RESEARCH NOTE

Development of a Y-STR 12-plex PCR system and haplotype analysis in a Korean population

SUN WHA PARK[1†], CHOON HONG HWANG[2†], EUN MIN CHO[1], JI HWAN PARK[1], BYUNG OK CHOI[3] and KI WHA CHUNG[1*]

[1]*Department of Biological Science, Kongju National University, 182 Gongju, Chungnam 314-701, Korea*
[2]*Dowgene Research Center, Seoul, Korea, and Department of Internal Medicine, Korea University, Seoul 136-705, Korea*
[3]*Department of Neurology and Ewha Medical Research Center, Ewha Womans University School of Medicine, 911-1 Mokdong Yangcheon-ku, Seoul 158-710, Korea*

Introduction

A Y-chromosomal short tandem repeat (Y-STR) dodecaplex PCR system for 12 loci has been developed, and using this system allele frequencies and haplotypes were determined in a Korean male population. From a study of 320 unrelated Korean males, 254 different haplotypes were identified. The haplotype diversity and discrimination capacity were estimated to be 0.996 and 0.794, respectively. Gene diversities ranged from 0.196 to 0.764 (mean gene diversity 0.557). From four loci, five mutations were identified. All cases were single repeat mutations, and mean mutation rate was determined to be 2.64×10^{-3}/locus/meiosis. The 12-plex Y-STR PCR system exhibited reliable typing results for the male–female mixed samples at ratios up to 1 : 100 and for several forensic materials from assaulted women. The system was highly sensitive even with minimal amounts of genomic DNA (0.1 ng). We believe that the newly developed 12-plex Y-STR PCR system will provide more discriminative power and will be suitable for practical forensic uses.

Short tandem repeat (STR) loci have shown to be powerful markers for the identification of human individuals due to their hypervariability and wide distribution throughout the genome (Tautz 1989). Of these, Y-chromosomal STRs are effective in determination of paternal lineage as well as in forensic application of male/female cell admixture. It has been suggested that Y-STRs are useful in comparison between closely related populations due to their different distribution of region-specific allele frequencies (de Knijff *et al*. 1997). The Y chromosome does not undergo meiotic recombination except for the pseudoautosomal region, thus Y-STR haplotype studies are useful for tracing human migration and identification of each ethnic group. Y-STRs are particularly helpful in detecting the male DNA fraction from sexual assault stains and swabs (Roewer 2009). Population genetic data for Y-STRs in various ethnic groups have been reported in the last decade (Butler *et al*. 2002; Iida *et al*. 2002; Kayser *et al*. 2002; Redd *et al*. 2002; Dai *et al*. 2004; Schoske *et al*. 2004; Asamura *et al*. 2008; Frank *et al*. 2008). The International Forensic Y Chromosome User Group has established the Y-chromosome haplotype reference database (YHRD; http://www.ystr.charite.de) which includes the minimal haplotype (minHt) comprising 10 loci (*DYS19*, *DYS389I*, *DYS389II*, *DYS390*, *DYS391*, *DYS392*, *DYS393*, *DYS385a,b*, *DYS438* and *DYS439*) (Roewer *et al*. 2001). Recently, Y-chromosomal SNP studies have revealed an association between binary SNP and Y-STR haplotypes (Nonaka *et al*. 2007; Robino *et al*. 2008).

The multiplex PCR analysis system is useful for rapidly analysing real forensic-case samples, as it can be easily used to analyse the genotypes in a large number of samples. Recently, Y-STR typing up to 16-plex in a single reaction has become commercially available e.g., AmpFlSTR Yfiler (Applied Biosystems, Foster City, USA), Powerplex Y system (Promega, Madison, USA), and Mentype Argus Y-MH PCR amplification kit (Biotype AG, Dresden, Germany). However, all loci in Powerplex (12 loci) and Mentype kit

*For correspondence. E-mail: kwchung@kongju.ac.kr.
†These authors contributed equally to this work.

Keywords. Y-STR; multiplex PCR; Korean; *DYS388*; *DYS443*; *DYS445*; *DYS447*; *DYS448*; *DYS450*; *DYS458*; *DYS460*; *DYS513*; *DYS552*; *DYS561*; *DYS587*; human genetics.

(eight loci) completely overlapped with the Yfiler loci. Thus, additional Y-STR multiplex PCR systems are required to insure additional discriminative power, particularly when mutation events occur.

