Int J Legal Med (2004) 118 : 178–183
DOI 10.1007/s00414-004-0436-5

TECHNICAL NOTE

John M. Butler · Richard Schoske

Duplication of DYS19 flanking regions in other parts of the Y chromosome

Received: 25 November 2003 / Accepted: 2 March 2004 / Published online: 7 April 2004

Abstract During the testing of alternative primers for the Y chromosome short tandem repeat marker DYS19, a duplicated region of the Y chromosome was discovered. The duplicated sequence is contained within GenBank accession AC006335 and has a high degree of homology with the DYS19 flanking region (GenBank accession AC017019) but without the polymorphic TAGA repeat. Bioinformatic approaches have been taken to try and understand the implications of this homolog to enable improved primer design for DYS19. Sequence alignments and careful placement of primers in order to obtain specific amplification of the DYS19 locus are discussed in the context of all previously published primer sets. Since the DYS19 locus is part of the widely used minimal haplotype, its robust amplification is highly desirable particularly in multiplex reactions. The discovery of this duplicated region of the Y chromosome shows the value of newly available human genome sequence information for assay design and the importance of using sequence queries and alignments in the primer design process.

Keywords Y Chromosome · Short tandem repeat · PCR amplification · DYS19 · Primer design · Sequence alignment

Introduction

The paternally inherited and subsequent haploid state of the Y chromosome makes it useful for applications in human identity testing and evolutionary studies [1, 2, 3]. The first Y chromosome short tandem repeat (STR) marker to be discovered and amplified with the polymerase chain reaction (PCR) was the GATA repeat marker DYS19 [4, 5]. The DYS19 marker is included in the "minimal haplotype" for the online Y-chromosomal STR haplotype reference database [6], which now contains information on over 18,000 males from around the world. DYS19 is commonly included in Y-STR multiplexes [7, 8, 9, 10, 11, 12] because it is part of the core set of loci examined in population studies [13, 14, 15]. Additionally, DYS19 is present in several commercially available Y-STR multiplex kits [16, 17].

During the development of new PCR primer sets for the purpose of amplifying DYS19 in a large multiplex, we discovered that a non-specific product arose with a new set of primers [10]. Using human genome sequence information, we investigated a possible cause for this additional amplicon with sequence alignments and discovered that at least two homologous regions to DYS19 exist on different regions of the Y chromosome. This finding resonates with the recent discovery of large-scale duplication on the Y chromosome [18]. In this study, three different primer sets were used to amplify the DYS19 locus with or without its Y homologs. Previously published primers were examined in the context of these DYS19 homologous regions.

Materials and methods

Sequence queries and alignments

Sequence queries with short oligonucleotides were performed using a standard BLAST search available at http://www.ncbi.nlm.nih.gov/BLAST/. Additional queries with larger sequence blocks, typically more than 200 contiguous bases, were performed using the human BLAT search (http://genome.ucsc.edu/cgi-bin/hgBlat) and the April 2003 reference sequence. Sequence alignments were performed using the Baylor College of Medicine (BCM) search launcher http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html.

Primer design and synthesis

Primer design was performed as described by Schoske et al. [12]. Oligonucleotides were purchased from MWG Biotech (High Point,

J. M. Butler (✉) · R. Schoske
Biotechnology Division,
National Institute of Standards and Technology,
100 Bureau Drive, Mail Stop 8311, Gaithersburg, MD 20899, USA
Tel.: +1-301-9754049, Fax: +1-301-9758505,
e-mail: john.butler@nist.gov

R. Schoske
Armed Forces Institute of Pathology, Washington DC

Table 1 Different primer sets used to amplify the Y-chromosome STR marker DYS19 in this study

