Validation of a single tube fluorescent multiplex assay for simultaneous typing of 20 Y-STR loci

Sanghamitra Sahoo¹,² and V K Kashyap¹,²*
¹National DNA Analysis Centre, Central Forensic Science Laboratory, 30 Gorachand Road, Kolkata, India

Received 12 June 2006; revised 27 April 2007; accepted 23 July 2007

This paper describes a single tube assay for 20 fluorescent labelled Y-STR loci, which is rapid and robust state-of-art multiplex system. Reaction conditions, including annealing temperature, concentration of primers, magnesium and template DNA, DNA polymerase and reaction volume, were optimised to yield robust amplification. The assay could withstand moderate fluctuations in reaction parameters with no affect on haplotype analysis. Sensitivity and robustness of the multiplex system is revealed from the examination of different matrices and environmental conditions wherein complete haplotypes could be obtained in all the samples with no failures except those exposed to severe environmental stresses. Genotyping of the 20 Y-STR loci showed that all loci included in the multiplex are highly polymorphic in the Indian populations and its inclusion increases the discriminatory power of the marker system. It could be summarized that the developed Y-STR multiplex assay is a simple, sensitive and robust system highly suitable for concurrent analysis of 20 polymorphic Y-STR loci.

Keywords: multiplex PCR, Y-STR, DNA profiling, validation

Introduction

Multiplex polymerase chain reaction (PCR) is a system in which two or more loci are simultaneously amplified in the same reaction. This method has been successfully applied in many areas of DNA typing, including analyses of deletions, mutations¹² and polymorphisms at short tandem repeat (STR) loci¹³⁸ since its first report¹. STRs or microsatellites contain 2-6 bp repeat units which are tandemly arranged and widely dispersed throughout the eukaryotic genome. Analysis of these tandemly repeated DNA sequences is of fundamental importance not only in population biology, such as human identification, understanding origin of man, migratory histories of populations, phylogenetic distribution of various plant and animal species, and other evolutionary processes, but also in linkage analysis of genetic traits. Multiplex STR analysis, therefore, has emerged as an indispensable assay as it requires only sub-nanogram amounts of DNA, is less labour intensive, provides easy interpretation of data, high discriminatory power and reliable results. Although a large number of multiplexed commercial kits are available for human autosomal STR markers [Powerplex® 2.1, Powerplex® 16 system (Promega Corp. Madison, WI), ABI AmpFl STR® Profiler™, COFiler™ and Identifiler™], the same is not the case for the microsatellites of Y-chromosome. The reason for this lacuna is because until recent past, the number of STRs known on the Y chromosome was limited¹⁰⁻¹⁴. In addition, lack of suitable Y-STR mutiplex system is also another reason why the Y-chromosome has not been as extensively used as the autosomal STRs in human identification and population genetic studies. Recently, however, three commercial Y-STR multiplex systems: Y-PLEX™ 6 kit (ReliaGene Technologies Inc.), Powerplex Y System (Promega Corp.) and AmpFLSTR Y filer kit (Applied Biosytems) are available; these include only 6, 12 and 17 Y-STR loci, respectively¹⁵,¹⁶. Although the Y chromosome harbors diversity which is only 20% as that of the autosomes¹⁷, a large number of Y-STRs have been identified in recent years, and a few of them have also been used for human identification, particularly paternity testing, and evolutionary, genealogical and anthropological studies¹⁸,¹⁹.
Majority of the STRs on the Y-chromosome are located in the non-recombining region and thus are passed down the generations without recombination. In addition, because the Y-chromosome is present in a haploid state, it is exceptional and unparalleled compared with any of the known markers, for tracing male movements either of the present-day or events of the past\(^{20}\). However, lack of recombination also suggests that individual markers cannot be combined by the product rule and to achieve power of discrimination similar to those of autosomal STR markers\(^{21,22}\), a large number of Y-STR loci should be included in the analysis. The Y-STR markers find application in a wide range of forensic investigations, especially in cases where father of the male child is deceased\(^{17,23}\), identification of human remains in mass disasters\(^{24}\) and offenders, particularly in sexual assault cases\(^{25}\), where the male component needs to be identified from the mixed male and female fractions of the body fluids\(^{26}\) and where methods such as differential lysis is unsuccessful, as in azoospermic semen samples. In addition, these STRs are equally important in evolutionary biology, archeological and anthropological studies pertaining to male population\(^{17,18,27-29}\). Few studies, using Y-STRs, were primarily based either on monoplexes or multiplexes carried out as 2-3 tube reactions. Since these are laborious and time-consuming protocols, single tube multiplex methods are much in demand\(^{30}\). The development of a large multiplex will lead to newer possibilities for studying which Y-STRs have the most significant impact on a particular population by enabling more rapid collection of the data from population samples. Understanding the genetic structure of individual populations will help in maintaining national DNA databases, which is increasingly essential\(^{31-33}\) for assessing the strength of match evidence in DNA profiling cases. Therefore, development of a Y-STR multiplex system is not only of immense value in justice delivery system\(^{34}\) but also in understanding the Indian scenario of male evolutionary and population genetics. We have, therefore, attempted to develop, standardize and validate a set of 20 Y-STR loci for a single tube multiplex reaction that will provide a sensitive, reproducible, robust and rapid PCR assay.

