

ORIGINAL ARTICLE

Genetic affinity and admixture of northern Thai people along their migration route in northern Thailand: evidence from autosomal STR loci

Wibhu Kutanan¹, Jatupol Kampuansai¹, Vincenza Colonna², Supaporn Nakbunlung³, Pornpilai Lertvicha⁴, Mark Seielstad⁵, Giorgio Bertorelle² and Daoroong Kangwanpong¹

The Khon Mueang (KM) are the largest group of northern Thai people. Our previous mtDNA studies have suggested an admixture process among the KM with the earlier Mon-Khmer-speaking inhabitants of this region. In this study, we evaluate genetic affinities and admixture among 10 KM populations in northern Thailand lying along the historical Yuan migration route, and 10 neighboring populations belonging to 7 additional ethnic groups: Lawa, Mon (Mon-Khmer-speaking groups), Shan, Yuan, Lue, Khuen and Yong (Tai-speaking groups) by analyzing 15 hypervariable autosomal short tandem repeat loci. The KM exhibited close relationships with neighboring populations, especially the Tai-speaking groups, reflecting an admixed origin of the KM. Admixture proportions were observed in all KM populations, which had a higher contribution from the parental Tai than the Mon-Khmer groups. Different admixture patterns of the KM along the migration route might indicate high heterogeneity among the KM. These patterns were not directly associated with geographical proximity, suggesting other factors, like variation in the timing of admixture with the existing populations may have had an important role. More genetic data from different marker systems solely transmitted through the male or female lineages are needed to complete the description of genetic admixture and population history of the KM.

Journal of Human Genetics (2011) 56, 130–137; doi:10.1038/jhg.2010.135; published online 25 November 2010

Keywords: autosomal STR; genetic affinity; genetic admixture; khon mueang; migration route

INTRODUCTION

Northern Thailand is in the landlocked heart of the golden peninsula of southeast Asia, characterized river plains, separated by steep mountainous ranges. It shares a border with Myanmar and Laos, and lies in close proximity to southern China. This territory was primarily inhabited by the indigenous Lawa (LW), who probably descended from younger prehistoric peoples around the fifth century AD.^{1–5} Subsequently, the area became ethnically, culturally and linguistically diversified through migration and settlement of the Mon and several Tai-speaking populations. The MO (MO) were among the oldest settlers in Thailand and established an advanced civilization in central Thailand around the fifth century AD. During the second half of the eighth century AD, they extended their kingdom as far as today's Lamphun province in northern Thailand.¹

The Yuan (YU), whose origins remain mysterious, are regarded as the first Tai-speaking population to reach Thailand. By the eighth century, they had established the first Tai kingdom, Chiang Sean, located in present-day Chiang Rai province.² By the end of thirteenth century AD,

the Yuan migration began with the invasion of Mon-speaking lands in Lamphun. They founded a new city at Chiang Mai and then extended their kingdom, named Lan Na, southward to Lampang.^{1–5} Apart from the Mon-Khmer and the Yuan groups, residing along the historical Yuan migratory route (Figure 1), this area was also populated by other Tai peoples: Lue, Yong, Khuen and Shan.

The Lue (LU) kingdom named Chiang Hung was founded in southern China around the twelfth century. At the present time, the Lue are distributed over parts of southern China, northern Myanmar, northern Laos and northern Thailand.^{2,6} People from two Lue villages (LU1, LU2) in this study are thought to have migrated from different cities in Xishuangbanna through Laos into northern Thailand about 200–300 years ago.⁷ The Yong (YO) were residents of Yong city (Mong Yawng) in Shan State of Myanmar, and were probably originally Lue. They were taken captive by war, and forcibly migrated into Thailand in 1805 AD.^{4,7} The city of the Khuen (KH), located at Keng Tung in the Shan State of Myanmar, was built by the Yuan king in 1267 AD.⁴ Their migration into Thailand was part of the same set of events that

¹Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand; ²Dipartimento di Biologia ed Evoluzione, Università di Ferrara, Ferrara, Italy; ³Department of Sociology and Anthropology, Faculty of Social Sciences, Chiang Mai University, Chiang Mai, Thailand; ⁴Tarnpanya Co.Ltd., Chiang Mai, Thailand and ⁵Population Genetics and Genetic Epidemiology, Genome Institute of Singapore, Biopolis, Singapore
Correspondence: Dr D Kangwanpong, Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50202, Thailand.
E-mail: scidkngw@chiangmai.ac.th

Received 9 June 2010; revised 12 September 2010; accepted 22 September 2010; published online 25 November 2010

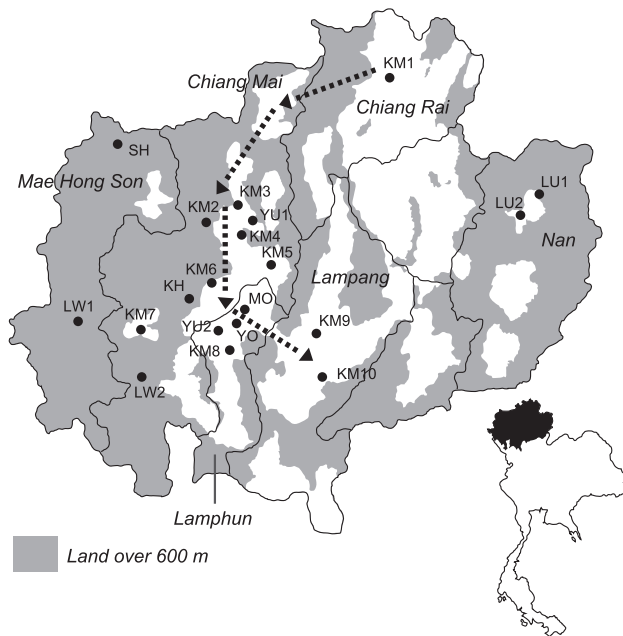


Figure 1 Geographical distribution of studied populations. See the meaning of population abbreviations in Table 1. The migration route of Khon Mueang is represented by dash arrow.

