J Forensic Sci, March 2006, Vol. 51, No. 2
doi:10.1111/j.1556-4029.2006.00052.x
Available online at: www.blackwell-synergy.com

Technical Note

Chien-Wei Chang,[1] *Ph.D.; Julio J. Mulero,*[1] *Ph.D.; Bruce Budowle,*[2] *Ph.D.; Lisa M. Calandro,*[1] *M.P.H.; and Lori K. Hennessy,*[1] *Ph.D.*

Identification of a Novel Polymorphism in the X-Chromosome Region Homologous to the DYS456 Locus

ABSTRACT: During an extensive multipopulation study with Y-short tandem repeat (STR) loci, amplified using the AmpFℓSTR® Yfiler™ PCR amplification kit, amplification of a 71 bp fragment was observed in 2.32% of the male samples analyzed ($N = 3141$). By direct sequencing of this fragment, it was determined that the primer binding sequences were identical to those of the DYS456 locus. A T to G single-nucleotide polymorphism (SNP) enabled amplification of the 71 bp fragment. The SNP is located within an X–Y homologous region at Xq21.31 and was observed with the highest frequency within the African American and Sub-Saharan African populations in our study. Presence of SNP on the X chromosome did not interfere with the reliability of typing the DYS456 locus and the other Y-STR loci typeable using the AmpFℓSTR® Yfiler™ PCR amplification kit. Full profiles in a mixture of male:female at 1:4000 were obtained using the current configuration of the AmpFℓSTR Yfiler kit even in the presence of female DNA containing the G variant.

KEYWORDS: forensic science, DNA typing, single-nucleotide polymorphism (SNP), short tandem repeat (STR), polymerase chain reaction (PCR), Yfiler, Y chromosome, Y-STR, DYS456, homolog

Biological evidence from sexual assault cases may contain mixtures of low levels of male DNA among a high background of female DNA. The differential extraction method, a protocol for the separation of DNA present in epithelial cells (or non-sperm cells) from sperm cell DNA, is used routinely on sexual assault evidence but is not applicable to all cell type mixtures encountered in forensic cases (1). If a sample extract cannot be enriched for male origin DNA and the major contribution of DNA is female in origin, autosomal short tandem repeat (STR) marker typing could fail to type the male component. The major component of the mixture (female DNA) can mask the genetic profile of the male contributor by competing for the reagents of the polymerase chain reaction (PCR). With current autosomal STR typing systems, male DNA in mixture samples can be interpreted when it comprises 10% or more of the DNA in the mixture (2–5). In order to meet this challenge, multiplex PCR assays utilizing Y-chromosome-specific markers have been implemented to analyze male/female mixture samples (6–10).

Multiplex STR assay development requires careful primer design to obtain specific and efficient amplification of STR loci. Single-nucleotide polymorphisms (SNPs) in primer binding sites may cause allele dropout (11–18). Nonspecific amplification may occur in an unintended chromosomal location if the region is highly homologous to the targeted STR primer binding sites. Initial selection of primer sequences is facilitated by searching public databases, such as dbSNP and human genome variation base (HGVbase), which contain useful information on human genomic variations (19,20). Although extremely useful, this *in silico* approach for primer design must be verified *in vitro* with validation studies on samples of different population groups. However, if the allele containing a polymorphism is rare or prevalent in a particular population group(s), identifying such genomic variants requires testing a sufficient number of individuals.

In this report, we describe a T to G polymorphism located within a Y homologous region on the X chromosome that resulted in the amplification of a 71 bp fragment outside the allele sizes of all Y-STR loci typeable with the AmpFℓSTR® Yfiler™ PCR amplification kit (Applied Biosystems, Foster City, CA). An allele-specific amplification assay was developed to detect this SNP, and its allele frequency was determined in different populations groups. In addition, the performance of the AmpFℓSTR® Yfiler™ kit was evaluated in male/female mixtures containing female genomic DNA with the SNP variant.