Materials and Methods
Reagents for PCR
10X PCR gold buffer containing 500 mM KCl, 150 mM Tris-HCl, pH 8.3, 25 mM MgCl\(_2\) and AmpliTaq Gold\(^\circledR\) DNA polymerase were procured from Applied Biosystems, Foster City, USA. The nucleotides (dNTP- 100 mM each) and glycerol was purchased from Promega Corp. (Madison, WI, USA) and SRL (Mumbai, India), respectively. Customised fluorescent labelled and non-fluorescent primers were procured from Isogen Bioscience BV, The Netherlands and reconstituted in TE buffer to a final concentration of 1 nmol/µL each. The working stocks of all primers were prepared in milli-Q water for a concentration of 20 pmol/µL.

Selection and Description of Y-STR loci
Our previous standardization experiments included eight Y-STR loci\(^{35}\), which constituted the minimal haplotype required for human identification recommended by the ISFG guidelines (protocol unpublished). The “minimal haplotype” includes tetranucleotide Y-STR markers DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS393 and a trinucleotide, DYS392. However, with the availability of information on polymorphism for new Y-STR loci\(^{10,36-40}\), we considered their inclusion as an asset in enhancing the power of discrimination manifold. The loci incorporated in the new multiplex include one dinucleotide locus YCAIIa/b, two other trinucleotide loci DYS388 and DYS426, and four tetranucleotide repeat loci DYS460, H4, DYS437 and DYS439. In addition to these, there are two pentanucleotide repeat loci DYS438 and DYS447 and one hexa repeat nucleotide locus DYS448. Three primer pairs, DYS389, DYS385 and YCAII generate two amplicons each, resulting in a total of 20 Y-STRs. The detailed information on loci used in the 20-Y-STR multiplex is given in Table 1. We carried out standardization and in-house validation of 20 Y-STR loci in a single tube multiplex reaction by using these primers described previously\(^{11}\). The multiplex system would also help in examining whether the Indian populations harbour similar level of polymorphism in all the selected Y-STR markers as those described by the European populations so that a uniform set of markers could be easily used in routine forensic investigations to provide a high power of discrimination.

Description of Primers
The amplification of the selected loci was carried out using NIST primer sequences obtained from the Y-STR database site (http://www.cstl.nist.gov/biotech/strbase/). Although
the genetic recombination between X and Y chromosome is low, certain degree of homology exists between the two chromosomes because of their common evolutionary history. Therefore, the objective was to incorporate those primers which have enough specificity to ensure preferential amplification of Y-chromosome sequences and minimize or eliminate any interference from X chromosome. The primers were aligned with each other and also with the sequence database at the National Center for Biotechnology Information (NCBI) using the Basic Local Alignment Search Tool (BLAST) program to rule out any possibility of dimerization and to determine the extent of homology with the X chromosome or to any other chromosome. The primers are labelled with four different fluorescent dyes, 6′-FAM (blue), VIC (green), NED (yellow) and PET (red), where different fluorochrome would detect overlapping fragment sizes.