Primer name	Sequence (5'-to-3')	Primer3 Tm (°C)	Size range for alleles 10–19
1F	6FAM-CTACTGAGTTTCTGTTATAGT	43.0	176–212 bp
1R	ATGGCATGTAGTGAGGACA	54.7	
1R'	ATGG**C**CATGTAGTGAGGACA	59.0	
2F	6FAM-AGGTATGAGATCAAATTGACTGTG	57.3	211–247 bp
2R	GTGAGGACAAGGAGTCCATCTG	61.1	
3F	NED-ACTACTGAGTTTCTGTTATAGTGTTTTT	55.0	233–269 bp
3R	GTCAATCTCTGCACCTGGAAAT	60.5	
DYS394 F [7], GDB	FAM-GACTACTGAGTTTCTGCAATAGTG	55.2	229–265 bp
DYS394 R [7], GDB	AGCACTGCACCTGGAAATAG	58.0	
Szibor et al. F [21]	GTTATATATATATAGTGTTTAG (silver stain)	36.0	84–120 bp
Szibor et al. R [21]	GTTAAGGAGAGTGTCACTA	44.2	
Wiegand and Kleiber F [22]	GTGTTATATATATATAGTGTTTTA (silver stain)	41.2	88–124 bp
Wiegand and Kleiber 3R [22]	GGTTAAGGAGAGTGTCACTA	49.2	

The forward primer in each set has a fluorescent dye covalently attached to its 5' end (6FAM or NED).
Predicted primer melting temperatures (T_m) were calculated using a total primer concentration=0.05 µM and [+]=50 mM and Primer3 [20].

NC), Invitrogen (Carlsbad, CA), or Applied Biosystems (Foster City, CA) and tested for quality control purposes as described previously [11, 12].

PCR reaction and thermal cycling conditions

PCR amplifications were performed in reaction volumes of 20 µl with 1 unit of AmpliTaq Gold DNA polymerase (Applied Biosystems, Foster City, CA), 10 mM Tris-HCl (pH 8.3), 50 mM KCl[1], 1.5 mM $MgCl_2$, 200 µM deoxynucleotide triphosphates (dNTPs: dATP, dCTP, dGTP, dTTP), 1.8 µM primer mix, and 2 µl DNA template at a concentration of ~2.5 ng/µl.

Thermal cycling conditions were as follows: 95°C for 10 min, then 28 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 1 min, followed by 60°C for 45 min and 25°C hold.

PCR primers used for this study are listed in Table 1. DNA samples analyzed were the same Y chromosome consortium (YCC) panel and commercial blood bank samples reported previously [11].

Detection and analysis of PCR products

The separation and detection of PCR products were accomplished with the ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, CA) following the manufacturer's protocols using the matrix filter set F to detect the four dyes with 6FAM (blue), JOE (green), NED (yellow), and ROX (red). Following data collection, samples were analyzed with Genescan 3.7 (Applied Biosystems) and allele designations were made in Genotyper 3.7 (Applied Biosystems).[2]

[1] The accepted SI unit of concentration, mol/l has been represented by the symbol M in order to conform to the conventions of this journal.

[2] Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology or the US Department of Defense, nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

Three additional published primer sets are also included for comparison purposes.

Results and discussion

The majority of published DYS19 assays use primers first described by Roewer et al. [4] (Table 1, primers 1F/1R). While these primers work fine in singleplex PCR reactions and small multiplexes, the annealing characteristics of the original primers are not optimal for large multiplexes (e.g., [11]) primarily because primer 1F has a low T_m. Poor amplification performance with primer set 1F/1R has been previously reported for a Y-STR pentaplex [8] and a decaplex [12]. An alternative approach has been to use DYS394 primers to improve amplification of the DYS19 region [7]. A note on the Y-STR Haplotype Reference Database site's primer and protocols section (http://www.ystr.org) suggests that the insertion of an additional C that is 5 bases from the 5'-end of primer 1R improves DYS19 amplification. Table 1 includes this primer as primer 1R'.

In the process of exploring alternative primers for more robust amplification of the DYS19 locus, we discovered that an extra PCR product was being generated with one of the primer pairs (Table 1, 2F/2R). This primer pair produced a peak that consistently sized at 187 bp in over 100 samples tested regardless of the normal DYS19 allele size variation, which occurred in the range of 211–247 bp (Fig. 1). This seemingly non-specific amplicon was male-specific (data not shown).

While primer set 2F/2R worked well in a multiplex assay with up to 19 other amplicons [10], the 187 bp non-specific product fell in the size range of DYS437 in our assay with the possibility of obscuring DYS437 allele 15, which usually sized at approximately 186 bp. Perhaps more importantly, the competition of another primer binding region lowered the overall amplification efficiency for generating the DYS19 alleles in the multiplex reaction. Thus, we began looking into a method to eliminate this 187 bp PCR product through re-design of PCR primers.

