

# Sequence polymorphisms of mtDNA HV1, HV2, and HV3 regions in the Malay population of Peninsular Malaysia

Abd Rashid Nur Haslindawaty · Sundararajulu Panneerchelvam ·  
Hisham Atan Edinur · Mohd Nor Norazmi · Zainuddin Zafarina

Received: 2 December 2009 / Accepted: 5 May 2010 / Published online: 26 May 2010  
© Springer-Verlag 2010

**Abstract** The uniparentally inherited mitochondrial DNA (mtDNA) is in the limelight for the past two decades, in studies relating to demographic history of mankind and in forensic kinship testing. In this study, human mtDNA hypervariable segments 1, 2, and 3 (HV1, HV2, and HV3) were analyzed in 248 unrelated Malay individuals in Peninsular Malaysia. Combined analyses of HV1, HV2, and HV3 revealed a total of 180 mtDNA haplotypes with 149 unique haplotypes and 31 haplotypes occurring in more than one individual. The genetic diversity was estimated to be 99.47%, and the probability of any two individuals sharing the same mtDNA haplotype was 0.93%. The most frequent mtDNA haplotype (73, 146, 150, 195, 263, 315.1C, 16140, 16182C, 16183C, 16189, 16217, 16274, and 16335) was shared by 11 (4.44%) individuals. The nucleotide diversity and mean of pair-wise differences were found to be  $0.036063 \pm 0.020101$  and  $12.544022 \pm 6.230486$ , respectively.

**Keywords** Mitochondrial DNA · Hypervariable regions · Malay population

## Introduction

The entire human mtDNA genome was first sequenced in 1981 [1] and subsequently revised with a few changes in

base composition later known as the revised Cambridge Reference Sequence (rCRS) [1, 2]. The maternal inheritance, high copy number per cell, high rate of mutation, and lack of recombination have made mtDNA a valuable tool in forensic identification [3]. The mtDNA control region of approximately 1.1 kb consists of three hypervariable regions (HV1, HV2, and HV3) and was found to be highly polymorphic in humans providing a high degree of discrimination between unrelated individuals.

Peninsular Malaysia or West Malaysia is situated in the southeastern tip of the Asian mainland, bordering Thailand at the north and separated from East Malaysia (Sabah and Sarawak) by the South China Sea [4]. The Malays represent about 50.4% of the total Malaysian population and regarded as one homogenous group [5], although we have shown that they can be quite different in their genetic makeup [6]. Mitochondrial DNA data on Malay population is so far very scanty. Zafarina and Goodwin [5] have reported only the HV1 data for the Modern Malays in Peninsular Malaysia, while Wong et al. [7] have published the HV1 and HV2 data for the Malays living in Singapore. Recently, Maruyama et al. [8] have published the HV1, HV2, and HV3 mtDNA database for Malay individuals living in Kuala Lumpur. In this study, we report a comprehensive mtDNA data with higher number of samples than the previously reported data as well as taking great care in ascertaining their background for at least three generations without mix-marriage.

## Methods and materials

### Samples

A total of 248 Malay individuals were sampled from several locations in Peninsular Malaysia: 60 from Kelantan,

**Electronic supplementary material** The online version of this article (doi:10.1007/s00414-010-0469-x) contains supplementary material, which is available to authorized users.

A. R. Nur Haslindawaty · S. Panneerchelvam · H. A. Edinur ·  
M. N. Norazmi · Z. Zafarina (✉)  
School of Health Sciences, Universiti Sains Malaysia,  
Health Campus,  
Kubang Kerian, Kelantan 16150, Malaysia  
e-mail: zafarina@kck.usm.my

34 from Negeri Sembilan, 61 from Johor, 56 from Perak, and 37 from Kedah (Fig. 1). To ensure that their parents were of Malay origin, their family history was taken prior to blood collection. Aboriginal populations were not included in this study.

#### Genomic DNA extraction

DNA extraction was performed using QIAquick Blood kit (Qiagen Inc) according to the manufacturer's protocol.

#### PCR amplification of hypervariable region

Three sets of oligonucleotide primer—L15997 (5'-CAC CAT TAG CAC CCA AAG CT-3') and H16410 (5'-GAG GAT GGT GGT CAA GGG AC-3') for HV1, L048 (5'-CTC ACG GGA GCT CTC CAT GC-3') and H408 (5'-CTG TTA AAA GTG CAT ACC GCC A-3') for HV2, and L316 (5'-GCT TCT GGC CAC AGC ACT TA-3') and H619 (5'-GGT GAT GTG AGC CCG TCT AA-3') for HV3 were used to amplify the corresponding regions. Polymerase chain reaction (PCR) amplification was performed in 20  $\mu$ l

reaction mixture consisting 1 $\times$  PCR buffer ( $\text{NH}_4(\text{SO}_4)_2$  and 2.5 mM  $\text{MgCl}_2$ ), 200 mM of each dNTPs, 10 pmol of each primer, and 5 U of Taq polymerase (Invitrogen). The following PCR thermal cycle conditions were performed at 95°C for 3 min, followed by 30 cycles of 95°C for 30 s, 60°C for 30 s, 72°C for 45 s, 72°C for 5 min, and final hold at 4°C.

PCR amplification was conducted on the GeneAmp PCR System 9700 (Applied Biosystem). Amplicon were purified using QIAquick PCR purification kit (Qiagen Inc). PCR products were quantified, and 20 ng was used in each sequencing reaction.