Materials and Methods

Samples

Participants in the AmpFℓSTR® Yfiler™ haplotype database project contributed population sample data (21). Population origin was established through self-reporting. DNA samples were also obtained from SeraCare Life Sciences (Oceanside, CA), and the polymorphism discovery resource panel (PDR) was purchased from Coriell cell repositories (Camdon, NJ). DNA samples from SeraCare or Coriell cell repositories were quantified on the ABI PRISM® 7000 SDS instrument (Applied Biosystems) prior to amplification using Quantifiler™ Y Human Male DNA and/or

[1] Applied Biosystems, Foster City, CA, 94404.
[2] Federal Bureau of Investigation Laboratory, Quantico, VA 22135.

Received 6 July 2005; and in revised form 27 Sept. 2005; accepted 1 Oct. 2005; published 9 Feb. 2006.

Quantifiler™ Human DNA Quantification kits (Applied Biosystems) according to the standard protocols provided by the manufacturer.

PCR Amplification

The AmpFℓSTR® Yfiler™ PCR was performed in a total volume of 25 µL comprising of 9.2 µL AmpFℓSTR® Yfiler™ buffer, 5 µL of AmpFℓSTR® Yfiler™ primer mix, 4 units of AmpliTaq Gold® DNA polymerase (5 units/µL), and 10 µL of target DNA (Applied Biosystems). All PCRs were performed in an ABI PRISM® GeneAmp® 9700 Gold-plated or Silver block Thermal Cycler (Applied Biosystems) using the 9600 emulation mode with an 11-min activation step at 95°C followed by 30 cycles of denaturation at 94°C for 1 min, annealing at 61°C for 1 min, and extension at 72°C for 1 min. A final extension was performed at 60°C for 80 min. PCR products were separated and detected on an ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems) following the manufacturer's recommendations. Prior to electrophoresis, 1 µL of the amplified product or allelic ladder and 0.3 µL of GeneScan™-500 LIZ® size standard (Applied Biosystems) were added to 8.7 µL of deionized Hi-Di™ formamide (Applied Biosystems), denatured at 95°C for 3 min, and then chilled on ice for 3 min. Samples were injected for 10 sec at 3 kV in performance optimized polymer (POP-4™) (Applied Biosystems) using the GeneScan36vb_POP4DyeSetG5Module. The data were collected using the ABI PRISM® 3100 Data Collection Software v1.1 (Applied Biosystems). Electrophoresis results were analyzed with GeneMapper® ID software v3.2 (Applied Biosystems). Allele peaks were called when the peak heights were greater than or equal to 50 relative fluorescence units (RFUs).

Sequencing

Sequencing reactions were performed to analyze the observed 71 nucleotide (nt) fragment. A region highly homologous to the DYS456 locus on the X chromosome was amplified using the forward primer 5′-GGACCTTGTGATAATGTAAGATA and the reverse primer 5′-TAGAGGGACAGAACTAATGGAATATA. PCR amplifications were carried out using the protocol described above, except that 150 nM of each primer were used. The resulting amplicon sizes ranged from 60 to 70 bp, because of the number of AT dinucleotide repeats adjacent to this SNP. The PCR products were cloned using the TOPO TA Cloning® kit (Invitrogen, Carlsbad, CA) and sequenced with BigDye® Terminator v 1.1 Cycle Sequencing Kit (Applied Biosystems) using M13 forward and reverse primers following the recommendations of the manufacturer. Capillary electrophoresis was performed on an ABI PRISM® 3100 Genetic Analyzer with Data Collection Software v1.1. The resulting sequencing data were analyzed with the software DNA Sequencing Analysis 3.7 (Applied Biosystems).

