OFFICIAL JOURNAL

ORIGINAL ARTICLE

Molecular Characterization of Hypervariable Region I and II in Mitochondrial DNA of Nusantara Malays

SharifahNany RahayuKarmilla SyedHassan[1], Ayunni Zainul Bahri[1], Md Salleh Yaapar[2], Reevany Bustami[3], Bin Alwi Zilfalil[1,*]

[1]Human Genome Centre, School of Medical Sciences, Universiti Sains Malaysia, Health Campus, Kubang Kerian, 16150 Kelantan, Malaysia

[2]School of Humanities, Universiti Sains Malaysia, 11800 Penang, Malaysia

[3]Centre of Policy Research and International Studies (CenPRIS), Universiti Sains Malaysia, 11800, Penang, Malaysia

Abstract

Previously, population and genetic studies on the Nusantara Malays are not widely known and discovered despite their dispersal pattern. The Malay-Archipelago, also known as Nusantara, is a part of the Austronesian (Malayo-Polynesian) linguistic family. The aim of this study is to identify the mtDNA profile for the Nusantara Malays. Blood samples were collected from maternally unrelated individuals (n=26) from five sub-ethnic groups of Nusantara Malays residing in Peninsular Malaysia were analysed to establish the sequence polymorphism and genetic diversity. MtDNA parameters were estimated, and the observed polymorphisms with their respective haplogroups in comparison to rCRS were inferred, respectively. The mtDNA sequences of HVI and HVII in 26 unrelated Nusantara Malay individuals were evaluated. Molecular diversity indices and differentiation tests were computed. HVI haplotypes were unique for each Nusantara Malay sample while three HVII haplotypes were shared. All Nusantara Malay individuals were grouped in haplogroup H. This data will enhance the DNA database of the Malay population, which can be used for elucidating the history of Nusantara Malays expansion in Peninsular Malaysia.

Keywords: Mitochondrial DNA, Nusantara Malay, sequence polymorphism, genetic diversity, mtDNA haplogroup

Introduction

Nusantara is the Indonesian/Malay name of Maritime Southeast Asia. It is derived from the word *Nusa* (nation or island) and *Antara* (in between or intermediate) (Anwar, 2016; Zoetmulder *et al.*, 1982). It includes countries that have similar culture and linguistic influence including Peninsular Malaysia, Borneo, the Indonesian archipelago, Singapore, Cambodia, Brunei and the southern part of the Philippines and Thailand (Anas *et al.*, 2019). In fact, the geopolitical coverage is almost identical to the Southeast region of Asia (Anwar, 2016).

Received: 10 November 2020**; accepted revised manuscript:** 23 December 2020 **Published online:** 31 December 2020

***Corresponding author:** Prof. Dr. Zilfalil Bin Alwi, Human Genome Centre, School of Medical Sciences, Universiti Sains Malaysia, Health Campus, Kubang Kerian, 16150 Kelantan, Malaysia

Tel: +609767 6531 **email**: zilfalil@gmail.com

The word 'Nusantara' deeply conveys historical, cultural and sentimental values of the natives and fertilizes a form of a sense of belonging that is often pronounced as 'a compatriot kin' (*bangsa serumpun*) based on their common ancestors, spoken language, beliefs, way of life, and skin colours, besides their domiciles (Anwar, 2016). The language spoken in Nusantara falls under the Nusantara cluster called Malayo-Polynesian. Malayo-Polynesian is a linguistic subgroup of the Austronesian family that is widely dispersed throughout Maritime Southeast Asia, Madagascar and the islands of the Pacific Ocean, as well as a few regions on continental Asia (Ross, 1996).

The ownership of Nusantara ("Archipelago") culture has been disputed by Malaysia and Indonesia as both occupy the same archipelago and have overlapping history and language (Kremer, 2011). The governments of both

countries have used Nusantara heritage as the majority of ethnic Malays of Malaysia, and the Indonesians are culturally and ethnically quite similar (Kremer, 2011). Malaysia and Indonesia have been locked in a contentious *adik-abang* ("younger brother-older brother") relationship (Kremer, 2011). Malaysia has become the so-called older brother of the *adik-abang* relationship, because of the history of the human migration process from Indonesia to Malaysia which has prevailed since pre-colonial times (Kremer, 2011). Furthermore, the concept of "Nusantara" could perhaps be a measure of human migration due to various factors such as economics, political, demographic, environmental, and similarities of social measures such as language, culture, and ethnicity (Bustami *et al.*, 2016).

Various genetic markers are used in population genetics and evolutionary studies, i.e., Autosomal Short Tandem Repeat (STR), Y chromosome and mitochondrial DNA (mtDNA). These lineage markers have their own criteria in explaining the molecular anthropology. Mitochondrial DNA are found in hundreds to thousands cellular organelle named mitochondria, which is known as the 'power cells' (White *et al.*, 2008). Each mitochondrion harboured 2-10 mtDNA in circular, double stranded form, having ~16.5 kb sequence. The mtDNA genetic variation shows matrilineal inheritance unique features, lack of recombination, more variable compared to the nuclear genome with higher rate of mutation and specific hypervariability within the D-loop and therefore is a widely used tool (Kivisild, 2015). The D-loop comprises of three short regions with high variability at population level: hypervariable sequence (HV): HVI (nucleotide position (np) 16024 to 16365), HVII (np 73 to 340), and HVIII (np 438 to 576) (Stoneking, 2000).

