Am. J. Hum. Genet. 53:391-400, 1993

The Gene for Autosomal Dominant Spinocerebellar Ataxia (SCA1) Maps Centromeric to D6S89 and Shows No Recombination, in Nine Large Kindreds, with a Dinucleotide Repeat at the AM10 Locus

Thomas J. Kwiatkowski, Jr.,* Harry T. Orr,‡ Sandro Banfi,† Alanna E. McCall,† Carla Jodice,§ Francesca Persichetti,§ Andrea Novelletto,§ Françoise LeBorgne-DeMarquoy,# Lisa A. Duvick,‡ Marina Frontali,|| S. H. Subramony,** Arthur L. Beaudet,*,† Luciano Terrenato,§ Huda Y. Zoghbi,*,† and Laura P. W. Ranum‡

*Institute for Molecular Genetics and †Department of Pediatrics, Baylor College of Medicine, Houston; ‡Department of Laboratory Medicine and Pathology, and Institute of Human Genetics, University of Minnesota, Minneapolis; §Dipartimento di Biologia, Tor Vergata University, and ||Istituto di Medicina Sperimentale, CNR, Rome; #Genethon, Paris; and **Department of Neurology, University of Mississippi School of Medicine, Jackson

Summary

Spinocerebellar ataxia type 1 (SCA1) is an autosomal dominant disorder which is genetically linked to the short arm of chromosome 6, telomeric to the human major histocompatibility complex (HLA) and very close to D6S89. Previous multipoint linkage analysis using HLA, D6S89, and SCA1 suggested that SCA1 maps centromeric to D6S89. Data from this study using nine large kindreds indicate a maximum lod score between SCA1 and D6S89 of 67.58 at a maximum recombination fraction of .004. To localize SCA1 more precisely, we identified five dinucleotide polymorphisms near D6S89. Genotypic analyses at these polymorphic loci were carried out in nine multigeneration SCA1 kindreds and in the Centre d'Étude du Polymorphisme Humain reference families. A new marker, AM10GA, demonstrates no recombination with SCA1. The maximum lod score for AM10GA linkage to SCA1 is 42.14 at a recombination fraction of 0. Linkage analysis and analysis of recombination events confirm that SCA1 maps centromeric to D6S89 and establish the following order: CEN-D6S109-AM10GA/SCA1-D6S89-LR40-D6S202-TEL.

Introduction

The spinocerebellar ataxias are a heterogeneous group of degenerative neurological disorders with variable clinical features resulting from degeneration of the cerebellum, brain stem, and spinocerebellar tracts. The clinical features of spinocerebellar ataxia type 1 (SCA1) include ataxia, dysarthria, ophthalmoparesis, and variable degrees of muscle wasting and neuropathy. Dominantly inherited SCA1 has been mapped to the short arm of chromosome 6 on the basis of linkage to the human major histocompatibility complex (HLA) (Yakura et al. 1974; Jackson et al. 1977). SCA1 has been shown to be tightly linked to the marker D6S89 on the short arm of chromosome 6, telomeric to HLA (Ranum et al. 1991*b;* Zoghbi et al. 1991). Recently, two families with dominantly inherited ataxia failed to show detectable linkage with HLA markers but were found to have SCA1 when studied for linkage to D6S89, demonstrating the superiority of the latter marker for study of ataxia families (Keats et al. 1991; R. Ramesar, personal communication).

Substantial efforts have been made to localize the SCA1 gene by using genetic and physical mapping methods. Genetically, SCA1 is flanked on the centromeric side by D6S88 at a recombination fraction (θ) of approximately .08 (on the basis of marker-marker distances, by using the Centre d'Étude du Polymorphisme

Received February 5, 1993; revision received April 14, 1993.

Address for correspondence and reprints: Thomas J. Kwiatkowski, Jr., Institute for Molecular Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030.

0002-9297/93/5302-0011$02.00

Humain [CEPH] reference families) and on the telomeric side by F13A at a θ of .19 (Ranum et al. 1991*b*). Both of these markers are quite distant and are not practical for use in efforts aimed at cloning the SCA1 gene. The D6S89 marker is much closer to SCA1 than any other marker. Data from a previous study favored a centromeric position for SCA1 relative to D6S89 by odds of 10^3:1 (Ranum et al. 1991*b*). To date, after careful analysis of clinical material on all potential recombination events between SCA1 and D6S89, only one definite recombination event has been documented between SCA1 and D6S89 among all nine kindreds evaluated.

To confirm the position of SCA1 with respect to D6S89 and to identify closer flanking markers, we used the two dinucleotide repeat polymorphisms D6S109 and D6S202 and sought new markers in the candidate SCA1 region. Using yeast artificial chromosome (YAC) clones isolated in the D6S89 region, we have identified three additional dinucleotide repeat polymorphisms, one of which (AM10GA) shows no recombination with SCA1 and confirms that D6S89 is telomeric to SCA1. Linkage analysis, physical mapping data (S. Banfi, M.-y. Chung, T. J. Kwiatkowski, Jr., L. P. W. Ranum, A. E. McCall, A. C. Chinault, H. T. Orr, and H. Y. Zoghbi, unpublished data), and analysis of recombination events demonstrate that the order of markers is as follows: CEN–D6S109–AM10GA/SCA1–D6S89–SB1–LR40–D6S202–TEL.

Subjects, Material, and Methods

SCA1 Kindreds

Nine large SCA1 families were used in the present study. Clinical findings and linkage data demonstrating that these families segregate SCA1 have been reported elsewhere (Jackson et al. 1977; Keats et al. 1991; Ranum et al. 1991*b;* Zoghbi et al. 1991). Analysis of polymorphisms at the loci D6S109, AM10GA, SB1, LR40, and D6S202 was performed on individuals from these kindreds.

