



Journal of Biological Sciences

ISSN 1727-3048

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Mitochondrial DNA Polymorphism and Phylogenetic Relationships of Proto Malays in Peninsular Malaysia

¹L.S. Lim, ¹K.C. Ang, ¹M.C. Mahani, ²A.W. Shahrom and ¹B.M. Md-Zain

¹School of Environmental and Natural Resources Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysia

²Forensic Unit, Hospital Universiti Kebangsaan Malaysia, Jalan Ya'acob Latif, Bandar Tun Razak, 56000 Cheras, Kuala Lumpur, Malaysia

Abstract: This study focus on the phylogenetic relationships among six Proto-Malay tribes (Jakun, Temuan, Semelai, Kuala, Seletar and Kanaq) and the effectiveness of using HVSI D-loop region segment and a 9 bp deletion of the intergenic region of COII/tRNA^{Lys} of mtDNA in portraying the phylogenetic relationships. The analysis showed high pairwise differences among Kanaq, Jakun and Semelai. Thirty two haplotypes were formed from 89 D-loop sequences of Proto Malay individuals. Deletion on the 9 bp tandem repeats of COII/tRNA^{Lys} was detected in Semelai and Orang Kuala. Neighbor-Joining, Maximum Parsimony and Maximum Likelihood analysis revealed Jakun and Semelai haplotypes are the earliest to split from ingroups and showed that Jakun may be an ancestor for the Malay populations in the Malay Peninsular, which also support the anthropological findings. Nine basepair deletion of COII/tRNA^{Lys} intergenic region had occurred in two clades. Results indicates a close relationships among Semelai, Temuan and Jakun. Most Kanaq and Seletar sequences were merged into tribe specific haplotypes, showing close relationships within tribes. The HVS I D-loop region of most Proto Malays are highly variable as 27 of 32 haplotypes were subgroup-specific. The HVS I of D-loop successfully revealed the close relationships among the Proto-Malays but was less effective in discriminating each tribe. Detection of ancestries for the Proto Malays using 9 bp deletion of COII/tRNA^{Lys} intergenic region reveals the Asian origins for Kuala and Semelai.

Key words: Mitochondrial DNA polymorphism, Proto Malays, HVS I, Orang Asli, COII/tRNA^{Lys} intergenic region

INTRODUCTION

Peninsular Malaysia is a region with a great diversity of human population formed by Malays, Chinese, Indian and the minority aboriginal (Orang Asli) populations (Lye, 2001; Hood, 2006). The Orang Asli of Peninsular Malaysia is separated into three main tribal groups, the Negrito, the Senoi and the Proto Malays based on the physical appearance and sociological differences (Fig. 1) (Nicholas, 2000; Bellwood, 1997; Hood, 2006; JHEOA, 2002).

Negrito is believed to be the earliest to arrive in Peninsular Malaysia, in about 25, 000 years ago. To date, Negrito has the least populations among the three Orang Asli groups. Their settlements are isolated and scattered but mainly distributed in the Northern and middle part of the peninsular. The Negritos are physically similar to Andaman islanders, Aeta in the Philippines, Melanesians and Tasmamians. They were predicted to have originated

from Africa and spread throughout Southeast Asia (Fix, 1995; JHEOA, 2002; Macaulay *et al.*, 2005).

Senoi is the largest Orang Asli group in Peninsular Malaysia (Nicholas, 2000) and mainly settle from the Middle to Northern part of the peninsular. Senoi is estimated to reach Peninsular Malaysia during the second wave of migration in about 8,000 years ago from South Asia, the mountain areas of Cambodia and Vietnam (Nicholas, 1996; Baer, 1999). Senoi has Mongoloid physical characteristics and speaks Khmer dialects (Nicholas, 1996). However, some believe Senoi are descendants of Australoid from Australia and Veddoid from South India (Fix, 1995).

The second largest race of Orang Asli, the Proto Malays were separated into six tribes, including Jakun/Orang Hulu, Temuan, Semelai, Kuala, Kanaq and Seletar (Fig. 1). Fix (1995) had categorized the Proto Malays into three categories: the first category consists of Tribes of Melayu Asli, who speak Malay and wear

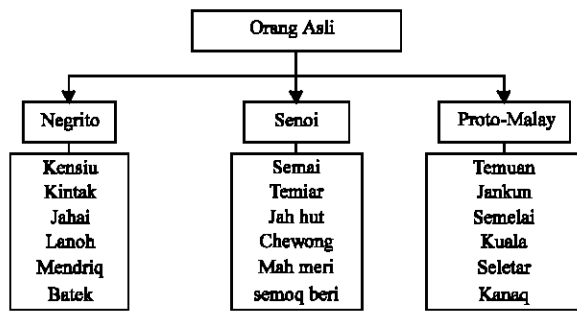


Fig. 1: Three main groups and 18 subgroups of Orang Asli in Peninsular Malaysia

Malay costume, such as Temuan. The second category consists of tribes with the combination of Proto Malays-Senoi from the linguistic and cultural aspect. The last category consists of tribes settling at coastal areas. The third category was mainly Muslim and speaks Sumatra dialects. Early findings indicated Proto Malays' main tribe (Jakun) were believed to have migrated to Peninsular Malaysia, after the arrival of Negritos and Senois, about 4 000 years ago. Fix (1995) believed Proto Malays migrate from the middle part of Asia (Yunnan) and came to Peninsular Malaysia through Peninsular Indo-China. This was made based on the cultural, linguistic and artifacts. Later finding based on archeology and linguistic suggested the proto-Austronesian speakers settled in Taiwan about 4000 B.C. before migrated Southwards to Southeast Asia region through Philippines into Borneo, Sulawesi, Central Java and Eastern Indonesia 2500 years ago. The second wave of migration occurred from central Java back to North into Peninsular Malaysia via the Straits of Malacca between 1500 and 500 B.C. (Andaya, 2001). On the other hand, Orang Kuala is believed to have migrated from Sumatra Island and Kanaq was migrated from the Riau Islands. Most of the Proto Malays are settled in the middle and Southern parts of Peninsular Malaysia (Fig. 1).

Proto Malays are similar to Deutero- Malays not only from the morphological aspect, but they also share similarity in their culture and languages. Therefore, they were named Proto Malays and predicted as one of the ancestral groups for Deutero-Malays (Kasimin, 1991). Deutero-Malays are believed to have reached Peninsular Malaysia about 1500-2000 years ago, through the Southern part of China, which is also the starting point for migration of Proto Malays to Peninsular Malaysia (Fix, 1995). Deutero-Malays were the earliest to have settled in Peninsular Malaysia compared to Chinese and Indians (Kasimin, 1991). It is also believed that marriages between Proto Malays with other populations from Arab,

Chinese, Indians and Saimese that migrated into the Malay peninsula that formed the diverse recent Deutero Malay populations (Sainuddin, 2003; Hussein *et al.*, 2007). Therefore, the origins of Proto Malays and Deutero Malays were only based on speculation. The Proto Malays were less well understood compared to Negrito and Senoi because of the similarities between Proto Malays and Deutero-Malays. Besides, most of the Proto Malays populations are located at the border of urban or suburban areas, creating a wrong impression that they were Deutero-Malays.

