Bull. Natl. Mus. Nat. Sci., Ser. D, **35**, pp. 25–34, December 22, 2009

Mitochondrial DNA diversity and population differentiation in ethnic groups in Laos

Ken-ichi Shinoda[1], Kazuo Umetsu[2] and Yoshiharu Sekino[3]

[1] Department of Anthropology, National Museum of Nature and Science, 3–23–1 Hyakunincho, Shinjuku-ku, Tokyo 169–0073, Japan
[2] Department of Forensic Medicine, Faculty of Medicine, Yamagata University, 2–2–2 Iida-Nishi, Yamagata-shi, Yamagata 990–9585, Japan
[3] Department of Liberal Arts, Musashino Art University, 1–736 Kodaira-shi, Tokyo 187–8505, Japan

Abstract Laos is a country inhabited by more than 100 indigenous ethnic groups and subgroups, many of which permeate neighboring countries. Hence, studying the genetic history of the ethnic populations in Laos can provide crucial clues to the peopling of Southeast Asia. Among the ethnic groups, the Lao Lum (lowland Lao) is the largest and most widespread. The Kham are believed to be the earliest inhabitants of Laos. The Lao Sun live at high altitudes in northern Laos. Ethnohistorical data suggest that the Lao Lum migrated to Southeast Asia from the north and arrived in present-day Laos by the 10th century. The Kham migrated from the north, probably from the Yangtze River basin, in prehistoric times. The Mon, individuals of one of the Lao Sun subgroups, migrated to Laos from southern China and share their ethnicity with the Muong of Vietnam.

To address this hypothesis, we obtained sequences of hypervariable segment (HVS) 1 and HVS2 from a total of 70 samples of individuals belonging 3 ethnic groups. Along with these, we used 16 diagnostic variants found in the coding regions and compared these with sequences obtained from other Southeast Asian populations. Among the individuals analyzed by HVS1 and HVS2 sequencing, we observed 37 different haplotypes defined by 82 polymorphic sites. The mitochondrial DNA (mtDNA) data show that the Lao Sum display a close affinity for both the North Vietnamese and the Thai, whereas the Lao Lum display a close affinity with the original indigenous groups of Southeast Asia. The Kham are unique because they exclusively share a large number of haplogroups with original inhabitants of the south of Southeast Asia and the north of Southeast Asia.

This is the first systematic study on the genetic structure of Laos's populations. The study aims at studying the matrilineal ancestry of the people of Laos and providing more data for evaluating the genetic structure of East Asians.

Key words : Mitochondrial DNA, Haplogroup, Laos, Ethnic group, Population genetics

Introduction

Largely on the basis of geography, Southeast Asia is commonly divided into 2 regions: the islands and the mainland. The islands comprise Indonesia, east Malaysia, and the Philippines, whereas the mainland comprises Burma (Myanmar), Cambodia, Laos, Thailand, Vietnam, and west Malaysia. Because of its geographical position, Southeast Asia is considered to have been an important route used by the first modern humans who migrated from the southern part of Asia to the northeast of Asia. Individuals belonging to the so-called "Australo-Melanesian" race arrived in the southeast of Asia about 50,000 years ago (Barker et al., 2007). They were the first inhabitants of both Sahul and the ancient continent of Sundaland, which formed most of the islands and the mainland of Southeast Asia. Moreover, recent migrations have been by individuals of the Austronesian populations, who are believed to have left Taiwan or formerly the coast

of South China about 6,000 years ago; these populations have spread throughout the Pacific via the coastal region of Southeast Asia (Bellwood, 1997).

Hence, the population history of the indigenous people of Southeast Asia may provide crucial insights into human migration in East Asia and the Pacific region. The origin and genetic affinities of the Southeast Asian population have been examined from many perspectives in anthropological studies (Barker et al., 2007; Matsumura et al., 2008).

Recently, because of molecular anthropological studies, there has been significant progress in understanding the peopling of the Southeast Asian islands (Hill et al., 2006, 2007; Li et al., 2007; Soares et al., 2008; Moodley et al., 2009). However, little is known regarding the peopling of the mainland because of the lack of fundamental DNA data. There have been reports on mitochondrial DNA (mtDNA) data from Thailand and Vietnam (Fucharoen et al., 2001; Irwin et al., 2008; Lertrit et al., 2008; Patcharee et al., 2008); however, increasingly volatile political situations in some countries, including Laos and Myanmar, pose difficulties in obtaining DNA data.

Although there are 49 officially recognized ethnicities in Laos, more than a hundred indigenous ethnic groups and subgroups inhabit its landscape, many of which permeate neighboring countries but retain their own culture. In this study, we investigated genetic relationships among 3 ethnic groups whose samples we were able to obtain. In 2007, one of the authors (Y.S.) conducted an ethnological survey among individuals inhabiting the region along the Mekong River in Laos, and obtained DNA samples from 3 ethnic groups.

