Am. J. Hum. Genet. 48:1051–1057, 1991

Fragile X Syndrome: Diagnosis Using Highly Polymorphic Microsatellite Markers

Robert I. Richards, Yang Shen,[1] Katherine Holman, Helen Kozman, Valentine J. Hyland,[2] John C. Mulley, and Grant R. Sutherland

Department of Cytogenetics and Molecular Genetics, Adelaide Children's Hospital, North Adelaide

Summary

We describe two highly polymorphic microsatellite AC repeat sequences, VK23AC and VK14AC, which are closely linked to the fragile X at Xq27.3. Both VK23AC (*DXS297*) and VK14AC (*DXS292*) are proximal to the fragile site. Two-point linkage analysis in 31 fragile X families gave (*a*) a recombination frequency of 1% (range 0.00%–4%) with a maximum lod score of 32.04 for *DXS297* and (*b*) a recombination frequency of 7% (range of 3%–15%) with a maximum lod score of 12.87 for *DXS292*. Both of these polymorphisms are applicable to diagnosis by linkage in families with fragile X syndrome. A multipoint linkage map of genetic markers at Xq27.3 was constructed from genotyping these polymorphisms in the CEPH pedigrees. The *DXS292* marker is in the *DXS98-DXS297* interval and is 3 cM proximal to *DXS297*.

Introduction

The rare fragile site at Xq27.3 (*FRAXA*) is associated with the most common familial form of mental retardation (Sutherland and Hecht 1985). Prenatal diagnosis and carrier detection can be performed cytogenetically; however, incomplete penetrance of the fragile site renders this technology, in isolation, inaccurate. Diagnosis for individuals who do not express the fragile site therefore relies on polymorphic DNA markers closely linked to it (Sutherland and Mulley 1990).

The majority of DNA polymorphisms currently used for risk estimation are detected by Southern blot analysis of RFLPs (Suthers et al. 1991*a*). These markers are less common on the X than on the autosomes (Hofker et al. 1986), and laboratory analysis is labor intensive. In addition, the majority have only two alleles and therefore a maximum heterozygosity—and, in the case of X-linked markers, PIC—of 50%.

Recently, two laboratories have described polymorphisms associated with length variation in dinucleotide microsatellite repeats (Litt and Luty 1989; Weber and May 1989). These polymorphisms can be rapidly typed and are usually highly informative (compared with RFLPs). They are quite common; there are about 50,000 copies of the (AC)n repeat in the mammalian genome—i.e., approximately one every 50 kbp, if one assumes uniform distribution.

We have screened cloned DNA from the vicinity of Xq27.3 for dinucleotide microsatellite sequences to identify polymorphisms useful for linkage analysis near *FRAXA*. The established physical order for markers proximal to *FRAXA* is *DXS98* (4D-8)–*DXS292* (VK14)–*DXS369*(RN1)–*DXS297*(VK23)–*FRAXA* (Suthers et al. 1990). Two polymorphisms, VK23AC and VK14AC, are described, together with both their genetic mapping in the Centre d'Etude du Polymorphisme Humain (CEPH) pedigrees and the results of their application to 31 fragile X syndrome pedigrees.

Material and Methods

Synthetic poly (AC.GT) (Pharmacia) was radioac-

Received November 29, 1990; revision received January 24, 1991.

Address for correspondence and reprints: Dr. R. I. Richards, Department of Cytogenetics and Molecular Genetics, Adelaide Children's Hospital, North Adelaide, South Australia 5006.

1. On leave from the International Peace Maternity and Child Health Hospital of China Welfare Institute, Shanghai.

2. Present address: Kolling Institute of Medical Research, Royal North Shore Hospital, Saint Leonards, New South Wales, Australia.

0002-9297/91/4806-0004$02.00

tively labeled with alpha-^{32}P-dCTP in a random-primed reaction (Multiprime; Amersham). AC repeat–containing DNA sequences were identified by hybridization to this probe in 0.5 M sodium phosphate pH 7.0, 7% SDS (without carrier DNA) at 65°C for 16 h and by washing at 65°C for 1 h in 2 × SSC. Clones tested constituted the VK series which had been mapped to Xq26-qter (Hyland et al. 1989).

