J Hum Genet (2001) 46:115–125

ORIGINAL ARTICLE

Goonnapa Fucharoen · Supan Fucharoen · Satoshi Horai

Mitochondrial DNA polymorphisms in Thailand

Received: October 23, 2000 / Accepted: December 4, 2000

Abstract Nucleotide sequences of the D-loop region of human mitochondrial DNA from six small ethnic groups of Thailand i.e., Hill tribes (Lisu and Mussur), Phuthai, Lao Song, Chong, and aboriginal Sakai, were analyzed. The sequences were compared with those of native Thai populations from two provinces, Chiang Mai and Khon Kaen. Based on a comparison of the 563-bp sequences in 215 Thai individuals, 137 different sequence types were observed. Of these, 124 were unique to their respective populations, whereas 13 were shared between two to five populations. The intergenic COII/tRNA^{Lys} 9-bp deletion was observed in every Thai population examined, except for the Sakai, with varying frequencies ranging from 18% to 40%. The D-loop sequences variation, and phylogenetic analysis, suggested that the 9-bp deletion had occurred in a very ancient ancestry of Southeast Asians, although multiple origins of the deletion cannot be ruled out. Genetic distances, based on net nucleotide diversities, between populations revealed that the Sakai were distantly related to the other Thai populations, while the Lao Song and Chong were closely related to each other. Close genetic affinities were also observed among the Hill tribes, Phuthai, and native northeast Thai (Khon Kaen), indicating that they may share some degree of the common ancestral maternal lineages.

Key words Mitochondrial DNA · D-loop region · Sequence polymorphism · Intergenic COII/tRNA^{Lys} 9-base pair deletion · Thai populations · Nucleotide diversity · Phylogenetic tree

G. Fucharoen
Department of Clinical Microscopy, Faculty of Associated Medical Sciences, Khon Kaen University, Khon Kaen, Thailand

S. Fucharoen
Department of Clinical Chemistry, Faculty of Associated Medical Sciences, Khon Kaen University, Khon Kaen, Thailand

S. Horai (✉)
Department of Biosystems Science, The Graduate University for Advanced Studies, Hayama, Kanagawa 240-0193, Japan
Tel +81-468-58-1575; Fax +81-468-58-1544
e-mail: horai@soken.ac.jp

Introduction

In Thailand there are many small ethnic groups inhabiting various geographical regions of the country, who still keep their original cultural heritage and practices. In this study, we focused on some of these ethnic groups to infer genetic relationships among them.

The first group is the Hill tribes or Tribal peoples, who live in the northern part of Thailand among the myriad mountains ranging down from Tibet and China. Two subgroups of tribes; namely, the Lisu and the Mussur, from Chiang Dao Hill, Chiang Mai province, were included. Both the Lisu and the Mussur have their own languages, which are classified in the central division of the Lalo (Yi) branch of the Tibeto-Burman family. Both of the tribes probably originated in Tibet and migrated to the north of Thailand in the nineteenth century (Nawigamune 1992).

The second group is the Phuthai, who live widely spread along the Mekong river basin in the northeastern part of Thailand. They are considered to be the "wealthiest" group, in term of culture and way of life, and have a combination of popular Buddhist and animist beliefs (Lohitkun 1995).

The third group is the Lao Song, who live in the central part of Thailand. Historically, it is believed that the Lao Song and the Phuthai have the same origin. They speak a language that is within the Tai linguistic family, and they have their own civilisation, which originated in the area near the border of China's Kwangsi province and the city of Dien Bien Phu or Muang Thang, Vietnam — the kingdom known as "Sipsongjutai or Sipsongchutai"(Lohitkun 1995; Vallibhotama 1991).

The fourth group is the Chong, who live in the eastern part of Thailand near Cambodia. The Chong were classified as a group of Mon-Khmer people of Austro-Asiatic origin who have their own "Chong language"(Wilai 1995). Little is known about the historical background of this ethnic group.

They usually live in the forest between hills along the eastern part of Thailand.

The last group that we examined is an aboriginal Sakai, or Negrito, group, who live in the south of Thailand. They were classified as Negritos in the Austro-Asiatic group (Brandt 1965). They have short, kinky hair, and dark skin, like Negros, and live in the forest. Anthropologically, it is believed that the Sakai is one of the oldest of the population groups of Thailand, and they moved to the southern part of Thailand from India and Malaysia.

In order to study the genetic background and relationship among these ethnic groups, we examined mtDNA polymorphisms, including D-loop sequence variation and intergenic COII/tRNA^{Lys} 9-bp deletion. These results were analyzed together with those for individuals from the large native Thai populations from Chiang Mai (northern part of Thailand) and Khon Kaen (northeastern part of Thailand).

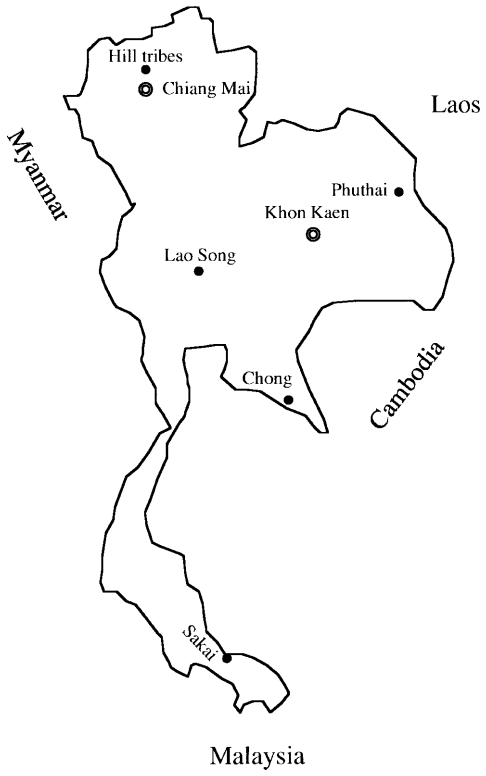

Fig. 1. Geographic distribution of eight Thai populations whose ethnic groups were examined. The Hill tribes in this study consist of two ethnic groups, the Lisu and Mussur

Subjects and methods

Subjects

Blood samples were obtained from 46 individuals from the Hill tribes (25 from the Lisu and 21 from the Mussur peoples) in Chiang Mai province (north), 25 Phuthai in Mukdahan province (northeast), 25 Lao Song in Supanburi province (central), 25 Chong in Chantaburi province (east), and 20 aboriginal Sakai in Trang province (south). Samples were also collected from native Thai people from Chiang Mai (30) and Khon Kaen (44) provinces. All the individuals gave their informed consent prior to their inclusion in this study. Figure 1 depicts the geographical distributions of these ethnic groups in Thailand. Genomic DNA was prepared from peripheral blood leukocytes using standard method with treatment of SDS and proteinase K followed by ethanol precipitation (Fucharoen et al. 1989).

