Am. J. Hum. Genet. 66:819–829, 2000

# An Unstable Trinucleotide-Repeat Region on Chromosome 13 Implicated in Spinocerebellar Ataxia: A Common Expansion Locus

John B Vincent,[1,*] Maria L. Neves-Pereira,[1] Andrew D. Paterson,[1] Etsuko Yamamoto,[1] Sagar V. Parikh,[2] Fabio Macciardi,[1] Hugh M.D. Gurling,[3] Steve G. Potkin,[4] Carlos N. Pato,[5] Antonio Macedo,[6] Maria Kovacs,[7] Marilyn Davies,[8] Jeffrey A. Lieberman,[9] Herbert Y. Meltzer,[10] Arturas Petronis,[1] and James L. Kennedy[1]

[1]Neurogenetics Section and [2]Bipolar Clinic, Clarke Division, Centre for Addiction and Mental Health, University of Toronto, Toronto; [3]Molecular Psychiatry Laboratory, Windeyer Institute of Medical Sciences, Department of Psychiatry and Behavioural Sciences, University College London Medical School, London; [4]Department of Psychiatry, University of California at Irvine, Irvine; [5]Department of Psychiatry, State University of New York at Buffalo, Buffalo; [6]Department of Psychiatry, Faculty of Medicine, and Center for Neuroscience, University of Coimbra, Coimbra, Portugal; [7]University of Pittsburgh School of Medicine, Department of Psychiatry and Western Psychiatric Institute and Clinic, Pittsburgh; [8]Department of Psychiatry, Case Western Reserve University, Cleveland; [9]Department of Psychiatry, University of North Carolina, Chapel Hill; and [10]Department of Psychiatry, Vanderbilt University, Nashville

**Larger CAG/CTG trinucleotide-repeat tracts in individuals affected with schizophrenia (SCZ) and bipolar affective disorder (BPAD) in comparison with control individuals have previously been reported, implying a possible etiological role for trinucleotide repeats in these diseases. Two unstable CAG/CTG repeats, *SEF2-1B* and *ERDA1*, have recently been cloned, and studies indicate that the majority of individuals with large repeats as detected by repeat-expansion detection (RED) have large repeat alleles at these loci. These repeats do not show association of large alleles with either BPAD or SCZ. Using RED, we have identified a BPAD individual with a very large CAG/CTG repeat that is not due to expansion at *SEF2-1B* or *ERDA1*. From this individual's DNA, we have cloned a highly polymorphic trinucleotide repeat consisting of $(CTA)_n$ $(CTG)_n$, which is very long (~1,800 bp) in this patient. The repeat region localizes to chromosome 13q21, within 1.2 cM of fragile site FRA13C. Repeat alleles in our sample were unstable in 13 (5.6%) of 231 meioses. Large alleles (>100 repeats) were observed in 14 (1.25%) of 1,120 patients with psychosis, borderline personality disorder, or juvenile-onset depression and in 5 (.7%) of 710 healthy controls. Very large alleles were also detected for Centre d'Etude Polymorphisme Humaine (CEPH) reference family 1334. This triplet expansion has recently been reported to be the cause of spinocerebellar ataxia type 8 (SCA8); however, none of our large alleles above the disease threshold occurred in individuals either affected by SCA or with known family history of SCA. The high frequency of large alleles at this locus is inconsistent with the much rarer occurrence of SCA8. Thus, it seems unlikely that expansion alone causes SCA8; other genetic mechanisms may be necessary to explain SCA8 etiology.**

## Introduction

Trinucleotide-repeat expansion (TRE) is associated with a number of neurological disorders and, in most of these diseases, provides a molecular basis for the observation of genetic anticipation. This has led many researchers to look for evidence of TRE as a possible etiological cause for neuropsychiatric diseases such as bipolar affective disorder (BPAD) and schizophrenia (SCZ). Although anticipation has been reported for both BPAD and SCZ in many studies (McInnis and Margolis 1998), the pervasive presence of ascertainment biases and lack of an appropriate statistical test means that it would be imprudent to draw any firm conclusions from these findings (Paterson et al. 1998). Although early reports have suggested that DNA from BPAD and SCZ patients contains significantly longer stretches of CAG/CTG repeats, as determined by repeat-expansion detection (RED; Lindblad et al. 1995; Morris et al. 1995; O'Donovan et al. 1995, 1996), several studies have contradicted these findings (Petronis et al. 1996*a;* Vincent et al. 1996, 1998, 1999*b;* Laurent et al. 1998; Li et al. 1998; Zander et al. 1998; Parikh et al., 1999). It is interesting to note that the sizes of trinucleotide repeat reported to be associated with psychosis are, in general, larger than repeat expansions associated with Huntington disease or spinocerebellar ataxias (SCAs). More recently it has been

Received April 20, 1999; accepted for publication December 13, 1999; electronically published March 6, 2000.

Address for correspondence and reprints: Dr. John Vincent, Department of Genetics, Room 9102, Hospital for Sick Children, 555 University Avenue, Toronto, ON M5G 1X8, Canada. E-mail: jvincent@genet.sickkids.on.ca

* Present affiliation: Department of Genetics, Hospital for Sick Children, Toronto.

© 2000 by The American Society of Human Genetics. All rights reserved.
0002-9297/2000/6603-0007$02.00

**Table 1**

**Patient Sample Sources, Numbers, and Diagnostic Criteria**

| Collection Center | Diagnostic Group | *n* | Diagnostic Instrument and Criteria |
|---|---|---|---|
| Toronto | Schizophrenia | 103 | SCID/DSM-IIIR |
| Italy | Schizophrenia | 34 | SADS/RDC |
| Portugal | Schizophrenia | 85 | OPCRIT |
| Cleveland | Schizophrenia | 98 | SADS/RDC |
| Long Island | Schizophrenia | 161 | SADS/RDC |
| Irvine | Schizophrenia | 46 | DSM-IIIR |
| Toronto | Schizophrenia | 25 | SCID/DSM-IV |
| Toronto | Bipolar affective disorder | 340 | SCID/DSM-IV |
| London | Bipolar affective disorder | 96 | SADS/RDC |
| Cleveland | Borderline personality disorder | 21 | SCID/DSM-IV |
| Cleveland | Juvenile-onset psychosis | 55 | SCID/DSM-IV |
| Pittsburgh | Juvenile-onset depression | 60 | K-SADS/RDC |

demonstrated that large repeats at two specific loci, *SEF2-1B* at 18q21.1 (Breschel et al. 1997) and *ERDA1* at 17q21.3 (Nakamoto et al. 1997; Ikeuchi et al. 1998), are responsible for the majority of the large-repeat tracts detected by RED, and the distribution of large repeats at these two loci is similar in BPAD, SCZ, and unaffected populations (Lindblad et al. 1998; Sidransky et al. 1998; Vincent et al. 1999*b*). Although there is no clear evidence that expansion at either locus may be pathogenic, because expansions in either moderate or large range at *SEF2-1B* do not segregate in SCZ or BPAD families (Breschel et al. 1997; Sirugo et al. 1997), to date, expansions >100 repeats at *ERDA1* have been identified in a single family with childhood-onset depression (Vincent et al. 1999*a*). These cases represent only a small fraction of families examined. To eliminate any possible role of other large TREs in BPAD and SCZ, it is necessary to identify and analyze the remaining large repeats that occur in our BPAD and SCZ sample.

