

East Asian mtDNA haplogroup determination in Koreans:

Haplogroup-level coding region SNP analysis and subhaplogroup-level control region sequence analysis

Hwan Young Lee, Ji-Eun Yoo,
Myung Jin Park, Ukhee Chung,
Kyoung-Jin Shin, Chong-Youl Kim

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



YONSEI UNIVERSITY
COLLEGE OF MEDICINE

Korean mtDNA database establishment and haplogroup assignment

- ❖ A high quality mtDNA control region sequence database was established in **593 Koreans** (<http://forensic.yonsei.ac.kr/>)
- ❖ Based on **shared haplogroup-specific polymorphisms in control region sequence**, 592 mtDNAs (99.8%) were classified into various East Asian haplogroups or subhaplogroups
- ❖ Statistical parameters were calculated using **“mtDNA Star”**

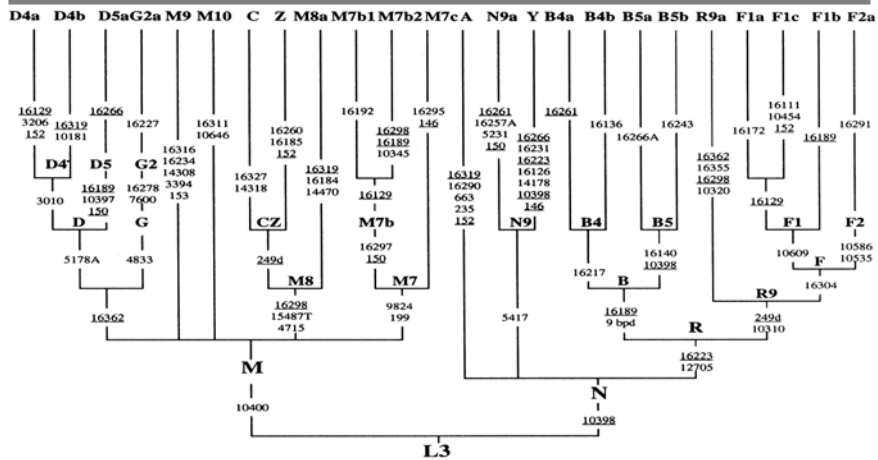
Statistics	Mutations	Transitions	Transversions	Deletions	Insertions	Length (nt)	GC (%)	GC (nt)	GC (nt)	GC (nt)
Target Region	193	22	14	7	11	486	0.990	0.32	10.44	
Control Region	284	236	26	21	19	111	0.992	0.25	13.40	
Proportion (%)	77.46	61.79	79.97	66.67	36.94	0.00	100.00	94.31	99.93	126.28

(K-J Shin, Yonsei University, unpublished)



YONSEI UNIVERSITY
COLLEGE OF MEDICINE

mtDNA haplogroup determination has practical value in forensic field



The ideal approach is confirmation of diagnostic coding region SNP

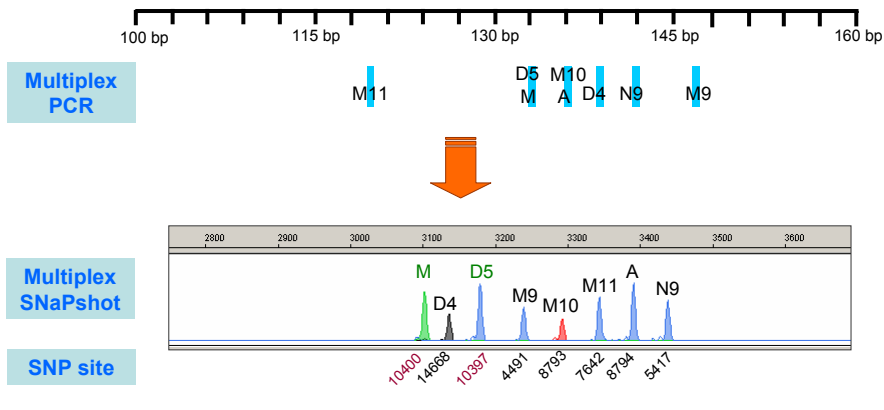
- ❖ Previously identified control region mutation motifs cannot exactly define major haplogroups and their subhaplogroups without **complementation of coding region information**
- ❖ As an example, **D4, G and M9** mtDNA are not distinguishable only with control region sequence polymorphisms

