Published in final edited form as:

Ann Hum Genet. 2010 July ; 74(4): 316–325. doi:10.1111/j.1469-1809.2010.00582.x.

Genetic diversity of the fragile X syndrome gene (*FMR1*) in a large Sub-Saharan West African population

Emmanuel K. Peprah^{1,*}, Emzily G. Allen¹, Scott M. Williams², Laresa M. Woodard¹, and Stephanie L. Sherman¹

¹Department of Human Genetics, Emory University, Atlanta, Georgia, USA

²Department of Molecular Physiology and Biophysics and Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee, USA

Abstract

Fragile X-associated disorders (FAD) are caused by the expansion of a CGG trinucleotide repeat found in the 5' untranslated region of the X-linked *FMR1* gene. Although examinations of characteristics associated with repeat instability and expansion of the CGG repeat upon transmission from parent to offspring has occurred in various world populations, none has been conducted in large Sub-Saharan African populations. We have examined the *FMR1* CGG repeat structure in a sample of general population males drawn from Ghana. We found that Ghanaians and African Americans have similar allele frequency distributions of CGG repeat and its flanking STR markers, DXS548 and FRAXAC1. However, the distribution of the more complex marker, FRAXAC2, is significantly different. The haplotype structure of the *FMR1* locus indicated that Ghanaians share several haplotypes with African Americans and Caucasians that are associated with the expanded full mutation. In Ghanaians, the majority of repeat structures contained two AGG interruptions, however, the majority of intermediate alleles (35–49) lacked AGG interruptions. Overall, we demonstrate that allelic diversity of the *FMR1* locus among Ghanaians is comparable to African Americans, but includes a minority of CGG array structures not found in other populations.

Keywords

FMR1; African population; haplotype; trinucleotide repeat

Introduction

Fragile X associated disorders (FAD) are caused by the expanded CGG trinucleotide repeat found in the 5' untranslated region of the X-linked *FMR1* gene. Expansion of CGG repeats to >200 copies (i.e., the full mutation), induces the methylation of the repeats and repression of gene transcription of *FMR1*. The absence of fragile-X mental retardation protein (FMRP) is now the well-established cause of fragile X syndrome (FXS), a common inherited form of intellectual disabilities (Fu *et al.*, 1991, Sutcliffe *et al.*, 1992, Verkerk *et al.*, 1991). FXS has

*To whom correspondence should be addressed: Emmanuel Peprah, Department of Human Genetics, Emory University School of Medicine, 615 Michael Street, Rm 335, Atlanta, GA 30322, (404) 727-9396, epeprah@emory.edu.

been studied extensively in European populations and occurs in about 1/4000 males and 1/8000 females (Crawford *et al.*, 2001). In addition to FXS, individuals who carry 55-200 repeats (i.e., the premutation) are at risk for fragile X-associated tremor/ataxia syndrome (FXTAS) (Hagerman *et al.*, 2008) and fragile X-associated primary ovarian insufficiency (FXPOI) (De Caro *et al.*, 2008). Overall, analyses of *FMRI* and its gene product, primarily in Caucasian populations, have led to significant knowledge of the function of FMRP and the putative pathway leading to expansion of the CGG repeats.

Comparison of the haplotype background of chromosomes with the fragile X mutation (referred to as fragile X chromosomes) and those without has revealed several factors that increase CGG repeat instability, causing expansion from generation to generation. These factors include i) number of repeats in the CGG repeat region, ii) number of AGG interruptions within the sequence of CGG repeats, iii) CGG repeat purity of the repeat at the 3' end, iv) the 5' position of the first AGG interruption, and v) haplotype background of the mutation (Gunter *et al.*, 1998, Zhong *et al.*, 1995, Eichler *et al.*, 1994b, Eichler *et al.*, 1996, Crawford *et al.*, 2000c, Kunst & Warren, 1994, Snow *et al.*, 1993, Crawford *et al.*, 2000b). Based on these factors and the observed distribution of alleles among individuals with and without fragile X mutations, models have been proposed hypothesizing a multi-step mutational process from the stable common allele to the unstable fully expanded allele (Morton & Macpherson, 1992, Chakravarti, 1992, Chiurazzi *et al.*, 1996b, Eichler *et al.*, 1996, Snow *et al.*, 1993, Ashley & Sherman, 1995). One simplified model proposed by Morton *et al.* (1992) illustrates the idea of this multi-step expansion process. They proposed four states of the repeat mutation that occur in the following mutational sequence: the normal allele state (N), the small unstable CGG repeat (S), followed by larger unstable premutation expansion state (Z), and finally the large full mutation variant (L), which causes FXS. The expansion of ancestral, pre-disease alleles would also follow the multi-allelic model and progress from the normal stable allele ($N \rightarrow S \rightarrow Z$) to the full mutation state (L). This expansion model predicts that mutated founder chromosomes lead to alleles with the number of CGG repeats in the upper-normal range (35-49 repeats), from which recurrent multistep expansion mutations could arise (Oudet *et al.*, 1993a, Oudet *et al.*, 1993b, Buyle *et al.*, 1993, Malmgren *et al.*, 1994b). Based on the significant difference in the haplotype distributions observed between normal chromosomes and fragile X chromosomes, several groups have concluded that FXS chromosomes were derived from a few ancestral alleles, suggesting a differential susceptibility among normal alleles (Oudet *et al.*, 1993a, Oudet *et al.*, 1993b, Buyle *et al.*, 1993, Malmgren *et al.*, 1994b, Tranebjaerg *et al.*, 1994, Matilainen *et al.*, 1995, Syrrou *et al.*, 1996, Arrieta *et al.*, 1999).

