Am. J. Hum. Genet. 58:513–522, 1996

# *FMR1* in Global Populations

Catherine B. Kunst,[1] Chris Zerylnick,[1] Laurie Karickhoff,[1] Evan Eichler,[2] Jennifer Bullard,[1] Maryse Chalifoux,[3] Jeanette J. A. Holden,[3] Antonio Torroni,[4] David L. Nelson,[2] and Stephen T. Warren[1]

[1]Howard Hughes Medical Institute, Departments of Biochemistry and Pediatrics, Emory University School of Medicine, Atlanta; [2]Department of Molecular and Human Genetics and the Human Genome Center, Baylor College of Medicine, Houston; [3]Department of Psychiatry, Queens University, and Cytogenetics and DNA Research Lab, Kingston, Ontario; and [4]Dipartmento di Genetica e Biologia Molecolare, Universita La Sapienza, Rome

## Summary

**Fragile X syndrome, a frequent form of inherited mental retardation, results from the unstable expansion of a cryptic CGG repeat within the 5′ UTR region of the FMR1 gene. The CGG repeat is normally polymorphic in length, and the content is frequently interrupted by AGG triplets. These interruptions are believed to stabilize the repeat, and their absence, leading to long tracts of perfect CGG repeats, may give rise to predisposed alleles. In order to examine the stability of normal FMR1 alleles, the repeat length of 345 chromosomes from nine global populations was examined with the content also determined from 114 chromosomes as assessed by automated DNA sequencing. The FMR1 alleles, defined by the CGG repeat, as well as by the haplotypes of nearby polymorphic loci, were very heterogeneous, although the level of variation correlated with the age and/or genetic history of a particular population. Native American alleles, interrupted by three AGG repeats, exhibited marked stability over 7,000 years. However, in older African populations, parsimony analysis predicts the occasional loss of an AGG, leading to more perfect CGG repeats. These data therefore support the suggestion that AGG interruptions enhance the stability of the FMR1 repeat and indicate that the rare loss of these interruptions leads to alleles with longer perfect CGG-repeat tracts.**

## Introduction

Fragile X syndrome, an X-linked dominant disorder with reduced penetrance, is the most frequent inherited cause of mental retardation in humans (Warren and Nelson 1994; Warren and Ashley 1995). The molecular basis of fragile X syndrome has been shown to be the unstable expansion of a CGG trinucleotide repeat in the 5′ UTR of the gene FMR1, which resides at map position Xq27.3, coincident with the cytogenetic fragile site (FRAXA) which characterizes the syndrome (Kremer et al. 1991; Oberlé et al. 1991; Verkerk et al. 1991). The CGG repeat is normally cryptic in nature, with interrupting AGG triplets, and is normally polymorphic in length, ranging from 6 to 52 repeats with a mode of 30 (Brown et al. 1993; Snow et al. 1993; Eichler et al. 1994; Hirst et al. 1994; Kunst and Warren 1994). Penetrant fragile X syndrome patients exhibit repeat lengths $>\sim$230 triplets, often approaching 1,000 repeats (Kremer et al. 1991). When the CGG repeat is expanded, the FMR1 gene is abnormally methylated and transcriptionally silent (Pieretti et al. 1991; Sutcliffe et al. 1992). The absence of the RNA-binding protein encoded by the FMR1 gene is believed to be responsible for the clinical phenotype associated with the syndrome (Ashley et al. 1993; Siomi et al. 1993).

Nonpenetrant fragile X syndrome carriers have CGG-repeat lengths intermediate in size between patients and normal individuals (Oberlé et al. 1991). These alleles, termed "premutations," are extremely unstable when transmitted, often increasing in repeat length within the premutation range. When maternally transmitted, these unstable premutations have a probability of undergoing the massive expansions into the affected size range that is directly related to the maternal CGG-repeat length (Fu et al. 1991). However, a transition from a normal allele into the premutation range (i.e., a new mutation) has not been described, despite the frequent incidence of fragile X syndrome and a significant reduction in fitness of penetrant individuals (Smits et al. 1993). Thus, speculation has focused on the concept of predisposed normal alleles, relatively prevalent in the population, which comprise the pool of ancestral alleles from which fragile X chromosomes are derived (Chakravarti 1992; Morton and Macpherson 1992).

Support for predisposed normal alleles has been pro-

Received October 18, 1995; accepted for publication December 14, 1995.

Address for correspondence and reprints: Dr. Stephen T. Warren, Howard Hughes Medical Institute, Emory University School of Medicine, 1510 Clifton Road, Room 4035 RRC, Atlanta, GA 30322. E-mail: swarren@bimcore.cc.emory.edu



0002-9297/96/5803-0010$02.00

vided by studies demonstrating linkage disequilibrium between normal and fragile X chromosomes, by using polymorphisms near the FMR1 repeat (Richards et al. 1992; Buyle et al. 1993; Oudet et al. 1993*a*, 1993*b*; Arinami et al. 1993; Macpherson et al. 1994). From such studies, a limited pool of founder chromosomes has been postulated for all fragile X chromosomes throughout the world, consisting of those normal alleles predisposed toward expansion over many future generations. Recent studies have strongly implicated not only the overall repeat length as possibly contributing toward this predisposition, but more important, the purity of the CGG repeat (or lack of interrupting AGG triplets) as playing a major role. Kunst and Warren (1994) have suggested that those FMR1 alleles with >24 perfect CGG repeats may constitute the pool of predisposed chromosomes, prone to ever-increasing instability as the repeat slightly increases in length with each generation. Eichler et al. (1994) strongly supported this concept with the finding that normal FMR1 alleles with >34 perfect CGG repeats have attained a mutation rate sufficiently elevated to lead to observable instability when transmitted.

