Origins of the fragile X syndrome mutation

M C Hirst, S J L Knight, Z Christodoulou, P K Grewal, J P Fryns, K E Davies

Abstract

The fragile X syndrome is a common cause of mental impairment. In view of the low reproductive fitness of affected males, the high incidence of the syndrome has been suggested to be the result of a high rate of new mutations occurring exclusively in the male germline. Extensive family studies, however, have failed to identify any cases of a new mutation. Alternatively, it has been suggested that a selective advantage of unaffected heterozygotes may, in part, explain the high incidence of the syndrome. Molecular investigations have shown that the syndrome is caused by the amplification of a CGG trinucleotide repeat in the FMR-1 gene which leads to the loss of gene expression. Further to this, genetic studies have suggested that there is evidence of linkage disequilibrium between the fragile X disease locus and flanking polymorphic markers. More recently, this analysis has been extended and has led to the observation that a large number of fragile X chromosomes appear to be lineage descendants of founder mutation events. Here, we present a study of the FRAXAC1 polymorphic marker in our patient cohort. We find that its allele distribution is strikingly different on fragile X chromosomes, confirming the earlier observations and giving further support to the suggestions of a fragile X founder effect.

(*J Med Genet* 1993;30:647–50)

The fragile X syndrome is the most frequent familial form of mental retardation and may be responsible for up to 50% of all X linked mental retardation.[1 2] It is characterised clinically by an IQ typically in the range 35 to 50, elongated facies (associated with oedema, tissue thickening, and prognathism) with large everted ears and macro-orchidism. The syndrome is also associated with the expression of a rare folate sensitive fragile site at Xq27.3.[3 4]

Extensive genetic and physical mapping has led recently to the isolation of a gene, FMR-1, in Xq27 which is defective in the fragile X syndrome.[5] Molecular studies have identified a stepwise amplification of a trinucleotide CGG:CCG repeat element within the 5' of the FMR-1 gene on the fragile X chromosome.[6–12] In unaffected carriers, the amplification is small and the gene functions normally. Upon further amplification, the CGG array expands dramatically and as a consequence the FMR-1 gene promoter becomes highly methylated and gene activity is lost.[13 15] Confirmatory evidence that the FMR-1 gene is indeed responsible for the clinical features of the syndrome has come from identifying a de novo microdeletion in a male with the typical characteristics of the syndrome in the absence of fragile X expression cytogenetically.[16]

The fragile X syndrome is found with a frequency of 0·3 per 1000 males and 0·2 to 0·6 per 1000 females and segregation studies have suggested an overall carrier rate of 1/800.[17 18] As the reproductive fitness of affected hemizygous and heterozygous females is low, and the mutation is constantly being lost from the population, this high frequency requires an explanation. Sherman *et al*[19] suggested that the high incidence of the fragile X mutation was the result of a high new mutation rate in the order of $2\cdot4 \times 10^{-4}$ in sperm. However, molecular analyses have failed to identify any sporadic cases suggesting that the mutation is being carried through generations by normal transmitting males and normal carrier females. An alternative hypothesis is that a selective advantage in unaffected carriers may play a role in maintaining the high incidence of the syndrome.[20 21] In this case, compensation for the loss of mutant alleles through reproductive failure is suggested to be achieved by a higher reproductive ability of unaffected and mildly affected subjects.

With the isolation of the FMR-1 gene and flanking polymorphic markers, it is now possible to map genetic haplotypes on the fragile X chromosome. Richards *et al*[22] have recently reported evidence for linkage disequilibrium between two dinucleotide markers and the fragile X mutation. Linkage disequilibrium mapping is based upon the expectation that in close proximity to the mutant gene, chromosomes descended from a common ancestral mutation will show a common haplotype reflecting that of the original ancestral chromosome. These observations suggest that most of the fragile X mutations we see today are the result of one or more founder mutations. In this study we have analysed the alleles of the closely linked AC dinucleotide marker FRAXAC1 in a large cohort of normal and fragile X subjects of European origin. The distribution of alleles in the fragile X population is significantly different from that in the normal population, confirming the observations of Richards *et al*.[22] Fragile X mutations on other haplotypes are also found, and these may be understood in the light of new evidence of mutation rates in simple repeat elements.

