Development of Y STR Megaplex Assays

Richard Schoske[1, 2], John M. Butler[1], Peter M. Vallone[1], Margaret C. Kline[1], Mechthild Prinz[3], Alan J. Redd[4], and Michael F. Hammer[4]

[1]National Institute of Standards and Technology, Biotechnology Division, 100 Bureau Drive, Mail Stop 8311, Gaithersburg, MD 20899 USA
[2]American University, Department of Chemistry, Washington, DC 20016 USA
[3]New York City Office of Chief Medical Examiner, Department of Forensic Biology, 520 First Avenue, New York, NY 10016
[4]University of Arizona, Division of Biotechnology, Tucson, AZ 85721

Corresponding Author: John M. Butler, National Institute of Standards and Technology, Biotechnology Division, 100 Bureau Drive, Mail Stop 8311, Building 227, Room A243, Gaithersburg, MD 20899; Tel: 301-975-4049; Fax: 301-975-8505; email: john.butler@nist.gov

ABSTRACT

Y chromosome short tandem repeat markers have a number of applications in human identity testing including typing the perpetrator of sexual assault cases without differential extraction and tracing paternal lineages for missing person investigations. In order for Y STR systems to become more widely accepted within the forensic DNA typing community, population studies and robust assays are required. We have focused on developing new Y STR multiplexes that can provide a high degree of discrimination between unrelated males.

The Y STR markers we are working with include the hexanucleotide repeat DYS448, pentanucleotides DYS438, DYS446, DYS447, DYS450, tetranucleotide repeat DYS19, DYS385 I/II, DYS389I/II, DYS390, DYS391, DYS393, DYS437, DYS439, DYS441, DYS442, GATA A7.1, GATA H4, G09411, trinucleotides DYS388, DYS392, DYS426, and the dinucleotide repeat YCAII. Primers for the markers DYS385, DYS389, and YCAII target duplicated regions of the Y chromosome and thus can produce two polymorphic peaks with each primer set. The multiplexes we are developing are the first to include all of the European "extended haplotype" (11 PCR products) in a single reaction.

We have compared our results to the commercially available Y-Plex™ 6 kit from Reliagene that amplifies 6 Y STRs. In order to improve the power of discrimination for Y chromosome tests, we have developed strategies for rapidly preparing multiplex PCR assays that utilize both four and five dye chemistries for detection and permit simultaneous amplification of 20 or more Y STR markers in a single reaction. Primer design issues will be reviewed, as will our efforts to avoid any homology with X chromosome sequences. Primers have been redesigned from previously published work with these Y STR markers in order to make them more compatible in a multiplex amplification. Results will be shown using our new Y STR multiplex from 3 different laboratories. In addition, allele ranges for each of the Y STR markers have been well characterized in a diverse set of world samples.

INTRODUCTION

Since first being described in 1988 [1], PCR multiplexing has been applied in many areas of DNA testing, including the analysis of deletions, mutations, and short tandem repeats [2-4]. Multiplex PCR primer design and optimization is more of a challenge than designing singleplex PCR because multiple primer annealing events need to occur at the same annealing conditions without interfering with one another. Additionally, extensive optimization is normally required to obtain a good balance between amplicons of the various loci being amplified [5,6].

Multiplex PCR assays must be run under the same PCR conditions. Thus, the primers used should have similar annealing characteristics (e.g. melting temperatures (T_m)) and not interact with one another. Excessive regions of complementarity between primers should be avoided to prevent the formation of primer dimers and hairpins that will cause the primers to bind one another instead of the template DNA. Through stringent initial primer selection we minimize the time consuming and often costly process of optimization. Following careful primer design, the only experimental condition in this study that is used to

optimize the multiplex PCR system is the primer concentration. Using the multiplex primer mixture design and optimization strategy presented here we have successfully developed and tested a 20plex for short tandem repeats (STRs) located on the Y chromosome.

The Y chromosome has been recognized as a useful tool in forensic science, evolutionary studies and human identity testing [7-9]. The literature reports a number of polymorphic short tandem repeat loci [9-16]. Due to the fact that a majority of the Y chromosome does not recombine during meiosis, the Y chromosome is transferred unchanged, barring mutations, from generation to generation establishing a paternal lineage. Thus, the product rule cannot be used to increase the power of discrimination as when using multiple autosomal markers. A useful way to increase the discriminatory power between unrelated male samples is to use multiplex PCR assays, in which a number of Y chromosome STR loci can be simultaneously examined. Over the past 4 years Y STR multiplexes have been designed and tested that at most only amplify five or six loci simultaneously [14,17]. We have made an effort to increase the level of multiplexing for Y STRs and recently have successfully demonstrated a multiplex capable of the simultaneous amplification of 10 Y STR loci [18]. We present here work on a multiplex capable of simultaneously amplifying 20 unique Y STR PCR products using 17 primer pairs. The Y STR markers DYS389, DYS385, and YCAII are duplicated along the Y chromosome and produce two amplicons from each individual, locus-specific primer pair. Our Y STR 20 plex includes all of the Y STRs that make up the "extended haplotype" used in Europe (DYS19, DYS385I/II, DYS389I/II, DYS390, DYS391, DYS392, DYS393, and YCAII [9]) plus additional polymorphic Y STRs DYS437 [11], DYS438 [11], DYS439 [11], DYS447 and DYS448 [Redd *et al*. in preparation], DYS388 [9,12], GATA A7.1 [10], and GATA H4 [10].

Data from a number of studies performed using the Y STR 20 plex will be presented. These studies include sensitivity and mixture analysis, an intralaboratory comparison to the commercially available Y-PLEX™ 6 kit from ReliaGene Technologies, an interlaboratory comparison of identical male DNA samples, and a prediction of the discriminatory power of the new Y STR 20plex. New Y STR multiplexes will help foster the expansion of current DNA databases and may prove vital to solving forensic cases more expeditiously.

MATERIALS AND METHODS

The allele size ranges of the Y STR 20plex listed in Table 1 and illustrated in Figure 1 were defined through extensive literature searches and testing of the Y chromosome consortium (YCC) panel. The YCC panel is made up of 74 male samples composed of individuals from diverse populations from around the world. A comprehensive listing of Y chromosome literature may be found in STRBase available at http://www.cstl.nist.gov/biotech/strbase [19]. Primers were then designed using Primer 3 [20] with defined PCR product sizes to match the required allele range. Candidate primers were screened for potential primer-cross reactions using a custom designed program written in Visual Basic [21]. Primers were purchased from MWG Biotech (High Point, NC) or Applied Biosystems (Foster City, CA) and were quality control tested prior to further use to confirm proper synthesis [22].