The first group is Lao Lum (lowland Lao), a diverse group that accounts for 66% of the population and comprises groups that live at lower altitudes, profess Buddhism, and practice wet-rice cultivation. Among the 3 groups, the Lao Lum is the largest in number and the most widespread. It is believed that they migrated to Southeast Asia from the north and arrived in present-day Laos by the 10th century. The second group comprises the Kham. Individuals in this group live at medium altitudes, practice slash-and-burn agriculture, and profess animism. It is believed that they are the earliest inhabitants of Laos who migrated from the north (probably from the Yangtze River basin) during prehistoric times. The third group Lao Sun includes the Punoi, Pusan, and Mon people. It is believed that the Mon people migrated to Laos from southern China and share their ethnicity with the Muong of Vietnam. They live at high altitudes in northern Laos and also use slash-and-burn methods of farming.

The present-day natives of mainland Southeast Asia can be classified according to their linguistic affiliations. The indigenous languages of Laos can be categorized into 4 major groups: the Daic or Tai-Kadai language, Mon-Khmer (a subgroup of the Austro-Asiatic language family), Tibeto-Burman (a subgroup of the Sino-Tibetan language family), and Hmong-Mien. The Lao Lum and Kham speak Tai-Kadai, whereas the Mon speak Mon-Khmer (Lewis, 2009).

mtDNA has been extensively studied in recent yeas as part of population genetics studies. However, no genetic investigations have been carried out for the population of Laos. The present study is the first to provide mtDNA sequencing data and haplogroup frequencies in 3 Lao ethnic groups. Their genetic affinities have been studied and compared with other Southeast Asian populations. This study will be useful not only for understanding the complex history of the peopling of the mainland of Southeast Asia but also for improving our knowledge on the distribution of mtDNA variation in the mainland Southeast Asian population.

Materials and Methods

Cheek swabs of 70 anonymous, maternally unrelated individuals were collected. These samples included those of individuals belonging to the 3 ethnic groups (locations shown in Figure 1): 12 Lao Lum, 27 Kham, and 31 Lao Sun (Mon, 23 samples; Punoi, 2 samples; Pusan, 6 samples).

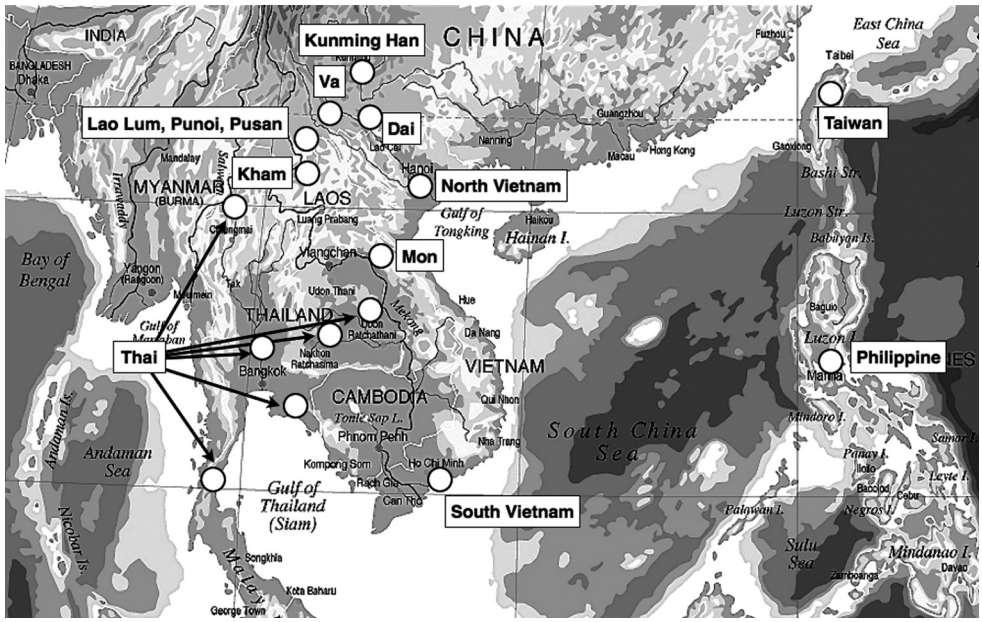

Fig. 1. The map of East and Southeast Asia. The circles indicate the sampling points of the present study and additional published data points used in the analyses.

Appropriate informed consent was obtained from all participants. This study was also formally approved by the relevant national level administrative bodies in Laos. These samples were exported to Japan, and DNA was extracted using the Puregene buccal cell DNA isolation kit (Gentra Systems, Minneapolis, Minnesota). This study was performed in keeping with the ethic guidelines of the Japan Anthropological Society.

