Haplotype Profiling of Y-STRs among Northern Population in Thailand

Supakit Khacha-ananda[1,2], Phatcharin Mahawong[3], Thida Kaewkod[4]

[1]Instructor, Department of Forensic Medicine, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand, 50200, [2]Instructor, Research Center in Bioresources for Agriculture, Industry and Medicine, Chiang Mai University, Chiang Mai, Thailand, [3]Medical Technologist, Department of Forensic Medicine, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand, [4]Postdoctoral Researcher, Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand

Abstract

Background: The investigation of Y-short tandem repeats (Y-STRs) has been widely performed in forensic caseworks. Due to variation between populations, understanding of genetic information within individual a population could provide the database and scientifically reliable results. This study was to investigate allele and haplotype frequencies of Y-STRs among localized people who lived in northern Thailand.

Methods: A retrospective descriptive study was conducted by gathering medical reports of Y-STRs typing. The allele frequency and gene diversity as well as haplotype frequency and diversity were calculated. Pairwise genetic distance (R_{st}) was also calculated based on haplotype pattern by using AMOVA and Multidimensional Scaling (MDS) tools.

Results: The result showed that DYS389II, DYS439, and DYS392 represented the highest number of allele patterns. The highest and lowest allele frequency was found to be allele 14 of DYS437 and allele 13 of DYS437, respectively. The highest and lowest gene diversity was observed in DYS389II and DYS437, respectively. One-hundred and sixty different haplotypes were defined where 144 carried a unique haplotype and 16 carried a replicate haplotype. A significant R_{st} value was obtained between the studied population and central Thai population.

Conclusion: This study provided genetic database of Y-STRs among localized population in northern region of Thailand. The genetic structure of our subjects also represented the significant close relationship with some other populations in Southeast Asia.

Key words: *Allele frequency, Forensic genetic, Haplotype frequency, Y-STRs*

Introduction

DNA fingerprinting performed on the specific region on the DNA strand known as short tandem repeats (STRs) has been widely used for human identification in forensic genetic casework[1]. Because of variation in STRs lenght and high meutation rate, it promoted the genetic polymorphism of STRs markers and therefore useful in human identity testing. The investigation of Y-STRs is a useful reliable marker for personal identity in forensic genetic cases including paternity testing in paternal lineage, historical cases, special cases of the missing person or disaster victim identification involving men, ancestor study, as well as sexual assault with mixed DNA profile between male and female DNA[2, 3]. Moreover, Y-STRs analysis can be crucial for sex confirmation in the case of amelogenin dropout[4].

Corresponding author:
Supakit Khacha-Ananda,
Department of Forensic Medicine, Faculty of Medicine, Chiang Mai University, 110 Intawaroros, Sriphum, Chiang Mai, 50200, Thailand.
Email: supakit.kh@cmu.ac.th

According to the Scientific Working Group on DNA Analysis Methods (SWGDAM), the interpretation of Y-STRs is based on the comparison of haplotype between question and known samples. The SWGDAM further recommended to calculate Y haplotype frequency based on the ratio of number of observed haplotypes and the total number of haplotypes in a specific population database. The additional statistical parameter might be performed by calculation of the upper confidence limit for the probability of the Y haplotype within a population and match probability[5, 6]. Therefore, the localized population allele and haplotype frequencies are essential for more accurate calcualtion of Y-STRs result.

Previous studies demonstrated the diversity of Y-STRs on different populations between Africa, Asia, Europe, Latin America, and North America including the different ethnic groups[7, 8]. Since Thailand is located on Mainland Southeast Asia which was a route for people migration from southern China and other nationalities during historical migration and settlement, it promoted ethnic diversity. As expected, it was confirmed by the finding of 70 different recognized languages in people who lived in Thailand[9]. Therefore, it was possible that factors for example different ethnic groups and geographic regions could affect a diversity in population probably influenced genetic variation in Y-STRs. Likewise, there is insufficient information of Y-STRs data on certain specific northern Thai populations. This study was to evaluate the genetic profile of Y-STRs in the northern Thai population in order to discover allele and haplotype frequency database for use in related populations. The understanding of genetic profiles within individual a population could provide the precise database and scientifically reliable results.

