

Ancestry informative SNPs and mtDNA analysis using the Ion PGM system

Masaki Hashiyada Ph.D. Dept. Legal Medicine Kansai Medical University Japan



 Japanese SNP database for racial discrimination (Precision ID Ancestry Panel)

mtDNA analysis of old human bones
 (300 ~ 150 years earlier)
 (Precision ID mtDNA Whole Genome Panel)

Japanese SNP database to get biogeographic ancestry information (for racial discrimination)

Precision ID Ancestry Panel (Applied Biosystems)

This kit can provide biogeographic ancestry information about samples.

Target	Amplicon length	Input DNA	13 shared 123 SNPs SNP (Dr. M. Seldin)	1
165 SNPs	Ave. 130 bp (Dr.K.Kidd) Ave. 120 bp (Dr.M.Seldin)	1 ng	123 55 55 55 SNPs (Dr. K. Kidd)	

DNA samples

225 unrelated Japanese volunteers' DNA

Construct Library

- Precision ID Ancestry Panel
- Precision ID Library Kit

Prepare Templete

- Ion Chef[™] Instrument
- Ion PGM[™] Hi-Q[™] Chef Kit

Run Sequence

- Ion PGM[™] Sequencer
- · Ion 318[™] chip (59 samples/chip)



AB applied biosystems





♦Arlequin ver3.5

Correlational analysis of SNPs located on the same chromosome

Comparing JPT data (1000 genome project) and our data

♦Structure 2.3.4 ♦Structure Harvester

Computing the proportion of the genome of an individual originating from each inferred population based on a set of individuals genotyped at multiple markers.

∻R

For principal component analysis



Sample: Sample 1 Admixture Prediction - Set of 151 AISNPs Barcode: IonXpress_001

100 90 80 90 80 90 80 90 80 90 90 90 90 90 90 90 90 90 90 90 90 90			
	Population Name	Percentage	
	Europe	0.0	
Plug-in:	Oceania	0.0	
HID SNP Genotyper	East Asia	100.0	
	Africa	0.0	
	South Asia	0.0	
	America	0.0	
	Southwest Asia	0.0	

Confidence: **HIGH.** Log-likelihood value: 115.56. The likelihood value of the sample is within the 95% confidence interval obtained from 10,000 random samples with the same admixture proportions as the test sample

Race prediction by the population likelihoods

Population Likelihoods - Set of 151 AISNPs

Koreans

Population Name

Koreans

Japanese

Ba

Likelihood

1.14E-43

1.40E-44

Geo Region

EastAsia

EastAsia

The results of more detailed race prediction as follows;

Race	Num.	Race	Num.	
Japanese	149	Taiwanese	9	
Koreans	36	Cambodians	4	
Hakka	12	Lao Loum	3	
Han	11	Ami	1	

Accuracy: 149/225=66.2%

Population Likelihoo	ds - Set of 151 Als	SNPs Ba	Population Likelihoods - Set of 151 AISNPs Ba			Population Likelihoods - Set of 151 AISNPs			
Han-Hap	мар		Taiwar	nese	6	Cambodia	ins		
	Se .	Si;		Go -	a,		Be .		
Populatio Name Han-HapMap Koreans	Geo Region EastAsia EastAsia	Likelihood 5.30E-46 3.20E-46	Population vame Taiwanese Han Koreans	Geo Region EastAsia EastAsia	Likelihood 1.59E-46 7.71E-48	Cambodians, Khmer Japanese-HapMap	Geo Region EastAsia EastAsia	1.78E-48 1.05E-49	

Extraction of 165 SNPs data from 1000 genome project





The detailed information on the area (1)

Africa (total 661)

Memo	Number of samples
African Ancestry in Southwest US	61
African Caribbean in Barbados	96
Mandinka in the Gambia	113
Mende in Sierra Leone	85
Yoruba in Ibadan, Nigeria	108
Esan in Nigeria	99
Luhya in Webuye, Kenya	99
	Memo African Ancestry in Southwest US African Caribbean in Barbados Mandinka in the Gambia Mende in Sierra Leone Yoruba in Ibadan, Nigeria Esan in Nigeria



The detailed information on the area (2)

Europe (total 503)

Abbreviation	Memo	Number of samples	
IBS	Iberian populations in Spain	107	
TSI	Toscani in Italy	107	
CEU	Utah residents with Northern and Western European ancestry	99	
GBR	British in England and Scotland	91	
FIN	Finnish in Finland	99	



The detailed information on the area (3)

Latin America (total 347)

Abbreviation	Memo	Number of samples
MXL	Mexican Ancestry in LA California	64
CLM	Colombian in Medellin, Colombia	94
PEL	Peruvian in Lima, Peru	85
PUR	Puerto Rican in Puerto Rico	104



