

ORIGINAL ARTICLE

Effect of migration patterns on maternal genetic structure: a case of Tai–Kadai migration from China to Thailand

Jatupol Kampaunsai¹, Wibhu Kutanan², Francesca Tassi³, Massupa Kaewgahya¹, Silvia Ghirotto³ and Daorong Kangwanpong¹

The migration of the Tai–Kadai speaking people from southern China to northern Thailand over the past hundreds of years has revealed numerous patterns that have likely been influenced by routes, purposes and periods of time. To study the effects of different migration patterns on Tai–Kadai maternal genetic structure, mitochondrial DNA hypervariable region I sequences from the Yong and the Lue people having well-documented histories in northern Thailand were analyzed. Although the Yong and Lue people were historically close relatives who shared Xishuangbanna Dai ancestors, significant genetic differences have been observed among them. The Yong people who have been known to practice mass migration have exhibited a closer genetic affinity to their Dai ancestors than have the Lue people. Genetic heterogeneity and a sudden reduced effective population size within the Lue group is likely a direct result of the circumstances of the founder effect.

Journal of Human Genetics (2017) 62, 223–228; doi:10.1038/jhg.2016.112; published online 8 September 2016

INTRODUCTION

Migration is a microevolutionary process that influences genetic variations within populations along with the differences that exist among them. Human migrations have been recorded throughout world history in all regions for various reasons, and have led to a variety of changes in population characteristics with regard to socio-cultural patterns, economies and genetic structures. When people move, their genetic traits are carried along with them, and then begin to change as a result of evolutionary forces such as drift and admixture.¹ The prevalence of some diseases is rather high among ethnic people who face the impact of migration and the founder effects, for example, the Usher syndrome type 2 among the French Canadian people and Alzheimer disease among the Amish populations.^{2,3}

The exemplified migratory scenario in Southeast Asia involves the migration of Tai–Kadai speaking populations. Southern China is known to have been the homeland of these people since prehistoric times.⁴ During the first millennium AD, the expansion of the powerful Han Chinese people prompted some groups of Tai–Kadai to migrate southward. Their gradual inland migration from southern China followed the course of rivers so as to allow them to find suitable conditions for their rice production. Eventually, the Tai–Kadai immigrants settled in northern Thailand and then became dominant over the native Austro–Asiatic speaking populations from the 13th century AD. Since then, the Tai–Kadai intracontinental migration

across southern China to northern Thailand has continued, driven by various political, cultural, environmental and other relevant factors.⁵

Among various Tai–Kadai speaking populations who have migrated from southern China into northern Thailand, the majority is the Lue or the Xishuangbanna Dai people. During the last few hundred years, some small groups of Lue migrated southward along the Mekong river and into northern Thailand via the Thai/Laos border. The migration patterns of the Lue people in each village were complex because they adopted various migration routes for different purposes over different periods of time. Some groups sought new settlements that would be suitable for agricultural cultivation, but some left their former homes because of political conflict or civil war. Currently, there are many Lue villages located in northern Thailand, for example, in the provinces of Nan, Phayao and Chiang Mai.⁶

One original homeland of the Lue people is located in the Yong city, formerly belonged to Xishuangbanna, China, but is now officially in the Shan state, Myanmar. In 1805 AD, after the period of Burmese decolonization, about 10 000 Lue people from Yong City, both royal family members and local residents, moved their households and resettled in the Pa Sang District of Lamphun Province in northern Thailand, following the King of Siam's (Thailand) resettlement campaign.⁷ After the period of migration, this Lue population defined themselves by a new ethnic name, Yong, to show their voluntary mass migration and distinction from the migration history of the other Lue people. Over a 200-year resettlement period, the Yong dispersed to

¹Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand; ²Department of Biology, Faculty of Science, Khon Kaen University, Khon Kaen, Thailand and ³Department of Life Science and Biotechnology, University of Ferrara, Ferrara, Italy
Correspondence: Dr J Kampaunsai, Department of Biology, Faculty of Science, Chiang Mai University, 239 Huay Kaew Road, Suthep, Muang Chiang Mai, Chiang Mai 50202, Thailand.

