Y-CHROMOSOMAL SHORT TANDEM REPEAT LOCUS DYS385 VARIATIONS IN HMONG ETHNIC GROUP OF NORTHERN THAILAND

Jatupol Kampuansai* and Kullaporn Totsparin

Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand
*e-mail: jatupol_k@hotmail.com

Abstract
DYS385 is a polymorphic marker on the Y chromosome. The distribution of DYS385 genotypes was studied in five Hmong populations residing in the northern part of Thailand. Genotype 13–21 was the most frequently observed in the Hmong. The haplotype diversity of the Hmong (0.8916, n = 90) was lower than the previously investigated in the Northern Thai population. The genetic relationship using DYS385 variations does not correlate with the Hmong cultural variety (White Hmong and Black Hmong) or locality relatedness. The DYS385 high polymorphism, compares to other loci, is very useful for identifying individuals and paternity test.

Keywords: short tandem repeat, DYS385, genetic variation, Hmong, northern Thailand

Introduction

Short tandem repeat (STR) or Microsatellite is a short sequence of nucleotide repeats, consists of 2 to 6 base pair repeating units. There are many STRs loci spread throughout the human genome. They have been used as the genetic markers to identify individuals and paternity relationship for decades. The STRs that locate in the Non Recombining Portion of Y-chromosome (NRPY) are proven to be a powerful tool for father-son paternity test (Carey and Mitnik 2002). Escaping from recombination, Y-chromosomal STRs (Y-STRs) provides a unique system of haplotype (the combination of allelic states of markers along the chromosome) and transmit unchanged from father to his son, except for rare novel changing in STR length.

Each Y-STR locus has different power of discrimination due to its degree of polymorphism. The DYS385 is one of the high polymorphic Y-STR loci, as it consists of two duplicated linked tandem repeat array. Whereas most of the Y-STRs produce a single amplified fragment for one primer pair in Polymerase Chain Reaction (PCR) technique, the DYS385 gives two variable fragments (Schneider et al. 1998). Thus, the pattern of DYS385 in one male individual is a combination between two subloci, DYS385a and DYS385b, which regards as a single haplotype during transmission. The distributions of DYS385 variants are ethnic-specific orientation. Haplotype 13-13 appears in a high frequency in Chinese, while 13-17 in Japanese, 13-14 in Filipinos, 11-14 in German, and 15-16 in Black (West Africa) (Schneider et al. 1998; Lessig and Edelmann 1998; Miranda et al. 2001; Gamero et al. 2001). Among Thai populations, different most frequent haplotype was observed between Northern Thai (haplotype 13-18) and Bangkok Thai (haplotype 14-18) (Bhoopat et al. 2003). In this study, the DYS385 polymorphism and its power of discrimination were investigated among the Hmong populations residing in northern Thailand.

The Hmong is the second largest group of hilltribes in Thailand with the estimated population number about 111,677 people. The majority of the Hmong had slowly migrated southward from China into Laos, then from Laos into northern Thailand during the first half of the twentieth century. The latest exodus of a Hmong migration into Thailand occurred recently about 50-60 years ago. There are now many Hmong villages located in the provinces of Northern Thailand especially in Nan and Chaing Rai. The term Miao or Meo is generally used by Thai and Laotian to identify this hilltribe group, but it is derogatory to the Hmong and is in no way acceptable to them. The Hmong has their own unique culture and uses the language which belongs to the Hmong-Mien linguistics family. They are patrilocal, a wife moves into her husband's family after married. According to the slightly different costume, there are two Hmong varieties living in northern Thailand, i.e. White Hmong and Black Hmong (Schliesinger 2000).

Methodology

The studied samples were 90 unrelated male volunteers from five Hmong villages in the northern part of Thailand (Figure 1). The white blood cell lysates of each individual were obtained with informs consent from our previous study (Besaggio et al. 2007). Total genomic DNA was extracted using inorganic salting out protocol (Seielstad et al. 1999). The DYS385 fragments were amplified as previously describe (Bhoopat et al. 2003). Amplicons were then separated by 8.5% polyacrylamide gel electrophoresis and compared with the DYS385 allelic ladder.

