Int J Legal Med (2012) 126:589–599
DOI 10.1007/s00414-012-0703-9

ORIGINAL ARTICLE

Understanding the Y chromosome variation in Korea—relevance of combined haplogroup and haplotype analyses

Myung Jin Park · Hwan Young Lee · Woo Ick Yang · Kyoung-Jin Shin

Received: 11 January 2012 / Accepted: 26 April 2012 / Published online: 9 May 2012

Abstract We performed a molecular characterization of Korean Y-chromosomal haplogroups using a combination of Y-chromosomal single nucleotide polymorphisms (Y-SNPs) and Y-chromosomal short tandem repeats (Y-STRs). In a test using DNA samples from 706 Korean males, a total of 19 different haplogroups were identified by 26 Y-SNPs including the newly redefined markers (PK4, KL2, and P164) in haplogroup O. When genotyping the SNPs, phylogenetic nonequivalence was found between SNPs M117 and M133, which define haplogroup O3a3c1 (O3a2c1a according to the updated tree of haplogroup O by Yan et al. (European Journal of Human Genetics 19:1013–1015, 2011)), suggesting that the position of the M133 marker should be corrected. We have shown that the haplotypes consisted of DYS392, DYS393, DYS437, DYS438, DYS448, and DYS388 loci, which exhibit a relatively lower mutation rate, can preserve phylogenetic information and hence can be used to roughly distinguish Y-chromosome haplogroups, whereas more rapidly mutating Y-STRs such as DYS449 and DYS458 are useful for differentiating male lineages. However, at the relatively rapidly mutating DYS447, DYS449, DYS458, and DYS464 loci, unusually short alleles and intermediate alleles with common sequence structures are informative for elucidating the substructure within the context of a particular haplogroup. In addition, some deletion mutations in the DYS385 flanking region and the null allele at DYS448 were associated with a single haplogroup background. These high-resolution haplogroup and haplotype data will improve our understanding of regional Y-chromosome variation or recent migration routes and will also help to infer haplogroup background or common ancestry.

Keywords Y chromosome · Haplogroup · Single nucleotide polymorphism · Short tandem repeat · Atypical allele · Korean

Introduction

The paternally inherited non-recombining portion of the Y chromosome (NRY) is targeted in forensic and medical genetics, human evolutionary studies, and genealogical reconstruction [1–6]. Two classes of markers on the NRY, single nucleotide polymorphism (SNP) and short tandem repeat (STR), are widely used. Due to their low mutation rate, Y-chromosome SNPs are useful genetic markers for reconstructing male lineages through hierarchically arranged allelic sets known as haplogroups, whereas the high mutation rates of Y-chromosomal STRs (Y-STRs) make them valuable for differentiating between unrelated males and inferring affinities among related populations [4–8].

Y-chromosomal haplogroups are distributed non-randomly in the population and geographical regions, allowing us to infer the ethnic or geographical origin of unknown samples [5, 8, 9]. It has been noted that Y-STR variability is highly partitioned by differences among haplogroups, suggesting the possibility of haplogroup prediction based on Y-STR

Electronic supplementary material The online version of this article (doi:10.1007/s00414-012-0703-9) contains supplementary material, which is available to authorized users.

M. J. Park · H. Y. Lee · W. I. Yang · K.-J. Shin
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

H. Y. Lee · K.-J. Shin (✉)
Human Identification Research Center, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-752, South Korea
e-mail: kjshin@yuhs.ac

haplotype information [10–12]. In addition, studies of the relationship between Y-chromosomal haplogroups and Y-STR variants, such as relationships of haplogroup affiliations for partial deletion/insertion mutations or intermediate alleles in Y-STRs, have been reported [13–16]. Thereby, the Y-chromosome haplotype reference database (YHRD) is expanding the amount of information for Y-chromosomal single nucleotide polymorphism (Y-SNP) and Y-STR haplotypes [17], and the number of studies integrating Y-SNP and Y-STR data is growing [16, 18–20].

In our previous studies, we performed molecular characterizations of 22 Y-STR markers and their haplotypes (DYS19, DYS385a/b, DYS388, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS446, DYS447, DYS448, DYS449, DYS456, DYS458, DYS464, DYS635, and GATA H4), which included atypical alleles, deletions in the DYS385 flanking region, and null alleles associated with DYS448 [24–27]. Analyzing the correlation between Y-STR haplotypes (or each Y-STR allele) and their haplogroup membership in a Korean population will be helpful for understanding the substructure of Korean haplogroups.

Meanwhile, the Y-chromosomal haplogroup tree has been updated. In particular, the phylogeny of haplogroup O-M175 has been recently revised to include the phylogenetic positions of redefined markers, L127, KL1, KL2, P164, and PK4 [21, 22]. A previous study showed that L127/KL1/KL2 and P164 are highly informative for separating substantial samples belonging to haplogroup O3a-M324 in the Han Chinese population [22]. The haplogroup O-M175 is one of the major clades in the Korean population, and the haplogroup O3a-M324 accounts for 43.9 % of Korean males [20, 23]. It is necessary to confirm the relative positions of the redefined markers and to determine the distribution of subhaplogroups in the Korean population. Therefore, to provide distributions of Korean haplogroup, we developed multiplex allele-specific PCR assays for the simultaneous detection of Y-SNP genotypes in a hierarchical order and to classify the Korean haplotypes into their corresponding haplogroups, especially with in-depth resolution for haplogroup O-M175 according to the revised phylogenetic tree. Then, we evaluated the haplogroup affiliation for each usual and variant allele to elucidate their relationship with the binary haplogroup and to subsequently identify Y-STR alleles potentially representing a substructure within the haplogroup tree.

Materials and methods

DNA samples

Our study protocol was approved by the Institutional Review Board of Severance Hospital, Yonsei University in Seoul, Korea. We analyzed DNA samples from 706 unrelated Korean males who had been previously typed for 22 Y-STRs (DYS19, DYS385a/b, DYS388, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS446, DYS447, DYS448, DYS449, DYS456, DYS458, DYS464, DYS635, and GATA H4) [24–27].