Three mutational pathways have been proposed for the stepwise progression to the full mutation allele based on further examination of the repeat structures, both repeat length and AGG interruption pattern, along with the haplotype background of the mutation (Crawford *et al.*, 2000c, Eichler *et al.*, 1996). In brief, the haplotype 2-1-3 (based on the CGG flanking markers DXS548, FRAXAC1 and FRAXAC2 and using the nomenclature presented in Fig. 1) was associated with a highly interrupted repeat $(CGG)_9AGG(CGG)_9AGG(CGG)_9$ (i.e., 9+9+9) (Eichler & Nelson, 1996a). This repeat structure tended to retain the interruptions and slowly expand into the intermediate allele range through additions of CGGs at the 3' end

of the repeat track (Eichler *et al.*, 1996). In contrast, the 6-4-5 haplotype was associated with “asymmetrical” repeat patterns (e.g., 9+12+9 or 9+10+9) and was hypothesized to “leap frog” toward expansion due to the loss of the 3' AGG interruption. Such mutations were thought to bypass intermediate alleles (Eichler *et al.*, 1996). In addition, Crawford and colleagues found evidence in a sample of chromosomes from African Americans suggesting that the absence of the proximal-most 5' interruption within the repeat increased repeat instability (Crawford *et al.*, 2000c). This repeat structure, unique to the African American population, was associated with the haplotype 4-4-5 and accounted for almost half of the fragile X chromosomes in individuals of African American ancestry (Crawford *et al.*, 2000c). Thus, fragile X chromosomes are hypothesized to result from different mutational steps depending on the ancestral normal alleles, possibly influenced by the repeat structure and/or cis-acting factors.

To further understand the mutational mechanisms leading to CGG expansion and determine whether other pathways for expansion exist, characterization of the *FMR1* locus in various populations is valuable. Here, we have analyzed the repeat structure and haplotype background in males from a Sub-Saharan African population. This study population was beneficial because: i) Sub-Saharan Africans are ethnically and genetically diverse, ii) Africa is believed to be the ancestral homeland of all modern humans, thus African populations have a longer mutational history compared to other populations and iii) African populations have been understudied and may reveal new mutational pathways leading to full mutation.

Materials and Methods

Ghanaian study sample

In a previous study, population-based blood samples from adults were collected in Sunyani, Ghana to study the role of genetic variation on plasma t-PA and PAI-1 concentrations (Williams *et al.*, 2007). From this cohort we analyzed 350 adult males randomly selected from this population sample. All males were assumed to be unaffected, which was confirmed by genotyping. Sample sizes for each marker and for sequencing studies differed due to the ability to obtain accurate genetic results.

African American and Caucasian comparison data

Our group previously reported all data from African American and Caucasian males previously (Crawford *et al.*, 2002, Crawford *et al.*, 2000a, Crawford *et al.*, 2000c). Briefly, unaffected males of African American and Northern European ancestry were ascertained from the five-county metropolitan area of Atlanta, Georgia. Those affected with FXS were ascertained from Atlanta and other cities within the United States.

CGG repeat sequencing

CGG repeat sequencing and characterization of the AGG interspersed pattern in Ghanaians was conducted as previously described (Crawford *et al.*, 1999b, Crawford *et al.*, 2002, Crawford *et al.*, 2000a, Crawford *et al.*, 2000c). In brief, PCR amplification of the *FMR1* CGG repeat was performed using forward Primer A and reverse Primer 571R for the CGG repeat (Crawford *et al.*, 1999a). Amplification was carried out in a 25ul reaction volume

with 1x *Pfu* Buffer Plus (Roche), 0.2mM dNTPs, 100% DMSO, 1 U *Pfu* polymerase, 10mM Primer A and Primer 571R. The thermalcycling program used was 98°C for 5min, 35 cycles at 98°C for 1min, 65°C for 1min, 75°C for 2min and one cycle at 75°C for 10min. The amplified product was purified using Microcon tubes (Millipore) and prepared for sequencing. For sequencing Primer C was used. Sequencing was performed on the ABI 3730 at Center for Medical Genomics (CMG) (Emory University School of Medicine) in 96 well plate format. Primers used for amplification were:

Primer A 5'-GGAACAGCGTTGATCACGTGACGTGGTTTC-3'

Primer 571R 5'-CCCCGGACGGGATCTCGGTTTCATGGAACA-3'

Primer C 5'-GCTCAGCTCCGTTTCGGTTTCACTT-3'

AGG Interspersion Determination

AGG interruptions were determined after sequencing *FMR1* using established protocols in which two individuals scored the chromatograms independently. For each sequence, Phred score was evaluated as a method of QC for base calling accuracy (Ewing & Green, 1998, Ewing *et al.*, 1998). Reading of the sequence was accurate through approximately 30-40 CGG repeats depending on the number of AGG interruptions. The remaining number of 3' CGG repeats after the last interruption was inferred to be repeated CGGs based on the repeat length obtained from genotyping.

STR Genotyping

Genotyping calls for FRAXA CGG repeat length, DXS548, FRAXAC1 and FRAXAC2 in the Ghanaian sample was analyzed with sample controls from African American and Caucasian samples (Crawford *et al.*, 1999b, Crawford *et al.*, 2002, Crawford *et al.*, 2000a, Crawford *et al.*, 2000c). Briefly, genotyping of FRAXA was carried out in a 15ul reaction volume with 1x Buffer 1 (Roche), 10mM ATP, 10mM CTP, 10mM TTP, 2.5mM GTP, 10mM 7-deaza GTP (Roche), 10% DMSO, 50ng/ul T4 Gene 32 protein (Roche), 5 U/ul Expand Long Tag polymerase, and 10mM of primers. The thermalcycling program used was 1 cycle at 95°C for 2min, 10 cycles at 95°C for 30s, 65°C for 45s, 68°C for 4min, 20 cycles at 95°C for 30s, 65°C for 45s, 68°C for 4min 20s and finally 1 cycle for 68°C for 7min. Primers used for amplification were:

FRAXA-F 5'-GCTCAGCTCCGTTTCGGTTTCACTTC-3'

FRAXA-R 5'-AGCCCCGCACTTCCACCACCAGCTC-3'

Genotyping STR markers DXS548, FRAXAC1 and FRAXAC2 were multiplexed, using different 5' labeled fluorescent dyes for each. Amplification was carried out in a 15ul reaction volume with 1x Buffer 3 (Roche), 0.2mM dNTPs, 25mM MgCl₂, 100% DMSO, 5 U/ul Expand Long Tag polymerase (Roche), 10mM of primers. The thermalcycling program used was 1 cycle of 94°C for 5min followed by 30 cycles, 94°C for 30sec, 55°C for 30sec, and 72°C for 30sec and one cycle at 72°C for 5min.

