04 ธ.ค. 2548

รายงานโครงการวิจัย

(1 กรกฎาคม 2545 – 30 ธันวาคม 2547)

โครงการ: การวิเคราะห์ลำดับนิวคลิโอไทด์จากไมโตคอนเดรียของโครงกระดูกโบราณในแหล่งโบราณคดีที่สูงในป่างมะผ้า จังหวัดแม่ฮ่องสอน

ผศ.ภก.ดร.ฉัตชัย ฉิ่นไพศาล ภาควิชาเภสัชวิทยาและพิษวิทยา คณะเภสัชศาสตร์ มหาวิทยาลัยศิลปากร
ผศ. ดร. รัศมี ชูทรงเดช ภาควิชาโบราณคดี คณะโบราณคดี มหาวิทยาลัยศิลปากร กรุงเทพฯ
รศ.พญ.ดร.พัชรีย์ วิชยานุวัติ ภาควิชาชีวเคมี คณะแพทยศาสตร์ ศิริราชพยาบาล มหาวิทยาลัยมหิดล

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย

Analyses of Mitochondrial DNA HV-I hypervariable segments of ancient teeth specimens from an archeological burial site in Pang Ma Pha, Mae Hong Son Province, Thailand

Executive summary

การทำการสกัด DNA จากฟันที่พบจากการขุดค้นถ้ำในโครงการโบราณคดีที่สูง ปางมะผ้า แม่ฮ่องสอน จำนวนทั้งสิ้น 52 ตัวอย่าง ที่เก็บได้จากการสำรวจแหล่งโบราณคดี ถ้ำยาป่าแหน 2 ในช่วงระหว่าง พ.ศ. 2541-2544 พบว่าสามารถทำการสกัด และเพิ่มจำนวน DNA ในส่วน hypervariable region I ของ mitochondrial DNA และหาลำดับนิวคลิโอไทด์ได้เป็นจำนวน 28 ตัวอย่าง ซึ่งในจำนวนทั้งหมด 28 ตัวอย่างนี้ สามารถทำการหาลำดับนิวคลิโอไทด์ ที่ตำแหน่ง bp 16209-16390 ได้ ซึ่งยังไม่ครอบคลุม HV-I ทั้งบริเวณและ ในจำนวนนี้ สามารถหาลำดับนิวคลิโอไทด์เพิ่มเติมจากตำแหน่ง bp 16111 ถึง 16208 ได้อีกเป็นจำนวน 11 ตัวอย่าง ซึ่งยังไม่ครบทั้ง 28 ตัวอย่าง

เมื่อเปรียบเทียบจำเพาะแต่ในบริเวณตำแหน่ง bp 16209-16390 ของ DNA fragment ที่ได้จาก ancient teeth พบลักษณะของ DNA sequence ที่แตกต่างกันเป็นจำนวนทั้งหมด 7 ชนิด โดยกลุ่มที่ใหญ่ที่สุดมีตัวอย่างจำนวน 18 ตัวอย่างที่มีลำดับนิวคลิโอไทด์เหมือนกัน (ซึ่งผู้วิจัยยังคงตั้งคำถามกับกลุ่มตัวอย่างนี้ และจะได้ดำเนินการ verify ความถูกต้องต่อไป) และมี 2 กลุ่มๆละจำนวน 3 ตัวอย่าง ที่มีนิวคลิโอไทด์เหมือนกัน และที่เหลืออีก 4 ตัวอย่างนั้นมีลำดับนิวคลิโอไทด์ที่แตกต่างกัน พบว่าในช่วงนี้ มี polymorphic sites อย่างน้อย 13 ตำแหน่ง แต่เมื่อพิจารณาตั้งแต่ bp 16111 ถึง 16390 จะพบว่ามีตำแหน่งที่เกิด polymorphism ต่างๆ ทั้งหมดอย่างน้อยที่สุด 24 ตำแหน่ง การวิเคราะห์ด้วยการทำ phylogenetic tree ระหว่าง ancient DNA ด้วยกันพบว่า สามารถแบ่งได้เป็น 2 กลุ่มใหญ่ๆด้วยกัน กลุ่มแรก (C1) มี 23 ตัวอย่าง ซึ่งรวมถึง samples ที่มี sequence เหมือนกัน จำนวน 18 ตัวอย่างนั้นด้วย ส่วนกลุ่มที่ 2 (C2) นั้นมี 5 ตัวอย่างด้วยกัน

ความตั้งใจของผู้วิจัยนั้นต้องการที่จะวิเคราะห์ลำดับ nucleotide ตั้งแต่ 16111-16390 เพื่อหาว่า ancient DNA เหล่านี้มีความใกล้ชิดกับชนกลุ่มใดในปัจจุบันมากที่สุด แต่เนื่องจาก sequenceที่ได้นั้นยังไม่ครบ 28 samples การวิเคราะห์ลำดับ nucleotide ที่ได้จึงใช้ช่วงความยาว 182 bp (ตั้งแต่ bp 16209-16390) เปรียบเทียบกับ mtDNA sequences ในช่วงเดียวกัน โดยเปรียบเทียบ 2 กลุ่ม กลุ่มเปรียบเทียบแรก ใช้ sequence จาก กลุ่มชาวมลาบรี (ตองเหลือง)จากน่าน 19 คนและชาวละว้า หรือลัวะจากแม่ฮ่องสอน 18 คน ซึ่งพูดภาษาในตระกูลออสโตรเอเชียติค กลุ่มมอญ-เขมร ซึ่งประชากรกลุ่มนี้เป็นกลุ่มที่เชื่อว่าเป็นประชากรที่ตั้งถิ่นฐานอยู่ในบริเวณภาคเหนือของประเทศไทยมาตั้งแต่สมัยก่อนประวัติศาสตร์ และ ชาวไทยเหนือ (ไทยญวน)139 คน กลุ่มเปรียบเทียบที่สองเป็น sequence จากชาวเขาเผ่ามูเซอ 21 คน ลีซอ 25 คน ชาวไทยเชียงใหม่ 30 คน และ ชาวขอนแก่น 50 คน ผลการวิเคราะห์ในกลุ่มประชากรแรกนั้น พบว่า cluster C1 ของ ancient DNA นั้น ส่วนใหญ่สามารถจัดอยู่รวมกับลัวะ และ ชาวไทยญวณ (NThai) ที่นำมาเปรียบเทียบ มีเพียงตัวอย่างเดียวที่มี sequence เหมือนมลาบรี ส่วนการเปรียบเทียบในกลุ่มที่ 2 นั้นพบว่าทั้ง cluster C2 นั้น ค่อนข้างมีความสัมพันธ์ใกล้ชิดกับชาวเขาทั้งในส่วน ลีซอ และมูเซอ มากกว่า cluster ที่ 1 ซึ่งพบแทรกอยู่ทั่วไปในกลุ่ม เชียงใหม่ มูเซอ และลีซอ รวมถึง ชาวขอนแก่น