The Houston (TX-SCA1) kindred includes 106 individuals, of whom 57 (25 affected) were genotyped (Zoghbi et al. 1988). Patients symptomatic at the time of exam, as well as asymptomatic individuals who have both a symptomatic child and a symptomatic parent, are classified as "affected." In this kindred, a deceased individual previously assigned as affected (on the basis of family history data) has been reassigned as "unknown" after review of medical records. This reassignment eliminates what was previously thought to be a recombination event between SCA1 and D6S89 in the TX-SCA1 kindred. To maximize the amount of information available for linkage analysis, we employed the strategy of separating the two chromosomes 6 in somatic cell hybrids for 15 affected individuals and for 1 unaffected individual from the TX-SCA1 kindred (Zoghbi et al. 1989). The Louisiana (LA-SCA1) kindred includes 50 individuals, of whom 26 (8 affected) were genotyped (Keats et al. 1991). The Minnesota (MN-SCA1) kindred includes 175 individuals, of whom 106 (17 affected) were genotyped (Haines et al. 1984; Ranum et al. 1991*b*). The Michigan (MI-SCA1) kindred includes 201 individuals, of whom 127 (25 affected) were genotyped (Nino et al. 1980). The Mississippi (MS-SCA1) kindred includes 84 individuals, of whom 37 (17 affected) were genotyped (Jackson et al. 1977).

Four Italian families segregating SCA1 were analyzed; their clinical phenotype and HLA linkage data were reported elsewhere (Spadaro et al. 1992). Three families originated in the Calabria region (southern Italy): family IT-P, with 135 members, of whom 80 (21 affected) were genotyped (for computational reasons, the family was subdivided into three different pedigrees [RM, VI, and FB], and only one of the three consanguinity loops was considered); family IT-PS, with 43 members, of whom 27 (7 affected) were genotyped; and family IT-NS, with 51 members, of whom 16 (3 affected) were genotyped. The fourth family, IT-MR, originated in Latium and consists of 17 individuals, of whom 10 (4 affected) were genotyped.

CEPH Families

The 40 CEPH reference families were genotyped at the D6S109, LR40, and D6S202 loci in order to provide a large number of informative meioses for marker-marker linkage analyses. Markers AM10GA and SB1 flank D6S89, having been isolated from a YAC contig built bidirectionally from D6S89 (S. Banfi, M.-y. Chung, T. J. Kwiatkowski, Jr., L. P. W. Ranum, A. E. McCall, A. C. Chinault, H. T. Orr, and H. Y. Zoghbi, unpublished data). A subset of 18 CEPH families which defines 26 recombinants between D6S109 and D6S89 was genotyped at AM10GA and SB1 in order to determine the order of AM10GA, D6S89, and SB1 with respect to D6S109.

Cloning of Sequences Containing Dinucleotide Repeats

The identification and description of polymorphic dinucleotide repeats at the D6S109 and D6S202 loci have been reported elsewhere (LeBorgne-Demarquoy et al. 1991; Ranum et al. 1991*a*). DNA fragments containing dinucleotide repeats were cloned at LR40 and SB1 from YAC clones at the LR40 and FLB1 loci, respectively (S. Banfi, M.-y. Chung, T. J. Kwiatkowski, Jr., L. P. W. Ranum, A. E. McCall, A. C. Chinault, H. T.

Orr, and H. Y. Zoghbi, unpublished data). DNA from each YAC clone was amplified in a 50-μl reaction containing 20 ng DNA, a single Alu primer (see below), 50 mM KCl, 10 mM Tris-Cl pH 8.3, 1.25 mM $MgCl_2$, 200 or 250 μM dNTPs, 0.01% (w/v) gelatin, and 1.25 units *Thermus aquaticus* DNA polymerase (*Taq* polymerase; Perkin Elmer Cetus, Norwalk, CT). For amplification of FLB1 YAC DNA, a primer (5′-AGGAGTGAGC-CACCGCACCCAGCC-3′) complementary to the 5′ end of the Alu consensus sequence (Susan Airhardt; Oncor Laboratories), designated "SAL1" was used at a final concentration of 0.6 μM. For amplification of LR40 YAC DNA, 0.2-μM primer PDJ34 was used (Breukel et al. 1990). Samples were overlaid with mineral oil, denatured at 94°C for 5 min, then subjected to 30 cycles of 1 min at 94°C for denaturation, 1 min at 55°C for annealing, and 5 min at 72°C for extension. The last extension step was lengthened to 10 min. Electrophoresis of 15 μl PCR products was performed on a 1.5% agarose gel, which was Southern blotted and hybridized with a probe prepared by random-hexamer–primed labeling (Feinberg and Vogelstein 1984) of synthetic poly(dG-dT)-poly(dA-dC) (Pharmacia; Piscataway, NJ), using [α-^{32}P]dCTP. Fragments hybridizing with the dinucleotide repeat probe were identified and were subsequently purified by electrophoresis on a low-melt agarose gel. Fragments were excised and reamplified by PCR as above.

For LR40, reamplified DNA was repurified by low-melt gel electrophoresis, and DNA was extracted from excised bands by passage through a glass-wool spin column (Heery et al. 1990). A purified 1.2-kb fragment was cloned into pBluescript plasmid modified as a "T-vector" (Marchuck et al. 1990). From this clone, a 0.6-kb *Hin*cII restriction fragment containing a GT repeat was subcloned into pBluescript plasmid and was sequenced on an automated sequencer (Applied Biosystems; Foster City, CA).

For SB1, a reamplified 1-kb fragment was ethanol precipitated and blunt-end cloned into pBluescript plasmid. Plasmid DNA was isolated and PCR amplified in a first reaction with M13 reverse primer and BamGT primer (5′-CCCGGATCCTGTGTGTGTGTGTGT-GTG-3′) and in a second reaction with M13 universal primer and BamCA primer (5′-CCCGGATCCACACA-CACACACACACAC-3′) (Feener et al. 1991). PCR conditions were as above except that primers were used at 1-μM concentration; 2.5 units *Taq* polymerase and approximately 30 ng DNA were used per reaction, with final reaction volumes of 100 μl and an annealing temperature of 50°C. Products were precipitated, resuspended, and digested with *Bam*HI (product of the universal primer reaction) or *Bam*HI and *Hin*dIII (product of the reverse primer reaction). These two fragments were cloned into pBluescript plasmid and were sequenced as above.