DXS548 contains an insertion/deletion polymorphism (DXS548-P In/del) in the G/T track that flanks the CA repeats. This polymorphism has been characterized previously in African

Americans and Caucasians (Crawford *et al.*, 2002, Crawford *et al.*, 2000a, Crawford *et al.*, 2000c, Chiurazzi *et al.*, 1996c). To characterize this polymorphism in the Ghanaian population, DXS548-F1 and DXS548-P was used to amplify the CA repeat of DXS548. The amplicon that only contains the CA repeat was used to determine the insertion/deletion polymorphism for the marker. Primers used for amplification were:

DXS548 F1 5'-GTACATTTAGAGTCACCCTGTGGTGA-3'

DXS548 R1 5'-AGACTTCACTATGCAATGGAATC-3'

DXS548-P 5'-GAATAGTCTCTGGGGTGGATCTC-3'

FRAXAC1 F1 5'-GCTCAGCTCCGTTTCGGTTTCACTTC-3'

FRAXAC1 R1 5'-AGCCCCGCACTTCCACCACCAGCTC-3'

FRAXAC2 F1 5'-GACTGCTCCGGAAGTTGAATCCTCA-3'

FRAXAC2 R1 5'-CTAGGTGACAGAGTGAGAGTCCTGTC-3'

For analysis, each amplicon of FRAXA and all other STRs was resolved via capillary electrophoresis on the ABI 3100 and analyzed via Genescan v3.7 (Applied Biosystems) and each fragment was assigned an allele corresponding to the previously established nomenclature noted in Figure 1 (Crawford *et al.*, 2000a) using Genotyper v3.7 (Applied Biosystems).

Statistical Analysis

Allele frequency distributions for FRAXA, DXS548, FRAXAC1, and FRAXAC2 were compared among study samples using a χ^2 test. For each marker, alleles were collapsed into the next lower category when any one sample had fewer than five individuals for any allele count. Expected heterozygosities were calculated for each marker using the population statistics software Genepop V4.0 (<http://genepop.curtin.edu.au/>) and Fstat V2.9.3 under the assumption of random mating.

Results

STR loci flanking FMR1 in the Ghanaian population

The Ghanaian population was genotyped for the *FMR1* CGG repeat allele (FRAXA) and the three flanking STRs (Fig. 1). For the flanking markers, the Ghanaian sample had a lower or comparable number of alleles compared to the other populations, but a similar level of expected heterozygosity compared with African Americans and greater expected heterozygosity compared with Caucasians (Table 1). This pattern results from the more uniform distribution of frequencies at each marker in the Ghanaian sample as shown in Figure 2. With the exception of FRAXAC2, the modal class was the same in all samples: allele 7 for DXS548, allele 0 for DXS548-P, and allele 3 for FRAXAC1. With the exception of FRAXAC2 and DXS548-P, the flanking marker allelic distributions were not statistically different between the Ghanaian and African American samples (DXS548: $\chi^2 = 2.16$, $P = 0.539$, $df = 3$; FRAXAC1: $\chi^2 = 2.14$, $P = 0.543$, $df = 3$). However these markers did differ, between the Ghanaian and the Caucasian samples (DXS548: $\chi^2 = 96.7$, $P < 0.0001$, $df = 3$;

FRAXAC1: $\chi^2 = 56.1$, $P < 0.0001$, $df = 3$). As reported previously, the allele distributions for the African American sample differed statistically from that in the Caucasian sample (DSX548: $\chi^2 = 115$, $P < 0.0001$, $df = 3$; FRAXAC1: $\chi^2 = 80.6$, $P < 0.0001$, $df = 3$) (Crawford et al. 2000).

FRAXAC2 is a more complicated polymorphism because it includes three variable sites (Figure 1). The fragment size that we assessed is essentially a “phenotype” of the underlying variation and thus more difficult to compare among populations. As was done in previous studies, we simply took the fragment size as the genotype for comparisons. As mentioned above, despite the comparable expected level of heterozygosity in the Ghanaian and African American samples, there were fewer alleles in the Ghanaian sample, suggesting a more equal allele frequency distribution. This is seen in Figure 2D and demonstrated by the statistical comparisons where the allele distribution of the Ghanaians was significantly different from the other study samples (Ghanaians vs African Americans: $\chi^2 = 64.2$, $P < 0.0001$, $df = 7$; Ghanaians vs Caucasians: $\chi^2 = 294$, $P < 0.0001$, $df = 7$; Caucasians vs African Americans: $\chi^2 = 318$, $P < 0.0001$, $df = 7$).

To compare our results to previous studies, we constructed haplotypes using only DXS548, FRAXAC1 and FRAXAC2. We identified 72 different haplotypes in the Ghanaian population compared to 71 and 202 haplotypes identified in Caucasians and African Americans, respectively (Crawford et al 2000). The most frequent haplotypes found in the Ghanaian population (7-3-5 (0.11), 7-3-3+ (0.09) and 7-4-4 (0.06)) were not frequent in the other two study samples. For the African Americans and Caucasians, the 7-3-4+ as the most frequent haplotype (0.13 and 0.39 respectively) and was found at < 2% in the Ghanaian study sample.

Distribution of the CGG repeat and AGG interspersions patterns in Ghanaian population

We examined the *FMRI* CGG repeat polymorphism (FRAXA) in the Ghanaian study sample, both with respect to the repeat length and AGG interspersions pattern, and compared them to the African American and Caucasian samples. The expected heterozygosity for the FRAXA repeat for all three populations was similar, although the observed number of alleles differed: 23 for Ghanaians, 37 for African Americans and 42 for Caucasians (Table 1). Similar to the comparisons among samples for the other flanking markers, the allele frequency distributions did not differ significantly between Ghanaians and African Americans ($\chi^2 = 10.3$, $P = 0.172$, $df = 7$), but both differed from Caucasians (Ghanaian vs. Caucasians: $\chi^2 = 56.5$, $P < 0.0001$, $df = 7$; African Americans vs. Caucasians $\chi^2 = 320$, $P < 0.0001$). Repeat sizes of 29 and 30 were the more common alleles in the Ghanaian sample, similar to the other two study samples (Fig. 3A). As Crawford et al (2000) showed a significant distribution difference between African Americans and Caucasians at the smaller and larger CGG repeat size groups, we partitioned our data excluding the most frequent alleles of 29, 30 and 31 repeats, dividing our samples into smaller (< 28) and larger (> 32) allele groups. The distribution of smaller and larger allele size groups was not statistically different between the Ghanaian and African American samples ($\chi^2 = 1.03$, $P = 0.311$, $df = 1$) nor between Ghanaian and Caucasians ($\chi^2 = 0.476$, $P = 0.490$, $df = 1$).

