



## Y-SNP miniplexes for East Asian Y-chromosomal haplogroup determination in degraded DNA

Myung Jin Park<sup>a</sup>, Hwan Young Lee<sup>a,b</sup>, Na Young Kim<sup>a</sup>, Eun Young Lee<sup>a</sup>, Woo Ick Yang<sup>a</sup>,  
Kyoung-Jin Shin<sup>a,b,\*</sup>

<sup>a</sup> Department of Forensic Medicine and Brain Korea 21 Project for Medical Science, Yonsei University College of Medicine, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-752, South Korea

<sup>b</sup> Human Identification Research Center, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-752, South Korea

### ARTICLE INFO

#### Article history:

Received 10 November 2011

Received in revised form 14 May 2012

Accepted 26 June 2012

#### Keywords:

Y chromosome haplogroup

Single nucleotide polymorphism

Degraded DNA

Small size amplicon

East Asians

### ABSTRACT

Four multiplex PCR systems followed by single base extension reactions were developed to score 22 single nucleotide polymorphisms (SNPs) and identify the most frequent East Asian Y chromosome haplogroups. Select Y chromosome SNPs allowed hierarchical testing for almost all of the major East Asian haplogroups along the revised Y chromosome tree. The first multiplex consists of six SNPs defining world-wide major haplogroups (M145, RPS4Y<sub>711</sub>, M89, M9, M214, and M175). The second multiplex includes six SNPs of subhaplogroup O (M119, P31, M95, SRY<sub>465</sub>, 47z, and M122). The third multiplex contains six SNPs that subdivide the subhaplogroup O3 (M324, P201, M159, M7, M134, and M133). The fourth multiplex comprises four SNPs of subhaplogroup C (M217, M48, M407, and P53.1). The sizes of the PCR amplicons ranged from 70 to 100 bp to facilitate their application to degraded forensic and ancient samples. Validation experiments demonstrated that the multiplexes were optimized for analysis of low template DNA and highly degraded DNA. In a test using DNA samples from 300 Korean males, 16 different Y chromosome haplogroups were identified; haplogroup O2b\* was the most frequently observed (29.3%), followed by haplogroups C3 (xC3c, C3d, C3e) (16.0%) and O3a3c1 (11.0%). These multiplex sets will be useful tools for Y-chromosomal haplogroup determination in anthropological and forensic studies of East Asian populations.

© 2012 Elsevier Ireland Ltd. All rights reserved.

### 1. Introduction

Genetic variations in the non-recombining portion of the Y chromosome (NRY) are analyzed in diverse disciplines including anthropological, forensic and medical genetics [1–5]. Because of a lack of recombination and low mutation rates, Y chromosome single nucleotide polymorphisms (Y-SNPs) are the most useful genetic markers for reconstructing male lineages. Therefore, Y chromosome haplogroups, which are defined by combinations of allelic states at hierarchically arranged Y-SNPs and small indels, have been extensively studied to infer the origins, evolution, and histories of migrations of modern human populations [6,7].

A number of changes have been made to the Y-chromosomal haplogroup tree, and a total of 311 distinct haplogroups have been defined with increased resolution [8]. The haplogroup O was

considerably rearranged during this revision; the L1 retroposon insertion (LINE 1) polymorphism, which had conflicted with the N7 polymorphism, was excluded from the list of markers used to define subhaplogroup O3. Persistent commercial or in-house development efforts for Y-SNP typing protocols have been made, but most developed methods involve typing of European Y haplogroups or world-wide major haplogroups with low resolution [9–12] and are not suitable to subdivide subhaplogroup O3, a major haplogroup in East Asians [13,14], into internal derivatives following the revised Y haplogroup tree. In addition, as the demand for inferring geographic origin is increasing in forensic DNA analysis as well as in ancient DNA analysis (e.g., identification of Korean War and Vietnam War victims, and genetic characterization of ancient remains), the development of sensitive and efficient methods for the Y-chromosomal haplogroup determination in degraded DNA is required.

Therefore, in the current study, a multiplex single base extension (SBE) method was developed to score Y-SNPs of East Asian haplogroups following a small size amplicon strategy that is suitable for application to degraded DNA. Y-SNPs were selected from SNPs that are hierarchically located along the

\* Corresponding author at: Department of Forensic Medicine, Yonsei University College of Medicine, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-752, South Korea. Tel.: +82 2 2228 2481; fax: +82 2 362 0860.

E-mail address: [kjshin@yuhs.ac](mailto:kjshin@yuhs.ac) (K.-J. Shin).

revised topology of Y chromosome haplogroups, focusing on East Asian-specific haplogroups. To assess the utility of our method for the analysis of highly degraded samples, the sensitivity and efficiency of the multiplex set were validated in samples of serially diluted DNA, artificially degraded DNA, and DNA extracted from 55-year-old skeletal remains. Finally, a Korean sample was analyzed using the newly developed multiplex systems, and the distribution of Y haplogroups was studied, since the geographic origins and history of migration of a population can be inferred from the distributions and ages of its haplogroups [13–15].