The occurrence of fragile X syndrome in numerous ethnic populations could be attributed to such founder alleles that either predate the radiation of human races or are recurrent within many diverse genetic backgrounds. Indeed, recent studies at other trinucleotide repeat loci (Ashley and Warren 1995) suggest that there may be significant differences in repeat-length alleles and haplotype backgrounds among various populations (Edwards et al. 1992; Zerylnick et al. 1995). In particular, Burke et al. (1994) surveyed normal alleles at the dentatorubral-pallidoluysian atrophy (DRPLA) locus in three populations and found that alleles with >24 perfect CAG repeats were most prevalent among the Japanese, where DRPLA is most frequent, and absent among Caucasians, where DRPLA is extraordinarily rare. African Americans exhibit a single allelic class >24 perfect repeats, which may be the precursor allele for Haw River syndrome, infrequently found in this population. Taken together, these data suggest that normal allelic frequencies may vary among populations and, in some instances, have an impact on the relevant disease incidence in that population. We report the detailed analysis of the FMR1 locus, including repeat length, cryptic content, and haplotype background, on global populations as well as introduce a method for utilizing automated sequencing technology in the analysis of FMR1.

## Material and Methods

### *DNA Samples*

DNA samples representative of unrelated Caucasians, Africans, Asians, and Native Americans were examined at the FMR1 locus. Our Caucasian population comprised 110 normal chromosomes from the United States and Canada. The Wolof and Mandenka of Senegal constituted the representative African groups. The primarily urban Wolof constitute 36% (Scozzari et al. 1988) of the Senegalese population, while the Mandenka, who are concentrated in rural regions, comprise 9% of the population (Chen et al. 1995). The Asians include samples from Tibet and the Kadazan population from the Sabah state of Borneo. The samples from Tibet are derived from individuals of ascertained Tibetan ancestry (Torroni et al. 1994*b*). The Kadazan represent the largest ethnic group in Sabah and are believed to have arisen from an Austronesian migration originating in southern China (Ballinger et al. 1992). Native American samples include the Mataco from northern Argentina, the Navajo from the southwestern United States, and four geographically and genetically similar populations from southern Mexico. These samples from southern Mexico, noted as "Mixtecs," comprise the Zapotecs, Mixe, and Mixtecs from the Alta and the Baja. Both the Mataco and the Mixtecs arose from the same Amerind migration from Asia to the Americas ≥21,000 years ago (Torroni et al. 1993). The Navajo arose from the more recent Nadene migration from Asia between 7,000 and 10,000 years ago (Torroni et al. 1992; Torroni and Wallace 1994). Caucasian admixture in the Mataco, Mixtecs, and Navajo is estimated as between 4% and 6%, on the basis of mtDNA analysis and Y chromosome haplotypes (Torroni et al. 1992, 1994*a*). In total, 345 chromosomes from non-Caucasians were analyzed at FMR1, including 101 Native Americans (29 Mataco, 38 Mixtec, and 34 Navajo), 97 Asians (66 Tibetan and 31 Bornean), and 147 Africans (16 African American, 50 Wolof, and 81 Mandenka). Because some of these samples were limited in availability, only 302 of these chromosomes were analyzed at FRAXAC1, including 111 Native Americans (29 Mataco, 46 Mixtec, and 36 Navajo), 57 Asians (31 Tibetan and 26 Bornean), and 134 African (15 African American, 41 Wolof, and 78 Mandenka).

### *Haplotype Analysis*

The FRAXAC1 locus was amplified using the primers designed by Richards et al. (1991) with one primer end-labeled with $\gamma$-$^{32}$P-ATP and T4 polynucleotide kinase as described elsewhere (Kunst and Warren 1994). All alleles were sized against alleles of known size run as controls, a $^{32}$P-labeled *Hpa*II pBR322 size standard, and an M13 sequencing ladder (Sequenase, USB). Amplification of the FMRa and FMRb alleles were performed simultaneously as described elsewhere (Kunst and Warren 1994). All FMRa and FMRb PCR products were analyzed against previously sequenced alleles as well as a $^{32}$P-labeled *Hpa*II pBR322 ladder. FMR1 CGG repeats

were analyzed by PCR as described elsewhere (Kunst and Warren 1994). Following the addition of loading dye, the PCR products were electrophoresed for 2.5–3.5 h at 80 W in a 5% Long Ranger polyacrylamide gel (AT Biochem) and were sized against sequenced CGG repeats of known sequence as well as a $^{32}$P-labeled *Hpa*II pBR322 ladder.

### *DNA Sequencing*

The FMR1 CGG repeat was amplified in a 50-μl reaction with primer FMR-c and biotinylated primer FMR-f (Fu et al. 1991) without the addition of α-$^{32}$P-dCTP. PCR products were isolated with streptavidin magnetic beads (Dynal) and sequenced either with an internal labeled primer as described elsewhere (Kunst and Warren 1994) or with a modified procedure on the ABI 373 automated sequencer. Samples with limiting amounts of DNA were subject to a nested PCR and sequencing procedure. An initial round of 25–30 PCR cycles were performed using primer A and primer 571R (Chong et al. 1994) in a 10-μl reaction containing 10–50 ng of DNA under standard FMR PCR conditions (Kunst and Warren 1994). A 1-μl aliquot of this initial reaction was reamplified in a 25-μl reaction with primer FMR-c and biotinylated FMR-f without the addition of α-$^{32}$P-dCTP as described elsewhere (Kunst and Warren 1994). The biotinylated strand was isolated as recommended by ABI (user bulletin 21), by using streptavidin magnetic beads, and sequenced with 3.2-pmol primer 1 (Brown et al. 1993), 5% dimethyl sulfoxide, and the ABI Prism Ready Reaction Dyedeoxy Terminator Cycle Sequencing Kit. Twenty-five cycles of 30 s at 94°C, 15 s at 60°C, and 1 min at 70°C were performed in a Perkin Elmer thermocycler 480. Sequencing products were purified with Centri-Sep columns (Princeton Separations), as recommended by the manufacturers, and analyzed on the ABI 373 automated sequencer.

