# CHAPTER 4
# DISCUSSION

## DNA Samples for control and FXS groups

The normal distribution of CGG repeat in the Thai population ranges from 19-50 repeats, with the three most common alleles 46.8% of 29 repeats, 31.4% of 30 repeats and 9.5% of 36 repeats (Appendix B, Figure 10; Limprasert et al., 1999). The DNA samples of the common CGG alleles (29, 30 and 36 repeats) were randomly selected from plentiful sample pools of these alleles according to that CGG distribution, while the DNA samples of the rare alleles (19-28 and 31-35 repeats), the intermediate alleles (37-56 repeats), and full mutation alleles (>200 repeats) were collected from two laboratories (the Human Genetic Units at Songklanagarind Hospital and Ramathibodi Hospital) due to their scarcity and thus limited sample numbers. Thus, these scarce DNA samples included mainly individuals from the Central and Southern parts of the country. The proportions of DNA samples in the control groups were unequal with a higher number of DNA samples from the Southern part (88/133), whilst the proportions of FXS samples were almost equal divided between both parts of the country. The 133 normal individuals and 50 FXS patients in this study, 111 normal X chromosomes (83.5%) and 25 FXS chromosomes (50.0%), matched the number of samples from the prior study most referred to (Limprasert et al., 2001).

## Selection of polymorphic markers

Haplotype analysis of the *FMR1* gene in many previous reports employed several commonly investigated microsatellites, DXS548, FRAXAC1, FRAXAC2 and FRAXE. In this study, we examined only 2 microsatellite markers, DXS548 and FRAXAC1, for the haplotype comparisons among ethnic groups. We did not choose FRAXAC2 due to its complex microsatellite structure of $(GT)_x$-C-$(TA)_y$-$(T)_z$ and high mutation rate of 3.3% (Zhong et al., 1993), and we also decided to ignore FRAXE, since it is located ~600 Kb distal to the FMR1-CGG region and is known to have high polymorphism (heterozygosities of 82.9% and 89.6% in Thai normal and FXS chromosomes, respectively), leading to difficulty in grouping many diverse haplotypes

(Limprasert et al., 2001). Due to the low mutation rate of single nucleotide polymorphisms (SNPs), several recent studies included SNP such as ATL1, FMRb and IVS10 for alternative polymorphic markers in their haplotypes. In the end, we decided to use all 6 SNPs which disperse along the *FMR1* gene. Two SNP loci, ATL1 (intron 1) and IVS10 (intron 10), had previously been used in a number of studies to analyze haplotypes among different ethnic groups (ATL1 studied by Gunter et al., 1998; Crawford et al., 2000; Limprasert et al., 2001; Sharma et al., 2003; Curlis et al., 2005; Zhou et al., 2006, IVS10 reviewed by Wang et al., 1997; Vincent and Gurling, 1998 and studied by Xu et al., 1999; Limprasert et al., 2001; Zhou et al., 2006). Although the study reported by Limprasert (2001) from Thai subjects revealed no linkage disequilibrium between these SNPs and FXS mutations, we still included such SNPs in the present study for the haplotype comparisons not only with prior study but also with distinct populations. We also selected 4 novel SNPs, WEX5 (5' UTR) as determined by Brightwell (2002b); and rs25731 (intron 3), rs25702 (intron 13) and rs25723 (intron 16) from the dbSNP from NCBI and HapMap web site, with high heterozygosities (the value close to 0.5) and near frequencies of two SNP alleles in two studied Asian populations (Chinese and Japanese). We did not consider the FMRb SNP locus as studied by Kunst and Warren (1994) as well as Curlis (2005) because the allele frequency of its G nucleotide has a value of 1 in Asian samples.

## Distribution of microsatellite and SNP markers

In order to consider the similarities and the differences in allele frequencies and distributions of all studied markers, we compared each marker between our present study and various reports from different ethnic groups (Tables 11-13).

In Table 11, the DXS548 polymorphic marker had a wide range of possible alleles (17-26 AC repeats) with a predominance of the DXS548-20 allele among normal X chromosomes in every ethnic group (65.0-90.2%) (Buyle et al., 1993; Oudet et al., 1993; Haataja et al., 1994; Macpherson et al., 1994; Malmgren et al., 1994; Zhong et al., 1994a, 1996; Bonaventure et al., 1998; Mingroni-Netto et al., 1999; Zhong et al., 1999; Sharma et al., 2003; Tzeng et al, 2005). A multimodal distribution of DXS548 alleles was observed on normal Indian chromosomes, with a major peak at allele 20 (46.6%) which revealed a relatively lower prevalence than other ethnic

groups, and minor peaks at allele 19 (22.6%) and allele 21 (14.0%), while the normal Thai, Chinese and Taiwanese chromosomes exhibited clearly unimodal distribution of the DXS548-20 allele. This indicates that the Asian populations of Thailand, China and Taiwan have a stronger skewed DXS548-20 allele frequency distribution, which ranged from 81-90% of the control samples, resulting in a lower heterozygosity of this locus (Zhong et al., 1999; Limprasert et al., 2001; Tzeng et al., 2005). However, there was no apparent difference of the DXS548 allele distribution between the control and FXS chromosomes in Thai (Table 6, Chi-square = 2.1, df = 4, P = 0.72), Taiwanese and Indian populations. In contrast to Chinese and Caucasians, the distribution of the DXS548 alleles differed significantly in the normal and FXS cohorts. The fragile X chromosomes showed preferential association with specific DXS548 alleles, especially the DXS548-21 and DXS548-25 alleles. The frequency of DXS548 allele 25 in Brazilian FXS chromosomes was the highest ever reported (Mingroni-Netto et al., 1999). In most studies, both the DXS548-21 and DXS548-25 alleles appeared in similar frequencies among affected groups, except that the DXS548-21 allele was predominantly found in 90-95% of mutated chromosomes in Finnish samples (Haataja et al., 1994; Zhong et al., 1996). This result provided strong evidence of a founder fragile X mutation in the Finnish population. In addition, a positive association with the syndrome was observed with another DXS548 allele at 26 (9.52% of fragile X chromosomes vs. 0.6% of normal chromosomes) in Argentines (Bonaventure et al., 1998).