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

A fragment of mtDNA was amplified by the PCR method described by Saiki et al. (1988). A set of two primers (A and E primers; see below) was designed to amplify a DNA fragment from the D-loop region of mtDNA (Horai et al. 1996; Kocher et al. 1989). PCR was carried out under the following thermal cycle conditions: denaturation at 94°C for 15 s, primer annealing at 45°C for 15 s, and extension at 72°C for 30 s, for a total of 30 cycles. The amplified fragments were subjected to direct DNA sequencing using the ABI PRISM 377 Dye Terminator Cycle Sequencing FS Kit and ABI PRISM 377 DNA sequencer (Applied Biosystems, Foster City, CA, USA). The following oligonucleotide primers were designed for amplification and sequencing:
A: 15897-5′-GTATAAACTAATACACCAGTCTTGT-3′-15921;
B: 15985-5′-AGCACCCAAAGCTAAGATTC-3′-16004;
C: 16204-5′-AGCAAGTACAGCAATCAACC-3′-16223;
D: 16413-5′-GCGAGGAGAGTAGCACTCTT-3′-16432;
E: 100-5′-CAGCGTCTCGCAATGCTATCGCGTG-3′-76.
(The notation of Anderson et al. (1981) is used for the numbering of bases.)

Detection of the COII/tRNA^{Lys} intergenic deletion

A set of oligonucleotide primers was designed, as follows, to amplify a COII/tRNA^{Lys} intergenic region of mtDNA: 8211-5′-TCGTCCTAGAATTAATTCCC-3′-8230 and 8310-5′-AGTTAGCTTTACAGTGGGCT-3′-8291. This set of primers amplifies 100-bp and 91-bp fragments of mtDNA with two and one copies of the above 9-bp tandem repeat (CCCCCTCTA), respectively. The amplified fragments were separated by electrophoresis on 4% NuSieve agarose (FMC BioProducts, Rockland, ME, USA) gels and detected fluorographically after staining with ethidium bromide.

Data analyses

The number of nucleotide substitutions per site between individual sequences was estimated using the two-parameter model of nucleotide substitutions (Kimura

1980). On the basis of the estimated number of nucleotide substitutions, phylogenetic trees were constructed using the neighbour-joining (NJ) method (Saitou and Nei 1987) and the unweighted pair-group method with arithmetic mean (UPGMA) (Nei 1987).

Results and discussion

Sequence differences and identities in Thai populations

We determined the nucleotide sequence of a 563-bp fragment of the D-loop region (positions 16048–16569, followed by positions 1–41 in the reference sequence of Anderson et al. 1981) for 215 Southeast Asians from eight Thai populations. There were 137 distinct types of sequences defined by 108 polymorphic sites (Fig. 2). Of these, 124 were unique to their respective populations, whereas 10 were shared between two populations, and 1 each was found in common in three to five populations. Among the 124 unique types, 17 types (14%) were shared by more than two individuals within each population, whereas the remaining 107 types (86%) were each observed in a single individual. The number of shared sequence types and of unique types observed in the eight Thai populations are shown in Table 1. The smaller number of sequence types observed in the Sakai was a result of the many identical sequences in that population. In fact, the most frequent sequence type was shared by 14 Sakai individuals (also see Fig. 3). Of the 10 sequence types shared between two populations (types 10, 22, 78, 79, 84, 92, 110, 112, 131, and 137), 5 (types 78, 79, 92, 110, 137) found in northeast Thai (Khon Kaen) were shared by 2 Phuthai (types 92 and 137), 1 Lisu (type 78), 1 Chong (type 110), and 1 north Thai (Chiang Mai; type 79), respectively. Also, three types found in the Phuthai (types 10, 22, and 84) were shared by north Thai from Chiang Mai (type 10), the Lao Song (type 22), and the Lisu (type 84). The remaining 2 sequence types were found in common in the Lao Song and north Thai (type 112), and in the Lisu and Mussur (type 131). Of the 3 sequence types found in common in more than three populations, type 49 was shared by five populations (northeast Thai, north Thai, Lisu, Mussur, and Phuthai). Type 42 was shared by four populations (northeast Thai, north Thai, Chong, and Mussur), and type 87 was found in common in three populations (northeast Thai, Lisu, and Mussur), respectively. All the four types observed in Sakai were unique among the Thai populations.

As shown in Table 1, the northeast Thai (Khon Kaen) possessed eight sequence types that were shared with other populations. The Phuthai, who also live in the northeastern part of Thailand, exhibited six types that are common with other populations, while for the rest of the populations, the number of shared types was five or less. These data indicate that the northeastern part of Thailand may be the center of human dispersion in the country. For the Mussur group, the number of unique types was only six in the ten types observed in that population, indicating that the Mussur have experienced some extent of gene flow from surrounding populations. By contrast, in the Chong and the Lao Song, 92% of individuals exhibited 13 and 15 unique types, respectively, suggesting that these two ethnic groups are maintaining their own ethnicity, with a low degree of genetic influence from other groups. In particular, among the small ethnic groups in the present study, the Sakai did not possess any types shared with others, suggesting the unique genetic position of this ethnic group.

Table 1. Unique and common sequence types observed in eight Thai populations

Population	No. of individuals	No. of sequence types	No. of unique sequence types (% individuals)[a]	No. of common types	
				Shared between two populations	Shared among move than three populations
Thai Chiang Mai	30	26	21 (73.3%)	3 (t10,t79,t112)[b]	2 (t49,t42)
Thai Khon Kaen	44	43	35 (81.8%)	5 (t78,t79,t92,t110,t137)	3 (t49,t42,t87)
Lisu	25	20	15 (68.0%)	3 (t78,t84,t131)	2 (t49,t87)
Mussur	21	10	6 (42.8%)	1 (t131)	3 (t49,t42,t87)
Chong	25	15	13 (92.0%)	1 (t110)	1 (t42)
Lao Song	25	17	15 (92.0%)	2 (t22,t112)	0
Phuthai	25	21	15 (64.0%)	5 (t10,t22,t84,t92,t137)	1 (t49)
Sakai	20	4	4 (100%)	0	0
Total	215	156	124 (79%)	—	—

[a] Percentages of individuals who possess unique types in each population
[b] Shared types, in parenthesis, are represented by t, followed by the type number