The AGG interspersed pattern within the CGG repeat, for the Ghanaian sample, revealed 41 interspersed structures (Table 2). Among these structures, the 9+9+9 structure was the most frequent pattern found in the Ghanaian population followed by 10+9+9. Another way to summarize the varied repeat structures is to examine the number of AGG interruptions. The majority of Ghanaian structures contained two interruptions, which was similar to other populations characterized (Table 3). We also examined the repeat structure and haplotype associations among the intermediate size alleles (defined as 35-49 repeats), as these are prone to instability. Although the sample size was small, majority plurality of Ghanaian intermediate repeat alleles lacked interruptions (44%, Table 3).

Repeat structures in unaffected Ghanaians found on fragile X-associated haplotypes

Table 4 examines the overlap of haplotypes found in the Ghanaian sample with those found on chromosomes with and without the fragile X mutation among Caucasians and African Americans. Ghanaians shared seven haplotypes that have been associated with FXS chromosomes in Caucasians and five that have been associated with FXS in African Americans (Table 4). Three fragile X chromosome haplotypes (6-4-5, 7-3-4+ and 7-4-5) were shared among the three populations. However, the combined frequency of these three haplotypes on fragile X chromosomes was 26% and 22% among Caucasian and African American populations, respectively. Among chromosomes without the mutation, the combined frequencies were 44% and 7%, respectively. These compare with the combined frequency of only 2% in the Ghanaian population. The most common haplotypes in the Ghanaian sample among this group shared with fragile X chromosomes were 7-3-3+ and 6-4-4. These were in low frequency in the unaffected African American (<1%) and Caucasian (3%) samples, and slightly higher frequencies in the Ghanaian unaffected population (9% and 4% respectively).

We also examined the CGG/AGG pattern among the Ghanaian alleles that were found on haplotypes backgrounds associated with fragile X chromosomes as well as the pattern associated with the most common haplotype background in the Ghanaian sample (7-3-5). We identified structures that lacked AGG interruptions on both the FXS-associated haplotypes as well as the common haplotype (Fig. 4A-B). On the 7-3-3+ background, 3/11 had pure repeats compared with 1/12 on the common 7-3-5 haplotype. The largest pure repeat structure was found on the 7-3-4+ FXS-associated haplotype (36 pure repeats). On this same background, one allele had a large 3' pure repeat, which potentially indicates a risk for instability (9+26).

Discussion

This study represents the first detailed characterization of the *FMRI* CGG repeat structure and its flanking STR markers, DXS548, FRAXAC1, and FRAXAC2 in a Ghanaian population-based sample. The primary goal was to compare well characterized *FMRI* repeat alleles from a West African population sample, one that represents a portion of the African American gene pool, with previously characterized alleles that are known to be at risk for repeat instability in Caucasian and African American populations. This comparison should help determine factors that affect repeat stability.

With respect to flanking markers, the distributions of DXS548 and FRAXAC1 were similar for Ghanaian and African Americans populations, but both differed from the Caucasian population. This is expected based on the ages of the different populations. In addition, the allele frequency distributions in the Ghanaian sample were more platokurtotic than those from the other samples, again indicating a longer population history of these alleles. FRAXAC2 distribution was significantly different among all three populations, but because this flanking marker includes three sites of variation it may be too complex to interpret with respect to population dynamics of the *FMRI* repeat.

Interestingly in each population, the majority of repeat length alleles of FRAXA were comprised of 30 and 29 CGG repeats, even though the distribution of the alleles varied between the Caucasians and both populations of African descent (Fig 3A). These common alleles were also observed as most common in other West African populations including the Bamileke (Chiurazzi *et al.*, 1996a) Wolof, Mandinka (Kunst *et al.*, 1996) and Mbuti pygmy of Central Africa (Eichler & Nelson, 1996b). Similarly in other world populations, 30 and 29 repeats were the most common in Caucasians (Malmgren *et al.*, 1994a, Haataja *et al.*, 1994, Kunst *et al.*, 1996, Chiurazzi *et al.*, 1996b), South Americans comprising Brazilians with varied ancestry (Mingroni-Netto *et al.*, 2002), Chilean populations (Jara *et al.*, 1998), several Asian populations including Chinese, Malays (Zhou *et al.*, 2006) and Taiwanese (Chiu *et al.*, 2008). However, this pattern is not always the case: in a Japanese sample, 27 repeat alleles were the most common (Otsuka *et al.*, 2009) and in a Mexican sample, 32 repeats were the most common (Rosales-Reynoso *et al.*, 2005). However these populations also contained the 30 and 29 peaks in lower frequencies. Thus, although it is likely that the 29 or 30 repeat allele is the ancestral allele, founder effects due geographic and cultural isolation or complex interactions of populations have altered the frequency distributions (Eichler & Nelson, 1996b).

