide (DYF371, DYS388, DYS392, DYS425, DYS426, DYS436), 23 tetranucleotide (DYS19, DYS385, DYS389I, DYS389AB, DYS390, DYS391, DYS393, DYS434, DYS435, DY437, DYS439, DYS441, DYS442, DYS443, DYS444, DYS445, DYS460, DYS461, DYS462, G10123, A10, C4, and H4), and two pentanucleotide repeats (DXYS156Y, and DYS438). Most of the Y-STR primer pairs amplify one PCR product, while some Y-STR primer pairs amplify two or more PCR products (YCAI, YCAII, YCAIII, DYF371, DYS385, G10123). We refer to the former Y-STRs as single-copy, and to the latter as multi-copy because they are present in more than one copy on the NRY.

With the exception of one multicenter-study using 13 Y-STRs [1], global population screens using the majority of the published Y-STRs have not been performed. One recent study [15] compared a collection of 19 Y-STRs in a large sample of individuals from the Iberian Peninsula. Much progress has been made in establishing a large database of many European populations using the following Y-STRs: DYS19, DYS385, DYS389I, DYS389II-I, DYS390, DYS391, DYS392, and DYS393. These Y-STRs define the "minimal haplotype", while the addition of YCAII to the these Y-STRs has been termed the "extended haplotype" [1,16,17] (Y-STR Haplotype Reference Database http://www.ystr.charite.de/index_gr.html).

The identification of additional Y-STRs is warranted for several reasons. First and foremost, increasing the number of highly polymorphic markers will improve the ability to distinguish paternal lineages. There are shared Y-STR haplotypes in populations because either males share identity by descent or because a particular set of Y-STRs does not distinguish closely related, but different, paternal lineages. In a sample of 41 European populations the discrimination capacities were 52% ($n = 4688$ individuals) and 71% ($n = 1957$ individuals) using the minimal and extended sets of Y-STRs, respectively [17]. The discrimination capacity in the Iberian Peninsula study [15] was 83% using 19 Y-STRs. The sharing of paternal lineages is likely to be more common in isolated populations where there is a higher degree of genetic drift, such as Native American populations. Second, there is a need to identify more Y-STRs that have longer repeats units. For example, YCAII is a polymorphic dinucleotide marker that suffers from "stutter" products during the PCR process due to polymerase slippage [18]. Stutter products are pronounced in dinucleotide repeats. Stutter bands are often reduced in longer repeat motif STRs, these loci can provide additional resolution in sample mixtures of multiple-male DNA profiles [19]. Third, a large pool of Y-STRs will provide a diverse sample of markers from which one can select tailored sets of STRs with distinct characteristics for multiplex design for particular applications. A small multiplex of the most informative Y-STRs could more efficiently distinguish Y-chromosome lineages than a set of a dozen or more less informative Y-STRs. Finally, increasing the number of Y-STRs will improve the estimation of the time to the most recent common ancestor (TMRCA). The TMRCA between two Y-STR haplotypes provides a natural metric to describe the relatedness between two individuals and could be used to make exclusions in forensics [20]. By including more Y-STRs, estimates of the TMRCA become more precise [21,20]and the ability to exclude paternal relatives increases.

Here we describe variation among 14 novel Y-STRs and make comparisons with 22 previously identified Y-STRs. This is accomplished by comparing the gene diversities of these 36 Y-STRs in the same panel of 73 cell line DNAs. In addition, we demonstrate the forensic value of the new Y-STRs in a European–American population from South Dakota.

2. Materials and methods

2.1. DNA samples

A repository of 74 lymphoblastoid cell lines was established in 1991 in collaboration with Dr. N Ellis of the Sloan Kettering Cancer Research Institute. The geographic and ethnic origins of these Y-Chromosome Consortium (YCC) cell lines were previously described [22] (http://www.ycc.biosci.arizona.edu/nomenclature_system/table1.html). The donor males include 11 !Kung, 5 Pygmy, 10 Bantu speakers, 14 Europeans, 5 Middle Easterners, 3 Pakistanis, 3 Chinese, 3 Japanese, 5 North Asians, 1 Cambodian, 2 Melanesians, and 12 Native Americans. We genotyped 73 of these 74 samples. Two additional female cell lines (YCC1 and YCC54) were included in the genotyping. For examining geographic variation in gene diversity we combined our samples into the following three groups: Africans (AFR, $n = 25$), Europeans/Middle Easterners (EUR/MEA, $n = 22$), and Asians (ASN, $n = 26$). The EUR/MEA group included Middle Easterners and Pakistanis, while the ASN group included Melanesians and Native Americans.

Blood samples from 148 European–American ("Caucasian") individuals were provided by Rex Riis of the State of South Dakota Forensic Laboratory. An amount of 500 μl of blood was digested with Proteinase K overnight, followed by a standard Phenol–Chloroform extraction using Eppendorf Phase-Lock Gel™ from Eppendorf Scientific, Inc. DNA was quantified using a FLx800 Microplate Fluorescence Reader from Bio-Tek Instruments Inc.

2.2. Identification of novel Y-STRs

The human Y chromosome draft sequence in GenBank [23] was searched for novel Y-STRs using the programs Tandem Repeats Finder [24] (v. 2.02), and RepeatMasker (v. 04/04/2000, A. Smit http://ftp.genome.washington.edu/RM/RepeatMasker.html). We searched for Y-STRs with stretches

of nine or more repeats, with a focus on tetranucleotide, pentanucleotide, and hexanucleotide repeats.

2.3. Primers

Primers were designed to have similar annealing temperatures using OLIGO primer analysis software (P Rychlik and W Rychlik; v 6.1). All of the primers were submitted to the BLAST [25] facility at NCBI to examine them for uniqueness and for NRY specificity. Table 1 lists the primers used for typing the 14 novel Y-STRs reported here. The forward primers were 5′ end-labeled with either 6FAM or HEX.

2.4. Amplification conditions

2.4.1. Uniplex PCR

For each of the novel Y-STRs (DYS446-450; DYS452-456, DYS458-459; DYS463-464) a uniplex PCR was performed in a 15 μl reaction in a MJ Research PTC-100 thermocycler. The PCR reactions included: 10 ng of DNA, 1.5 μl of 10X PCR buffer (100 mM Tris–HCL; pH 8.3,8.8 or 9.2; 250, 500 or 750 mM KCl; 15, 25, or 35 mM $MgCl_2$; see Table 1 for locus-specific buffer compositions), 0.5 units of Taq from Qiagen or Gibco (now Invitrogen Life Technologies), 0.11 ng of Clonetech antibody (BD Biosciences), 0.8 μl of Clonetech Antibody Dilution Buffer (50 mM KCl, 10 mM Tris–HCl, pH 7.0, 50% glycerol), 200 μM of each dNTP, and 0.13 μM of each primer. Initial denaturation at 94 °C for 3 min was followed by 30 cycles of 94 °C for 30 s, 56 °C for 30 s, 72 °C for 35 s, followed by a final extension at 72 °C for 3 min. Two female YCC samples were included to test for male specificity.

DYS434 and DYF371 were amplified according to the conditions described above. DYS438 was amplified using a touchdown PCR profile: annealing was at 60 °C (−0.5 °C for each of eight cycles) for 30 s, followed by an annealing

Table 1
Primers sequences and Y-STR characteristics in YCC panel

Y-STR	Primer sequence	Repeat motif	Repeat size[a]	Size (bp)[b]	10X PCR buffer[c]
DYS446	F-TATTTCAGTCTTGTCCTGTC; R-AAATGTATGGCCAACATAGCAAAACCA	$[TCTCT]_n$	10–18	288–328	*A–C–C*
DYS447	F-GGTCACAGCATGGCTTGGTT; R-GGGCTTGCTTTGCGTTATCT	$[TAATA]_n[TAAAA]_1$; $[TAATA]_n[TAAAA]_1[TAATA]_n$	22–29	206–241	*A–C–A*
DYS448	F-TGTCAAAGAGCTTCAATGGAGA; R-TCTTCCTTAACGTGAATTTCCTC	$[AGAGAT]_nN_{42}[AGAGAT]_n$	17–23	282–318	*A–C–A*
DYS449	F-TGGAGTCTCTCAAGCCTGTTCTA; R-CCTGGAAGTGGAGTTTGCTGT	$[TTTC]_nN_{50}[TTTC]_n$	26–33; 35–36	343–383	*A–B–B*
DYS450	F-CCAGTGATAATTCAGATGATATG; R-GCCTTTCCAATTTCAATTTCTGA	$[TTTTA]_n$	8–11	356–371	*A–A–C*
DYS452	F-GTGGTGTTCTGATGAGGATAAT; R-TTTACATGATGTAGCAAATAGGTT	$[TATAC]_2[TGTAC]_2[TATAC]_n$; $[CATAC]_1[TATAC]_1[CATAC]_1$; $[TATAC]_3[CATAC]_2[TATAC]_3$; $[CATAC]_1[TATAC]_3$	27–33	221–251	*A–B–C*
DYS453	F-GGGTAACAGAACAAGACAGT; R-CTAAAAGTATGGATATTCTTCG	$[AAAT]_n$	9–13	123–139	*C–A–C*
DYS454	F-GACATGTAGCTCTTCACTTCAC; R-GACTGACCTCACATTGTTGTTA	$[AAAT]_n$	10–12	196–204	*A–B–B*
DYS455	F-ATCTGAGCCGAGAGAATGATA; R-GGGGTGGAAACGAGTGTT	$[AAAT]_n$	8–12	168–184	*A–B–C*
DYS456	F-GGACCTTGTGATAATGTAAGATA; R-CCCATCAACTCAGCCCAAAAC	$[AGAT]_n$	13–18	141–161	*B–B–C*
DYS458	F-AGCAACAGGAATGAAACTCCAAT; R-CCACCACGCCCACCCTCC	$[GAAA]_n$	13–20	111–139	*A–C–C*
DYS459	F-CAGGTGAACTGGGGTAAATAAT; R-TTGAGCAACAGAGCAAGACTTA	$[TAAA]_n$	7–10	140–152	*A–B–B*
DYS463	F-AATTCTAGGTTTGAGCAAAGACA; R-ATGAGGTTGTGTGACTTGACTG	$[AAAGG]_n[AAGGG]_n[AAGGA]_2$	18; 20–27	224–269	*A–B–A*
DYS464	F-TTACGAGCTTTGGGCTATG; R-CCTGGGTAACAGAGAGACTCTT	$[CCTT]_n$	11–19	250–282	*C–A–A*