displaced the Yong. The Lue, Yong and Khuen's colloquialisms, script and their spoken dialects are similar to each other.⁷ The Shan (SH) or Tai Yai, which means great Tai, predominantly reside in the Shan State of Myanmar that bears their name. After their many migrations into northern Thailand, beginning around 1400 AD, the Shan today reside in the mountainous areas of Mae Hong Son province. With each successive generation, some of these people retained their original ethnicity, while those who assimilated became the present-day Khon Mueang (KM).²

Among the people living in present day northern Thailand, the Khon Mueang, speaking a Tai-Kadai language¹ comprise the majority. 'Khon Mueang' is the local name commonly applied by the people of northern Thailand to themselves,⁸⁻⁹ but this word now applies mostly to a social and political category rather than a distinct population, and it is not the ethnic name of any specific group. Nevertheless, the biological ancestry of the Khon Mueang has been variously proposed to originate from local Mon-Khmer-speaking peoples; the Lawa and Mon, assimilated by the Tai rulers; or via direct descent from the Tai-speaking groups who had migrated from southern China.^{3,10-11} Our previous analysis of mtDNA variation in the Khon Mueang and other Tai-speaking groups supported this latter hypothesis: that the Khon Mueang ancestors originated from Tai-speaking groups in south and southeast China, with subsequent admixture with Mon-Khmer populations.⁷

Admixture, the genetic consequence of mixing between genetically differentiated populations, resulting from geographical, ecological or cultural separation, is common in human evolution. Pioneering research on genetic admixture of human populations was mainly applied to African-Americans,¹²⁻¹⁵ Latin-Americans,¹⁶ Europeans¹⁷⁻¹⁸ and East Asians,¹⁹⁻²¹ but has been limited in Southeast Asia.

Geography had an important role in the location of Tai settlements. River plains, surrounded by mountain peaks, were chosen because of their high agricultural productivity. Three geographically separated basins; the Chiang Rai Basin in the Kok River valley, the Chiang Mai-Lamphun Basin in the Ping River valley and the Lampang Basin in the

Wang River valley, were the most suitable areas, along their migration route from Chiang Rai to Lampang province (Figure 1).⁴

Microsatellites, also known as short tandem repeats (STRs) have been broadly used for inferring human population history²²⁻²³ and relationships among continental populations²⁴ as well as between geographically contiguous populations.²⁵⁻²⁷ In this research work, autosomal STR variation analysis of 10 present-day Khon Mueang villages from the areas of Yuan settlement, as well as 10 other populations, who had an important role in northern Thai history (and, hence, could potentially be parental populations of the Khon Mueang) were performed, to investigate the population affinity of the studied populations and to explore the level of admixture in the Khon Mueang.

MATERIALS AND METHODS

Samples

Blood samples were collected from 943 unrelated healthy individuals (436 from Khon Mueang and 507 from neighboring populations), with informed consent, after interviewing them about their linguistic, village and personal history. Ten Khon Mueang (KM) populations reside along the Yuan migratory route in three different basins. KM1 reside in the Chiang Rai Basin, whereas KM2-KM6 and KM8 are in the Chiang Mai-Lamphun Basin, except KM7, who reside in the Mae Cham Basin of Chiang Mai (which is not on the Yuan migration route). KM9 and KM10 settled in the Lampang Basin (Figure 1). Based on historical sources, 10 neighboring populations, belonging to 7 ethnic groups, were also sampled: the Lawa (LW1, LW2), Mon (MO), Shan (SH), Yuan (YU1, YU2), Khuen (KH), Lue (LU1, LU2) and Yong (YO), who have had historical contact with the KM. DNA samples of the YU, KH, LU and YO from the previous study⁷ were also included in this study. General information about the studied populations is listed in Table 1.

DNA isolation and STR typing

Genomic DNA was extracted, using a standard inorganic salting-out method.²⁸ Multiplex PCR amplification was performed, using a commercial AmpFISTR Identifier kit (Applied Biosystem, Foster City, CA, USA), including primers for 15 autosomal STR loci: D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, vWA, TPOX, D18S51, D5S818, FGA, D19S433 and D2S1338, by reducing the total reaction volume to 5 μ l. Amplicons were separated by multicapillary electrophoresis in an ABI 3100 genetic analyzer (Applied Biosystem). Results were then analyzed by GeneMapper software v.3.7 (Applied Biosystem).

Statistical analyses

STR allele frequencies, Hardy-Weinberg *P*-values, observed and expected heterozygosity (H_O and H_E , respectively), total allele and gene diversity indices were calculated with ARLEQUIN 3.1.²⁹ The observed heterozygosity, averaged within the populations, was tested for heterogeneity with the Kruskal-Wallis test,³⁰ using the commercially available STATISTICA 7.0 software package (StatSoft, Padova, Italy). The relative amount of gene flow into each population was calculated according to the model of Harpending and Ward (1982),³¹ which uses the theoretic regression of population heterozygosity and distance to the centroid. As these markers are widely used for forensic study, several parameters of forensic and population genetic importance including power of discrimination, matching probability, polymorphic information content, power of exclusion and typical paternity index were obtained with the Excel PowerStats spread sheet (www.promega.com/geneticidtools/powerstats).

To clarify population affinity, pair-wise genetic distances based on allele frequency variance (F_{st}), were computed by ARLEQUIN 3.1. The distance matrix was then plotted in two dimensions, by means of multidimensional scaling (MDS), using the STATISTICA 7.0 software.

The correlation among genetic and geographical distances was assessed by the Mantel test³² employing ARLEQUIN 3.1. Geographical distances between the approximate locations of each population were computed as great-circle distances calculated from their latitudinal and longitudinal coordinates.

