GENOMICS 21, 394–402 (1994)

A Multicopy Dinucleotide Marker That Maps Close to the Spinal Muscular Atrophy Gene

ARTHUR H. M. BURGHES,*[,†,‡,1] SUSAN E. INGRAHAM,* MICHAEL MCLEAN,§ TERRI G. THOMPSON,¶ JOHN D. MCPHERSON,¶ ZSOFIA KOTE-JARAI,*[,†] JOHN D. CARPTEN,† CHRISTINE J. DIDONATO,‡ JOH-E. IKEDA,|| LINDA SURH,§ BRUNHILDE WIRTH,** CAROLE A. SARGENT,†† MALCOLM A. FERGUSON-SMITH,†† PAUL FUERST,‡ ROBERT K. MOYZIS,‡‡ DEBORAH L. GRADY,‡‡ KLAUS ZERRES,** ROBERT KORNELUK,§ ALEX MACKENZIE,§ AND JOHN J. WASMUTH¶

*Department of Medical Biochemistry and †Department of Neurology, College of Medicine, and ‡Department of Molecular Genetics, College of Biological Sciences, Ohio State University, Columbus, Ohio 43210; §Molecular Genetics Laboratory, Children's Hospital of Eastern Ontario, Ottawa, Ontario K1H 8L1 Canada; ¶Department of Biochemistry, University of California, Irvine, California 92717; ||IGSP/ERATO, Tokhai University School of Medicine, Bohseidai, Isehara City, Kanagawa 259-11 Japan; **Institute for Human Genetics, 53111 Bonn, Germany; ††Department of Pathology, University of Cambridge, Cambridge CB2 1QP, United Kingdom; and ‡‡Los Alamos National Laboratory, MS-M886, Los Alamos, New Mexico 87545

Received December 13, 1993; revised March 2, 1994

Spinal muscular atrophy (SMA) is a common autosomal recessive disorder resulting in loss of motor neurons. The interval containing the SMA gene has been defined by linkage analysis as 5qcen–D5S435–SMA–D5S557–5qter. We have isolated a new dinucleotide repeat marker, CATT1, that lies between these two closest markers. The marker CATT1 has 16 alleles and is highly polymorphic. The marker can have 1 to 4 (or more) copies per chromosome, giving rise to individuals with up to 8 (or more) alleles. All of the subloci map between the markers D5S557 and D5S435 and lie in close proximity to one another. The marker CATT1 is linked to the SMA gene with a lod score of Z_{max} = 34.42 at $\theta = 0$ and crosses all available recombinants. Certain alleles occurred more frequently in either the SMA or normal populations, indicating significant allelic association between CATT1 and the SMA locus. Haplotype analysis combining U.S. and Canadian SMA families reveals that one haplotype group (VII) occurs significantly more frequently in the SMA population than in the normal. This confirms the allelic association of CATT1 with the SMA locus. © 1994 Academic Press, Inc.

INTRODUCTION

Childhood spinal muscular atrophy (SMA) is a common autosomal recessive disorder. The disorder is characterized by loss of the anterior horn cells of the spinal cord. The juvenile SMAs can be divided into three clinical groups (Dubowitz, 1978; Pearn, 1982; Brooke, 1985). A fourth class of SMA is characterized by adult onset. All three forms of juvenile SMA have been mapped to 5q11.2–q13.3 (Brzustowicz *et al.*, 1990; Melki *et al.*, 1990a,b; Gilliam *et al.*, 1990; Simard *et al.*, 1992; Daniels *et al.*, 1992; Wirth *et al.*, 1993; MacKenzie *et al.*, 1993). Recombinants have positioned the gene between the markers D5S435 and D5S557 (Soares *et al.*, 1993; Burghes *et al.*, 1994; Francis *et al.*, 1993; Wirth *et al.*, 1994). Recently a physical map of the region containing the SMA gene has been constructed using both radiation hybrid mapping (Thompson *et al.*, 1993) and contiguous sets of YAC clones (Francis *et al.*, 1993; Kleyn *et al.*, 1993; Carpten *et al.*, 1994). The data confine the SMA gene to an interval of approximately 1.4 Mb. In this paper we present a dinucleotide repeat marker, CATT1, that maps between the two closest flanking markers. The marker is highly polymorphic and has one to four (or more) alleles per chromosome, giving rise to individuals with up to eight (or more) alleles. All of the subloci of CATT1 map between the two markers that flank the SMA gene. This marker crosses all available recombinants and shows linkage disequilibrium with the SMA locus.

[1] To whom correspondence should be addressed at the Department of Neurology, 452 Means Hall, 1654 Upham Drive, Columbus, OH 43210. Telephone: (614) 293-8434. Fax: (614) 293-4688.

MATERIALS AND METHODS

Family material. Spinal muscular atrophy was diagnosed by the standard criteria presented by the international SMA consortium (Munsat, 1991, 1992). The SMA patients were classified as type 1, 2, or 3 according to the criteria outlined by the international SMA consortium. For analysis the families were divided into chronic (types 2 and 3) and acute (type 1).

The details of most families used in this study, including the recombinant families, have been reported previously (Burghes *et al.*,

0888-7543/94 $6.00
Copyright © 1994 by Academic Press, Inc.
All rights of reproduction in any form reserved.

1994; MacKenzie *et al.*, 1993; Wirth *et al.*, 1993, 1994). Additional families with only one affected individual were included for allelic association studies.