Analysis of Y-chromosome single nucleotide polymorphism

Two multiplex allele-specific PCR assays were developed to identify alleles of M7, M9, M95, M117, M119, M134, M174, M175, M122, M231, P31, P201, JST002611, $RPS4Y_{711}$, SRY_{465}, and 47z (multiplex I), and M134, M324, P164, P201, JST002611, and KL2 (multiplex II). Some allele-specific primers were designed to have a tail at the 5′ end, thereby allowing different alleles to produce amplicons of different sizes. The performance of the two multiplex allele-specific primer sets was assessed by analyzing 300 DNA samples with known haplogroups using a single-base extension (SBE) reaction. Multiplex PCR reactions for all markers except the 47z marker were performed in a final volume of 10 μl containing 1 ng of template DNA, 1.0 μl of Gold ST*R 10× buffer (Promega, Madison, WI, USA), and 2.5 U of AmpliTaq Gold® DNA polymerase (Applied Biosystems, Foster City, CA, USA) for multiplex I or 1.5 U for multiplex II, and appropriate concentrations of primers (Table S1). Because of the difficulty in simultaneous amplification of the 47z marker with other markers in the multiplex I, monoplex allele-specific PCR was performed for the 47z marker with the same PCR conditions as above except that 0.5 U of AmpliTaq Gold® DNA polymerase (Applied Biosystems) was used. Thermal cycling was done on a Veriti 96-Well Thermal Cycler (Applied Biosystems) with the following conditions: 95°C for 11 min; 30 cycles of 94°C for 20 s, 59°C for 1 min, and 72°C for 30 s; and a final extension of 60°C for 45 min. The PCR product of the 47z marker was mixed with an equal amount of product from the multiplex I before being separated by capillary electrophoresis using an ABI PRISM 310 Genetic Analyzer (Applied Biosystems) after mixing with GeneScan™ 500 LIZ® size standard (Applied Biosystems). Automated allele calling was performed using GeneMapper® ID software 3.2 (Applied Biosystems). An allelic ladder containing both ancestral and derived alleles for all markers was used to perform allelic designation in the multiplex allele-specific PCR assays.

Monoplex PCR followed by SBE reaction was also used to analyze M110, M133, M159, M207, M242, M267, P203, and PK4 markers. The monoplex PCR reactions were performed in a final volume of 25 μl which contained 1 ng of template DNA, 2.5 μl of Gold ST*R 10× buffer (Promega), 2.0 U of AmpliTaq Gold® DNA polymerase (Applied

Biosystems), and appropriate concentrations of primers (Table S2). Thermal cycling was done with 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 30 s; and a final extension of 72°C for 7 min. Before the 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 inactivated at 80°C for 15 min. The SBE reaction was carried out with a SNaPshot™ Multiplex kit (Applied Biosystems), the purified PCR product, and SBE primer mix (Table S3) according to the manufacturer's instructions. Thermal cycling conditions were as follows: 25 cycles at 96°C for 10 s, 50°C for 5 s, and 60°C for 30 s. After the SBE reaction, 1 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 incubating at 80°C for 15 min. The purified SBE products were separated by capillary electrophoresis using an ABI PRISM 310 Genetic Analyzer (Applied Biosystems) after mixing with GeneScan™ 120 LIZ® size standard (Applied Biosystems). Automated allele calling was made using GeneMapper® ID software 3.2 (Applied Biosystems).

Statistical analysis

The haplogroup frequencies were determined by direct counting. Standard diversity parameters including gene/haplotype diversity and analysis of molecular variance (AMOVA) were calculated with the software package Arlequin 3.5.1.3. [28]. AMOVA was performed for STR allele frequencies among haplogroups to test simple hierarchical partitioning of haplotypes in the haplogroup. A median-joining (MJ) network was constructed for the Y-STR haplotypes within specific haplogroups by the program NETWORK 4.6.0.0 (http://www.fluxus-engineering.com). Y-STR weighting was applied in accordance with Gomes et al. [29].

Results

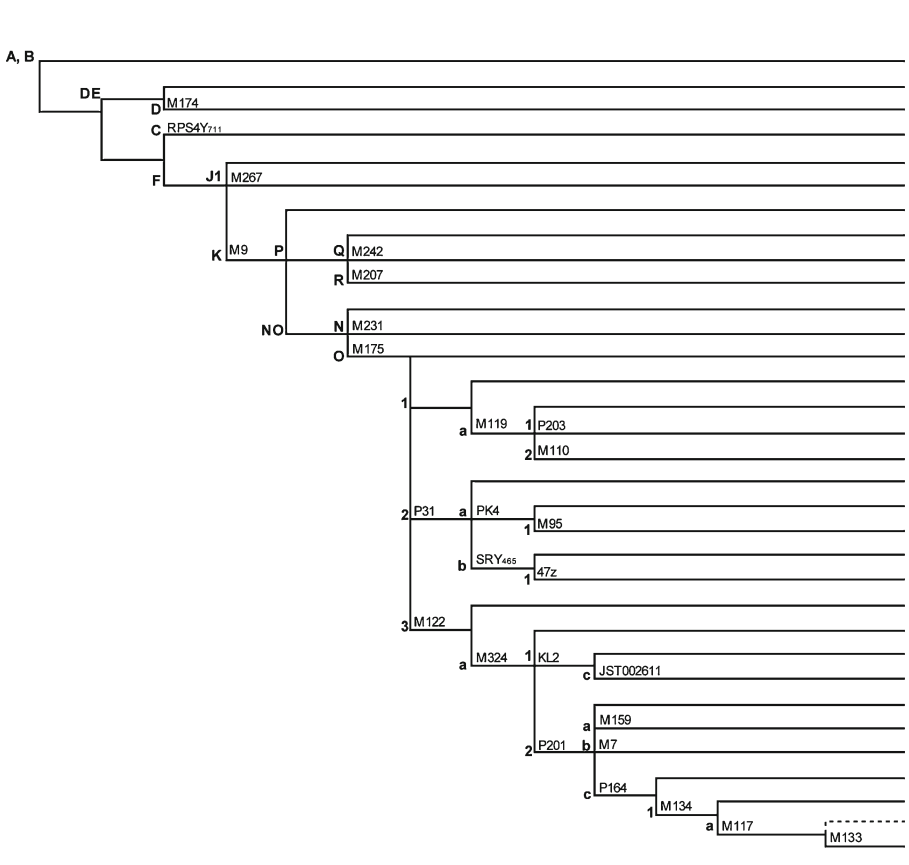
Y-chromosome haplogroups of Koreans

Twenty-six Y-SNP markers were selected to explore the most frequent Korean haplogroups of the Y chromosome based on the newly revised Y chromosome tree topology [21, 22]. In particular, newly defined or relocated markers (JST002611, KL2, P164, and PK4) were selected according to the updated tree of haplogroup O [22].

Therefore, a set of 16 Y-SNPs (M7, M9, M95, M117, M119, M134, M174, M175, M122, M231, P31, P201, JST002611, RPS4Y$_{711}$, SRY$_{465}$, and 47z) was initially analyzed in all samples using multiplex I to determine the haplogroups frequent in East Asians. According to the results obtained from multiplex I, the samples belonging to haplogroup O3-M122 were then typed using multiplex II to further divide the subhaplogroups O3 according to the revised tree of haplogroup O [22]. Three markers (M134, P201, and JST002611) are common to multiplexes I and II, which enabled to check for sample switch and to confirm the typing results. Representative electropherograms of the two developed multiplex allele-specific PCR assays are shown in Figs. S1 and S2. In addition, to detect rare haplogroups and to confirm the position of the newly updated markers, subsequent typing of the remaining samples was performed hierarchically using a monoplex SBE reaction according to the haplogroup designated results using multiplex II. Representative electropherograms of the monoplex SBE reactions are shown in Fig. S3. Naming for haplogroups primarily followed the nomenclature proposed by Karafet et al. [21] according to the criteria for publication of population data in *International Journal of Legal Medicine* [30], but the revised names of the O subhaplogroups according to the nomenclature proposed by Yan et al. [22] were also indicated in parenthesis. A total of 19 different haplogroups were identified (when not including new lineage with M117+, M133−); haplogroup O2b*-SRY$_{465}$ was most frequently observed (22.0 %), followed by haplogroups C-RPS4Y$_{711}$ (12.9 %) and O3a3c1 (O3a2c1a)-M117 (12.8 %) (Fig. 1). The haplogroup diversity was 0.8830, and the discriminatory capacity was 2.69 %.