One of the most important factors in admixture analysis is the level of divergence between the parental populations,³³ thus, both distance-based and

Table 1 General information of the studied populations

Populations	Code	Sample size	Linguistic affiliation		Location	Geographic coordination	
			Family	Subfamily		District, province	Latitude (° N)
Khon Mueang1	KM1	51	Tai-Kadai	Tai	Mueang, Chaing Rai	19° 58'	99° 51'
Khon Mueang2	KM2	42	Tai-Kadai	Tai	Mae Rim, Chiang Mai	19° 00'	98° 47'
Khon Mueang3	KM3	36	Tai-Kadai	Tai	Mae Tang, Chiang Mai	19° 09'	98° 55'
Khon Mueang4	KM4	50	Tai-Kadai	Tai	San Sai, Chiang Mai	19° 02'	98° 58'
Khon Mueang5	KM5	43	Tai-Kadai	Tai	San Kam Pang, Chiang Mai	18° 44'	99° 06'
Khon Mueang6	KM6	46	Tai-Kadai	Tai	San Pa Tong, Chiang Mai	18° 37'	98° 53'
Khon Mueang7	KM7	47	Tai-Kadai	Tai	Mae Cham, Chiang Mai	18° 29'	98° 21'
Khon Mueang8	KM8	45	Tai-Kadai	Tai	Ban Hong, Lamphun	18° 18'	98° 94'
Khon Mueang9	KM9	46	Tai-Kadai	Tai	Koh Ka, Lampang	18° 16'	99° 23'
Khon Mueang10	KM10	30	Tai-Kadai	Tai	Hang Chat, Lampang	18° 05'	99° 23'
Mon	MO	36	Austro-Asiatic	Mon-Khmer	Pa Sang, Lamphun	18° 31'	98° 53'
Lawa1	LW1	47	Austro-Asiatic	Mon-Khmer	Mae La Noi, Mae Hong Son	18° 23'	97° 56'
Lawa2	LW2	50	Austro-Asiatic	Mon-Khmer	Hod, Chiang Mai	18° 08'	98° 20'
Shan	SH	44	Tai-Kadai	Tai	Pang Ma Pa, Mae Hong Son	19° 37'	98° 14'
Lue1	LU1	51	Tai-Kadai	Tai	Pua, Nan	19° 09'	100° 56'
Lue2	LU2	41	Tai-Kadai	Tai	Tha Wang Pha, Nan	19° 05'	100° 47'
Khuen	KH	48	Tai-Kadai	Tai	Mae Wang and San Pa Tong, Chiang Mai	18° 38'	98° 51'
Yuan1	YU1	87	Tai-Kadai	Tai	Mae Taeng and SanSai, Chiang Mai	19° 00'	98° 59'
Yuan2	YU2	48	Tai-Kadai	Tai	Ban Hong, Lamphun	18° 24'	98° 45'
Yong	YO	55	Tai-Kadai	Tai	Pa Sang, Lamphun	18° 24'	98° 56'

model-based approaches were employed to investigate genetic differentiation among the potential parental populations. Pair-wise genetic distances based on allele frequency variances (F_{st}) were computed using ARLEQUIN 3.1. Significance levels for these values were adjusted by sequential Bonferroni correction.³⁴ Population structure was also investigated by means of the Bayesian clustering method implemented in STRUCTURE 2.3^{35–37} under assumptions of admixture, correlated allele frequencies and either with or without (LOCPRIOR model³⁷) assistance of sampling location as prior information. When the LOCPRIOR model was used, the value of the parameter r was also reported. An r value lower than 1, indicates that sampling location in the model is effective.³⁷ For each number of clusters (K) from 1 to 10, five runs were performed, using an MCMC chain burn-in length of 100 000 iterations followed by 1 000 000 iterations. Chain convergence was assessed by comparing results from five different chains. For each K , the posterior probability of clustering was estimated from the average logarithmic probability of data across runs. Finally, second order rate of change of logarithmic probability data between subsequent K values was estimated according to Evanno *et al*, 2005³⁸ to identify the optimal number of clusters in the data. Outputs from STRUCTURE were graphically modified by DISTRUCT.³⁹

Using results obtained from distance-based and model-based clustering method, significantly differentiated populations were chosen to represent source populations in the admixture calculation. ADMIX 2.0⁴⁰ was used to evaluate the estimator of admixture coefficient (m_Y) based on average coalescence times between pairs of genes sampled within and between populations.³³ This method can use either the molecular distance between alleles or assume that all alleles are equally divergent. By using allele frequencies, the estimated coefficients are less affected by the stochasticity of the mutation process and the short time-scale of human admixture process is influenced by genetic drift rather than accumulated mutation.^{18,41} Therefore, in this study, no divergence of alleles and no mutation model were used to estimate the genetic contribution of parental populations to the KM

RESULTS

Genetic diversity

Although a detailed analysis of allele frequencies was not the objective of this study, some interesting points were observed (Table 2). The

Table 2 Genetic diversities of 15 short tandem repeats for the 20 studied populations

	Total alleles	Average heterozygosity	Gene diversity
KM1	122	0.7360	0.7827 ± 0.3957
KM2	115	0.7476	0.7709 ± 0.3910
KM3	110	0.7482	0.7575 ± 0.3853
KM4	126	0.8173	0.7878 ± 0.3982
KM5	120	0.7504	0.7775 ± 0.3940
KM6	120	0.7812	0.7816 ± 0.3956
KM7	110	0.7645	0.7607 ± 0.3856
KM8	117	0.8030	0.7810 ± 0.3955
KM9	115	0.7884	0.7851 ± 0.3973
KM10	107	0.7689	0.7733 ± 0.3940
MO	113	0.7704	0.7900 ± 0.4009
LW1	103	0.7630	0.7510 ± 0.3809
LW2	114	0.7800	0.7669 ± 0.3883
SH	117	0.7485	0.7834 ± 0.3967
LU1	112	0.7817	0.7652 ± 0.3874
LU2	104	0.7756	0.7619 ± 0.3868
KH	114	0.7417	0.7589 ± 0.3846
YU1	126	0.7663	0.7807 ± 0.3931
YU2	119	0.7708	0.7825 ± 0.3959
YO	125	0.7685	0.7758 ± 0.3922

Abbreviations: GD, Gene diversity; KH, Khuen; KM, Khon Mueang; LU, Lue; LW, Lawa; MO, Mon; SH, Shan; YO, Yong; YU, Yuan.

