Am. J. Hum. Genet. 55:1218–1229, 1994

Association between Ag1-CA Alleles and Severity of Autosomal Recessive Proximal Spinal Muscular Atrophy

Christine J. DiDonato,[1,2] Kenneth Morgan,[4] John D. Carpten,[1,2] Paul Fuerst,[1] Susan E. Ingraham,[3] Gary Prescott,[5] John D. McPherson,[6] Brunhilde Wirth,[7] Klaus Zerres,[7] Orest Hurko,[8] John J. Wasmuth,[6] Jerry R. Mendell,[2] Arthur H. M. Burghes,[1,2,3] and Louise R. Simard[5]

Departments of [1]Molecular Genetics, [2]Neurology, and [3]Medical Biochemistry, Ohio State University, Columbus; [4]Departments of Epidemiology and Biostatistics, Human Genetics, and Medicine, McGill University, and [5]Génétique Médicale, Hôpital Sainte-Justine, Montreal; [6]Department of Biochemistry, University of California, Irvine; [7]Institute for Human Genetics, Bonn; and [8]Department of Neurology, Johns Hopkins Medical School, Baltimore

Summary

The gene for autosomal recessive proximal spinal muscular atrophy (SMA) has been mapped to an 850-kb interval on 5q11.2-q13.3, between the centromeric D5S823 and telomeric D5S557 markers. We report a new complex marker, Ag1-CA, that lies in this interval, whose primers produce one, two, or rarely three amplification-fragment-length variants (AFLVs) per allele. Class I chromosomes are those which amplify a single AFLV allele, and class II chromosomes are those which amplify an allele with two or three AFLVs. Ag1-CA shows highly significant allelic association with type I SMA in both the French Canadian (Hôpital Sainte-Justine [HSJ]) and American (Ohio State University [OSU]) populations ($P<.0001$). Significant association between the Ag1-CA genotype and disease severity was also observed. Type I patients were predominantly homozygous for class I chromosomes ($P=.0003$ OSU; $P=.0012$ HSJ), whereas the majority of type II patients were heterozygous for class I and II chromosomes ($P=.0014$ OSU; $P=.001$ HSJ). There was no significant difference in Ag1-CA genotype frequencies between type III patients ($P=.5$ OSU; $P=.25$ HSJ) and the paired normal chromosomes from both carrier parents. Our results indicate that Ag1-CA is the most closely linked marker to SMA and defines the critical candidate-gene region. Finally, we have proposed a model that should be taken into consideration when screening candidate SMA genes.

Introduction

Proximal spinal muscular atrophy (SMA) is an autosomal recessive disorder resulting in loss of α-motor neurons in the spinal cord. Affected individuals are classified into three groups, depending on the age at onset and progression of the disease. Children with acute type I SMA are the most severely affected. They have onset of symptoms prior to 6 mo, are never able to sit, and rarely live beyond the age of 2 years. Types II and III (chronic) SMA are milder forms and show onset of symptoms between 6 mo and 17 years.

SMA affects 1 in 10,000 live births, representing an overall carrier frequency of 1:40 (Pearn 1980). Although the primary defect responsible for SMA is unknown, all three forms of autosomal recessive proximal SMA have been mapped to 5q11.2-q13.3 by linkage analysis (Brzustowicz et al. 1990; Gilliam et al. 1990; Melki et al. 1990*a*, 1990*b*; Simard et al. 1992; MacKenzie et al. 1993; Wirth et al. 1993, 1994; Burghes et al. 1994*a*). Recently, a physical map of the region was assembled using a contiguous set of YAC clones (Francis et al. 1993; Kleyn et al. 1993; Carpten et al., in press) and radiation hybrids (Thompson et al. 1993). Although recombinants have localized the SMA gene to an interval of ~850 kb (Francis et al. 1993; B. Wirth, unpublished data), the isolation of the gene has been hampered by complex genomic structures found in this region. These include chromosome 5–specific repetitive loci (Francis et al. 1993; Thompson et al. 1993), multicopy sequences unique to 5q11.2-q13.3 that are represented multiple times (Brahe et al. 1994; Burghes et al. 1994*b*), and a high incidence of pseudogenes (authors' unpublished observation). Owing to the overall complexity of this region and the high incidence of pseudogenes, it is imperative to further refine the SMA candidate-gene interval. Simard et al. (1994) have shown the presence of linkage disequilibrium in the French Canadian population. Recently, we have reported a multicopy marker, CATT-1, which shows allelic association in the American and Canadian populations (Burghes et al. 1994*b*).

In this paper, we present data from a new multicopy marker, Ag1-CA, which lies between D5S823 and D5S557. It demonstrates highly significant allelic association with SMA in both the American and French Canadian populations and acts to core two consensus haplotypes observed in French Canadian families. Given the strong allelic asso-

Received June 24, 1994; accepted for publication August 22, 1994.

Address for correspondence and reprints: Dr. Arthur H. M. Burghes, Department of Neurology, 452 Means Hall, 1654 Upham Drive, Ohio State University, Columbus, OH 43210.

0002-9297/94/5506-0020$02.00

ciation between SMA and Ag1-CA, as well as the significant segregation of Ag1-CA class genotypes with respect to clinical severity, the Ag1-CA marker identifies the critical region that should be searched for candidate cDNAs of the SMA gene.

Families, Material, and Methods

Families

All patients fulfilled the diagnostic criteria for proximal SMAs defined by the international SMA consortium (Munsat 1991, 1992). DNA samples from families of SMA patients were analyzed in the laboratories of the Clinical Neuroscience Center, Ohio State University (OSU), and the Service de Génétique Médicale, Hôpital Sainte-Justine (HSJ). The HSJ families are, with two exceptions, of French Canadian origin. The OSU families attended a number of different U.S. clinics, and in addition there were families from Hungary. Two families from Germany, previously reported by Wirth et al. (1993, 1994), were included. Although the ethnic origins of the OSU families were not precisely defined, the OSU sample was more heterogeneous than was the HSJ sample. We do not know of any close kinship between the OSU and HSJ families.

Sibships were classified as type I, II, or III SMA. Where there were siblings of two types, the sibship was classified with the more severe type. The HSJ families consisted of 37 sibships from 36 families. All but two families were French Canadian. One sibship had type I and II SMA siblings and was classified as type I. One sibship had type II and III SMA siblings and was classified as type II. The extended family had a patient with type II SMA and an aunt and uncle with type III SMA; each sibship was separately classified. The OSU sample consisted of 58 sibships from 55 families. The three extended families consisted of two pairs of first-cousin sibships and one pair of second-cousins once-removed sibships. We classified one sibship with type I and II siblings as type I; one sibship with types II and III siblings as type II; two sibships that were either type II or type III as type II; and one sibship with type III and type IV siblings as type III SMA.

Somatic and Radiation Hybrid Lines

The somatic cell hybrid HHW105, which contains an intact chromosome 5 (Dana and Wasmuth 1982) as its only human material in a Chinese hamster background, was obtained from the National Institute of General Medical Sciences (NIGMS) Human Genetic Mutant Cell Repository (Corriel Institute for Medical Research) and grown in α-minimal essential medium (αMEM) that had been supplemented with proline as described by the supplier. The cell line HHW1064 (Gilliam et al. 1989), which contains a human chromosome 5 deleted for 5q11.2-q13.3 in a Chinese hamster background, was grown under the same conditions as HHW105. A Chinese hamster cell line was also obtained from NIGMS and grown in αMEM according to the supplier's instructions. The radiation hybrid, XZH28, which has been previously described, contains as its sole human content ~2 Mb of DNA from the SMA region (Thompson et al. 1993).

