Int J Legal Med (2009) 123:495–501
DOI 10.1007/s00414-009-0373-4

ORIGINAL ARTICLE

Forensic and phylogeographic characterization of mtDNA lineages from northern Thailand (Chiang Mai)

Bettina Zimmermann · Martin Bodner · Sylvain Amory · Liane Fendt · Alexander Röck · David Horst · Basil Horst · Torpong Sanguansermsri · Walther Parson · Anita Brandstätter

Received: 25 May 2009 / Accepted: 13 August 2009 / Published online: 2 September 2009

Abstract The immigration of diverse ethnic groups over the past centuries from surrounding countries into Thailand left footprints in the genetic composition of Thai mitochondrial DNA (mtDNA) lineages. The entire mtDNA control region (1,122 bp) was typed in 190 unrelated male volunteers from the northern Thailand province of Chiang Mai following highest quality standards. For a more precise haplogroup classification, selected single nucleotide polymorphisms from the mtDNA coding region were genotyped. We found several new, so far undescribed mtDNA lineages. Quasi-median networks were constructed for visualisation of character conflicts. The data were put into population-genetic relationships with other Southeast Asian populations. Although the frequencies of the Thai haplogroups were characteristic for Southeast Asia in terms of haplotype composition and genetic structure, the Thai population was significantly different from other Southeast Asian populations. This necessitates establishing regional databases, especially for forensic applications. The population data have been submitted to the EMPOP database (www.empop.org) and will be available on publication.

Keywords mtDNA population data · Southeast Asia · Network analysis · Haplogroup assignment · Phylogeographic analyses · Forensic science

Introduction

The Kingdom of Thailand is located in the heart of Southeast Asia. Due to its geographical location, Thai culture has always been under strong influence from China and India. Thailand is the only Southeast Asian country

Electronic supplementary material The online version of this article (doi:10.1007/s00414-009-0373-4) contains supplementary material, which is available to authorized users.

B. Zimmermann · M. Bodner · S. Amory · L. Fendt · W. Parson (✉)
Institute of Legal Medicine, Innsbruck Medical University,
Müllerstrasse 44,
Innsbruck, Austria
e-mail: walther.parson@i-med.ac.at

S. Amory
Laboratory of Molecular Anthropobiology,
Institute of Legal Medicine,
Strasbourg, France

S. Amory
International Commission on Missing Persons,
Sarajevo, Bosnia and Herzegovina

A. Röck
Institute of Mathematics, University of Innsbruck,
Innsbruck, Austria

D. Horst
Pathologisches Institut, Ludwig-Maximilians-Universität,
Munich, Germany

B. Horst
Department of Surgical Pathology, Columbia University,
New York, NY, USA

T. Sanguansermsri
Faculty of Medicine, Chiang Mai University,
Chiang Mai, Thailand

A. Brandstätter (✉)
Division of Genetic Epidemiology,
Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University,
Schöpfstrasse 41,
Innsbruck, Austria
e-mail: anita.brandstaetter@i-med.ac.at

never colonised by a European power [1]. Thailand's population is dominated by various peoples speaking Tai-Kadai, which is one of the major language families of Southeast Asia [2]. Other resident ethnic groups in Thailand include Malays, Mon, Khmer, and various hill tribes [3].

The Chinese colonisation of Thailand started during the early second millennium AD when people migrated from southern China into Southeast Asia. During the twelfth and thirteenth centuries, Thai people were subjugated under the Khmers leading to an evident mitochondrial DNA (mtDNA) exchange between the Tai-Kadai and the Khmer groups [4]. In addition, Thailand has a long tradition of granting political asylum to religious or ethnically hunted refugees from neighbouring countries. Vietnamese Christians, Mon people from Burma, and political dissidents from Cambodia have sought and received shelter in Thailand since hundreds of years [5]. The genetic diversity of northern Thailand hill tribes, in particular, was additionally affected by sex-specific migration rates and cultural factors [3].

