

Preliminary Analysis of Y-STR Polymorphism among Temiar sub-tribe of *Orang Asli* in Kelantan

Marlia Marzukhi¹, Sofia Sakina M Yussup¹, Siti Saizah Md Said¹, Farida Zuraina M Yusof¹

¹Faculty of Applied Sciences, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, MALAYSIA

Abstract: *Orang Asli* is the Bumiputera group that can be found in Peninsular Malaysia and they are divided into various sub-tribes based on their physical appearance, culture and their settlement. Temiar is one of the *Orang Asli* sub-tribes that can be found mostly in Perak, Pahang and Kelantan. Fifty unrelated males individuals from the Temiar sub-tribes from two villages in Kelantan were analyzed using five Y-STR loci (DYS385a/b, DYS389I/II and DYS393). Among the 50 individuals studied, 20 different haplotypes were observed and 11 haplotypes were found to be unique, while nine haplotypes were shared by more than one individual. The overall haplotype diversity for the five Y-STR tested was 0.902 ($SE \pm 0.015$) and the discrimination capacity was 0.400. The locus single locus diversity ranged from the highest in DYS389a/b with a value of 0.861 to the lowest of 0.343 ($SE \pm 0.054$) in DYS393. DNA sequences from the selected samples were then used to develop two types of phylogenetic trees; a character-based method and a distance-based method.

Keywords: DYS385a/b, DYS389I/II, DYS393, *Orang Asli*, Temiar, Kelantan

1. Introduction

Orang Asli or aborigines are the indigenous minority people of Peninsular Malaysia and they are divided into three main tribes which are Negrito, Senoi and Proto-Malay. Each of the main group is further divided into six different sub-tribes [1]. Senoi is the largest tribe of *Orang Asli* and consists of six sub-tribes namely Temiar, Semai, Semoq Beri, Jahut Mah Meri and Che Wong. They are found along the slope of the Titiwangsa Mountain in the rural areas of Perak, Pahang and Kelantan [2]. They use an Aslian language in their communication suggesting that they are historically linked with the tribal people of Burma, Thailand and Indo-China [3]. Members of the Temiar sub-tribes mostly make a living through activities such as hunting and gathering, swidden agriculture and free trade activity. They are famous for their dream theory where they use their dreams to guide them in their social and political life. However, after British colonization, they begin to abandon their traditional ways and start to embrace different religions such as Islam and Christianity [4].

There are many markers that has been used in the studies of population genetics. One of the most common used marker is Y-chromosome Short Tandem Repeats (Y-STR). In human genome, Y-chromosome is only present in a single copy in males and passes from father to son while being absence in females. Most of the regions on the Y-chromosome escape recombination, and the important feature of escaping the recombination is the haplotypes that usually remain constant from generation to generation except if there is a mutation [5]. To date, there are more than 100 Y-STR markers that have been used to study paternal lineage [6]. They are used in differentiation between individuals due to their high polymorphic nature. However, there has yet any studies to be done on the phylogeny of *Orang Asli* using Y-STR markers.

This study was carried out to explore the potential of Y-STR markers in the phylogenetic study of *Orang Asli*, specifically the Temiar population in Kelantan. It is also to study the pattern of polymorphism of selected Y-STR markers namely DYS385a/b, DYS389I/II and DYS393 in an isolated population. However, more markers and subjects should be included in future studies.

2. Material and methods

2.1. DNA sample

Samples were collected after obtaining an ethical approval from the Institute Research Management and Innovation (IRMI) of Universiti Teknologi MARA (UiTM), Department of *Orang Asli* Development (JAKOA) and the National Medical Research Register (NMRR). Formal and informed consent were acquired from the *Orang Asli* subjects from two villages in Kelantan and only the one with a pure lineage of *Orang Asli* were selected. Buccal cell samples were collected from 50 unrelated individuals of the Temiar sub-sub-tribes. Sterile swab was rubbed against the inner side of the right cheek for approximately 30 seconds and pressed on the FTA[®] card (Whatman Bioscience). The process was repeated with the left side of the cheek using a fresh swab. FTA[®] cards were left to dry for at least an hour before storage for further use in the lab.

