

## HAPLOTYPE ANALYSES OF *ORANG ASLI* POPULATION IN TAMAN NEGARA PAHANG USING Y-STR MARKERS

SOFIA, S.M.Y.<sup>1</sup>, BADRUL, M.M.Z.<sup>2</sup> and FARIDA, Z.M.Y.<sup>1\*</sup>

<sup>1</sup>Faculty of Applied Sciences, Universiti Teknologi MARA (UiTM),  
40450 Shah Alam, Selangor, Malaysia

<sup>2</sup>Faculty of Science and Technology, Universiti Kebangsaan Malaysia,  
43600 Bangi, Selangor

\*Email: [fzuraina@salam.uitm.edu.my](mailto:fzuraina@salam.uitm.edu.my)

### ABSTRACT

There are many maternal line studies conducted for *Orang Asli* using mitochondrial DNA. This study focuses on the paternal line, specifically Y-STR and three markers were considered, namely DYS19, DYS390 and DYS391. These markers were used to compare 21 subjects comprising of 7 from Senoi tribes and 14 from Negrito tribes. The samples were collected from Kampung Kuala Atok, Kg. Sungai Tiang, Kg. Dedari, Kg. Krom, Kg. Tereseke and Kg. Gam located in Taman Negara Pahang and villages nearby. From these, 14 haplotypes were gained with one of it shared by both tribes while the rest were unique to the tribe. The analyses conducted from the haplotype were gene diversity (GD), locus diversity (DL), haplotype diversity (HD) and discrimination capacity (DC). From the analysis, the GD values range from 0.2480 to 0.7108 whilst the DL values range from 0.2500 to 0.7446. The haplotype diversity and discrimination capacity values were 0.8875 and 0.6364, respectively. Since this work only studied the comparison between Senoi and Negrito, further analysis is going to include the other sub tribes of *Orang Asli*.

**Key words:** *Orang Asli*, Y-STR, DYS19, DYS390, DYS391, Senoi, Negrito

### INTRODUCTION

The aborigines of Peninsular Malaysia, also known as *Orang Asli* makes up 0.5% of the total population. They are traditionally divided into three main groups which are Negrito, Senoi and Proto-Malay. Each of the tribes can be divided further into six subtribes, making a total of 18 subtribes of *Orang Asli* in Peninsular Malaysia (Nicholas, 2003). Studies from different point of views have been conducted (Lye, 2011). Some non-molecular studies include the historical approach by Bellwood and the educational feature of *Orang Asli* by Abdullah and colleagues (Bellwood, 1993; Abdullah *et al.*, 2013). Molecular studies that have been done on *Orang Asli* populations include the work on mitochondrial DNA sequences and polymorphism (Zainuddin and Goodwin, 2004; Lim *et al.*, 2010). However, not many studies have been done on Y-STR of *Orang Asli* (Alshuridin *et al.*, 2011). Although there is lack of studies on Y-STR of aborigines in Malaysia, they are quite common worldwide. Some of the studies of Y-STR were done on local aborigines such as the Paiwan people of

Taiwan, Papuans and Australian aborigines (Forster *et al.*, 2008; Wu *et al.*, 2011). Microsatellites are made up of repeat motifs arranged tandemly and considered to be one of the most commonly used PCR-based markers (Jarne and Lagoda, 1996). Y-STRs are microsatellites on the non-recombining region of Y chromosome that do not code for proteins (Willard, 2003). Since Y-STR escapes recombination, nucleotide differences between a father and son can only be gained through mutation (Underhill *et al.*, 2000). Compared to recombining loci, it is important for Y-STR to be analyzed using more locus to increase discrimination capacity (Bosch *et al.*, 2002). In this work, three markers were used namely DYS 19, DYS 390 and DYS 391 as shown in Table 1 (Butler, 2003; Schoske, 2003). These three markers were part of the SWGDAM-recommended minimal haplotype markers.

### MATERIALS AND METHODS

#### Sample collection and FTA card washing

The raw materials used in the study were buccal swab samples of *Orang Asli*. Samples were collected after ethical approval was obtained from the

\* To whom correspondence should be addressed.