### *Parsimony Analysis*

Sequences were encoded (C = CGG; and A = AGG) and aligned using ClustalW software (default setting: gap penalty 15.0, gap extension 1.0). Parsimony analysis (Swofford 1993) was performed using branch-and-bound and heuristic searches (PAUP 3.1). Only haplotypes which had at least four different sequences were analyzed, because of PAUP's requirement of at least four taxa. Cladograms were generated both with and without a defined ancestral state. Ancestral states were chosen on the basis of the most frequent allele within a particular haplotype for each ethnic group.

### *Statistical Analysis*

All distributions were assessed using standard $\chi^2$ analysis.

## Results

### *Haplotype Analysis*

Population groups representative of Africans, Asians, and Native Americans were examined at FMR1, FRAXAC1, FMRa, and FMRb and compared to a pool of 110 North American Caucasian chromosomes described elsewhere (Kunst and Warren 1994). FRAXAC1 is a well-characterized dinucleotide repeat polymorphism located 7 kb upstream of the FMR1 CGG repeat, which in Caucasians demonstrates linkage disequilibrium between normal and fragile X chromosomes (Richards et al. 1992). The distribution of FRAXAC1 alleles was determined for each population and compared to each other and to the normal Caucasian distribution (table 1). All populations had distributions of FRAXAC1 alleles that were distinct from Caucasians ($P < .05$). Among the Navajo, Mixtecs, and Mataco, the near-even split between alleles C and D reflects the common origin of these Native Americans. The D allele was the most frequent allele of the Tibetan and Bornean populations, in contrast to the C allele, which is most frequent among Caucasians. Greater heterogeneity was observed in the African populations (Wolof and Mandenka) as well as in a small sample of African Americans. The A allele of FRAXAC1, which is preferentially associated with fragile X chromosomes in Caucasians, was not observed in either the Native American or Asian populations ($P < .05$). Two diallelic intragenic polymorphic loci, FMRa and FMRb (Kunst and Warren 1994), were also analyzed in these populations, and only the A alleles of FMRa and FMRb were observed in the non-Caucasian samples (data not shown).

### *Distribution of FMR1 Alleles*

The distribution of FMR1 alleles was determined for all populations, with alleles being sized against CGG repeats of known sequence. In Caucasians, the mode repeat number is 30 (Brown et al. 1993; Hirst et al. 1994; Kunst and Warren 1994; Snow et al. 1994). Unique distributions of CGG-repeat lengths were observed in all populations (table 2). It is interesting that the mode repeat number in non-Caucasian populations was either 29 or 30 or was evenly split between these repeat alleles. In the Mataco, Mixtec, and Navajo, markedly less heterogeneity was observed than in Caucasians or other populations (e.g., Wolof and Mandenka). These data are consistent with the known occurrences of genetic bottlenecks in the ancestral Native American populations that occurred during their migrations from Asia (Wallace and Torroni 1992). While heterogeneity within the CGG repeat was most likely lost in the bottlenecks experienced during their migrations to the Americas, it is important to observe how little variation in repeat length has accumulated in the >21,000 and >7,000

## Table 1

**Distribution of AC1 Alleles**

| | AC1 Alleles | | | | | | No. of Chromosomes |
|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | |
| Mataco | | | .48 | .52 | | | 29 |
| Mixtecs | | | .52 | .48 | | | 44 |
| Navajo | | | .50 | .50 | | | 36 |
| Tibetans | | | .23 | .77 | | | 31 |
| Borneans | | | .27 | .73 | | | 26 |
| Wolof | .05 | .02 | .41 | .49 | .02 | | 41 |
| Mandenka | .03 | .04 | .53 | .39 | | .03 | 78 |
| African American | | .13 | .47 | .20 | .13 | .07 | 15 |
| Caucasian | .09 | .01 | .75 | .15 | | | 110 |

years, respectively, since the Amerind and Nadene migrations occurred.

*Sequence Analysis of FMR1*

The position of AGG interruptions within the FMR1 CGG repeat is highly polymorphic and appears to be involved in repeat stability (Eichler et al. 1994; Kunst and Warren 1994). For this reason, sequence analysis on the FMR1 repeat was performed on alleles from the representative populations. FMR1 alleles were PCR amplified and sequenced directly from the PCR products using internal primers as described elsewhere (Kunst and Warren 1994), or as modified for the ABI 373 automated sequencer as described in Material and Methods (fig. 1). Because this is a direct sequencing method, only males are useful for the sequence analysis. The cognate sequence of the FMR1 CGG repeat was determined for 114 of the 150 haplotyped males in our sample (fig. 2), including 28 Native American (21 Mataco and 7 Navajo), 37 Asian (15 Tibetan and 22 Bornean), and 49 African (29 Mandenka, 8 Wolof, and 12 African American) males. Thirty-six unique FMR1 alleles were seen in these populations, which is less than the heterogeneity observed in the Caucasian population assessed by Kunst and Warren (1994), but which is similar to that found in a random sampling of 102 chromosomes by Hirst et al. (1994). These differences may be due in part to the selection by Kunst and Warren (1994) of larger normal FMR1 repeats, which are highly heterogeneous within the normal population. Two alleles, 83 $(CGG)_9AGG(CGG)_9AGG(CGG)_9$ and 40 $(CGG)_nAGG(CGG)_9AGG(CGG)_9$ (see the appendix) represent 57% of the alleles observed within the non-Caucasian populations. The remaining alleles tended to be unique to individuals or populations; for example, allele 68 was present in three Navajo samples but absent from other non-Caucasian populations.