#### Direct sequencing

A total of 10  $\mu$ l sequencing reaction was prepared consisting of 20 ng of purified PCR product, 3.3 pmol of primers, and 1:8 ABI BigDye® Terminator versions 3.1. Amplification was performed on GeneAmp PCR System 9700 (Applied Biosystem). The following thermal cycle condition was used at 96°C for 1 min, followed by 25 cycles of 96°C for 10 s, 50°C for 5 s, 60°C for 4 min, and final hold at 4°C. The sequencing reaction was purified by ethanol precipitation prior to sequencing on ABI the 3130 $\times$ 1 Genetic Analyzer.

#### Analysis of data

Each sample was sequenced in both directions 5' and 3' to avoid ambiguities in sequence determination. Data were analyzed using the ABI Prism Sequencing Analysis software version 5.3.1. The sequence samples were aligned and compared with rCRS [2] using a Vector NTI Advance™ 9.0 (Informax™, MD, USA) to determine the polymorphisms. The C-stretches region in HV1 and HV2 were verified by performing additional sequencing.

The genetic diversity and probability of random match were calculated using the following formulae  $h(1 - \sum X^2)^n / n - 1$  [9] and  $\sum X^2$  [10] respectively, where  $\sum X^2$  is the sum of the square of the haplotype frequencies, and  $n$  is the sample size. Other statistical parameters such as nucleotide diversity and mean number of pair-wise differences were generated by Arlequin version 3.1 [11].

## Results and discussions

The “Malaysian Constitution” refers to a Malay as an individual who speaks the Malay language, professes Islam, and follows the Malay customs [12]. In Malaysia, the majority of Malays reside in Peninsular Malaysia. Previous studies have shown that the “Constitutional” Malay race comprises a population admixture of different populations



**Fig. 1** Map showing specific locations of the Malay samples collected in Peninsular Malaysia. Malaysia (modified from: [commons.wikimedia.org/wiki/File:Malaysia\\_stat](http://commons.wikimedia.org/wiki/File:Malaysia_stat))

**Table 1** Mitochondrial DNA polymorphisms for HV1, HV2, and HV3 and haplogroups in 248 unrelated Malay individuals living in Peninsular Malaysia

Sample	HG	HV1 <sup>a</sup> (15997–16410)	HV2 <sup>b</sup> (048–408)	HV3 (438–619)
101PT	A	126, 223, 290, 319	152, 235	523d, 524d
49SM	B	51, 182C, 183C, 189		CRS
69S	B	51, 183C, 189	195, 309.1C	CRS
52LG	B	51, 183C, 189, 194C, 195	356.1C	CRS
11Y	B	51, 183C, 189, 239	356.1C	CRS
154PB	B4a	147, 183C, 184A, 189, 217, 261	309.1C, 309.2C	CRS
153PB	B4a	178, 182C, 183C, 189, 217, 261	146, 309.1C, 309.2C	523d, 524d
174GP	B4a	182C, 183C, 189, 217, 230R, 261	146, 309.1C	523d, 524d
122BN	B4a	182C, 183C, 189, 217, 261	146	523d, 524d
2M	B4a	182C, 183C, 189, 217, 261	146, 309.1C	523d, 524d
36SM	B4a	182C, 183C, 189, 217, 261	146, 309.1C, 309.2C	523d, 524d
31SM	B4a	182C, 183C, 189, 217, 261	146, 309.1C, 309.2C	523d, 524d
97PT	B4a	182C, 183C, 189, 217, 261	146, 309.1C, 309.2C	523d, 524d
162GP	B4a	182C, 183C, 189, 217, 261	146, 309.1C, 309.2C	523d, 524d
168GP	B4a	182C, 183C, 189, 217, 261	146, 309.1C, 309.2C	523d, 524d
103PT	B4a1a1	182C, 183C, 189, 217, 247, 261	146	523d, 524d
119BN	B4a1a1	182C, 183C, 189, 217, 247, 261	146	523d, 524d
23R	B4a1a1	182C, 183C, 189, 217, 247, 261	146, 309.1C, 309.2C	523d, 524d
135KK	B4b1	93, 136, 182C, 183C, 189, 194C, 195, 217	207	499
3LB	B4c	182C, 183C, 189, 210, 217	146, 309.1C, 309.2C	523d, 524d
156PB	B4c	182C, 183C, 189, 217, 235	146, 309.1C, 309.2C	523d, 524d
26Mb	B4c1	182C, 183C, 189, 311	309.1C, 309.2C	523d, 524d
27Mb	B4c1	182C, 183C, 189, 311	309.1C, 309.2C	523d, 524d
62S	B4c1b3	140, 182C, 183C, 189, 214, 217, 274	146, 150, 152, 195, 309.1C, 309.2C	CRS
30SM	B4c1b3	140, 154, 182C, 183C, 189, 217, 274, 335	146, 150, 195, 309.1C, 309.2C	CRS
94PT	B4c1b3	111G, 140, 182C, 183C, 189, 217, 274, 335	129, 146, 150, 195, 309.1C	CRS
10LB	B4c1b3	140, 181C, 182C, 183C, 189, 217, 274, 335	146, 150, 195, 309.1C, 309.2C	CRS
2R	B4c1b3	140, 182C, 183C, 189, 213, 217, 274, 335	146, 150, 195, 309.1C, 309.2C	CRS
43SM	B4c1b3	140, 182C, 183C, 189, 217, 240, 274, 335	146, 150, 195, 309.1C, 309.2C	CRS
61LG	B4c1b3	140, 182C, 183C, 189, 217, 240, 274, 335	146, 150, 195, 309.1C	CRS
113BN	B4c1b3	140, 182C, 183C, 189, 217, 274, 278, 335	146, 150, 195, 309.1C, 378	CRS
133KK	B4c1b3	140, 182C, 183C, 189, 217, 274, 278, 335	146, 150, 309.1C, 309.2C	CRS
179GP	B4c1b3	140, 182C, 183C, 189, 217, 274, 311, 335	146, 150, 195	CRS
28SM	B4c1b3	140, 182C, 183C, 189, 217, 274, 335	146, 150, 195, 309.1C, 309.2C	CRS
19M	B4c1b3	140, 182C, 183C, 189, 217, 274, 335	146, 150, 195, 309.1C, 309.2C	CRS
33SM	B4c1b3	140, 182C, 183C, 189, 217, 274, 335	146, 150, 195, 309.1C, 309.2C	CRS

