

Understanding the Y chromosome variation by haplogroup and haplotype analyses in Koreans

Kyoung-Jin Shin, Ph.D.

Department of Forensic Medicine
Yonsei University College of Medicine

Value of Y-Chromosome Markers

Y



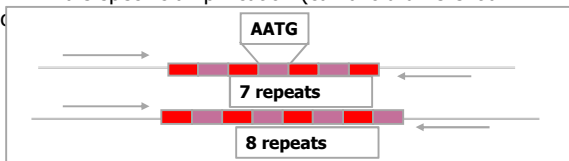
Non-Recombination
Portion of Y chromosome
(NRY)

Application

Forensic case work on sexual assault evidence
Paternity testing
Missing persons investigations
Human migration and evolutionary studies
Historical and genealogical research

Advantage

Male-specific amplification (can avoid differential



Y-STRs
(short tandem repeats)

Lack of recombination enables comparison of male individuals separated by large periods of time

Surnames usually retained by males; can make links where paper trail is limited

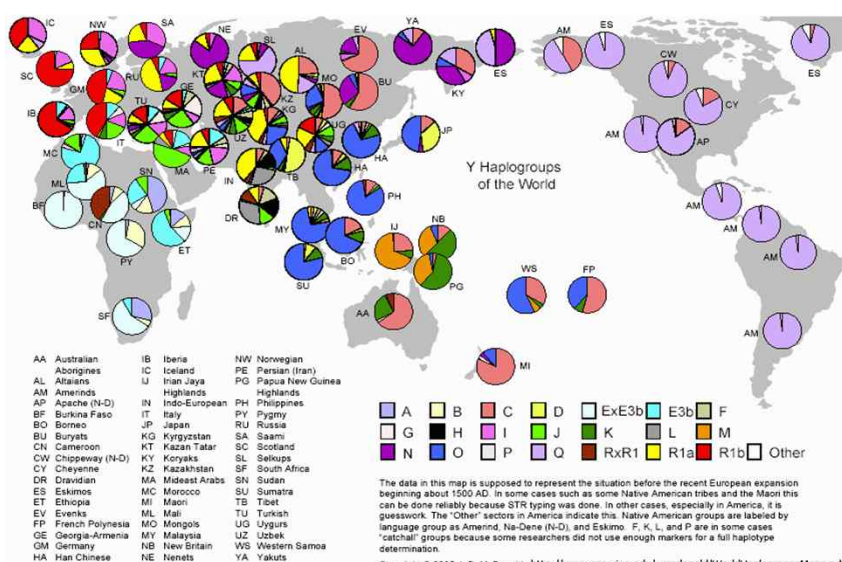
Y-SNPs
(single nucleotide polymorphisms)

Focusing on Y-SNPs

- Increasing requests for **information regarding ethnic and geographical origin** from specimens left at a crime scene.
 - Globalization of crime suspects and victims
 - Increase of movement from East-southern population
 - Increase of excavation of Korean War victims and ancient mummies
 - Individual identification in mass disaster such as airplane crashes, tsunamis or terrorist attacks where people from various geographical areas are involved

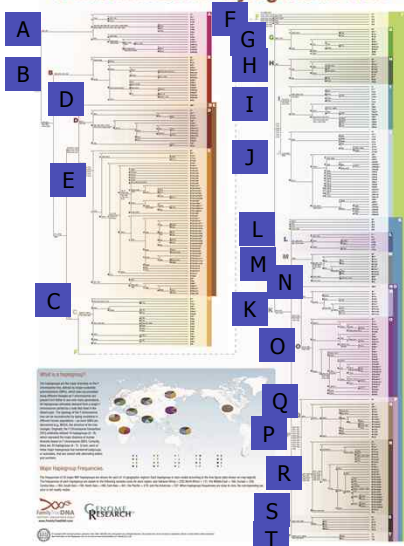
- Y-haplogroup ← combination of allelic states at SNPs
 - Set of haplotypes defined by slowly mutating markers (mainly SNPs) which have more phylogenetic stability
 - Unique event polymorphisms (UEP) record history of Y chromosome