In this study, 12 polymorphic Y-STR loci were chosen (*DYS388*, *DYS443*, *DYS445*, *DYS447*, *DYS448*, *DYS450*, *DYS458*, *DYS460*, *DYS513*, *DYS552*, *DYS561* and *DYS587*), and a new 12-plex PCR system was developed. Although this dodecaplex system included two overlapping loci with the Yfiler kit, it completely excluded the minHt, Powerplex and Mentype loci; thus, it could provide an additional discriminative power of Y-STR system. Using this established 12-plex PCR systems, the allele frequencies and haplotypes were analysed in a Korean male population.

Materials and methods

Samples

Whole blood samples were collected from 320 unrelated Korean male individuals, including students and faculty from Kongju National University, and visitors (patients) of Ewha Medical Research Center, Ewha Womans University School of Medicine. We tried to exclude relatives by asking familial history and reviewing students and patients records for each participant. We did not consider surname composition ratio, because the rate of top five surnames is up to 60% among more than 100 surnames in Korea. We also analysed 120 family samples including father/son pairs to estimate the mutational rate. Paternity of the families with mutation was confirmed by the genotyping of 15 autosomal STR markers using an AmpFlSTR Idnetifiler kit (Applied Biosystems, Foster City, USA). All participants included in this study provided written informed consent.

DNA extraction

DNA extraction from whole blood was performed using a QIAamp DNA blood mini kit (Qiagen, Hilden, Germany). Underwear stains and vaginal swabs from sexually assaulted women were treated with lysis buffer containing proteinase K (final concentration of 0.3 mg/mL) at 56°C for 1 h. Supernatant (epithelial layer) was discarded after centrifugation at 20,000 xg (14,200 r.p.m.) for 10 min, and the precipitate (sperm layer) was dissociated with lysis buffer containing proteinase K and dithiothreitol (DTT). DNA was finally purified using a QIAamp DNA micro kit (Qiagen, Hilden, Germany) after incubation at 56°C for 2 h. DNA extraction from hair root sample was performed using the same kit after treatment of proteinase K.

Dodecaplex PCR amplification

Most primer sequences were those used in previous studies (Butler *et al*. 2002; Iida *et al*. 2002; Dai *et al*. 2004; Schoske *et al*. 2004), while those for *DYS447* and *DYS561* were modified from the original sequences (Butler *et al*. 2002; Dai *et al*. 2004). Fluorescent dyes were labelled at the 5-end of the forward primers to create blue (6-FAM), green (VIC) and red products (PET), respectively. The Liz-500 was used for the internal standard (Applied Biosystems, Foster City, USA). The reaction mixture contained 5 ng of genomic DNA, 1× Gold STR buffer, 0.5 unit of Ampli *Taq* Gold DNA polymerase (Applied Biosystems, Foster City, USA), and different concentrations of 12 primer pairs in a reaction volume of 20 μL. Primer sequences, labelling dyes and final concentrations of primers are listed in table 1 of electronic supplementary material at http://www.ias.ac.in/jgenet/. Thermal cycling was conducted on a PTC-200 DNA engine (MJ Research, Waltham, USA) using the following condition: 95°C for 11 min followed by 32 cycles at 94°C for 1 min, 59°C for 1 min, 72°C for 1 min, and a final extension at 60°C for 1 h.

Allelic nomenclature and genotyping

Exact allele size and repeat structure were determined by sequencing each allele for all loci. Alleles were basically designated by the number of repeats according to the recommendations of the DNA commission of the International Society of Forensic Genetics (ISFG) (Gusmão *et al*. 2006). Standard allelic ladders were prepared for all the loci. PCR co-amplified products were separated by an ABI 3100 genetic analyzer, Foster City, USA and were then analysed using the Genotyper software v 3.7 (Applied Biosystems, Foster City, USA). We analysed haplotype profile on DNA from established cell line, 9948 (Applied Biosystems, Foster City, USA) as an international standard (see table 2 in electronic supplementary material).