PCR primer sequences and amplification conditions for the Y STR 20 plex will be described elsewhere (Schoske, et al., in preparation). The Y-PLEX™ 6 kit amplification conditions were performed as described by the manufacturer in the Y-PLEX™ 6 instruction manual version 1.0 (ReliaGene Technologies, New Orleans, LA). The samples amplified with the Y-PLEX™ kit were run on the ABI 310 and those samples amplified with the Y STR 20plex were run on the ABI 3100 Genetic Analyzer (Applied Biosystems) according to the following sample set-up and analysis protocols.

Sample Set-up for analysis using the Y-PLEX™ 6 kit on the ABI 310
4 dye chemistry
Sample vessel contains
0.75 µL GS 500 ROX (Applied Biosystems, P/N 401734
18.25 µL Hi-Di™ formamide (Applied Biosystems, P/N 4311320)
1.0 µL PCR product

Sample Set-up for analysis using the Y STR 20 plex on the ABI 3100
5dye chemistry
Sample vessel contains
0.40 µL GS500 LIZ (Applied Biosystems, P/N 4322682)
18.60 µL Hi-Di formamide
1.0 µL PCR product

Sample Analysis for ABI 310
Use POP™ –4 (Applied Biosystems)
Use Module [GS STR POP 4 A (1ml)
Filter set A
Injection: 5 seconds @ 15kv
Sample run time is at least 26 minutes per sample

Sample Analysis for ABI 3100
Set-up G5 matrix file
Use POP™ –4
BioLims Project: 3100_Project
Dye Filter Set: G5
Run Module: GeneScan36_POP4
Analysis Module: GS500Analysis

Design Strategy

The process of multiplex PCR primer mixture design testing and optimization is illustrated in Figure 2. It starts with the selection of the loci to be examined and ends with the empirical testing of the multiplex PCR primer mixture. This strategy focuses on the careful design of the primers using in part publicly available software.

Selection of Loci and Allele Range Determination

Over 200 publications from the literature were examined in order to determine which Y STR markers were most commonly used and therefore should be included in our multiplex. The Y STR Markers DYS447 and DYS448 were newly discovered as part of an ongoing effort at the University of Arizona to characterize polymorphic Y STR markers [Redd *et al*., in preparation]. The loci used in the construction of the Y STR 20plex are given in Table 1. After selecting the loci, the known allele ranges were determined by examining the literature for the latest population data for the respective loci [9-16]. The preliminary allele ranges for DYS447, and DYS448 were experimentally determined using the YCC panel [Redd *et al*. in preparation]. The allele ranges for the other loci were reviewed and/or adjusted due to the results obtained from the YCC panel. Following Y STR selection, the sequence for each locus was obtained from GenBank® (www.ncbi.nlm.nih.gov) by performing a standard nucleotide BLAST (Basic Local Alignment Search Tool) search. A search for each locus was completed using previously published primer sequences obtained from the references listed in Table 1. Approximately half of the loci used in the construction of the Y STR 20plex returned multiple GenBank® accession numbers. For example, a BLAST search using the DYS391 forward primer described in the Genome Data Base (http://gdbwww.gdb.org/) indicated significant homology for regions of both the X and Y chromosome as has been reported previously [23]. A sequence alignment of the X and Y homologs was performed using the Baylor College of Medicine (BCM) search launcher located at http://searchlauncher.bcm.tmc.edu:9331/multi-align/multi-align.html. Analysis

of this DYS391 sequence alignment revealed that several differences existed between the X and Y homologs. Primers were designed to target the Y chromosome portion of the DYS391 sequence by exploiting differences between the X and Y sequences so that the PCR would be male-specific.

Once all of the sequences for all of the chosen loci were examined and the allele ranges and size ranges were determined a schematic of the multiplex was laid out (Figure 1). The allele and size ranges for each locus were determined using the GenBank® accession sequences as a standard reference point. For example, GenBank® accession number AC002992, which contains the DYS439 sequence, has 13 GATA repeats. Template DNA containing 13 GATA repeats amplified using the primers reported by Ayub and coworkers should yield an amplicon 252 bp in length. The allele range from the literature for DYS439 is 9-14 repeats. Thus, if the Ayub *et al.* [11] primers were used, amplicons in the size range of 236-256 bp would be expected. Since the size ranges for DYS447 and DYS448 were not available in the literature, the size range was determined empirically by running the YCC samples [Redd *et al.* in preparation].

Primer Design

The Primer 3 software available at www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi [20] was used to design the primers in the Y STR 20plex. Primers were designed to have similar calculated melting temperatures (T_m) so that all primers would anneal with a similar efficiency during the annealing step of PCR. Following primer selection using Primer 3, primer-primer comparisons were performed looking for potential cross reactivity. Multiplex primer comparison was accomplished using an in-house Visual Basic computer algorithm that enables a pair wise comparison of each primer in the multiplex PCR primer mixture [21]. The number of possible primer checks that need to be completed is equal to $(2n^2+n)$ where n is the number of primer pairs to be tested [21]. In the case of the Y STR 20plex there are 17 primer pairs and thus 595 possible primer interactions (Figure 3). An alignment score is used to reflect the number of complementary base pairs minus the number of mismatched base pairs between two primers. No effort is made to weigh G-C pairs higher than A-T pairs.

Fluorescent Dye Labels

To maintain consistency the forward primers are labeled with a fluorescent dye to generate a labeled PCR product for detection purposes. For the work with the Y STR 20plex on the ABI Prism® 3100 Genetic Analyzer 16 capillary array system, 6-FAM (blue), VIC (green), NED (yellow), and PET (red) were the fluorescent dyes covalently linked to the 5' end of the forward primers and the fluorescent dye LIZ (orange) for the labeling of the internal standard. The above dyes were chosen because they are the same one's used in the ABI Identifiler™ kit used to type autosomal STR markers, making the construction of the appropriate matrices less difficult. If 4-dye chemistry is desired, the red dye ROX can be used as the internal standard with 6FAM (blue), VIC (green), and NED (yellow) being used to label the PCR products.