Cell culture and somatic cell hybrid lines. Somatic cell hybrid HHW105, the sole human component of which is chromosome 5 (Dana and Wasmuth, 1982), was obtained from the NIGMS Human Genetic Mutant Cell Repository (Corriel Institute for Medical Research, Camden, NJ) and grown in proline-supplemented αMEM as described by the supplier. DNA was also obtained from the cell line HHW1064 containing a der5 chromosome, which has a deletion of 5q11.2–q13.3 (Gilliam *et al.*, 1989). The somatic cell hybrid cell lines A5, A10, B6, and B3 (Funanage *et al.*, 1984) were the generous gift of V. Funanage. A10 contains a translocated human chromosome 5, der5t(8:5)(8pter–q11::5q11–q31) and human chromosome 16. A5 contains a human chromosome 5 that consists solely of the long arm (5q), as well as human chromosomes 7, 13, 14, 15, 20, 21, 22, and X. B3 contains a dicentric human chromosome 5, dic5(pter–q22::q22–pter), as well as human chromosomes 3, 8, 10, 14, 15, 16, and 22. B6 contains an intact human chromosome 5, in addition to a number of other human chromosomes (Funange *et al.*, 1984). The cell lines A10, A5, B3, and B6 were grown in F12 medium without hypoxanthine (JRH Biosciences, Lenexa, KS).

Isolation and mapping of CATT1. Cosmids isolated from the SMA region often do not hybridize to a single locus when used as FISH probes. The hybridization patterns are complex but chromosome 5-specific (Thompson *et al.*, 1993). The CATT locus is derived from a chimeric cDNA clone (Sargent *et al.*, 1991; Habeebu *et al.*, 1989). Cosmids isolated by screening with the chimeric cDNA showed FISH hybridization to 5p13, 5p14, 5q13, and 5q15 (Sargent *et al.*, 1991). This pattern is similar to those of cosmids known to derive from the SMA region (Thompson *et al.*, 1993). To refine the mapping of this clone, primers were developed from the chimeric cDNA. The primers were initially chosen by examination of the sequence and selection of an area with approximately 50% (G+C) content, no predicted secondary structure, and no extended stretch of a single residue. When these primers gave multiple bands, the sequence was reanalyzed using the program PRIMER (Lincoln *et al.*, 1991), and the primers indicated in Fig. 1A were selected. All subsequent experiments were performed with this second primer set. A cosmid containing the CATT1 sequence was also identified from the chromosome 5 cosmid library (kindly provided by Dr. L. Deaven, Life Sciences Division, Los Alamos National Laboratory) in the Ottawa laboratory. The primers defining the CATT1 locus were used on a panel of radiation hybrids (Warrington *et al.*, 1991) and mapped using the radiation hybrid programs of Cox *et al.* (1990). The primer sequences used are 5′-GAACGGCAGAGTCGGAGGAGAAT-3′ (forward primer) and 5′-GAGAAGGCTTCCTCCTGAGTATGC-3′ (reverse primer).

PCR analysis. The conditions for the markers D5S76 (Sherrington *et al.*, 1991), D5S507 (Wirth *et al.*, 1994), D5S125 (Mankoo *et al.*, 1991), D5S435 (Soares *et al.*, 1993), D5S351 (Hudson *et al.*, 1992), and MAP1B (RB104-106, RB110-111; Lien *et al.*, 1991) were as described previously (MacKenzie *et al.*, 1993; Burghes *et al.*, 1994; Wirth *et al.*, 1994). The CATT1 forward primer was endlabeled with [γ-^{32}P]ATP using a standard T4 polynucleotide kinase reaction (Sambrook *et al.*, 1989). The marker CATT1 was amplified using 50 to 100 ng genomic DNA in 25-μl reactions containing 0.3 m*M* dATP, 0.3 m*M* dCTP, 0.3 m*M* dGTP, 0.3 m*M* TTP, 50 m*M* KCl (pH 9.0), 10 m*M* Tris (pH 8.3), 1.5 m*M* $MgCl_2$, 0.01% Triton X-100, 0.01% Tween-20, 0.01% gelatin, 1 unit *Taq* polymerase, 50 ng ^{32}P-end-labeled forward primer, and 50 ng reverse primer. Cycling conditions were 94°C, 3 min initial denaturation; 94°C, 30 s, 66°C, 30 s, 72°C, 1 min, for 35 cycles; 72°C, 7 min final extension. The PCR products were denatured by adding 1 vol formamide stop mix and heating for 3 min at 90°C. The samples were then loaded onto an 8% denaturing polyacrylamide gel and electrophoresed at 1500 V for 5–7 h. Best resolution was obtained using a wide-tooth (8-mm) comb.

Haplotype and linkage analyses. Linkage analysis was performed using the LINKAGE 5.1 package (Lathrop *et al.*, 1984). Spinal muscular atrophy is an autosomal recessive disorder, and full penetrance was assumed in all cases. A gene frequency of 0.0063 was used (Brzustowicz *et al.*, 1990; Burghes *et al.*, 1994). Lod scores were calculated using the MLINK program. As the marker can have up to four alleles (or more) per chromosome, it was necessary to code the alleles. For cases in which the chromosome haplotype could be assigned without doubt, a number was assigned to that haplotype (so that it could be represented as an allele in the linkage program), and the frequency with which that haplotype occurred was used for the allele frequency. For cases in which the haplotype could not be unambiguously assigned, data from a single informative allele were used, and the frequency with which that allele occurred in the population was used as the allele frequency. Note that the frequency of an allele is taken as the number of times the allele occurs in the population divided by the total number of alleles scored. The latter system was also used for the families for which full haplotypes could be determined. There was no significant difference in the lod scores obtained by the two methods.

The haplotypes (group of CATT1 alleles representing all subloci on a chromosome) were determined for affected and normal chromosomes by examination of pedigrees. Carrier status of unaffected siblings was confirmed by the previous typing with markers flanking the SMA gene (Burghes *et al.*, 1994; MacKenzie *et al.*, 1993). Because the multiallelic CATT1 marker can have one or more alleles that are uninformative, some alleles cannot be assigned with certainty. This results in multiple possible haplotypes in some instances for the normal and affected chromosomes. To account for this, each potential haplotype was assigned an equal probability; for example, if there were two possibilities, each was assigned a weight of 0.5, or if there were four possibilities, each was assigned 0.25. These weights were then used to calculate haplotype frequencies.