**Table 1.** Information of SWGDAM recommended Y-STR markers used in the study

Marker name	Repeat motif	Allele range	PCR product sizes	GenBank® accession
DYS19	TAGA	10 – 19	214–250 bp	AC017019
DYS390	(TCTA) (TCTG)	17 – 28	191–227 bp	AC011289
DYS391	TCTA	6 – 14	91–119 bp	AC011302

**Table 2.** Forward and reverse primer sequences for minimal haplotype Y-STR

Locus	Primer Sequences (5'-to-3')	T <sub>m</sub> (°C)	T <sub>a</sub> (°C)
DYS19	F: ACTACTGAGTTTCTGTTATAGTGTTTTT	55.0	58.3
	R: GTCAATCTCTGCACCTGGAAAT	60.5	
DYS390	F: TATATTTTACACATTTTGGGCC	57.2	46.8
	R: GTGACAGTAAAATGAAAACATTGC	57.0	
DYS391	F: TTCAATCATACCCCATATCTGTC	57.9	48.0
	R: GATAGAGGGATAGGTAGGCAGGC	60.0	

Research Management Institute (RMI) of Universiti Teknologi MARA (UiTM), Jabatan Kemajuan *Orang Asli* (JAKOA) and National Medical Research Register (NMRR). Formal and informed consent were acquired from *Orang Asli* subjects from the selected areas which were Kampung (Kg.) Kuala Atok, Kg. Sungai Tiang, Kg. Dedari, Kg. Krom, Kg. Teresek and Kg. Gam. Only male *Orang Asli* with pure lineage was selected for this project. Cheek cells were obtained using buccal swab applicator (Whatman, USA). The tip of the applicator was placed inside the mouth. The inside of the right cheek was swabbed using the applicator for 30 seconds. The tip was then rubbed on the tongue and under it to soak as much saliva as possible so the cells can be transferred on the card. The tip was pressed gently on the marked area of the FTA card. The steps were repeated using new sterile applicator for the left side of the cheek. The cards were kept dry and stored until they were transferred back to lab for further analysis. Before use for any molecular work, FTA card discs were washed using 200 µL FTA reagents and rinsed with TE buffer, pH 8.0 before being dried for an hour.

#### **Polymerase Chain Reaction(PCR) amplification and DNA visualization**

An amount of 25 µL of PCR component mix (Solis BioDyne, Estonia and New England Biolabs) was added to the card and run through a thermocycler (BioRad). During PCR, DNA were denatured for three minutes at 94°C followed by 30 cycles starting with 94°C for 30 seconds, appropriate annealing temperature that might differs

for each marker for another 30 seconds, and 90 seconds of extension at 72°C. Then, it was followed by final extension at 72°C for five minutes. The samples were stored at 4°C until further used. Primer sequence, melting temperature (T<sub>m</sub>) and annealing temperature (T<sub>a</sub>) used are shown in Table 2. The PCR products went through 2.0% agarose gel electrophoresis for half an hour at 100 V. After electrophoresis, the gels were visualized under UV. The images obtained were saved for recording purpose.

#### **DNA sequencing and alignments**

The PCR products that show positive and clear bands during visualization were sent for DNA sequencing (MyTACG Bioscience Enterprise). After both forward and reverse sequences of samples were obtained, the sequences were aligned using online software, ClustalW (Ver 2.0). The aligned sequences obtained were analysed with BLAST to make sure that the samples were correct.

#### **Statistical analysis**

The numbers of repetition of the repeat motifs were recorded to construct haplotype data. Haplotype frequencies were obtained using simple calculation. Gene diversity, locus diversity, haplotype diversity and discrimination capacity were calculated.

Gene diversity (GD) is the degree of probability for two randomly chosen groups to be non-identical. It was calculated as the formula  $GD = 1 - \sum p_i^2$  where  $p_i$  represents the frequency of  $i^{\text{th}}$  allele in a population (Nei, 1987).

Locus diversity (LD) is the average heterozygosity or GD per locus and can be used to measure genetic variability in a population. LD was calculated as the formula  $LD = \frac{n}{n-1}(1 - \sum p_i^2)$  where  $n$  is the number of individuals in a given population and  $p_i$  is corresponding allele frequency (Nei, 1987).