Sample	Haplogroup	Control region sequence
BF4229	D4	16223 16362 16519 73 263 309.1C 315.1C 489 523d 524d
385	G	16223 16362 16519 73 263 309.1C 315.1C 489
476	G	16078T 16179 16223 16234 16362 16519 73 152 263 309.1C 309.2C 315.1C 489
BF4102	M9	16223 16234 16274 16362 73 153 263 315.1C 489
409	G	16189 16223 16269 16278 16362 73 260 263 284 309.1C 309.2C 315.1C 489
BF4271	D4	16172Y 16189 16223 16278 16362 73 263 309.1C 315.1C 489 573.1C 573.2C 573.3C 573.pC



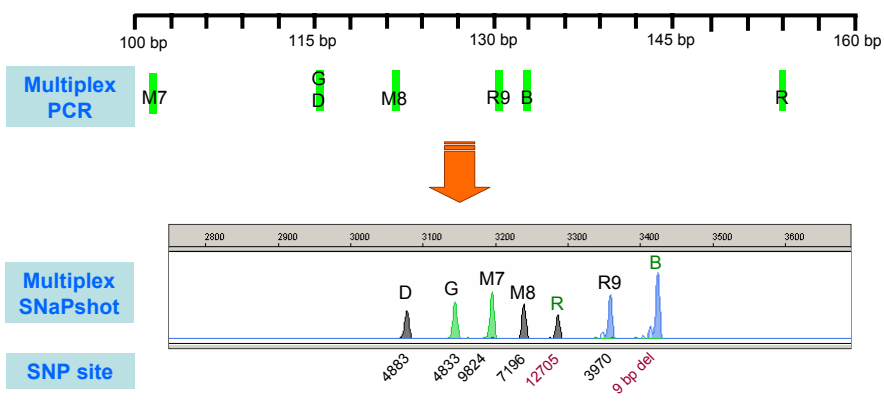
Design of three multiplex systems for coding region SNP scoring

❖ Multiplex I



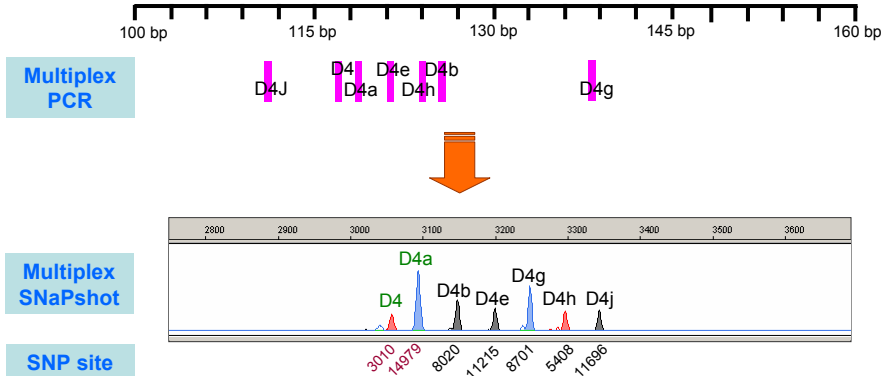
Design of three multiplex systems for coding region SNP scoring

❖ Multiplex II



Design of three multiplex systems for coding region SNP scoring

❖ Multiplex III



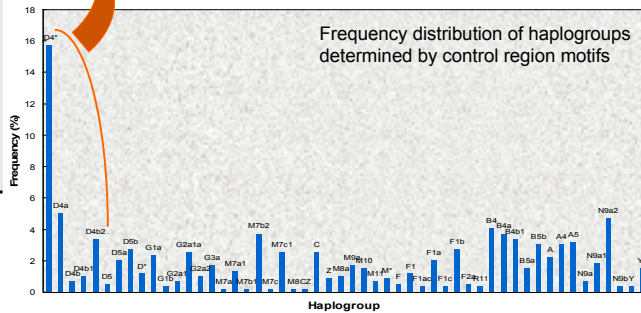
Control region motifs for East Asian haplogroups were identified