## 2. Materials and methods

### 2.1. DNA samples

Our study protocol was approved by the Institutional Review Board of Severance Hospital, Yonsei University in Seoul, Korea. DNA samples from 300 unrelated Korean males were obtained from the National Biobank of Korea. DNA concentrations were measured using a NanoDrop<sup>®</sup> ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and the final sample concentrations were adjusted to 1.0 ng/μl. For the sensitivity test, 9948 male DNA (Promega, Madison, WI, USA) was serially diluted to concentrations of 1000, 500, 250, 125, 62, 31, and 15 pg/μl. Artificially degraded DNA was prepared by digesting 1.2 μg of human genomic DNA with 0.02 U of DNase I (NEB, Ipswich, MA, USA) at 37 °C for 40 min. DNA degradation to fragment sizes of around 100 bp was confirmed by ethidium bromide staining after agarose gel electrophoresis (data not shown). Ten DNA samples that were extracted from 55-year-old skeletal remains during a previous study [16] were analyzed to evaluate multiplex SBE reactions. Concentrations of DNA obtained from skeletal remains were determined using the Quantifiler<sup>™</sup> Human DNA Quantification Kit and 7500 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) [16].

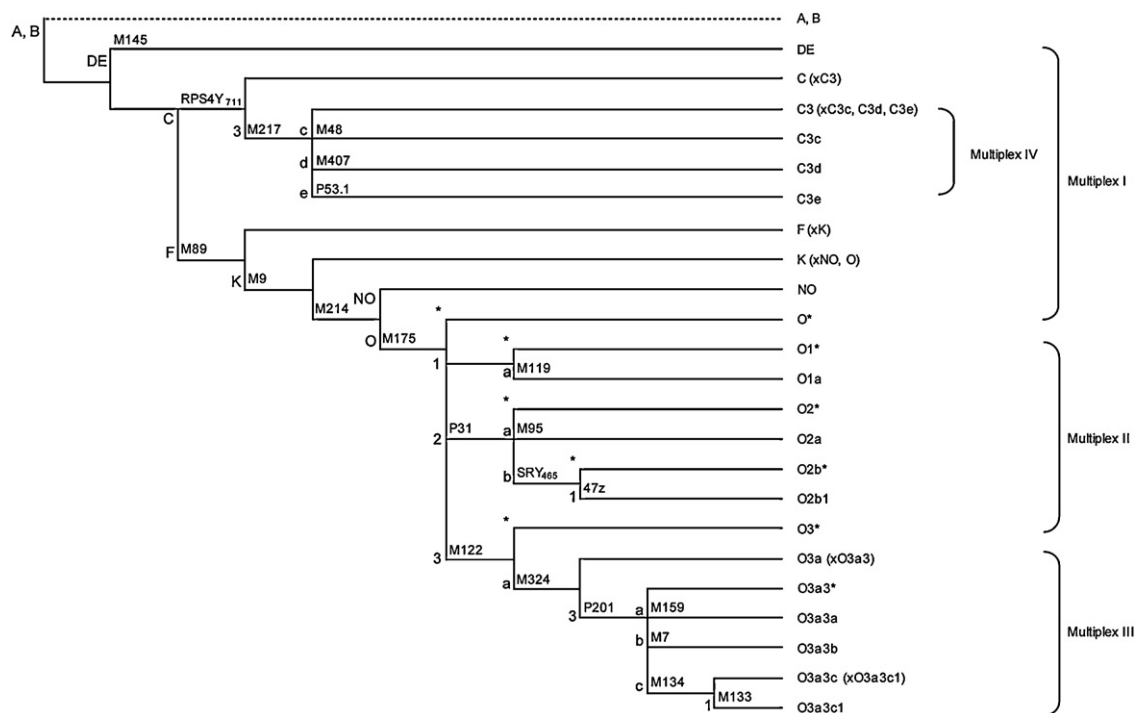
### 2.2. Y-SNP selection and primer design for PCR amplification and SBE

A set of 22 biallelic Y chromosome markers (M7, M9, M48, M89, M95, M119, M122, M133, M134, M145, M159, M175, M214, M217, M324, M407, P31, P53.1, P201, RPS4Y<sub>711</sub> (M130), SRY<sub>465</sub> (M176), and 47z) was selected to determine the world-wide major haplogroups, subhaplogroups O, and subhaplogroups C that are present in East Asian populations, including Koreans [13–15,17]. The 22 Y-SNPs and the haplogroup tree defined by these markers are shown in Fig. 1. The nomenclature and topology of the Y chromosome haplogroups followed those of Karafet et al. [8].

Primers for PCR amplification and subsequent SBE were designed using web-based programs Primer 3 (<http://frodo.wi.mit.edu/primer3/>) and Batchprimer 3 (<http://batchprimer3.bioinformatics.ucdavis.edu/index.html>), respectively, based on sequences obtained from GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) or dbSNP (<http://www.ncbi.nlm.nih.gov/snp/>). PCR primers were then selected to produce amplicons smaller than 100 bp and to amplify male-specific fragments in tests using 1 ng of each male and female DNA. SBE primers were designed to provide even peak heights and appropriate peak intervals by adjusting  $T_m$  values and adding poly-T and/or poly-A tails to the 5' ends of primers. PCR and SBE primer sequences are shown in Supplementary material Tables S1 and S2, respectively.

### 2.3. Multiplex PCR amplification and PCR product purification

A total of four multiplex PCR systems were developed to type 22 Y-SNPs (Fig. 1). PCR amplifications were performed in a final volume of 25 μl that contained 1 ng of template DNA, 2.5 μl of Gold ST<sup>®</sup>R 10× buffer (Promega), 2.0 U of AmpliTaq Gold<sup>®</sup> DNA polymerase (Applied Biosystems), and appropriate concentrations of primers (Supplementary material Table S1). Regarding multiplex IV, 1.5 U of AmpliTaq Gold DNA Polymerase were used. Thermal cycling was performed on a Veriti 96-Well Thermal Cycler (Applied Biosystems) under the following conditions: 95 °C for 11 min; 33 cycles of 94 °C for 20 s, 60 °C for 1 min and 72 °C for



**Fig. 1.** Phylogenetic tree of the 22 Y-chromosomal binary polymorphisms analyzed in this study. The analyzed Y-SNPs are shown in each branch, and the corresponding haplogroups and multiplexes are shown at the end of each branch according to Karafet et al. [8]. Broken-line branches correspond to haplogroups not found in East Asians.

