# Genetic Characteristics of 22 Y-STR loci in Koreans

Myung Jin Park, Hwan Young Lee, Na Young Kim,
Jeong Eun Sim, Woo Ick Yang, Sang-Ho Cho,
Kyoung-Jin Shin

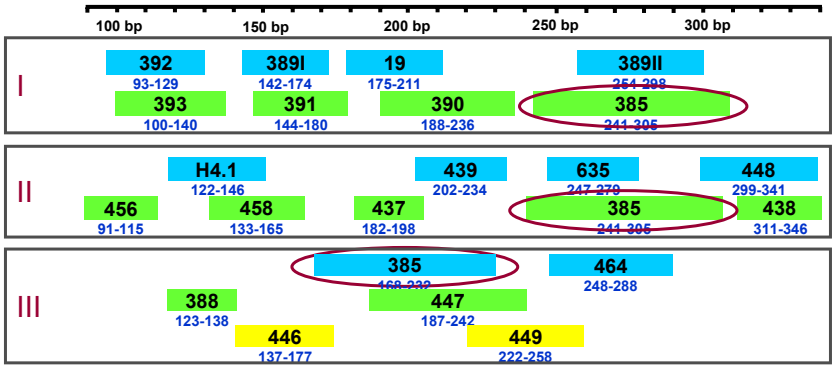
Department of Forensic Medicine, Yonsei University College of Medicine
Human Identification Research Center, Yonsei University



## Establishment of a Korean Y-STR database for 22 Y-STRs

- **22 Y-STRs**
  - DYS19, DYS385, DYS388, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS446, DYS447, DYS448, DYS449, DYS456, DYS458, DYS464, DYS635 and GATA H4.1
- **3 multiplex PCR systems**
- **708 unrelated Korean males**
- **Haplotype analysis and characterization of atypical alleles**

# Schematic of 3 multiplex PCRs for 22 Y-STRs




- DYS385 was included in each multiplex to detect sample switching
- Small amplicon was designed for **DYS385** in multiplex III (**168-232 bp vs. 241-305 bp**)



# Y-STR analysis in 708 Koreans

| Y-STR haplotype | No. of haplotype | Haplotype diversity | Discriminatory power |
|---|---|---|---|
| **Minimal Haplotype loci** | **485** | **0.9966** | **0.6850** |
| **17 Y-filer loci haplotype** | **657** | **0.9995** | **0.9280** |
| **22 Y-STRs haplotype** | **693** | **0.9999** | **0.9788** |

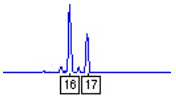
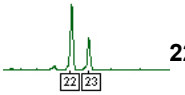
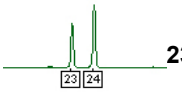
- Atypical alleles
  - **Intermediate alleles** at DYS447, DYS449, DYS458 and DYS464
  - **Duplicated alleles** at DYS19, DYS390 and DYS447
  - **Different allele designation at DYS385** depending on the primer binding site
  - **Null allele** at DYS448