In the Ghanaian sample, the common interruption patterns were 10+9+9 (30 repeats) and 9+9+9 (29 repeats), similar to those previously reported for other African and Caucasian populations (Eichler & Nelson, 1996b, Kunst *et al.*, 1996). These structures contain two AGG interruptions that are less likely to expand compared to other AGG interspersed patterns (Eichler *et al.*, 1994a, Gunter *et al.*, 1998). Examination of intermediate repeat allele structures, or those at higher risk for instability, is instructive. The Ghanaian, Hellenic and Tunisian populations have higher frequencies of pure repeats among intermediate allele structures in contrast to African Americans and Caucasians (Table 3). In the Hellenic, the reported prevalence of FXS is similar to Caucasians and African Americans (Crawford *et al.*, 2002). However, the frequency of FXS is reported to be higher in the Tunisian, most likely due to a founder effect (Falik-Zaccari *et al.*, 1997, Patsalis *et al.*, 1999). At this point, it is not possible to accurately predict the frequency of FXS in Ghana based on these repeat structures. Other factors may also come into play beyond the repeat structure (Crawford *et al.*, 2000c). As Ghana represents an older and more diverse population, haplotype analyses may not provide additional hints with respect to susceptibility factors due to weak linkage disequilibrium (LD), as suggested previously for studies among African Americans (Crawford *et al.*, 2000c, Eichler *et al.*, 1994b). Further, recently identified cis-elements may be difficult to identify if LD is reduced (Beilina *et al.*, 2004, Entezam & Usdin, 2008, Ennis

et al., 2007). A complementary study of the frequency and haplotype backgrounds of chromosomes with the expanded fragile X mutation in the Ghanaian population would be useful to further characterize potential mutational pathways. Understanding these pathways would be important not only to understand fragile X associated disorders, but also important to gain insight to all repeat expansion disorders that arose in the human lineages

Acknowledgments

We the authors would like to kindly thank Dr. Dana Crawford for her help in manuscript preparations and also discussion. This work was supported by National Institutes of Health grants R01HD029909 and G12GM00680.

References

- Arrieta I, Gil A, Nunez T, Telez M, Martinez B, Criado B, Lostao C. Stability of the FMR1 CGG repeat in a Basque sample. *Hum Biol.* 1999; 71:55–68. [PubMed: 9972098]
- Ashley A, Sherman S. Population dynamics of a meiotic/mitotic expansion model for the fragile X syndrome. *Am J Hum Genet.* 1995; 57:1414–1415. [PubMed: 8533772]
- Beilina A, Tassone F, Schwartz PH, Sahota P, Hagerman PJ. Redistribution of transcription start sites within the FMR1 promoter region with expansion of the downstream CGG-repeat element. *Hum Mol Genet.* 2004; 13:543–9. [PubMed: 14722156]
- Buyle S, Reyniers E, Vits L, De Boulle K, Handig I, Wuyts F, Deelen W, Halley D, Oostra B, Willems P. Founder effect in a Belgian-Dutch Fragile X population. *Hum Genet.* 1993; 92:269–272. [PubMed: 8406435]
- Chakravarti A. Fragile X founder effect? *Nature Genetics.* 1992; 1:237–238. [PubMed: 1302018]
- Chiu H, Tseng Y, Hsiao H, Hsiao H. The AGG interruption pattern within the CGG repeat of the FMR1 gene among Taiwanese population. *J of Genetics.* 2008; 87:275–277. [PubMed: 19147914]
- Chiurazzi P, Destro-Bisol G, Genuardi M, Oostra BA, Spedini G, Neri G. Extended gene diversity at the FMR1 locus and neighbouring CA repeats in a sub-Saharan population. *Am J Med Genet.* 1996a; 64:216–9. [PubMed: 8826479]
- Chiurazzi P, Genuardi M, Kozak L, Giovannucci-Uzielli M, Bussani C, Dagna-Bricarelli F, Grasso M, Perroni L, Sebastio G, Sperandio M, Oostra BA, Neri G. Fragile X founder chromosomes in Italy: A few initial events and possible explanation for their heterogeneity. *Am J Hum Genet.* 1996b; 64:209–215.
- Chiurazzi P, Genuardi M, Kozak L, Giovannucci-Uzielli M, Bussani C, Dagna-Bricarelli F, Grasso M, Perroni L, Sebastio G, Sperandio M, Oostra B, Neri G. Fragile X founder chromosomes in Italy: A few initial events and possible explanation of their heterogeneity. *Am J Med Genet.* 1996c; 64:209–215. [PubMed: 8826478]
- Crawford D, Meadows K, Newman J, Taft L, Pettay D, Gold L, Hersey J, Hinkle E, Stanfield M, Holmgreen P, Yeargin-Allsopp M, Boyle C, Sherman S. Prevalence and Phenotype consequence of FRAXA and FRAXE alleles in a large, ethnically diverse, special education-needs population. *Am J Hum Genet.* 1999a; 64:495–507. [PubMed: 9973286]
- Crawford DC, Acuna JM, Sherman SL. FMR1 and the fragile X syndrome: human genome epidemiology review. *Genet Med.* 2001; 3:359–71. [PubMed: 11545690]
- Crawford DC, Meadows KL, Newman JL, Taft LF, Pettay DL, Gold LB, Hersey SJ, Hinkle EF, Stanfield ML, Holmgreen P, Yeargin-Allsopp M, Boyle C, Sherman SL. Prevalence and phenotype consequence of FRAXA and FRAXE alleles in a large, ethnically diverse, special education-needs population. *Am J Hum Genet.* 1999b; 64:495–507. [PubMed: 9973286]
- Crawford DC, Meadows KL, Newman JL, Taft LF, Scott E, Leslie M, Shubek L, Holmgreen P, Yeargin-Allsopp M, Boyle C, Sherman SL. Prevalence of the fragile X syndrome in African-Americans. *Am J Med Genet.* 2002; 110:226–33. [PubMed: 12116230]
- Crawford DC, Schwartz CE, Meadows KL, Newman JL, Taft LF, Gunter C, Brown WT, Carpenter NJ, Howard-Peebles PN, Monaghan KG, Nolin SL, Reiss AL, Feldman GL, Rohlf EM, Warren ST, Sherman SL. Survey of the fragile X syndrome CGG repeat and the short-tandem-repeat and

