Annals of Human Biology, 2012; Early Online: 1–6

ISSN 0301-4460 print/ISSN 1464-5033 online
DOI: 10.3109/03014460.2012.719548

informa
healthcare

RESEARCH PAPER

Population data of six *Alu* insertions in indigenous groups from Sabah, Malaysia

B. P. Kee[1], K. H. Chua[1], P. C. Lee[2] & L. H. Lian[1]

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

***Background and aim*: The present study is the first to report the genetic relatedness of indigenous populations of Sabah, Malaysia, using a set of *Indel* markers (HS4.32, TPA25, APO, PV92, B65 and HS3.23). The primary aim was to assess the genetic relationships among these populations and with populations from other parts of the world by examining the distribution of these markers.**

***Subjects and methods*: A total of 504 volunteers from the three largest indigenous groups, i.e. Kadazan-Dusun, Bajau and Rungus, were recruited for the study. Six *Alu* insertions were typed by PCR with specific primer sets.**

***Results*: All insertions were found to present at different frequencies, ranging from 0.170–0.970. The heterozygosity of most of the markers was high (>0.4), with the exception of HS3.23 and APO. A genetic differentiation study revealed that these populations are closely related to each other ($G_{ST} = 0.006$). A principle component plot showed that these populations have higher affinity to Mainland South East Asia/East Asia populations, rather than Island Southeast Asia (ISEA) populations.**

***Conclusion*: In summary, these indigenous groups were closely associated in terms of their genetic composition. This finding also supports the colonization model of ISEA, which suggests that the inhabitants of this region were mostly descendants from Southern China.**

Keywords: *Alu* insertion, Kadazan-Dusun, Bajau, Rungus

INTRODUCTION

The *Alu* element, one of the most prominent examples of repetitive DNA, is a member of the short interspersed element (SINE), which is estimated to exist in more than 10% of the entire human genome (Houck et al. 1979; Smit 1996). Each *Alu* element is ~300 nucleotides in length and is believed to derive from 7SL RNA (Weiner et al. 1986). The *Alu* element is mobilized within the genome by a gene jumping mechanism known as retroposition—a RNA-mediated transcription process. This mechanism has contributed to the random yet wide distribution of *Alu* elements, with varied density, throughout the genome. The insertion of *Alu* elements began as early as 65 million years ago (Shen et al. 1991). The current amplification rate of *Alu* insertions, at a rate of one new insertion in 200 newborns, has been found to be much lower than in the past, being merely 1% of previous peak rates (Deininger and Batzer 1999). There are more than one million accumulated copies of *Alu* elements being reported in the human genome (Lander et al. 2001). Most *Alu* insertions are constant in all human beings, regardless of the origin of population. However, a small number of them (~5%) are polymorphic. These insertions arose recently during global colonization by modern humans and have great potential to reveal information about modern human expansion and migration events. Most *Alu* insertions are shared by both human and primate genomes, but ~7000 of these insertions are unique to humans (The Chimpanzee Sequencing and Analysis Consortium 2005).

Over decades of extensive studies, *Alu* insertions have been found useful in different aspects of scientific research, especially in cancer and evolutionary studies (Flint et al. 1996; Deininger and Batzer 1999; Batzer and Deininger 2002). Scientists have postulated that *Alu* elements are the major contributor to the evolutionary process throughout primate history, including that of humans (Batzer and Deininger 2002). Massive insertions of *Alu* elements have caused genomic instability that has facilitated the process of speciation (Challem and Taylor 1998). Recently, the study of *Alu* elements has focused on the search for population-specific markers. Researchers have demonstrated the possibility of inferring the origin of a population using a combination of polymorphic *Alu* markers (Ray et al. 2005).

Correspondence: Dr L. H. Lian, Department of Molecular Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia. Tel: (+60)3 7967 5740. Fax: (+60)3 7967 4957. E-mail: lhlian@um.edu.my

(*Received 16 January 2012; accepted 31 July 2012*)

Figure 1. Geographical location of Sabah, situated at the North end of Borneo Island (adapted from Kee et al. 2011).