In addition to the haplotype analysis, affected individuals were scored for the presence or absence of single alleles. The frequency of single alleles in the normal population was determined from each parent's normal chromosome to ensure a similar population for both normal and affected and to guard against any selection bias.

The association of alleles or haplotypes with affected or normal disease state was tested using a χ^2 contingency analysis, following the procedures of Sokal and Rohlf (1969).

RESULTS

Mapping of CATT1

Initially the CATT1 locus was identified by the similarity of *in situ* hybridization patterns obtained with CATT1-containing cosmids and other cosmids obtained from the SMA region. Primers to the CATT1 locus (Fig. 1A, dashed lines) were designed and used to map this locus using a panel of previously characterized radiation hybrids (Warrington *et al.*, 1991).

The retention frequency for the CATT1 locus in this set of radiation hybrids was 28%. Using both two-point and four-point analyses (Cox *et al.*, 1990), the CATT1 locus was determined to be 20 cR_{6500} distal to the proximal flanking marker for SMA, D5S435, and 6 cR_{6500} proximal to the previously described monomorphic marker TT28.8 (Thompson *et al.*, 1993). Using the previously determined value of 1 cR_{6500} equal to approximately 17 to 35 kb, we determined that CATT1 is approximately 340 to 600 kb distal to D5S435 (Fig. 1B).

Data from a YAC contig (Carpten *et al.*, 1994) place CATT1 centromeric of the marker D5S557. YAC 22 is

A
TCACAAAATGTTCTCCCCCAGAATGAAGAGTTTCACTGGATGGA
ACGGCAGAGTCGGAGGAGAATTATTCCTTTAATTTTTATTTCTG
TTGATGTTGCAATTGTTTTTATGCAGTGCAAGCAAACATACACA
CACACACACACACACACACACACACNNATNCAAGCTGTGAATGN
TTTATGCATACTCAGGAGGAAGCCTTCTCAGGGTCACNGTTTCC
GGAAACTGACCTTGAAAACAGACCGNTATGTCAACAAATCCC

B

CEN	a	b	c
D5S125			
	10	18.67	1226
LAS96			
	23	13.68	38531
D5S435/TT106.5			
	20	14.8	9E+10
CATT1			
	6	20.12	14271
TT28.8			
	15	15.55	2E+9
Y4T/38.3(YJK850)			
	12	16.93	2E+8
MAP1B			
TEL			

SMA

C
y97
y98
y111
y22
CATT1 D5S557 D5S351

FIG. 1. (**A**) Sequence from CATT1 cDNA, indicating primer sequences used for initial mapping (dashed line) and for PCR amplification (solid line). (**B**) Radiation hybrid map of the SMA region, including CATT1. The distances for other markers have been reported previously (Thompson *et al.*, 1993). To the right and between each pair of markers are indicated (**a**) the RH map distance separating the markers (in cR_{6500}; 1 cR_{6500} = 17–35 kb); (**b**) the lod score for the likelihood of linkage vs nonlinkage; and (**c**) the odds against inversion of adjacent loci. Loci on the same line separated by a slash showed only a single or no discordancy in any of the 109 RH hybrids. (**C**) Location of CATT1 with respect to YAC clones and nearby markers.

positive for D5S351 and D5S557 but not CATT1. This YAC is 680 kb. In addition, YACs 97 and 98 from the contig of Kleyn *et al.* (1993) are positive for CATT1, whereas D5S557 is present on YACs 98 and 111 but not 97 (Fig. 1C). This clearly places CATT1 centromeric of D5S557, toward the SMA gene.

Characterization of CATT1 Alleles in Hybrids

Figure 2 shows the results of CATT1 amplification of a series of somatic cell hybrids that contain a single intact or derivative chromosome 5. It is clear that there is more than one CATT1 sublocus per chromosome, as the haplotype shows up to four alleles (Fig. 2). These subloci all map to 5q11.2–q13.3, as only the radiation hybrids containing this genomic region are positive. Furthermore, the CATT1 subloci are present in radiation hybrid 28, the sole human component of which is the 2-Mb region containing the SMA gene (Thompson *et al.*, 1993). These subloci must lie close to each other, as they are not separated among the radiation hybrid panel. This is also consistent with the YAC contig and cosmids isolated from the region (Carpten *et al.*, 1994 and unpublished data).

Analysis of CATT1 in SMA Families

CATT1 has multiple copies that physically map close together. The frequency of occurrence of the different alleles is shown in Table 1. The table indicates the frequency of alleles as a fraction of total alleles seen.

The CATT1 marker was used to analyze all available SMA families. The marker crossed all recombinant families in the Columbus and Ottawa populations and is linked to the SMA locus with a lod score of Z_{max} = 34.42 at θ = 0. Figure 3 shows an example of the marker on an SMA family. The number of CATT1 subloci varies among different chromosomes. Genotyping of key centromeric and telomeric recombinants is shown in Fig. 4. It is clear that CATT1 crosses these recombinants and segregates with SMA.