ผลการศึกษาสรุปได้ว่า ตัวอย่างฟันที่ได้จากซากที่ค้นพบจากการสำรวจถ้ำในแหล่งโบราณคดีที่สูง อำเภอปางมะผ้า จังหวัดแม่ฮ่องสอนนั้น อาจมีส่วนสัมพันธ์กับประชากรที่อาศัยอยู่ในบริเวณภาคเหนือของไทย โดยเฉพาะ ชนกลุ่มน้อย และชาวเขาบางกลุ่ม สิ่งที่น่าสนใจที่ควรจะทำการศึกษาต่อไป คือการเปรียบเทียบกับ ชนชาติต่างๆที่มีอยู่ทางเหนือ กลุ่มชนโบราณ ในช่วงอายุใกล้เคียงกัน รวมถึงการนำตัวอย่างของประชากรจีนในส่วนต่างๆของประเทศ มาเปรียบเทียบ แต่ด้วยจำนวน nucleotides ที่หาได้นั้นค่อนข้างมีจำกัด และเป็นส่วนท้ายของ hypervariable region –I ซึ่งพบ polymorphism น้อยกว่าส่วนต้น ทำให้ power ของการวิเคราะห์นั้นยังไม่สูงมากนัก อาจต้องมีการหาลำดับ nucleotide เพิ่มเติมต่อไป ซึ่งผู้วิจัยจักดำเนินการต่อไปจนกว่าจะได้ข้อมูลเพิ่มเติมซึ่งทำให้การวิเคราะห์มีประสิทธิภาพมากที่สุด

Abstract

การสกัด DNA จากฟันของโครงกระดูกคนในโครงการโบราณคดีที่สูง ในอำเภอปางมะผ้า จังหวัดแม่ฮ่องสอน จำนวน 52 ตัวอย่าง จากถ้ำยาป่าแหน 2 ในช่วงระหว่าง ปี 2541-2544 ซึ่งพบว่ามีอายุเก่าแก่ถึง กว่า 2000 ปีมาแล้ว สามารถทำการสกัด และเพิ่มจำนวน DNA ในส่วนของไมโตคอนเดรีย จากลำดับนิวคลิโอไทด์ ตำแหน่งที่ 16209-16390 (182 bp) เป็นจำนวน 28 ตัวอย่าง อย่างไรก็ดี ลำดับของนิวคลิโอไทด์ที่ได้นั้นเป็นส่วนท้ายๆของ hypervariable region-I (HV-I) ซึ่งอาจมี polymorphism ที่น้อยกว่าส่วนต้น จึงได้ทำการเพิ่มจำนวน DNA ในส่วนต้นของ HV-I ต่อ และใน 28 ตัวอย่างนี้ สามารถเพิ่มจำนวน DNA ไปทาง 5' จนถึงตำแหน่ง 16111 ได้อีก 12 ตัวอย่าง เมื่อนำลำดับนิวคลิโอไทด์ในบริเวณนี้ (16209-6390) มาเปรียบเทียบกับ ชนกลุ่มน้อยบางกลุ่ม คือ มลาบรี ลัวะ ซึ่งพูดภาษาในตระกูลออสโตรเอเชียติค กลุ่มมอญ-เขมร ที่เชื่อกันว่าตั้งถิ่นฐานอยู่ในบริเวณภาคเหนือของประเทศไทยมาตั้งแต่สมัยก่อนประวัติศาสตร์ และกับชาวไทยญวณทางเหนือ พบว่า DNA ที่ได้จากคนโบราณนี้สามารถแบ่งได้เป็น 2 กลุ่มใหญ่ คือ C1 ซึ่งพบว่ามีความใกล้ชิดกับลัวะ และชาวไทยญวณ ส่วน C2 ในการวิเคราะห์เปรียบเทียบกับ ชาวเขาเผ่า ลีซอ มูเซอ ชาวเชียงใหม่ และชาวขอนแก่น พบว่า C2 นั้น ค่อนข้างอยู่รวมตัวกันนิวคลิโอไทด์จากชาวเขาเผ่า ลีซอ มูเซอ ในขณะที่ C1 พบกระจายในคนเชียงใหม่ และ ขอนแก่นบ้าง อย่างไรก็ตาม อาจต้องมีการทำการวิเคราะห์เพิ่มเติมโดยใช้ลำดับนิวคลิโอไทด์ที่ยาวขึ้นของส่วนนี้ และเปรียบเทียบกับตัวอย่างจากคนโบราณในสมัยเดียวกันที่พบในภูมิภาคนี้ และรวมถึงชาวจีนที่สันนิษฐานว่าอาจมีส่วนในการเป็นบรรพบุรุษของชาวไทย

Extraction of mtDNA from human teeth from the remains (~2000 years before the present) from the Ya Pa Nae 2 cave in Ma Hong Song province of Thailand was performed. Amplification and sequencing of 28 out of 52 samples was successful, yielding only 182-bp fragment from position 16209-16390 of the hypervariable segment-I (HV-I) of mitochondria. The sequences were compared with those of modern northern ethnic groups as well as sequences from Thai Chaing Mai, and Khon Kaen. 7 sequence types were observed among 28 teeth DNA, and two clusters were observed by phylogenetic analysis (C1, N=23 and C2, N=5). Two phylogenetic trees were constructed to determine genetic affinity to those of modern Thai. C1 were founded to be closely related to Lawa and Mlabri ethnic groups and C2 to the northern Thai (NThai). Second tree comparing ancient Thai with Lisu, Mussur, Thai Chiang Mai, and Khon Kaen revealed that C2 were also close to the two Hill tribes, whereas C1 were scattered within Chiang Mai as well as some modern Khon Kaen. However, sequence information within this latter part of HV-I is known to have less polymorphism. It remains to be determined if these remains are related to other ancient remains or other modern subpopulations thought to be ancestors of Thai.

เนื้อหางานวิจัย
ในที่นี้จะขอรายงานการวิจัยเป็นภาษาอังกฤษ เพื่อเตรียมการสำหรับ ส่ง **manuscript** ตีพิมพ์ หลังจากการแก้ไขการทดลอง ตามที่ผู้ทรงคุณวุฒิเห็นสมควร ดังนั้นส่วน **appendix** ท้ายตามที่สกว.กำหนดจึงเน่ส่วนเดียวกับ เนื้อหางานวิจัย

Introduction

The genetic relationship between human populations of the Thai and their ancestors has not been the subject of any study to date. Besides discovery of many promising archeological sites where artifacts as well as skeletal remains of people are found, most of the works from these areas are mainly centered on anthropological, cultural and historical, and archeological studies only. In addition, hypotheses of early settling of Thai people in Thailand at present have long been a subject of much debate among anthropologists around the country. There are two main hypotheses regarding where the Thai ancestors may have come from. The "**Immigrant hypothesis**" acknowledges that Thai ancestors, Tai, migrated from China, moved southwards, conquered the native populations on the southern route to reach the Indo-China region. The "**endogenous hypothesis**", in contrast, recognizes that Thai ancestors were not Tai from China, but the native people of the Suwannapoom, such as Lawa, Mon, and Khmer, living in scatters around the country.

With the advent of genetic and recombinant DNA technology we are now beginning to see more research focusing on the origin of people and relationship among modern sub-populations within a specific region. However, a literature survey of such studies was unable to find a published, peer-reviewed study related to genetic relationship among modern and ancient Thai population. We, therefore, attempt to determine the genetic background of the ancient Thai by extracting DNA from approximately 2000-year old human teeth from the expedition set forth during 1998-2001 of the Pang Ma Pha highland archeology project[1] and determine the nucleotide sequences of their mitochondrial (mt) D-loop regions. Exploration of the caves and rock shelters in this area also reveals the ancient burial arrangement where various styles of the log coffins were used in funeral services. Artifacts and carbon dating of samples in various Pang Ma Pha sites were determined and found to range from 22,190 to 2200 year before present time (ybp), indicating a wide range of prehistoric to historic inhabitation.

