A suggested Khmuic origin of the hunter-gatherer Mlabri in northern Thailand: evidence from maternal DNA lineages

Jatupol Kampuansai[1]*, Wibhu Kutanan[2], Suphanun Phuphanitcharoenkul[1] and Daoroong Kangwanpong[1]

[1]*Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand*

[2]*Department of Biology, Faculty of Science, Khon Kaen University, Khon Kaen 40002, Thailand*

**Corresponding author: jatupol_k@hotmail.com*

ABSTRACT

Khamu, an ethnic group residing in the northern part of Thailand, are suggested to be the descendents of the Southeast Asian prehistoric inhabitants. Their language belongs to the Khmuic branch of the Mon-Khmer subfamily within the Austro-Asiatic family. This linguistic clustering suggests a shared common ancestor among the Khamu, H'tin, and hunter-gatherer Mlabri. Mitochondrial hypervariable region I of 39 Khamu and 19 Mlabri individuals, living in Nan province, were analyzed and compared with other ethnic groups in northern Thailand. The investigation on their maternal inherited marker revealed that they were one of the Southeast Asian indigenous populations. Their current genetic structures were shaped by the intense drift effect due to the ancient migrations along the Thai-Lao border. The Khamu are genetically closely related to the H'tin-Mal, a putative Mlabri's ancestral stock. Thus, this study proposed that the Khamu and the H'tin-Mal are possible ancestral origins of the Mlabri.

Keywords: mitochondrial DNA, hypervariable region I, genetic structure, Khamu, Mlabri

INTRODUCTION

After the interesting genetic pattern of the Mlabri, the last enigmatic hunter-gatherer group of northern Thailand, was reported in 2005, the tentative suggestions on their origin have been made (Oota *et al.,* 2005). Genetic evidences suggested that the Mlabri were recently founded from a very small number of individuals of an agricultural group, but they subsequently adopted their current hunting-gathering lifestyle (Oota *et al.,* 2005). The origin of the Mlabri and their cultural revision were more refined that they share recent common ancestor with the H'tin, their neighboring primitive agricultural groups in northern Thailand (Xu *et al.,* 2010). Their close genetic

affiliation was also found to be consistent with the linguistic classification. Languages used by both the Mlabri and the H'tin are grouped to the Khmuic branch of the Mon-Khmer sub-family in the Austro-Asiatic linguistic family (Lewis, 2009). Based on a calibration of the chronology of sound changes in H'tin, linguists proposed a divergence pattern of the Khmuic language, prior to the genetic affinity revelation, as following: the H'tin branched from the Khamu ethnic group about 600 years ago, then branched into two varieties (Mal and Prai) some 200-300 years ago. However, the divergence between Mlabri and H'tin languages cannot be determined, but most likely by not more than a few centuries (Filbeck, 1978; Filbeck, 1987). Even though, genetic evidences could reveal possibly origin of the Mlabri from the H'tin, there was no report on the genetic structure of the Khamu, the main speakers of Khmuic language, who might be the ancestral stock of the H'tin and the Mlabri. Thus, in this study, sequences from the mitochondrial DNA hypervariable region I (mtDNA HVR-I) of the Khamu were analyzed to determine their genetic structure, as well as the genetic relationship with their neighboring populations residing in northern Thailand. This mtDNA marker is widely used to address questions concerning the specific intra- and inter-population variation. Due to its maternal uniparental inheritance pattern and relatively high mutation rate, it has been extensively used to trace back biological ancestors, as well as to reconstruct the genetic history of populations.

The Khamu are the indigenous inhabitants of northern Thailand and northwestern Laos. Some sources state that they have occupied this area since at least one thousand years ago (Schliesinger, 2000). Most of the Khamu people reside mainly in dispersed area near the Thailand-Laos border. Some live in northern Vietnam and Xishuangbanna of southern China. There is also a Khamu community in California, USA, which was formed by refugees since the Vietnam War. In Thailand, the Khamu villages are found mainly in Nan and Chiang Rai provinces in the north. They also scatter in other provinces as laborers (Lewis, 2009). The Khamu in Huay Sataeng village, Thung Chang district, Nan province had migrated from Laos about 150 years ago. These people somehow believe that their ancestors were the original inhabitants of Nan province. Therefore, it is possible that they might have moved into Laos from Thailand in the first place, and came back recently along an ancient migration route around the outset of the Vietnam War (personal interviews). Nowadays, the Khamu in this village are still practicing their unique original traditions and speak Khamu language. Their subsistence economy is mainly based on small farm agriculture supplemented by hunting, trapping, fishing, and gathering of forest products.