[a] Repeat size includes adjacent variant and invariant tandem-repeat blocks and was determined by sequencing (see Table 2).
[b] The sequence and the measured size differ slightly; sequenced controls were used for sizing.
[c] The letters *A–B–C* refer to the pH ($A = 8.3$, $B = 8.8$, and $C = 9.2$), KCl ($A = 250$ mM, $B = 500$ mM, and $C = 750$ mM), and $MgCl_2$ ($A = 15$ mM, $B = 25$ mM, and $C = 35$ mM) of the 10X PCR buffers, respectively.

temperature of 56 °C for 30 cycles. DXYS156 was amplified and genotyped following published methods [9]. Four Y-STRs (DYS460, DYS461, DYS462, and G10123) were typed in the laboratory of Peter de Knijff. These four Y-STRs were amplified in uniplex reactions. The 25-μl PCR reactions included: 5 ng of DNA, 2.5 μl of GENEAMP PCR buffer from Perkin-Elmer (100 mM Tris–HCL; pH 8.3, 500 KCl; 15 mM $MgCl_2$), one unit of Taq from Applied Biosystems, 200 μM of each dNTP, and 0.13 μM of each primer. Initial denaturation at 94 °C for 2 min was followed by 5 cycles of 94 °C for 15 s, 58 °C for 15 s, 72 °C for 20 s, followed by 30 cycles of 94 °C for 15 s, 54 °C for 15 s, 72 °C for 20 s, and a final extension at 72 °C for 10 min.

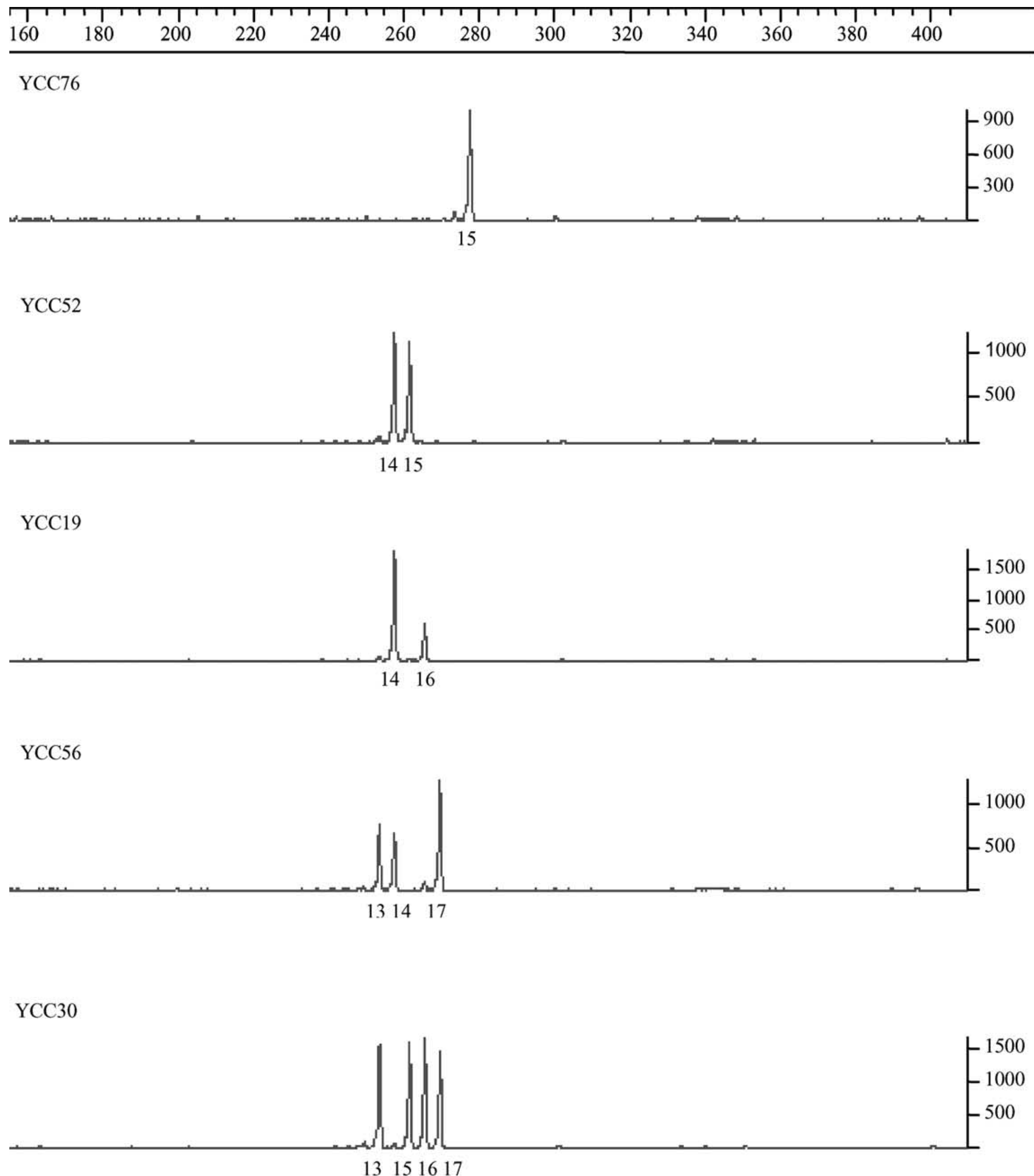

Fig. 1. Electropherogram for DYS464 in five YCC samples. The DYS464 primers target four duplicated regions on the NRY and the relative peak heights of the PCR products are consistent with four independent amplifications. An alternative method of genotype assignment than those shown would treat peak height as a quantitative indicator of copy number such that: YCC76 = 15,15,15,15; YCC52 = 14,14,15,15; YCC19=14,14,14,16; YCC56 = 13,14,17,17; and YCC30 = 13,15,16,17.

2.4.2. Multiplex PCR

Multiplex I includes the following five Y-STRs: DYS394 (DYS394 primer sequences are alternative primers for DYS19 primer sequences that amplify the DYS19 locus), DYS390, DYS391, DYS393, and DYS385. The 11-μl PCR reactions contained: 10 ng of DNA, 1.0 μl of 10X PCR buffer (100 mM Tris–HCl; pH 8.3, 750 mM KCl; 35 mM $MgCl_2$), 0.2 units of Taq (Qiagen or Gibco), 0.044 ng of Clonetech antibody, 0.32 μl of Clonetech Antibody Dilution Buffer (50 mM KCl, 10 mM Tris–HCl, pH 7.0, 50% glycerol), 182 μM of each dNTP, 0.34 μM DYS394 Forward (F) and Reverse (R) primer sequences, 0.06 μM of DYS390 F and R, 0.05 μM of DYS391 F and R, 0.05 μM of DYS393 F and R, and 0.06 μM of DYS385 F and R. The forward primer sequences of DYS394, DYS391, and DYS385 were 5′ end-labeled with the fluorescent dye 6FAM, while forward primer sequences for DYS393 and DYS390 were 5′ end-labeled with HEX. PCR cycling conditions were the same as described above in the uniplex PCR for the novel Y-STRs except that the annealing temperature was 57 °C and the extension time was 30 s.