Table 1 (continued)

Sample	HG	HV1 <sup>a</sup> (15997–16410)	HV2 <sup>b</sup> (048–408)	HV3 (438–619)
59LG	B4c1b3	140, 182C, 183C, 189, 217, 274, 335	146, 150, 195, 309.1C, 309.2C	CRS
29SM	B4c1b3	140, 182C, 183C, 189, 217, 274, 335	146, 150, 195, 309.1C, 309.2C	CRS
118BN	B4c1b3	140, 182C, 183C, 189, 217, 274, 335	146, 150, 195, 309.1C, 309.2C	CRS
34SM	B4c1b3	140, 182C, 183C, 189, 217, 274, 335	146, 150, 195, 309.1C	CRS
57LG	B4c1b3	140, 182C, 183C, 189, 217, 274, 335	146, 150, 195, 309.1C	CRS
50LG	B4c1b3	140, 182C, 183C, 189, 217, 274, 335	146, 150, 195, 309.1C	CRS
128KK	B4c1b3	140, 182C, 183C, 189, 217, 274, 335	146, 150, 195, 309.1C	CRS
121BN	B4c1b3	140, 182C, 183C, 189, 217, 274, 335	146, 150, 195, 309.1C	489
127KK	B4e2	147, 183C, 184A, 189, 217, 234, 235	309.1C, 309.2C	CRS
126KK	B4e2	147, 183C, 184A, 189, 217, 235	309.1C	CRS
132KK	B4e2	147, 183C, 184A, 189, 217, 235, 261	309.1C	CRS
164GP	B5	140, 183C, 189	210, 309.1C, 309.2C	523d, 524d
14M	B5	129, 172, 183C, 189, 304	152, 210	523d, 524d
3M	B5a	117, 140, 183C, 189, 266A, 293	210	503, 523d, 524d
31R	B5a	140, 167, 183C, 189, 266A	210, 309.1C	523d, 524d
89MR	B5a	140, 169A, 183C, 189, 209, 266A	210	523d, 524d
145PB	B5a	140, 182C, 183C, 189, 214, 261, 266A	152, 199, 210, 309.1C, 309.2C	523d, 524d
87MR	B5a	140, 182C, 183C, 189, 261, 266A	152, 210, 309.1C, 309.2C	523d, 524d
16R	B5a	140, 183C, 189, 266A	210	523d, 524d
142PB	B5a	140, 183C, 189, 266A	210, 309.1C, 309.2C	523d, 524d
13R	B5a	140, 183C, 189, 266A	210, 309.1C	523d, 524d
14R	B5a	140, 183C, 189, 266A	210, 309.1C	523d, 524d
17R	B5a	140, 183C, 189, 266A	210, 309.1C	523d, 524d
134KK	B5a	140, 183C, 189, 266A	210, 309.1C, 324G	523d, 524d
85MR	B5a	140, 183C, 189, 266A, 293	210	523d, 524d
83MR	B5b	67, 140, 183C, 189, 243	103, 152	523d, 524d
159GP	B5b	140, 183C, 189, 243	103, 152	523d, 524d
161GP	B5b	140, 183C, 189, 243	103, 152	523d, 524d
165GP	B5b	140, 183C, 189, 243	103, 152	523d, 524d
14LB	C	86, 223, 242, 256, 298, 327	146, 249d, 297	489
22R	D	75, 93, 129, 223, 240, 291, 362	152, 309.1C	472, 489
20LB	D	93, 223, 309, 362	152, 199	489
21M	D4a3	129, 223, 249, 278, 311, 362	152, 309.1C	489
115BN	E	223, 324, 362, 390	309.1C	489
102PT	E	223, 362, 390	309.1C	489
112PT	E1a1	37, 223, 291, 362, 390	195	489