Table 2 shows the number of times each particular

A
1 2 3 4 5 6
3 5 6 7 9 10 13 15 16

B
1 2 3 4 5 6
6 7 8 11 12 15

FIG. 2. Amplification of the CATT1 locus in a series of chromosome 5 hybrids. (**A**) Lane **1:** PCR blank reaction contains no DNA. Lane **2:** amplification of human genomic DNA (alleles present: 3, 5, 6, 10, 13). Lane **3:** amplification of DNA from hybrid A5 containing 5q (alleles present: 5, 7, 13, 15, 16). Analysis with other chromosome 5 markers indicates that this hybrid contains more than one chromosome 5. Lane **4:** amplification of DNA from hybrid A10 (5q11–q31) (alleles present: 15, 16). Lane **5:** amplification of DNA from hybrid B3 (dic5 pter–q22::q22–pter) (alleles present: 9, 15, 16). Lane **6:** DNA from hybrid B6 containing chromosome 5 (alleles present: 5, 7, 13, 15, 16). Analysis with other chromosome 5 markers indicates that this hybrid may contain more than one chromosome 5. Differences in band intensities of alleles within a lane are most likely due to dosage differences. (**B**) Lane **1:** amplification of DNA from hybrid HHW105 containing chromosome 5 as its sole human component (alleles present: 6, 7, 8). Lane **2:** amplification of DNA from hybrid 1064 containing a deletion of 5q11.2–q13.3. Lane **3:** amplification of DNA from radiation hybrid 83 containing the SMA region (alleles present: 7, 11, 12, 15). Lane **4:** amplification of DNA from radiation hybrid 86 containing the SMA region (alleles present: 7, 11, 12, 15). Lane **5:** amplification of DNA from radiation hybrid 28 that contains the SMA region as its sole human component (alleles present: 7, 11, 12, 15). Lane **6:** PCR blank reaction contains no DNA.

TABLE 1

Frequency of Alleles in both the American (Columbus) and Canadian (Ottawa) Populations

		Columbus population: Frequency in				Ottawa population: Frequency in			
Allele	Size (bp)	Normals	Acute SMA	Chronic SMA	Total SMA	Normals	Acute SMA	Chronic SMA	Total SMA
1	188	0.0040	0.0000	0.0000	0.0000	0.0000	0.0000	0.0095	0.0052
2	182	0.0080	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	180	0.0040	0.0104	0.0000	0.0042	0.0095	0.0000	0.0000	0.0000
4	178	0.0119	0.0104	0.0000	0.0042	0.0095	0.0000	0.0190	0.0104
5	176	0.1202	0.0730	0.0423	0.0546	0.0671	0.1264	0.0190	0.0677
6	174	0.0987	0.1040	0.0833	0.0916	0.0943	0.1034	0.0571	0.0781
7	172	0.1466	0.0835	0.1669	0.1335	0.1547	0.1609	0.1619	0.1615
8	170	0.1215	0.1562	0.1461	0.1501	0.0827	0.1609	0.1524	0.1563
9	168	0.0324	0.0730	0.0478	0.0579	0.0237	0.0115	0.0000	0.0052
10	166	0.0873	0.1457	0.1117	0.1253	0.0950	0.1609	0.1238	0.1406
11	164	0.0577	0.0626	0.0763	0.0708	0.1044	0.0345	0.0763	0.0574
12	162	0.1201	0.1770	0.1730	0.1746	0.1374	0.1379	0.2286	0.1875
13	160	0.0000	0.0000	0.0000	0.0000	0.0095	0.0000	0.0000	0.0000
14	158	0.0457	0.0104	0.0209	0.0167	0.0570	0.0230	0.0190	0.0208
15	156	0.1142	0.0936	0.1317	0.1164	0.1553	0.0805	0.1333	0.1094
16	152	0.0278	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Total alleles		251.60	95.84	143.30	239.14	210.66	87.00	105.00	192.00
Total *chromosomes*		*90*	*40*	*51*	*91*	*74*	*34*	*40*	*74*

Note. The frequencies are given per allele, as more than one allele can occur on a chromosome.

allele occurs on SMA and normal chromosomes for all SMA groups. In acute (type 1) SMA, alleles 11 and 15 occurred less often on the SMA chromosome in the Canadian and total populations. In chronic (types 2 and 3) SMA, allele 5 occurred less commonly on the SMA chromosome in the American and total populations. Analysis of SMA chromosomes from both populations and all SMA types combined revealed that alleles 5 and 14 were less common in SMA chromosomes, whereas alleles 8 and 11 were more common in SMA chromosomes than in normal chromosomes. Because of nonindependence between alleles in the multiway table of allelic associations (Table 2), the indicated significance levels for individual tests may not be strictly true. Assuming the Bonferroni procedure for nonindependence, *P* values for each allelic association can be multiplied by the number of informative alleles to obtain a corrected level of significance (Weir, 1990). When this correction is applied, the negative allelic association between allele 11 and SMA in the Ottawa population remains significant ($P = 0.030$), and the negative association of allele 5 and chronic SMA in the total population remains significant ($P = 0.029$). Other associations become nonsignificant. It is possible, however, that a strict application of the Bonferroni procedure in this manner may actually underestimate the significance of allelic association in a complex data set of the sort considered here (Zelterman, 1993). Indeed, preliminary data in the Ottawa population using a system that reduces the complexity of the alleles indicate that this is the case (McLean and MacKenzie, unpublished data).

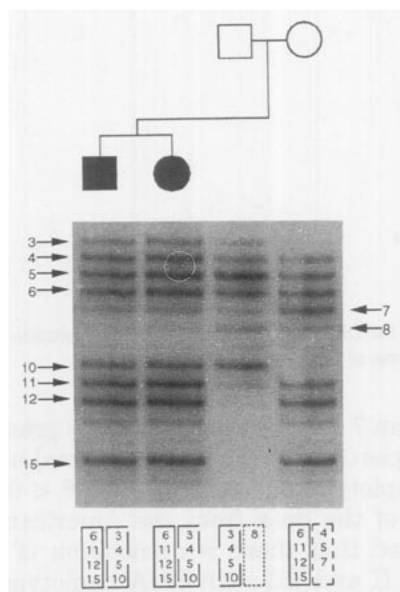

FIG. 3. An example of CATT1 analysis on an SMA family. (**top**) The pedigree. (**middle**) Results of amplification with CATT1 primers. Alleles read from affected son are indicated to the left. Allele 7, present in the mother but not in the children and allele 8 present in the father but not in the children, are indicated on the far right. (**bottom**) Haplotypes read from the gel. Note that alleles 4 and 5 are uninformative; the haplotypes indicated are one of four possibilities.