In Table 12, the FRAXAC1 alleles in Thais, Chinese, Taiwanese and Japanese can be seen as being less polymorphic than the other Indian and Caucasian ethnic groups (Hirst et al., 1993; Jacobs et al., 1993; Macpherson et al., 1994; Richards et al., 1994; Zhong et al., 1994a; Chiurazzi et al., 1996; Zhong et al., 1996; Bonaventure et al., 1998; Mingroni-Netto et al., 1999; Zhong et al., 1999; Sharma et al., 2003; Tzeng et al., 2005). Only two FRAXAC1 alleles with 18 and 19 repeats were detected in Chinese, Taiwanese and Japanese chromosomes, while there were three different alleles, FRAXAC1-17, FRAXAC1-18 and FRAXAC1-19, observed in Thai chromosomes, with the most frequent FRAXAC1 allele among Thais being the FRAXAC1-17 allele followed by the FRAXAC1-18 allele. There was no correspondence between these findings in Thais and Chinese, Taiwanese and Japanese populations which showed the FRAXAC1-18 allele followed by the FRAXAC1-19 allele. With the exception the Asian populations of Thailand, China, Taiwan and Japan, all the other reports of Indians and Caucasians show

FRAXAC1-19 to be the modal allele among control X chromosomes. The Indian population revealed a high prevalence of the FRAXAC1-19 allele among the fragile X cohort (65.8%), similar to the finding of the Finnish fragile X chromosomes (84%) but in contrast to the other Caucasian fragile X samples presenting in only 25-40% (Zhong et al., 1996; Sharma et al., 2003). We found no significantly different frequency distributions of the FRAXAC1 marker on fragile X chromosomes compared to Thai controls (Table 6, Chi-square = 0.77, df = 2, P = 0.68). In a reverse way, the allele frequencies of FRAXAC1 in Chinese and Caucasian fragile X chromosomes significantly differed from the normal controls. FRAXAC1-18 and FRAXAC1-21 were the most prevalent alleles in affected Chinese and Caucasians, respectively, indicating a strong association with the syndromes of such alleles.

In Table 13, two different alleles of all SNP markers in normal Thai chromosomes showed frequencies similar to those of other normal Asian chromosomes (Chinese and Japanese) and in contrast to those of normal Caucasian and African chromosomes except that the frequencies of ATL1 in Thai and African American subjects appeared similar (Gunter et al., 1998; Xu et al., 1999; Crawford et al., 2000; HapMap database). The allele distributions of all SNP markers between the normal and FXS groups in Thais showed no statistically significant differences (Table 6, $P > 0.05$). Unlike in white American and African American populations, significant differences of allele distributions of ATL1 between normal controls and FXS patients were observed. The ATL1-G allele was frequently found in FXS chromosomes (82.9% in the case of white Americans and 88% in the case of African Americans) compared with the lower frequencies in normal chromosomes (39.7% for white Americans and 74.0% for African Americans). Overall they found a significant linkage disequilibrium between the fragile X mutation and some alleles of the neighboring microsatellites and SNP markers in central Chinese and Caucasians (DXS548-21, DXS548-25, DXS548-26, FRAXAC1-18 for Chinese, FRAXAC1-21 for Caucasians and ATL1-G), whereas the Thai fragile X patients were found to have no linkage disequilibrium between the *FMR1* gene and any allele of all adjacent polymorphic markers (both microsatellites and SNPs). This study, therefore, suggests no founder effect of the fragile X mutation within the Thai populations, in contrast with the presence of a founder effect of the FXS in the Caucasian populations (Richards et al., 1992; Buyle et al., 1993).

**Table 11. Comparison of DXS548 markers among different populations.**

| Population (N/FXS) | Percentages of allele frequencies (N/FXS) | | | | | | | | | | Reference |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | |
| Thai (133/50) | - | 0.8/0 | - | 90.2/96.0 | 6.8/4.0 | 0.8/0 | - | - | 1.5/0 | - | Present study |
| Taiwanese (100/28) | 1.0/0 | - | 2.0/3.6 | 90.0/96.4 | 7.0/0 | - | - | - | - | - | Tzeng et al. (2005) |
| Chinese (227/27) | - | - | 4.0/4.0 | 81.0/22.0* | 13.0/74.0* | 2.0/0 | 1.0/0 | - | - | - | Zhong et al. (1999) |
| Indian (350/42) | 0.3/0 | 1.1/0 | 22.6/21.4 | 46.6/42.8 | 14.0/14.3 | 1.1/2.3 | 0.6/2.3 | 3.1/0 | 8.6/16.6 | 2.0/0 | Sharma et al. (2003) |
| French (162/106) | - | - | 2.0/0 | 72.0/39.0* | 14.0/30.0* | - | 1.0/1.0 | 2.0/1.0 | 9.0/27.0* | 2.0/1.0 | Oudet et al. (1993) |
| Belgian-Dutch (134/68) | - | 0.8/0 | 0/1.5 | 73.1/39.7* | 10.5/20.6* | - | 1.5/1.5 | 0.8/0 | 10.5/36.8* | 3.0/0 | Buyle et al. (1993) |
| British (188/44) | - | - | 1.1/9.1 | 73.3/34.1* | 14.9/36.3* | - | - | - | 6.4/18.2 | 4.3/2.3 | Macpherson et al. (1994) |
| Swedish (28/28) | - | - | - | 78.6/50.0 | 14.3/46.4* | - | - | - | 7.1/3.6 | - | Malmgren et al. (1994) |