MATERIALS AND METHODS

Five milliliters of peripheral blood samples from 39 unrelated Khamu volunteers from Ban Huay Sataeng, Thung Chang district,

Nan province were collected, with informed consent, using anticoagulant-EDTA vacutainers. Information on linguistic, cultural aspects, village history, and personal lineage was obtained by interviewing. Total genomic DNA was extracted from whole blood sample according to a standard inorganic salting out protocol (Seielstad *et al.,* 1999). The first hypervariable region (HVR-I) segment of the mtDNA control region (nucleotide positions 16048-16383; 336bp) of all Khamu individuals, together with 19 Mlabri DNA samples from our previous studies (The HUGO Pan-Asian SNP Consortium, 2009; Xu *et al.,* 2010), were amplified and directly sequenced, as previously described (Schurr *et al.,* 1999; Fucharoen *et al.,* 2001). The sequences were edited and aligned against the Cambridge Reference Sequence (Andrew *et al.,* 1999) using SeqScape software v2.0 (Applied Biosystem, Foster City, CA). PCR analysis of the 9-bp deletion between the cytochrome oxidase subunit II (COII) and lysine transfer RNA gene ($tRNA^{Lys}$) intergenic region was carried out as previously described (Fucharoen *et al.,* 2001).

Polymorphic sites and mtDNA haplotypes among Khamu samples were identified using DnaSP v5 software (Librado and Rozas, 2009). A neighbor joining tree of all observed haplotypes was constructed by MEGA4 software (Tamura *et al.,* 2007). Following the global haplogroup nomenclature (Schurr and Wallace, 2002; Yao and Zhang, 2002; Tanaka *et al.,* 2004), each haplotype was classified into one of the haplogroup. The demographic history was examined by mismatch distribution and neutrality approaches. DnaSP v5 software was used to plot the observed pairwise differences against the expected mismatch distribution under both constant size population and sudden population expansion models. Sum of squared deviation (*SSD*) and raggedness index (*r*) (Harpending, 1994) of the distribution were calculated and tested for their goodness-of-fit to expansion model using the Arlequin 3.5 program (Excoffier and Lischer, 2010). The calculation of neutrality statistics i.e., *D* (Tajima, 1989), R_2 (Ramos-Onsins and Rozas, 2002), and F_s (Fu, 1997), was performed and followed by coalescent simulated significance tests as implement in the DnaSP v5 software.

To investigate the relationship among the populations residing in northern Thailand, a joint dataset of the same length mtDNA HVR-I from the neighboring populations was assembled to the analysis, encompassing three main linguistic groups: the Mon-Khmer subfamily of the Austro-Asiatic family, the Tai subfamily of the Tai-Kadai family, and the hill-tribe (Table 1). Note that the hill-tribes, for this analysis, were the official classified populations as the hill-tribe by the Thai government. Their languages belong to the Sino-Tibetan (Akha, Karen, Lisu, and Musser) and the Hmong-Mien (Hmong and Yao) linguistic families (Lewis, 2009). The hierarchical Analysis of Molecular Variance (AMOVA) was performed to examine the genetic structure in 3 level i.e., (1) within population, (2) among populations within each

groups (Mon-Khmer, Tai, and hill tribes), and among groups. The Arlequin 3.5 package (Excoffier and Lischer, 2010) was used to calculate the genetic distance between each pair of populations. An unrooted neighbor joining tree, based on these pairwise difference distances, was constructed by MEGA4 software (Tamura *et al.,* 2007).

Table 1 General information of the studied populations and their genetic diversities.