Materials and Methods

Study design

A retrospective descriptive study was conducted by gathering information from of Y-STRs typing report from Faculty of Medicine, Chiang Mai University, Thailand. This study was performed in accordance with the Code of Ethichs of the World Medical Association (Declaration of Helsinki). Ethical approval for this study was obtained from the Research Ethics Committee of Faculty of Medicine, Chiang Mai University, Thailand (Number 6109/2019).

Data collection

Y-STRs loci containing DYS19, DYS385a/b, DYS389-I, DYS389-II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, and DYS439 were recorded. The criteria for data enrollment consisted of 1) male living in 17 provinces of northern Thailand according to Department of Provincial Administration who were declared ethnicity as northern region of Thai 2) each subject must not be blood-related.

Allele frequency and gene diversity

Y-STRs data was transferred to calculate the allele frequency and gene diversity (GD) using the counting method according to the formula supplied by Nei, 1987[10] as follow equations:

$$\text{Allele frequency} = \frac{\text{Number of observed allele}}{\text{Number of total allele}}$$

$$\text{Gene diversity (GD)} = \frac{n(1 - \Sigma Pi^2)}{(n-1)}$$

where n: number of populations, Pi: allele frequency in tested population

Haplotype frequency and diversity

The number of haplotypes was estimated by counting method. Haplotype frequency and diversity were calculated as follow equations[11]:

$$\text{Haplotype frequency} = \frac{\text{Number of observed haplotype}}{\text{Number of total haplotype}}$$

$$\text{Haplotype diversity} = \frac{n(1 - \Sigma Pi^2)}{(n-1)}$$

where n: number of populations, Pi: haplotype frequency in tested population

Genetic Distance

Pairwise genetic distance (Rst) was also calculated based on haplotype pattern by using AMOVA and Multidimensional Scaling (MDS) tools of YHRD (www.yhrd.org).

Results

Totally 53 alleles were detected at the all tested Y-STRs loci in 180 samples. The most recorded allele was found to be DYS389II, DYS439, and DYS392 loci with 7 patterns. The most allele frequency was observed at allele 14 of DYS437, whereas the least allele frequency was observed at allele 15 of DYS389I, allele 21 of DYS390, allele 11 of DYS393, allele 12 of DYS391, allele 15 of DYS439, allele 16 of DYS392, and allele 13 of DYS437. The greatest polymorphic marker was DYS389II with a genetic diversity (GD) value of 0.74057 and the least polymorphic was DYS437 with a GD value of 0.41239. Moreover, nine loci including DYS389I, DYS390, DYS389II, DYS19, DYS393, DYS391, DYS439, DYS392. and DYS438 were classified to high polymorphism (gene diversity > 0.5) as shown in Table 1.

Regarding DYS385a/b loci, the dominant allele combination at DYS385 locus (13/19) had a frequency of 0.0778. Gene diversity for the entire sample was 0.97058. We found the tri-allelic pattern of DYS385a/b locus with 16/17/20 as shown in Table 2. This tri-alleic pattern was no matching with reported information form two databases (YHRD database (www.yhrd.org) and NIST: National Institute of Standards and Technology).

A 160 distinct haplotypes were observed in the total data set, 144 (90%) of which were unique haplotypes and 16 (10%) of which were replicated haplotypes. The replicated haplotypes can be classified into 13 haplotypes (8.125%) of them were shared by two individuals, 2 haplotypes (1.25%) of them were shared by three individuals, and 1 haplotype (0.625%) was shared by four individuals. Haplotype diversity for the entire sample was 0.9984 (99.84%) meaning that the probability of finding an individual sharing this haplotype is 1 in 625 (0.16%). The common haplotype observed in this study on DYS389I, DYS390, DYS389II, DYS19, DYS385a/b, DYS393, DYS391, DYS439, DYS392, DYS437, and DYS438 loci was 14, 23, 31, 14, 11/12, 13, 11, 12, 14, 14, 11. This haplotype pattern was matched 52 times in total 266,542 haplotypes in global populations, and 3 times in total 1,170 haplotypes in Thai populations according to haplotype pattern in YHRD database.