E-mail: jatupol.k@cmu.ac.th

Received 3 June 2016; revised 25 July 2016; accepted 30 July 2016; published online 8 September 2016

Table 1 General information of the studied populations and summary statistics

| Population | Subdistrict, district, province | N | H | π | F (prob.) ^a | R ^b |
|---------------------|----------------------------------|----|---------------|---------------|------------------------|----------------|
| Yong1 | Pa Sang, Pa Sang, Lamphun | 31 | 0.981 ± 0.014 | 0.020 ± 0.011 | -12.67 (0.000) | 0.008 |
| Yong2 | Pak Bong, Pa Sang, Lamphun | 40 | 0.978 ± 0.010 | 0.018 ± 0.010 | -13.90 (0.000) | 0.019 |
| Yong3 | Tha Pla Duk, Mae Tha, Lamphun | 46 | 0.935 ± 0.024 | 0.024 ± 0.012 | 7.16 (0.019) | 0.033 |
| Yong4 | Huay Sai, Ban Thi, Lamphun | 41 | 0.974 ± 0.013 | 0.019 ± 0.010 | -14.20 (0.000) | 0.019 |
| Yong5 | Law Yaw, Ban Hong, Lamphun | 49 | 0.973 ± 0.010 | 0.021 ± 0.011 | -12.80 (0.000) | 0.013 |
| Yong6 ¹⁰ | Makok, Pa Sang, Lamphun | 62 | 0.965 ± 0.009 | 0.021 ± 0.011 | -10.41 (0.000) | 0.013 |
| Lue1 ¹⁰ | Hia, Pua, Nan | 51 | 0.915 ± 0.027 | 0.018 ± 0.010 | -5.34 (0.044) | 0.033 |
| Lue2 ¹⁰ | Nong Bua, Tha Wang Pha, Nan | 44 | 0.878 ± 0.026 | 0.020 ± 0.010 | 0.49 (0.639) | 0.174 |
| Lue3 ¹⁰ | Koh Chang, Mae Sai, Chiang Rai | 50 | 0.988 ± 0.007 | 0.021 ± 0.011 | -29.24 (0.000) | 0.007 |
| Lue4 ¹⁰ | Luang Nua, Doi Saket, Chiang Mai | 46 | 0.932 ± 0.020 | 0.018 ± 0.010 | -2.88 (0.187) | 0.031 |
| Dai1 ⁴ | —, Xishuangbanna, Yunnan | 21 | 0.995 ± 0.017 | 0.022 ± 0.012 | -12.68 (0.000) | 0.020 |
| Dai2 ¹¹ | —, Jinghong, Yunnan | 38 | 0.994 ± 0.008 | 0.021 ± 0.011 | -29.59 (0.000) | 0.015 |
| Dai3 ¹² | —, Jinghong, Yunnan | 56 | 0.996 ± 0.004 | 0.018 ± 0.010 | -52.61 (0.000) | 0.016 |

Abbreviations: π , nucleotide diversity; F, Fu's *F_s* statistic; H, haplotype diversity; N, number of samples; R, raggedness index.

^aBold letter: statistically significant at $P < 0.01$.

^bBold letter: raggedness index < 0.03 .

many areas of northern Thailand where most of them continue to reside and maintain their unique culture, tradition and dialect.⁸

Because of the striking importance of the migration patterns on the genetic structure, we have conducted a descriptive study on the maternal genetic structure of the Yong and Lue people to answer two main questions. First, is there any genetic differentiation between the Yong and Lue that is owing to their different migration patterns? Second, how did their ancestors, the Dai people in southern China, contribute their genetic history to these two ethnic groups?

MATERIALS AND METHODS

With informed consent, we collected 5 ml of peripheral blood samples from 207 unrelated individuals belonging to five Yong populations from northern Thailand (Table 1), using anticoagulant-EDTA vacutainers. Information on linguistics, cultural aspects, village history and personal lineage was also obtained through interviews. Total genomic DNA was extracted according to a standard inorganic salting out protocol.⁹

Mitochondrial DNA (mtDNA) D-loop fragment was amplified using published primer pairs.¹³ The purified PCR product was sequenced for hypervariable region I (HVR-I) as previously described^{13,14} by Macrogen, South Korea. The 360 bp HVR-I sequences (position 16024–16363) of five Yong populations together with another set of data on the Yong population from a previous study,¹⁰ were edited and aligned against the Cambridge Reference Sequence¹⁵ using SeqScape software v2.0 (Applied Biosystem, Foster City, CA, USA). DNASP v5 software¹⁶ was used to identify polymorphic sites and mtDNA haplotypes.