Multiplex II includes the following seven Y-STRs: DYS19, DYS388, DYS389(I and II), DYS392, DYS426, and DYS439. DYS19 was included in multiplex II as means of insuring sample continuity between PCR reactions in multiplex I and II. The 9-μl PCR reactions contained: 10 ng of DNA, 1.0 μl of 10X PCR buffer (100 mM Tris–HCl; pH 8.3, 500 mM KCl; 35 mM $MgCl_2$), 0.25 units of Taq (Qiagen or Gibco), 0.055 ng of Clonetech antibody, 0.4 μl of Clonetech Antibody Dilution Buffer (50 mM KCl, 10 mM Tris–HCl, pH 7.0, 50% glycerol), 222 μM of each dNTP, 0.44 μM DYS19 F and R, 0.08 μM DYS388 F and R, 0.22 μM DYS389 F and R, 0.10 μM of DYS392 F and R, 0.04 μM of DYS426 F and R, and 0.18 μM of DYS439 F and R. The forward primer sequences of DYS19, DYS388, and DYS426 were 5′ end-labeled with the fluorescent dye 6FAM; the forward primer sequences for DYS392 and DYS439 were 5′ end-labeled with NED; and the forward primer sequences for DYS389 (I and II) was 5′ end-labeled with HEX. We used novel primer sequences for DYS439 (F was AATTAATAGATTCAAGGTGA, and R was CCCATCATCTCTTTACTATT) because we re-discovered this Y-STR in our search for novel Y-STRs (see Section 3.1 further). PCR cycling conditions were the same as described above in the uniplex PCR for the novel Y-STRs except that the denaturation time was 25 s and the extension time was 30 s.

Multiplex III includes the following nine Y-STRs: DYS447, DYS448, DYS449, DYS454, DYS455, DYS457 (DYS457 primers are alternate primer sequences for the DYS437 primer sequences that amplify the DYS437 locus; DYS457 F was TGCAGCCTCAATTTCCTGGT, and R was TATAGATAGATAGATAACCACAG), DYS458, DYS459, and DYS464. The 15-μl PCR reactions contained: 10 ng of DNA, 1.0 μl of 10X PCR buffer (100 mM Tris–HCl; pH 8.3, 500 mM KCl; 20 mM $MgCl_2$), 0.5 units of Taq (Qiagen or Gibco), 0.066 ng of Clonetech antibody, 0.8 μl of Clonetech Antibody Dilution Buffer (50 mM KCl, 10 mM Tris–HCl, pH 7.0, 50% glycerol), 200 μM of each dNTP, 0.20 μM DYS447 F and R, 0.20 μM DYS448 F and R, 0.20 μM DYS449 F and R, 0.20 μM of DYS454 F and R, 0.17 μM of DYS455 F and R, 0.13 μM of DYS457 F and R, 0.09 μM of DYS458 F and R, 0.13 μM of DYS459 F and R, 0.13 μM of DYS464 F and R. The forward primer sequences of DYS447, DYS448, DYS449, DYS458, and DYS459 were 5′ end-labeled with the fluorescent dye 6FAM; the forward primer sequences for DYS454, DYS455, DYS457, and DYS464 were 5′ end-labeled with HEX. PCR cycling conditions were the same as described above in the uniplex PCR for the novel Y-STRs.

Multiplex IV includes the following four Y-STRs: DYS446, DYS452, DYS453, and DYS456. The 15-μl PCR reactions contained: 10 ng of DNA, 1.0 μl of 10X PCR buffer (100 mM Tris–HCl; pH 8.3, 750 mM KCl; 35 mM $MgCl_2$), 0.5 units of Taq (Qiagen or Gibco),

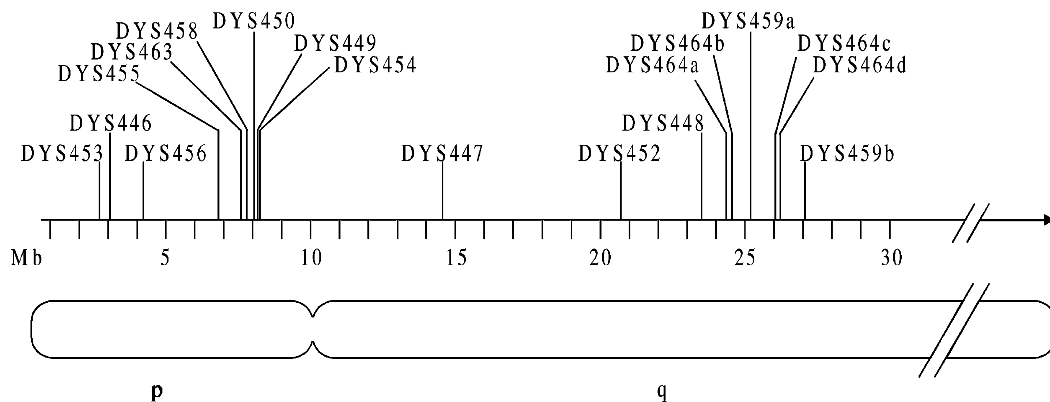

Fig. 2. Approximate locations of 14 novel Y-STRs on the Y chromosome. The physical locations were determined from BLAT searches (see http://www.ucsc.abcc.ncifcrf.gov/) that included the Y-STR sequences and about 100 bp of flanking sequence. Note that DYS459 (a and b) and DYS464 (a, b, c, and d) are located in two and four positions on the map, respectively.

0.066 ng of Clonetech antibody, 0.8 μl of Clonetech Antibody Dilution Buffer (50 mM KCl, 10 mM Tris–HCl, pH 7.0, 50% glycerol), 200 μM of each dNTP, 0.20 μM DYS446 F and R, 0.20 μM DYS452 F and R, 0.20 μM DYS453 F and R, 0.20 μM of DYS456 F and R. The forward primer sequences of DYS446, DYS452, DYS456 were 5′ end-labeled with the fluorescent dye 6FAM; the forward primer sequence for DYS453 was 5′ end-labeled with HEX. PCR cycling conditions were the same as described above in the uniplex PCR for the novel Y-STRs.

DYS446

AC006152.3 F.cca**(TCTCT)**$_{14}$tctcctctcttcagacagagtctcactctgtcgcccaggcttgagtgcagtggtacaagct cagctcactgcaacctccatctcccaggttcaagtgattctcctgcctcagcctcctgagtatctggaattaac agcatgcaccaccacgcccagctaatgtttttgtatttttttttgtagagatg.R

Allele (bp)	YCC ID	
12(298)	YCC19	24bp **(TCTCT)**$_{12}$ 214bp
16(318)	YCC24	24bp **(TCTCT)**$_{16}$ 214bp
13(303)	YCC26	24bp **(TCTCT)**$_{13}$ 214bp
13(303)	YCC33	24bp **(TCTCT)**$_{13}$ 214bp

DYS447

AC005820 F.ttatacattttagggagacataaggcat**(TAATA)**$_{6}$**(TAAAA)**$_{1}$**(TAATA)**$_{9}$ **(TAAAA)**$_{1}$**(TAATA)**$_{6}$taaacattagttctgtccagaaaggcag.R

Allele (bp)	YCC ID	
23(216)	YCC19	48bp **(TAATA)**$_{6}$(TAAAA)$_{1}$**(TAATA)**$_{12}$(TAAAA)$_{1}$**(TAATA)**$_{3}$ 48bp
26(221)	YCC24	48bp **(TAATA)**$_{6}$(TAAAA)$_{1}$**(TAATA)**$_{12}$(TAAAA)$_{1}$**(TAATA)**$_{6}$ 48bp
24(216)	YCC26	48bp **(TAATA)**$_{7}$(TAAAA)$_{1}$**(TAATA)**$_{16}$ --------------- 48bp
25(221)	YCC33	48bp **(TAATA)**$_{7}$(TAAAA)$_{1}$**(TAATA)**$_{08}$(TAAAA)$_{1}$**(TAATA)**$_{8}$ 48bp

DYS448

AC025227.6 F.ttagaaatagagatcgcgagacagaaagggagatagagacatggataa**(AGAGAT)**$_{11}$atagagatag**(AGAGAT)**$_{3}$ agatagatagagaa**(AGAGAT)**$_{8}$agagaggtaaagatagagataaatttccagaccggccagaaatat.R

Allele (bp)	YCC ID	
20(300)	YCC19	70bp **(AGAGAT)**$_{11}$atagagatag(AGAGAT)$_{3}$agatagatagagaa**(AGAGAT)**$_{9}$ 68bp
21(306)	YCC24	70bp **(AGAGAT)**$_{12}$atagagatag(AGAGAT)$_{3}$agatagatagagaa**(AGAGAT)**$_{9}$ 68bp
19(294)	YCC26	70bp **(AGAGAT)**$_{11}$atagagatag(AGAGAT)$_{3}$agatagatagagaa**(AGAGAT)**$_{8}$ 68bp
21(306)	YCC33	70bp **(AGAGAT)**$_{13}$atagagatag(AGAGAT)$_{3}$agatagatagagaa**(AGAGAT)**$_{8}$ 68bp

DYS449

A051663.9 F.tgaatattttcccttaacttgtgtgatttttttgtttaaaaagttcccccatgtttttttcttgctctt tttcttttctctctt**(TTTC)**$_{15}$tctctctcctcctc**(TTTC)**$_{2}$cttc**(TTTC)**$_{2}$ttttcctctttccttc **(TTTC)**$_{14}$tctctttgtttctgtcttactcttgttgtccaacctagagtgcaatggtgcaatcttagctt.R