To determine the genetic flow among populations, pairwise genetic distance (R_{st}) for each locus were estimated between previously reported populations using YHRD calculation online. The observed haplotype frequency in this study was compared with databases in YHRD consisting of population from central Thailand[12], Hongkong[13], Japan[14-16], Laos and Nepal[17, 18], South Korea[19-21], Taiwan[22], and Vietnam[7, 23]. Haplotypes of the northern Thai population were significantly different from those of the other 8 populations ($P < 0.05$). The R_{st} values for genetic distance was ranging from 0.0118 – 0.5092. It indicated that the genetic differentiation of each populations was classified as low to very high differentiation according to the previous report[24]. The lowest R_{st} value indicating the smallest genetic distances was observed between central Thai population and Vietnam population; whereas the highest R_{st} value indicating the largest genetic distances was found between population from Laos and Taiwan. It is clear that the our studied population was found to cluster more closely with populations from the central Thai population ($R_{st} = 0.0154$) as shown in Table 3. As demonstrated in the MDS plot, it showed the close relationship between populations from northern Thailand, central Thailand, Hongkong, Vietnam, and Laos. These clusters form a conspicuous cluster standing far apart from populations from South Korea, Japan, Nepal, and Taiwan as shown in Figure 1.

Table 1 Allele frequency and gene diversity of Y-STRs loci

Alleles	DYS389I	DYS390	DYS389II	DYS19	DYS393	DYS391	DYS439	DYS392	DYS437	DYS438
9						0.0500	0.0278			0.0111
10						0.6167	0.0444	0.0167		0.5944
11					0.0056	0.3278	0.3167	0.1500		0.3278
12	0.3500				0.2611	0.0056	0.3444	0.0222		0.0667
13	0.3722			0.0611	0.3167		0.1944	0.4444	0.0056	
14	0.2722			0.3167	0.3389		0.0667	0.3500	0.7222	
15	0.0056			0.4222	0.0778		0.0056	0.0111	0.2611	
16				0.1722				0.0056	0.0111	
17				0.0278						
21		0.0056								
22		0.0167								
23		0.2444								
24		0.3889								
25		0.3333								
26		0.0111								
27			0.0278							
28			0.2611							
29			0.3167							
30			0.3000							
31			0.0611							
32			0.0167							
33			0.0167							
Gene diversity	0.66855	0.68128	0.74057	0.69112	0.71456	0.51254	0.74018	0.66025	0.41239	0.53765

Table 2 Haplotype pattern and frequency of DYS385a/b locus

Alleles	Frequency	Alleles	Frequency	Alleles	Frequency	Alleles	Frequency	Alleles	Frequency
7/16	0.0056	12/13	0.0056	13/19	0.0778	14/22	0.0056	16/18	0.0167
11/11	0.0056	12/16	0.0111	13/20	0.0222	14/23	0.0056	16/19	0.0444
11/12	0.0500	12/17	0.0167	13/22	0.0056	15/15	0.0222	16/20	0.0056
11/13	0.0278	12/18	0.0111	14/14	0.0111	15/16	0.0333	16/21	0.0056
11/14	0.0222	12/19	0.0333	14/15	0.0056	15/17	0.0111	17/17	0.0056
11/16	0.0056	12/20	0.0167	14/16	0.0056	15/18	0.0278	17/18	0.0056
11/17	0.0056	12/22	0.0056	14/17	0.0278	15/19	0.0500	17/19	0.0056
11/18	0.0056	13/13	0.0111	14/18	0.0333	15/20	0.0222	17/21	0.0056
11/19	0.0167	13/14	0.0111	14/19	0.0222	15/21	0.0333	18/18	0.0111
11/20	0.0056	13/17	0.0722	14/20	0.0056	16/17/20	0.0056	18/19	0.0056
12/12	0.0111	13/18	0.0611	14/21	0.0056	16/17	0.0278	21/21	0.0056

Table 3 The genetic comparison by pairwise R_{st} value between different populations

Populations	Northern Thai	Central Thai	Hongkong	Japan	Laos	Nepal	South Korea	Taiwan	Vietnam
Northern Thai	-								
Central Thai	0.0154	-							
Hongkong	0.0512	0.0198	-						
Japan	0.1120	0.0899	0.0977	-					
Laos	0.0244	0.0492	0.1044	0.1437	-				
Nepal	0.0931	0.0800	0.0661	0.0809	0.1689	-			
South Korea	0.1395	0.1027	0.0970	0.0252	0.1681	0.1035	-		
Taiwan	0.3792	0.3092	0.2218	0.3545	0.5092	0.3345	0.3337	-	
Vietnam	0.0239	0.0118	0.0565	0.0882	0.0278	0.1180	0.1076	0.3905	-