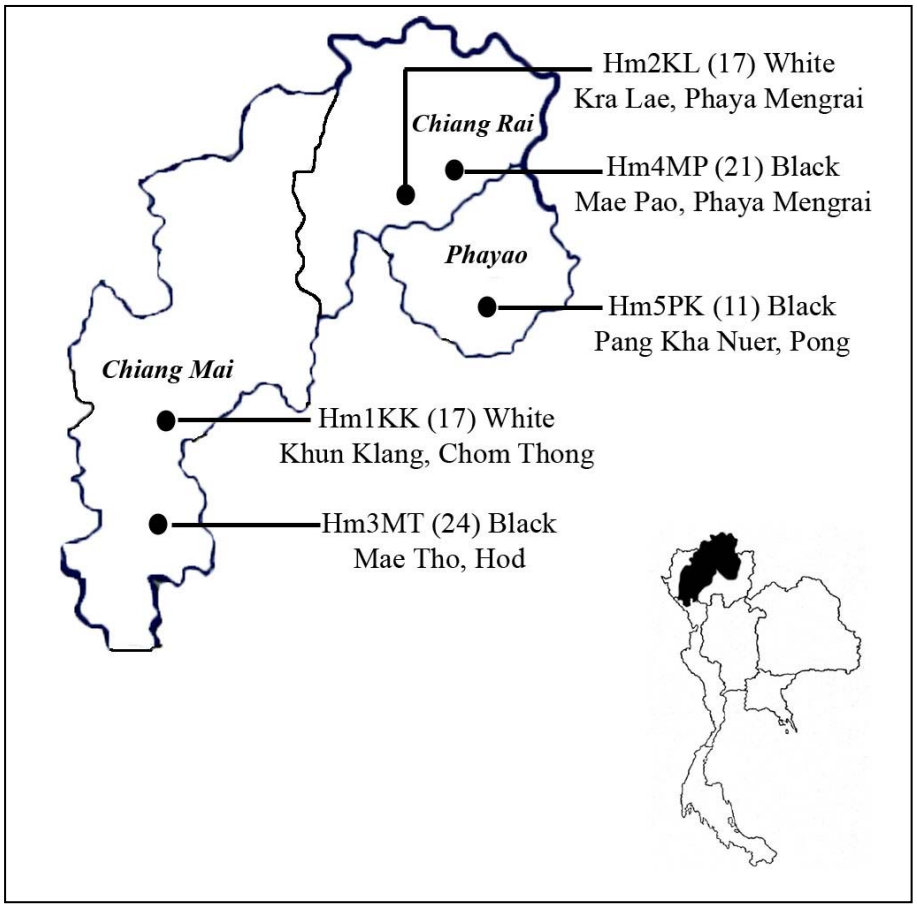
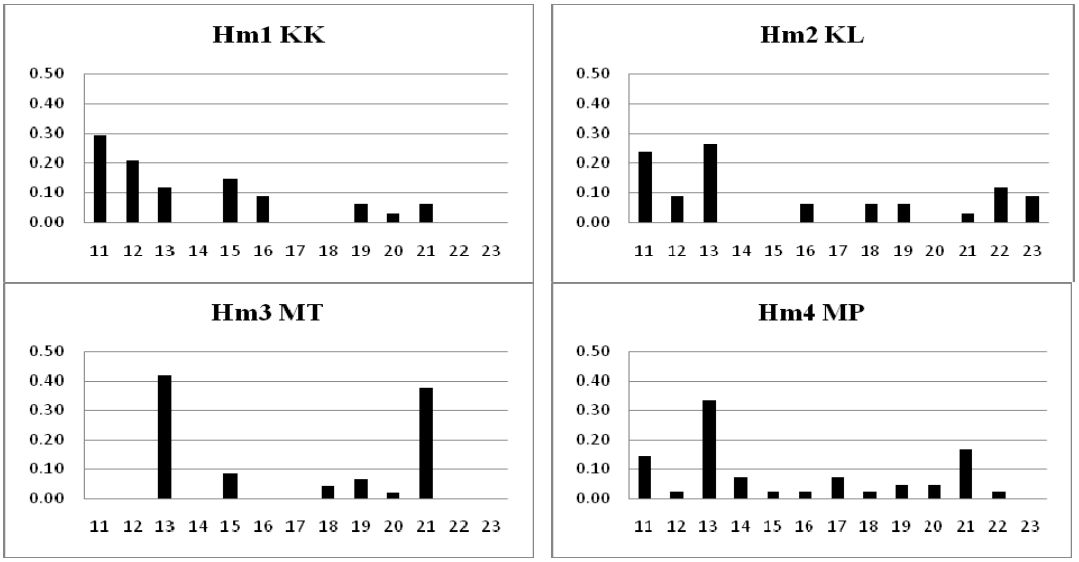