Sensitivity and forensic application

For the sensitivity test, serial dilutions of male DNA (0.05–2 ng) were amplified using the multiplex PCR system. For the testing of the male–female mixed samples, the amount of male DNA was kept constant (2 ng) with increasing amounts of female DNA with ratios of 1:1, 1:10, 1:50, 1:100 and 1:200. The examinations for the serial dilution and male–female DNA mixture were performed three times using different samples. For the testing of forensic cases, several practical forensic samples (one underwear and two vaginal swabs from sexually assaulted women) were obtained from the National Institute of Scientific Investigation in Korea, and they were applied to our multiplex PCR system. For the underwear, three stained regions were cut out up to 0.5 cm^2, and then used for DNA purification.

Statistical analysis

Allele frequencies were calculated by direct gene counting. The gene diversity (GD) was determined using the formula $N(1 - \sum P_i^2)/(N - 1)$, where N represents the number of individuals, and P_i is the frequency of the ith allele. Haplotype diversity (HD) was calculated by a formula similar to the GD given as $N(1 - \sum X_i^2)/(N - 1)$, where X_i are haplotype frequencies. Discrimination capacity was determined

by N_{Hp}/N, where N_{Hp} is observed haplotype number. The confidence intervals (c.i.) for mutation rates were estimated from the binomial standard deviation (s.d.).

Results and discussion

Performance of Y-STR dodecaplex PCR

In this study, we established a 12-plex Y-STR PCR system using different concentrations of each primer pair. Most loci showed similar levels of amplification; although some alleles (e.g., *DYS458* and *DYS443*) had slightly small peaks compared to others. As shown in figure 1, multiplex coamplification of 12 loci enabled unambiguous allele typing. The minimal male genomic DNA requirement to obtain reliable results was 0.1 ng (figure 2,A). Although genomic DNA less than 0.1 ng was also likely to work, but several markers (e.g., *DYS388*, *DYS443*, *DYS460*, *DYS513* and *DYS552*) showed too small peaks (less than 50 RFU) to determine