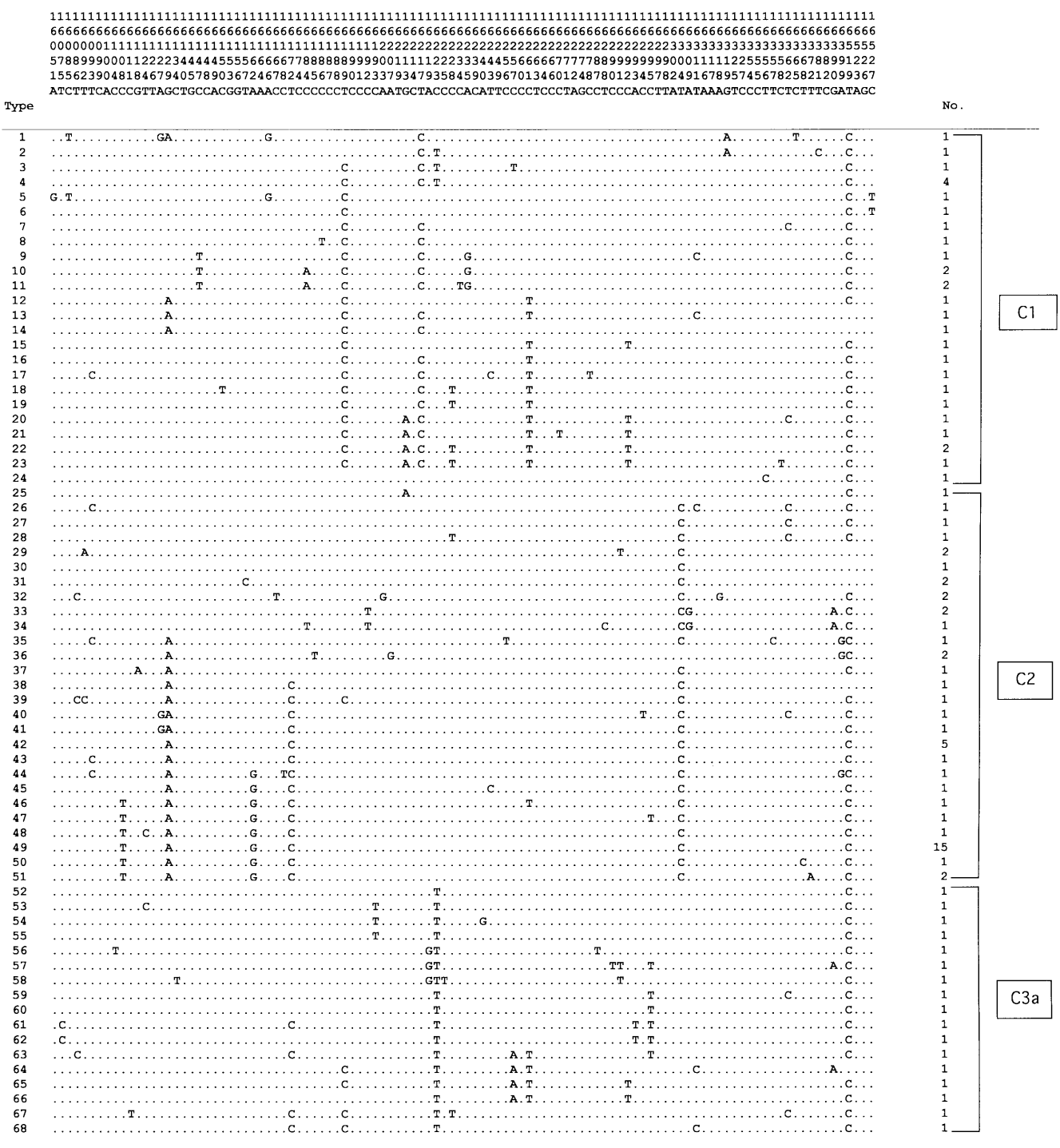

Fig. 2. Nucleotide sequence differences in the D-loop region of mtDNA for 215 Thai people. Presented are nucleotides at 108 polymorphic sites (ps) based on a comparison of 563-bp in the reference sequence of Anderson et al. (1981), *shown below the order for polymorphic sites*. For other sequences, only differences from the reference sequence are indicated. The number of individuals with each sequence type is shown in *Arabic numerals* at *the right of each sequence*. The *boxed letters* indicate the clusters characterized in Fig. 3

Phylogeny of mtDNA sequences

The 563-bp sequences from a total of 215 individuals from Thailand were aligned and compared, and the number of nucleotide substitutions between each pair of sequences was estimated. Nucleotide diversity among the 215 individuals was estimated to be 1.40%, which is slightly higher than the corresponding value estimated for East Asian populations (1.34%; Horai et al. 1996). Using the estimated number of nucleotide substitutions between individual se-