Primer Quality Control

Once the process of primer design was complete, the primers were purchased from one of several oligonucleotide manufacturers. Primer purity is important and can affect PCR amplification and detection (Figure 4). The PCR product shown in the top panel of Figure 4 was produced from a NED-labeled primer that had been reverse-phase HPLC purified. The PCR product shown in the bottom panel of Figure 4 was produced from a TAMRA labeled primer and shows dye-blobs that failed to be removed through affinity matrix purification. Note that the peak height was much less for the TAMRA-labeled PCR product even though the primers were prepared at the same concentration.

Reverse-phase HPLC purified primers appear to be free of the undesirable dye blobs. However, in order to keep the cost down during the initial construction of this multiplex, not all of the primers were purchased with HPLC purification. Even after ordering HPLC-purified primers, an additional quality control check was performed using time-of-flight mass spectrometry (TOF-MS) [22]. The TOF-MS molecular weight observed for the purchased oligonucleotide is compared to the expected molecular weight, which is determined from the mass of the dye (if present) and the sum of the masses of the nucleotides. Primer synthesis failure products may also be observed if a primer is of poor purity.

Primer Testing

All primer pairs were tested under identical amplification conditions including the same DNA template concentration. If amplification of a particular locus was poor, the primer concentration for that locus was increased and the test was repeated. Unlike previously published multiplex PCR primer mixture design protocols [5,6] in which many other experimental conditions, such as $MgCl_2$ concentration, *Taq* concentration, dNTP concentration, buffer concentration, annealing temperature adjustments, and PCR cycle numbers were modified to achieve better amplification, only the primer concentrations were adjusted in our approach. After running multiple male test samples, the primer concentrations were adjusted. The primer concentrations used in this study are in some cases five fold higher than those in other multiplex work [14,17]. Primer concentrations are being used as the primary means to increase the amount of PCR product balance. Following the successful individual runs of each primer pair they were combined into a multiplex.

For the initial construction of the multiplex all of the primer pairs were combined at a concentration of 1.0 µM. If one PCR product was higher in peak height relative to the other amplicons in the multiplex, the appropriate primer pair concentrations was decreased to try and generate a more balanced yield between the various PCR products. Relative peak heights from the amplicons in the Y STR 20plex were used to estimate the needed primer concentrations adjustments in order to improve the amplification balance. After several adjustments of primers concentrations, a final primer mix was decided upon, which generates fairly balanced PCR products (Figure 5).

RESULTS AND DISCUSSION

Results of Screening for Potential Primer Interactions

Of the 595 possible primer interactions, none of the alignment scores were greater than 7. A summary of all of the possible primer interactions with alignment scores of 5 or greater is displayed in Figure 3. An alignment score of greater than 8, the default value in the Primer 3 program for single primer comparison, has previously been shown to lead to significant primer dimer formation under typical PCR amplification conditions [24]. Thus, the lack of significant primer-dimers along with the similar primer T_m values would suggest that the Y STR 20plex has been generally optimized in terms of primer design. A standard nucleotide BLAST search was also performed on the primers to ensure that there are not any primer binding sites in the human genome other than the Y STR marker of interest.

Testing of YCC Panel using the Y STR 20plex

The Y STR 20plex was tested against the YCC panel and a number of unknown male samples including the candidates for a Y chromosome standard reference material (SRM) under development at NIST. Figure 5 shows the electropherogram analyzed in GeneScan® for one of the male DNA samples. Genescan® results were then imported in the Genotyper® software in order to identify all of the alleles based on the number of repeats. The result of one of the male DNA samples analyzed in Genotyper® using a newly developed macro for the Y STR 20plex is shown in Figure 6A. The Y chromosome haplotypes for all of the samples within the YCC panel were determined using this new macro. Unlike, macros used for allele calling in autosomal STR work (i.e. Kazam), this genotyping macro does not rely on allelic ladders but instead on categories with allele bin windows of ± 0.75 bp to ±1.25 bp based on sequence information obtained from GenBank. Sequencing of alleles from the loci amplified in the Y STR 20plex is underway so that the exact number of repeats in standard samples can be confirmed.

Sensitivity and Mixture Studies

Sensitivity studies and mixture studies were performed using both male/male mixtures and male/female mixtures. The sensitivity studies at NIST show that the male component can be detected with as little as 250 pg of DNA using 28 cycle PCR (data not shown). However, the presence of dye blobs makes the calling of certain alleles problematic. The presence of dye blobs in the Y STR 20plex electropherograms underscores the need for HPLC purified primers. Future work will include results using primer mixtures that have been fully HPLC purified. For the male/female DNA mixtures it was possible to detect the male

component up to the highest tested ratio of 1:150 (data not shown). For the male/male mixtures we were able to detect the male component up to the tested ratio of 1:10 (data not shown). Further sensitivity and mixture studies will be conducted at the New York City Office of the Chief Medical Examiner (NYC-OCME) and will be reported in future communications.

The NYC-OCME has been supplied with a Y STR 16plex that is identical to the Y STR 20plex except that the primer pairs for DYS447, DYS448, and YCAII have been removed. DYS447 and DYS448 have been removed because they are labeled with PET and can only be detected with new ABI 5-dye filter sets. YCAII was removed because of the high amount of stutter present with dinucleotide repeats [9]. These stutter products make typing of forensic samples more difficult, particularly with mixed DNA samples.

Comparison to the Y-PLEX™ 6 Kit

An intralaboratory comparison study was performed to determine the accuracy of the Y STR 20plex. The Y chromosome haplotypes for all of the YCC samples generated using the Y STR 20plex and accompanying macro and the Y-PLEX™ 6 kit from Reliagene were examined. The Y STR markers used in the comparison were those that make up the Y-PLEX™ 6 kit and include DYS19, DYS393, DYS389II, DYS390, DYS391, and DYS385II. Figure 6A depicts a Genotyper® from a male DNA sample run with the Y STR 20 plex on the ABI 3100. Figure 6B shows a Genotyper® result from the same male sample but with the Y-PLEX™ 6 kit on the ABI 310. The Genotyper® macro provided by ReliaGene Technologies makes use of allelic ladders while the Y STR 20plex uses catergories with allele bin windows of ±0.75 bp to ±1.25 bp based on sequence information and allele sizing data.