We identified a BPAD individual, A7, with a large RED product but no large repeats at *SEF2-1B* or *ERDA1* (Parikh et al. 1999). RED evidence from this individual suggested the presence of a large stretch of CAG/CTG repeats >690 bp. In the present study, we describe the cloning and characterization of this large trinucleotide repeat, the screening for large repeats at this locus in psychosis and control populations, and association and linkage disequilibrium studies for SCZ and BPAD at this locus. Since this repeat was cloned, it has emerged that expansion at this repeat region has recently been reported to be the cause of spinocerebellar ataxia type 8 (SCA8 [MIM 603680]; Koob et al. 1999). Data presented in this article, however, suggest that expansion at this locus is common in the background population.

## Subjects and Methods

### *Patient and Control Sample Selection*

Description of patient source and diagnostic methods are given in table 1. All patients were screened for absence of major medical and neurological disorders. Of the 1,120 patients screened for large-repeat alleles, 82% were Caucasian, 9% Asian, 2% black, and 7% either of mixed ethnicity or with no information available. Fifty-three percent were female, 43% male, and for 4% information was not available. Mean age was 36.8 years (± 11.2 years, SD). The control DNA samples were obtained from staff members and students at the Clarke Institute (Toronto), Case Western Reserve University (Cleveland), University of Lexington (Kentucky), University of Coimbra (Portugal), and members of the public responding to advertisements, the majority of whom had been assessed for absence of psychiatric illness. Of the 710 control individuals analyzed, 72% were Caucasian, 6% Asian, 15% black, and 7% either of mixed ethnicity or for whom information on ethnicity was unavailable. Forty-six percent were female, 34% male, and 20% sex unknown, and the mean age was 30.2 years (±11.3 years). Ninety-three proband-mother-father trios from the Toronto BPAD sample and 54 trios from the Toronto and Italy SCZ samples were available for genotyping for transmission disequilibrium analysis. Local ethical committee approval was obtained, blood was drawn after written informed consent was obtained from each subject, and DNA was extracted according to standard procedures. DNA from lymphoblastoid cell lines for CEPH pedigrees 884, 1331, 1333, 1340, 1345, 13291, 13292, 13293, and 13294 were purchased from BIOS Corp. DNA samples for the Old Order Amish pedigree 884 and for CEPH pedigree 1334 were pur-

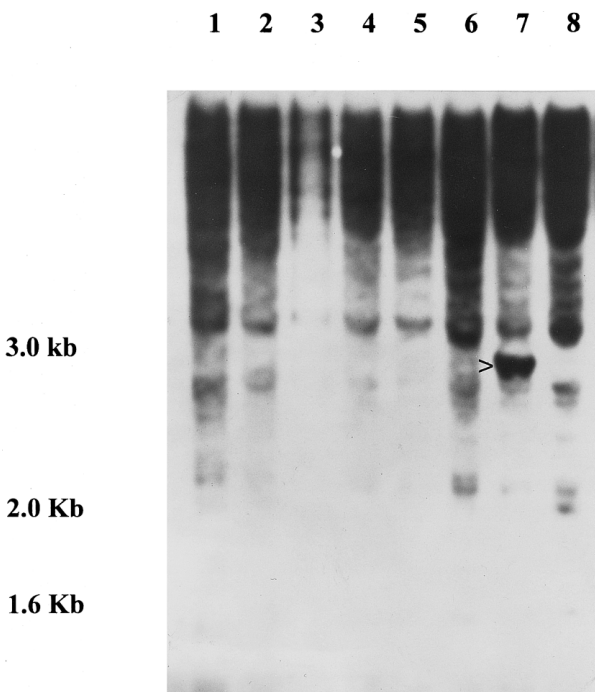


**Figure 1** Southern blot of *Eco*RI-digested genomic DNA hybridized with a large (1.5–3-Kb), nonspecific CAG/CTG repeat probe. Lane 7 shows a strong band at 2.8–3 Kb (*arrow*) for individual A7. Lanes 1–6 and 8 contain DNA from other BPAD and SCZ individuals.

chased from the Coriell Institute, and the most recent diagnostic information was made available by E. I. Ginns and J. Egeland.

### *Genomic Library Construction and Screening*

The method used for cloning was similar to the DIRECT strategy (Sanpei et al. 1996), except that a very-large-repeat probe (1.5–3 Kb) generated by asymmetric PCR (Petronis et al. 1996*b*) was used for screening at very high stringency. Genomic DNA from BPAD individual A7 (previously identified as giving large RED products for the triplet CAG/CTG but without large alleles at either of two commonly expanded repeats at *SEF2-1B* or *ERDA1*; Vincent et al. 1999*b*) was digested with *Eco*RI and cloned into λgt10 (Stratagene). The ligated vector was packaged by means of MaxPlax packaging extracts (Epicentre Technologies) and plated with NM514 *Escherichia coli* cells to give ~$1.4 \times 10^6$ pfu, with average insert size 3.4 Kb. The library was screened with a 1.5–3-Kb CAG/CTG probe generated by template-independent PCR using the complementary primers $[CAG]_7$ and $[CTG]_7$, as described elsewhere (Petronis et al. 1996*b*). Southern hybridization of *Eco*RI-digested genomic DNA with use of the same probe revealed a strong band for A7 at 2.8–3 Kb (fig. 1). After secondary screening, 10 positive plaques were selected and DNA prepared according to standard procedures. RED analysis (Schalling et al. 1993; Vincent et al. 1996) was performed on the clones, confirming the presence of a large CAG/CTG repeat in clone λ7a (fig. 2). A second library was constructed in the same way from genomic DNA from another individual with large RED products (without large alleles at *SEF2-1B* or *ERDA1*). Clone λ90-4a1 was identified as an unexpanded version of λ7a.

### *Sequence Analysis*

The 1.45-Kb *Eco*RI insert from clone λ7a was subcloned into M13 and sequenced by use of a Li-Cor Long Reader 4200. Sequence analysis of clone λ90-4a1 and of alleles at this locus was performed on an ABI Prism 310 (Applied Biosystems), with λgt10 forward and re-