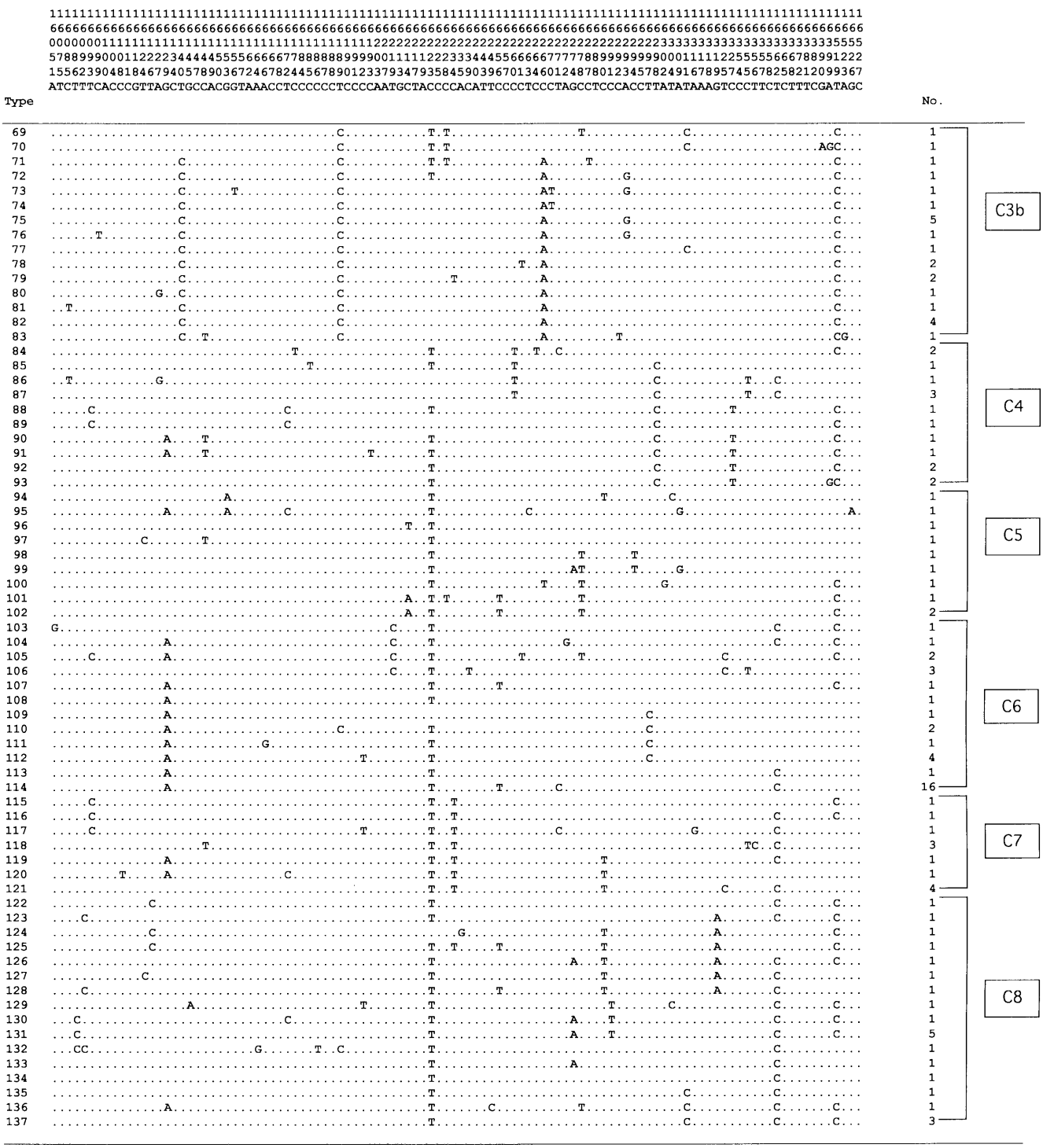

Fig. 2. Continued

quences, a phylogenetic tree was constructed, using the neighbor-joining method (Saitou and Nei 1987), as shown in Fig. 3. On the basis of branching patterns in the tree, sequences were classified into eight monophyletic clusters, shown by brackets with cluster numbers (C1–C8) in Fig. 3. C3 was further divided into C3a and C3b, which were informative for lineages with the intergenic COII/ $tRNA^{Lys}$ 9-bp deletion (discussed later). To ensure the robustness of these clusters, we examined shared polymorphic sites which appeared in more than 40% lineages within each cluster, as shown in Table 2. The majority of members in each cluster exhibited a specific combination of one to four polymorphic sites, with the exceptions of members in C3a and C8, who did not have any specific polymorphic sites. A C-to-T transition at 16223 (nucleotide position in the reference sequence of Anderson et al. 1981)

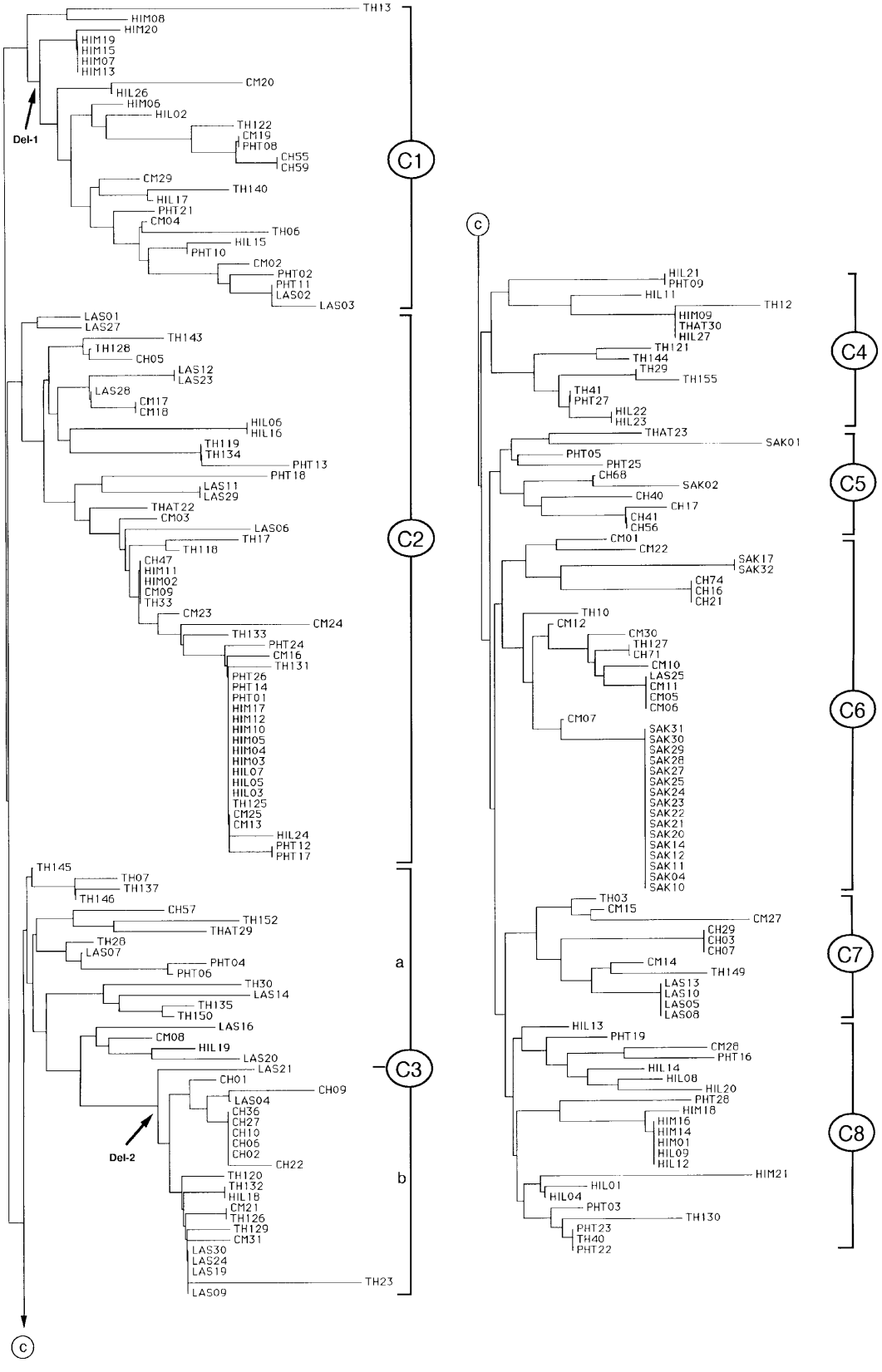

Fig. 3. Phylogenetic tree showing the 215 Thai mtDNA lineages from the eight populations. The phylogenetic tree was constructed by the neighbor-joining (NJ) method (Saitou and Nei, 1987), based on the pairwise number of nucleotide substitutions in the 563-bp D-loop region. The eight distinct clusters in the tree are indicated by *brackets with cluster numbers C1–C8*. The locations of ethnic groups sampled are indicated by the *letters at the tip of each branch*: *TH* and *THAT* for Thai from Khon Kaen, *CM* for Thai from Chiang Mai, *CH* for Chong, *HIL* for Lisu, *HIM* for Mussur, *LAS* for Lao Song, *PHT* for Phuthai, and *SAK* for Sakai. The *two digits after the population names* are the individual sample numbers. All individuals with the 9-bp deletion appear after the branching points *Del-1* and *Del-2*