A comparison of the Y chromosome haplotypes indicate there were 3 discrepant calls out of a possible 1,036 allele designations. This equates to 99.7% agreement between the Y STR 20plex and the commercially available Y-PLEX™ 6 kit. The alleles that were the subject of the discrepant calls were all deemed "off-ladder" alleles by the Y-PLEX™ 6 kit, while the Y STR 20plex identified them as full repeats. In all three cases the amplicons size in base pairs generated by the Y STR 20plex are smaller than those generated by the Y-PLEX™ 6 kit. Thus, it is possible that these microvariants could be due to deletions that exist in regions lying outside of the Y STR 20plex primer binding sites and therefore not be detected with the new primer sets. We plan to sequence the effected samples in order to ascertain the exact cause of the discrepancies.

Interlaboratory Study with the Y STR 20plex

An aliquot of the Y STR 20plex primer mixture was prepared at NIST and sent to the University of Arizona for analysis. Both laboratories analyzed the YCC panel samples on an ABI 3100 and used Genotyper® software to generate the resulting Y chromosome haplotypes. The results show significant agreement between the two laboratories. Out of the 2,418 allele designations examined over 98.5% were in agreement. The majority of the discrepant calls were for Y STR marker DYS385. An investigation as to the discrepancies involved with DYS385 is underway.

Power of Discrimination

Over recent years there have been many studies published describing the discriminatory power of Y STR markers [9, 25-26]. It has been demonstrated that a set of seven Y chromosome STRs consisting of DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393 can discriminate between most of the male individuals in various European populations [9]. These studies typically focus around a single population (i.e. Swiss, Dutch, Korean). The population used in our initial study consists of 97 male samples including 74 samples from the YCC panel and 21 additional male DNA samples. No attempt was made to separate these samples into appropriate ethnic categories for population analysis. Table 2 summarizes the calculated discriminatory power of the Y STR 20plex and various subsets of it based on equations provided by Dr. Charles Brenner at http://dna-view.com/haplofreq.htm.

In the case of Y STRs, the discriminatory power of a set of markers is defined as the probability of obtaining a non-unique haplotype based on the markers used. A unique haplotype is defined as one that only occurs once in a given population. The subsets chosen were as follows. DYS385I/II was the most polymorphic Y STR marker observed for the population studied and there were 18 unique haplotypes observed, with a 1 in 56 chance of obtaining a non-unique haplotype. The subset, which included markers

DYS19, DYS389I/II, DYS390, DYS393 was 4 times more polymorphic than DYS385. This set was chosen because these markers are commonly used as a multiplex in Europe [14]. The Y-PLEX™ 6 kit from ReliaGene was chosen because it is the only commercially Y STR kit available. It includes the markers DYS19, DYS389II, DYS390, DYS393, and DYS385I/II. Adding DYS385I/II more than doubled the discriminatory power of this multiplex increasing the chance of finding a non-unique haplotype to 1 in 470. The minimal and extended haplotype was included in this analysis because of their wide use by the European community. The Y STR minimal haplotype when amplified in two or three multiplex reactions has been recently recommended for court use in Europe [27]. The discriminatory power of the extended haplotype, which includes the addition of the dinucleotide repeat YCAII, reached 1 in 900. The Y STR 20plex discriminatory power reached 1 in 1,100 for the limited sample set analyzed. However, unlike the minimal and extended haplotypes, which currently require two or three or four multiplexes to obtain results, the Y STR 20plex can achieve a high power of discrimination in a single amp lification.

ACKNOWLEDGEMENTS

This work was funded by National Institute of Justice research grants 2000-IJ-CX-K006 to Mike Hammer and 1997-LB-VX-0003 to John Butler through the NIST Office of Law Enforcement Standards. Richard Schoske is a PhD candidate in chemistry at American University and is funded by the United States Air Force, through the Air Force Institute of Technology. The authors thank Professor Jim Girard from American University for his academic advisement of Richard Schoske. The technical assistance of Al Agellon and Veronica Kearney is also greatly appreciated. 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.

REFERENCES

[1] Chamberlain, JS., Gibbs RA., Rainer JE., Nguyen PN., and Casey CT. 1988. Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA Amplification. *Nucleic Acids Res.* **16:** 11141-11156.

[2] Henegariu, O., Hirschman, P., Killian K., Kirsch, C., Lengauer, R., and Maiwald, K. 1994. Rapid Screening of the Y Chromosome in idiopathic sterile men, diagnostic for deletion in AZF, a genetic Y factor expressed during spermatogenesis. *Andrologia.* **26:** 97-106.

[3] Shuber, AP., Skoletsky, J., Stern R., and Handelin, B.L. 1993. Efficient 12-mutation testing in the CFTR gene: a general model for complex mutational analysis. *Hum. Mol. Genet.* **2:** 153-158.

[4] Kimpton, CP., Fisher, D., Watson, S., Adams, M., Urquhart, A., Lygo, J., and Gill, P. 1994. Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. *Int. J. Leg. Med.* **106:** 302-311.

[5] Henegariu, O., Heerema, SR., Vance, GH., and Vogt, PH. 1997. Multiplex PCR: Critical Parameters and Step-by-Step Protocol. *BioTechniques* **23:** 504-511.

[6] Zangenberg, G., Saiki RK., and Reynolds R. 1999. In: Innis MA, Gelfand DH., Snisky JJ. (eds.) PCR Applications: Protocols for Functional Genomics, Academic Press: San Diego, pp. 73-94.

[7] Prinz, M., Boll, K., Baum, H., and Shaler, B. 1997. Multiplexing of Y chromosome specific STRs and performance for mixed samples. *Forensic Sci. Int.* **85:** 209-218.

[8] Underhill, PA., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, WH., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, JR., Mehdi, SQ., Seielstad, MT., Wells, RS., Piazza, A., Davis, RW., Feldman, MW., Cavalli-Sforza, LL., and Oefner, PJ. (2000) Y chromosome sequence variation and the history of human populations. *Nat.Genet.* 26(3): 358-361.

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

[10] White, PS., Tatum, O. L., Deaven, LL., and Longmire, JL. 1999. New, male-specific microsatellite markers from the human Y chromosome. *Genomics*. **57:** 433-437.

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

[12] Deknijff, P. 2001. available at http://www.medfac.leidnuniv.nl/fldo .