**Table 11. (continued)**

| Population (N/FXS) | Percentages of allele frequencies (N/FXS) | | | | | | | | | | Reference |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | |
| Finnish (283/60) | - | - | - | 69.0/2.0* | 16.0/90.0* | - | - | 3.0/0 | 10.0/8.0 | 2.0/0 | Haataja et al. (1994) |
| Finnish (54/37) | - | - | - | 69.0/0* | 17.0/95.0* | - | - | 4.0/0 | 7.0/5.0 | 4.0/0 | Zhong et al. (1996) |
| American (190/125) | - | - | 3.2/7.2 | 73.1/39.2* | 9.4/20.8* | 1.1/0.8 | 0/4.8 | 3.7/8.8 | 7.9/16.8 | 1.6/1.6 | Zhong et al. (1994a) |
| Argentine (168/42) | - | - | 0.6/0 | 77.4/61.9 | 10.71/9.52 | 1.78/0 | - | 2.38/0 | 6.54/19.0* | 0.6/9.52* | Bonaventure et al. (1998) |
| Brazilian (60/70) | - | 1.7/0 | 0/1.4 | 65.0/10.0* | 15.0/28.6 | 3.3/2.9 | 1.7/0 | 0/1.4 | 10.0/55.7* | 3.3/0 | Mingroni-Netto et al. (1999) |

* Significant differences between normal (N) and FXS groups.

**Table 12. Comparison of FRAXAC1 markers among different populations.**

| Population (N/FXS) | Percentages of allele frequencies (N/FXS) | | | | | | | Reference |
|---|---|---|---|---|---|---|---|---|
| | 16 | 17 | 18 | 19 | 20 | 21 | 22 | |
| Thai (133/50) | - | 61.7/62.0 | 36.8/38.0 | 1.5/0 | - | - | - | Present study |
| Taiwanese (100/28) | - | - | 70.0/82.1 | 30.0/17.9 | - | - | - | Tzeng et al. (2005) |
| Chinese (216/27) | - | - | 71.0/93.0* | 29.0/7.0* | - | - | - | Zhong et al. (1999) |
| Japanese (142/40) | - | - | 69.0/38.0 | 34.0/59.0 | - | - | - | Richards et al. (1994) |
| Indian (380/41) | - | 0.3/0 | 16.5/19.5 | 70.2/65.8 | 0.8/0 | 12.1/14.6 | - | Sharma et al. (2003) |
| Finnish (54/37) | - | - | 28.0/11.0 | 61.0/84.0 | 2.0/0 | 9.0/5.0 | - | Zhong et al. (1996) |
| English (130/73) | - | 1.0/0 | 16.0/27.0 | 78.0/38.0* | 0/5.0 | 5.0/27.0* | - | Hirst et al. (1993) |
| British (304/137) | 0/0.7 | 0.3/0.7 | 17.3/27.0 | 74.0/39.5* | 0.7/0.7 | 7.0/31.4* | 0.7/0 | Jacobs et al. (1993) |
| British (188/44) | - | - | 19.2/34.0 | 75.5/29.6* | 0/9.1 | 5.3/27.3* | - | Macpherson et al. (1994) |

**Table 12. (continued)**

| Population (N/FXS) | Percentages of allele frequencies (N/FXS) | | | | | | | Reference |
|---|---|---|---|---|---|---|---|---|
| | 16 | 17 | 18 | 19 | 20 | 21 | 22 | |
| Italian (235/137) | 0/0.7 | 0/0.7 | 23.0/27.0 | 68.9/39.4* | 2.1/0.7 | 6.0/31.4* | - | Chiurazzi et al. (1996) |
| American (190/125) | - | 1.6/0 | 22.1/34.4 | 66.8/33.6* | 5.8/3.2 | 3.7/28.8* | - | Zhong et al. (1994a) |
| Argentine (168/42) | - | 0.6/0 | 30.35/14.28 | 63.1/35.71* | 0.6/2.38 | 5.35/47.61* | - | Bonaventure et al. (1998) |
| Brazilian (64/72) | 1.6/2.8 | 1.6/0 | 18.8/27.8 | 67.2/25.0* | - | 10.9/44.4* | - | Mingroni-Netto et al. (1999) |

* Significant differences between normal (N) and FXS groups.

**Table 13. Comparison of SNP markers among different populations.**

| Population | WEX5 (N/FXS) | | ATL1 (N/FXS) | | rs25731 (N/FXS) | | IVS10 (N/FXS) | | rs25702 (N/FXS) | | rs25723 (N/FXS) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | C | G | A | G | A | T | C | T | A | G | A | C |
| Thai | 63.2/62.0 | 36.8/38.0 | 36.1/36.0 | 63.9/64.0 | 36.1/38.0 | 63.9/62.0 | 39.1/38.0 | 60.9/62.0 | 60.9/62.0 | 39.1/38.0 | 60.9/62.0 | 39.1/38.0 |
| Chinese | 52.3/- | 47.7/- | 37.5/- | 62.5/- | 48.9/- | 51.1/- | 37.8/- | 62.2/- | 62.5/- | 37.5/- | 62.2/- | 37.8/- |
| Japanese | 52.3/- | 47.7/- | 40.9/- | 59.1/- | 47.7/- | 52.3/- | 48.8/- | 51.2/- | 52.3/- | 47.7/- | 52.3/- | 47.7/- |
| European | 95.0/- | 5.0/- | 65.0/- | 35.0/- | 7.5/- | 92.5/- | 91.7/- | 8.3/- | 9.2/- | 90.8/- | 8.3/- | 91.7/- |
| White American | - | - | 60.3/17.1* | 39.7/82.9* | - | - | 90.9/92.0 | 9.1/8.0 | - | - | - | - |
| African American | - | - | 26.0/12.0* | 74.0/88.0* | - | - | - | - | - | - | - | - |
| Sub-Saharan African | 96.6/- | 3.4/- | 15.0/- | 85.0/- | 0/- | 100/- | 90.0/- | 10.0/- | 10.0/- | 90.0/- | 24.2/- | 75.8/- |

* Significant differences between normal (N) and FXS groups.