By typing of the newly defined or relocated SNPs KL2, JST002611, P164, and PK4, their phylogenetic positions were confirmed using Korean haplogroup O-M175 samples. The KL2 mutation was found in all samples belonging to O3a-M324 (xP201). Moreover, the chromosomes with the KL2 mutation were divided into two groups based on the presence or absence of the JST002611 mutation. On the other hand, none of the samples shared mutations KL2 and P201, so their relative position could be determined. In addition, the mutation P164 was observed in all samples with the M134 mutation but was not found in the samples with the M7 mutation. We also found that some samples belonging to O3a3 (O3a2)-P201 (xM159, M7, M134) had the P164 mutation, thereby corresponding to paragroup (O3a2c*)-P164. Therefore, the positions of SNPs KL2, JST002611, and P164 could be confirmed in our surveyed samples. However, the relative position of the PK4 marker could not be determined because of a lack of positive samples.

The addition of KL2 and P164 markers to the phylogeny of haplogroup O improved the resolution of the Korean haplogroup O3a-M324, which had the same result as the Han Chinese population [22]. However, the distribution of the subhaplogroups differed significantly from those in the east, north, and south Han Chinese populations [22]



Haplogroup		N	Freq
Karafet et al. (2008)	Yan et al. (2011)		
A,B			
DE (xD)			
D		18	0.0255
C		91	0.1289
G,H,I			
J1		1	0.0014
K (xP, NO)			
P			
Q		13	0.0184
R		1	0.0014
NO*			
N		27	0.0382
O*	O*		
O1*	O1*		
O1a*	O1a*	3	0.0042
O1a1	O1a1	19	0.0269
O1a2	O1a2		
O2*	O2*	10	0.0142
O2a1a	**O2a***		
O2a*	**O2a1**		
O2b*	O2b*	155	0.2195
O2b1	O2b1	71	0.1006
O3*	O3*	2	0.0028
O3a*	O3a*		
-	**O3a1 (xO3a1c)**	11	0.0156
O3a4	**O3a1c**	72	0.1020
O3a3*	**O3a2***	42	0.0595
O3a3a	**O3a2a**		
O3a3b	**O3a2b**	3	0.0042
O3a3b2	**O3a2c***	9	0.0127
O3a3c	**O3a2c1**	68	0.0963
-	-	8	0.0113
O3a3c1	**O3a2c1a**	82	0.1161

Fig. 1 Phylogenetic tree of the 26 Y-chromosomal binary polymorphisms analyzed in this study. The analyzed Y-SNPs are shown in each branch, and the corresponding haplogroups and observed frequencies are shown at the end of each branch according to Karafet et al. (*left*) and Yan et al. (*right*) [21, 22]. The lineages renamed by Yan et al. are indicated in *bold*, and a new lineage (M117+, M133−) found in the present study is indicated by *broken line*

($p<0.0001$). Specifically, there were significant differences in the distributions of sublineages inside haplogroups O1 and O2. While haplogroup O1a1-P203 was present in only 2.7 % of Koreans, it was found in 13.0 % of the Han Chinese population [22]. Haplogroup O2-P31, another derived sublineage of O-M175, was divided into subgroup with PK4 mutation and subgroup with SRY_{465} mutation. Haplogroup (O2a) defined by revised phylogenetic position of PK4 in the updated tree of haplogroup O was absent in the Korean population but is abundant in the south Han Chinese population. Meanwhile, O2b-SRY_{465} and its derived sublineage O2b1-47z were found frequently (37.7 %) in the Korean population, but they were nearly absent in the Han Chinese population. The subhaplogroups were concentrated in Korean and Japanese populations [23], and the results of our current study are consistent with those findings.

Non-equivalence between M117 and M133 markers

Unexpectedly, we found that the M117 marker (allele-specific PCR) was not always observed phylogenetically equivalent to the M133 (SBE reaction) with the multiplex allele-specific PCR compared to SBE reaction (data not shown). All but eight of the analyzed samples with the M117 mutation had the M133 mutation (M117+, M133+), whereas the eight samples were detected without the M133 mutation (M117+, M133−). Sequence structures of the eight samples were also confirmed by direct sequencing analysis (data not shown).

In order to evaluate the relationships of these samples (M117+, M133−) with those of the other samples (M117+/M133+ and M117−/M133−) within haplogroup O3a3c (O3a2c1)-M134, a network was constructed using Y-STR haplotypes (Fig. 2). Haplotypes produced two major clusters, clearly separated by the presence or absence of both M117 and M133 mutations. A cluster formed by the haplotypes (M117+, M133−) was distinct from the major cluster formed by the haplotypes (M117+, M133+), which indicated non-equivalence of M133 to M117.

Y-STR haplotypes and their relationship with haplogroups

A list of observed haplotypes and corresponding haplogroups has been submitted to the YHRD (YA003406