Table 2. Shared polymorphic sites within clusters

Cluster	No. of lineages	Shared polymorphic sites within a cluster[a]					
C1	**29**	16189:T/C (93%)	**16217:T/C** (86%)	**16261:C/T** (41%)	16519:T/C (93%)		
C2	**52**	**16108:C/T** (40%)	16129:G/A (71%)	**16162:A/G** (44%)	**16172:T/C** (63%)	**16304:T/C** (92%)	16519:T/C (88%)
C3a	**19**	16223:C/T (100%)	16519:T/C (95%)				
C3b	**22**	**16140:T/C** (100%)	16189:T/C (100%)	**16266:C/A** (100%)	16519:T/C (100%)		
C4	**15**	16223:C/T (67%)	**16260:C/T** (47%)	**16298:T/C** (87%)	**16327:C/T** (47%)	16519:T/C (67%)	
C5	**10**	16223:C/T (100%)	**16278:C/T** (60%)	16519:T/C (40%)			
C6	**34**	16129:G/A (88%)	16223:C/T (97%)	**16256:C/T** (50%)	**16271:T/C** (47%)	16362:T/C (56%)	
C7	**12**	16223:C/T (100%)	**16234:C/T** (100%)	**16290:C/T** (50%)	16362:T/C (83%)		
C8	**22**	16223:C/T (95%)	16362:T/C (91%)	16519:T/C (73%)			

Numbers in parentheses represent the percentage of lineages (more than 40%) that exhibited the polymorphism
Polymorphic sites shown in boldface are specific to each cluster
[a] Each polymorphic site corresponds to the site and replacement in the reference sequence reported by Anderson et al. (1981)

Table 3. Distribution of eight ethnic groups in each monophyletic cluster of the phylogenetic tree

Cluster	Specificity	Chiang Mai	Khon Kaen	Lisu	Mussur	Chong	Lao Song	Phuthai	Sakai	Total	π (%)[a]
C1	Mussur-1	17.2	13.8	13.8	**24.1**	6.9	6.9	17.2	0	29	0.93
C2	Khon Kaen-1	17.3	**21.2**	11.5	15.4	3.8	15.4	15.4	0	52	0.81
C3a	Khon Kaen-2	5.3	**52.6**	5.3	0	5.3	21.2	10.5	0	19	1.02
C3b	Chong-1	9.1	22.7	4.5	0	**36.4**	27.3	0	0	22	0.39
C4	Khon Kaen-3	0	**46.7**	33.3	6.7	0	0	13.3	0	15	1.00
C5	Chong-2	0	10.0	0	0	**50.0**	0	20.0	20.0	10	1.00
C6	Sakai-1	26.5	5.9	0	0	11.8	2.9	0	**52.9**	34	0.75
C7	Lao Song-1	25.0	16.6	0	0	25.0	**33.4**	0	0	12	0.78
C8	Lisu-1	4.5	9.1	**36.4**	22.7	0	0	27.3	0	22	0.91
Total										215	1.40

Percentages of individuals in each cluster are shown. Maximum value in the cluster is shown in boldface
[a] π (%) denotes nucleotide diversity within a cluster

was shared by most members in clusters C3a and C4 through C8, while this polymorphism was virtually absent in clusters C1, C2, and C3b. A T-to-C transition at 16519 was an overwhelming polymorphism in the entire population, and this polymorphism occurred in six clusters, excluding C6 and C7. A G-to-A transition at 16129 occurred in the majority of lineages from C2 (71%) and C6 (88%), whereas these two clusters appeared at discrete positions in the phylogenetic tree. Thus, the above three polymorphisms are due to either recurrent mutations or ancient polymorphisms. A T-to-C transition at 16362 was predominant in C6 (56%), C7 (83%), and C8 (91%), whereas this polymorphism occurred at a frequency of less than 7% in the other clusters, except for C4 (26%). Although this polymorphism appeared to have occurred in parallel with mitochondrial lineage divergence, it is informative for late diverging lineages (C6 through C8) in the phylogenetic tree. However, in some clusters, most of the members exhibited a specific polymorphism which was virtually absent in other clusters. For example, a T-to-C transition at 16304 was observed in 48 members (92%) of C2, which was unique to this cluster and was virtually absent in other clusters (only three individuals in the remainder of the clusters). Unique polymorphisms were also seen in at least six other clusters, as shown in Table 2. Therefore, it is likely that most of the clusters reflected their ancestral states of lineage divergence. However, it is apparent that lineages from the eight Thai populations studied were intermingled in the phylogenetic tree, although individuals from single populations dominated in some of the clusters.

To evaluate the features of clustering patterns further, the composition of the eight clusters and the ethnic origins of the 215 Thai lineages were summarized, as shown in Table 3. Although the numbers of individuals sampled from the eight study populations were not the same, we assigned "specificity" for each cluster based on the population from which the maximum percentage of individuals was derived. In this way, we were able to assign specificity for nine clusters/sub-clusters. Although this assignment of specificity seems to be somewhat arbitrary, the rule for assignment is a simple "majority-rules" voting procedure. However, it may be useful for understanding the relationships of mtDNA sequences among very closely related human populations such as Southeast Asians from Thailand, because ancient migrations between populations are anticipated. For example, in C5, the Chong were dominant (50%), with two members each from Phuthai and Sakai and one from northeast Thai (Khon Kaen). The Chong were also dominant in C3b (36%). We therefore assigned these clusters as being specific to the Chong and named them Chong-2 and Chong-1 (Table 3). In this way, we assigned specificity for the nine clusters/sub-clusters. It is interesting to note that native northern Thai individuals (Chiang Mai) were not dominant in any clusters, although

this large northern population may have influenced gene flow toward the south. However, Thai from Khon Kaen exhibited three dominant clusters (C2, C3a, and C4). For other ethnic groups, the Lisu, Mussur, Lao Song, and Sakai were dominant in C8, C1, C7, and C6, respectively. The Phuthai did not exhibit a dominant cluster and were dispersed into other group-specific clusters.