Previously, population and genetic studies on the Nusantara Malays have not been undertaken despite of their dispersal pattern. This present study aims to utilize the mtDNA hypervariable regions for dissecting the genetic profile of the Nusantara Malays as it allows tracing of a direct genetic line. To the best of our knowledge, there has been no study reported on the mtDNA of the Nusantara Malays and the relationship among them. In this present data, variations of mitochondrial DNA hypervariable region I (HVI) and hypervariable region II (HVII) were analysed in a total of five sub-ethnic groups of Nusantara Malays residing in Peninsular Malaysia to establish the sequence polymorphism and genetic diversity among the Nusantara Malays population.

Nusantara Malays are the descendants of Austronesian speakers which forms part of the greater 'Austronesian Diaspora' into Peninsular Malaysia (Chambers and Edinur, 2013). Peninsular Malaysia was once a very strategic trading centre, connecting Indochina and the Indonesian archipelago (Hatin *et al.*, 2011). Previously, Nusantara Malays from various regions have migrated and assimilated with the native Malay population in the Malay Peninsula. Even though the Nusantara subjects selected in this study were recruited from Peninsular Malaysia, they represent their places of origin from various locations in South East Asia, as indicated in the name of their Malay sub-groups. These selected subjects would carry the genetic information of the Nusantara Malays in view of their previous modern human migration pattern into the Malay Peninsula which is the natural focus of the whole region.

Aceh Malays are people who originated from the northernmost region of Sumatra once ruled by the Sultanate of Aceh Darussalam (1496-1903) (Din, 2011). Bugis Malays are people who originated from the coastal area in the southwestern region of Sulawesi Island (NurWaliyuddin, 2018). Jawa Malays are people who originated from the central and eastern regions of Java island of Indonesia. They formed the largest ethnic group in Indonesia and migrated to the Malay Peninsula mostly during the first decade of Malaysia's Independence (Mas'ud, 2008). Kedah Malays are people originated from the core region of the Kedah Tua kingdom, which became an important trading centre that provided an alternative inland route via Kedah-Yarang trans-peninsula route which connected the Bay of Bengal and Gulf of Thailand (Ramli and Rahman, 2012; Samsudin *et al.*, 2010). Through trading with the people from Arabia and India, Kedah Malays were also influenced by the Arab-Muslim and Indian civilizations. Kelantan Malays are people who reside in the north eastern region of the Malay Peninsula and were influenced by the traders mainly from China, India as well as the Middle East (Hussin, 2004).

Materials and Methods

Sample collection

The approval to conduct this study was obtained from the Research Ethics Committee (Human), Universiti Sains Malaysia (JEPeM Code: USM/JEPeM/17020073). Subjects were selected from five sub-ethnic groups of Nusantara Malays based on very strict criteria to establish the authenticity of their ancestral lineages (must be at least three generations of the respective sub-

 | ISSN (Online): 2716-649X

ethnic groups, no history of mixed-marriage and must have resided in the respective location/village for at least three generations). The 26 healthy subjects were from Acheh, Bugis, Kedah, Kelantan and Jawa sub-ethnic groups (Table 1). Subjects were chosen based on the Nusantara cluster and linguistic influence which is Malayo-Polynesian and known family history.

Table 1: Description of the sample collection for Nusantara Malays.

Sub-ethnic group in Nusantara Malays	City/Town/Region
Acheh	Kampung Acheh, Yan, Kedah, Malaysia
Bugis	Kampung Parit Tengah, Benut, Johor, Malaysia
Kedah	Kampung Hilir, Merbok, Kedah, Malaysia
Kelantan	Kampung Tok Uban, Pasir Mas, Kelantan, Malaysia
Jawa	Kampung Lubok, Semerah, Johor, Malaysia

DNA extraction and quantification

Genomic DNA was extracted from participants' whole blood sample using a commercially available kit, QIAamp® DNA Mini Kit, according to the manufacturer's guidelines (QIAGEN Ag., Germany). The extracted DNA samples were analysed for DNA integrity using 2% agarose gel electrophoresis. The purity and concentration of DNA were determined by using the Infinite 200 NanoQuant machine (TECAN, Magellan, Austria, GmBh) with the value of pure DNA was A260/A280 ratio of 1.8 to 2.0.

PCR amplification of the mtDNA hypervariable region

Two PCR primers were used to amplify the position 15876–639 of mtDNA (1332 bp): 15876F (5'-TCAAATGGGCCTGTCCTTGTAG-3') and 639R (5'-GGGTGATGTGAGCCCGTCTA-3') (Integrated DNA Technologies, Inc., USA). PCR amplification was performed in a 50 µL reaction volume using 1.0U of FastStart™ *Taq* DNA Polymerase (5 U/µL) (Sigma-Aldrich, Inc., Germany). Each PCR reaction contained 1X of PCR buffer with $MgCl_2$ (10X), 0.1mM of dNTPs (2.5 mM), 0.4 µM of each primer (10 µM) and ~ 100 ng/uL of the DNA template. Thermal cycling was performed using a GeneAmp® PCR System 9700 (Applied Biosystems, USA) starting with 5 min at 95 °C followed by 35 cycles of 30 s at 95 °C, 30 s at 55 °C, and 2 min at 72 °C; the final extension was performed at 72 °C for 10 min. The PCR product was analysed by electrophoresis on a 2% agarose gel and visualized by SYBR® Green I (Life Technologies, USA) with staining under a UV transilluminator (Alpha Innotech, USA).