Direct sequencing of PCR products from the D-loop region of mtDNA

A segment of hypervariable region (HVR) 1 (nucleotide positions 16121 to 16402, according to the revised Cambridge reference sequence (rCRS); Andrews et al., 1999) and HVR2 (128 to 408) was sequenced for all samples. Aliquots (2 μl) of the extracts were used as templates for polymerase chain reaction (PCR). Amplifications were carried out in a reaction mixture (total volume, 25 μl), containing 1 unit of Taq DNA polymerase (HotStarTaq™ DNA polymerase; Qiagen), 0.1 μM of each primer, and 100 μM deoxyribonucleoside triphosphates (dNTPs) in 1× PCR buffer (supplied by the manufacturer). The PCR conditions were as follows: incubation at 95°C for 15 min; 30 cycles of heat treatment at 94°C for 20 s, 50–56°C for 20 s, and 72°C for 15 s; and a final extension step at 72°C for 1 min. The following primers were used to amplify the HVR1 and HVR2 region: (L16120) 5′-TTACTGCCAGCCACCATGAA-3′, (H16403) 5′-TTGATTTCACGGAGGATGGTG-3′, (L127) 5′-AGCACCCTATGTCGCAGTAT-3′, and (H409) 5′-TGTTAAAAGTGCATACCGCC-3′. The PCR products were subjected to agarose gel electrophoresis on a 1.5% gel and were recovered by using a QIAEX II agarose gel extraction kit (Qiagen, Germany). Aliquots of the samples were prepared for sequencing using a BigDye cycle sequencing kit (Applied Biosystems, Foster City, CA, USA), which was performed using forward and reverse primers. The primers that were used for PCR amplification were also used in the sequencing reaction. Sequencing was performed using a single primer in both directions so as to enable identification of polymorphisms or ambiguous bases. The sequencing reactions were performed on a DNA sequencer (ABI model no. 3130) equipped with the SeqEd software.

PCR-amplified product length polymorphism analysis

In this study, a simple and rapid amplified product length polymorphism (APLP) method has been employed for the detection of mtDNA haplogroups (Umetsu et al., 2001, 2005). Using APLP, diagnostic polymorphic sites can be examined directly. Thus, with this method, it is possible to avoid ambiguity (caused by insufficient enzymatic digestion) with respect to the actual mutation site, which is one of the disadvantages of the RFLP method.

The amplification reaction was conducted in a DNA thermal cycler 9700 (Applied Biosystems, Foster City, CA, USA) by using 4 sets of 9-plex PCR assays. Multiplex A was designed to analyze the target sites 10398A, intergenic COII/tRNALys 9-bp deletion, 5178A, 3010A, 14979C, 8020A, 13104G, 11215T, and 11959G; Multiplex B, to analyze target sites 10400T, 8793C, 4833G, 8200C, 3394C, 14178C, 3970T, 5417A, and 13183G; Multiplex C, to analyze target sites 3594T, 11969A, 11696A, 6455T, 4386C, 12811C, 15487T, 8684T, and 1736G; Multiplex D, to analyze target sites 10873T, 8994A, 4580A, 10550G, 12308G, 1719A, 15607G, 7028C, and 12612G; and Multiplex E, to analyze target sites 12705T, 8563G, 15535T, 12882T, 11084G, 14318C, 752T, 2766T, and 9296T.

Each reaction was performed in a total volume of 20 μL containing 5–15 ng genomic DNA, optimum concentrations of each primer (Umetsu et al. 2005), and reagents of the multiplex PCR kit (Qiagen, Hilden, Germany). The concentration of the primers was adjusted to obtain balanced amounts of PCR products. After an initial denaturation step of 15 min at 95°C, 32 amplification cycles were carried out, each consisting of 10 s at 94°C, 10 s at 52°C, and 5 s at 72°C. This was followed by a final extension step of 3 min at 72°C. If the analysis of mitochondrial single nucleotide polymorphisms (mtSNPs) failed because of a base substitution in a primer or a problem with the DNA quality, the mtSNP would be analyzed by a single PCR with the corresponding primer set.

The sizes of the PCR products ranged from 50 to 151 bp. A 2 mL aliquot of the PCR product was separated by electrophoresis in a 6.5 cm native polyacrylamide gel (10%T, 5%C), containing 375 mM Tris-HCl buffer (pH 8.9), with running buffer (12.5 mM Tris and 96 mM glycine; pH 8.3). DNA bands were detected fluorographically after staining with SYBR Gold (Molecular Probes, Eugene, OR, USA).

Data analysis

The 278-bp HVS1 sequence and the 283-bp HVS2 sequence were edited and aligned against the rCRSs using the Genetyx-Mac software (version 15). The nucleotide diversity and the mean number of pairwise differences between the mitochondrial D-loop sequences were computed using the Arlequin software package (version 3.0; Excoffier et al., 2005) based on Tamura-Nei distances and a gamma parameter value of 0.26 (Mayer et al., 1999). The length polymorphisms of the A and C stretches in region from 16180 to 16188 (triggered by the substitution of T with C at position 16189) were ignored in this analysis. The differences between the population were also computed using the Arlequin software (Raymond and Rousset, 1995). Neighbor-joining (NJ) trees were constructed on the basis of the pairwise Fst values by using the Mega 3.0 program (Kumar et al., 2004) in order to study the relationships between the populations.

For haplotypes defined in the region from 16121 to 16402, the haplotype diversity within groups was estimated by $1-i\sum x^2i$, where xi is the relative frequency in the sample of the ith haplotype (Torroni et al., 2001). Comparative data, often comprising HVS1 sequence data (Sykes et al. 1995; Betty et al. 1996; Fucharoen et al., 2001; Oota et al. 2002; Yao et al., 2002a,b,c; Tajima et al. 2004; Trejaut et al. 2005; Irwin et al., 2008), were obtained from the literature. The data used included samples from Thailand, Taiwan (aboriginal population), the Philippines, Han Chinese, and China and Vietnam (aboriginal populations). The geographical locations of these populations are shown in Figure 1.