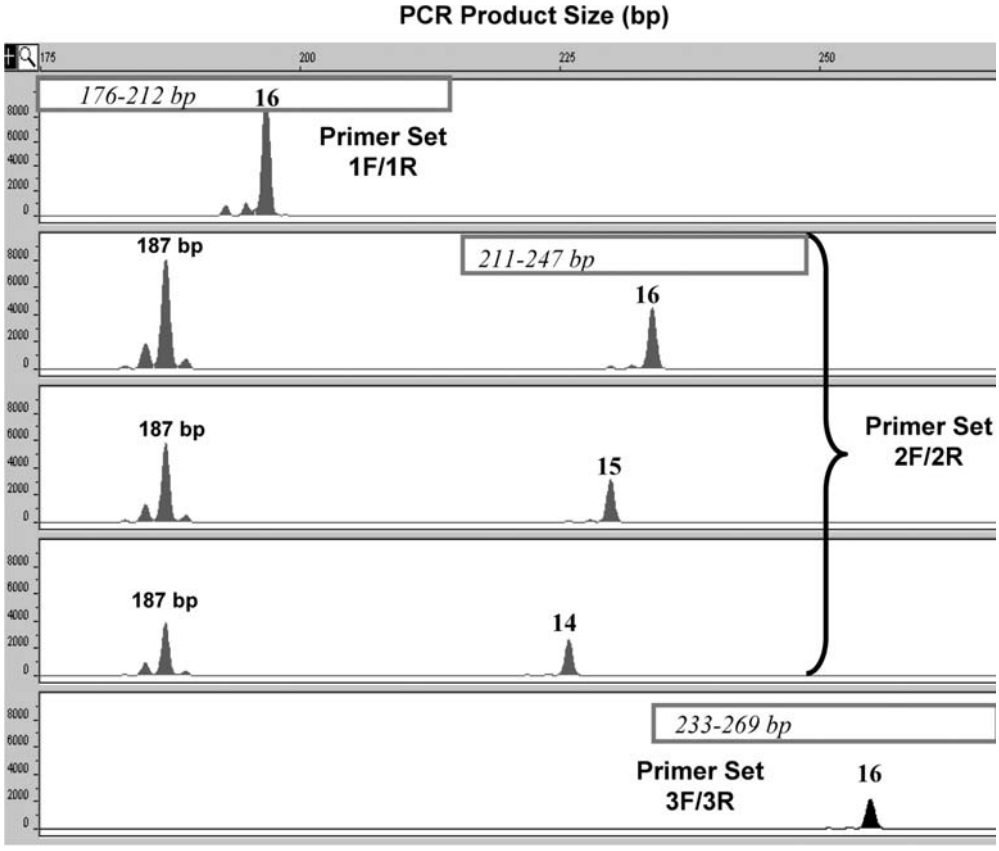

Fig. 1 Electropherograms of samples amplified with primer sets 1F/1R, 2F/2R, 3F/3R listed in Table 1. Amplification of male templates using sets 1F/1R and 3F/3R resulted only in the amplification of the DYS19 locus while primer set 2F/2R produced, in addition to the DYS19-specific product, an invariant amplicon at 187 bp that was specific to the AC006335 region. *Numbers* above the peaks indicate the DYS19 allele call

Information learned from sequence queries and alignments

A BLAST search was conducted with primer set 2F/2R. Primer 2F produced significant alignments with GenBank accessions AC006335, AC007320, AC009491, AC009947, AC010891, AC017019, and AC023342. Primer 2R matched AC006335, AC006991, AC007320, AC010891, AC016694, AC017019, AF140632, and HSY27H39. Further BLAST searches revealed that AC017019, HSY27H39, and AF140632 matched primer 1F and only AC017019 produced a significant alignment with primer 1R'. Thus changing from primer set 1F/1R' to 2F/2R picked up additional GenBank hits. The next step was to see where these sequence regions lay on an assembly of the Y chromosome.