[13] Ida, R., Tsubota, E., Matsuki, T. 2001. Identification and characterization of two novel human polymorphic STRs on the Y chromosome. *Int. J. Legal Med.* **115:** 54-56.

[14] Gusmao, L., Gonzalez-Neira, A., Pestoni, C., Brion, M., Lareu, M. V., and 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.

[15] Thomas, MG., Bradman, N., and Flinn, HM. 1999. High throughput analysis of 10 microsatellite and 11 dialleic polymorphisms on the human Y-chromosome. *Hum Genet.* **105:** 577-581.

[16] Roewer, L., Kayser, M., Nagy M., and de Knijff, P., 1996. Male identification using Y-chromosomal STR polymorphisms. *Advances in Forensic Haemogenetics.* **6:** 124-126.

[17] Redd, AJ., Clifford, SL., and Stone-King, M. 1997. Multiplex DNA typing of short tandem repeat loci on the Y-chromosome. *Biol. Chem.* **378:** 923-927.

[18] Ruitberg CM, Butler JM. New primer sets for Y chromosome and CODIS STR loci. Poster presented at the Eleventh International Symposium on Human Identification, Biloxi, MS (2000).

[19] Ruitberg, CM., Redder DJ., Butler, JM. 2001. STRBase: a short tandem repeat DNA Database for the human identity testing community. *Nucleic Acids Res.* **29:** 320-322.

[20] Rozen, S., Skaletsky HJ. 1998. Primer3 (primer 3_www.cgi v 0.2); Code available at http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi.

[21] Butler JM., Ruitberg, CM., and Vallone, PM. 2001. Capillary electrophoresis as a tool for optimization of multiplex PCR reactions. *Fresenius J. Anal. Chem.* **369:** 200-205.

[22] Butler, JM., Devaney, JM., Marino, MA., Vallone, PM. 2001. Quality control of PCR primers used in multiplex STR amplification reactions. *Forensic Sci. Int.* **119:** 87-96.

[23] Gusmao, L., Gonzalez-Neira A., Sanchez-Diz, P., Lareu, MV., Amorim, A, and Carracedo. 2000. Alternative primers for DYS391 typing: advantages of their applications to forensic genetics. *Forensic Sci. Int.* **112:** 49-57.

[24] Beasley E.M., Myers, R.M., Cox, D.R., and Lazzeroni L.C. 1999. In: Innis MA, Gelfand D.H., Snisky J.J. (eds.) PCR Applications: Protocols for Functional Genomics, Academic Press: San Diego, pp. 55-71.

[25] Gehrig, C., Biol, L., Hochmeister, M., Budowle, B. 2000. Swiss allele frequencies and haplotypes of 7 Y-specific STRs. *J. Forensic Sci.* **45:** 436-439.

[26] Kim, YJ., Shin, DK., Kim, JM., Jin, HJ., Kwak, KD., Han, MS., Chol, SK, Kim, W. 2001. Y-chromosome STR haplotype profiling in the Korean population. *Forensic Sci. Int.* **115:** 231-237.

[27] Pascali, VL., Dobosz, M., and Brinkman, B. 1998. Coordinating Y-chromosomal STR research for the courts. *Int. J. Legal Med.* **112:** 1-1.

Table 1. Information on the markers used in the Y STR 20plex. Size ranges were calculated using GenBank® sequences as reference alleles for each locus. Information on the newly discovered markers DYS447 and DYS448 remains to be fully determined (TBD) and will be covered in a future publication. A total of 17 primers pairs are used to generate 20 unique PCR products. Primers for the markers DYS389, DYS385, and YCAII target duplicated portions of the Y chromosome and can generate multiple PCR products.

Marker Name	Repeat Motif	Allele Range	PCR Product Sizes	Reference
DYS19	TAGA	7-16	214-250 bp	Roewer 1992
DYS385	GAAA	7-23	240-304 bp	Schneider 1998
DYS389 I	(TCTG) (TCTA)	I: 7-16	133-169 bp	Kayser 1997
DYS389 II	(TCTG) (TCTA)	II: 25-33	257-289 bp	Kayser 1997
DYS390	(TCTA) (TCTG)	18-27	191-227 bp	Kayser 1997
DYS391	TCTA	7-14	91-119 bp	Kayser 1997
DYS392	TAT	6-16	242-272 bp	Kayser 1997
DYS393	AGAT	9-16	108-136 bp	Kayser 1997
YCAII	CA	17-27	145-165 bp	Kayser 1997
DYS437	TCTA	14-17	184-196 bp	Ayub 2000
DYS438	TTTTC	6-13	299-334 bp	Ayub 2000
DYS439	AGAT	9-14	204-224 bp	Ayub 2000
Y-GATA-A7.1	ATAG	7-12	99-119 bp	White 1999
Y-GATA-H4	TAGA	11-17	119-143 bp	White 1999
DYS388	ATT	11-17	151-169 bp	Kayser 1997
DYS426	TGT	10-13	88-97 bp	Jobling 1996
DYS447	*TBD*	*TBD*	*TBD*	Redd *et al*. in prep
DYS448	*TBD*	*TBD*	*TBD*	Redd *et al*. in prep

Table 2. Discriminatory power possible when using different combinations of Y STR markers tested against 97 male DNA samples including the YCC panel. The calculated values for the discriminatory power for these Y STR haplotypes were determined using the formula provided at http://**dna-view.com/haplofreq.htm** by Dr. Charles Brenner. In this case, discriminatory power is defined as the chance of obtaining a non-unique haplotype. A unique haplotype is defined as one that occurs only once for a given population. Both the minimal and extended haplotypes are widely used in the European forensic community (http://ystr.charite.de).