Y chromosome haplogroup



Y chromosome phylogenetic tree

Y-Chromosome Phylogenetic Tree



- In 2008, Karafet TM et al.
 - Genome Res 2008;18:830-838
 - 20 major clades
 - 311 haplogroups based on 599 binary markers
- In 2011, Yan et al.
 - Eur J Hum Genet 2011;19:1013-1015
 - An updated tree of haplogroup O
 - Redefined marker **KL2, JST002611, PK4 and P164**



Need for development of Y-SNP typing method according to the revised tree

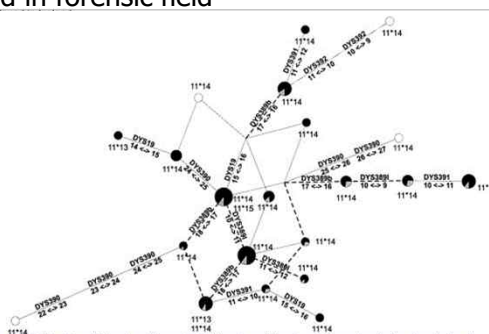
Relationships between haplogroup and haplotype

- Y-STR haplotype mainly used in forensic field



<http://www.yhrd.org>

- The Y-STR variability with



Number	Name	Haplogroup	DYS 9	DYS 13	DYS 15	DYS 19	DYS 23	DYS 24	DYS 29	DYS 30	DYS 31	DYS 33	DYS 35	DYS 37	DYS 39	DYS 42	DYS 43
N2341	Khan	J2	13	13	30	24	9	14	14	14	12	11					
N2350	Khan	J2	13	14	29	24	9	15	14	14	12	11					
N2351	Khan	J2	14	13	29	24	9	14	14	14	12	12					
N2984	Khan	J2	13	14	30	26	9	15	14	14	12	12					
N2990	Khan	J2	13	14	30	23	9	14	14	14	12	11					

Relationships between haplogroup and haplotype

- Haplogroup prediction from Y-STR values

The screenshot shows the 'Haplogroup Predictor' website interface. It includes a navigation menu, a 'Click your program choice:' section with buttons for '21 Haplogroup Program', '23 Haplogroup Data Program', '21 Y-STR Order', '23 Y-STR Order', 'Numerical Order', and 'Alphabetical Order'. Below this is a table for entering Y-STR marker values, with columns for 'Area Selection' (Northern European, West European, Mediterranean, East European) and 'Result'. The table contains various Y-STR markers and their corresponding values. To the right of the table is a 'Results Table' with columns for 'Haplogroup', 'Final score', and 'Probability (%)'. A blue arrow points from the text box to the 'Results Table'.

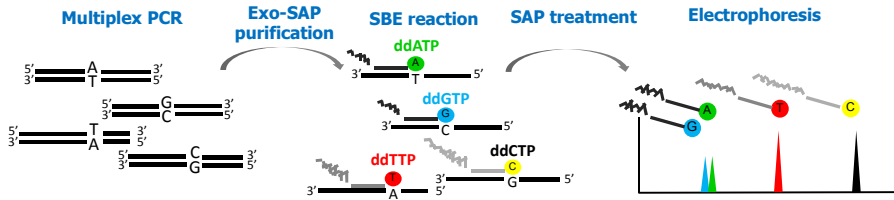
The possibility of haplogroup prediction based on the Y-STR haplotype information

Presentation Outline

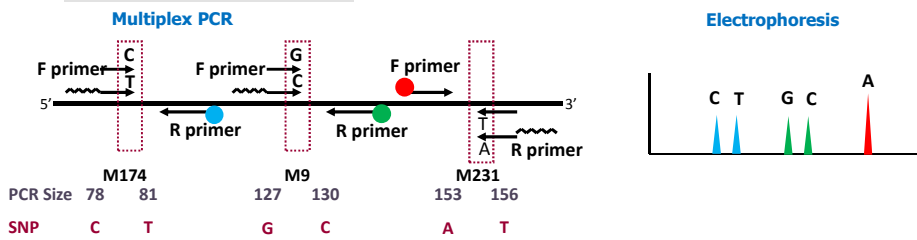
- Development of multiplex systems for determination of East Asian Y haplogroups for large [reference samples](#)
- Presentation of Y-haplogroup distributions in Koreans
- Elucidation of relationships between Y-haplogroups and Y-STR haplotypes in a Korean population
 - Evaluation of haplogroup affiliation for Y-STR allele/haplotype
 - Representation of the substructure within haplogroup
 - Evaluation of possibility for haplogroup prediction based on Y-STR haplotypes