Nucleotide sequencing of the mtDNA hypervariable region

Products were purified via enzymatic digestion prior to and used as a template for cycle sequencing. The reagents for post-PCR purification include 1.4 µL of Shrimp Alkaline Phosphatase (1U/µL), 0.2 µL Exonuclease 1 (20U/µl), 2.4 µL ddH_2O and 10 µL of PCR product which were incubated in the thermal cycler at 37 °C for 1 hour and the enzyme was inactivated at 80 °C for 20 min. Two different primers were used in cycle sequencing reaction: (1) 15946F (5'-CAAGGACAAATCAGAGAAAA-3') and (2) H408 (5'-CTGTTAAAAGTGCATACCGCCA-3'). The purified DNA sample was sent for sequencing by 1st BASE DNA Sequencing Services using a Big-Dye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, USA).

Data analysis

The interpretation of the HVI and HVII chromatogram was made as per the guidelines to improve the quality of the data (Carracedo *et al.*, 2000; Holland and Parsons, 1999; SWGDAM, 2003; SWGDAM, 2013) using DNA sequencing softwares i.e., BioEdit version 7.2.5 (Hall, 1999) and Chromas version 2.6.6 (https://technelysium.com.au/wp/chromas/).

The sequences were matched and aligned with the revised Cambridge reference sequences (rCRS) (Andrews *et al.*, 1999) by using Mega 7 (Kumar *et al.*, 2016). The coding for heteroplasmic sites was done according to the IUPAC codes in the interpretation guideline to interpret the mtDNA data analysis (SWGDAM, 2003). The haplotypes, haplotype diversity (HD), nucleotide diversity, and other molecular diversity

indices were determined using Arlequin software version 3.5.1.2 (Excoffier and Lischer, 2010).

Results

The haplotypes and frequency distributions of each HVI and HVII among the Nusantara Malays population were highly diverse. In this study, HVI was entirely unique for each Nusantara Malays while there is a slight discrepancy in HVII as three haplotypes were shared among them. Out of the 26 Nusantara Malays, 22 and 19 different haplotypes were observed for HVI and HVII, respectively. The highest frequency of the HVII haplotype was shared among three Nusantara Malays while another two HVII haplotypes were shared among two Nusantara Malays. No haplotype was shared when comparing the HVI within the Nusantara Malays.

Molecular diversity indices and differentiation tests were computed. The gene diversity was calculated according to Tajima (Tajima, 1989). Nucleotide diversity, haplotype diversity, mean pairwise difference, a number of haplotypes, mismatch distributions, Harpending's raggedness index, Ewens-Watterson test, Chakraborty's test, Tajima's D test and Fu's Fs statistics were calculated by using Arlequin software version 3.5.1.2 (Excoffier and Lischer, 2010) as shown in Table 2.

Table 2: Molecular diversity indices for the HVI region and HVII region among Nusantara Malays population.

Diversity indices	HVI region	HVII region
No. of polymorphic sites	61	25
No. of observed transitions	46	16
No. of observed transversions	17	3
No. of observed substitutions	63	19
No. of observed indels	2	8
Nucleotide composition (%)		
C	33.52	28.91
T	22.08	26.06
A	33.36	29.01
G	11.04	16.02
Mean number of pairwise differences	8.843077 ± 4.213411	4.132308 ± 2.124881
Heterozygosity/sample	0.02578 ± 0.06659	0.01514 ± 0.05857
No. of haplotypes	26	22
Gene diversity	1.0000 ± 0.0107	0.9846 ± 0.0160
Nucleotide diversity	0.025782 ± 0.013685	0.015137 ± 0.008671
Theta (Hom) ± S.D	N.A.	62.121889 ± 67.640975
Theta (k) [95 % confidence interval limits]	N.A.	64.712648 [26.580908, 170.646968]
Theta (S) ± S.D	15.985500 ± 5.399331	4.979090 ± 1.907217
Theta (π) ± S.D	8.843077 ± 4.693795	4.132308 ± 2.367145
Alleles frequency (Mean ± S.D)	0.14037 ± 0.09008	0.16529 ± 0.11414
Sum of square deviation	0.00658446 ± 0.34	0.02445394 ± 0.04
Harpending's raggedness index	0.01056568 ± 0.82	0.01777041 ± 0.95
Mismatch distribution observed mean	8.843	4.132
Mismatch observed variance	34.756	5.448
Ewens-Watterson test / Slatkin's exact test P-value	N.A.	0.91300
Chakraborty's test	N.A.	8.66101
Tajima's D test	-1.77611	-1.46831
Fu's FS test	-21.06227	-18.84279

A total of 61 and 25 polymorphic sites were observed in HVI and HVII, respectively. Haplotypes for HVI in Nusantara Malays were highly discrete compared to HVII despite more polymorphic sites that were detected in HVI as compared to HVII. The observed number of segregating sites (Theta (S)) and the mean number of pairwise differences (Theta (π)) for HVI were greater than HVII as the gene, and nucleotide diversity was greater due to the uniqueness of the haplotypes. Neutrality tests of Tajima's D and Fu's Fs statistics were carried out to examine the population history and natural selection of the Nusantara Malays. The deviation from neutrality was estimated based on the expectation of constant population size at mutation-drift equilibrium. Tajima's D test measures the allele frequency distribution of nucleotide sequence data (Joshi *et al.*, 2013). Negative values of Tajima's D test showed that the Nusantara Malays as a purifying selection or, alternately, as a population size expansion (e.g.,

after a bottleneck or a selective sweep) (Tajima, 1989).