Considering the shared ancestral linkages between Yong, Lue and Xishuangbanna Dai people, a joint data set of the same length mtDNA HVR-I from four Lue and three Dai populations (Table 1) was integrated into the analysis. Haplotypes among the studied populations were counted by a simple counting scheme. The 'unique type' was identified for a haplotype that was only found within a given population, but not in the other populations. The haplotype shared between possible pairs of the Yong people, but was not observed in Lue or Dai people, was designated as the 'Yong private type', and the same criterion was applied to identify the 'Lue private type' as well as the 'Dai private type'. The haplotype shared between one of the possible different ethnic group pairs was defined as the 'ethnic-sharing type'.

Signatures of population demographic changes that were owing to bottleneck or expansion scenarios were examined by two different methods. First, mismatch distribution analysis under the constant size population model together with the raggedness index and the neutrality Fu's *F_s* statistic¹⁷ were computed using the Arlequin 3.5 package.¹⁸ Second, Bayesian skyline plots¹⁹ were employed using the MCMC algorithm²⁰ by the program BEAST version 1.8.²¹ The jModel test 2.1.7²² was employed to choose the suitable model for

each population for creating input files of BEAST by BEAUTi v1.8. A strict molecular clock with a fixed rate of 1.64×10^{-7} substitutions per site per year was applied.²³ Each MCMC sample was run for 1×10^7 steps, sampled every 1000 steps, with the first 1×10^6 steps regarded as the burn-in. Tracer 1.5 (<http://tree.bio.ed.ac.uk/software/tracer>) was used to visualize the Bayesian skyline plot.

A hierarchical analysis of molecular variance²⁴ was performed by Arlequin 3.5 to examine the genetic structure in three levels, that is, (1) within a population, (2) among populations within each ethnic group (Yong, Lue and Dai) and (3) among ethnic groups. To summarize the diversity of the analyzed populations, a Discriminant Analysis of Principal Component (DAPC) was applied using the *dapc* function within the *adeigenet* R package.²⁵ The DAPC is a multivariate method, free of assumptions with regard to Hardy-Weinberg equilibrium or linkage disequilibrium.²⁶ To assess the structuring patterns, the discriminant functions were calculated for three groups, corresponding to the Yong, Lue and Dai people. The first two discriminant functions were plotted with the scatter function of the *adeigenet* R package. Genetic distances between populations were estimated using Φ statistics based on the pairwise differences identified between haplotypes using the Arlequin 3.5 package. The genetic distance matrix was then used to plot the multidimensional scaling, performed by STATISTICA 10.0 (Statsoft, Tulsa city, OK, USA).

RESULTS

Among mtDNA HVR-I sequences of a total of 269 samples, 118 distinct haplotypes, defined by 89 variable sites, were identified (Supplementary Figure S1). When the published Lue and Dai data were integrated into an analysis, a total of 234 haplotypes were identified. Among them, 172 haplotypes have been acknowledged as unique to their specific populations. The percentages of unique haplotypes were relatively low in the Lue4 (26.3%) and Yong5 (26.7%) populations, whereas in the other samples, the percentages were higher, ranging from 35.7 to 85.0%. There were 18 private haplotypes that were shared only in their specific ethnic groups. The percentages of the private types were higher in most Yong populations than in the Lue and Dai populations. There were 44 haplotypes shared between two or more different ethnic groups (Figure 1).