Allele (bp)	YCC ID	
28(351)	YCC19	106bp **(TTTC)**$_{12}$tctctctcctcctc(TTTC)$_{2}$cttc(TTTC)$_{2}$ttttcctctttccttc**(TTTC)**$_{16}$ 83bp
28(351)	YCC24	106bp **(TTTC)**$_{14}$tctctctcctcctc(TTTC)$_{2}$cttc(TTTC)$_{2}$ttttcctctttccttc**(TTTC)**$_{14}$ 83bp
29(355)	YCC26	106bp **(TTTC)**$_{15}$tctctctcctcctc(TTTC)$_{2}$cttc(TTTC)$_{2}$ttttcctctttccttc**(TTTC)**$_{14}$ 83bp
31(363)	YCC33	106bp **(TTTC)**$_{16}$tctctctcctcctc(TTTC)$_{2}$cttc(TTTC)$_{2}$ttttcctctttccttc**(TTTC)**$_{15}$ 83bp

DYS450

AC051663 F.agagaccagcttggtatttctattgcagcattttgagagaacgtgaaacataggtgtgtcctgtt**(GT)**$_{8}$ aagcatgtatacataaccggttgaacaaactaggatcactcacacaatggaatatgatgcagctgtttgtaga tctggtagaaattattctgaatctatcggttcctgggccgtt**(TTTTA)**$_{9}$tttcattttata**(TTTTA)**$_{3}$t **(TTTTA)**$_{2}$tttttggctggtgggttatttattactgattccatt.R

Allele (bp)	YCC ID	
10(366)	YCC19	219bp **(TTTTA)**$_{10}$tttcattttata(TTTTA)$_{3}$t(TTTTA)$_{2}$ 59bp
9(361)	YCC24	219bp **(TTTTA)**$_{09}$tttcattttata(TTTTA)$_{3}$t(TTTTA)$_{2}$ 59bp
9(361)	YCC26	219bp **(TTTTA)**$_{09}$tttcattttata(TTTTA)$_{3}$t(TTTTA)$_{2}$ 59bp
8(356)	YCC33	219bp **(TTTTA)**$_{08}$tttcattttata(TTTTA)$_{3}$t(TTTTA)$_{2}$ 59bp

Fig. 3. Sequence data and allele nomenclature for novel Y-STRs: DYS446, DYS447, DYS448, DYS449, DYS450, DYS452, DYS453, DYS454, DYS455, DYS456, DYS458, DYS459, DYS463, and DYS464. Repeat units are shown in uppercase bold in reference sequence. The subscript number outside the repeat unit parentheses indicates the number of consecutive repeats. Variant regions are bold in sequenced samples.

DYS452

	AC010137.3	F.t**(TATAC)$_2$(TGTAC)$_2$(TATAC)$_{12}$(CATAC)$_1$(TATAC)$_1$(CATAC)$_1$(TATAC)$_3$(CATAC)$_2$ (TATAC)$_3$(CATAC)$_1$(TATAC)$_3$**aaccaattaattagctgagtataataaatctttaaaaa.R
Allele (bp)	YCC ID	
31(241)	YCC19	23bp (TATAC)$_2$(TGTAC)$_2$**(TATAC)$_{12}$**(CATAC)$_1$(TATAC)$_1$(CATAC)$_1$**(TATAC)$_3$(CATAC)$_2$(TATAC)$_3$**(CATAC)$_1$(TATAC)$_3$ 63bp
28(226)	YCC24	23bp (TATAC)$_2$(TGTAC)$_2$**(TATAC)$_{13}$**(CATAC)$_1$(TATAC)$_1$(CATAC)$_1$**(TATAC)$_4$**--------------(CATAC)$_1$(TATAC)$_3$ 63bp
30(236)	YCC26	23bp (TATAC)$_2$(TGTAC)$_2$**(TATAC)$_{11}$**(CATAC)$_1$(TATAC)$_1$(CATAC)$_1$**(TATAC)$_3$(CATAC)$_2$(TATAC)$_3$**(CATAC)$_1$(TATAC)$_3$ 63bp
33(251)	YCC33	23bp (TATAC)$_2$(TGTAC)$_2$**(TATAC)$_{14}$**(CATAC)$_1$(TATAC)$_1$(CATAC)$_1$**(TATAC)$_3$(CATAC)$_2$(TATAC)$_3$**(CATAC)$_1$(TATAC)$_3$ 63bp

DYS453

	AC006157.2	F.gtctcaaa**(AAAT)$_{11}$**aaaataagctatctgcagggctggaggctcttgactt.R
Allele (bp)	YCC ID	
12(135)	YCC19	28bp **(AAAT)$_{12}$** 59bp
11(131)	YCC24	28bp **(AAAT)$_{11}$** 59bp
11(131)	YCC26	28bp **(AAAT)$_{11}$** 59bp
11(131)	YCC33	28bp **(AAAT)$_{11}$** 59bp

DYS454

	AC025731.12	F.agcccagcaacatatcacaatctccctgtggtcggggcacaggcaaaagca**(AAAT)$_{11}$** aacctaggtgctaatccaagtgatatgttacaatgtttcctgttgacacaacccaacctggg.R
Allele (bp)	YCC ID	
12(204)	YCC19	73bp **(AAAT)$_{12}$** 83bp
11(200)	YCC24	73bp **(AAAT)$_{11}$** 83bp
11(200)	YCC26	73bp **(AAAT)$_{11}$** 83bp
11(200)	YCC33	73bp **(AAAT)$_{11}$** 83bp

DYS455

	AC012068	F.ctgcctaagcccacaaggtcaaggctgcagtgagctgtgatcacccgagggca ctccagcctgggcaacactgtgagaccatatatcta**(AAAT)$_{11}$**aacggaag.R
Allele (bp)	YCC ID	
10(176)	YCC19	110bp **(AAAT)$_{10}$** 26bp
11(180)	YCC24	110bp **(AAAT)$_{11}$** 26bp
11(180)	YCC26	110bp **(AAAT)$_{11}$** 26bp
11(180)	YCC33	110bp **(AAAT)$_{11}$** 26bp

DYS456

		F (23mer) has first five base pairs of repeat
	AC010106.2	F.**(AGAT)$_{15}$**attccattagttctgtccctctagagaaccctaatacatcagtttaagaa.P2
Allele (bp)	YCC ID	
14(145)	YCC19	18bp **(AGAT)$_{14}$** 71bp
16(153)	YCC24	18bp **(AGAT)$_{16}$** 71bp
17(157)	YCC26	18bp **(AGAT)$_{17}$** 71bp
15(149)	YCC33	18bp **(AGAT)$_{15}$** 71bp

DYS458

	AC010902.4	F.**(GAAA)$_3$**aggaag**(GAAA)$_{16}$**.R
Allele (bp)	YCC ID	
15(119)	YCC19	41bp (GAAA)$_3$aggaag**(GAAA)$_{15}$** 18bp
17(127)	YCC24	41bp (GAAA)$_3$aggaag**(GAAA)$_{17}$** 18bp
16(123)	YCC26	41bp (GAAA)$_3$aggaag**(GAAA)$_{16}$** 18bp
16(123)	YCC33	41bp (GAAA)$_3$aggaag**(GAAA)$_{16}$** 18bp

Fig. 3. (*Continued*).

A beta-test of the Y-20plex [26] was performed using the YCC panel in the laboratory of Michael F. Hammer at the University of Arizona. Two Y-STRs among the Y-20plex (YCAII, and H4) were not previously typed in the YCC panel. Thus, many of the Y-STR alleles in the YCC panel were verified twice.

2.5. Analysis of PCR products

An amount of 1 μl aliquot of each PCR product was combined with 0.25 μl of GS-500 (ROX) and 8.75 μl of Hi-Di-formamide and electrophoresed on an 3100 Genetic Analyzer (Applied Biosystems) using a 36-cm array and

DYS459		
	AC010682.2	F.atcacaccccactgcaggccaaaaatagg**(TAAA)**$_{9}$taagaaaaaaggatcaaattaac tgcaaatatatcaggt.R
Allele (bp)	YCC ID	
10(152)	YCC51	51bp **(TAAA)**$_{10}$ 61bp
9(148)	YCC24	51bp **(TAAA)**$_{09}$ 61bp
9(148)	YCC26	51bp **(TAAA)**$_{09}$ 61bp
8(144)	YCC42	51bp **(TAAA)**$_{08}$ 61bp
DYS463		
	AC007275.4	F.tgtaccacctttcacagcaaattctaatttacactgatgtagactaagagccacagagcttg atcaaccatgaagaaag**(AAAGG)**$_{7}$**(AAGGG)**$_{15}$**(AAGGA)**$_{2}$aaagaggagt.R
Allele (bp)	YCC ID	
27(269)	YCC19	102bp **(AAAGG)**$_{6}$**(AAGGG)**$_{19}$(AAGGA)$_{2}$ 32bp
22(244)	YCC24	102bp **(AAAGG)**$_{6}$**(AAGGG)**$_{14}$(AAGGA)$_{2}$ 32bp
24(254)	YCC26	102bp **(AAAGG)**$_{7}$**(AAGGG)**$_{15}$(AAGGA)$_{2}$ 32bp
20(234)	YCC33	102bp **(AAAGG)**$_{6}$**(AAGGG)**$_{12}$(AAGGA)$_{2}$ 32bp
DYS464		
	X17354.1	F.ctcagtttaaaatacatgcctg**(CCTT)**$_{13}$cttt**(CCTT)**$_{2}$cttcccttctcttcc ttcctccctccatccctc**(CCTT)**$_{4}$ccct**(CCTT)**$_{4}$cttt**(CCTT)**$_{2}$tctct**(CCTT)**$_{2}$ cctccttctctcttctcttttttcttttcttccctga.R
Allele (bp)	ID	
15(266)	YCC10	41bp **(CCTT)**$_{15}$cttt(CCTT)$_{2}$cttcccttctcttccttcctccctccatccctc(CCTT)$_{4}$ ccct(CCTT)$_{4}$cttt(CCTT)$_{2}$tctct(CCTT)$_{2}$ 59bp
13(258)	YCC53	41bp **(CCTT)**$_{13}$cttt(CCTT)$_{2}$cttcccttctcttccttcctccctccatccctc(CCTT)$_{4}$ ccct(CCTT)$_{4}$cttt(CCTT)$_{2}$tctct(CCTT)$_{2}$ 59bp
18(278)	YCC76	41bp **(CCTT)**$_{18}$cttt(CCTT)$_{2}$cttcccttctcttccttcctccctccatccctc(CCTT)$_{4}$ ccct(CCTT)$_{4}$cttt(CCTT)$_{2}$tctct(CCTT)$_{2}$ 59bp
14(262)	DKT183	41bp **(CCTT)**$_{14}$cttt(CCTT)$_{2}$cttcccttctcttccttcctccctccatccctc(CCTT)$_{4}$ ccct(CCTT)$_{4}$cttt(CCTT)$_{2}$tctct(CCTT)$_{2}$ 59bp