The nucleotide sequences of HVI (nucleotide position 16024–16365) and HVII (nucleotide position 73–340) in mtDNA were sequenced in both 5' and 3' directions. Table 3 lists the variable nucleotide sites and state frequencies of HVI and HVII observed in the Nusantara Malays population. Three states of polymorphism were detected in six nucleotide positions in total, two sites in HVI and four sites in HVII. The results show that nucleotide variations at positions 16262, 16363, 16365, 153 and 312 were highly polymorphic. Addition at nucleotide position (np) 16312, 16365, 152, 262, 283 and 319 while deletion at np 250 and 318 was observed in Nusantara Malay samples.

Haplogroups classification and a phylogenetic tree was performed by using HaploGrep2 software (Weissensteiner *et al.*, 2016) where the haplogroup data were compatible with PhyloTree Build 17 (van Oven, 2015) as shown in Table 4 and Figure 1. As shown in the results, all Nusantara Malays were distributed in haplogroup H.

Table 3. Polymorphic allele database contents of HVI and HVII observed in the Nusantara Malays.

Hypervariable region	Nucleotide position	State Frequencies		
I	16052	A : 0.9615	G : 0.0385	
	16087	T : 0.8846	C : 0.1154	
	16093	T : 0.9615	C : 0.0385	
	16094	T : 0.9231	C : 0.0769	
	16109	C : 0.9615	T : 0.0385	
	16125	T : 0.9615	C : 0.0385	
	16127	T : 0.9615	C : 0.0385	
	16130	G : 0.7308	A : 0.2692	
	16146	G : 0.9231	A : 0.0769	
	16149	C : 0.9615	T : 0.0385	
	16163	A : 0.9615	G : 0.0385	
	16170	C : 0.9615	T : 0.0385	
	16173	T : 0.8462	C : 0.1538	
	16182	A : 0.8846	G : 0.1154	
	16183	A : 0.9231	C : 0.0769	
	16184	A : 0.9231	C : 0.0769	
	16190	T : 0.8846	C : 0.1154	
	16192	C : 0.9615	A : 0.0385	
	16193	C : 0.9615	T : 0.0385	
	16195	A : 0.9231	C : 0.0769	
	16205	G : 0.9231	A : 0.0769	
	16210	T : 0.9615	C : 0.0385	
	16215	C : 0.9615	A : 0.0385	
	16218	T : 0.9231	C : 0.0769	
	16220	A : 0.9615	C : 0.0385	
	16224	T : 0.6154	C : 0.3846	
	16225	T : 0.9231	C : 0.0769	
	16232	T : 0.9615	C : 0.0385	
	16235	C : 0.9615	T : 0.0385	
	16246	C : 0.9231	G : 0.0769	
	16256	G : 0.9231	A : 0.0769	
	16257	C : 0.9231	T : 0.0769	
	16259	A : 0.9231	C : 0.0769	
	16260	C : 0.9615	T : 0.0385	
	16261	C : 0.9231	A : 0.0385	T : 0.0385
	16262	T : 0.0769	C : 0.9231	
	16266	A : 0.9615	C : 0.0385	
	16268	C : 0.9615	A : 0.0385	
	16270	A : 0.9231	C : 0.0769	
	16272	T : 0.8846	C : 0.1154	

 | ISSN (Online): 2716-649X

	16273	A : 0.9615	G : 0.0385	
	16277	T : 0.9231	A : 0.0769	
	16279	C : 0.9615	T : 0.0385	
	16283	C : 0.9231	A : 0.0769	
	16288	C : 0.9615	T : 0.0385	
	16289	T : 0.9231	C : 0.0769	
	16291	C : 0.9615	T : 0.0385	
	16292	C : 0.8846	T : 0.1154	
	16294	A : 0.9231	C : 0.0769	
	16296	C : 0.9231	A : 0.0385	T : 0.0385
	16298	T : 0.9231	C : 0.0769	
	16299	T : 0.9615	C : 0.0385	
	16305	T : 0.8462	C : 0.1538	
	16312	T : 0.8846	C : 0.1154	
	16312.1	- : 0.9615	T : 0.0385	
	16313	A : 0.9615	G : 0.0385	
	16319	A : 0.9615	T : 0.0385	
	16320	G : 0.9231	A : 0.0769	
	16325	T : 0.9615	C : 0.0385	
	16356	C : 0.9615	T : 0.0385	
	16357	T : 0.9615	C : 0.0385	
	16363	C : 0.3846	T : 0.6154	
	16365.1	- : 0.0769	C : 0.9231	
	73	G : 0.9615	A : 0.0385	
	95	G : 0.9615	A : 0.0385	
	147	T : 0.7692	C : 0.2308	
	151	C : 0.9615	T : 0.0385	
	152.1	- : 0.9615	C : 0.0385	
	153	C : 0.1154	T : 0.8846	
	196	T : 0.9231	C : 0.0769	
	200	T : 0.8077	C : 0.1923	
	208	G : 0.9231	A : 0.0769	
	211	A : 0.9615	G : 0.0385	
	226	G : 0.9615	A : 0.0385	
	242	A : 0.9615	G : 0.0385	
II	250Del	A : 0.8077	- : 0.1923	
	262	C : 0.9231	G : 0.0769	
	262.1	- : 0.8846	G : 0.0769	C : 0.0385
	263	G : 0.8846	- : 0.0385	A : 0.0769
	264	C : 0.9615	G : 0.0385	
	283.1	- : 0.9615	T : 0.0385	
	287	C : 0.9615	T : 0.0385	
	303	A : 0.9615	G : 0.0385	
	312	T : 0.5385	C : 0.4615	
	313	C : 0.9231	- : 0.0385	T : 0.0385
	318Del	C : 0.8846	- : 0.1154	
	319.1	- : 0.9231	A : 0.0385	C : 0.0385
	328	C : 0.9231	G : 0.0769	

Table 4. Sequence polymorphism of the three hypervariable regions and their respective haplogroups in the Nusantara Malays population.