Figure 1 Map of Chiang Mai, Chiang Rai, and Phayao provinces shows code of studied Hmong population, number of samples (in parentheses), Hmong variety (White or Black), and locality (village, district)

The frequencies of observed DYS385 alleles and haplotypes were calculated by simple counting schemes. Haplotype diversities with their standard deviation (S.D.) were computed by ARLEQUIN 3.5 package (Excoffier and Lischer 2010). Considering both type and frequency, the DYS385 haplotype sharing between villages was estimated by the matching probability equation, $m = \sum p_i p_j$ where p_i and p_j were the frequencies of haplotype in population *i* and *j*, respectively. Pairwise genetic distances between poulation pairs, based on sum of squared number of alleles difference (linearization *Rst*) (Slatkin 1995), were calculated and tested for their significances by ARLEQUIN 3.5 (Excoffier and Lischer 2010). The analysis of molecular variance (AMOVA) (Excoffier et al. 1992) was performed as implements of the ARLEQUIN 3.5 software to quantify the genetic variation at three hierarchical levels, namely within population, between populations of the same group, and between groups. Note that two grouping pattern were concerned, 1) groups of Hmong varieties (White Hmong and Black Hmong), 2) groups of geographic relatedness (Chiang Mai and Chiang Rai/Phayao). The significance of the fixation indices was tested using a non-parametric permutation approach. The diversity of the pool Hmong samples was compared to those of the other 15 Y-STRs data set, i.e. DYS19, DYS388, DYS389a, DYS389b, DYS390, DYS391, DYS392, DYS393, DYS426, DYS436, DYS437, DYS439, DYS460, DYS461, and Y-GATA-A10, obtained from the same Hmong individuals in previous study (Besaggio et al. 2007).

Results

Allelic frequencies distribution of DYS385 in each Hmong population and pooled samples are illustrated in Figure 2. The most frequent allele consists of 13 repeats (allele 13), except in Hm1KK which allele 11 is the most common. The distributions of the observed alleles and frequencies are different for all populations. Pooling all samples shows that allele 13 appears in the highest frequency in the Hmong, follows by allele 21 and allele 11.



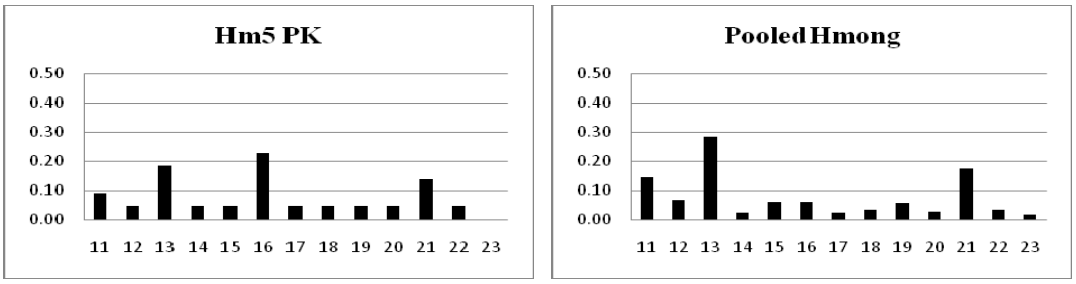


Figure 2 DYS385 allelic frequencies distributions in Hmong (X-axis is allele in repeat, Y-axis is observed frequency)

Observed haplotype frequencies of 90 Hmong men from 5 villages are shown in Table 1. Thirty different haplotypes are found and 7 haplotypes are shared among two or more populations. Number of observed haplotypes and diversity of each population are different. The Hm5PK exhibits the highest haplotype diversity, while the Hm3MT is the lowest. The most frequent haplotype in Hmong is 13-21. The overall haplotype diversity for pooled Hmong samples is 0.8916±0.0259 (Table 1). The highest matching probability, reflecting both type and frequency sharing, is observed between Hm3MT and Hm4MP (Table 2). Among 10 pairwise population comparisons, 6 pairs are significantly differences at $p<0.05$. The highest genetic distance value is observed between Hm1KK and Hm3MT (Table 2).