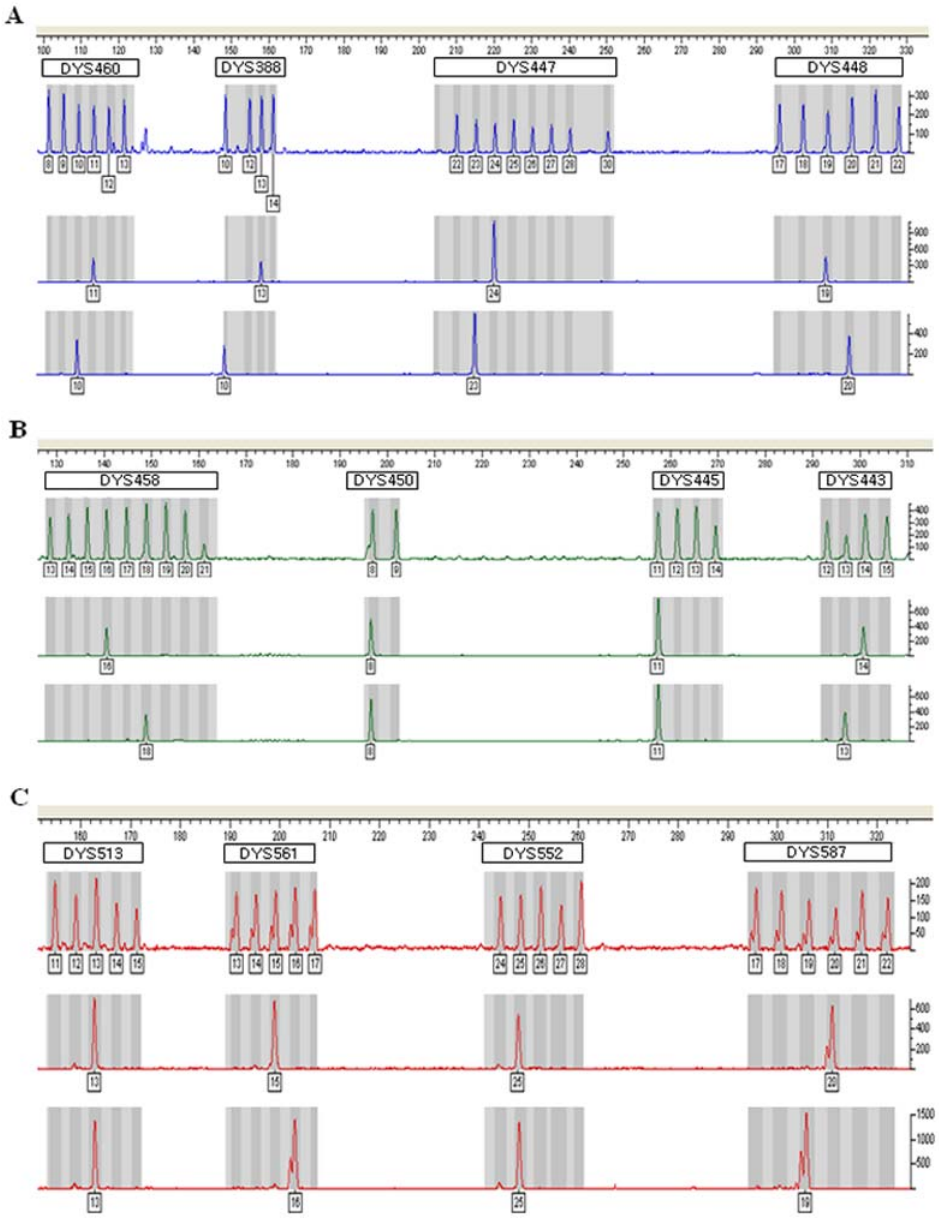

Figure 1. Coamplification of Y-STR markers using the dodecaplex Y-STR PCR system. The multiplex PCR was performed in a 20 μL of reaction volume containing 5 ng of genomic DNA and primer pairs with different concentrations shown in table 1 of electronic supplementary material. (A) Electropherograms of the 6FAM-labelled loci *DYS460*, *DYS388*, *DYS447* and *DYS448*, (B) the VIC-labelled loci *DYS458*, *DYS450*, *DYS445* and *DYS443*, and (C) the PET-labelled loci *DYS513*, *DYS561*, *DYS552* and *DYS587*.