In Caucasians, all FMR1 repeats associated with the FRAXAC1-A allele have the first AGG at triplet position 10, and 96% have the second AGG at triplet position 20 (Kunst and Warren 1994). Only three alleles were observed with the FRAXAC1-A allele in our non-Caucasian sample. These three African alleles all had the first AGG at triplet position 11, and two had a second AGG at triplet position 21. These data, taken with the absence of the B alleles of FMRa and FMRb in non-Caucasian

## Table 2

**Distribution of *FMR1* Alleles**

| | *FMR1* CGG-Repeat Length | | | | | | | | | | | | | | | | | | | | | | | | | | | | | No. of Chromosomes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 13 | 14 | 15 | 16 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 49 | |
| Mataco | | | | | | | | | | | | | | .55 | .45 | | | | | | | | | | | | | | | | | 29 |
| Mixtecs | | | | | | .03 | .05 | .03 | | | | | | .42 | .47 | | | | | | | | | | | | | | | | | 38 |
| Navajo | | | | | | | | | | | | | .03 | .53 | .32 | | | | | | | | | .12 | | | | | | | | 34 |
| Tibetan | | | | .02 | .02 | | .02 | .03 | | | .05 | | .02 | .64 | .14 | .02 | | | | .02 | .05 | | | | | .02 | | | | | | 66 |
| Bornean | | | | | | | .03 | .06 | | | .03 | | | .39 | .39 | | | | | | .03 | .03 | | | | | | | | | .03 | 31 |
| African American | | | .06 | | | | .06 | | | | | | | .19 | .19 | .19 | .06 | | | | | .06 | | | | | .06 | | | .06 | | 16 |
| Wolof | | .02 | | | | | | .02 | .04 | .03 | .04 | | | .32 | .30 | .08 | .04 | | .06 | .02 | .02 | .02 | .04 | | | | | | | | | 50 |
| Mandenka | | | | | .01 | .03 | .06 | | .01 | | .01 | | .01 | .18 | .28 | .16 | .01 | .06 | | | .03 | .03 | .04 | .03 | .03 | | | | .01 | | | 80 |
| North American Caucasian | .01 | | | | .03 | .07 | .04 | .02 | .04 | | .01 | .01 | .04 | .06 | .34 | .16 | .05 | .05 | .02 | .03 | | | .01 | | | .01 | .01 | .02 | | | | 110 |

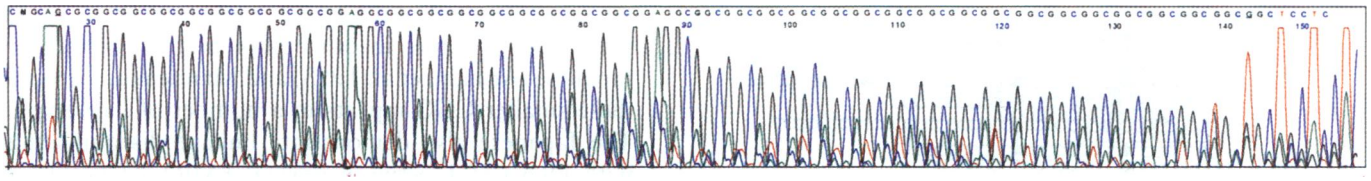

**Figure 1** ABI 373A sequencing of the FMR1 CGG repeat

populations, suggest that the FRAXAC1-A allele arose independently in Caucasians and Africans.

In our study, the Mataco demonstrated only two FMR1 CGG-repeat lengths. Sequencing revealed that these were comprised of three distinct alleles. It is interesting that the FRAXAC1 haplotype was very tightly associated with repeat sequence in the Mataco. All individuals with haplotype 1 had 29 CGG repeats and interruptions at position 10 and 20 (Mataco alleles 1–11, fig. 2). Only one chromosome with haplotype 2 (Mataco allele 12) had this repeat structure. The majority of haplotype 2 chromosomes had 30 CGG repeats with AGGs at positions 11 and 21 (Mataco alleles 14–21, fig. 2). Mataco allele 13 (fig. 2) likely resulted from the loss of an AGG triplet from triplet position 11, resulting in a CGG-repeat allele with 29 repeats and a single AGG interruption at position 20.

The Navajo appear to be stably segregating an FMR1 CGG repeat with 39 CGG triplets present in 12% (table 2) of the population. These FMR1 alleles, associated with the FRAXAC1-D allele, were observed in three individuals. In Caucasians, the FMR1 CGG repeat typically has interrupting AGG triplets at position 10 or 11 (90%; 74 of 82), and 20 or 21 (63%; 50 of 82) which seem to confer stability to the CGG repeat (Kunst and Warren 1994). Because haplotype and sequence analysis of Caucasians suggests that FMR1 repeats with >24 uninterrupted CGG triplets may be prone to instability (Kunst and Warren 1994), we were interested in determining the cryptic nature of the 39 repeat alleles in the Navajo. Each of the Navajo chromosomes with 39 CGG repeats has an identical cryptic structure with interruptions at triplet positions 10, 20, and 30 (Navajo alleles 1–3, fig. 2), making these alleles stable not only in length but in cryptic nature for at least 7,000 years.