30 s; and a final extension of 72 °C for 7 min. For the following SBE reaction, 5.0 µl of the PCR product was purified by incubating at 37 °C for 45 min with 1.0 µl of ExoSAP-IT (USB, Cleveland, OH, USA). The enzyme was then inactivated at 80 °C for 15 min.

#### 2.4. Multiplex SBE reactions and SNP scoring

Four multiplex SBE reactions were carried out with each purified multiplex PCR product, SBE primer mix (Supplementary material Table S2) and a SNaPshot™ Multiplex kit (Applied Biosystems) according to the manufacturer's instructions. Thermal cycling was performed on a Veriti 96-Well Thermal Cycler with 25 cycles at 96 °C for 10 s, 50 °C for 5 s and 60 °C for 30 s. After the SBE reaction, 1.0 U of SAP (USB) was added to the extension product, and the mix was incubated at 37 °C for 45 min to remove the unincorporated ddNTPs. SAP was inactivated by incubation at 80 °C for 15 min.

The final products were mixed with GeneScan™ 120 LIZ® size standard (Applied Biosystems) and analyzed by capillary electrophoresis with an ABI PRISM 310 Genetic Analyzer (Applied Biosystems) and GeneMapper® ID software 3.2 (Applied Biosystems). SNPs were scored when their peak heights were above the interpretational threshold of 100 relative fluorescent units.

#### 2.5. PCR amplification with serially diluted DNA and degraded DNA

Five replicates of serially diluted 9948 male DNA were amplified independently under the same PCR conditions described above. Additional five replicates of serially diluted 9948 male DNA were then amplified independently under the same PCR conditions except for the number of PCR cycles: 33 cycles for 1 ng, 500 pg, and 250 pg of template DNA; 35 cycles for 125 and 62 pg; and 37 cycles for 31 and 15 pg. The genotype results for each diluted DNA sample were compared to known genotypes that were determined with 1 ng of standard DNA. To test the effect of input DNA amount on SNP scoring, the peak height of each amplicon was obtained with GeneMapper software, and average and standard deviation values were calculated at each concentration.

One microliter of artificially degraded DNA was independently amplified twice under the same PCR conditions as multiplex PCR amplification of 1 ng DNA. The genotyping results were then confirmed to be identical to those of 1 ng non-degraded human genomic DNA. Ten DNA samples that were extracted from 55-year-old skeletal remains [16] were also amplified with at least 60 pg of template DNA under the same PCR conditions except for 35 PCR cycles. Then, following the low copy number DNA interpretation rule (i.e., replicate analyses with duplicate results prior to reporting alleles) [18,19], SNPs were scored only when the SNP was observed in common from two replicate reactions.

### 3. Results and discussion

#### 3.1. Y-SNP selection and primer design

The selected 22 Y-SNP markers were amplified with four multiplex PCR systems to explore East Asian Y chromosome haplogroups. Multiplex I was composed of the six Y-SNPs, M145, RPS4Y<sub>711</sub>, M89, M9, M214, and M175, which distinguish all worldwide major clades except for the African lineages A and B. Among the six major haplogroups in Multiplex I, clade O is the most common haplogroup in East Asians, and accordingly, a more detailed analysis of the haplogroup O lineage is necessary to differentiate East Asian Y haplogroups. Therefore, Multiplex II (M119, P31, M95, SRY<sub>465</sub>, 47z, and M122) and Multiplex III (M324, P201, M159, M7, M134, and M133) were constructed to subdivide