We extended this analysis by examination of the haplotypes occurring on the SMA and normal chromosomes. The haplotypes associated with normal and affected were determined by examination of the pedigree as described under Materials and Methods. In cases

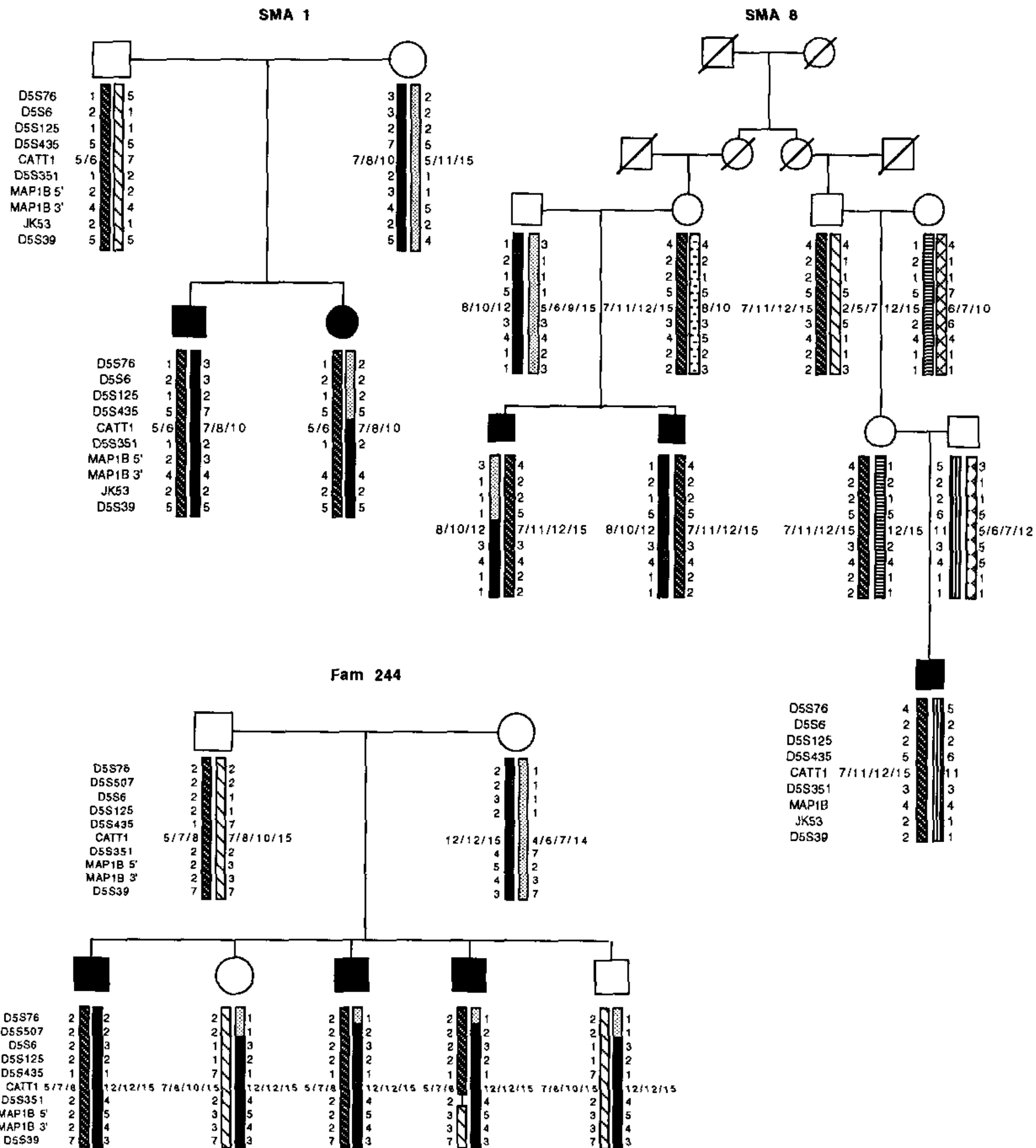

FIG. 4. Recombinant families typed with CATT1. SMA1 and SMA8 are recombinant for D5S435. Family 244 is recombinant for MAP1B. Note that in SMA 1, CATT1 alleles 5 and 7 are uninformative; haplotypes shown are one of four possibilities.

in which some alleles were uninformative, all possible haplotypes were constructed and assigned a weight inversely proportional to the number of possible haplotypes (e.g., if one allele is uninformative, there are two possibilities, each with a 0.5 probability). The haplotypes were grouped into 12 classes, as indicated in Table 3. The Canadian families and the American families are shown separately. The overall allele distributions in the two population groups were not significantly different. Consequently, the two family groups were combined for further analysis. In the combined data set, the frequency of group VII (allele 12 present with alleles other than 7 or 15) was significantly greater in the SMA haplotypes (33.82/165) than in normal haplotypes (18.58/164 haplotypes; χ^2 5.22 at 1 *df*; $P < 0.025$). An examination of the data from the American families alone indicated that there is association of the combined groups II and VII with SMA haplotypes (24.43/91) compared to normals (12.24/90; χ^2 4.99 at 1 *df*; $P < 0.05$). No other haplotypes were significantly associated, either positively or negatively, with SMA. Therefore, haplotyping of all SMA families indicates the presence of allelic association between CATT1 and SMA in this population.