Sample ID	HVI region	HVII region	Haplogroup
Nusantara 1	16223T 16261T 16362C	73G 152C 263G 310.1C	H2a2a
Nusantara 2	16129A 16209C 16223T 16272G	73G 152C 225A 248d 263G 309.1CCT 310C 316A	H2a2a
Nusantara 3	16189C 16223T 16297C	73G 263G 310C	H2a2a1g
Nusantara 4	16086C 16129A 16297C 16324C	73G 199C 263G 310.1C	H2a2a
Nusantara 5	16126C 16231C 16311C	73G 263G 310.1C	H2a2a1
Nusantara 6	16145A 16223T 16291T 16362C	73G 263G 310.1C	H2a2a
Nusantara 7	16223T 16291T 16362C	73G 195C 263G 310C	H2a2a
Nusantara 8	16108T 16129A 16162G 16172C 16304C	73G 248d 263G 310.1C	H2a2a
Nusantara 9	16013.1C 16086C 16148T 16223T 16259T 16278T 16319A	73G 150T 263G 310.1C	H2a2a
Nusantara 10	16223T 16362C	73G 146C 199C 263G 309.1CT 310C	H2a2a
Nusantara 11	16191A 16260A 16267A 16295A 16318T	73G 207A 248d 263G 309.1CT 310C	H2a2a1
Nusantara 12	16093C 16129A 16223T 16256T 16271C 16362C	73G 263G 309.1CT 310C	H2a2a
Nusantara 13	16260T 16298C 16355T 16362C	73G 207A 248d 263G 301G 309.1CT 310C 324G	H2a2a
Nusantara 14	16129A 16172C 16223T 16234T 16290T 16312G	73G 146C 263G 309.1CCT 310C	H2a2a
Nusantara 15	16223T 16271C 16287T 16319A 16356C 16362C	73G 241G 263G 310.1C	H2a2a
Nusantara 16	16223T 16311C	73G 146C 199C 263G 310.1C	H2a2a
Nusantara 17	16223T 16295T 16362C	73G 146C 199C 263G 310.1C	H2a2a
Nusantara 18	16181G 16182C 16183C 16189C 16194C 16204A 16217C 16219C 16224C 16245G 16255A 16258C 16261T 16265C 16269C 16276A 16282A 16288C 16293C	73G 263G 310.1C	H2a2a1g
Nusantara 19	16223T 16304C	73G 199C 262.1G 263G 285T	H2a2a
Nusantara 20	16181G 16182C 16183C 16189C 16194C 16204A 16214A 16217C 16224C 16245G 16255A 16258C 16269C 16276A 16282A 16288C 16293C	73G 146C 263G 309.1CCT 310C 324G	H2a2a1g
Nusantara 21	16145A 16181G 16192T 16223T 16291T 16304C	73G 210G 263G 309.1CT 310C	H2a2a
Nusantara 22	16092C 16223T 16362C	73G 94A 263G 310.1C	H2a2a
Nusantara 23	16011d 16311C	152C 263G 283T	H2a2a
Nusantara 24	16129A 16172C 16304C	73G 248d 263G 310.1C	H2a2a
Nusantara 25	16051G 16086C 16124C 16169T 16172C	73G 146C 152C 195C 263G 309.1CT 310C	H2a2a1d
Nusantara 26	16093C 16129A 16223T 16256T 16271C 16362C	73G 263G 309.1CT 310C	H2a2a

 | ISSN (Online): 2716-649X

Discussions

The number of variations found in the amplified products was quite limited, possibly due to the number of subjects used and the length of the amplified products. The designed primers were one of the significant elements to consider for target amplification. The length and precision of the amplified products were determined by these designed primers. Moreover, they moderately determine the area of amplification in an entire genomic DNA. Negative results from any issues with the specificity of a primer utilized notwithstanding poor optimization of PCR reaction can bring about primers annealing apart from the desired region (Bustin and Huggett, 2017).

In this manner, the advancement of the oligonucleotides must be precise to improve the specificity towards the targeted sequence to be amplified. The mis-binding of primer may be the result of low primer specificity (Ishar *et al.*, 2019). An efficient primer would have the correct sequence to compliment the template (Ye *et al.*, 2012). Furthermore, it is essential to run BLAST sequencing. DNA degradation during transportation or storage may also result in such errors in which can be excluded by gel electrophoresis (Ishar *et al.*, 2019).

However, this study might show some limitations due to the limited sample size and sampling location. Therefore, molecular diversity indices and differentiation tests computed in this study might show genetic drift in the two mtDNA hypervariable regions. Furthermore, the observed homozygosity H (Theta (Hom)) and observed number of alleles (Theta (k)) were unable to be computed as all HVI gene copies are different.

This study reveals that all Nusantara Malay individuals were grouped in haplogroup H. Interestingly, the European mtDNA lineages H were observed in Nusantara Malays. Haplogroup H is believed to have originated in southwest Asia, around 20,000 to 25,000 years ago (Metspalu *et al.*, 2004). Haplogroup H might have been introduced via recent and historical admixture with European and Indian populations.