Fig. 3. (*Continued*).

filter set D. PCR products from Multiplex I and II were mixed and then injected into a single run. The data were analyzed with Genescan (v. 3.7, Applied Biosystems) and Genotyper (v. 1.1, Applied Biosystems).

2.6. DNA sequencing and nomenclature

Four samples of different ethnic origins were sequenced for each novel Y-STR. The primers for DYS464 produced multiple bands in all but three of the YCC samples, these three samples were sequenced, as well as one non-YCC sample. PCR products were purified using Qiaquick PCR Purification spin columns from Qiagen. Sequence reactions were obtained using a BigDye terminator Cycle Sequencing FS Ready Reaction kit (Applied Biosystems). The suggested repeat nomenclature of novel Y-STR alleles follows the guidelines of the DNA commission of the International Society for Forensic Genetics (ISFG) [27], in that alleles were named according to the total number of directly-adjacent variant repeat-blocks and non-variant repeat-blocks. Invariant repeat-blocks that were not directly adjacent to the variant Y-STR locus were not counted so that future alternative primer designs, aimed at producing smaller PCR fragments that exclude the additional repeat-blocks, would not change the repeat nomenclature [28]. Single repeats were counted when they were directly flanked by variant repeat-blocks in a compound Y-STR.

2.7. Statistical analyses

We computed simple diversity statistics to compare the variation of both novel and previously published Y-STRs in the YCC panel. The number of alleles was counted for each Y-STR in the set of YCC samples. Gene diversity, h, or the probability that two alleles, chosen at random, are different, was calculated following Nei [29]. Discrimination capacity was determined by dividing the number of haplotypes by the number of samples. DYS389I was subtracted from DYS389II because the latter contains the former (see http://www.medfac.leidenuniv.nl/fldo/dys389.htm) so that our analysis included two distinct sections of this compound Y-STR which we will refer to as DYS389I and DYS389II-I.

3. Results

3.1. Identification of Y-STRs

Seventeen candidate Y-STRs were identified, three of which were previously described: DYS393, DYS437, and

DYS439. DYS439 and DYS437 were included in Multiplex II and III, respectively, using novel primers (see Section 2.4.2). Table 1 lists the remaining fourteen Y-STRs that were novel and male specific. The new Y-STRs include eight tetranucleotide repeats, five pentanucleotide repeats, and one hexanucleotide repeat. Two Y-STRs produced multiple PCR products; DYS459 primer sequences produced one or two PCR products, while DYS464 primer sequences produced 1–4 peaks (see Fig. 1). The number of PCR products observed at DYS459 and DYS464 is consistent with the number of physical locations along the Y chromosome per Y-STR (Fig. 2).

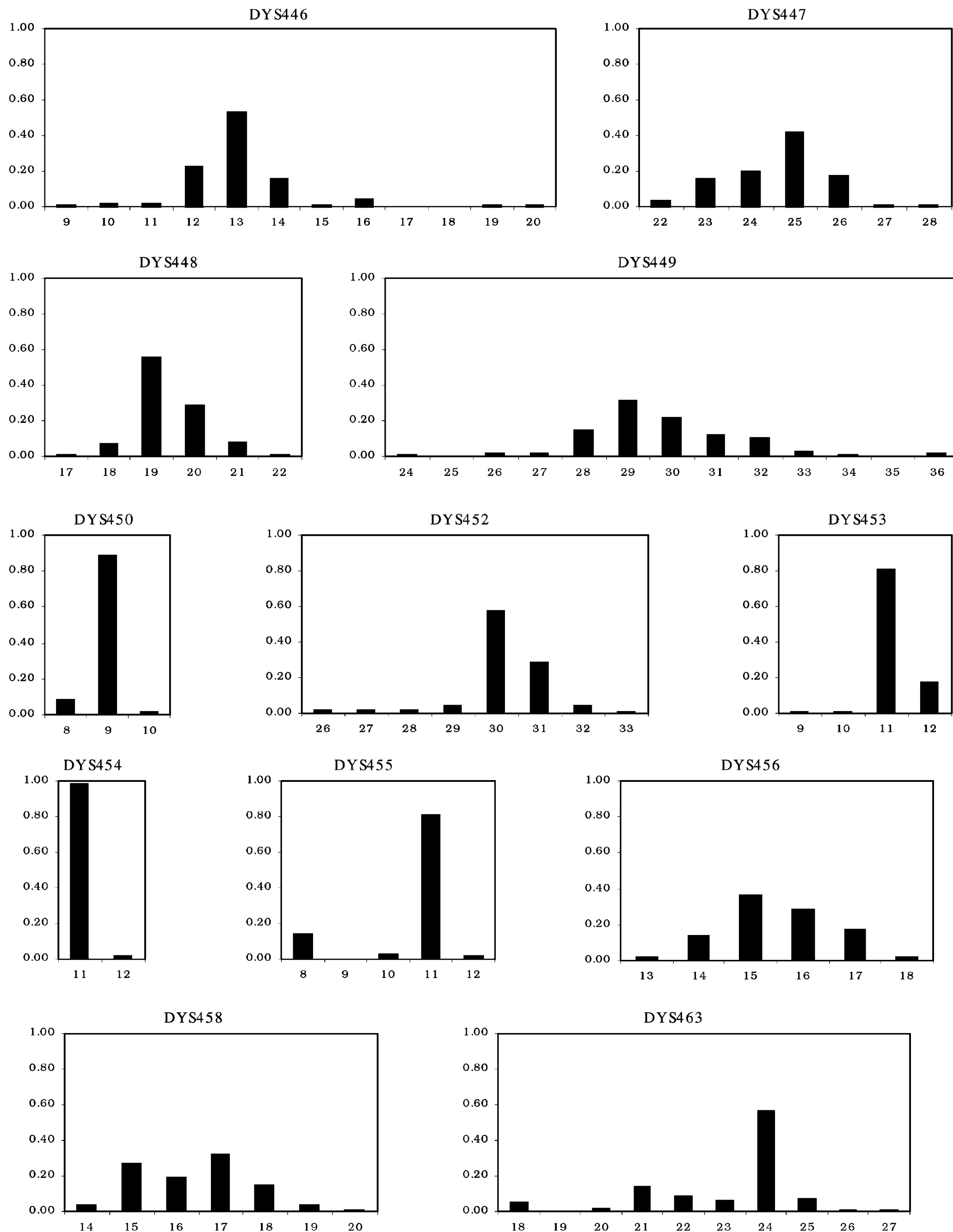

Fig. 4. Histograms of the observed allele frequencies for 12 novel Y-STRs in the European–American sample. Multi-copy Y-STRs, DYS459 and DYS464, were excluded. The ascending allele numbers indicate the number of repeats as inferred by direct sequencing (see Fig. 3).

3.2. Nomenclature

The sequence data for the 14 Y-STRs are shown in Fig. 3. Nine Y-STRs contained variation in a single repeat region (DYS446, DYS450, DYS453, DYS454, DYS455, DYS456, DYS458, DYS459, DYS464) and thus can be categorized as simple repeats [30]. Six of these nine Y-STRs (DYS446, DYS453, DYS454, DYS455, DYS456, DYS459) do not have any additional tandem-repeats (i.e. blocks greater than one consecutive repeat) anywhere in the amplified region. For these simple repeat Y-STRs the allele nomenclature follows directly from the number of repeats in the single variable block. The remaining three simple repeat Y-STRs (DYS450, DYS458, and DYS464) contain additional repeat blocks in various positions within the amplified region (shown in bold font in the clone sequences, see Fig. 3), particularly DYS464 which has five blocks of repeat units (three blocks of two repeat count, and two blocks of four repeat count) dispersed throughout the amplified region. For these simple repeats we excluded the non-variant repeat blocks that were not directly adjacent to the simple repeat.

