Human Molecular Genetics, 2002, Vol. 11, No. 4 371–378

Premutation and intermediate-size *FMR1* alleles in 10 572 males from the general population: loss of an AGG interruption is a late event in the generation of fragile X syndrome alleles

C. Dombrowski, S. Lévesque, M. L. Morel, P. Rouillard, K. Morgan[1] and F. Rousseau*

Unité de Recherche en Génétique Humaine et Moléculaire, Hospital St-François d'Assise Research Center, Department of Medical Biology, Faculty of Medicine, Laval University, Quebec City, Quebec, Canada and [1]Departments of Human Genetics and Medicine, McGill University, and Research Institute of the McGill University Health Centre, Montréal, Canada

Received September 21, 2001; Revised and Accepted December 11, 2001

We previously reported a 1:259 prevalence of female carriers of *FMR1* premutation-size alleles (greater than 54 triplet repeats) in the general population. We now have screened 10 572 independent males from the same population for similar alleles using high-throughput Southern blotting. We identified 13 male carriers of an allele with more than 54 repeats. This corresponds to a prevalence of 1:813 males (95% confidence interval 1:527 to 1:1781). Haplotype analysis of four markers flanking the triplet array revealed that the prevalence of the major fragile X mutation-associated haplotype was increased among *FMR1* alleles of 40–54 repeats. Although sequencing of highly unstable premutation alleles from fragile X families revealed only pure CGG tracts, this was not the case for alleles of similar size that were identified in males from the general population. Forty-eight out of forty-nine alleles of 40 or more triplets had one or two AGG interruptions. This observation, combined with the observation of the enrichment of major fragile X syndrome haplotypes in all alleles of this size, is evidence that the loss of an AGG interruption in the triplet repeat array is not necessary for expansion of normal alleles of 29–30 triplets to intermediate size. The loss of AGG interruptions thus appears to be a late event that leads to greatly increased instability and may be related to the haplotype background of specific *FMR1* alleles.

INTRODUCTION

Fragile X syndrome, an important cause of inherited mental retardation (1), is due to the expansion of a (CGG)n triplet repeat (2–4) located in the 5′-untranslated region (5′-UTR) of the first exon of the *FMR1* gene (3). This expansion is associated with the presence of abnormal methylation of the CpG island (2,3,5). The vast majority of cases of the syndrome, estimated to be ~95% (6), are due to the typical triplet repeat expansion; in the remaining cases where no expansion could be detected, either deletions or point mutations were observed (6). The prevalence of the disease was evaluated using direct DNA analysis and estimated to be lower than initially proposed using cytogenetic testing. However, these are minimal estimates since they were obtained in individuals who showed some level of mental retardation or learning difficulties and mildly affected males may not have been included in the sample (7,8). Up to now no direct population-based prevalence estimate of the full mutation has been obtained. Most likely, this is because of the large size of the sample (around 100 000 individuals) that would be required. The appearance of a full fragile X mutation, generally greater than 230 repeats and methylated, in a family has always been preceded by the transmission of a premutation, which is an unmethylated expanded allele of usually between 55 and 230 repeats. In fact, in fragile X families, premutations are generally more frequent than full mutations (9) and are all at risk of expanding to a full mutation in a single female transmission. The bulk of *FMR1* alleles in the general population have 29–30 repeats and are usually stable upon transmission over a few generations (10). Alleles of an intermediate size between frequent normal alleles and the smallest 55-triplet premutations (i.e. between 40 and 54 repeats) have been shown to be slightly unstable upon transmission. However, they have never been observed to expand to a full mutation in a single step, i.e. without first attaining premutation size (greater than 54 triplets). These intermediate-size alleles have been also termed 'gray-zone' alleles (11) but current evidence suggests that they do not confer risk of fragile X syndrome in the generation which follows immediately.

The mutational process from *FMR1* normal alleles to full mutations is believed to be complex and to involve multiple

*To whom correspondence should be adressed at: Unité de Recherche en Génétique Humaine et Moléculaire, 10, rue de l'Espinay, Québec City, Québec, Canada G1L 3L5. Tel: +1 418 525 4470; Fax: +1 418 525 4195; Email: francois.rousseau@crsfa.ulaval.ca

steps (12–14). It has been observed that: (i) no direct transition from a normal-size allele to a full mutation has been detected; (ii) in all populations studied, *FMR1* mutations are in linkage disequilibrium with closely flanking genetic markers. This association is stronger in populations with known founder effects for other diseases such as in the Quebec French-Canadian population (15) and in the Finnish population (16); (iii) within fragile X families, the risk of expansion from a premutation to a full mutation varies with the size of the premutation (17,18); (iv) expansion of the *FMR1* (CGG)n triplet repeat is polarized i.e. it occurs only at the 3′ end of the triplet array. There are various patterns of AGG interruptions of the array that are believed to be responsible for 'stabilizing' the alleles (6,10,18).

Reliable estimates of the prevalence of premutations and intermediate-size alleles, along with the distribution of *FMR1* haplotypes relative to allele sizes in the general French-Canadian population, should contribute to understanding how the full (or large-expanded) mutation arises from particular alleles. Previously, we identified 41 premutation-size alleles in a large cohort of 10 624 unrelated women from the general population (15). We obtained accurate estimates of the prevalence of premutation-size alleles that apparently apply to other populations as well (19). Models of *FMR1* mutation in the normal population have proposed that the prevalence of premutations may be larger in females than in males (ratio of 1.59) (13).

We report here the results of screening 10 572 males from the general population for premutation-size *FMR1* alleles. We found a 1:813 [95% confidence interval (CI) 1:527 to 1:1781] prevalence of normal transmitting males with more than 54 triplets. We also identified a number of alleles of intermediate size (between 40 and 54 triplets). We analyzed the haplotypes associated with these large alleles using flanking microsatellites (DXS548, FRAXAC1 and FRAXAC2) and the ATL1 single nucleotide polymorphism (20), and we sequenced the triplet repeat array of alleles larger than 40 triplets to determine the AGG interruption pattern. These results provide new insights into the molecular and population genetics of the fragile X syndrome and suggest that the loss of an AGG repeat is preceded by an earlier event, which confers instability.