5R	E1a1	223, 288, 291, 362, 390	146, 150	489, 523d, 524d
107PT	E1a1	223, 291, 362, 390	146, 309.1C	489
92PT	E1a1	223, 291, 362, 390	195, 329	489
95PT	E1a1	223, 291, 362, 390	309.1C	489
100PT	E1a1	223, 291, 362, 390	207, 309.1C	489
105PT	E1a1	223, 291, 362, 390	309.1C, 309.2C	489
123KK	E1a1	223, 291, 362, 390	309.1C, 309.2C	489
116BN	E1a1a	145, 223, 291, 362, 390	207, 309.1C	489
18LB	E1b	93, 223, 261, 362, 390	152, 309.1C	489
63S	E1b	223, 261, 359, 390	152, 309.1C	489
4Y	E1b	223, 261, 362, 390	152	489
167GP	E1b	223, 261, 362, 390	152, 309.1C, 309.2C	489
9M	E1b	223, 261, 362, 390	152, 309.1C	489
19R	E1b	223, 261, 362, 390	152, 309.1C	489
98PT	E2	51, 184, 223, 325, 362, 390	195	489
93PT	E2	51, 184, 223, 325, 362, 390	195	489
114BN	E2	51, 184, 223, 325, 362, 390	195	489
96PT	E2b	51, 185, 223, 362, 390	195	489
24M	F1a	129, 172, 291, 304	249d, 309.1C	523d, 524d
3Y	F1a	129, 172, 293, 304	249d	523d, 524d
56LG	F1a	129, 172, 304	151, 152, 249d, 309.1C	523d, 524d
4M	F1a	129, 172, 304	249d	523d, 524d
176GP	F1a	129, 172, 304	249d	523d, 524d
15M	F1a	129, 172, 304	249d, 309.1C	523d, 524d
18M	F1a	129, 172, 304, 362	249d, 309.1C	523d, 524d
1Y	F1a1	129, 162, 172, 224, 304	249d	CRS
6LB	F1a1	129, 162, 172, 304	249d, 309.1C	523d, 524d, 548
26LB	F1a1a	92, 108, 129, 162, 172, 234, 299, 304	309.1C	523d, 524d
18R	F1a1a	108, 129, 147, 162, 172, 304	249d	523d, 524d
147PB	F1a1a	108, 129, 162, 172, 183C, 189, 234	249d	523d, 524d
25R	F1a1a	108, 129, 162, 172, 189, 304	249d, 309.1C	523d, 524d
8LB	F1a1a	108, 129, 162, 172, 239, 304, 327	249d, 309.1C	523d, 524d
66S	F1a1a	108, 129, 162, 172, 304	249d	523d, 524d
71S	F1a1a	108, 129, 162, 172, 304	249d, 309.1C	523d, 524d
84MR	F1a1a	108, 129, 162, 172, 304	249d, 309.1C	523d, 524d
124KK	F1a1a	108, 129, 162, 172, 304	249d, 309.1C	523d, 524d
125KK	F1a1a	108, 129, 162, 172, 304	249d, 309.1C	523d, 524d
16LB	F1a1a	108, 129, 162, 172, 304	249d, 309.1C	523d, 524d
20R	F1a1a	108, 129, 162, 172, 304, 357	249d, 309.1C	CRS
			249d	523d, 524d

Table 1 (continued)

Sample	HG	HV1 <sup>a</sup> (15997–16410)	HV2 <sup>b</sup> (048–408)	HV3 (438–619)
143PB	F1a1a	108, 129, 162, 172, 304, 362	249d	CRS
8M	F1a3	129, 172, 304, 311	249d, 309.1C	CRS
131KK	F1a3	129, 172, 304, 311	249d, 309.1C	523d, 524d
155PB	F1a4	129, 172, 294, 304	152, 249d, 309.1C, 309.2C	521d-524d
39Mb	F1a4	129, 172, 294, 304, 362	152, 249d, 309.1C	CRS
1M	F3b	220C, 265, 298, 311, 362	150, 152, 249d, 309.1C	CRS
172GP	F4b	170, 218, 304, 311	249d, 309.1C	CRS
173GP	F4b	170, 218, 304, 311	249d, 309.1C	CRS
177GP	G	223, 256, 278	151, 152	489, 538C
35SM	G1a2	124, 214, 223, 362	309.1C	489
10M	G2a	214A, 223, 256, 278	152	489
109PT	G2a	214A, 223, 256, 278	152	489
16M	G2a1	214A, 223, 227T, 256, 278	152	489
80MR	M	37, 129, 184, 223, 278, 372	249d	456, 489
79MR	M	69, 223, 278, 311	150, 152, 309.1C	489
67S	M	86, 223, 278, 319	150	489
91MR	M	92, 93, 140, 169, 172, 183C, 189, 223, 278	249, 309.1C, 319	489, 523d, 524d
10R	M	93, 129, 183C, 189, 213, 218, 223	194, 309.1C	489, 523d, 524d
22LB	M	93, 129, 223, 309, 319	146, 195, 215, 309.1C	489
9R	M	93, 209, 223, 224, 263, 278, 319	146, 150, 151, 309.1C, 309.2C	489, 523d, 524d
3R	M	93, 209, 223, 224, 263, 278, 319	146, 152	489
7R	M	93, 223	152, 195, 309.1C	489, 523d, 524d
137KK	M	93, 223, 278	146	489, 523d, 524d
13LB	M	111, 129, 140, 172, 189, 223, 278	146, 249	489, 523d, 524d
37SM	M	111, 184, 189, 223, 295, 296, 311		489, 523d, 524d, 551
19LB	M	129, 153, 172, 223, 263, 309	204, 207, 309.1C	466, 489
90MR	M	129, 223, 263, 274, 311	309.1C	CRS
32SM	M	168, 223, 278	199	489
47SM	M	172, 223, 261	309.1C, 309.2C	489
68S	M	183C, 189, 223, 271, 311	151, 152, 309.1C, 309.2C	489
13M	M	183C, 189, 223, 278	150, 309.1C, 309.2C	489
27R	M	189, 192, 223, 278	150, 228	489
8R	M	209, 223, 233, 274, 304	143	489, 523d, 524d
7Y	M	209, 223, 233, 274, 304	143	489, 523d, 524d
6M	M	209, 223, 233, 274, 304		489, 523d, 524d
5M	M	209, 223, 224, 263, 278, 319	146, 199, 309.1C	489, 523d, 524d