Haplotype diversity, HD is the measure of diversity for a sample haplotype. The formula is  $HD = \frac{n}{n-1}(1 - \sum p_i^2)$  where  $n$  is the number of individuals in a given population and  $p_i$  represents the haplotype frequency (Coble *et al.*, 2013).

Discrimination capacity (DC) is the measure of ability of a haplotype to differentiate from one another. The formula used is  $DC = \frac{H}{n}$  where  $H$  is the number of haplotypes obtained while  $n$  is the number of samples in the population studied.

## RESULTS AND DISCUSSION

### Haplotype of *Orang Asli* using three markers of Y-STR

After all the number of repeats for each samples were found, haplotypes for *Orang Asli* using Y-STR were compiled. For two samples to have the same haplotype, all the number of repeats for each marker for the two samples has to be the same. Since Y-STR is highly polymorphic, large numbers of haplotypes were expected. Table 3 represents the haplotypes of *Orang Asli* based on the samples analyzed.

From the total of 21 samples used, 14 haplotypes were found. There were ten of the haplotypes that were unique to individual, three were shared by two individuals and one was shared by five individuals. From the shared haplotype, three were unique for Negrito tribe namely haplotype 2, 3 and 10. More markers should be used to make more specific

conclusion but from the present data, it can be said that number of alleles of Y-STR can be used to discriminate tribes. The number of haplotype acquired is low due to the small number of markers. The higher the number or markers used in constructing a haplotype profile, the more unique haplotypes will be produced (Calderon *et al.*, 2013).

### Statistical analysis

The summary of the results for gene diversity (GD), allele frequencies (p), locus diversity (LD), standard error (SE) are shown in Table 4. From the calculation of haplotype diversity (HD) and discrimination capacity (DC), the values were  $0.8875 \pm 0.0369$  and  $0.6364$  respectively.

The alleles are in the range of worldwide studies that have been recorded in YHRD database. However, the highest allele frequency for DYS 19 was for allele with 16 repeats while the highest in the database is for repeat 14. Similarly for DYS 390 and DYS 391, the allele with the highest frequency in this study, repeat 28 and repeat 8 respectively are of the lowest worldwide (Willuweit and Roewer, 2007). This is possibly because of the DNA uniqueness of *Orang Asli*. However, more samples are required for more accurate values and to reduce small sample number bias.

Both genetic and locus diversity have similar values and can be discussed together. The GD values from this study were 0.2480, 0.7108 and 0.4917 for DYS19, DYS390 and DYS391 respectively. GD of DYS 19 in this study can be thought as very low relative to the other populations. When the mutation rate is taken into consideration, DYS390 has the lowest mutation rate followed by DYS19 and finally DYS391 (Willuweit and Roewer, 2007). Since the rate for DYS390 is the lowest compared to DYS 19, it should provide lesser heterozygosity since the

**Table 3.** Haplotypes of *Orang Asli* for three of Y-STR minimal haplotype markers

Haplotype	DYS19	DYS390	DYS391	Number of sample (n)	Frequency (f)
1	16	29	9	1	0.0454
2	16	29	8	2	0.0909
3	16	30	8	2	0.0909
4	16	32	8	1	0.0454
5	15	28	8	1	0.0454
6	16	27	8	1	0.0454
7	16	28	8	5	0.2273
8	14	28	9	1	0.0454
9	16	27	9	1	0.0454
10	16	30	9	2	0.0909
11	16	31	8	1	0.0454
12	16	28	10	1	0.0454
13	16	28	9	1	0.0454
14	18	30	8	1	0.0454

**Table 4.** Summaries of the statistical analysis

Allele	n	p	GD	LD	SE
DYS 19					
14	1	0.0454			
15	1	0.0454			
16	19	0.8636	0.2480	0.2500	0.0916
18	1	0.0454			
DYS 390					
27	2	0.0909			
28	10	0.4545			
29	3	0.1364			
30	5	0.2273	0.7108	0.7446	0.0777
31	1	0.0454			
32	1	0.0454			
DYS 391					
8	14	0.6364			
9	7	0.3182	0.4917	0.5151	0.1050
10	1	0.0454			

DNA should be more preserved. To add, they also have longer repeat motif, suggesting them to be more stable than DYS 19. However this finding is in harmony with the findings of Santos and colleagues where in their study, DYS19 was unusually stable relative to the high mutation nature of microsatellites (Santos *et al.*, 1996).