RESULTS

Prevalence of male carriers of premutation-size FMR1 alleles

The 11 495 blood samples analyzed constituted 2299 pools of five men each and included 923 duplicate samples that were tagged prior to the DNA analyses. Thus, 10 572 samples from different men were analyzed by *Dpn*II+*Eag*I mini-Southern blotting. 2199 pools tested negative whereas 100 pools showed a restriction fragment of larger than normal size. Individual analysis by mini-Southern blot of the samples in positive pools identified 117 samples which were candidates for containing a significantly larger than normal *FMR1* allele as compared to the constant control restriction fragment also detected by the QcPX probe (Fig. 1). These were all analyzed by PCR to determine the exact number of repeats and only 13 alleles were identified as having 55 or more triplet repeats (range 55–95). The size of alleles in the remaining 104 samples ranged between 35 and 54 triplet repeats. Therefore, the estimate of the prevalence of male carriers of a premutation-size allele is 13 per 10 572 or 1:813 males (95% CI, 1:527 to 1:1781). Thus,

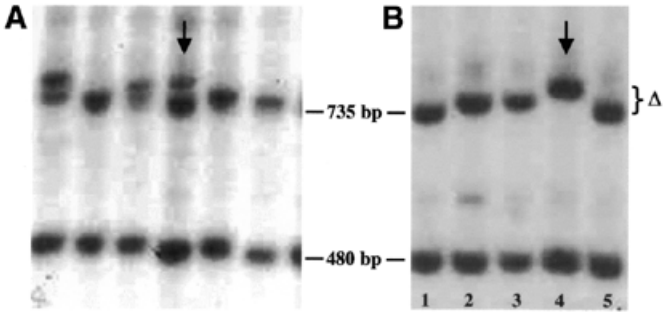

Figure 1. Southern blot analysis of pools of five samples each (**A**) and individual samples of a positive pool (**B**) by *Dpn*II+*Eag*I digestion followed by hybridization to probe QcPX (see Materials and Methods). The positive pool and its corresponding positive sample are indicated with arrows. PCR analysis of the positive sample revealed 51 triplet repeats.

premutation-size allele frequencies were not statistically different ($\chi^2 = 2.3$, 1 df, P-value = 0.13) between males in this study and females in our previous study (1:518) (15). It is worth mentioning that the pooling method used here for screening for premutation-size alleles was validated in our previous female prevalence study (15) by reanalyzing 1000 samples from 200 pools which showed that very few if any allele of 54 triplets and more was missed.

Of the 104 samples showing an apparently larger than normal fragment by Southern blot but which was not in the premutation size-range, 36 were in the intermediate size-range, i.e. containing between 40 and 54 repeats. This represents only a fraction of the 460 alleles of 40–54 triplets that would be expected based on the prevalence of intermediate-size alleles of 1/23 in Special Education Needs Schools children (19). Thus, although there is no published population-based estimate of this prevalence in normal individuals, it is likely that some intermediate-size alleles were not identified in the initial screening step due to the fact that our screening method, when applied to pools of DNAs, does not have a very high sensitivity for intermediate-size alleles. We did not compute an estimated prevalence for intermediate-size alleles since this was neither the appropriate method for detecting such alleles nor the primary objective of this study. However, 36 identified intermediate-size alleles were used to study the haplotypes most frequently encountered in this size-range as well as for determination of their AGG interruption pattern.

Control samples from 102 consecutive newborn males had alleles varying between 13 and 38 triplets (Table 1 and Fig. 2) with the two most prevalent alleles being 29 triplets (29/102) and 30 triplets (29/102). There was a smaller mode at 19 repeats (6/102).

Haplotype analysis

We previously reported strong linkage disequilibrium in the French-Canadian population between *FMR1* mutations in fragile X families and flanking polymorphic markers DXS548, FRAXAC1 and FRAXAC2 (15). In the present study of male samples, we analyzed the haplotypes in each category of *FMR1* alleles. Among 68 unrelated Québec fragile X males with complete DXS548-FRAXAC1-FRAXAC2 haplotypes,

Table 1. Distribution of *FMR1* haplotypes by size category and array structure

FMR1 allele						*FMR1* allele					
Triplet repeat array			Haplotype	Frequency		Triplet repeat array			Haplotype	Frequency	
Range and Sample	Repeat number	Array structure [2]	[1]	n	% [3]	Range and Sample	Repeat number	Array structure [2]	[1]	n	% [3]
> 230 triplets Fragile-X males (n=68)	n/a	Pure CGGs	X	34	50	< 40 triplets general population newborn males (n=102)	37	9+9+17	X	1	
	n/a	Pure CGGs	X'	17	25		34	9+9+14	X	1	
	n/a	Pure CGGs	O	6	9		33	9+9+13	X	1	6
	n/a	Pure CGGs	N	6	9		30	9+9+10	X	1	
	n/a	Pure CGGs	A	5	7		28	11+16	X	1	
55 to < 230 triplets general population adult males (n=13)	95	Pure CGGs	X	1			28	9+9+8	X	1	
	80	9+9+60	X	1			38	9+9+18	O	1	
	66	9+56	X	1	46		36	9+9+8+7	O	1	
	64	9+54	X	1			36	9+9+7+8	O	1	
	60	9+9+40	X	1			35	9+25	O	1	
	56	9+9+36	X	1			32	9+9+12	O	2	
	82	9+9+62	O	1	15		32	9+8+13	O	1	
	59	9+7+41	O	1			31	10+9+10	O	2	
	61	10+16+33	N	1	8		31	9+11+9	O	1	
	62	9+52	A	1			31	9+12+8	O	1	
	57	9+7+39	A	1	31		31	9+9+11	O	1	
	56	9+7+38	A	1			30	10+9+9	O	11	
	55	9+7+37	A	1			30	9+11+8	O	1	
40-54 triplets general population adult males (n=36)	51	9+10+30	X	1			30	9+10+9	O	3	
	51	9+9+31	X	2			29	10+9+8	O	11	62
	49	9+9+29	X	2			29	9+10+8	O	1	
	48	9+9+28	X	2			29	9+9+9	O	6	
	47	9+9+27	X	3	44		29	18+10	O	1	
	44	9+9+24	X	2			28	9+9+8	O	6	
	43	9+9+23	X	1			25	9+15	O	1	
	41	9+9+21	*X'	1			23	13+9	O	1	
	41	9+9+21	X	1			22	Pure CGGs	O	1	
	40	9+9+20	X	1			22	11+10	O	1	
	50	9+9+30	O	1			22	13+8	O	1	
	47	9+9+27	O	1			21	10+10	O	1	
	45	10+9+24	O	1			21	11+9	O	1	
	42	9+32	O	1	19		19	10+8	O	2	
	42	9+9+22	O	1			19	9+9	O	1	
	41	9+9+21	O	1			13	Pure CGGs	O	1	
	40	9+9+20	O	1			31	10+9+10	N	1	
	54	10+43	N	1			30	10+9+9	N	13	
	54	10+9+33	N	1			29	10+9+8	N	8	
	53	10+9+32	N	1			29	9+9+9	N	2	
	50	10+9+29	N	1			28	10+9+7	N	1	
	49	10+9+28	N	1	31		28	9+9+8	N	1	31
	46	10+9+25	N	2			25	14+10	N	1	
	45	10+9+24	N	1			22	9+12	N	1	
	43	10+9+22	N	2			20	10+9	N	1	
	41	10+9+21	N	1			19	10+8	N	3	
	48	9+10+27	A	1			31	9+12+8	A	1	1
	42	9+32	A	1	6						