Three of the remaining Y-STRs (DYS447, DYS452, and DYS463) can be classified as compound repeats since they consist of two different repeat motifs that are directly adjacent [30]. DYS463 contained two variable regions that were directly adjacent and made up of two different repeat motifs, namely $(AAAGG)_n$ and $(AAGGG)_n$, as well as a directly adjacent non-variable $(AAGGA)_2$ motif. DYS447 contained three variable $(TAATA)_n$ motifs directly interspersed with two single $(TAAAA)_1$ motifs, that is, $(TAATA)_n$-TAAAA-$(TAATA)_n$-TAAAA-$(TAATA)_n$. For DYS447 we included the single $(TAAAA)_1$ motifs that were directly abutted between the variable regions in the tally of repeat number. The long compound structure of DYS452 is made up of three types of directly contiguous and alternating motifs: $(TATAC)_n$, $(TGTAC)_n$, and $(CATAC)_n$; all of the DYS452 motifs were included in the tally of the repeat number. Two variable regions were found in DYS452, one $(TATAC)_n$ region varied in repeat length, and another stretch of $(TATAC)_n$-$(CATAC)_2$-$(TATAC)_3$ varied because of a deletion. The remaining two Y-STRs, DYS448 and DYS449, were more complex in structure since they contained two variable blocks interspersed with intervening non-variable sequence. Both of these Y-STRs have additional non-variant repeat blocks between the variable regions that were not counted in the repeat length nomenclature.

Deletion or transition variants were observed in the sequence data for two of the three compound structure Y-STRs, DYS452 and DYS447. In sample YCC24 at DYS452 there may have been a deletion of the $(CATA)_2$ motif as well as two repeats from the adjacent $(TATAC)_3$ motif. In sample YCC26 at DYS447 there may have been an A to T substitution in the fourth base of the second $(TAAAA)_1$ motif, or a complete deletion of $(TAAAA)_1$, resulting in the merging of the second and third $(TAATA)_n$ motifs (Fig. 3).

3.3. Allele frequency distributions

The observed distributions of allele frequencies for the European–Americans for 12 single-copy Y-STRs are shown in Fig. 4. The number of alleles varied from 2 (DYS454) to 11 (DYS449). Four Y-STRs (DYS450, DYS453, DYS454, and DYS455) have high and spiked distributions for a single allele with few additional alleles. Broad and lower-lying

Table 2
Allelic patterns in multi-copy Y-STRs in European–Americans

DYS459		DYS464	
Allele[a]	Frequency	Allele[a]	Frequency
7, 9	0.007	15	0.007
8, 9	0.196	16	0.007
9	0.230	12, 14	0.020
8, 10	0.061	12, 15	0.014
9, 10	0.493	13, 14	0.007
9, 11	0.007	13, 15	0.020
10	0.007	13, 17	0.014
		14, 15	0.020
		14, 16	0.014
		15, 16	0.047
		15, 17	0.135
		15, 18	0.027
		16, 17	0.027
		11, 14, 15	0.020
		11, 14, 16	0.014
		12, 13, 14	0.007
		12, 14, 15	0.047
		12, 14, 16	0.007
		12, 14, 17	0.007
		12, 15, 16	0.034
		12, 15, 17	0.014
		13, 14, 15	0.014
		13, 15, 16	0.007
		14, 15,16	0.014
		14, 15, 17	0.014
		14, 15, 18	0.007
		14, 16, 17	0.027
		14, 16, 18	0.014
		15, 16, 17	0.176
		15, 16, 18	0.027
		15, 17, 18	0.041
		11, 12, 15, 16	0.007
		11, 13, 15, 16	0.007
		12, 13, 14, 15	0.007
		12, 13, 14, 16	0.007
		12, 13, 15, 16	0.007
		12, 14, 15, 16	0.054
		12, 14, 15, 17	0.014
		12, 14, 16, 17	0.007
		13, 15, 16, 17	0.007
		14, 15, 16, 17	0.027
		14, 15, 17, 18	0.007
		15, 16, 17, 18	0.007

[a] Allelic pattern produced by multi-copy Y-STR.

Table 3
Diversity among 36 Y-STRs in YCC panel

Y-STR (N)[a]	YCC no. of alleles (73)	YCC gene diversity		AFR gene diversity		EUR/MEA gene diversity		ASN gene diversity	
		(73)	Rank	(25)	Rank	(22)	Rank	(26)	Rank
DYS464*	41	0.973	1	0.967	2	0.965	1	0.926	2
DYS385*	36	0.973	2	0.980	1	0.922	2	0.926	3
YCAII*	17	0.908	3	0.867	4	0.840	4	0.855	5
DYF371*	19	0.891	4	0.623	22	0.814	5	0.929	1
DYS449	10	0.874	5	0.887	3	0.779	7	0.880	4
DYS446	9	0.836	6	0.857	5	0.849	3	0.763	6
DYS463	9	0.814	7	0.787	8	0.766	8	0.665	14
DYS390	8	0.789	8	0.773	10	0.740	11	0.677	13
DYS448	9	0.782	9	0.803	7	0.693	15	0.603	18
DYS447	7	0.781	10	0.677	18	0.801	6	0.717	9
DYS392	8	0.768	11	0.420	28	0.714	14	0.760	7
DYS458	8	0.748	12	0.813	6	0.723	12	0.686	12
DYS19	6	0.747	13	0.733	15	0.636	20	0.702	11
DYS459*	8	0.730	14	0.737	14	0.606	23	0.548	21
DYS389II-I	6	0.724	15	0.740	13	0.745	10	0.582	20
DYS456	6	0.706	16	0.607	23	0.723	13	0.702	10
DYS439	5	0.694	17	0.693	17	0.623	22	0.754	8
DYS452	7	0.681	18	0.637	20	0.762	9	0.625	16
DYS389I	5	0.669	19	0.743	12	0.636	19	0.615	17
DYS393	5	0.664	20	0.783	9	0.671	16	0.532	23
DYS461	4	0.637	21	0.757	11	0.628	21	0.446	30
H4	5	0.599	22	0.657	19	0.541	26	0.588	19
DYS438	4	0.598	23	0.420	29	0.654	18	0.471	26
DYS450	4	0.593	24	0.407	30	0.437	30	0.655	15
DYS437	4	0.565	25	0.517	26	0.654	17	0.280	33
DXYS156Y	5	0.550	26	0.333	32	0.178	32	0.459	29
DYS460	5	0.544	27	0.520	25	0.606	24	0.520	24
DYS391	4	0.532	28	0.637	21	0.481	28	0.471	27
DYS426	4	0.516	29	0.547	24	0.481	29	0.508	25
DYS462	4	0.504	30	0.280	33	0.524	27	0.542	22
DYS453	5	0.454	31	0.707	16	0.329	31	0.222	35
DYS388	7	0.360	32	0.227	34	0.576	25	0.271	34
DYS454	3	0.336	33	0.380	31	0.091	34	0.465	28
DYS434	3	0.269	34	0.227	35	0.091	35	0.428	31
DYS455	5	0.230	35	0.473	27	0.091	33	0.077	36
G10123*	4	0.157	36	0.153	36	0.000	36	0.286	32

[a] The asterisk denotes multi-copy Y-STRs.

distributions were observed at DYS446, DYS463, and DYS449. The allele frequency distributions of the multi-copy Y-STRs are shown in Table 2.

3.4. Gene diversity

The gene diversities of 36 Y-STRs in 73 male YCC samples are shown in Table 3. Diversity values ranged from 0.1572 (G10123) to 0.9730 (DYS464 and DYS385) and the number of alleles ranged from 3 (DYS454 and DYS434) to 41 (DYS464). The multi-copy Y-STRs had the highest gene diversity values and the largest number of alleles. The gene diversity values were mostly determined by the particular properties of a given Y-STR rather than variation among different geographic groups in the YCC panel. For example, gene diversities observed at DYS464, DYS385, and DYS449 were very high among Africans, Europeans/Middle Easterners, and Asians, while gene diversities observed at DYS426, G10123, and DYS460 were low among the three continental groups. In general, there was good concordance in the rankings of the most diverse to the least diverse Y-STRs between the three groups: the correlation between gene diversities in the Africans, Europeans/Middle Easterners, and Asians were high and significant (Africans–European/Middle Easterners $r = 0.74$; Africans–Asians $r = 0.61$; and European/Middle Easterners–Asians $r = 0.75$; all comparisons $P < 0.001$). Most of the variance observed in gene diversity values among geographic groups occurs in Y-STRs

of low variability, such as DYS453, DYS455, and DYS454. However, there are some gene diversity differences across geographic groups in Y-STRs of high gene diversity. For example, DYF371 was very diverse in the Asians, slightly less diverse in the Europeans/Middle Easterners, and much less diverse in the Africans. Similarly, DYS392 has the highest diversity, and longest mean allele length, in the Asians and Europeans/Middle Easterners, and much lower gene diversity, and mean allele length, in Africans.