Allele-Specific Primer Extension Assay

A primer extension assay was developed to further evaluate the G mutation in females. The PCR product was generated from nine females initially identified to have the G mutation. Ten microliters of the PCR product were incubated with 4 µL of ExoSAP-IT® (USB, Cleveland, OH) for 15 min at 37°C followed by 15 min at 80°C for enzyme inactivation. The ABI PRISM® SNaPshot® kit reaction was performed in a GeneAmp® PCR System 9700 thermocycler following the recommendations of the manufacturer (Applied Biosystems). Each reaction contained 5 µL of SNaPshot ready reaction mix, 1 µL of purified PCR products and 1 µL of the 0.2 µM extension primer 5′-GGACCTTGTGATAATGTAAGATA. The reaction mixture was subjected to 25 single base extension cycles of denaturation at 96°C for 10 sec, annealing at 59°C for 5 sec, and extension at 60°C for 30 sec. The reactions were then treated with 1 µL of shrimp alkaline phosphatase for 60 min at 37°C, followed by 15 min at 80°C for enzyme inactivation. The PCR products (0.5 µL) were mixed with 9 µL of HiDi™ formamide and 0.5 µL of GeneScan®-120 LIZ® dye-labeled size standard (Applied Biosystems) and electrophoresis was performed on an ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems) using the SNP36_POP4 default module. The results were analyzed using ABI PRISM® GeneScan® Analysis Software v3.7.1 (Applied Biosystems).

Male:Female Mixture Experiments

To evaluate the effect of this G mutation on Y-STR amplification in mixture samples, male control DNA 007 and various female DNA samples, with or without this G variant on the X chromosome, were combined in a 1:4000 ratio. Mixtures containing 125 pg of male control DNA 007 and 500 ng of female DNA were prepared. Female control 9947A and two female samples #187 and #333, heterozygous for the SNP variant, were used for the mixture study. The mixture samples were amplified and analyzed as described in the PCR amplification section.

Results and Discussion

A novel FAM dye-labeled peak with an electrophoretic mobility consistent with that of 71 nt fragment (70 nt plus one nt because of