High GD value for DYS390 is in agreement with worldwide study of Y-STR. The high diversity of this particular marker might be caused by the combined repeat motif and the long repetition of repeat motifs. The GD value for DYS391 is a lot higher than DYS19 in this study and that was not in congruence with the previous studies. However, among those three markers, DYS 391 has the highest mutation rate, hence the variation of GD worldwide (Hashiyada *et al.*, 2006; Coble *et al.*, 2013). To condense, DYS19 has the lowest GD and DL values followed by DYS391 and DYS390. GD of DYS19 is not in agreement with other populations but is comparable to the previous study of *Orang Asli*. This proposes unique characteristic of Malaysian aborigines. However, to answer the relationship between this study and the previous study, aside from the environmental factors, the discrepancies between the two are most probably due to small sample size.

The overall HD was 0.8875 while the HD for Negrito and Senoi were 0.8673 and 0.8163 respectively. It shows that the overall HD was higher compared to per tribe. This supports the idea that subjects in a small tribe is more similar to one another compared to in a larger population. The HD values were very low and might be attributed to the small sample size and Y-STR markers (Alam *et al.*, 2010).

The DC of the haplotype profile was very low. The value could be greater by increasing the number of markers used while building the haplotype (Wu *et al.*, 2011). This is because the more number of markers used, the more unique haplotypes can be obtained (Pontes *et al.*, 2007).

## CONCLUSIONS

To summarize, the analyses done on the three markers of Y-STR for two tribes of *Orang Asli* in Pahang, Negrito and Senoi showed that different markers can provide different information about a population. Since the result of this study does not follow the linear relationship between mutation rate and GD, it is unlikely that mutation rate is the only cause of GD. To get more significant result and reduce bias, more markers should be used for comparisons and more samples from different locations should be used to find out whether the distinction between individuals are more influenced by the genetic composition (tribes) or gene flow (migration). It would be valuable if the sample size can be increased in future studies and to make sure it is more exhaustive so that analysis can be done not only between three tribes but also between 18 subtribes.

## ACKNOWLEDGEMENTS

This study was supported by research grant from Ministry of Science, Technology and Innovation (MOSTI), Malaysia [600-RMI/FRGS TD 5/3 (1/2013)]. The authors would like to thank RMI for

granting the ethical approval, Jabatan Kemajuan Orang Asli (JAKOA), National Medical Research Register (NMRR) and Universiti Teknologi MARA for their continuous support. Special thanks to villagers and all *Orang Asli* who have helped in this research.