Dinucleotide repeats were cloned at AM10 from a YAC containing this locus. A λFixII library was constructed by using DNA from this yeast clone, and human clones were identified by filter hybridization, using human placental DNA as a probe. A gridded array of these human clones was grown, and filters containing DNA from these clones were hybridized with a ^{32}P-labeled poly(dG-dT)-poly(dA-dC) probe as described above. DNA was prepared from positive clones, digested with various restriction enzymes, and analyzed by agarose gel electrophoresis. Southern blotting and hybridization were carried out with the poly(dG-dT)-poly(dA-dC) probe. A 1-kb fragment hybridizing with the dinucleotide repeat probe was identified, cloned into M13, and sequenced.

PCR Analysis

Primer sequences and concentrations, as well as PCR cycle times used for amplification of dinucleotide repeat sequences from human genomic DNA, are presented in table 1. For the LR40 polymorphism, primer set A was used for analysis of the TX-SCA1, LA-SCA1, and MS-SCA1 kindreds, while primer set B was used for all other kindreds. Buffer compositions were as follows: 50 mM KCl, 10 mM Tris-Cl pH 8.3, 1.25 mM $MgCl_2$ (1.5 mM $MgCl_2$ for AM10GA), 250 μM dNTPs (200 μM dNTPs for AM10GA), 0.01% (w/v) gelatin, and 0.5–0.625 unit *Taq* polymerase. For the LR40 analysis, 2% formamide was included in the PCR buffer. When primer set B was used for LR40 analysis, 125 μM dNTPs, 1.5 mM $MgCl_2$, and 1 unit *Taq* polymerase were used. All reaction volumes were 25 μl and contained 40 ng genomic DNA. Four microliters of each reaction was mixed with 2 μl formamide loading buffer, denatured at 90–100°C for 3 min, and cooled on ice, and 2–4 μl was used for electrophoresis on a 4% or 6% polyacrylamide/7.65-M urea sequencing gel for 2–3 h at 1,100 V. PCR assay conditions have been reported elsewhere for D6S202 and D6S109 (LeBorgne-Demarquoy et al. 1991; Ranum et al. 1991*a*).

SCA1 Linkage Analysis

The D6S109, AM10GA, D6S89, SB1, LR40, and D6S202 markers were analyzed for linkage to SCA1 by using the computer program LINKAGE, version 5.1 (Lathrop et al. 1984), which includes the MLINK, ILINK, LINKMAP, CLODSCORE, and CMAP programs. Age-dependent penetrance classes were assigned independently for each of the families included in the

Table 1

Primers and PCR Conditions for Amplification of Dinucleotide Repeat Sequences

Marker/(Type)	Primers[a]	PCR Steps	PCR Cycles
AM10GA/$(GA)_n$	5′-AAGTCAGCCTCTACTCTTTGTTGA-3′ 5′-CTTGGAGCAGTCTGTAGGGAG-3′	94°C for 30 s 55°C for 30 s 72°C for 30 s	30
SB1/$(GT)_n$	5′-TGAAGTGATGTGCTCTGTTC-3′ 5′-AAAGGGGTAGAGGAAATGAG-3′	94°C for 60 s 60°C for 60 s 72°C for 60 s	30
LR40$(GT)_n$[b]	5′-AGGAGAGGGGTCATGAGTTG-3′ 5′-GGCTCATGAATACATTACATGAAG-3′	94°C for 60 s 58°C for 60 s 72°C for 60 s	25
LR40/$(GT)_n$[c]	5′-CTCATTCACCTTAGAGACAAATGGATAG-3′ 5′-ATGGTATAGGGATTTTNCCAAACCTG-3′	94°C for 60 s 60°C for 60 s 72°C for 45 s	27

[a] The first primer of each pair was end labeled with γ-^{32}P ATP and polynucleotide kinase. Primer concentrations were 1 μM.
[b] Set A.
[c] Set B.

analysis. Marker alleles were recoded to reduce the number of alleles segregating in a family to four, five, or six alleles to simplify the analysis. The allele frequencies for the various markers were based on the frequencies of the alleles among the spouses in each family and were determined separately for the two American black kindreds, for the Italian kindreds, and for the Caucasian kindreds from Minnesota, Michigan, and Mississippi, with the following exception: the allele frequencies for D6S109 in the MI-SCA1 and MN-SCA1 kindreds were based on the frequencies of the alleles in the CEPH families.

Maximum Lod scores (Z_{max}) for the various markers were calculated with the MLINK program by running each of the analyses separately for the various families, at θ values with increments of .0005 to .001, and then by adding the values of each of the kindreds. The analyses were done separately to ensure that the allele frequencies for the various markers were representative for each of the ethnically diverse families. As a control, the θ values at the Z_{max} between each marker and SCA1 were calculated by using the ILINK program after the allele frequencies for each marker were set equal to one another. In all cases the $\hat{\theta}$ values were the same as, and the Z_{max} values were very similar to, those reported in table 5.