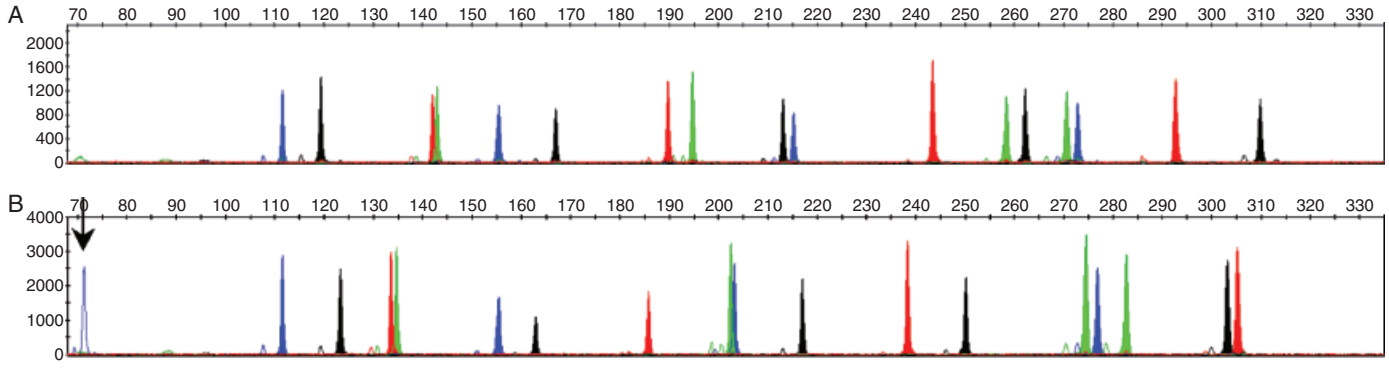

FIG. 1—*Detection of a 71 nucleotide (nt) fragment in a male sample amplified with the Yfiler kit. (A) Male control 007 (B) African American male DNA sample. One nanogram of DNA was amplified with AmpFℓSTR® Yfiler™ PCR amplification kit, and electrophoresis was performed on a 3100 Genetic Analyzer. The arrow indicates the 6-FAM dye-labeled peak with a mobility of 71 nts. The 17 Y-short tandem repeat (STR) allele peaks in the read region are highlighted.*

```
DYS456 locus          GGACCTTGTGATAATGTAAGATAGATAGATAGATAGATAGATAGATAGATA....GATATTCCATTAGTTCTGTCCCTCTA
                                          *
Allele G on X Chr.    GGACCTTGTGATAATGTAAGATAGATATATATATATATA.....................TATATTCCATTAGTTCTGTCCCTCTA
Allele T on X Chr.    GGACCTTGTGATAATGTAAGATATATATATATATATATA.....................TATATTCCATTAGTTCTGTCCCTCTA
```

FIG. 2—*Nucleotide sequence alignment of the DYS456 locus and its homologous region on the X chromosome. The asterisk indicates the polymorphic position in the homologous region. The forward and reverse primer sequences used for amplification of the DYS456 locus are underlined.*

TABLE 1—*Allele frequencies in different populations.*

Allele	African American ($N = 802$)	Sub-Saharan African ($N = 59$)	Caucasian ($N = 1156$)	Hispanic ($N = 480$)	Native American ($N = 106$)	Asian ($N = 330$)	Filipino ($N = 105$)	Vietnamese ($N = 103$)
T	0.923	0.898	0.998	0.994	1	1	1	1
G	0.077	0.102	0.002	0.006	0	0	0	0

N, number of chromosomes.

nontemplate adenylation) (Fig. 1) was identified in some male samples during the course of a population study using the AmpFℓSTR® Yfiler™ PCR amplification kit. Primer subtraction experiments identified that the DYS456 primers generated this fragment (data not shown). Singleplex amplification reactions with the DYS456 specific primers confirmed this finding (data not shown). The 71 nt PCR product from 13 individuals was cloned and sequenced. A T/G SNP was discovered residing within the X–Y homology region at Xp21.3 (position: 89,882,184) (20,22,23). Its flanking sequence has a high degree of homology with the DYS456 locus-flanking region. However, this site contains a polymorphic AT dinucleotide repeat sequence instead of the tetranucleotide repeat TAGA sequence found on the Y chromosome. The sequence alignment in Fig. 2 shows that the G nucleotide resides at the 3′ end of the DYS456 forward primer. This demonstrates why the primer was extended in individuals carrying this variant on the X chromosome. The T nucleotide, found in the majority of the samples analyzed to date using the AmpFℓSTR® Yfiler™ kit, prevents amplification of this X-chromosome region.

In the analysis of 3141 male DNA samples from different populations genotyped with the AmpFℓSTR® Yfiler™ kit, the 71 nt fragment was observed in the profiles of 73 individuals. The allele frequency for the minor allele G was estimated at 7.7% and 10.2% in the African American and Sub-Saharan African populations, respectively (Table 1). Caucasian and Hispanic populations displayed much lower allele frequencies (0.2% and 0.6%, respectively). The presence of the G allele in Caucasian and Hispanic groups may be because of population admixture. The polymorphism was not observed in the Asian and Native American populations studied, possibly because of the smaller sample size compared with Caucasians.

To further investigate the diversity of this polymorphism, we tested 214 female samples from the polymorphism discovery resource panel consisting of U.S. residents with ancestry from all the major regions of the world (24). Nine out of the 214 individuals tested had the variant G, and based on our allelic-specific assay, (Fig. 3) all were heterozygous for this SNP. The unequal peak heights observed for the two alleles in heterozygous individuals are