2.2. FTA[®] card purification

The card was punched into a disc of 2mm diameter. The disc was then washed with 200µl of FTA purification reagent. After 5 minutes, the reagent was discarded and followed by rinsing of the disc using 200µl of TE buffer (pH 8.0). The discs were left to dry for at least an hour at room temperature for further analysis.

2.3. PCR amplification

The selected markers which are DYS385a/b, DYS389I/II and DYS393 were amplified using Polymerase Chain Reaction (PCR). Definite pair of markers were used for each of the selected marker. An amount of 25µl total volume per reaction was used. The mix includes 5µl 5X HOT Firepol[®] Blend Master Mix (Solis BioDyne), 0.4µl of each forward and reverse primer, FTA card disc and nuclease-free water. The sample was then amplified in the thermocycler (Bio-Rad). The DNA profile for the PCR reaction were; denaturation of DNA for 13 minutes and 30 seconds at 95°C followed by 30 cycles starting with 95°C for 15 seconds, optimum temperature for each marker for another 45 seconds, and 150 seconds of extension at 72°C. It was followed by a final extension at 72°C for seven minutes and 30 seconds. The samples were stored at 4°C for further use.

2.4. DNA fragment analysis and DNA sequencing

The PCR products were run using 2% of agarose gel in (80V) for size confirmation. All samples were sent for fragment analysis. The fragment analysis was conducted using the Applied Biosystem Genetic Analyzer and interpreted using a GeneMapper[®] 4.0 analysis software. The DNA sequencing was conducted using the Applied Biosystems 3730XL Genetic Analyzer and were analysed using the Applied Biosystems DNA Sequencing Analysis Software v5.2.

2.5. Statistical analysis

Data from the DNA fragment analysis were used to calculate the allelic frequency (p_i), gene diversity (GD), haplotype diversity (D_H), locus diversity (D_L) and discrimination capacity (DC) for the selected markers. Allelic and haplotype frequencies were obtained using a simple frequency formula. D_H and D_L were calculated using the formula by Nei [7] as shown in (1) where the p_i values for D_H and D_L were haplotype frequency and allele frequency respectively and n is the number of individuals.

$$D = \frac{n}{n-1} \left(1 - \sum p_i^2 \right) \quad (1)$$

Gene diversity or heterozygosity were calculated using the formula by Nei [7] as shown in (2) where p_i represents the haplotype frequency.

$$GD = 1 - \sum p_i^2 \quad (2)$$

Discrimination capacity was estimated as the number of individual-specific (unique) haplotypes observed divided by the number of individuals typed.

2.6. Phylogenetic tree construction

Forward and reverse DNA sequences were first aligned using online program Clustal Omega and then analyzed with Basic Local Alignment Search Tool (BLAST). The aligned sequences were then used to develop

Neighbor-Joining and Maximum-Likelihood tree using MEGA6 software. The trees constructed were then subjected to 1000 bootstrap replication to test tree reliability.

3. Results and Discussion

3.1. Statistical analysis

From the 50 unrelated Temiar individuals analysed, there were 20 different haplotypes observed. Among those haplotypes, 11 were found to be unique, which means they were only found in one individual. The other nine haplotypes were found to be shared by more than one individual. The most common shared haplotype was Haplotype 6 as this haplotype was found in 11 individuals. The haplotypes were shown in detail in Table I, where Ht is the haplotype, n is the number of individual and p_i is the haplotype frequencies.

TABLE I: The Haplotype and Haplotype Frequency of DYS385a/b, DYS389I/II and DYS393

Ht	DYS385a/b	DYS389I	DYS389II	DYS393	n	p_i
1	9,12	13	30	13	2	0.040
2	10,11	13	28	14	1	0.020
3	10,11	14	29	14	1	0.020
4	12,13	13	29	14	1	0.020
5	12,17	12	28	12	3	0.060
6	12,18	13	29	14	11	0.220
7	12,18	13	30	14	1	0.020
8	13,16	13	29	13	1	0.020
9	13,16	13	30	14	1	0.020
10	13,16	14	30	13	2	0.040
11	13,17	13	29	13	1	0.020
12	13,17	13	29	14	2	0.040
13	13,17	13	30	14	1	0.020
14	13,17	13	31	13	1	0.020
15	13,18	13	29	14	5	0.100
16	13,18	13	30	14	4	0.080
17	13,19	12	28	14	1	0.020
18	13,19	12	30	14	1	0.020
19	14,18	13	28	14	7	0.140
20	14,18	14	29	14	3	0.060

The allele frequency, GD, D_L and SE for DYS389I/II and DYS393 are shown in Table II. The allele frequency, GD, D_L and SE for DYS385a/b are shown separately in Table III.