Dai3 showed the highest haplotype diversity (0.996 ± 0.004), whereas the lowest value was observed in that of Lue2 (0.878 ± 0.026). In general, the differences of haplotype diversities within the Yong and Dai groups varied within narrow ranges, whereas those values fluctuated among different Lue villages (Table 1). When we pooled populations of the same ethnic group together, the haplotype diversity was the highest among the Dai (0.996 ± 0.002),

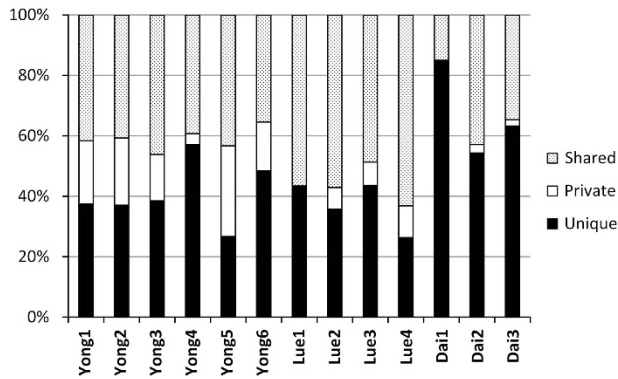


Figure 1 Percentage of haplotype observed in each population.

followed by the Yong (0.986 ± 0.002) and Lue (0.975 ± 0.004). At the nucleotide level, diversities ranged from 0.024 ± 0.012 in the Yong3 to 0.018 ± 0.010 in the Yong2, Lue1, Lue4 and Dai3 populations (Table 1). Each ethnicity (Yong, Lue and Dai) showed the same overall nucleotide diversities (0.020 ± 0.010), which corresponded with previous estimations in the Dai, Zhuang and Thai populations.¹¹

To reveal the signatures of the population demographic changes, we performed mismatch distribution analysis based on the fact that the smoothness of the unimodal mismatch distribution and small raggedness values (<0.03) indicated population expansion, but in a converse trend, that is, multimodal distribution and high raggedness values reflected stationary or shrinkage populations.^{27–29} We detected population expansion signals in the pool samples of each ethnic group (Figure 2). When we consider population expansion in each population, there are four populations, that is, Yong3, Lue1, Lue2 and Lue4 that show multimodal mismatch distributions, whereas the remaining populations have unimodal distributions (Supplementary Figure S2). The F_s statistics agreed well with the mismatch analysis, with insignificant test results ($P > 0.01$) for all multimodal populations (Figure 2 and Table 1). The Bayesian skyline plots show a continuous population expansion in the Dai population and it is likely to have grown rapidly over the last 15 000 years (kya). The Yong and Lue's effective population sizes gradually expanded during the Paleolithic and Neolithic periods, and then began to lessen around 5 kya (Figure 2). Recently sharp reductions (~ 1 – 2 kya) have been observed in all Lue populations, except Lue3, which has been unchanged in population size (Supplementary Figure S3).

We have investigated genetic variations of the Yong, Lue and Dai data sets by analysis of molecular variance. When all 13 populations were grouped together, 96.68% of the genetic variations within the populations were found, whereas 3.32% of the variations were estimated to be among them. The genetic variations between populations of the Lue ethnic group (5.15%) were larger than those of the Yong (1.97%) or Dai (2.00%) populations. The proportion of genetic variations attributed to the differences among the 3 ethnic groups is 0.45% with nonsignificance ($P > 0.05$). However, when we consider each pair of the ethnic groups, Yong/Lue is found to be significantly different, whereas Yong/Dai and Lue/Dai do not differ (Table 2). The overlapping ethnic clusters in the DAPC plot points out that the genetic variations between groups were not sufficient enough to recognize the Yong, Lue or Dai ethnic groups (Figure 3).

The multidimensional scaling biplot shows the genetic clustering of populations belonging to the Yong and Dai ethnic groups, which is distant from the heterogeneous Lue populations. Only the Lue3 is closely related to the Yong/Dai cluster (Figure 4).