Materials and methods

Multiplex SBE reaction

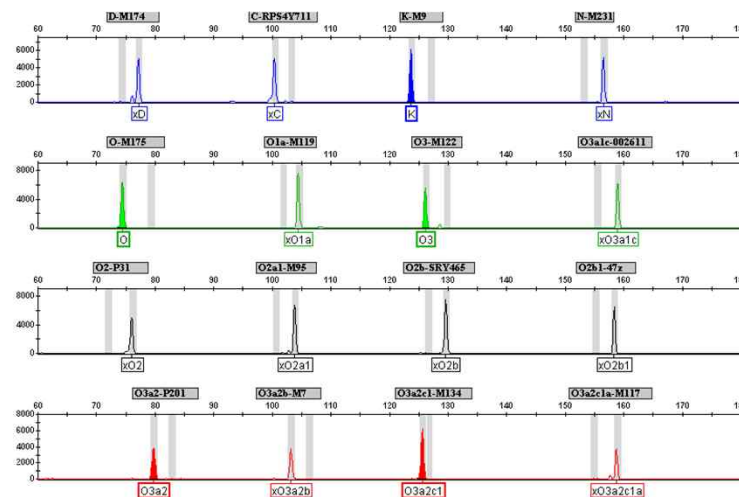


Multiplex allele-specific PCR



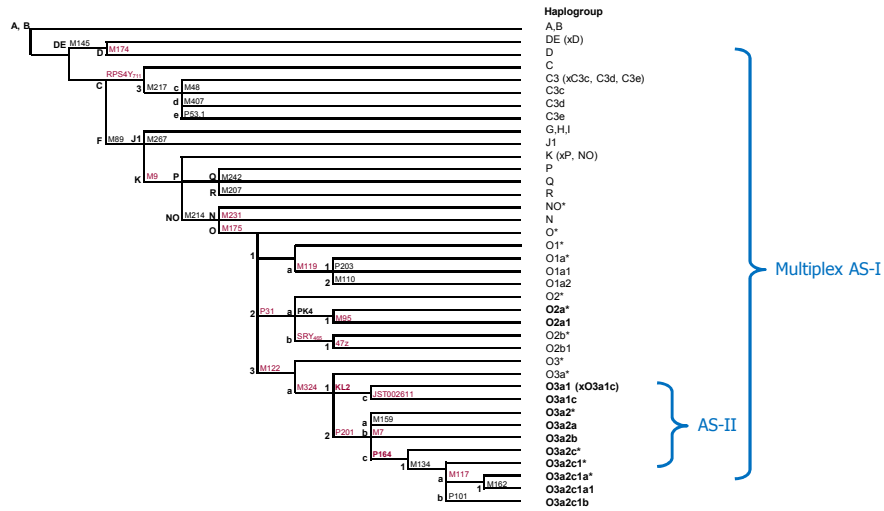
Results – Development of multiplex PCR systems

■ Multiplex AS-I



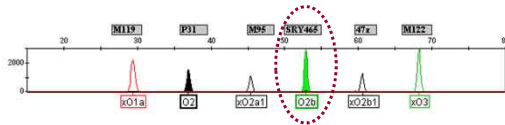
Results – Development of multiplex PCR systems