CEPH Linkage Analysis

Forty CEPH families were typed for the GT repeat markers D6S109, D6S202, and LR40. The original alleles were recoded to five alleles. The SB1 and AM10 markers were typed in a subset of the CEPH panel which defines 26 recombinants, from 18 different families, between D6S109 and D6S89. The CLODSCORE program was used for the two-point analyses, and CMAP was used for the three-point and four-point analyses. For the three-point and four-point analyses, the interval between the mapped markers was fixed on the basis of the two-point $\theta_m = \theta_f$ results. The likelihood of the location of the test locus (SCA1) was calculated at 10 different positions within each interval. The test for sex difference in the θ values was performed by using a χ^2 statistic, with $\chi^2 = 2(\ln 10)[Z(\theta_m,\theta_f) - Z(\theta = \theta_m = \theta_f)]$, where $Z(\theta_m,\theta_f)$ is the overall Z_{max} for arbitrary θ_m and θ_f, while $Z(\theta = \theta_m = \theta_f)$ is the Z_{max} constrained to $\theta_m = \theta_f$. Under homogeneity, χ^2 approximates a χ^2 with 1 df. Rejection of homogeneity occurs when $\chi^2 > 3.84$.

Results

Dinucleotide Repeat Cloning and Sequencing and Analysis

Dinucleotide repeats SB1 and LR40 were amplified directly from YAC clones by Alu-primed PCR, and the dinucleotide repeat–containing fragments were identified by hybridization. Figure 1 shows the identification of a GT repeat–containing 1.2-kb PCR product from an amplification of LR40 YAC DNA; the SB1 GT re-

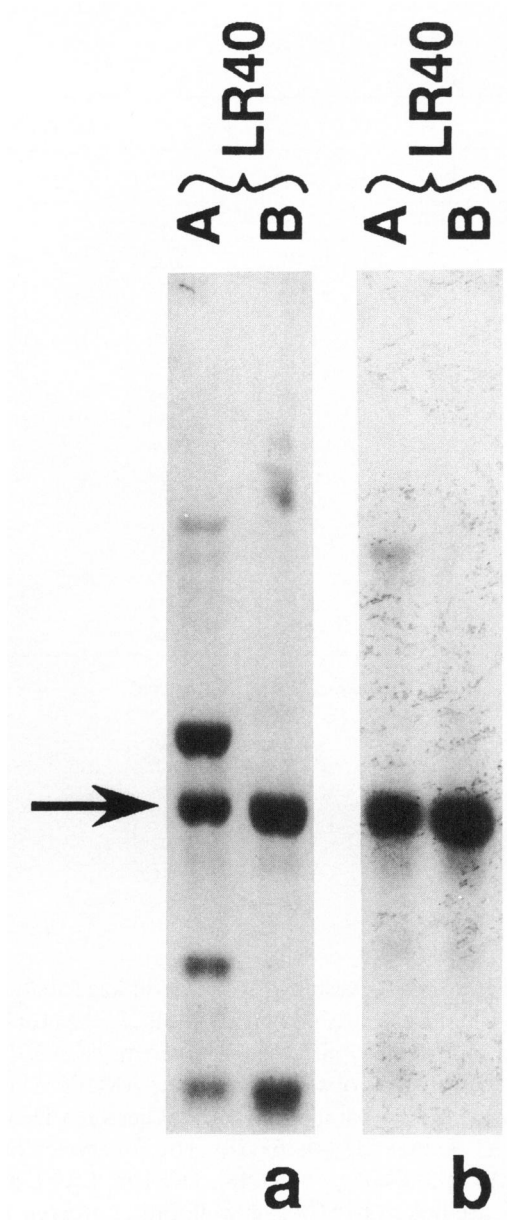

Figure 1 Identification of dinucleotide repeat sequences in Alu-PCR products from YACs. *a*, Ethidium bromide–stained agarose gel showing the Alu-PCR products of two LR40 YACs. The Alu primer was PDJ34. *b*, Hybridization of the synthetic poly(GT)-poly(CA) probe to a Southern transfer of the gel in panel *a*. A dinucleotide repeat–containing fragment was identified on the basis of the strong hybridization signal. The dinucleotide-positive fragment, identified with an arrow, was subsequently cloned and sequenced.

peat was identified in a similar fashion. These PCR products were cloned either directly or by further amplification by using tailed poly(GT) or poly(CA) primers paired with an Alu primer. In addition, two dinucleotide repeats were subcloned from a lambda phage clone from a library constructed from a YAC at the AM10 locus.

Dinucleotide repeats from the SB1, LR40, and AM10 loci were sequenced. At LR40, the cloned repeat sequence was $(CA)_{16}TA(CA)_{10}$. The AM10 fragment contained two repeat sequences separated by 45 bp of nonrepeat sequence. The first repeat, designated "AM10GA," was $(GA)_2ATGACA(GA)_{11}$. The second repeat, designated "AM10GT," was not used in this study, because, on analysis of the TX-SCA1 kindred, it yielded the same information as the AM10GA repeat. The AM10GT repeat consists of $(GA)_2AA(GA)_6GT-GA(GT)_{16}AT(GT)_5$. Primer information for AM10GT is available through the Genome Data Base. At SB1, the repeat tract was not sequenced; only flanking sequence was determined.

As there are differences in allele distributions of markers among the different races, allele frequencies are reported here separately for the CEPH kindreds (Caucasian) and the TX-SCA1 kindred (American black) (table 2). CEPH allele frequencies were based on 72 independent chromosomes for SB1, on 82 independent chromosomes for AM10, and on the full set of 40 families for D6S109 and LR40. TX-SCA1 allele frequencies were based on 45 independent chromosomes for LR40, 43 independent chromosomes for SB1, 45 independent chromosomes for AM10, and 42 independent chromosomes for D6S109.

Genetic Linkage Data

CEPH families.—In order to establish a well-defined genetic map for the SCA1 region, newly isolated DNA markers were mapped by using the CEPH reference families. Results of pairwise linkage analyses in CEPH kindreds are shown in table 3. No recombination was observed between AM10GA and D6S89 ($\theta = 0$; $Z_{max} = 15.1$) by using a subset of the CEPH panel which defines 26 recombinants between D6S109 and D6S89. The markers D6S109 and LR40 are close to D6S89, with θ values of .067 ($Z_{max} = 71.4$) and .04 ($Z_{max} = 84.5$), respectively.