For all sequence accessions listed above, approximately 1,000 bp on either side of the primer matching sites was retrieved from GenBank and used to perform a human BLAT search using the April 2003 assembly of the human genome. These BLAT searches revealed that the primary DYS19 locus region – GenBank accession AC017019, AF140632, and HSY27H39 – occurs at 9.437 mb along the assembled human Y chromosome. However, several other matches of almost 100% homology occurred on other parts of the Y chromosome near the 6 mb region and the 23 mb region (Fig. 2). These searches also uncovered that AC006335 and AC010891 were equivalent as were AC007320 and AC023342. In addition, the AC006335 and AC007320 sequence segments are duplicated and oriented in opposite directions (Fig. 2).

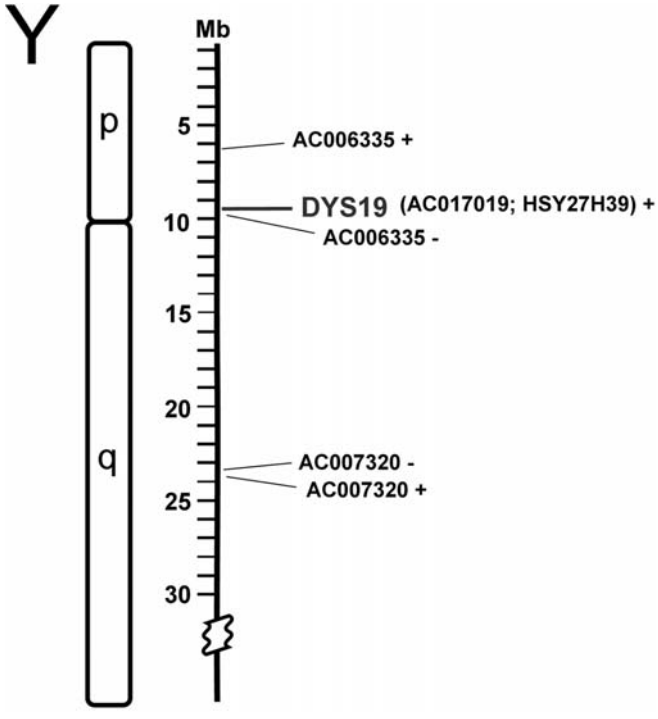

Fig. 2 Illustration of Y-chromosome positions for DYS19 locus and homologous regions discovered during a human BLAT search against the human genome reference sequence. The AC numbers describe sequence contigs present in GenBank with either forward (+) or reverse (–) orientations