[1] DXS548-FRAXAC1-FRAXAC2 haplotype classified as X, major fragile X haplotype, T50-T42-T62; X', second major fragile X haplotype, T40-T42-T62; N, most common normal haplotype, T40-T38-T61; A, 'alternative' haplotype, T42-T36-T60; O, haplotypes other than X, X', N and A. [2] The triplet array structure for alleles in fragile X families were inferred from premutated males from the same family. [3] Proportion of each haplotype in each allele-size category. The asterisk indicates the only occurrence of the T40-T42-T62 haplotype observed outside of fragile X families (see also asterisk in Figure 2).

the most common haplotype in the normal allele size range (T40-T38-T61) was found in only 9% (6/68) of disease alleles and what we designated as the 'alternative haplotype' (T42-T36-T60) was found in only 7% (5/68) of fragile X chromosomes. The major fragile X haplotype (T50-T42-T62) was present in 50% (34/68) of the samples. Another haplotype (T40-T42-T62) apparently related to the major fragile X haplotype was present in another 25% (17/68) of the chromosomes analyzed. Thus, we call the T50-T42-T62 haplotype the major fragile X haplotype (X in Fig. 2 and Table 1) and the T40-T42-T62 haplotype the second major fragile X haplotype (X' in Fig. 2 and Table 1).

Among alleles in the premutation size-range the haplotype observed most frequently was the major fragile X haplotype (6/13) (T50-T42-T62) while the most common normal haplotype

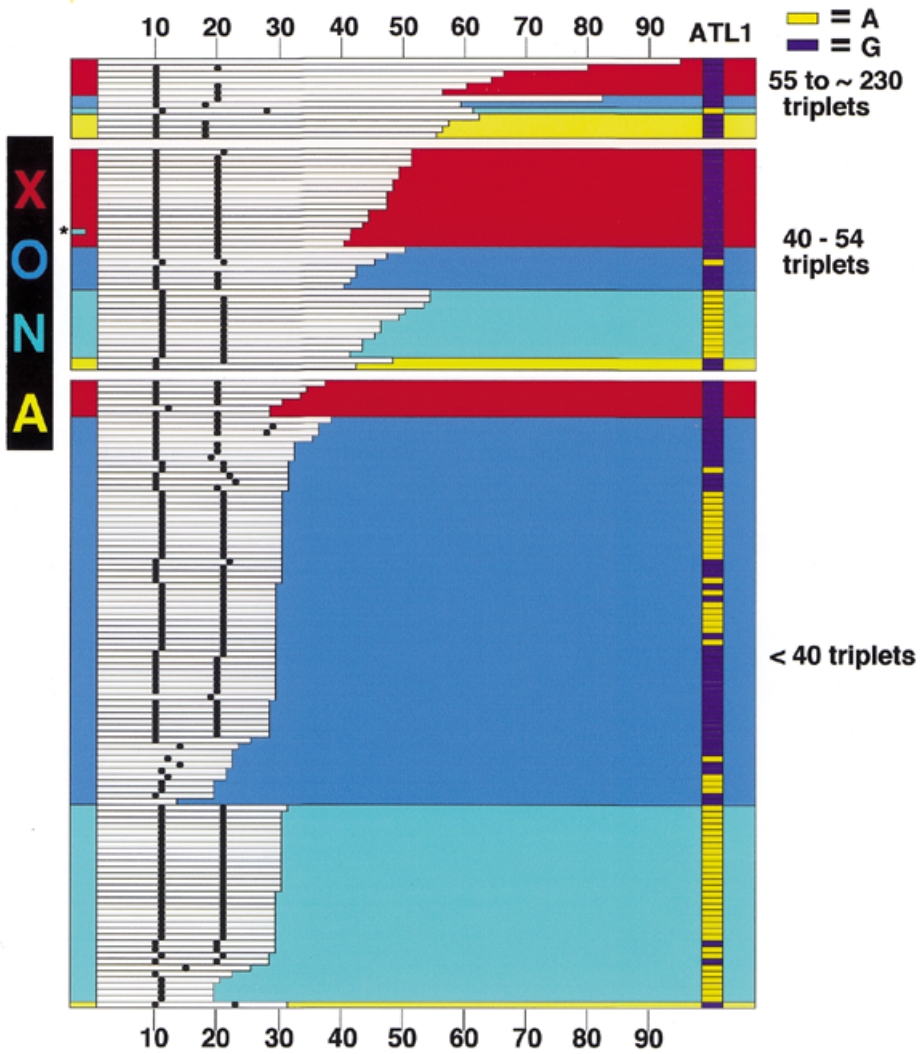

Figure 2. Size, array structure and haplotypes of *FMR1* alleles of 0–39, 40–54, greater than 54 identified in males from the general population. Each individual is represented by a bar. White blocks including black dots represent the number of triplets (size scale at top and bottom of figure). Black dots represent the AGG interruptions and their position in the CGG array. ATL1 genotypes are on the right side of the figure (purple and yellow blocks). The three-locus microsatellite haplotype (DXS548-FRAXAC1-FRAXAC2) of each chromosome is identified by the large colored regions: X, red, for the major fragile X haplotype; N, green, for the most frequent normal haplotype; A, yellow, for the 'alternative' haplotype; O, blue, for all other haplotypes, including those that are not complete for all markers. The asterisk identifies the only observed occurrence (apart from fragile X families) of the T40-T42-T62 haplotype.