The consequences of this ethnic amalgamation on the genetic structure of Thailand are widely unknown. Available mitochondrial genetic data of Thailand's population is scarce, restricted by small sample sizes and limited to mitochondrial regions HVS-I [3, 4, 6–8] and HVS-I and HVS-II [9–11]. In this population study, 190 unrelated individuals living in the province of Chiang Mai (Fig. 1) were analysed, and their genetic structure was put into a population-genetic context with surrounding Asian populations.

Fig. 1 Map of Thailand. Note: The province of Chiang Mai is *shaded in grey*

Materials and methods

Samples, DNA extraction, and control region sequence analysis

Blood samples were taken from 190 unrelated male volunteers from small rural villages throughout the northern Thai province of Chiang Mai. DNA was isolated from peripheral blood lymphocytes according to standard procedures [12]. The entire mtDNA control region (CR; nps 16024–16569; 1–576) was amplified, sequenced, and evaluated following EMPOP recommendations [13, 14] and updated nomenclature guidelines for mtDNA [15].

Haplogroup assignment and analysis of coding region SNPs

MtDNA CR haplotypes were affiliated to haplogroups based on patterns of shared haplogroup-specific or -associated polymorphisms [16–23]. To confirm and specify the haplogroup assignment derived from the CR patterns, we analysed selected coding region single nucleotide polymorphisms (SNPs) targeting haplogroup-specific mutations in Southeast Asian lineages using a modified version of the SNaPshot assay presented in [24]. Furthermore, we developed a sequencing-based screening method for haplogroup M, N, and C-specific coding region SNPs. The analysed SNPs were selected on the basis of published phylogenetic trees [16–23, 25].

In paragroup N* samples, we examined two coding region fragments ranging from nps 4450–5860 and from nps 6340–7770. For unresolved or ambiguous paragroup M* associated samples, we determined the sequences of

Table 1 Diversity measures from various ethnic populations from Thailand

Population statistics	Chiang Mai	Akha	Chiang Rai Lisu	Black Lahu	Mae Hong Son Lisu	Red Karen	White Karen	Various ethnic groups of Thailand
N	190	91	53	39	42	39	40	215
Reference	This study	Oota 2001 [6]	Oota 2001 [6]	Oota 2001 [6]	Oota 2001 [6]	Oota 2001 [6]	Oota 2001 [6]	Fucharoen 2001 [9]
Number of haplotypes	124	28	28	11	19	14	17	131
Number of unique haplotypes	95	12	20	6	7	8	10	125
Haplotype diversity	0.985	0.928	0.909	0.834	0.932	0.839	0.891	0.992
Mean pairwise differences	7.498±3.518	5.616±2.720	6.307±3.039	6.0256±2.934	5.843±2.849	4.626±2.319	5.108±2.529	7.016±3.302

Analysed range, 16048–16365

four fragments within the coding region (nps 1820–2450, nps 4450-5580, nps 8670–9850, and nps 11450–12820). Furthermore, one fragment was investigated ranging from nps 5820–6660 for samples assigned to lineage C*.

These target fragments were amplified in singleplex polymerase chain reaction (PCR) reactions in a total volume of 25 µl under following conditions: 95°C for 2 min followed by 35 cycles of 95°C for 15 s, 56°C for 30 s, and 72°C for 90 s, followed by an extension phase at 72°C for 10 min. Purification of PCR products, cycle sequencing, and purification of cycle sequencing products were carried out following standard procedures [14]. The amplification and sequencing primer sequences applied are listed in Table S1a and S1b, respectively.

Random match probability

The random match probability was calculated as the sum of squared haplotype frequencies [26] based on mtDNA CR sequences (C-insertions in length-heteroplasmic regions around nps 16193, 309, and 573 were disregarded).

Population-genetic analyses

Molecular diversity indices, pairwise differences between and within populations, and an analysis of molecular variance (AMOVA) were calculated using ARLEQUIN (version 3.11). The Chiang Mai samples from northern Thailand were compared with hill tribes and other ethnic groups of Thailand (Table 1). All sequences were aligned and trimmed to a greatest common range of nps 16048–16365, and length-heteroplasmic C-insertions around np 16193 were disregarded.