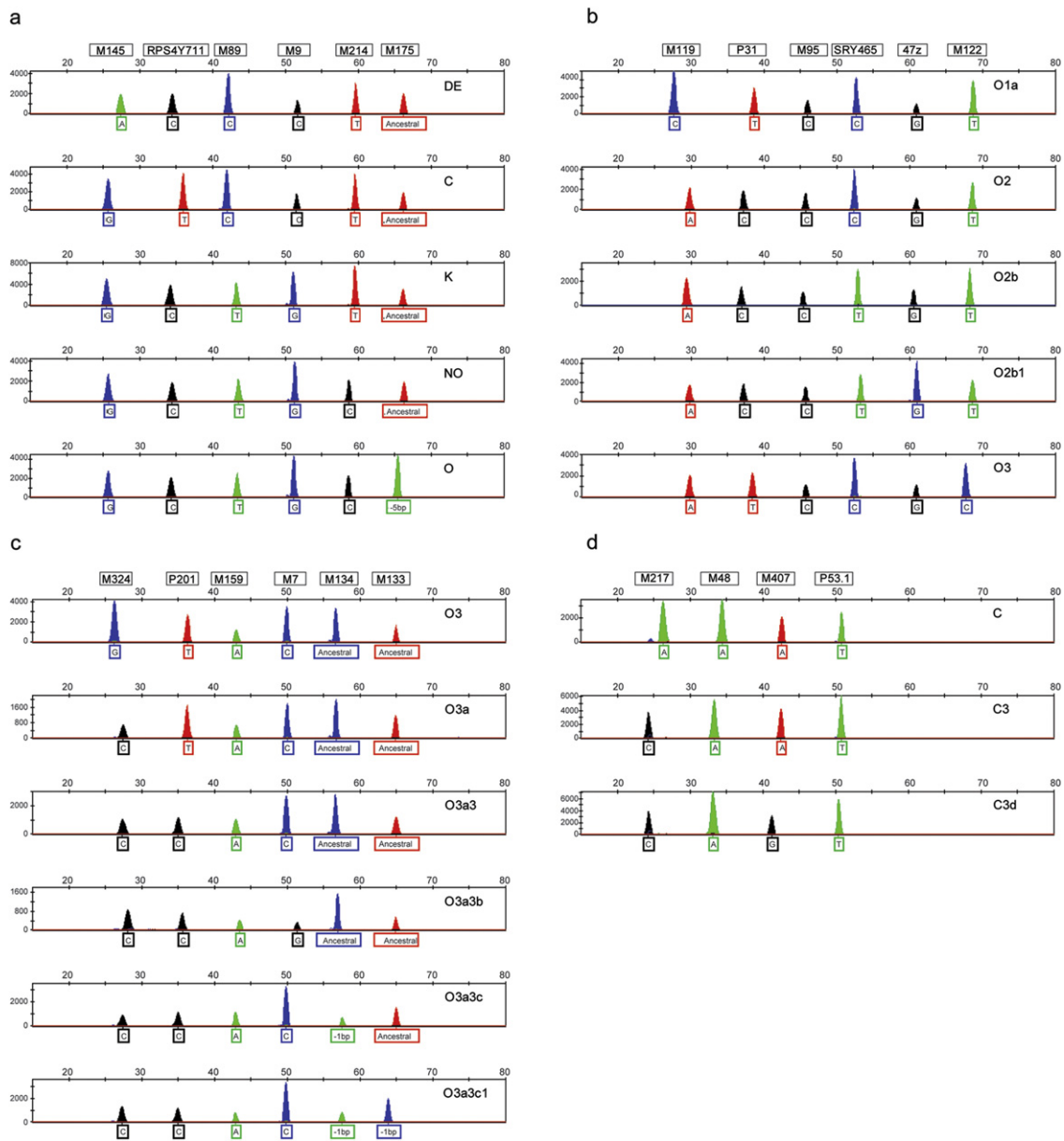
haplogroup O into subhaplogroups O1–O3 and their internal derivatives (Fig. 1). Finally, Multiplex IV (M48, M217, M407, and P53.1) was developed to further define subhaplogroups C, especially haplogroup C3 and its derivatives which are widespread in East Asia [15].

All PCR primers were designed to produce small amplicons between 70 and 100 bp in length to facilitate the analysis of degraded forensic and ancient samples (Supplementary material Table S1) [20,21]. For amplification of the 47z marker (DXYS5) that is located in a Y-chromosomal region with high sequence similarity to the X chromosome (DXYS5X), the 3' end of the reverse PCR primer was designed to be located at a nucleotide position that shows a mismatch between DXYS5Y and DXYS5X sequences, thereby allowing specific amplification of the target sequence on the Y chromosome only (DXYS5Y). In addition, a degenerate forward primer was used to allow the amplification of alternate assembly (based on HuRef, sequence accession number NW001842422.1) as well as the Genbank reference sequence (sequence accession number AC019058.4). Male-specific amplification of the 47z marker on the Y chromosome (DXYS5Y) was confirmed in several male and female DNA samples (data not shown).

SBE primers were designed to produce extension products ranging in size from 22 to 67 bases. Each extension product had a size difference of more than 7 bases from all others, so that the extension products could be easily discriminated during electrophoresis. For this reason, the 5' ends of certain primers were tailed with poly-T nucleotides. If the complementary nucleotide for the tail was A, nucleotide A was used instead of nucleotide T (Supplementary material Table S2). In case of the M133 marker primer with a 3' end AAAAAA sequence, a 5' poly-A tail was used instead of a poly-T tail to avoid self-complementary binding of the primer (Supplementary material Table S2). Representative electropherograms of the four developed multiplexes are shown in Fig. 2.

#### 3.2. Sensitivity and efficiency of multiplex SBE reactions

To evaluate the newly developed multiplexes, their sensitivities were tested with five replicates of DNAs at various concentrations. Average peak height from five replicates showed no linear correlation with DNA concentrations under the condition of 33 PCR cycles (Supplementary material Fig. S1a). This was explained by the fact that the amount of amplified products in plateau phase after approximately 30 cycles was not proportional to the amount of input DNA [22]. When the increased PCR cycles were used for the diluted DNA samples to improve the amplification yield (Supplementary material Fig. S2a), the peak intensities were increased and allele losses showed a tendency to decrease (Supplementary material Figs. S1b vs. S2b). However, the less input DNA used, the more standard deviation increased, which was caused by the stochastic effect [23]. As a result, all 22 SNPs were successfully typed with DNA amounts as low as 62 pg and 35 PCR cycles, at which only P31 marker failed to generate signal on only one occasion (Supplementary material Fig. S2b). For 31 pg of template DNA with 37 PCR cycles, allele losses were observed at RPS4Y<sub>711</sub> and M119 markers in one of five (20%) and at M214, M175, M7, M159, and M217 markers in two of five (40%). For 15 pg of template DNA, most markers showed allele losses except for the RPS4Y<sub>711</sub> and M407 markers with 37 PCR cycles. The alleles of M89, M175 and M7 markers were lost most frequently (80%). Based on these results, these multiplexes were proven to be able to successfully type 1 ng to 62 pg, and Multiplex IV was the most sensitive SBE reaction (average loss; 4.3%), followed by Multiplex II (6.2%), Multiplex III (8.1%) and Multiplex I (10.0%).