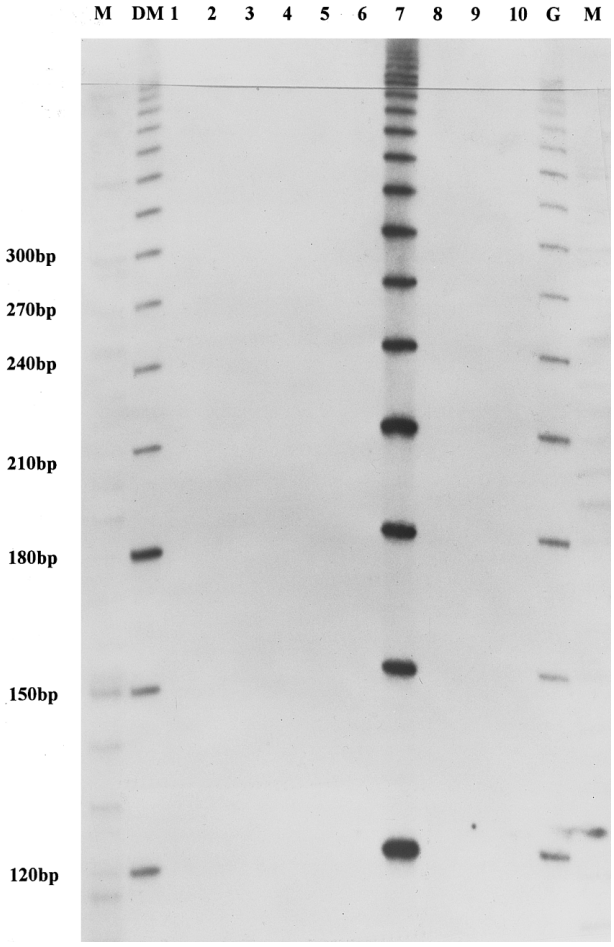


**Figure 2** RED on primary λgt10 clones (*a–j*) picked from genomic library (screened with CAG/CTG probe) generated from individual A7 (RED from genomic DNA: lane number). Myotonic dystrophy positive control is shown in lane DM. Size marker (Sequamark size ladder; Research Genetics) is shown in lane M.

verse primers and specific primers 7aCAG, 7aCTG, and 7aEXT (GenBank accession AF087653, 5′ to 3′ nucleotides 314–335, 846–822, and 1065–1084, respectively).

*PCR genotyping.*—PCR genotyping was performed with the primers 7aCAG and 7aCTG; 95°C for 3 min followed by 30 cycles of 95°C for 45 s, 52°C for 45 s, 72°C for 45 s under standard conditions, followed by 6% polyacrylamide gel electrophoresis, blotting onto Hybond-N+ (Amersham), hybridization using $^{32}$P-end-labeled $[CTG]_{10}$ oligonucleotide followed by autoradiography.

*Southern hybridization analysis.*—To screen efficiently for the presence or absence of large expansions that may not amplify sufficiently for PCR detection, all DNA samples that were apparent homozygotes and all those that failed to amplify were checked by Southern hybridization. This was performed according to standard procedures, with *Eco*RI-restricted genomic DNA (5 μg), and the 0.95-Kb *Eco*RI insert from clone λ90-4a1 as the hybridization probe. This insert contains a much shorter repeat stretch than clone λ7a, thus permitting much less cross-hybridization with other repeat loci.

*Allele-specific oligonucleotide (ASO) analysis.*—Filters from the genotype analysis were stripped (0.5% SDS, 100°C) and reprobed with either ASO1 (5′-TACTACTGCTGC-3′) or ASO2 (5′-TACTGCTACTGC-3′). ASO 1 and ASO2 were 5′ end-labeled with polynucleotide kinase and $^{32}$P-γATP. Hybridization was performed with use of Amasino buffer at 27°C, and filters were washed at 37.5°C in 1 × SSC, 0.1% SDS.

### Statistical Analysis

Allele distributions for patient and control samples were compared by means of a rank-sum test (Mann-Whitney). Preferential transmission of alleles in the trios was analyzed by means of an extended transmission/disequilibrium test (ETDT; Sham and Curtis 1995). Linkage analysis for the Old Order Amish pedigree was performed by MLINK from the FASTLINK suite (Terwilliger and Ott 1994). Test for association was performed by means of $\chi^2$ analysis (SPSS 7.0).

### Chromosomal Localization

The trinucleotide repeat was localized to chromosome 13 by PCR screening of the NIGMS somatic cell hybrid panel 2 (rodent/human hybrid). Subchromosomal localization was performed by PCR screening the GeneBridge 4 radiation hybrid panel (Research Genetics) and anchored CEPH YACs.

### Northern Blot and cDNA Screening

Human multiple-tissue northern blots and human adult and fetal brain cDNA libraries HL3002b and HL3003a (Clontech) were screened with probes flanking the repeat region according to the manufacturer's instructions.

## Results

### Cloning and Characterization of λ7a

Southern hybridization, followed by stringent washing, of *Eco*RI-digested genomic DNA with a large, nonspecific CAG/CTG repeat probe (1.5–3 Kb) revealed a strong signal band at 2.8–3 Kb (fig. 1). *Eco*RI-digested genomic DNA from this patient was used to generate a λgt10/genomic library, which was then screened for repeat containing clones by use of the large CAG/CTG repeat probe. Of 10 clones identified, one contained a large CAG/CTG repeat (fig. 2). The RED analysis displayed in figure 2 shows ligation products upward of $[CAG]_{200}$; however, this analysis used a large excess of the λ7a template DNA. Titration of the template DNA was performed, and, at 100 pg, the ligation product size is closer to $[CAG]_{90}$. This clone, λ7a, contained a 1.45-Kb *Eco*RI fragment. The 1.45-Kb fragment was subcloned into M13 and sequenced and showed a stretch of 18 CTA repeats followed directly by 85 CTG repeats (nt 451–759; GenBank AF087653). A second clone, λ90-4a1, was identified from a second genomic library from another individual, which contains the same flanking sequence, and was used as confirmation of the sequence. The single insert from this clone was only 959 bp long and contained only 25 CTA/CTG repeats. The ~300-bp discrepancy in size arises from an apparent *Eco*RI polymorphism (AF087653, nt 1193), which was born out by genomic Southern hybridization for A7, which showed a 2-Kb size difference between normal and expanded alleles for *Eco*RI and only a 1.7-Kb size difference for *Pst*I and *Hin*dIII digests (fig. 3*a*, 3*b*). The Southern hybridization evidence also confirmed that the repeat size is much larger in the genomic DNA (~600 CTA/CTG repeats) than in λ7a (103 repeats: $CTA_{18}CTG_{85}$ repeats), suggesting that contraction of the repeat occurred during the cloning procedure. No size mosaicism was observed for the large repeat allele in lymphocyte DNA. In normal alleles, the CTA repeat has either eight or nine copies, whereas the CTG repeat varies from 9 to 25 copies. PCR genotyping was performed for 1,400 Caucasian, 141 Asian, and 125 black unrelated individuals from the combined patient and control sample. Distribution of alleles (scored as the sum of CTA and CTG repeats) is shown in figure 4*a* and shows interethnic differences. Patient A7, as well as a diagnosis of BPAD I, also suffers from familial tremor, asthma, eczema, and thyroiditis. The sole sibling of A7, who also has tremor but is unaffected by BPAD, does not possess an expanded allele. Clone λ90-4a1 has several single-base-pair dis-