Table 1. Sequence variation in the 70 Laos individuals analyzed in the present study

Haplo-type	Haplo-group	Mutations in the segments[a]		Ethnic Group[c]		
		HVS 1 16121–16402 (16000+)	HVS 2 128-408[b]	Lao Lum	Kham	Lao Sun
1	A*	223, 235, 290, 311, 319, 362	152, 183, 198, 233, 234, 309+C, 315+C			2 (Pusan)
2	A1	126, 223, 234, 235, 290, 319	153, 234, 309+C			1 (Punoi)
3	B4*	182C, 183C, 189, 217, 261, 357	309+C, 315+C			2 (Mon)
4	B4*	179, 182C, 183C, 189	150, 310, 315+C	1		
5	B5	183C, 189, 262, 266A	209, 315+C			4 (Mon)
6	B5	182C, 183C, 189, 266A	209 ,252, 315+C			5 (Mon)
7	B5	182C, 183C, 189,197G,	210, 310, 315+C	1		
8	C	189, 223, 298, 327, 389	207 ,249d, 309+C	1		
9	C	189, 223, 298, 327	N.D.			1 (Mon)
10	D4b2b	145, 223	195, 309+C		1	
11	D4e	223, 362	146, 227, 269, 309+C, 315+C			1 (Mon)
12	D4j	223, 362	227, 309+C	1		
13	D5a	182C, 183C, 189, 223, 362	205, 309+C		1	
14	D5a	182C, 183C, 172, 189, 223, 362	150, 309+C, 315+C			1 (Mon)
15	F*	192, 288, 304, 309, 389	N.D.		3	
16	F*	260, 298, 355, 362	204, 207, 249d, 279, 309+C, 315+C			1 (Mon)
17	F*	260, 298, 355, 362	204 ,207, 249d, 309+C, 315+C	1		
18	F1	129, 162, 172, 304	249d, 309+C			2 (Puni, Pusan)
19	F1	129, 162, 172, 304	249d, 309+C, 315+C			6
20	F1	129, 172, 304	152, 249d, 309+C	1	4	
21	F1	183C, 189	206, 249d, 309+C			1 (Mon)
22	F1	129, 162, 172, 304, 311	152, 249d, 309+C			1 (Mon)
23	F1	129, 304	N.D.	2	3	
24	F1	189, 304	146, 249d, 309+C, 315+CC, 326			1 (Pusan)
25	M*	129, 223, 272	152, 249d, 309+C, 316		1	
26	M*	271	151, 309+C		1	
27	M*	129, 166, 217, 223, 298A ,319, 381	152, 309+C, 315+C		2	
28	M*	183C, 189, 223, 311	146, 150, 152, 310, 315+C		3	
29	M*	129, 203, 223, 271	143, 146, 151,309+C		1	
30	M*	223, 311, 362, 381	215, 309+C		4	
31	M*	214A, 223, 256, 298	152, 309+C	1		
32	M*	215, 223, 274	130, 143, 150, 309+C, 315+C			1 (Pusan)
33	M7b	129, 192, 223, 297	150, 179, 198, 309+C	1		
34	M9	158, 223, 234, 311, 362	150, 152, 153, 309+C	1		
35	N*	167 ,223, 249	152, 227, 309+C, 315+C	1		
36	N*	(R) 182C, 183C, 189, 311, 389, 398	184, 189, 309+C, 315+CC, 366		1	
37	N9a	223, 257A, 261, 311	150, 227, 309+C, 315+C			2 (Mon)

[a] All polymorphic sites are numbered according to the revised Cambridge Reference Sequence (Andrews et al. 1999). CRS denotes that the sequence of the segment is identical to the revised Cambridge reference sequence, and N.D. indicates 'Not Determined'. The suffix A indicates a transversion, and d indicates a deletion. Deletions are recorded at the last possible site.

[b] Nucleotide change at position 263 in the segment 128–408 was observed in all specimens, therefore, it is omitted.

Results and Discussion

The sequence variation in HVS1 and HVS2 along with haplogroup classifications by PCR-APLP methods is shown in Table 1. HVS1 and HVS2 sequences were used to gain information on the internal population diversity. There were 37 distinct types of sequences defined by 82 polymorphic sites. Of these, 35 were unique to their respective populations, whereas only 2 (types 20 and 23) were shared by 3 Lao Lum individuals and 7 Kham individuals. Among the 35 unique types, 11 (31.4%) were shared by more than 2 individuals within each population, where-

as the remaining 24 types (68.6%) were observed in a single individual. Sequencing analysis and APLP testing indicated that mtDNA haplotypes in the Laos samples fell into the 2 major non-African haplogroups, M and N. However, 8 haplotypes could not be further classified into specific haplogroups of M that have been identified as M* in Table 1.

The 278-bp sequences of HVS1 from the 70 individuals were aligned and compared, and the number of nucleotide substitutions between each pair of sequences was estimated. Nucleotide diversity among the 70 individuals was estimated to be 0.02, which is slightly higher than the corresponding value estimated for the Taiwanese, East Asian, and Oceanian populations (0.014 to 0.019), and lower than that estimated for the south Vietnamese population (0.022) (Trejaut et al., 2005).