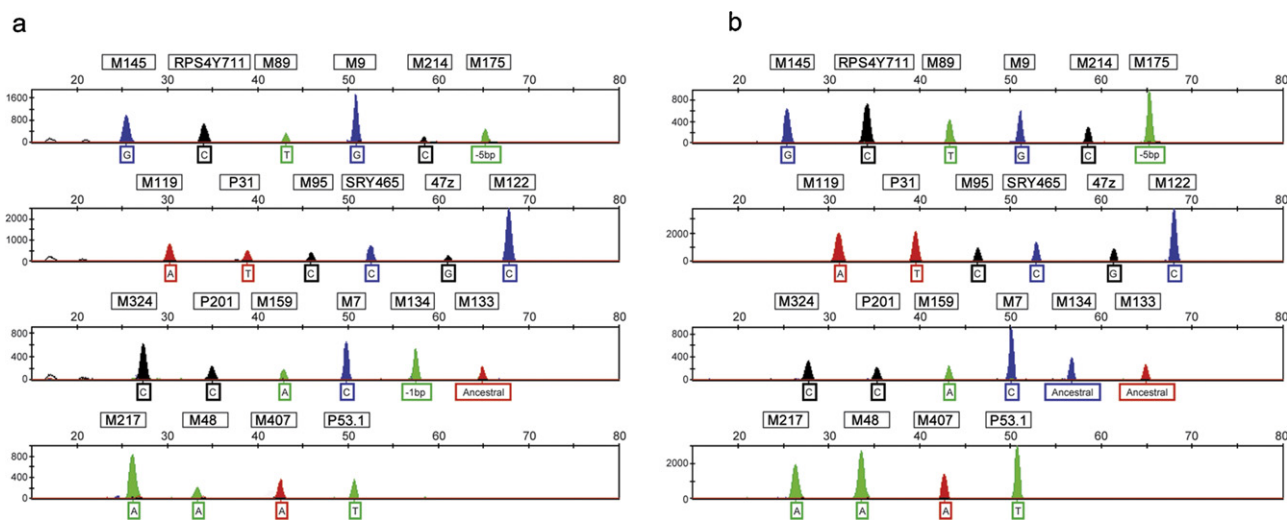


**Fig. 2.** Representative electropherograms of multiplex SBE reactions I (a), II (b), III (c), and IV (d) obtained from male donors belonging to each haplogroup observed in this study. Each labeled peak represents a SNP and the right-hand side of the each panel shows the relevant haplogroup.

To test the efficiency of the multiplex systems, the 22 SNPs were also typed in one artificially degraded DNA sample and 10 DNA samples extracted from 55-year-old skeletal remains. Representative profiles from the DNA samples are presented in Fig. 3. The results from the replicate analyses of the artificially degraded DNA sample and its intact DNA sample showed full allele concordance although the signal intensities for all SNP alleles were decreased by a percentage ranging from 66.7% to 93.6% (Fig. 4). For the artificially degraded DNA sample, peak intensities among the loci lost the balance due to poor amplifications of some markers such as RPS4Y<sub>711</sub>, M214, SRY<sub>465</sub>, and M48. The correct allelic determinations with no dropout or drop-in and small values of standard deviation from average peak height between independent replicates demonstrate that the multiplexes are reliable and reproducible. The 10 DNA samples extracted from 55-year-old skeletal remains were also successfully amplified and could be assigned to the relevant haplogroups using the multiplexes (Fig. 3b and Table 1). For

amplification of the DNA samples from the skeletal remains, a higher PCR cycle number was used to enhance amplification efficiency. By increasing the PCR cycle number, however, unwanted phenomena such as allele dropout, heterozygous peak imbalance, increase in stutter peaks, or pull-up peaks have been reported in STR-PCR assay [18]. In addition, the reduction of PCR target sizes in these developed multiplexes also increases the chances of amplifying contaminating DNA, especially in highly degraded DNA samples such as old skeletal remains DNA. However, Y-SNP fortunately has no heterozygous allele and stutter peak associated with each allele so that the unwanted effect of increased PCR cycle number is limited to the SNP analysis. No allele dropout was observed in the 10 DNA samples from the skeletal remains using 35 PCR cycles and as little as 62 pg of template DNA. Allele drop-in was observed in some markers such as RPS4Y<sub>711</sub>, M214, P31, and SRY<sub>465</sub>, but allele drop-in interfering with determination of relevant haplogroups was not observed in the consensus profiles obtained by





**Fig. 3.** Representative electropherograms of artificially degraded DNA (a) and a DNA sample obtained from 55-year-old skeletal remains (b), which belong to haplogroups O3a3c and O3a3\*, respectively. The each panel from top to bottom shows electropherogram of multiplex SBE reactions I–IV.