Materials and methods

Genomic DNA was prepared from peripheral blood lymphocytes and cell lines by standard methods. The non-fragile X samples were made up of unrelated subjects including members of the CEPH collection of DNA. Analysis
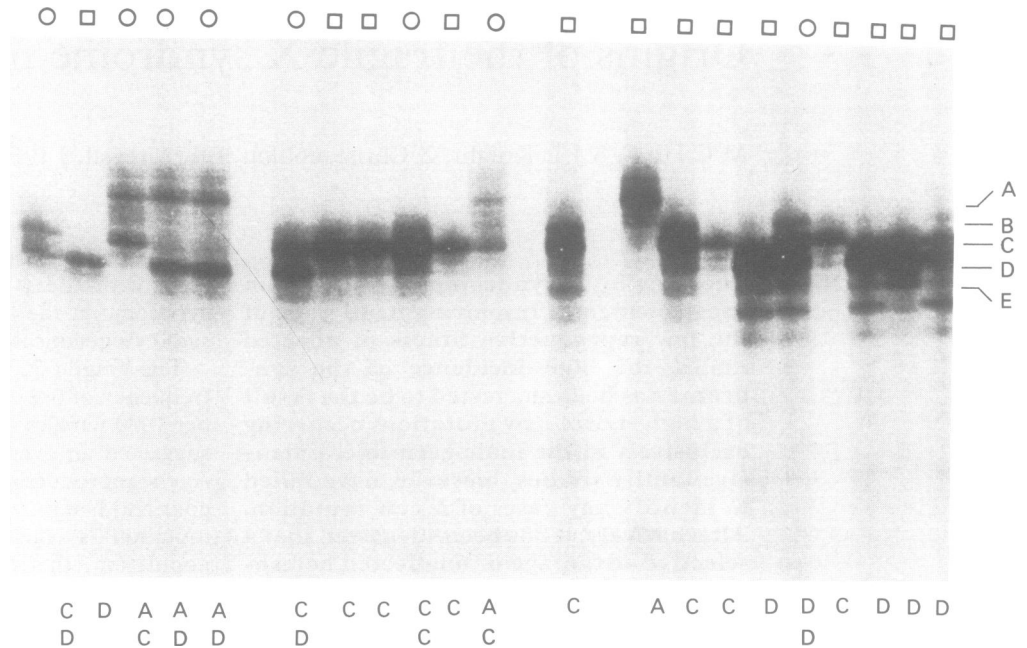
Molecular Genetics Group, Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK.
M C Hirst
S J L Knight
Z Christodoulou
P K Grewal
K E Davies

Division of Human Genetics, University Hospital Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium.
J P Fryns

Correspondence to Professor Davies.

Typing of the FRAXAC1 alleles in 21 subjects. Individual alleles are scored below each track. The migration positions of the five alleles are shown for reference. For the modified conditions of PCR amplification and polyacrylamide gel analysis see Material and methods. Squares = males, circles = females.

of the dinucleotide repeat marker FRAXAC1 was carried out using the polymerase chain reaction with oligonucleotide primers,[23] but the method was modified as follows: 100 ng DNA in a 10 µl final volume containing 50 mmol/l KCl, 10 mmol/l Tris HCl (pH 8·4), MgCl$_2$, 200 µmol/l dTGP, 200 µmol/l dTTP, 200 µmol/l dCTP, 25 µmol/l dATP, 0·5 µmol/l each oligonucleotide, 1 unit *Taq* polymerase (Amplitaq, Cetus) and 10 µCi α-[^{35}S]-dATP (Amersham International, SJ1304). Reactions were cycled through 10 cycles of 94°C (one minute), 60°C (1·5 minutes), 72°C (1·5 minutes) followed by 25 cycles of 94°C (one minute, 55°C (1·6 minutes), 72°C (1·5 minutes). Products were heated to 95°C for two minutes in formamide loading buffer, resolved by electrophoresis through 6% polyacrylamide, and visualised by autoradiography with Kodak XAR5 film for 24 to 60 hours at room temperature (figure).

Results

Alleles of the FRAXAC1 repeat were studied in normal and affected fragile X subjects. Five alleles were identified (A, B, C, D, and E, figure) each differing by one AC repeat unit. The table shows the distribution of these five alleles in both normal and fragile X subjects. The expected allele frequencies were calcu-

Allele frequencies at the FRAXAC1 dinucleotide repeat.

Allele	Size (bp)	Fragile X observed	Expected*
A	112	20	3·37
B	110	4	0
C	108	28	57·28
D	106	21	11·79
E	104	0	0·56
Total		73	73

*Based upon allele frequencies found in 130 unrelated X chromosomes.
Statistical analysis was carried out giving a total χ^2 value of 105·12, which is significant at p = 0·001.

lated from a survey of allele frequencies on 130 unrelated normal X chromosomes from a random DNA bank of British subjects and the CEPH consortium DNA bank. The 73 unrelated fragile X males analysed are of British (n = 53) and Belgian (n = 20) origin.