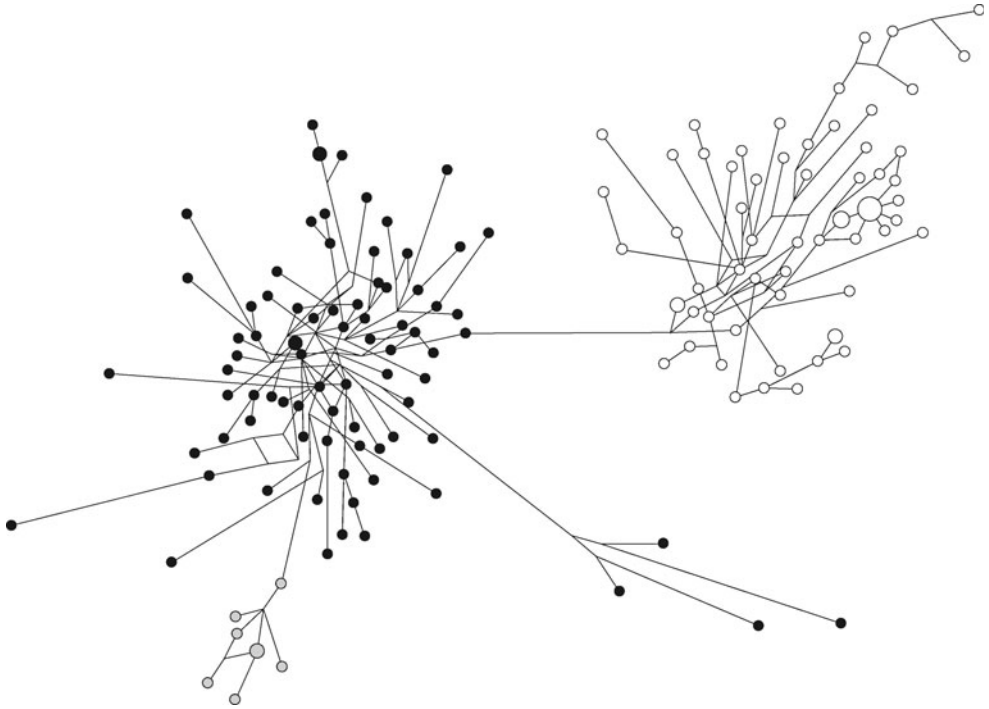


Fig. 2 Relationships between Y haplotypes within haplogroup O3a3c (O3a2c1)-M134. The median-joining network was based on information from 17 Y-STR loci (DYS19, DYS388, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS446, DYS447, DYS448, DYS449, DYS456, DYS458, DYS635, and GATA H4). The haplotypes carrying both M117 and M133 mutations are indicated by *filled circles*. The haplotype carrying neither M117 nor M133 mutations is indicated by an *open circle*. The haplotypes carrying the M117 mutation, but not M133 mutation, are indicated by *gray circles*

and YA003407) and was shown in Table S4. The Korean 22 Y-STRs haplotypes were classified into Y-chromosomal haplogroups. We did not observe any individuals sharing the same 22 Y-STRs haplotype among different haplogroups, even when 17 Y-STRs haplotype was applied. However, when considering minimal haplotypes, three cases of haplotype sharing were observed between haplogroup O2b-SRY_{465} and O2b1-47z and one between haplogroup O3a4 (O3a1c)-JST002611 and O3a3c1 (O3a2c1a)-M117. The apportionment of 22 Y-STRs haplotype variability among and within the 19 observed haplogroups was determined by AMOVA; only 38.75 % ($p<0.0001$) of the genetic variation was attributable to the difference among haplogroups in the Korean population. To evaluate which Y-STR contributes more to the difference among the haplogroups, a separate AMOVA analysis was performed based at Y-STR (Table 1), and DYS392, DYS393, DYS437, DYS438, DYS448, and DYS388 loci showed more than 60 % of their total genetic variation between haplogroups. In this context, the combined haplotypes consisting only of DYS392, DYS393, DYS437, DYS438, DYS448, and DYS388 loci were constructed and assessed by AMOVA analysis. The analysis revealed 70.93 % of the total variation among haplogroups for the combined haplotypes, with 38.75, 41.37, and 41.92 % of total variation among haplogroups for all 22 Y-STRs, 17 Y-STRs, and minimal haplotypes, respectively.

To identify the haplogroup affiliation of the combined haplotype with empirical data, the haplotypes observed frequently were determined for each Korean haplogroup, and we searched for the shared haplotypes, excluding DYS388, in the YHRD release 37 (June 16, 2011) [17] (Table 2). Using the six Y-STRs, most presented haplotypes were classified based on each haplogroup background, except for the non-differentiated haplotype (13-13-13-14-18-12) between haplogroup O2b*-SRY_{465} and O2b1-47z and a haplotype 14-13-10-14-18-12 between haplogroup O1a1-P203 and haplogroup N-M231. However, the addition of DYS390 and DYS389I to the six Y-STR haplotypes could differentiate common haplotypes between the O2b*-SRY_{465} (allele 23-24) and O2b1-47z (most allele 22) and between the N-M231 (allele 13-14) and O1a1-P203 (allele 12), respectively. In the YHRD database search, the matched haplotypes (except DYS388) tended to belong to the same haplogroups as those in Koreans, although some sublineages of haplogroup O3 were not further divided in the YHRD, in comparison to this study. These findings indicate that the six Y-STR loci with lower mutation rates seem to be more strongly structured in the haplogroup background.

Relationships between Y-STR variants and haplogroups

Certain haplogroup memberships also seem to be determined by the presence of atypical alleles (Table S5). The unusually short allele at DYS447, intermediate allele 30.2 at DYS449 and allele 14.3 at DYS464 were present in each single haplogroup background, haplogroups (O3a1)-KL2 (xJST002611), O1a1-P203, and N-M231, respectively. Unusually short alleles at DYS447, which are caused by partial deletion of sequences TAAAA$(TAAATA)_n$, only occurred in haplogroup (O3a1)-KL2 (xJST002611). At DYS449, the intermediate alleles caused by TC partial repeat insertion in

Table 1 Diversity, mutation rate, and AMOVA analysis for each Y-STR marker and haplotype

Y-STR	Gene/haplotype diversity	Mutation rate ($\times 10^{-3}$)[a]	% Variance	
			Among haplogroups	Within haplogroups
DYS19	0.7115	1.76	44.26	55.74
DYS385	0.9595	2.09	14.82	85.18
DYS389-I	0.6675	2.44	45.83	54.17
DYS389-II	0.7244	2.60	18.84	81.16
DYS390	0.6733	2.29	44.55	55.45
DYS391	0.2667	3.11	25.07	74.93
DYS392	0.6789	0.68	76.27	23.73
DYS393	0.6281	0.81	65.70	34.30
DYS437	0.4320	2.02	65.46	34.54
DYS438	0.6330	0.33	84.90	15.10
DYS439	0.6329	5.37	12.69	87.31
DYS448	0.7510	0.00	60.82	39.18
DYS456	0.5108	5.59	31.40	68.60
DYS458	0.7779	8.38	10.40	89.60
DYS635	0.6857	5.66	26.33	73.67
GATA H4	0.6115	3.43	38.34	61.66
DYS388	0.4850	0.00	71.84	28.16
DYS446	0.7886	2.71	31.59	68.41
DYS447	0.7520	5.41	42.39	57.61
DYS449	0.8523	18.97	14.50	85.50
DYS464	0.9668	3.99	16.53	83.47
6 Y-STRs[b]	0.9317	–	70.93	29.07
9 Y-STRs[c]	0.9966	–	41.92	58.08
17 Y-STRs[d]	0.9995	–	41.37	58.63
22 Y-STRs[e]	0.9999	–	38.75	61.25

[a]Mutation rate information was reported by Lee et al. [25]

[b]The six Y-STRs are a combination of DYS392, DYS393, DYS437, DYS438, DYS448, and DYS388 loci.

[c]The nine Y-STRs are a combination of DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, and DYS393

[d]The 17 Y-STRs are a combination of DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, and GATA H4

[e]The 22 Y-STRs are DYS19, DYS385a/b, DYS388, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS446, DYS447, DYS448, DYS449, DYS464, DYS456, DYS458, DYS635, and GATA H4

the first TTTC track, named 27.2 and 29.2, were all detected in haplogroup O2b1-47z. Another type of intermediate allele is a TT partial repeat insertion in the first TTTC and the second TTTC track, respectively. A sample with allele 28.2 from the first track insertion of TT belonged to haplogroup C-RPS4Y_{711}, and all samples with allele 30.2 by the second track insertion belonged to haplogroup O1a1-P203. Finally, intermediate alleles, named 12.3 and 14.3 at DYS464, occurred in haplogroups O3a3* (O3a2*)-P201 and N-M231, respectively.