Selected multipoint analyses were performed to position the newly isolated markers D6S109, LR40, and D6S202 with respect to markers previously mapped by using the CEPH panel. The CMAP program was used for three-point and four-point linkage analyses to position D6S109 relative to D6S88 and D6S89 and to position LR40 and D6S202 relative to each other and to D6S89 and F13A. For the three-point analyses, the D6S88–D6S89 interval was fixed on the basis of the two-point θ in CEPH, and Z was calculated at various θ values. The order D6S88–D6S109–D6S89 is favored

Table 2

Allele Frequencies of New Markers

	D6S109[b]	AM10GA		SB1		LR40		D6S202[c]
ALLELE[a]	TX-SCA1	TX-SCA1	CEPH	TX-SCA1	CEPH	TX-SCA1	CEPH	TX-SCA1
A0	...	...	.012[d]	.070[e]	...	...	...	...
A1	.048[f]	.022	.024	.163	.027	.244[g]	.022[h]	.05[i]
A2	.024	.289	.220	.186	.166	.045	.043	.11
A3	.119	...	.024	.070	.333	.111	.065	.11
A4	.024	.333	.232	.023	...	.133	.033	.13
A5	.071	.267	.488	.186	.097	.111	.272	.11
A6	.261	...	...	.093	.111	...	.098	.03
A7	.024	.089	...	.093	.153	.022	.054	.22
A8	.095	...	...	.093	.083	.045	.076	.13
A9	.143	...	...	...	.014	.089	.054	.08
A10	...	...	...	...	...	.022	.065	.03
A11	.048	...	...	.023	...	.133	.011	...
A12	.048	...	...	...	...	.045	.054	...
A13	.048	...	...	...	.014	...	.097	...
A14	.071	...	...	...	...	...	.033	...
A15	...	...	...	...	...	...	.023	...

[a] Alleles are numbered such that the largest allele is assigned the lowest number and each successive allele is 2 bp smaller.
[b] CEPH data published by Ranum et al. (1991*a*).
[c] CEPH data published by LeBorgne-Demarquoy et al. (1991).
[d] 123 bp.
[e] 220 bp.
[f] 215 bp.
[g] 241 bp (Set A, table 1).
[h] 267 bp (Set B, table 1).
[i] 154 bp.

over the next most likely order by odds of 4×10^3:1 (table 4). For the four-point analyses, both the D6S89–D6S202–F13A and the D6S89–LR40–F13A intervals were fixed on the basis of the two-point θ values; Z's were then calculated for LR40 and D6S202 at various θ values on the respective fixed maps. The order D6S89–LR40–D6S202–F13A is favored over the next most likely order in both analyses; odds in favor were 400:1 when the position of LR40 was varied and 10^6:1 when the position of D6S202 was varied (table 4).

The order of AM10GA and D6S89 could not be determined by using the D6S109/D6S89 CEPH recombinants. However, the order AM10GA–D6S89–SB1 was deduced by characterization of overlapping YAC clones containing these markers (S. Banfi, M.-y. Chung, T. J. Kwiatkowski, Jr., L. P. W. Ranum, A. E. McCall, A. C. Chinault, H. T. Orr, and H. Y. Zoghbi, unpublished data). Furthermore, one end of this contig is present in a well-characterized radiation-reduced hybrid known to contain D6S109 and other centromeric markers, indicating the order D6S109–AM10GA–D6S89–SB1.

SCA1 kindreds.—Results of pairwise linkage analyses in SCA1 kindreds are shown in table 5. AM10GA, D6S89, and SB1 are all closely linked to SCA1. No recombination was observed between AM10GA and SCA1; Z_{max} is 42.14 at a $\hat{\theta}$ of 0. The $\hat{\theta}$ between D6S89 and SCA1 is .004 (Z_{max} = 67.58). The $\hat{\theta}$ between SB1 and SCA1 is .007 (Z_{max} = 39.46). D6S109, LR40, and D6S202 are linked to SCA1 as well, but at greater distances ($\hat{\theta}$ = .04, .03, and .08, respectively). On the basis of genetic mapping in nine large kindreds, the SCA1 locus is very close to D6S89 and AM10GA, with a Z_{max} − 1 support interval ⩽.02 in both cases.

Analysis of Key Recombinants

To date, only one recombination event between D6S89 and SCA1 has been confirmed in an affected individual. This patient, individual MI-2 in figure 2 is also recombinant at SB1, although uninformative at LR40 and D6S202. He carries a disease haplotype at the HLA, D6S109, and AM10 loci, demonstrating that SCA1 is centromeric to D6S89, as indicated by the rightmost arrow in figure 2. To eliminate the possibility