To date, studies on the origins and the migration history as well as the relationships among Orang Asli in Peninsular Malaysia are limited (Lian, 2001; Hill *et al.*, 2006; Macaulay *et al.*, 2005; Ang, 2009) with most of the records being contributed from the anthropological point of view (Nicholas, 1996; Kasimin, 1991). In the past two decades, the migration history of Malays and Orang Asli were revised from the molecular evolution approach. MtDNA has been widely used in the studies of human migration and population studies (Al-Zahery *et al.*, 2003; Kashyap *et al.*, 2003; Hill *et al.*, 2006; Lutz *et al.*, 1998; Macaulay *et al.*, 2005). More recently the existence of a Southern route has been supported by analysis of mtDNA restriction enzyme from New Guinea (Forster *et al.*, 2001) and control region from mainland India and the Andaman Islands (Endicott, 2003). The mtDNA presents variation among individuals. It evolves much faster than single copy nuclear genes. It has been widely used in population and evolutionary studies for the last two decades (Lim *et al.*, 2004).

Therefore, the objectives of this research were to study the mtDNA polymorphism and phylogenetic relationships among six Proto-Malay tribes (Jakun, Temuan, Semelai, Orang Kuala, Orang Seletar, Orang Kanaq). In addition the effectiveness of using the hypervariable site I (HVS I) of D-loop region segment and a 9-bp deletion of the intergenic region of COIII/tRNA^{Lys} of mtDNA in elucidating the phylogenetic relationships among the Proto Malays can be assessed.

MATERIALS AND METHODS

Subjects: Blood samples representing Deutero-Malay and six tribes of Proto Malays (Jakun, Temuan, Semelai, Kuala, Seletar, Kanaq) populations were collected at Hospital Universiti Kebangsaan Malaysia, Hospital Jabatan Hal Ehwal Orang Asli and the settlements for Proto Malays from 2003 to 2006 (Fig. 2). The origins and ethnicity of each individual was confirmed for at least three generations through a simple interview. Genomic DNA was extracted using the standard method with the

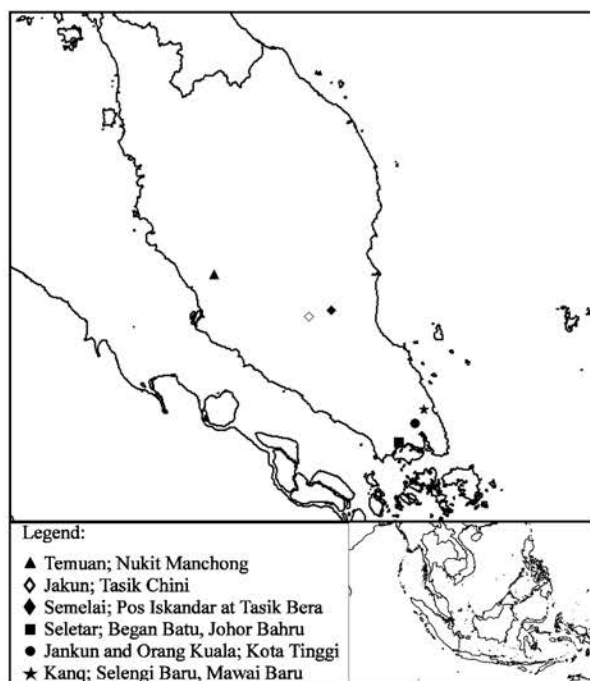


Fig. 2: Map of the locations for sampling. Name of tribe and the sampling site were listed according to the symbols on the map. Small map at the bottom right corner shows location of Peninsular Malaysia in Southeast Asia region (Nicholas, 2000)

treatment of SDS 2%, Proteinase K 10 mg mL⁻¹ and phenol/chloroform extraction (Hillis *et al.*, 1996; Sambrook and Rusell, 2001).

MtDNA analysis: HVS I of D-loop region and tandem repeats in the intergenic region of COII/tRNA^{Lys} in mtDNA was amplified by PCR method (Saiki, 1989) using Invitrogen Taq Polymerase, recombinant. Primers A and E (Table 1) designed by Fucharoen *et al.* (2001) was used to amplify a fragment from the HVS I D-loop region (15897 bp-100 bp according to mitochondrial map proposed by Anderson *et al.* (1981) whereas primers F and G were designed by Horai *et al.* (1996) to amplify tandem repeats of COII/tRNA^{Lys} intergenic region (8211 bp-8310 bp).

Amplification for both of the regions were performed using 50 µL of total volume per reaction with 50-120 ng µL⁻¹ DNA template, 20 pmol µL⁻¹ primers, 3.0 mM Magnesium Chloride, 0.2 mM dNTP mix, 20 mM Tris-Cl in pH 8.4, 50 mM Potassium Chloride and 0.05 U Taq Polymerase. Temperature profile for 30 cycles of amplification was 94°C of denaturation for 15 sec, 45°C of annealing for 15 sec and 72°C of extension for 30 sec in Perkin Elmer GeneAmp[®]PCR System 2400 (Horai *et al.*, 1996; Fucharoen *et al.*, 2001). Amplified products were

Table 1: Primers used in amplification and sequencing of HVS I D-loop region (Horai *et al.*, 1996; Fucharoen *et al.*, 2001)

Primers	Primer's sequence
A (Forward)	15897-5' GTATAAACTAATACACCAGTCTTGT 3'-15921
C (Forward)	16204-5' AGCAAGTACAGCAATCAACC 3'-16223
E (Reverse)	100-5' CAGCGTCTCGCAATGCTATCGCGTG 3'-76
F (Forward)	8211-5' TCGTCCTAGAATTAATCCC 3'-8230
G (Reverse)	8310-5' AGTTAGCTTACAGTGGGCT 3'-8291

(The notation of Anderson *et al.* (1981) is used for numbering of bases)

purified before sequencing. All of the primers in Table 1 were used in sequencing. Sequencing was carried out in ABI 3730 xL machine (Applied Biosystems).

Data analysis: The nucleotide sequences for both regions were aligned by CLUSTAL X 8.13, CLUSTAL W and manually. HVS I DNA sequences from every subject were edited and combined to form haplotypes using MacClade 3.0 (Maddison and Maddison, 1992). HVS I haplotypes were analyzed for the variable sites, ratio of transition to transversion and sequences diversity aspects using PAUP 4.0b8 (Swofford, 2002). Total C-T transition sites and pairwise differences were also analyzed using MEGA 3.0 (Kumar *et al.*, 2004). Gene diversity index, F_{st} and gene flow, N_m were calculated using DNASP 4.0 (Rozas *et al.*, 2003). Intergenic region of COII/tRNA^{Lys} sequences were aligned and were scanned for the 9 bp (CCCCCTCTA) deletion, 9 bp partial deletion and triple copy of the 9 bp manually. Individuals with any of the mutations were recorded.