- single-nucleotide-polymorphism haplotypes in an African American population. *Am J Hum Genet.* 2000a; 66:480–93. [PubMed: 10677308]
- Crawford DC, Wilson B, Sherman SL. Factors involved in the initial mutation of the fragile X CGG repeat as determined by sperm small pool PCR. *Hum Mol Genet.* 2000b; 9:2909–2918. [PubMed: 11092767]
- Crawford DC, Zhang F, Wilson B, Warren ST, Sherman SL. Fragile X CGG repeat structures among African-Americans: identification of a novel factor responsible for repeat instability. *Hum Mol Genet.* 2000c; 9:1759–69. [PubMed: 10915764]
- De Caro J, Dominguez C, Sherman S. Reproductive health of adolescent girls who carry the FMR1 premutation: expected phenotype based on current knowledge of fragile X-associated primary ovarian insufficiency. *Ann N Y Acad Sci.* 2008; 1135:99–111. [PubMed: 18574214]
- Eichler E, Holden JJ, Popovich BW, Reiss AL, Snow K, Thibodeau S, Richards CS, Ward P, Nelson DL. Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nat Genet.* 1994a; 8:88–94. [PubMed: 7987398]
- Eichler E, Holden JJ, Popovich BW, Reiss AL, Snow K, Thibodeau S, Richards CS, Ward P, Nelson DL. Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nat Genet.* 1994b; 8:88–94. [PubMed: 7987398]
- Eichler E, Nelson L. Genetic variation and evolutionary stability of the FMR1 CGG repeat in six closed human populations. *Am J Hum Genet.* 1996a; 64:220–225.
- Eichler E, Nelson L. Genetic variation and evolutionary stability of the FMR1 CGG repeat in six closed human populations. *Am J Hum Genet.* 1996b; 64:220–225.
- Eichler EE, Macpherson JN, Murray A, Jacobs PA, Chakravarti A, Nelson DL. Haplotype and interspersed analysis of the FMR1 CGG repeat identifies two different mutational pathways for the origin of the fragile X syndrome. *Hum Mol Genet.* 1996; 5:319–30. [PubMed: 8852655]
- Ennis S, Murray A, Brightwell G, Morton NE, Jacobs PA. Closely linked cis-acting modifier of expansion of the CGG repeat in high risk FMR1 haplotypes. *Hum Mutat.* 2007; 28:1216–1224. [PubMed: 17674408]
- Entezam A, Usdin K. ATR protects the genome against CGG-repeat expansion in Fragile X premutation mice. *Nucleic Acids Res.* 2008; 36:1050–1056. [PubMed: 18160412]
- Ewing B, Green P. Base-calling of automated sequencer traces using Phred. II error probabilities. *Genome Res.* 1998; 8:186–194. [PubMed: 9521922]
- Ewing B, Hillier L, Wendl M, Green P. Base-calling of automated sequencer traces using Phred. I. Accuracy Assessment. *Genome Res.* 1998; 8:175–185. [PubMed: 9521921]
- Falick-Zaccari TC, Shachak E, Yalon M, Lis Z, Borochoff Z, Macpherson JN, Nelson DL, Eichler EE. Predisposition to the fragile X syndrome in Jews of Tunisian descent is due to the absence of AGG interruptions on a rare Mediterranean haplotype. *Am J Hum Genet.* 1997; 60:103–12. [PubMed: 8981953]
- Fu Y, Kuhl D, Pizzuti A, Pieretti M, Sutcliffe J, Richards S, Verkert A, Holden JJ, Fenwick R, Warren ST, Oostra BA, Nelson DL, Caskey C. Variation of the CGG repeat at the fragile X site results in genetic instability: Resolution of the Sherman paradox. *Cell.* 1991; 1047–1058. [PubMed: 1760838]
- Gunter C, Paradee W, Crawford DC, Meadows K, Newman J, Kunst C, Nelson D, Schwartz C, Murray A, Macpherson J, Sherman S, Warren S. Re-examination of factors associated with expansion of CGG repeats using a single nucleotide polymorphism in FMR1. *Hum Mol Genet.* 1998; 7:1935–1946. [PubMed: 9811938]
- Haataja R, Vaisanen ML, Li M, Ryyanen M, Leisti J. The fragile X syndrome in Finland: demonstration of a founder effect by analysis of microsatellite haplotypes. *Hum Genet.* 1994; 94:479–83. [PubMed: 7959680]
- Hagerman R, Hall D, Coffey S, Leehey M, Bourgeois J, Gould J, Zhang L, Seritan A, Berry-Kravis E, Olichney J, Miller J, Fong A, Carpenter R, Bodine C, Gane L, Rainin E, Hagerman H, Hagerman P. Treatment of fragile X-associated tremor ataxia syndrome (FXTAS) and related neurological problems. *Clin Interv Aging.* 2008; 3:251–261. [PubMed: 18686748]