(T40-T38-T61) was observed only once (Table 1 and Fig. 2). Interestingly, no occurrence of the T40-T42-T62 haplotype (also highly prevalent in fragile X families) was observed in premutation-size alleles. In our previous study of female samples (15) we inferred an 'alternative' haplotype (T42-T36-T60) that was the most frequent in premutation-size alleles (57%) but was rare in both normal-range *FMR1* alleles as well as in fragile X syndrome associated alleles. In the present study, this alternative haplotype was present in four out of 13 males with a premutation-size allele.

Among the 36 intermediate-size *FMR1* alleles (40–54 triplets), the major fragile X haplotype was the most common (15/36). But the second major fragile X haplotype (T40-T42-T62) was observed only once (Table 1 and Fig. 2). The most common normal haplotype (T40-T38-T61) was detected in 11 out of 36 alleles whereas the alternative haplotype (T42-T36-T60) was found in only 2/36 chromosomes, i.e. already rarer than among alleles over 54 triplets (Fisher's exact test, P-value = 0.04).

The T40-T38-T61 haplotype was present in 32/102 of the control chromosomes from newborns, whereas the major fragile X haplotype (T50-T42-T62) was found in only six samples. The 'alternative' haplotype was present in one individual and the remaining individuals carried various haplotypes, the most frequent of which (T40-T38-T60) occurred in 16/102 individuals. Interestingly, the second major fragile X haplotype (T40-T42-T62), common in fragile X chromosomes in the French-Canadian population, was absent from all chromosomes that had less than 40 triplet repeats. This suggests that a recombination between the most common normal haplotype T40-T38-T61 and the major fragile X haplotype

T50-T42-T62 occurred sometime after the expansion of the T50-T42-T62 *FMR1* allele to intermediate size. Although recombination between DXS548 and FRAXA loci, 140 kb apart, is believed not to be rare (21), this suggests that, at least in this population, the mutational event leading to the appearance of *FMR1* alleles with more than 40 triplet repeats occurred before the recombination event between FRAXA and DXS548 loci.

FMR1 triplet array structure

It has been shown by several groups that *FMR1* triplets are generally interrupted by AGG repeats (10,22–24) and the current interpretation is that very unstable alleles are derived from an increase in the length of the 3′ pure CGG tract and/or by loss of an AGG interruption (6). Indeed, in several studies which analyzed normal-size alleles and a few intermediate-size alleles, the majority of normal *FMR1* alleles reported had one or two AGG interruptions, whereas in fragile X families usually no interruptions have been detected (10,23,25).

In order to assess the various stages of progression of *FMR1* alleles from normal size-range to disease alleles, we determined by genomic sequencing the *FMR1* triplet array structure of alleles in the general population sample of males. This could not be done in our previous study of 10 624 women (15) because of the technical difficulty of sequencing these arrays in women. Identification of premutation-size and intermediate-size alleles in males allowed us to perform these sequencing experiments and unexpected results were obtained. As shown in Table 1 and Figure 2, determination of the AGG interruption pattern of the 13 premutation-size alleles revealed nine alleles with two AGG repeats, three with one AGG interruption and only one with no AGG interruption. Furthermore, 33 of the 36 alleles with 40–54 repeats had two AGG interruptions and the remaining three alleles had one interruption. Interestingly, we observed *FMR1* alleles with the first AGG interruption positioned after the 11th triplet (12th, 14th, 15th, 19th) only among alleles of less than 40 triplets. Previous studies of AGG interruption patterns of *FMR1* triplet arrays have not reported the sequence of premutation-size alleles that were not derived from fragile X families (10,19,23–30). Our results clearly differ from the *FMR1* array structure observed in normal transmitting males from fragile X families which all have one or no AGG interruption (24,25,27). Indeed, we observed no AGG interruption in all 12 normal transmitting males that were first or second degree relatives from affected males in the 68 families studied: six were on the major fragile X haplotype, five on the second major fragile X haplotype and one on the normal haplotype. It is interesting to underline our observation that although alleles of 40 triplets and more identified in the general population do not show the same array structure in terms of AGG interruptions as premutations in fragile X families, they are associated with the major fragile X haplotype found in fragile X families (see Discussion).

ATL1 polymorphism

Gunter *et al.* (20) reported a polymorphism that lies between the CGG array and FRAXAC2. This polymorphism was observed to be in strong linkage disequilibrium with the triplet array interruption pattern. We thus analyzed this marker. The results are shown in Figure 2. In newborn samples, the 'G' allele was observed in 48/102 *FMR1* alleles. All 54 remaining 'A' alleles without exception were associated with the T38 allele of FRAXAC1 and the majority of them had an AGG repeat in the 11th position of the *FMR1* triplet array (47/54). This applies as well to all other *FMR1* alleles with ATL1*A observed in our study, either fragile X, premutated or intermediate-size alleles. The 'A' allele of the ATL1 polymorphism was present only in 7/68 fragile X alleles, six of which were on the most frequent normal haplotype. In premutation-size and intermediate-size alleles, ATL1*A was also mostly associated with the normal haplotype (12/13). The relative frequencies of ATL1*G and ATL1*A were associated with the *FMR1* allele sizes in the general population ($\chi^2 = 11.9$, 2 df, *P*-value = 0.001). In our samples, ATL1*G was associated with several other haplotypes and AGG-interruption patterns, including several 10+ alleles. This is in keeping with the conclusion of Gunter *et al.* (20) that ATL1*G is older than the 'A' allele. However, this marker did not yield much additional information to our analyses, given that it was so tightly associated with a single FRAXAC1 allele in the studied population.