TABLE II: Allele frequency, GD, D_L and SE for DYS389I/II and DYS393

Allele	DYS389I	p_i	DYS389II	p_i	DYS393	p_i
12	5	0.100			3	0.060
13	39	0.780			7	0.140
14	6	0.120			40	0.800
28			12	0.240		
29			25	0.500		
30			12	0.240		
31			1	0.020		
GD	0.368		0.634		0.336	
D_L	0.376		0.647		0.343	
SE	0.056		0.028		0.054	

TABLE III: Allele frequency, GD, D_L and SE for DYS385a/b

Allele	n	p_i
--------	-----	-------

9,12	2	0.040
10,11	2	0.040
12,13	1	0.020
12,17	3	0.060
12,18	12	0.240
13,16	4	0.080
13,17	5	0.100
13,18	9	0.180
13,19	2	0.040
14,18	10	0.200
GD	0.844	
D_L	0.861	
SE	0.015	

The overall haplotype diversity for the five Y-STR tested in 50 individuals, was 0.902 ($SE \pm 0.015$) and the discrimination capacity was 0.400. The low discrimination capacity value within this sub-tribe was expected due to the small number of subjects and Y-STR markers used in the study. Only 20 haplotypes were observed from the 50 individuals, suggesting that more markers would provide a higher DC. The discrimination capacity value obtained is considered low compared to the DC value of 16 Y-STR markers used in a study of three native populations in Sarawak; Iban, Bidayuh and Melanau [8]. However, this finding can also be attributed to the Temiars' lifestyle, culture and isolated area of living, causing more similarities between each individuals. The duplicated marker DYS385a/b showed the greatest diversity with the value of locus diversity of 0.861 ($SE \pm 0.015$). It was also found to have the highest locus diversity in three main Malaysia ethnics; Malay, Chinese, and Indian [9] as well as in the Thai population of Yala, Thailand [10]. The least polymorphic locus was DYS393 with a value of 0.343($SE \pm 0.054$) where only three alleles were found in the locus. This value was very similar to the result of a study in the Japanese and Central Poland population, showing that the diversity of this locus is almost consistent throughout different populations all over the world [11,12]. The D_L value for DYS389I was more similar to other populations compared to DYS389II. The D_L values for the DYS389II varied more in different populations although its mutation rate is the same with DYS389I. Although those two loci have the same length for their repeat motifs, the numbers of repeats in DYS389II were much higher, increasing the chance of polymorphism [13].

3.2. Phylogenetic analysis

A total of six phylogenetic trees have been developed using the preliminary results from this study where two trees, Neighbor-Joining (NJ) and Maximum-likelihood (ML) trees were constructed for DYS389I, DYS389II and DYS393. These trees were then subjected to 100 replication bootstrap analysis. Sequences retrieved from database (NCBI) were used for both trees [14,15]. Analysis for three trees for DYS389I and DYS389II (Figure 1, Figure 4 and Figure 5) showed that samples 56, 60, 70 and 80 were grouped together in clade A, except in Figure 4 where sample 80 was outside clade A. This suggests that trees using DYS389I markers to compare between samples are reliable enough to distinguish subjects from different locations. The low bootstrap value on most of the branches might be due to the lack of differences between the nucleotides of the sequences. It can also be due to the short sequence, merely around 150 base pairs, not enough to provide reliable variances to be translated into high bootstrap values.

On the other hand, in the Neighbour-Joining tree for DYS389II (Figure 2), the groupings of subjects into clades are not relative to their location. Although this locus has longer sequences, the differences in them are not enough to group them according to location. More samples are required but it can be said for now that DYS389II is not a good locus for tree building compared to DYS389I. Trees constructed using DYS393 (Figure 3 and Figure 6) show positive groupings of subjects according to their respective villages. However, the bootstrap values of DYS393 were also the lowest, indicating that, although it showed to be a good locus to compare subjects phylogenetically, the results might be less reliable.