Loci Tested	Unique Haplotypes	Discriminatory Power
DYS385	18	1 in 56
Gusmao pentaplex [ref. 14] DYS19, DYS389I/II, DYS390, and DYS393	62	1 in 210
Reliagene Y-PLEX™ 6 Kit DYS19, DYS389II, DYS385I/II, DYS390, DYS391, and DYS393	79	1 in 470
Minimal Haplotype DYS19, DYS389I/II, DYS385I/II, DYS390, DYS391, DYS392, and DYS393	83	1 in 600
Extended Haplotype DYS19, DYS389I/II, DYS385I/II, DYS390, DYS391, DYS392, DYS393, and YCAII	87	1 in 900
Y STR 20plex DYS19, DYS385I/II, DYS389I/II, DYS390, DYS391, DYS392, DYS393, YCAII, DYS437, DYS438, DYS439, DYS447, DYS448, DYS388, GATA A7.1, and GATA H4	89	1 in 1,100

Figure 1. Schematic of size ranges and dye label colors used in the designing of the Y STR 20 plex assay. The length of the box containing the locus name represents the size ranges for the know alleles; the base pair scale is shown above. The color of the box reflects the dye color used to label the amplification products. Four different fluorescent dyes are used in this Y STR 20plex that are spectrally distinguishable using Filter Set G5 on the ABI 3100: 6-FAM (blue), VIC (green), NED (yellow), and PET (red). An internal size standard is labeled with LIZ dye (orange). All the loci outlined in red have newly developed primer sets.

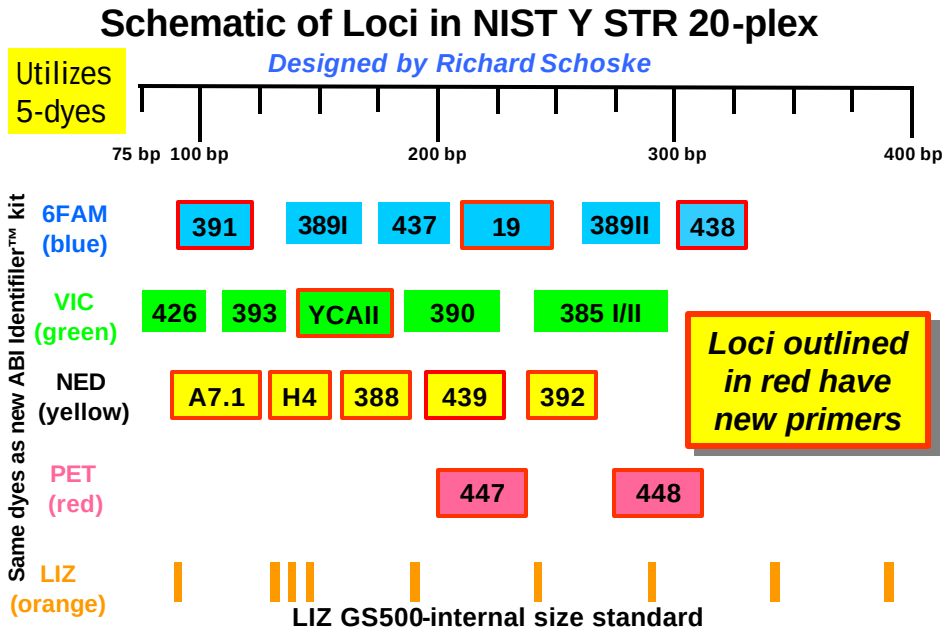


Figure 2. Flow-chart of multiplex design strategy presented in this paper

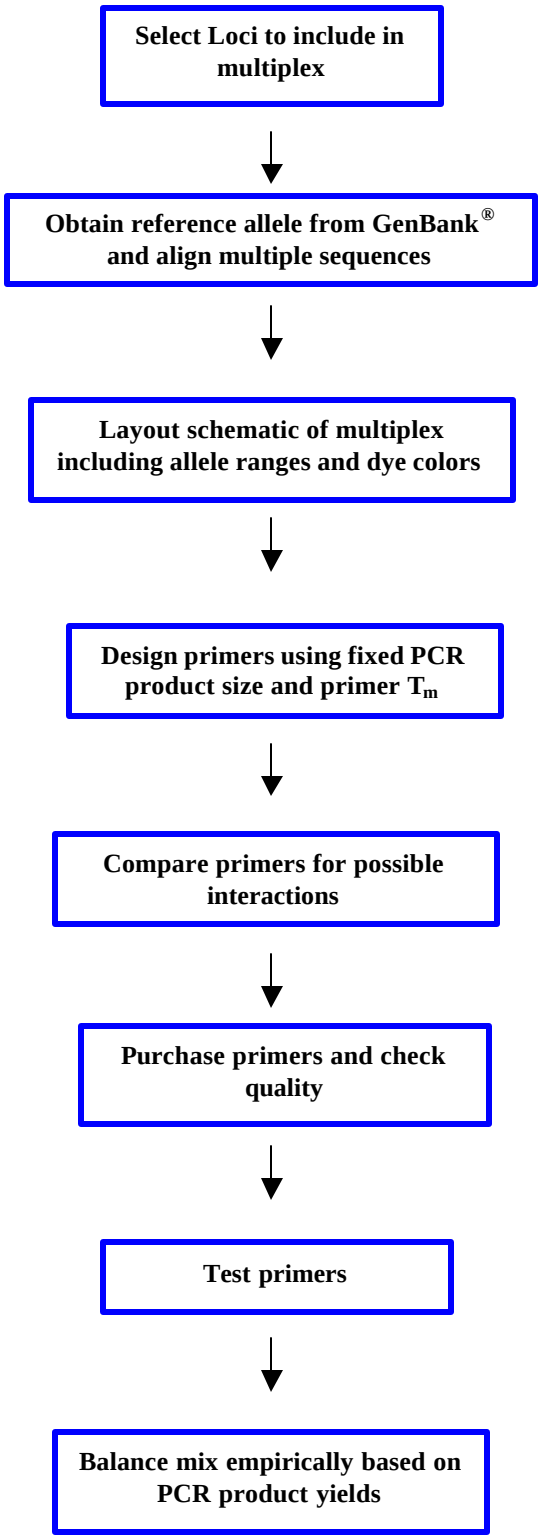


Figure 3. Check for primer cross-reactivity between the 34 primers in the Y STR 20plex primer mixture. The numbers along the bottom of the plot correspond to the primer name and number listed on the right hand side. The numbers within the plot are pair wise "alignment" scores for predicting potential primer alignments and cross-reactivity between primers. The alignment score is defined as the number of complementary base pairs minus the number of mismatched base pairs. The highest primer-to-primer interactions had an alignment score of 7, which occurred in 7 of the 595 possible interactions. Portions of the grid having no value displayed contain alignment scores of less than 5.