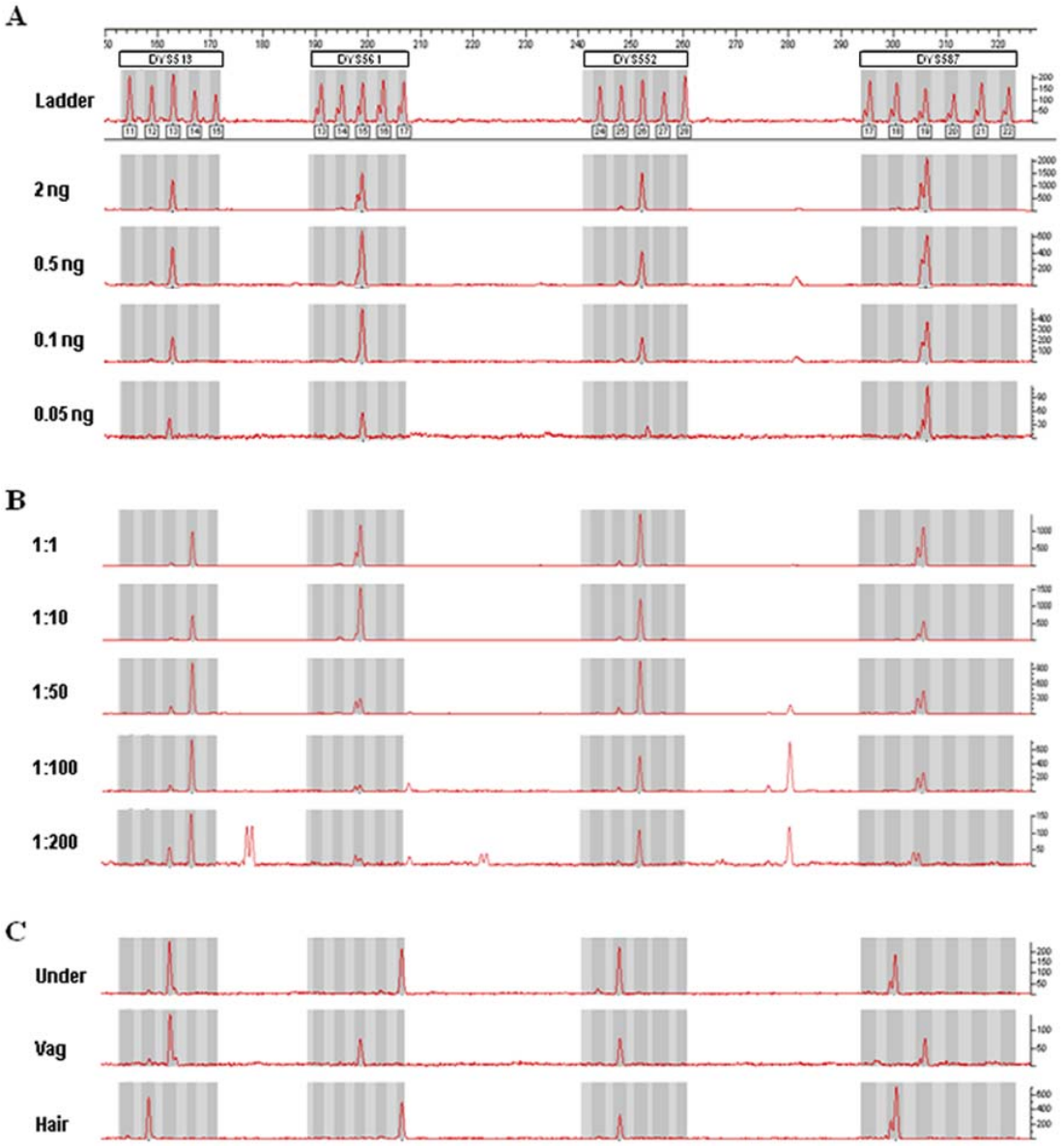

Figure 2. Sensitivity and forensic application of the Y-STR multiplex PCR system. The PCR reactions were performed by the same method as figure 1, except for different condition of template DNA in each reaction. Each electropherogram shows the PET-labelled four loci (*DYS513*, *DYS561*, *DYS552* and *DYS587*). (A) Amplification of serial diluted male DNA. Male DNA was serially diluted (2, 0.5, 0.1 and 0.05 ng) and used for multiplx PCR. (B) Testing of male–female mixed samples. The male DNA was kept constant (2 ng) with increasing amounts of female DNA with ratios of 1:1, 1:10, 1:50, 1:100 and 1:200. (C) Testing of forensic samples. DNA purification from the underwear (Under) and vaginal swab (Vag) were performed after separation of sperm layer from epithelial layer. Hair root (Hair) was used for DNA purification without layer separation.

exact allele type. This system also showed reliable allele typing in the male–female mixed DNA samples up to ratio of 1:100 (figure 2,B). When the male–female mixed sample of 1:200 was used, *DYS388* and *DYS561* were poorly amplified, and non-specific peaks appeared nearby *DYS388*, *DYS448* and *DYS460* peaks. No amplification was achieved by female DNA alone up to 50 ng as template (data not shown).

The allelic names were followed basically by the number of repeats (Gusmão *et al*. 2006), or by the previous studies (Butler *et al*. 2002; Iida *et al*. 2002; Redd *et al*. 2002; Dai *et al*. 2004; Schoske *et al*. 2004). Alleles of *DYS448* and *DYS458* have been named in two different ways. For example, the allele $(AGAGAT)_{11}N_{10}(AGAGAT)_3N_{14}(AGAGAT)_8$ of *DYS448* was named 22 by Butler *et al*. (2002) and 19 by Redd *et al*. (2002). Allele $(GAAA)_3N_6(GAAA)_{14}$ of *DYS458* was named 17 by Tang *et al*. (2003) and 14 by Redd *et al*. (2002). In this study, we designated each allele as 19 and 14, respectively, because these names have been widely used by

the AmpFlSTR Yfiler kit (Applied Biosystems, Foster City, USA). Alleles observed in 9948 male DNA cell line standard are presented in table 2 of electronic supplementary material.

Population data and forensic application

In this study, the total number of alleles observed were 70, and the mean allelic number per locus was 5.83. The largest allelic number was observed at *DYS447* with 10 alleles, and followed by *DYS458* with nine alleles, whereas the smallest number was observed at *DYS450* (only three alleles). GD ranged from 0.196 at *DYS450* to 0.764 at *DYS458*, with mean value of 0.557. Allele frequencies and GDs for the 12 Y-STRs are shown in table 1. By combined analysis of 12 loci studied in a sample of 320 unrelated males, 254 different haplotypes could be obtained. Of them, 223 haplotypes were unique. The complete haplotype analysis data are available from the corresponding author upon request. Three similar haplotypes were most frequently observed from 10 samples as followed: 12-13-11-25-18-8-17 or 18-11-13 or 14-25-15-19 (locus order; same as table 1 of electronic supplementary material). The values of haplotype diversity and discrimination capacity were calculated to be 0.996 and 0.794, respectively.