TABLE 2

Allele Counts of CATT1 in SMA and Normal Chromosomes

	Ottawa chronic SMA		Ottawa acute SMA		Columbus chronic SMA		Columbus acute SMA		Total chronic SMA		Total acute SMA		Total	
Allele	Normal	SMA	Normal	SMA	Normal	SMA	Normal	SMA	Normal	SMA	Normal	SMA	Normal	SMA
1	0	1	0	0	1	0	0	0	1	1	0	0	1	1
2	0	0	0	0	2	0	0	0	2	0	0	0	2	0
3	2	0	0	0	1	0	0	1	3	0	0	1	3	1
4	1	2	0	0	0	0	3	1	1	2	3	1	4	3
5	8	2	5	11	15	6***	14	7	23	8*	19	18	42	26***
6	11	6	10	9	15	12	8	10	26	18	18	19	44	37
7	18	17	14	14	20	24	14	8	38	41	28	22	66	63
8	9	16	8	14	15	21	13	15	24	37	21	29	45	66†
9	3	0	1	2	7	7	1	7	10	7	2	9	12	16
10	7	13	11	14	12	16	8	14	19	29	19	28	38	57†
11	8	8	14	3*	10	11	5	6	18	19	19	9†	37	28
12	14	24	14	12	21	25	9	17	35	49	23	29	58	78
13	2	0	0	0	0	0	0	0	2	0	0	0	2	0
14	5	2	7	2	7	3	4	2	12	5	11	4	23	9**
15	14	14	18	7**	15	19	12	8	29	33	30	15**	59	48
16	0	0	0	0	3	0	4	0	3	0	4	0	7	0
Total chromosomes	35	40	35	34	50	51	38	40	85	91	73	74	158	165

Note. The bottom line (Total chromosomes) indicates the number of chromosomes in each group. Each allele count was compared by a χ^2; the statistically significant differences are indicated. In some cases an allele shows dosage compatible with two or three copies in an individual. For the groups that show significant differences, this occurred in one Columbus acute family for allele 15, three Columbus chronic families for allele 8, and one Columbus acute family for allele 8. For allele 15, there were two copies in one parent, one copy in the other parent, and two copies in the affected individual. Therefore, two alleles were assigned to the SMA group and one to the normals. One chronic family presented the same situation with allele 8. In the remaining chronic families and an acute family, allele 8 occurred twice in one parent and once in the other parent with a single copy in the affected; therefore, one copy was assigned to the SMA group and two to the normal group. The P values of significant differences are as follows:

* $P < 0.005$.
** $P < 0.01$.
*** $P < 0.025$.
† $P < 0.05$.

DISCUSSION

The gene for spinal muscular atrophy (SMA) has been mapped between the markers D5S435 and D5S557 (Soares *et al.,* 1993; Francis *et al.,* 1993; Burghes *et al.,* 1994; Wirth *et al.,* 1994). The marker CATT1 maps between these two markers and crosses all available SMA recombinants, giving a lod score of 34.42 at $\theta = 0$. The CATT1 locus is unusual in that it is composed of more than one sublocus, all of which can amplify, giving rise to individuals with up to eight or more alleles. The separate subloci have been cloned from the chromosome 5 cosmid library. In this case there are four alleles. Sequence data of the CATT1 locus from the four distinct subloci show that the sequence is similar for 300 bp in all of these subloci (MacKenzie *et al.,* unpublished data). These subloci also appear to be separated by unique sequence (Burghes *et al.,* unpublished data). Therefore, it does not appear that the CATT1 locus exists as a tandem duplication but as a series of subloci. Markers that flank the SMA gene, such as D5S435 and D5S112, do not show this characteristic. Other dinucleotides that physically lie close to the CATT1 locus also show multiple alleles (C. Brahe, Universita Cattolica del Sacro Cuore, Rome, Italy, pers comm., May 1993; Kleyn *et al.,* 1993). These usually have multiple subloci, and the subloci appear to lie very close together as if in a tandem organization (DiDonato and Burghes, unpublished observation). It is also clear that there are region-specific repetitive DNAs in the area containing the SMA locus. One such repeat identified in YAC end clones and cosmids derived from the region appears to be located on 5p, 5q31, and 5q11.2–q13.3 (Francis *et al.,* 1993; Thompson *et al.,* 1993).

Although highly informative, the multiple subloci of CATT1 make analysis of allelic association difficult. We do not feel that it is possible to calculate disequilibrium coefficients in the conventional manner. Instead, we have compared the presence or absence of alleles directly and used a χ^2 to compare the values obtained. These data clearly indicate differences in the frequency of alleles present in the SMA population. This was extended by analysis of haplotypes. In this case the haplotypes were weighted according to the probability of their occurrence, as certain alleles are uninformative.