In addition, the Thailand samples were compared with surrounding Southeast Asian populations (Table 2). All sequences were aligned and trimmed to a greatest common range of nps 16024–16365 and nps 73–340; length-heteroplasmic C-insertions around nps 16193 and 309 were disregarded.

Results and discussion

Haplogroup structure and most common haplotypes

Within a total of 190 samples from Chiang Mai, 145 distinct CR haplotypes (disregarding C-insertions around positions 16193, 309, and 573) were found (Table S2). The proportions of the macrohaplogroups M, N, and R were 39.5%, 7.9%, and 52.6%, respectively. The most common haplotype was F1a (9.5%; 16129A-16172C-16304C-16519C-73G-249DEL-263G-315.1C-523DEL-524DEL),

Table 2 Diversity measures from Southeast Asian populations

Population statistics	Thailand	Japan	N-China	Korea	Taiwan	Vietnam	Malaysia
N	190	162	232	593	640	187	205
Reference	This study	Imaizumi 2002 [27]	Kong 2003 [28]	Lee 2006 [29]	Trejaut 2005 [30]	Irwin 2008 [31]	Wong 2007 [32]
Number of haplotypes	137	131	129	408	339	154	152
Number of unique haplotypes	108	119	80	325	231	136	115
Haplotype diversity	0.989	0.988	0.988	0.995	0.994	0.991	0.991
Mean pairwise differences	10.213±4.684	8.878±4.114	8.427±3.914	9.594±4.405	9.790±4.489	10.366±4.750	10.680±4.883

Analysed range, 16024–16365 and 73–340

followed by F1a1a (7.9%; 16108T-16129A-16162G-16172C-16304C-16519C-73G-249DEL-263G-315.1C-523DEL-524DEL) and M7b1 (6.8%; 16129A-16192T-16223T-16297C-73G-150T-199C-263G-315.1C-489C).

Haplogroups C, G, A, and Z each occurred at frequencies lower than 3%. We found five remarkable lineages within macrohaplogroups M and N that were new to the current literature; until their phylogenetic position is clarified, we denote those lineages as "macrohaplogroup-polymorphic position" (e.g., "N-5120"). A detailed list of all haplogroups and their frequencies are given in Table 3 and Fig. 2, respectively.

Effectiveness of targeted coding region sequencing

All samples were sequenced in the entire CR and SNP-genotyped with an extended version of the multiplex assay presented in [24]. For the majority of samples (72.6%), the haplogroup information from the SNP multiplex genotyping assay confirmed the haplogroup status of CR sequencing but did not lead to a higher resolution. For 12.6% of samples, however, the SNP multiplex provided a more precise haplogroup affiliation. In the majority of remaining samples (14.8%), targeted coding region sequencing enabled a more detailed haplogroup assignment.

Random match probability and mean number of pairwise differences

The estimated probability of a random match between two unrelated individuals from the Chiang Mai dataset was 1.04%, corresponding to a power of discrimination of 98.96% for the entire CR, when ignoring length variants around positions 16193, 309, and 573. The mean number of pairwise differences for this dataset was calculated 11.56± 5.26 ignoring hotspot insertions.

Comparison with other ethnic groups from Thailand

To shed more light on the genetic composition in Thailand, we compared the current sample set from Chiang Mai, northern Thailand, with six distinct ethnic groups and a compiled sample set of various ethnic groups of Thailand (Table 1). The highest intra-population diversity was found within our sample set from Chiang Mai with an average number of pairwise differences around 7.5 followed by the sample set from [9] (7.0). The lowest intra-population diversity was found in the matrilocal group of Red Karen with an average number of pairwise differences around 4.6. Twenty-seven haplotypes (21.8%) of Chiang Mai were found in other populations from Thailand (Table S3a). All six ethnic groups (Akha, Lisu from Chiang Rai and from Mae Hong Son, Black Lahu, Red Karen, and White Karen)

Table 3 Haplogroup frequencies of 190 samples from northern Thailand (Chiang Mai)