Preparation and Isolation of Sequence-tagged Sites (STSs) between D5S435 and D5S351

To amplify the human DNA from XZH28 and subsequently clone it, Alu-vectorette PCR was performed. XZH28 DNA (1–5 μg) was digested to completion with *Eco*RV, *Hin*cII, and *Rsa*I (BRL) in separate reactions with a total of 10 U of restriction enzyme each. Vectorette adapters (Riley et al. 1990) were ligated to the digested XZH28 DNA by standard methods at a 20:1 molar ratio for adapter:DNA. The subsequent vectorette libraries of XZH28 were diluted to 5–10 ng/μl and used as template for Alu-vectorette PCR. PCR was performed on the XZH28 vectorette libraries by using Alu559 (Ledbetter et al. 1990) and the universal vectorette primer, UVP224 (Riley et al. 1990). Each PCR reaction was carried out in a total volume of 50 μl, using 5–10 ng of vectorette library DNA; 6.7 mM $MgCl_2$; 67 mM Tris-HCl pH 8.0; 16.6 mM $(NH_4)_2SO_4$; 28 mM β-mercaptoethanol; 10% dimethylsulfoxide (v/v); 0.3 mM each of dATP, dCTP, dGTP, and dTTP; and 2.5 U of *Taq* DNA polymerase (Perkin Elmer Cetus). After an initial denaturation of 3 min at 94°C, 35 cycles of 94°C for 45 s, 55°C for 45 s, and 72°C for 3 minutes were performed with a final 7-min extension at 72°C in a Perkin Elmer Cetus thermal cycler. An aliquot was analyzed on a 2% agarose gel.

The Alu-vectorette PCR products were purified using a Promega PCR prep column according to the manufacturer's instructions. Twenty micrograms of glycogen (Boehringer Mannheim) were added to the eluate, and the samples were then ethanol precipitated. The products were resuspended in 10 mM Tris-HCl/1 mM EDTA (pH 8.0) buffer at a concentration of ~100 ng/μl. One hundred nanograms of the Alu-vectorette PCR products was ligated to 25 ng of the T-tailed vector, pT7 (Novagen), using T4 ligase and the manufacturer's conditions. Ligations were electroporated into electrocompetent D5Hα cells at 2.5 mV, 200 Ohms, and 2.5 μFD. The colonies obtained were master plated in duplicate. One plate was lifted onto Hybond N^+ (Amersham) and was hybridized with the UVP224 primer which had been end labeled with [γ-^{32}P]ATP by using a standard T4 polynucleotide kinase (USB) reaction (Sambrook et al. 1989) to identify clones that contained inserts. The hybridization was carried out using Rapid Hyb (Amersham) according to the manufacturer's protocol. Positive clones were amplified by PCR using the Alu559 and UVP224 primers to determine the size of the inserts. Clones that contained inserts >120 bp

were sequenced using the dsDNA Cycle Sequencing System (BRL) with miniprep DNA that had been purified using Promega miniprep columns. Products of the sequencing reactions were analyzed on 6% denaturing polyacrylamide gels. Primer sets were designed and then synthesized using a PCR mate 1000 DNA synthesizer (Beckman).

One of the STSs identified was Ag1, which mapped back to XZH28. The primer set for Ag1 is (forward, 5′ ACT gCA CTC CAT gCA CTC AgC; reverse, 5′ TCC TTg TAA CCT CAg AgT gTg T), and the PCR conditions included an initial denaturation of 3 min at 94°C followed by 35 cycles at 94°C for 30 s, at 58°C for 20 s, and at 72°C for 20 s. The product size of Ag1 was 128 bp.

Isolation of an Ag1 Cosmid

The chromosome 5 cosmid library (Los Alamos) was replica plated into microtiter plates containing Luria broth (Sambrook et al. 1989) and kanamycin (25 μg/ml) and grown overnight. Each plate was then converted into DNA by using a modified alkaline lysis miniprep (Sambrook et al. 1989). To identify genomic clone(s) from Ag1, each plate pool was screened with the Ag1 STS. After identification of the positive plate(s), row and column DNA from that microtiter plate(s) was prepared in order to identify the coordinates of the positive cosmid (108H11). The Ag1 cosmid was digested with *Sau*3AI (BRL) and was ligated into dephosphorylated *Bam*HI-digested Puc18, using T4 ligase (Boehringer-Mannheim). The ligation mixture was transformed into DH5α competent cells (BRL) and the resulting colonies were master plated in duplicate. One plate was lifted onto Hybond N^+ and screened with a $(CA)_9$ oligonucleotide probe that was end labeled as described above. The hybridization was carried out overnight at 42°C in 6 × SSC/0.5% Carnation milk/1% SDS and was washed sequentially at 42°C for 5–10 min in 2 × SSC/0.1% SDS, 1 × SSC/0.1% SDS, 0.5 × SSC/0.1% SDS, and 0.2 × SSC/0.1% SDS. Between buffer changes, filters were monitored and compared with positive and negative controls. Positive clones were picked and sequenced (BRL dsDNA cycle sequencing). The subclone, Ag1-CA, contains a $(CA)_{21}$ dinucleotide repeat.

PCR of Ag1-CA and Other Polymorphic Primer Sets

Primers that flank the Ag1-CA repeat sequence were designed (forward, 5′ ATT TAC TTT TCC AAg ggg gAg g; reverse, 5′ CAT gTT gCT TAg gCC TCg TCT) and were used for linkage analysis. This was done by first end-labeling the forward primer. The PCR was carried out in a volume of 25 μl that contained 0.3 mM dNTPs, 50 mM KCl (pH 9.0), 10 mM Tris-HCl (pH 8.3), 2.5 mM $MgCl_2$, 0.01% triton X 100 (v/v), 0.01% tween-20 (v/v), 0.01% gelatin (w/v), 1 U of *Taq* DNA polymerase, 50–100 ng of genomic DNA, 50 ng [$\gamma-^{32}P$] end-labeled forward primer, and 50 ng reverse primer. Cycling conditions included an initial denaturation of 3 min at 94°, followed by 35 cycles of 30 s at 94°C, 45 s at 65°C, 1 min at 72°C, and a final extension of 7 min at 72°C. The PCR products were denatured by adding one volume of formamide stop mix and heating for 3 min at 90°C. The samples were loaded onto an 8% denaturing polyacrylamide gel and electrophoresed at 1,600 V for ~4 h. Best resolution was obtained using a wide-tooth (8 mm) comb. The conditions for the markers D5S125 (Mankoo et al. 1991), D5S435 (Soares et al. 1993), D5S557 (Francis et al. 1993), D5S351 (Hudson et al. 1992), MAP1B-3′ (Lien et al. 1991), and JK53 (Morrison et al. 1992) have been described previously.

Linkage Analysis and Genotyping of Ag1-CA

The linkage analysis was performed using the LINKAGE 5.1 package (Lathrop et al. 1984). SMA was analyzed as a completely penetrant autosomal recessive disorder with a gene frequency of .0063 (Brzustowicz et al. 1990; Burghes et al. 1994*a*). LOD scores were calculated using the MLINK program.

The primer set Ag1-CA amplifies one, two, or rarely three PCR fragments simultaneously. For clarity, we have designated these PCR fragments as "amplification-fragment-length variants" (AFLVs). Ag1-CA alleles are composed of one, two, or three AFLVs, as determined after segregation analysis. "Genotype" refers to an individual's alleles at the Ag1-CA locus, while "haplotype" refers to all markers in the SMA region. This has been done to eliminate confusion of terminology in the extended haplotype analysis of French Canadian families. One or more AFLVs comprise the alleles of the highly polymorphic Ag1-CA locus. The AFLVs lie close together, as the Ag1-positive cosmid contains two AFLVs. Alleles were assigned independently in the two laboratories by inspection of segregation in the families, taking into account any differences in signal intensity of the AFLVs for evidence of variation in copy number. Fifteen AFLVs were observed of sizes 90–122 bp (table 1). Ag1-CA alleles with one AFLV per chromosome were classified as class I alleles. Alleles with two or three AFLVs per chromosome were classified as class II alleles. Alternative Ag1-CA genotypes were possible for two HSJ and seven OSU families (e.g., see fig. 1*B*). Since alleles with three AFLVs were rarely observed, we only considered class I and class II alleles with two AFLVs in the instances of alternative genotypes. Three alleles in HSJ and five alleles in OSU normal chromosomes were unknown because of missing parental data. In total, 58 different Ag1-CA alleles were identified.