TABLE 3

Haplotypes for CATT1 and SMA Chromosomes

		Columbus population				Ottawa population				Total			
		Number of chromosomes		Proportion of chromosomes		Number of chromosomes		Proportion of chromosomes		Number of chromosomes		Proportion of chromosomes	
Group	Group definition	Normals	SMA	Normals	SMA	Normals	SMA	Normals	SMA	Normals	SMA	Normals	SMA
I	Single allele	11.87	13.97	0.1319	0.1535	7.90	10.14	0.1068	0.1370	19.77	24.10	0.1206	0.1461
II	7/12/15 together	2.63	8.57	0.0292	0.0942	7.56	5.15	0.1022	0.0696	10.18	13.72	0.0621	0.0832
III	12/15 together without 7	9.56	7.81	0.1062	0.0859	7.40	4.83	0.1000	0.0653	16.96	12.64	0.1034	0.0766
IV	7/12 together without 15	6.36	5.56	0.0706	0.0611	4.29	6.83	0.0580	0.0923	10.65	12.40	0.0649	0.0751
V	7/15 together without 12	7.50	4.28	0.0833	0.0470	6.45	5.97	0.0872	0.0807	13.95	10.25	0.0851	0.0621
VI	15 present with alleles other than 7 or 12	7.66	6.46	0.0851	0.0710	8.21	4.28	0.1109	0.0578	15.87	10.73	0.0967	0.0651
VII	12 present with alleles other than 7 or 15	9.61	15.86	0.1067	0.1743	8.97	17.96	0.1212	0.2426	18.57	33.82*	0.1132	0.2050
VIII	7 present with alleles other than 12 or 15	18.38	12.63	0.2042	0.1388	13.37	11.91	0.1806	0.1610	31.75	24.55	0.1936	0.1488
IX	6/8 together without 7 or 12 or 15	2.39	4.17	0.0265	0.0458	3.34	3.03	0.0451	0.0409	5.73	7.19	0.0349	0.0436
X	8 present with alleles other than 6 or 7 or 12 or 15	6.38	5.66	0.0709	0.0623	0.74	2.29	0.0099	0.0310	7.12	7.96	0.0434	0.0482
XI	6 present with alleles other than 7 or 8 or 12 or 15	4.29	1.73	0.0476	0.0190	3.35	0.19	0.0452	0.0026	7.63	1.92	0.0465	0.0116
XII	Other	3.38	4.30	0.0376	0.0472	2.44	1.42	0.0329	0.0192	5.82	5.72	0.0355	0.0347
	Total	90.00	91.00			74.00	74.00			164.00	165.00		

Note. The haplotypes were determined and weighted as described under Materials and Methods.
* Data significant at the 0.025 level in a χ^2 with 1 *df.*

We feel that this method will result in a conservative estimate of whether allele association exists, as the weighting could attenuate the degree of linkage disequilibrium observed. Even with this system the combined data showed evidence for allele association. We therefore concluded that the CATT1 locus contains multiple subloci that lie close to the SMA locus and that CATT1 shows allelic association with SMA.

The CATT1 marker crosses all SMA recombinants and shows allelic association with SMA, indicating that CATT1 lies close to the SMA gene. The proximity of the marker to the SMA gene and the high informativity associated with the multiple alleles per chromosome should make CATT1 useful in prenatal diagnosis and ultimately the cloning of the SMA gene.

ACKNOWLEDGMENTS

We are grateful for the support provided by MDA, MDAC, MRC-Canada ERATO Genosphere, and Deutsche Forschungsgemeinschaft, which funded this research. Z. Kote-Jarai was the recipient of an MDA postdoctoral fellowship.

REFERENCES

Brooke, M. H. (1985). "A Clinician's View of Neuromuscular Diseases," 2nd ed., Williams and Wilkins, Baltimore.

Brzustowicz, L. M., Lehner, T., Castilla, L. H., Penchaszadeh, G. K., Wilhelmsen, K. C., Daniels, R., Davies, K. E., Leppert, M., Ziter, F., Wood, D., Dubowitz, V., Zerres, K., Hausmanowa-Petrusewicz, I., Ott, J., Munsat, T. L., and Gilliam, T. C. (1990). Genetic mapping of chronic childhood-onset spinal muscular atrophy to chromosome 5q11.2–13.3. *Nature* **344:** 540–541.

Burghes, A. H. M., Ingraham, S. E., Kote-Jarai, Z., Rosenfeld, S., Herta, N., Nadkarni, N., DiDonato, C. J., Carpten, J., Hurko, O., Florence, J., Moxley, R. T., Cobben, J. M., and Mendell, J. R. (1994). Linkage mapping of the spinal muscular atrophy gene. *Hum. Genet.* **93:** 305–312.

Carpten, J. D., DiDonato, C. J., Ingraham, S. E., McPherson, J. D., Wagner-McPherson, C., Wasmuth, J. J., Nieuwenhuyser, B., Fischbeck, K., and Burghes, A. H. M. (1994). A YAC contig of the region containing the spinal muscular atrophy gene: Identification of an unstable region. Submitted for publication.

Cox, D. R., Burmeister, M., Price, E. R., Kim, S., and Myers, R. M. (1990). Radiation hybrid mapping: A somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* **250:** 245–250.

Dana, S., and Wasmuth, J. 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. *Somatic Cell Genet.* **8:** 245.

Daniels, R. J., Thomas, N. H., MacKinnon, R. N., Lehner, T., Ott, J., Flint, T. J., Dubowitz, V., Ignatius, J., Donner, M., Zerres, K., Rietschel, M., Cookson, W. O. C., Brzustowicz, L. M., Gilliam, T. C., and Davies, K. E. (1992). Linkage analysis of spinal muscular atrophy. *Genomics* **12:** 335–339.

Dubowitz, V. (1978). "Muscle Disorders of Childhood," Saunders, Philadelphia.

Francis, M. J., Morrison, K. E., Campbell, L., Grewal, P. K., Christodoulou, Z., Daniels, R. J., Monaco, A. P., Frischauf, A., McPherson, J., Wasmuth, J., and Davies, K. (1993). A contig of non-chimeric YACs containing the spinal muscular atrophy gene in 5q13. *Hum. Mol. Genet.* **2:** 1161–1167.

Funanage, V. L., Myoda, T. T., Moses, P. A., and Lowell, H. R. (1984). Assignment of the human dihydrofolate reductase gene to the q11–q22 region of chromosome 5. *Mol. Cell. Biol.* **4:** 2010–2016.

Gilliam, T. C., Freimer, N. B., Kaufmann, C. A., Powchik, P. P., Bassett, A. S., Bengtsson, U., and Wasmuth, J. J. (1989). Deletion mapping of DNA markers to a region of chromosome 5 that cosegregates with schizophrenia. *Genomics* **5:** 940–944.

Gilliam, T. C., Brzustowicz, L. M., Castillo, L. H., Lehner, T., Penchaszadeh, G. K., Daniels, R. J., Byth, B. C., Knowles, J., Hislop, J. E., Shapira, Y., Munsat, T., Ott, J., and Davies, K. E. (1990). Genetic homogeneity between acute and chronic forms of spinal muscular atrophy. *Nature* **345:** 823–825.