# Microvariant and intermediate alleles in DYS447, DYS449, DYS458 and DYS464

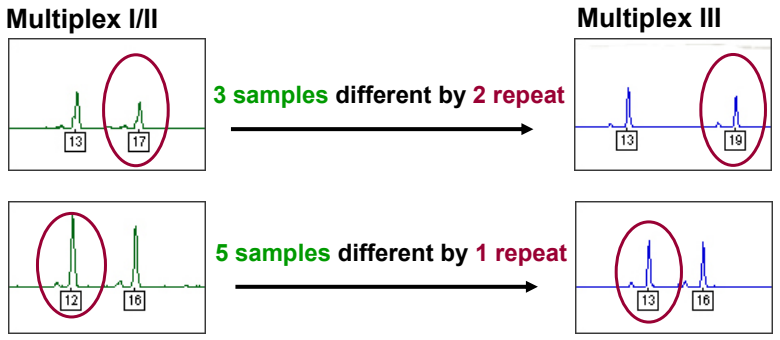
| Locus | Allele | Sequence structure | N |
|---|---|---|---|
| DYS447 | 18 | $(TAATA)_7TAAAA_1(TAATA)_{10}$~~$TAAAA(TAATA)_n$~~ | 2 |
| | 19 | $(TAATA)_7TAAAA_1(TAATA)_{11}$~~$TAAAA(TAATA)_n$~~ | 4 |
| DYS449 | 26 (25)* | $(TTTC)_{12}$-~~tctc~~-N10-N32-cttc-$(TTTC)_{14}$ | 1 |
| | 30 (30.1)* | $(TTTC)_{16}$-tctc-N10-N32-cttc-$(TTTC)_{14}$ | 1 |
| | 33 (42)* | $(TTTC)_{15}$-tctc-N10-$[N32\text{-}cttc]_2$-$(TTTC)_{18}$ | 1 |
| | 27.2 | $(TTTC)_3TC(TTTC)_{10}$-tctc-N10-N32-cttc-$(TTTC)_{14}$ | 1 |
| | 28.2 | $(TTTC)_3TT(TTTC)_{11}$-tctc-N10-N32-cttc-$(TTTC)_{14}$ | 1 |
| | 29.2 | $(TTTC)_3TC(TTTC)_{10}$-tctc-N10-N32-cttc-$(TTTC)_{14}$ | 1 |
| | 30.2 | $(TTTC)_{16}$-tctc-N10-N32-cttc- $(TTTC)_{10}TT(TTTC)_4$ | 12 |
| DYS458 | 14.1 | $(GAAA)_{14}G$ | 1 |
| | 17.2 | $(GAAA)_{15}AA(GAAA)_2$ | 1 |
| DYS464 | 12.3 | $(CCTT)_7CTT(CCTT)_5$ | 1 |
| | 14.3 | $(CCTT)_3CTT(CCTT)_{11}$ | 5 |

*** Ostensible alleles at DYS449**

# Duplicated alleles at DYS19, DYS390 and DYS447

| Locus | | Allele | Sequence structure |
|---|---|---|---|
| DYS19 |   | 16,17 | $(TAGA)_3tagg(TAGA)_{13,14}$ |
| DYS390 |   | 22,23 | $(TCTG)_8(TCTA)_{9,10}(TCTG)_1(TCTA)_4$ |
| DYS447 |   | 23,24 | $(TAATA)_7TAAAA(TAATA)_{7,8}TAAAA(TAATA)_7$ |

# Different allele designation at DYS385 in multiplex I/II and multiplex III



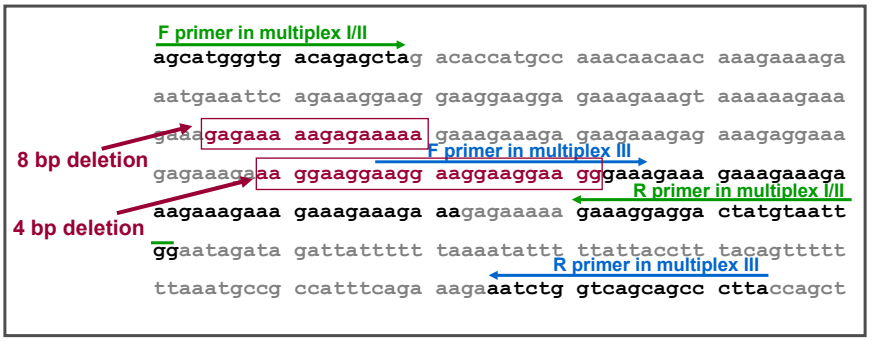



# Characterization of allele designation difference at DYS385

- Due to the deletion mutation in 8 samples, DYS385 allele designation in multiplex I and II was differently observed from that in multiplex III.