**Test Samples**

Genomic DNA was obtained from blood or tissue samples from both human and non-human sources following organic extraction protocol.

### Table 1—General information on loci used in the 20 Y-STR multiplex system

<table>
<thead>
<tr>
<th>Marker Name</th>
<th>Repeat motif</th>
<th>Gen Bank accession</th>
<th>Reference allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>DYS19</td>
<td>TAGA</td>
<td>AC017019 (r&amp;c)</td>
<td>15</td>
</tr>
<tr>
<td>DYS385 a/b</td>
<td>GAAA</td>
<td>AC022486 (r&amp;c)</td>
<td>11</td>
</tr>
<tr>
<td>DYS389 I</td>
<td>(TCTG) (TCTA)</td>
<td>AC004617 (r&amp;c)</td>
<td>12</td>
</tr>
<tr>
<td>DYS389 II</td>
<td>(TCTG) (TCTA)</td>
<td>AC011289</td>
<td>29</td>
</tr>
<tr>
<td>DYS390</td>
<td>(TCTA) (TCTG)</td>
<td>AC011289</td>
<td>24</td>
</tr>
<tr>
<td>DYS391</td>
<td>TCTA</td>
<td>AC011302</td>
<td>11</td>
</tr>
<tr>
<td>DYS392</td>
<td>TAT</td>
<td>AC011745 (r&amp;c)</td>
<td>13</td>
</tr>
<tr>
<td>DYS393</td>
<td>AGAT</td>
<td>AC006152</td>
<td>12</td>
</tr>
<tr>
<td>YCAI A/B</td>
<td>CA</td>
<td>AC015978</td>
<td>23</td>
</tr>
<tr>
<td>DYS388</td>
<td>ATT</td>
<td>AC004810</td>
<td>12</td>
</tr>
<tr>
<td>DYS426</td>
<td>GAT</td>
<td>AC007034</td>
<td>12</td>
</tr>
<tr>
<td>DYS437</td>
<td>TCTA</td>
<td>AC002992</td>
<td>16</td>
</tr>
<tr>
<td>DYS438</td>
<td>TTTTC</td>
<td>AC002531</td>
<td>10</td>
</tr>
<tr>
<td>DYS439</td>
<td>AGAT (TAATA)</td>
<td>AC002992</td>
<td>13</td>
</tr>
<tr>
<td>DYS447</td>
<td>(TAAA)</td>
<td>AC005820</td>
<td>23</td>
</tr>
<tr>
<td>DYS448</td>
<td>AGAGAT</td>
<td>AC025227</td>
<td>22</td>
</tr>
<tr>
<td>DYS460</td>
<td>(A7.1)</td>
<td>AC009235 (r&amp;c)</td>
<td>10</td>
</tr>
<tr>
<td>Y-GATA-H4</td>
<td>TAGA</td>
<td>AC011751 (r&amp;c)</td>
<td>12</td>
</tr>
</tbody>
</table>

Reference allele refers to the number of repeats found in the GenBank sequence, which must sometimes be made reverse and complement (r&c) in order to maintain consistency with previously used repeat motifs.

**Basic PCR Protocol**

The basic PCR (20 µL reaction volume) included: autoclaved milli-Q water, 1X PCR gold buffer, 1.5 mM MgCl₂, 200 µM each dNTP, 1U AmpliTaq Gold™, 5% v/v glycerol (if used), primer (0.2-1.2 µM each) and genomic DNA. The amplicons were checked by electrophoresis on Nusieve® Agarose gels (FMC Bioproducts, Rockland, ME, USA) containing Ethidium bromide in 1X TBE [0.09 M Tris-Borate; 0.002 M EDTA (pH 8.0)] buffer and detected under the UV transilluminator. All the amplification reactions included a known DNA sample as male positive control, a female DNA as negative control and an amplification negative control.