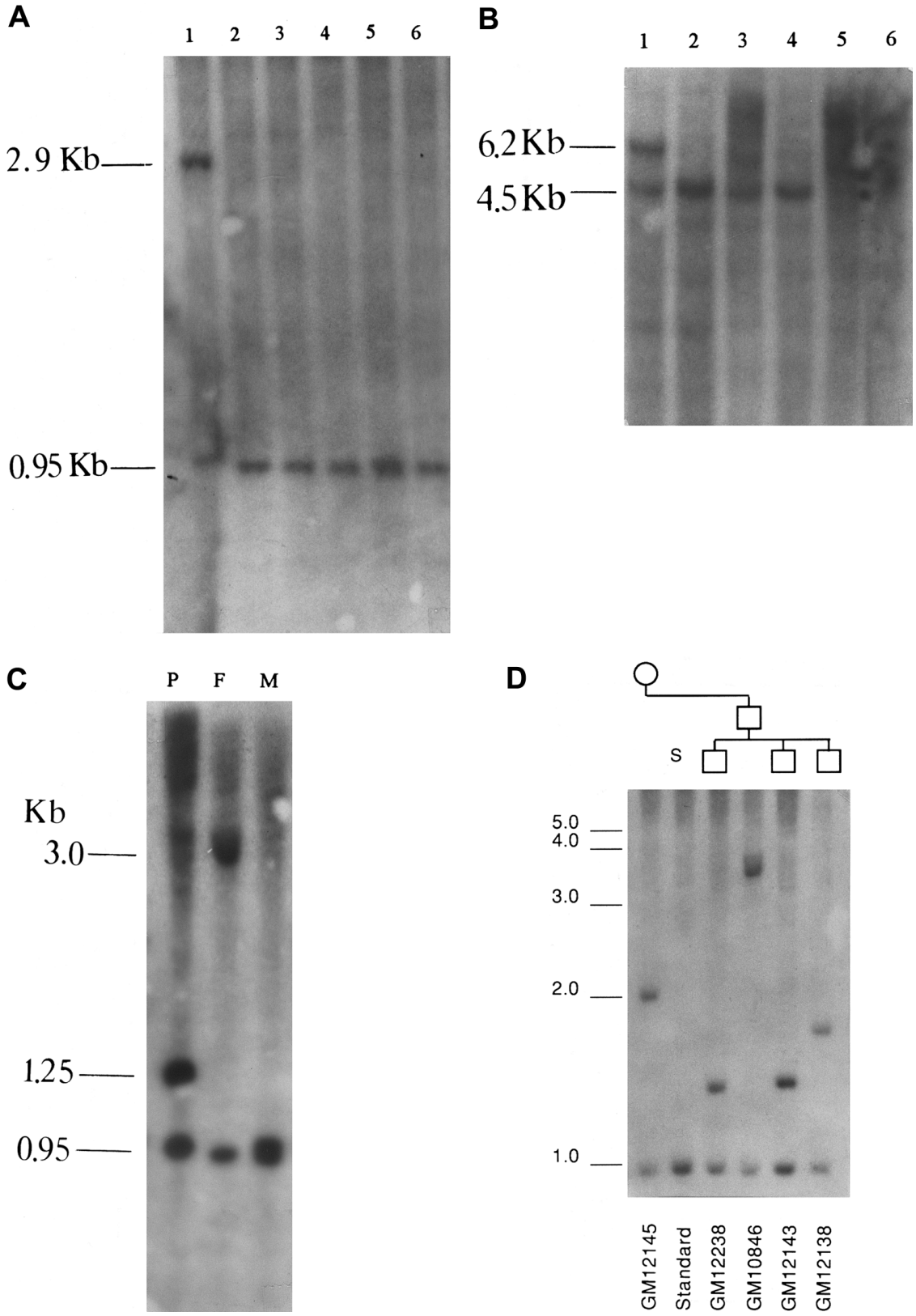


**Figure 3** Southern hybridization analysis, which was performed with use of (*A*) *Eco*RI and (*B*) *Hin*dIII on BPAD proband A7 (lane 1), sibling of A7 (lane 2), and unaffected individuals (lanes 3–6) and *Eco*RI (*C*) on SCZ trio, proband (lane P), father (lane F), and mother (lane M) and (*D*) on CEPH pedigree 1334. Family member CEPH numbers shown (*bottom*). Lane S represents standard control DNA.