In order to examine gene diversities of a subset of the Y-STRs in a population sample, we genotyped 26 of the aforementioned Y-STRs in a sample of 148 European–Americans from South Dakota. Table 4 shows the gene diversity values among 26 Y-STRs in the European–Americans. There was a good correspondence between the gene diversities in the European–Americans and the YCC panel ($r = 0.80$; $P < 0.001$) and between the European–Americans and the Europeans/Middle Easterners in the YCC panel ($r = 0.82$; $P < 0.001$). Ten of the novel Y-STRs were among the most diverse Y-STRs, and DYS464 was the most diverse Y-STR.

3.5. Haplotype resolution

In order to test the value of various Y-STRs for distinguishing male lineages, haplotype analysis was carried out in the YCC panel and in the European–American sample. We began both analyses with the set of Y-STRs included in the minimal haplotype. In the YCC panel, 65 of the 73

Table 4
Y-STR diversity in European–Americans

Y-STR[a]	No. of alleles	Gene diversity
DYS464*	43	0.939
DYS385*	28	0.823
DYS449	11	0.812
DYS458	7	0.771
DYS456	6	0.740
DYS447	7	0.732
DYS390	5	0.732
DYS459*	7	0.667
DYS439	5	0.666
DYS446	10	0.651
DYS463	9	0.645
DYS448	6	0.603
DYS452	8	0.589
DYS437	4	0.584
DYS392	5	0.579
DYS19	6	0.557
DYS391	5	0.533
DYS389I	3	0.521
DYS389II-I	5	0.501
DYS426	3	0.446
DYS388	6	0.423
DYS455	4	0.324
DYS453	4	0.314
DYS393	5	0.258
DYS450	3	0.198
DYS454	2	0.027

[a] The asterisk denotes multi-copy Y-STRs.

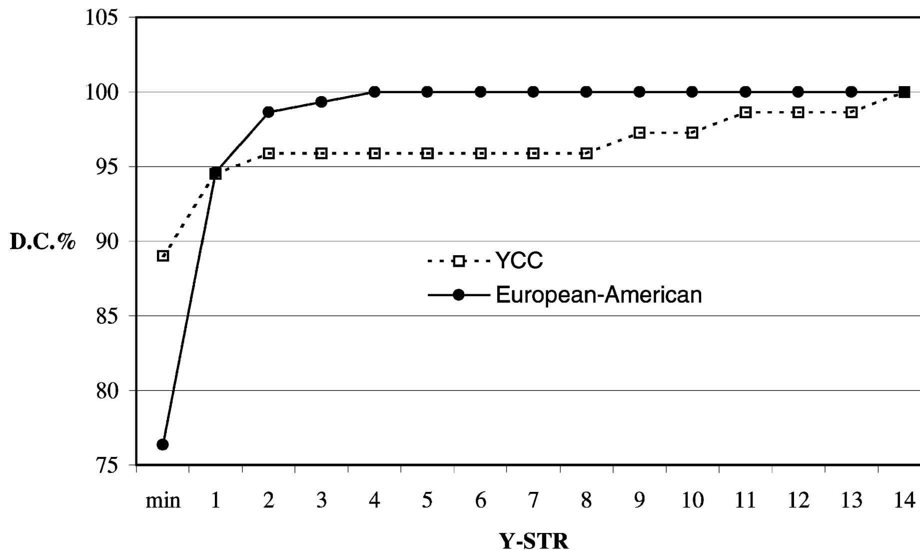

Fig. 5. Discrimination capacity obtained in the YCC panel and in the European–American sample obtained by adding Y-STRs to the minimal haplotype Y-STRs. The x-axis indicates the number of Y-STRs included in the analysis, beginning with the set of Y-STRs that define the minimal haplotype (DYS19, DYS385, DYS389I, DYS389II-I, DYS390, DYS391, DYS392, and DYS393) and incrementally and cumulatively adding a single Y-STR in order of highest gene diversity to the lowest gene diversity (see Tables 3 and 4). The y-axis indicates the discrimination capacity (D.C.) in percent. All of the Y chromosomes in the YCC panel were unique after including 14 Y-STRs to the minimal haplotype Y-STRs, while four Y-STRs in addition to the minimal haplotype Y-STRs resolved all of the European–American Y chromosomes.

samples (89%) were distinguished with the minimal set of Y-STRs (Fig. 5). Eight pairs of individuals were still identical. Six of the eight shared Y-STR haplotypes derive from isolated and indigenous populations, including pairs of Africans (Biaka, !Kung), Native Americans (Karitiana, Surui), and North Asians (Yakuts; two pairs). Another identical haplotype was shared between a Mbuti Pygmy and a Bantu speaker. The last identical haplotype was shared between two Europeans (English and German). We then sequentially and cumulatively added each of the tested Y-STRs (in descending order of gene diversity) to the minimal haplotype Y-STRs and recalculated discrimination capacity of the resulting haplotype (e.g. see [15]). The addition of DYS464 to the minimal set of Y-STRs distinguished both pairs of Native Americans, the pair of Europeans, as well as a pair of Africans (Mbuti Pygmy and Bantu speaker). YCAII distinguished one of the pairs of Yakut haplotypes, leaving three pairs of identical Biaka Pygmy, !Kung, and Yakut hapolotypes. The inclusion of DYS458 and DYS456 eventually distinguished the African hapolotypes, while the last pair of Yakut haplotypes were distinguished with the fourteenth Y-STR (DYS461).

In the European–American sample, 113 out of 148 samples (76%) were distinct using the minimal haplotype Y-STRs. The shared haplotypes include: 11 pairs; 3 triplicates; 1 haplotype shared by 8 individuals; and another haplotype shared by 9 individuals. These two most frequent haplotypes are 14-(11,14)-13-16-23-11-13-13 (6.1%; DYS19-DYS385-DYS389I-DYS389(II-I)-DYS390-DYS391-DYS392-DYS393) and 14-(11,14)-13-16-24-11-13-13 (5.4%). The inclusion of DYS464 to the minimal haplotype Y-STRs resolved the vast majority (94.6%) of the shared European–American paternal lineages. Only three additional Y-STRs (DYS449, DYS458, and DYS456) were necessary to completely distinguish all of the Y chromosomes in the European–American sample.

4. Discussion and conclusions

The 14 novel Y-STRs discovered here appear useful in forensic casework for two reasons: (1) they provide additional power to resolve Y chromosome haplotypes, and (2) they are all tetra-, penta- and hexa-nucleotide repeats. The novel Y-STRs have more than doubled the number of known pentanucleotide markers and they include the first hexanucleotide repeat on the NRY (DYS448). These longer repeat motif STRs may be useful for improving the interpretation of sample mixtures [31,32]. Depending on the particular STR, the stutter peak heights of dinucleotide and trinucleotide repeats can be higher than 30% of their corresponding STR allele, while stutter products of tetranucleotides are approximately 15%, and pentanucleotide repeats may have stutter products of less than 1–2% [19]. These novel Y-STRs with longer repeat motifs should be evaluated for use in mixture studies.

The YCC panel served as a resource for identifying highly informative Y-STRs. These cell lines represent a globally diverse set of DNAs and provide standard reference material for direct comparisons among known and novel Y-STRs. We have shown that Y-STRs identified on the basis of their high diversity in the YCC panel are likely to be highly informative in population samples. Interestingly, these results also suggest that different Y-STRs may vary in their gene diversity depending on the particular population sample under investigation, and perhaps the length of the alleles. While most of the variation in diversity was Y-STR dependent, there was some variation among different geographic groups in the rank order of Y-STR diversity (Tables 3 and 4). Although we did not type 13 other published Y-STRs (DYS288, YCAI, YCAIII, DYS425, A10, C4, DYS435, DYS436, DYS441, and DYS442, DYS443, DYS444, and DYS445), this comparative analysis of 36 Y-STRs indicated that out of our 14 novel STRs, 7 were among the 10 most polymorphic single-copy Y-STRs (DYS449, DYS446, DYS463, DYS448, DYS447, DYS458, and DYS459) while one (DYS464) appeared to be the most polymorphic Y-STR yet described.

Variation in peak heights associated with DYS464 is consistent with four independent amplifications. A BLAT search indicated that this entire region is present on the long arm of the NRY in four copies, and is part of the DAZ gene region [33]. Although we assigned 41 allelic patterns based on the presence or absence of different size peaks, Fig. 2 shows that it may be possible to assign additional allelic classes on the basis of different peak heights (e.g. if peak heights can be reliably related to copy number differences). In the YCC panel, 51 haplotypes were found when peak heights were considered. More work is needed to quantify the reproducibility of the peak height variation and the additional resolving potential of this Y-STR in forensic studies. We have observed some allelic patterns (data not shown) that could be consistent with duplications (5–6 peaks; two tall peaks and two short peaks) or deletions (three peaks of equal height). Nevertheless, it is simple to assign allelic patterns to the presence or absence of peaks and to ignore peak height as we have done in all of the analyses presented here.