average observed heterozygosity (H_o) range (0.7360 (KM1)–0.8173 (KM4)) was similar to that previously reported for many Thai populations.⁴² This range was not statistically different among all the populations ($\chi^2=12.5206$; $P=0.8622$; Kruskal–Wallis test). The highest average heterozygosity of KM4 may have been shaped by the gene flow process, while the minimum average heterozygosity, found in KM1, might be the result of genetic drift. To confirm the relative

effect of gene flow and genetic drift, a regression plot between average heterozygosity and distance from the centroid based on the model of Harpending and Ward (1982) reveals that the KM4 are the most displaced from the theoretical linear regression line, suggesting they have experienced the greatest amount of gene flow. KM1 were plotted lowest below from the line reflecting genetic drift, hence, less heterozygosity⁴³ (Supplementary Figure 1). Three loci; D18S51 in KM1, D16S539 in LU1 and TH01 in SH; deviated from Hardy–Weinberg equilibrium, after applying Bonferroni correction. Allele

frequency distributions, including several parameters such as power of discrimination, matching probability, polymorphic information content, power of exclusion and typical paternity index, are provided in supplementary materials (Supplementary Table 1–20).

Genetic affinity and genetic structure

To visualize the relationships among the populations, the MDS plot, derived from a distance matrix of F_{st} , is shown in Figure 2. The KM populations, except for KM7, are clustered together in the center of the MDS, with most of the Tai-speaking populations, indicating a close genetic relationship among Tai-speaking populations. The most likely explanation for the KM (except KM7) being located intermediately among the Tai-speaking populations is that the KM are admixed populations. For the outlier populations, LU2 and KM7 are the only 2, which separately plotted, from the 17 Tai populations, while the 3 Mon-Khmer populations (MO, LW1 and LW2) are segregated away from the center cluster as well as from each other, indicating high genetic differentiation. As found in the model-based clustering result, presented later in the text, the LW1, LW2 and LU2 initially emerge when the lower values of K are considered, confirming the great extent of genetic differentiation from any other populations (Figure 3).

The model of isolation by distance predicts an increase in genetic differentiation with geographical distance, thus, correlation between genetic and geographical distances is expected.⁴⁴ A Mantel test reveals an absence of correlation between genetic and geographical distances ($r=0.2727$, $P=0.077$), suggesting that factors in addition to geography describe the major trends of genetic variations among our populations.

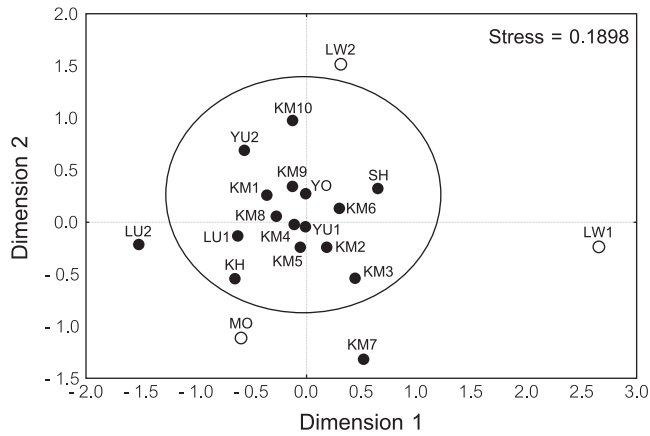


Figure 2 Multidimensional scaling plot of studied populations based on F_{st} (filled circles: Tai-speaking populations, blank circles: Mon-Khmer-speaking populations).

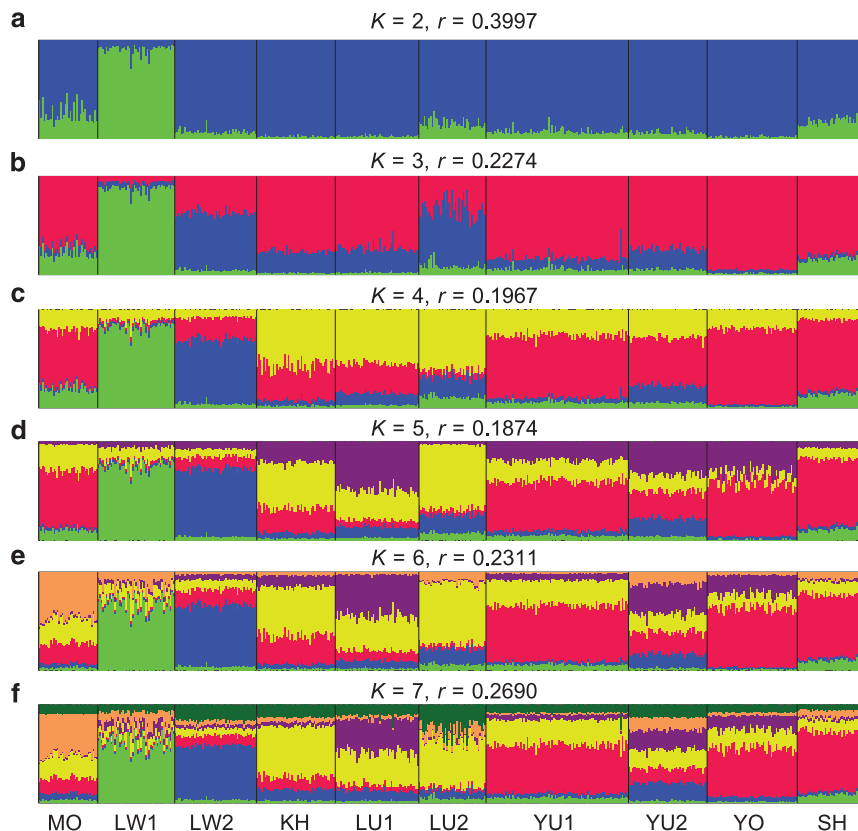


Figure 3 Bar plot estimation figures of 10 putative parental populations, with sequential K from 2 to 7 (a–f) and their r values, inferred from the STRUCTURE 2.3 analyses. These plots were produced by the average of five replications, using DISTRUCT.