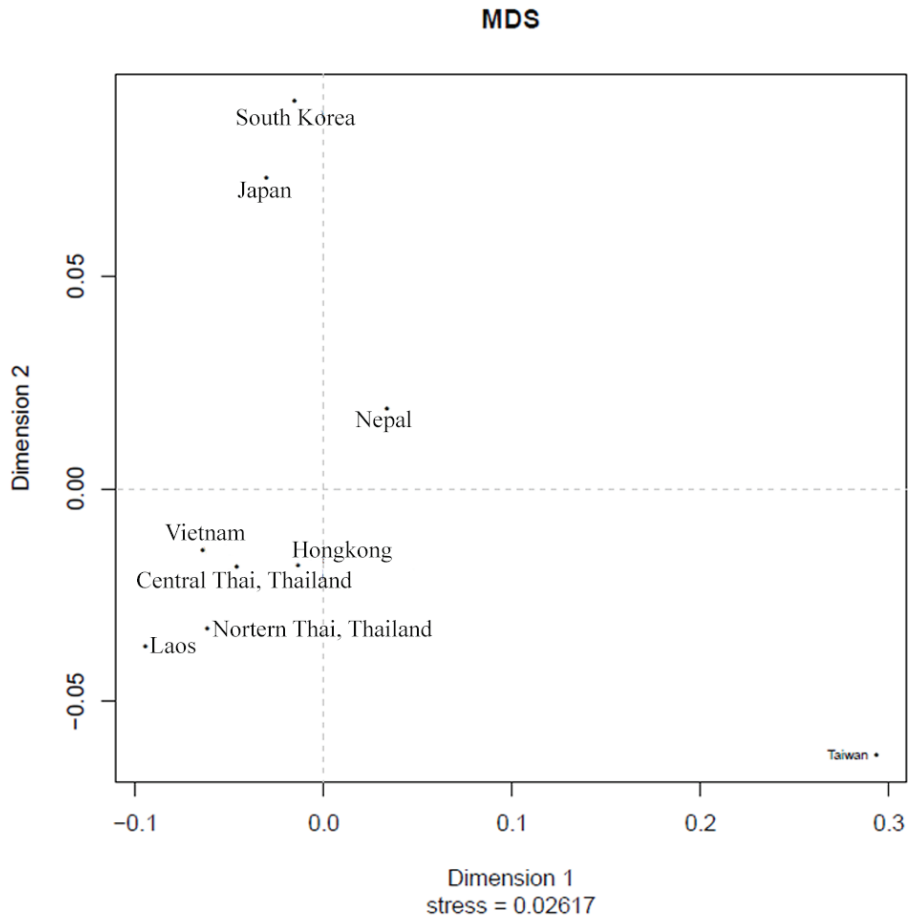

Figure 1 A MDS plot based on R_{st} between northern Thai population and eight reference populations.

Discussion

Our study discovered the genetic structure of localized population in northern region of Thailand in terms of Y-STRs. Such information is of essential importance for forensci casworks. We found that allele 14 at DYS437 locus had the highest frequency which was in concordance with previously report observation in the population from lower northern and central Thailand as well as China[12, 25]. We observed the highest allele frequency on DYS391, followed with DYS549 and DYS437 loci that was concordant with the population from central, west, and southern Thai populations[26]. The value and highest locus of gene diversity among studied Y-STRs loci were similar to the gene diversity of central Thai populations which was previously described[27]. The gene diversity of Y-STRs among different continents and geographic regions was caused from mutation, including random genetic drift. It affected allele frequency in small populations without natural selection resulting in some allele being unable to transmit into the next generation. Additionally, the factors related with social and cultural activity were reported to affect the allele frequency of Y-STRs[28]. It is interesting that one sample of DYS385a/b allele was found to be a tri-allelic pattern (16/17/20). The tri-allelic patterns of DYS385a/b were sometimes observed into 9/13/14, 11/12/13, 11/12/14, 11/14/15, 12/16/17, 13/14/15, 13/14/18, 13/18/19, 13/19/20, 14/16/17, and 15/16/17[29]. Howoever, the allelic pattern of DYS385a/b could be found to have a five-allelic pattern and a six-allelic pattern in the population from China[30]. In the present study, we found that studied population had a close genetic relationship with the population from central Thai, Laos, and Vietname; whereas the difference between the northern Thai population and Taiwanese population was large. It is probable that the patrilineal genetic structure could be from the same ancestral population. Thai population