Table 1 Observed DYS385 haplotype frequencies in 5 Hmong populations

Haplotype	Hm1KK	Hm2KL	Hm3MT	HmMP4	Hm5PK	Total
11-11	0.1176	0.2353		0.0952		0.0889
11-12	0.3529					0.0667
11-15				0.0476		0.0111
11-16					0.1818	0.0222
11-17				0.0476		0.0111
12-16				0.0476	0.0909	0.0222
12-18		0.0588				0.0111
12-19		0.1176				0.0222
12-20	0.0588					0.0111
13-13				0.0476		0.0111
13-16					0.0909	0.0111
13-17				0.0952		0.0222
13-18		0.0588	0.0833	0.0476		0.0444
13-19	0.1176		0.0417	0.0476		0.0444
13-20				0.0952		0.0222
13-21	0.1176	0.0588	0.7083	0.2381	0.1818	0.3000
13-22		0.2353		0.0476	0.0909	0.0667
13-23		0.1765				0.0333
14-14				0.0476		0.0111
14-19				0.0476		0.0111
14-21					0.0909	0.0111
15-15	0.0588					0.0111
15-16	0.1765					0.0333

15-19			0.0833			0.0222
15-20			0.0417		0.0909	0.0222
15-21			0.0417			0.0111
16-16		0.0588				0.0111
16-17					0.0909	0.0111
18-19					0.0909	0.0111
21-21				0.0476		0.0111
Diversity	0.8456	0.8824	0.5000	0.9381	0.9636	0.8916
S.D.	0.0639	0.0468	0.1206	0.0388	0.051	0.0259

Table 2 Matching probabilities between Hmong populations (below the diagonal) and *Rst* genetic distance values (above the diagonal, *significant difference at p<0.05)

	Hm1KK	Hm2KL	Hm3MT	Hm4MP	Hm5PK
Hm1KK		0.1704	0.6499*	0.1598*	0.2879*
Hm2KL	0.0346		0.1608*	0.0000	0.0000
Hm3MT	0.0882	0.0466		0.2702*	0.2442*
Hm4MP	0.0448	0.0504	0.1746		0.0000
Hm5PK	0.0214	0.0321	0.1326	0.0519	

The genetic variation of the DYS385 dataset in Hmong was investigated by AMOVA (Table 3). When all 5 populations were grouped together, 78.38% of the genetic variation is found within population, whereas 21.62% attributes among them. The genetic variation between White Hmong and Black Hmong is 6.06% and this amount is not significant. When the populations are grouped according to their geographic relatedness (Chiang Mai and Chiang Rai/Phayao), the minus value percentage is observed. This minus percentage is a program error occurs when a by far smaller of the variation among groups than within population or among populations/within group is computed. Comparing among 16 Y-STR loci, the DYS385 shows the highest haplotype diversity follow by the DYS19 and DYS460. Twelve Y-STRs loci exhibit high diversity with more than 0.5 value. The DYS426 and DYS436 are monomophic loci in Hmong samples as only one allele is observed for each locus (Figure 3).

Table 3 Results of the Analysis of Molecular Variance (AMOVA)

Grouping pattern	**No. of pop.**	**Within population**		**Among populations Within group**		**Among groups**	
		%Variation	***Fst***	**%Variation**	***Fsc***	**%Variation**	***Fct***
All populations	5	78.38	**0.2162**	21.62			
Variety*	2/3	84.72	**0.1528**	9.22	**0.0982**	6.06	0.0606
Locality**	2/3	83.11	**0.1689**	33.15	**0.2852**	-16.26	-0.1626