**8 bp deletion**: (**gagaaaaa**)$_{2\rightarrow1}$-N38-(aagg)$_6$**(GAAA)$_n$**

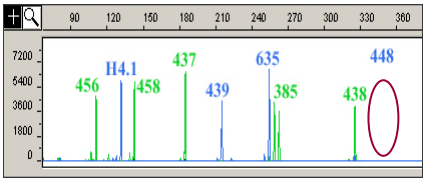
**4 bp deletion**: (gagaaaaa)$_2$-N38- (**aagg**)$_{6\rightarrow5}$**(GAAA)$_n$**



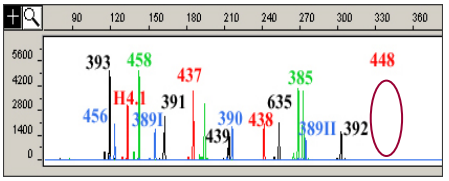
# Null allele at DYS448

- In 6 samples, null allele at DYS448 were commonly observed using multiplex II and Y-filer

**Multiplex II**

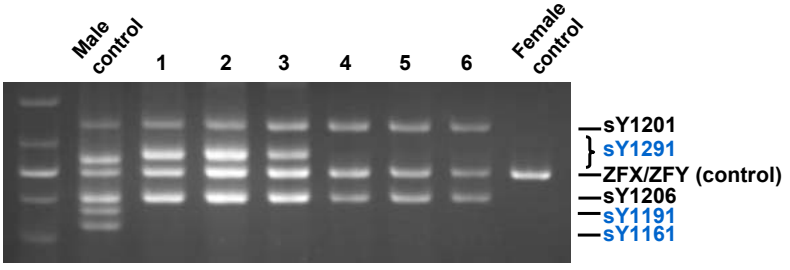



**AmpFISTR® Yfiler™ kit**



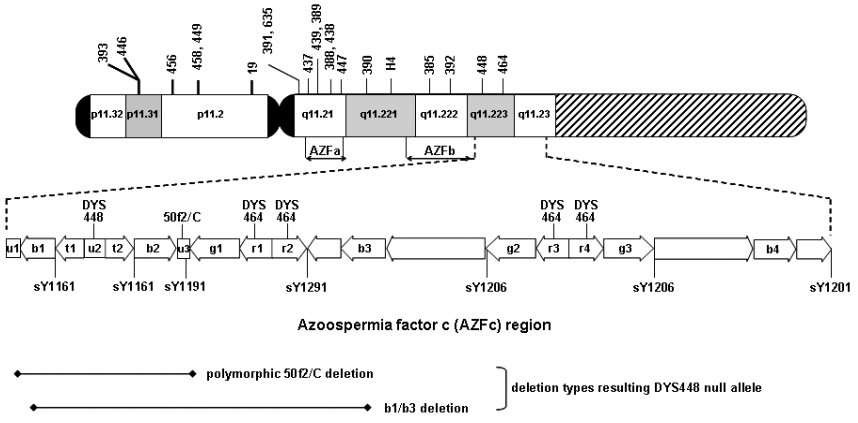



# Multiplex PCR assay of *AZFc* markers



# Characterization of the null allele at DYS448

- Schematic representation of Y-STRs and STS markers




**Null alleles at DYS448 were occurred with partial AZFc deletion.**

# Frequencies of null allele at DYS448 in various ethnic groups

| | | | |
|---|---|---|---|
| Korean | 6 in 708 | 0.85 % | Present study |
| Japanese | 10 in 1079 | 0.93 % | Mizuno et al. FSI In press |
| Nepalese | 3 in 769 | 0.39 % | Parkin et al. FSI 2006 |
| Malays, Chinese and Indian in Malaysia | 3 in 980 | 0.31 % | Chang et al. FSI 2006 |
| Kalmyk | 7 in 99 | 7.07 % | Roewer et al. FSI In press |
| Mexican | 1 in 326 | 0.31 % | Gutiérrez-Alarcón et al. Leg Med 2007 |
| Spanish | 1 in 247 | 0.41 % | Sánchez et al. FSI In press |
| Asian | 2 in 330 | 0.61 % | AmpFISTR® Yfiler™ database |
| African American | 2 in 985 | 0.20 % | AmpFISTR® Yfiler™ database |
| Caucasian (USA) | 2 in 1276 | 0.16 % | AmpFISTR® Yfiler™ database |