The main purposes of this study are to create the hypervariable-I (HV-I) sequence information from mtDNA sequence polymorphism obtained from these ancient remains and compared with modern Thai as well as some of the local ethnic groups who lives in the northern part of Thailand. The hypervariable regions of mtDNA have been utilized to provide population diversity as well as migration patterns of ancient people by virtue of its abundance in the cells, and thus the remains, as opposed to chromosomal DNA. They also represent a powerful tool to reconstruct phylogenetic trees for related populations since mtDNA is passed on to offspring only through maternal inheritance. In addition, polymorphism of the D-loop or hypervariable region and restriction fragment length polymorphisms (RFLPs) of mtDNA have long been recognized and accepted as a tool for establishing ethnicity and lineages as well as to reconstruct the genetic, geographic history, and

migratory patterns among subpopulations of people in the past. We evaluate the nucleotide diversity among ancient Asian from Pang Ma Pha and also examine phylogenetic relationships among these ancient Asian and modern populations by constructing mtDNA trees to understand the genetic relationship of the people living in this area. As question remains of whose culture of log coffins in this highland archaeology belongs to, the integration of such information into archaeological and historical studies will hopefully enable a useful link as to who these remains were.

We now show here for the first time of the successful extraction and amplification of mtDNA from ancient remains in Thailand. Extraction and amplification of 52 teeth samples, most of which are from molar 2 or 3, results in 28 amplifiable DNA within part of the HV-I of the mtDNA.

Materials and Methods

Samples

Teeth specimens from human remains come from the exploratory expedition site in Ya Pa Nae 2 cave in Pang Ma Pha district of Mae Hong Son during 1998-2001. This site has been dated to the period of late Holocene (~2200 ybp). For modern mtDNA sequences, Lawa, Mlabri (Tong Lueng), and northern Thai sequences are provided by Assoc. Prof. Daorung Kangwanpong from the Department of Biology, Faculty of Science, Chiang Mai University. Lisu and Mussur, Thai Chiang Mai and Khon Kaen mtDNA sequences were retrieved from a literature search[2].

Extraction and purification of ancient DNA from old teeth specimens

A modified protocol using a combination of conventional proteinase K digestion followed by incubation with chaotropic agent, guanidinium isothiocyanate (GUSCN), according to Boom et al[3].was used for DNA extraction. Briefly, the surface of teeth specimens was decontaminated by scouring with a sterile, disposable lancet followed by dusting the surface with a hand-drill until clean. Samples were then rinsed thoroughly with sterile water. Subsequently, the samples were soaked in 10% sodium hypochlorite for 20-30 min., rinsed, and then submerge in 30% H_2O_2 for 20-30 min. After cleaning, air dry, and exposed to UV light for 20 min, teeth samples were pulverized by crushing in a sterile, autoclavable, tailor-made mortar and pestle. A portion of clean, pulverized teeth powder in a sealed microcentrifuge tube was immersed in liquid nitrogen and quickly thawed for at least three times. 0.5 mL of proteinase K digestion buffer (10 mM Tris-HCl, 2 mM EDTA, 10 mg/mL DTT, 0.5 mg/mL proteinase K, 0.1% SDS) was then added to the powder, and overnight incubation was performed at 56°C with gentle shaking. The next morning the sample was centrifuged at 3000 g and the supernatant was saved, purified by phenol/chloroform extraction followed by ethanol precipitation. The cell pellet was later incubated with 0.5 mL of lysis buffer containing 10 ml of 0.1 M Tris HCl pH 6.4, 12 g of guanidinium thiocyanate, 2.2 ml of 0.02 M EDTA and 0.5 ml of Triton X-100 at 65°C with gentle agitation for 48 hours. After spinning and collection of the supernatant, incubate the pellet with 0.5 mL of extraction buffer (2.5 M potassium phosphate pH 6.4) at 65°C for another 15-30 minutes. Combine all supernates and add 40

μL of 100% w/v silicon dioxide solution, incubate with gentle mixing in the dark for 15 minutes. Wash silica pellet twice with 70% EtOH and once with acetone. Air-dry the pellet and elute DNA twice using 60°C TE.

mtDNA PCR amplification and sequencing of ancient DNA

Parts of the hypervariable region I, i.e., 5' upstream of MT1, MT1 and MT4 according to Oota et al[4] were amplified in three overlapping fragments, subcloned into a TA vector system (Promega) and individually sequenced for at least three clones to confirm the nucleotide sequences (Fig. 1). The condition for PCR was 95°C 1 min., 55°C 1 min., and 72°C 1 min. for 35 cycles with hot start. However, second round PCR were necessary in order to see a distinct band. Both DNA extract from proteinase K digestion and GUSCN/SiO_2 extract of the same sample were used in PCR reactions. Often, successful amplification was only from GUSCN/SiO_2 extract. MT4 forward primer for the first PCR was CC3; F16190 [5' CCC CAT GCT TAC AAG CAA G 3'], and reverse primer was CC4; R16422 [5' ATT GAT TTC ACG GAG GAT G 3'] (numbering is as in Cambridge Reference Sequence (CRS)). Primers pair for second, nested forward PCR primer was CC3, and reverse primer was CC2; R16410 [5' GAG GAT GGT GGT CAA GGG AC 3']. MT1 and 5' upstream sequence fragment was amplified into 2 small fragments. Primer pairs for the first PCR were CC1: F15971 [5' TTA ACT CCA CCA TTA GCA CC 3'] and CC6: R16258 [5' TGG CTT TGG AGT TGC A 3'] and the primer pairs for the second PCR were: CC1 with CC7; R16126 [5' ACA ATA TTC ATG GTG G 3'] and CC5; F16095 [5' CGT ACA TTA CTG CCA G 3'] with CC6. Sequencing reactions were performed using a kit (Bid Dye) for automated sequencer.

Nucleotide diversity, sequence and phylogenetic analysis

Two set of comparisons were made to construct phylogenetic trees of the HV-I sequences. First set of analysis was done by comparing modern DNA sequences of interest groups kindly provided by Dr. Daorung. Among these subpopulations were the 19 Mlabri (Tong Leong) from Nan province, 18 Lawa from Ma Hong Son province, and 139 northern Thai. Pairwise difference distance method was performed and phylogenetic tree was constructed by Neighbor-joining analysis[5]. Second comparison of mtDNA sequences was performed in a total of 126 mtDNA sequences including 2 hill tribes (21 Mussur and 25 Lisu), 30 Thai Chiang Mai, and 50 Thai Khon Kaen[2]. Phylogenetic network of mtDNA analyses were constructed by means of parsimony using PAUP 4.0 program. We also calculate nucleotide diversity of the ancient mtDNA sequence using DNASP program.

Results

Amplification, sequencing, and polymorphism of ancient mtDNA

After second PCR amplification 28 from the total of 52 teeth samples gave a specific band encompassing a 223 (16190 to 16410) bp of MT4 of the hypervariable region I. In order to avoid ambiguity frequently encountered from direct PCR sequencing the PCR fragments were then sub-cloned into a TA vector and selected for sequencing. A minimum number of three clones was used to verify the concordance of the

sequences. Further amplification to obtain 5' upstream sequence in MT1 region of all samples was also attempted. However, at the time of this report we were able to extend the nucleotide sequence for only some of the specimens (Fig. 3).