These data indicate a strikingly different AC1 allele distribution in the fragile X population when compared to our normal controls, with the A and D alleles both present in greater numbers than expected. The A allele frequency is highly significant being found on only 5% of normal chromosomes and 27% of fragile X chromosomes. The D allele is present on 29% of fragile X chromosomes as compared to 16% of normal ones. The C allele is less frequent on the fragile X chromosome (38%) compared to normal (78·5%).

Examination of the alleles of FRAXAC1 showed a significant difference between the expected and observed allele frequencies, with a strong association between several alleles and expansion of the CGG repeat to a full mutation. In the case of FRAXAC1, we found an overall increase in the frequency of the A and D alleles, with a decrease in frequency of the most common C allele. This is also true of the data of Richards *et al.*[22] They reported haplotype analysis with the other flanking dinucleotide repeat probe, FRAXAC2. The haplotype most commonly associated with the fragile X chromosome is AC1 A/AC2 F (the AF haplotype), but the AC1 D allele was also found in excess in the DD (Australian) and DB (New York) haplotypes. In our data overall, the alleles A and D make up 56% of the fragile X population whereas they are only found in 20% of the normal population. This compares very well with the observations that AF, DD, and DB make up 58% of the fragile X population, but only 15% of the normal population in the study of Richards *et al.*[22] These observations suggest that fragile X mutations have

arisen on at least two specific allelic backgrounds and that those subjects with the same alleles are lineage descendant from the ancestral mutations.

Discussion

Family studies have failed to show any new fragile X mutation involving expansion of the FMR-1 CGG repeat element. All carriers and affected subjects have been shown to have inherited the mutation from a parent also carrying an expanded allele. Thus, in order to explain the high population frequency of the syndrome, there must be a pre-existing population of mutation carriers. Where extensive family records are available, it has been possible to trace the mutation back over many generations.[24] Unfortunately, as genetic material is unavailable for more than four or five generations, we do not know whether the mutation is carried as a 'premutation' sized CGG element or as another form of the mutation. Alternatively, it may be possible that the expansion of the CGG element is independent of the repeat itself, but is controlled by a very closely linked flanking marker. Alterations in length of minisatellite loci are known to be controlled by flanking modulating elements.[25]

While there appears to be a common genetic background on which FMR-1 CGG expansions have occurred, 44% of fragile X chromosomes carry other haplotypes, suggesting that mutations may have arisen independently on other genetic backgrounds. This diversity of haplotypes may, however, also be the result of mutations in the AC dinucleotide markers themselves. A recent study[26] in the isolated Finnish population has shown that the mutation rates of simple sequence repeat markers were found to have a high mutation rate, in one case approaching 1%. Thus, the mutation rates of the simple repeats are sufficient to allow the appearance of new alleles on the descendants of the ancestral haplotype, resulting in a decrease in the excess of the ancestral haplotype. As linkage disequilibrium suggests that the fragile X mutation is old, one might expect such further mutations to have occurred. This is being investigated by the analysis of less mutable polymorphic markers such as RFLPs.

Morton and MacPherson[27] have recently proposed a model in which the fragile X mutation is postulated to be occurring as a multistep process. This attractive model provides a framework in which the seemingly contradictory observations of a mutation old enough to establish a founder effect and an apparently high new mutation rate are united.[28] The model suggests that alleles of over 50 copies of CGG arise from normal alleles, but these are non-phenotypic and stable for many generations. This longevity would allow the required time for the establishment of an ancestral haplotype, resulting in the observed linkage disequilibrium. The second important aspect of this model is that it is proposed that these larger CGG alleles convert to premutations at a frequency of 1 to 2% per generation. Once in

this state, alleles are highly unstable and progress rapidly to what are seen as full fragile X mutations. This rapid expansion would therefore account for the high mutation rate observed in family studies. Thus a pool of pre-existing, or predisposed carriers is constantly giving rise to new fragile X cases. The nature of predisposing alleles and the event which triggers the dramatic expansion of the CGG repeat is currently unknown. This may be in the nature of the flanking DNA elements or in the very nature of the CGG repeat element itself. Evidence that other simple repeat elements in the human genome have frequencies of mutation approaching 1%[26] suggests that this phenomenon may not be a rare occurrence.

In summary, we have confirmed the observations of Richards *et al*[22] that there is evidence of a common genetic background in the fragile X syndrome. The diversity of haplotypes at the fragile X locus, while it may reflect genetic heterogeneity, may also be explained by mutations in the AC markers themselves.