Genetic differentiation of the parental populations

The 10 neighboring populations, which belonged to 7 ethnic groups, were considered as parental populations in admixture analysis, that is, indigenous Lawa (LW1 and LW2), the Mon (MO) and Tai-speaking populations (YU1, YU2, LU1, LU2, KH, YO and SH) which are ethnically, culturally and linguistically diversified through their historical migration and settlement. As the level of divergence between the parental populations is a crucial factor in admixture analysis,³³ therefore, the clustering analysis was performed to choose the highly diverged populations among these 10 populations.

Based on pair-wise F_{st} after adjusting the significance level according to the sequential Bonferroni method (Supplementary Table 21), KH-LU1, YU1-LU1, YU1-YO, YU1-SH and YO-SH were not significantly different. As STRUCTURE was unable to correctly identify the number of subpopulations at low levels of population differentiation ($F_{st} < 0.02$),⁴⁵ this limitation could be partially overcome by using the sampling location parameter of STRUCTURE.³⁷ Because of the low level of genetic divergence (F_{st} ranges from 0.0028 to 0.0285) among our putative parental populations, the cluster analysis with the assistance of sampling information, which is more informative than those without ($r_{MAX}=0.39$ in Figure 3), was chosen, and hence, only results from this model for $K=1-7$ are shown in Figure 3. By utilizing the *ad hoc* statistic ΔK , based on the rate of change in the log probability of data between successive K values,³⁸ a strong modal peak at $K=6$ was found (data not show). Both populations with multiple sources of ancestry as well as populations with ancestry predominantly derived from one of the inferred components, are presented. In each K , individuals from the same population with similar ancestry showed high within-population genetic homogeneity.

At $K=6$, three distinct clusters of the Mon-Khmer-speaking groups were observed, although the MO cluster was not clearly delineated. The green and blue components in Figure 3 belonged to the highly diverged LW1 and LW2, respectively, while the orange component separated the MO from other populations. However, the MO cluster appears equally mixed by the yellow and red components from the Tai. Among the Tai-speaking populations, three main components, the yellow, purple and red, prevailed (Figure 3). The yellow component was commonly observed in KH and LU2, whereas purple characterized the LU1 population although LU1 was equally mixed with yellow. Finally, the red component distinguished the YU1, YO and SH from other populations. The YU2 seems to be more admixed with various Mon-Khmer and Tai components.

Admixture estimation

Based on genetic distance and STRUCTURE results, six highly differentiated populations were selected and divided into two linguistic groups; the Mon-Khmer (LW1, LW2 and MO) and the Tai (LU1, LU2 and YO), to represent variation present in parental populations.

The admixture proportions in different KM populations (Table 3) revealed that, when examining the genetic contribution from two parental populations, the Mon-Khmer contributed a lower proportion to almost all KM than did the Tai (Table 3 and Figure 4). Among the Mon-Khmer groups, the MO and LW2 contributed more genes to most of the KM than the LW1, which agrees with other historical evidence.² Focusing on the contribution of the parental Tai, the YO arise as the major genetic contributor to most of the KM except KM3, KM5 and KM7, which have been more influenced by the LU. Variation in the KM genetic admixture patterns was observed. The study of admixture patterns suggested that gene flow from Tai-speaking populations to the KM was influenced more by the mass migration around 1800 AD, than by the migration in 1300 AD.

DISCUSSION

This study aimed to analyze the population affinity among the present-day KM and their neighboring populations, and the genetic admixture of KM along the historical Yuan migration route by using a battery of 15 highly informative bi-parental STR loci. We hypothesize that the admixed origin of the KM was from the neighboring Mon-Khmer and Tai populations. In general, the admixed populations show genetic characteristics that are intermediate between the parental populations, although some admixed populations did not show this characteristic because of the genetic drift, as it was observed in previous reports.^{33,46} The genetic diversity of all KM (Table 2) were within the same range as those of the neighboring populations, except for the average heterozygosity values of KM4 and KM1 which can be explained by the effect of gene flow and genetic drift, respectively. The MDS plot derived from allele frequency variance revealed that the KM populations, except for KM7, were located intermediately among the Tai-speaking populations (Figure 2), supporting their admixed origin.

The MDS plot also showed that the KM are more closely related to the neighboring Tai than the Mon-Khmer. The clustering as shown in the center of the plot between KM and their neighboring Tai indicated the linguistic population grouping in the Tai, but not in the Mon-Khmer-speaking groups. No clear geographical population grouping in the studied populations, and the absence of a correlation between genetics and geography was also observed. This implies that geographical factors have had less influence in shaping genetic variation in these populations.

The model-based clustering method, implemented in STRUCTURE, has been recognized as a powerful tool in characterizing population differentiation.⁴⁵ We applied it here to estimate the extent of genetic differentiation among the parental populations. Generally speaking, samples from the Mon-Khmer groups (MO, LW1 and LW2), were remarkably genetically differentiated from each other as well as from the Tai-speaking populations. A possible explanation relies on their history and geographical location. The MO are more admixed than the LW, as shown by the presence of the Tai components (yellow and red in Figure 3). The MO geographical location, which was closer to the Tai than the LW might be served as a reason why the MO gene pool was highly admixed (Figure 1). The western Highland LW (LW1) have been living in thickly forested mountains in Mae Hong Son province and are characterized by limited mobility and the strong preservation of their LW identity.⁴⁷ Hence, inbreeding might have occurred, as shown by their low genetic variation, reflected by total allele and diversity indices values (Table 2). The eastern Lowland LW (LW2) seem to have had some contacts with the YU, as shown by the presence of some components (yellow and red in Figure 3) typical to the YU, as they used to live in the city of Chiang Mai before they migrated to the Bo Luang plateau in Hod, Chiang Mai province. Because of linguistic and geographical barriers, this might possibly lead to restricted genetic exchange among both of the LW groups.