**The relatively high frequencies of the DYS448 null allele in Asians suggest giving careful consideration to the use of DYS448 for commercial genotyping and further database construction in Asians.**

# Mutation analysis of 22 Y-STRs

Int J Legal Med
DOI 10.1007/s00414-006-0130-x

ORIGINAL ARTICLE

## Haplotypes and mutation analysis of 22 Y-chromosomal STRs in Korean father–son pairs

**Hwan Young Lee · Myung Jin Park · Ukhee Chung · Han Young Lee · Woo Ick Yang · Sang-Ho Cho · Kyoung-Jin Shin**

Received: 29 May 2006 / Accepted: 11 September 2006
© Springer-Verlag 2006

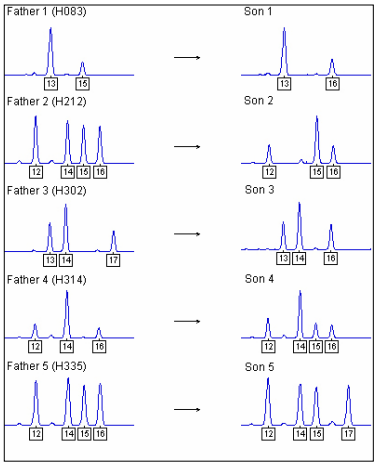
YONSEI UNIVERSITY COLLEGE OF MEDICINE

*Dept. of Forensic Medicine*

# Sequence information for 35 mutations observed at 16 among 22 Y-STRs

| Locus | Allele[a] | | Father haplotype[b] | Repetitive sequence structure → mutation | Type of mutation | No. of steps | Father's Age[c](years) |
|---|---|---|---|---|---|---|---|
| | Father | Son | | | | | |
| DYS19 | 15 | 16 | H121 | $(TAGA)_3TAGG(TAGA)_{12\to13}$ | Gain | One | ND |
| DYS19 | 17 | 16 | H346 | $(TAGA)_3TAGG(TAGA)_{14\to13}$ | Loss | One | 33 |
| DYS385 | 10-**18** | 10-**17** | H154 | $(GAAA)_{18\to17}$ | Loss | One | 35 |
| DYS385 | 13-**20** | 13-**19** | H025 | $(GAAA)_{20\to19}$ | Loss | One | 28 |
| DYS389I | 14 | 13 | H349 | $(TCTG)_4(TCTA)_{11\to10}$ | Loss | One | ND |
| DYS389II | 29 | 30 | H291 | $(TCTG)_4(TCTA)_{11\to12}(TCTG)_3(TCTA)_{11}$ | Gain | One | 27 |
| DYS389II | 32 | 31 | H147 | $(TCTG)_4(TCTA)_{13\to14}(TCTG)_3(TCTA)_{10}$ | Loss | One | 27 |
| DYS390 | 24 | 25 | H011 | $(TCTG)_4(TCTA)_{11\to12}(TCTG)_3(TCTA)_8$ | Gain | One | 34 |
| DYS393 | 16 | 15 | H133 | $(AGAT)_{16\to15}$ | Loss | One | 31 |
| DYS437 | 14 | 15 | H285 | $(TCTA)_{8\to9}(TCTG)_2(TCTA)_8$ | Gain | One | 37 |
| DYS439 | 13 | 12 | H098 | $(GATA)_{13\to12}$ | Loss | One | 25 |
| DYS439 | 13 | 12 | H254 | $(GATA)_{13\to12}$ | Loss | One | 31 |