DISCUSSION

We screened 10 572 males from the general population for *FMR1* premutation-size alleles by Southern blotting. We identified 13 such alleles. Flanking mircosatellite markers were genotyped and *FMR1* alleles were sequenced to determine the triplet array structure for AGG interruptions. Our estimate of the prevalence of premutation-size alleles is 1 in 813 males (95% CI, 1:527 to 1:1781) which is not statistically different from our previous estimate of the allele frequency in females (15). This is the first study in males from the general population which yields a reliable estimate of the *FMR1* premutation-size allele frequency in males although the confidence interval is still large. If we had used the 60 triplet repeat cut-off for the minimal size of 'premutations' that has been used in some other recent studies, the prevalence of *FMR1* alleles of 60+ repeats would have been 8:10572 or 1:1322 (95% CI, 1:781 to 1:4300); still not significantly different than the estimated frequency of such alleles in women (15). Thus, the estimated female to male premutation-size carrier frequency ratio in the French-Canadian population is 3.1 (95% CI, 1.4 to 9.0) (as calculated using $2q_{females}/q_{males}$), which is not statistically different from 1.59 predicted by mathematical simulation using an infinite population model (13).

Previous studies of premutation-size alleles in males have been subject to various biases due to the fact that they were not population based (18). Our previous population-based study of 10 624 women did not allow us to make a definitive assessment of the haplotypes nor to determine the AGG interruption pattern. The present study of population-based samples of males allowed us to accomplish these objectives.

Recently, Toledano-Alhadef *et al.* (31) reported a much higher prevalence of 1/113 female carriers of *FMR1* alleles with more than 54 triplets in Israel. This is very different from the female carrier prevalence of 1:259 we reported earlier (15) ($\chi^2 = 22.5$, *P*-value < 0.0001). However, prevalence estimates in other Caucasian population studies are similar to our findings in French-Canadians (19,32,33). Several observations we make here provide insights into the dynamics of *FMR1* alleles in the general population.

Firstly, *FMR1* alleles found in French-Canadian fragile X males are relatively homogenous with 75% being on two

apparently related major fragile X haplotypes (T50-T42-T62 and T40-T42-T62). The T50-T42-T62 haplotype is not the most frequent in Northern European populations, it is associated with a 9 + 9 + *n* structure and was hypothesized to expand slowly from normal size to fragile X mutation through a multistep mutational process (34) leading to increase in size without loss of AGG repeats.

Secondly, the second most common haplotype (T42-T36-T60) in the 55–95 triplet size-range was observed only once in the sample of normal-size alleles. This haplotype was present in 7% of the fragile X chromosomes. Thus, *FMR1* alleles between 55 and 95 triplets are not homogenous but are associated with at least one other major haplotype (the 'alternative' major haplotype; designated 'A' in Table 1 and Figure 2), which is not frequent in fragile X families and in the population sample we studied. This haplotype has been observed in the Northern European population and may have a different mutation pathway and does not retain AGG repeats, as suggested by Eichler *et al.* (34). Our observations are thus compatible with this model.

Thirdly, the T50-T42-T62 major fragile X haplotype is the predominant haplotype in all *FMR1* alleles of 40 triplets or more. In the French-Canadian population the predominance of the major fragile X haplotype among premutation-size and intermediate-size alleles from the general population in such a notable proportion is striking. This provides evidence of a rare, early event that is responsible for the expansion of the *FMR1* triplet array on this haplotype from the normal range to the 40+ triplets range. This has been predicted and is probably the case for most of the major fragile X haplotypes found in various countries, as suggested by Chiurazzi *et al.* (35).

Finally, and most notable, is our observation that all but one of the 13 premutation-size alleles had two ($n = 9$) or one ($n = 3$) AGG interruptions as well as all intermediate-size alleles of 40–54 triplets ($n = 36$), even if they were associated with the major fragile X haplotype. All alleles tested from normal transmitting males in fragile X families contained no AGG interruption which is in keeping with published reports that most frequently observed no interruption in normal transmitting males from fragile X families (24,25,27). We conclude from these observations that the loss of an AGG interruption is not the initial event which confers instability for triplet-repeat expansion of the *FMR1* gene. More likely this loss is a late event in the expansion process and possibly the last barrier before the hyperexpansion that is observed in fragile X families and in the most recent generations preceding the cases.

This study provides strong evidence in favor of the hypothesis of a precursor *FMR1* haplotype with a greater likelihood for triplet repeat expansion than other haplotypes. The elements involved in this process are clearly not the AGG interruptions but may be particular sequences close to the triplet-repeat array as suggested by Gunter *et al.* (20). Alternatively, if the hypothesis of a greater likelihood of expansion of *FMR1* alleles on the major fragile X haplotype is incorrect, then it may be that there was a population bottleneck in the past when most of the normal alleles on that haplotype were lost, leaving the artifactual impression that they progressed rapidly from normal to intermediate.

It will be important to undertake other population-based studies in different populations in order to determine whether similar states of fragile X precursor alleles are present or are limited to the French-Canadian population, and attributable to its history and genetic structure.

MATERIALS AND METHODS

Samples

15 288 hemoglobin/hematocrit leftover samples from male outpatients were collected from the hematology laboratory of a general hospital in Québec during a period of 48 months. Twenty-five percent of the samples were discarded randomly to ensure that any given patient from the hospital had a significant chance of not being included in this anonymous study. This left a working sample of 11 495 whole blood samples. Each sample was assigned a unique number and handled in an unlinked anonymous protocol precluding the association of DNA results with personal identifying information. 923 duplicates were identified because of file number matches prior to anonymization and DNA analysis, and the anonymous sample numbers of the replicates were tagged as duplicates to be excluded from the subsequent study. This study was approved by the CHUQ ethical review board including procedures for sample and information collection and coding, analysis of samples and recording and reporting of results.

