GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Short Communication

Genetic data for 15 STR loci in a Kadazan-Dusun population from East Malaysia

B.P. Kee[1], L.H. Lian[1], P.C. Lee[2], T.X. Lai[2] and K.H. Chua[1]

[1]Department of Molecular Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
[2]Biotechnology Program, School of Science and Technology, Universiti Sabah Malaysia, Kota Kinabalu, Sabah, Malaysia

Corresponding author: L.H. Lian
E-mail: lhlian@um.edu.my

Genet. Mol. Res. 10 (2): 739-743 (2011)
Received September 15, 2010
Accepted November 25, 2010
Published April 26, 2011
DOI 10.4238/vol10-2gmr1064

ABSTRACT. Allele frequencies of 15 short tandem repeat (STR) loci, namely D5S818, D7S820, D13S317, D16S539, TH01, TPOX, Penta D, Penta E, D3S1358, D8S1179, D18S51, D21S11, CSF1PO, vWA, and FGA, were determined for 154 individuals from the Kadazan-Dusun tribe, an indigenous population of East Malaysia. All loci were amplified by polymerase chain reaction, using the Powerplex 16 system. Alleles were typed using a gene analyzer and the Genemapper ID software. Various statistical parameters were calculated and the combined power of discrimination for the 15 loci in the population was calculated as 0.999999999999999. These loci are thus, informative and can be used effectively in forensic and genetic studies of this indigenous population.

Key words: Short tandem repeats; Population data; East Malaysia; Powerplex 16 system

INTRODUCTION

Short tandem repeats (STR) have been widely used in different aspects of modern scientific research, such as human evolution, forensics, anthropology, and disease development (Calafell et al., 1998; Blessmann et al., 2008; Divine et al., 2010; Lewis Jr., 2010). Therefore, it is important to profile the STRs for local populations, which are helpful for further genetic characterization. Sabah is the second largest State in Malaysia, and is located at the northern pole of Borneo Island (Figure 1). There are approximately 28 indigenous groups that account for 60% of the local population. Among them, the Kadazan-Dusun tribe is the largest group, which makes up 18% of the total population. Their origins are still debated until today. The settlement of Island Southeast Asia (ISEA) by modern humans was described by researchers with the "Out of Taiwan" model (Bellwood, 2007). The dispersal was believed to occur in two tiers - the first tier was marked by the arrival of "Australo-Melanesian" about 50,000 years ago. This was followed by the colonization by "Autronesians", who came by sea from Southern China through Taiwan, about 5000 years ago for agricultural purposes (Bellwood, 2007). Hence, in the present study, we examined 15 STR loci in the Kadazan-Dusun population. To the best of our knowledge, this is the first report of population data with regard to the STRs of this indigenous tribe from East Malaysia.

Figure 1. Location of the Sabah State, East Malaysia (Source: edited from Google Maps).

MATERIAL AND METHODS

One hundred and fifty-four unrelated healthy individuals from the Kadazan-Dusun tribe residing in the State of Sabah in East Malaysia were randomly sampled. Ethical approval for the project was obtained from the University of Malaya Medical Centre's Medical Ethics Committee (No. 770.21), which operates according to the ICH-GCP guidelines and the Declaration of Helsinki. Genomic DNA was extracted from peripheral blood samples via a modified conventional phenol-chloroform extraction method. The quality and quantity of the extracted DNA was assessed spectrophotometrically. The amplification process was carried out using the Promega Powerplex 16 system according to the cycling parameters recommended by the manufacturer. Amplified products were resolved by native 2% (w/v) agarose gel electrophoresis to ensure suc-

cessful amplification of the desired loci. Typing was performed by capillary electrophoresis using an ABI3100 genetic analyzer. Allelic call and genotyping were carried out via the Genemapper ID software version 3.2 (ABI), by referring to the allelic ladder included in the kit.

Quality assurance of the test was conducted via laboratory internal controls and kit DNA controls. Allelic frequencies, the observed heterozygosity, power of discrimination, power of exclusion, and polymorphism information content (PIC) were calculated using PowerStats V1.2 (http://www.promega.com) (Tereba, 1999). The estimation of Hardy-Weinberg equilibrium (HWE) and expected heterozygosity was done with the exact test using the Arlequin V3.1 software (Schneider et al., 2000). The level of significance of the exact test was set at 5%.

RESULTS AND DISCUSSION

Allelic frequencies and statistical parameters of the 15 STR loci examined in this study are shown in Tables 1 and 2. All loci showed a high degree of heterozygosity (>0.7), with exception of loci D7S820 and TPOX (0.623 and 0.526, respectively). The loci were observed to have high discriminating power, as the power of discrimination of each loci varied from 0.721 (TPOX) to 0.968 (Penta E), whereas the PIC ranged from 0.464 (TPOX) to 0868 (Penta E). The combined power of discrimination for the 15 STR loci studied is 0.999999999999999. All loci were found to show no deviation from HWE ($P > 0.05$). In terms of gene diversity, the Penta E locus was identified as the most polymorphic marker in the Kadazan-Dusun population, with 89.6% heterozygosity. In our study, the 15 STR loci exhibited high levels of polymorphism in the Kadazan-Dusun population. Therefore, these loci can be used effectively as genetic markers for forensic identification purposes and in various other genetics-based studies.