Statistical Analysis

Some of the "observed" numbers of certain Ag1-CA alleles and genotypes were fractional counts as a consequence of averaging frequencies of alternative genotypes. The Pearson χ^2 statistic was used to test homogeneity of allele-class genotypes between SMA and normal chromosomes for each of

Table 1

Counts of AFLVs in Both the OSU and HSJ Populations

AFLV	Size (bp)	OSU				HSJ			
		Normal	Acute SMA	Chronic SMA	Total SMA	Normal	Acute SMA	Chronic SMA	Total SMA
1	122	3	0	0	0	0	0	0	0
2	120	2	0	0	0	0	0	0	0
3	118	7	2	1	3	3	2	2	4
4	116	10	4	2	6	6	2	4	6
5	114	25	3	7	10	22	0	6	6
6	112	16	1	5	5	11	1	2	3
7	110	19	9	14	23	15	2	7	9
8	108	37	8	19	26	27	5	15	20
9	106	14	5	3	8	14	5	9	14
10	104	2	3	6	9	3	0	4	4
11	102	9	6	7	12	6	0	1	1
12	100	11	15	8	23	10	13	18	31
13	98	1	0	0	0	0	0	0	0
14	96	0	2	0	2	0	0	0	0
15	90	1	0	0	0	0	0	0	0
Total AFLVs		157	58	72	127	117	30	68	98
Total chromosomes		95	51	48	99	67	26	43	69

NOTE.—The AFLVs from 2 HSJ and 5 OSU normal chromosomes were excluded because of missing parental data. In addition, 2 HSJ and 7 OSU families were excluded, as phase could not definitively be assigned to SMA or normal chromosomes. Note that the total number of AFVLs differs from the total number of chromosomes, since more than one AFLV can occur on a chromosome.

the three clinical types of SMA. In type III HSJ families, the Ag1-CA allele-class genotypes 1/1 and 1/2 were combined into one category. The small sample distribution of χ^2 may be reasonably close to the asymptotic χ^2 approximation under the null hypothesis (Fienberg 1980).

Many Ag1-CA alleles were infrequent and contingency tables of SMA and normal chromosomes for each type of SMA were sparse, with the number of alleles ranging from 20 to 38. A χ^2 test for small expected values proposed by Smith (1986) was used for tests of homogeneity of allele frequencies. The test is a normal approximation to the distribution of χ^2 based on the exact mean and variance of χ^2 under the null hypothesis. We used the second test statistic of Smith (1986), which is based on $\sqrt{\chi^2}$, to test for the presence of overall allelic association. To identify Ag1-CA alleles that were significantly more frequent than expected on SMA chromosomes, we used a sequential procedure and the Freeman-Tukey deviate as the test criterion. The Freeman-Tukey deviate is $\sqrt{x_i} + \sqrt{x_i + 1} - \sqrt{4m_i + 1}$, where x_i is the observed frequency and m_i is the expected frequency. The Freeman-Tukey deviate is approximately normally distributed with mean 0 and variance 1, where x_i follows a Poisson distribution with mean m_i (Bishop et al. 1975). We considered a large value of the Freeman-Tukey deviate for an SMA Ag1-CA allele to be >1.65, nominally a probability of <.05. The sequential test procedure was as follows. First, the Ag1-CA allele with the largest Freeman-Tukey deviate was identified by fitting the log-linear model of independence to the contingency table of alleles from families of a specific SMA type. If the value of the deviate met the criterion, then that cell was treated as a structural zero and the resulting incomplete contingency table was fit using the log-linear model of quasi independence (Bishop et al. 1975; Fienberg 1980). The next SMA Ag1-CA allele with the largest deviate was identified, and if the value met the criterion then the relative excess of that allele was considered to be significant, the cell was treated as a structural zero, and the resulting incomplete contingency table (now with two zeroes) was fit using the log-linear model of quasi independence. This procedure was terminated at any step in which the largest value of the Freeman-Tukey deviate of the remaining Ag1-CA alleles did not meet the criterion. In only one case, that of the analysis of type I SMA in OSU, was it necessary to fit an incomplete contingency table using the model of quasi independence. This was because all other Ag1-CA alleles, judged to be significantly in excess in SMA chromosomes of a particular group of families, were not observed in the normal chromosomes. Data analysis and graphics were carried out using the S language (Becker et al. 1988) as implemented in S-PLUS (Statistical Sciences 1992).

Results

Isolation and Physical Mapping of Ag1

The Ag1 STS marker was identified by Alu-vectorette PCR of radiation hybrid DNA (XZH28), which contained