The addition of the novel Y-STRs to the minimal haplotype improved the ability to distinguish Y chromosomes in the European–American sample. The most common European haplotype, 14-(11,14)-13-16-24-11-13-13, is found in 3.1% of 9972 individuals in the Y-STR HRDatabase [17]. This haplotype was found at 5.4% in the European–American sample. DYS464 alone distinguished 75% of these haplotypes in the European–American sample. The remaining shared haplotypes were distinguished using only three additional Y-STRs. All but a single pair of the European–American haplotypes differ at three or more Y-STRs among the 26 tested. The single pair of most closely related chromosomes differed by two Y-STRs out of 26 tested. Thus, these European–American Y chromosomes are not

identical by descent, as the minimal haplotype suggested. The closely related Y chromosomes in the YCC panel were more difficult to distinguish. When considering all 36 Y-STRs tested there were four pairs of individuals that differed at a single Y-STR, including a pair of Surui, two pairs of Yakuts, and a pair of Biaka Pygmies. Autosomal genotyping of three of these pairs with 15 STRs [26] suggests that these males are patrilineal relatives, perhaps cousins; the pairs of haplotypes exhibited differences at both alleles of a locus for 2–5 of the 15 STRs. Moreover, when the Bayesian method of Walsh [20] was used to estimate the TMRCA between these pairs of Y chromosomes, assuming a mutation rate of 2.0×10^{-3} we found a TMRCA of only 11.8 generations (95%CI = 1.7–39.2 generations). Future studies of Y-chromosomes from well-characterized pedigrees could test the potential to distinguish between patrilineal relatives of various degrees using highly variable Y-STRs.

Acknowledgements

This research was supported by the National Institute of Justice Grant 2000-IJ-CX-K006 to MFH. Rex Riis kindly provided blood samples from the European–Americans. We thank Richard Schoske for sharing the beta-version of the Y-STR 20plex. We appreciate the duplicate Y-STR typing data in the YCC panel that were provided by Mark Jobling. Bruce Walsh computed the TMRCA estimates. We acknowledge Shari Garcia and Matthew Kaplan and for technical assistance and Jared Ragland for marker identification.

References

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

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

[3] A. Betz, G. Bassler, G. Dietl, X. Steil, G. Weyermann, W. Pflug, DYS STR analysis with epithelial cells in a rape case, Forensic Sci. Int. 118 (2001) 126–130.

[4] M. Prinz, A. Ishii, A. Coleman, H.J. Baum, R.C. Shaler, Validation and casework application of a Y chromosome specific STR multiplex, Forensic Sci. Int. 120 (2001) 177–188.

[5] L. Roewer, M. Kayser, P. Dieltjes, M. Nagy, E. Bakker, M. Krawczak, P. de Knijff, Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations, Hum. Mol. Genet. 5 (1996) 1029–1033.

[6] L. Roewer, J. Arnemann, N.K. Spurr, K.H. Grzeschik, J.T. Epplen, Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts, Hum. Genet. 89 (1992) 389–394.

[7] N. Mathias, M. Bayes, C. Tyler-Smith, Highly informative compound haplotypes for the human Y chromosome, Hum. Mol. Genet. 3 (1994) 115–123.

[8] H. Chen, W. Lowther, D. Avramopoulos, S.E. Antonarakis, Homologous loci DXYS156X and DXYS156Y contain a polymorphic pentanucleotide repeat (TAAAA)n and map to human X and Y chromosomes, Hum. Mutat. 4 (1994) 208–211.

[9] T. Karafet, P. de Knijff, E. Wood, J. Ragland, A. Clark, M.F. Hammer, Different patterns of variation at the X- and Y-chromosome-linked microsatellite loci DXYS156X and DXYS156Y in human populations, Hum. Biol. 70 (1998) 979–992.

[10] M.A. Jobling, V. Samara, A. Pandya, N. Fretwell, B. Bernasconi, R.J. Mitchell, T. Gerelsaikhan, B. Dashnyam, A. Sajantila, P.J. Salo, Y. Nakahori, C.M. Disteche, K. Thangaraj, L. Singh, M.H. Crawford, C. Tyler-Smith, Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males, Hum. Mol. Genet. 5 (1996) 1767–1775.

[11] P.S. White, O.L. Tatum, L.L. Deaven, J.L. Longmire, New, male-specific microsatellite markers from the human Y chromosome, Genomics 57 (1999) 433–437.

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

[13] R. Iida, E. Tsubota, T. Matsuki, Identification and characterization of two novel human polymorphic STRs on the Y chromosome, Int. J. Legal Med. 115 (2001) 54–56.

[14] R. Iida, E. Tsubota, K. Sawazaki, M. Masuyama, T. Matsuki, T. Yasuda, K. Kishi, Characterization and haplotype analysis of the polymorphic Y-STRs DYS443, DYS444, and DYS445 in a Japanese population, Int. J. Legal Med. 116 (2002) 191–194.

[15] E. Bosch, A.C. Lee, F. Calafell, E. Arroyo, P. Henneman, P. de Knijff, M.A. Jobling, High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions, Forensic Sci. Int. 125 (2002) 42–51.

[16] P. de Knijff, M. Kayser, A. Caglia, D. Corach, N. Fretwell, C. Gehrig, G. Graziosi, F. Heidorn, S. Herrmann, B. Herzog, M. Hidding, K. Honda, M. Jobling, M. Krawczak, K. Leim, S. Meuser, E. Meyer, W. Oesterreich, A. Pandya, W. Parson, G. Penacino, A. Perez-Lezaun, A. Piccinini, M. Prinz, C. Schmitt, P.M. Schneider, R. Szibor, J. Teifel-Greding, G. Weichhold, L. Roewer, Chromosome Y microsatellites: population genetic and evolutionary aspects, Int. J. Legal Med. 110 (1997) 134–149.

[17] L. Roewer, M. Krawczak, S. Willuweit, M. Nagy, C. Alves, A. Amorim, K. Anslinger, C. Augustin, A. Betz, E. Bosch, A. Caglia, A. Carracedo, D. Corach, A.F. Dekairelle, T. Dobosz, B.M. Dupuy, S. Furedi, C. Gehrig, L. Gusmao, J. Henke, L. Henke, M. Hidding, C. Hohoff, B. Hoste, M.A. Jobling, H.J. Kargel, P. de Knijff, R. Lessig, E. Liebeherr, M. Lorente, B. Martinez-Jarreta, P. Nievas, M. Nowak, W. Parson, V.L. Pascali, G. Penacino, R. Ploski, B. Rolf, A. Sala, U. Schmidt, C. Schmitt, P.M. Schneider, R. Szibor, J. Teifel-Greding, M. Kayser, Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes, Forensic Sci. Int. 118 (2001) 106–113.

[18] P.S. Walsh, N.J. Fildes, R. Reynolds, Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA, Nucleic Acids Res. 24 (1996) 2807–2812.
[19] J.M. Butler, Forensic DNA Typing, Academic Press, San Diego, 2001.
[20] B. Walsh, Estimating the time to the most recent common ancestor for the Y chromosome or mitochondrial DNA for a pair of individuals, Genetics 158 (2001) 897–912.
[21] M.P.H. Stumpf, D.B. Goldstein, Genealogical and evolutionary inference with the human Y chromosome, Science 291 (2001) 1738–1742.
[22] The Y chromosome consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups, Genome Res. 12 (2002) 339–348.
[23] D.A. Benson, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, B.A. Rapp, D.L. Wheeler, GenBank, Nucleic Acids Res. 30 (2002) 17–20.
[24] G. Benson, Tandem repeats finder: a program to analyze DNA sequences, Nucleic Acids Res. 27 (1999) 573–580.
[25] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[26] J. M. Butler, R. Schoske, P. M. Vallone, M. C. Kline, A. J. Redd, M. F. Hammer, A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers, Forensic Sci. Int. 129 (2002) 10–24.
[27] P. Gill, C. Brenner, B. Brinkmann, B. Budowle, A. Carracedo, M.A. Jobling, P. de Knijff, M. Kayser, M. Krawczak, W.R. Mayr, N. Morling, B. Olaisen, V. Pascali, M. Prinz, L. Roewer, P.M. Schneider, A. Sajantila, C. Tyler-Smith, DNA commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs, Int. J. Legal Med. 114 (2001) 305–309.
[28] L. Gusmão, A. González-Neira, C. Alves, M. Lareu, S. Costa, A. Amorim, A. Carracedo, Chimpanzee homologous of human Y specific STRs. A comparative study and a proposal for nomenclature, Forensic Sci. Int. 126 (2002) 129–136.
[29] M. Nei, Molecular Evolutionary Genetics, Columbia University Press, New York, 1987.
[30] A. Urquhart, C.P. Kimpton, T.J. Downes, P. Gill, Variation in short tandem repeat sequences—a survey of twelve microsatellite loci for use as forensic identification markers, Int. J. Legal Med. 107 (1994) 13–20.
[31] P. Gill, R. Sparkes, R. Pinchin, T. Clayton, J. Whitaker, J. Buckleton, Interpreting simple STR mixtures using allele peak areas, Forensic Sci. Int. 91 (1998) 41–53.
[32] T.M. Clayton, J.P. Whitaker, R. Sparkes, P. Gill, Analysis and interpretation of mixed forensic stains using DNA STR profiling, Forensic Sci. Int. 91 (1998) 55–70.
[33] R. Reijo, T.Y. Lee, P. Salo, R. Alagappan, L.G. Brown, M. Rosenberg, S. Rozen, T. Jaffe, D. Straus, O. Hovatta, A. de la Chapelle, S. Silber, D.C. Page, Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene, Nat. Genet. 10 (1995) 383–393.