Network analyses for Y-STR haplotypes carrying the observed atypical alleles were performed to evaluate the relationships among these haplotypes within a haplogroup context (Fig. 3). The haplotypes excluding the Y-STRs that displayed atypical alleles were used in the network analysis to eliminate bias from that Y-STR. The haplotypes possessing the short allele variant at DYS447 formed a distinct cluster in the haplogroup (O3a1)-KL2 (xJST002611) (Fig. 3a), suggesting that this variant likely defines a sublineage within the haplogroup (O3a1)-KL2 (xJST002611). For DYS449 variants, only the haplotypes with allele 30.2 were closely related to every other variant, indicating shared ancestry (Fig. 3b). Another network was constructed to approximate the relationship among the haplotypes with or without allele 30.2. The network was generated based on the analysis of Y-STR haplotypes belonging to haplogroup O1a1-P203 in the present study and a Japanese haplotype with allele 30.2 obtained from the Sorenson Molecular Genealogy Foundation by manual searching (Fig. 3c). The haplotypes carrying allele 30.2 formed a distinct cluster, indicating the possibility of defining a sublineage within haplogroup O1a1-P203. Unlike DYS447 and DYS449, the haplotypes with allele 14.3 at DYS464 did not form a distinct branch (Fig. 3d). This finding suggests that the variant allele 14.3 is partitioned across more than two sublineages within haplogroup N-M231.

Interestingly, an intermediate allele 17.2 at DYS458, which is known to be associated with haplogroup J1 [14], was found in our samples. The corresponding haplogroup J1 was confirmed by an additional SBE reaction (Fig. S3e).

Y-chromosomal haplogroups were determined for samples with deletion mutations in the DYS385 flanking region

Table 2 Representative haplotypes for each haplogroup and the number of matched haplotypes in the YHRD database

Haplotype[a] (392-393-438-437-448-388)	N[b]	%[b]	Haplogroup	YHRD database release 37																
				DE 1,251[c]	D 513[c]	E 718[c]	C 355[c]	F 5,547[c]	Q 188[c]	N 246[c]	O 1,603[c]	O1a 66[c]	O2 275[c]	O2b 216[c]	O2b1 85[c]	O3 795[c]	O3a 633[c]	O3a3 463[c]	O3a3c 182[c]	O3a3c1 65[c]	Total 7,165[c]
11-13-10-14-17-12	10	55.6	D-M174	77	77			1													78
11-13-10-14-19-12	2	11.1		214	206	5	3	2													219
11-13-11-14-17-12	2	11.1		26	26																26
11-15-10-14-21-12 (or 13)	23	25.3	C-RPS4Y$_{711}$				73														73
11-14-10-14-21-12 (or 13)	19	20.9				1	45														46
11-14-10-14-22-13	12	13.2					19														19
11-13-11-14-0-13[d]	4	4.4					6	2													8
14-14-12-14-19-12	7	53.8	Q-M207					12	11												12
15-14-12-14-19-12	3	23.1						2	2												2
14-13-10-14-20-13	5	18.5	N-M231					9		8	1										9
14-13-10-14-19-12 (or 13)	4	14.8						28		23	5										28
15-13-10-14-19-12 (or 13)	4	14.8						10		10											10
16-14-11-14-19-12[e]	3	11.1						7		7											7
14-13-10-14-18-12	3	11.1						37		5	32	24	1								37
14-13-10-14-18-12[f]	13	68.4	O1a1-P203					37		5	32	24	1								37
16-13-10-14-18-12[f]	2	10.5						2		1	1	1									2
13-13-10-14-18-12	3	30.0	O2*-P31					11			11	1	7								11
13-14-10-14-18-12	2	20.0						41			41		33								41
13-13-13-14-18-12 (or 13)	124	80.0	O2b*-SRY$_{465}$								467		151	151	61						467
13-13-13-14-18-12	46	64.8	O2b1-47z								467		151	151	61						467
13-12-10-15-21-*12*[g]	6	54.5	(O3a1)-KL2[h]					6			6					6	4	2			6
12-12-10-15-21-*12*	3	27.3						4			4					4	3				4
13-12-10-14-19-12	24	33.3	O3a4-JST002611					88			88					88	55	8			88
13-12-10-14-20-12	16	22.2						41			41					41	27	8			42
13-12-10-14-20-*13*	7	16.7	O3a3*-P201					41			41					41	27	8			42
13-12-10-15-20-*13*	25	59.5						78			78					78	50	34			78
13-12-10-15-20-*12*	3	100	O3a3b-M7					78			78					78	50	34			78
13-12-9-15-19-12	3	33.3	(O3a2c*)-P164[h]					1			1					1	1	1			1
13-13-10-15-19-12	2	22.2						23	4		19					19	15	13			23
12-12-10-15-19-12	36	52.9	O3a3c-M134					74			74					74	74	67	29		74

Table 2 (continued)

Haplotype[a] (392-393-438-437-448-388)	N[b]	%[b]	Haplogroup	YHRD database release 37															
				DE 1,251[c]	D 513[c]	E 718[c]	C 355[c]	F 5,547[c]	Q 188[c]	N 246[c]	O 1,603[c]	O1a 66[c]	O2 275[c]	O2b 216[c]	O2b1 85[c]	O3 795[c]	O3a 633[c]	O3a3 463[c]	O3a3c 182[c]	O3a3c1 65[c]	Total 7,165[c]
12-13-10-15-19-12	9	13.2						20			19					19	19	17	9		20
14-12-11-15-20-10	53	58.9	O3a3c1-M117					137			137					137	137	128	76	37	137

[a] This is a set of DYS392, DYS393, DYS438, DYS437, DYS448, and DYS388 loci. The searched haplotype in the YHRD database is indicated in bold and the same haplotypes belonging to different haplogroups are underlined. The numbers in parentheses indicated the alternative allele of the DYS388 marker in each haplotype. The DYS388 allele in italics indicates an allele only observed in the corresponding Korean haplogroups

[b] N and % represent the number of matched haplotypes and the proportion of haplotypes in the corresponding Korean haplogroups, respectively

[c] The number indicates the number of haplotype entries in the corresponding haplogroups from the searched YHRD database

[d] Three haplotypes of 11-13-11-14-0-13 carried the null allele of DYS448 caused by polymorphic 50f2/C deletion and one carried that caused by b1/b3 deletion

[e] Three haplotypes of 16-14-11-14-19-12, one of 16-14-11-14-20-12, and one of 14-14-10-14-19-12 carried allele 14.3 at DYS464

[f] Nine haplotypes of 14-13-10-14-18-12, one of 14-13-11-14-18-12, and two of 16-13-10-14-18-12 carried allele 30.2 at DYS449

[g] Six haplotypes of 13-12-10-15-21-12 carried an unusually short allele at DYS447

[h] Haplogroup affiliations referring to the nomenclature suggested by Yan et al. [22] were indicated in parenthesis

and with reported null alleles at DYS448 [27] (Tables S6 and S7). Some of the deletion events at DYS385 and DYS448 occurred on a single haplogroup background, similar to an atypical allele. Among the two kinds of mutations at DYS385, the deletion of GAGAAAAA in the upstream of core repeat unit $(GAAA)_n$ only occurred within haplogroup O3a3b (O3a2b)-M7. Network analysis of the haplotypes with the two deletion mutations revealed a distinct cluster formed by the haplotypes with an 8-bp deletion mutation (Fig. 4a). This suggests that the haplotypes with the 8-bp deletion have a common ancestry. Meanwhile, the null allele at DYS448 occurred in two different haplogroups, C-RPS4Y$_{711}$ and O3a3* (O3a2*)-P201. The null alleles are caused by a polymorphic 50f2/C deletion or b1/b3 deletion by rearrangement of the azoospermia factor c (*AZFc*) region in our samples [27].