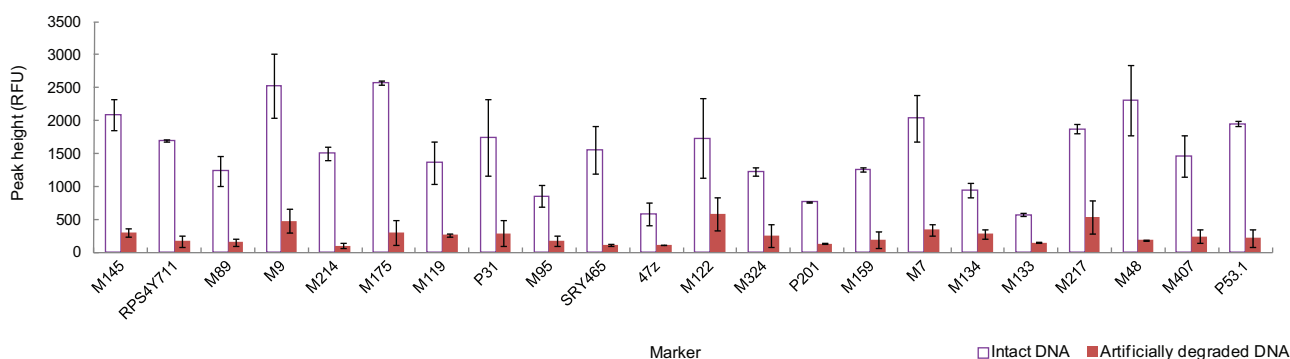
duplicate amplification. Although a single incident of allele drop-in was detected in the negative controls and reagent blanks, they were not reproduced upon re-amplification. In addition, when comparing the genotyping success of Y-SNP markers with those of STR markers from a previous study [16] in the DNA samples, the Y-SNP markers provided higher success than STR markers, especially in lower quality DNA samples (Table 1). This result suggests that the multiplexes may be very useful for analyzing degraded forensic and ancient samples by applying the consensus profiles obtained from replication of analysis containing negative control and reagent blank to illustrate that there has been a source of contamination in the analytical method [18,19].

Collectively, the newly developed multiplexes described in the present study are well suited for Y-SNP analysis of total DNA amounts as low as 62 pg, and of degraded DNA with fragment sizes as small as 100 bp. The characteristics of Y-SNP analysis, which include a lack of heterozygous alleles or stutter artifacts, should allow the interpretation of Y-SNP profiles to be much simpler than that of autosomal SNP or STR analyses, especially when DNA is scarce and substantially fragmented. Thus, our multiplex systems are expected to assist researchers to work with more ease and efficiency when dealing with degraded DNA [24]. However, it is important to be attentive to multiple peaks in the case of interpretation of mixed samples or drop-ins due to DNA contaminants.

### 3.3. Y chromosome haplogrouping of Koreans

DNA samples from 300 Korean males were analyzed using the four multiplex systems. A total of 16 different Y-SNP haplogroups were identified; haplogroup O2b\* was most frequently observed (29.3%), followed by haplogroups C3 (x3C, C3d, C3e) (16.0%) and O3a3c1 (11.0%) (Table 2). The haplogroup information with 17 Y-STR haplotype data has been contributed to Y-Chromosome Haplotype Reference Database (<http://www.yhrd.org>, accession number; YA003728). The haplogroup diversity was 0.8505, and the discrimination power was 5.33%. The distributions of the major haplogroups did not differ from previous haplogroup frequencies reported for Koreans [14,25–27] ( $p > 0.01$ ). Haplogroup O, the most prevalent haplogroup (78.3%) in Koreans, was subdivided into several subhaplogroups by Multiplex II, and the subhaplogroup O3, which comprises 36.3% of Koreans, was further subdivided into internal derivatives by Multiplex III. The haplogroup nomenclature was designated according to the revised Y chromosome tree, and does not include the LINE 1 mutation from the list of markers defining subclades of haplogroup O3 [8].

Multiplexes II and III allow the detection of 12 sublineages of haplogroup O, but only 10 were found in Koreans. The frequencies of subhaplogroups O\*, O1a, O2\*, O2b\*, O2b1, and O3 in Koreans were not significantly different from those reported by Xue et al. [14] and Kim et al. [26,27], but some differences were observed between the present study and Hong et al. [25] that may be



**Fig. 4.** Peak height comparison between intact and artificially degraded DNA at each Y-SNP locus. Intact and artificially degraded DNA was amplified in replicates and values are expressed as the average peak height at each locus. The error bar indicates the standard deviation from the average.

**Table 1**

Y haplogroup affiliations, success rates of STR typing and DNA concentrations of 10 old skeletal remains.

Sample	Y haplogroup	Success of STR typing <sup>a</sup> (%)	Concentration <sup>a</sup> (pg/μl)
1	NO	93.3	114.8 ± 10.20
2	O2*	100	205.7 ± 2.75
3	O2b*	66.7	55.9 ± 5.94
4	O2b*	100	106.5 ± 3.56
5	O3a3*	100	766.1 ± 39.03
6	O3a3c	80.0	149.8 ± 11.30
7	O1a	33.3	27.8 ± 0.13
8	O2b*	100	169.9 ± 10.96
9	O3a3*	100	84.9 ± 14.71
10	O2b*	100	275.2 ± 57.04

<sup>a</sup> Success of STR typing using the AmpFISTR<sup>®</sup> Identifier<sup>®</sup> PCR Amplification kit and DNA concentrations reported in a previous study [16].

explained by differences in sample size or population stratification. Among the derivatives of haplogroup O2, subhaplogroup O2b\* and its derived subhaplogroup O2b1 are known to be concentrated in Korean and Japanese populations with opposite frequencies (33.3% and 4.0% in Korea vs. 7.7% and 22.0% in Japan) [28], and the results of our current study (29.3% and 8.3%) are consistent those findings. Haplogroup O3 and its subhaplogroups are found in most modern East Asian populations, and their distributions in East Asia implicate a migratory route of modern humans from south to north. Among the derivatives of haplogroup O3, subhaplogroup O3a3c occurs most frequently in East Asians [13,14], and its frequency was 18.7% in Koreans in the present study. Subhaplogroup O3a3c1, a subclade of O3a3c, accounted for 58.9% of the O3a3c Korean samples, and its frequency and proportion in subhaplogroup O3a3c are known to vary depending on the population investigated [13,14]. Sino-Tibetan populations are characterized by frequent occurrence of subhaplogroups O3a3c and O3a3c1 [29], while other East Asians and Southeast Asians show a moderate distribution of these haplogroups [13,14]. On the other hand, haplogroup O2a was not found in our sample, and the O3a3b haplogroup was observed only once, which is consistent with the fact that haplogroups O2a and O3a3b are rare and restricted to the southern populations of East Asia.