### *Parsimony Analysis*

Cladograms for FMR1 alleles were generated for each FRAXAC1 haplotype in each population that had more than four unique alleles, both with and without defining an ancestral allele as described in the Material and Methods. Samples from different populations of a single FRAXAC1 haplotype were not pooled, to avoid the risk of FRAXAC1 revertants. Because genetically closed populations are likely to be descended from a small pool of founder chromosomes, it is probable that all FRAXAC1 alleles of a particular haplotype are derived from the same progenitor X chromosome. In most cases, the definition of an ancestral allele altered the resulting cladogram. Identical cladograms, both with and without the definition of an ancestral FMR1 allele, were generated for the FRAXAC1-D haplotype of the Borneans and the FRAXAC1-C haplotype of the Mandenka (fig. 3). It is interesting that each cladogram predicts the loss of one or two AGG interruptions from progenitor alleles containing two AGG triplets.

## Discussion

The structure of the FMR1 CGG repeat is highly heterogeneous. As shown in the appendix, 83 unique FMR1 alleles have been observed in the 234 alleles presented in this study. It is interesting that three of these (alleles 39, 40, and 83) represent 47% of the observed alleles. The remaining alleles tend to be unique to either individuals or populations. There has been some uncertainty as to whether FMR1 alleles containing a single AGG interruption—or lacking interruptions altogether—result from the loss of AGG triplets from a progenitor allele containing two interruptions or arose from independent founder alleles. The homogeneity of the Mataco alleles suggests that Mataco allele 13 (fig. 2) resulted from the loss of an AGG interruption. Parsimony analysis of the Mandenka and Borneans (fig. 3) also suggests that FMR1 alleles with a single AGG interruption or no AGG interruptions can arise from CGG repeats containing two AGG triplets. Therefore, these data indicate that loss of an AGG occurs in humans and gives rise to alleles with more perfect repeats. These findings support the earlier suggestion (Eichler et al. 1994; Kunst and Warren 1994) that FMR1 alleles predisposed toward expansion may occur via this mechanism.

North American Caucasian chromosomes display marked linkage disequilibrium between normal and fragile X chromosomes as well as between normal chromosomes of different repeat lengths (Richards et al. 1992; Jacobs et al. 1993; Kunst and Warren 1994). All chromosomes with ≤27 CGGs in Caucasians were ob-

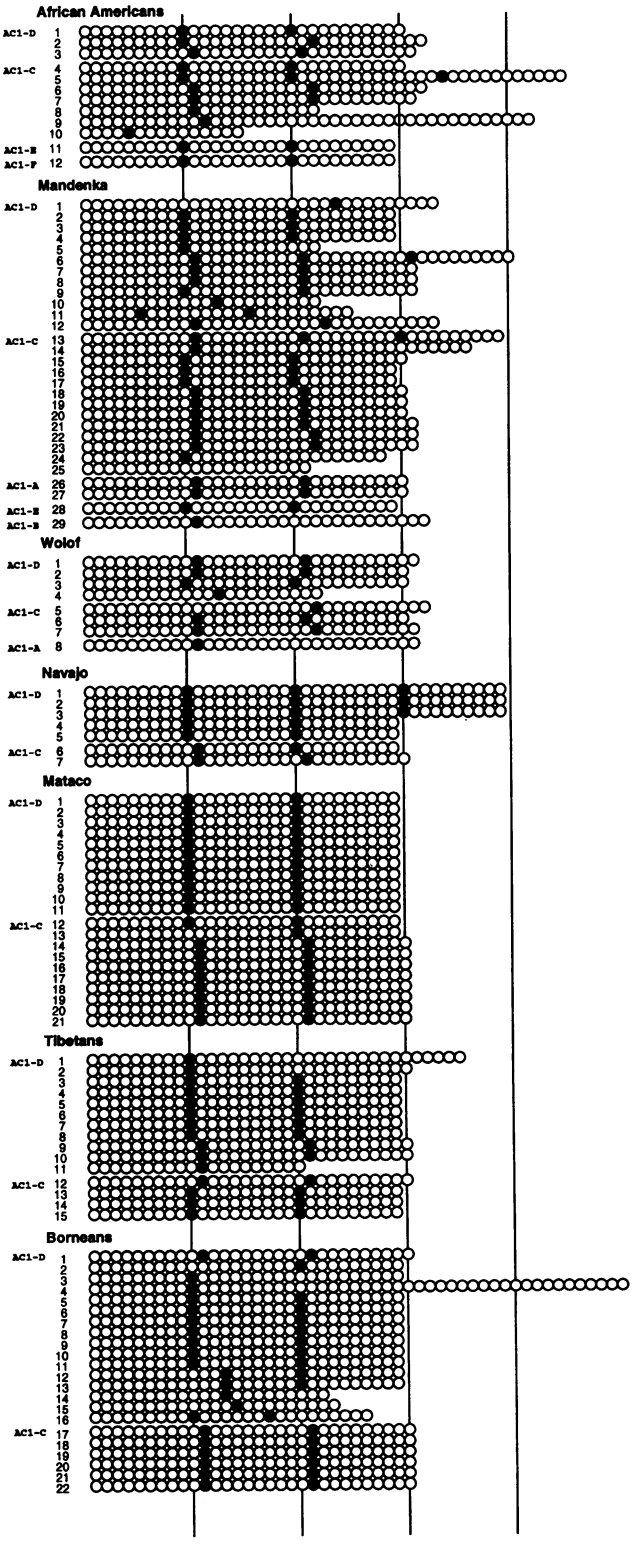


**Figure 2** Sequence analysis of FMR1 in global populations. Unblackened circles represent CGG triplets, while blackened circles represent AGG triplets. Haplotypes shown are the FRAXAC1 haplotypes as described by Jacobs et al. (1993).