Population	Province, district	No.	Reference	Group[b]	H[c]	π[d]
Akha1	Northern Thailand[a]	91	Oota et al., 2001	H	0.933±0.011	0.017±0.009
Akha2	Chiang Rai, Mueang	26	Besaggio, 2007	H	0.920±0.041	0.019±0.011
Blang1	Chiang Rai, Mae Sai	38	Kutanan, 2011	M	0.979±0.010	0.023±0.012
Blang2	Chiang Rai, Mae Chan	45	Kutanan, 2011	M	0.968±0.013	0.023±0.012
Hmong	Chiang Mai, Mae Rim	64	Besaggio, 2007	H	0.943±0.016	0.021±0.011
H'Tin1	Nan, Pua	37	Kutanan, 2011	M	0.740±0.062	0.016±0.009
H'Tin2	Nan, Thung Chang	25	Kutanan, 2011	M	0.643±0.071	0.013±0.007
H'Tin3	Nan, Chiang Klang	38	Kutanan, 2011	M	0.690±0.071	0.019±0.010
Karen1	Northern Thailand[a]	39	Oota et al., 2001	H	0.831±0.044	0.015±0.008
Karen2	Northern Thailand[a]	40	Oota et al., 2001	H	0.876±0.036	0.016±0.009
Karen3	Mae Hong Son, Mae Sarieng	74	Besaggio, 2007	H	0.907±0.021	0.020±0.010
Khamu	**Nan, Thung Chang**	**39**	**present study**	**M**	**0.881±0.028**	**0.021±0.011**
Khuen	Chiang Mai, San Pa Tong	60	Kampuansai, 2007	T	0.967±0.010	0.025±0.013
Lawa1	Mae Hong Son, Mae La Noi	46	Kutanan, 2010	M	0.959±0.013	0.019±0.010
Lawa2	Chiang Mai, Hod	50	Kutanan, 2010	M	0.913±0.018	0.019±0.010
Lisu1	Chiang Rai	53	Oota et al., 2001	H	0.925±0.024	0.019±0.010
Lisu2	Mae Hong Son	42	Oota et al., 2001	H	0.955±0.013	0.018±0.010
Lisu3	Chiang Mai	25	Fucharoen, 2001	H	0.980±0.018	0.021±0.011
Lisu4	Chiang Rai, Mueang	20	Besaggio, 2007	H	0.916±0.038	0.022±0.012
Lua1	Nan, Pua	20	Kutanan, 2010	M	0.642±0.087	0.015±0.009
Lua2	Nan, Bo Klea	43	Kutanan, 2010	M	0.173±0.072	0.003±0.002
Lue1	Nan, Pua	51	Kampuansai, 2007	T	0.915±0.027	0.020±0.010
Lue2	Nan, Ta Wang Pha	44	Kampuansai, 2007	T	0.878±0.026	0.021±0.011
Lue3	Chiang Rai, Mae Sai	50	Kampuansai, 2007	T	0.988±0.007	0.022±0.012
Lue4	Chiang Mai, Doi Sa Ket	46	Kampuansai, 2007	T	0.932±0.020	0.019±0.010
Mlabri	**Nan, Wiang Sa**	**19**	**present study**	**M**	**0.000±0.000**	**0.000±0.000**
Mon	Lamphun, Pa Sang	41	Kutanan, 2011	M	0.921±0.022	0.022±0.012
Mussur1	Chiang Mai	21	Fucharoen, 2001	H	0.881±0.047	0.018±0.010
Mussur2	Northern Thailand[1]	39	Oota et al., 2001	H	0.856±0.030	0.018±0.010
Mussur3	Chiang Rai, Mueang	25	Besaggio, 2007	H	0.980±0.018	0.024±0.013
Paluang	Chiang Mai, Fang	51	Kutanan, 2011	M	0.917±0.024	0.017±0.009
Yao	Chiang Rai, Mueang	44	Besaggio, 2007	H	0.987±0.008	0.030±0.016

Table 1 (continued).

Population	province, district	No.	Reference	Group[b]	H[c]	π[d]
Mussur3	Chiang Rai, Mueang	25	Besaggio, 2007	H	0.980±0.018	0.024±0.013
Paluang	Chiang Mai, Fang	51	Kutanan, 2011	M	0.917±0.024	0.017±0.009
Yao	Chiang Rai, Mueang	44	Besaggio, 2007	H	0.987±0.008	0.030±0.016
Yong	Lamphun, Pa Sang	62	Kampuansai, 2007	T	0.965±0.009	0.022±0.012
Yuan1	Chiang Mai, San Sai	39	Kampuansai, 2007	T	0.969±0.015	0.021±0.011
Yuan2	Chiang Mai, Mae Tang	50	Kampuansai, 2007	T	0.974±0.009	0.023±0.012
Yuan3	Lamphun, Ban Hong	50	Kampuansai, 2007	T	0.966±0.012	0.022±0.012

[a]The province was not specified by the references.

[b]Linguistic group: M-Mon-Khmer, T-Tai, H-Hill tribe.

[c]Haplotype diversity.

[d]Nucleotide diversity.

Bold letter: the samples collected in this study.