The basic PCR protocol included (1) Amplification of each loci as monoplexes and (2) Multiplexing of the Y-STR loci through a series of standardization steps:

i) Standardization of annealing temperature

ii) Optimization of primer concentrations

iii) Optimization of DNA polymerase and magnesium concentrations

iv) Validation of multiplex assay

**Results and Discussion**

### Single Locus PCR

Each of the primer pairs was amplified individually at the same annealing temperature (55ºC) in 1x reaction condition with 0.5 µM primer concentration and without glycerol (Table 2). A high primer concentration was maintained for checking the efficacy of primers under similar amplification conditions. Fig. 1 shows the agarose gel image for successful amplification of all the single Y-STR loci.

### Multiplex PCR

Combining the primers together for amplifying many loci simultaneously requires alteration/optimization of certain parameters of the reaction. Initially, all the primers were added in equimolar amounts in various combinations under single locus PCR conditions in addition to 5%v/v glycerol, to test the individual primer concentrations and other parameters for a successful multiplex reaction. For the initial standardization, all primers were mixed in 0.5 µM concentrations, of which only 8.0 µL of Y-STR primer mix was added to the 20 µL reaction mix. Fig. 2 shows the 4% agarose gel photograph for positive amplification with Y-STR multiplex. The final primer concentrations for each of the primer pairs are given in Table 3. Performance of the PCR-based STR assays can be influenced by numerous factors, including the difference in
instrumentation, variability in quality and quantity of DNA template being amplified, variations in reaction and analysis conditions. Therefore, further standardization of the multiplex was carried out after a few preliminary rounds of detection of the amplicons in a genetic analyzer for a high throughput format to obtain accurate and precise size calling of the alleles.

**Detection and Analysis of PCR Products**

Detection of the amplicons was achieved with ABI Prism® 3100 Genetic Analyzer 16-capillary array system (Applied Biosystems) using the G5 matrix filter for the 5-dye chemistry. GS500 Liz™ (Applied Biosystems), which produced orange colour fluorescence, was used as size standard for sizing length of the amplicons in basepairs. Samples were prepared in Hi-Di™ formamide (Applied Biosystems) with GS500 Liz™ size standard (mixed in 25:1 ratio) and 1 µL of PCR product. Electrophoresis were performed at 15 kV for 44 min with a run temperature of 60ºC using 3100 POP™-4 (Applied Biosystems), 1X Genetic Analyzer Buffer with EDTA and a 36 cm array. Following data collection, samples were analysed with Genescan® 3.7 (Windows NT, Applied Biosystems) for sizing of the products. Allele designations were made based on sizing bin windows of up to ±0.15 bp.

**Optimization of Reaction Conditions**

Depending on the difference in peak heights of the amplicons, appropriate adjustments were made in the concentration of the critical reagents and thermocycling conditions including reaction volume, template and primer concentrations, annealing temperature to obtain more balanced signal intensity.
**Reaction Volume and Template Concentration**

Change in the reaction volume can alter the concentration of PCR reaction components. Varied reaction volumes (5, 10, 12.5, 15, 20 and 25 µL) were evaluated to determine the effects on amplification. When the concentrations were kept uniform, amplification was seen at >10 µL volumes. To test the sensitivity of the multiplex system, a serial dilution of the test samples was performed to quantify the optimal concentration of DNA template required for a robust multiplex amplification. Template concentration of 10 ng per reaction volume gave more than 100 RFU and was found to be sufficient for robust amplification. Larger loci drop-outs were observed at concentrations less than 5 ng template DNA.