served on a single FRAXAC1 haplotype ($P < .025$), while chromosomes in other populations, such as the Japanese (Arinami et al. 1993) and Mandenka, segregate these smaller CGG repeats in a number of haplotypes. The Japanese also demonstrate linkage disequilibrium with FRAXAC1 between normal FMR1 CGG-repeat-length classes and between some of these repeat-length classes and fragile X chromosomes (Arinami et al. 1993). However, unlike Caucasians, no linkage disequilibrium with FRAXAC1 is observed in Japanese chromosomes between normal chromosomes of all repeat lengths and fragile X chromosomes. In contrast to Caucasians and Japanese, no linkage disequilibrium between flanking FRAXAC1 alleles and the CGG repeat was observed in either the Wolof or Mandenka. Because there were no statistically significant differences in the distributions of FMR1 and FRAXAC1 alleles in these Africans, their haplotype data were pooled for reanalysis to increase the sample size to 121 chromosomes. Again, no statistically significant differences were observed in the distribution of FRAXAC1 alleles within different normal FMR1-repeat-length classes. The presence of linkage disequilibrium between normal CGG-repeat-length classes in Caucasians and the absence of this type of disequilibrium in the Wolof and Mandenka may be due in part to the relatively more recent origin of Europeans, as compared with Africans.

We previously suggested that Caucasian haplotype 3

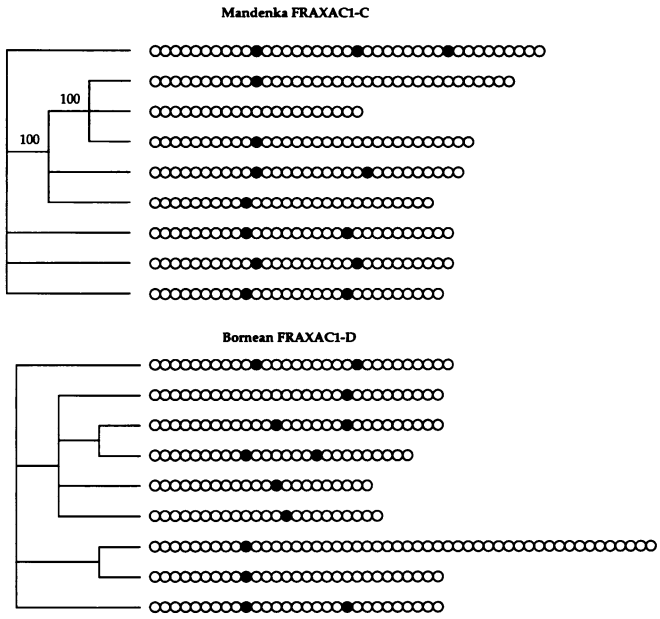


**Figure 3** Parsimony analysis of FMR1 in the Mandenka and Borneans. Unblackened circles represent CGG triplets, while blackened circles represent AGG triplets. Numbers represent the percentage of trees having a particular branch point. The Mandenka cladogram is an average of 26 trees of equal length. The Bornean alleles generated only a single tree.

chromosomes may not yet be in equilibrium in fragile X syndrome. In Caucasians, haplotype 3 represents the greatest percentage of CGG repeats with >24 uninterrupted CGG triplets, while representing only 25% of fragile X chromosomes (Kunst and Warren 1994). This and other data led us to suggest that in the future a greater proportion of fragile X chromosomes would arise on haplotype 3 chromosomal backgrounds. The absence of haplotype 3 alleles in non-Caucasian populations further supports the recent evolution of this haplotype and suggests that the disequilibrium between normal and fragile X chromosomes within this haplotype is unique to the Caucasian population.

If CGG-repeat alleles with >24 uninterrupted CGG repeats are the pool of normal chromosomes from which fragile X chromosomes evolve, the prevalence of predisposing normal alleles in the non-Caucasian populations studied is 3.5% ± 1.7%. This prevalence is within the range (0.5%–4%) predicted by Morton and Macpherson (1992) if fragile X syndrome is in equilibrium and similar to the 2% observed in Caucasian studies (Kunst and Warren 1994). However, these putative predisposing alleles were not observed in all populations, especially the Native American populations. This may be due either to the small sample size or to the reduction of genetic variation that these populations have undergone. Our results are similar to those observed by Zerylnick et al. (1995) for the myotonic dystrophy CTG-repeat locus where predisposing CTG-repeat alleles were either absent or reduced in frequency in Native Americans when compared with other populations. It is possible that predisposing FMR1 alleles were lost from Native American populations during the bottlenecks experienced in their migrations to the Americas (Wallace and Torroni 1992) and that the remaining alleles, with a marked cryptic nature, have remained relatively stable during the past 7,000 years. These data support the suggestion that the cryptic interruptions can serve to stabilize trinucleotide repeats and that fragile X syndrome might be extremely rare in these populations. Further data is needed on the frequency of fragile X syndrome in these populations, in order to compare disease frequency with the presence of predisposing normal alleles.

## Acknowledgments

We are indebted to Dr. S. L. Sherman, Dr. N. Arnheim, Dr. A. S. Santachiara-Benerecetti, Dr. L. Excoffier, Dr. C. R. Scott, Dr. L. G. Moore, Dr. C. M. Vullo, and Dr. Y. Y. Gan for providing part of the population samples. J.J.A.H was supported in part by the Ontario Mental Health Foundation. C.B.K. is an Associate and S.T.W. an Investigator of the Howard Hughes Medical Institute. This work was supported in part by NIH grants HD20521 and HD29256.