- Jara L, Aspillaga M, Avendano I, Obreque V, Blanco R, Valenzuela C. Distribution of (CGG)_n and FMR1 associated microsatellite alleles in a normal Chilean populat. *Am J Med Genet.* 1998;277–282. [PubMed: 9475597]
- Kunst CB, Warren ST. Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell.* 1994; 77:853–61. [PubMed: 7911740]
- Kunst CB, Zerylnick C, Karickhoff L, Eichler E, Bullard J, Chalifoux M, Holden JJ, Torroni A, Nelson DL, Warren ST. FMR1 in global populations. *Am J Hum Genet.* 1996; 58:513–22. [PubMed: 8644711]
- Malmgren H, Gustavson K, Oudet C, Holmgren G, Pettersson U, Dahl N. Strong founder effect for the Fragile X syndrome in Sweden. *Eur J Hum Genet.* 1994a; 2:103–109. [PubMed: 8044655]
- Malmgren H, Gustavson K, Oudet C, Holmgren G, Pettersson U, Dahl N. Strong founder effect for the Fragile X syndrome in Sweden. *European Hum Genet.* 1994b; 2:103–109.
- Matilainen R, Airaksinen E, Mononen T, Launiala K, Kaariainen R. A population-based study of the causes of mild and severe mental retardation. *Acta Paediatrica.* 1995; 84:261–266. [PubMed: 7780246]
- Mingroni-Netto R, Angeli C, Auricchio T, Leal-Mesquita E, Ribeiro-Dos-Santos A, Ferrari I, Hutz M, Salzano F, Hill K, Hurtado A, Vianna-Morgante A. Distribution of CGG repeats and FRAXAC1/DXS548 alleles in South American populations. *Am J Med Genet.* 2002; 111:243–252. [PubMed: 12210320]
- Morton NE, Macpherson JN. Population genetics of the fragile X syndrome: Multiallelic model for the FMR1 locus. *Proc Natl Acad Sci.* 1992; 89:4215–4217. [PubMed: 1570349]
- Otsuka S, Sakamoto Y, Siomi H, Itakura M, Yamamoto K, Matumoto H, Sasaki T, Kato M, Nanba E. Fragile X carrier screening and FMR1 allele distribution in the Japanese population. *Brain Dev.* 2009
- Oudet C, Mornet E, Serre J, Thomas F, Lentes-Zengerling S, Kretz C, Deluchat C, Tejada I, Boue J, Boue A. Linkage disequilibrium between the Fragile X mutation and two closely linked CA repeats suggest that Fragile X chromosome are derived from a small number of founder chromosomes. *Am J Hum Genet.* 1993a; 52:297–304. [PubMed: 8094266]
- Oudet C, Von Koskull H, Nordstrom A, Peippo M, Mandel J. Striking Founder effect for the Fragile X syndrome in Finland. *Eur J Hum Genet.* 1993b; 1:181–189. [PubMed: 8044645]
- Patsalis PC, Sismani C, Hettinger JA, Boumba I, Georgiou I, Stylianidou G, Anastasiadou V, Koukoulli R, Pagoulatos G, Syrrou M. Molecular screening of fragile X (FRAXA) and FRAXE mental retardation syndromes in the Hellenic population of Greece and Cyprus: incidence, genetic variation, and stability. *Am J Med Genet.* 1999; 84:184–90. [PubMed: 10331587]
- Rosales-Reynoso M, Mendoza-Carrera F, Troyo-Saroman R, Medina C, Barros-Nunez P. Genetics diversity at the FMR1 locus in Mexican Population. *Archives of Medical Research.* 2005; 36:412–417. [PubMed: 15950084]
- Snow K, Doud L, Hagerman R, Pergolizzi R, Erster S, Thibodeau S. Analysis of a CGG sequence at the FMR-1 locus in Fragile X families and in the general population. *Am J Hum Genet.* 1993; 53:1217–1228. [PubMed: 7902673]
- Sutcliffe J, Nelson DL, Zhang F, Pieretti M, Caskey C, Saxe D, Warren ST. DNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum Mol Genet.* 1992:397–400. [PubMed: 1301913]
- Syrrou M, Patsalis P, Georgiou I, Hadjimarcou M, Constantinou-Deltas C, Pagoulatos G. Evidence for high-risk haplotypes and (CGG)_n expansion in Fragile X syndrome in the Hellenic population of Greece and Cyprus. *Am J Med Genet.* 1996; 64:234–238. [PubMed: 8826482]
- Tranebjaerg L, Hilling S, Jessen J, Lind D, Hansen M. Prevalence of Fra(X) in the county of Funen in Denmark is lower than expected. *Am J Med Genet.* 1994; 51:423–427. [PubMed: 7943011]
- Verkerk A, Pieretti M, Sutcliffe J, Fu Y, Kuhl D, Pizzuti A, Reiner O, Richards S, Victoria M, Zhang F, Eussen B, Van Ommen G, Blonden L, Riggins G, Chastain J, Kunst CB, Galjaard H, Caskey C, Nelson DL, Oostra BA, Warren ST. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell.* 1991; 65:905–914. [PubMed: 1710175]

- Williams S, Stocki S, Jiang L, Brew K, Gordon S, Vaughan D, Brown N, Poku K, Moore J. A population-based study in Ghana to investigate inter-individual variation in plasma t-PA and PAI-1. *Ethn Dis.* 2007; 17:492–497. [PubMed: 17985503]
- Zhong N, Yang W, Dobkin C, Brown WT. Fragile X gene instability: anchoring AGGs and linked microsatellites. *Am J Med Genet.* 1995; 57:351–361.
- Zhou Y, Tang K, Law HY, Ng IS, Lee CG, Chong SS. FMR1 CGG repeat patterns and flanking haplotypes in three Asian populations and their relationship with repeat instability. *Ann Hum Genet.* 2006; 70:784–96. [PubMed: 17044853]

Figure 1.

Location of the short tandem repeats (STRs) of the fragile X locus and the FRAXA interspersions patterns adopted from (Crawford *et al.*, 2000a, Zhou *et al.*, 2006). All DNA elements are arranged from centromere (CEN) to telomere (TEL) with filled arrows represent microsatellites and *FMRI* exons represented by filled boxes. For each STR the structure is represented below the STR name designation with allele designation and base pairs in parentheses above. DXS548 insertion/deletion polymorphism (DXS548-P) designated (+) for insertion and (-) for deletion, with base pair indicated by number. Exon 1 contains the polymorphic CGG repeat allele (represented by the white insert) with the AGG interspersions represented by (+) and total number of pure CGG arrays for each segment represented by numbers.

Figure 2. A, DXS548 ($n = 234$); B, DXS548-P in/del ($n = 156$); C, FRAXAC1 ($n = 234$); D, FRAXAC2 ($n = 234$); allele distributions for Ghanaians (black bars), Caucasians ($n = 721$; white bars) (Crawford *et al.*, 2000a) and African Americans ($n = 637$; gray bars) (Crawford *et al.*, 2000a).

Figure 3.

A, Distribution of *FMRI* CGG repeat alleles in Ghanaians ($n = 300$; black bars), Caucasians ($n = 721$; white bars) (Crawford *et al.*, 2000a) and African Americans ($n = 637$; Gray bars) (Crawford *et al.*, 2000a). Distributions of FRAXA CGG repeat alleles in West African populations (3B), Ghanaians ($n = 300$; black bars), Mandinka ($n = 80$; light gray bars) (Kunst *et al.*, 1996), Wolof ($n = 50$; white bars) (Kunst *et al.*, 1996) and Bamileke ($n = 82$; dark gray bars) (Chiurazzi *et al.*, 1996a).

Figure 4.