Thai - Present study

Chinese, Japanese, European, Sub-Saharan African - HapMap (www.hapmap.org)

White American - Gunter et al. (1998) for ATL1

White American - Xu et al. (1999) for IVS10

African American - Crawford et al. (2000)

**Haplotype association among normal groups**

Several studies in different ethnic groups had evaluated the variety of combined haplotypes that can be formed from the two or three markers from the combination of polymorphic loci of DXS548, FRAXAC1, FRAXAC2 and FRAXE. In order to compare our findings with the majority of studies that had used two-marker haplotypes, we compared the most frequently studied haplotype combination of DXS548-FRAXAC1 (Table 14). The most common haplotypes found in both control and FXS groups of Thais were 20-17 followed by 20-18, in contrast to the major haplotype 20-18 among normal individuals of central Chinese and Taiwanese (Zhong et al., 1999; Tzeng et al., 2005). When compared with the control X chromosomes, no apparent differences in distribution of a certain haplotype were found among the fragile X chromosomes of this study (Table 7, Chi-square = 4.51, df = 7, P = 0.72). This indicates that there is no distinct founder haplotype prevalent in the patient groups of the Thai population. Our findings contradict other reports in different ethnic groups, notably Hirst (1994), Macpherson (1994), Zhong (1994b), Chiurazzi (1996b), Zhong (1996a), Gunter (1998), Bonaventure (1999), Mingroni-Netto (1999), Pekarik (1999), Zhong (1999), Sharma (2003) and Tzeng (2005). The 20-19 was the modal haplotype observed among control X chromosomes from the Indian and Caucasian populations studied. These studies also demonstrated significantly different frequency distributions of combined haplotypes on fragile X chromosomes compared to their controls, and also the fragile X mutation in these studies was, according to their evidence, thought to be associated with founder haplotypes (21-18 and 25-21) in the Chinese and several Caucasian populations, along with founder haplotypes of 20-21 and 21-19 which were found exclusively in the Argentine and Finnish fragile X chromosomes, respectively. Although the haplotype 21-18 was found among the Thai FXS chromosomes, the percentage of this haplotype was similar to the normal group. Another common founder haplotype, 25-21, as well as the other founder haplotypes 20-21 and 21-19, were not found in Thai FXS patients. This clearly suggests that the existence of a striking founder effect is evident in the Chinese and Caucasian fragile X chromosomes, but no comparative founder effect is apparent in the Thai fragile X chromosomes. There have been a few other studies indicating no founder effect, such as in the Ashkenazi Jews and Taiwanese (Pesso et al., 1997; Tzeng et al., 2005).

**Table 14. Comparison of haplotype (DXS548-FRAXAC1) among different populations.**

| Population (N/FXS) | DXS548-FRAXAC1 Haplotype (percentage in N/FXS) | | | | | | | | | | Reference |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 20-17 | 20-18 | 20-19 | 20-21 | 21-17 | 21-18 | 21-19 | 25-19 | 25-20 | 25-21 | |
| Thai (133/50) | 55.6/62.0 | 33.1/34.0 | 1.5/0 | - | 3.8/0 | 3.0/4.0 | - | - | - | - | Present study |
| Taiwanese (100/28) | - | 64.0/78.6 | 26.0/17.8 | - | - | 3.0/0 | 4.0/0 | - | - | - | Tzeng et al. (2005) |
| Chinese (206/24) | - | 54.9/20.8* | 26.2/4.2* | - | - | 9.7/62.5* | 2.4/8.3 | - | - | - | Zhong et al. (1999) |
| Indian (262/40) | - | 2.3/10.0 | 33.2/30.0 | 0/2.5 | - | 3.0/5.0 | 6.5/10.0 | 0/2.5 | - | 4.2/12.5 | Sharma et al. (2003) |
| Caucasian (102/70) | - | 7.8/2.9 | 67.7/20.0* | 1.0/4.3 | - | 2.0/24.3* | 7.8/10.0 | 4.9/5.7 | - | 2.9/22.9* | Hirst et al. (1994) |
| Caucasian (157/70) | - | 19.7/20.0 | 66.2/34.3* | - | - | 5.1/24.3* | 4.5/8.6 | - | - | - | Zhong et al. (1994b) |
| White American (564/152) | - | 6.9/2.6 | 62.2/20.4* | 1.4/3.2 | - | 7.1/32.9* | 6.1/4.6 | 1.9/3.2 | 0.2/0 | 3.9/23.0* | Gunter et al. (1998) |
| British (188/44) | - | 7.9/0 | 64.9/18.2* | 0.5/6.8 | - | 9.0/31.8* | 5.9/0 | 3.2/2.3 | - | 2.7/15.9* | Macpherson et al. (1994) |