DNA purification and Southern blot analysis

Samples were grouped sequentially in pools of five males and 40 µl of whole blood was transferred to a 1.5 ml Eppendorf™ tube for a total of 200 µl of whole blood per pool. The remaining portion of each individual sample was kept frozen at –20°C until further study. DNA was purified from whole blood pools using a mini-prep method (36) and then analyzed using high-throughput mini-Southern blotting as described (15) with the following differences: (i) the restriction enzyme used was *Dpn*II+*Eag*I and (ii) the 1.5 mm-thick 6% polyacrylamide gel (1:30 bisacrylamide) in 1× TBE was migrated for 16 h at 35 V. After ethidium bromide staining the DNA was electroblotted on a Qiagen nylon membrane (Qiagen) at 80 mA for 4 h. The blot was then hybridized to a new probe designated QcPX specifically designed to detect the *Dpn*II+*Eag*I fragments of interest in the same hybridization conditions as StB12.3 (37). After washing in 0.5× SSC at 65°C for 2× 15 min, the blot was exposed overnight and until slight overexposure before searching for premutation-size bands.

Screening strategy

Positive pools of male DNA (candidates for containing an *FMR1* allele greater than 40 triplets) were identified by the presence of a *Dpn*II+*Eag*I restriction fragment that hybridized to probe QcPX and was significantly larger than bands of the other alleles in the pool (Fig. 1). A *Dpn*II+*Eag*I enzyme digestion was chosen because in males, it generates a constant control band which is slightly smaller than the band bearing the *FMR1* triplet repeats, and allows for relatively accurate sizing of the *FMR1* alleles. Practically, 60 bp (20 triplet repeats) correspond to a difference of migration of 3 mm between the mean size of the other alleles in the pool and the outlier band(s). Pools with discrete individual bands that were <3 mm from the bulk of the other alleles were classified as containing 'large alleles'. For pools where the presence of a larger than normal band was

suspected, DNA from the five different individual samples was repurified and individually reanalyzed by mini-Southern blotting (Fig. 1). DNA from the samples containing the large band was re-extracted for PCR determination of the number of triplet repeats and for haplotype analysis (see below). The sensitivity of this screening strategy for detection of *FMR1* alleles with more than 54 triplets is estimated to be greater than that of the *Bcl*I digests in women (15) because pools of samples from five men contain only five X chromosomes as compared to the 10 in pools of five women, and the size of the *Dpn*II+*Eag*I restriction fragment containing the *FMR1* triplet repeats is ~25% smaller than the restriction fragment generated by a *Bcl*I digest (i.e. 735 bp versus 1 kb).

Genotyping *FMR1* repeats and flanking markers

The number of trinucleotide repeats at the *FMR1* locus was determined for all positive samples by PCR following the protocol of Fu *et al.* (38) with inclusion of 50% 7-deaza-dGTP (Boehringer Mannheim) and 15% DMSO in the exo$^-$ *pfu* DNA polymerase amplification mix (Stratagene, San Diego, CA) but with direct incorporation of α-dCT^{32}P to reveal the PCR products. Sizing of the trinucleotide repeat was done by comparison to a sequence ladder loaded on the gel. The triplet array structure was determined as follows: the trinucleotides were preamplified from 50 ng of genomic DNA as described above (but only with cold dNTPs). The PCR products were purified with the QIAquick PCR purification Kit (Qiagen) and the PCR products eluted in 35 µl which were then sequenced with primer 170 of Hirst *et al.* (23) labeled with 20 µCi γ-AT^{33}P and 0.5 U/µl T4 polynucleotide kinase 1X (New England Biolabs). Sequencing reactions were set up with the Promega fmol DNA sequencing system using 1 ng of purified sequencing product and the labeled primer using one cycle of 5 min at 98°C followed by 30 cycles of 15 s at 98°C and 1 min at 70°C in a Perkin-Elmer DNA Thermal Cycler 180. Sequencing products were migrated on a sequencing gel as described above, but for only 75 mins, and exposed for 3 days. The *FMR1* triplet-repeat array structure of fragile X alleles in families was determined from 12 normal transmitting males from these families that were first or second degree relatives of the fragile X cases. *DXS548* (3), *FRAXAC1* and *FRAXAC2* (39) microsatellites were analyzed as described (15) and standardized according to the nomenclature proposed by Chiurazzi *et al.* (40) using the standardized panel of cell line DNAs for FRAXA haplotype standardization available through http://www.rmga.qc.ca/. The ATL1 polymorphism (alleles: A/G) was analyzed by allele-specific oligonucleotide PCR. Conditions used were as follows: in a final reaction volume of 25 µl, 0.5 µM of common primer (ATL1F 5′-TCATCAGTCCTGGTAATAGAA-3′) and 0.5 µM allele-specific primer (ATL1A 5′-GACACAGAATCATAA-ATGT-3′ or ATL1G 5′-GACACAGAATCATAAATGC-3′), 200 µM of dNTPs (New England Biolab), 70 ng of genomic DNA, 1.25 U *Taq* DNA polymerase (Qiagen) and 1× PCR buffer containing 1.5 mM $MgCl_2$ (Qiagen). Thermocycler conditions consisted of an initial 5 min at 95°C followed by 30 cycles of: 30 s at 95°C, 30 s at 55°C and 1 min at 72°C. Final extension was performed by 10 min at 72°C. Products were visualized on 1.2% agarose gel using ethidium bromide staining.

Allele frequencies

The sample included unrelated fragile X chromosomes observed in 68 males from fragile X families. Allele frequencies from the normal population were derived from an anonymous unlinked control sample of 102 consecutive newborn males.

We call the most frequent haplotype found in fragile X syndrome males in Québec, the major fragile X haplotype. The 'alternative' major haplotype inferred for premutation carriers was comprised of the most frequent allele at each of four marker loci estimated for *FMR1* alleles with CGG repeats in the range of 55–64 in the female screening study (15).

ACKNOWLEDGEMENTS

We thank Richard Réhel and Jacqueline Dionne for their expert technical assistance. We also thank the personnel of the hematology laboratory of St-Francois-d'Assise Hospital and Dr Jacques Boulay for their precious collaboration in collecting the samples. This project was funded by the Canadian Institutes for Health Research and the Canadian Genetic Diseases Network of Centres of Excellence. The Research Centre and Research Unit is supported by the Fonds de la Recherche en Santé du Québec. C.D., M.L.M., P.R. and S.L. were supported by studentship awards from either the Research Centre, Canadian Institutes for Health Research, National Science and Engineering Research Council of Canada, the Fonds de la Recherche en Santé du Québec or from FCAR. F.R. is a Canadian Institutes for Health Research Scientist.