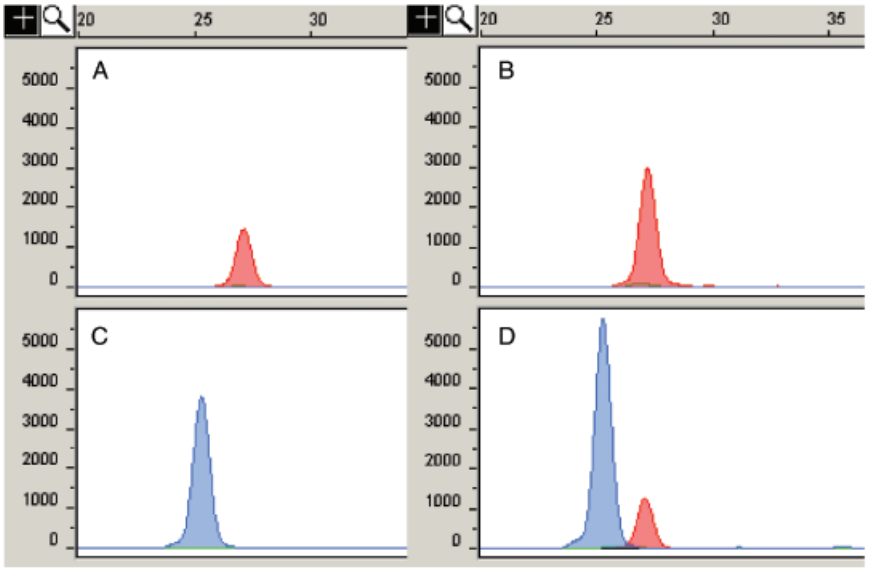

FIG. 3—*Electropherograms of the allele-specific primer extension assays for the T/G single-nucleotide polymorphism (SNP). Blue peak = G allele, red peak = T allele: (A) male control 007 with one copy of the allele T; (B) female control 9947A with two copies of the allele T; (C) male sample #1 with one copy of the allele G; (D) female sample #187 with the alleles G and T.*