RESULTS

Among the 39 Khamu HVR-I sequences, 12 distinct haplotypes, defined by 22 substitutions, were found. Of these substitutions, 17 were transitions and 5 were transversions. Among 12 observed haplotypes, 4 were found in single individuals while others were shared by two or more people. Each mtDNA haplotype was identified as one of the reference haplogroup (Schurr and Wallace, 2002; Yao and Zhang, 2002; Tanaka *et al.,* 2004) according to their variation from the Cambridge Reference Sequence (Andrew *et al.,* 1999). Fourteen samples (35.9%) were classified to haplogroup F by T16304C. Twelve of these could be further defined as members of the F1a subgroup by a T16172C transition. The 9-bp COII/tRNA^{Lys} intergenic deletions were found in 17 samples (43.6%). All these 9-bp deletion samples contained a motif of T16140C-T16189C-C16266A polymorphism characterizing the subgroup B5a of haplogroup B. The C16223T and T16362C transitions were observed in 6 samples (15.4%), which were then classified in the subgroup D of haplogroup M. Two other samples (5.1%) also contained a C16223T which is mostly found in the haplogroup M sequences. However, this variation was also reported in some non-M haplogroup such as N9a (Yao and Zhang, 2002). Thus, as no coding-region information was available in this study, these two samples were left unidentified (Fig. 1). All 19 Mlabri sequences were identical based on the T16140C-A16182C-A16183C-T16189C-C16266A substitutions, together with the 9-bp COII/tRNA^{Lys} intergenic deletions. This Mlabri single motif is shared with one of the B5a Khamu sequences.

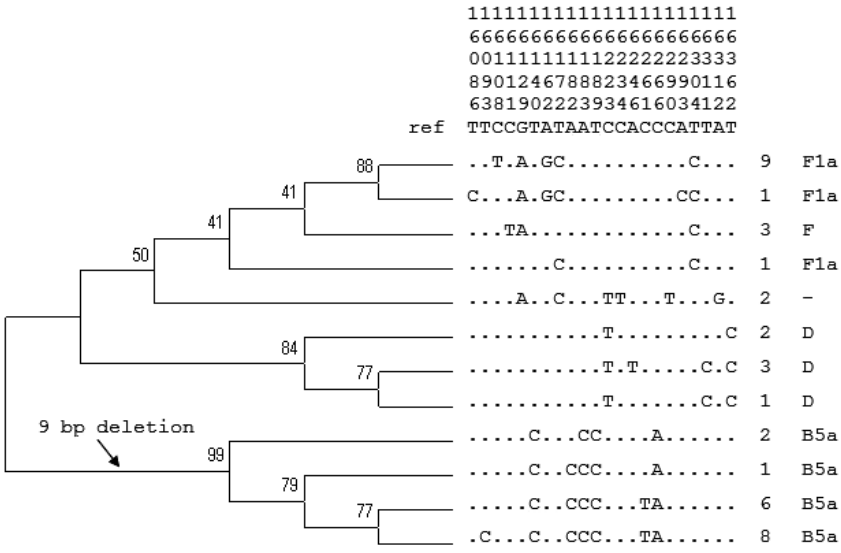

Figure 1 Nucleotide sequence differences in the HVR-I of mtDNA for 39 Khamu samples. Presented are nucleotides at 22 polymorphic sites based on a comparison of 336-bp in the reference sequence of Andrew *et al.* (1999), shown below the order for polymorphic sites. For other sequences, only differences from the reference sequence are indicated. The number of individuals with each sequence type is shown in Arabic numerals to the right of each sequence, together with their haplogroup (- is an unidentified haplogroup). The neighbor-joining (NJ) tree, based on the pairwise difference distances, among each sequence types is shown on the left. The bootstrap values are indicated above each branch. All sequence types with the 9-bp deletion appear after the branching point indicated by the arrow.

Demographic history of the Khamu was reconstructed using the mismatch distribution analysis and the neutrality tests. The observed mismatch distribution showed multimodal-like pattern, the characteristic of population with constant size or in demographic equilibrium (Rogers and Harpending, 1992) (Fig. 2). Even though, the null hypothesis of goodness-of-fit test under population expansion model could not be rejected (*SSD*=0.015 p=0.25 and *r*=0.024 p=0.43), the population growth was not detected with the neutrality statistics (*D*=1.14 p=0.89; R_2=0.16 p=0.91 and F_s=1.26 p=0.73). The contradictory of the expansion detecting parameters based on mismatch distribution (*SSD* and *r*) and mutation spectrum neutrality (*D*, R_2, and F_s) had been described previously (Ramos-Onsins and Rozas, 2002). Using the coalescent simulations of non-recombining DNA sequences, the statistics based on the frequency spectrum of mutations, especially R_2 and F_s, are the most powerful test for detecting population growth and must provide a large significant negative

value if sudden population expansion occur. The non-significance ($p>0.05$) of neutrality statistics observed in Khamu indicated no large-scale expansion signal and the number of residences might be in a stationary phase of mutation drift equilibrium or probably shrinkage (Depaulis *et al.,* 2003).