was reported to have a close genetic relationship with East Asian populations who are native Austroasiatic and Sino-Tibetan speakers[31]. The native Austroasiatic speakers have covered and spread over Asia, India, and Southeast Asia[32]. According to the report of Kutanan *et al.*, 2017, the ancient origin of the Thai and Laotian populations was derived from Austroasiatic groups and some sub-population from Taiwan was separated from the Thai population[33]. The native Sino-Tibetan speakers originated among people in northern China and spread over on the east, west and south of China[34]. The migration of people from Southern China into the northern region of Thailand influenced the relationship between the Thai population and other Southeast Asian populations[35, 36].

Conclusion

We provided the precise database of genetic information of Y-STRs loci in specific northern Thai population to be applied for forensic genetic casework. The limitation of this study is that it does not represent the whole genetic structure of all populations in Thailand. To clear the understanding about the different genetic structure between other populations, the information of population among other regions of Thailand should be collected to compare the genetic polymorphism between individual people.

Conflict of Interest: The authors declare that they have no conflicts of interest.

Source of Funding: This work was supported by the Faculty of Medicine, Chiang Mai University, Thailand [grant no. 102-2562]. This research work was partially supported by Chiang Mai University.

References

1. Pena S, Chakraborty R, Epplen J, Jeffreys A. DNA Fingerprinting: State of the Science. Switzerland: Birkhauser Verlag, Basel; 1993.
2. Kayser M. Forensic use of Y-chromosome DNA: a general overview. Hum Genet. 2017;136(5):621-35.
3. Roewer L. Y-chromosome short tandem repeats in forensics—Sexing, profiling, and matching male DNA. WIREs. 2019;1(4):e1336.
4. Yang X, Shi MS, Yuan L, Lu D. Application of multiple genetic markers in a case of determination of half sibling. Fa Yi Xue Za Zhi. 2016;32(1):45-8.
5. SWGDAM. Y-chromosome short tandem repeat (Y-STR) interpretation guidelines. FSC. 2009;11(1).
6. Gusmão L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, et al. DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. Forensic Sci Int. 2006;157(2-3):187-97.
7. Miranda-Barros F, Romanini C, Peréz LA, Nhu ND, Phan TD, Carvalho EF, et al. Y Chromosome STR haplotypes in different ethnic groups of Vietnam. Forensic Sci Int Genet. 2016;22:e18-e20.
8. Purps J, Siegert S, Willuweit S, Nagy M, Alves C, Salazar R, et al. A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. Forensic Sci Int Genet. 2014;12:12-23.
9. Kutanan W, Kampuansai J, Brunelli A, Ghirotto S, Pittayaporn P, Ruangchai S, et al. New insights from Thailand into the maternal genetic history of Mainland Southeast Asia. Eur J Hum Genet. 2018;26(6):898-911.
10. Nei M. Molecular evolutionary genetics. New York: Columbia University Press; 1987.
11. Gusmão L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, et al. DNA commission of the international society of forensic genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. Int J Legal Med. 2006;120(4):191-200.
12. Siriboonpiputtana T, Jomsawat U, Rinthachai T, Thanakitgosate J, Shotivaranon J, Limsuwanachot N, et al. Y-chromosomal STR haplotypes in Central Thai population. Forensic Sci Int Genet. 2010;4(3):e71-e2.
13. Ip S, Lin S, Lam T. Haplotype data of 27 Y-STR loci in Hong Kong Chinese. Forensic Sci Int Genet. 2019;38:e14-e5.
14. Watahiki H, Fujii K, Fukagawa T, Mita Y, Kitayama T, Mizuno N. Polymorphisms and microvariant sequences in the Japanese population for 25 Y-STR markers and their relationships to Y-chromosome haplogroups. Forensic Sci Int Genet. 2019;41:e1-e7.
15. Kwak KD, Jin HJ, Shin DJ, Kim JM, Roewer L, Krawczak M, et al. Y-chromosomal STR haplotypes and their applications to forensic and

population studies in east Asia. Int J Legal Med. 2005;119(4):195-201.