- Multiplex allele-specific PCR assay

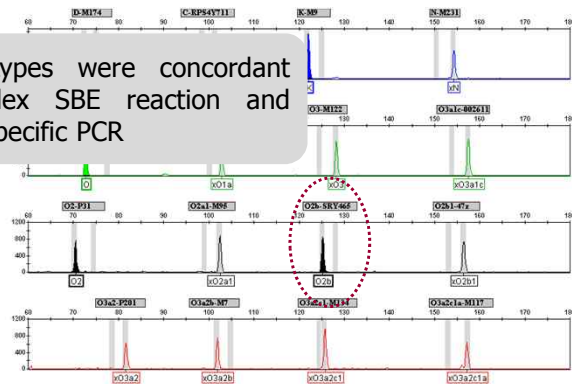


Results – concordance test

Multiplex SBE reactions



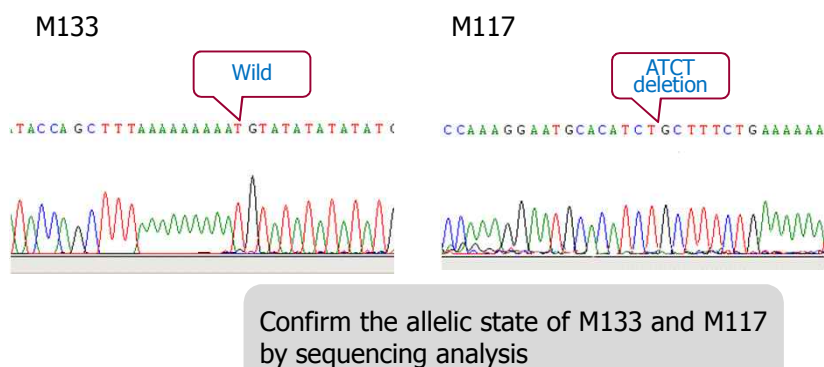
Multiplex allele-specific PCR



All SNPs genotypes were concordant between multiplex SBE reaction and multiplex allele-specific PCR

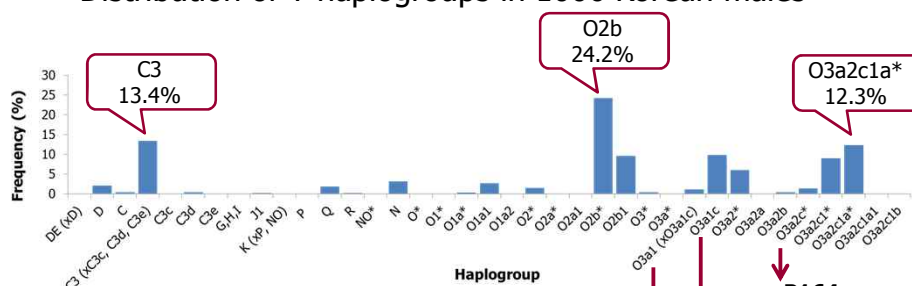
Results – concordance test

- Discordance haplogroup O3a2c1a designation between M133 (multiplex SBE reaction) and M117 (multiplex allele-specific PCR) in one sample



Results – distribution of haplogroups

- Distribution of Y haplogroups in 1006 Korean males



21 different haplogroups were identified by 33 Y-SNPs

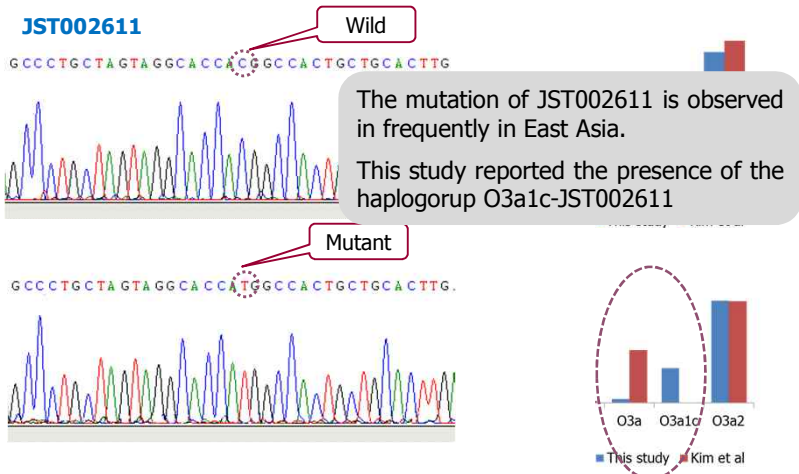
The haplogroup diversity : 0.8830

Discriminatory capacity : 2.1 %

The phylogenetic position of newly redefined markers KL2, JST002611 and P164 was confirmed