Haplogroup	HV1	HV2	HV3 etc
B4	16183C-16189-16217		
B4a	16182C-16183C-16189-16217-16261		(16519), 523d-524d
B4b1	16136-16183C-16189-16217		16519, 499
B4d	16183C-16185-16189d-16217-16234		546
B4c1a	16183C-16189-16217-16311		16519
B4c1b	16140-16183C-16189-16217-16274-16335	150	
B4c1c	16183C-16189-16217-16311	150-195-214	
B4f	16168-16172-16183C-16189-16217-16249-16325	200	16390
B5a1	16129-16140-16187-16189-16266R	93-210	16519, 523d-524d
B5b	16140-16183C-16189-16243		16519, 523d-524d, (or 513d-514d)
A	16223-16290-16319	235	
A4	16223-16290-16319-16362	235	523d-524d
A5	16187-16223-16290-16319	235	523d-524d
N9a	16223-16257A-16261	150	
N9a1	16129-16223-16257A-16261	150	
N9a2	16172-16223-16257A-(16261)	150	
N9a2a	16172-16223-16257A-16261	150	16497
N9b	16183C-16189-16223		16519
Y1b	16126-16231-16266	146	16519
Y2	16126-16231-16311		482

Coding region SNP scoring is useful for molecular dissection of D4 haplogroup

Haplogroup	Freq.(%)
D4*	8.26
D4a	5.06
D4b	0.84
D4b1	1.69
D4b2	3.71
D4e	2.52
D4g	1.00
D4h	1.35
D4j	2.35

Coding region SNP scoring using Multiplex III



Control region motifs for D4 subhaplogroups are identified

Haplogroup	HV1	HV2	HV3 etc
D4	16223-16362		489
D4a	16129 -16223-16362	152	(16519)-489
D4b1	16223- 16319 -16362		489-523d-524d
D4b2*	16223-16362		489-523d-524d
D4b2b	(16223)-16362	194	16519-489-523d-524d
D4d	16245 -16362		489
D4e*	16223-16362		489
D4e1	16223-16362	94	489
D4g1	16223- 16278 -16362		489-573.pC
D4h*	16223-16362		489
D4h1	16174 -16223-16362	146-183	489
D4h2	16174 -16223- 16311 -16362	152	489
D4i	16223- 16294 -16362		489
D4j*	16223-16362		489
D4j1	16184 -16223- 16311 -16362		489
D4k1	16192 -16223	195	489
D4k2	16223- 16274 - 16290 - 16319 -16362	195	489
D4m	16244 -16362		489
D4n	16223- 16355A -16362		489

Coding region SNP scoring is indispensable in some haplogroups

- ❖ One of **G2a1** haplotype according to control region sequence was found to be **D4g** haplogroup, and 8 and 1 of **D4** haplotypes turned out to be **G** and **M9** haplogroups, respectively
- ❖ **D4 paragroups**, e.g., **D4***, **D4b2***, **D4e*** and **D4j***, which have a mutation motif 16223-16362-489, need **coding region SNP scoring for exact haplogroup determination**
- ❖ Complementation of coding region SNP information to control region polymorphisms will lead to **mtDNA data quality control** and **molecular dissection of haplogroups**

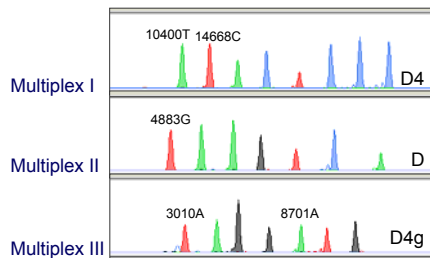


YONSEI UNIVERSITY
COLLEGE OF MEDICINE

Multiplex systems are proved to be efficient in skeletal remain analysis

- ❖ Efficiency test was performed in **101 skeletal remains from Korean War (1950~1953) victims**
- ❖ **Small amplicon sizes** enabled SNP score in old skeletal remains to be successfully analyzed without artifact

