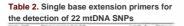
## East Asian mtDNA haplogroup determination in Koreans: Haplogroup-level coding region SNP analysis and subhaplogroup-level control region sequence analysis

Hwan Young Lee<sup>1</sup>, Ji-Eun Yoo<sup>1</sup>, Myung Jin Park<sup>1</sup>, Ukhee Chung<sup>1</sup>, Chong-Youl Kim<sup>1,2</sup> and Kyoung-Jin Shin<sup>1,2</sup> <sup>1</sup> Department of Forensic Medicine, Yonsei University College of Medicine, Seoul, Korea <sup>2</sup> Human Identification Research Center, Yonsei University, Seoul, Korea

Recently, a high-quality mtDNA control region sequence database for 593 Koreans was established through a redundant approach to data generation and analysis. For data quality control, Korean mtDNAs (99.8%) were also classified into various East Asian mtDNA haplogroups or subhaplogroups based on previously reported patterns of shared haplogroup-specific or haplogroup-associated polymorphisms in the control region. However, many haplogroup-diagnostic SNPs are located in mtDNA coding regions, and in some haplogroups, scoring of coding region SNPs is required for exact haplogroup determination due to the lack of information in their control region sequences. In addition, the ideal approach to haplogroup determination is the direct confirmation of diagnostic coding region SNPs. Accordingly, 21 SNP markers and 1 deletion motif from the coding region were selected, and three multiplex systems applying single base extension methods were developed in the present study. Using two of the multiplex systems, all 593 Korean mtDNAs were allocated into 15 East Asian major haplogroups: M, D, D4, D5, G, M7, M8, M9, M10, M11, R, R9, B, A, and N9. Using the other multiplex reaction, haplogroup D4 was further determined into six subhaplogroups: D4a, D4b D4e, D4g, D4h, and D4j. Using these techniques, the absence of major systematic errors in the data was confirmed. In addition, control region mutation motifs important for the assignment of East Asian mtDNA haplogroups and subhaplogroups were identified by collating informative control region SNPs on the basis of coding region SNP information. As the small amplicon sizes used in the three multiplex systems are expected to produce good results in degraded samples, the efficiency of the systems were tested in 101 samples from skeletal remains obtained from Korean War (1950-1953) victims.

Table 1. Selected SNPs and primers for the

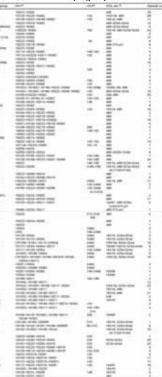
-	-		Print access of TOP	10000	7,11	Anna Ant	Figure 1. Electropherograms of the multiplex I and multiplex II
						-	obtained from a sample belonging to the haplogroup D5 (
61	4400	15	UNCETATIACALCULATED.	0.41	22	141	
2				0.44		1.11	sample belonging to the haplogroup B (B). Electropherogra
•	-	1	UNCENTIACOD/RCF1MADD	1.41	111	144	
	100		and the station section and	8.11	10.0	100	multiplex III from samples belonging to D4a and D4b, respectiv
			INVESTIGATION OF THE OWNER	4.10	56.0		
es .	4754		SEACECTAMASSACISM/COTUR	1.49	96.7	100	
	10794		autoconsistentia	0.44	98.5		
	+9984	12	THE AND TRACK THE PERSON NEW YORK	4.88	10.0	1.84	A
<u>.</u>		121	INCOLOUGA AND CONTRACT	0.00	10.0	100	A Hart Hart Mart Mart
۰.		÷.,	PROBABILA PROPARATION	1.00	- 22		14468C 44010 min 1
6	and the		institution and attracted advised	6.48	10.4	100	
			UNIDER ANY MUSIC ANY ADDRESS OF A DECK	5.41	16.1		AND ADD ADD ADD
	400		STOCCASH AND TRUCK MARK	5.46	10.0	1.16	ATTIA TOAC LINET
	4400		-boothilacontributions and	1.46	14.4		She del
	1100		ataocasacritecipoceasa.	5.40	16.0	198	Figure 2. Electrophero
			541030001Ath000A0966	1.41	14.7		
	6-8-1	- 1	HIGH CLUET STYTACCORT TTACT TO ACCORT	22	- 66.4	1.00	the three multiplexe
	82996				10.4	1.11	the three multiplexe
	-	- 2	WITHOUT DOLLARS AND A	5.46	22.2	100	B skeletal remain samp
	11	121	mittuatis/ats/matca*	1.00		i interiori	INNEC THE AND THE SKETCER TETRAIT Samp
			RENEWAL AND	1.00	100		belonging to D4g
-	1000						beionging to b4g
	March 1		INDEX ADDRESS OF ANY OCTA	1.0	14.4	www.	
	-		RIGHTSANGANA (SANCOTT	0.49	0410		-10-44
	14070		-BODATISENCTINCTELADERAM	6.01	10.0	1.0108	HEAT Interest and a second sec
			AGAKIN/TSA/SOCIADIA/TS	0.44	08.4		MENT AND THE INTER AND THE AND THE
	400		UNACTAGO DI MATORI	0.04	96.7	1016	alloc stores stores
		. *	TRADUCIAN PRODUCTTON	2.44	96.7		
	11278		OCACECTIN/CODEACCTV1	12.01	- 16 er	1971	
			without the state of the state	1,20		100	
×	810	12.	EINACTAR/CARDINE/ICGARGE	0.04	96.6		C Hitt
1.1	1.00	121	numeror and an international	0.04		Contra 1	C LIFTSC AND AND AND AND AND AND
-	-	12	many latter and a second work as	2.45	- 22.2	100	NUMA BUILD
		12	dis bio a the line in the second	1.00	-	1.00	MEDA LITTLE SHET HEROG
		12.1	department and the activity	1.00	- 21		