Bold letter: statistical significance at $p < 0.05$.
Fst, *Fsc*, and *Fct* are the fixation index of within population, within group, and among groups, respectively.
*Between White Hmong and Black Hmong **Between Chiang Mai and Chiang Rai/Phayao

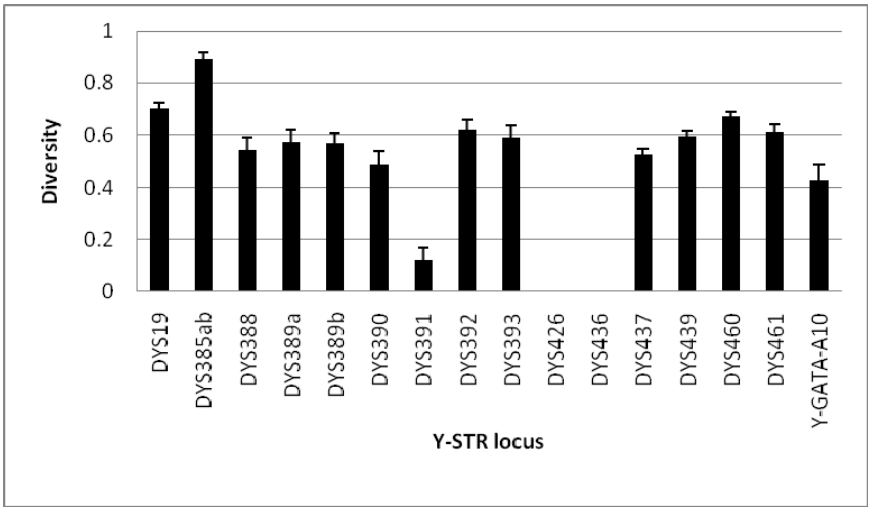


Figure 3 Haplotype diversities of 16 Y-STR loci in Hmong

Discussion and Conclusion

DNA typing using Y-chromosomal short tandem repeat loci has become the powerful method in father-son paternity test. Polymorphism of the Y-STR may arise only by rare mutation and do not reshuffling as other chromosome experience. Thus, the entire Y chromosome preserves a record of male lineage and transmits to his son. However, this special property of the Y-chromosome makes a higher sensitivity to the genetic drift than the autosomal ones. Founder effect, the phenomenon that erases some male lineages, can be observed by interpreting the allelic distribution. Theoretically, mutations of the STR are occurred by the stepwise slippage mispairing during the DNA replication. This process results the increasing or decreasing in a single unit of STR repeat (Levinson and Gutman 1987). If the slippage mutations had taken place within particular population, the continuous of single repeat variants is expected to be found. The discontinuous repeat variants, as observed in Hm1KK, Hm2KL, and Hm3MT, reflects the founder effect in these populations. Some ancestral male lineages must be lost during small founder group migrated to the present day localities.

Haplotype diversity of the Hmong (0.8916) was lower than the previous investigated in the Northern Thai (0.9430) (Bhoopat et al. 2003). The consanguineous marriage together with the patrilocal postmarital residence cultures may be the important factors that shape low diversity in the Hmong, comparing to the inter-ethnic marriage in Northern Thai. However, there is no specific DYS385 variant to distinguish the Hmong from other ethnic groups. Allele 13, the most frequent allele in Hmong, also appears at high frequency among the East-Asian populations. Moreover, the haplotype 13-17, 13-18, 13-19 which counted for one-third of the Northern Thai (Bhoopat et al. 2003), are observed in most of the Hmong populations. Even though, haplotype 13-21 is the highest frequency in pooled Hmong samples, it is not the highest one in Hm1KK and Hm2KL. Thus, there is no DYS385 Hmong-specific pattern observed in this study.

According to a similar culture practice, close relationship in each Hmong variety was first expected to be seen. The AMOVA results (Table 3) show that neither close cultural practice nor locality relatedness correlate with the genetic relationship. Most of the genetic variances found within villages. The significant fixation index between populations suggests that each Hmong villages had constructed their own genetic structure, may be though the founder effect, consanguineous marriage, and patrilocal postmarital residence. Although the

mentioned processes show their effects in lowering of the genetic diversity, the high haplotype diversity of Hm5PK (0.9636) does not fit to this scenario. The Hmong in Pang Kha Nuer village, Pong district, Phayao province (Hm5PK) migrated from different agricultural regions, i.e. Chiang Rai, Phayao, and Nan provinces (Srikummool 2005). The combination of many ancestral groups reflects in Hm5PK high diversity, unlike other Hmong populations which move from one specific area (Srikummool 2005).