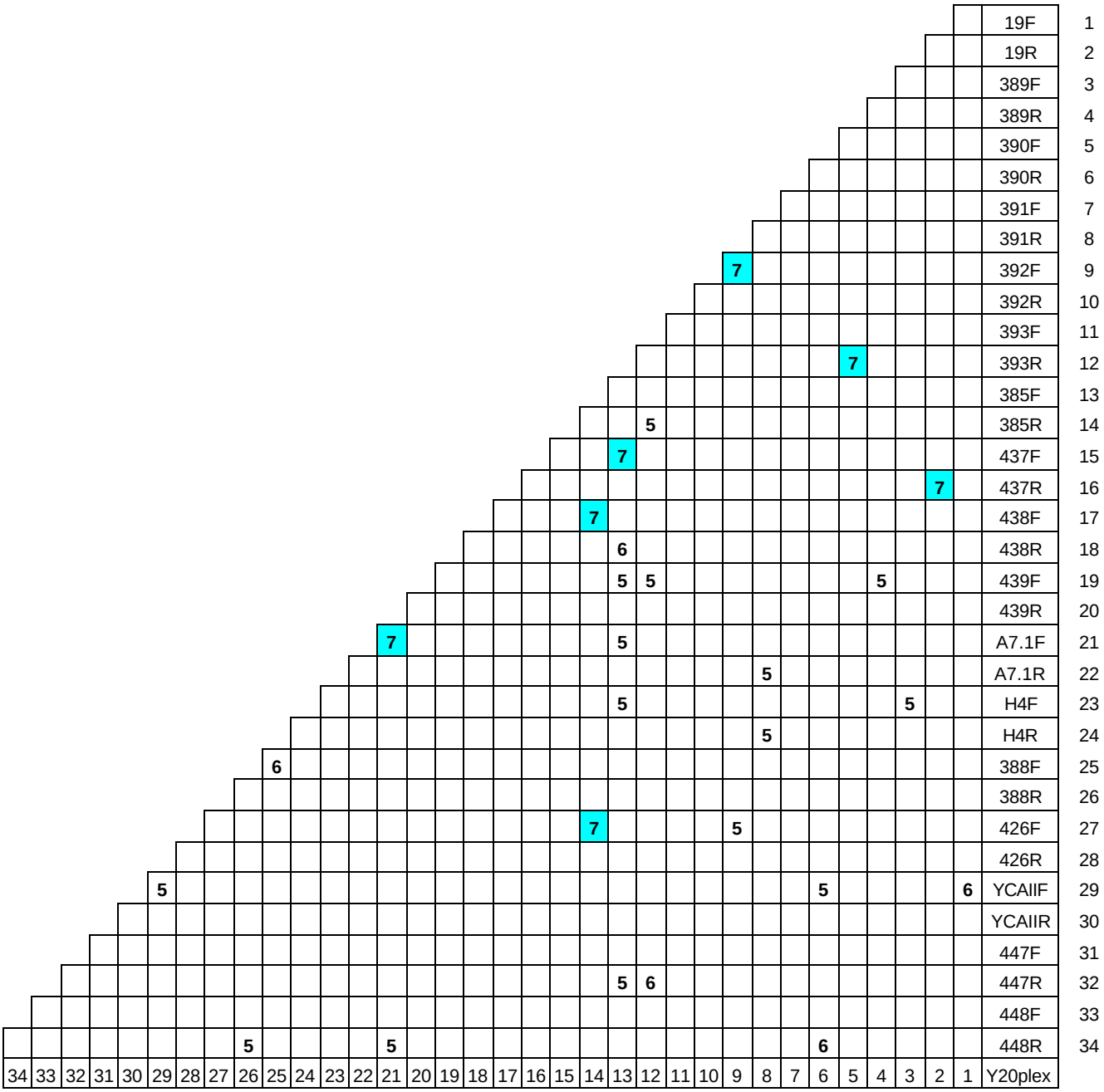


Figure 4. Comparison of ABI 310 electropherograms from singleplex PCR reactions of the Y STR GATA H4 marker amplified with dye-labeled primers that were purified differently. PCR products were generated using 10 ng genomic DNA and 1 µM primers for H4 with the fluorescent dye NED or TAMRA. Samples analyzed were 1 µL PCR product in 20 µL deionized formamide containing 0.75 µL GS500 ROX size standard (not displayed). Electrophoresis conditions were as reported in the Material and Methods section.