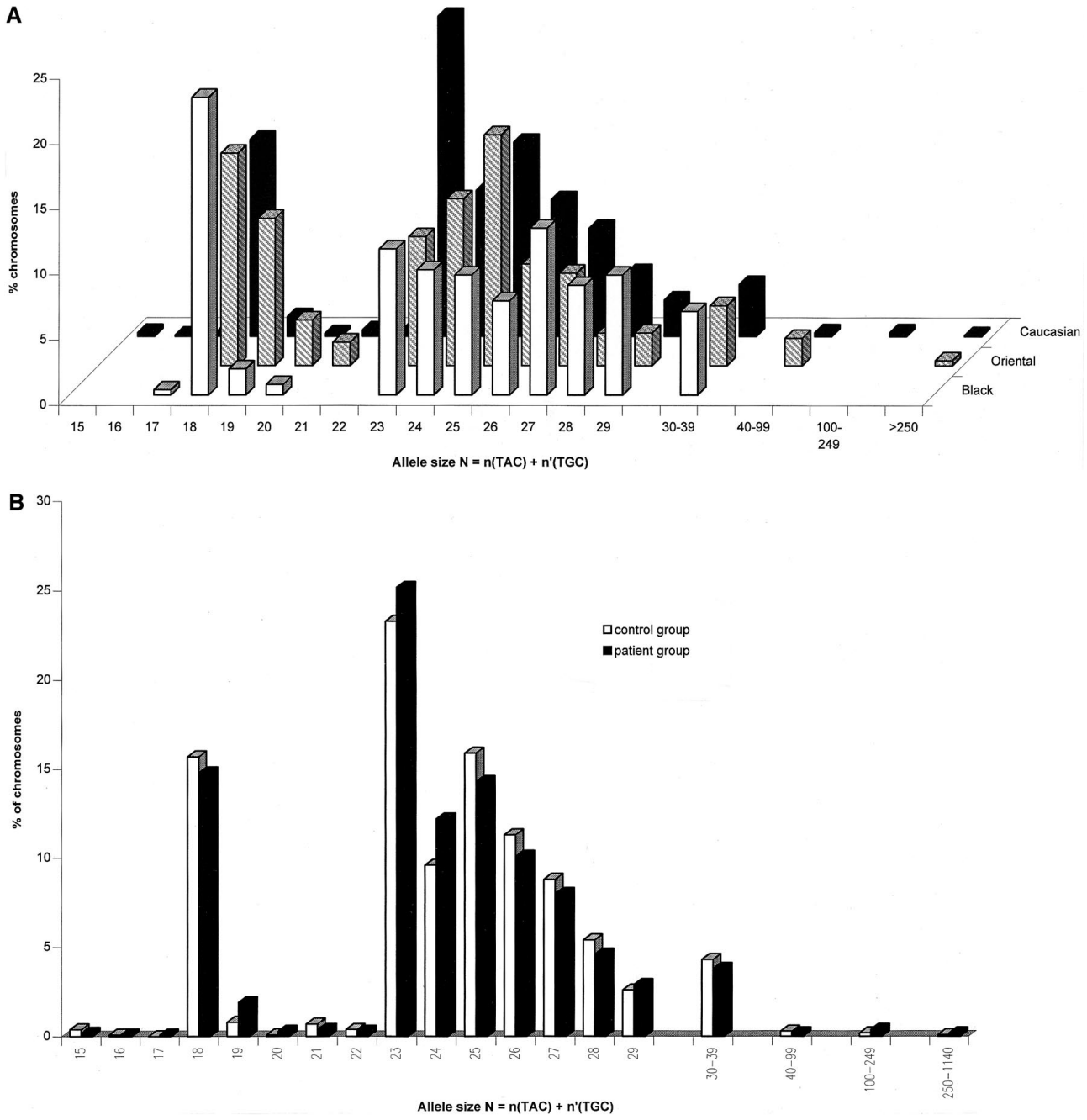


**Figure 4** *A,* Distribution of alleles for different ethnic groups (1,400 Caucasian, 141 Asian, and 125 black unrelated individuals from the combined patient-control sample). *B, SCA8* allele distribution for 497 unaffected Caucasian control individuals and 901 Caucasian psychosis and depression individuals (SCZ [*n* = 390], schizoaffective disorder [*n* = 15], BPADI [*n* = 357], BPADII [*n* = 56], borderline personality disorder [*n* = 14], juvenile-onset depression [*n* = 40], and major psychosis [*n* = 29]).

crepancies from sequence AF087653: a C instead of T at nt 1123 and a C instead of T at nt 1193. The sequence for *SCA8,* AF126748 (Koob et al. 1999) , which appears to be same repeat region, according to sequence, localization, and allelic distribution, also has Cs instead of Ts at these positions and also lacks a T at nt 398 of AF087653. Our repeat region is referred to as *SCA8,* for consistency.

### *CEPH 1334*

Control CEPH pedigree 1334 was screened for the repeat by use of PCR. The paternal grandmother, father,

and three sons appeared to have only a single allele, although the father and sons were clearly obligate heterozygotes. Very large alleles for these individuals were shown by Southern hybridization (fig. *3d*). Repeat sizes were determined by means of semi-log calibration curves for the 1-Kb ladder (Gibco BRL). The paternal grandmother (GM12145, aged 70 years) has ~370 repeats, which expands to ~900 repeats in her son (GM10846; aged 43 years) and then contracts to ~160, ~180, and ~290 repeats in his sons (GM12238, GM12143, GM12138, aged 9, 8, and 17 years, respectively). PCR amplification and sequencing showed GM12143 to have 9 CTA uninterrupted repeats followed by 159 uninterrupted CTG repeats. Because the CEPH pedigree DNAs are extracted from lymphoblastoid cell lines and it is not known how stable the repeats are over many passages, the relationship to repeat length in lymphocyte DNA is unclear.

### *Localization of the Repeat Region and Linkage Analysis for BPAD in the Old Order Amish*

The repeat region was localized to within 1.92 cR of STS marker WI-2964. A tiling path of YAC clones around WI-2964 was screened for the repeat region by PCR. CEPH YACs 810g9 and 744f11 were positive for the repeat. The RPCI-11 BAC library was screened by hybridization with the 7aCAG/7aCTG PCR product and was positive for clone H_NH0121J06. This region maps to 13q21.2-21.31. According to the GB4 map (GeneMap '98), this maps between anchored markers *D13S275* and *D13S152,* 52.7–56.6 cM or (211.38–213.71 cR) from the p telomere. A number of recent studies have shown positive results for parametric and nonparametric linkage analyses for BPAD and SCZ on 13q (Barden and Morissette, 1998; Blouin et al. 1998; maximum LOD score [MLS] 4.18 at 13q32, ~30 cM distal to *SCA8*). The serotonin receptor *HTR2A,* which is a strong candidate gene and for which association to BPAD and SCZ has been reported (Gutierrez et al. 1995; Williams et al. 1996), maps to 13q14, 5.8 cM and 67 cR proximal to the repeat. Linkage to BPAD in the pedigrees from the Old Order Amish ($Z_{max}$ dominant = 1.4 at *D13S1*, ~26 cM proximal to *SCA8;* Ginns et al. 1996) and in the National Institute of Mental Health Genetics Bipolar Initiative pedigrees ($P$ = .02 at *D13S793,* ~28 cM distal to *SCA8;* Stine et al. 1997) has been reported; however, we found no evidence for linkage of *SCA8* to either BP I (dominant: MLS = −0.18; recessive: MLS = −0.08) or BP I and II (dominant: MLS = −0.15; recessive: MLS = −0.06) in Amish pedigree 884.

### *Expansion Screening in Psychosis and Control Populations*

We screened 1,120 DNAs from unrelated patients diagnosed with psychosis (SCZ spectrum or bipolar disorder), juvenile-onset depression, or borderline personality disorder (table 1) and 710 unrelated controls unaffected with psychiatric illness for expansion at *SCA8*. PCR genotyping was used initially and, for confirmation of expanded alleles, apparent homozygotes or cases of failed amplification, Southern hybridization analysis was used. Six apparent homozygotes from the patient population, and five from the controls, could not be excluded for expansion, because of poor restriction digestion of DNA. Seven patients and eight controls that failed to PCR-amplify were excluded for expansion. Fourteen patients (1.25%) were identified as having large alleles (⩾100 repeats [age at interview, in years]: 100 [49], 103 [24], 106 [46], 107 [35], 116 [37], 130 [35], 130 [age not available], 180 [33], 257 [34], 550 [45], 600 [38], 600 [32], 1,140 [31], and 1,300 [30] repeats) and 9 (.8%) with intermediate-sized alleles (⩾45 repeats: 46 [25], 49 [38], 50 [25], 50 [28] 51 [29], 53 [28], 57 [43], 82 [39], and 83 [48] repeats). Five controls (0.7%) were shown to have large alleles (⩾100 repeats: 103 [21], 117 [33], 230 [21], 550 [22], and 970 [age not available] repeats) and two (0.3%) with intermediate alleles (⩾45 repeats: 50 [45] and 65 [age not available] repeats). Analysis of the various subgroups according to diagnosis and ethnic group shows the highest clustering of large alleles in SCZ Caucasians (⩾100 repeats: 103, 106, 107, 116, 257, 550, 600, and 1,140 repeats; 8 [2.1%] of 390). $\chi^2$ comparison of frequency of large alleles in control and affected groups did not reach significant levels.

### *Association and Transmission/Disequilibrium Analysis for SCZ and BPAD at* SCA8

Patients and controls for each major ethnic group were analyzed separately, because interethnic difference in allele distribution is evident (fig. 4*a*). There was no significant difference in distribution of alleles for 901 unrelated Caucasian individuals with psychosis (including SCZ [$n$ = 390], schizoaffective disorder [$n$ = 15], BPAD I [$n$ = 357], BPAD II [$n$ = 56], borderline personality disorder [$n$ = 14], juvenile onset depression [$n$ = 40], and juvenile-onset major psychosis [$n$ = 29]) and 497 unrelated Caucasian control individuals (fig. 4*b;* Mann-Whitney rank sum test: 2-tailed $P$ = .24). Subdivision of the patient group according to diagnosis revealed no significant differences in distribution in comparison with the control group. In a smaller but more closely matched subgroup, 100 BPAD individuals and 100 control individuals matched pairwise for age, sex, and ethnicity, the difference in distribution of alleles nearly reaches

significance (Mann-Whitney rank sum test: 2-tailed $P = .06$). We tested 93 BPAD trios and 54 SCZ trios for evidence of transmission disequilibrium at *SCA8* (table 2). We observed no significant preferential transmission of alleles for either BPAD ($\chi^2 = 18$, 19 df, $P = .49$) or SCZ ($\chi^2 = 11$, 11 df, $P = .46$) using an extended transmission/disequilibrium test (Sham and Curtis 1995); however, much larger numbers would be required to exclude transmission disequilibrium for the rarer alleles.