Phylogenetic trees were constructed based on the haplotypes and the 9 bp deletion intergenic region of COII/tRNA^{Lys} feature, using PAUP version 4.0b8. Three methods of the analysis are: a) Neighbor-joining (NJ) with Tamura-Nei algorithm (Tamura and Nei, 1993; Swofford, 2002; b) Maximum Parsimony (MP) with Stepwise Addition 1000 replicates in heuristic search (Swofford, 2002; Nei and Kumar, 2000); and c) Maximum Likelihood (ML) with HKY 85 model (Hasegawa *et al.*, 1985). The 9 bp deletion in COII/tRNA^{Lys} intergenic region was considered as one of the characters in the phylogenetic trees. All of the trees were subjected to bootstrap analysis with 1000 replicates to get the bootstrap value support.

RESULTS

Analysis on HVS I in D-loop region: About 529 bp of HVS I from nucleotide 16048 to 16569 was involved in this analysis (Anderson *et al.*, 1981). A total of 32 haplotypes were formed from 89 nucleotide sequences (Table 2). Fifty of 529 bp nucleotides were variable with two nucleotides substitution (47 transitions sites and 3 A-C transversion). Thirty-six of 47 transition characters were C-T transition

Table 2: Composition of haplotypes detected in the six tribes of Proto Malays

Haplotypes	Proto Malay Tribes	Haplotypes	Proto Malay Tribes
1	Temuan (1)	20	Seletar (5)
2	Kuala (1)	21	Jakun (1), Semelai (4), Temuan (1)
3*	Kuala (1)	22	Temuan (3)
4	Temuan (1)	23	Semelai (2)
5	Temuan (1)	24	Temuan (1)
6	Kuala (2)	25	Jakun (1)
7	Jakun (1)	26	Kuala (1)
8	Temuan (1)	27	Jakun (2)
9*	Semelai (1)	29	Jakun (1)
11	Jakun (1), Semelai (3)	32	Semelai (3)
12	Kanaq (9), Kuala (1)	37	Semelai (2)
13	Seletar (1)	39	Seletar (7)
14*	Kuala (1)	40	Kuala (7)
15	Temuan (1)	41	Kuala (1)
17	Jakun (2), Temuan (5)	42	Jakun (2)
19	Jakun (1)	43	Jakun (3), Kanaq (5), Seletar (2)

*Represent sequence with 9 bp deletion in COII/tRNA^{Lys} intergenic region. Numbers in brackets indicates individuals

Table 3: Polycytosine tracts found in six tribes of Proto Malays

Polycytosine tract	Jakun	Kanaq	Kuala	Seletar	Semelai	Temuan	Total
AAAA-CCCCCTCCCC	8	14	11	15	9	9	66
AAAA-CCCCCTCCCT	2				2	6	10
AA - - - CCCCCCCCCC	2		1		1		4
AAAA-CCCCCCCCC	1				3		4
AAA - -CCCCCCCCC	2		1				3
AAAA-CCCCCCCCC			2				2
Total	15	14	15	15	15	15	89

and the remaining 11 characters were A-G transition, giving transition to transversion ratio of 31.16. Nucleotide diversity was 1.47%.

Length polymorphism occurred in HVS I (Polycytosine tract) within the range of nucleotide 16148-16193. Six polymorphisms were found among six tribes of Proto Malays are shown in Table 3. Subjects of 74.16% involved in this study had polymorphism type I, which is AAAA-CCCCCTCCCC. Seletar and Kanaq populations are two tribes that only have type I Polycytosine tract. Polycytosine tract of Jakun, Semelai and Kuala were highly variable. Only two types of Polycytosine tract were found in Temuan, one of the main tribe of Proto Malays.

Pairwise differences between tribes were calculated using MEGA 3.0. Tamura-Nei algorithm was chosen in the calculation. Average sequence divergence in each was low, within the range 1 to 2% (Table 4). Lowest sequence divergence was found between Temuan-Kuala (1%) whereas highest occurred between Kanaq-Jakun (2%) and Kanaq-Semelai (2%). Mean distance among tribes were within 1.3 to 2.3%.

Most of the haplotypes (27 of 32 haplotypes) in this study were ethnic specific. Therefore, only six haplotypes were shared among tribes. Haplotype 43 was the most common haplotype shared among three tribes: Jakun, Kanaq and Seletar. Ten haplotypes were traced in Jakun populations and 9 detected in Temuan. However, there were only two haplotypes detected in Seletar and Kanaq

Table 4: Pairwise differences (percentage) among six tribes of Proto Malays based on Tamura-Nei (1993) algorithm

Tribes	Jakun	Kanaq	Kuala	Seletar	Semelai	Temuan
Jakun						
Kanaq	2.0					
Kuala	1.5	1.1				
Seletar	1.3	1.4	1.1			
Semelai	1.5	2.0	1.4	1.4		
Temuan	1.4	1.5	1.0	1.2	1.4	

(15 and 14 individuals, respectively). The two haplotypes detected in Kanaq individuals were not ethnic specific.

Analysis on COII/tRNA^{Lys} intergenic region: The mutations occurred between nucleotide 7586-8294 based on the mtDNA map suggested by Anderson *et al.* (1981). Out of 89 individuals from Proto Malays analyzed, only three individuals (one Semelai and two Kuala) were detected with the deletion of the 9 bp tandem repeats (CCCCCTCTA). As a whole, there were only 3.37% of total subjects in the study having the 9 bp deletion.

Three individuals detected with 9 bp deletion in COII/tRNA^{Lys} intergenic region formed unique HVS I D-loop haplotypes, respectively, which are haplotype 3, haplotype 9 and haplotype 14. C-T transition at the sites 16268 and 16318 were detected for haplotype 3 and haplotype 9 whereas C-T transition at sites 16194, 16230, 16285 and 160 were detected in haplotype 14 but not in haplotype 3 and haplotype 9. Therefore, haplotype 14 was grouped in a separate clade from the other two haplotypes in the phylogenetic trees constructed.

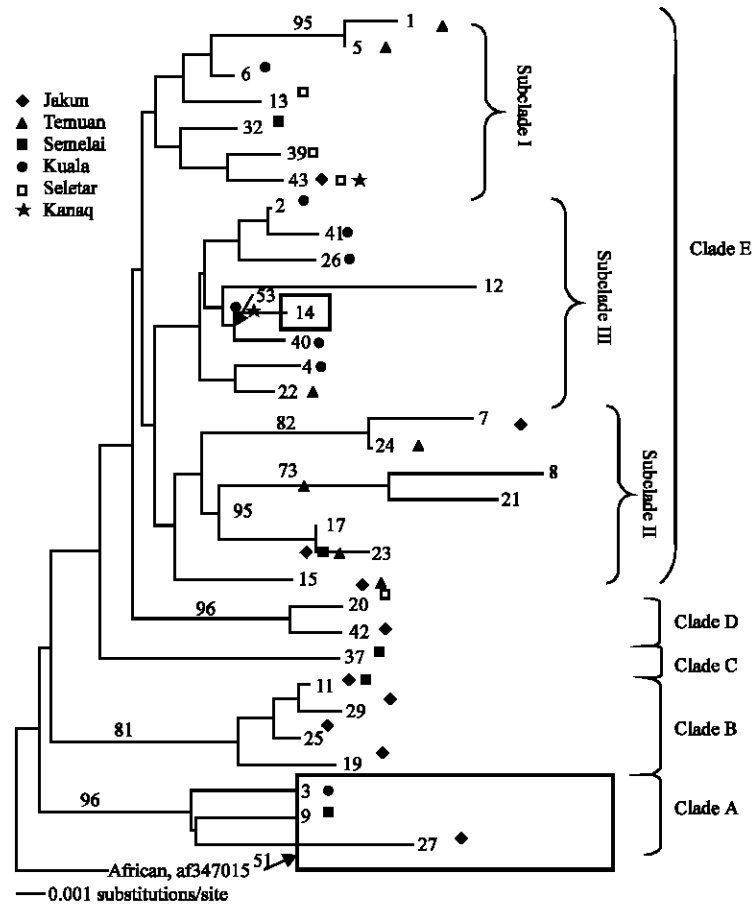


Fig. 3: NJ Phylogram constructed based on Tamura-Nei algorithm showed relationships among Proto Malays haplotypes. Numbers on the branch represent Bootstrap value with analysis of 1000 replicates. Boxed subclades consist of haplotypes with 9 bp deletions in COII/tRNA^{Lys} intergenic region