DNA from positive lambda clones was digested with *Sau*3A and was subcloned into *Bam*HI-cut M13 mp18 for sequence analysis. Synthetic oligodeoxyribonucleotide primers suitable for PCR were designed from apparently unique sequences flanking the microsatellite AC repeats. Length polymorphism of the AC repeats was typed in a PCR using the reaction conditions of Kogan et al. (1987), except for the addition of the 1 μCi of alpha-^{32}P-dCTP to each reaction. These PCR conditions were used because they have been successfully applied to multiplex PCR (Chamberlain et al. 1988, Richards et al., submitted).

PCR incubations were performed in 10-μl volumes in a Perkin Elmer-Cetus thermal cycler for 10 cycles at 94°C for 60 s, at 60°C for 90 s, and at 72°C for 90 s, followed by 25 cycles at 94°C for 60 s, at 55°C for 90 s, and at 72°C for 90 s. The volume was adjusted to 40 μl with formamide loading buffer (95% formamide, 1 mM EDTA, 0.01% bromophenol blue, 0.01% xylene cyanol). After denaturation at 90°C for 3 min, 2.5-μl aliquots of each reaction mixture were subjected to electrophoresis in 6% polyacrylamide denaturing (7 M urea) gels. Genotypes were determined after autoradiography for 18–48 h.

Linkage analysis were based on 31 kindreds. Five small kindreds with isolated cases of fragile X were excluded from the analysis because of uncertainty about relative mutation rates in males and females. Parameters used were as given elsewhere (Suthers et al. 1991*a*), except for allele frequencies of the marker loci, VK23AC and VK14AC, frequencies which were determined from unrelated individuals in the informative fragile X families and are given in table 1.

Modification of the phenotypic coding of family members was made in those cases in which closely linked flanking markers clearly indicated carrier status different from that determined by the conventional definition of phenotype. In previous studies the phenotype has been defined only by cytogenetic expression of the fragile X and by mental retardation. Now that there are many markers closely linked on either side of *FRAXA* (Suthers et al. 1991*a*), the definition of phenotype for the purposes of the present study was extended to include the genotypes of closely linked surrounding markers. If a cytogenetically negative nonretarded individual has marker alleles both proximal and distal to *FRAXA*, clearly demonstrating that the individual carries the fragile X chromosome (if the possibility of double crossover is disregarded), then that individual will contribute more information to a linkage analysis if he or she is coded as a carrier than if he or she is coded as unaffected in a defined penetrance class. Conversely, if it could be demonstrated that such an individual does not carry the fragile X chromosome (if the possibility of double crossover is disregarded), then that individual also will contribute more information to a linkage analysis if he or she could be coded as a definite noncarrier. Determination of such carrier status by DNA markers does not rely on map distances estimated in the present study by using VK23AC and VK14AC. The genetic map distances on which carrier status is based were previously established in an independent analysis of linkage data (Suthers et al. 1991*a*). For the present study, first the unaffected individuals of either sex were coded as if they were affected, when closely linked informative markers flanking the fragile X demonstrated that they were carriers. Second, a penetrance class of 100% was assigned to unaffected individuals of either sex in whom closely linked informative markers flanking the fragile X demonstrated that they were not carriers. This both removed the option, available under incomplete pene-

Table 1

Alleles, Allele Frequency, and Heterozygosity for VK23AC and VK14AC in Fragile X Pedigrees

Marker and Allele	No. of Chromosomes	Frequency	Heterozygosity
VK23AC:	65 (60)		.74 (.67)
+4		.00 (.02)	
+2		.08 (.03)	
0		.34 (.53)	
−2		.32 (.13)	
−4		.11 (.17)	
−6		.15 (.12)	
−12		.00 (.02)	
VK14AC:	44 (60)		.53 (.58)
+6		.00 (.02)	
+4		.00 (.05)	
+2		.02 (.13)	
0		.64 (.62)	
−2		.23 (.08)	
−4		.11 (.10)	

NOTE.—Data in parentheses refer to CEPH pedigrees.