The *SCA8* repeat is highly polymorphic (observed heterozygosity .86, in 497 Caucasian control individuals). Fourteen intergenerational instabilities were observed (14/231), including CEPH pedigree 1334, for *SCA8:* 7 from maternal transmissions (+530, +4, +3, +2, +2, +1, −2; mean change +77 repeat units), 5 from paternal transmissions (−1, −596, −610, −720, −740; mean change −533 repeat units), and 2 where parental origin of the unstable allele was unclear. The smallest allele for which intergenerational instability was observed was 24 repeats (maternal transmission). In one trio, an unaffected father was identified with a stretch of >700 repeats, which was transmitted to a son diagnosed with SCZ with a decrease in repeat number to ~115 repeats (fig. 3*c*). Another trio was identified in which a daughter affected with BPAD received an allele with 82 repeats from the mother (unaffected) with 81 repeats.

### *ASO Analysis*

All PCR-amplified alleles tested (1262 control alleles, 1632 patient alleles) hybridized with the ASO1 oligonucleotide (AF087653 variant). No positive hybridization was observed for the ASO2 oligonucleotide (AF126748 variant). Alleles that were too large to PCR amplify could not be checked by this approach.

### *Northern Blot Analysis and cDNA Screening*

No RNA bands or cDNA clones corresponding to the *SCA8* repeat region were identified from a wide range of tissues, including heart, brain, lung, liver, pancreas, kidney, and skeletal muscle.

## Discussion

We have identified, cloned, and characterized an unstable trinucleotide repeat that, along with *SEF2-1B* and *ERDA1*, is responsible for the major proportion of the large RED products that have been observed in our (and other) studies of CAG/CTG repeats in BPAD and SCZ populations. This repeat locus was cloned independently by Koob et al. (1999) and named *SCA8*. Large alleles at *SCA8* are, however, relatively infrequent, at ~1% in comparison with ~5% and 15% for *SEF2-1B* and *ERDA1,* respectively. Sequence analysis of the 10 clones isolated from the genomic library from individual A7 revealed a high enrichment for large CAG/CTG repeats and included two clones for *SEF2-1B* (Breschel et al. 1997) and two clones for *CAGH39* (Margolis et al. 1997) as well as the new unstable repeat *SCA8*. It is clear that this method of cloning represents a useful approach for identification of new CAG/CTG repeats from the genome.

Large alleles at *SEF2-1B* have been demonstrated in two of the three CEPH pedigrees (1420 and 1344; Breschel et al. 1997) that were reported to have expansion at the *RED1* locus, which was identified and mapped to chromosome 18 by linkage analysis (Schalling et al. 1993). The third CEPH pedigree thought to have CAG/CTG expansion at the *RED1* locus, 1334, does not have large alleles at *SEF2-1B*, and, in fact, we have observed

**Table 2**

**Extended Transmission/Disequilibrium Test (ETDT; Sham and Curtis, 1995) for BPAD and SCZ Trios at *SCA8***

| Group and Value ($n$) | *SCA8* alleles: $N = n$[TAC] + $n'$[TGC] | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 15 | 16 | 18 | 19 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30–37 | 81 |
| BPAD trios (93) | | | | | | | | | | | | | | | |
| Transmitted | 0 | 1 | 24 | 1 | 2 | 0 | 34 | 11 | 20 | 16 | 12 | 12 | 5 | 5 | 1 |
| Untransmitted | 1 | 1 | 34 | 0 | 1 | 1 | 29 | 16 | 25 | 14 | 7 | 6 | 4 | 5 | 0 |
| $\chi^2$ | | | 1.72 | | | | .40 | .93 | .56 | .13 | 1.32 | 2.0 | | | |
| $P$ value[a] | | | .19 | | | | .53 | .34 | .46 | .72 | .25 | .16 | | | |

| | *SCA8* alleles: $N = n$[TAC] + $n'$[TGC] | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 18 | 19 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 70 |
| SCZ trios (54): | | | | | | | | | | | | |
| Transmitted | 18 | 1 | 1 | 23 | 10 | 7 | 5 | 7 | 5 | 4 | 3 | 1 |
| Untransmitted | 9 | 4 | 1 | 21 | 9 | 11 | 11 | 6 | 6 | 6 | 1 | 0 |
| $\chi^2$ | 3.0 | | | .09 | .05 | .89 | 2.25 | .08 | .09 | .4 | | |
| $P$ value[a] | .08 | | | .76 | .82 | .35 | .13 | .78 | .76 | .53 | | |

Note.—Statistics are not shown for alleles for which fewer than 10 observations were made.

[a] $P$ values are not corrected for multiple testing.

very large alleles at *SCA8* in this pedigree for three of four children tested, the father, and the paternal grandmother. Thus, it appears that the original *RED1* locus consists of two loci, one on chromosome 18 and the other on chromosome 13.

The *SCA8* locus appears to be frequently unstable (6%; 14/231 transmissions), even for relatively small alleles. The majority of increases in repeat size occur during transmission of maternal alleles, and contractions occur predominantly in male transmissions. In CEPH pedigree 1334, a maternal transmission results in an increase from ~370 to ~900 repeats, and three paternal unstable transmissions result in contractions from ~900 to ~160, 180, and 290 repeats. One unstable paternal transmission observed in an SCZ trio resulted in a contraction from ~700 to ~115 repeats. It is difficult to gauge the respective contributions of the CTG repeats and CTA repeats in the expanded alleles, although both are enlarged in clone λ7a ($[CTA]_{18}[CTG]_{85}$). We assume that the CTG repeat is most likely the more dynamic of the two repeats, judging by the available evidence (the high degree of heterogeneity of the CTG repeat size compared with that for CTA at *SCA8* and the relative lengths of the two repeats in clone λ7a). The CTA repeat is enlarged to 18 copies in the clone λ7a, but it is unclear whether this is expanded further in the large allele of proband A7.