Haplogroup H displays two unique features i.e A very high frequency in most of its range and an extremely wide geographic distribution (Achilli *et al.*, 2004), and this could explain its presence in Southeast Asia, i.e., in the Indian population and Cape Malay Muslims in Cape Town Metropolitan area (Isaacs *et al.*, 2013; Thangaraj *et al.*, 2008).

The presence of haplogroup H in the Nusantara Malays in this study might corroborate the relationship of the human migration pattern. The spreading theory of the Nusantara Malays could be proven with other research and comparison studies across the South East Asia region and population. The genetic data provided would also be linked with the cultural, historical and linguistic information to strengthen the Nusantara Malays diaspora.

Interestingly, the Nusantara Malays in this study were not distributed in haplogroup M such as revealed by most of the studies for Malay population in Southeast Asia especially those sampled within Peninsular Malaysia (Eng, 2014; Macaulay *et al.*, 2005; Rashid *et al.*, 2009; Zainuddin, 2004). The presence of haplogroup H in Nusantara Malays were almost similar as the Cape Malay Muslim in Cape Town, South Africa (Isaacs *et al.*, 2013). Certain admixture i.e., geographical marriage or ethnic preferred marriages, in previous generations and ancestors for the maternal line might be the main contributor in the mtDNA haplogroup distribution. The results of this study showed a resemblance of Cape Malays with the Nusantara Malays in the contribution of maternal lineage haplogroups inducing the origin of Cape Malays based on the genetic affiliation of the Malay population in the Malayo-Polynesian region. Parallel resemblance of mitochondrial DNA haplogroup H provided some insight into the genetic ancestry of both Nusantara and Cape Malays population. Thus, it should be interesting to further investigate the contribution of other Nusantara Malay using PCR-RFLP protocols.

 | ISSN (Online): 2716-649X

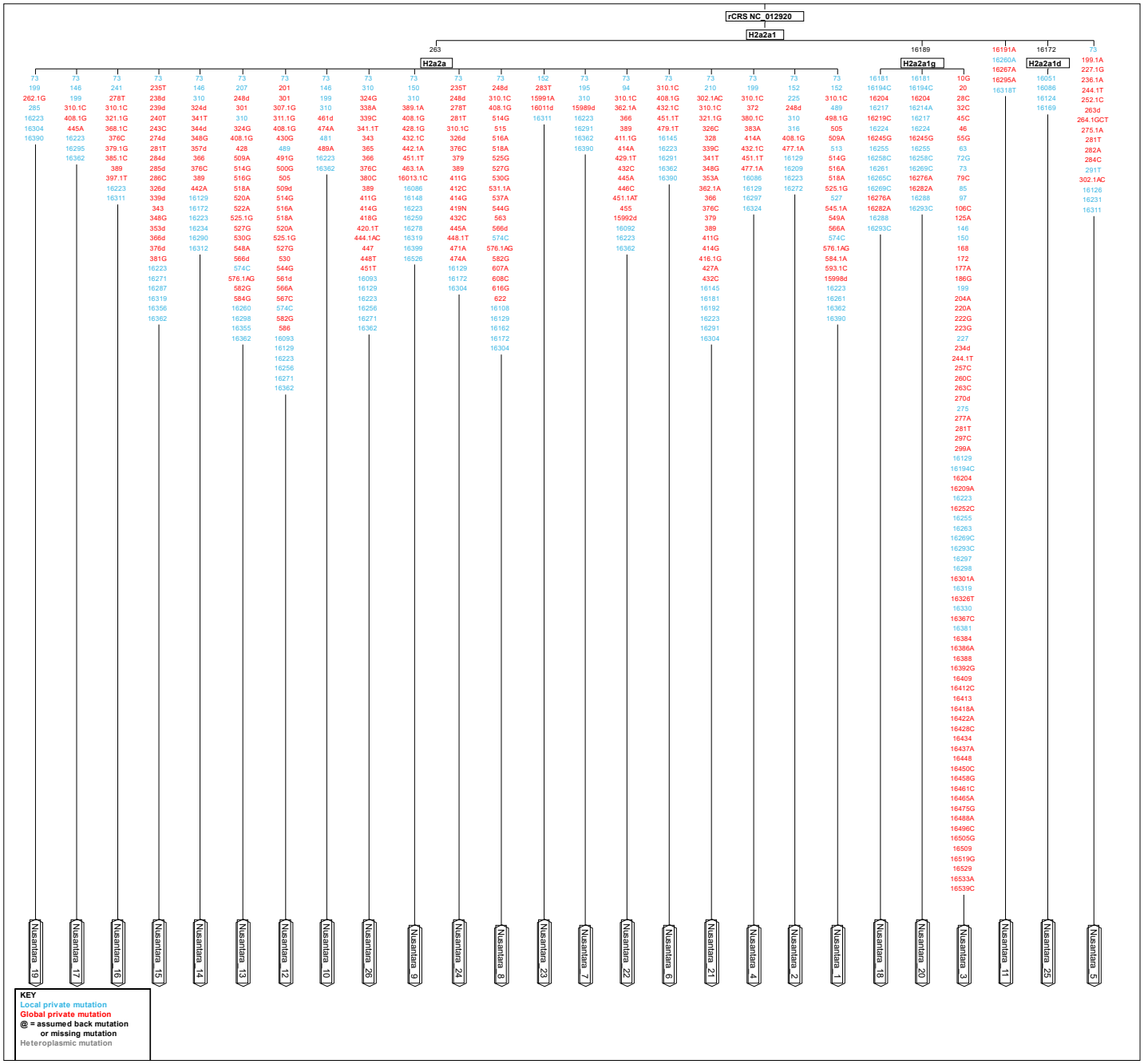

Figure 1. Phylogenetic tree of the Nusantara Malays.