73S	M	223	146, 199, 309.1C	489, 523d, 524d
136KK	M	223	146, 199, 309.1C	489, 523d, 524d
129KK	M	223, 246T, 311	195, 309.1C, 309.2C	489, 593
139KK	M	223, 246T, 311	195, 309.1C, 309.2C	489, 593
148PB	M	223, 246T, 311, 362	146, 195, 309.1C	489
38SM	M	223, 269, 271	150, 151, 309.1C	489
39SM	M	223, 269, 271	150, 151, 309.1C	489
99PT	M	223, 269, 345, 367	309.1C, 373	489
9Y	M2	172, 183C, 189, 223, 287, 319	207	489
111LB	M2b	129, 187, 189, 223, 274, 319, 320, 362	143, 195, 337	447G, 489
76MR	M5	86, 129, 209, 223, 272	152, 225, 249d, 316	489, 521d-524d
74S	M5	86, 129, 209, 223, 272	152, 225, 249d, 316	489, 521d-524d
7M	M5	86, 129, 209, 223, 272	152, 225, 249d, 316	489, 523d, 524d
82MR	M5	129, 209, 223, 249, 272	152, 225, 249d, 309.1C, 316	489, 523d, 524d
17LB	M5	129, 209, 223, 272	143, 152, 225, 249D, 316	489, 523d, 524d
5Y	M5	129, 209, 223, 272	152, 225, 249d, 309.1C, 309.2C, 316	489, 523d, 524d
12M	M5	129, 209, 223, 272	152, 225, 249d, 309.1C, 316	489, 523d, 524d
33R	M5	129, 209, 223, 272, 311	199, 249d, 309.1C, 309.2C, 316	489, 523d, 524d
6Y	M7b	129, 189, 223, 297	150, 199, 309.1C, 332	489
120BN	M7b1	129, 189, 192, 223, 297, 311	150, 199, 309.1C	489
60LG	M7b1	129, 183C, 189, 192, 223, 297	150, 199, 260, 309.1C, 332	489
48SM	M7b1	129, 183C, 189, 192, 223, 297	150, 199, 260, 309.1C, 332	489
45SM	M7b1	129, 183C, 189, 192, 223, 297	150, 199, 260, 309.1C, 332	489
58LG	M7b1	129, 183C, 189, 192, 223, 297	150, 199, 260, 309.1C, 332	489
51LG	M7b1	129, 183C, 189, 192, 223, 297	150, 199, 260, 309.1C, 332	489
44SM	M7b1	129, 183C, 189, 192, 223, 297	150, 199, 260, 309.1C, 332	489
46SM	M7b2	129, 183C, 189, 192, 223, 293, 297, 298, 362	150, 199, 260, 309.1C, 332	489
104PT	M7b3	86, 129, 223, 297, 324	199	489
110PT	M7b3	86, 129, 223, 297, 324	199	489
4R	M7c1c	93, 223, 295, 362	146, 199, 204, 309.1C, 309.2C	489, 523d, 524d
144PB	M7c1c	93, 223, 295, 362	146, 199, 309.1C	489, 523d, 524d
26R	M7c1c	176G, 223, 295, 362	146, 199, 209.1C, 309.2C	489, 523d, 524d
77MR	M7c1c	223, 295, 362	146, 150, 199	489, 523d, 524d
75MR	M7c1c	223, 295, 362	146, 150, 199	489, 523d, 524d
11R	M7c1c	223, 295, 362	146, 152, 199, 309.1C	489, 523d, 524d
25LB	M7c1c	223, 295, 362	146, 199	489, 523d, 524d
29R	M7c1c	223, 295, 362	146, 199	489, 523d, 524d
23M	M7c1c	223, 295, 362	146, 199, 309.1C	489, 523d, 524d
28R	M7c1c	223, 295, 362	146, 199, 309.1C	489, 523d, 524d

Table 1 (continued)