Table 3

Pairwise Linkage Results in the CEPH Families

Marker Pair	$\theta_m = \theta_f$	Z_{max}	θ_m	θ_f	Z_{max}	χ^2
HLA and:						
D6S88	.128	26.4	.103	.168	26.8	1.86
D6S109	.126	48.4	.062	.176	51.0	12.1[a]
AM10	.608	.0440	.301	.500	.246	.929
D6S89	.158	43.3	.091	.225	46.6	15.2[a]
SB1	.574	.0190	.299	.500	.400	1.75
LR40	.213	25.5	.116	.306	30.0	20.8[a]
HZ30	.251	21.6	.191	.318	23.6	8.95[a]
F13A	.291	8.81	.255	.326	9.14	1.52
D6S88 and:						
D6S109	.017	48.6	.024	.009	48.8	.846
AM10	.654	.0290	.499	.696	.047	.0820
D6S89	.086	36.1	.076	.098	36.2	.461
SB1	.203	1.09	.136	.687	1.36	1.27
LR40	.088	31.1	.078	.104	31.2	.350
HZ30	.135	30.4	.124	.152	30.4	0
F13A	.180	10.2	.158	.217	10.3	.626
D6S109 and:						
AM10	.730	.933	.170	.502	1.67	3.39
D6S89	.067	71.4	.035	.090	72.5	5.15[a]
SB1	.742	1.95	.113	.501	4.32	10.9[a]
LR40	.109	50.6	.050	.152	52.9	10.5[a]
HZ30	.162	36.6	.147	.174	36.7	.515
F13A	.207	14.4	.211	.204	14.4	0
AM10 and:						
D6S89	.000	15.1	.000	.000	15.1	.000
SB1	.000	13.2	.000	.000	13.2	.000
LR40	.021	8.74	.000	.050	9.11	1.74
HZ30	.000	13.8	.000	.000	13.8	.000
F13A	.135	3.48	.042	.253	4.39	4.16[a]
D6S89 and:						
SB1	.000	25.0	.000	.000	25.0	.000
LR40	.040	84.5	.030	.049	84.7	.925
HZ30	.078	76.0	.075	.077	76.0	0
F13A	.151	30.7	.139	.160	30.7	0
SB1 and:						
LR40	.033	14.4	.022	.044	14.5	.350
HZ30	.026	17.5	.032	.020	17.5	0
F13A	.136	4.80	.119	.155	4.84	.170
LR40 and:						
HZ30	.079	64.8	.092	.050	65.0	1.09
F13A	.131	29.1	.121	.140	29.2	.461
HZ30 and:						
F13A	.109	38.4	.122	.106	38.4	0

[a] Statistically significant differences were observed in the θ values when the assumption of homogeneity ($\theta_m = \theta_f$) was rejected; i.e., the likelihood that $\chi^2 > 3.84$ with 1 df should occur by chance is $P < .05$.

of sample mix-up, the patient's DNA was reextracted from a hair sample and was retyped for D6S109, D6S89, D6S202, LR40, AM10GA, and SB1. The results from the hair sample matched those from the cell line originally established from the patient's blood. The patient's medical records were carefully reexamined, and we confirmed that he did indeed have ataxia. In addition, his haplotypes are consistent with those of a sister and a daughter.

D6S109 lies centromeric to D6S89; six recombination events have been observed between D6S109 and SCA1, as shown in figure 2. At this point, D6S109 is the

Table 4

Three-Point and Four-Point Linkage Analyses in the CEPH Families

Order	Z_{max}	Relative Odds	Odds in Favor
D6S109-D6S88-D6S89	90.6	2×10^8	
D6S88-D6S109-D6S89	94.2	8×10^{11}	4×10^3:1
D6S88-D6S89-D6S109	82.3	1	
LR40-D6S89-D6S202-F13A	96.1	1×10^{34}	
D6S89-LR40-D6S202-F13A	98.6	4×10^{36}	400:1
D6S89-D6S202-LR40-F13A	73.9	8×10^{11}	
D6S89-D6S202-F13A-LR40	62.0	1	
D6S202-D6S89-LR40-F13A	89.5	1×10^{32}	
D6S89-D6S202-LR40-F13A	57.5	1	
D6S89-LR40-D6S202-F13A	95.5	1×10^{38}	10^6:1
D6S89-LR40-F13A-D6S202	77.6	1×10^{20}	

centromeric marker closest to SCA1. The arrows in figure 2 denote the maximum region common to all affected chromosomes and, therefore, the maximum possible region containing the SCA1 gene, which extends from D6S89 to D6S109.

No additional marker-SCA1 recombination events have been observed between D6S89 and SB1. Markers further telomeric to SB1 show additional recombination with SCA1—one recombination event between SCA1 and LR40 and three recombination events between SCA1 and D6S202. These events are depicted in figure 2 (all recombination events shown are in affected individuals).

Discussion

SCA1 has been mapped telomeric to the HLA loci on the basis of close linkage to D6S89 (Ranum et al. 1991*b*; Zoghbi et al. 1991). Identification of families affected with SCA1 is now accomplished by testing for linkage to D6S89, as HLA markers have failed to detect linkage in at least two SCA1 families that demonstrate linkage to D6S89 (Keats et al. 1991; R. Ramesar, personal communication). Data from one study mapped SCA1 centromeric to D6S89 on the basis of multipoint linkage analysis by using HLA, D6S89, and SCA1 in two kindreds (Ranum et al. 1991*b*). To confirm this finding and to develop a high-resolution genetic map of the putative SCA1 region, we sought new DNA markers of the dinucleotide repeat type.

Three markers, designated "AM10," "LR40," and "FLB1," were isolated from a radiation-reduced hybrid known to contain D6S89 as well as D6S88 and F13A, two markers known to flank the SCA1 locus. YACs were isolated at the AM10, LR40, and FLB1 loci. Three newly described repeats were isolated by using DNA either from these YAC clones or from phage subclones of one of the YACs. The dinucleotide repeat sequences LR40 and SB1 were identified by amplifying inter-Alu sequences, using DNA from YACs at LR40 and FLB1, respectively, and subsequently hybridizing the Alu-PCR products with a poly(dG-dT)-poly(dC-dA) probe. This approach proved rapid but was not always successful, as it required the dinucleotide repeat sequence to be flanked by properly oriented Alu sequences. When this approach failed (as in the case of AM10 YACs), we resorted to screening phage subclones from an appropriate nonchimeric YAC for the presence of dinucleotide repeat sequences. This approach, although tedious and time consuming, is more likely to yield a dinucleotide sequence from a target region. Using these three dinucleotide polymorphisms, as well as two additional dinucleotide polymorphisms (D6S109 and D6S202) developed in our laboratories, we performed genotypic analysis on individuals from nine SCA1 kindreds. Because genotypic analysis using D6S89 revealed a single

Table 5

Pairwise Lod Scores for SCA1 and Dinucleotide Repeat Markers

	Z AT θ =									
	0	.001	.05	.1	.2	.3	.4	Z_{max}	$\hat{\theta}$	SUPPORT INTERVAL[a]
SCA1:D6S109	$-\infty$	22.68	33.81	32.03	25.19	16.56	7.24	33.82	.04	.02–.09
SCA1:AM10GA	42.14	42.06	38.48	34.51	25.86	16.63	7.30	42.14	0	0–.02
SCA1:D6S89	$-\infty$	67.35	62.78	56.39	42.51	27.56	12.09	67.58	.004	0–.02
SCA1:SB1	$-\infty$	39.02	37.33	33.92	26.16	17.53	8.33	39.46	.007	0–.03
SCA1:LR40	$-\infty$	27.80	31.77	29.73	23.61	16.11	7.77	32.08	.03	.001–.07
SCA1:D6S202	$-\infty$	4.41	25.80	26.47	22.12	14.77	6.51	26.61	.08	.04–.14

[a] $Z_{max} - 1$ = support interval of θ (Conneally et al. 1985).