When the practical forensic samples (an underwear and two vaginal swabs from sexually assaulted women) were examined, the dodecaplex system showed distinguishable Y-STR typing results, although some alleles failed to amplify according to samples (e.g., *DYS458*, *DYS460* and *DYS513*). The dodecaplex system also provided exact typing data in the analysis of a hair root sample (figure 2,C).

Mutational analysis

Five cases of mutations were identified from four markers (once in each *DYS447*, *DYS561* and *DYS513*, and twice in

Table 1. Allele frequencies and gene diversity for 12 Y-STR loci in a Korean population (n = 320).

Locus	Allele	Frequency	GD*	Locus	Allele	Frequency	GD*
DYS388	10	0.091	0.403	*DYS450*	8	0.891	0.196
	12	0.753			9	0.106	
	13	0.153			10	0.003	
	14	0.003		*DYS458*	13	0.006	0.764
DYS443	12	0.019	0.546		14	0.006	
	13	0.628			15	0.091	
	14	0.194			16	0.141	
	15	0.153			17	0.359	
	16	0.006			18	0.266	
DYS445	10	0.003	0.549		19	0.100	
	11	0.553			20	0.028	
	12	0.378			21	0.003	
	13	0.063		*DYS513*	11	0.038	0.556
	14	0.003			12	0.200	
DYS447	20	0.003	0.727		13	0.625	
	21	0.003			14	0.116	
	22	0.013			15	0.019	
	23	0.122			17	0.003	
	24	0.241		*DYS552*	24	0.034	0.600
	25	0.431			25	0.572	
	26	0.113			26	0.250	
	27	0.053			27	0.091	
	28	0.019			28	0.053	
	30	0.003		*DYS561*	13	0.009	0.404
DYS448	17	0.047	0.728		14	0.009	
	18	0.397			15	0.744	
	19	0.259			16	0.206	
	20	0.203			17	0.031	
	21	0.075		*DYS587*	17	0.013	0.602
	22	0.019			18	0.166	
DYS460	8	0.003	0.603		19	0.584	
	9	0.178			20	0.166	
	10	0.234			21	0.056	
	11	0.559			22	0.016	
	12	0.022					
	13	0.003					

*Gene diversity.

DYS460) in a sample of 158 father/son pairs. All mutations were single repeat mutations, and they were observed only once per gametogenesis. Three cases exhibited repeat gain mutation, and other two cases exhibited repeat loss (*DYS447*, 25 > 26, *DYS460*, 11 > 10 and 11 > 12, *DYS561*, 15 > 16; *DYS513*, 13 > 12). Although *DYS460* exhibited two mutation events, we could not tell that the *DYS460* has particularly high mutation rate, because analysed meiosis number is not enough to reach a conclusion. The average mutation rate of 12 markers was estimated to be 2.64×10^{-3}/locus/meiosis (95% c.i.: 0.33×10^{-3} to 4.95×10^{-3}), which was similar to the Y-STR mutation rates in populations of Germany (Hohoff *et al*. 2007), and of Spain and Portugal (Gusmão *et al*. 2005).

Conclusion

Most markers in this system are highly informative except for *DYS450*. The parameters of haplotype diversity (0.996) and discrimination capacity (0.794) are also suitable for the Y-STR study. Although HD and discrimination capacity showed that 12-plex PCR system is somewhat less discriminative than the Yfiler kit (0.9996 and 0.9087, respectively) in Korean population (Kim *et al*. 2008), we thought that it could provide additional discriminative power of the Y-STR system, because most loci did not overlap with the Yfiler kit. The defined multiplex system showed high sensitivity in minimal genomic DNA, and provided clear male allele typing results for DNA from male–female mixtures and several forensic cases. In conclusion, we believe that dodecaplex Y-STR system has a high discriminating power to analyse kinship tests and practical forensic cases.

Acknowledgements

This study was supported by a grant of Korea Science and Engineering Foundation funded by the Korean Government (Ministry of Education, Science and Technology) (KOSEF Grant no: R01-2008-000-20604-0).