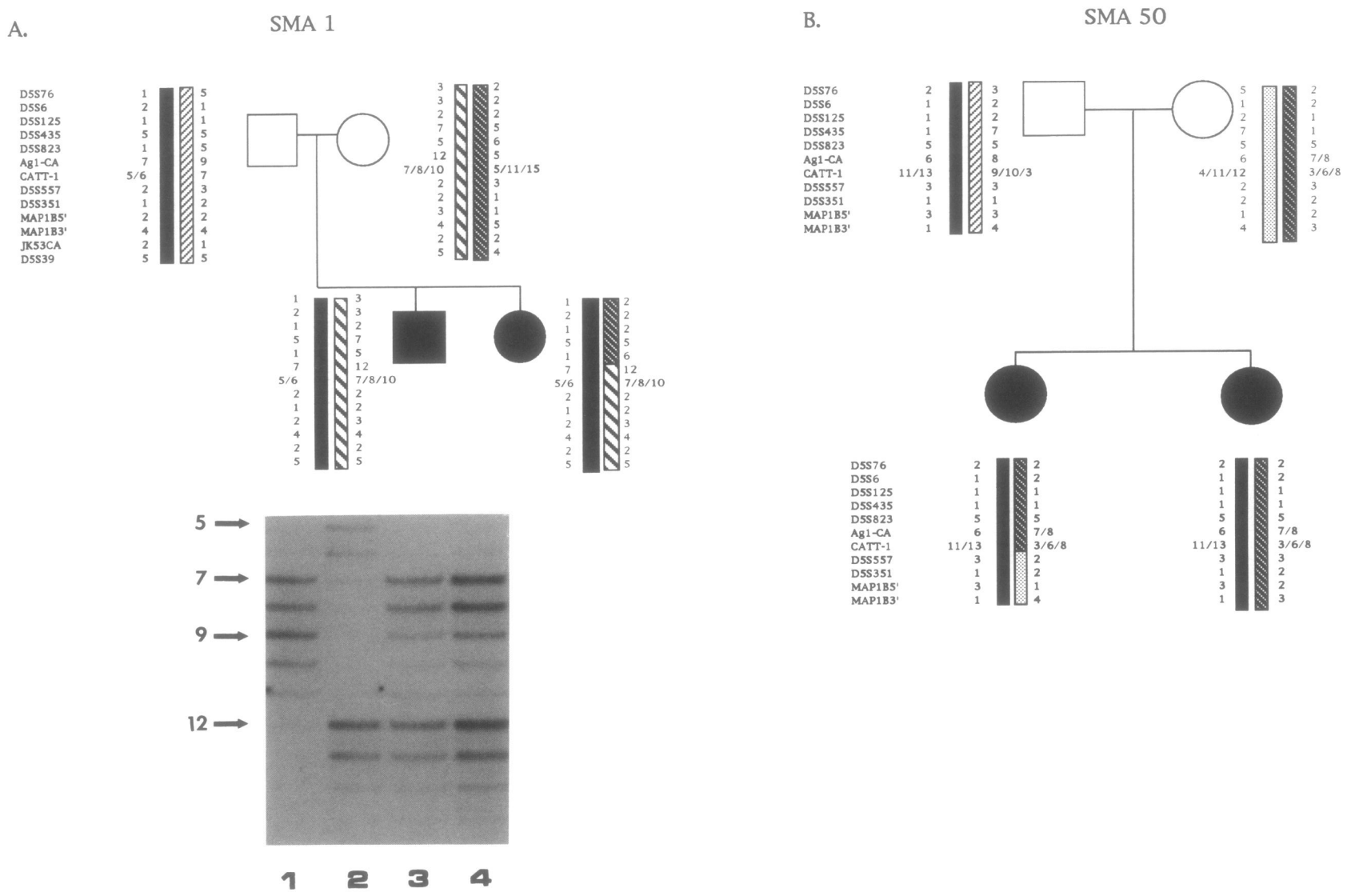

Figure 1 *A*, Family 1 (SMA1) is recombinant with the marker D5S823. This family is fully informative for the Ag1-CA locus. Lane 1, Father. Lane 2, Mother. Lane 3, Affected male. Lane 4, Affected female. Examples of an AFLV in the father are 7 and 9. This individual has two alleles that are each composed of single AFLVs (7 and 9). His genotype is 7 and 9, and his haplotype is all the markers in the SMA region. *B*, Family 50 (SMA50) is recombinant with the marker D5S557. This is an example of a family in which the phase of Ag1-CA alleles could not be unambiguously assigned. There are three possible genotypes for this locus and each has been given an equal weight of .33. Only one genotype has been shown; the other two possibilities are either "6/8" (PN), "6/8" (PSMA), "6/8" (MN), and "7" (MSMA) or "6" (PN), "8" (PSMA), "8" (MN), and "6/7" (MSMA). Abbreviations are as follows: (PN) = paternal normal, (PSMA) = paternal SMA, (MN) = maternal normal, and (MSMA)= maternal SMA.

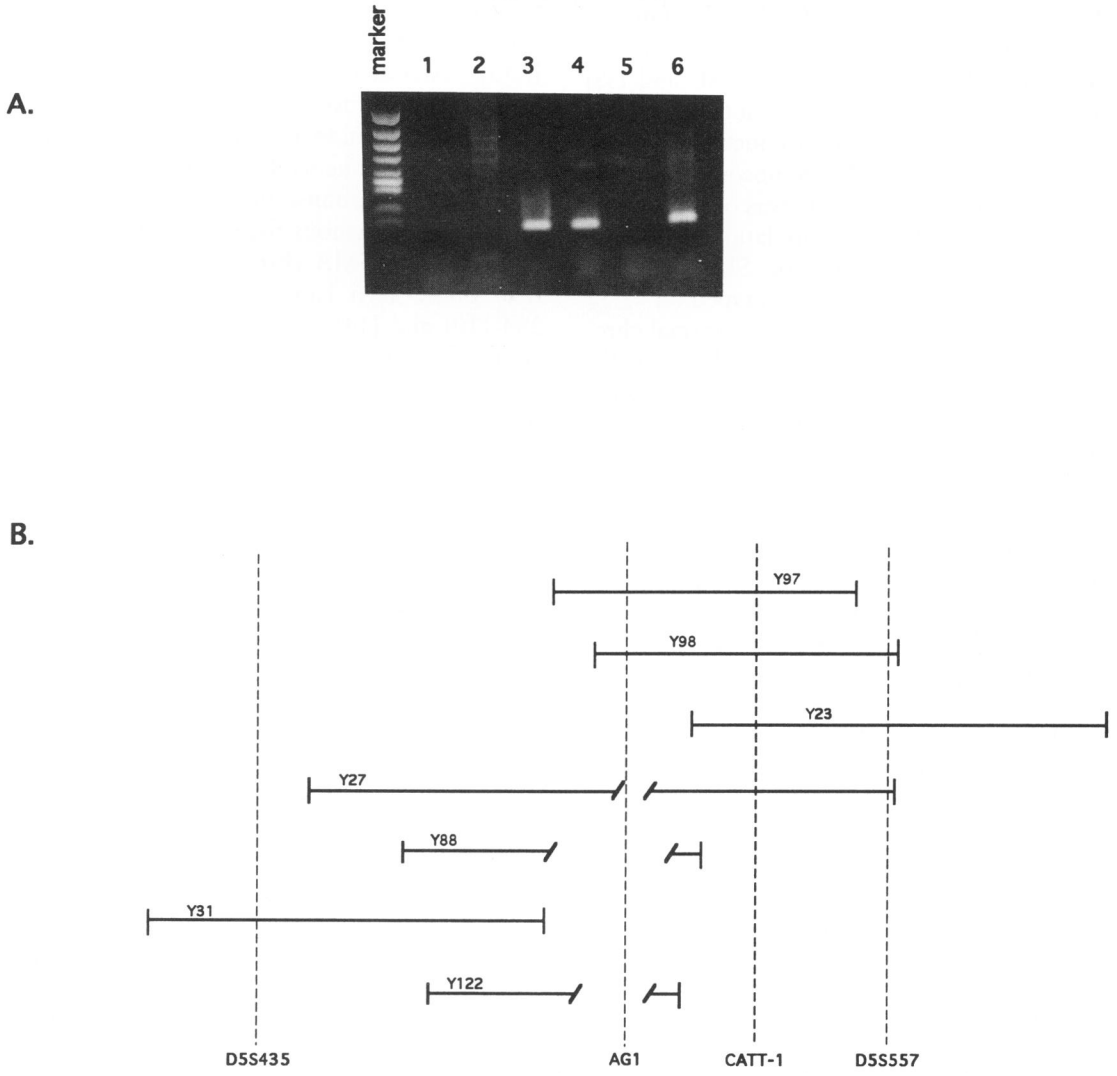

Figure 2 *A*, Mapping panel of Ag1. Lane 1, Negative control. Lane 2, Amplification of chinese hamster DNA. Lane 3, Amplification of human genomic DNA. Lane 4, Amplification of DNA from hybrid HHW105 containing only human chromosome 5. Lane 5, Amplification of DNA from hybrid HHW1064, which is deleted for 5q11.2-q13.3. Lane 6, Amplification of XZH28, which contains the SMA region as its only human DNA. *B*, Location of Ag1 with respect to YAC clones and nearby markers. Broken lines denote deletions.

the SMA region (~2 Mb of DNA) as its only human component. The STS was initially mapped using somatic cell hybrids. Ag1 was present in radiation hybrid XZH28 and the hybrid HHW105 but absent in hybrid HHW1064, indicating that it maps to the region containing the SMA gene (fig. 2*A*). The STS marker was also mapped in YACs. It was positive on YACs 97 and 98 and negative on YACs 88 and 122 in the contig reported by Kleyn et al. (1993). With the YACs reported by Carpten et al. (in press), Ag1 was negative on YACs 23, 27, and 31 but positive on all the other YACs extending centromeric of CATT-1. This indicates that the most likely physical position for Ag1 is within the deletion of YAC 27 (fig. 2*B*).