**Table 14. (continued)**

| Population (N/FXS) | DXS548-FRAXAC1 Haplotype (percentage in N/FXS) | | | | | | | | | | Reference |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 20-17 | 20-18 | 20-19 | 20-21 | 21-17 | 21-18 | 21-19 | 25-19 | 25-20 | 25-21 | |
| Finnish (54/36) | - | 13.0/0 | 55.6/0* | - | - | 13.0/11.1 | 3.7/83.3* | - | - | 5.6/5.6 | Zhong et al. (1996a) |
| Italian (202/125) | - | 9.9/4.0 | 63.9/20.8* | - | - | 7.4/16.8* | 5.4/1.6 | - | - | 2.5/24.0* | Chiurazzi et al. (1996b) |
| Czech (33/35) | - | 3.0/0 | 66.7/8.6* | - | - | 0/17.1* | 6.1/11.4 | 0/8.6 | - | 6.1/22.9* | Pekarik et al. (1999) |
| Brazilian (60/71) | - | 3.3/2.8 | 58.3/4.2* | - | - | 8.3/23.9* | 3.3/1.4 | 0/12.7 | - | 3.3/40.8* | Mingroni-Netto et al. (1999) |
| Argentine (168/42) | - | 20.8/7.1* | 54.8/28.6* | 0.6/26.2* | - | 6.0/7.1 | 4.2/2.4 | 1.8/4.8 | - | 3.0/14.3* | Bonaventure et al. (1998) |

* Significant differences between normal (N) and FXS groups.

In this present examination, even though we included more polymorphic markers, both microsatellites and SNPs (four additional SNPs), and sample pools of FXS patients (from 25 to 50) than the previous report of Limprasert (2001), we still found no statistically significant differences of both allele frequencies and haplotype associations between the control and FXS groups. DXS548 was not useful for haplotype analysis in this study, as it had a very low polymorphic frequency in the Thai (Het = 18.1% for controls and Het = 7.7% for FXS groups). When we omitted this marker, we found three common haplotypes with similar percentages in both control and FXS groups indicating no founder haplotype association with FXS chromosomes. This finding suggests that there was no founder effect at the *FMR1* gene in Thai subjects but, we could not completely exclude it since the FXS mutation in the Thai population may occur on these common haplotypes. Analysis of the combined haplotype among microsatellite and SNP markers in the 6 subgroups of normal alleles (19-28, 29, 30, 31-35, 36 and 37-56) showed a positive significant association between a common CGG alleles (29, 30 and 36) and haplotypes (FRAXAC1-WEX5-ATL1-rs25731-IVS10-rs25702-rs25723) (Table 9, P = 0.00). The CGG-29 allele was associated with haplotype 17-G-G-A-T-A-A (Hap A), the CGG-30 allele was associated with haplotype 18-C-A-T-C-G-C (Hap B), the CGG-36 allele was associated with haplotype 17-C-G-T-T-A-A (Hap C), while the uncommon CGG allele in the 19-28, 31-35, and 37-56 groups was not associated with any haplotype. These findings suggest that there is a haplotype association of FRAXAC1-WEX5-ATL1-rs25731-IVS10-rs25702-rs25723 among common alleles of normal Thai chromosomes. Moreover, all alleles with 36 CGG repeats occurred on Hap A or Hap C and the most common alleles in the 37-56 CGG repeats also occurred on Hap A or Hap C (77%) suggesting that Hap A and Hap C are associated with 36-56 CGG repeats. However, we could not prove that Hap A and Hap C are high risk for repeat expansion because this association was not observed in the FXS groups.

Recently, the examination of haplotype analysis among unselected populations of Asians studied by Zhou et al. (2006) employed polymorphic markers from both microsatellites (DXS548, FRAXAC1 and FRAXAC2) and SNPs (ATL1 and IVS10). Due to the common Asian ethnicity and the similarity of the investigated markers, we compared the haplotypes determined by typing the markers FRAXAC1, ATL1 and IVS10 between our results of normal groups in Thais and then reanalyzed the data from the Chinese, Malay and Indian studies (Table 15). The most frequent common haplotype found in Thais, Chinese and Malays was 17-G-T (accounting for 60.9%, 75.7% and 59.6% of normal X chromosomes, respectively), followed by 18-A-C (responsible for 35.3%, 20.3% and 37.1% of normal X chromosomes, respectively). In direct reverse to the Indian study, our study found that the most prevalent haplotype was the 18-A-C (46.5%), followed by 17-G-T (18.2%), 18-G-C (17.2%) and 20-G-C (12.1%). The statistical analysis of chi-square test of all races revealed that there was no significant difference between Thais and Malays (P = 0.44) but there were slightly significant differences between Thais and Chinese (P = 0.02) as well as Thais and Indians (P = 0.00). The results of the racial comparison between Thais and Indians was expected because the extant Indian populations are very ethnically complex with varied genetic affinities, especially those with some Caucasian ethnic background (Sharma et al., 2003). The unexpected outcomes from Thais, Chinese and Malays lead us to retrace our anonymous normal samples comprised of individuals from the central and the Southern parts of the country. The proportion of such samples from the Southern part, in which more people are related to Malays, was found to be about 2/3 of the total samples, which helps explain comparison between Thais and Malays of no significant difference, while the ethnical comparison from Thais and Chinese shows a slightly significant difference.