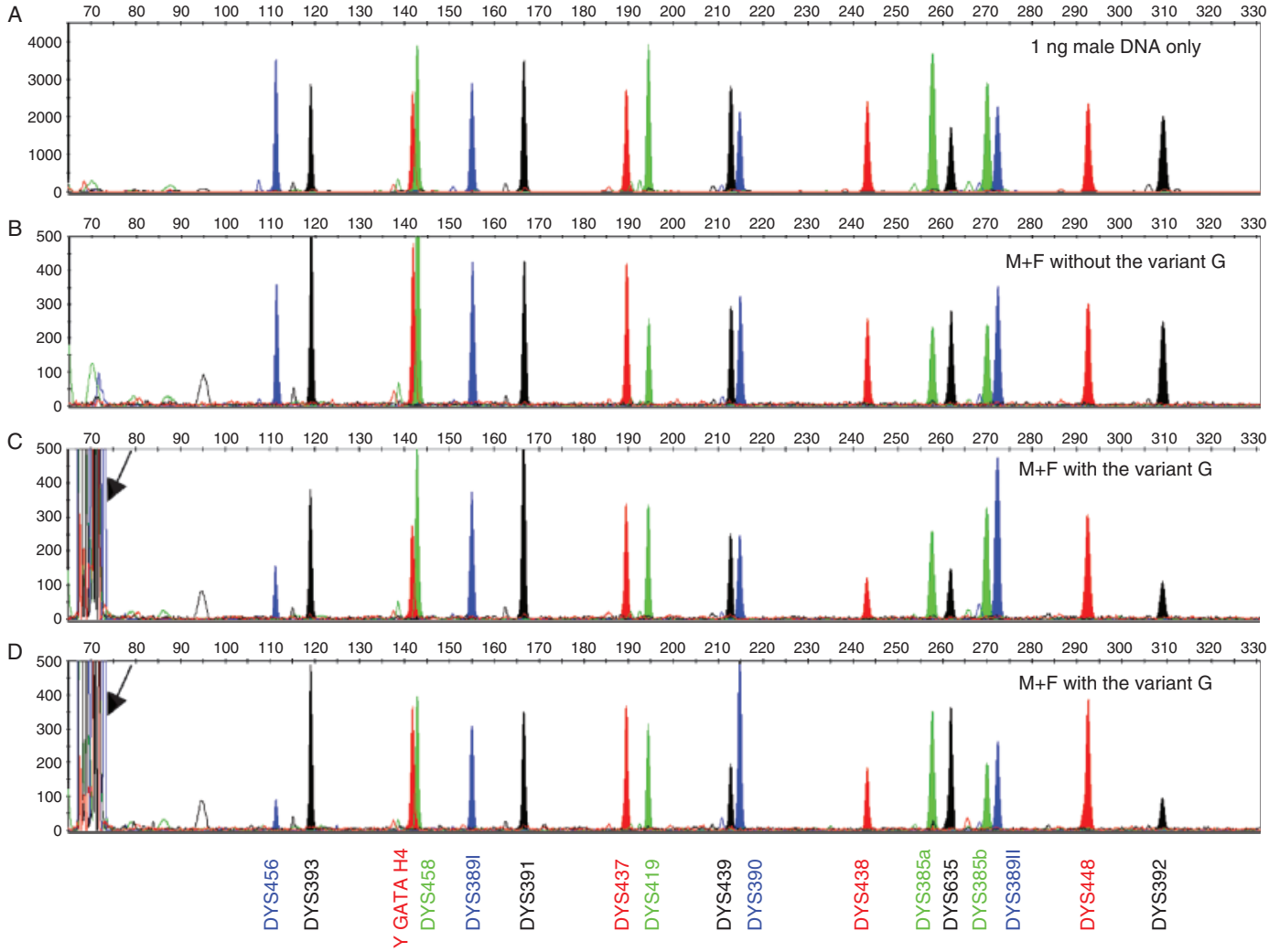

FIG. 4—*Male:female (1:4000) mixture studies using the AmpFℓSTR® Yfiler kit. Electrophoretic profiles of 1 ng male control DNA 007 only (A) or 125 pg male control DNA 007 in the presence of 500 ng female DNA 9947A (B), #187 (C), and #333 (D). Female control DNA 9947A is homozygous for the allele T. Female DNAs #187 and #333 are heterozygotes containing the G mutation. Arrows indicate the off-scale FAM dye-labeled 71-nt length fragment peaks. Y-STR allele peaks are highlighted in the red region.*

most likely because of differences in efficiency of incorporation of the dye-labeled ddNTPs with the primer extension assay (25).

All males tested carrying the G variant yielded a 71 nt long fragment, which contains 10 repeats of the dinucleotide motif. However, the X homology with the allele T showed length variation among females from the polymorphism discovery resource panel. The number of AT repeats varied from 10 to 14. Thus, it is probable that the G variant arose in a 10 AT repeat allele. Although in this study only a 71 nt length fragment was observed in individuals with the G allele, some people with the G allele may harbor a slightly longer or shorter fragment because of variation in the number of dinucleotide repeats.

TABLE 2—*DYS456 peak heights in male/female (M/F) mixture.**

M/F Mixture	Female DNA Used in M/F Mixture†			
	9947	#187	#333	#404
0.5 ng male 500 ng female	1223	442	472	390
0.25 ng male 500 ng female	409	273	238	222
0.125 ng male 500 ng female	374	181	110	139

*Average peak heights in relative fluorescence units from duplicate measurements.

†9947A, female control DNA without the variant G; #187, #333, #404, female DNAs with the variant G.

Based on the allele frequency estimates for the SNP, approximately 14.3% and 0.6% of African American females would be expected to be heterozygous and homozygous, respectively, for G variant. Therefore, there will be cases of male:female mixtures where the female component is large and it contains the SNP. In such situations, reagents will be consumed beyond expectations compared with samples that do not contain the SNP. Undesired consumption of primers and other reagents in the PCR may affect sensitivity or signal quality of multiplex STR analysis; therefore, the performance of the AmpFℓSTR® Yfiler™ PCRs was evaluated for typing male DNA in the presence of high copy numbers of this G variant by performing male:female mixture studies. Even in the presence of high copy numbers of the variant G ($\sim 8.3 \times 10^4$ copies/500 ng female DNA from heterozygotes),