The nucleotide sequences show that the nucleotides are of human origin when compared to standard human mtDNA reference sequence (CRS)[6]. Since the 5' upstream sequence and MT1 has not been determined in all samples. We then evaluate the polymorphisms using the data only from nucleotide 16209 to 16390. There are at least 7 haplotypes of sequences with at least 13 polymorphic sites in the region determined (Fig. 3). If all the upstream sequences are considered, there are at least 24 polymorphic sites. Group 1 is observed the most frequent (18 samples) whereas 4 other sequence types (2, 3, 4, and 5) only have one member each. Group 6 and 7 are observed in 3 samples each.

The largest group 1 is characterized by T/C transition at position 16224 and 16304, and C/T transition at position 16242. Group 2 is characterized by several (5) C/T transitions at position 16223, 16261, 16264, 16321, 16327, and one T/C transition at 16297. Group 3 is characterized by C/T transition at 16223, C/A transversion at position 16236, and T/C transition at position 16362. Group 4 is characterized by a C/A transversion at position 16266. Group 5 is characterized by C/T transition at position 16256, and T/C transition at position 16304. Group 6 is characterized by a C/T transition at position 16223, and a T/C transition at position 16297. The last group is characterized by having only a T/C transition at position 16304 only.

Nucleotide diversity, sequence and phylogenetic analysis

In order to determine the nucleotide diversity of the sequences the DNASP program is used in analysis of nucleotide 16209-16390 of all 28 positive samples. The average number of differences is 2.431 and the within population nucleotide diversity, Pi, is 0.01336 (Fig. 4) Figure 5 shows the phylogenetic tree for the 28 ancient mtDNA, 182 bp of HV-I sequences using Neighbor-Joining (NJ)/ UPGMA method version 3.6a2.1 by program Bioedit version 7.0.4.1 available through the internet (URL: http://www.mbio.ncsu.edu/BioEdit/bioedit.html). This NJ tree reveals that the 28 mtDNA lineages could be classified into 2 clusters (C1 and C2). C1 clade share the T/C polymorphism at nucleotide position 16304 from group 1 (N=18), 4 (N=1), 5 (N=1), and group 7 (N=3). The C2 clade members consist of 5 individuals from group 2 (N=1), 3 (N=1), and group 6 (N=3), in which they all share the C/T transition at position 16223. The high number of halplotypes from 28 individuals is noted as compared to 4 haplotypes from 18 Lawa and only 1 type from 19 Mlabri individuals, whereas the northern Thai mtDNA sequences from 139 individuals reveal 50 haplotypes (Fig. 6).

To infer phylogenetic relationships among the 7 ancient Thai groups and other modern populations, using pairwise difference distance method we constructed a Neighbor-joining (NJ) tree for these 4 populations (Figure 7). Interestingly, analysis of the tree also indicates that the 7 haplotypes for ancient Thai appears to fall into 2 main clusters, C1 and C2. The C1 clade are found in 23 of ancient DNA (group 1, 4, 5, 7), Mlabri, Lawa and 36% of northern Thai whereas C2 clade are clustered among the remaining (64%) northern Thai

and 5 ancient Thai (group 2, 3, 6). The shared site of all 4 haplotypes from Lawa and group 1, 5, and 7 is T/C transition at position 16304, whereas the shared polymorphic site between group 4 (T58 in fig. 7) is C/A transversion at position 16266. The observation of ancient Thai sequences in C1 suggests close genetic affinities between ancient Thai and Mlabri and Lawa whose inhabition are believed to be originated in the northern Thailand with dominant use of Austro-Asiatic language. Interestingly, in C2 clade sequences in the analytical area from group 6 are exactly the same as Nthai8 haplotype.

In order to determine genetic relationships among other modern Thai in different part of the country a second set of phylogenetic tree was constructed by PAUP program. In this analysis a total of 126 individuals including 2 hill tribes (21 Mussur and 25 Lisu)[2], 30 Thai Chiang Mai, and 50 Thai Khon Kaen (altogether 100 haplotypes) as well as 28 samples from ancient teeth were included into the construction of an NJ tree. As shown in figure 8 and figure 9 (unrooted tree) the genetic differences based on these 182 bp mtDNA suggest that the ancient sequences show two clusters with the exception, however, of group 3 (T24) which, interestingly, shows affinity to both Lisu and Mussur ethnicity classified as cluster C8 in Fuchareon et al[2]. Within the C2 cluster (group 2, 3, 6) in this study sequences from group 6 (T180, 173, 138) are also in close affinity to 2 Chiang Mai haplotypes as well as some other Khon Kaen DNA. Nonetheless, sample in group 2 (T3) is uniquely close to 2 Lisu individuals in C4 cluster of Fuchareon et al[2]. Of note the tree also shows that C2 cluster in this study is highly linked to the Chiang Mai subpopulation, whereas C1 is more scattered among both Chiang Mai and Khon Kaen subpopulations.

Conclusion

This study demonstrates for the first time the successful extraction and amplification of ancient DNA from human skeletal remains in Thailand. In this initial report we extract DNA from teeth samples from remains within a cave in a highland archeology site which is demonstrated to be at least as old as 2200 ybp. We use nucleotide sequences of MT1 and MT4 as the longest sequence possible to construct the phylogenetic tree to examine genetic relationships with modern Thai subpopulations and some ethnic groups within the northern Thailand. We first ask whether the ancient Thai would have a genetic affinity to some of the ethnic groups inhabiting the northern Thailand. However, when Lawa and Mlabri are compared the tree shows two clusters of mtDNA with the first cluster being more closely related to Lawa and Mlabri than the other. These 2000 year old sequences in the other analysis also form two clusters with two selected northern hill tribes, and some with the a specific group of Khon Kaen.

There are diverse theories with respect to the legends of how Thai ancestors might have come from. These include: immigrant theory i) Central China centered in Sichuan province and along the Yanxi River gradually migrated southwards and finally into the Indo-China region, ii) the arrival of Tai from Altai mountain areas to the Sichuan province, iii) Tai has long inhabited very wide geographic distribution in the southeast and south of China, north of Indo-China as well as part of the northern India, and the endogenous hypotheses iv) Thais

originally stay in the Kingdom of Siam long ago, and v)suggestion that Thai were native, indigenous in the northeast of Thailand, later migrated to the west of Indo-China as well as south China [ref]. It has also been shown that native northeast Thai are closely related to the Hill tribes and Phuthai[2]. In our study we also found that ancient Thai mtDNAs show close genetic relationship with Lisu and Mussur while some, however, are also scattered through Khon Kaen population (C1). This may reflect the fact that the 3' sequences of HV-I are less hypervariable than the 5' sequences in which we have yet to continue to retrieve. It has been speculated that Mlabri and Lawa whom spoken languages are related to Austro-Asiatic language are in the same group of Mon-Khmer which inhabit the northern Thai since prehistoric time. It would be of interest to compare the longer HV-I sequences as well as HV-II with those of modern and ancient selected Chinese groups that are thought to be the ancestor Thai to further evaluate the origin of Thai.