POPULATION OF INTEREST

In this study, we aimed to assess the genomic diversity of six polymorphic *Alu* markers in the indigenous populations residing in the state of Sabah in Malaysia. Sabah is located at the northern end of Borneo Island (Figure 1). It has a population size of ~3.3 million, comprising of 32 ethnic groups, of which 28 are indigenous (Wise 2008). Indigenous populations make up more than 60% of the local population. Kadazan-Dusun, Bajau and Rungus are three of the largest indigenous groups in Sabah; others include the Murut, Orang Sungai, Lotud, Dumpas, etc. The peopling pattern of the region of Island Southeast Asia (ISEA) has been studied and elucidated by researchers. The 'Out of Taiwan' model, which was established based on linguistic approaches, is one of the prevailing theories and suggests that the colonization of ISEA occurred in two tiers (Bellwood 2007). The first dispersal was marked by the arrival of 'Australo-melanesians' ~50 000 years ago. These people were then replaced or assimilated by 'Austronesians', who came by sea from Southern China through Taiwan for climate or agricultural reasons (Bellwood 2007). Thus far, there has been no published study reporting on the genetic relevance of these populations. Therefore, this present research could provide the very first insight into the degree of relatedness among these indigenous groups.

MATERIALS AND METHODS

Sample collection and preparation

A total of 504 indigenous individuals (150 Kadazan-Dusun; 228 Bajau; 126 Rungus) from Sabah were recruited for this study. In order to ensure all samples originated from pure genetic lineage, both parents and grandparents of the recruited candidate must also originate from the same indigenous group. Prior to venous blood collection, informed consent (approved by ethical committee of University Malaya Medical Centre: 770.21) was obtained from volunteers. Genomic DNA was extracted from blood samples via a conventional extraction method, as stated previously (Chua et al. 2009; Kee et al. 2011). Both the quality and quantity of extracted DNA were determined by spectrophotometry via a Nanophotometer (Implen, Germany).

Amplification of Alu insertions

Six *Alu* insertion polymorphisms, HS4.32, TPA25, APO, PV92, B65 and HS3.23, were selected and typed in all samples with primer sets reported in previous studies (Batzer et al. 1994; Arcot et al. 1996; Tishkoff et al. 1996). Amplification of *Alu* inserted regions was carried out in a thermal cycler (Veriti, Applied Biosystems, California, USA). The 10 μl amplification mixture contained 100 ng of template DNA, 1 × DreamTaq buffer, 0.2 mM of dNTP mix,

Table I. Chromosomal locations and primers sequences used for amplification of the markers.

Alu markers	Chromosome	Forward primer (5' – 3')	Reverse primer (5' – 3')	TA (°C)	Reference
HS4.32	12	GTTTATTGGGCTAACCTGGG	TGACCAGCTAACTTCTACTTTAACC	63	Arcot et al. (1996)
TPA25	8	GTGAAAAGCAAGGTCTACCAG	GACACCGAGTTCATCTTGAC	63	Tishkoff et al. (1996)
APO	11	TGTGAGCCTAGGAGTTTGAG	CTGGCTGATTTTAGGAGGGA	65	Batzer et al. (1994)
PV92	16	AACTGGGAAAATTTGAAGAAAAGT	TGAGTTCTCAACTCCTGTGTGTTAG	60	Batzer et al. (1994)
B65	11	ATATCCTAAAAGGGACACCA	AAAATTTATGTCATGGGTAT	54	Batzer et al. (1994)
HS3.23	7	GGTGAAGTTTCCAACGCTGT	CCCTCCTCTCCCTTTAGCAG	65	Arcot et al. (1996)

TA, Annealing temperature.

Table II. Insertion frequencies of *Alu* markers in the PC analysis (the world population data was obtained from ALFRED).

Alu insertion marker		Insertion frequency					
Region	Population	HS4.32	TPA25	APO	PV92	B65	HS3.23
Africa	Alur	0.500	0.250	0.708	0.125	0.750	NA
	Nguni	0.114	0.200	0.600	0.240	0.600	0.910
	Kungsan	0.240	0.140	0.880	0.200	0.500	0.950
	San	0.321	0.200	0.821	0.300	0.654	NA
East Asia	Han	0.438	0.551	0.929	0.832	0.587	NA
Mainland SEA	Japanese	0.438	0.576	0.844	0.857	0.412	NA
	Cambodian	0.546	0.400	0.792	1.000	0.417	NA
ISEA	Vietnamese	0.556	0.278	0.944	0.875	0.444	NA
	Moluccas	0.280	0.550	0.750	0.690	0.260	NA
	Nuru Tenggaras	0.330	0.380	0.780	0.510	0.400	0.590
	Kadazan-Dusun	0.170	0.537	0.970	0.593	0.530	0.710
	Bajau	0.307	0.474	0.936	0.675	0.476	0.937
	Rungus	0.329	0.603	0.952	0.694	0.476	0.948
European	Finnish	0.579	0.444	0.974	0.156	0.421	NA
	French	0.700	0.725	0.950	0.275	0.625	0.890
Oceania	PNG	0.340	0.160	0.680	0.240	0.180	0.230
North America	Alaskan Natives	0.210	0.260	0.917	0.619	0.490	0.790
	Maya	0.270	0.650	0.940	0.790	0.270	0.610
	Hispanic American	0.460	0.560	0.970	0.510	0.020	0.740