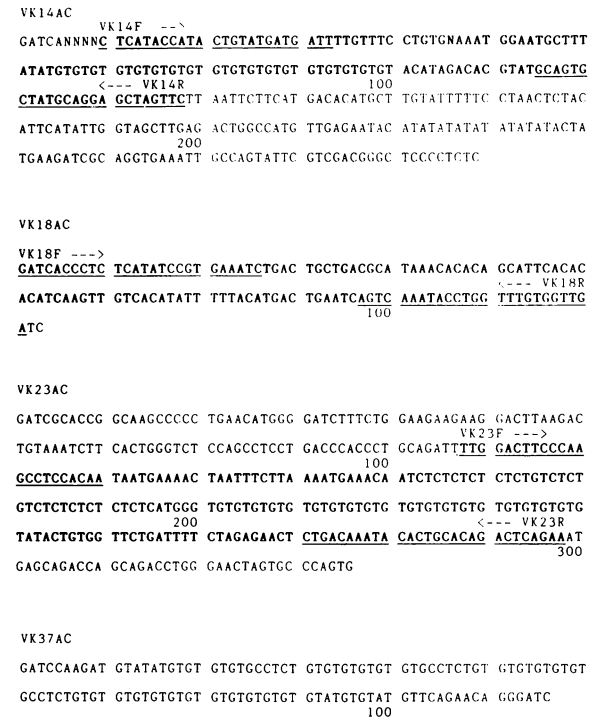

Figure 1 AC/GT repeat sequences determined from VK14, VK18, VK23, and VK37. Nucleotide sequences containing AC.GT repeats from each of the markers near Xq27.3 are shown. Sequences used to design primers for genotyping polymorphism of repeat length are underlined. PCR products are shown in boldface.

trance, that they might carry the fragile X and allowed full utilization of data from these individuals.

Twenty-seven and 14 of the 31 fragile X families were at least partially informative for VK23AC and VK14AC, respectively. One additional family was informative for the VK23B *Hin*dIII RFLP but not for VK23AC. The analysis of fragile X families was carried out by using the LINKAGE (version 5.04) package of computer programs and incorporated the data from the *Hin*dIII RFLP of VK23B from the earlier study of Suthers et al. (1991*a*). Confidence intervals were obtained by the lod – 1 method (Conneally et al. 1985). Multipoint analysis was based on 40 normal families from CEPH (Dausset et al. 1990). Analyses of the CEPH families were carried out by using the LINKAGE (version 4.9) package for use with CEPH three-generation families.

Results

Identification of (AC.GT) in Repeats at Xq27.3

Nineteen human genomic DNA clones (lambda VK7, 9–11, 14, 16–18, 21, 23–25, 29, 34, 37, 40, 41, 44, and 47) which map to the interval Xq26-qter (Hyland et al. 1989) were screened for the presence of AC-repeat microsatellite sequences, and nine of these were positive. The four positive clones VK14, VK18, VK23, and VK37, which physically map closest to *FRAXA* (Suthers et al. 1990) were sequenced to determine both the length of AC repeats and the composition of unique flanking sequences. The relevant sequences from each of these regions are shown in figure 1, as are the location and sequences of PCR primers designed from them to type the AC-repeat-length polymorphism. No primers could be designed from the VK37 sequence, because of the close proximity between the *Sau*3AI cloning sites and the AC repeat.

Characterization of AC-Repeat Microsatellite Polymorphism

Primers were specifically designed at distances sufficient to allow simultaneous typing of both VK23AC and VK14AC. Twenty unrelated individuals were typed for length variation of the AC repeats at the VK14, VK18, and VK23 loci. The VK18 repeat showed no polymorphism and was therefore not analyzed further. Both the VK14 and VK23 AC repeats revealed length variation inherited in a Mendelian fashion (fig. 2) in fragile X families and had heterozygosities greater than 50% (table 1). Key carriers from all available fragile X–linked mental retardation pedigrees were genotyped, and informative families were completed. Heterozygosities for unrelated individuals in the fragile X pedigrees were 53% VK14, and 74% VK23, compared with the 58% VK14 and 67% VK23 for unrelated individuals from unaffected (CEPH) pedigrees.

Two-Point Linkage Analysis

Genotypes of AC-repeat length at the *DXS297* and *DXS292* loci were used to calculate genetic distance from *FRAXA* (table 2). *DXS292* was found to have a peak lod score of 12.87 at a recombination fraction of .07 (confidence interval .03–.15), and *DXS297* had a peak lod score of 32.04 at a recombination fraction of .01 (confidence interval .00–.04).