Results – distribution of haplogroups

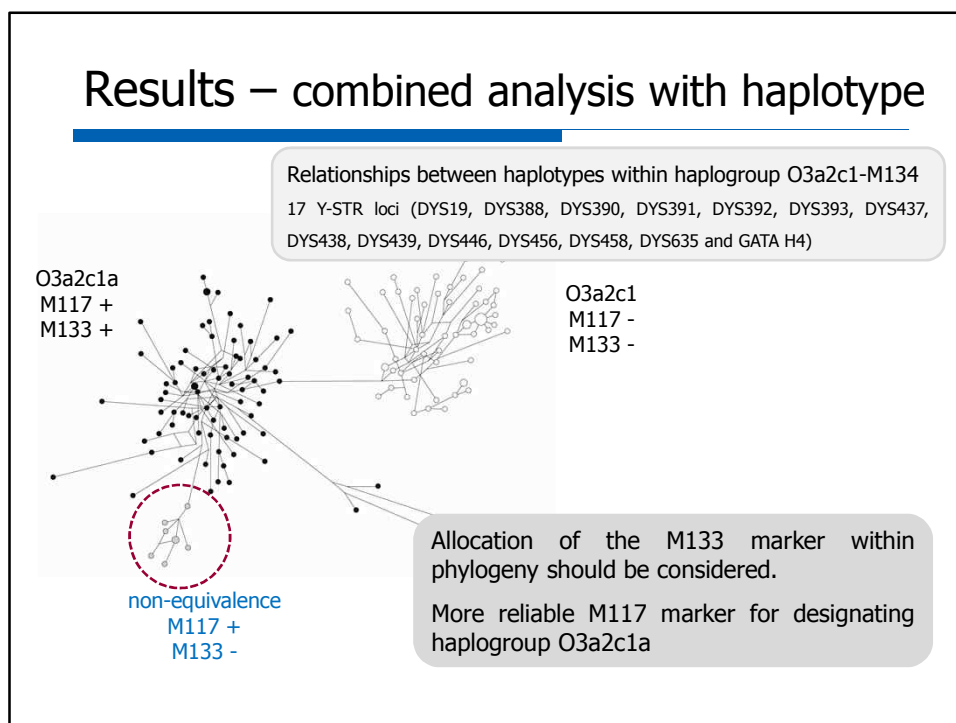
- Comparing known haplogroup frequencies from other study for Korean



Results – combined analysis with haplotype

- Non-equivalence between M117 and M133 markers
 - Combined haplogroup and haplotypes analyses in 706 samples with known 22-YSTR genotypes.
 - M117 and M133 designate the haplogroup O3a2c1a
 - Non-equivalence M117(+) M133(-) samples were found.
 - A median-joining network was constructed to evaluate the relationships of these samples based on 17 Y-STR haplotype information

Results – combined analysis with haplotype



Results – combined analysis with haplotype

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

Y-STR	Gene/haplotype diversity	Mutation rate (x10 ⁻³) ^a	% Variance	
			Among	Within haplogroups
DYS392-DYS393-DYS438-DYS437-DYS448-DYS388			55.18	85.00
		% Variance		
	Gene diversity	(x10 ⁻³) ^a	Among haplogroups	Within haplogroups
6 Y-STRs	0.9317	-	71.15	28.85
9 Y-STRs	0.9966	-	42.09	57.91
17 Y-STRs	0.9995	-	41.52	58.48
22 Y-STRs	0.9999	-	38.95	61.05
DYS138	0.7772	0.30	20.21	69.79
DYS635	0.6857	5.66	26.24	73.76
GATA H4	0.6115	3.43	38.28	61.72
DYS388	0.4850	0.00	72.33	27.67
DYS446	0.7886	2.71	31.95	68.05
DYS447	0.7520	5.41	42.78	57.22
DYS449	0.8523	18.97	14.86	85.14
DYS464	0.9668	3.99	16.71	83.29