Comparing among the Y-STRs, DYS385 shows the highest polymorphism. Its specialized feature as duplicated linked subloci creates lot number of haplotype variations. Even though, distinct DYS385 haplotype frequencies distributions are observed among the Hmong, not all population pairs are statistical significant different, bases on linearization *Rst* values (Table 2). Only the Hm3MT is genetically different from all other Hmong. Most of the populations showed close related to at least one of the other villages. This result suggests that the diversity of DYS385 can not distinguish the genetic different among Hmong populations. High polymorphic Y-STR loci, such as DYS19, DYS460, DYS461, dataset must be combined to increase the power of discrimination. Interestingly, the Y-STR Haplotype Reference Database (YHRD) (available online at http://www.yhrd.org) suggests an informative forensic Y-STR core set, consists of DYS19, DYS389a/b, DYS390, DYS391, DYS392, DYS393, DYS385a/b, but the DYS391 locus presents a low diversity in the Hmong (Figure 3) and should not be used to distinguish the Hmong individuals. This observation lead to our suggestion that, as the Y-STRs distributions in each ethnic group are different, the survey of genetic marker diversities is still important before establishing specific Y-STRs dataset for forensic purpose in particular populations.

Acknowledgements

We thank all the blood donors and village chiefs for their participation in interviewing and samples collecting. This study was financial supported by the Junior Researcher Fellowship 2010, Chiang Mai University, Thailand.

References

1. Besaggio D, Fuselli S, Srikummool M, Kampuansai J, Castrì L, Tyler-Smith C, Seielstad M, Kangwanpong D, Bertorelle G (2007) Genetic variation in Northern Thailand Hill Tribes: origins and relationships with social structure and linguistic differences. BMC Evolutionary Biology 7, Suppl 2.
2. Bhoopat T, Hohoff C, Steger HF (2003) Identification of DYS385 allele variants by using shorter amplicons and Northern Thai haplotype data. J Forensic Sci 48: 1108-12.
3. Carey L, Mitnik L (2002) Trends in DNA forensic analysis. Electrophoresis 23:1386–97.
4. Excoffier L, Lischer HL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Eco Res 10: 564-567.
5. Excoffier L, Smouse P, Wuattro J (1992) Analysis of molecular variance inferred from metric distance among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491.
6. Gamero JJ, Romero JL, Gonzalez JL, Carvalho M, Anjos MJ, Corte-Real F, Vieira DN, Vide MC (2001) A population-genetic study of the DYS385 haplotypes in two Spanish populations and the African immigrant population in Spain. J Forensic Sci 46:193.
7. Lessig R, Edelmann JY (1998) Chromosome polymorphisms and haplotypes in west Saxony (Germany). Int J Legal Med 111:215–8.

8. Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203-221.
9. Miranda JJ, Benecke M, Hidding M, Schmitt C (2001) Y-chromosomal short tandem repeat haplotypes at the loci DYS393, DYS19, DYS392, and DYS385-I/II, DYS390, DYS389-I/II, and DYS391 in a Filipino population sample. J Forensic Sci 46:1250–3.
10. Schliesinger J (2000) Ethnic groups of Thailand: Non-Tai-speaking peoples. White Lotus Press, Bangkok, Thailand.
11. Schneider PM, Meuser S, Waiyawuth W, Seo Y, Rittner C (1998) Tandem repeat structure of the duplicated Y-chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations. Forensic Sci Int 97:61–70.
12. Seielstad M, Bekele E, Ibrahim M, Toure A Traore M (1999). A view of modern human origins from Y chromosome microsatellite variation. Genome Res 9:558-567.
13. Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457-462.
14. Srikummool M (2005) X-, Y-chromosomal and mitochondrial DNA variations of the Karen, Hmong and Iu Mien in the upper northern part of Thailand. Ph.D Thesis, Chiang Mai University, Chiang Mai, Thailand.