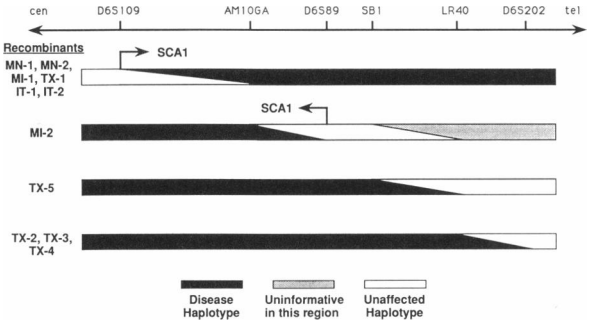

Figure 2 Summary of SCA1 recombination events which led to the precise mapping of the SCA1 locus. Recombinant disease-carrying chromosomes are shown for the markers shown at the top. A schematic diagram of the relevant region of 6p22 (not drawn to scale) is shown at the top of the figure. Families are coded as follows: TX = Houston; MN = Minnesota; MI = Michigan; and IT = Italy. Each recombination event is assigned a number following the family code.

recombination event between D6S89 and SCA1, and because the other flanking markers (D6S88 and F13A) are quite far from the SCA1 locus, we needed to analyze a large number of individuals for the identification and mapping of recombination events in the vicinity of the SCA1 locus. In addition, individuals from the CEPH reference families were genotyped at these five loci, for accurate pairwise genetic distance calculations and multipoint analyses.

No recombination was observed between the markers AM10GA, D6S89, and SB1 in CEPH families. The order of these three markers was determined to be AM10GA–D6S89–SB1, by characterization of a contig of YAC clones spanning these loci (S. Banfi, M.-y. Chung, T. J. Kwiatkowski, Jr., L. P. W. Ranum, A. E. McCall, A. C. Chinault, H. T. Orr, and H. Y. Zoghbi, unpublished data). In addition, physical mapping data from using well-characterized radiation-reduced hybrids (S. Banfi, M.-y. Chung, T. J. Kwiatkowski, Jr., L. P. W. Ranum, A. E. McCall, A. C. Chinault, H. T. Orr, and H. Y. Zoghbi, unpublished data) allowed the orientation of this contig with the AM10GA end closest to the centromere. Furthermore, D6S202 is present in a well-characterized radiation-reduced hybrid known to contain only markers telomeric to D6S89. The order CEN–D6S109–D6S89–F13A–TEL was reported elsewhere (NIH/CEPH Collaborative Mapping Group 1992). Multipoint linkage analyses in the current study have expanded this regional map to include LR40 and D6S202 and have positioned D6S109 with respect to D6S88, as follows: CEN–D6S88–D6S109–D6S89–LR40–D6S202–F13A–TEL. Marker-marker genetic distances were calculated from the CEPH family data as part of this study.

Analysis of recombination events within the SCA1 kindreds supports the order of markers determined genetically in CEPH. The marker D6S89 shows only a single recombination with SCA1 in the nine SCA1 kindreds studied, as shown in figure 2. The new marker SB1 demonstrates the same single recombinant, individual MI-2. This individual is nonrecombinant, however, at AM10GA. This indicates an order of AM10GA–D6S89–SB1, an order further supported by physical mapping data (S. Banfi, M.-y. Chung, T. J. Kwiatkowski, Jr., L. P. W. Ranum, A. E. McCall, A. C. Chinault, H. T. Orr, and H. Y. Zoghbi, unpublished data). Ordering of these three markers using CEPH family data was not possible, as no recombination was observed among them. LR40 demonstrates an additional telomeric recombinant with SCA1 (individual TX-5), and D6S202 demonstrates three telomeric recombinants beyond LR40 (individuals TX-2, TX-3, and TX-4). Individual MI-2 is uninformative at LR40 and D6S202 (fig. 2). Another new marker, AM10GA, demonstrated no recombination with SCA1. The individual (MI-2) who shows the recombination event with D6S89 is informative at AM10GA and inherits from his affected parent the allele associated with SCA1. This individual has also inherited the HLA and D6S109 alleles associated with the disease. Six recombinants were observed between SCA1 and D6S109, which is centromeric to the disease locus. These events, along with CEPH linkage and physical mapping data, support the order D6S109–AM10GA/SCA1–D6S89–SB1–LR40–D6S202. In order to rule out sample mix-up, the critical recombinant individual MI-2 was posthumously regenotyped at D6S109, D6S89, D6S202, LR40, AM10GA, and SB1, by PCR of a hair sample, and the above findings were confirmed. Thus, this single recombinant supports a centromeric location for SCA1 relative to D6S89 (Ranum et al. 1991*b*). The identification of a new marker, AM10GA, allowed us to cross over the recombination event in individual MI-2 and to orient the walk toward the SCA1 gene.

Pairwise linkage analysis and evaluation of recombination events by using SCA1 and six highly informative dinucleotide repeats from chromosome 6p in nine SCA1 families has placed SCA1 very close to AM10GA, in the 7-cM interval defined by D6S109 and D6S89. The $\hat{\theta}$ values with SCA1 are .004 and 0 for D6S89 and AM10GA, respectively, with a $Z_{max} - 1$ support interval ≤ 0.02 in both cases.