COII/tRNA^{Lys} intergenic 9-bp deletion

The presence of a 9-bp deletion in the COII/tRNA^{Lys} intergenic region of mtDNA is one of the characteristics not only of Asians (Horai and Matsunaga 1986; Horai 1987; Stoneking and Wilson 1989; Horai 1991a, 1991b; Ballinger et al. 1992; Harihara et al. 1992; Passarino et al. 1993; Horai et al. 1996) but also of populations of Asian origin, including Polynesians (Hertzberg et al. 1989; Hagelberg and Clegg 1993; Lum et al. 1994; Redd et al. 1995) and Native Americans (Schurr et al. 1990; Ward et al. 1991; Shields et al. 1992, 1993; Torroni et al. 1992; Horai et al. 1993). However, some studies have suggested multiple origins for the deletion in Asia (Schurr et al. 1990; Ballinger et al. 1992; Torroni et al. 1994; Barrientos et al. 1995; Redd et al. 1995). We screened the present Thai populations from seven localities for the 9-bp deletion. The frequency of the 9-bp deletion was 32% for the Lao Song, 40% for the Chong, 18% for the northeast Thai, 23% for the north Thai, 20% for the Phuthai, and 24% for the Hill tribes (combined data for Lisu and Mussur). However, in the Sakai, the 9-bp deletion was completely absent. The frequency of the deletion in Thai populations ranges from 18% to 40%, except for the Sakai, in which the frequency is similar to that observed in native Taiwanese and Chinese (Horai 1991b). Some randomly selected individuals from the Thai populations were examined for the mtDNA D-loop sequence variation, as mentioned above. A total of 49 Thai people exhibited the 9-bp deletion among the individuals sequenced; 27 were in the C1 cluster and 22 were in the C3 cluster in the phylogenetic tree based on D-loop sequences (Fig. 3). In Fig. 3, in C1, all the individuals with this deletion appear after the branching point "Del-1", and in C3, all individuals with the deletion are derived from lineages after the "Del-2" point (C3b). This suggests that the deletion event may have occurred twice in the ancestry of Southeast Asian lineages. Most lineages after the branching point "Del-1" (Fig. 3) showed a specific combination of two polymorphic sites (16217T/C; 16261C/T), whereas lineages after the point "Del-2" (Fig. 3) exclusively exhibited a different set of polymorphic sites (16140T/C; 16266C/A). However, 16189T/C and 16519T/C were dominant (93%–100%) in both C1 and C3b. Therefore, the possibility of a single event of the 9-bp deletion cannot be excluded. If this were to be the case, ancient lineages with both 16189T/C and 16519T/C could have first experienced the deletion once, and after that, additional new polymorphisms could have occurred in the two lineages; this would have led to the different clusters (C1 and C3b) seen at present. To evaluate the two possibilities (single or multiple events of the deletion), we estimated the coalescence time for "Del-1" and "Del-2" lineages.

Horai et al. (1995) analyzed complete mtDNA sequences from three humans (African, European, and Japanese) and four species of hominoids in order to infer modern human origins. Based on these data for humans, they estimated the substitution rate for the present D-loop region as 8.6×10^{-8} /site per year (Horai et al. 1996). The average nucleotide diversities for the coalescence of the clusters C1 (excluding two individuals without the 9-bp deletion) and C3b (including individuals with the 9-bp deletion) were estimated as 0.846% and 0.387%, respectively. Assuming the average rate of the D-loop sequence substitution of 8.6×10^{-8} /site per year yields mean estimates of a coalescence time of 49,200 years for "Del-1" and 22,500 years for "Del-2". It is interesting to note that East Asians with the 9-bp deletion are derived from "Del-1" lineages, because most of them exhibited the specific polymorphic site 16217T/C together with two common sites (16189T/C and 16519T/C), but no 16261C/T. From the coalescence analysis, it is likely that the 9-bp deletion has occurred independently twice in Southeast Asian ancestry. To confirm the possibility of multiple deletion events, we further examined the net nucleotide diversity (d_A distance) between clusters. First, we estimated the nucleotide diversity within and between the nine clusters/subclusters. Then, from these values, we obtained the net nucleotide diversity between each pair of the nine clusters (see Table 4 caption). A neighbour joining tree, drawn on the basis of d_A distances, is depicted for the nine clusters/subclusters (Fig. 4). Although bootstrap values in the tree were not high, C1 and C3b formed a monophyletic clade with the highest bootstrap value (82%). This result indicates that the 9-bp deletion event appears to have occurred only once in the ancestry of Southeast Asian lineages. Moreover, it is likely that the deletion event occurred a relatively long time ago, because both the C1 and the C3b lineages have accumulated different sets of additional polymorphisms. Therefore, because of the discrepancy between the coalescence and net nucleotide diversity analyses, we are not able to conclude whether the deletion event occurred once or twice in Southeast Asian lineages.

Nucleotide diversity and phylogeny in the Thai population

The intrapopulational nucleotide diversity (d_x or d_y), interpopulational nucleotide diversity (d_{xy}), and net nucleotide diversity (d_A) among eight ethnic groups of the Thai population were calculated, and are shown in Table 4. The value for nucleotide diversity within populations (d_x or d_y) ranged from 1.174% to 1.450%, while that in the Sakai was found to be 0.565%. However, the interpopulational nucleotide diversity (d_{xy}) and net nucleotide diversity (d_A) between the Sakai and other groups were high, indicating that the Sakai have a unique genetic position among the Thai populations studied here. Low d_A values were observed between the Lisu and Mussur (0.008%), between

the Phuthai and the northeast Thai (0.001%), and between the Lisu and Phuthai (0.002%), indicating close genetic relationships between them. On the basis of d_A distances, phylogenetic trees were constructed, using the NJ and UPGMA methods (Fig. 5A,B). Both trees indicated that the Sakai diverged first from the rest of the ethnic groups. The two ethnic groups of Lao Song and Chong showed a close genetic relationship, supported by the high bootstrap values (83% for NJ and 66% for UPGMA). In the NJ tree, the two Hill Tribes (Lisu and Mussur) formed a monophyletic cluster together with the Phuthai, while native Thai from Chiang Mai diverged first. Native Thai from Khon Kaen appeared in the middle part between the Lao Song—Chong and Lisu—Mussur—Phuthai clusters. In the UPGMA tree, five ethnic groups (Thai from Chiang Mai, the Mussur, Lisu, Phuthai, and Thai from Khon Kaen) formed a rather tight cluster (bootstrap value; 86%)