Polycytosine tract is a unstable tract (Malyarchuk *et al.*, 2002). Haplotype 14 was not group with the other three haplotypes with 9 bp deletions in COII/tRNA^{Lys} intergenic region due to different polycytosine pattern (Haplotype 14: AAAACCCCTCCC; Haplotype 3 and 9: AACCCCCCCCCC). There are a few substitutions which only detected in Haplotype 3 and 9 but not Haplotype 14. However, none of the substitution that segregated Haplotype 14 from Haplotype 3 and 9 were considered as mutational hotspot sites for D-loop region as listed by Malyarchuk *et al.* (2002). Haplotype 27 which did not faced 9 bp deletions in COII/tRNA^{Lys} intergenic region was clustered in the clade consists of Haplotype 3 and 9 due to the similarity of polycytosine tract.

Phylogenetic analysis: NJ, MP and ML trees were constructed based on 32 haplotypes obtained. The 9 bp deletions of tandem repeats in COII/tRNA^{Lys} intergenic

region was considered as one character in the analysis and represented by numerical numbers: 1 for no deletion detected and 0 for deletion detected. Out of a total of 529 HVS I sequences studied, 51 (inclusive of outgroup) characters were variable and 29 of the variable characters were parsimony informative. An African (AF347015) individual was chosen as outgroup because all living humans are believed to be of African descent based on Out of Africa theory (Cann *et al.*, 1987; Viligant *et al.*, 1991).

Proto Malays haplotypes were divided into five main clades (Fig. 3). Clade E was the biggest among five clades. All Temuan and most of the Kuala haplotypes only clustered in this clade. Subclade II of Clade E only consists of haplotypes from the middle of Peninsular Malaysia. Clade A, Clade B and Clade D were supported by bootstrap value of 96, 81 and 96%, respectively. Four Jakun haplotypes were clustered to form Clade B and three of the haplotypes were Jakun-specific. Jakun and

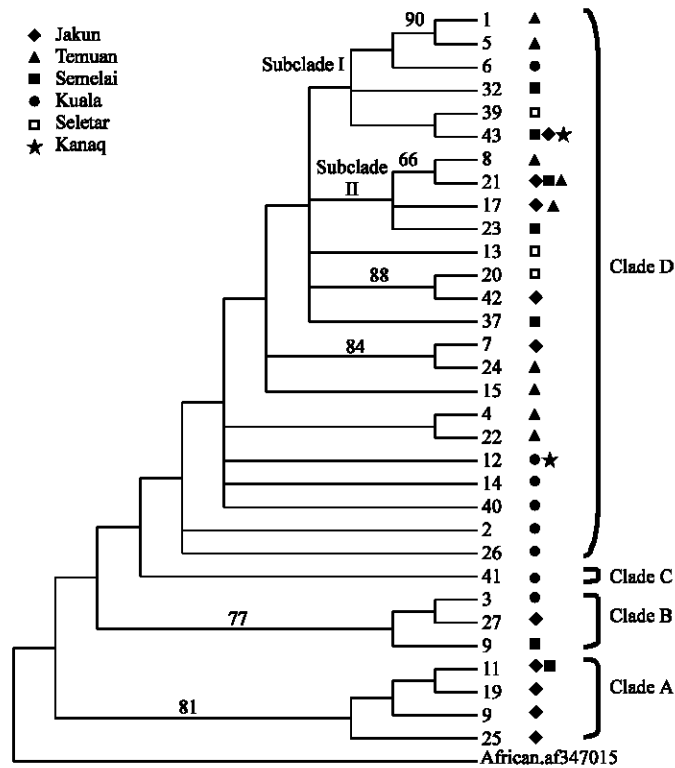


Fig. 5: ML phylogeny tree constructed based on HYK85 substitution model with heuristic search showed relationships among Proto Malays haplotypes studied. Numbers on the branch represent Bootstrap value with analysis of 1000 replicates. Subclades with β consist of haplotypes with 9 bp deletions in COII/tRNA^{Lys} intergenic region

phylogram was constructed (Fig. 6). From Fig. 6, it is also observed that Subclade I (mainly formed by Semelai, Seletar and Kanaq) and Subclade III (mainly form by Kuala, Temuan and Kanaq) form close relationships within the taxa in each subclade, respectively. The relationships can also be inferred from the pairwise differences (Table 4) and the low number of steps to form a branch in MP tree (Fig. 4). In Fig. 6, Subclade II of Clade E on NJ phylogram did not show close relationships within the taxa but the subclade was mainly made up of Jakun, Temuan and Semelai haplotypes.

From social sciences aspects, there are obvious differences in cultural background, linguistic and the religious beliefs for every subgroup of Proto Malays when compared with the other two Orang Asli ethnic groups, Negrito and Senoi. Therefore, anthropological findings further group Proto Malays into three main categories: Temuan which speaks Malay language group, Semelai which speaks own language and which allows intermarriage with Senoi and Kuala and Seletar which settle in the Southern coastal area of Peninsular Malaysia and speak Sumatran language (Idris, 1968).

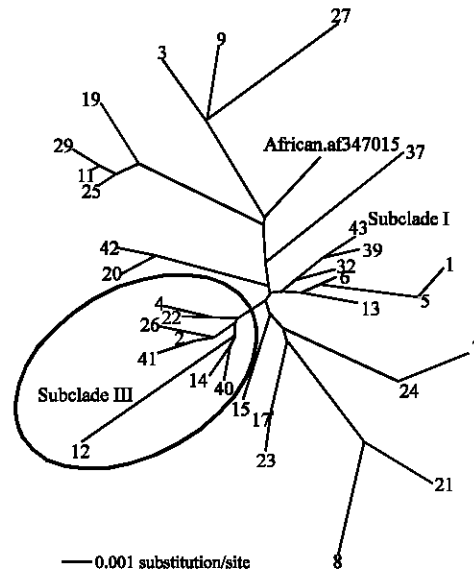


Fig. 6: Unrooted NJ phylogram shows the distribution of the closely related haplotypes in Subclade I and III. Out group is African. The phylogram is constructed based on Tamura-Nei (1993) algorithm