Sample	HG	HV1 <sup>a</sup> (15997–16410)	HV2 <sup>b</sup> (048–408)	HV3 (438–619)
12R	M7c1c	223, 295, 362	146, 199, 309.1C	489, 523d, 524d
78MR	M7c1c	223, 295, 362, 372	150, 199, 309.1C	489, 523d, 524d
70S	M10	223, 311, 359	150	489
21R	M10	223, 311, 362	146, 152, 195	489, 523d, 524d
8Y	M10	223, 311, 362		489
1LB	M12	129, 172, 220, 223, 234, 290, 312	146, 309.1C	489
7LB	M12	129, 172, 223, 234, 290, 312	146, 309.1C, 309.2C	489
55LG	M21a	93, 129, 172, 223, 256, 271		489
170GP	M21a	93, 129, 223, 256, 271, 362		489
4LB	M21a	93, 129, 223, 256, 271, 362	309.1C	489
171GP	M21a	129, 192, 223, 256, 271	146, 309.1C	489
21LB	M21b	223, 263, 381	182	489
23LB	M21b	223, 263, 381	182	489
10Y	M21d	129, 192, 223, 291		489, 517
9LB	M21d	145, 181, 223, 291, 304, 390	210, 309.1C	489
152PB	M22	223, 290, 304	152, 309.1C	489
12LB	M22	223, 290, 304	159, 195, 309.1C	489
20M	M45b	182C, 183C, 189, 209, 223	309.1C, 309.2C	489
22M	M45b	183C, 189, 209, 223, 233, 274, 304	152, 309.1C, 309.2C	481, 489
65S	N	182C, 183C, 189, 223, 362	150, 309.1C	456, 489
32R	N9a	223, 257A, 261, 297, 330	150, 309.1C, 379	CRS
41SM	N9a	223, 257A, 261, 311	150	CRS
1R	N9b	111, 168, 172, 183C, 189, 223, 311	152	513
149PB	N9b	111, 168, 172, 183C, 189, 223, 311, 362	152	513
151PB	N9b	111, 168, 172, 183C, 189, 223, 311, 362	152	513
138KK	N9b	111, 168, 172, 183C, 189, 223, 311, 362	152	513, 523d, 524d
24R	N21	193, 223, 344	150, 195, 337d	523d, 524d
54LG	N21	193, 291	150, 195, 309.1C, 337d	CRS
141PB	N21	193, 223	150, 195, 309.1C, 337d	CRS
25M	N22	93, 168, 223, 249		CRS
40SM	N22	93, 168, 223, 249	195	CRS
108PT	P	129, 176, 266, 270, 311, 357	143, 212, 309.1C, 309.2C	CRS
117BN	Q1	129, 144, 189, 223, 241, 265C, 311	89, 146	489, 596
175GP	R	37	152	CRS
180GP	R	37	152	CRS
181GP	R	37	152	CRS



15LB	R	47, 51, 168, 184A, 201, 304	146, 309.1C	CRS
140KK	R	189, 221, 298, 311, 362	249d, 309.1C	523d, 524d
24LB	R	214, 238	152, 339	417
53LG	R	261, 319, 362		CRS
12Mb	R9	304, 325, 362		CRS
42SM	R9a	260, 298, 355, 362	249d, 309.1C	CRS
2LB	R9a	260, 298, 355, 362	207, 249d, 309.1C	CRS
5LB	R9a	260, 298, 355, 362	207, 249d, 309.1C	CRS
4Mb	R9b	192, 288, 304, 309	143, 146, 183	523d, 524d, 573.1C
86MR	R10	235, 249, 288	195, 309.1C, 329	CRS
81MR	R10	235, 249, 288	195, 309.1C, 329	CRS
72S	R10	235, 249, 288, 372	195, 309.1C, 329, 374	CRS
64S	R10	249, 288, 359	195, 329, 374	CRS
30R	R10	172, 189, 249	151, 195, 329, 374	CRS
11M	R11	86, 182C, 183C, 189, 239, 309, 311	185, 189	CRS
6R	R21	295, 296, 304, 354	146, 154, 199, 249d, 309.1C	471
88MR	R22	129, 145, 182C, 183C, 189, 249, 288, 304, 311	309.1C, 309.2C, 329	CRS
157PB	R22	148, 223, 249, 288, 301, 304, 390	152, 186, 309.1C, 329	513, 593
160GP	U7	187, 207, 243, 309, 318T	151, 152, 188, 309.1C	523d, 524d
163GP	U7	187, 207, 243, 309, 318T	151, 152, 188, 309.1C	523d, 524d
166GP	U7	187, 207, 243, 309, 318T	151, 152, 188, 309.1C	523d, 524d
2Y	Y2	126, 231, 311	228, 234, 309.1C	482, 523d, 524d
146PB	Y2	126, 231, 311	228	482
111PT	Y2	126, 231, 311		482
169GP	Y2	126, 231, 311	309.1C	482, 524.1A, 524.2C
178GP	Y2	126, 231, 311	309.1C	482, 524.1A, 524.2C

Each sequence was compared with revised Cambridge reference sequence and the presence of the substitution, deletion, or insertion was recorded following the International Union of Pure and Applied Chemistry (IUPAC) designations

Samples are numbered according to the sampling locations (Fig. 1)

<sup>a</sup> Sequences were stated by omitting the 16,000 bases; thus, 16037 polymorphism is stated as 37

<sup>b</sup> Polymorphisms at nucleotide position 73, 263, and 315.1C were not listed since they were present in all samples

Kelantan: *Mb* Kota Bharu, *M* Machang, *R* Rantau Panjang

Kedah: *Y* Yan, *LB* Lembah Bujang

Negeri Sembilan: *SM* Sri Menanti, *LG* Lenggeng

Johor: *BN* Benut, *PT* Pontian, *MR* Muar, *S* Semerah

Perak: *KK* Kuala Kurau, *PB* Parit Buntar, *GP* Gopeng

**Table 2** Comparison of the most frequent Malay mtDNA haplotypes observed in this study compared to previous data

mtDNA haplotypes	Frequencies (%)		
	Malaysia Malays [8] n = 124	Singapore Malays [7] n = 205	This study n = 248
<b>HV1:</b>			
16140, 16182C, 16183C, 16189, 16217, 16274, 16335	4.84	0.49	4.44
16223, 16295, 16362	4.03	4.39	3.23
16182C, 16183C, 16189, 16217, 16261	0.81	-	2.82
16129, 16183C, 16189, 16192, 16223, 16297	-	-	2.42
16108, 16129, 16162, 16172, 16304	1.61	1.95	2.42
16140, 16183C, 16189, 16266A	2.42	3.41	2.42
<b>Combined HV:</b>			
73, 146, 150, 195, 263, 315.1C, 16140, 16182C, 16183C, 16189, 16217, 16274, 16335	2.42		4.44
73, 146, 263, 315.1C, 523d, 524d, 16182C, 16183C, 16189, 16217, 16261	0.81		2.82
73, 199, 260, 263, 315.1C, 332, 489, 16129, 16183C, 16189, 16192, 16223, 16297	-	N. A	2.42
73, 249d, 263, 315.1C, 523d, 524d, 16108, 16129, 16162, 16172, 16304	0.81		2.02
73, 146, 199, 263, 315.1C, 489, 523d, 524d, 16223, 16295, 16362	2.42		2.02
73, 210, 263, 315.1C, 523d, 524d, 16140, 16183C, 16189, 16266A	0.81		2.02