The lod score for the *FRAXA:DXS297* comparison remains positive at a recombination fraction of zero, despite the observation of a recombinant between *FRAXA* and *DXS297* in a definite carrier female. This result presumably arises from the incorporation of mutation into the analysis: there is a small chance that the apparent "recombinant" is instead a second mutation in a family already segregating for the fragile

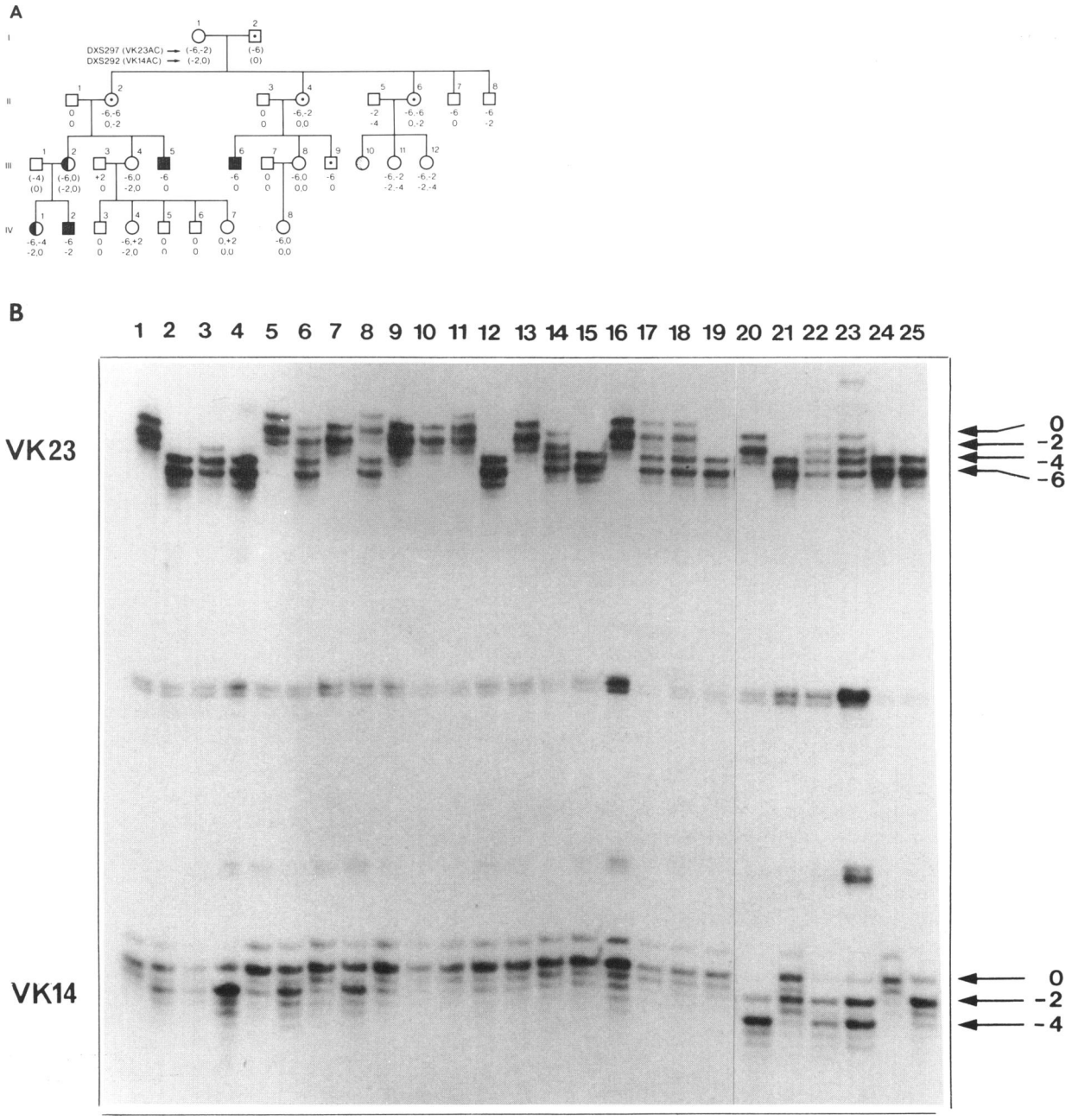

Figure 2 Genotype analysis of VK14AC and VK23AC repeats in a fragile X syndrome pedigree. *Right,* Fragile X syndrome pedigree with DXS297 (AC) and DXS292 (AC) genotypes. ⊡ and ⊙ = transmitter based on pedigree and/or DNA results; ◐ and ■ = affected individual with cytogenetically characterized fragile X; ◐ = affected individual without cytogenetic analysis; ( ) = genotype inferred. Carrier status was determined in conjunction with analysis of the distal flanking marker St14. *Below,* Autoradiograph of DXS297 (AC) and DXS292 (AC) genotypes. Individuals in the affected pedigree are as follows: lane 1, II-1; lane 2, II-2; lane 3, IV-1; lane 4, IV-2; lane 5, III-3; lane 6, III-4; lane 7, IV-3; lane 8, IV-4; lane 9, IV-5; lane 10, IV-6; lane 11, IV-7; lane 12, III-5; lane 13, II-3; lane 14, II-4; lane 15, III-6; lane 16, III-7; lane 17, III-8; lane 18, IV-8; lane 19, III-9; lane 20, II-5; lane 21, II-6; lane 22, III-11; lane 23, III-12; lane 24, II-7; and lane 25, II-8.

Table 2

Two-Point Lod Scores for Fragile X Kindreds

Linkage Comparison	Lod Score at Recombination Fraction of							Maximum Recombination Fraction	Maximum Lod Score
	.0	.01	.05	.1	.2	.3	.4		
FRAXA:DXS297.....	31.45	32.04	30.36	27.73	21.38	14.35	6.91	.01 (.00–.04)	32.04
FRAXA:DXS292.....	...	9.95	12.70	12.72	10.67	7.51	3.81	.07 (.03–.15)	12.87
DXS297:DXS292.....	...	7.82	11.57	12.06	10.48	7.52	3.85	.09 (.04–.17)	12.08

X. In contrast, the lod score for the *FRAXA:DXS292* comparison approaches minus infinity at a recombination fraction of zero, because the two recombinants involve unaffected individuals who have been diagnosed as carriers by closely linked flanking markers. Recombination events within the *DXS292–DXS297–FRAXA* linkage group were consistent with the