The haplotype diversity (H) and nucleotide diversity (p) estimated for each group are presented in Table 2. In the ethnic populations of Laos, H ranges from 0.878 to 0.91, and p from 0.018 to 0.02. These diversity ranges are similar to those observed in other ethnic groups in East Asia (Yao et al., 2002a; Wen et al., 2005). As shown in Table 2, the proportion of population-specific haplotypes ranges from 63% to 75%. The proportion of population-specific haplotypes is generally high in the Laos populations; this suggests that these ethnic groups continue to maintain their own ethnicity with a lesser degree of genetic influence from other ethnic groups. Significant differences based on haplotype frequencies were observed among all population pairs at a confidence level of 0.05. Tajima's D values of these populations were not significantly different from 0 at a confidence level of 0.05, indicating no signal for demographic expansions. Because the sample sizes in the data of Lao Lum were relatively small, the differences observed here may be because of insufficient sampling or potential regional differences, and hence should be considered with caution. However, the mismatch in distribution was unequivocally multimodal (data not shown), thus contradicting the hypothesis of population growth and pointing toward the strong effects of drift and/or small sample sizes.

Phylogenetic analysis conducted using coding region information clustered the observed haplotypes into 16 distinct haplogroups and subhaplogroups (Table 1). The 2 familiar and widespread Southeast Asian mtDNA haplogroups are haplogroups B and R9, the latter encompassing haplogroup F. Haplogroup F1a, which is common and widespread in Southeast Asia, is found largely in all 3 groups, especially the Kham. Although haplogroup B is only present at low frequencies in the Lao Lum groups and absent in the Kham population that we sampled, it has a high frequency in the Lao Sun group. This haplogroup, especially haplogroup B4, has been observed at high frequencies in the islands in Southeast Asia. Thus, the frequency of this haplogroup reflects the genetic affinity of this haplogroup to the islands in Southeast Asia.

Although, 3 basic haplogroups (B, D, and F) accounted for more than 75% of the variation observed in the 3 ethnic groups, 20% of the individuals were classified into haplogroups M* and N*. Because the diagnostic variants in the coding regions are mainly used for identifying the East Asian population, it is difficult to classify the Southeast Asian population further. Judging from the HVS1 and HVS2 sequences, the unclassified haplogroups M and N, which are found in Laos,

Table 2. Divesity Indices of Lao populations calculated from HVS1 region

Population	Size	Number of haplotype	Number of unique haplotype	Haplotype diversity	Nucleotide diversoty	Gene diversity	Mean number of pairwise differences	Tajima's D
Lao Lum	12	11	7 (63.6%)	0.903	0.019±0.011	0.985±0.040	5.394±2.796	−1.294
Kham	27	12	9 (75.0%)	0.878	0.018±0.010	0.912±0.030	5.043±2.527	−0.559
Lao Sun	31	16	12 (75.0%)	0.91	0.020±0.010	0.940±0.022	5.557±2.744	−0.519

appear to be only very distantly related to each other. If the remaining unclassified M* and N* haplogroups do represent indigenous haplogroups, then it would imply that about one-fourth of the modern inhabitants of Laos can trace their maternal ancestry back to the first anatomically modern inhabitants of Southeast Asia.

Although haplogroup N9a is widespread in mainland Southeast Asia and may have a southern origin, it is absent in the Lao Lum and Kham. Haplogroup M7b, which also is widespread in Southeast Asia, is only found in Lao Lum. The frequencies of haplogroup F in mainland and island Southeast Asia are the highest in Kham.

Haplogroup E is the most common indigenous haplogroup in island Southeast Asia. It is also common in Taiwanese aboriginals but is almost absent in China (Hill et al., 2007). Moreover, it is also absent in the samples examined in the present study. This means that there is only some genetic connection between the island areas and inland regions like Laos. In the present study, the results of the mtDNA analyses demonstrate that the populations of the islands in Southeast Asia have a different ancestry from most of the Laos population. The high frequencies of haplogroups in the north, such as A, C, and D (Yao et al. 2002a; Kong et al. 2003), found in the Lao Sun samples were consistent with their Chinese or Vietnamese origins.

To clarify the genetic characteristics of the regional population, the distribution of haplogroup frequencies among the sampled groups was compared to information available from other Asian populations in nearby regions (Figure 1). The populations in the nearby regions show great genetic and linguistic heterogeneity; further, the haplogroup frequencies varied among these populations (Table 3). Genetic differences (Fst values) among these 12 populations are shown in Table 4. With regard to the population of Laos, Fst values are highest between Kham and all other populations; this indicates that the Kham is

Table 3. The Haplogroup Distribution Frequencies (%) in the 12 Populations Considered in the Present Study