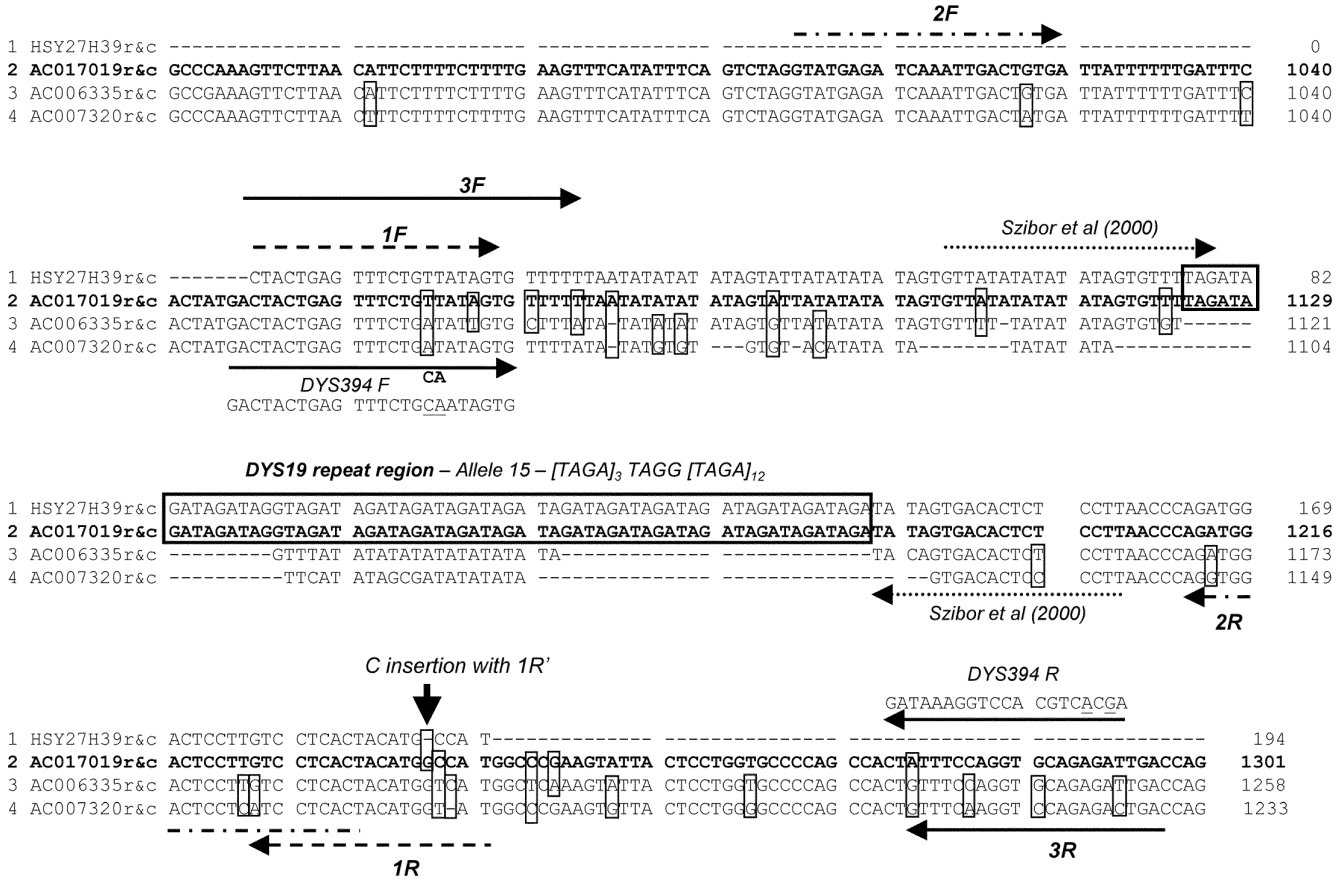

Fig. 3 Sequence alignments for the DYS19 locus (HSY27H39 and AC017019) and its Y homologs (AC006335 and AC007320). Positions of primers used in this study (see Table 1) or other published ones are illustrated as *horizontal arrows*. The sequences are displayed as reverse and complement (r&c) so that "forward" and "reverse" primers and the repeat structure are in a familiar format. The homologs are nearly identical to the flanking regions of DYS19 but are missing the TAGA repeat region (*indicated by dashes*). Sequence differences between the various alignments are *boxed*. The *vertical arrow* near the 5'-end of primer 1R indicates an insertion of a single G in the AC017019 version of DYS19 relative to the original HSY27H39 sequence

An alignment of AC017019, AC006335, and AC007320 against HSY27H39 revealed a high degree of similarity between these sequences (Fig. 3). Primer pairs 1F/1R and 2F/2R are plotted as dashed arrows next to their annealing sites on the aligned sequences. Primers 2F/2R amplify both the DYS19 locus (AC017019) and the Y-homolog from AC006335, which is 42 bp smaller without the TAGA allele 15 repeat region being present. The alignment of AC017019 and HSY27H39 also demonstrates why the 1R' primer with its C insertion improves amplification of the DYS19 locus as mentioned on the Y-STR Haplotype Database website (see http://www.ystr.org).

Primers to target only DYS19

By exploiting sequence differences between the AC006335 and AC007320 homologs we were able to design primer set 3F/3R to specifically target the DYS19 repeat region. These are the primers present in the recently published NIST 20-plex [11] as well as a commercial multiplex [17] (personal communication Benjamin Krenke, Promega Corporation, 15 October 2003). The bottom panel of Fig. 1 demonstrates that only the DYS19 alleles are amplified with primer set 3F/3R. Thus, careful primer design so that the 3'-end of one or both primers match the intended target exactly but has destabilizing mismatches relative to unwanted homologous sequences, can permit allele-specific PCR of the desired locus. Nucleotide differences between the four homologous GenBank entries described here are highlighted by vertical boxes in Fig. 3 to indicate potential differential binding sites for the 3'-ends of DYS19 allele-specific primers.