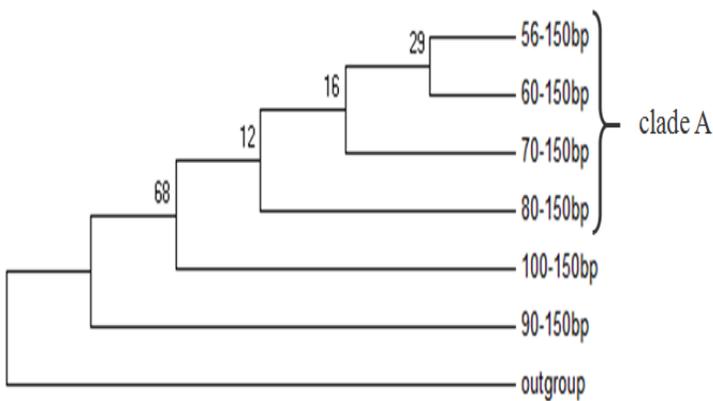


Fig 1: NJ tree for DYS389I

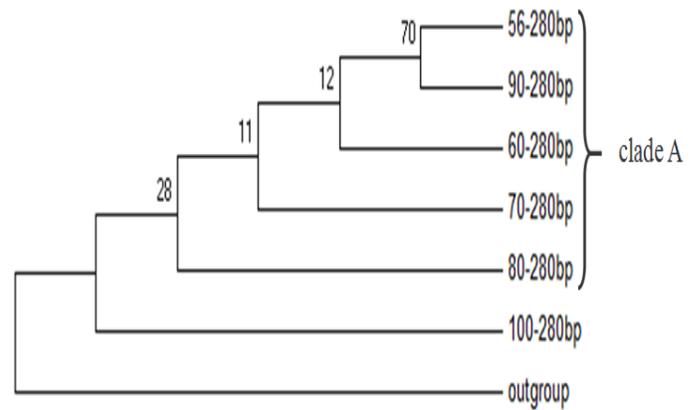


Fig 2: NJ tree for DYS389II

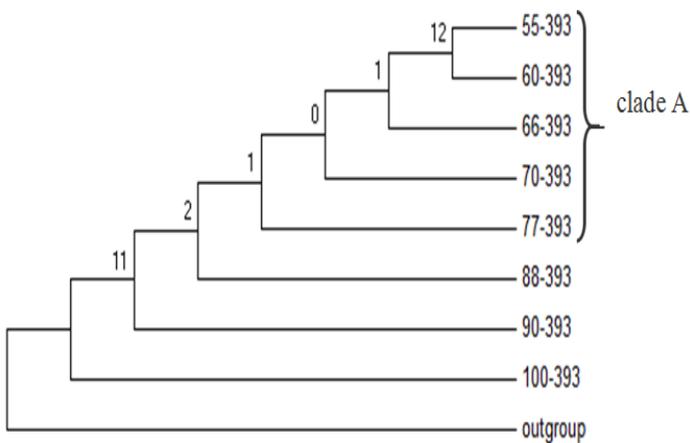


Fig 3: NJ tree for DYS393

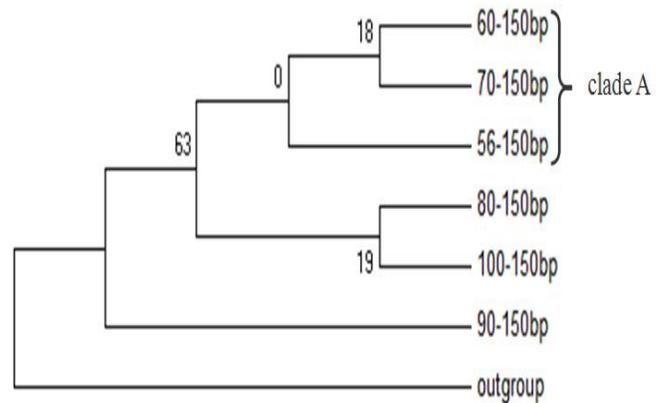


Fig 4: ML tree for DYS389I

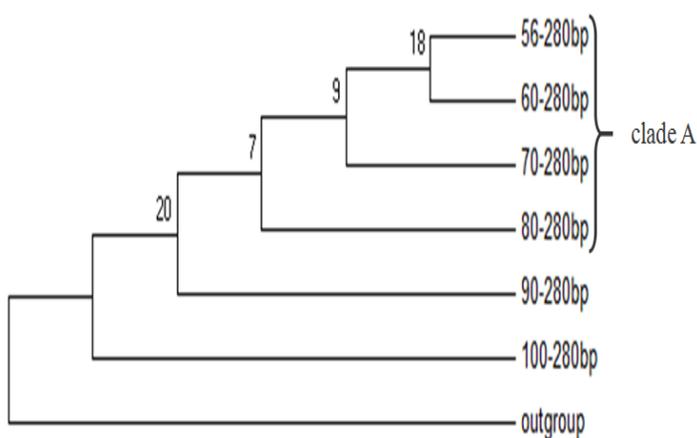


Fig 5: ML tree for DYS389II

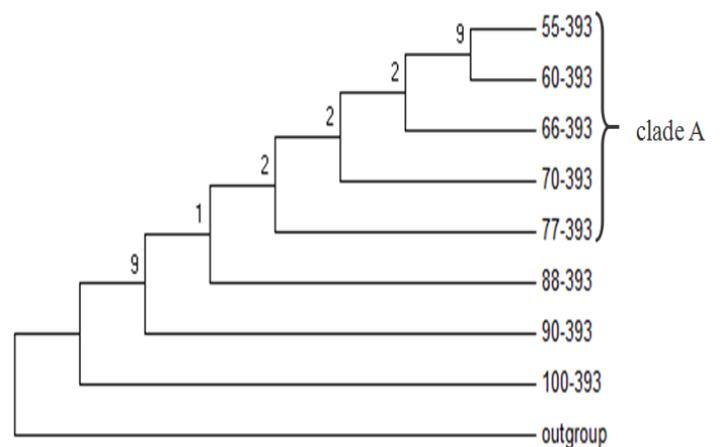


Fig 6: ML tree for DYS393

4. Conclusion

Preliminary results showed the polymorphism that occurred in the Temiar's sub-tribe of *Orang Asli* in Kelantan. Combinations of five Y-STR markers; DYS385a/b, DYS389I/II and DYS393 have low discrimination capacity suggesting that this combination would not be good enough to distinguish between the male individuals in a population. DYS385a/b showed the highest single locus diversity and can be useful to determine the male lineage. DYS393 however, was the least polymorphic locus and it might not be useful to distinguish between individuals. DYS389I showed a promising result to be used in the phylogenetic analysis while DYS393 might not be a good locus to be used as it lacks reliability. From the preliminary results, it can be suggested that more markers should be used and the study should be extended to other *Orang Asli*'s sub-tribes so that it can provide a high discrimination capacity and be useful in forensic application as well as in phylogeny field.

5. Acknowledgment

The study was funded by the FRGS research grant 600-RMI/FRGS TD 5/3 (1/2013). The authors would like to thank IRMI of Universiti Teknologi MARA (UiTM) and the Ministry of Higher Education, Malaysia for this financial contribution and support. The authors would also like to thank the staff of the Department of *Orang Asli* Development (JAKOA) and the National Medical Research Register (NMRR) for their assistance in the sample collections as well as all the volunteers who have kindly donated buccal samples for this study.