Results – combined analysis with haplotype

Haplotype (392-393-438-437-448-388)	N	%	Haplogroup	YHRD database release 39								
				DE	D	C	C3	F	Q	N	O	
11-13-10-14-17-12	10	55.6	D-M174	88	88	1		1				
11-13-10-14-19-12	2	11.1		222	208	3		2				
11-13-11-14-17-12	2	11.1		29	28							
11-15-10-14-21-12 (or 13)	23	26.4	C3-M217			115	38					
11-14-10-14-21-12 (or 13)	19	21.8		2		74	29					
11-14-10-14-22-13	12	13.8				45	15	1				
11-13-11-14-0-13	4	4.6			11	5		3				
14-14-12-14-19-12	7	53.8	Q-M207					20	18			
15-14-12-14-19-12	3	23.1						8	5			
14-13-10-14-20-13	5	18.5	N-M231					17		14	2	
14-13-10-14-19-12 (or 13)	4	14.8						39		27	10	
15-13-10-14-19-12 (or 13)	4	14.8						15		14		
16-14-11-14-19-12	3	11.1						10		10		
14-13-10-14-18-12	3	11.1						60		8	52	

Results – combined analysis with haplotype

Haplotype (392-393-438-437-448-388)	N	%	Haplogroup	YHRD database release 39					
				O	O1a	O1a1	O2	O2b	O2b1
14-13-10-14-18-12	13	68.4	O1a1-P203	52	44	13	1		
16-13-10-14-18-12	2	10.5		3	3	2			
13-13-10-14-18-12	3	30.0	O2*-P31	15	1		11		
13-14-10-14-18-12	2	20.0		45			37		
13-13-13-14-18-12 (or 13)	124	80.0	O2b*-SRY ₄₆₅	714			398	398	128
13-13-13-14-18-12	46	64.8		O2b1-47z	714			398	398

Results – combined analysis with haplotype

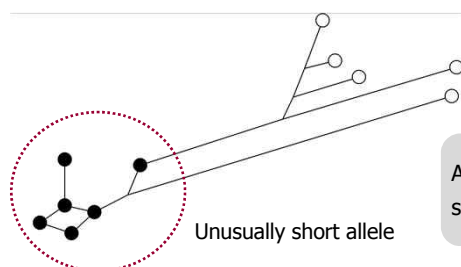
Haplotype (392-393-438-437-448-388)	N	%	Haplogroup	YHRD database release 39					
				O3	O3a	O3a2	O3a2b	O3a2c1	O3a2c1a
13-12-10-15-21-12	6	54.5	O3a1-KL2	12	10	2			
12-12-10-15-21-12	3	27.3		7	6				
13-12-10-14-19-12	24	33.3	O3a1c-JST002611	129	96	10			
13-12-10-14-20-12	16	22.2		70	55	16			
13-12-10-14-20-13	7	16.7	O3a2*-P201	70	55	16			
13-12-10-15-20-13	25	59.5		121	93	77	4	1	
13-12-10-15-20-12	3	100	O3a2b-M7	121	93	77	4	1	
13-12-9-15-19-12	3	33.3	O3a2c*-P164	4	4	4			
13-13-10-14-19-12	2	22.2		22	18	15			
12-12-10-15-19-12	36	52.9	O3a2c1-M134	122	122	115		77	
12-13-10-15-19-12	9	13.2		30	30	28		20	
14-12-11-15-20-10	53	58.9	O3a1c1a-M117	211	211	202		150	110

Results – combined analysis with haplotype

- Relationships between Y-STR variants and haplogroups
 - DYS447 unusually short allele

Y-STR	Allele	Structure	Haplogroup	N
DYS447	18	(TAATA) ₇ TAAAA(TAATA) ₁₀ TAAAA(TAATA) _n	O3a1-KL2	2
	19	(TAATA) ₇ TAAAA(TAATA) ₁₁ TAAAA(TAATA) _n	O3a1-KL2	4