Haplo-group	Lao Lum	Kham	Lao Sun	Thai[a]	Vietnam (north)[b]	Vietnam (south)[c]	Aboriginal Taiwanese[d]	Philip-pine[e]	South China (Han)[f]	Kunming (Han)[g]	Dai[h]	Va[h]
A	0	0	9.7	2.8	1.7	0	0	0	0.8	4.7	3.2	1.7
B4	8.3	0	6.5	14.1	10.7	20	25.9	25.4	22.7	11.7	11.3	1.7
B5	8.3	0	29	15.3	12.4	5.7	5.9	8.5	2.5	7	8.1	1.7
C	8.3	0	3.2	4	4	2.9	0	0	1.7	4.7	9.7	0
D4	8.3	3.7	3.2	2.8	2.8	0	0	0	15.1	18.6	8.1	20.7
D5	0	3.7	3.2	1.1	1.1	0	4.8	0	5	4.6	4.8	1.7
E	0	0	0	0	0	0	11.4	16	0	0	0	0
F	33.3	59.3	22.6	28.7	22	11.4	26.6	10.6	23.6	23.3	24.1	39.6
M*	8.3	29.6	16.1	14.2	15.8	20	1.9	12.8	8.4	2.1	4.9	26
M7b	8.3	0	0	4.5	14.1	11.4	9	6.4	6.7	16.3	11.3	0
M7c	0	0	0	2.8	3.4	0	9	12.8	2.5	0	1.6	0
M9	8.3	0	0	6.8	0.6	0	0	0	0.8	0	0	5.2
N*	8.3	3.7	0	0.6	6.3	0	2.9	4.3	7.6	0	9.7	1.7
N9a	0	0	6.5	2.3	5.1	5.7	1.2	0	2.5	7	3.2	0
Y	0	0	0	0	0	22.9	1.4	3.2	0	0	0	0
n	12	27	31	177	187	35	640	94	119	43	60	70

a: Data from Fucharoen et al. (2001)
b: Irwin et al. (2008)
c: Oota et al. (2002)
d: Trejaut et al. (2005)
e: Phillippine include Luzon (Sykes et al. 1995) and Philippine (Tajima et al. 2004)
f: South China (Han) included Guangdong and Hong Kong Han Chinese. (Betty et al. 1996, Yao et al. 2002c)
g: Yao et al. (2002c)
h: Yao et al. (2002b)

Table 4. Pairwise Fst values between each pair of population

	Lao Lum	Kham	Lao Sun	Thai	N. Vietnam	S. Vietnam	Taiwan	Phillippin	South China	Kunming	Dai	Va
Lao Lum	0											
Kham	0.05018	0										
Lao Sun	0.00103	0.1256	0									
Thai	0.02856	0.07789	0.00673	0								
N. Vietnam	0.02183	0.09458	0.01633	0.00859	0							
S. Vietnam	0.01449	0.15081	0.04281	0.02189	0.00587	0						
Taiwan	0.01035	0.13461	0.07258	0.03529	0.03964	0.04207	0					
Phillippin	0.03746	0.17921	0.06671	0.04665	0.03916	0.02112	0.02042	0				
South China	0.01561	0.11494	0.05695	0.0253	0.02367	0.0253	0.02491	0.03983	0			
Kunming	0.02182	0.13417	0.04202	0.02729	0.01733	0.03935	0.04069	0.06357	0.0068	0		
Dai	0.03659	0.11176	0.02825	0.01418	0.00356	0.02845	0.02924	0.04531	0.00602	0.00396	0	
Va	0.00407	0.02345	0.07815	0.04824	0.06147	0.09742	0.11017	0.13081	0.05991	0.06513	0.06705	0

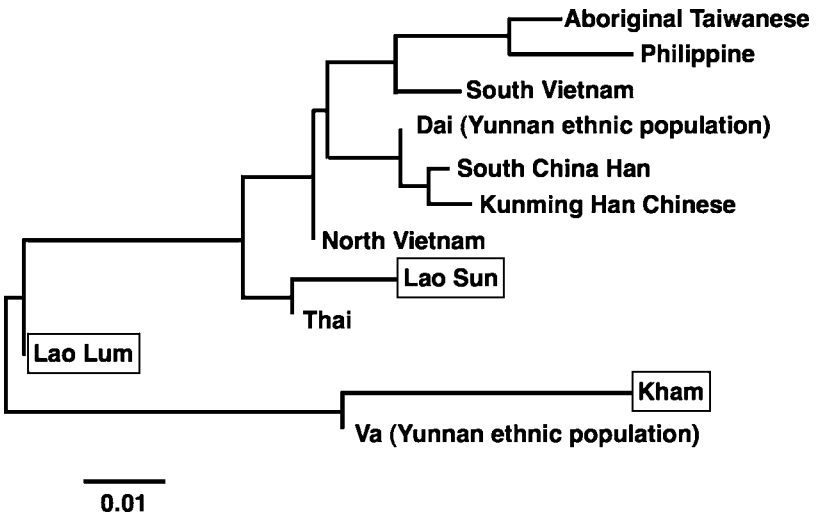

Fig. 2. Neighbor-joining tree based on the Fst values determined for 12 populations.

genetically distinct from the other populations living in Southeast Asia.

These genetic affinities are reflected in the NJ trees based on both the pairwise Fst values (Figure 2). This analysis of the haplogroup profiles presents a general view of the genetic differences. As demonstrated in Figure 2, the aboriginal Taiwanese and Philippine samples showed a close genetic relationship, and it stands to reason that the South Han Chinese and Kunming Han Chinese populations are closely related to each other. Further, populations in the northern part of Southeast Asia and the southern part of South China displayed a relatively short genetic distances. As expected, the Lao Sun was also included in this cluster. This is in agreement with the hypothesis that these populations are descendants of the immigrants from southern China or Vietnam.