DISCUSSION

From their homeland in southern China, over hundreds of years, the Tai–Kadai people gradually migrated southward and became the majority population of northern Thailand. The demic diffusion of the Tai–Kadai language is substantiated by linguistic sharing that occurs between the Tai–Kadai speaking people in China and Thailand.³⁰ Although, their accents have changed somewhat owing to the temporal and spatial factors, both of them still share many words with the same meaning and similar pronunciations. In this research study, we have explored the effects of migration patterns on the genetic structure of two close linguistic relatives, the Yong and Lue ethnic groups, from northern Thailand, using mtDNA HVR-1 variations. The results show a significant genetic difference between the ethnic groups of Yong and Lue. The large ‘within group’ variation of the Lue people (Table 2), together with their high distinction in multidimensional scaling (Figure 4), reveals genetic heterogeneity in the Lue groups who inhabited different areas. This was possibly a result of the founder effect that shows the impact of a sudden reduction of the effective population size (Figure 2) and the shaping of Lue's genetic structure (Figure 4). Each small group of migrated Lue people resettled in northern Thailand, and therefore was cut off from their parental stock and each other by geographic barriers such as hilltops and dense forests. Over a short course of time, the distinctive genetic structure of each Lue population were shaped. Differences among the Lue populations also include the possibility of admixture with the other ethnic groups in Laos during their process of migration.⁶

Although the Yong and Lue show different genetic structure, they still share mtDNA HVR-1 haplotypes with their common Dai ancestors (Figure 1). Indications of nonsignificant between-group variations in analysis of molecular variance results (Table 2) and the overlapped clustering in DAPC analysis (Figure 3) indicate common genetic ancestry among the Yong, Lue and Dai people, in agreement with the historical record.⁸ However, the Yong people seem to maintain their maternal Dai genetic structure more effectively than the Lue people do as shown by several demographic parameters (Figure 2 and Table 1). Mass migration, which is a scenario in which a population moves directly from one place to another place within a short period of time, has possibly promoted the maintenance of the ancestral Dai genetic structure in the Yong population. The founder effect has an influence on the reduced effective population size in both the Yong and Lue people, but its impact has a lesser influence on the Yong maternal effective size than the Lue has had (Figure 2). A relatively large number of the Yong immigrants, owing to their mass migration from southern China to northern Thailand, may have debilitated the founder effect's impact on the Yong genetic structure. Interestingly, most of the Yong populations exhibit genetic homogeneity (Figure 3) even though they have lived in different localities. It is noteworthy to mention that all Yong villages are located in an enclosed area surrounded by mountains known as the Chiang Mai–Lamphun Basin. Geographic proximity within this basin provides for easier contact between the Yong populations. Thus, the Yong people's close genetic relatedness may be owing to the gene flow that exists among them as shown by a high degree of Yong private haplotypes (Figure 1).

One of the Lue populations, Lue3 from Chiang Rai, showed constant population size (Figure 2), and is more closely related to the Yong and Dai populations than other Lue populations (Figure 4). To find out why the Lue3 was not affected by the founder scenario that other Lue practiced, we went back to the Lue village history. Through personal interviews with the villagers, we found that this Lue