full profiles and concordant genotyping results for the 17 Y-STR loci were obtained with 125 pg of male DNA (Fig. 4). As expected, the mixture samples containing high amounts of female DNA heterozygous for the SNP had off-scale peaks approximating 71 nts, representing the expected PCR product amplified from the DYS456 homologous region. This consumption of primers caused, on an average, a 55% reduction in peak heights of the DYS456 allele in the mixture studies (Table 2). However, even at the highest mixture ratio of 1:4000, which contained only 125 pg of male DNA, DYS456 peak heights still ranged from 110 to 181 RFUs and were sufficient for peak detection and accurate genotyping results. This effect is expected to be slightly greater if the female component is homozygous for the SNP. It is possible that at DNA concentrations less than or equal to 125 pg or in the presence of high concentration of female DNA containing the G variant the DYS456 may fall below the peak detection threshold, which depends on instrument sensitivity, stochastic effects, and consumption of primers. However, consumption of primers and reagents did not affect the peak heights of the other 16 Y-STR loci. The average peak heights of the FAM™-, VIC®-, NED™-, and PET®-labeled loci (excluding DYS456) in the 1:4000 mixtures were very similar to the mixtures containing female control DNA 9947A (344, 389, 299, and 321 RFUs, respectively, compared with 305, 400, 296, and 370 RFUs). Therefore, even without the DYS456 locus the remaining 16 loci can be used to search a haplotype database and are still capable of providing high discriminatory capacity.

In summary, our results demonstrate that the presence of the 71 nt fragment has no impact on the interpretation of a Y-STR haplotype in males. Even when a small mass of male DNA is admixed with a large mass of female DNA containing the G variant, the sample will still yield the correct Y STR haplotype using the current configuration of the AmpFℓSTR® Yfiler™ kit. Notably, in the presence of 4000-fold excess female DNA that carries the SNP, 125 pg of DNA were typed correctly at the DYS456 locus, as well as at all other Y-STR loci. Lastly, in samples containing male DNA only, this dinucleotide repeat on the X chromosome can be used as an additional marker, providing more information for exclusion or for assessing statistical weight of evidence samples.