REFERENCES

1. Nussbaum,R.L. and Ledbetter,D.H. (1995) The fragile-X syndrome. In Scriver,C.R., Beaudet,A.L., Sly,W.S. and Valle,D. (eds), *The Metabolic and Molecular Bases of Inherited Disease*, 7th edn. McGraw-Hill, New York,Vol. 1, pp. 795–810.
2. Oberle,I., Rousseau,F., Heitz,D., Kretz,C., Devys,D., Hanauer,A., Boue,J., Bertheas,M.F. and Mandel,J.L. (1991) Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science*, **252**, 1097–1102.
3. Verkerk,A.J., Pieretti,M., Sutcliffe,J.S., Fu,Y.H., Kuhl,D.P., Pizzuti,A., Reiner,O., Richards,S., Victoria,M.F., Zhang,F.P. *et al.* (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, **65**, 905–914.
4. Yu,S., Pritchard,M., Kremer,E., Lynch,M., Nancarrow,J., Baker,E., Holman,K., Mulley,J.C., Warren,S.T., Schlessinger,D. *et al.* (1991) Fragile X genotype characterized by an unstable region of DNA. *Science*, **252**, 1179–1181.
5. Vincent,A., Heitz,D., Petit,C., Kretz,C., Oberle,I. and Mandel,J.L. (1991) Abnormal pattern detected in fragile-X patients by pulsed-field gel electrophoresis. *Nature*, **349**, 624–626.
6. Imbert,G., Feng,Y., Nelson,D.L., Warren,S.T. and Mandel,J.L. (1998) *FMR1* gene and mutations in fragile-X syndrome: Molecular biology, biochemistry, and genetics. In Wells,R.D. and Warren,S.T. (eds), *Genetic Instabilities and Hereditary Neurological Diseases*. Academic Press, San Diego, CA, pp. 27–53.
7. Hagerman,R.J., Hull,C.E., Safanda,J.F., Carpenter,I., Staley,L.W., O'Connor,R.A., Seydel,C., Mazzocco,M.M., Snow,K., Thibodeau,S.N. *et al.* (1994) High functioning fragile X males: demonstration of an unmethylated fully expanded FMR-1 mutation associated with protein expression. *Am. J. Med. Genet.*, **51**, 298–308.
8. Nolin,S.L., Glicksman,A., Houck,G.E.,Jr, Brown,W.T. and Dobkin,C.S. (1994) Mosaicism in fragile X affected males. *Am. J. Med. Genet.*, **51**, 509–512.
9. Rousseau,F., Heitz,D., Tarleton,J., MacPherson,J., Malmgren,H., Dahl,N., Barnicoat,A., Mathew,C., Mornet,E., Tejada,I. *et al.* (1994) A multicenter study on genotype-phenotype correlations in the fragile X