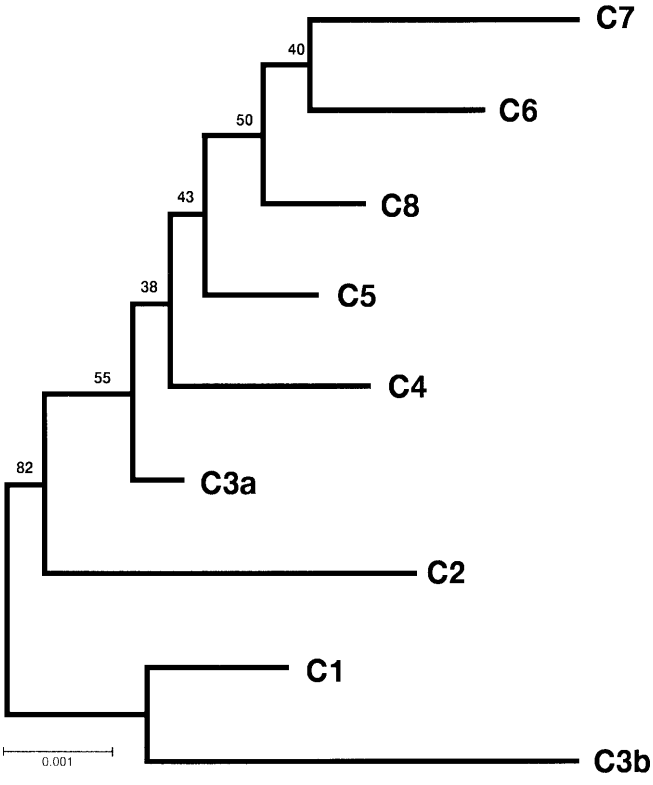

Fig. 4. Neighbour-joining (NJ) tree showing the relationships of clusters/subclusters found in the Thai populations based on net nucleotide diversity (d_A distances). The numbers shown for each interior branch are the bootstrap probabilities (Felsenstein 1986). The d_A distances are indicated on the scale *below the tree*

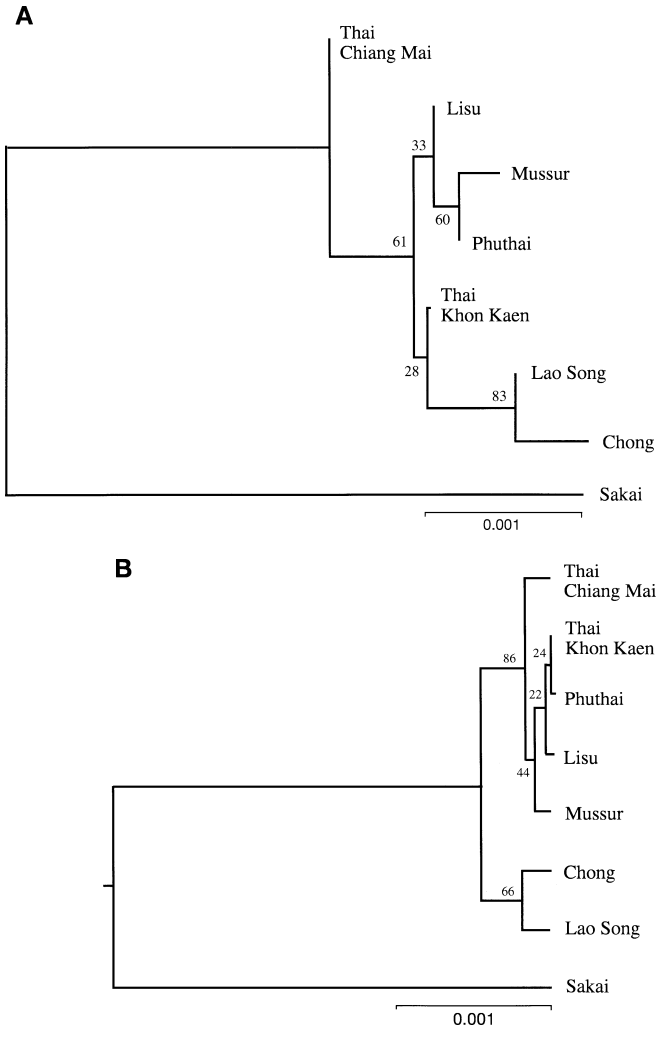

Fig. 5. A Neighbour-joining (NJ) tree and **B** Unweighted pair group method with arithmetic mean (UPGMA) tree showing the relationships of the eight Thai populations based on d_A distances. The numbers shown for each interior branch are the bootstrap probabilities. The d_A distances are indicated on the scale *below the tree*

Table 4. Estimates of interpopulational nucleotide diversity (d_{xy}), intrapopulational nucleotide diversity (d_x or d_y), and net nucleotide diversity (d_A) among eight ethnic groups

Population	Chiang Mai (n = 30)	Khon Kaen (n = 44)	Lisu (n = 25)	Mussur (n = 21)	Chong (n = 25)	Lao Song (n = 25)	Phuthai (n = 25)	Sakai (n = 20)
Chiang Mai	**1.269**	1.343	1.367	1.277	1.458	1.348	1.380	1.376
Khon Kaen	0.019	**1.378**	1.397	1.320	1.482	1.389	1.413	1.529
Lisu	0.038	0.013	**1.389**	1.290	1.492	1.390	1.421	1.471
Mussur	0.055	0.044	0.008	**1.174**	1.449	1.352	1.324	1.476
Chong	0.120	0.090	0.095	0.160	**1.406**	1.393	1.554	1.576
Lao Song	0.058	0.045	0.040	0.110	0.035	**1.311**	1.445	1.542
Phuthai	0.020	0.001	0.002	0.011	0.126	0.064	**1.450**	1.571
Sakai	0.459	0.557	0.494	0.606	0.590	0.604	0.563	**0.565**

Note: All values are multiplied by 100. The bold face numbers on the diagonal refer to (d_x or d_y), and those above the diagonal refer to d_{xy}. The numbers below the diagonal represent the values of $d_A = d_{xy} - [(d_x + d_y)/2]$ (Nei and Miller 1990)

separating from the Chong—Lao Song clade with a bootstrap value of 66%. Although it has been suggested that the Lao Song and Phuthai have a close affinity, and were originally the same population group in the kingdom of Sipsongchutai but migrated to Thailand at different times in the past (Lohitkun 1995; Vallibhotama 1991), the present result was found to conflict with this suggestion. Using β-globin gene haplotype analysis on nuclear DNA, we have also found that the Chong population had different β-globin gene haplotypes from those observed among the Phuthai and Lao Song (Fucharoen et al. 1997). The other interesting finding in this study is that native northeast Thai are closely related to the Hill tribes and Phuthai, whose ancestors probably originated in Tibet and moved to Thailand through China, Myanmar, and Laos.