Further dinucleotide repeat cloning efforts are under way in order to identify a single recombination event

which will establish a centromeric flanking marker very close to SCA1. D6S109 lies centromeric to D6S89, but at a $\hat{\theta}$ of .04 with SCA1 (table 5). Meanwhile, YACs extending centromeric from D6S89 will be very valuable in the search for candidates for the SCA1 gene.

Acknowledgments

We thank the SCA1 families for their interest and cooperation, without which this study would not have been possible. We thank Dr. Robert Currier of the University of Mississippi School of Medicine for facilitating the study of the MS-SCA1 family and for many helpful discussions. We thank Dr. Lawrence J. Schut of the University of Minnesota and the United Hospital (St. Paul) for his valuable contribution of the clinical material for the MN- and MI-SCA1 kindreds. We thank Dr. Stephen Rich of the University of Minnesota and Drs. Jurg Ott and Chantal Mérette of Columbia University for many helpful discussions and suggestions regarding the linkage analyses. We are grateful to the staff of the Baylor Nucleic Acids core (Dr. R. Gibbs, director) for sequencing and oligonucleotide synthesis. This work was supported by National Institutes of Health grants RO1 NS27699 (to H.Y.Z.) and NS22920 (to H.T.O.), by grants from the Muscular Dystrophy Association (to H.T.O., L.P.W.R., and H.Y.Z.), by a grant from the National Ataxia Foundation (to H.Y.Z.), by Mental Retardation Research Center grants 5P30HD024064, and by National Center for Human Genome Research grant HG00210 (to the Baylor College of Medicine). Portions of this work were supported by Consiglio Nazionale delle Richerche, P.F. "Biotecnologie e' Biostrumentazioni" contract 92.00126.PF70 and by P.F. "Ingegneria Genetica" (to M.F.). A.L.B. is an investigator of the Howard Hughes Medical Institute.

References

Breukel C, Wijnen J, Tops C, Klift Hv/d, Dauwerse H, Khan PM (1990) Vector-Alu PCR: a rapid step in mapping cosmids and YACs. Nucleic Acids Res 18:3097

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

Feener CA, Boyce FM, Kunkel LM (1991) Rapid detection of CA polymorphisms in cloned DNA: application to the 5′ region of the dystrophin gene. Am J Hum Genet 48:621–627

Feinberg AP, Vogelstein B (1984) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137:266–267

Haines JL, Schut LJ, Weitkamp LR (1984) Spinocerebellar ataxia in a large kindred: age at onset, reproduction, and genetic linkage studies. Neurology 34:1542–1548

Heery DM, Gannon F, Powell R (1990) A simple method for subcloning DNA fragments from gel slices. Trends Genet 6:173

Jackson JF, Currier RD, Terasaki PI, Morton NE (1977) Spinocerebellar ataxia and HLA linkage: risk prediction by HLA typing. N Engl J Med 296:1138–1141

Keats BJB, Pollack MS, McCall A, Wilensky MA, Ward LJ, Lu M, Zoghbi HY (1991) Tight linkage of the gene for spinocerebellar ataxia to D6S89 on the short arm of chromosome 6 in a kindred for which close linkage to both HLA and F13A1 is excluded. Am J Hum Genet 49:972–977

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446

LeBorgne-Demarquoy F, Kwiatkowski TJ, Zoghbi HY (1991) Two dinucleotide repeat polymorphisms at the D6S202 locus. Nucleic Acids Res 19:6060

Marchuck D, Drumm M, Saulino A, Collins FS (1990) Construction of T-vectors: a rapid and general system for direct cloning of unmodified PCR products. Nucleic Acids Res 19:1154

NIH/CEPH Collaborative Mapping Group (1992) A comprehensive genetic linkage map of the human genome. Science 258:67–86

Nino HE, Noreen HJ, Dubey DP (1980) A family with hereditary ataxia: HLA typing. Neurology 30:12–20

Ranum LPW, Chung M-y, Duvick LA, Zoghbi HY, Orr HT (1991*a*) Dinucleotide repeat polymorphism at the D6S109 locus. Nucleic Acids Res 19:1171

Ranum LPW, Duvick LA, Rich SS, Schut LJ, Litt M, Orr HT (1991*b*) Localization of the autosomal dominant HLA-linked spinocerebellar ataxia (SCA1) locus, in two kindreds, within an 8-cM subregion of chromosome 6p. Am J Hum Genet 49:31–41

Spadaro M, Giunti P, Lulli P, Frontali M, Jodice C, Cappellacci S, Morellini M, et al (1992) HLA-linked spinocerebellar ataxia: a clinical and genetic study of large Italian kindreds. Acta Neurol Scand 85:257–265

Yakura H, Wakisaka A, Fujimoto S, Itakura K (1974) Hereditary ataxia and HLA genotypes. N Engl J Med 291:154–155

Zoghbi HY, Jodice C, Sandkuijl LA, Kwiatkowski TJ Jr, McCall AE, Huntoon SA, Lulli P, et al (1991) The gene for autosomal dominant spinocerebellar ataxia (SCA1) maps telomeric to HLA complex and is closely linked to the D6S89 locus in three large kindreds. Am J Hum Genet 49:23–30

Zoghbi HY, Pollack MS, Lyons LA, Ferrell RE, Daiger SP, Beaudet AL (1988) Spinocerebellar ataxia: variable age of onset and linkage to human leukocyte antigen in a large kindred. Ann Neurol 23:580–584

Zoghbi HY, Sandkuyl LA, Ott J, Daiger SP, Pollack M, O'Brien WE, Beaudet AL (1989) Assignment of autosomal dominant spinocerebellar ataxia (SCA1) centromeric to the HLA region on the short arm of chromosome 6, using multilocus linkage analysis. Am J Hum Genet 44:255–263