NA, Not available.

0.4 U of DreamTaq DNA Polymerase, 2 mM of $MgCl_2$ and 0.4 mM of each forward and reverse primers. Chromosomal locations and primer sequences used for the study are listed in Table I. Each reaction was subjected to a cycle of denaturation at 94°C for 5 min, 35 cycles of 94°C for 30 s, appropriate optimized annealing temperature (54–65°C) for 30 s, 72°C for 30 s and a final extension at 72°C for 5 min. Amplified products were subsequently resolved on 2% (w/v) native agarose gels, stained with ethidium bromide and visualized under a UV transilluminator. Allelic scoring was carried out by the direct counting method.

Statistical analysis

AMOVA and departure of the examined markers from Hardy-Weinberg Equilibrium (HWE) were determined via Arlequin package V3.11 based on exact test, with significance level set at 5% (Excoffier et al. 2005). Population and forensic parameters, such as heterozygosity, frequencies, polymorphism information content (PIC), power of discrimination (PD), matching probability and power of exclusion (PE) were computed using Powerstat V1.2 (Promega, USA, Wisconsin, USA). Population differentiation was estimated by a set of measurements using FSTAT (Goudet 1995). Principal component (PC) analysis was conducted to investigate the genetic association among the studied populations by XLSTAT (Addinsoft, New York, USA). PC analysis also included available *Alu* insertion data of other global populations from the Allele Frequency Database (ALFRED) to depict relationships between the indigenous groups and global populations, which may reveal information on historical movement within this region (Rajeevan et al., 2005). Insertion frequencies, as in Table II, of five out of six markers (except HS3.23) in this study were utilized to construct the PC plot.

RESULTS

Alu insertions examined in the present study were polymorphic in the Kadazan-Dusun, Bajau and Rungus populations. The allelic and genotypic frequencies of

Table III. Insertion frequency, population differentiation analysis and AMOVA of six *Alu* markers in Kadazan-Dusun, Bajau and Rungus populations.

Alu insertion	HS4.32	TPA25	HS3.23	PV92	B65	APO	Average
Population							
Kadazan-Dusun ($n = 150$)	0.170	0.537	0.937	0.713	0.530	0.970	—
Bajau ($n = 228$)	0.307	0.474	0.936	0.721	0.476	0.936	—
Rungus ($n = 126$)	0.329	0.603	0.948	0.706	0.480	0.952	—
Differentiation analysis							
H_T	0.394	0.498	0.112	0.409	0.501	0.090	0.334
H_S	0.384	0.493	0.112	0.410	0.500	0.090	0.332
D_{ST}	0.009	0.005	0.000	0.001	0.000	0.000	0.002
G_{ST}	0.023	0.009	0.002	0.002	0.000	0.002	0.006
G_{IS}	0.093	0.033	0.052	0.114	0.057	0.032	0.068
AMOVA differentiation							
Among populations	3.167	1.404	0	0	0.053	0.363	0.938
Within populations	96.833	98.596	100	100	99.947	99.637	99.062
p	0	0.004	NS	NS	NS	NS	—

H_T, Total gene diversity; H_S, Intra-population diversity; D_{ST}, Inter-population diversity; G_{ST}, Coefficient of gene differentiation; G_{IS}, Inbreeding coefficient; NS, Not significant, $p > 0.05$.