References

1. Shoocongdej, R. *Overview of social and cultural development of Highland Archeology in Pang Ma Pha, Mae Hong Song province.* In *Man, culture, and paleoenvironment of the Highland Archeology in Pang Ma Pha, Mae Hong Son province.* 2003. Silpakorn University, Tha Pra Campus.
2. Fucharoen, G., S. Fucharoen, and S. Horai, *Mitochondrial DNA polymorphisms in Thailand.* J Hum Genet, 2001. **46**(3): p. 115-25.
3. Boom, R., et al., *Rapid and simple method for purification of nucleic acids.* J Clin Microbiol, 1990. **28**(3): p. 495-503.
4. Oota, H., et al., *Molecular genetic analysis of remains of a 2,000-year-old human population in China- and its relevance for the origin of the modern Japanese population.* Am J Hum Genet, 1999. **64**(1): p. 250-8.
5. Saitou, N. and M. Nei, *The neighbor-joining method: a new method for reconstructing phylogenetic trees.* Mol Biol Evol, 1987. **4**(4): p. 406-25.
6. Anderson, S., et al., *Sequence and organization of the human mitochondrial genome.* Nature, 1981. **290**(5806): p. 457-65.

Output ที่ได้จากโครงการ

1. การนำเสนอผลงาน poster presentation เรื่อง การวิเคราะห์ลำดับนิวคลิโอไทด์จากไมโตคอนเดรียของโครงกระดูกโบราณในแหล่งโบราณคดีที่สูงในปางมะผ้า จังหวัดแม่ฮ่องสอน การประชุมนักวิจัยรุ่นใหม่ พบเมธีวิจัยอาวุโส สกว. 9-11 มกราคม 2547 โรงแรมเฟลิกซ์ ริเวอร์แคว จ.กาญจนบุรี

APPENDIX

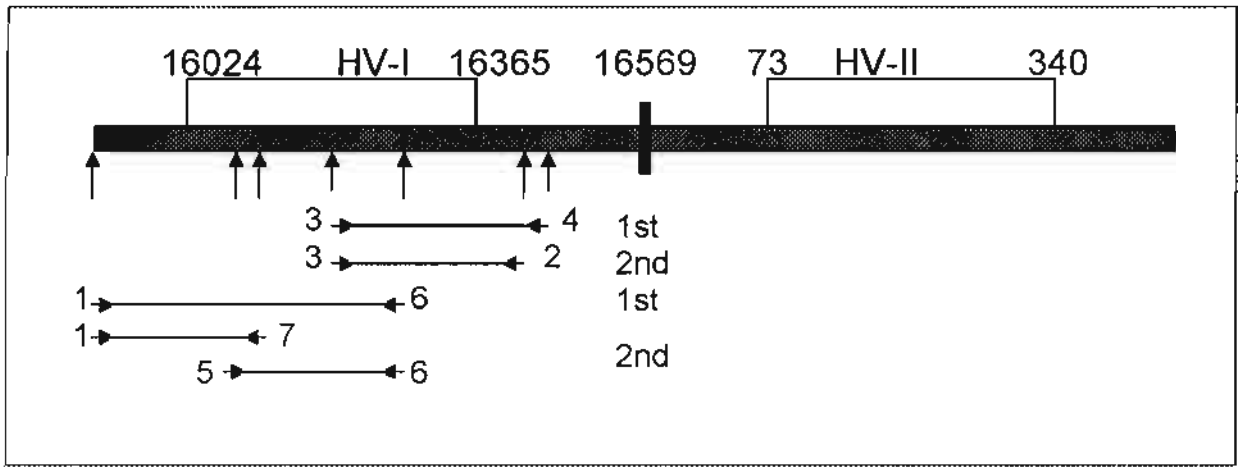

Figure. 1 *Schematic diagram of the mtDNA D-loop region containing the hypervariable region-I (HV-I) amplified in mtDNA testing. First and second PCR are indicated. Horizontal arrows indicate forward and reverse primers. Numbers next to the primers are CC primer numbers.*

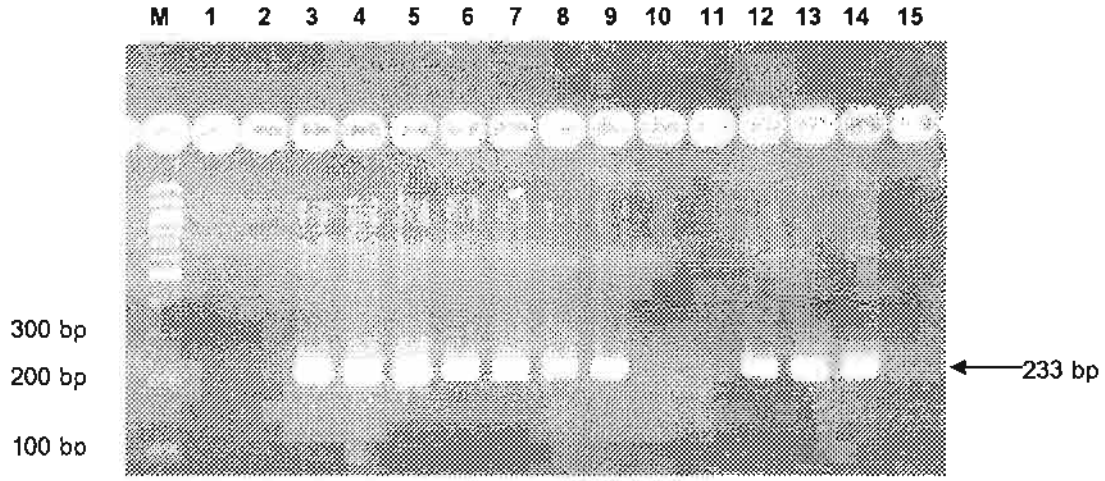

Figure 2 *Some of the positive PCR amplification from second PCR reaction using primer CC3/CC2 (lane3 to 15). M: 100 base pair ladder, lane 1: blank control, lane 2: negative control*