| DYS446 | 13 | 12 | H279 | $(TCTCT)_{13\to12}$ | Loss | One | 34 |
| DYS447 | **23**-24 | 24 | H010 | $(TAATA)_7TAAAA(TAATA)_{7\to6}TAAAA(TAATA)_7$ or deletion | – | – | 28 |
| DYS447 | 26 | 25 | H289 | $(TAATA)_7TAAAA(TAATA)_{8\to6}TAAAA(TAATA)_8$ | Loss | One | ND |
| DYS449 | 29 | 28 | H204 | $(TTTC)_3N50(TTTC)_{4\to3}$ | Loss | One | 28 |
| DYS449 | 30 | 31 | H306 | $(TTTC)_{5\to6}N50(TTTC)_5$ | Gain | One | 36 |
| DYS449 | 31 | 32 | H040 | $(TTTC)_{5\to6}N50(TTTC)_6$ | Gain | One | ND |
| DYS449 | 32 | 31 | H112 | $(TTTC)_6N50(TTTC)_{7\to6}$ | Loss | One | 34 |
| DYS449 | 32 | 33 | H084 | $(TTTC)_{5\to6}N50(TTTC)_7$ | Gain | One | ND |
| DYS449 | 33 | 34 | H155 | $(TTTC)_7N50(TTTC)_{8\to9}$ | Gain | One | 30 |
| DYS449 | 34 | 35 | H267 | $(TTTC)_8N50(TTTC)_{8\to9}$ | Gain | One | ND |
| DYS456 | 15 | 16 | H258 | $(AGAT)_{15\to16}$ | Gain | One | 30 |
| DYS456 | 18 | 17 | H057 | $(AGAT)_{18\to17}$ | Loss | One | 32 |
| DYS458 | 18 | 17 | H294 | $(GAAA)_{18\to17}$ | Loss | One | 27 |
| DYS458 | 18 | 19 | H260 | $(GAAA)_{18\to19}$ | Gain | One | 26 |
| DYS458 | 19 | 20 | H118 | $(GAAA)_{19\to20}$ | Gain | One | 28 |
| DYS464 | 12-**14-16** | 12-14-**15**-16 | H314 | $(CCTT)_{14\to15}$ or $(CCTT)_{16\to15}$ | – | One | ND |
| DYS464 | 12-**14**-15-16 | 12-**15**-16 | H212 | $(CCTT)_{14\to15}$ | Gain | One | 33 |
| DYS464 | 12-14-15-**16** | 12-14-15-**17** | H335 | $(CCTT)_{16\to17}$ | Gain | One | 30 |
| DYS464 | 13-**15** | 13-**16** | H083 | $(CCTT)_{15\to16}$ | Gain | One | 29 |
| DYS464 | 13-14-**17** | 13-14-**16** | H302 | $(CCTT)_{17\to16}$ | Loss | One | 35 |
| DYS635 | 21 | 22 | H177 | $(TCTA)_4(TGTA)_2(TCTA)_2(TGTA)_2(TCTA)_{11\to12}$ | Gain | One | 32 |
| DYS635 | 22 | 21 | H203 | $(TCTA)_4(TGTA)_2(TCTA)_2(TGTA)_2(TCTA)_{12\to11}$ | Loss | One | 33 |
| DYS635 | 22 | 21 | H247 | $(TCTA)_4(TGTA)_2(TCTA)_2(TGTA)_2(TCTA)_{12\to11}$ | Loss | One | 40 |
| GATA H4.1 | 21 | 20 | H076 | $(AGAT)_2CTAT(AGAT)_2(AGGT)_2(AGAT)_{11\to10}$ | Loss | One | 34 |