References

1. Gill P, Jeffreys AJ, Werrett DJ. Forensic application of DNA 'fingerprints'. Nature 1985;318:577–9.
2. Frank WE, Llewellyn BE, Fish PA, Riech AK, Marcacci TL, Gandor DW, et al. Validation of the AmpFℓSTR® Profiler Plus™ PCR Amplification Kit for use in forensic casework. J Forensic Sci 2001;46:642–6.
3. LaFountain MJ, Schwartz MB, Svete PA, Walkinshaw MA, Buel E. TWGDAM validation of the AmpFℓSTR® Profiler Plus™ and AmpFℓSTR® COfiler™ STR multiplex systems using capillary electrophoresis. J Forensic Sci 2001;46:1191–8.
4. Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ, et al. Validation of a 16-locus fluorescent multiplex system. J Forensic Sci 2002;47:773–85.
5. Collins PJ, Hennessy LK, Leibelt CS, Roby RK, Reeder DJ, Foxall PA. Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFℓSTR® Identifiler™ PCR amplification kit. J Forensic Sci 2004;49:1265–77.
6. Grignani P, Peloso G, Fattorini P, Previdere C. Highly informative Y-chromosomal haplotypes by the addition of three new STRs DYS437, DYS438 and DYS439. Int J Legal Med 2000;114:125–9.
7. Parson W, Niederstatter H, Brandstatter A, Berger B. Improved specificity of Y-STR typing in DNA mixture samples. Int J Legal Med 2003;117: 109–14.
8. Budowle B, Sinha SK, Lee HS, Chakraborty R. Utilization of Y-chromosome short tandem repeat haplotype in forensic application. Forensic Sci Rev 2003;15:153–62.
9. Shewale JG, Nasir H, Schneida E, Gross AM, Budowle B, Sinha SK. Y-chromosome STR system, Y-PLEX™ 12, for forensic casework: development and validation. J Forensic Sci 2004;49:1278–90.
10. Budowle B, Adamowicz M, Aranda XG, Barna C, Chakraborty R, Cheswick D, et al. Twelve short tandem repeat loci Y chromosome haplotypes: genetic analysis on populations residing in North America. Forensic Sci Int 2005;150:1–15.
11. Gusmao L, Amorim A, Prata MJ, Pereira L, Lareu MV, Carracedo A. Failed PCR amplifications of MBP-STR alleles due to polymorphism in the primer annealing region. Int J Legal Med 1996;108:313–5.
12. Kline MC, Jenkins B, Rodgers S. Non-amplification of a vWA allele. J Forensic Sci 1998;43:250–1.
13. Boutrand L, Egyed B, Furedi S, Mommers N, Mertens G, Vandenberghe A. Variations in primer sequences are the origin of allele drop-out at loci D13S317 and CD4. Int J Legal Med 2001;114:295–7.
14. Han GR, Song ES, Hwang JJ. Non-amplification of an allele of the D8S1179 locus due to a point mutation. Int J Legal Med 2001;115: 45–7.
15. Nelson MS, Levedakou EN, Matthews JR, Early BE, Freeman DA, Kuhn CA, et al. Detection of a primer-binding site polymorphism for the STR locus D16S539 using the Powerplex® 1.1 system and validation of a degenerate primer to correct for the polymorphism. J Forensic Sci 2002;47: 345–9.
16. Heinrich M, Muller M, Rand S, Brinkmann B, Hohoff C. Allelic drop-out in the STR system ACTBP2 (SE33) as a result of mutations in the primer binding region. Int J Legal Med 2004;118:361–3.
17. Leibelt C, Budowle B, Collins P, Daoudi Y, Moretti T, Nunn G, et al. Identification of a D8S1179 primer binding site mutation and the validation of a primer designed to recover null alleles. Forensic Sci Int 2003;133:220–7.
18. Budowle B, Sprecher CJ. Concordance study on population database samples using the PowerPlex™ 16 Kit and AmpFℓSTR® Profiler Plus™ Kit and AmpFlSTR® COfiler™ Kit. J Forensic Sci 2001;46:637–41.
19. Fredman D, Munns G, Rios D, Sjoholm F, Siegfried M, Lenhard B, et al. HGVbase: a curated resource describing human DNA variation and phenotype relationships. Nucleic Acids Res 2004;32(Database issue): D516–9.
20. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Church DM, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2005;33(Database issue):D39–45.
21. Mulero JJ, Chang CC, Calandro LM, Green RL, Li Y, Johnson CL, et al. Development and validation of the AmpFℓSTR® Yfiler™ PCR Amplification Kit: a male specific, single amplification 17 Y-STR multiplex system. J Forensic Sci 2006;51:64–75.
22. Mumm S, Molini B, Terrell J, Srivastava A, Schlessinger D. Evolutionary features of the 4-Mb Xq21.3 XY homology region revealed by a map at 60-kb resolution. Genome Res 1997;7:307–14.
23. Sargent CA, Boucher CA, Blanco P, Chalmers IJ, Highet L, Hall N, et al. Characterization of the human Xq21.3/Yp11 homology block and conservation of organization in primates. Genomics 2001;73:77–85.
24. Collins FS, Brooks LD, Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variation. Genome Res 1998;8:1229–31.
25. Brandis JW. Dye structure affects Taq DNA polymerase terminator selectivity. Nucleic Acids Res 1999;27:1912–8.

Additional information and reprint requests:
Chien-Wei Chang, Ph.D.
Applied Biosystems
850 Lincoln Centre Dr., M/S 404-3
Foster City, CA 94404
E-mail: changcw@appliedbiosystems.com