namely Arabs, Chinese, Indians, Siamese, and Proto-Malays. Available mtDNA data on the Malay population of Peninsular Malaysia is so far limited and was only represented by Malay individuals from a particular area (for example, Maruyama et al. [8]: studied Malays from Kuala Lumpur only). In this study, we reported on a more diverse population of the Malays, representing Malays from the East (Kelantan), the North (Kedah and Perak), and the South (Negeri Sembilan and Johor) of Peninsular Malaysia. The Malay individuals involved in this study were also carefully selected by establishing their ancestry for at least three generations of Malay without any mix-marriage. None of the previously published data has this inclusion criterion in selecting their samples [7, 8].

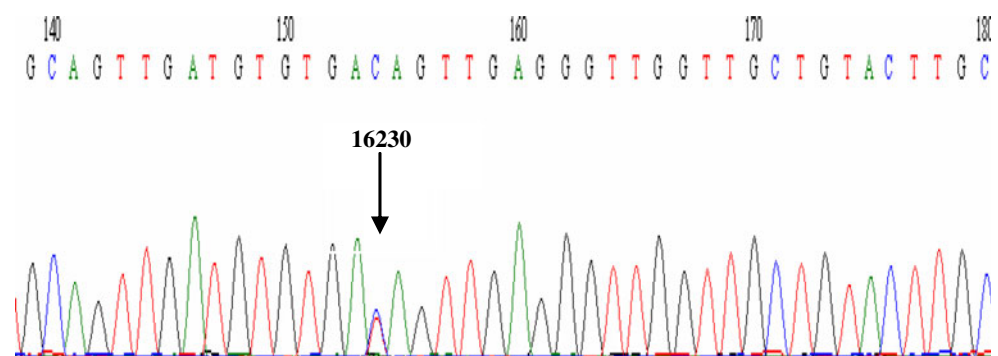
Haplotype determination was done by excluding the C tracts polymorphisms in HV2 region following the procedure from previous studies [13, 14]. The nucleotide poly-

morphisms in 248 Malay individuals were displayed in Table 1. A total of 157 HV1 haplotypes, 73 HV2 haplotypes, and 28 HV3 haplotypes were obtained. The genetic diversity calculated was 99.24% for HV1, 95.14% for HV2, and 79.93% for HV3. The random match probability was calculated to be 1.16% for HV1.

The most frequent HV1 haplotype (16140, 16182C, 16183C, 16189, 16217, 16274, and 16335) was shared by 11 individuals (4.44%), HV2 common haplotype (73, 249d, 263, 315.1C) was shared by 27 individuals (10.89%), and transition at position 489 was the most frequently observed polymorphism in HV3 (27.02%). On combination of HV1, HV2, and HV3 sequences, a total of 180 haplotypes were identified, 149 of which were unique, and 31 were present in more than one individual.

The genetic diversity calculated was 99.47%, and the probability of random match of two individuals sharing the

**Fig. 2** Electrochromatogram showing point heteroplasmy at nucleotide position 16230 for sample number 174RW. The point heteroplasmy was detected through the heavy strand sequencing due to the occurrence of poly-C stretch from nucleotide position 16182 to 16193 on the light strand



same mtDNA haplotype was 0.93%. Statistical parameters such as nucleotide diversity and the mean of pair-wise differences for combined mtDNA haplotypes were calculated as  $0.036063 \pm 0.020101$  and  $12.544022 \pm 6.230486$ , respectively.

The most frequent mtDNA haplotype in combined analysis of the HV regions (73, 146, 150, 195, 263, 315.1C, 16140, 16182C, 16183C, 16183C, 16189, 16217, 16274, and 16335) was shared by 11 individuals (4.44%). This haplotype was found at a frequency of 2.42% by Maruyama et al. [8] but was not observed in Vietnamese [15], Korean [16], Hong Kong [17], and northeast Chinese Han [18] populations but was relatively frequent in Island Southeast Asian populations (2.8%) [19] and Aboriginal Taiwanese (5.5%) [20]. In addition, polymorphism at positions 73, 263, 315.1C, 489, and 16223 were frequently observed in most of the samples.

The second most frequent haplotype observed in this study (73, 146, 263, 315.1C, 523d, 524d, 16182C, 16183C, 16189, 16217, and 16261), which was shared by seven individuals (2.82%) was only found in one individual by Maruyama et al. [8]. This haplotype was also widely distributed in East and Southeast Asian populations [19, 21], but was not reported in Vietnamese [15], Hong Kong [17], and northeast Chinese Han [18] populations. Comparison of the most common haplotypes observed in our study with previously published Malay data is shown in Table 2. Most of the frequent haplotypes of the combined HV regions in this study was found at lower frequencies in the previously published Malay data [7, 8]. In fact, one of our most common haplotypes (73, 199, 260, 263, 315.1C, 332, 489, 16129, 16183C, 16189, 16192, 16223, and 16297) which was shared by six individuals (2.42%), was not observed in other studies.