Haplogroup	*N*	Frequency (%)
A	3	1.6
B*	1	0.5
B4*	1	0.5
B4a	4	2.1
B4b	4	2.1
B4c	4	2.1
B4d	4	2.1
B4g	2	1.1
B5a	10	5.3
C7	5	2.6
D4	8	4.2
D5	2	1.1
F*	2	1.1
F1*	1	0.5
F1a*	18	9.5
F1a1*	1	0.5
F1a1a	15	7.9
F1a1c	1	0.5
F1a2	3	1.6
F1b′d	1	0.5
F1c	1	0.5
F2	4	2.1
F3a	8	4.2
G2	5	2.6
M*	8	4.2
M14	1	0.5
M-5054	4	2.1
M7b*	8	4.2
M7b1	13	6.8
M7b'd	1	0.5
M7c	2	1.1
M-8718*	4	2.1
M-8718-9548	8	4.2
M9	5	2.6
N*	1	0.5
N-5120	1	0.5
N-5581	5	2.6
N9a	3	1.6
pre-N22	2	1.1
R*	1	0.5
R11	1	0.5
R2	2	1.1
R22	1	0.5
R9	10	5.3
Z	1	0.5

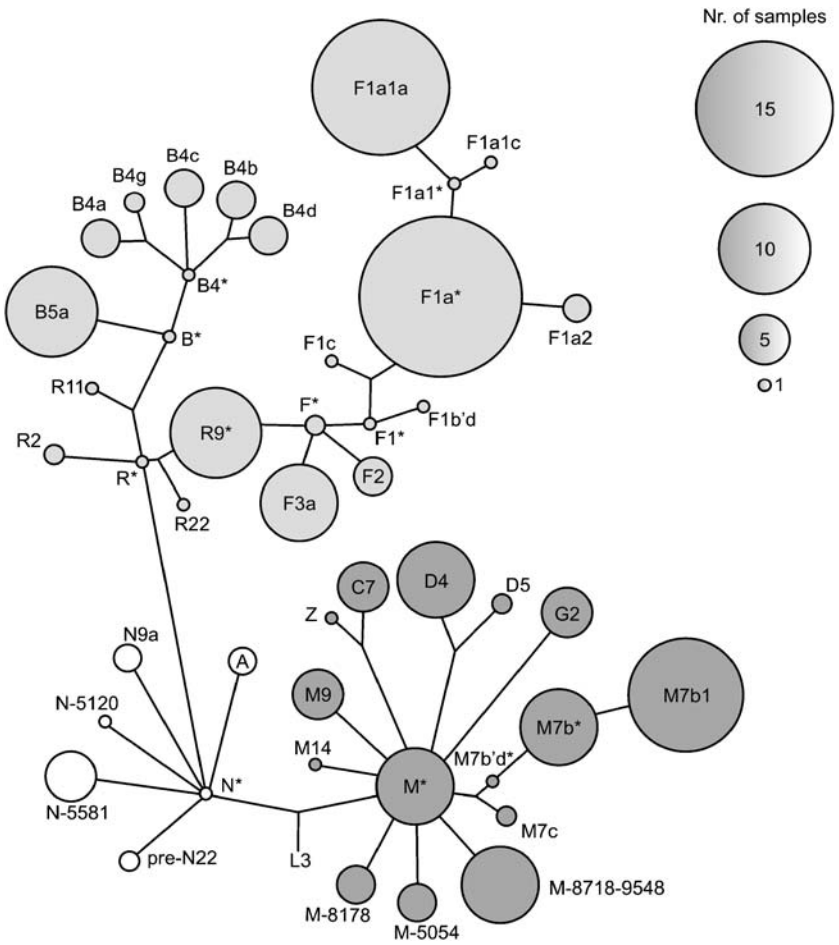

Fig. 2 Schematic tree of mitochondrial lineages in a population sample from Chiang Mai, Thailand (N=190). Note: The size of the spheres portrayed corresponds to haplogroup frequencies. *Dark grey*: descendants from haplogroup M; *light grey*: descendants from haplogroup R; *white*: descendants from haplogroup N. The tree is rooted in L3