Conclusion

In conclusion, mtDNA sequences of HVI and HVII in 26 unrelated Nusantara Malay individuals were evaluated. HVI haplotypes were unique for each Nusantara Malay sample while three HVII haplotypes were shared. The results show that nucleotide variations at positions 16262, 16363, 16365, 153 and 312 were highly polymorphic. Further analysis of additional control and coding region mtDNA markers and increased sample size may provide an extensive view of the population structure of Nusantara Malays. In spite of that, the mtDNA control region analyzed here can improve the existing database at the ethnic level as well as deliver valued data for forensic and population genetic epidemiology of the Malayo-Polynesian population. The present data will enhance the DNA database of the Malay population, which can be used for elucidating the history of Nusantara Malays expansion in Peninsular Malaysia and calculating probabilities of the match based on the control region of the mtDNA. This report could also be used by evolutionary biologists to study genetic variations in order to understand the possible relationships of the Nusantara Malays population with other populations and the mtDNA haplogroups generated in this data report can be used for tracing the migration and ancestry of the Nusantara Malays population.

Acknowledgements

The research team would like to thank the individuals, organizations and participants who that took part in the study and enabled this research to be possible. This work was supported by the Universiti Sains Malaysia: Research University Individual (RUI) Grant: 1001/PPSP/8012301 and Bridging grant: 304/PPSP/6316166; and Nanyang Technological University Grant (Muhammed Ariff Research Grant (MAS)): 304.PPSP.6150148.N119. All authors contributed to aspects of the conception or design of the experimental work, analysis of data, reviewing and approving the final content of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Scozzari, R., Cruciani, F., Zeviani, M., Briem, E., Carelli, V., Moral, P., Dugoujon, J.-M., Roostalu, U., Loogväli, E.-L., Kivisild, T., Bandelt, H.-J., Richards, M., Villems, R., Santachiara-Benerecetti, A. S., Semino, O. & Torroni, A. (2004). The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *American journal of human genetics,* **75(5)**, 910-918. doi: 10.1086/425590

Anas, N., Alwi, E. A. Z. E., Yaacob, Z., Ramlan, A. N. M., Ishak, A. H. & Manan, S. K. A. (2019). The Nusantara Ethnic Groups From Historical and Genomic Perspectives. *International Journal of Civil Engineering and Technology (I,* **10(4)**, 287-301.

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M. & Howell, D. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics,* **23**, 147.

Anwar, O. M. (2016). Contextualizing Nusantara Studies. *Journal of Nusantara Studies,* **1(1)**, 1-6.

Bustami, R., Suyatno & Wekke, I. S. (2016). Migrants In Nusantara: Indonesian And Malaysian Perspectives. *Journal of Nusantara Studies* **1(2)**, 63-74.

Bustin, S. & Huggett, J. (2017). qPCR primer design revisited. *Biomolecular detection and quantification,* **14**, 19-28. doi: 10.1016/j.bdq.2017.11.001

Carracedo, A., Bär, W., Lincoln, P. J., Mayr, W., Morling, N. & Olaisen, B. (2000). DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. *Forensic Sci. Int.,* **110**, 79-85.

Chambers, G. K. & Edinur, H. A. (2013). *Genetic relationships between Malays and Maori*. Wellington, New Zealand: Chair of Malay Studies & Victoria University Foundation.

Din, M. A. O. (2011). Asal-Usul Orang Melayu: Menulis Semula Sejarahnya (The Malay Origin: Rewrite Its History). *Jurnal Melayu,* **7**, 1-82.

Eng, K. K. (2014). *Complete mitochondrial DNA genome variation in Peninsular Malaysia.* (Dissertation). University of Leeds.pages 334

Excoffier, L. & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour,* **10**. doi: 10.1111/j.1755-0998.2010.02847.x

Hall, T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.,* **41**, 95-98.

Hatin, W. I., Nur-Shafawati, A. R., Zahri, M.-K., Xu, S., Jin, L., Tan, S.-G., Rizman-Idid, M., Zilfalil, B. A. & Consortium, T. H. P.-A. S. (2011). Population Genetic Structure of Peninsular Malaysia Malay Sub-Ethnic Groups. *PLoS ONE,* **6(4)**. doi: 10.1371/journal.pone.0018312.t001

Holland, M. M. & Parsons, T. J. (1999). Mitochondrial DNA sequence analysis - validation and use for forensic casework. *Forensic Sci. Rev.,* **11**, 21-50.

Hussin, H. (2004). Orang Melayu Kelantan Dari Sudut Budaya Dan Perspektif Sejarah Lampau. *Jebat,* **31**, 15–38.