References

Asamura H., Fujimori S., Ota M., Oki T. and Fukushima H. 2008 Evaluation of miniY-STR multiplex PCR systems for extended 16 Y-STR loci. *Int. J. Leg. Med*. **122**, 43–49.

Butler J. M., Schoske R., Vallone P. M., Kline M. C., Redd A. and Hammer M. F. 2002 A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Sci. Int*. **129**, 10–24.

Dai H. L., Wang X. D., Li Y. B., Wu J., Zhang J., Zhang H. J. *et al*. 2004 Characterization and haplotype analysis of 10 novel Y-STR loci in Chinese Han population. *Forensic Sci. Int*. **145**, 47–55.

de Knijff P., Kayser M., Caglià A., Corach D., Fretwell N., Gehrig C. *et al*. 1997 Chromosome Y microsatellites: population genetic and evolutionary aspects. *Int. J. Leg. Med*. **110**, 134-140.

Frank W. E., Ralph H. C. and Tahir M. A. 2008 Y chromosome STR haplotypes and allele frequencies in a southern Indian male population. *J. Forensic Sci*. **53**, 248–251.

Gusmão L., Sánchez-Diz P., Calafell F., Martín P., Alonso C. A., Alvarez-Fernández F. *et al*. 2005 Mutation rates at Y chromosome specific microsatellites. *Hum. Mutat*. **26**, 520–528.

Gusmão L., Butler J. M., Carracedo A., Gill P., Kayser M., Mayr W. R. *et al*. 2006 DNA commission of the international society of forensic genetics: an update of the recommendations on the use of Y-STRs in forensic analysis. *Int. J. Leg. Med*. **120**, 191–200.

Hohoff C., Dewa K., Sibbing U., Hoppe K., Forster P. and Brinkmann B. 2007 Y-chromosomal microsatellite mutation rates in a population sample from northwestern Germany. *Int. J. Leg. Med*. **121**, 359–363.

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

Kayser M., Brauer S., Willuweit S., Schädlich H., Batzer M. A., Zawacki J. *et al*. 2002 Online Y-chromosomal short tandem repeat haplotype reference database (YHRD) for U.S. populations. *J. Forensic Sci*. **47**, 513–519.

Kim S. H., Kim N. Y., Hong S. B., Cho N. S., Kim J. J., Han M. S. and Kim W. 2008 Genetic polymorphisms of 16 Y chromosomal STR loci in Korean population. *Forensic Sci. Int. Genet*. **2**, e9–10.

Nonaka I., Minaguchi K. and Takezaki N. 2007 Y-chromosomal binary haplogroups in the Japanese population and their relationship to 16 Y-STR polymorphisms. *Ann. Hum. Genet*. **71**, 480–495.

Redd A. J., Agellon A. B., Kearney V. A., Contreras V. A., Karafet T., Park H. Y. *et al*. 2002 Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Sci. Int*. **130**, 97–111.

Robino C., Crobu F., Di Gaetano C., Bekada A., Benhamamouch S., Cerutti N. *et al*. 2008 Analysis of Y-chromosomal SNP haplotypes and STR haplotypes in an Algerian population sample. *Int. J. Leg. Med*. **122**, 251–255.

Roewer L. 2009 Y chromosome STR typing in crime casework. *Forensic Sci. Med. Pathol*. **5**, 77–84.

Roewer L., Krawczak M., Willuweit S., Nagy M., Alves C., Amorim A. *et al*. 2001 Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *Forensic Sci. Int*. **118**, 106–113.

Schoske R., Vallone P. M., Kline M. C., Redman J. W. and Butler J. M. 2004 High-throughput Y-STR typing of U. S. populations with 27 regions of the Y chromosome using two multiplex PCR assays. *Forensic Sci. Int*. **139**, 107–121.

Tang J. P., Hou Y. P., Li Y. B., Wu J., Zhang J. and Zhang H. J. 2003 Characterization of eight Y-STR loci and haplotypes in a Chinese Han population. *Int. J. Leg. Med*. **117**, 263-270.

Tautz D. 1989 Hypervariability of simple sequence as a general source for polymorphic DNA markers. *Nucleic Acids Res*. **17**, 6463–6471.

Received 7 January 2009, in final revised form 24 June 2009; accepted 16 July 2009
Published on the Web: 6 October 2009