In addition, haplogroup C could be further subdivided into four sublineages (C3, C3c, C3d, and C3e) by Multiplex IV. Most of the

**Table 2**

Y-chromosomal haplogroup distribution among 300 Korean males.

Haplogroup	Terminal marker	No. of sample	Frequency
DE	M145	3	0.0100
C (xC3)	RPS4Y <sub>711</sub>	2	0.0067
C3 (xC3c, C3d, C3e)	M217	48	0.1600
C3c	M48	0	0.0000
C3d	M407	2	0.0067
C3e	P53.1	0	0.0000
F (xK)	M89	0	0.0000
K (xNO, O)	M9	5	0.0167
NO	M214	5	0.0167
O*	M175	0	0.0000
O1a	M119	8	0.0267
O2*	P31	5	0.0167
O2a	M95	0	0.0000
O2b*	SRY <sub>465</sub>	88	0.2933
O2b1	47z	25	0.0833
O3*	M122	2	0.0067
O3a (xO3a3)	M324	27	0.0900
O3a3*	P201	23	0.0767
O3a3a	M159	0	0.0000
O3a3b	M7	1	0.0033
O3a3c (xO3a3c1)	M134	23	0.0767
O3a3c1	M133	33	0.1100

Koreans belonged to subhaplogroup C3 (xC3c, C3d, C3e) and the distribution was not different from previous reports [14,27]. Haplogroup C3 lineages are known to be informative in revealing the pattern of prehistoric migrations of regional population because of its extensive distribution in East Asia [15]. Therefore, the observed frequencies and distributions of Korean haplogroups may provide information that is useful for inferring the history of Koreans [30] as well as for forensic applications.

Reportedly, the majority of East Asians (55.9%), including Koreans, belong to haplogroup O, followed by haplogroups C (19.9%), NO (xO) (7.3%), and DE (5.1%) [14]. Since our newly developed multiplexes contain the 22 SNPs that define haplogroup O and its internal derivatives as well as world-wide major haplogroups, these multiplex sets are applicable to most East Asian populations for fine haplogrouping according to the revised topology of Y chromosome haplogroups [8].

#### 4. Conclusion

Four multiplex systems for scoring 22 Y-SNPs were developed to identify the most frequent East Asian Y-chromosomal haplogroups, and subsequent validation tests showed that the multiplexes are sensitive and efficient for analyzing low template and highly degraded DNA. The newly developed multiplexes allowed not only subdivision of haplogroup C, but also subdivision of haplogroup O and its derived subhaplogroup O3 according to the revised Y chromosome tree, thereby demonstrating their suitability for analyzing East Asian lineages, which are characterized by high occurrence of haplogroups O and C. These multiplex sets will be useful tools for Y-chromosomal haplogroup determination in anthropological and forensic studies of East Asian populations.

#### Acknowledgements

The biospecimens for this study were provided by National Biobank of Korea, Korea Centers for Disease Control and Prevention, supported by the Korean Ministry of Health and Welfare. This study was supported by a faculty research grant from Yonsei University College of Medicine in Seoul, Korea for 2008 (6-2008-0266).

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.fsigen.2012.06.014>.

#### References

- [1] M.A. Jobling, A. Pandya, C. Tyler-Smith, The Y chromosome in forensic analysis and paternity testing, *Int. J. Legal Med.* 110 (1997) 118–124.
- [2] M.A. Jobling, C. Tyler-Smith, New uses for new haplotypes the human Y chromosome, disease and selection, *Trends Genet.* 16 (2000) 356–362.
- [3] M.A. Jobling, In the name of the father: surnames and genetics, *Trends Genet.* 17 (2001) 353–357.
- [4] P.A. Underhill, P. Shen, A.A. Lin, L. Jin, G. Passarino, W.H. Yang, E. Kauffman, B. Bonn -Tamir, J. Bertranpetit, P. Francalacci, M. Ibrahim, T. Jenkins, J.R. Kidd, S.Q. Mehdi, M.T. Seielstad, R.S. Wells, A. Piazza, R.W. Davis, M.W. Feldman, L.L. Cavalli-Sforza, P.J. Oefner, Y chromosome sequence variation and the history of human populations, *Nat. Genet.* 26 (2000) 358–361.
- [5] M.F. Hammer, T.M. Karafet, A.J. Redd, H. Jarjanazi, S. Santachiara-Benerecetti, H. Soodyall, S.L. Zegura, Hierarchical patterns of global human Y-chromosome diversity, *Mol. Biol. Evol.* 18 (2001) 1189–1203.
- [6] M.A. Jobling, C. Tyler-Smith, The human Y chromosome: an evolutionary marker comes of age, *Nat. Rev. Genet.* 4 (2003) 598–612.
- [7] P.A. Underhill, G. Passarino, A.A. Lin, P. Shen, M. Miraz n Lahr, R.A. Foley, P.J. Oefner, L.L. Cavalli-Sforza, The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations, *Ann. Hum. Genet.* 65 (2001) 43–62.
- [8] T.M. Karafet, F.L. Mendez, M.B. Meilerman, P.A. Underhill, S.L. Zegura, M.F. Hammer, New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree, *Genome Res.* 18 (2008) 830–838.