To assess the relationship of the haplotypes carrying the deletion by rearrangement, a network was constructed using both the haplotypes from our study and the reported haplotypes with the DYS448 deletion mutation [31] (Fig. 4b). Three haplotypes with a polymorphic 50f2/C deletion and the reported haplotypes with similar deletion mutation (called non-b1/b3 class II) all belonged to haplogroup C-RPS4Y$_{711}$, although they formed two subgroups. Network results suggested the possibility of one common origin for this mutation. In contrast, haplotypes with the b1/b3 deletion mutation were wide spread in haplogroup C-RPS4Y$_{711}$, O3a3* (O3a2*)-P201 (in this study), C3c-M148, C*-RPS4Y$_{711}$, and O3a3c (O3a2c1)-M134, indicating an independent origin for the deletion event. These results were consistent with those from the report by Balaresque et al. [31].

Discussion

This study elucidated the genealogy of Korean Y-chromosomal haplogroups according to the revised phylogenetic tree, specifically reflecting the updated phylogeny of haplogroup O related to 22 Y-STRs haplotypes. In order to type the Y-chromosomal haplogroups, we used two different techniques, allele-specific PCR assays and SBE reactions. The allele-specific PCR assays were optimized for simultaneous detection of Y-SNPs followed by fragment analysis on an automatic DNA analyzer like general forensic STR typing method. So, they would be useful for simple and rapid identification of the haplogroups frequent in East Asian for large number of samples. Whereas, the SBE reactions were additionally used to further divide rare sub-haplogroups or to confirm relatively hierarchical positions because allele-specific PCR assays are often tedious and time-consuming to establish, requiring lengthy optimization procedures. After all, most East Asian haplogroups could be

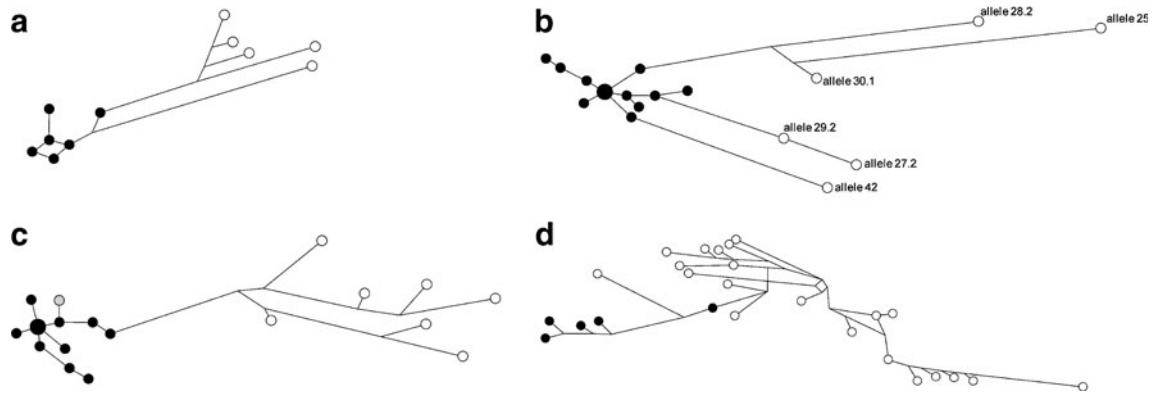


Fig. 3 Median-network analysis of haplotypes carrying atypical alleles based on 17 Y-STR (DYS19, DYS388, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS446, DYS447, DYS448, DYS449, DYS456, DYS458, DYS635, and GATA H4) or 16 Y-STR loci except for the Y-STR displaying the atypical allele. **a** Relationships between haplotypes belonging to haplogroup (O3a1)-KL2 (xJST002611). The haplotypes possessing the unusually short allele variant at DYS447 are indicated by *filled circles*. **b** Relationships between haplotypes carrying DYS449 atypical alleles. **c** Relationships between haplotypes within haplogroup O1a1-P203. The haplotypes possessing the allele 30.2 at DYS449 are indicated by *filled circles*. A Japanese haplotype with the allele 30.2 is indicated by *gray circle*. **d** Relationships between haplotypes belonging to haplogroup N-M231. The haplotypes possessing the allele 14.3 at DYS464 are indicated by *filled circles*

successfully determined using two multiplex allele-specific PCR assays in the present study.

In particular, recently redefined markers, KL2, JST002611, and P164, provide enhanced phylogenetic resolution of the Korean haplogroup O3a-M324. Although the absence of the JST002611 mutation has been reported in a Korean population [23], a considerable number of samples were found to have that mutation in our surveyed Korean males (10.2 %). The fact that this mutation is observed frequently in East Asia, including Japan, China, and mainland Southeast Asia (up to 21 %) [19, 32], supports the reliability of our data, and the presence of that mutation was also confirmed by sequence analysis (data not shown).

Interestingly, we were able to hierarchically infer non-equivalence of M117 and M133 from the presence of M117 +/M133− samples. The network analysis seems to suggest a possible founder lineage by the ancestral state of M133 out of haplogroup O3a3c1 (O3a2c1a) (M117+, M133+). However, the possibility for restoring the M133 mutation, e.g., back-mutation, is much more restricted due to the character of the 1-bp deletion mutation. Therefore, allocation of the M133 marker within phylogeny should be considered again and replaced with the more reliable binary marker M117 for designating haplogroup O3a3c1 (O3a2c1a), although further analysis is needed.