# Electropherograms for DYS464 in five father-son pairs with mutation events






# Comparison on Y-STR mutation rates between the present study and previous literature summary

| Locus | This work | | | Previous literature summary[a] | | | Total | |
|---|---|---|---|---|---|---|---|---|
| | No. of mutations | No. of allele transmissions | Mutation rate ($\times 10^{-3}$) | No. of mutations | No. of allele transmissions | Mutation rate ($\times 10^{-3}$) | Mutation rate ($\times 10^{-3}$) | 95% CI ($\times 10^{-3}$) |
| DYS19 | 2 | 369 | 5.4 | 12 | 7,585 | 1.58 | 1.76 | 0.96–2.95 |
| DYS385 | 2 | 738 | 2.7 | 22 | 10,749 | 2.05 | 2.09 | 1.34–3.11 |
| DYS388 | 0 | 369 | 0.0 | 0 | 109 | 0.00 | 0.00 | 0.00–6.27 |
| DYS389I | 1 | 369 | 2.7 | 14 | 5,789 | 2.42 | 2.44 | 1.36–4.02 |
| DYS389II | 2 | 369 | 5.4 | 14 | 5,776 | 2.42 | 2.60 | 1.49–4.23 |
| DYS390 | 1 | 369 | 2.7 | 16 | 7,063 | 2.27 | 2.29 | 1.33–3.66 |
| DYS391 | 0 | 369 | 0.0 | 23 | 7,015 | 3.28 | 3.11 | 1.97–4.67 |
| DYS392 | 0 | 369 | 0.0 | 5 | 6,981 | 0.72 | 0.68 | 0.22–1.59 |
| DYS393 | 1 | 369 | 2.7 | 4 | 5,770 | 0.69 | 0.81 | 0.26–1.90 |
| DYS437 | 1 | 369 | 2.7 | 5 | 2,598 | 1.92 | 2.02 | 0.74–4.40 |
| DYS438 | 0 | 369 | 0.0 | 1 | 2,638 | 0.38 | 0.33 | 0.01–1.85 |
| DYS439 | 2 | 369 | 5.4 | 14 | 2,612 | 5.36 | 5.37 | 3.07–8.72 |
| DYS446 | 1 | 369 | 2.7 | – | – | – | 2.71 | 0.07–15.10 |
| DYS447 | 2 | 370 | 5.4 | – | – | – | 5.41 | 0.65–19.53 |
| DYS448 | 0 | 369 | 0.0 | 0 | 344 | 0.00 | 0.00 | 0.00–4.20 |
| DYS449 | 7 | 369 | 19.0 | – | – | – | 18.97 | 7.63–39.09 |
| DYS456 | 2 | 369 | 5.4 | 2 | 346 | 5.78 | 5.59 | 1.52–14.32 |
| DYS458 | 3 | 369 | 8.1 | 3 | 347 | 8.65 | 8.38 | 3.08–18.24 |
| DYS464 | 5 | 1,476 | 3.4 | 2 | 280 | 7.14 | 3.99 | 1.60–8.21 |
| DYS635 | 3 | 369 | 8.1 | 6 | 1,221 | 4.91 | 5.66 | 2.59–10.75 |
| GATA H4.1 | 1 | 369 | 2.7 | 5 | 1,382 | 3.62 | 3.43 | 1.26–7.46 |
| Total | 36 | 9,226 | 3.9 | 148 | 68,605 | 2.16 | 2.36 | 2.03–2.73 |