Impact of duplicated regions on other published primers

Table 2 provides a summary of mismatches between all previously published primer sequences including the ones in this study and the DYS19 locus and its two Y chromosome homologous regions. Most of the previously used primers have perfect matches with respect to the DYS19 locus (GenBank accession AC017019) and mismatches

Table 2 Summary of PCR primer binding characteristics to various regions of the Y-chromosome

Primer	Template		
	AC017019 (r&c) (DYS19)	AC006335 (r&c)	AC007320 (r&c)
1F	Perfect match	A-T (3), T-A (7)	T-A (7)
1R'	Perfect match	C-T (17)	T-C (1), G-A (2), C-T (17), C-del (18)
2F	Perfect match	Perfect match	G-A (3)
2R	Perfect match	Perfect match	A-G (4), T-C (14), G-A (15)
3F	Perfect match	**T-A (1),** T-C (5), A-T (9), T-A (13)	**T-A (1),** T-A (13)
3R	Perfect match	**A-G (1)**	**A-G (1),** C-A (6), G-C (11), T-C (18)
DYS394 F [7]	A-T (7), C-T (8)	A-T (4), A-T (7), C-A (8)	A-T (7), C-A (8)
DYS394 R [7]	T-A (17), C-A (19)	A-G (3), T-A (17), C-A (19)	A-G (3), C-A (8), G-C (13), T-A (17), C-A (19)
Szibor et al. F [22]	del-T (6)	G-T (1), A-T (2), T-G (3), T-G (5), del-T (6), A-del (18), A-T (20)	T-A (4), del-A (6), G-C (7), G-T (9), T-A (22), G-T (23)
Szibor et al. R [22]	Perfect match	T-C (1)	T-C (12)
Wiegand and Kleiber F	Perfect match	A-T (1), T-G (2), T-G (4), A-del (17)	T-A (3), T-A (5), G-C (6), G-T (8), T-A (19), T-A (21), G-T (22), T-A (23), G-T (24)
Wiegand and Kleiber R	Perfect match	T-C (1)	T-C (12)

The three regions of the Y-chromosome under consideration are listed according to their GenBank accession as reverse and complement (r&c) sequences (see Figs. 2 and 3). Mismatches between primer and DNA template are listed as primer nucleotide-template nucleotide (number of bases from 3'-end of primer). Several insertions (ins) and deletions (del) are present as well as nucleotide mismatches. Reverse primer information is described in terms of the top GenBank strand to enable easier comparison to Fig. 3. The closer the mismatch is to the 3' end of a primer, the easier it is to develop allele-specific PCR conditions to prevent amplification of the DYS19 homologous regions AC006335 and AC007320. In set 3F/3R, the 3' nucleotides (*bold font*) for both the forward and reverse primers were designed to be mismatched with AC006335 and AC007320 templates thus promoting amplification of the DYS19 locus

with the homologous regions AC006335 and AC007320, which is likely the reason that no reports have been made regarding these Y chromosome homologous regions before. The closer a mismatch is to the 3'-end of a primer, the more significant it is in terms of generating allele-specific amplification. Note that in the intentional design of the 3F/3R primers to specifically target the DYS19 locus [11], mismatches with AC006335 and AC007320 are seen at the 3'-nucleotide for both the forward and reverse primers. If additional primers are designed for the DYS19 locus in the future, it will be important to keep these homologous regions in mind to avoid unwanted amplification of other regions of the Y-chromosome.

In the literature duplications and even triplications [14] of PCR products for the DYS19 locus have been reported with previously published primers. Since these extra DYS19 amplicons size as true alleles, it is doubtful that they are amplifications of the homologous regions reported here that do not contain the repeat region. However, it is certain that the DYS19 locus and its flanking regions occur more than once in various Y chromosome lineages making it a more complex marker than has been previously thought.

Conclusions

The results presented here illustrate the importance of properly identifying possible areas of homology when attempting to design specific and robust PCR primers. Duplicated regions on the Y chromosome can impact primer binding and result in either a less robust amplification or, as in the case presented here with primers 2F/2R, the amplification of undesired segments of the Y-chromosome. We were able to avoid amplification of Y chromosome homologs by designing primers that permit specific amplification of the DYS19 region in the same manner as previously performed with DYS391 to avoid X chromosome homologs [11, 19]. These new DYS19 primers, designed using information uncovered as part of this study, have been successfully applied to a large Y-STR multiplex assay [11]. In addition, issues with previously published DYS19 primers have been discussed in the context of similar Y chromosome regions.