Among the seven Tai-speaking populations, there were three primary clusters identified by STRUCTURE. Surprisingly, despite the geographical proximity of the LU1 and LU2, they appear to form two weakly differentiated clusters. Both LU populations also have shared genetic components (yellow in Figure 3), but LU2 seems to have maintained a degree of genetic identity by social isolation, which can be seen from their location in the MDS plot and is mirrored in the low number of total alleles and a low diversity indices value (Table 2 and Figure 2), whereas LU1 might have mixed with other Tai during their migration through Laos.⁷

The study of admixture patterns may reflect our understanding of historical and anthropological aspects of demographic migrations.²⁰

Table 3 Weighted average across loci and s.d. of the estimated contributions of parental populations to the 10 studied Khon Mueang populations

Parental populations	Khon Mueang populations									
	KM1	KM2	KM3	KM4	KM5	KM6	KM7	KM8	KM9	KM10
<i>Two groups</i>										
<i>Mon-Khmer</i>										
Average	36.59	28.67	18.94	40.14	34.02	59.60	47.50	45.11	56.54	46.39
s.d.	23.75	17.93	23.29	28.89	17.07	20.71	19.37	22.40	25.42	22.79
<i>Tai</i>										
Average	63.41	71.33	81.06	59.86	65.98	40.40	52.50	54.89	43.46	53.62
s.d.	23.75	17.93	23.29	28.89	17.07	20.71	19.37	22.40	25.42	22.79
<i>Six groups</i>										
<i>MO</i>										
Average	27.38	4.09	-4.77	33.15	20.44	32.05	26.54	25.06	33.71	15.4
s.d.	17.69	15.19	17.6	21.9	14.8	16.45	15.28	17.42	18.34	18.17
<i>LW1</i>										
Average	-3.12	13.17	21.74	3.53	-1.01	0.3	14.47	9.28	1.53	-2.91
s.d.	11.51	9.68	11.91	15.03	9.57	10.65	11.12	11.5	12.01	11.67
<i>LW2</i>										
Average	9.84	10.09	6.12	3.34	18.16	27.58	9.8	6.29	19.97	36.89
s.d.	16.6	12.88	16.12	22.47	12.76	13.39	14.21	16.08	16.67	16.47
<i>LU1</i>										
Average	2.45	19.43	46.34	9.16	28.54	1.35	22.29	8.61	4.9	15.83
s.d.	28.68	21.94	28.54	42.06	20.66	23.58	24.32	28.19	29.88	28.52
<i>LU2</i>										
Average	12.2	11.75	12.67	0.21	2.24	-7.54	3.36	3.12	-7.84	-7.19
s.d.	20.38	15.98	21.93	27.75	14.94	16.61	17.27	20.57	21.56	20.37
<i>YO</i>										
Average	51.25	41.46	17.9	50.6	31.63	46.25	23.55	47.65	47.73	41.98
s.d.	28.02	20.89	26.49	39.67	19.18	21.26	22.01	26.57	27.37	25.01

As already mentioned, there are three main geographical regions along the YU migratory route. The Chiang Rai basin was inhabited by the native LW and it was later occupied by the YU.⁴ Although the MO have never been reported as having lived in this area, the KM in Chiang Rai (KM1) seem to show a detectable genetic contribution from the MO. The recent backward migration of KM1 from Chiang Mai to Chiang Rai might explain this phenomenon.

The demographic scenario in the Chiang Mai-Lamphun basin is more complex. Before the YU movement, the upper and lower parts of this basin were previously dominated by the LW and MO respectively, in different periods.⁴ Thus, the admixture results were in agreement with ethno-historical sources. The KM, residing in the upper and lower parts of the Chiang Mai-Lamphun basin, had rather high admixture proportions from the LW and MO, respectively. In the plain of the Mae Cham basin, where the KM (KM7) migrated from Chiang Mai 200 years ago, the KM7 displayed a contribution from all parental groups similar to other KM samples. However, as shown in the MDS plot (Figure 2), KM7 has greater genetic divergence relative to other KM samples. The distinct gene pool of the Mae Cham population might be explained by admixture between them and other

surrounding populations. As there are many populations, particularly the hill-tribes, living in the Mae Cham basin, admixture estimation may not be reliable. The hill-tribes might be parental sources for KM7, which would make them highly differentiated, but unfortunately there is no such genetic data set available to analyze.

Different admixture patterns were observed in two KM villages from Lampang. KM9 inevitably shows the highest fraction of MO admixture, who ruled the region between 750 and 1300 AD,³⁻⁴ whereas the KM10 show the highest contribution from the native LW. The KM10, who migrated from the city of Chiang Saen, settled in the present Chiang Rai province, within the last 200 years.⁴⁸ They could have mixed with the LW in Chiang Rai before their migration and then later mixed with the MO in Lampang.

The admixture observations in this study are largely consistent with historical records from northern Thailand. The variability in the KM genetic admixture patterns even among villages located in the same geographical region indicates a high heterogeneity among KM populations (Figure 5). This might reflect variation in the timing of village formation, and the parental populations of the region, when admixture started, had a greater influence on shaping the current gene pool

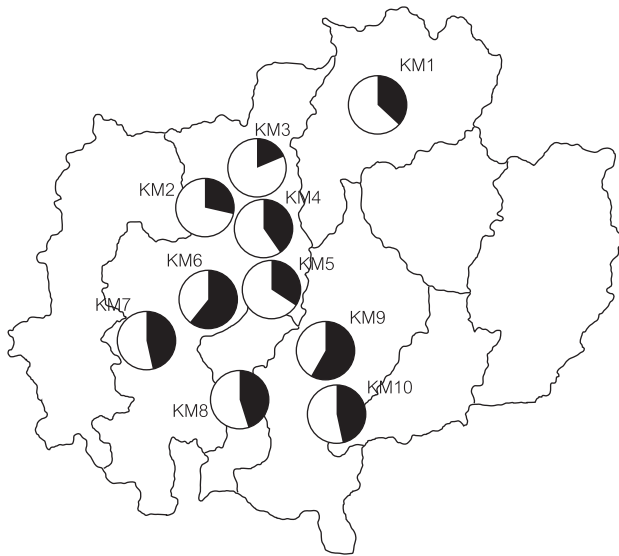


Figure 4 Pie diagrams showing the distribution of Mon-Khmer-speaking population (black) and Tai-speaking population (white) contributing to the Khon Mueang samples.