## Numbers of haplotypes and haplotype diversities obtained by adding each marker to the extended SWGDAM haplotype

| Haplotype | Gene diversity[a] | No. of haplotypes | Discriminatory capacity (%) | Haplotype diversity |
|---|---|---|---|---|
| Extended SWGDAM haplotype | | 558 | 78.8 | 0.9982 |
| Extended SWGDAM haplotype + DYS449 | 0.8521 | 612 | 86.4 | 0.9992 |
| Extended SWGDAM haplotype + DYS446 | 0.7894 | 598 | 84.5 | 0.9989 |
| Extended SWGDAM haplotype + DYS458 | 0.7783 | 616 | 87.0 | 0.9992 |
| Extended SWGDAM haplotype + DYS447 | 0.7523 | 583 | 82.3 | 0.9986 |
| Extended SWGDAM haplotype + DYS448 | 0.7508 | 565 | 79.8 | 0.9983 |
| Extended SWGDAM haplotype + DYS635 | 0.6858 | 591 | 83.5 | 0.9988 |
| Extended SWGDAM haplotype + GATA H4.1 | 0.6112 | 579 | 81.8 | 0.9985 |
| Extended SWGDAM haplotype + DYS456 | 0.5098 | 574 | 81.1 | 0.9984 |
| Extended SWGDAM haplotype + DYS388 | 0.4855 | 566 | 79.9 | 0.9983 |
| Extended SWGDAM haplotype + DYS437 | 0.4324 | 568 | 80.2 | 0.9983 |
| Extended SWGDAM haplotype + DYS458 + DYS635 | | 639 | 90.3 | 0.9994 |
| Extended SWGDAM haplotype + DYS458 + DYS447 | | 630 | 89.0 | 0.9994 |
| Extended SWGDAM haplotype + DYS458 + GATA H4.1 | | 631 | 89.1 | 0.9993 |
| Extended SWGDAM haplotype + DYS458 + DYS635 + DYS447 | | 650 | 91.8 | 0.9996 |
| Extended SWGDAM haplotype + DYS458 + DYS635 + GATA H4.1 | | 649 | 91.5 | 0.9995 |
| Extended SWGDAM haplotype + DYS458 + DYS635 + DYS447 + GATA H4.1 | | 658 | 92.9 | 0.9996 |
| Y-filer™ 17 Y-STR haplotype | | 657 | 92.8 | 0.9995 |

[a] Gene diversity of added marker to SWGDAM haplotype loci



## Concluding remarks

1. 18 microvariant alleles at DYS449, 6 intermediate alleles at DYS464, 2 intermediate alleles at DYS458, and 1 duplicated allele at DYS19, DYS390 and DYS447 have been discovered and characterized.
2. At DYS385, allele size differences were one or two repeats dependent on the primer set used for PCR amplification and two widely used commercial kits amplify DYS385 so as to include the mutable sites.
3. Arrangement analysis of sequence tagged sites demonstrated that the deletion patterns at DYS448 (and DYS464) were associated with arrangements of the *AZFc* gene and the DYS448 deletion appears relatively frequent in Asians.
4. The combined haplotypes of DYS447, DYS458, DYS635, GATA H4.1, and SWGDAM Y-STR loci was comparable to haplotypes of 17 loci in the AmpFlSTR® Yfiler™ kit in Koreans.

**Genetic characteristics of 22 Y-STRs in Koreans: Atypical alleles, deletions in the DYS385 flanking regions, and null alleles associated with AZFc microdeletions**

Myung Jin Park[1], Hwan Young Lee[1], Na Young Kim[1], Soong Deok Lee[2], Woo Ick Yang[1], Sang-Ho Cho[1,3], Kyoung-Jin Shin[1,3]

[1]Department of Forensic Medicine, Yonsei University College of Medicine, Seoul, Korea.
[2]Department of Forensic Medicine, Seoul National University College of Medicine, Seoul, Korea.
[3]Human Identification Research Center, Yonsei University, Seoul, Korea

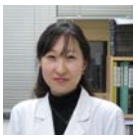
**22nd Congress of the International Society for Forensic Genetics**
**Copenhagen, Denmark, August 2007.**

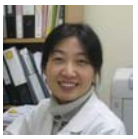


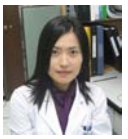

# Acknowledgement

- Funding
  - National R&D program of Ministry of Science and Technology (MOST)
  - Korea Science and Engineering Foundation (KOSEF)

- To our lab members


**Hwan Young Lee, Ph.D**


**Myung Jin Park, MSc.**


**Na Young Kim, MSc.**


**Jeong Eun Sim, BSc.**