1 Kerr B, Turner G, Mulley J, Gedeon A, Partington M. Non-specific X linked mental retardation. *J Med Genet* 1991;28:378–82.
2 Glass IA. X linked mental retardation. *J Med Genet* 1991;28:361–71.
3 Lubs HA. A marker X chromosome. *Am J Hum Genet* 1969;21:231–44.
4 Sutherland GR. Fragile sites on human chromosomes: demonstration of their dependence on the type of tissue culture medium. *Science* 1977;197:256–66.
5 Verkerk AJMH, Pieretti M. Sutcliffe JS, *et al*. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in the fragile X syndrome. *Cell* 1991;65:905–14.
6 Yu S, Pritchard M, Kremer E, *et al*. Fragile X genotype characterised by an unstable region of DNA. *Science* 1991;252:1179–81.
7 Kremer EJ, Pritchard M, Lynch M, *et al*. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science* 1991;252:1711–18.
8 Nakahori Y, Knight SJL, Holland J, *et al*. Molecular heterogeneity of the fragile X syndrome. *Nucleic Acids Res* 1991;19:4355–9.
9 Oberlé I, Rousseau F, Heitz D, *et al*. Instability of a 550 bp DNA segment and abnormal methylation in fragile X syndrome. *Science* 1991;252:1097–102.
10 Hirst MC, Nakahori Y, Knight SJL, *et al*. Genotype prediction in the fragile X syndrome. *J Med Genet* 1991;28:824–9.
11 Knight S, Hirst MC, Roche A, *et al*. Molecular studies of the fragile X syndrome. *Am J Med Genet* 1992;43:217–23.
12 Fu YH, Kuhl DPA, Pizzuti A, *et al*. Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* 1991;67:1–20.
13 Bell MV, Hirst MC, Nakahori Y, *et al*. Physical mapping across the fragile X: hypermethylation and clinical expression of the fragile X syndrome. *Cell* 1991;64:861–6.
14 Vincent A, Heitz D, Petit C, *et al*. Abnormal pattern detected in fragile X patients by pulsed field gel electrophoresis. *Nature* 1991;329:624–6.
15 Pieretti M, Zhang F, Fu YH, *et al*. Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* 1991;66:817–22.
16 Worhle D, Kotzot D, Hirst M, *et al*. Microdeletion of less than 250 kb including the proximal part of the FMR-1 gene and the fragile site, in a male with clinical phenotype of fragile X syndrome. *Am J Hum Genet* 1992;51:299–306.
17 Webb T. The epidemiology of the fragile X syndrome. In: Davies KE, ed. *The fragile X syndrome*. Oxford: Oxford University Press, 1989:40–5.
18 Brown WT. The fragile X: progress toward solving the puzzle. *Am J Hum Genet* 1990;47:175–80.
19 Sherman SL, Morton NE, Jacobs PA, Turner G. The marker (X) syndrome: a cytogenetic and genetic analysis. *Ann Hum Genet* 1984;48:21–37.
20 Vogel F. Mutation and selection in the marker (X) syndrome. *Ann Hum Genet* 1984;48:327–32.
21 Vogel F, Crusio WE, Kovac C, *et al*. Selective advantage of fra(X) heterozygotes. *Hum Genet* 1990;86:25–32.
22 Richards RI, Holman K, Friend K, *et al*. Evidence of founder chromosomes in fragile X syndrome. *Nature Genet* 1992;1:257–60.
23 Richards RI, Holman K, Kozman H, *et al*. Fragile X syndrome: genetic localisation by linkage mapping of two

microsatellite repeats *FRAXAC1* and *FRAXAC2* which immediately flank the fragile site. *J Med Genet* 1991;**28**:818–23.

24 Drugge U, Holmgren G, Blomquist H, *et al.* Study of individuals possibly affected with fragile X syndrome in a large Swedish family in the 18th and 20th centuries. *Am J Med Genet* 1992;**43**:353–4.

25 Jeffreys AJ, Neumann R, Wison V, *et al.* Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. *Cell* 1990;**60**:473–85.

26 Hastbacka J, de la Chapelle A, Kaitila I, *et al.* Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genet* 1992;**2**:204–11.

27 Morton NE, MacPherson J. Population genetics of the fragile X syndrome: multi-allelic model for the FMR1 locus. *Proc Natl Acad Sci USA* 1992;**89**:4215–17.

28 Chakravarti C. Fragile X founder effect ? *Nature Genet* 1992;**1**:237–8.