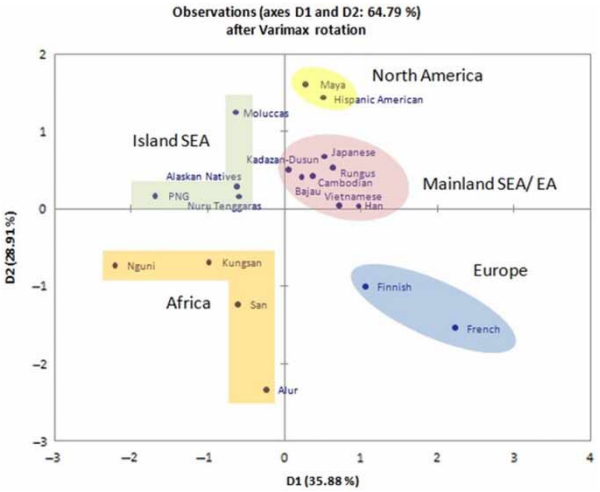

Figure 2. Principal component analysis, based on insertion frequency, in the indigenous groups, in comparison to other populations (SEA, South East Asia; EA, East Asia).

the insertion markers are summarized in Table III. The frequencies of inserted allele (+) ranged from 0.170–0.970. None of these markers were completely fixed to any of the populations, despite the fact that two markers (HS3.23 and APO) were present with insertion frequencies of higher than 0.9 in all populations. On the contrary, HS4.32 insertion was found to be present at the lowest frequency, with an average of 0.269. All populations showed high levels of diversity (>0.4) for most markers, with the highest attainable heterozygosity for a biallelic marker being 0.5. However, the heterozygosity for HS3.23 and APO insertion was low in all populations, with average of 0.112 and 0.089, respectively. Out of 18 HWE tests conducted, only one deviation from the equation was observed (PV92 insertion in Bajau population). The deviation was regarded as a random statistical fluctuation as it does not appear to be locus- or population-specific in our cohort of study.

Total gene diversity (H_T) is accounted for by two fractions of genomic variability that arise within and among populations, which can be measured by intra-population (H_S) and inter-population (D_{ST}) gene diversity. The coefficients of gene differentiation (G_{ST}) of all three populations in our study were low (Table III), where the highest value, $G_{ST} = 0.023$, was observed in HS4.32 insertion. The mean value of G_{ST} in all the six markers was 0.006. The inbreeding coefficient (G_{IS}) ranged from 0.032–0.114, with an average of 0.068. The AMOVA showed that genetic variations among populations were significant in two markers, i.e. HS4.32 (3.17%, $p < 0.01$) and TPA25 (1.40%, $p < 0.01$). On average, over 99% of the genetic variations were contributed for within populations.

In order to further elaborate on the genetic association of the examined groups, a PC plot was constructed (Figure 2). The two main principal components contributed to 64.79% of the variance observed. The first component accounted for up to 35.88% of the variation, whereas the second component made up 28.91%. All groups were huddled into distinctive clusters, i.e. Africa, Europe, North America, Mainland Southeast Asia/East Asia (SEA/EA) and Island SEA. The first dimension separated African populations to the lower left part of the plot and European populations to the lower right. On the other hand, the second dimension differentiated all Asian populations, together with North American populations from both African and European populations. Kadazan-Dusun, Bajau and Rungus populations were grouped closely within the Mainland SEA/EA cluster.

As shown in Table IV, the matching probability (MP) of all markers in the present study differed from 0.359–0.898, with an average value of 0.5429. The power of discrimination (PD) of the *Alu* insertions varied greatly from 0.102–0.646, with averages of 0.434, 0.477 and 0.460 in Kadazan-Dusun, Bajau and Rungus populations, respectively. When accessing the discriminating power of the six loci altogether, the combined PDs for all markers were 0.978 for the Kadazan-Dusun, 0.986 for the Bajau and 0.983 for the Rungus. The polymorphism information content (PIC) values varied from 0.06–0.370. The lowest PE value (0.02) was observed in the APO insertion in the Kadazan-Dusun population, while the highest PE value of 0.209 was noted in TPA25 insertion of the Rungus. The combined powers of exclusion (PE) for the Kadazan-Dusun, Bajau and Rungus populations were 0.398, 0.436 and 0.475, respectively. All populations had an average typical paternity index (TPI) of higher than 0.7.

DISCUSSION

This work establishes the genetic relationship between three major indigenous groups, i.e. Kadazan-Dusun, Bajau and Rungus in the state of Sabah, Malaysia and with other populations in regions of SEA and East Asia by utilizing genetic data obtained from six polymorphic *Alu* insertions. These markers were also evaluated for their usefulness in

Table IV. Power of discrimination (PD), Power of exclusion (PE), Polymorphism information content (PIC), Typical paternity index (TPI) and deviation from Hardy-Weinberg Equilibrium (HWE) of *Alu* markers in Sabahan indigenous populations.