and the population compilation from [9] share various haplotypes at similar proportions with the sample from Chiang Mai, thus affirming its representativeness for northern Thai populations. AMOVA was used to test for significant variation in the mtDNA genetic structure among the various Thai populations (Table S4). Most of the observed genetic variation was attributable to differences within populations (95.0%). Variance among populations accounted for 5.0% (Table S4a). Although all of these populations were from the same geographical region, they differed highly significantly in their genetic structure (Table S4b), which can be explained by their population history (agricultural versus partly nomadic life style) and different social structures. The population average pairwise differences (Table S4c) showed that the mean number of pairwise differences for the Chiang Mai sample was lowest with the White Karen population. However, when using corrected values, the observed and the expected number of pairwise differences was closest between Chiang Mai and the composition of various ethnic groups from [9], thus further demonstrating the similarity in genetic structure of these two complex random population samples.

Comparison with other Southeast Asian populations

We found 19 shared haplotypes (approximately 14%) in the Thai and other Southeast Asian populations (Table S3b). AMOVA was used to test for significant variation in the mtDNA distributions among the various Southeast Asian populations (Table S5). The vast majority of the observed variance (96.3%) within these six populations was attributable to differences within populations, and 3.7% represented differences among populations (Table S5a).

One of the most common haplotypes (comparison range, 16048–16365), a haplogroup F1a lineage (16129A-16172C-16304C-16519C) found in the population from Chiang Mai was also found in several Thai groups (Akha, Black Lahu, Red Karen, White Karen, and in the data set from [9]), as well as in Vietnam and Malaysia. Also, one haplogroup M7b1 lineage (16129A-16192T-16223T-16297C) was found several times in the dataset from [9] and in the datasets from northern China and Vietnam. Interestingly, haplogroup F1a1a, clearly defined by CR polymorphisms 16108T and 16162G, could only be found in the Thai populations and not in the surrounding Asian

populations. A haplotype that only occurred once in our Chiang Mai sample (16223T-16311C-16362C-16519C) seemed to be common for Southeast Asia since it was found in Malaysia (twice) and in the distant Korean population (once), as well as in the group of Akha (observed 12 times), in the White Karen (observed five times), and in the compilation of [9] (observed four times; see Tables S3b).

Conclusions

The mtDNA CR sequences presented in this study were generated according to high-quality laboratory standards. The frequencies of the different haplogroups were characteristic for Southeast Asian populations. However, the Thai population was significantly different both in terms of haplotype composition and genetic structure. This necessitates establishing regional databases especially for forensic database searches in order to get reliable frequency estimates. From a phylogenetic point of view, it was interesting to find new, so far undescribed lineages. More sequences from such haplotypes need to be collected and typed for the entire mtDNA genome in order to expand the Southeast Asian mtDNA tree. The haplotypes reported in the present study will be made available from the EMPOP database (www.empop.org) on publication.

This publication follows the recommendations of the International Society for Forensic Genetics on the use of mtDNA in forensic analysis.

Acknowledgements The authors would like to thank Gabriela Huber and Daniela Niederwieser for excellent technical assistance (Institute of Legal Medicine, Innsbruck Medical University). The study was supported by the FWF Austrian Science Fund Translational Research Programme (L397).