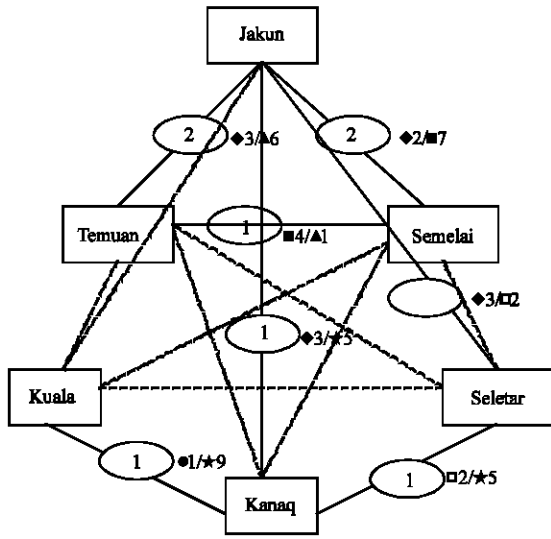


Fig. 7: Number of shared haplotypes observed between each pair of two populations. The figure in a circle on the solid line connecting two populations is the number of shared haplotypes. Pairs of populations that have no shared types are shown by dotted lines. Proto Malay subgroups following the symbol are total individuals sharing haplotypes from the respective two populations

The mitochondrial NJ, MP, ML tree topologies that reveal the relationships for Proto Malays in Fig. 3-5 support the clade with Haplotypes 3 (Kuala), 9 (Semelai) and 27 (Jakun) (i.e., haplotype with 9 bp deletion of COII/tRNA^{Lys} intergenic region) and the clade which contained Haplotypes 11 (Jakun, Semelai), 29 (Jakun), 25 (Jakun) and 19 (Jakun) as the earliest clades to have diverge from ingroup. The matrilineal result supports the anthropological findings that believed in the migration of Jakun from Yunan about 5,000 years ago, the third wave of successful migration to the Malay Peninsular (Kasimin, 1991; JHEOA, 2002). The date predicted was believed to be the earliest arrival date for Proto Malays in Peninsular Malaysia. Figure 7 shows the shared haplotypes between each pair of populations. Jakun shares two haplotypes with each of the land Proto Malay subgroups (Temuan and Semelai), respectively. Jakun also share a haplotype with Seletar and Kanaq. However, there was no common haplotype between Jakun and Kuala.

Semelai haplotypes were among the earliest to split from ingroup in all the phylogeny trees. Not much information has been available on the Semelai origins. Semelai was grouped under Proto Malays based on their

physical appearance and similar culture. Semelai language is different from the Malay language and is believed to originate from Mon-Khmer language, like the Senoi language. To date, Semelai still settle around Bera Lake (Hoe, 2001). A number of genetic studies showed Semelai is closely related to Temuan based on Adenosine Deaminase-2 and Peptidase B-6 allele frequencies, closely related to Semelai based on Hemoglobin E allele frequencies, closely related to Jakun based on Glucose-6-Phosphate Dehydrogenase, Immunoglobulin G 1;21 and Immunoglobulin G 1,2;21 allele frequencies. However, Semelai is different from other Orang Asli subgroups based on Duffy A allele frequencies (Baer, 1999). Figure 7 shows there are two shared haplotypes between Semelai and Jakun and one shared haplotype between Semelai and Temuan, revealing closer relationship between Jakun-Semelai than Jakun-Temuan. Phylogeny trees show Semelai haplotypes fell in clades with Jakun and Temuan haplotypes. Despite their linguistic difference, these prove the close relationships of Jakun and Semelai.

Anthropological records indicated that Kuala migrated from Sumatra Island a few hundred years ago and speaks Indonesian dialect. Phylogeny trees (Fig. 3-5) show most of the Kuala haplotypes cluster in the last clade of the phylogeny trees. Figure 7 revealed that Kuala only has one shared haplotype with Kanaq, but seven ethnic-specific haplotypes (Table 2). Highly diverse haplotypes from 15 individuals and most distant relationships with other Proto Malay subgroups suggest the possibility of different origins of Kuala from Jakun, although, it is still not proven that Kuala originate from Sumatra Island. However pairwise differences (Table 3) shows Kuala form closer relationships with Temuan and coastal Proto Malays (Kanaq, Seletar). Social relationships between Kuala and other Proto Malays subgroups are not close. Inter-marriage with other ethnics is not encouraged in Kuala societies but the bonds between Kuala villages are strong. Therefore, it can be assume that Kuala had settled in the peninsular for a sufficiently long enough of time to allow the genetic evolved, or that the subgroup never subject genetic drift like most of the Orang Asli subgroups (Negrito and Senoi) since their arrival (Hill *et al.*, 2006).

Results from this study found that Kuala is closer to Temuan compared to Seletar or Kanaq. This finding, based on maternal inherited DNA is beyond anthropological expectation. High genetic affinity between Kuala and Temuan can be observed from the trees in Fig. 3-5 and the low pairwise differences (Table 3). However, there was no shared haplotypes between Kuala-Temuan. Temuan settlements were found to be

focused in the middle part of the peninsular. Their settlements were always surrounded by Malay and Chinese settlements in urban areas. Mixed marriages with other ethnics are also not encouraged. Could the Kuala be direct descendents from Temuan who were separated for several generations because of geographical barriers or living styles? The reason for the Kuala-Temuan relationship is yet to be proved in other aspects. Temuan and Semelai were reported to have highly diverse HVS I than other Orang Asli subgroups by Hill *et al.* (2006). Here, we detected 9 haplotypes with seven ethnic-specific haplotypes from Temuan populations, which support Hill *et al.* (2006) findings.

Kanaq actually has a closer relationship with Kuala from the maternal aspect, especially when compared with relationship of Kuala-Seletar. However, pairwise differences for the two relationships are the same. Kanaq is well-known for their isolation from the outside world and strictly prohibit mixed marriages with other ethnic group. Kanaq is presently the smallest Orang Asli tribe in Peninsular Malaysia with a population of 83 individuals. Average pairwise differences show the relationships between Kanaq and other Proto Malays are distant. opined that Kanaq is facing degeneration or genetic drift as a consequence of intramarriages within a group of only 50 of them since, World War II. The number of Kanaq increased from 50 in 1970s to 83 in 2004, but the strict intramarriage rule is still practiced today. There has been no record on intermarriage of Kanaq with any other tribe or race, including Kuala which had been resettled by the government from Batu Pahat to Kota Tinggi, near to the Kanaq settlement. The close relationship between Kuala and Kanaq support the anthropological view that Kanaq migrated from Riau Island of Indonesia more than a century ago (JHEOA, 2002). The studies of human migration based on Flint *et al.* (1989), Lum and Cann (1998) and Disotell (1999) showed human populations migrated from mainland Asia to Hawaii via Peninsular Malaysia, the Indonesian islands, Philippines, Papua New Guinea and other islands in the Pacific Ocean. Therefore, there is a possibility that Kanaq are the descendants of Kuala that had previously migrated to Riau Islands before Kanaq returning to Peninsular Malaysia during Bugis attacks in 18th century (Fig. 8).