Acknowledgements This work was supported by research funds from the US National Institute of Justice through Interagency Agreement 1999-IJ-R-094 with the NIST Office of Law Enforcement Standards. Much of this work was performed while R.S. was a graduate student at NIST and American University through funding from the US Air Force. Points of view are those of the authors and do not necessarily represent the position of the U.S. Departments of Justice or Defense. The technical assistance and sugges-

tions of Peter Vallone, David Duewer, Margaret Kline, and Jan Redman from our group at NIST and Alan Redd from the University of Arizona are gratefully acknowledged.

References

1. Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Genet Rev 4:598–612
2. Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. Int J Legal Med 110:118–124
3. Honda K, Roewer L, Knijff P de (1999) Male DNA typing from 25-year-old vaginal swabs using Y chromosomal STR polymorphisms in a retrial request case. J Forensic Sci 44:868–872
4. Roewer L, Arnemann J, Spurr NK, Grzeschik KH, Epplen JT (1992) Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. Hum Genet 89:389–394
5. Roewer L, Epplen JT (1992) Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in casework. Forensic Sci Int 53:163–171
6. 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
7. Redd AJ, Clifford SL, Stoneking M (1997) Multiplex DNA typing of short-tandem-repeat loci on the Y chromosome. Biol Chem 378:923–927
8. Gusmao L, Gonzalez-Neira A, Pestoni C, Brion M, Lareu MV, Carracedo A (1999) Robustness of the Y STRs DYS19, DYS389 I and II, DYS390 and DYS393: optimization of a PCR pentaplex. Forensic Sci Int 106:163–172
9. Shin DJ, Jin HJ, Kwak KD et al. (2001) Y-Chromosome multiplexes and their potential for the DNA profiling of Koreans. Int J Legal Med 115:109–117
10. Schoske R, Butler JM, Vallone PM, Kline MC, Prinz M, Redd AJ, Hammer MF (2001) Development of Y STR megaplex assays. In: Proceedings of the Twelfth International Symposium on Human Identification, 2001. Promega Corporation, Madison WI (http://www.promega.com/geneticidproc/ussymp12proc/contents/butler.PDF)
11. 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
12. Schoske R, Vallone PM, Ruitberg CM, Butler JM (2003) Multiplex PCR design strategy used for the simultaneous amplification of 10 Y chromosome short tandem repeat (STR) loci. Anal Bioanal Chem 375:333–343
13. Kayser M, Caglia A, Corach D et al. (1997) Evaluation of Y-chromosomal STRs: a multicenter study. Int J Legal Med 110: 125–133
14. Santos FR, Gerelsaikhan T, Munkhtuja B, Oyunsuren T, Epplen JT, Pena SDJ (1996) Geographic differences in the allele frequencies of the human Y-linked tetranucleotide polymorphism DYS19. Hum Genet 97:309–313
15. Wang W, Kishida T, Fukuda M, Tamaki Y (1996) The Y-27H39 polymorphism in a Japanese population. Int J Legal Med 109: 157–158
16. Sinha SK, Budowle B, Arcot SS et al. (2003) Development and validation of a multiplexed Y-chromosome STR genotyping system, Y-PLEX 6, for forensic casework. J Forensic Sci 48: 93–103
17. Krenke BE, Fulmer PM, Driftmier-Miller K, Sprecher CJ (2003) The PowerPlex Y system. Profiles in DNA (Promega Corporation) 6:6–9
18. 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
19. Gusmao L, Gonzalez-Neira A, Sanchez-Diz P, Lareu MV, Amorim A, Carracedo A (2000) Alternative primers for DYS391 typing: advantages of their application to forensic genetics. Forensic Sci Int 112:49–57.
20. Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa NJ, pp 365–386
21. Szibor R, Kayser M, Roewer L (2000) Identification of the human Y-chromosomal microsatellite locus DYS19 from degraded DNA. Am J Forensic Med Pathol 21:252–254
22. Wiegand P, Kleiber M (2001) Less is more – length reduction of STR amplicons using redesigned primers. Int J Legal Med 114:285–287