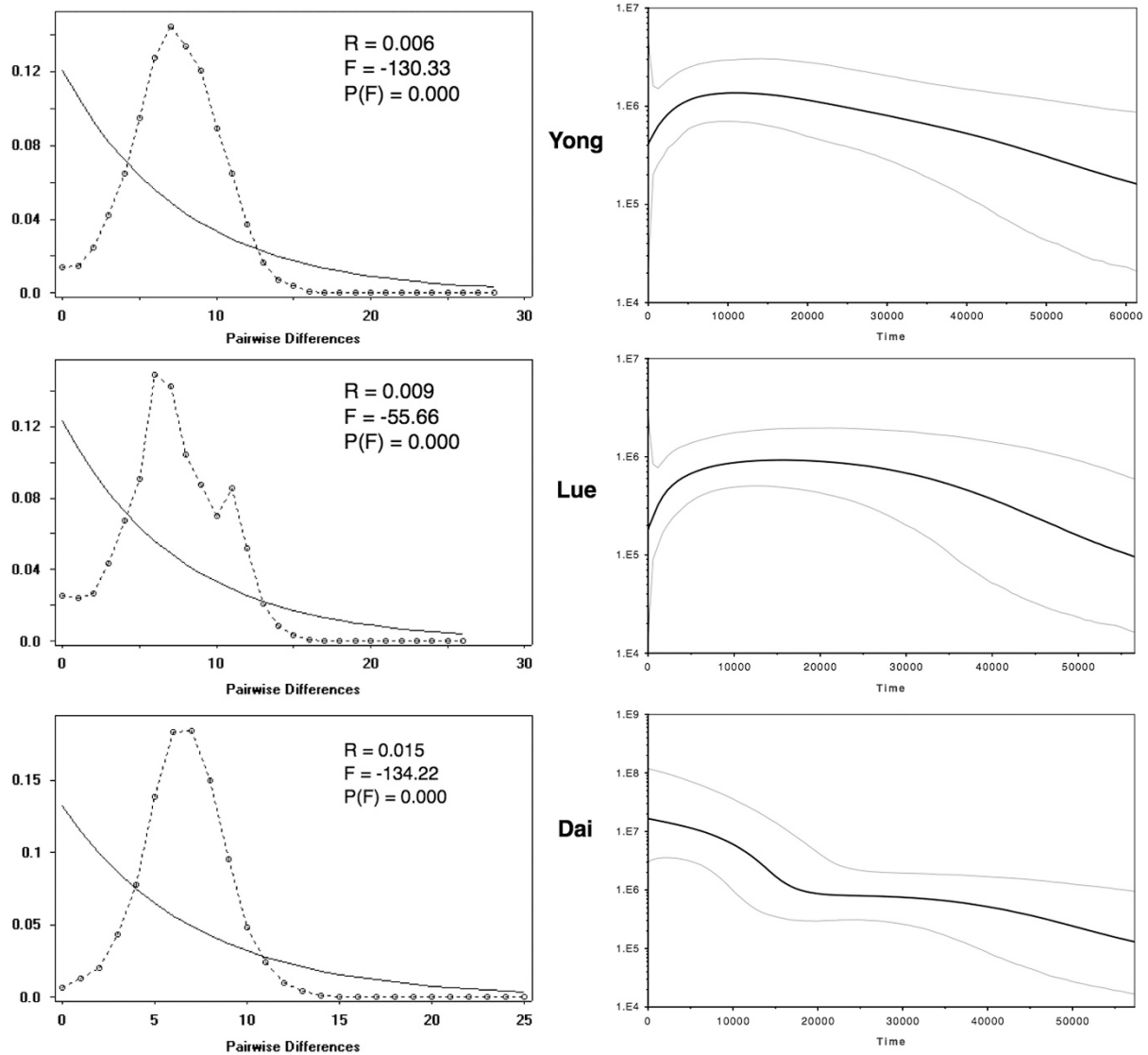


Figure 2 Mismatch distributions of the Yong, Lue and Dai populations (left). The x axis displays the number of nucleotide differences between pairs of sequences and the y axis reveals the frequency of each number of nucleotide differences. Dotted lines indicate the observed distributions and solid lines indicate the expected distributions. Bayesian skyline plot showing size changes of each ethnic group (right). Maternal effective size and time from present in years are displayed in the y axis and x axis, respectively. The solid line indicates the median estimate and the thin lines are the 95% highest posterior density edge. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

Table 2 Analysis of molecular variance results

| Groups | No. of population in each group | Within populations | | Among populations within groups | | Among groups | |
|--------------|------------------------------------|--------------------|-------------|---------------------------------|-------------|--------------|-----------------------|
| | | Variance (%) | Φ_{st} | Variance (%) | Φ_{sc} | Variance (%) | Φ_{ct} |
| All samples | 13 | 96.68 | 0.0332* | 3.32 | | | |
| Yong/Lue/Dai | 6/4/3 | 96.55 | 0.0301* | 3.00 | 0.0346* | 0.45 | 0.004 ^{ns} |
| Yong/Lue | 6/4 | 95.88 | 0.0320* | 3.17 | 0.0412* | 0.95 | 0.0095* |
| Yong/Dai | 6/3 | 98.06 | 0.0199* | 1.99 | 0.0194* | -0.05 | -0.0005 ^{ns} |
| Lue/Dai | 4/3 | 95.93 | 0.0411* | 4.11 | 0.0407* | -0.04 | -0.0004 ^{ns} |
| Yong | 6 | 98.03 | 0.0197* | 1.97 | | | |
| Lue | 4 | 94.85 | 0.0515* | 5.15 | | | |
| Dai | 3 | 98.00 | 0.0200* | 2.00 | | | |

*Statistically significant at $P < 0.05$.