A point mutation heteroplasmy at nucleotide position 16230 (Fig. 2) was observed in sample 174RW. This heteroplasmy was observed as a purine transition (A→G). Since the presence of homopolymeric tract of cytosine between nucleotide positions 16184–16193 had affected the downstream sequence quality, this point heteroplasmy was detected from the sequencing of the heavy strand and was confirmed by re-sequencing [22].

## Conclusion

We have generated the mtDNA database for HV1, HV2, and HV3 in 248 unrelated Malay individuals of Peninsular Malaysia. The database can be used for reference in forensic identification purposes as well as in evolutionary studies.

**Acknowledgments** Special thanks to the Ministry of Higher Education, Malaysia, for funding this research under the Fundamental Research Grant Scheme (203/PPSK/6171005).

**Ethical standards** The conducted experiments comply with the current law practice in Malaysia.

**Conflict of interest** The authors have no conflict of interest.

## References

1. Anderson S, Bankier AT, Barrell BG, Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence mitochondrial genome. *Nature* 290:457–464
2. Andrew RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147
3. Pakendorf B, Stoneking M (2005) Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet* 6:165–183
4. Nik Hassan Suhaimi NAR (1998) The encyclopedia of Malaysia: early history, vol 4. Archipelago Press, Singapore, pp 6–11
5. Zafarina Z, William G (2004) Mitochondrial DNA profiling of modern Malay and Orang Asli populations in Peninsular Malaysia. *Int Congr Ser* 1261:428–430
6. Edinur HA, Zafarina Z, Spínola H, Nurhaslindawaty AR, Panneerchelvam S, Norazmi MN (2009) HLA polymorphism in six Malay subethnic groups in Malaysia. *Hum Immunol* 70:518–526
7. Wong HY, Tang JS, Budowle B, Allard MW, Syn CKC, Tan-Siew WV, Chow ST (2007) Sequence polymorphism of the mitochondrial DNA hypervariable regions I and II in 205 Singapore Malays. *Legal Med* 9:33–37
8. Maruyama S, Nohira-Koike C, Minaguchi K, Nambiar P (2010) mtDNA control region sequence polymorphisms and phylogenetic analysis of Malay population living in or around Kuala Lumpur in Malaysia. *Int J Legal Med* 124(2):165–170
9. Tajima F (1989) Statistical methods for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595
10. Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA (1991) Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am J Hum Genet* 48:370–382
11. Excoffier L, Laval G, Schneider S (2006) Arlequin version 3.1: An integrated software package for population genetics data analysis. *Computational and Molecular Population Genetics Lab (CMPG) Switzerland: Institute of Zoology, University of Berne*
12. Syed HA (1979) *Orang Melayu: Masalah dan masa depannya*. Penerbit Adabi Kuala Lumpur, pp 1–3
13. Melton T (2004) Mitochondrial DNA heteroplasmy. *Forensic Sci Rev* 16:1
14. Forster L, Forster P, Gurney SMR, Spencer M, Huang C, Röhl A, Brinkmann B (2010) Evaluating length heteroplasmy in the human mitochondrial DNA control region. *Int J Legal Med* 124(2):133–143
15. Irwin JA, Saunier JL, Strouss KM, Diegoli TM, Sturk KA, O'Callaghan JE, Paintner CD, Hohoff C, Brinkmann B, Parsons TJ (2008) Mitochondrial control region sequences from a Vietnamese population sample. *Int J Legal Med* 122:257–259
16. Lee HY, Yoo JE, Park MJ, Chung U, Shin KJ (2006) Mitochondrial DNA control region sequences in Koreans: identification of useful variable sites and phylogenetic analysis for mtDNA data quality control. *Int J Legal Med* 120:5–14
17. Irwin JA, Saunier JL, Beh P, Strouss KM, Paintner CD, Parsons TJ (2009) Mitochondrial DNA control region variation in a population sample from Hong Kong, China. *Forensic Sci Int* 3(4): e119–e125

18. Zhang YJ, Xu QS, Zheng ZJ, Lin HY, Lee JB (2005) Haplotype diversity in mitochondrial DNA hypervariable region I, II and III in northeast China Han. *Forensic Sci Int* 149:267–269
19. Hill C, Soares P, Mormina M, Macaulay V, Clarke D, Blumbach PB, Vizuete-Forster M, Forster P, Bulbeck D, Oppenheimer S, Richards M (2007) A mitochondrial stratigraphy for Island Southeast Asia. *Am J Hum Genetics* 80:29–43
20. Tajima A, Sun CS, Pan IH, Ishida T, Saitou N, Horai S (2003) Mitochondrial DNA polymorphisms in nine aboriginal groups of Taiwan: implications for the population history of aboriginal Taiwanese. *Hum Genet* 113:24–33
21. Kivisild T, Tolk H, Parik J, Wang Y, Papiha S, Bandelt HJ, Villems R (2002) The emerging limbs and twigs of the East Asian mtDNA tree. *Mol Bio Evol* 19(10):1737–1751
22. Carracedo A, Bär W, Lincoln P, Mayr W, Morling N, Olaisen B, Schneider P, Budowle B, Brinkmann B, Gill P, Holland M, Tully G, Wilson M (2000) DNA commission of the international society for forensic genetics: Guidelines for mitochondrial DNA typing. *Forensic Sci Int* 110:79–85