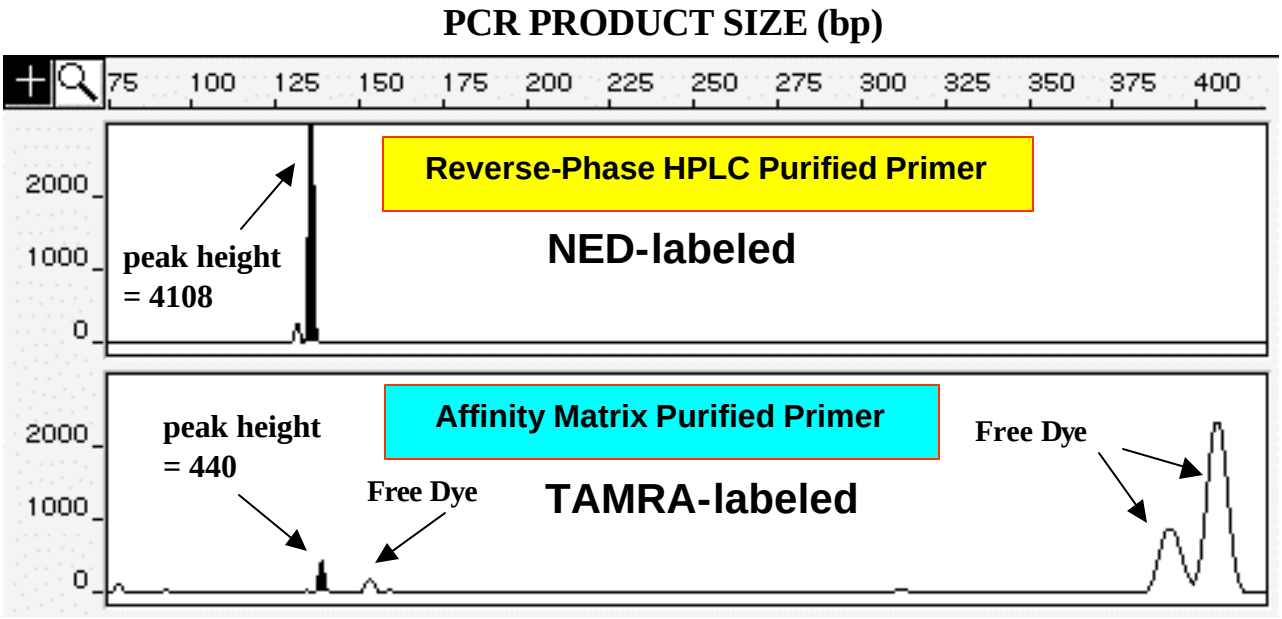


Figure 5. Electropherogram of Y STR 20plex PCR products generated using the strategy described in this work. Results produced on the ABI 3100 Genetic Analyzer as described in the Materials and Methods. The 391, 389I, 437, 19, 389II peaks are labeled with 6-FAM (blue). The 426, 393, YCAII, 390, and 385I/II peaks are labeled with VIC (green). The A7.1, H4, 388, 439, and 392 peaks are labeled with NED (yellow). The 447, and 448 peaks are labeled with PET (Red). A GS500 LIZ (orange) internal standard was used to generate the calculated sizes.

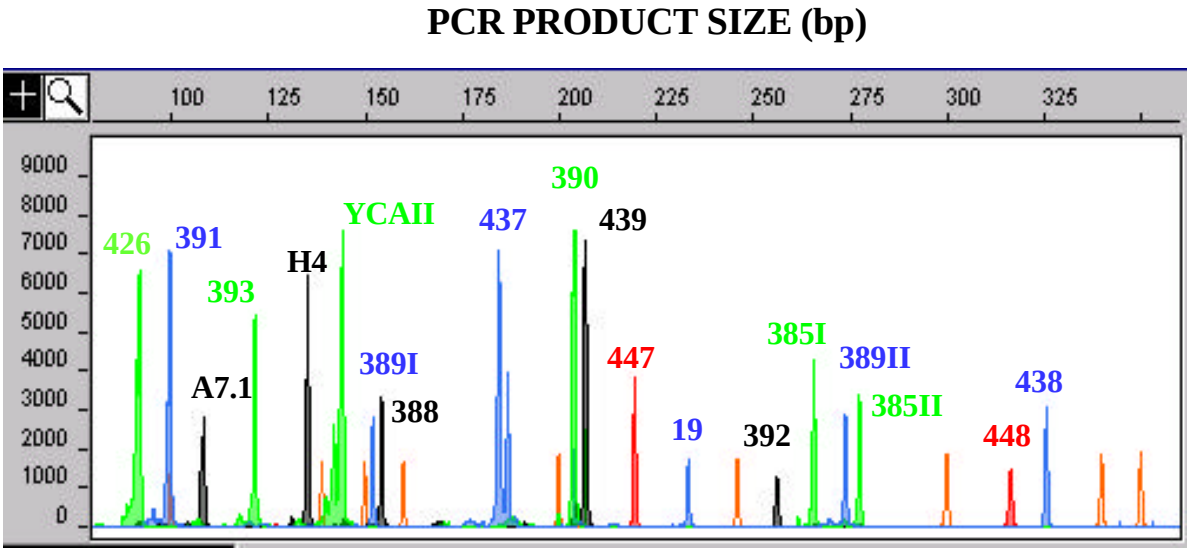
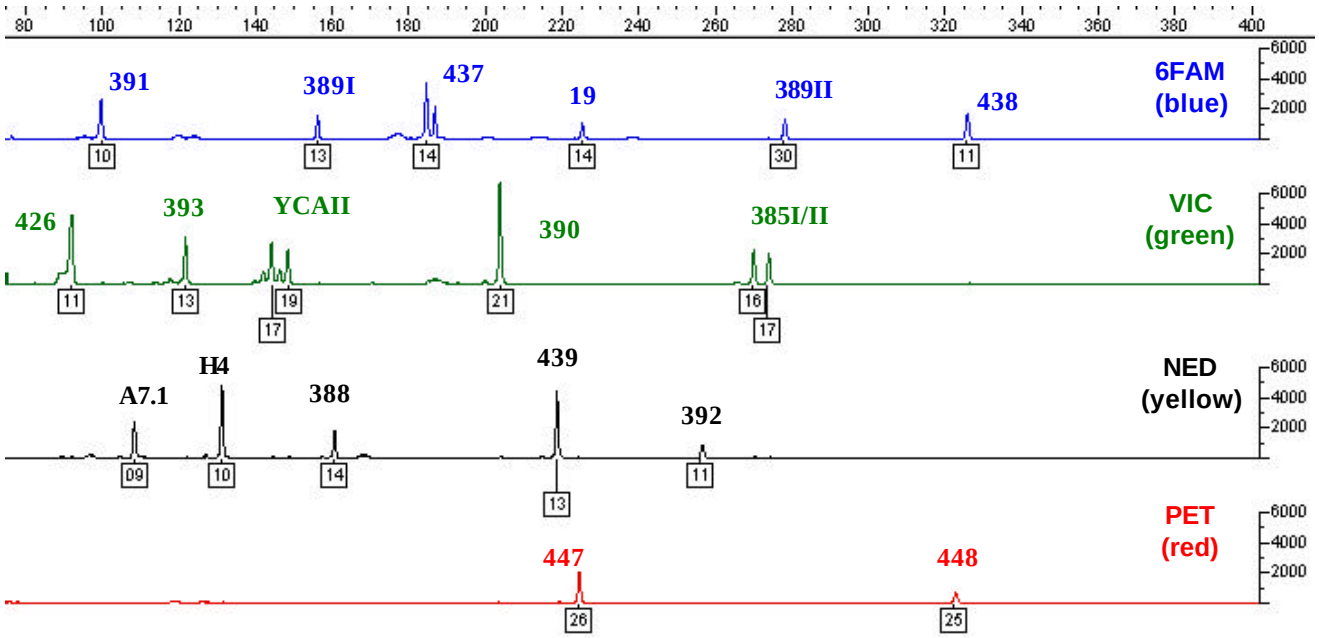


Figure 6. Genotyper® results for the same male DNA sample obtained using (A) the Y STR 20plex and (B) the Y-PLEX™ 6 kit from ReliaGene.

(A) Y STR 20plex



(B) Y-PLEX™ 6 Kit