**Table 15. Distribution of FMR1 flanking haplotypes (FRAXAC1-ATL1-IVS10) in four Asian populations.**

| Haplotype | Thai | | Chinese | | Malay | | Indian | |
|---|---|---|---|---|---|---|---|---|
| AC-AT-IV | NO. | % | NO. | % | NO. | % | NO. | % |
| 16-G-T | 0 | 0.0 | 2 | 1.1 | 0 | 0.0 | 0 | 0.0 |
| 17-G-T | 81 | 60.9 | 134 | 75.7 | 106 | 59.6 | 18 | 18.2 |
| 17-G-C | 1 | 0.8 | 0 | 0.0 | 0 | 0.0 | 2 | 2.0 |
| 17-A-C | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 1 | 1.0 |
| 18-A-C | 47 | 35.3 | 36 | 20.3 | 66 | 37.1 | 46 | 46.5 |
| 18-G-T | 0 | 0.0 | 3 | 1.7 | 2 | 1.1 | 1 | 1.0 |
| 18-G-C | 2 | 1.5 | 2 | 1.1 | 4 | 2.2 | 17 | 17.2 |
| 19-A-C | 1 | 0.8 | 0 | 0.0 | 0 | 0.0 | 1 | 1.0 |
| 19-G-C | 1 | 0.8 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 |
| 20-G-C | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 12 | 12.1 |
| 20-A-C | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 1 | 1.0 |
| total | 133 | 100.0 | 177 | 100.0 | 178 | 100.0 | 99 | 100.0 |

(AC = FRAXAC1, AT = ATL1, IV = IVS10)

Thai vs Chinese: Chi-square = 16.61, df = 7, P = 0.02 (significant)

Thai vs Malay: Chi-square = 5.81, df = 6, P = 0.44 (not significant)

Thai vs Indian: Chi-square = 64.68, df = 9, P = 0.00 (significant)

Our study employed more SNP markers (total 6 SNP loci) than other reports (only 2 SNP loci). Haplotype analysis using the total 6 SNPs in the present study provided more information than previous study of Limprasert (2001) which could not completely separate the CGG-29 allele and the CGG-36 allele to different haplotype backgrounds; our study also showed the appearance of haplotype block in this studied region. Since all pairs of SNP polymorphisms within a block would be in strong linkage disequilibrium, we could genotype only a few SNP loci that were chosen carefully to provide enough information for prediction of the remaining of the common SNPs in that region. As a result, only a few of these tag SNPs were required to identify each of the common haplotypes in a region (Cardon and Abecasis, 2003; The International HapMap Consortium, 2003). From the genotyping of individual and family samples, we noticed that

only three tag SNPs of the total 6 SNPs could provide enough information to identify each common haplotype in our study. We could select ATL1, a SNP from the group of WEX5 and rs25731, and a SNP from the group of IVS10, rs25702 and rs25723. For the group of WEX5 and rs25731, however, we proposed that the Biallelic-ARMS PCR technique of WEX5 was more convenient than the PCR-RFLP technique of rs25731.

## Variation of AGG interruption pattern

During analysis of the length variation of uninterrupted tracts of CGG repeats, we noticed that the 3' tract of repeats distal to the last interruption was more variable than the 5' or the middle tracts of repeats, a finding clearly consistent with previous studies of polarized variability (Eichler et al., 1994, 1995). Furthermore, we found four chromosomes with 43 CGG repeats had an AGG pattern of 9A9A6A6A9. It was interesting that the CGG tract length variability among these highly interspersed alleles had never been observed to exceed 10 repeats for the majority of examined samples. This suggests that the multiple AGG interruptions may play a crucial role in maintaining repeat stability.

The general FMR1 CGG repeat substructure has two AGG interruptions occurring with a periodicity of once every 9 or 10 CGG repeat units (Eichler et al., 1994; Hirst et al., 1994; Macpherson et al., 1994; Snow et al., 1994; Eichler et al., 1995; Zhong et al., 1995). In our study, we found an allele with an AGG pattern of 20A9 occurring in 1/95 cases, or 1.1%. This may have been due to the loss of the first AGG interspersion at the 5' end of the repeats during the A to C conversion mechanism. However, this allele occurred less frequently than the allele with the AGG pattern of 9/10An ($n \geq 20$) which showed the loss of the second AGG interspersion (9/95 or 9.5%). These about 10-fold significant differences may indicate that there has been a mutational bias in the loss of AGG interruptions (Eichler et al., 1995).

When we compared the AGG interruption patterns among the different ethnic groups, there were notable differences in the distribution of various interspersed FMR1 CGG repeat alleles (Appendix A, Table 22) (Hirst et al., 1994; Eichler et al., 1995; Hirst et al., 1997; Larsen et al., 1999; Carwford et al., 2000; Faradz et al., 2001; Zhou et al., 2006). The Caucasian and African American populations displayed a greater number of CGG repeat patterns than Asians (Eichler et al., 1995; Crawford et al.,

2000). However, certain similarities of some general features in the distribution of these patterns also became apparent after we reanalyzed these raw data again (Table 16). In all populations, AGG interruptions were punctuated in the FMR1 CGG repeat with a periodicity of once every 9 or 10 CGG repeats. The 9A9A9 and 10A9A9 were the predominate patterns found in all ethnic groups, resulting in a universal mode at 29 and 30 repeats, respectively, in the interruptions. This suggests that such AGG configurations are highly conserved (Eichler et al., 1995, 1996). Despite these similarities in the overall distribution and compositions of FMR1 CGG repeats among the races, a few distinct differences were observed. Particularly in the Asians, we noticed the existence of a $(CGG)_6AGG$ insertion within the CGG triplet repeat region and the prevalence of the 9A9A6A9 allele (75-91.7%) which was not found in the Caucasians and African Americans (Kunst et al., 1996; Chen et al.1997; Hirst et al., 1997; Larsen et al., 1999; Crawford et al., 2000; Faradz et al., 2001; Zhou et al., 2006) In addition, due to the different CGG distributions among races (the multimodal distributions of 20, 23, 29, 30, and 31 repeats for Caucasians, the trimodal distributions of 29, 30, 31 repeats for African Americans, and the trimodal distributions of 29, 30, and 36 repeats for Asians), the CGG-31 alleles with the patterns of 10A9A10 and 10A10A9 exhibited higher frequencies both in Caucasians (79.7%) and African Americans (77.5%) than Asians (Kunst et al., 1996; Chen et al., 1997). The 10A9 allele occurred almost exclusively among the Caucasians (96.7%), while in contrast it was rare or absent in the other populations indicating that this allele became fixed and then expanded in the Caucasian population (Eichler et al., 1995).