Y-STRs, rapidly mutating genetic markers, rather than Y-SNPs, are used as highly informative markers in the study of recent evolutionary events [6]. On the other hand, STR variability is structured in a haplogroup background so that the haplotypes can be used to predict haplogroup status, thereby enabling the inference of the geographic or ethnic origin of unknown samples [10–12]. Specifically, we found that genetic and haplotype variations of DYS392, DYS393, DYS437, DYS438, DYS448, and DYS388 with a relatively lower mutation rate were highly structured by haplogroup background. When assessing the haplogroup affiliation of 22 Y-STRs including the Yfiler loci, we found that the fraction of STR variability observed among haplogroups varied significantly depending on the mutation rate. This may be explained by the fact that the STRs with a low mutation rate preserved any signal of haplogroup history or founder effects and generated a small allele range due to their low mutation rate, thereby maintaining a non-homogenous allele distribution across haplogroups. This finding is consistent with the finding that the distributions of allele frequencies at DYS392 and DYS438 reflect evolutionary lineages [33]. In addition, it was recently shown that rapidly mutating Y-STRs would detect considerably less genetic population substructure [34]. Among the reported rapidly mutating Y-STRs, DYS449 was also found to have a weaker relationship with haplogroup affiliation in the Korean population. Instead, the fast mutating STRs such as DYS449 and DYS458 can increase the discriminatory capacity when added to a certain haplotype context [16, 24] such that the potential for distinguishing between paternal lineages would increase, thereby enabling high-resolution differentiation [34]. Therefore, Y-STRs with relatively low mutation rates can be more informative in terms of population or haplogroup substructure reflecting geographic region, in contrast with fast mutating Y-STRs differentiating paternal lineages.

On the other hand, atypical alleles, such as unusually short alleles or partial insertion/deletion events, were found to have certain haplogroup affiliations even though those STRs have relatively faster mutation rates. We showed that the unusually short alleles (17 and 18) at DYS447 and a variant allele 30.2 at DYS449 arose on haplogroups (O3a1)-KL2 (xJST002611) and O1a1-P203, and likely define sub-lineages within those haplogroups, respectively. It should be

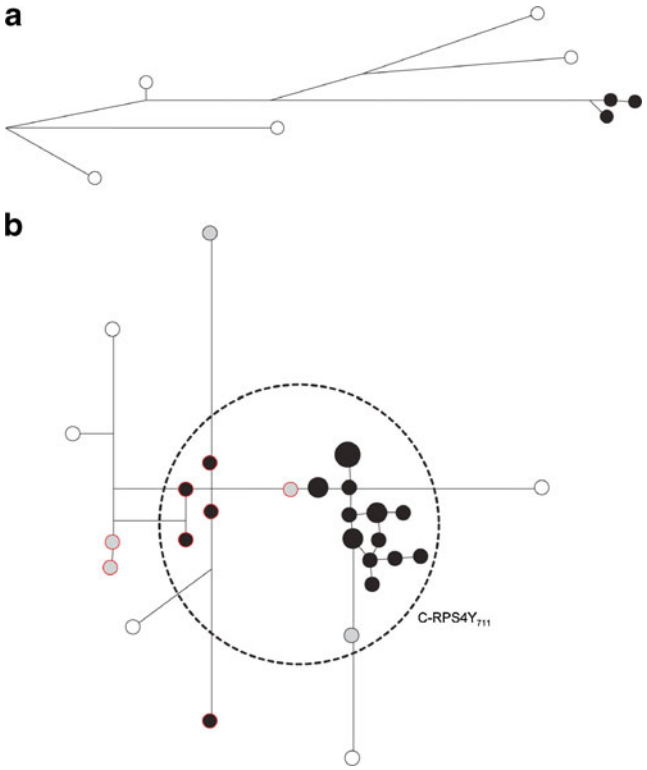


Fig. 4 Median-network analysis of haplotypes carrying a deletion mutation in flanking or entire regions around the Y-STR marker. **a** Relationship between the haplotypes carrying deletion mutations in the DYS385 flanking region. The haplotypes possessing the 8-bp deletion mutation are indicated by *filled circles*. The haplotypes possessing the 4-bp deletion mutation are indicated by *open circles*. **b** Relationship between haplotypes carrying DYS448 null alleles. The network was constructed using the haplotypes from the present study and Balaresque et al. [31] based on information from 11 Y-STR loci (DYS19, DYS388, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS446, DYS447, DYS449, DYS456, DYS458, DYS635, and GATA H4). The haplotypes with polymorphic 50f2/C deletion (or non-b1/b3 class II) are indicated by *filled circles*. The haplotypes with b1/b3 deletion are indicated by *gray circles*. The haplotypes with the other deletion type [31] are indicated by *open circles*. The haplotypes from the present study are indicated by additional *red circles*

noted that most intermediate alleles at DYS449 occur on multiple haplogroup backgrounds, whereas the allele 30.2 observed in the Korean samples and in one Japanese sample showed a strong association with a subclade within haplogroup O1a1-P203, similarly to alleles 32.2 and 32.3, which were related with haplogroup A-P97 in Cameroon [15]. Unlike atypical alleles of DYS447 and DYS449, the allele 14.3 at DYS464 might occur within as many as two sublineages in haplogroup N-M231. Haplogroup N is a major clade accompanying its derived SNPs, so higher resolution analysis using the derived SNP is needed to define the specific nature of the substructure for intermediate allele 14.3 at DYS464. Interestingly, a single haplotype sample with allele 17.2 at DYS458 along with allele 15 at DYS388 was found in our Korean population; the pattern of intermediate allele at DYS458 and alleles with ≥15 repeats at DYS388 is confined to a subclade in haplogroup J1-M267 [14, 35]. Additionally, the repeat sequence structure $(GAAA)_{15}AA(GAAA)_2$ was coincident with the intermediate allele at DYS458 in the haplogroup background J1-M267 [14]. Overall, while the DYS447, DYS449, DYS458, and DYS464 markers have a higher mutation rate, their atypical alleles seem to reflect an independent single mutation event that induces lost or decreased mutagenicity, thereby acting like binary markers. Therefore, our findings indicate that the non-consensus alleles can be useful for achieving a further level of resolution within binary Y-haplogroup tree [14, 15].

Additionally, the 8-bp deletion mutation in the DYS385 flanking region and null alleles at DYS448 caused by the polymorphic 50f2C deletion (or non-b1/b3 class II) were also associated with haplogroups O3a3b (O3a2b)-M7 and C-RPS4Y_{711}, respectively. The 8-bp deletion mutation could be mapped to Y-chromosome position 19261093–19261108 (DYS385a) or 19301828–19301843 (DYS385b) using human genome build NCBI36.3/hg18, and it has not been reported in dbSNP. The mutation only occurred in haplogroup O3a3b (O3a2b)-M7 in our samples, so further analysis of Y-SNPs derived from M7 is needed to identify the relationship between the mutation and sublineages of haplogroup O3a3b (O3a2b)-M7 in the Southeastern Asian population with notable frequencies of M7 [32, 36].

Based on our network analysis, the null alleles caused by the polymorphic 50f2C deletion (or non-b1/b3 class II) seem to have originated from a single mutation event in haplogroup C-RPS4Y_{711}. In the network-based cluster, two groups could be subdivided by Korean and Han Chinese males and by other Asians including Kyrgyzstani and Kalmyk. This finding suggests different evolutionary trajectories for each group. It is known that the deletion type of non-b1/b3 class II has risen to a high frequency in haplogroup C3*-M217 (xC3a, C3c) in the Asian population [31]. Therefore, the deletion mutation may also be useful for revealing substructure within the haplogroup C3*-M217, which occurs at a moderate frequency in Asia.