Considering the combined 36 population dataset, the Lue3 showed the highest haplotype diversity ($h = 0.988\pm0.007$), while the Lua2 showed the lowest value of 0.173 ± 0.072.

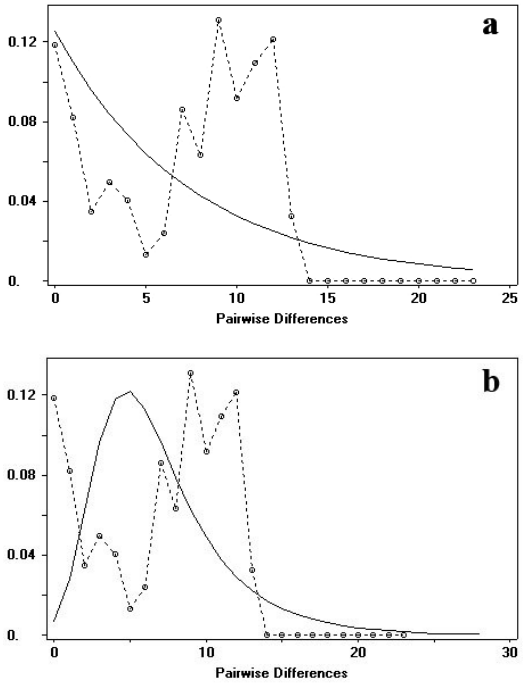

Figure 2 Mismatch distributions of the HVR-I sequences of the 39 Khamu samples. The dotted lines indicate the observed distributions and the solid lines indicate the expected distributions under constant size population model (a) and sudden population expansion model (b).

Nucleotide diversities (π) varied in different ethnic populations, ranging from 0.003 ± 0.002 (Lua2) to 0.025 ± 0.013 (Khuen). The intra-population diversities of the Khamu were 0.881 ± 0.028 and 0.021 ± 0.011 for haplotype and nucleotide diversities, respectively. No diversity was seen in the Mlabri, since there was only one mitochondrial HVR-I haplotype observed (Table 1).

Genetic structure of the combined dataset was investigated by AMOVA (Excoffier *et al.,* 1992). When all 36 populations were grouped together, 87.72% of the genetic variation was found within populations, whereas the variation among populations was 12.28%. The genetic variation between populations within the Mon-Khmer group (22.76%) was by far larger than that of the Tai (4.55%) or Hill tribe (7.67%). The proportion of genetic variation attributed to the difference among 3 linguistic groups was at 1.16% ($p<0.05$) (Table 2).

An unrooted neighbor joining tree of the 36 populations, residing in northern part of Thailand, was constructed using pairwise different distances (Fig. 3). All Tai and Hill- tribe samples formed a tight cluster, except the Hmong who seem to be closely related to Mon-Khmer group. Within the Mon-Khmer speaking group, 2 clades were identified: the Khamu, H'tin, Mlabri, Paluang, and Mon in one clade (Mon-Khmer A), and the rest in the other (Mon-Khmer B) (Fig. 3). These 2 Mon-Khmer clades were significantly different ($p<0.05$), as tested by AMOVA (Table 2).

Table 2 Results of the Analysis of Molecular Variance (AMOVA).

Groups	No. of population	Within populations		Among populations Within groups		Among groups	
		Variance (%)	Φ_{st}	Variance (%)	Φ_{sc}	Variance (%)	Φ_{ct}
All samples	36	87.72	**0.1228**	12.28			
Mon-Khmer	13	77.24	**0.2276**	22.76			
Mon-Khmer a/b*	5/8	75.99	**0.2401**	20.69	**0.2140**	3.32	**0.0332**
Tai	9	95.45	**0.0455**	4.55			
Hill tribe	14	92.33	**0.0767**	7.67			
Language (M/T/H)**	13/9/14	87.39	**0.1261**	11.45	**0.1158**	1.16	**0.0116**

Bold letter: statistical significant at $P < 0.05$

*Grouped according to the NJ tree (Fig. 3)