The trinucleotide-repeat locus has now been ascribed to SCA8 (Koob et al. 1999). Although no homologies were found for the sequence flanking the repeat, and our northern blot analysis and cDNA library screening did not show any evidence of transcripts, Koob et al. (1999) have identified a transcript from cerebellar polyA RNA that contains the CTA;CTG repeat in the 3′ UTR. The inheritance pattern of disease in SCA8 appears complex. Penetrance of expanded alleles appears to be dependent on size of repeat and, probably, age, and pathogenic alleles appear to be mainly of maternal origin, possibly because of parental origin effect, whereby the paternal transmissions tend toward contraction and maternal transmissions tend toward expansion. Despite the differential parental effect on repeat instability leading to an apparent maternal penetrance bias in the large kindred, imprinting has been ruled out, because four cases in other families are reported in which SCA8 is transmitted paternally and biallelic expression of the repeat is also demonstrated in the cerebellum (Koob et al. 1999). It is, however, worth noting that the *HTR2A* gene, which is believed to exhibit polymorphic imprinting (Bunzel et al. 1998), is close to *SCA8* (as close as 3.4 cM in male meioses; Genetic Location Database). The authors conclude that the bias in maternal versus paternal transmissions of disease alleles is due to the differential instability for maternal and paternal meioses, which we also observed among our trio samples and CEPH pedigrees. The repeat sizes in CEPH pedigree 1334 (fig. *3d*) neatly demonstrate the mutation dynamics during maternal and paternal transmissions. The rate of very large alleles for *SCA8* is significantly higher in our study (1%; 19 independent cases with alleles ⩾100 repeats of 1,800 independent cases and controls studied) in comparison to the frequency of SCAs (0.01%, or 1/10,000; Koob et al. 1999). Even in the very narrow pathogenic repeat range defined by Koob et al (1999) of 107–127 CTG repeats or 110–130 combined repeats, our data still reveal 0.3% of affecteds and 0.1% of controls with repeats in this range. These data suggest that an etiologic factor other than expansion at *SCA8* is required for the development of ataxia. The GenBank sequences for the *SCA8* trinucleotide repeat from Koob et al. (1999) and from the present study (AF126748 and AF087653) differ at the junction between the two repeats (AF126748: $[TAC]_n$ TGCTAC$[TGC]_n$; AF087653: $[TAC]_n[TGC]_n$). Our sequence analysis and also ASO analysis failed to identify any alleles, either normal or expanded, with the AF126748 version. This suggests that the AF126748 variant at this position may be associated with ataxia; however, no patients with SCA8 have been tested to confirm this theory. Another possibility might be that the ratio of expansion of the CTA and CTG triplets, as well as expansion itself, is important for onset of disease. It is of note that the *SCA8* sequence (Koob et al. 1999; AF126758) has only 11 uninterrupted CTA triplets and 79 CTG triplets, whereas clone λ7a (AF087653) has 18 CTA triplets and 85 CTG triplets. Thus, CTA repeats represent 12% of the total repeat in AF126748 compared with 17.5% in AF087653. The observation of a large *SCA1* allele, 44 repeats long, with several interruptions and stably transmitted from an unaffected parent (Quan et al. 1995) suggests that repeat content, as well as size, may be important where genotype-phenotype correlations are inconsistent. Sequence analysis of repeats in expansions in SCA patients compared with expansions in unaffected individuals at the *SCA8* locus may be required.

Another alternative could be that all the very-large-repeat individuals detected in this study are nonpenetrant for SCA8, because of epigenetic modification, so that the expanded allele is not fully expressed. This could result from either (1) *SCA8* undergoing methylation of large repeat alleles, causing gene silencing, as occurs at *FMR1* in fragile X syndrome or (2) chromatin rearrangement or nucleosome repositioning around the expanded allele, preventing or impeding transcription of the gene. Nucleosome repositioning/chromatin rearrangement caused by large TRE raises the possibility that large expansion at *SCA8* could affect expression of other nearby genes. The finding of a higher frequency of expansion alleles for *SCA8* among individuals with psychosis in comparison with controls may imply a role

for *SCA8* or a nearby gene as a susceptibility locus for psychosis, and further studies are implicated. Because the expansion alleles are relatively rare in affected and control populations (frequency ~.01) and the predicted effect relatively modest (odds ratio ~1.8), a very large *N* (>10,000) will be required to achieve statistical significance at an 80% power level. Studies of families with psychosis and large *SCA8* repeats may be a more realistic approach for determination of the relative risk of large repeats for onset of psychosis.

*SCA8* is clearly anomalous in comparison to the other TRE spinocerebellar ataxias in that (1) the triplet is CTG rather than CAG, (2) the repeat is noncoding, (3) there is a bias toward expansion in maternal transmission and contraction in paternal transmission, (4) a much larger range of expansions is observed, in patients and controls, and (5) expansion is relatively frequent in the background population. Further work is necessary to determine whether, and to what degree, TRE at *SCA8* actually plays a role in SCA and whether expansion is a susceptibility factor for psychosis.

## Acknowledgments

Much gratitude is owed to Jennifer Skaug, Seema Khan, Barbara Kallam, and Jo-Anne Herbrick, at the Centre for Applied Genomics, and the Hospital for Sick Children, for help with the sequencing and mapping work. We also thank Dr. Vural Ozdemir for providing patient DNA samples. The efforts of Tasha Cate at Centre for Addiction and Mental Health in compiling much of the demographic data are much appreciated. This work was supported by funding from the Scottish Rite Schizophrenia Research Program, the Ontario Mental Health Foundation (OMHF), the Medical Research Council of Canada (MRC; grant MT15007), and the National Alliance for Research on Schizophrenia and Depression. Partial support was also obtained from National Institute for Mental Health grants MH33990 and POIMH56193. The BPAD trio collection was funded by Axys Pharmaceuticals. J.B.V. is a Medical Research Council/ Schizophrenia Society of Canada Research Fellow; A.D.P. is an MRC Research Fellow, and A.P. holds an OMHF New Investigator Fellowship.

## Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

GenBank database, http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html

GeneMap'98, http://www.ncbi.nlm.nih.gov/genemap98/

Genetic Location Database, http://cedar.genetics.soton.ac.uk/public_html/ldb.html

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for SCA8 [MIM 603680])

## References

Barden N, Morissette J (1998) Chromosome 13 workshop. Psychiatr Genet 8:93–96

Blouin J-L, Dombroski BA, Nath SK, Lasseter VK, Wolyniec PS, Nestadt G, Thornquist M, et al (1998) Schizophrenia susceptibility locus on chromosomes 13q32 and 8p21. Nat Genet 20:70–73

Breschel TS, McInnis MG, Margolis RL, Sirugo G, Corneliussen B, Simpson SG, McMahon FJ, et al (1997) A novel, heritable, expanding CTG repeat in an intron of the SEF2-1 gene on chromosome 18q21.1. Hum Mol Genet 6: 1855–1863

Bunzel R, Blumcke I, Cichon S, Normann S, Schramm J, Propping P, Nothen MM (1998) Polymorphic imprinting of the serotonin-2A (5HT2A) receptor gene in human adult brain. Brain Res Mol Brain Res 59:90–92

Ginns EI, Ott J, Egeland JA, Allen CR, Fann CS, Pauls DL, Weissenbachoff J, et al (1996) A genome-wide search for chromosomal loci linked to bipolar affective disorder in the Old Order Amish. Nat Genet 12:431–435

Gutierrez B, Arranz M, Fananas L, Valles V, Guillamat R, van Os J, Collier D (1995) 5HT2A receptor gene and bipolar affective disorder. Lancet 346:969