**Table 16. Comparison of the AGG interspersion patterns among different populations.**

| CGG Repeats | AGG Pattern | Thai (%) | Chinese (%) | Japanese (%) | Malay (%) | Javanese (%) | Eskimo (%) | Indian (%) | Caucasian (%) | African American (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 10A9 | 2/2 | | | | | 3/3 | | 29/30 (96.7) | 5/5 |
| 29 | 9A9A9 | 15/17 (88.2) | 93/97 (95.9) | | 80/86 (93.0) | | 3/3 | 30/30 (100.0) | 34/37 (91.9) | 65/70 (92.9) |
| 30 | 10A9A9 | 16/18 (88.9) | 31/33 (93.9) | | 49/51 (96.1) | | 45/46 (97.8) | 33/41 (80.5) | 81/93 (87.1) | 55/82 (67.1) |
| 31 | 10A9A10 | | 1/9 (11.1) | | | | 1/1 | | 19/36 (52.8) | 23/49 (46.9) |
| | 10A10A9 | | 1/9 (11.1) | | | | | | 7/26 (26.9) | 15/49 (30.6) |
| 36 | 9A9A6A9 | 15/18 (83.3) | 15/19 (79.0) | 15/20 (75.0) | 11/12 (91.7) | 16/20 (80.0) | 16/18 (88.9) | 7/8 (87.5) | | |

References - Asians: Thai, Present study; Chinese, Zhou et al., 2006; Japanese, Hirst et al., 1997; Malay, Zhou et al., 2006; Javanese, Faradz et al., 2001; Eskimo, Larsen et al., 1999; Indian, Zhou et al., 2006

- Caucasians - Hirst et al. (1994) and Eichler et al. (1995)

- African Americans - Eichler et al. (1995) and Crawford et al. (2000)

**Relationship among CGG repeat number, flanking haplotype and AGG configuration**

When we considered the relationship among CGG repeat numbers, haplotype (FRAXAC1-ATL1) and AGG interruption pattern in various ethnic groups, each race showed a variation of relationships (Appendix A, Table 23) (Gunter et al., 1998; Crawford et al., 2000; Zhou et al., 2006). Nevertheless, the similarities and differences of these relationships were obviously observed after we reanalyzed these raw data again (Table 17). In the Thais, Chinese and Malays, we found a strong association among CGG repeat numbers, the combined haplotype (FRAXAC1-ATL1) and AGG configuration. The CGG-29 allele with the 9A9A9 pattern and the CGG-36 allele with the 9A9A6A9 pattern were positively associated with haplotype 17-G. The CGG-30 allele with 10A9A9 pattern was positively associated with haplotype 18-A. In Indians, the CGG-29 allele with the 9A9A9 pattern occurred on three distinct haplotypes 17-G, 18-G and 20-G, with similar frequencies of in 30%, 40% and 26.7%, respectively, while the CGG-36 allele with the 9A9A6A9 pattern and the CGG-30 allele with the 10A9A9 pattern were positively associated with the same haplotypes 17-G and 18-A, respectively, as observed in other Asian groups. Also, the CGG-29 allele with the 9A9A9 pattern found in Caucasians was associated with haplotypes 18-G (37.5%), 19-G (25%) and 21-G (16.7%) but the CGG-30 allele with the 10A9A9 pattern was associated with a different haplotype than observed in Asians, 19-A. The relationships in African Americans were similar to those of Caucasians, as the CGG-29 allele with the 9A9A9 pattern and the CGG-30 allele with the 10A9A9 pattern were associated with the same haplotypes observed in Caucasians, 18-G (66.7%) and 19-A (62.5%), respectively.

**Table 17. Comparative analysis of haplotype (FRAXAC1-ATL1) and AGG configuration among different populations.**

| FRAXAC1-ATL1 | AGG Pattern | Thai (%) | Chinese (%) | Malay (%) | Indian (%) | Caucasian (%) | African American (%) |
|---|---|---|---|---|---|---|---|
| 17-G | 9A9A9 | 12/17 (70.6) | 90/97 (92.8) | 76/86 (88.4) | 9/30 (30.0) | | |
| 17-G | 9A9A6A9 | 15/18 (83.3) | 14/19 (73.7) | 9/12 (75.0) | 5/8 (62.5) | | |
| 18-A | 10A9A9 | 16/18 (88.9) | 29/33 (87.9) | 49/51 (96.1) | 31/41 (75.6) | | |
| 18-G | 9A9A9 | | | | 12/30 (40.0) | 9/24 (37.5) | 14/21 (66.7) |
| 19-A | 10A9A9 | | | | | 55/71 (77.5) | 20/32 (62.5) |
| 19-G | 9A9A9 | | | | | 6/24 (25.0) | |
| 20-G | 9A9A9 | | | | 8/30 (26.7) | | |
| 21-G | 9A9A9 | | | | | 4/24 (16.7) | |

Thai - Present study

Chinese, Malay, Indian - Zhou et al. (2006)

Caucasian - Gunter et al. (1998)

African American - Crawford et al. (2000)