established order derived from physical mapping. Only one recombination event was observed between *DXS297* and *FRAXA,* in a female unambiguously expressing the fragile X. Since recombination was observed between *DXS292* and *FRAXA* in the same individual, *DXS297* and *DXS292* are on the same side of *FRAXA*. In other pedigrees, a noncarrier female and a male transmitter (as determined by closely linked informative markers flanking *FRAXA*) demonstrated recombination between *DXS292* and *FRAXA* but not between *DXS297* and *FRAXA*. Hence, of the two markers, *DXS297* is confirmed as being closer to *FRAXA*. Two additional individuals, both carrier females, were recombinants between *DXS292* and *FRAXA*. They were uninformative for recombination between *DXS297* and *FRAXA*.

Figure 2 demonstrates Mendelian inheritance of both AC-repeat polymorphisms. Some of the difficulties in this large affected pedigree were resolved through the analysis of the AC-repeat markers, which were both partially informative. A recombinant between *FRAXA* and *DXS292* was detected between individuals II-2 and III-2. Male transmitter status for I-2 and III-9 was based on informative flanking markers VK23AC and St14, and III-8 and IV-8 are likely carriers.

The degree of linkage disequilibrium between VK23AC, the VK23B *Hin*dIII RFLP, and the VK23B *Xmn*I RFLP was not quantified. Only one family not informative for VK23AC was found to be informative for the *Hin*dIII RFLP, suggesting that the RFLP does not significantly increase informativeness in families already typed for VK23AC. Very few families have been typed for the *Xmn*I RFLP, so the extent to which this RFLP might increase informativeness in families already typed for VK23AC is not known.

Multipoint Linkage Analysis

The result of multipoint analysis of VK23AC and VK14AC in CEPH families is given in figure 3. This shows the two-point lod scores and the multipoint recombination frequencies. The new marker *DXS292* (VK14AC) is placed 3% proximal to *DXS297* in the interval between *DXS297* and *DXS98*. This confirms the established physical order and is consistent with anecdotal evidence for the order derived above from the fragile X families.