Genetic Analysis of Ag1-CA in SMA Families

Ag1-CA is a multicopy marker that has 15 AFLVs and 58 alleles. This marker may have one, two, or rarely three AFLVs per chromosome. As is clear from the data presented above, all copies map to the genomic region containing the SMA gene. Ag1-CA maps between D5S823 and D5S557, and no recombinants were observed between Ag1-CA and SMA in the OSU, HSJ, and German families (fig. 1). In the OSU population, Ag1-CA was linked to the SMA locus with a LOD score of 16.52 at $\theta = 0$. The observed counts of the various AFLVs in phase-known families are shown in table 1. Following segregation analysis, chromosomes were designated as being either class I or class II. Class I chromosomes contain an Ag1-CA allele with a single AFLV, whereas class II chromosomes have an allele that contains two or three AFLVs. The distribution of particular Ag1-CA alleles is summarized for both the OSU and HSJ populations in figure 3. Obligate heterozygote carriers and their affected children were used to assign

normal and SMA chromosomes; any fractional counts are due to equal weighting of alternative genotypes. The normal alleles represented by Ag1-CA in the OSU and HSJ families were generally the same, except that certain alleles predominated in one population. For instance, the "9/12" allele was present in >11% of HSJ chromosomes and only 5% of OSU chromosomes. Allele "14" was observed only in Hungarian families in the OSU population. However, allele "12" occurred predominantly on SMA chromosomes in both populations; this allele was typed on 41.5/190 SMA chromosomes and only on 2/182 normal chromosomes. Furthermore, allele "11," while absent in the HSJ population, was present on 10% of the OSU SMA chromosomes and 1% of the OSU normal chromosomes. Finally, alleles "7" and "8" were present on 20% SMA and only 9% normal chromosomes in the OSU population.

In light of obvious differences in the distribution of various alleles between SMA and normal chromosomes, we tested for allelic association between SMA and Ag1-CA, using a χ^2 heterogeneity test for small expected values that was proposed by Smith (1986). We observed significant allelic association between Ag1-CA and SMA for type I families in both patient populations (z=4.14, P<.0001 and z=6.45, P<.0001 in HSJ and OSU, respectively) and for type II families in the HSJ population (z=4.90, P<.0001). There was lack of significant association for type II families in the OSU population (z=1.08, P= .14) and for type III families in both patient populations (z=1.32, P=.09 and z=1.42, P=.08 in HSJ and OSU, respectively). To identify the SMA alleles that occurred significantly more frequently than expected, we used a sequential test procedure and a criterion based on the size of the Freeman-Tukey deviate. The SMA chromosomes carrying Ag1-CA allele "12" were significantly more frequent than expected in type I SMA families in both patient populations (Freeman-Tukey deviates 1.98 and 2.13 in HSJ and OSU, respectively) and in type II families in the HSJ population (Freeman-Tukey deviate 1.71). In addition, SMA chromosomes carrying either Ag1-CA allele "11" or "7" were significantly more frequent than expected in the OSU type I families (Freeman-Tukey deviates 1.67 and 1.75, respectively).

Finally, there were significant differences in the Ag1-CA allele-class genotype frequencies between SMA and normal chromosomes for type I and II, but not for type III SMA (table 2). The estimated class 1 allele frequencies showed a marked trend of decreasing frequency with decreasing severity in HSJ SMA chromosomes (.85, .68, and .18 in types I, II, and III, respectively). The frequencies of class I alleles in OSU SMA chromosomes were also greater in acute than in chronic SMA (.85, .44, and .54 in types I, II, and III SMA, respectively).

Extended Haplotype Analysis in the French Canadians

Owing to their common and recent origins, the French Canadian families were analyzed in detail with the markers [D5S125-D5S435-(Ag1-CA)-D5S557-D5S351-MAP1B-JK53]. Extended haplotype data were available for 63 SMA chromosomes and 61 normal chromosomes (table 3). As can be seen, there were a total of 110 unique haplotypes, and only one haplotype was common to both normal and SMA chromosomes. Overall, SMA chromosomes are more homogenous than are normal chromosomes, in that 28.6% of SMA chromosomes were observed more than once. Two haplotypes, [145-124-(9/12)-164-198-234-110] and [145-124-(12)-164-200-232-114], each accounted for 9.5% of the SMA chromosomes. This indicates that 19% of the SMA chromosomes may be ancestrally related to two different haplotypes.

In type III SMA, the extended haplotype containing the Ag1-CA "9/12" allele predominates (fig. 4). This haplotype was present on 10 SMA chromosomes. Of these, 6 were identical [145-124-(9/12)-164-198-234-110]. Four additional chromosomes have been included, as they may be related. SMA chromosome FC24 may have diverged from the consensus haplotype at JK53 as a result of a single mutation or recombination event. In contrast, a single recombination between Ag1-CA and D5S557 or two mutations, one at D5S557 and the other at MAP1B, could explain how SMA chromosomes FC8 and FC35 could have diverged from the consensus. If there was a single recombination, this event would place the SMA locus proximal to D5S557, consistent with the current genetic mapping data. Finally, if SMA chromosome FC16 is related, a minimum of two recombination events would be required, placing the SMA locus distal to D5S435. Family FC7 (two type III individuals) was homozygous for the "9/12" haplotype. Consequently, if there is a single SMA mutation segregating with this haplotype, it is most likely one that results in the least severe form of the disease.

Assembling the 16 SMA chromosomes carrying allele "12" at the Ag1-CA locus was more difficult. However, a subset of these do appear to be related to each other. Six SMA chromosomes from five families were identical. Family FC36, with a history of type I SMA, was homozygous for this extended haplotype. This observation suggests that, if these chromosomes carry an identical SMA mutation, then this haplotype segregates with type I SMA. SMA chromosome FC33 differs at a single site, D5S435. FC28 maps SMA proximal to D5S351. We could not establish any obvious relationship between the eight remaining SMA chromosomes carrying allele "12" and the consensus haplotype; these are listed in figure 4.

Discussion

Recombination analysis has mapped the SMA gene to an 850-kb interval between the markers D5S823 and D5S557 (Francis et al. 1993; B. Wirth, unpublished data). Recently, using the markers D5S435, D5S125, D5S351,