16. Mizuno N, Nakahara H, Sekiguchi K, Yoshida K, Nakano M, Kasai K. 16 Y chromosomal STR haplotypes in Japanese. Forensic Sci Int. 2008;174(1):71-6.
17. Gayden T, Mirabal S, Cadenas AM, Lacau H, Simms TM, Morlote D, et al. Genetic insights into the origins of Tibeto-Burman populations in the Himalayas. J Hum Genet. 2009;54(4):216-23.
18. Parkin EJ, Kraayenbrink T, Opgenort JRML, van Driem GL, Tuladhar NM, de Knijff P, et al. Diversity of 26-locus Y-STR haplotypes in a Nepalese population sample: Isolation and drift in the Himalayas. Forensic Sci Int. 2007;166(2):176-81.
19. Park MJ, Lee HY, Yoo JE, Chung U, Lee SY, Shin KJ. Forensic evaluation and haplotypes of 19 Y-chromosomal STR loci in Koreans. Forensic Sci Int. 2005;152(2-3):133-47.
20. Lee HY, Park MJ, Chung U, Lee HY, Yang WI, Cho SH, et al. Haplotypes and mutation analysis of 22 Y-chromosomal STRs in Korean father-son pairs. Int J Legal Med. 2007;121(2):128-35.
21. Kim SH, Kim NY, Hong SB, Cho NS, Kim JJ, Han MS, et al. Genetic polymorphisms of 16 Y chromosomal STR loci in Korean population. Forensic Sci Int Genet. 2008;2(2):e9-e10.
22. Wu F, Ho C, Pu C, Hu K, Willuweit S, Roewer L, et al. Y-chromosomal STRs haplotypes in the Taiwanese Paiwan population. Int J Legal Med. 2011;125(1):39-43.
23. Ha HH, Nguyen TH, Tran LH, Nguyen HTH, Hoang H, Chu HH. Genetic characteristics of 23 Y-chromosomal STRs in the Kinh population in Northern Vietnam. Int J Legal Med. 2019;133(5):1403-4.
24. Wright S. Evolution and the genetics of populations. Chicago, USA: University of Chicago Press; 1978.
25. Hao S, Zhang X, Liu Y, Lu D. Genetic diversities of 23 Y-Chromosome short tandem repeat loci in a han population in the Beijing Region. JFSM. 2018;4(2):111-4.
26. Sathirapatya T, Sukawutthiya P, Vongpaisarnsin K. Massively parallel sequencing of 24 Y-STR loci in Thai population. Forensic Sci Int Genet Suppl Ser. 2017;6:310-3.
27. Suppajariyawat P. The analysis of allele frequencies and haplotype profiling of five Y chromosome STR loci in Thai male population and forensic application [Master's thesis]: Bangkok, Thailand, Mahidol University; 2003.
28. Kutanan W, Kanfwanpong D. Y chromosome and the study of human evolution. Thai J Genet. 2014;7(2):69-86.
29. Butler JM, Decker AE, Kline MC, Vallone PM. Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation. J Forensic Sci. 2005;50(4):853-9.
30. Li F, Zhao P, Xiao C, Feng C, Chen L, Du W. Identification of extra alleles in DYS385a/b multi-allelic patterns. Leg Med. 2019;37.
31. Willuweit S, Roewer L. Y chromosome haplotype reference database (YHRD): Update. Forensic Sci Int Genet. 2007;1(2):83-7.
32. Riccio ME, Nunes JM, Rahal M, Kervaire B, Tiercy JM, Sanchez-Mazas A. The Austroasiatic Munda population from India and Its enigmatic origin: a HLA diversity study. Hum Biol. 2011;83(3):405-35.
33. Kutanan W, Kampuansai J, Srikummool M, Kangwanpong D, Ghirotto S, Brunelli A, et al. Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai–Kadai languages. Hum Genet 2017;136(1):85-98.
34. Su B, Xiao C, Deka R, Seielstad M, Kangwanpong D, Xiao J, et al. Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. Hum Genet. 2012;107:582-90.
35. Wangkumhang P, Shaw PJ, Chaichoompu K, Ngamphiw C, Assawamakin A, Nuinoon M, et al. Insight into the peopling of Mainland Southeast Asia from Thai population genetic structure. PLoS One. 2013;8(11):e79522-e.
36. Brunelli A, Kampuansai J, Seielstad M, Lomthaisong K, Kangwanpong D, Ghirotto S, et al. Y chromosomal evidence on the origin of northern Thai people. PLoS One. 2017;12(7):1-13.