Isaacs, S., Geduld-Ullah, T. & Benjeddou, M. (2013). Reconstruction of major maternal and paternal lineages of the Cape Muslim population. *Genetics and molecular biology,* **36(2)**, 167-176. doi: 10.1590/S1415-47572013005000019

Ishar, S. M., Parameswaran, K., Masduki, N. S. & Rus Din, R. D. (2019). Variations in mitochondrial DNA control region among Malay and Chinese subpopulations (sequence 16000–16200). *Mitochondrial DNA Part A,* **30(8)**, 843-847. doi: 10.1080/24701394.2019.1687693

Joshi, J., Salar, R. K., Banerjee, P., S, U., Tantia, M. S. & Vijh, R. K. (2013). Genetic variation and phylogenetic relationships of Indian buffaloes of uttar pradesh. *Asian-Australasian journal of animal sciences,* **26(9)**, 1229-1236. doi: 10.5713/ajas.2012.12669

Kivisild, T. (2015). Maternal ancestry and population history from whole mitochondrial genomes. *Investigative Genetics,* **6**, 3. doi: 10.1186/s13323-015-0022-2

Kremer, C. (2011). Malaysia and Nusantara Heritage in the Adik-Abang Relationship. *Stanford Journal of International Relations,* **13(2)**, 28-29.

Kumar, S., Stecher, G. & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution,* **33(7)**, 1870-1874. doi: 10.1093/molbev/msw054

Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., Taha, A., Shaari, N. K., Raja, J. M., Ismail, P., Zainuddin, Z., Goodwin, W., Bulbeck, D., Bandelt, H.-J., Oppenheimer, S., Torroni, A. & Richards, M. (2005). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science,* **308(5724)**, 1034.

Mas'ud, R. N. (2008). *Peran Demak Terhadap Runtuhnya Majapahit (1478-1527 M)*: UIN Sunan Kalijaga Yogyakarta.

Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., Serk, P., Karmin, M., Behar, D. M., Gilbert, M. T. P., Endicott, P., Mastana, S., Papiha, S. S., Skorecki, K., Torroni, A. & Villems, R. (2004). Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genetics,* **5**, 26-26. doi: 10.1186/1471-2156-5-26

NurWaliyuddin, H. Z. A. (2018). *Understanding the genetic history of the Malay and Orang Asli populations in Peninsular Malaysia by Killer Cell Immunoglobulin-Like Receptors (KIR).* (Dissertation). Universiti Sains Malaysia, Kelantan, Malaysia

Ramli, Z. & Rahman, N. H. S. N. A. (2012). Hubungan Antara Semenanjung Tanah Melayu Dengan China Sejak Awal Abad Masihi. *International Journal Of The Malay World And Civilisation,* **30(1)**, 171–196.

Rashid, N. H. A., Panneerchelvam, S. & Zainuddin, Z. (2009). Analysis of mitochondrial DNA (mtDNA) polymorphism in Kelantan Malays. *Proceedings of the 8th Malaysia Congress on Genetics,* **OB06**, 139-143.

Ross, M. (1996). On the Origin of the Term "Malayo-Polynesian". *Oceanic Linguistics,* **35(1)**, 143-145. doi: 10.2307/3623036

Samsudin, A., Masron, T., Saidin, M., Abdullah, J. & Shahidan, S. (2010). *Lembah Bujang, Sebuah Kerajaan Lama Yang Berpusat: Dari Perspektif Sistem Maklumat Geografi (GIS)*: Penanda Arasan Penyelidikan Arkeologi Di UKM.

Stoneking, M. (2000). Hypervariable Sites in the mtDNA Control Region Are Mutational Hotspots. *Am. J. Hum. Genet.,* **67**, 1029-1032.

 | ISSN (Online): 2716-649X

SWGDAM (2003). Guidelines for mitochondrial DNA (mtDNA) nucleotide sequence interpretation. *Forensic Sci. Commun.,* **5(2)**.

SWGDAM (2013). Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics,* **123**, 585-595.

Thangaraj, K., Chaubey, G., Kivisild, T., Selvi Rani, D., Singh, V. K., Ismail, T., Carvalho-Silva, D., Metspalu, M., Bhaskar, L. V. K. S., Reddy, A. G., Chandra, S., Pande, V., Prathap Naidu, B., Adarsh, N., Verma, A., Jyothi, I. A., Mallick, C. B., Shrivastava, N., Devasena, R., Kumari, B., Singh, A. K., Dwivedi, S. K. D., Singh, S., Rao, G., Gupta, P., Sonvane, V., Kumari, K., Basha, A., Bhargavi, K. R., Lalremruata, A., Gupta, A. K., Kaur, G., Reddy, K. K., Rao, A. P., Villems, R., Tyler-Smith, C. & Singh, L. (2008). Maternal Footprints of Southeast Asians in North India. *Human Heredity,* **66(1)**, 1-9. doi: 10.1159/000114160

van Oven, M. (2015). PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Science International: Genetics Supplement Series,* **5**, e392-e394. doi: https://doi.org/10.1016/j.fsigss.2015.09.155

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A. & Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Research,* **44(W1)**, W58-W63. doi: 10.1093/nar/gkw233

White, D. J., Wolff, J. N., Pierson, M. & Gemmell, N. J. (2008). Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology,* **17(23)**, 4925-4942. doi: 10.1111/j.1365-294X.2008.03982.x

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. & Madden, T. L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics,* **13(1)**, 134. doi: 10.1186/1471-2105-13-134

Zainuddin, Z. (2004). *The analysis of human mitochondrial DNA in peninsular Malaysia.* (Dissertation). University of Glasgow

Zoetmulder, P. J., Robson, S. O. & Koninklijk Instituut voor Taal-, L.-e. V. N. (1982). *Old Javanese-English dictionary*. s-Gravenhage: M. Nijhoff.

 | ISSN (Online): 2716-649X