The detailed information on the area (4)

South-Eastern Asia (total 489)

Abbreviation	Memo	Number of samples
PJL	Punjabi in Lahore, Pakistan	96
GIH	Gujarati Indian in Houston, TX	103
ITU	Indian Telugu in the UK	102
STU	Sri Lankan Tamil in the UK	102
BEB	Bengali in Bangladesh	86

The detailed information on the area (5)

East Asia (total 729)

Abbreviation	Memo	Number of samples	
KHV	Kinh in Ho Chi Minh City, Vietnam	99	
CDX	Chinese Dai in Xishuangbanna, China	93	
CHS	Southern Han Chinese, China	105	
СНВ	Han Chinese in Beijing, China	103	
JPT	Japanese in Tokyo, Japan	104	
Our data	Japanese in north area, Japan	225	

There were no differences between JPT and our data by Arlequin analysis.

The result of Structure analysis (1)

The Structure Harvester software can calculate the number ' K' based on the allele frequency distributions. (K is the number of ancestry groups) K=3 (blue, green and red)





The result of principal component analysis (1)





The result of principal component analysis (3)



The result of principal component analysis (4)



The result of principal component analysis (5)



Summary of ancestry panel analysis

- We constructed a Japanese SNP database using the Precision ID Ancestry Panel Kit.
- European, African, and East Asian data could be discriminated clearly by this kit based on the principal component analysis.
- East Asian data (Vietnamese, Chinese and Japanese)
 could not be discriminated clearly. The more SNPs may be required to classify East Asians.

mtDNA analysis of old human bones (300~150 years earlier)



Three major features

 There are thousands of copies of mtDNA in a cell
 Inherited from a maternal lineage
 There are many SNPs in the mtDNA genome

> • Hyper Variable Regions (HVR1,2 and 3)

Haplogroups



https://www.mitomap.org/foswiki/pub/MITOMAP/MitomapFigures/simple-tree-mitomap-2012.pdf





• The three major cemeteries were found in the *Hatanai* site.

Hatanai site

 \sim

 Circled numbers indicate excavated skeletons.

Eurasian Continent

J

6

- Black-circled skeletons could not be analyzed reliably.
- Red-circled skeletons were analyzed in this study.



Cemetery No.	Specimen No.	Age	Sex	
1	54	41~over 60	Female	
2	51	41~over 60	Unknown	
3	44	about 23~40	Female	

mtDNA haplogroup : D4a



DNA samples

Bone and tooth sample : DNA solution 2 µl
Researcher A and B : 1 ng

Construct Library

- Precision ID mtDNA Whole Genome Panel
- Precision ID Library Kit

Prepare Templete

- Ion Chef[™] Instrument
- Ion PGM[™] Hi-Q[™] Chef Kit

Run Sequence

- Ion PGM[™] Sequencer
- · Ion 318[™] chip
- Plug-in: HID_SNP_Genotyper



AB applied biosystems"

The comparison of mtDNA whole genome between specimens of Hatanai and rCRS. Three skeletons had the same mtDNA whole genome sequence which was classified as D4a1c.

	Ref	Position	Variant			ł	naplog	roup			Ref	Position	١
#1	Α	73	G							#20	Т	9540	
#2	Т	152	С				D4a			#21	Α	10398	
#3	А	263	G							#22	С	10400	
#4	-	302	CC							#23	Т	10410	
#5	Т	489	С	Μ						#24	Т	10873	
#6	А	750	G							#25	G	11719	
#7	А	1438	G							#26	С	12705	
#8	Α	2706	G							#27	С	14668	
#9	G	3010	Α			D4				#28	С	14766	
#10	С	3206	Т				D4a			#29	Т	14783	
#11	А	4769	G							#30	Т	14979	
#12	С	4883	Т							#31	G	15043	
#13	С	5178	Α		D					#32	G	15301	
#14	С	7028	Т							#33	А	15326	
#15	Α	7822	G						D4a1c	#34	G	16129	
#16	С	8414	Т			D4				#35	С	16223	
#17	Т	8473	С				D4a			#36	Т	16362	
#18	А	8701	G							#37	Т	16519	
#19	А	8860	G							#38	Α	16642	

	Ref	Position	Variant		haplogroup					
#20	Т	9540	С							
#21	Α	10398	G							
#22	С	10400	Т	М						
#23	Т	10410	С					D4a1		
#24	Т	10873	С							
#25	G	11719	Α							
#26	С	12705	Т							
#27	С	14668	Т			D4				
#28	С	14766	Т							
#29	Т	14783	С	М						
#30	Т	14979	С				D4a			
#31	G	15043	Α	М						
#32	G	15301	Α							
#33	А	15326	G							
#34	G	16129	Α				D4a			
#35	С	16223	Т							
#36	Т	16362	С		D					
#37	Т	16519	С							
#38	Α	16642	G							

The comparison between researcher A and rCRS. The researcher A was typed as D4a1.