HV1-HV2-HV3 region sequence
16223-16278-16320-16362
73-263-309.1C-315.1C
489



YONSEI UNIVERSITY
COLLEGE OF MEDICINE

East Asian HG can be determined using “mtDNA Sequence Manager”

❖ We have developed the haplogroup determining program, “mtDNA Sequence Manager” based on the collated control region mutation motifs for East Asian haplogroups or subhaplogroups

❖ By using this program, 593 Korean mtDNAs and 101 Korean War victim mtDNAs can be classified into various East Asian haplogroups or subhaplogroups

Sample	Expected HG	Determined HG	Sequence
C5881	M7b2	M7b2	16129C 16129A 16183C 16189C
C5882	D4a1	D4a1	16192T 16223T 73G 195C 263G
C5883	I5b	I5b	16140C 16182C 16183C 16189C
C5884	M10b	M10b	16066G 16223T 16311C 73G 29G
C5886	H5a1	H5a1	16111T 16129A 16223T 16257A
C5887	I4a	I4a	16182C 16183C 16189C 16194C
C5888	I5b	I5b	16140C 16183C 16189C 16194C
C5889	I5a1	I5a1	16129A 16140C 16187T 16189C
C5810	M7b2	M7b2	16129A 16183C 16189C 16223T
C5811	M7c	M7c	16223T 16319A 16519C 73G 14G
C5812	M7b2	M7b2	16129A 16172C 16189C 16223T
C5813	Y1b	Y1b	16126C 16231C 16266T 16319A

(K-J Shin, Yonsei University, unpublished)



YONSEI UNIVERSITY
COLLEGE OF MEDICINE

“mtDNA Sequence Manager” is applicable to mtDNA DB quality control

❖ By assigning new mtDNA sequences to appropriate East Asian mtDNA haplogroups using this program, it is possible to pinpoint errors present in mtDNA sequences

❖ Discordance between expected HG and determined HG implies problems with given data

Y1b mutation motif:
16126-16231-16266-146

Sample	Expected HG	Determined HG	Sequence
129	D4	D4	16223T 16362C 73G 146C 263G
129	D4	D4	16223T 16362C 73G 152C 287A
130	C	C	16195C 16223T 16298C 16299C
131	M0	M0	16182T 16223T 16298C 16319A
132	A	A	16195T 16223T 16298T 16319A
133	D4 M10	D4 M10	16223T 16256T 16311C 16362C
134	D4	D4	16223T 16362C 73G 152C 263G
136	C	C	16219A 16223T 16298C 16227T
137	D4 (D4b)	D4 (D4b)	16223T 16319A 16362C 73G 152G
140	Y1b	Y1b	16231C 16266T 16319A 73G 146G
143	D4 (D4a)	D4 (D4a)	16223T 16356A 16362C 73G 152G
145	F	F	16182T 16223T 16298C 16319A

(K-J Shin, Yonsei University, unpublished)



YONSEI UNIVERSITY
COLLEGE OF MEDICINE

Concluding remarks

- ❖ East Asian haplogroup determination is efficiently carried out using **haplogroup-level coding region SNP analysis** and **subhaplogroup-level control region sequence analysis**
- ❖ **Identification of control region mutation motif and molecular dissection of haplogroups** can be achieved by coding region SNP analysis
- ❖ The 3 multiplex systems work well even in **degraded samples** and it will present a promising means for forensic and human genetics involving East Asian mtDNA haplogroups



YONSEI UNIVERSITY
COLLEGE OF MEDICINE

Acknowledgement

- ❖ *to our lab members and to Kokuryo Research Foundation for research fund support*



YONSEI UNIVERSITY
COLLEGE OF MEDICINE