**Three linguistic groups: Mon-Khmer (M), Tai (T), and Hill tribes (H)

DISCUSSION

Genetic structure of the Khamu

In the present study, 336-bp sequences in the mtDNA hypervariable region I (nucleotide position 16048-16383) and the deletion in the COII/ $tRNA^{Lys}$ intergenic region of 39 Khamu samples from individuals residing in northern Thailand, were analyzed. The 9-bp deletion in the COII/$tRNA^{Lys}$ intergenic region of mtDNA is one of the characteristics of Asians, and ethnic groups of Asian origin including Polynesians and Native Americans (Ballinger *et al.,* 1992; Yao and Zhang, 2002). In this study, this marker was detected in the Khamu population with the frequency of 44%, which was higher than that from the previous investigation among Thai populations (18% to 40%, Fucharoen *et al.,* 2001). Although some studies have implied multiple events of this deletion in some populations (Ballinger *et al.,* 1992; Schurr *et al.,* 1999; Fucharoen *et al.,* 2001), our phylogenetic analysis showed a single origin of the deletion in Khamu lineages, suggesting that there was no admixture process within this group. Using this 9-bp deletion together with the substitutions of HVR-I sequences, the haplogroups of the Khamu lineages could be investigated. Most of their maternal lineages were classified into the haplogroup B, F, and M. These haplogroups are commonly presented in high frequency among Southeast Asian populations (Schurr and Wallace, 2002). Thus, it might be tentatively inferred that the Khamu are one of the Southeast Asian indigenous peoples, or they share their common ancestor with other prehistoric Southeast Asian ethnic groups.

The Pleistocene expansion is one characteristic of the prehistoric human lineages. The growth of population members, due to the decline of the ice age, left its imprint in the

smooth bell-shape mismatch distribution. The estimate expansion times were different in each region of the world starting from 110,000 B.P in east Africa to 57,000 B.P. in America (Ward *et al.,* 1991). For the East and Southeast Asian populations, the expansion time was between 66,000 to 60,000 B.P. (Yao *et al.,* 2002). Even though the expansion scenario was expected to be seen in every population inheriting the prehistoric genetic structure, this signal might be erased by the drift effect and left only a few major lineages in present day population. There was no certain expansion signal in Khamu and, moreover, the neutrality tests suggested a shrinkage or very little change in number of residences. If we referred to the Khamu Huay Sataeng village's history that they had been wandering between Thai-Laos border, and then a small group of them resettled in northern Thailand around 150 years ago (personal interviews), it would be reasonable to interpret that the Khamu might pass a period of expansion, but the signal was wiped out by the intense founder effect during their migration and settlement.

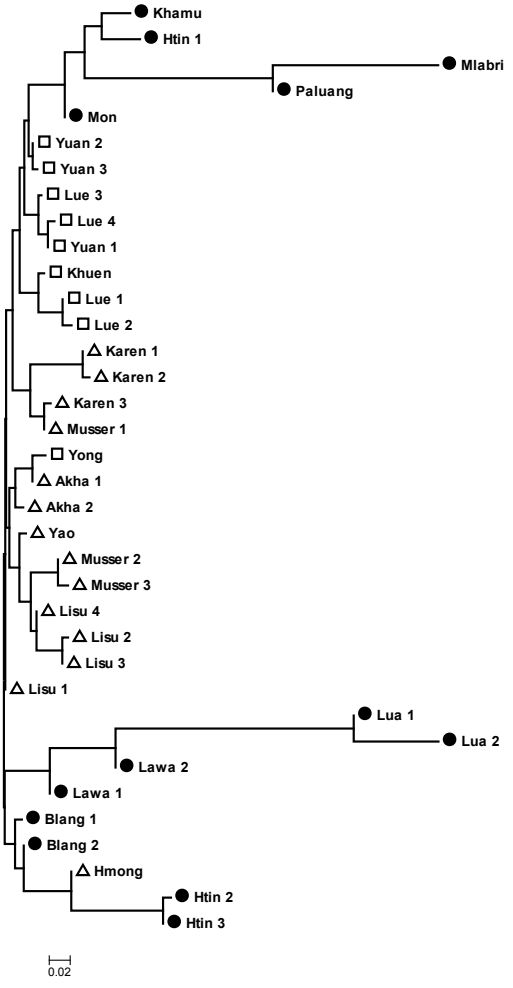

Figure 3 Unrooted neighbor-joining tree of the 36 populations based on number of pairwise difference distances. ● Mon-Khmer speaking group, □ Tai speaking group, △ Hill-tribe group.

Genetic relationship of Khamu and other ethnic groups

Theoretically, languages in the same family might originate in one area and could distribute to be the local languages in several places. The shared features of languages in one group can be due to shared ancestors. Therefore, close genetic relationship is expected to be seen among the populations using closely related languages (Barbujani and Sokal, 1990). In this study, the languages of populations fall into three groups: Tai, Mon-Khmer, and hill-tribes. Although the difference among linguistic groups was significant as shown by the AMOVA analysis, it explained only 1.16% of the total variation. It seems that, on average, the genetic structures of these 3

speaking groups are similar. The differentiations are more likely to occur within each population (87.39%) and between populations within group (11.45%). Thus, this can be interpreted that the linguistic boundary had little, but still important, influence on the distribution of genetic relationship among peoples residing in northern Thailand. Other forces, such as drift, had much more effect on their genetic structures.