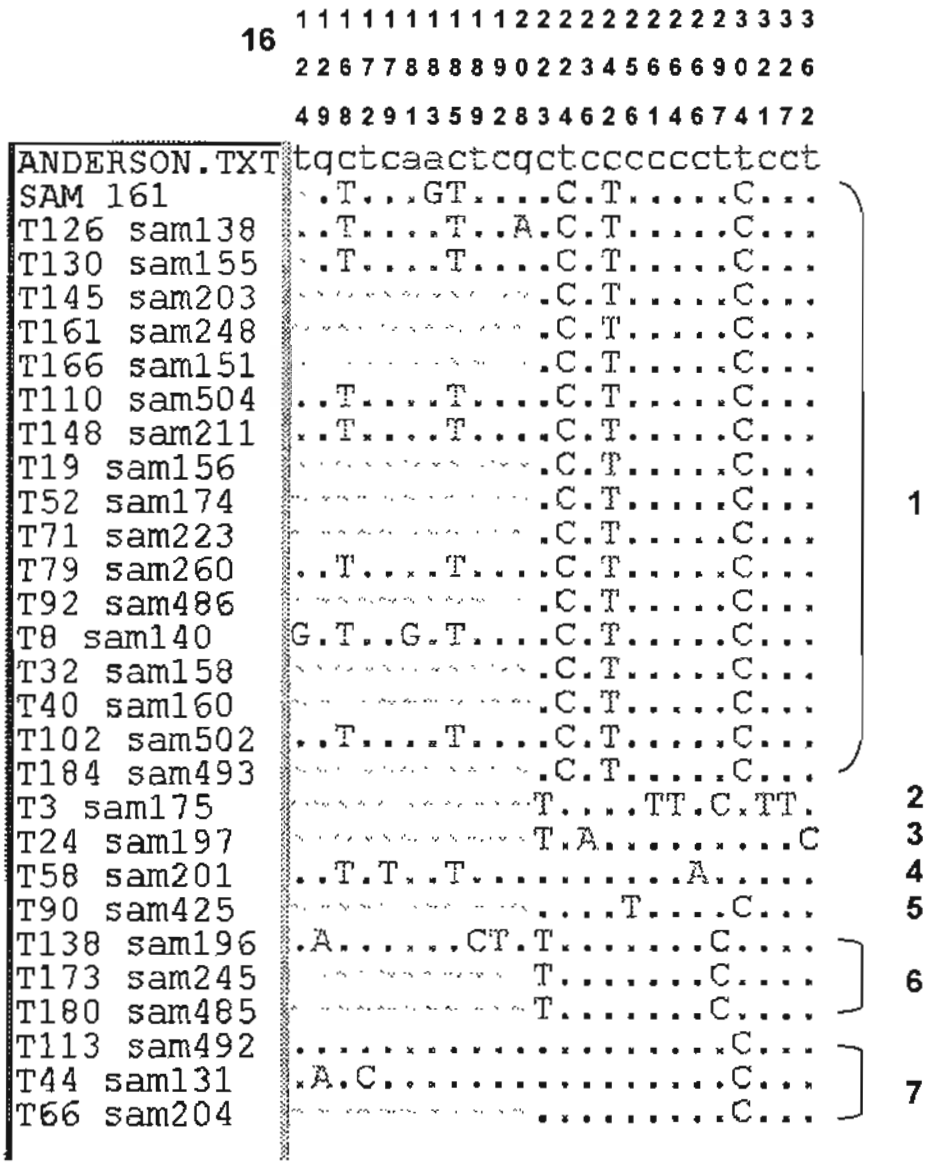

Figure 3 *Nucleotide sequence differences in the D-loop region of mtDNA for 28 ancient teeth. A total of at least 24 polymorphic sites are found in comparison with the D-loop region of mtDNA in the reference sequence of Anderson et al. (1981). Only the differences from the reference sequence are shown. T stands for each single clone of the plasmid DNA represent that sample number on the right. Dots(.) represent the same nucleotide as in reference sequence whereas ~ indicates unfinished sequence. Seven clusters of polymorphisms are shown on the right.*

OUTPUT from DNASP program

Input Data File: D:\...\ALIGN16209-16390 no anderson.fas

Number of sequences: 28 Number of sequences used: 28

Selected region: 1-182 Number of sites: 182

Total number of sites (excluding sites with gaps / missing data): 182

Pairwise Deletion option:

Sites with alignment gaps or missing data were considered

Analysis in Pairwise Comparisons

Gaps/missing information were excluded ONLY in pairwise comparisons

Number of pairwise comparisons: 378

Average number of sites analyzed: 182.00

Average number of differences: 2.431

Nucleotide diversity, Pi: 0.01336

Analysis at Individual Sites

Number of sites analyzed: 182.00

Number of polymorphic sites, S: 13

Average number of differences: 2.431

Nucleotide diversity, Pi: 0.01336

Theta-W, per sequence: 3.34065

Theta-W, per site: 0.01836

Figure 4 *Nucleotide diversity output of the 28 mtDNA sequence from ancient teeth from DNASP program.*

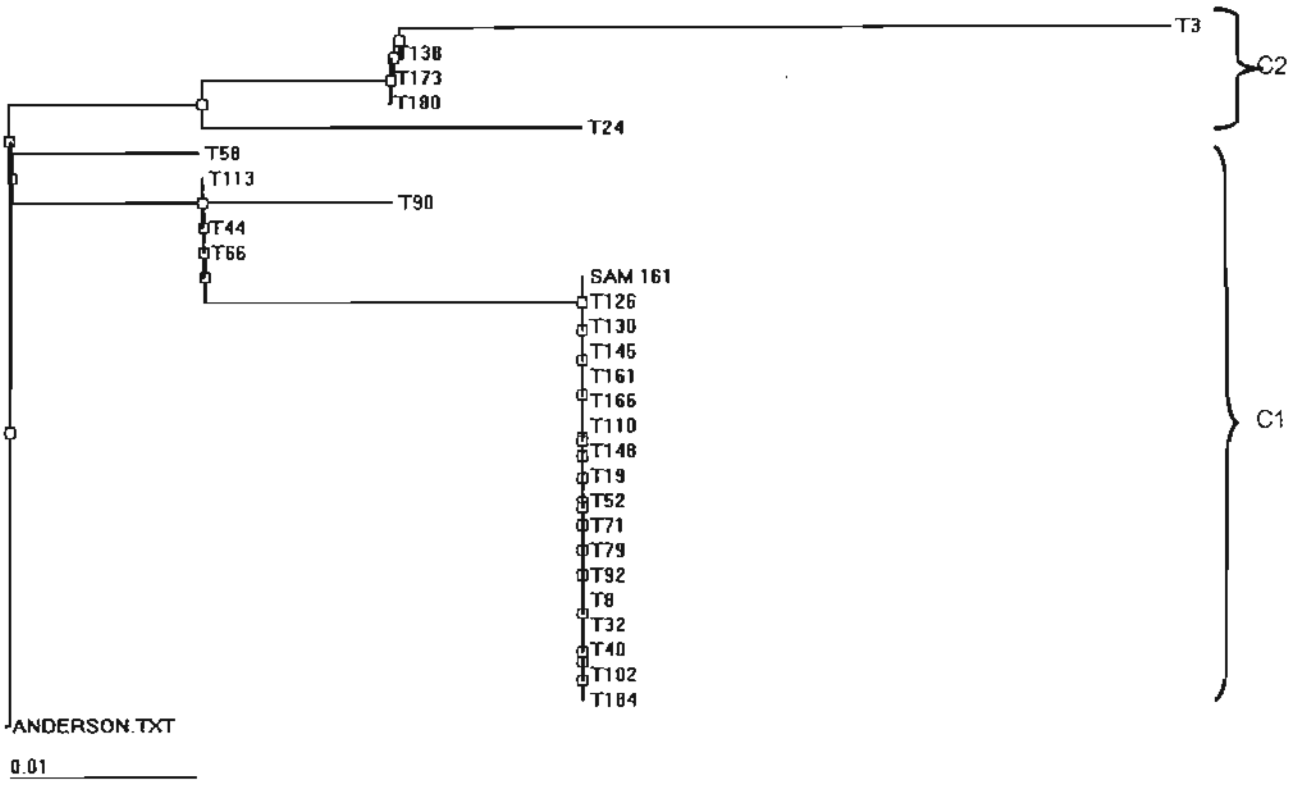

Figure 5 *Phylogenetic tree for the 28 mtDNA sequences using Neighbor-Joining/UPGMA method version 3.6a2.1. The two cluters in the tree are indicated by brackets (C1, C2).*