## REFERENCES

- Abdullah, R., Mamat, W.H.W., Amir Zal, W.A. & Ibrahim, A.M. 2013. Teaching and Learning Problems of the Orang Asli Education: Students' Perspective. *Asian Social Science*, **9(12)**.
- Alam, S., Ali, M.E., Ferdous, A., Hossain, T., Hasan, M.M. & Akhteruzzaman, S. 2010. Haplotype diversity of 17 Y-chromosomal STR loci in the Bangladeshi population. *Forensic Science International: Genetics*, **4(2)**, e59-e60.
- Alshuridin, S.H.N., Yusof, F.Z.M., Hanafi, W.N.W. & Sohari, R.R. 2011. Tracing Human Origin Among Orang Asli Population in Taman Negara Through Y-Chromosome Variation. *IEEE Colloquium on Humanities, Science and Engineering Research*, 892-896.
- Bellwood, P. 1993. Cultural and biological differentiation in Peninsular Malaysia: the last 10,000 years. *Asian Perspectives*, **32(1)**: 37-60
- Bosch, E., Lee, A.C., Calafell, F., Arroyo, E., Henneman, P., de Knijff, P. & Jobling, M.A. 2002. High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions. *Forensic Science International*, **125(1)**: 42-51.
- Butler, J.M. 2003. Recent Development in Y-Short Tandem Repeat and Y-Single Nucleotide Polymorphism Analysis. *Forensic Science Review*, **15(2)**: 91-111.
- Calderon, S., Perez-Benedico, D., Mesa, L., Guyton, D., Rowold, D.J. & Herrera, R.J. 2013. Phylogenetic and forensic studies of the Southeast Florida Hispanic population using the next-generation forensic PowerPlex® Y23 STR marker system. *Legal Medicine*, **15(6)**: 289-292.
- Coble, M.D., Hill, C.R. & Butler, J.M. 2013. Haplotype data for 23 Y-chromosome markers in four US population groups. *Forensic Science International: Genetics*, **7(3)**, e66-e68.
- Forster, P., Kayser, M., Meyer, E., Roewer, L., Pfeiffer, H., Benkmann, H. & Brinkmann, B. 1998. Phylogenetic resolution of complex mutational features at Y-STR DYS390 in aboriginal Australians and Papuans. *Molecular Biology and Evolution*, **15(9)**: 1108-1114.
- Hashiyada, M., Nagashima, T., Itakura, Y., Sakai, J., Kanawaku, Y., Kanetake, J. & Funayama, M. 2006. 12 Y-chromosomal STR haplotypes in Japanese. *Forensic Science International*, **158(2)**: 204-212.
- Jarne, P. & Lagoda, P.J.L. 1996. Microsatellites, from Molecules to Population and Back. *Trends in Ecology and Evolution*, **11(10)**: 424-429.
- Lim, L.S., Ang, K.C., Mahani, M.C., Shahrom, A.W. & Md-Zain, B.M. 2010. Mitochondrial DNA Polymorphism and Phylogenetic Relationships of Proto Malays in Peninsular Malaysia. *Journal of Biological Sciences*, **10(2)**: 71-83.
- Lye, T-P. 2011. A history of Orang Asli Studies: Landmarks and Generations. *Kajian Malaysia*, **29(1)**: 23-52.
- Nei, M. 1987. Genetic distance and molecular phylogeny. *Population Genetics and Fishery Management*. Seattle. USA, 193-223.
- Nicholas, C. 2004. The Orang Asli: First on the Land, Last in the Plan. *Kajian Malaysia*, **21(1&2)**: 315-329.
- Pontes, M.L., Cainé, L., Abrantes, D., Lima, G. & Pinheiro, M.F. 2007. Allele frequencies and population data for 17 Y-STR loci (AmpF/STR® Y-filer™) in a Northern Portuguese population sample. *Forensic Science International*, **170(1)**: 62-67.
- Santos, F.R., Gerelsaikhon, T., Munkhtuja, B., Oyunsuren, T., Epplen, J.T. & Pena, S.D. 1996. Geographic differences in the allele frequencies of the human Y-linked tetranucleotide polymorphism DYS19. *Human Genetics*, **97(3)**: 309-313.
- Schoske, R. 2003. *The Design, Optimization and Testing of Y Chromosome Short Tandem Repeat Megaplexes*. Unpublished doctoral dissertation, American University, Washington D. C.
- Underhill, P.A., Shen, P., Lin, A.A., Jin, L., Passarino, G., Yang, W.H. & Oefner, P.J. 2000. Y chromosome sequence variation and the history of human populations. *Nature Genetics*, **26(3)**: 358-361.
- Willard, H.F. 2003. Genome biology: Tales of the Y chromosome. *Nature*, **423(6942)**: 810-813.
- Willuweit, S. & Roewer, L. 2007. Y chromosome haplotype reference database (YHRD): update. *Forensic Science International: Genetics*, **1(2)**: 83-87.
- Wu, F.C., Ho, C.W., Pu, C.E., Hu, K.Y., Willuweit, S., Roewer, L. & Liu, D.H. 2011. Y-chromosomal STRs haplotypes in the Taiwanese Paiwan population. *International Journal of Legal Medicine*, **125(1)**: 39-43.
- Zainuddin, Z. & Goodwin, W. 2004. Mitochondrial DNA profiling of modern Malay and Orang Asli populations in peninsular Malaysia. *International Congress Series*, **1261**: 428-430.