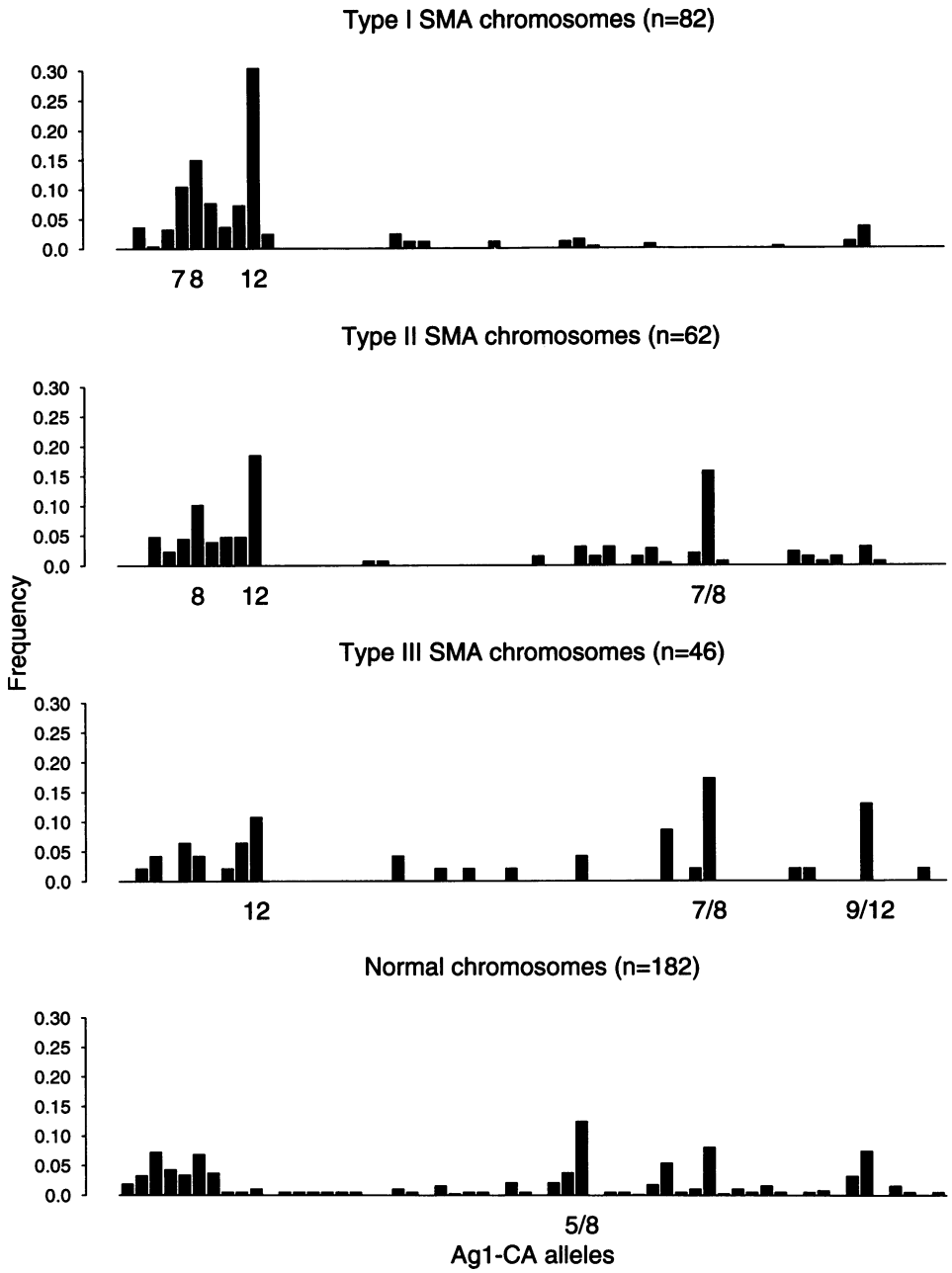

Figure 3 The combined distribution of Ag1-CA alleles in both the OSU and HSJ population. Ag1-CA alleles and their order in the histograms of allele frequencies in pooled data are as follows: "3," "4," "5," "6," "7," "8," "9," "10," "11," "12," "14," "15," "1/3," "1/5," "1/8," "2/3," "2/4," "2/8," "2/12," "3/4," "3/5," "3/7," "3/8," "3/12," "4/5," "4/6," "4/7," "4/8," "4/9," "4/10," "5/6," "5/7," "5/8," "5/9," "5/10," "5/11," "5/12," "6/7," "6/8," "6/11," "6/12," "7/8," "7/9," "7/10," "7/11," "7/12," "8/9," "8/10," "8/11," "8/12," "9/10," "9/11," "9/12," "10/12," "11/12," "11/13," "6/7/11," "9/11/12."

and JK53 in 22 French Canadian families, Simard et al. (1994) reported statistically significant linkage disequilibrium. In addition, linkage disequilibrium has also been observed with the marker D5S435 in type I German SMA families (B. Wirth, unpublished data). In both instances, linkage analysis indicated that certain haplotypes occurred only on SMA chromosomes.

The 850-kb region that contains the SMA gene is unusual in that it contains a number of multicopy markers (Kleyn et al. 1993; Brahe et al. 1994; Burghes et al. 1994*b*). We have extensively characterized one of these, CATT-1 (Burghes et al. 1994), and were able to demonstrate significant allelic association between CATT-1 and SMA in the American and Canadian SMA populations. In this pa-

Table 2

Ag1-CA Allele-Class Genotype Frequencies According to Type of SMA

Ag1-CA Allele-Class Genotype	Type I[a]				Type II[b]				Type III[c]			
	SMA		Normal		SMA		Normal		SMA		Normal	
	HSJ Families											
1/1	9	{.69}	1	{.08}	4.67	{.36}	1	{.08}	0		.5	{.06}
1/2	4	{.31}	5	{.38}	8.33	{.64}	3	{.23}	4	{.36}	6	{.67}
2/2	0		7	{.54}	0		9	{.69}	7	{.64}	2.5	{.28}
Total	13	{1.0}	13	{1.0}	13	{1.0}	13	{1.0}	11	{1.0}	9	{1.01}
	OSU Families											
1/1	20.67	{.74}	4.67	{.19}	0		2.67	{.16}	2	{.17}	1	{.08}
1/2	6	{.21}	11	{.46}	16	{.89}	5	{.29}	9	{.75}	8	{.67}
2/2	1.33	{.05}	8.33	{.35}	2	{.11}	9.33	{.55}	1	{.08}	3	{.25}
Total	28	{1.0}	24	{1.0}	18	{1.0}	17	{1.0}	12	{1.0}	12	{1.0}

NOTE.—The SMA genotype is that of one patient in a family, and the normal genotype consists of the normal chromosomes of the parents. Curly braces ({ }) indicate the relative frequencies of the Ag1-CA genotypes, by SMA genotype. The results of χ^2 tests of the difference in the distribution of Ag1-CA allele-class genotypes between SMA and normal genotypes are given in footnotes a–c below.

[a] HSJ families—$\chi^2 = 13.51$, df = 2, $P = .0012$; OSU families—$\chi^2 = 16.43$, df = 2, $P = .0003$.

[b] HSJ families—$\chi^2 = 13.88$, df = 2, $P = .001$; OSU families—$\chi^2 = 13.16$, df = 2, $P = .0014$.

[c] HSJ families—$\chi^2 = 1.32$, df = 1, $P = .25$ (Ag1-CA genotypes 1/1 and 1/2 were combined); OSU families—$\chi^2 = 1.39$, df = 2, $P = .5$.

per, we describe the identification and genetic analysis of a new multicopy marker, Ag1-CA, which lies centromeric of CATT-1. Both Ag1-CA and STS 98T (Kleyn et al. 1993) map to a single P1 clone (authors' unpublished data). Ag1-CA—having one, two, or rarely three AFLVs per allele—is less complex than CATT-1. We have demonstrated highly significant association between Ag1-CA and acute SMA and provide evidence for a correlation between the Ag1-CA genotype and the severity of the disorder. Allele "12," which is significantly associated with SMA, was typed on 22% SMA chromosomes and on only 1% of normal chromosomes, corresponding to frequencies of .218 and .005, respectively. Interestingly, this association was significant in type I SMA of both populations and type II SMA in the HSJ population and was not significant in either of the type III SMA populations. With respect to Ag1-CA genotypes, 72% of type I SMA patients (compared to 15% normals) were class I homozygotes. In type II SMA, 78% of SMA patients (compared to 27% normals) were class I/class II heterozygotes. There was no significant difference in the Ag1-CA genotype between type III SMA and normal individuals. Overall, our observations indicate that data

Table 3

Extended Haplotypes Observed in the French Canadian Population

Haplotype D5S125-D5S435-(Ag1-CA)-D5S557-D5S351-MAP1B-JK53	Normal Chromosomes		SMA Chromosomes	
	No.	%	No.	%
145-124-(9/12)-164-198-234-110	0	0	6	9.5
145-124-(12)-164-200-232-114	0	0	6	9.5
141-140-(3/4)-158-198-232-104	1	1.6	2	3.2
141-136-(8)-158-196-232-112	0	0	2	3.2
145-136-(7/8)-164-198-232-112	0	0	2	3.2
105 other haplotypes	60	98.4	45	71.4
Total	61	100	63	100