On the other hand, Kham and Va are separated from all other populations by a large genetic distance. Further, it is evident that the Lao Lum occupies an intermediate position. It is interesting to note that the Va are regarded as the earliest inhabitants of southwest Yunnan, and they speak the Mon-Khmer language that belongs to the family of south Asian languages (Du and Yip, 1993). Yao et al. (2002b) pointed out that the Va may receive genetic contributions from north or northwest China, possibly during the southwest

migration of the ancient Di-Qiang tribe. The maternal genetic structure of the Kham and Va population bears the signatures of both southern and northern populations. Both ethnic groups have relatively high frequencies of F1 lineages (south) and D lineages (north). In addition, the Kham contains relatively high frequencies of indigenous Southeast Asian M* lineages. According to ethnological data, individuals of the Kham population are believed to have migrated from the Yangtze River basin in prehistoric times. It is also likely that they acquired genetic contributions during the southwest migration from their ancestral area.

The aim of this study was to gain insights into variations in mtDNA and determine the origins of the 3 ethnic groups in Laos. In particular, we wanted to test the hypothesis based on ethnological data that states that the 3 ethnic groups stem from a different ancestral population. In some cases, the mtDNA analyses revealed clear results on this issue. However, the origin and development of populations of ethnic groups are very complex processes. Therefore, hypotheses that are based on different data sets or different methods of analysis should contradict each other for the reconstruction of the population history. In this study, we focused on the 3 ethnic populations of Laos to infer genetic relationships among the individuals. It is necessary to conduct further studies for many of the remaining ethnic groups in order to reconstruct a more realistic depiction of the peopling of Laos as well as that of Southeast Asia as a whole.

Acknowledgments

We are grateful to all the volunteers who donated DNA samples for this study. This research was supported in part by a Grant-in-Aid for Scientific Research (No. 20370096) from the Ministry of Education, Culture, Science and Technology of Japan.

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. *Nature Genetics*, **23**: 147.

Barker G., Barton H., Bird M., Daly P., Datan I., Dykes A., Farr L., Gilbertson D., Harrison B., Hunt C., Higham T., Kealhofer L., Krigbaum J., Lewis H., McLaren S., Paz V., Pike A., Piper P., Pyatt B., Rabett R., Reynolds T., Rose J., Rushworth G., Stephens M., Stringer C., Thompson J., and Turney C. (2007) The "human revolution" in lowland tropical Southeast Asia: the antiquity and behavior of anatomically modern humans at Niah Cave (Sarawak, Borneo). *Journal of Human Evolution*, **52**: 243–261.

Bellwood P. (1997) Prehistory of the Indo-Malaysian archipelago. University of Hawaii Press, Hawaii.

Betty D.J., Chin-Atkins A.N., Croft L., Sraml M., and Easteal S. (1996) Multiple independent origins of the COII/tRNA^{Lys} intergenic 9-bp mtDNA deletion in aboriginal Australians. *American Journal of Human Genetics*, **58**: 428–433.

Du R. and Yip V.F. (1993) Ethnic groups in China. Science Press, Beijing.

Excoffier L., Laval G., and Schneider S., 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**: 47–50.

Fucharoen G., Fucharoen S., and Horai S. (2001) Mitochondrial DNA polymorphisms in Thailand. *Journal of Human Genetics*. **46**: 115–125.

Hill C., Soares P., Mormina M., Macaulay V., Meehan W., Blackburn J., Clarke D., Raja J.M., Ismail P., Bulbeck D., Oppenheimer S., and Richards M. (2006) Phylogeography and ethnogenesis of aboriginal Southeast Asians. *Molecular Biology and Evolution*. **23**: 2480–2491.

Hill C., Soares P., Mormina M., Macaulay V., Clarke D., Blumbach P.B., Vizuete-Forster M., Forster P., Bulbeck D., Oppenheimer S., and Richards M. (2007) A Mitochondrial stratigraphy for island Southeast Asia. *American Journal of Human Genetics*, **80**: 29–43.

Irwin J., Saunier, J.L., Strouss K.M., Diegoli T.M., Strurk K.A., O'Callaghan J.E., et al. (2008) Mitochondrial control region sequences from a Vietnamese population sample. *International Journal of Legal Medicine*, **122**: 257–259.

Kong Q. P., Yao Y. G., Liu M., Shen S. P., Chen C., Zhu C. L., Palanichamy M. G., and Zhang Y. P. (2003) Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. *Human Genetics*, **113**: 391–405.

Kumar S., Tamura K., and Nei M. (2004) MEGA3: Inte-

grated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinformatics*, **5**: 150–163.

Lertrit P., Poolsuwan S., Thosarat R., Sanpachudayan T., Boonyarit H., Chinpaisal C., and Suktitipat B. (2008) Genetic history of Southeast Asian populations as revealed by ancient and modern human mitochondrial DNA analysis. *American Journal of Physical Anthropology*, **137**: 425–440.

Lewis M.P. (2009) Ethnologue: Languages of the world, Sixteenth edition. Dallas, Texas.