**Primer Titrations**

Concentrations of the primers play a crucial role in the amplification efficacy of loci in a multiplex reaction. Therefore, adjustments in the primer concentrations are critical to produce equivalent signals for each locus. Primer concentrations between 0.1 µM and 1.5 µM with 28-30 cycles and template between 1 and 25 ng were tested for determining relative signal balance between loci. When high primer concentrations were used, smaller loci were preferentially amplified. With the final primer concentrations as given in Table 3, peak heights decreased in smaller loci and increased in larger loci, giving a relative balance of signal intensity. Locus drop-out was seen at primer concentrations less than 0.2 µM and 5 ng of DNA. Final reaction volumes were adjusted to 10 µL to which primer volume was further reduced to 2.0 µL of primer mix.

**Variation of Annealing Temperature**

Changes in the annealing temperature can affect the specificity and balance of the amplified loci. To examine these effects, amplification reactions were performed at various annealing temperatures (55°C-60°C) with a minimum of 25 ng DNA template concentration. Variations in annealing temperatures can also cause cross-reactivity and allele drop-outs. The most commonly observed loci to drop-out at higher annealing temperature include DYS19, DYS389II, DYS447 and DYS388. Locus drop-out and additional artifacts were observed at temperatures above 58°C, whereas at lower annealing temperature an increased yield of the smaller loci caused a locus-to-locus imbalance. The final annealing temperature was maintained at 57°C (Table 2).

**Optimization of DNA polymerase and Magnesium Concentrations**

The DNA polymerase plays a key role in determining the proportion of amplification that could be achieved. Therefore, variations in the concentrations of AmpliTaq® Gold DNA polymerase were evaluated. Between 1 and 5U/20 µL DNA polymerase concentrations were examined with template concentration of 1-25 ng and 28-30 cycling reactions. Optimal balance of signal between loci were seen at as low as 1U per reaction volume. No locus drop-out was observed at this enzyme concentration but the yield of larger loci was low.

Magnesium has significant effect in polymerase activity and specificity, but EDTA in the samples can inadvertently alter effective concentrations of magnesium. Various titrations between 1-2.5 mM were examined, where the optimal balance between loci was seen at 1.5 mM magnesium concentration. Locus-to-locus imbalance increased on decreasing the concentration of magnesium to less than 1.2 mM. Lower concentration resulted in locus drop-outs while higher concentration led to the formation of too many non-specific products.

**Comparison of Thermal Cyclers**

Different models of thermal cyclers have slightly different heating and cooling properties. To evaluate the consistency of yield across loci on different thermal cyclers, two different instruments— Perkin-Elmer GeneAmp® PCR System 2400 and 9700 were examined. This exercise was carried out to rule out biased haplotypes due to the usage of different PCR machines. The signal intensity of each locus (RFU) was compared to the total signal across all the loci for a sample. No consistent performance differences were observed between the thermal cyclers (Table 4).

**Validation of the Multiplex**

STR analysis of forensic samples in particular, relies on two important parameters (i) specificity of primers for humans and (ii) sensitivity for detection. Thus, validation studies of the multiplex system were carried out on different samples and matrices to verify the species specificity, sensitivity, mixture analysis and effect of environmental stresses. The specificity of multiplexes for amplification was examined with five different animal species, including rat, frog, fish, chicken and bovine sp. None of the tested samples yielded positive amplification. This assay with high degree of human specificity is essential for accurate result interpretation in case of contamination from non-human sources. The sensitivity of the multiplex
system was tested by quantifying the optimal amount of template DNA required for reliable amplification in serially diluted DNA samples. An advantage of Y-STR analysis even with low template concentration is that while in the case of autosomal markers, allele drop-out creates potential for misinterpretation at heterozygous locus, drop-out of a Y-STR loci will produce no information except in cases of male/male mixtures or bi-allelic loci such as DYS385a/b and YCAII a/b. Further, to demonstrate that the sampling method does not produce biased haplotype profile for the Y-STRs, reproducibility with liquid blood, dried bloodstains, buccal swabs, semen sample and autopsy tissue was carried out on various male samples. In all instances, different samples produced complete, correct and identical profiles. Exposure to different environmental conditions, such as temperature and sunlight exposure is known to affect the quality of DNA resulting in degradation of larger templates and decreased yield of larger amplicons. To evaluate this, DNA was extracted from bloodstains exposed to various natural and laboratory-controlled conditions for different time periods. All the samples exposed to 50°C or lower temperatures and a range of light exposures produced complete haplotypes (Fig. 3). To confirm the effect of PCR inhibitors on substrate materials which can adversely affect the performance of amplification, we carried out amplification and generated complete profiles from various matrices such as, gauze, wood, glass, and soil, etc. One of the salient features of Y-STRs is its utility in analysis of mixed samples. Male and female fractions in various concentrations and combinations, from a ratio of 10:1 to 1:10 were analyzed. It was carried out to exclude the probability of any female sample getting amplified and secondly, to detect inhibitory concentrations of female DNA for Y-specific amplifications. Large concentrations of female fractions affect the amplification of higher loci. Even with varying ratios of the male samples in the male/male mixtures, complete and correct haplotypes could be generated. These data suggest that this multiplex has the level of specificity and sensitivity to reliably produce results for proportions to male/female mixtures generally obtained in forensic samples. Determining the number of donors in the male/male mixtures only becomes difficult when bi-allelic loci are taken during interpretation.