Alu insertion	HS4.32	TPA25	HS3.23	PV92	B65	APO
Kadazan-Dusun ($n = 150$)						
PD	0.438	0.640	0.213	0.571	0.641	0.102
PE	0.044	0.155	0.011	0.106	0.155	0.002
PIC	0.240	0.370	0.110	0.330	0.370	0.060
TPI	0.660	0.930	0.560	0.820	0.930	0.520
HWE	0.143	0.410	0.457	0.549	0.412	0.115
Bajau ($n = 228$)						
PD	0.587	0.646	0.216	0.569	0.630	0.216
PE	0.111	0.145	0.011	0.080	0.176	0.011
PIC	0.330	0.370	0.110	0.320	0.370	0.110
TPI	0.830	0.900	0.570	0.750	0.970	0.570
HWE	0.281	0.145	1.000	0.020	0.790	1.000
Rungus ($n = 126$)						
PD	0.601	0.591	0.174	0.580	0.639	0.172
PE	0.117	0.209	0.007	0.094	0.161	0.008
PIC	0.340	0.360	0.090	0.330	0.370	0.090
TPI	0.840	1.050	0.550	0.790	0.940	0.550
HWE	0.424	0.352	0.278	0.196	0.480	1.000

forensic and human identification investigations in the local context. No previous study has reported on the characterization of genetic association of these indigenous groups based on the distribution of polymorphic *Alu* insertions. High expected heterozygosity values were observed in most of the markers examined, except in the HS3.23 and APO insertions. The heterozygosity was comparable for all the three populations, suggesting that these populations may share a common source of genetic ancestry.

Gene diversity analysis reveals that all markers were present with a considerably high degree of diversity ($H_T > 0.39$), except HS3.23 and APO insertions ($H_T = \sim 0.1$). The majority of the genetic variability is contributed by differences between individuals of a population ($H_{S=}0.332$), whereas only a small part of the variability comes from genomic diversity between populations ($G_{ST} = 0.006$; $D_{ST} = 0.003$). The low G_{ST} implies that there is minimal degree of genomic differentiation between the Kadazan-Dusun, Bajau and Rungus populations. Concordantly, the AMOVA study also shows that these indigenous populations are genetically homogenous as only 0.938% of the differences are attributed to variations among the populations whilst only two out of six markers (HS4.32 and TPA25) exhibit significant heterogeneity in these examined populations.

When compared to frequency data available from the genetic database, we observed high insertion frequencies in both TPA25 and HS3.23 insertions in populations from different regions, with average frequencies of 0.86 and 0.77, respectively. This implies that these insertions may have occurred before the spread of modern humans to other parts of the world and were distributed to all descending populations at a fair rate. In contrast, the frequency of *Alu* insertion in TPA25 locus is lower in African populations (average frequency = 0.198) than that in other regions (average frequency = 0.518), except in Oceania and Alaskan Natives. This observation may indicate that the ancient group(s) that emigrated from Africa 80 000 years ago may have carried higher frequencies of the insertion, which were then passed on to their descendants during world colonization. In addition, the TPA *Alu* insertion, located in the Tissue Plasminogen Activator (PLAT) gene, may have favoured the survival of these ancient settlers against harsh environmental challenges during the great migration. Interestingly, PV92 insertion was reported with low frequencies in both Africans and Europeans (average frequency = 0.216) than in Asians and Americans (average frequency = 0.720). This reflects a possible isolation of these settlers in restricted populations due to catastrophic events (such as the Toba eruption ~70 000 years ago). The recovered populations then continued to colonize the world in different directions resulting in the genetic discrepancies in modern human populations.