Haplogroup status from Y-STR information is inferred in forensic or population genetic fields because many STRs are routinely genotyped in multiplex assays, but SNP genotyping typically requires additional cost, time, or considerable amounts of sample DNA. Many methods for haplogroup prediction such as Y-STR allele-frequency and machine-learning approaches have been developed [11, 12] based on a computational perspective. However, in our study, only the haplotype status of DYS392, DYS393, DYS437, DYS438, DYS448, and DYS388 could be used for rough major haplogroup prediction through matching the haplotypes with known haplogroups in the reference database. Additionally, the atypical alleles with shared common sequence structure, the 8-bp deletion mutation at DYS385 flanking region, and the polymorphic 50f2C deletion reflect

a single haplogroup background such that the haplogroup status could be inferred from the Y-STR variants. Our findings and the dataset of Y-STR/Y-SNP provide useful information for inferring haplogroup background as well as for forensics, population genetics, and resolving male genealogies in Asia.

Acknowledgments The authors would like to thank the anonymous reviewers for their valuable comments and suggestions on our paper.

This research was supported by Future-based Technology Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2010-0020631).

References

1. Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. Int J Legal Med 110:118–124
2. Jobling MA, Tyler-Smith C (2000) New uses for new haplotypes the human Y chromosome, disease and selection. Trends Genet 16:356–362
3. Jobling MA (2001) In the name of the father: surnames and genetics. Trends Genet 17:353–357
4. Underhill PA, Shen P, Lin AA et al (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361
5. Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4:598–612
6. Roewer L, Croucher PJ, Willuweit S et al (2005) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. Hum Genet 116:279–291
7. Roewer L, Kayser M, Dieltjes P, Nagy M, Bakker E, Krawczak M, de Knijff P (1997) Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related populations. Hum Mol Genet 5:1029–1033
8. Butler JM (2003) Recent developments in Y-short tandem repeat and Y-single nucleotide polymorphism analysis. Forensic Sci Rev 15:91–111
9. Jobling MA (2001) Y-chromosomal SNP haplotype diversity in forensic analysis. Forensic Sci Int 118:158–162
10. Bosch E, Calafell F, Santos FR et al (1999) Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. Am J Hum Genet 65:1623–1638
11. Schlecht J, Kaplan ME, Barnard K, Karafet T, Hammer MF, Merchant NC (2008) Machine-learning approaches for classifying haplogroup from Y chromosome STR data. PLoS Comput Biol 4:e1000093
12. Athey TW (2005) Haplogroup prediction from Y-STR values using an allele-frequency approach. J Genet Geneal 1:1–7
13. Kayser M, Brauer S, Weiss G, Schiefenhövel W, Underhill PA, Stoneking M (2001) Independent histories of human Y chromosomes from Melanesia and Australia. Am J Hum Genet 68:173–190
14. Myres NM, Ekins JE, Lin AA, Cavalli-Sforza LL, Woodward SR, Underhill PA (2007) Y-chromosome short tandem repeat DYS458.2 non-consensus alleles occur independently in both binary haplogroups J1-M267 and R1b3-M405. Croat Med J 48:450–459
15. Myres NM, Ritchie KH, Lin AA, Hughes RH, Woodward SR, Underhill PA (2009) Y-chromosome short tandem repeat intermediate variant alleles DYS392.2, DYS449.2, and DYS385.2 delineate new phylogenetic substructure in human Y-chromosome haplogroup tree. Croat Med J 50:239–349
16. Robino C, Crobu F, Di Gaetano C et al (2008) Analysis of Y-chromosomal SNP haplogroups and STR haplotypes in an Algerian population sample. Int J Legal Med 122:251–255
17. Willuweit S, Roewer L, International Forensic Y Chromosome User Group (2007) Y chromosome haplotype reference database (YHRD): update. Forensic Sci Int Genet 1:83–87
18. Onofri V, Alessandrini F, Turchi C, Fraternale B, Buscemi L, Pesaresi M, Tagliabracci A (2007) Y-chromosome genetic structure in sub-Apennine populations of Central Italy by SNP and STR analysis. Int J Legal Med 121:234–237
19. Nonaka I, Minaguchi K, 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
20. Kim SH, Han MS, Kim W, Kim W (2010) Y chromosome homogeneity in the Korean population. Int J Legal Med 124:653–657
21. Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18:830–838
22. Yan S, Wang CC, Li H, Li SL, Jin L, Genographic Consortium (2011) An updated tree of Y-chromosome Haplogroup O and revised phylogenetic positions of mutations P164 and PK4. Eur J Hum Genet 19:1013–1015
23. Kim SH, Kim KC, Shin DJ et al (2011) High frequencies of Y-chromosome haplogroup O2b-SRY465 lineages in Korea: a genetic perspective on the peopling of Korea. Investig Genet 2:10
24. Park MJ, Lee HY, Yoo JE, Chung U, Lee SY, Shin KJ (2005) Forensic evaluation and haplotypes of 19 Y-chromosomal STR loci in Koreans. Forensic Sci Int 152:133–147
25. Lee HY, Park MJ, Chung U, Lee HY, Yang WI, Cho SH, Shin KJ (2007) Haplotypes and mutation analysis of 22 Y-chromosomal STRs in Korean father-son pairs. Int J Legal Med 121:128–135
26. Park MJ, Lee HY, Kim NY, Sim JE, Yang WI, Cho SH, Shin KJ (2007) Genetic characteristics of 22 Y-STR loci in Koreans. Korean J Leg Med 31:162–170
27. Park MJ, Shin KJ, Kim NY, Yang WI, Cho SH, Lee HY (2008) Characterization of deletions in the DYS385 flanking region and null alleles associated with *AZFc* microdeletions in Koreans. J Forensic Sci 53:331–334
28. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567
29. Gomes V, Sánchez-Diz P, Amorim A, Carracedo A, Gusmão L (2010) Digging deeper into East African human Y chromosome lineages. Hum Genet 127:603–613
30. Parson W, Roewer L (2010) Publication of population data of linearly inherited DNA markers in the International Journal of Legal Medicine. Int J Legal Med 124:505–509
31. Balaresque P, Bowden GR, Parkin EJ et al (2008) Dynamic nature of the proximal AZFc region of the human Y chromosome: multiple independent deletion and duplication events revealed by microsatellite analysis. Hum Mutat 29:1171–1180
32. Karafet TM, Hallmark B, Cox MP, Sudoyo H, Downey S, Lansing JS, Hammer MF (2010) Major east-west division underlies Y chromosome stratification across Indonesia. Mol Biol Evol 27:1833–1844
33. Gusmão L, Krawczak M, Sánchez-Diz P et al (2003) Bimodal allele frequency distribution at Y-STR loci DYS392 and DYS438: no evidence for a deviation from the stepwise mutation model. Int J Legal Med 117:287–290
34. Ballantyne KN, Keerl V, Wollstein A et al (2012) A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. Forensic Sci Int Genet 6:208–218
35. Cinnioğlu C, King R, Kivisild T et al (2004) Excavating Y-chromosome haplotype strata in Anatolia. Hum Genet 114:127–148
36. Shi H, Dong YL, Wen B et al (2005) Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. Am J Hum Genet 77:408–419