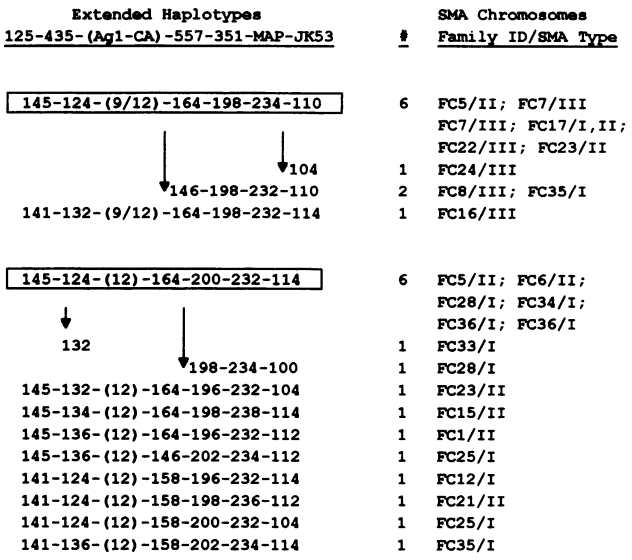

Figure 4 Two consensus haplotypes identified in the French Canadian population.

should be rigorously analyzed when estimating allelic association between SMA and closely linked loci, since association appears to be influenced by disease severity as well as ethnic origin.

Extended haplotype analysis in the French Canadian population indicated that the majority of SMA patients were likely compound heterozygotes for two different SMA mutations, a factor that probably plays a role in the observed clinical heterogeneity. Both the fact that >19% of these haplotypes may be related to two ancestral SMA chromosomes and the identification of two SMA families each homozygous for a different consensus haplotype allow us to predict the relationship between these particular SMA chromosomes and disease severity. When the [145-124-(9/12)-164-198-234-110] consensus haplotype was inherited from both parents (FC7), the affected individual was diagnosed with type III SMA. This implies that, if this haplotype marks a single SMA mutation, this chromosome is likely associated with a mild SMA mutant allele. In contrast, the homozygous individual (FC36) for the consensus haplotype [145-124-(12)-164-200-232-114] was diagnosed with type I SMA, suggesting that this chromosome marks a severe SMA mutation. However, the inheritance of the severe SMA chromosome with the mild SMA chromosome resulted in an intermediate phenotype (FC5, type II; FC23, type II). Consequently, these mutations appear to show incomplete dominance. The exception was family FC35, a type I SMA pedigree. Given the difference between the inherited SMA chromosome with respect to the consensus "12" haplotype, it may be that the SMA mutation on this chromosome is different from that carried on the consensus haplotype. In agreement with this possibility, one of the grandparents in this pedigree was of German descent. These observations support the claim that, while some predictions can be made, until the exact mutation is identified, the effects of intraallelic and interallelic complementation remain unclear. Finally, the lack of an obvious consensus haplotype in the OSU population, coupled with the significant association of allele "12" in both populations, suggests that SMA chromosomes carrying this allele are likely to be old (before the colonization of America). The observation of two consensus haplotypes in the French Canadian population likely reflects genetic drift among regions of Quebec, consistent with this population's early history (Simard et al. 1994).

Overall, the data presented in this paper are consistent with the association of a severe SMA mutation with the Ag1-CA locus. The mild SMA mutations do not show association. It should be noted that the association of Ag1-CA with disease severity could be a consequence of the mechanisms by which mutations arise. For example, deletions could remove one of the Ag1-CA AFLV copies as well as the SMA gene. In support of this hypothesis, analysis of German type I SMA families with Ag1-CA has shown the lack of inheritance of a parental Ag1-CA allele by SMA offspring, which is consistent with a deletion corrected (B. Wirth, unpublished data). In addition, Melki et al. (1994) have also reported deletions within SMA families.

As stated earlier, the genomic region surrounding the SMA locus is complex, containing multicopy markers as well as a high incidence of pseudogenes. In an attempt to explain our observations of the relation between Ag1-CA and SMA, we propose the following model, based on our knowledge of the α-globin gene cluster and the deletion mutations that lead to the α-thalassemias. This cluster contains five related genes: two α-globin genes, an α-pseudogene, an ζ-gene, and an ζ-pseudogene. The two α-globin genes are almost identical; their introns share a high degree of sequence similarity, and PCR amplification of portions of intronic sequence can actually occur at two locations (Weatherall 1992). Consequently, if a microsatellite was present in a block of intronic sequence that had been duplicated, PCR amplification of that microsatellite could give rise to a multicopy pattern similar to that observed for Ag1-CA. We propose that there are at least two copies of the SMA gene per chromosome and that the Ag1-CA AFLV copies are located nearby. As in severe α-thalassemia, most type I SMA patients would carry on each chromosome deletions that encompass both SMA genes as well as one of the Ag1-CA AFLV copies, or in certain instances both AFLV copies as in the German type I families (giving rise to class I chromosomes or nulls). Normal individuals with single-allele chromosomes contain deletions that remove one SMA gene and one Ag1-CA AFLV but do not extend to the second SMA locus or Ag1-CA

AFLV. As in the α-globins, deletion of one SMA locus would not be detrimental, because the second gene, located in cis to the first, would be up-regulated. This scenario as well as divergence of the Ag1-CA locus would account for 34% of normal chromosomes being class I. Type II SMA patients are heterozygotes; one chromosome is deleted for both copies of the SMA gene while the other contains one SMA gene that contains a point mutation. In this instance, the normal SMA gene in cis would not be up-regulated, since the mRNA from the aberrant gene is produced at normal levels, as in the α-globins (Weatherall 1992). In type III SMA, both chromosomes carry one SMA gene that has a point mutation, and the second gene, located in cis, cannot be up-regulated. Finally, in a manner similar to α- and β-globins, the phenotypic heterogeneity observed in type I/II or type II/III families could be explained by the interaction of the SMA gene product with another protein encoded by an unlinked locus. The structure or concentration of this interacting protein could influence the presentation of SMA. Definitive evidence for or against this hypothesis must await the isolation of the SMA gene(s). However, we present this model because it has important implications with respect to the isolation and confirmation of a candidate cDNA as being a product of the SMA locus.

In conclusion, the marker Ag1-CA shows highly significant association in type I SMA, and the Ag1-CA chromosome class marks the severity of the disorder. Consequently, this marker clearly identifies the critical region that should be searched for candidate genes.

Acknowledgments

We are grateful to L. Deaven of the Los Alamos National Laboratory for making the chromosome 5 cosmid library available. We would like to thank the numerous clinicians and genetic counselors who have played an extremely important role in identifying and following the SMA families used in this study. We would also like to acknowledge the Kennedy/Hopkins NICHD Mental Research Center Core Grant HD24061 for establishing some of the lymphoblastoid cell lines from SMA patients used in this study. This research was funded by the Muscular Dystrophy Association (USA), the Muscular Dystrophy Association of Canada, the Fonds de la Recherches en Santé du Québec, the Networks of Centres of Excellence Program (Canadian Genetic Diseases Network), and Deutsche Forschungsgemeinschaft. We are also grateful to Families of SMA for support given to C.J.D., a recipient of a predoctoral fellowship. J.D.C. is the recipient of an NIH predoctoral minority fellowship. Finally, we would like to give a heartfelt thank-you to all SMA families, as this work would have been impossible without their kind cooperation.