Discussion

The mutation responsible for fragile X syndrome has not yet been identified and characterized at the molecular level. DNA diagnosis relies on analysis of flanking genetic markers (Suthers et al. 1991*a*). The utility of markers is governed by their distance from the mutation and by their information content. In the present work, we have characterized two additional markers, both highly polymorphic and closely linked to *FRAXA*. The AC-repeat microsatellites identified at the *DXS297* and *DXS292* loci were applied to the 36 fragile X families available to our laboratory.

Multipoint recombination frequencies between markers near *FRAXA* were previously estimated from CEPH pedigrees (Suthers et al. 1991*b*); they are *DXS98* (12.3%) *DXS369* (0%) *DXS297* (5.7%) *DXS296* (0%) *IDS* (1%) *DXS304* (12%) *DXS52* (recombination frequency is shown in parentheses between the respective markers). *FRAXA* was located 3.7% distal to *DXS297* and 2% proximal to *DXS296* (Suthers et al. 1991*a*). In the present study the addition of *DXS297* (VK23AC) and *DXS292* (VK14AC) data from the CEPH families had little effect on the multipoint recombination frequencies between the

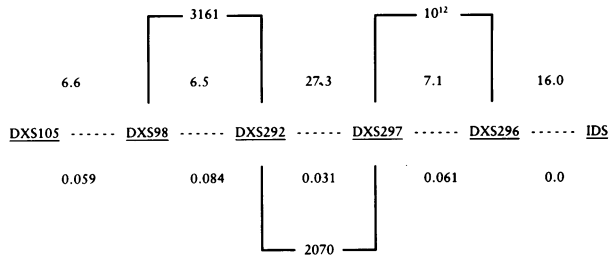

Figure 3 Partial multipoint map of region near *FRAXA*, based on CEPH analysis. *FRAXA* is located between *DXS297* and *DXS296*. The order of loci was derived from physical mapping and confirmed by the odds, as shown, against inverting adjacent loci. Numbers above the marker intervals are two-point lod scores, and numbers below the marker intervals are multipoint recombination fractions.

markers (fig. 2). The suggested recombination frequency for diagnosis using VK23AC therefore remains at 4%; however, when considered in light of subsequent data from the fragile X families themselves, as determined in the present study, this may be an overestimate (table 2). The previous report of 4% was based only on the VK23B *Hin*dIII RFLP typed in a subset of the Adelaide families. The data presented in table 2 were based on all the Adelaide families and used the more informative VK23AC marker. When data are available from additional markers currently being characterized near the fragile X, the *FRAXA* locus may be repositioned onto the background map, as previously described by Suthers et al. (1991*a*).

DXS292 has now been placed on the multipoint background map by using CEPH pedigrees. Since it is 3 cM proximal to *DXS297* on the background map, the suggested recombination frequency for diagnosis of fragile X syndrome is 7%. This corresponds exactly with the most likely recombination frequency derived from fragile X families by using two-point analysis (table 2). When it is used for diagnosis, *DXS292* should be used in conjunction with an informative marker distal to *FRAXA*.

The availability of numerous closely linked markers (Suthers et al. 1991*a*) can have considerable impact on the precision of linkage analysis. In the present study, sets of closely linked informative markers flanking *FRAXA* facilitated the identification with virtual certainty of male transmitters and carrier females. Although it has been shown that recombination values are insensitive to variation of the penetrance parameters used in linkage analysis (Oberlé et al. 1986), the exact coding of carrier status, wherever possible, had substantial impact on both the magnitude of lod scores and the associated lod – 1 confidence intervals.

Three mentally retarded individuals (not expressing the fragile X) within our set of fragile X families were shown (with a probability greater than 99%) not to have inherited the chromosomal segment containing *FRAXA*. This excluded the diagnosis of fragile X–linked mental retardation. The diagnosis of five nonretarded individuals expressing the "fragile X" in 1%–2% of cells was also clarified. None of these individuals had inherited the chromosomal segment containing the fragile X, implying that the detection of the common fragile site (Sutherland and Baker 1990) can lead to misdiagnosis of carrier status. In two cases this confirmed the conclusion tentatively made earlier when only loosely linked flanking markers were available (Mulley et al. 1988). The availability of numerous flanking markers can be invaluable for accurate diagnosis of some mentally retarded individuals and of individuals with low rates of expression of the "fragile X." These clarifications can have considerable impact on the determination of the potential carrier status of other family members in some sections of a pedigree.