syndrome, using direct diagnosis with probe StB12.3: the first 2,253 cases. *Am. J. Hum. Genet.*, **55**, 225–237.
10. Kunst,C.B. and Warren,S.T. (1994) Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell*, **77**, 853–861.
11. Nolin,S.L., Lewis,F.A.,III, Ye,L.L., Houck,G.E.,Jr, Glicksman,A.E., Limprasert,P., Li,S.Y., Zhong,N., Ashley,A.E., Feingold,E. *et al.* (1996) Familial transmission of the *FMR1* CGG repeat. *Am. J. Hum. Genet.*, **59**, 1252–1261.
12. Morton,N.E. and Macpherson,J.N. (1992) Population genetics of the fragile-X syndrome: multiallelic model for the *FMR1* locus. *Proc. Natl Acad. Sci. USA*, **89**, 4215–4217.
13. Morris,A., Morton,N.E., Collins,A., Macpherson,J., Nelson,D. and Sherman,S. (1995) An n-allele model for progressive amplification in the *FMR1* locus. *Proc. Natl Acad. Sci. USA*, **92**, 4833–4837.
14. Kolehmainen,K. (1994) Population genetics of fragile X: a multiple allele model with variable risk of CGG repeat expansion. *Am. J. Med. Genet.*, **51**, 428–435.
15. Rousseau,F., Rouillard,P., Morel,M.L., Khandjian,E.W. and Morgan,K. (1995) Prevalence of carriers of premutation-size alleles of the FMRI gene—and implications for the population genetics of the fragile X syndrome. *Am. J. Hum. Genet.*, **57**, 1006–1018.
16. Oudet,C., von Koskull,H., Nordstrom,A.M., Peippo,M. and Mandel,J.L. (1993) Striking founder effect for the fragile X syndrome in Finland. *Eur. J. Hum. Genet.*, **1**, 181–189.
17. Heitz,D., Devys,D., Imbert,G., Kretz,C. and Mandel,J.L. (1992) Inheritance of the fragile X syndrome: size of the fragile X premutation is a major determinant of the transition to full mutation. *J. Med. Genet.*, **29**, 794–801.
18. Ashley-Koch,A.E., Robinson,H., Glicksman,A.E., Nolin,S.L., Schwartz,C.E., Brown,W.T., Turner,G. and Sherman,S.L. (1998) Examination of factors associated with instability of the *FMR1* CGG repeat. *Am. J. Hum. Genet.*, **63**, 776–785.
19. Crawford,D.C., Meadows,K.L., Newman,J.L., Taft,L.F., Pettay,D.L., Gold,L.B., Hersey,S.J., Hinkle,E.F., Stanfield,M.L., Holmgreen,P. *et al.* (1999) Prevalence and phenotype consequence of FRAXA and FRAXE alleles in a large, ethnically diverse, special education-needs population. *Am. J. Hum. Genet.*, **64**, 495–507.
20. Gunter,C., Paradee,W., Crawford,D.C., Meadows,K.A., Newman,J., Kunst,C.B., Nelson,D.L., Schwartz,C., Murray,A., Macpherson,J.N. *et al.* (1998) Re-examination of factors associated with expansion of CGG repeats using a single nucleotide polymorphism in *FMR1*. *Hum. Mol. Genet.*, **7**, 1935–1946.
21. Warren,S.T. and Nelson,D.L. (1994) Advances in molecular analysis of fragile X syndrome. *J. Am. Med. Assoc.*, **271**, 536–542.
22. Eichler,E.E., Holden,J.J., Popovich,B.W., Reiss,A.L., Snow,K., Thibodeau,S.N., Richards,C.S., Ward,P.A. and Nelson,D.L. (1994) Length of uninterrupted CGG repeats determines instability in the *FMR1* gene. *Nat. Genet.*, **8**, 88–94.
23. Hirst,M.C., Grewal,P.K. and Davies,K.E. (1994) Precursor arrays for triplet repeat expansion at the fragile X locus. *Hum. Mol. Genet.*, **3**, 1553–1560.
24. Zhong,N., Yang,W., Dobkin,C. and Brown,W.T. (1995) Fragile X gene instability: anchoring AGGs and linked microsatellites. *Am. J. Hum. Genet.*, **57**, 351–361.
25. Snow,K., Tester,D.J., Kruckeberg,K.E., Schaid,D.J. and Thibodeau,S.N. (1994) Sequence analysis of the fragile X trinucleotide repeat: implications for the origin of the fragile X mutation. *Hum. Mol. Genet.*, **3**, 1543–1551.
26. Eichler,E.E. and Nelson,D.L. (1996) Genetic variation and evolutionary stability of the *FMR1* CGG repeat in six closed human populations. *Am. J. Med. Genet.*, **64**, 220–225.
27. Eichler,E.E., Hammond,H.A., Macpherson,J.N., Ward,P.A. and Nelson,D.L. (1995) Population survey of the human *FMR1* CGG repeat substructure suggests biased polarity for the loss of AGG interruptions. *Hum. Mol. Genet.*, **4**, 2199–2208.
28. Zhong,N., Ju,W., Pietrofesa,J., Wang,D., Dobkin,C. and Brown,W.T. (1996) Fragile X 'gray zone' alleles: AGG patterns, expansion risks, and associated haplotypes. *Am. J. Med. Genet.*, **64**, 261–265.
29. Murray,A., Macpherson,J.N., Pound,M.C., Sharrock,A., Youings,S.A., Dennis,N.R., McKechnie,N., Linehan,P., Morton,N.E. and Jacobs,P.A. (1997) The role of size, sequence and haplotype in the stability of FRAXA and FRAXE alleles during transmission. *Hum. Mol. Genet.*, **6**, 173–184.
30. Teague,J.W., Morton,N.E., Dennis,N.R., Curtis,G., McKechnie,N., Macpherson,J.N., Murray,A., Pound,M.C., Sharrock,A.J., Youings,S.A. *et al.* (1998) FRAXA and FRAXE: evidence against segregation distortion and for an effect of intermediate alleles on learning disability. *Proc. Natl Acad. Sci. USA*, **95**, 719–724.
31. Toledano-Alhadef,H., Basel-Vanagaite,L., Magal,N., Daviodov,B., Ehrlich,S., Drasinover,V., Taub,E., Halpern,G., Ginott,N. and Shohat,M. (2001) Fragile-X carrier screening and the prevalence of premutation and full-mutation carriers in Israel. *Am. J. Hum. Genet.*, **69**, 351–360.
32. Murray,J., Cuckle,H., Taylor,G. and Hewison,J. (1997) Screening for fragile X syndrome: information needs for health planners. *J. Med. Screen.*, **4**, 60–94.
33. Ryynanen,M., Heinonen,S., Makkonen,M., Kajanoja,E., Mannermaa,A. and Pertti,K. (1999) Feasibility and acceptance of screening for fragile X mutations in low-risk pregnancies. *Eur. J. Hum. Genet.*, **7**, 212–216.
34. Eichler,E.E., Macpherson,J.N., Murray,A., Jacobs,P.A., Chakravarti,A. and Nelson,D.L. (1996) Haplotype and interspersion analysis of the *FMR1* CGG repeat identifies two different mutational pathways for the origin of the fragile X syndrome. *Hum. Mol. Genet.*, **5**, 319–330.
35. Chiurazzi,P., Macpherson,J., Sherman,S. and Neri,G. (1996) Significance of linkage disequilibrium between the fragile X locus and its flanking markers. *Am. J. Med. Genet.*, **64**, 203–208.
36. Rousseau,F., Rehel,R., Rouillard,P., DeGranpre,P. and Khandjian,E.W. (1994) High throughput and economical mutation detection and RFLP analysis using a minimethod for DNA preparation from whole blood and acrylamide gel electrophoresis. *Hum. Mutat.*, **4**, 51–54.
37. Rousseau,F., Heitz,D., Biancalana,V., Blumenfeld,S., Kretz,C., Boue,J., Tommerup,N., Van Der Hagen,C., DeLozier-Blanchet,C., Croquette,M.F. *et al.* (1991) Direct diagnosis by DNA analysis of the fragile X syndrome of mental retardation. *N. Engl. J. Med.*, **325**, 1673–1681.
38. Fu,Y.H., Kuhl,D.P., Pizzuti,A., Pieretti,M., Sutcliffe,J.S., Richards,S., Verkerk,A.J., Holden,J.J., Fenwick,R.G.,Jr, Warren,S.T. *et al.* (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell*, **67**, 1047–1058.
39. Richards,R.I., Holman,K., Friend,K., Kremer,E., Hillen,D., Staples,A., Brown,W.T., Goonewardena,P., Tarleton,J., Schwartz,C. *et al.* (1992) Evidence of founder chromosomes in fragile X syndrome. *Nat. Genet.*, **1**, 257–260.
40. Chiurazzi,P., Pomponi,M.G., Sharrock,A., Macpherson,J., Lormeau,S., Morel,M.L. and Rousseau,F. (1999) DNA panel for interlaboratory standardization of haplotype studies on the fragile X syndrome and proposal for a new allele nomenclature. *Am. J. Med. Genet.*, **83**, 347–349.