The genetic relevance of the indigenous groups in comparison to 16 global populations from different regions is illustrated by a PC plot. The two main PC have contributed to substantial variations (64.79%) among the populations. All populations were found clumped into a few clusters according to global continents. In our study, Asian populations were clearly distributed into two distinct clusters, i.e. Mainland SEA/EA and ISEA. The ISEA cluster is made up of Moluccas and Nuru Tenggaras populations. Interestingly, the Alaskan Native was included in this cluster as well, which signifies a certain degree of genetic similarity of these natives to inhabitants of ISEA. The three indigenous populations were clustered close to each other in the PC plot, showing that these populations share a high degree of genetic similarity among them. This is in agreement with our interpretation obtained from population differentiation analysis. However, both the Kadazan-Dusun and Bajau populations were situated closer within the cluster, next to the Cambodian population, while the Rungus population was located closer to the Japanese population. The Papua New Guinea (PNG) was positioned to the left of the upper quarter in the PC plot, distant from both Asian and European populations. This observation is in agreement with a recent report, where PNG and Aboriginal Australians were believed to have separated from ancestral Eurasians before European and Asian populations split from each other (Rasmussen et al. 2011).

Our Kadazan-Dusun, Bajau and Rungus were grouped in the Mainland SEA/EA cluster, together with the Han Chinese, Japanese, Cambodian and Vietnamese. Furthermore, the Sabahan indigenous populations showed higher affinity to the Mainland SEA/EA cluster than the ISEA counterparts. In another investigation of genetic structure via mitochondrial DNA, Y-Chromosome and autosomal markers, Ibans (an indigenous group residing in the neighbouring state of Sabah, Sarawak) were also shown to have high genetic similarity to Mainland SEA populations (Simonson et al. 2011). It is therefore indicative that the indigenous populations in Borneo Island may be descendants of migrating settlers from Mainland SEA or EA. Similar remarks have been observed in a genetic study based on SNP array, where SEA populations were found to have significant influence on the genetic composition of the EA populations (The HUGO Pan-Asian SNP Consortium 2009). It was suggested that all populations in SEA and EA were direct descendants of migrants from the primary wave into the continent, expanding from Southern Asia (The HUGO Pan-Asian SNP Consortium 2009; Xing et al. 2010). On the other hand, a genetic study of Andaman Islanders has revealed a recent migration from SEA into the region that happened ~18 000 years ago (Thangaraj et al. 2005). This has again highlighted the influence of SEA populations on the peopling patterns of the neighbouring regions. Despite extensive and in-depth studies conducted by scientists, it is still very challenging to understand and obtain a clear picture of the paths travelled by humans in the past, largely due to the fact that interpretation is only based on limited sex-biased markers and the analysis method (Stoneking and Delfin 2010).

In our study, the six *Alu* markers showed a moderate degree of discriminating power in our populations. The average combined PD of these markers was 0.982 271. Although *Alu* markers do not generate discriminating power

as high as multi-allelic markers (such as STR), they are known for their ability to infer population origin owing to their identical-by-descent state (Shedlock et al. 2004). Studies have shown that *Alu* insertion polymorphisms are able to infer a population of origin accurately and the possibility can be extended to identification and differentiation of samples with mixed ancestry (Watkins et al. 2003; Ray et al. 2005; Mamedov et al. 2010). With the discovery of more population-indicative *Alu* insertions, they can be used efficiently in forensic and crime investigations by providing additional information pertaining to suspects' population of origin.

CONCLUSION

In summary, this study revealed a close genetic lineage between Kadazan-Dusun, Bajau and Rungus populations. These populations also showed higher affinity to Mainland SEA/EA populations, rather than other ISEA populations. This observation supports the view that the current settlement of ISEA populations originated from Southern China. However, we were not able to deduce a possible coastal or maritime route taken by these settlers during the colonization of SEA.

ACKNOWLEDGEMENTS

This research study was supported by grants from Ministry of Higher Education Malaysia High Impact Research Grant (E-000044–20001) and Ministry of Science, Technology and Innovation (MOSTI) Malaysia (02-01-03-SF0332).

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

REFERENCES

Arcot SS, Adamson AW, Lamerdin JE, Kanagy B, Deininger PL, Carrano AV, Batzer MA. 1996. Alu fossil relics–distribution and insertion polymorphism. Genome Res 6:1084–1092.

Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. Nat Rev Genet 3:370–379.

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, Deininger PL. 1994. African origin of human-specific polymorphic Alu insertions. Proc Natl Acad Sci USA 91:12288–12292.

Bellwood P. 2007. Prehistory of the Indo-Malaysian archipelago. Australia: The Australian National University E Press.

Challem JJ, Taylor EW. 1998. Retroviruses, ascorbate, and mutations, in the evolution of Homo sapiens. Free Radic Biol Med 25:130–132.