6. References

- [1] C. Nicholas, "The Orang Asli: first on land, last in the plan", *Kajian Malaysia*, vol. XXI(1,2), pp. 315-329, 2003.
- [2] Department of Orang Asli Development (2013). *Orang Asli Info*. In *Orang Asli*. Available: <http://www.jakoa.gov.my/>
- [3] K.C. Ang, M.S. Ngu, K.P. Reid, M.S. Teh, Z.S. Aida, D.X.R. Koh, A. Berg, S. Oppenheimer, H. Salleh, M.M. Clyde, B.M. Md-Zain, V.A. Canfield, and K.C. Cheng, "Skin color variation in Orang Asli tribes of Peninsular Malaysia", *PLoS ONE*, vol. 7(8), pp. 1-7, August 2012.
<http://dx.doi.org/10.1371/journal.pone.0042752>
- [4] T. Masron, F. Masami, and N. Ismail, "Orang Asli in Peninsular Malaysia : population, spatial distribution and socio-economic condition", *Journal of Ritsumeikan Social Sciences and Humanities*, vol. 6, pp. 75-115, November 2014.
- [5] M.A. Jobling, and C. Tyler-Smith, "The human Y-Chromosome: an evolutionary market comes of age", *Nat. Rev. Genet.*, vol. 4, pp. 518-612, August, 2003.
<http://dx.doi.org/10.1038/nrg1124>
- [6] S.H.N. Alshuridin, F.Z.M. Yusof, W.N.W. Hanafi, and R.R. Sohari, "Tracing Human Origin Among Orang Asli Population in Taman Negara Through Y-Chromosome Variation", presented at the IEEE Colloquium on Humanities, Science and Engineering Research, December 5-6, 2011.
- [7] M. Nei, *Molecular Evolutionary Genetics*, New York, U.S.A.: Columbia University Press, 1987, pp. 176-179.
- [8] Y.M. Chang, Y. Swaran, Y.K. Phoon, K. Sothirasan, H.T. Sim, K.B. Lim, and D. Kuehn, "Haplotype diversity of 17 Y-chromosomal STRs in three native Sarawak populations (Iban, Bidayuh and Melanau) in East Malaysia", *Forensic Science International : Genetics*, vol. 3, pp. 77-80, Jun 2009.
<http://dx.doi.org/10.1016/j.fsigen.2008.07.007>
- [9] Y.M. Chang, R. Perumal, P.Y. Keat, and D.L.C. Kuehn, "Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations (Malays, Chinese and Indians) in Malaysia", *Forensic Science International*, vol. 167, pp. 70-76, March 2007.
<http://dx.doi.org/10.1016/j.forsciint.2006.01.002>

- [10] S. Kantang, W. Waiyawuth, and T. Kusamran, "Allele frequencies and haplotype diversity of 12-loci male-specific Y-chromosome (STRs) among Thai population in Yala province of Thailand" in Proc. 2010 CIFS Academic Day, 2010, pp. 39-43.
- [11] M. Hashiyada, T. Nagashima, Y. Itakura, J. Sakai, Y. Kanawaku, J. Kanetake, and M. Funayama, "12 Y-chromosomal STR haplotypes in Japanese", *Forensic Science International*, vol. 158(2), pp. 204-212, May 2006.
<http://dx.doi.org/10.1016/j.forsciint.2005.04.035>
- [12] I. Soltyszewski, W. Pepinski, M. Spolnicka, E. Kartasinska, M. Konarzewska, and J. Janica, "Y-chromosomal haplotypes for the AmpFISTR Yfiler PCR Amplification Kit in a population sample from Central Poland", *Forensic Science International*, vol. 168, pp. 61-67, May 2007.
<http://dx.doi.org/10.1016/j.forsciint.2006.01.009>
- [13] K. N. Ballantyne, M. Goedbloed, R. Fang, O. Schaap, O. Lao, A. Wollstein, and et al., "Mutability of Y-Chromosomal Microsatellites: Rates, Characteristics, Molecular Bases, and Forensic Implications", *The American Journals of Human Genetics*, vol. 87, pp. 341-353, September 2010.
<http://dx.doi.org/10.1016/j.ajhg.2010.08.006>
- [14] A.W. Bergen, C.Y. Wang, J. Tsai, K. Jefferson, C. Dey, K.D. Smith, S.C. Park, S.J. Tsai, and D. Goldman, "An Asian-Native American paternal lineage identified by RPS4Y resequencing and by microsatellite haplotyping", *Ann Hum Genet.* vol. 63, pp. 63-80, January 1999.
<http://dx.doi.org/10.1046/j.1469-1809.1999.6310063.x>
- [15] Y.J. Kim, D.J. Shin, J.M. Kim, H. J. Jin, K. D. Kwak, M.S. Han, S.K. Choi, W. Kim, "Y-chromosome STR haplotype profiling in the Korean population", *Forensic Science International*, vol. 115(3), pp. 231-237, January 2001.
[http://dx.doi.org/10.1016/S0379-0738\(00\)00332-7](http://dx.doi.org/10.1016/S0379-0738(00)00332-7)