Seletar is the second smallest Proto Malay tribe and is believed to have arrived from Singapore. Variability of Seletar haplotypes is higher compared to Kanaq. Pairwise differences (Table 3) show close relationships between Seletar and Temuan as between Kuala and Temuan. However, no shared haplotypes are found between Seletar and Temuan (Fig. 7). Intermarriage between Seletar and Jakun also caused Seletar haplotypes to be clustered

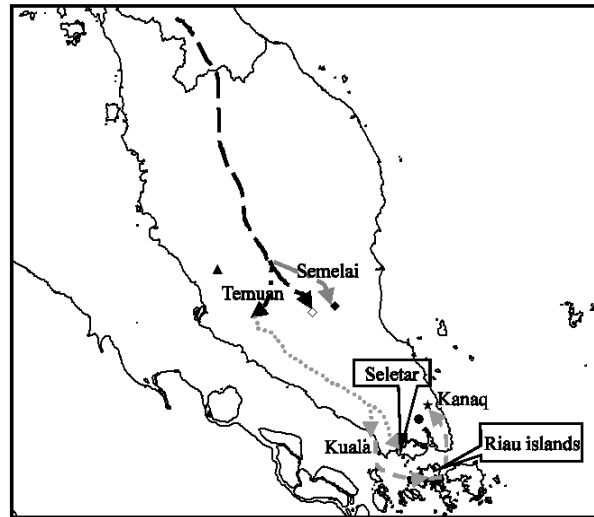


Fig. 8: The map of Peninsular Malaysia shows likely route of migration for Proto Malays

into the clade of land Proto Malays. Intermarriages with Chinese, Jakun, Semelai, Deutero-Malays and aborigines of Borneo have been recorded in Seletar populations. Seletar, also called sea nomads, once spent their whole life on the sea. To date, Seletar have been relocated to the Johor coastal areas.

The 9 bp deletions in COII/tRNA^{Lys} intergenic region had been detected in about 6 % of Semelai as well as 13.3% of Kuala. These figures approach the frequencies calculated by Ballinger *et al.* (1992) in their study to trace Mongoloid descendants based on mtDNA analysis. The study showed the 9 bp deletions in COII/tRNA^{Lys} intergenic region feature origins from middle part of China mainland which later spread through migration out of China by two main routes: one to the South through coastal areas and the other to the North across Siberia. The 9 bp deletion in COII/tRNA^{Lys} intergenic region was detected in every 32 Orang Asli (Negrito, Senoi and Proto Malays) in Peninsular Malaysia (7 Temiar, 5 Semai, 1 Jakun, 2 Jani, 17 other unknown tribes) in the study by Ballinger *et al.* (1992). Deletion rate discovered in this study is lower compared to Ballinger *et al.* (1992) which is 3:99 and only restricted to Proto Malays.

Temuan and Jakun have 7 and 6 ethnic-specific haplotypes, respectively. Jakun, Temuan and Semelai are the main tribes for Proto Malays. Temuan populations are distributed in Selangor and Negeri Sembilan whereas Jakun populations are distributed in South Pahang and Johor. Populations of both Temuan and Jakun are mainly focus in urban and suburban areas and show higher haplotype variability. Semelai, the third biggest Proto

Malay tribe with populations near to the Tasik Bera and Tasik Chini has lower haplotype variability. However, intermarriage of Jakun and Semelai with Chinese populations near both lakes is believed to have increased variability of these gene pools. Temuan on the other hand, has high haplotypes variability in even though intermarriage seldom occurred in the tribe. This may be due to intramarriages among Temuan villages from different places. High variation in Kuala haplotypes was unexpected as Kuala populations are only found near to the Tebrau Strait. However this may also be due to intramarriages and close relationships among Kuala villages situated along the south coast of the peninsular.

There have been other related studies including data based on alleles frequencies for five enzyme loci i.e., Phosphoglucosmutase I, Adenosine Deaminase, 6-Phosphoglutamate Dehydrogenase, Haptoglobin dan Transferrin done by Tan (2001) and other geneticists during 1960s. UPGMA trees constructed based on the allozyme data on 16 races from Southeast Asia placed Proto Malays in the Mongoloid main clade, as the sister taxon to Ifugao, Atayal and Bunun. The dendrogram divided 16 studied races into two main clades, with one descended from Mongoloid and the other as non-Mongoloid. Although the study did not cover Proto Malay tribes in detail, it agreed with the anthropological data that showed the Proto Malays was descended from Southern China.

The effectiveness of HVS I D-loop region segment and a 9-bp deletion of the intergenic region of COII/tRNA^{Lys} of mtDNA in detecting the phylogenetic relationships among Proto Malays: HVS I is one of the three variable regions in D-loop mtDNA (Lian, 2001; Malyarchuk *et al.*, 2002). Rapid evolution rate of mtDNA control region also caused transitions, differences in nucleotide frequencies and high variability in nucleotide substitution rate (Tamura and Nei, 1993). Polycytosine tract had been reported create bias by extend the closely related samples (Malyarchuk *et al.*, 2002).

Polycytosine tracts of Proto Malays were highly variable although about 74.04% of the samples have type I polycytosine tract (Table 2). The figure was more than that reported by Horai and Hayasaka (1990) for Caucasians, Negroid, Mongoloids and Japanese, which by 60% from 101 tested individuals posted type I polycytosine tract. Polycytosine tract type I was also detected in Deutero-Malays, Chinese and Indian populations in Peninsular Malaysia (Lim *et al.*, 2004). In addition, Jakun and Semelai formed higher variability of polycytosine tracts compared to Kanaq and Seletar (Table 2). These analysis showed polycytosine tracts

were not ethnic-specific and not suitable as sufficient informative characters are needed for MP analysis.

High ti/tv ratio was another feature of D-loop segment (Horai and Hayasaka, 1990). Analysis from the results showed two thirds of polymorphic characters from 32 haplotype sequences studied had undergone C-T transition with high ti/tv ratio, which is 31.36. The amount of parsimony informative polymorphic characters was low (5.5%). This led to the region being uninformative for maximum parsimony analysis. Bootstrap values were less due to the lack of informative characters. HVS I actually has low ti/tv ratio at higher level of divergent above species but the ratio is high among ingroup. Ti/tv ratio dropped to 7.21 when *Gorilla gorilla* (NC001645), *Pan troglodytes* (NC001643) and *Pan paniscus* (NC001644) sequences were included in the analysis as outgroups (data not shown). This result support former studies which stated that the bias by the transition-transversion will be high in the beginning but will be decreased after a certain period when tranversions have slowly accumulated (Brown *et al.*, 1982; Brown, 1985; Hixson and Brown, 1986). The same result was reported for Cytochrome b of mtDNA in *Xantusiid* lizards (Hedges *et al.*, 1991) and *Dorsophila* from *Hawaii* (DeSalle *et al.*, 1987).