Finally, the simultaneous detection of highly polymorphic VK23AC and VK14AC markers as described above now allows determination of at least one informative marker on the proximal side of *FRAXA*— within 3 d, in most families. Because these markers are highly polymorphic, they facilitate detection of both nonpaternity and sample error and often permit inference of missing parental marker genotypes when this cannot be achieved with the less informative diallelic RFLPs. The present experimental approach's failure to secure polymorphic markers from VK18 and VK37 necessitates either screening additional sequences distal to *FRAXA* as they become available or walking from the existing loci (e.g., with yeast artificial chromosomes), in order to identify polymorphic dinucleotide repeats. The characterization of these markers distal to *FRAXA* will enable a comprehensive approach to genetic linkage, an approach that will supersede present RFLP analysis.

Acknowledgments

R.I.R. wishes to thank Shelley Richards for support and encouragement during the course of these studies. This work was supported in part by grants from the Channel 7 Children's Research Foundation of South Australia, Incorporated, and from the National Health and Medical Research Council of Australia. Y.S. was supported by a fellowship from the World Health Organization.

References

Chamberlain JS, Gibbs RA, Ranier JE, Nguyen PN, Caskey CT (1988) Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. Nucleic Acids Res 16:11141–11156

Conneally PM, Edwards JH, Kidd KK, Lalouel J-M, Morton NE, Ott J, White R (1985) Report of the Committee on Methods of Linkage analysis and Reporting. Cytogenet Cell Genet 40:356–359

Daussett J, Cann H, Cohen D, Lathrop M, Lalouel J-M, White R (1990) Centre d'Etude du Polymorphisme Humain (CEPH): collaborative genetic mapping of the human genome. Genomics 6:575–577

Hofker MH, Skraastad MI, Bergen AAB, Wapenaar MC, Bakker E, Millington-Ward A, van Ommen GJB, Pearson PL (1986) The X chromosome shows less genetic variation at restriction sites than the autosomes. Am J Hum Genet 39:438–451

Hyland VJ, Fernandez KEW, Callen DF, MacKinnon RN, Baker E, Friend K, Sutherland GR (1989) Assignment of anonymous DNA probes to specific intervals of human chromosomes 16 and X. Hum Genet 83:61–66

Kogan SC, Doherty BS, Gitschier J (1987) An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. N Engl J Med 317:985–990

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401

Mulley JC, Turner G, Bain S, Sutherland GR (1988) Linkage between the fragile X and *F9, DXS52* (St14), *DXS98* (4D-8) and *DXS105* (cX55.7). Am J Med Genet 30:567–580

Oberlé I, Heilig R, Moisan JP, Kloepfer C, Mattei MG, Mattei JF, Boue J, et al (1986) Genetic analysis of the fragile X mental retardation syndrome with two flanking polymorphic DNA markers. Proc Natl Acad Sci USA 83: 1016–1020

Richards RI, Holman K, Shen Y, Kozman H, Harley H, Brook D, Shaw D. Human glandular kallikrein genes: genetic and physical mapping of the KLK1 locus using a highly polymorphic microsatellite PCR marker (submitted)

Sutherland GR, Baker E (1990) The common fragile site in band q27 of the human X chromosome is not coincident with the fragile X. Clin Genet 37:167–172

Sutherland GR, Hecht F (1985) Fragile sites on human chromosomes. Oxford University Press, Oxford

Sutherland GR, Mulley JC (1990) Diagnostic molecular genetics of the fragile X. Clin Genet 37:2–11

Suthers GK, Hyland VJ, Callen DF, Oberle I, Rocchi M, Thomas NS, Morris CP, et al (1990) Physical mapping of new DNA probes near the fragile X (*FRAXA*) by using a panel of cell lines. Am J Hum Genet 47:187–195

Suthers GK, Mulley JC, Voelckel MA, Dahl H, Väisänen ML, Steinbach P, Glass IA, et al (1991*a*) Genetic mapping of new DNA probes at Xq27 defines a strategy for DNA studies in the fragile X syndrome. Am J Hum Genet 48: 460–467

Suthers GK, Oberlé I, Nancarrow J, Mulley JC, Hyland VJ, Wilson PJ, McCure J, Morris CP, Hopwood JJ, Mandel JL, Sutherland GR (1991*b*) Genetic mapping of new RFLPs at Xq27-q28. Genomics 9:37–43

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396