## Appendix

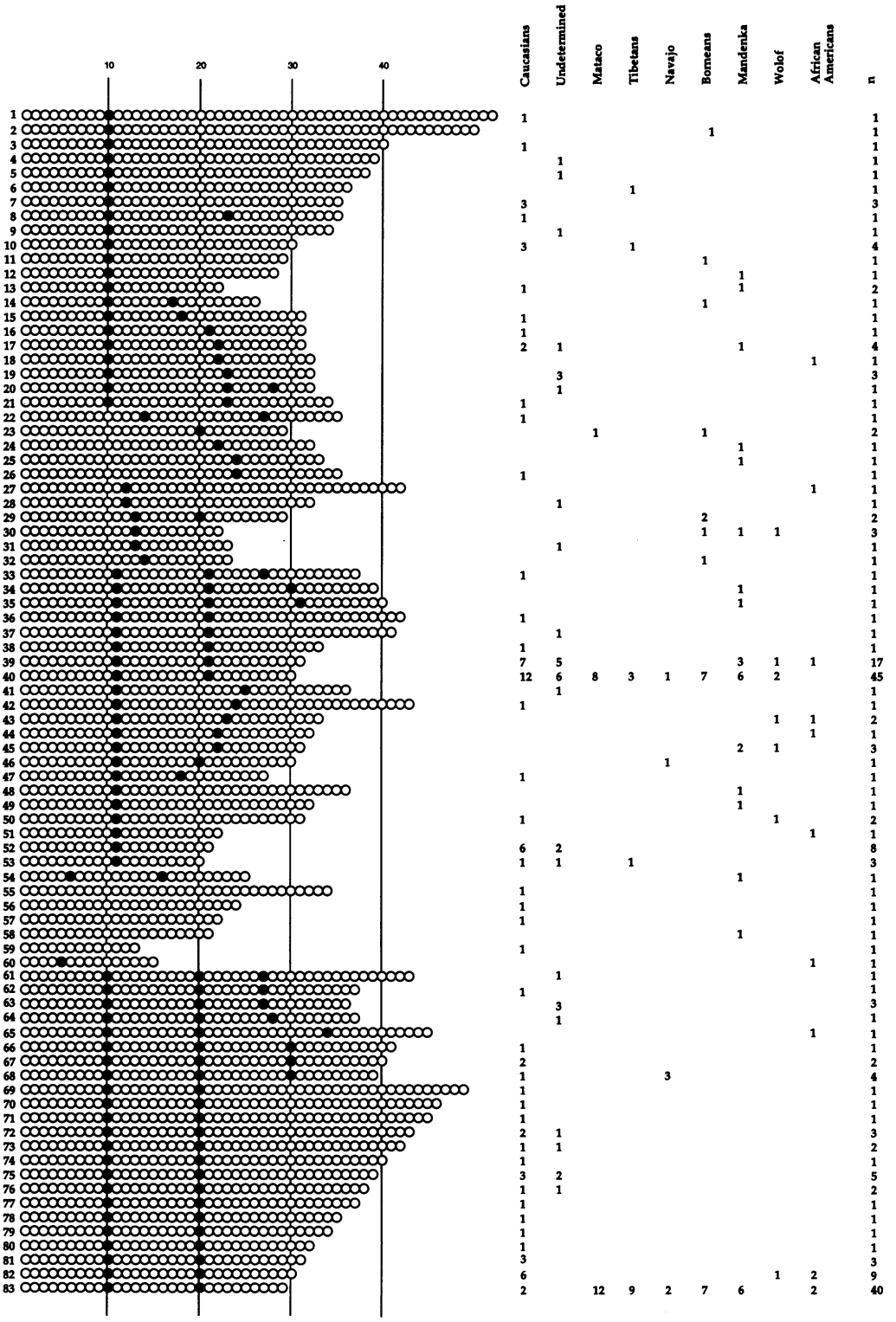


**Figure A1** Global FMR1 alleles. Unblackened circles represent CGG triplets; blackened circles represent AGG triplets; and stippled circles represent TGG triplets. Caucasians shown are those from Kunst and Warren (1994) and previously unpublished samples. Undetermined are sequenced alleles from individuals of unknown descent. The number of alleles is designated for each population. *n* is the total no. of alleles observed in all populations.

## References


Arinami T, Asano M, Kobayashi K, Yanagi H, Hamaguchi H (1993) Data on the CGG repeat at the fragile X site in the non-retarded Japanese population and family suggest the presence of a subgroup of normal alleles predisposing to mutate. Hum Genet 92:431–436

Ashley CT, Warren ST (1995) Trinucleotide repeat expansion and human disease. Annu Rev Genet 29:703–728

Ashley CT, Wilkinson KD, Reines D, Warren ST (1993) FMR1 protein: conserved RNP family domains and selective RNA binding. Science 262:563–566

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen KH, et al (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. Genetics 130:139–152

Brown WT, Houck GE, Jeziorowska A, Levinson FN, Ding X, Dobkin C, Zhong N, et al (1993) Rapid fragile X carrier screening and prenatal diagnosis using a nonradioactive PCR test. JAMA 13:1569–157527

Burke JR, Ikeuchi T, Koide R, Tsuji S, Yamada M, Pericak-Vance MA, Vance JM (1994) Dentatorubral-pallidoluysian atrophy and Haw River syndrome. Lancet 344:1711–1712

Buyle S, Reyniers E, Vits L, De Boulle K, Handig I, Wuyts FLE, Deelen W, et al (1993) Founder effect in a Belgian-Dutch fragile X population. Hum Genet 92:269–272

Chakravarti A (1992) Fragile X founder effect? Nat Genet 1:237–238

Chen Y-S, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. Am J Hum Genet 57:133–149

Chong SS, Eichler EE, Nelson DL, Hughes MR (1994) Robust amplification and ethidium-visible detection of the fragile X syndrome CGG repeat using *Pfu* polymerase. Am J Med Genet 51:522–526

Edwards A, Hammond HA, Jin L, Caskey CT, Chakroborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12:241–253

Eichler EE, Holden JJA, Popovich BW, Reiss AL, Snow K, Thibodeau SN, Richards CS, et al (1994) Length of uninterrupted CGG repeats determines instability in the *FMR1* gene. Nat Genet 8:88–94