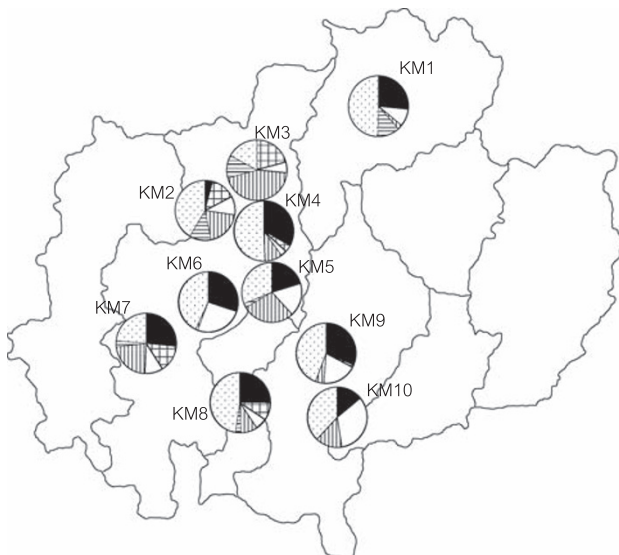


Figure 5 Pie diagrams showing the distribution of MO (black), LW1 (checkered), LW2 (white), LU1 (vertical stripe), LU2 (horizontal stripe) and YO (dot) contributing to the Khon Mueang samples.

of the KM, than geographical factors. An earlier genetic study suggested that the KM migrated from southern or south-eastern China.⁷ Several populations lived in proximity in an area of southern China, leading to extensive genetic exchange among them.^{19–21,49–51} Although we were unable to look for admixture and affinities with the modern descendants of the KM ancestral populations from outside Thailand, this study did reveal genetic admixture among the KM in northern Thailand. Moreover, the present results from our admixture analysis are consistent with the suggestion from previous studies by the same group,⁷ using a battery of 15 highly informative bi-parental STR loci. Further study using genetic markers with different modes of inheritance are needed to gain a deeper understanding of genetic admixture and the population history of the KM.

ACKNOWLEDGEMENTS

We thank all blood donors and village chiefs for their participation. We are also grateful to the Population Genetics and Genetic Epidemiology, the Genome Institute of Singapore, Singapore for genotyping support, as well as for training in statistical analyses from Dipartimento di Biologia ed Evoluzione, Università di Ferrara, Ferrara, Italy. This study was supported by Royal Golden Jubilee Ph.D. program (Grant No. PHD/0243/2547).

- Schliesinger, J. *Ethnic Groups of Thailand Non-Tai-Speaking Peoples*, (White Lotus Press: Bangkok, 2000).
- Schliesinger, J. *Tai group of Thailand, Volume 1: Introduction and overview*, (White Lotus Press: Bangkok, 2001).
- Penth, H. *A brief history of Lanna: civilizations of north Thailand*, (Silkworm Books: Chiang Mai, 2000).
- Ongsakul, S. *History of Lan Na*, (Silkworm Books: Chiang Mai, 2005).
- Condominas, G. *From Lawa to Mon, from Saa' to Thai*, (The Australian national University: Canberra, 1990).
- Gordon, R. G. *Ethnologue: Languages of the World*, 15th edn (SIL International: Dallas, 2005).
- Kampanasai, J., Bertorelle, G., Castri, L., Nakbunlung, S., Seielstad, M. & Kangwong, D. Mitochondrial DNA variation of Tai speaking peoples in Northern Thailand. *Sci. Asia* **33**, 443–448 (2007).
- Bhumisak, J. *The origin of Siam, Tai, Laos and Khmer* (in Thai) (Textbook project foundation: Bangkok, 1990).
- Charoenmuang, T. *Khon Mueang*, (Local government studied project) Faculty of Social Sciences, Chiang Mai University: Chiang Mai, 2001.
- Nakbunlung, S. *Origins and biological affinities of the modern Thai population: an osteological perspective*, (Ph.D Thesis, University of Illinois: Urbana, 1994).
- Poolswan, S. *Science and the reality in Thai culture* (in Thai) (Kob Fi: Bangkok, 2001).
- Parra, E. J., Kittles, R. A., Argyropoulos, G., Pfaff, C. L., Hiester, K., Bonilla, C. *et al*. Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. *Am. J. Phys. Anthropol.* **114**, 18–29 (2001).
- Parra, E. J., Marcini, A., Akey, J., Martinson, J., Batzer, M. A., Cooper, R. *et al*. Estimating African American admixture proportions by use of population-specific alleles. *Am. J. Hum. Genet.* **63**, 1839–1851 (1998).
- Bortolini, M. C., Da Silva, W. A., Junior, W., De Guerra, D. C., Remonato, G., Miranda, R. *et al*. African-derived South American populations: a history of symmetrical and asymmetrical matings according to sex revealed by bi- and uni-parental genetic markers. *Am. J. Human. Biol.* **11**, 551–563 (1999).
- Sans, M., Weimer, T. A., Franco, M. H., Salzano, F. M., Bentancor, N. & Alvarez, I. Unequal contributions of male and female gene pools from parental populations in the African descendants of the city of Melo, Uruguay. *Am. J. Phys. Anthropol.* **118**, 33–44 (2002).
- Sans, M. Admixture studies in Latin America: From the 20th to the 21st century. *Hum. Biol.* **72**, 155–177 (2000).
- Dupanloup, I., Bertorelle, G., Chikhi, L. & Barbujani, G. Estimating the impact of prehistoric admixture on the genome of Europeans. *Mol. Biol. Evol.* **21**, 1361–1372 (2004).
- Belle, E. M., Landry, P. A. & Barbujani, G. Origin and evolution of the Europeans's genome: evidence from multiple microsatellite loci. *Proc. R. Soc. B.* **273**, 1595–1602 (2006).
- Yao, Y. G., Lu, X. M., Luo, H. R., Li, W. H. & Zhang, Y. P. Gene admixture in the silk road of China: evidence from mtDNA and melanocortin 1 receptor polymorphism. *Genes Genet. Syst.* **75**, 173–178 (2000).
- Wen, B., Xie, X., Gao, S., Li, H., Shi, H., Song, X. *et al*. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *Am. J. Hum. Genet.* **74**, 856–865 (2004).
- Wen, B., Li, H., Lu, D., Song, X., Zhang, F., He, Y. *et al*. Genetic evidence supports demic diffusion of Han culture. *Nature* **431**, 302–305 (2004).
- Zhivotovsky, L. A., Underhill, P. A., Cinnioçlu, C., Kayser, M., Morar, B., Kivisild, T. *et al*. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am. J. Hum. Genet.* **7**, 50–61 (2004).
- Li, S. L., Yamamoto, T., Yoshimoto, T., Uchihi, R., Mizutani, M., Kurimoto, Y. *et al*. Phylogenetic relationship of the populations within and around Japan using 105 short tandem repeat polymorphic loci. *Hum. Genet.* **118**, 695–707 (2006).
- Deka, R., Shriver, M. D., Yu, L. M., Heidreich, E. M., Jin, L., Zhong, Y. *et al*. Genetic variation at 23 microsatellite loci in 16 human populations. *J. Genet.* **78**, 99–121 (1999).
- Reddy, B. M., Naidu, V. M., Madhavi, V. K., Thangaraj, K., Langstieh, B. T., Venkataramana, P. *et al*. STR data for the AmpF/STR Profiler Plus loci among 27 population of different social hierarchy from southern part of Andhra Pradesh, India. *Forensic Sci. Int.* **149**, 81–97 (2005).
- Pandu, G., Gandhi, K. P., Sharma, J. D., Chaubey, G. & Thangaraj, K. Genetic profile of nine STR loci among Goud and Padmashali populations of Andhra Pradesh, India. *Forensic Sci. Int.* **157**, 201–205 (2006).