High bootstrap values only obtained for sister taxa proved that HVS I evolved rapidly. HVS I was unstable and variable among species (Walberg and Clayton, 1981) as well as at individual level (Barrientos *et al.*, 1995). This can be observed from the total number of ethnic-specific haplotypes in the 89 samples. There are 18 ethnic specific haplotypes out of the total haplotypes (32) discovered from Proto Malays samples. However, HVS I from small tribes like Kanaq and Seletar were less variable. There were only 2 haplotypes from 14 Kanaq individuals and 3 haplotypes formed out of 15 Seletar individuals. Despite the bias caused by HVS I, the region successfully revealed the matrilineal relationships among six Proto Malay tribes. Bias caused by the nature of HVS I was minimized by using several approaches such as chose an appropriate algorithm and down-weighted unstable characters during phylogenetic analysis.

The mutation recovery system for mtDNA was less effective compared to other DNA molecules as it is in single copy. Phylogenetic and evolution analysis that depends on a single DNA region might not be accurate. Hence, we combined data from two regions of mtDNA for the phylogeny analysis. The 9 bp deletions occurred in COII/tRNA^{Lys} intergenic region is used to traced the Asian descendants such as Polynesian, Pacific islanders and Native Amerindians (Horai and Mutsunaga, 1986; Schurr and Wallace, 2002; Ballinger *et al.*, 1992; Fucharoen *et al.*,

2001). The deletion was also detected in African Pygmies (Viligant *et al.*, 1991; Chen *et al.*, 1995) and Brazilians with high frequencies. To date, the deletion is still considered useful in detecting the descendants of Asia, Africa and Amerindian. In this study, Proto Malays from Peninsular Malaysia which are the descendants from Asia did not show high frequencies in the 9 bp deletion in COII/tRNA^{Lys} intergenic region as expected. The deletion was only detected in a Semelai and two Kuala samples out of the 89 tested individuals. This deletion was a special feature for certain clades in the analysis for the relationships of tribes in Japan (Horai *et al.*, 1996), Thailand (Fucharoen *et al.*, 2001) and Yunan (Ya *et al.*, 2001). However, it was found to be less useful in this study.

CONCLUSION

This maternal phylogenetic study based on HVS I and the 9 bp deletions in COII/tRNA^{Lys} intergenic region of mtDNA revealed the relationships of Proto Malays is not only dictated by geographical factors but also influenced by their marriage system and the sociocultural behavior of the tribes but not by the linguistic and religion factors. Analysis of the haplotype variability indicated the tribes settled in the urban areas (Temuan) or outskirts areas (Jakun, Seletar) Peninsular Malaysia having higher variability of haplotypes.

The accuracy of the analysis can be improved by addition of data such as from the morphological and anatomy fields. Molecular data can be increased by selecting more loci including autosomal loci or through genome analysis. In addition, it is important to select the appropriate phylogeny approaches so as to increase the accuracy and reliability of the result.

Supplementary material: The sequence data for this study have been released in GenBank with accession numbers EU332755-EU332786.

ACKNOWLEDGMENTS

We thank Prof Dato' Hood Salleh from LESTARI, Dr. Adura Mohd Adnan and Dr Choong Chee Yen for technical advice, Center for Genetic Analysis and Technology UKM, Forensic Department of Hospital Universiti Kebangsaan Malaysia, Ministry of Health Malaysia, Department of Orang Asli Affairs, Malaysia. This project is sponsored by Malaysian Association for the Advancement of Medical Instrumentation and Imaging Technology (MAAMIIT) and funded by UKM research grants ST-007-2003, ST-008-2003, UKM-GUP-

ASPL-07-04-146 and IRPA 0802020019EA301 from the Ministry of Science Technology and Innovation, Malaysia.

REFERENCES

- Al-Zahery, N., O. Semino, G. Benuzzi, C. Magri, G. Passarino, A. Torroni and A.S. Santachiara-Benerecetti, 2003. Y chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-neolithic migration. *Mol. Phylo. Evol.*, 28: 458-472.
- Andaya, L.Y., 2001. The search for the origins of Melayu. *J. Southeast Asian Stud.*, 32: 315-330.
- Anderson, A.T., B.G. Bankier, M.H.L. Barrell, A.R. de Bruijin and J.D. Coulson *et al.*, 1981. Sequence and organization of the human mitochondrial genome. *Nature*, 290: 457-465.
- Ang, K.C., 2009. Sistematik molekul orang asli di semenanjung Malaysia. Ph.D. Thesis, Universiti Kebangsaan Malaysia, Bangi, Malaysia.
- Baer, A., 1999. Health, Disease and Survival: A Biomedical and Genetic Analysis of the Orang Asli of Malaysia. COAC, Kuala Lumpur.
- Ballinger, S.W., T.G. Schurr, A. Torroni, Y. Gan and J.A. Hodge *et al.*, 1992. Southeast Asia mitochondrial DNA analysis reveals genetics continuity of ancient Mongoloid migrations. *Genetics*, 130: 139-152.
- Barrientos, A., J. Casademont, A. Solans, P. Moral and F. Cardellach *et al.*, 1995. The 9-bp deletion in region V of mitochondria DNA: Evidence of mutation recurrence. *Hum. Genet.*, 96: 225-228.
- Bellwood, P., 1997. Prehistory of the Indo-Malaysian Archipelago. Univ. of Hawaii Press, Honolulu.
- Brown, W.M., E.M. Prager, A. Wang and A.C. Wilson, 1982. Mitochondrial DNA sequences of primates: Tempo and mode of evolution. *J. Mole. Evol.*, 18: 225-239.
- Brown, W.M., 1985. Evolution of the Animal mitochondrial DNA Genome. In: *Molecular Evolutionary Genetics*, Macintyre, R. (Ed.). Plenum Press, New York, pp: 95-130.
- Cann, R.L., M. Stoneking and A.C. Wilson, 1987. Mitochondrial DNA and human evolution. *Nature*, 325: 31-36.
- Chen, X., R. Prosser, S. Simonetti, J. Sadlock, G. Jagiello and E.A. Schon, 1995. Rearranged mitochondrial genomes are present in human oocytes. *Am. J. Hum. Genet.*, 57: 239-247.
- DeSalle, R., T. Freedman, E.M. Prager and A.C. Wilson, 1987. Tempo and mode of sequence evolution in mitochondrial DNA of hawaiian *Drosophila*. *J. Mol. Evol.*, 26: 157-164.