- [9] R. Lessig, M. Zoledziewska, K. Fahr, J. Edelmann, M. Kostrzewa, T. Dobosz, W.J. Kleemann, Y-SNP-genotyping – a new approach in forensic analysis, *Forensic Sci. Int.* 154 (2005) 128–136.
- [10] M. Brión, J.J. Sanchez, K. Balogh, C. Thacker, A. Blanco-Verea, C. Børsting, B. Stradmann-Bellinghausen, M. Bogus, D. Syndercombe-Court, P.M. Schneider, A. Carracedo, N. Morling, Introduction of a single nucleotide polymorphism-based “Major Y-chromosome haplogroup typing kit” suitable for predicting the geographical origin of male lineages, *Electrophoresis* 26 (2005) 4411–4420.
- [11] V. Onofri, F. Alessandrini, C. Turchi, M. Pesaresi, L. Buscemi, A. Tagliabracci, Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs, *Forensic Sci. Int.* 157 (2006) 23–35.
- [12] G. Berniell-Lee, K. Sandoval, I. Mendizabal, E. Bosch, D. Comas, SNPlexing the human Y-chromosome: a single-assay system for major haplogroup screening, *Electrophoresis* 28 (2007) 3201–3206.
- [13] H. Shi, Y.L. Dong, B. Wen, C.J. Xiao, P.A. Underhill, P.D. Shen, R. Chakraborty, L. Jin, B. Su, Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122, *Am. J. Hum. Genet.* 77 (2005) 408–419.
- [14] Y. Xue, T. Zerjal, W. Bao, S. Zhu, Q. Shu, J. Xu, R. Du, S. Fu, P. Li, M.E. Hurles, H. Yang, C. Tyler-Smith, Male demography in East Asia: a north–south contrast in human population expansion times, *Genetics* 172 (2006) 2431–2439.
- [15] H. Zhong, H. Shi, X.B. Qi, C.J. Xiao, L. Jin, R.Z. Ma, B. Su, Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia, *J. Hum. Genet.* 55 (2010) 428–435.
- [16] H.Y. Lee, M.J. Park, N.Y. Kim, J.E. Sim, W.I. Yang, K.J. Shin, Simple and highly effective DNA extraction methods from old skeletal remains using silica columns, *Forensic Sci. Int. Genet.* 4 (2010) 275–280.
- [17] I. Nasidze, D. Quinque, I. Dupanloup, R. Cordaux, L. Kokshunova, M. Stoneking, Genetic evidence for the Mongolian ancestry of Kalmyks, *Am. J. Phys. Anthropol.* 128 (2005) 846–854.
- [18] P. Gill, J. Whitaker, C. Flaxman, N. Brown, J. Buckleton, An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA, *Forensic Sci. Int.* 112 (2000) 17–40.
- [19] S. Cowen, P. Debenham, A. Dixon, S. Kutranov, J. Thomson, K. Way, An investigation of the robustness of the consensus method of interpreting low-template DNA profiles, *Forensic Sci. Int. Genet.* 5 (2011) 400–406.
- [20] P. Wiegand, M. Kleiber, Less is more – length reduction of STR amplicons using redesigned primers, *Int. J. Legal Med.* 114 (2001) 285–287.
- [21] J.M. Butler, Y. Shen, B.R. McCord, The development of reduced size STR amplicons as tools for analysis of degraded DNA, *J. Forensic Sci.* 48 (2003) 1054–1064.
- [22] P. Kainz, The PCR plateau phase – towards an understanding of its limitations, *Biochim. Biophys. Acta* 1494 (2000) 23–27.
- [23] P.S. Walsh, H.A. Erlich, R. Higuchi, Preferential PCR amplification of alleles: mechanisms and solutions, *PCR Methods Appl.* 1 (1992) 241–250.
- [24] C. Bouakaze, C. Keyser, S. Amory, E. Crubézy, B. Ludes, First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA, *Int. J. Legal Med.* 121 (2007) 493–499.
- [25] S.B. Hong, H.J. Jin, K.D. Kwak, W. Kim, Y-chromosome haplogroup O3-M122 variation in East Asia and its implications for the peopling of Korea, *Korean J. Genet.* 28 (2006) 1–8.
- [26] S.H. Kim, M.S. Han, W. Kim, W. Kim, Y chromosome homogeneity in the Korean population, *Int. J. Legal Med.* 124 (2010) 653–657.
- [27] S.H. Kim, K.C. Kim, D.J. Shin, H.J. Jin, K.D. Kwak, M.S. Han, J.M. Song, W. Kim, W. Kim, High frequencies of Y-chromosome haplogroup O2b-SRY465 lineages in Korea: a genetic perspective on the peopling of Korea, *Invest. Genet.* 2 (2011) 10.
- [28] M.F. Hammer, T.M. Karafet, H. Park, K. Omoto, S. Harihara, M. Stoneking, S. Horai, Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes, *J. Hum. Genet.* 51 (2006) 47–58.
- [29] T. Gayden, S. Chennakrishnaiah, J. La Salvia, S. Jimenez, M. Regueiro, T. Maloney, P.J. Persad, A. Bukhari, A. Perez, O. Stojkovic, R.J. Herrera, Y-STR diversity in the Himalayas, *Int. J. Legal Med.* 125 (2011) 367–375.
- [30] H.J. Jin, K.D. Kwak, M.F. Hammer, Y. Nakahori, T. Shinka, J.W. Lee, F. Jin, X. Jia, C. Tyler-Smith, W. Kim, Y-chromosomal DNA haplogroups and their implications for the dual origins of the Koreans, *Hum. Genet.* 114 (2003) 27–35.