	Ref	Position	Variant	haplogroup				
#1	А	73	G					
#2	Т	152	С				D4a	
#3	А	263	G					
#4	-	302	С					
#5	-	310	С					
#6	Т	489	С	Μ				
#7	Т	593	С					
#8	А	750	G					
#9	А	1438	G					
#10	А	2706	G					
#11	G	3010	Α			D4		
#12	С	3206	Т				D4a	
#13	А	4769	G					
#14	С	4883	Т					
#15	С	5178	Α		D			
#16	С	7028	Т					
#17	С	8414	Т			D4		
#18	Т	8472	С					
#19	А	8701	G					
#20	Α	8860	G					

Α

		Ref	Position	Variant		Image:				
-	#21	Т	9540	С						\sim
	#22	Α	10398	G						
	#23	С	10400	Т	М					
	#24	Т	10410	С					D4a1	
	#25	Т	10873	С						
	#26	G	11719	Α						
	#27	С	12705	Т						
	#28	С	14668	Т			D4			
	#29	С	14761	Т						
	#30	Т	14783	С	Μ					
	#31	Т	14979	С				D4a		
	#32	G	15043	Α	Μ					
	#33	G	15301	Α						
	#34	Α	15326	G						
	#35	G	16129	Α				D4a		
	#36	С	16223	Т						
	#37	Т	16362	С		D				
	#38	Т	16519	С						
	#39	А	16642	G						

The comparison between researcher B and rCRS. The researcher B was typed as D4a1b.

	Ref	Position	Variant	haplogroup				
#1	-	42	G					
#2	Α	73	G					
#3	Т	152	С				D4a	
#4	Α	263	G					
#5	-	302	С					
#6	Т	310	TC					
#7	Т	489	С	М				
#8	Α	750	G					
#9	Α	1438	G					
#10	Α	2706	G					
#11	G	3010	Α			D4		
#12	С	3206	Т				D4a	
#13	Α	4769	G					
#14	С	4883	Т					
#15	С	5178	Α		D			
#16	С	7028	Т					
#17	С	8414	Т			D4		
#18	Т	8473	С				D4a	
#19	Α	8701	G					
#20	Α	8860	G					
#21	Т	9540	С					
#22	Α	10398	G					

В

	Ref	Position	Variant	haplogroup					
#23	С	10400	Т	М					
#24	Т	10410	С					D4a1	
#25	Α	10709	G						
#26	Т	10873	С						
#27	G	11719	Α						
#28	С	12705	Т						
#29	Α	13651	G						D4a1b
#30	С	14668	Т			D4			
#31	Α	14755	G						
#32	С	14766	Т						
#33	Т	14783	С	Μ					
#34	Т	14979	С				D4a		
#35	G	15043	Α	М					
#36	G	15301	Α						
#37	Α	15326	G						
#38	G	16129	Α				D4a		
#39	С	16223	Т						
#40	Т	16362	С		D				
#41	Т	16519	С						
#42	-	16611	G						
#43	Α	16642	G						





The differences in mtDNA whole genome in each samples



Summary of whole mtDNA analysis

All of the three skeletons buried in separate cemeteries had mtDNA of the same haplogroup "D4a1c". The mtDNA whole genome sequences of the skeletons were identical perfectly.

annau

Haplogroups of researchers A and B were typed as "D4a1" and "D4a1b", respectively. There were several differences in mtDNA sequences in each samples.

- Where interpretent and HVR 3 were identical perfectly in skeletons and researchers. Only one variant was found between skeletons and researchers in HVR 2.
- This study proved the mtDNA whole genome sequencing using Ion PGM System is useful to analyze ancient human remains.

Acknowledgment

Department of Legal Medicine, Kansai Medical University

> Atsushi Akane, Tomohiro Matsumoto, Sumitaka Yoshimura and Masahiro Obayashi

Department of Forensic Medicine, Tohoku University Graduate School of Medicine

Yohko Nagura, Tsukasa Ohuchi and Masato Funayama

Department of Legal Medicine,

Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi

Noboru Adachi

ありがとうございました Grazie Thank you

"Speaker was provided travel and hotel support by Thermo Fisher Scientific for this presentation, but no remuneration

When used for purposes other than Human Identification or Paternity Testing the instruments and software modules cited are for Research Use Only. Not for use in diagnostic procedures. Thermo Fisher Scientific and its affiliates are not endorsing, recommending, or promoting any use or application of Thermo Fisher Scientific products presented by third parties during this seminar. Information and materials presented or provided by third parties are provided as-is and without warranty of any kind, including regarding intellectual property rights and reported results. Parties presenting images, text and material represent they have the rights"