- Disotell, T.R., 1999. Human evolution: The Southern route to Asia. *Curr. Biol.*, 9: 925-928.
- Endicott, K., 2003. Indigenous Rights Issues in Malaysia. In: *At Risk of Being Heard: Identity, Indigenous Rights and Post-Colonial States*, Dean, B. and J.M. Levy (Eds.). University of Michigan Press, USA.
- Fix, A.G., 1995. Malayan paleosociology: Implications for patterns of genetic variation among the orang asli. *Am. Anthropol.*, 97: 313-323.
- Flint, J., A.J. Boycem, J.J. Martinson and J.B. Clegg, 1989. Population bottlenecks in polynesia revealed by minisatellites. *Hum. Genet.*, 83: 257-263.
- Forster P., A. Torroni, C. Renfrew and A. Röhl, 2001. Phylogenetic stars contraction applied to Asian and papuan mtDNA evolution. *Mol. Biol. Evol.*, 18: 1864-1881.
- Fucharoen, G., S. Fucharoen and S. Horai, 2001. Mitochondrial DNA polymorphisms in Thailand. *J. Hum. Genet.*, 46: 115-125.
- Hasegawa, M., H. Kishino and T. Yano, 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.*, 22: 160-174.
- Hedges, S.B., R.L. Bezy and L.R. Maxson, 1991. Phylogenetic relationship and biogeography of xantusiid Lizards, inferred from mitochondrial DNA sequences. *Mol. Biol. Evol.*, 8: 767-780.
- Hill, C., P. Soares, M. Mormina, V. Macaulay and W. Meehan *et al.*, 2006. Phylogeography and ethnogenesis of aboriginal southeast Asians. *Mol. Biol. Evol.*, 23: 2480-2491.
- Hillis, D.M., C. Moritz and B.K. Mable, 1996. *Molecular Systematics*. Sinauer Associates, USA.
- Hixson, J.E. and W.M. Brown, 1986. A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: Sequence divergence, structure, evolution and phylogenetic implications. *Mol. Biol. Evol.*, 3: 1-18.
- Hoe, B.S., 2001. *Semelai Communities at Tasek Bera: A Study of the Structure of an Orang Asli Society*. Center for Orang Asli Concerns, Subang Jaya, Malaysia.
- Hood, S., 2006. *People and Tradition: The Orang Asli: Origins, Identity and Classification*. Archipelago Press, Kuala Lumpur.
- Horai, S. and E. Mutsunaga, 1986. Mitochondrial DNA polymorphism in Japanese II analysis with restriction enzymes of four or five base pair recognition. *Hum. Genet.*, 72: 105-117.
- Horai, S. and K. Hayasaka, 1990. Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *Am. J. Hum. Genet.*, 46: 828-842.
- Horai, S., K. Murayama, K. Hayasaka, S. Matsubayashi and Y. Hattori *et al.*, 1996. MtDNA polymorphism in east Asian populations, with special reference to the peopling of Japan. *Am. J. Hum. Genet.*, 59: 579-590.
- Hussein, T., S. Ibrahim and R. Alias, 2007. *Malaysia Negara Kita*. 2nd Edn., MDC Co., Kuala Lumpur.
- Idris, J., 1968. Distribution of orang asli in Malaysia. *JFMSM.*, 8: 45-48.
- JHEOA. (Jabatan Hal Ehwal Orang Asli), 2002. *Kehidupan, budaya dan pantang larang orang asli*. Department of Orang Asli, Kuala Lumpur.
- Kashyap, V.K., T. Sitalaximi, B.N. Sarkar and R. Trivedi, 2003. Molecular relatedness of the aboriginal groups of andaman and nicobar islands with similar ethnic populations. *Int. J. Hum. Genet.*, 3: 5-11.
- Kasimin, A., 1991. *Religion and Social Change Among the Indigenous People of the Malay Peninsular*. Dewan Bahasa dan Pustaka, Kuala Lumpur, ISBN 10: 9836222650, pp: 326.
- Kumar, S., K. Tamura and M. Nei, 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.*, 5: 150-163.
- Lian, L.H., 2001. *Genetic polymorphism at the non-coding segments in human mitochondria DNA from Malaysia*. Ph.D. Thesis, Universiti Malaya, Kuala Lumpur, Malaysia.
- Lim, L.S., B.M. Md-Zain, M.C. Mahami and A.W. Shahrom, 2004. HVSI as a tool in phylogenetic analysis among human populations in Peninsular Malaysia. *Malaysian J. Biochem. Mol. Biol.*, 10: 1-8.
- Lum, J.K. and R.L. Cann, 1998. MtDNA and language support a common origin of micronesians and polynesians in island southeast Asia. *Am. J. Phys. Anthropol.*, 105: 109-119.
- Lutz, S., H.J. Weisser, J. Heizmann and S. Pollak, 1998. Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int. J. Legal Med.*, 111: 67-77.
- Lye, T.P., 2001. *Orang Asli of peninsular Malaysia: A Comprehensive and Annotated Bibliography*. Kyoto University, Kyoto, Japan.
- Macaulay, V., C. Hill, A. Achilli, C. Rengo and D. Clarke *et al.*, 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genome. *Science*, 308: 1034-1036.
- Maddison, W.P. and D.R. Maddison, 1992. *McClade: Analysis of Phylogeneny and Character Evolution*. Sinauer Associates Inc., Massachusetts, pp: 159-194.
- Malyarchuk, B.A., I.B. Rogozin, V.B. Berikov and M.V. Derenko, 2002. Analysis of phylogenetically reconstructed mutational spectra in human mitochondria DNA control region. *Hum. Genet.*, 111: 46-53.

- Nei, M. and S. Kumar, 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
- Nicholas, C., 1996. The Orang Asli of Peninsular Malaysia. In: *Indigenous Peoples of Asia: Many Peoples, One Struggle*, Nicholas, C. and R. Singh (Eds.). Asia Indigenous Peoples Pact, Bangkok, pp: 157-156.
- Nicholas, C., 2000. The Orang Asli and the Contest for Resources: Indigenous Politics, Development and Identity in Peninsular Malaysia. International Work Group on Indigenous Affairs (IWGIA), Denmark.
- Rozas, J., J.C. Sánchez-Delbarrio, X. Messeguer and R. Rozas, 2003. DnaSP, DNA polymorphic analyses by the coalescent and other methods. *Bioinformatics*, 19: 2496-2497.
- Saiki, R.K., 1989. The Design and Optimization of the PCR. In: *PCR Technology: Principles and Applications for DNA Amplification*, Erlich, H.A. (Ed.). Stockton Press, New York, pp: 7-16.
- Sainuddin, S., 2003. *Titas Tamadun Melayu*. Quantum Books, Perak.
- Sambrook, J. and W.D. Russell, 2001. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York, pp: A8.9-A8.10.
- Schurr, T.G. and D.C. Wallace, 2002. Mitochondrial DNA diversity in southeast Asian populations. *Hum. Biol.*, 74: 431-452.
- Swofford, D.L., 2002. *Phylogenetic Analysis Using Parsimony and Other Methods** Version 4.0 Beta Version. Sinauer Associates, Sunderland.
- Tamura, K. and M. Nei, 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.*, 10: 512-526.
- Tan, S.G., 2001. Genetic Relationships Among Sixteen Ethnic Groups from Malaysia and Southeast Asia. In: *Genetic, Linguistic and Archeological Perspectives on Human Diversity in Southeast Asia*, Jin, Seielstad, Xiao, (Eds.). World Scientific, New Jersey, pp: 83-92.
- Viligant, L., M. Stoneking, H. Harpending, K. Hawkes and A.C. Wilson, 1991. African populations and the evolution of mitochondrial DNA. *Science*, 253: 1503-1507.
- Walberg, M.W. and D.A. Clayton, 1981. Sequence and properties of the human KB cell and mouse L Cell D-loop regions of mitochondrial DNA. *Nucleic Acids Res.*, 9: 5411-5421.
- Ya, P.Q., Z.T. Chu, Q. Dai, C.D. Wei, J.Y. Chu, A. Tajima and S. Horai, 2001. Mitochondrial DNA polymorphisms in Yunnan nationalities in China. *J. Hum. Genet.*, 46: 211-220.