CGG repeat interspersions among Ghanaian alleles with open circles representing CGG arrays and filled in circles representing AGG interruptions. Haplotypes associated with fragile X chromosomes found in African Americans and Caucasians (5A) are represented with the AGG interruptions found in Ghanaians arranged in ascending order with respect to FraxA CGG repeat size. Interspersion patterns found associated with the FXS haplotypes in Ghanaians (5B).

Table 1
Expected heterozygosity of FRAXA and flanking loci for the Ghanaian, Caucasians and African American study samples.

	FRAXA			DXS548			FRAXAC1			FRAXAC2		
	Alleles	Expected Heterozygosity (Variance)		Alleles	Expected Heterozygosity (Variance)		Alleles	Expected Heterozygosity (Variance)		Alleles	Expected Heterozygosity (Variance)	
Ghanaians ^a	23	0.801	(15.777)	8	0.719	(2.636)	6	0.615	(0.833)	15	0.878	(3.071)
Caucasians ^b	42	0.838	(44.106)	10	0.485	(3.131)	6	0.469	(0.639)	13	0.687	(0.856)
African Americans ^c	37	0.808	(17.863)	14	0.685	(3.355)	6	0.603	(0.912)	18	0.880	(4.419)

^a Ghanaian sample: for FRAXA, n=300; for DXS548, FRAXAC1 and FRAXAC2, n=234.

^b Caucasian sample: n=721 (taken from Crawford, Schwartz et al 2000)

^c African Americans sample: n=637 (taken from Crawford, Schwartz et al 2000)

Table 2
Frequency of CGG/AGG repeat interspersions found among unaffected Ghanaian sample (n=110)

Interspersion Pattern	No. (frequency)	Interspersion Pattern	No. (frequency)
9+9+9	20 (0.162)	10+9	1 (0.008)
10+9+9	15 (0.122)	10+11	1 (0.008)
30 Pure	10 (0.081)	10+13	1 (0.008)
9+9+10	8 (0.065)	10+15	1 (0.008)
9+12	6 (0.049)	10+9+8	1 (0.008)
9+19	4 (0.032)	10+9+19	1 (0.008)
9+20	4 (0.032)	10+9+20	1 (0.008)
10+9+10	3 (0.024)	10+10+18	1 (0.008)
10+10+8	2 (0.016)	11+9	1 (0.008)
10+19	2 (0.016)	12+9+10	1 (0.008)
10+20	2 (0.016)	12+9	1 (0.008)
10+22	2 (0.016)	14+9	1 (0.008)
13+10	2 (0.016)	21+9	1 (0.008)
31 pure	2 (0.016)	23 pure	1 (0.008)
36 pure	2 (0.016)	28 pure	1 (0.008)
9+11+9	1 (0.008)	29 pure	1 (0.008)
9+10+11	1 (0.008)	33 pure	1 (0.008)
9+10+10	1 (0.008)	35 pure	1 (0.008)
9+10+9	1 (0.008)	37 pure	1 (0.008)
9+21	1 (0.008)		
9+26	1 (0.008)		
9+25	1 (0.008)		

Table 3
Number of AGG interruptions by study sample (modified from Crawford et al 2000c)

Among unaffected	0	1	2	3	4
Ghanaian (n=110) ^a	0.18	0.16	0.65	0	0
African American (n=213) ^b	0.03	0.25	0.68	0.04	0
Caucasian (n=565) ^c	0.04	0.29	0.63	0.03	0.01
Hellenic (n=109) ^d	0.07	0.17	0.74	0.02	0
Tunisian Jews (n=148) ^e	0.19	0.12	0.66	0.02	0
Among intermediate repeats (35-49)	0	1	2	3	>4
Ghanaian (n=9) ^a	0.44	0.22	0.33	0	0
African American (n=52) ^b	0.04	0.42	0.15	0.35	0.04
Caucasian (n=58) ^c	0	0.21	0.53	0.22	0.03
Hellenic (n=19) ^d	0.37	0.26	0.26	0.11	0
Tunisian Jews (n=16) ^e	0.38	0.06	0.38	0.19	0

^aCurrent study
^bCrawford, Zhang et al 2000
^cSnow et al 1994, Hirst et al 1994, Zhong et al 1995 and Eichler et al 1996
^dPatsalis et al 1999
^eFalik-Zaccai et al 1997

Table 4
Ghanaian repeat alleles found on FXS associated haplotype backgrounds in Caucasian (CAU) and African American (AA)

FXS Haplotype	Number of FXS cases (Freq.) ^a	Number of unaffected (Freq.) ^b	Number of Ghanaian unaffected (Freq.) ^c	Ghanaian CGG Repeat Allele (n) ^c
Caucasian				
<i>a</i> 6-4-5	16(0.157)	25 (0.035)	1 (0.004)	31(1)
<i>a</i> 7-3-4+	10 (0.098)	284 (0.394)	3 (0.013)	40(1), 36(2)
6-4-4	8 (0.078)	24 (0.330)	9 (0.038)	37(1), 34(1), 29(3)
3-3-3	1 (0.010)	0	6 (0.026)	29(5)
6-4-4+	1 (0.010)	1 (0.001)	1 (0.004)	25(18)
7-3-3+	1 (0.010)	21 (0.029)	21 (0.090)	30(18), 24(1)
<i>a</i> 7-4-5	1 (0.010)	6 (0.008)	1 (0.004)	29(1)
African American				
<i>a</i> 7-3-4+	9 (0.143)	83 (0.017)	3 (0.013)	40(1), 36(2)
<i>a</i> 6-4-5	3 (0.048)	14 (0.022)	1 (0.004)	31(1)
<i>a</i> 7-4-5	2 (0.032)	19 (0.030)	1 (0.004)	29(1)
7-6-4+	2 (0.032)	0	6 (0.026)	48(2), 36(1), 34(1), 29(2)
6-3-6+	1 (0.016)	0	1 (0.004)	31(1)

^a Caucasian (n=102) and African American (n=63), taken from Crawford, Schwartz et al 2000

^b Caucasian (n=575) and African American (n=220), taken from Crawford, Schwartz et al 2000

^c Current study (n=234)