Li H., Cai X., Winograd-Cort E.R., Wen B., Cheng X., and Qin Z. (2007) Mitochondrial DNA diversity and population differentiation in Southern East Asia. *American Journal of Physical Anthropology*, **134**: 481–488.

Matsumura H., Oxenham M.F., Dodo Y., Domett K., Thuy N.K., Cuong N.L., Dung N.K. Huffer D., and Yamagata M. (2008) Morphometric affinity of the late Neolithic human remains from Man Bac, Ninh Binh Province, Vietnam: Key skeletons with which to debate the "Two layer" hypothesis. *Anthropological Science*, **116**: 123–148.

Mayer S., Weiss G., and von Haeseler A. (1999). Pattern of nucleotide substitution and rate of heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics*, **152**: 1103–1110.

Moodley Y., Linz B., Yamaoka Y., Windsor H.M., Breurec B., Wu J.Y., Maady A., Thiberge S., Phuanukoonnon S., Jobb G., Siba P., Graham D.Y., Marshall B.J., and Achtman M. (2009) The Peopling of the Pacific from a bacterial perspective. *Science*, **323**: 527–530.

Oota H., Kitano T., Jin F., Yuasa I., Wang L., Ueda S., Saitou N., and Stoneking M. (2002) Extreme mtDNA homogeneity in continental Asian populations. *American Journal of Physical Anthropology*, **118**: 146–153.

Patcharee L., Poolsuwan S., Thosarat R., Sanpachudayan T., Boonyarit H., Chinpaisal C., Suktitipat B. (2008) Genetic history of Southeast Asian populations as revealed by ancient and modern human mitochondrial DNA analysis. *American Journal of Physical Anthropology*, **137**: 425–440.

Raymond M. and Rousset F. (1995) An exact test for population differentiation. *Evolution*, **49**: 1280–1283.

Soares P., Trejaut J.A., Loo J.H., Hill C., Mormina M., Lee C.L., Chen Y.M., Hujashov G., Peter Forster, Macaulay V., Bulbeck D., Oppenheimer S., Lin M., and Richards M. (2008) Climate Change and postglacial human dispersals in Southeast Asia. *Molecular Biology and Evolution*, **25**: 1209–1218.

Sykes B., Leiboff A., Low-Beer J., Tetzner S., and Richards M. (1995) The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. *American Journal of Human Genetics*, **57**: 1463–1475.

Tajima A., Hayami M., Tokunaga K., Juji T., Matsuo M., Marzuki S., Omoto K., and Horai S. (2004). Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *Journal of Human Genetics*, **49**: 187–193.

Torroni A., Bandelt H.J., Macaulay V., Richards M., Cruciani F., Rengo C., Martinez-Cabrera V., Villems R., Kivisild T., Metspalu E., Parik J., Tolk H.V., Tambets K., Forster P., Karger B., Francalacci P., Rudan P., Janicijevic B., Rickards O., Savontaus M.L., Huoponen K., Laitinen V., Koivumäki S., Sykes B., Hickey E., Novelletto A., Moral P., Sellitto D., Coppa A., Al-Zaheri N., Santachiara-Benerecetti A.S., Semino O., and Scozzari R. (2001) A signal, from human mtDNA, of postglacial recolonization in Europe. *American Journal of Humman Genetics*, **69**: 844–852.

Trejaut J.A., Kivisild T., Loo J.H., Lee C.L., He C.L., Hsu C.J., Li Z.Y., and Lin M. (2005) Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biol.* DOI: 10.1371/journal.pbio. 0030247 (http://biology.plosjournals.org/)

Umetsu K., Tanaka M., Yuasa I., Saitou N., Takeyasu T., Fuku N., Naito E., Ago K., Nakayashiki N., Miyoshi A., Kashimura S., Watanabe G., and Osawa M. (2001) Multiplex amplified product-length polymorphism analysis for rapid detection of human mitochondrial DNA variations. *Electrophoresis*, **22**: 3533–3538.

Umetsu K., Tanaka M., Yuasa I., Adachi N., Miyoshi A., Kashimura S., Park K.S., Wei Y.H., Watanabe G., and Osawa M. (2005) Multiplex amplified product-length polymorphism analysis of 36 mitochondrial single-nucleotide polymorphisms for haplogrouping of East Asian populations. *Electrophoresis*, **26**: 91–98.

Wen B., Li H., Gao S., Mao X., Gao Y., Li F., Zhang F., He Y., Dong Y., Zhang Y., Huang W., Jin J., Xiao C., Lu D., Chakraborty R., Su B., Deka R., Jin L. (2005) Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. *Molecular Biology and Evolution*, **22**: 725–734.

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. *American Journal of Physical Anthropology*, **118**: 63–76.

Yao Y.G., and Zhang T.P. (2002b) Phylogeographic analysis of mtDNA variation in four ethnic populations from Yunnan province: new data and reappraisal. *Journal of Human Genetics*, **47**: 311–318.

Yao Y.G., Kong Q.P., Bandelt H.J., Kivisild T., Zhang Y.P. (2002c) Phylogenetic differentiation of mitochondrial DNA in Han Chinese. *American Journal of Human Genetics*, **70**: 635–651.