Acknowledgments This work was supported by the Thailand Research Foundation (S. Fucharoen) and the Ministry of Education, Science, and Culture of Japan (S. Horai). We would like to thank Mr. Phisit Somsak, Mr. Narong Upan, Mr. Kasama Ayukarn, and Associate Professor Phannee Chinoluck and her staff at the faculty of Sciences, Chulalongkorn University, Bangkok, Thailand for their help in collecting specimens used in this study.

References

Anderson S, Bankier AT, Barrell BG, De Bruijn MH, Coulson AR, Drouin J, Eperon IC, Neirlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen KH, Wallace DC (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. Genetics 130:139–152

Barrientos A, Casademont J, Solans A, Moral P, Cardellach F, Urbano-Marquez A, Estivill X, Nunes V (1995) The 9-bp deletion in region V of mitochondrial DNA: evidence of mutation recurrence. Hum Genet 96:225–228

Brandt JH (1965) The Southeast Asian Negritos: further notes on the Negrito of Thailand (part I). Siam Society, Bangkok

Felsenstein J (1986) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Fucharoen G, Fucharoen S, Wilai Y, Chinoluck P, Khunsuk S, Sanchaisuriya K, Sae-ung N (1997) Beta-globin gene haplotypes in some minor ethnic groups in Thailand. Southeast Asian J Trop Med Pub Health (Supp 13):115–119

Fucharoen S, Fucharoen G, Sriroongrueng W, Laosombat V, Jetsrisuparb A, Prasatkaew S, Tanphaichitr VS, Suvatte V, Tuchinda S, Fukumaki Y (1989) Molecular basis of β-thalassemia in Thailand: analysis of β-globin gene mutation using the polymerase chain reaction. Hum Genet 84:41–46

Hagelberg E, Clegg JB (1993) Genetic polymorphisms in prehistoric Pacific islanders determined by analysis of ancient bone DNA. Proc R Soc Lond [Biol] 252:163–170

Harihara S, Hirai M, Suutou Y, Shimizu K, Omoto K (1992) Frequency of a 9-bp deletion in the mitochondrial DNA among Asian populations. Hum Biol 66:161–166

Hertzberg M, Mickleson KNP, Serjeantson SW, Prior JF, Trent RJ (1989) An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians. Am J Hum Genet 44:504–510

Horai S (1987) Evolutionary implications of mitochondrial DNA polymorphism in human populations. In: Vogel F, Sperling K (eds) Human genetics, Proceedings of the 7th International Congress. Springer-Verlag, Heidelberg Berlin New York Tokyo, pp 177–181

Horai S (1991a) Molecular phylogeny and evolution of human mitochondrial DNA. In: Kimura M, Takahata N (eds) New aspects of the genetics in molecular evolution. Japan Scientific Societies Press, Tokyo/Springer-Verlag, Berlin, pp 135–152

Horai S (1991b) A genetic trail of human mitochondrial DNA. In: Mukohata Y (ed) New era of bioenergetics. Academic, Tokyo, pp 273–299

Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci USA 92:532–536

Horai S, Kondo R, N-Hattori Y, Hayashi S, Sonoda S, Tajima K (1993) Peopling of the Americas, founded by four major lineages of mitochondral DNA. Mol Biol Evol 10:23–47

Horai S, Matsunaga E (1986) Mitochondrial DNA polymorphism in Japanese. II. Analysis with restriction enzymes of four or five base pair recognition. Hum Genet 72:105–117

Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, Park KS, Omoto K, Pan IH (1996) mtDNA polymorphism in east Asian populations, with special references to the peopling of Japan. Am J Hum Genet 59:579–590

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:112–120

Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc Natl Acad Sci USA 86:6196–6200

Lohitkun T (1995) Tai in Southeast Asia. Eastern Printing, Bangkok

Lum JK, Rickards O, Ching C, Cann RL (1994) Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. Hum Biol 66:567–590

Nawigamune W (1992) Chiangmai and the Hill tribes. Sangdad Publication, Bangkok

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York, pp 293–298

Nei M, Miller JC (1990) A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data. Genetics 125:873–879

Passarino G, Semino O, Modiano G, Santachiara-Benerecetti AS (1993) CoII/tRNALys intergenic 9-bp deletion and other mtDNA markers clearly reveal that the Tharus (southern Nepal) have Oriental affinities. Am J Hum Genet 53:609–618

Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro ASM, Stoneking M (1995) Evolutionary history of the COII/$tRNA^{Lys}$ intergenic 9-base pair deletion in human mitochondrial DNAs from the Pacific. Mol Biol Evol 12:604–615

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi RG, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Saitou N, Nei M (1987) The neighbor-joining method: a new method for constructing phylogenetic trees. Mol Biol Evol 4:406–425

Schurr TG, Ballinger SW, Gan Y, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, Weiss KM, Wallace DC (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet 46:613–623

Shields GF, Hecker K, Voevoda MI, Reed JK (1992) Absence of the Asian-specific region V mitochondrial marker in Native Beringians. Am J Hum Genet 50:758–765

Shields GF, Schmiechen AM, Frazier BL, Redd A, Voevoda MI, Reed JK, Ward RH (1993) MtDNA sequences suggest a recent evolutionary divergence for Beringian and Northern North American populations. Am J Hum Genet 53:549–562

Stoneking M, Wilson AC (1989) Mitochondrial DNA. In: Hill A, Serjeantson S (eds) The colonization of the Pacific: a genetic trail. Clarendon, Oxford, pp 215–245

Torroni A, Schurr TG, Yang C, Szathmary EJE, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, Wallace DC (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 130:153–162

Torroni A, Chen YS, Semino O, Santachiara-Beneceretti AS, Santachiara-Beneceretti AS, Scott CR, Lott MT, Winter M, Wallace

DC (1994) MtDNA and Y-chromosome polymorphisms in four Native American populations from southern Mexico. Am J Hum Genet 54:303–318

Vallibhotama S (1991) Tai in Sibsongchutai of Southern China (in Thai). In: Vallibhotama S, Wongthes S, Thai Noi, Thai Yai, Thai Siam. Special edition of the Silpa — Wattanatham Magazine, Matichon, Pikanase Printing Center, Bangkok, pp 6–16

Ward RH, Frazier BL, Dew-Jager K, Pääbo S (1991) Extensive mitochondrial diversity within a single Amerindian tribe. Proc Natl Acad Sci USA 88:8720–8724

Wilai Y (1995) Characteristics of beta E-globin gene haplotypes in Sakai and Chong tribes. M.Sc. thesis, Chulalongkorn University, Bangkok