References

Becker RA, Chambers JM, Wilks AR (1988) The new S language: a programming environment for data analysis and graphics. Wadsworth & Brooks, Pacific Grove, CA

Bishop YMM, Fienberg SE, Holland PW (1975) Discrete multivariate analysis: theory and practice. MIT Press, Cambridge, MA

Brahe C, Velonà I, van der Steege G, Zappata S, van der Veen AY, Osinga J, Tops CM, et al (1994) Mapping of two new markers within the smallest interval harboring the spinal muscular atrophy locus by family and radiation hybrid analysis. Hum Genet 93:494–501

Brzustowicz LM, Lehner T, Castilla LH, Penchaszadeh GK, Wilhelmsen KC, Daniels R, Davies KE, et al (1990) Genetic mapping of chronic childhood-onset spinal muscular atrophy to chromosome 5q11.2-13.3. Nature 344:540–541

Burghes AHM, Ingraham SE, Kote-Jarai Z, Rosenfeld S, Herta N, Nadkarni N, DiDonato CJ, et al (1994*a*) Linkage mapping of the spinal muscular atrophy gene. Hum Genet 93:305–312

Burghes AHM, Ingraham SE, McLean M, Thompson TG, McPherson JD, Kóte-Jarai Z, Carpten JC, et al (1994*b*) A multicopy dinucleotide marker that maps close to the spinal muscular atrophy gene. Genomics 21:394–402

Carpten JD, DiDonato CJ, Ingraham SE, Wagner-McPherson C, Nieuwenhuijsen BW, Wasmuth JJ, Burghes AHM. A YAC contig of the region containing the spinal muscular atrophy gene: identification of an unstable region. Genomics (in press)

Dana S, Wasmuth J (1982) Linkage of the leuS, emtB and chr genes on chromosome 5 in humans and expression of human genes encoding protein synthesis components in human-Chinese hamster hybrids. Somat Cell Genet 8:245

Fienberg SE (1980) The analysis of cross-classified categorical data, 2d ed. MIT Press, Cambridge, MA

Francis MJ, Morrison KE, Campbell L, Grewal PK, Christodoulou Z, Daniels RJ, Monaco AP, et al (1993) A contig of nonchimeric YACs containing the spinal muscular atrophy gene in 5q13. Hum Mol Genet 2:1161–1167

Gilliam TC, Brzustowicz LM, Castillo LH, Lehner T, Penchaszadeh GK, Daniels RJ, Byth BC, et al (1990) Genetic homogeneity between acute and chronic forms of spinal muscular atrophy. Nature 345: 823–825

Gilliam TC, Freimer NB, Kaufmann CA, Powchik PP, Bassett AS, Bengtsson U, Wasmuth JJ (1989) Deletion mapping of DNA markers to a region of chromosome 5 that cosegregates with schizophrenia. Genomics 5:940–944

Hudson TJ, Engelstein M, Lee MK, Ho EC, Rubenfield MJ, Adams CP, Housman DE, et al (1992) Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms. Genomics 13:622–629

Kleyn PW, Wang CH, Lien LL, Vitale E, Pan J, Ross BM, Grunn A, et al (1993) Construction of a yeast artificial chromosome contig spanning the spinal muscular atrophy disease gene region. Proc Natl Acad Sci USA 90:6801–6805

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

Ledbetter SA, Nelson DL, Warren ST, Ledbetter DH (1990) Rapid isolation of DNA probes within specific chromosome regions by interspersed repetitive sequence polymerase chain reaction. Genomics 6:475–481

Lien LL, Boyce FM, Kleyn P, Brzustowicz LM, Menninger J, Ward DC, Gilliam TC, et al (1991) Mapping of human micro-

tubule-associated protein 1B in proximity to the spinal muscular atrophy locus at 5q13. Proc Natl Acad Sci USA 88:7873–7876

MacKenzie A, Roy N, Besner A, Mettler G, Jacob P, Korneluk R, Surh L (1993) Genetic linkage analysis of Canadian spinal muscular atrophy kindreds using flanking microsatellite 5q13 polymorphisms. Hum Genet 90:501–504

Mankoo BS, Sherrington R, DeLa Concha A, Kalsi G, Curtis D, Melmer G, Gurling HMD (1991) Two microsatellite polymorphisms at the D5S39 locus. Nucleic Acids Res 19:1963

Melki J, Abdelhak S, Sheth P, Bachelot MF, Burlet P, Marcadet A, Aicardi J, et al (1990*a*) Gene for chronic proximal spinal muscular atrophies maps to chromosome 5q. Nature 344:767–768

Melki J, Lefebvre S, Burglen L, Burlet P, Clermont O, Millasseau P, Reboullet S, et al (1994) De novo and inherited deletions of the 5q13 region in spinal muscular atrophies. Science. 264: 1474–1477

Melki J, Sheth P, Abdelhak S, Burlet P, Bachelot MF, Lathrop MG, Frezal J, et al (1990*b*) Mapping of acute (type I) spinal muscular atrophy to chromosome 5q12-q14. Lancet 336:271–273

Morrison KE, Daniels RJ, Suthers GK, Flynn GA, Francis MJ, Buckle VJ, Davies KE (1992) High-resolution genetic map around the spinal muscular atrophy (SMA) locus on chromosome 5. Am J Hum Genet 50:520–527

Munsat TL (1991) Workshop report: international SMA collaboration. Neuromuscul Disord 1:81

——— (1992) Meeting report: international SMA consortium meeting. Neuromuscul Disord 2:423–428

Pearn J (1980) Classification of spinal muscular atrophies. Lancet 1:919–922

Riley J, Butler R, Ogilvie D, Finniear R, Jenner D, Powwell S, Anand R, et al (1990) A novel, rapid method for the isolation of terminal sequence from yeast artificial chromosomes (YAC) clones. Nucleic Acids Res 18:2887–2890

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2d ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Simard LR, Prescott G, Rochette C, Morgan K, Lemieux B, Mathieu J, Melancon SB, et al (1994) Linkage disequilibrium analysis of childhood-onset spinal muscular atrophy (SMA) in the French-Canadian population. Hum Mol Genet 3:459–463

Simard LR, Vanasse M, Rochette C, Morgan K, Lemieux B, Melancon SB (1992) Linkage study of chronic childhood-onset spinal muscular atrophy (SMA): confirmation of close linkage to D5S39 in French Canadian families. Genomics 14:188–190

Smith CAB (1986) Chi-squared tests with small numbers. Ann Hum Genet 50:163–167

Soares VM, Brzustowicz LM, Kleyn PW, Knowles JA, Palmer DA, Asokan S, Penchaszadeh GK, et al (1993) Refinement of the spinal muscular atrophy locus to the interval between D5S435 and MAP1B. Genomics 15:365–371

Statistical Sciences (1992) S-PLUS, version 3.1, release 1 for Sun SPARC, SunOS 4.x

Thompson TG, Morrison KE, Kleyn P, Bengtsson U, Gilliam TC, Davies KE, Wasmuth JJ, et al (1993) High resolution physical map of the region surrounding the spinal muscular atrophy gene. Hum Mol Genet 2:1169–1176

Weatherall DJ (1992) Thalassemia. In: Stamatoyannopoulos G, Nienhuis AW, Majerus PW, Varmus H (eds) The molecular basis of blood disease. WB Saunders Philadelphia, pp 157-207

Wirth B, Pick E, Leutner A, Voosen B, Knapp M, Piechaczek-Wappenschmidt B, Rudnik-Schöneborn S, et al (1994) Large linkage analysis in 100 families with autosomal recessive spinal muscular atrophy (SMA) and 11 CEPH-families using 15 polymorphic loci in the region 5q11.2-q13.3. Genomics 20:84–93

Wirth B, Voosen B, Röhrig D, Knapp M, Piechaczek B, Rudnik-Schöneborn S, Zerres K (1993) Fine mapping and narrowing of the genetic interval of the genetic interval of the spinal muscular atrophy region by linkage studies. Genomics 15:113–118