Fu Y-H, Kuhl DPA, Pizzuti A, Pieretti M, Sutcliffe JS, Richards S, Verkerk AJMH, 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

Hirst MC, Grewal PK, Davies KE (1994) Precursor arrays for triplet repeat expansion at the fragile X locus. Hum Mol Genet 3:1553–1560

Jacobs PA, Bullman H, Macpherson J, Youings S, Rooney V, Watson A, Dennis NR (1993) Population studies of the fragile X: a molecular approach. J Med Genet 30:454–459

Kremer EJ, Pritchard M, Lynch M, Yu S, Holman K, Baker E, Warren ST, et al (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. Science 252:1711–1718

Kunst CB, Warren ST (1994) Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. Cell 77:853–861

Macpherson JN, Bullman H, Youings SA, Jacobs PA (1994) Insert size and flanking haplotype in fragile X and normal populations: possible multiple origins for the fragile X mutation. Hum Mol Genet 3:399–405

Morton NE, Macpherson JN (1992) Population genetics of the fragile-X syndrome: multiallelic model for the FMR1 locus. Proc Natl Acad Sci USA 89:4214–4217

Oberlé I, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer A, Boue J, et al (1991) Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science 252:1097–1102

Oudet C, Mornet E, Serre JL, Thomas F, Lentes-Zengerling S, Kretz C, Deluchat C, et al (1993*a*) Linkage disequilibrium between the fragile X mutation and two closely linked CA repeats suggests that fragile X chromosomes are derived from a small number of founder chromosomes. Am J Hum Genet 52:297–304

Oudet C, von Koskull H, Nordstrom AM, Peippo M, Mandel JL (1993*b*) Striking founder effect for the fragile X syndrome in Finland. Eur J Hum Genet 1:181–189

Pieretti M, Zhang F, Fu Y-H, Warren ST, Oostra BA, Caskey CT, Nelson DL (1991) Absence of expression of the FMR-1 gene in fragile X syndrome. Cell 66:817–822

Richards RI, Holman K, Friend K, Dremer E, Hillen D, Staples A, Brown WT, et al (1992) Fragile X syndrome: evidence of founder chromosomes. Nat Genet 1:257–260

Richards RI, Holman K, Kozman H, Dremer E, Lynch M, Pritchard M, Yu S, et al (1991) Fragile X syndrome: genetic localisation by linkage mapping of two microsatellite repeats FRAXAC1 and FRAXAC2 which immediately flank the fragile site. J Med Genet 28:818–823

Scozzari R, Torroni A, Semino O, Sirugo G, Brega A, Santachiara-Benerecetti AS (1988) Genetic studies on the Senegal population. I. Mitochondrial DNA polymorphisms. Am J Hum Genet 43:534–544

Siomi H, Siomi MC, Nussbaum RL, Dreyfuss G (1993) The protein product of the fragile X gene, FMR1, has characteristics of an RNA binding protein. Cell 74:291–298

Smits APT, Dreesen JCFM, Post JG, Smeets DFCM, de Die-Smulders C, Spaans-van der Bijl T, Govaerts LCP, et al (1993) The fragile X syndrome: no evidence for any recent mutations. J Med Genet 30:94–96

Snow K, Doud LK, Hagerman R, Pergolizzi RG, Erster SH, Thibodeau SN (1993) Analysis of a CGG sequence at the FMR-1 locus in fragile X families and in the general population. Am J Hum Genet 53:1217–1228

Snow K, Tester DJ, Krukeberg KE, Schaid DJ, Thibodeau SN (1994) Sequence analysis of the fragile X trinucleotide repeat: implications for the origin of the fragile X mutation. Hum Mol Genet 3:1543–1551

Sutcliffe JS, Nelson DL, Zhang F, Pieretti M, Caskey CT, Saxe D, Warren ST (1992) DNA methylation represses FMR-1 transcription in fragile X syndrome. Hum Mol Genet 1:397–400

Swofford D (1993) Phylogenetic analysis using parsimony (PAUP), Version 3.1.1. Illinois Natural History Survey, Champaign

Torroni A, Chen Y-S, Semino O, Santachiara-Benerecetti AS, Scott C R, Lott MT, Winter M, et al (1994*a*) mtDNA and Y-chromosome polymorphisms in four Native American

populations from southern Mexico. Am J Hum Genet 54:303–318

Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, Wallace DC (1994*b*) Mitochondrial DNA analysis in Tibet: implication for the origin of the Tibetan population and its adaptation to high altitude. Am J Phys Anthropol 93:189–199

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, et al (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53:563–590

Torroni A, Schurr TG, Yang C-C, Szathmary EJE, Williams RC, Schanfield MS, Troup GA, et al (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migration. Genetics 130:153–162

Torroni A, Wallace DC (1994) MtDNA variation in human populations and implications for detection of mitochondrial DNA mutations of pathological significance. J Bioenerg Biomemb 26:261–171

Verkerk AJMH, Pieretti M, Sutcliffe JS, Fu Y-H, Kuhl DPA, Pizzuti A, Reiner O, 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

Wallace DC, Torroni A (1992) American Indian prehistory as written in the mitochondrial DNA: a review. Hum Biol 64:403–416

Warren ST, Ashley (1995) Triplet repeat expansion mutations: the example of fragile X syndrome. Annu Rev Neurosci 18:77–99

Warren ST, Nelson DL (1994) Advances in molecular analysis of fragile X syndrome. JAMA 217:536–542

Zerylnick C, Torroni A, Sherman SL, Warren ST (1995) Normal variation at the myotonic dystrophy locus in global human populations. Am J Hum Genet 56:123–130