Habeebu, S. S. M., Gibson, J. E., Affara, N. A., and Ferguson-Smith, M. A. (1989). Localization of two zinc finger protein genes to (a) two loci on chromosome 5 at 5p13–p14 and to (b) the long arm of the X at Xq13–q21.1. *Cytogenet. Cell Genet.* **51:** 1009.

Hudson, T. J., Engelstein, M., Lee, M. K., Ho, E. C., Rubenfield, M. J., Adams, C. P., Housman, D. E., and Dracopoli, N. C. (1992). Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms. *Genomics* **13:** 622–629.

Kleyn, P. W., Wang, C. H., Lien, L. L., Vitale, E., Pan, J., Ross, B. M., Grunn, A., Palmer, D. A., Warburton, D., Brzustowicz, L. M., Kunkel, L. M., and Gilliam, T. C. (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, G. M., Lalouel, J. M., Julier, C., and Ott, J. (1984). Strategies for multilocus linkage analysis in humans. *Proc. Natl. Acad. Sci. USA* **81:** 3443–3446.

Lien, L. L., Boyce, F. M., Kleyn, P., Brzustowicz, L. M., Menninger, J., Ward, D. C., Gilliam, T. C., and Kunkel, L. M. (1991). Mapping of human microtubule-associated protein 1B in proximity to the spinal muscular atrophy locus at 5q13. *Proc. Natl. Acad. Sci. USA* **88:** 7873–7876.

Lincoln, S. E., Daly, M. J., and Lander, E. S. (1991). PRIMER: A computer program for selecting PCR primers, version 0.5. Whitehead Institute for Biomedical Research, Cambridge, MA.

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

Mankoo, B. S., Sherrington, R., De La Concha, A., Kalsi, G., Curtis, D., Melmer, G., and Gurling, H. M. D. (1991). Two microsatellite polymorphisms at the D5S39 locus. *Nucleic Acids Res.* **19:** 1963.

Melki, J., Abdelhak, S., Sheth, P., Bachelot, M. F., Burlet, P., Marcadet, A., Aicardi, J., Barois, A., Carriere, J. P., Fardeau, M., Fontan, D., Ponsot, G., Billette, T., Angelini, C., Barbosa, C., Ferriere, G., Lanzi, G., Ottolini, A., Babron, M. C., Cohen, D., Hanauer, A., Clerget-Darpoux, F., Lathrop, M., Munnich, A., and Frezal, J. (1990a). Gene for chronic proximal spinal muscular atrophies maps to chromosome 5q. *Nature* **344:** 767–768.

Melki, J., Sheth, P., Abdelhak, S., Burlet, P., Bachelot, M. F., Lathrop, M. G., Frezal, J., and Munnich, A. (1990b). Mapping of acute (type I) spinal muscular atrophy to chromosome 5q12–q14. *Lancet* **336:** 271–273.

Munsat, T. L. (1991). Workshop report: International SMA collaboration. *Neuromuscular Disord.* **1:** 81.

Munsat, T. L. (1992). Meeting report: international SMA consortium meeting. *Neuromuscular Disord.* **2:** 423–428.

Pearn, J. H. (1982). Infantile motor neuron diseases. *In* "Advances in Neurology" (L. P. Rowland, Ed.), pp. 121–130, Raven Press, New York.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sargent, C. A., Chalmers, J., Leversha, M., Affara, N. A., and Fergus-

on-Smith, M. A. (1991). Identification of a novel repeat element mapping to 4 loci on chromosome 5. *Cytogenet. Cell Genet.* **58:** 1901.

Sherrington, R., Melmer, G., Dixon, M., Curtis, D., Mankoo, B., Kalsi, G., and Gurling, H. (1991). Linkage disequilibrium between two highly polymorphic microsatellites. *Am. J. Hum. Genet.* **49:** 966–971.

Simard, L. R., Vanasse, M., Rochette, C., Morgan, K., Lemieux, B., Melancon, S. B., and Labuda, D. (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.

Soares, V. M., Brzustowicz, L. M., Kleyn, P. W., Knowles, J. A., Palmer, D. A., Asokan, S., Penchaszadeh, G. K., Munsat, T. L., and Gilliam, T. C. (1993). Refinement of the spinal muscular atrophy locus to the interval between D5S435 and MAP1B. *Genomics* **15:** 365–371.

Sokal, R. R., and Rohlf, F. J. (1969). "Biometry," p. 599, Freeman, San Francisco.

Thompson, T. G., Morrison, K. E., Kleyn, P., Bengtsson, U., Gilliam, T. C., Davies, K. E., Wasmuth, J. J., and McPherson, J. D. (1993). High resolution physical map of the region surrounding the spinal muscular atrophy gene. *Hum. Mol. Genet.* **2:** 1169–1176.

Warrington, J. A., Hall, L. V., Hinton, L. M., Miller, J. N., Wasmuth, J. J., and Lovett, M. (1991). Radiation hybrid map of 13 loci on the long arm of chromosome 5. *Genomics* **11:** 701–711.

Weir, B. R. (1990). "Genetic Data Analysis," Sinauer Associates, Sunderland, MA.

Wirth, B., Voosen, B., Röhrig, D., Knapp, M., Piechaczek, B., Rudnik-Schöneborn, S., and 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.

Wirth, B., Pick, E., Leutner, A., Dadze, A., Voosen, B., Knapp, M., Piechaczek-Wappenschmidt, B., Rudnik-Schöneborn, S., Schönling, J., Cox, S., Spurr, N. K., and Zerres, K. (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.

Zelterman, D. (1993). A semiparametric bootstrap technique for simulating extreme order statistics. *J. Am. Stat. Soc.* **88:** 477–485.