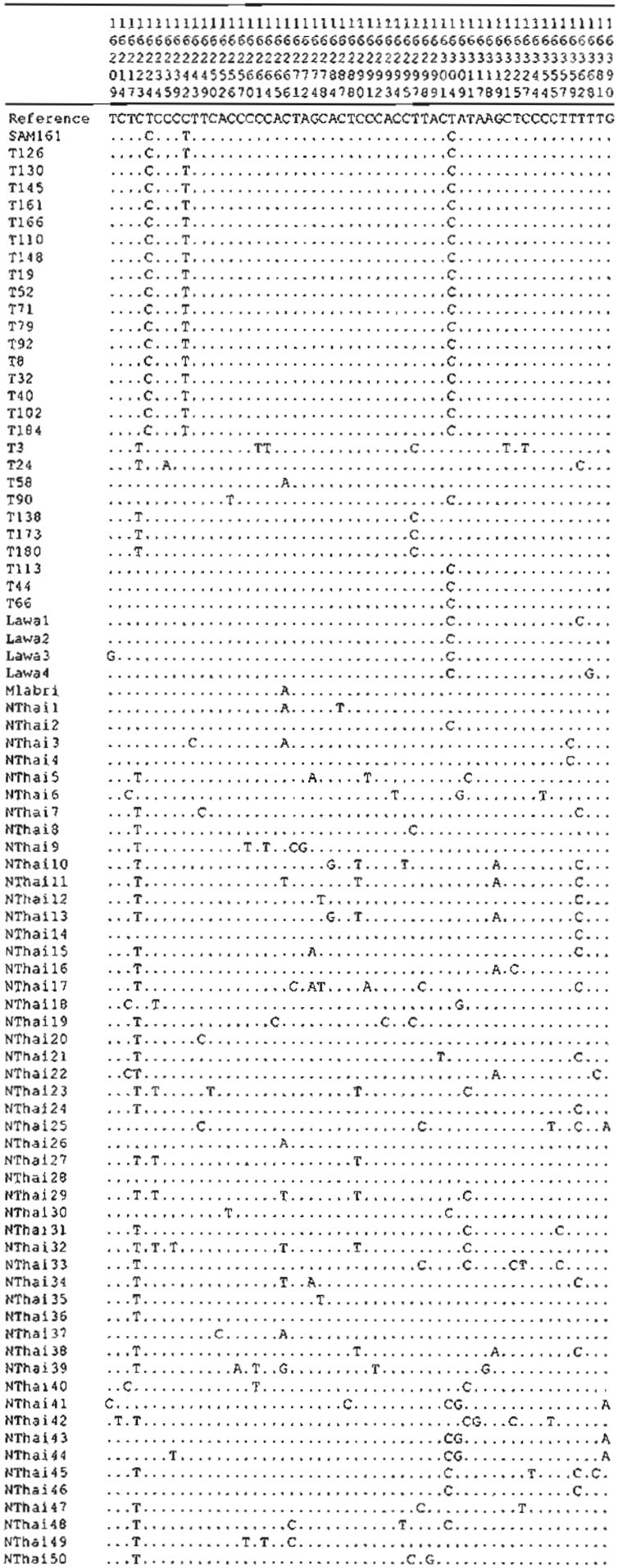

Figure 6 *Nucleotide sequence difference in D-loop region for 28 ancient teeth mtDNA sequences, 4 types of Lawa, 1 Mlabri, and 50 other northern Thai (NThai) based on 182-bp fragment. The positions of 55 polymorphic sites observed are numbered according to the published reference sequence for human mtDNA*

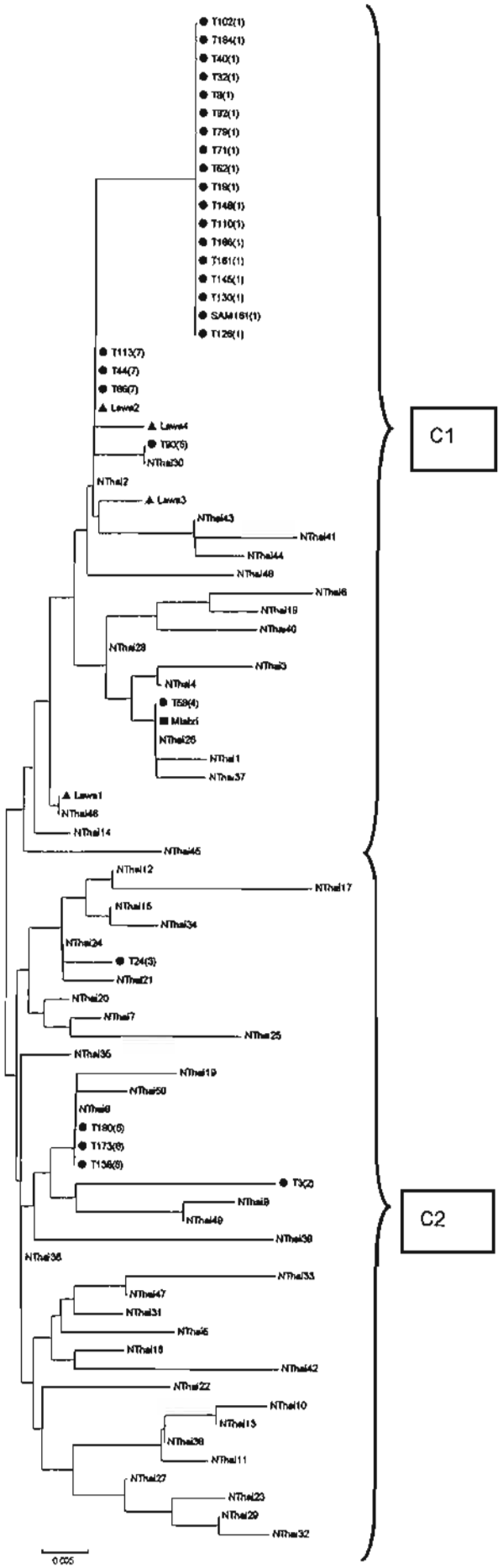

Figure 7 *Phylogenetic tree showing 28 ancient mtDNA lineages from Pang Ma Pha and three other populations. The phylogenetic tree was constructed, based on pairwise difference distant method, by the neighbor-joining method. The two clusters are indicated by brackets (C1, C2). The name of ethnic haplotypes (Lawa, Mlabri, and northern Thai or NThai) are indicated at the tips of each branch.*