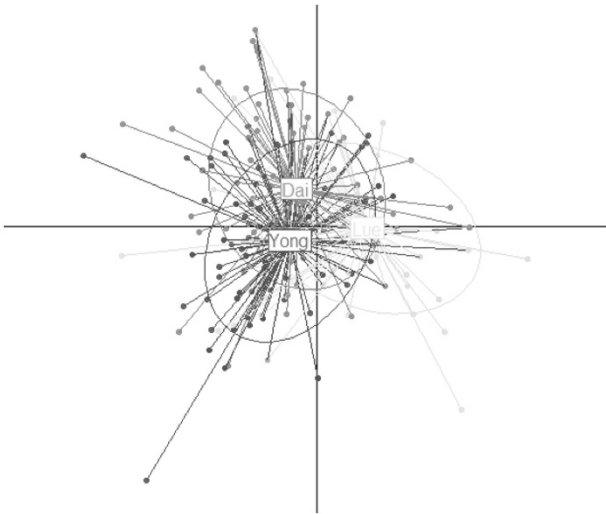


Figure 3 DAPC analysis shows the overlapped clustering patterns of the Yong (blue), Lue (yellow) and Dai (red). DAPC, Discriminant Analysis of Principal Component. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

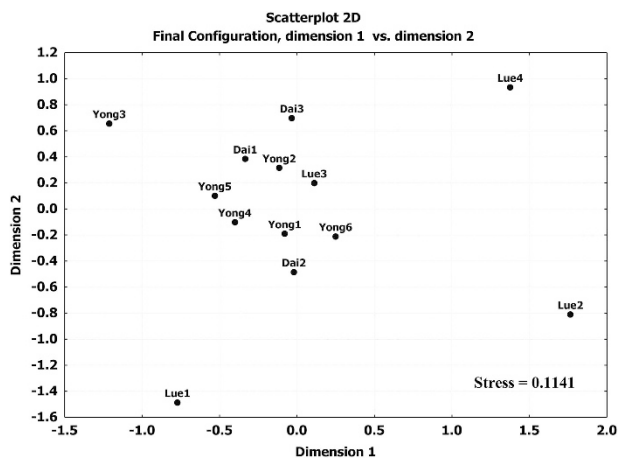


Figure 4 Biplots of dimensional 1 to 2 of the Yong, Lue and Dai, based on Φ_{st} distance matrix.

population had just migrated directly from Chiang Rung (Jinghong), Xishungbanna, southern China within the past 100 years. This very recent migration event may have maintained their ancestors' genetic structure.

Historically, migration has had as much to do with the movement of disease as with the movement of people.³¹ Although, there has been no recent report on the incidence of genetic diseases among the Tai–Kadai ethnic groups, our historical and genetic links between the Tai–Kadai people in China and Thailand may have proven to be fruitful with regard to future studies on the possibility of founder mutations in terms of genetic diseases or genetic predispositions to disease. The genetic fluctuations among the Tai–Kadai speaking groups in northern Thailand, as the results of different migration patterns, are also an important basis for forensic, medical genetic, epidemiological and pharmacogenomic studies on these people. Even though our results serve as evidence of the effect of migration patterns on populations' genetic structure, they rely on sequences acquired from only the HVR-I region of the mtDNA. Further investigations of haplogroup lineages from the rest of the mtDNA

genome, as well as autosomal and Y-chromosomal markers, could elucidate a more comprehensive insight of the complex migration history of the Tai–Kadai people and their neighboring ethnic groups.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We wish to thank all cooperating volunteers and village chiefs for their participation in the sample collection process. We are grateful to Russell Kirk Hollis, Faculty of Humanities, Chiang Mai University, Thailand, for English approval. JK has received support from the Office of the Higher Education Commission and the Thailand Research Fund (Grant No. MRG5180053). WK was supported by Khon Kaen University under the Incubation Researcher Project. FT and SG were funded by the ERC Advanced Grant Agreement No 295733, 'LanGeLin' project.