Table 1. Allelic frequencies and statistical parameters of D5S818, D7S820, D13S317, D16S539, TH01, TPOX, Penta D, and Penta E in the Kadazan-Dusun population, Sabah in East Malaysia (N = 154).

Allele	D5S818	D7S820	D13S317	D16S539	TH01	TPOX	Penta D	Penta E
23								0.023
22								0.010
21								0.026
20								0.007
19								0.026
18								0.065
17								0.078
16								0.049
15								0.231
14	0.003	0.010		0.010				0.068
13	0.068	0.007		0.052			0.208	0.023
12	0.172	0.136	0.247	0.114		0.003	0.127	0.084
11	0.393	0.539	0.412	0.312		0.481	0.091	0.185
10	0.338	0.123	0.081	0.260	0.088		0.133	0.026
9.3					0.046			
9	0.026	0.020	0.036	0.253	0.406	0.065	0.383	0.023
8		0.166	0.224		0.166	0.451	0.055	
7					0.218		0.003	
6					0.078			
5								0.075
4								0.003
Ho	0.714	0.623	0.721	0.760	0.792	0.526	0.766	0.896
He	0.699	0.650	0.713	0.758	0.747	0.563	0.768	0.882
PD	0.843	0.830	0.864	0.890	0.896	0.721	0.908	0.968
PIC	0.642	0.609	0.663	0.714	0.710	0.464	0.734	0.868
PE	0.451	0.320	0.461	0.527	0.585	0.211	0.538	0.787
P	0.437	0.066	0.546	0.125	0.389	0.442	0.094	0.052

Ho = observed heterozygosity; He = expected heterozygosity; PD = power of discrimination; PIC = polymorphism information content; PE = power of exclusion; P = probability value of the exact test for HWE.

Table 2. Allelic frequencies and statistical parameters of D3S1358, D8S1179, D18S51, D21S11, CSF1PO, vWA, and FGA in the Kadazan-Dusun population, Sabah in East Malaysia (N = 154).

Allele	D3S1358	D8S1179	D18S51	D21S11	CSF1PO	vWA	FGA
34.2				0.003			
33.2				0.110			
32.2				0.140			
32				0.023			
31.2				0.084			
31				0.136			
30.2				0.013			
30				0.260			
29				0.182			
28				0.049			
27							0.007
26							0.033
25.2							0.003
25							0.088
24							0.107
23							0.140
22.2							0.003
22			0.003				0.244
21							0.169
20						0.029	0.107
19.2							0.003
19	0.013		0.055			0.084	0.052
18.2							0.046
18	0.091		0.010			0.296	
17	0.192	0.013	0.068			0.292	
16	0.260	0.104	0.146			0.104	
15	0.432	0.149	0.334			0.081	
14	0.013	0.224	0.292			0.107	
13		0.289	0.02		0.020		
12		0.023	0.071		0.412		
11		0.068			0.308	0.003	
10		0.127			0.221	0.003	
9		0.003			0.039		
Ho	0.701	0.831	0.721	0.805	0.714	0.779	0.870
He	0.703	0.814	0.771	0.842	0.686	0.793	0.859
PD	0.857	0.932	0.913	0.951	0.829	0.928	0.958
PIC	0.652	0.787	0.735	0.820	0.625	0.762	0.841
PE	0.430	0.658	0.461	0.609	0.451	0.561	0.735
P	0.584	0.217	0.780	0.163	0.709	0.250	0.339

Ho = observed heterozygosity; He = expected heterozygosity; PD = power of discrimination; PIC = polymorphism information content; PE = power of exclusion; P = probability value of the exact test for HWE.

ACKNOWLEDGMENTS

Research supported by the Ministry of Science, Technology and Innovation, Malaysia (eScience Fund #12-02-03-2045), the Ministry of Higher Education, Malaysia (FRGS #FP024/2009), and the University of Malaya (PPP #PS234/2010A).

REFERENCES

Bellwood P (2007). Prehistory of the Indo-Malaysian Archipelago. ANU E Press, Australia.

Blessmann M, Kaifi JT, Schurr PG, Cihan A, et al. (2008). Short tandem repeat polymorphism in exon 4 of esophageal cancer related gene 2 predicts relapse of oral squamous cell carcinoma. *Oral Oncol.* 44: 143-147.

Calafell F, Shuster A, Speed WC, Kidd JR, et al. (1998). Short tandem repeat polymorphism evolution in humans. *Eur. J. Hum. Genet.* 6: 38-49.

Divine AM, Edlund H and Allen M (2010). Forensic analysis of autosomal STR markers using pyrosequencing. *Forensic Sci. Int-Gen.* 4: 122-129.

Lewis CM Jr (2010). Hierarchical modeling of genome-wide short tandem repeat (STR) markers infers native American prehistory. *Am. J. Phys. Anthropol.* 141: 281-289.

Schneider S, Roessli D and Excoffier L (2000). Arlequin: A Software for Population Genetics Data Analysis. Ver. 2.000. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva.

Tereba A (1999). Tools for Analysis of Population Statistics. *Profiles in DNA* 2: 14-16.