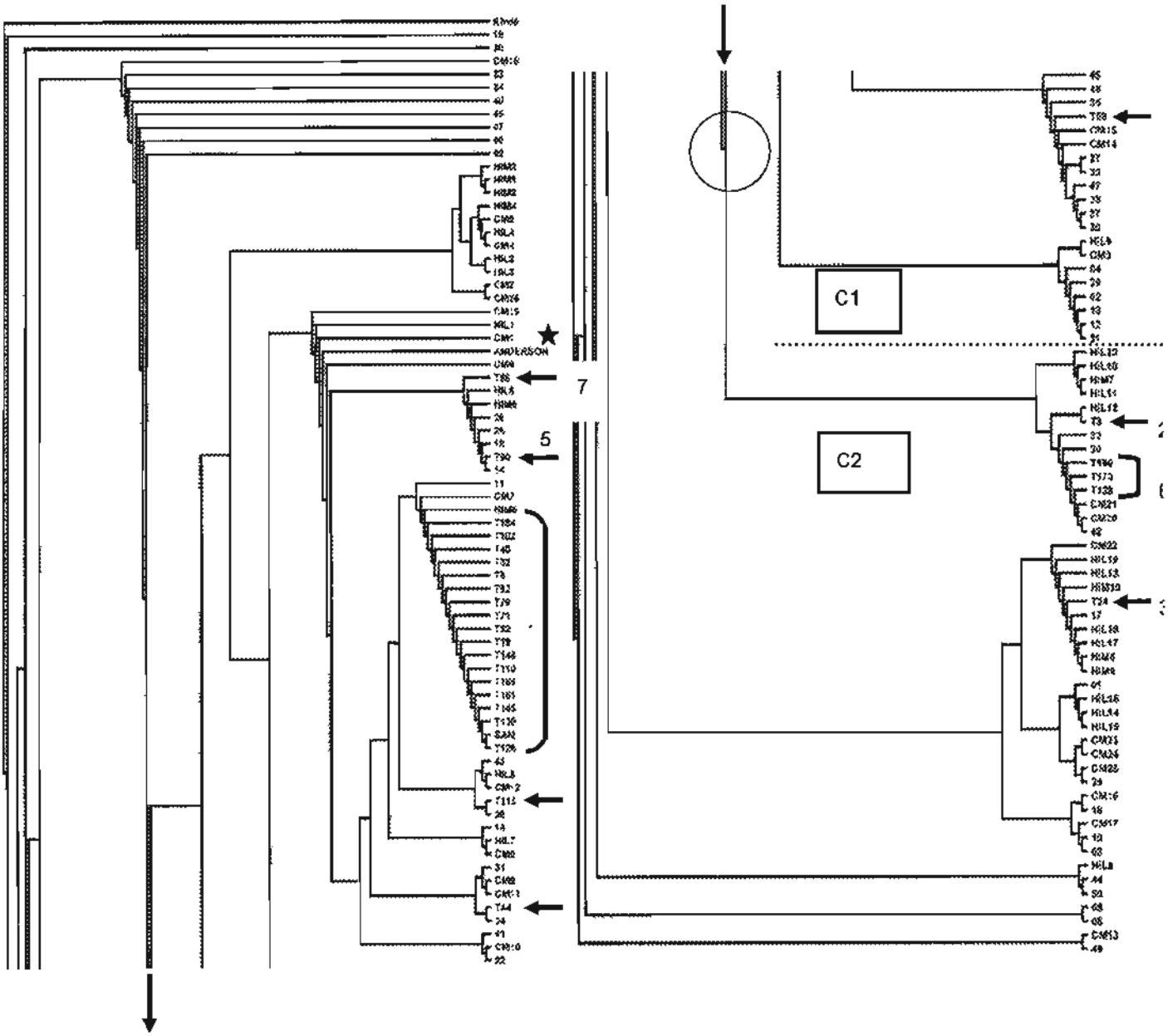

Figure 8 *Phylogenetic tree of the 28 ancient DNA comparing with 4 other subpopulations, Lisu (HIL), Mussur (HIM), Thai Chiang Mai (CM) and normal Khon Kaen (number) based on parsimony analysis by the neighbor-joining method. Two main clusters of 2000 year old DNA are observed (C1, C2) diverged at the circle point with the exception of T24 which branches first as shown in the rooted tree.*

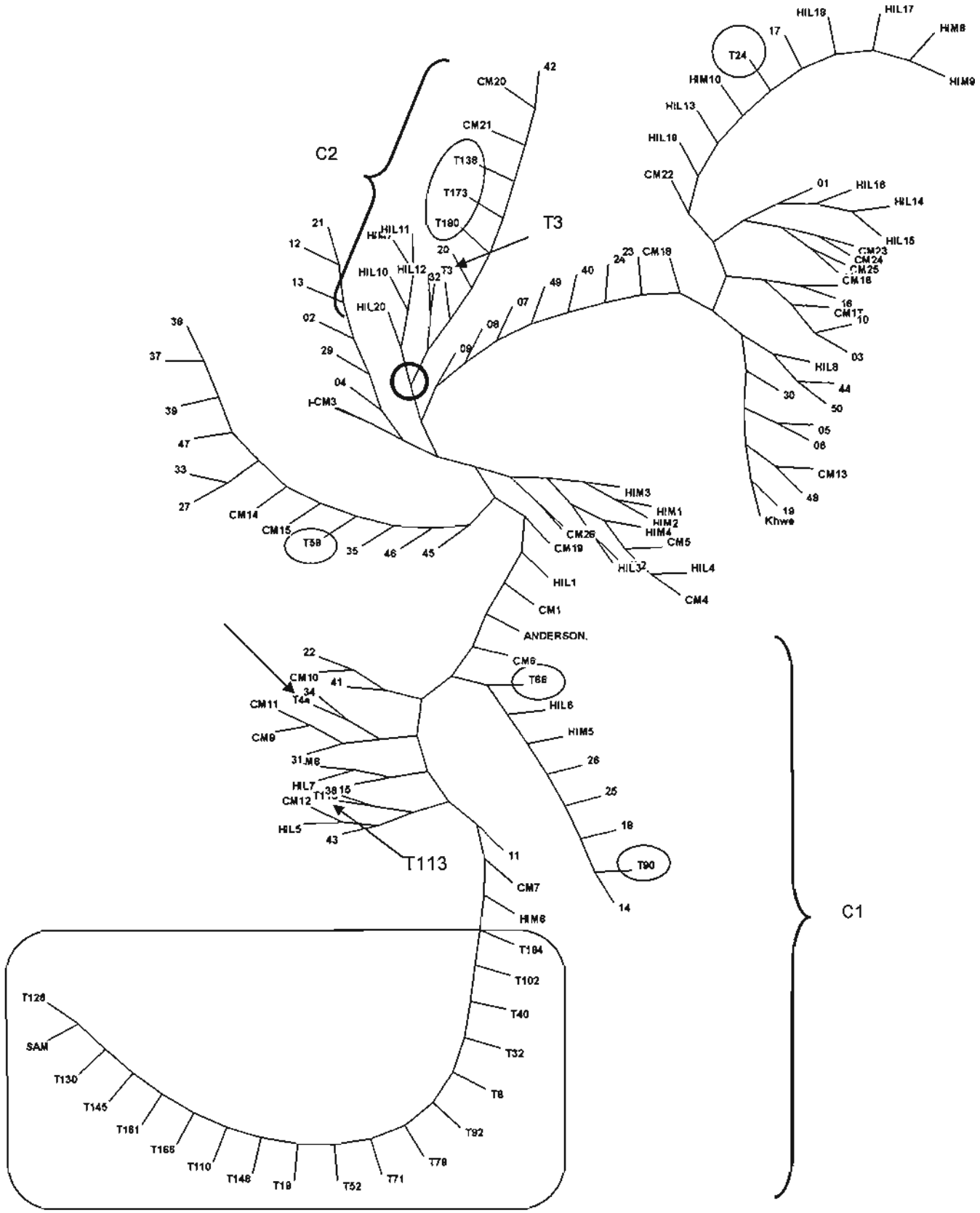

Figure 9 *Unrooted phylogenetic tree of the 28 ancient DNA comparing with 4 other subpopulations, Lisu (HIL), Mussur (HIM), Thai Chiang Mai (CM) and normal Khon Kaen (number) based on parsimony analysis by the neighbor-joining method.*