Ikeuchi T, Sanpei K, Takano H, Sasaki H, Tashiro K, Cancel G, Brice A, et al (1998) A novel long and unstable CAG/CTG trinucleotide repeat on chromosome 17q. Genomics 49:321–326

Koob MD, Moseley ML, Schut LJ, Benzow KA, Bird TD, Day JW, Ranum LPW (1999) An untranslated CTG expansion causes a novel form of spinocerebellar ataxia. Nat Genet 21:379–384

Laurent C, Zander C, Thibaut F, Bonnet-Brilhault F, Chavand O, Jay M, Samolyk D, et al (1998) Anticipation in schizophrenia: no evidence of expanded CAG/CTG repeat sequences in French families and sporadic cases. Am J Med Genet 81:342–346

Li T, Vallada HP, Liu X, Xie T, Tang X, Zhao J, O'Donovan MC, et al (1998) Analysis of CAG/CTG repeat size in Chinese subjects with schizophrenia and bipolar affective disorder using the repeat expansion detection method. Biol Psychiatry 44:1160–1165

Lindblad K, Nylander P-O, De Bryen A, Sourey D, Zander C, Engstrom C, Holmgren G, et al (1995) Detection of expanded CAG repeats in bipolar affective disorder using the repeat expansion detection (RED) method. Neurobiol Dis 2:55–62

Lindblad K, Nylander P-O, Zander C, Yuan Q-P, Stahle L, Engstrom C, Balciuniene J, et al (1998) Two commonly expanded CAG/CTG repeat loci: involvement in affective disorders? Mol Psychiatry 3:405–410

Margolis RL, Abraham MR, Gatchell SB, Li SH, Kidwai AS, Breschel TS, Stine OC, et al (1997) cDNAs with long trinucleotide repeats from human brain. Hum Genet 100: 114–122

McInnis MG, Margolis RL (1998) Anticipation and psychiatric disorders. In: Wells RD, Warren ST (eds) Genetic instabilities and hereditary neurological disorders. Academic Press, San Diego, pp 401–412

Morris AG, Gaitonde E, McKenna PJ, Mollon JD, Hunt DM

(1995) CAG repeat expansions and schizophrenia: association with disease in females and with early age-at-onset. Hum Mol Genet 4:1957–1961

Nakamoto M, Takebayashi H, Kawaguchi Y, Narumiya S, Taniwaki M, Nakamura Y, Ishikawa Y, et al (1997) A CAG/CTG expansion in the normal population. Nat Genet 17: 385–386

O'Donovan MC, Guy C, Craddock N, Bowen T, McKeon P, Macedo A, Maier W, et al (1996) Confirmation of association between expanded CAG/CTG repeats and both schizophrenia and bipolar disorder. Psychol Med 26:1145–1153

O'Donovan MC, Guy C, Craddock N, Murphy KC, Cardno AG, Jones LA, Owen MJ, et al (1995) Expanded CAG repeats in schizophrenia and bipolar disorder. Nat Genet 10: 380–381

Parikh SV, Vincent JB, Kennedy JL (1999) Clinical characteristics of bipolar disorder subjects with large CAG/CTG repeat DNA. J Affect Disord 55:221–224

Paterson AD, Naimark DMJ, Vincent JB, Kennedy JL, Petronis A (1998) Genetic anticipation in neurological and other disorders. In: Wells RD, Warren ST (eds) Genetic instabilities and hereditary neurological disorders. Academic Press, San Diego, pp 413–428

Petronis A, Bassett AS, Honer WG, Vincent J, Tatuch Y, Sasaki T, Ying DJ, et al (1996*a*) Search for unstable DNA in schizophrenia families with evidence for genetic anticipation. Am J Hum Genet 59:905–911

Petronis A, Heng HHQ, Tatuch Y, Shi X-M, Klempan Y, Tsui L-C, Ashizawa T, et al (1996*b*) Direct detection of expanded trinucleotide repeats using PCR and DNA hybridization techniques. Am J Med Genet 67:85–91

Quan F, Janas J, Popovich BW (1995) A novel CAG repeat configuration in the *SCA1* gene: implications for the molecular diagnostics in spinocerebellar ataxia type 1. Hum Mol Genet 4:2411–2413

Sanpei K, Takano H, Igarashi S, Sato T, Oyake M, Sasaki H, Wakisaka A, et al (1996) Identification of spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. Nat Genet 14: 277–284

Schalling M, Hudson TJ, Buetow KH, Housman DE (1993) Direct detection of novel expanded trinucleotide repeats in the human genome. Nat Genet 4:135–139

Sham PC, Curtis D (1995) An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. Ann Hum Genet 59:323–336

Sidransky E, Burgess C, Ikeuchi T, Linblad K, Long RT, Philibert RA, Rapoport J, et al (1998) A triplet repeat on 17q accounts for most expansions detected by the repeat-expansion-detection technique. Am J Hum Genet 62:1548–1551

Sirugo G, Pakstis AJ, Kidd KK, Matthysse S, Levy DL, Holzman PS, Parnas J, et al (1997) Detection of a large CTG/CAG trinucleotide repeat expansion in a Danish schizophrenia kindred. Am J Med Genet 74:546–548

Stine OC, McMahon FJ, Chen L-S, Xu J, Myers DA, MacKinnon DF, Simpson S, et al (1997) Initial genome screen for bipolar disorder in the NIMH genetics initiative pedigrees: chromosomes 2, 11, 13, 14, and X. Am J Med Genet 74:263–269

Terwilliger JD, Ott J (1994) Handbook of human genetic linkage. Johns Hopkins University Press, Baltimore

Vincent JB, Kalsi G, Klempan T, Tatuch Y, Sherrington RP, Breschel T, McInnis MG, et al (1998) No evidence of expansion of CAG or GAA repeats in schizophrenia families and monozygotic twins. Hum Genet 103:41–47

Vincent JB, Klempan T, Parikh SS, Sasaki T, Meltzer HY, Sirugo G, Cola P, et al (1996) Frequency analysis of large CAG/CTG trinucleotide repeats in schizophrenia and bipolar affective disorder. Mol Psychiatry 1:141–148

Vincent JB, Kovacs M, Krol R, Barr CL, Kennedy JL (1999*a*) Intergenerational CAG repeat expansion at ERDA1 in a family with childhood onset depression, schizoaffective disorder and recurrent major depression. Am J Med Genet 88: 79–82

Vincent JB, Petronis A, Strong E, Parikh SV, Meltzer HY, Lieberman J, Kennedy JL (1999*b*) Analysis of genome-wide CAG/CTG repeats, and at *SEF2-1B* and *ERDA1* in schizophrenia and bipolar affective disorder. Mol Psychiatry 4: 229–234

Williams J, Spurlock G, McGuffin P, Mallet J, Nothen MM, Gill M, Aschauer H, et al (1996) Association between schizophrenia and T102C polymorphism of the 5-hydroxytryptamine type 2a-receptor gene. Lancet 347:1294–1296

Zander C, Schurhoff F, Laurent C, Chavand O, Bellivier F, Samolyk D, Leboyer M, et al (1998) CAG repeat sequences in bipolar affective disorder: no evidence for association in a French population. Am J Med Genet 81:338–341