**CGG Repeat instability in Thai FXS**

Our analysis of haplotype and sequence content permit some coarse speculation as to the CGG repeat instability of predisposing fragile X alleles (Figure 9). When we considered three factors involving repeat instability (CGG repeat number, haplotype and AGG interruption pattern), we observed an association among these factors in three different backgrounds. Hap A and Hap C may be evolutionally derived since they have only two SNP marker differences in the haplotype (WEX5 and rs25731) and they have only slight differences in the AGG configuration of $(CGG)_6AGG$ insertion (the CGG-29 allele with Hap A may be changed to the CGG-36 allele with Hap C by the insertion of $(CGG)_6AGG$). The distinction between Hap A and Hap B, however, results from both differences in all SNP markers and the AGG substructure. The fragile X alleles occurred on Hap A and Hap B are likely similar frequency (38% of Hap A and 36% of Hap B), while the occurrence of fragile X alleles on Hap C found less than those haplotype backgrounds (24%) may be due to several observations on this genetic background, i.e., rare allele with 9/10An ($n \geq 20$), no long pure CGG allele and highly AGG interspersed alleles. Another important feature of $(CGG)_6AGG$ insertion mostly observed in Hap C enhances the stability of the intermediate alleles (Hirst et al., 1997; Faradz et al., 2001). Studies of the evolution of the FMR1 CGG repeat in various species of both non-primate and primate mammals have revealed that the majority of non-primates have small uninterrupted CGG repeats with a mean repeat length of ~8 repeats, while the repeats among primates are larger with a mean repeat length of ~20 repeats and more highly specific interruption (Eichler et al., 1995). In one study (similar to several evolutionary studies of other human genetic disorders), the range of the number of the CAG repeats in the *MJD1* gene of Machado-Joseph disease among different species of primates, macaques (13-14) and chimpanzees (14-20) showed smaller repeat sizes than Caucasians (14-40) (Rubinsztein et al., 1995; Limprasert et al., 1996). CAG repeat tracts in the SCA1 gene of spinocerebellar ataxia type 1 also show lower repeat numbers in non-human primates, 9-15 repeats for macaques and 20-26 repeats for chimpanzees, than in Caucasians, which ranged from 9-37 repeats (Limprasert et al., 1997). These studies provide strong evidence that the number of trinucleotide repeats has evolved from small repeat numbers to larger tracts. So, we hypothesize two distinct mutational pathways as shown in Figure 9. First, the FXS mutation may arise from three common haplotype backgrounds (Hap A, Hap B and Hap C) by

gradual replication slippage of large normal alleles (37-56 repeats). This pathway progresses relatively slow toward instability and the hyperexpansion thresholds associated with the disease. Hap A and Hap C are more likely alterable to high-end normal alleles than Hap B (23%) because we found enriched Hap A and Hap C in this group (77%). Second, the FXS mutation may independently occur on any of the three haplotype backgrounds with a recurrent mutational event involving the loss of 5' or 3' AGG interruptions, since the allele with the AGG configuration of 20A9, the allele with pattern 9/10An ($n \geq 20$) and the uninterrupted CGG allele were all observed. After the recurrent loss of AGG interspersions, the CGG allele is prone to progress rapidly toward the disease state by rapid slippage.

Further experiment with a greater number of large normal alleles will be required to fully understand the evolution of the fragile X mutation in Thais.

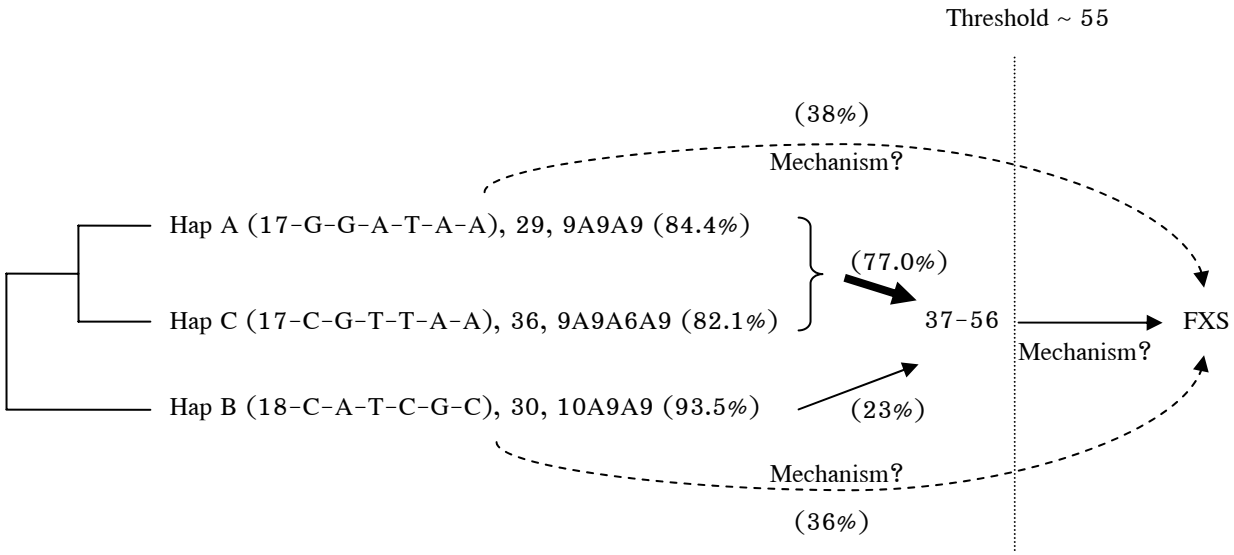


**Figure 9. A model suggesting the causes of repeat instability.** The distances and the lengths of each line are not according to scale.