-		-	- 19	1.44
1.20535-2	A TO OTT THE TRANSPORTATION OF A TO TO OTHER TO AN OWNER AN ADDRESS TO TO A TO AN ADDRESS OF A TO TO AN ADDRESS OF A TO A TO TO AN ADDRESS OF A TO A TO A DRESS OF A TO A TO A TO A TO A DRESS OF A TO A TO A TO A TO A TO A DRESS OF A TO A TO A TO A TO A TO A DRESS OF A TO A TO A TO A TO A TO A TO A DRESS OF A TO A TO A TO A TO A TO A TO A TO A DRESS OF A TO A	*******	11111111111	100000000000000000000000000000000000000
1	Mulacitiacawaenergi Mulacitiacawaenergi Mulacitiacamaenergi Mulacitiacamaenergi Mulacitiacamaenergi Mulacitiacamaenergi Mulacitiacamaenergi	11111111	1111111	11111111
1111111	Contract and contract the second of the seco	2211232	1222122	1112121

Ibeer IMPO Seco Seco Seco	
HINT ALLA FORT	Figure 2. Electropherograms of the three multiplexes from
1480C 1187A 4890 3400 <sup>2594</sup> 1020 1480C 1997 1020 14994d	skeletal remain sample, B042, belonging to D4g
ADC ADIA TOTO	IGANET LANGET AUTO NOOSTING 4250
MINA RUDO LITIS SHAT HERE	anor alla Tibe Mor
MEAA LATIVE LIZINE MARTING	MEAN HETTER AND A A A A
3 mtDNA Hanlogroup freque	ncies observed in 694 Koreans
S. mona napiogroup neque	

Table 3. Control region mutation motifs for East Asian mtDNA haplogroup determination



In conclusion, the data show that the identification of control region mutation motifs and the molecular dissection of haplogroups can be achieved by coding region SNP analysis. This study shows that East Asian mtDNA haplogroup determination is efficiently carried out using haplogroup-level coding region SNP analysis and subhaplogroup-level control region sequence analysis, and that East Asian mtDNA data quality control and error identification can be easily performed using the reliable control region mutation motifs. Also, the three multiplex systems produced good results even in degraded samples and are a promising tool for forensic and evolutionary genetics involving East Asian mtDNA haplogroup determination.

Published in Electrophoresis (2006) 27:4408-4418