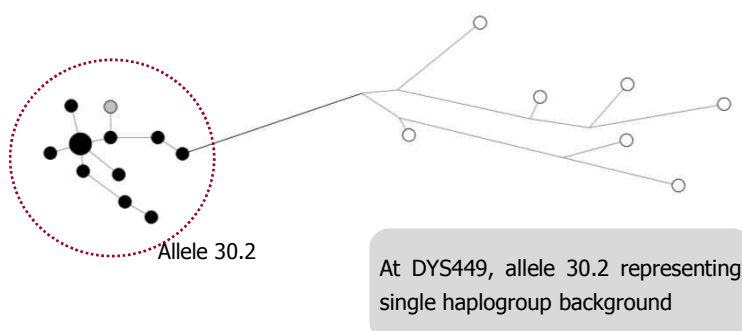
- Median-network in haplogroup O3a1-KL2



At DYS447 short allele representing single haplogroup background

Results – combined analysis with haplotype

- Relationships between Y-STR variants and haplogroups
 - Relationships between haplotypes within haplogroup O1a1-P203

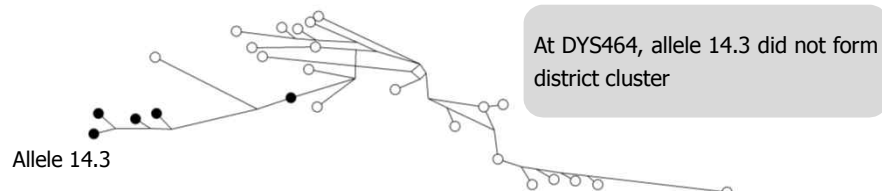


Results – combined analysis with haplotype

- Relationships between Y-STR variants and haplogroups
 - DYS464 variants

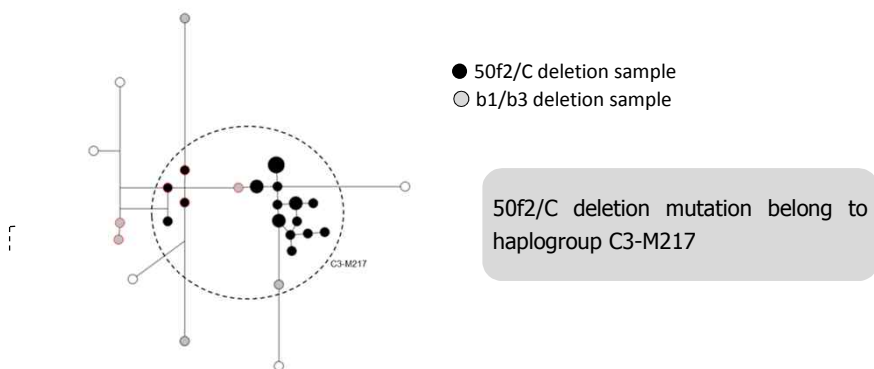
Y-STR	Allele	Structure	Haplogroup	N
DYS464	12.3	(CCTT) ₇ CTT(CCTT) ₅	O3a2*-P201	1
	14.3	(CCTT) ₃ CTT(CCTT) ₁₁	N-M231	5

- Median-network in haplogroup N-M231



Results – combined analysis with haplotype

- Relationships between Y-STR variants and haplogroups
 - Relationship between haplotypes carrying DYS448 null allele



Summary

- 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 and [useful for simple and rapid identification](#) of the haplogroups in large number of samples.
- A total of 21 different haplogroups were identified by 33 Y-SNPs using the multiplex allele-specific PCR, multiplex SBE reactions as well as additional monoplex SBE reactions.
- Phylogenetic non-equivalence was found between SNPs M117 and M133, suggesting that the position of the M133 should be corrected.

Summary

- **DYS392-DYS393-DYS438-DYS437-DYS448-DYS388**, which markers are showed the relatively **slower mutation rates**, could be used for rough **major haplogroup prediction** through matching the haplotypes with known haplogroups in the reference database.
- The **DYS447, DYS449 and DYS464** markers have a **higher mutation rate**, their atypical alleles seem to reflect an independent single mutation event, thereby **acting like binary markers**.
- Our finding 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 East Asia.

Acknowledgement

- This presentation material was prepared by Myung Jin Park in the course of her doctoral dissertation.