Focusing within each linguistic group, the Mon-Khmer exhibited the highest genetic divergence among their members, as shown by their higher genetic variation (22.76%) and the longer branches in the neighbor-joining tree (Fig. 3) than those of the Tai and hill-tribe groups. The increase of the genetic divergence within the Mon-Khmer speaking group was explained in a previous study (Kutanan *et al.,* 2011) that it might be due to the drift and the post-marital residence cultures. The Mon-Khmer speaking peoples, the indigenous inhabitants of northern Thailand, were fragmented and exiled to rural area during and after the Tai colonization in the thirteenth century A.D. (Penth, 2000). This scenario drove the founder effect on each Mon-Khmer populations and enhanced the genetic divergence among them. The lower mtDNA haplotype diversity in the matrilocal (males move to his wife's house after marriage) H'tin and Lua than the patrilocal (females move to her husband's house) Mon-Khmer samples (Table 1) also supported the post-marital residence effect in this group as previously reported (Kutanan *et al.,* 2011). The clustering of two Mon-Khmer clades (A and B) was not consistent with the branching of the Paluangic (Paluang, Blang, Lawa) and Khmuic (Khamu, Htin, Mlabri, Lua) languages within Mon-Khmer subfamily of Austro-Asiatic linguistic family. This result confirmed the small effect of the linguistic boundary on the genetic relationship within this linguistic group.

The Mlabri was suggested to be originated from a small agricultural group who then adopted a hunting-gathering subsistence mode, unlike other hunter-gatherer groups who practice the ancient lifestyle. This cultural revision in Mlabri was supported by the sharing of mtDNA haplotype between the Mlabri and other agricultural Southeast Asian populations (Oota *et al.,* 2005). The Khamu is another ethnic group who shares the Mlabri mtDNA single haplotype and is closely related to the H'tin. Note that there are two H'tin varieties, Mal and Prai, and only the Mal (H'tin1) are genetically closely related to the Khamu, while the Prai (H'tin2, 3) are not. This H'tin-Mal was a putative Mlabri's ancestral stock (Xu *et al.,* 2009) and lexically branched from the Khamu ethnic group about 600 years ago (Filback, 1978). From our finding, we propose an extended view about the origin of Mlabri that their ancestor is expected to be not only the H'tin-Mal, but also the Khamu. This view has to be focused on in further analyses. The ancestral origin of the Mlabri might be either

Khamu or H'tin-Mal, or both of them, if the Mlabri had embraced a cultural revision earlier than the divergence of these two ethnic groups.

Even though, our results, to our knowledge, are the first evidence showing the genetic structure of the Khamu and their relationship to other ethnic groups, only one uniparentally inherited marker was analyzed. Moreover, only a small number of Khamu samples was studied, compared to the large population size of Khamu (estimated at around 31,400 individuals), scattering in many provinces of Thailand (Lewis, 2009). More genetic markers, larger sample size, and more sampling areas should fill in the descriptions on their ethnic history, and make it possible to determine their and other prehistoric groups' ancient ancestries, especially the hunter-gatherer Mlabri.

ACKNOWLEDGEMENTS

We thank all blood donors and village chiefs for their participation in interviewing and sample collecting. This study was financially supported by the Junior Researcher Fellowship 2009, Chiang Mai University, Thailand.

REFERENCES

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. and Howell, N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23: 147.

Ballinger, S.W., Schurr, T.G., Toronni, A., Gan, Y.Y., Hodge, J.A., Hassan, K., Chen, K.H. and Wallace, D.C. 1992. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migration. *Genetics* 130: 139–152.

Barbujani, G. and Sokal, R. 1990. Zones of sharp genetic change in Europe are also linguistic boundaries. *Proc Natl Acad Sci USA* 87: 1816–1819.

Besaggio, D., Fuselli, S., Srikummool, M., Kampuansai, J., Castrì, L., Tyler-Smith, C., Seielstad, M., Kangwanpong, D. and Bertorelle, G. 2007. Genetic variation in Northern Thailand Hill Tribes: origins and relationships with social structure and linguistic differences. *BMC Evol Biol* 7: S12.

Depaulis, F., Mousset S. and Veuille M. 2003. Power of neutrality tests to detect bottlenecks and hitchhiking. *Mol Evol* 57: 190–200.