References

1. General Information About Thailand (2009) http://www.siamweb.org/content/Thailand/131/index_eng.php
2. LeBar M, Hickey GC, Musgrave JK (1964) Ethnic groups of mainland Southeast Asia. Human Relations Area Files Press, New Haven
3. Besaggio D, Fuselli S, Srikummool M et al (2007) Genetic variation in Northern Thailand Hill Tribes: origins and relationships with social structure and linguistic differences. BMC Evol Biol 7(Suppl 2):S12
4. Lertrit P, Poolsuwan S, Thosarat R et al (2008) Genetic history of Southeast Asian populations as revealed by ancient and modern human mitochondrial DNA analysis. Am J Phys Anthropol 137:425–440
5. The World Factbook: Thailand (2009) https://www.cia.gov/library/publications/the-world-factbook/geos/th.html
6. Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M (2001) Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. Nat Genet 29:20–21
7. Yao YG, Nie L, Harpending H, Fu YX, Yuan ZG, Zhang YP (2002) Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. Am J Phys Anthropol 118:63–76
8. Li H, Cai X, Winograd-Cort ER et al (2007) Mitochondrial DNA diversity and population differentiation in southern East Asia. Am J Phys Anthropol 134:481–488
9. Fucharoen G, Fucharoen S, Horai S (2001) Mitochondrial DNA polymorphisms in Thailand. J Hum Genet 46:115–125
10. Allard MW, Wilson MR, Monson KL, Budowle B (2004) Control region sequences for East Asian individuals in the Scientific Working Group on DNA Analysis Methods forensic mtDNA data set. Leg Med (Tokyo) 6:11–24
11. Jin HJ, Kwak KD, Hong SB et al (2006) Forensic genetic analysis of mitochondrial DNA hypervariable region I/II sequences: an expanded Korean population database. Forensic Sci Int 158:125–130
12. Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215
13. Parson W, Bandelt H-J (2007) Extended guidelines for mtDNA typing of population data in forensic science. Forensic Sci Int Genet 1:13–19
14. Brandstätter A, Niederstätter H, Pavlic M, Grubwieser P, Parson W (2007) Generating population data for the EMPOP database—an overview of the mtDNA sequencing and data evaluation processes considering 273 Austrian control region sequences as example. Forensic Sci Int 166:164–175
15. Bandelt H-J, Parson W (2007) Consistent treatment of length variants in the human mtDNA control region: a reappraisal. Int J Legal Med . doi:10.1007/s00414-006-0151-5
16. Kivisild T, Tolk HV, Parik J et al (2002) The emerging limbs and twigs of the East Asian mtDNA tree. Mol Biol Evol 19:1737–1751
17. Kong Q-P, Bandelt H-J, Sun C et al (2006) Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. Hum Mol Genet 15:2076–2086
18. Thangaraj K, Chaubey G, Singh VK et al (2006) In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup 'M' in India. BMC Genomics 7:151
19. Abu-Amero KK, Gonzalez AM, Larruga JM, Bosley TM, Cabrera VM (2007) Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. BMC Evol Biol 7:32
20. Hill C, Soares P, Mormina M et al (2007) A mitochondrial stratigraphy for island southeast Asia. Am J Hum Genet 80:29–43
21. Reddy BM, Langstieh BT, Kumar V et al (2007) Austro-Asiatic tribes of Northeast India provide hitherto missing genetic link between South and Southeast Asia. PLoS ONE 2:e1141
22. Chaubey G, Metspalu M, Karmin M et al (2008) Language shift by indigenous population: a model genetic study in South Asia. Int J Hum Genet 8:41–50
23. Soares P, Trejaut JA, Loo JH et al (2008) Climate change and postglacial human dispersals in southeast Asia. Mol Biol Evol 25:1209–1218
24. Álvarez-Iglesias V, Jaime JC, Carracedo A, Salas A (2007) Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups. Forensic Sci Int Genet 1:44–55
25. van Oven M, Kayser M (2008) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30:E386–E394
26. Brandstätter A, Parsons TJ, Parson W (2003) Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. Int J Legal Med 117:291–298
27. Imaizumi K, Parsons TJ, Yoshino M, Holland MM (2002) A new database of mitochondrial DNA hypervariable regions I and II

sequences from 162 Japanese individuals. Int J Legal Med 116:68–73
28. Kong Q-P, Yao Y-G, Liu M et al (2003) Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. Hum Genet 113:391–405
29. Lee HY, Yoo J-E, Park MJ, Chung U, Shin K-J (2006) Mitochondrial DNA control region sequences in Koreans: identification of useful variable sites and phylogenetic analysis for mtDNA data quality control. Int J Legal Med 120:5–14
30. Trejaut JA, Kivisild T, Loo JH et al (2005) Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. PLoS Biol 3:e247
31. Irwin JA, Saunier JL, Strouss KM et al (2008) Mitochondrial control region sequences from a Vietnamese population sample. Int J Legal Med 122:257–259
32. Wong HY, Tang JS, Budowle B et al (2007) Sequence polymorphism of the mitochondrial DNA hypervariable regions I and II in 205 Singapore Malays. Leg Med (Tokyo) 9:33–37