Chua KH, Kee BP, Tan SY, Lian LH. 2009. An association between Interleukin-6 (IL-6) Promoter polymorphisms (−174 G/C) and Systemic Lupus Erythematosus (SLE). Braz J Med Biol Res 42: 551–555.

Deininger PL, Batzer MA. 1999. Alu repeats and human disease. Mol Genet Metab 67:183–193.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50.

Flint J, Rochette J, Craddock CF, Dode C, Vignes B, Horsley SW, Kearney L, Buckle VJ, Ayyub H, Higgs DR. 1996. Chromosomal stabilisation by a subtelomeric rearrangement involving two closely related Alu elements. Hum Mol Genet 5:1163–1169.

Goudet J. 1995. FSTAT (Version 1.2): a computer program to calculate F-statistics. J Hered 86:485–486.

Houck CM, Rinehart FP, Schmid CW. 1979. A ubiquitous family of repeated DNA sequences in the human genome. J Mol Biol 132: 289–306.

Kee BP, Lian LH, Lee PC, Lai TX, Chua KH. 2011. Genetic data for 15 STR loci in a Kadazan-Dusun population from East Malaysia. Genet Mol Res 10:739–743.

Lander ES, Linton LM, Birren B, et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.

Mamedov IZ, Shagina IA, Kurnikova MA, Novozhilov SN, Shagin DA, Lebedev YB. 2010. A new set of markers for human identification based on 32 polymorphic Alu insertions. Eur J Hum Genet 18: 808–814.

Rajeevan H, Cheung KH, Gadagkar R, Stein S, Soundararajan U, Kidd JR, Pakstis AJ, Miller PL, Kidd KK. 2005. ALFRED: an allele frequency database for microevolutionary studies. Evol Bioinform Online 1:1–10.

Rasmussen M, Guo X, Wang Y, et al. 2011. An Aboriginal Australian genome reveals separate human dispersals into Asia. Science 334: 94–98.

Ray DA, Walker JA, Hall A, Llewellyn B, Ballantyne J, Christian AT, Turteltaub K, Batzer MA. 2005. Inference of human geographic origins using Alu insertion polymorphisms. Forensic Sci Int 153: 117–124.

Shedlock AM, Takahashi K, Okada N. 2004. SINEs of speciation: tracking lineages with retroposons. Trends Ecol Evol 19:545–553.

Shen MR, Batzer MA, Deininger PL. 1991. Evolution of the master Alu gene(s). J Mol Evol 33:311–320.

Simonson TS, Xing J, Barrett R, Jerah E, Loa P, Zhang Y, Watkins WS, Witherspoon DJ, Huff CD, Woodward S, Mowry B, Jorde LB. 2011. Ancestry of the Iban is predominantly Southeast Asian: genetic evidence from autosomal, mitochondrial, and Y chromosomes. PLoS One 6:e16338.

Smit AF. 1996. The origin of interspersed repeats in the human genome. Curr Opin Genet Dev 6:743–748.

Stoneking M, Delfin F. 2005. The human genetic history of East Asia: weaving a complex tapestry. Curr Biol 20:R188–R193.

Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, Singh L. 2005. Reconstructing the origin of Andaman Islanders. Science 308:996.

The Chimpanzee Sequencing and Analysis Consortium 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87.

The HUGO Pan-Asian SNP Consortium 2009. Mapping human genetic diversity in Asia. Science 326:1541–1545.

Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, Bonne-Tamir B, Santachiara-Benerecetti AS, Moral P, Krings M. 1996. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. Science 271:1380–1387.

Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AM, Carroll ML, Nguyen SV, Walker JA, Prasad BV, Reddy PG, Das PK, Batzer MA, Jorde LB. 2003. Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms. Genome Res 13:1607–1618.

Weiner DB, Watson SR, Babcock G, Keller SJ. 1986. Expression of human T antigens in interspecies hybridomas. Cell Immunol 100: 197–209.

Wise MR. 2008. Indigenous groups of Sabah: an annotated bibliography of linguistic and anthropological sources. Malaysia: The Natural History Publications.

Xing J, Watkins WS, Shlien A, Walker E, Huff CD, Witherspoon DJ, Zhang Y, Simonson TS, Weiss RB, Schiffman JD, Malkin D, Woodward SR, Jorde LB. 2010. Toward a more uniform sampling of human genetic diversity: a survey of worldwide populations by high-density genotyping. Genomics 96:199–210.