Excoffier, L. and Lischer, H.E.L. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Evol Res* 10: 564–567.

Excoffier, L., Smouse, P. and Wuattro, J. 1992. Analysis of molecular variance inferred from metric distance among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.

Filbeck, D. 1978. *Tin, a Historical Study*. Canberra, Australian National University.

Filbeck, D. 1987. New ethnic names for the Tin of Nan province. *J Siam Soc* 75:129–138.

Fu, Y.X. 1997. Statistical tests of neutrality of mutations against population growth,

hitchhiking and background selection. *Genetics* 147: 915–925.

Fucharoen, G., Fucharoen, S. and Horai, S. 2001. Mitochondrial DNA polymorphisms in Thailand. *J Hum Genet* 46: 115–125.

Harpending H.C. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biol* 66: 591–600.

Kampuansai, J., Bertorelle, G., Castri, L., Nakbunlung, S., Seielstad, M. and Kangwanpong, D. 2007. Mitochondrial DNA Variation of Tai Speaking Peoples in Northern Thailand. *Science Asia* 33: 443–448.

Kutanan, W., Kampuansai, J., Fuselli, S., Nakbunlung, S., Seielstad, M., Bertorelle, G. and Kangwanpong, D. 2011. Genetic structure of the Mon-Khmer speaking groups and their affinity to the neighbouring Tai populations in Northern Thailand. *BMC Genetics* 12: 56.

Kutanan, W. and Kangwanpong, D. Genetic structure of the ethnic populations Lua and H'tin in Northern Thailand. 2010. *Thai J Genet* 3: 160–171 (in Thai).

Lewis, M.P. 2009. *Ethnologue: Languages of the World, Sixteenth edition. Dallas, Tex.: SIL International*. Online version: http://www.ethnologue.com/.

Librado, P. and Rozas, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.

Oota, H., Pakendorf, B., Weiss, G., Haeseler, A., Pookajorn, S., Settheetham-Ishida, W., Tiwawech, D., Ishida, T. and Stoneking, M. 2005. Recent origin and cultural reversion of a hunter-gatherer group. *PLoS Biology* 3: e71.

Oota, H., Settheetham-Ishida, W., Tiwawech, D., Ishida, T. and Stoneking, M. 2001. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nat Genet* 29: 20–21.

Penth, H. 2000. *A Brief History of Lanna: Civilizations of North Thailand*. Silkworm books, Chiang Mai.

Ramos-Onsins, S.E. and Rozas, J. 2002. Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* 19: 2092–2100.

Rogers, A.R., Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic distances. *Mol Biol Evol* 9: 552–569.

Schliesinger, J. 2000. *Ethnic Groups of Thailand: Non-Tai-Speaking Peoples.* White Lotus Press, Bangkok.

Schurr, T.G., Sukernik, R.I., Starikovskaya, Y.B. and Wallace, D.C. 1999. Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk sea-Bering sea region during the Neolithic. *Am J Phys Anthropol* 108: 1–39.

Schurr, T.G. and Wallace, D.C. 2002. Mitochondrial DNA diversity in southeast Asian population. *Hum Biol* 74: 431–452.

Seielstad, M., Bekele, E., Ibrahim, M., Toure, A. and Traore, M. 1999. A view of modern human origins from Y chromosome microsatellite variation. *Genome Res* 9: 558–567.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596–1599.

Tanaka, M., Cabrera, V.M., González, A.M., Larruga, J.M., Takeyasu, T., Fuku, N., *et al*. 2004. Mitochondrial genome variation in eastern Asia and the Peopling of Japan. *Genome Res* 14: 1832–1850.

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

Ward, R.H., Frazier, B.L., Dew-Jager, K. and Paabo, S. 1991. Extensive mitochondrial diversity within a single Amerindian tribe. *Proc Natl Acad Sci USA* 88: 8720–8724.

Yao, Y.G., Nie, L., Harpending, H., Fu, Y.X., Yuan, Z.G. and Zhang, Y.P. 2002. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *Am J Phys Anthropol* 118: 63–76.

Yao, Y.G. and Zhang, Y.P. 2002. Phylo geographic analysis of mtDNA variation in four ethnic populations from Yunnan province: new data and a reappraisal. *J Hum Genet* 47: 311–318.

Xu, S., Kangwanpong, D., Seielstad, M., Srikummool, M., Kampuansai, J. and Jin, L. 2010. Genetic evidence supports linguistic affinity of Mlabri - a hunter-gatherer group in Thailand. *BMC Genetics* 11: 18.