- 27 Thangaraj, K., Chaubey, G., Singh, V. K., Reddy, A. G., Pavate, P. P. & Singh, L. Genetic profile of nine autosomal STR loci among Halakki and Kunabhi populations of Karnataka, India. *J. Forensic Sci.* **51**, 190–192 (2006a).
- 28 Sielstad, M., Bekele, E., Ibrahim, M., Toure, A. & Traore, M. A view of modern human origins from Y chromosome microsatellite variation. *Genome Res.* **9**, 558–567 (1999).
- 29 Excoffier, L., Laval, G. & Schneider, S. Arlequin ver. 3.0: an integrated software package for population genetic data analysis. *Evol. Bioinform. Online* **1**, 47–50 (2005).
- 30 Zar, J. H. *Biostatistical Analysis*, 4th edn (Prentice Hall: New Jersey, 1999).
- 31 Harpending, H. C. & Ward, R. In: *Biochemical Aspects of Evolutionary Biology*, (eds Nitecki M.) 213–256, (University of Chicago Press: Chicago, 1982).
- 32 Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220 (1967).
- 33 Bertorelle, G. & Excoffier, L. Inferring admixture proportions from molecular data. *Mol. Biol. Evol.* **15**, 1298–1311 (1998).
- 34 Rice, W. R. Analyzing tables of statistical tests. *Evolution* **43**, 223–225 (1989).
- 35 Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
- 36 Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).
- 37 Hubisz, M., Falush, D., Stephens, M. & Pritchard, J. Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resources* **9**, 1322–1332 (2009).
- 38 Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
- 39 Rosenberg, N. A. DISTRUCT: a program for the graphical display of population structure. *Mol. Ecol. Notes* **4**, 137–138 (2003).
- 40 Dupanloup, I. & Bertorelle, G. Inferring admixture proportions from molecular data: Extension to any number of parental populations. *Mol. Biol. Evol.* **18**, 672–675 (2001).
- 41 Barbujani, G. DNA variation and language affinities. *Am. J. Hum. Genet.* **61**, 1011–1014 (1997).
- 42 Listman, J. B., Malison, R. T., Sughondhabiro, A., Yang, B. Z., Raaum, R. L., Thavichachart, N. *et al.* Demographic changes and marker properties affect detection of human population differentiation. *BMC Genet.* **8**, 21 (2007).
- 43 Krithika, S., Maji, S. & Vasulu, T. S. Geographic Contiguity, Patterns of Gene Flow and Genetic Affinity among the Tribes of Arunachal Pradesh, India. *Int J Hum Genet.* **7**, 267–276 (2007).
- 44 Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The history and geography of human genes*, (Princeton University Press: Princeton, 1994).
- 45 Latch, E. K., Dharmarajan, G., Glaubitz, J. C. & Rhodes Jr, O. E. Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conserv. Genet.* **7**, 295–302 (2006).
- 46 Di Benedetto, G., Ergüven, A., Stenico, M., Castri, L., Bertorelle, G., Togan, I. *et al.* DNA Diversity and Population Admixture in Anatolia. *Am J Phys Anthropol* **115**, 144–156 (2001).
- 47 Nahhas, R. W. *Sociolinguistic survey of Lawa in Thailand*, (Survey Unit Department of Linguistics Faculty of Humanities Payap University: Chiang Mai, 2007).
- 48 Princess Maha Chaki Siridhorn anthropology center, 2005, [<http://www.sac.or.th/ethnic>].
- 49 Li, H., Cai, X., Winograd-Cort, E. R., Wen, B., Cheng, X., Qin, Z. *et al.* Mitochondrial DNA diversity and population differentiation in southern East Asia. *Am J Phys Anthropol.* **134**, 481–488 (2007).
- 50 Su, B., Xiao, C. & Jin, L. Genetic history of ethnic populations in southwestern China. *Recent Advances in Human Biology.* **8**, 57–67 (2001).
- 51 Li, H., Wen, B., Chen, S. J., Su, B., Pramoongjago, P., Liu, Y. *et al.* Paternal genetic affinity between western Austronesians and Daic populations. *BMC Evol Biol.* **8**, 146 (2008).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)