- 1 Wijsman, E. M. & Cavalli-Sforza, L. L. Migration and genetic population structure with special reference to humans. *Annu. Rev. Ecol. Evol. Syst.* **15**, 279–301 (1984).
- 2 van der Walt, J. M., Dementieva, Y. A., Martin, E. R., Scott, W. K., Nicodemus, K. K., Kroner, C. C. *et al.* Analysis of European mitochondrial haplogroups with Alzheimer disease risk. *Neurosci. Lett.* **365**, 28–32 (2004).
- 3 Ebermann, I., Koenekoop, R. K., Lopez, I., Bou-Khzam, L., Pigeon, R. & Bolz, H. J. An USH2A founder mutation is the major cause of Usher syndrome type 2 in Canadians of French origin and confirms common roots of Quebecois and Acadians. *Eur. J. Hum. Genet.* **17**, 80–84 (2009).
- 4 Qian, Y. P., Chu, Z. T., Dai, Q., Wie, C. D., Chu, J. Y., Tajima, A. *et al.* Mitochondrial DNA polymorphism in Yunnan nationalities in China. *J. Hum. Genet.* **46**, 211–220 (2001).
- 5 Penth, H. *A Brief History of Lanna: Civilizations of North Thailand* (Silkworm books, Chiang Mai, Thailand, 2000).
- 6 Schliesinger, J. *Tai Groups of Thailand: Volume 1 Introduction and Overview* (White Lotus Press, Bangkok, Thailand, 2000).
- 7 Cohen, P. T. Lue ethnicity in national context: a comparative study of Tai Lue communities in Thailand and Laos. *J. Siam. Soc.* **86**, 49–61 (1998).
- 8 Malasam, S. *On the Migration and Settlement of the Yong people in Lamphun, Thailand from 1805-1902* (Thammasat University Press, Bangkok, Thailand, 2001).
- 9 Seielstad, 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).
- 10 Kampuansai, J., Bertorelle, G., Castri, L., Nakhunlung, S., Seielstad, M. & Kangwanpong, D. Mitochondrial DNA variation of Tai speaking peoples in northern Thailand. *Sci. Asia* **33**, 443–448 (2007).
- 11 Yao, Y. G., Nie, L., Harpending, H., Fu, Y. X., Yuan, Z. G. & Zhang, Y. P. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *Am. J. Phys. Anthropol.* **118**, 63–76 (2002).
- 12 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).
- 13 Schurr, T. G., Sukernik, R. I., Starikovskaya, Y. B. & Wallace, D. C. Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk sea-Bering sea region during the Neolithic. *Am. J. Phys. Anthropol.* **108**, 1–39 (1999).
- 14 Fuchareon, G., Fuchareon, S. & Horai, S. Mitochondrial DNA polymorphisms in Thailand. *J. Hum. Genet.* **46**, 115–125 (2001).
- 15 Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M. & Howell, N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147 (1999).
- 16 Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
- 17 Fu, Y. X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925 (1997).
- 18 Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
- 19 Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).
- 20 Drummond, A. J., Nicholls, G. K., Rodrigo, A. G. & Solomon, W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* **161**, 1307–1320 (2002).
- 21 Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUTi and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
- 22 Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772 (2012).

- 23 Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A. *et al*. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* **84**, 740–759 (2009).
- 24 Excoffier, L., Smouse, P. & Wuattro, J. Analysis of molecular variance inferred from metric distance among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).
- 25 Jombart, T. & Ahmed, I. Adegnet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071 (2011).
- 26 Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94 (2010).
- 27 Rogers, A. R. & Harpending, H. Population growth makes waves in the distribution of pairwise genetic distances. *Mol. Biol. Evol.* **9**, 552–569 (1992).
- 28 Rogers, A. R. Genetic evidence for a pleistocene population explosion. *Evolution* **49**, 608–615 (1995).
- 29 Jobling, M. A., Hurler, M. E. & Tyler-Smith, C. Making Inferences from Diversity in Human Evolutionary Genetics: Origins, Peoples and Disease 115–116 (Garland publishing, USA, 2004).
- 30 Pittayaporn, P. Layers of Chinese loanwords in Proto-Southwestern Tai as evidence for the dating of the spread of Southwestern Tai. *Manusya* **20**, 47–68 (2014).
- 31 Davis, M. C., Novak, S. J. & Hampikian, G. Mitochondrial DNA analysis of an immigrant Basque population: loss of diversity due to founder effects. *Am. J. Phys. Anthropol.* **144**, 516–525 (2011).

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