Population Studies

As a part of the validation study and to determine the extent of polymorphism of the selected Y-STR markers such that they are able to differentiate between unrelated males with a high degree of significance, we tested the multiplex system on various population samples.\textsuperscript{46-50} We observed that a

![Fig. 3—Validation of the 20-Y-STRs multiplex PCR in liquid blood and stain samples in guaze and Whatman paper.](image-url)
total of 1240 distinct Y-STR haplotypes in 1297 unrelated, healthy Indian males analysed with the 20 Y-STR multiplex, with complete profiles were generated in more than 90% of the samples analysed. The power of discrimination between individuals with the 20 Y-STR was estimated to be 0.999. The number of alleles found for each locus ranged from 7-18, with the exception of DYS385 and YCAII loci, which gave 81 and 55 genotypes, respectively. All the Y-STR loci depicted gene diversity values greater than 0.5, even for those loci that harbored the least number of alleles (DYS426, DYS448). Validation of the 20 Y-STR multiplex system on population samples also made it possible to make “Allelic Ladder” so that correct and accurate alleles calls could be made when more samples are analyzed. For this, amplified products were pooled for making the ladder, thus facilitating further typing of all possible alleles by using Genotyping Macro. The allelic profile of the Y-STR multiplex run on 3100 genetic analyzer is represented in Fig. 4.

Conclusions
The validated Y-STR multiplex assay offers exclusively single tube reaction, which is rapid, reliable, reproducible, highly sensitive and cost-effective system with a high discriminatory potential. The increased number of loci in the multiplex improves the power of discrimination between individuals, which is one of the most important requirements for individual identification and paternity testing in forensic genetics. The simultaneous amplification of many loci in the same reaction not only saves time and resources but also provides a larger dataset for comparison as compared to tedious protocols of single locus amplifications or tetra- and pentaplexes. The existence of a large multiplex will also allow inclusion of statistical power to DNA evidences as rapid collection of Y-STR haplotype data from Indian populations is accomplished. This is because the utility of Y-chromosome in forensic casework requires large population genetic samples to be studied. Approximately 1300 male samples belonging to different regions of the country with varied ethnicity and linguistic affinity have already been examined with this multiplex system. This will not only be of assistance to justice delivery system but also provide time estimates of various demographic events that have been experienced by the Indian populations.

Fig. 4—Allelic ladder constructed from the multiplex PCR amplification of 20 Y-STRs in approx 900 male individuals.
thereby increasing our understanding of the Indian scenario of male evolutionary and population genetics.

Acknowledgement

We acknowledge the facility provided at Central Forensic Science Laboratory, Kolkata, and would specially like to thank Dr R Trivedi for providing technical support required in the study, and to the research scholars of the institute for carrying out validation studies on population samples. This study was funded by a research grant under the X Five Year Plan to CFSL, Kolkata. SS was supported by a fellowship from Ministry of Home Affairs, Govt of India.

References


