

## Myotonic Dystrophy Type 2: Human Founder Haplotype and Evolutionary Conservation of the Repeat Tract

Christina L. Liquori,<sup>1,2</sup> Yoshio Ikeda,<sup>1,2</sup> Marcy Weatherspoon,<sup>1,2</sup> Kenneth Ricker,<sup>4</sup> Benedikt G. H. Schoser,<sup>5</sup> Joline C. Dalton,<sup>1,2</sup> John W. Day,<sup>1,3</sup> and Laura P. W. Ranum<sup>1,2</sup>

<sup>1</sup>Institute of Human Genetics, Departments of <sup>2</sup>Genetics, Cell Biology, and Development, and <sup>3</sup>Neurology, University of Minnesota, Minneapolis; <sup>4</sup>Department of Neurology, University of Würzburg, Würzburg, Germany; and <sup>5</sup>Friedrich-Baur-Institute, Department of Neurology, Ludwig-Maximilians-University, Munich

Myotonic dystrophy (DM), the most common form of muscular dystrophy in adults, can be caused by a mutation on either chromosome 19 (DM1) or 3 (DM2). In 2001, we demonstrated that DM2 is caused by a CCTG expansion in intron 1 of the zinc finger protein 9 (*ZNF9*) gene. To investigate the ancestral origins of the DM2 expansion, we compared haplotypes for 71 families with genetically confirmed DM2, using 19 short tandem repeat markers that we developed that flank the repeat tract. All of the families are white, with the majority of Northern European/German descent and a single family from Afghanistan. Several conserved haplotypes spanning >700 kb appear to converge into a single haplotype near the repeat tract. The common interval that is shared by all families with DM2 immediately flanks the repeat, extending up to 216 kb telomeric and 119 kb centromeric of the CCTG expansion. The DM2 repeat tract contains the complex repeat motif (TG)<sub>n</sub>(TCTG)<sub>n</sub>(CCTG)<sub>n</sub>. The CCTG portion of the repeat tract is interrupted on normal alleles, but, as in other expansion disorders, these interruptions are lost on affected alleles. We examined haplotypes of 228 control chromosomes and identified a potential premutation allele with an uninterrupted (CCTG)<sub>20</sub> on a haplotype that was identical to the most common affected haplotype. Our data suggest that the predominant Northern European ancestry of families with DM2 resulted from a common founder and that the loss of interruptions within the CCTG portion of the repeat tract may predispose alleles to further expansion. To gain insight into possible function of the repeat tract, we looked for evolutionary conservation. The complex repeat motif and flanking sequences within intron 1 are conserved among human, chimpanzee, gorilla, mouse, and rat, suggesting a conserved biological function.

### Introduction

Myotonic dystrophy (DM) is an autosomal dominant, multisystemic disease with a consistent constellation of clinical features, including myotonia, muscular dystrophy, cardiac-conduction defects, posterior iridescent cataracts, and endocrine disorders (Harper 2001). In 1992, the mutation for myotonic dystrophy type 1 (DM1 [MIM 160900]) on chromosome 19 was shown to be caused by an expanded CTG repeat in the 3' UTR of the dystrophin myotonia-protein kinase (*DMPK*) gene (Brook et al. 1992; Buxton et al. 1992; Fu et al. 1992; Harley et al. 1992; Mahadevan et al. 1992). In 1998, we mapped the myotonic dystrophy type 2 (DM2 [MIM 602668]) locus to 3q21 (Ranum et al. 1998), and we

recently identified the mutation as a CCTG expansion in intron 1 of the zinc finger protein 9 (*ZNF9*) gene (Liquori et al. 2001). In Europe, the disease in these families has been called “proximal myotonic myopathy” (PROMM [MIM 600109]) (Ricker et al. 1994) or “proximal myotonic dystrophy” (PDM) (Udd et al. 1997), and, in the United States, these families were initially described as “having myotonic dystrophy with no CTG repeat expansion” (Thornton et al. 1994; Day et al. 1999). Clinical and molecular parallels between DM1 and DM2 indicate that CUG and CCUG microsatellite expansions in RNA can be pathogenic and can cause the multisystemic features of both diseases (Day et al. 1999, 2003; Ricker 1999; Liquori et al. 2001; Ranum and Day 2002).

The DM2 repeat tract is a complex repeat motif with an overall configuration of (TG)<sub>n</sub>(TCTG)<sub>n</sub>(CCTG)<sub>n</sub>. In the general population, the DM2 repeat tract is stable when transmitted from one generation to the next, and the CCTG portion of the repeat tract is interrupted, most commonly by both GCTG and TCTG motifs—that is, (TG)<sub>n</sub>(TCTG)<sub>n</sub>(CCTG)<sub>n</sub>GCTG CCTG TCTG (CCTG)<sub>n</sub> (Liquori et al. 2001). On the expanded alleles that have been sequenced, an uninterrupted variant of the CCTG

Received May 29, 2003; accepted for publication July 18, 2003; electronically published September 22, 2003.

Address for correspondence and reprints: Dr. Laura P. W. Ranum, Department of Genetics, Cell Biology and Development and the Institute of Human Genetics, MMC 206, 420 Delaware Street SE, University of Minnesota, Minneapolis, MN 55455. E-mail: ranum001@umn.edu

© 2003 by The American Society of Human Genetics. All rights reserved. 0002-9297/2003/7304-0013\$15.00

**Table 1****Primer Sequences for Newly Developed STR Markers**

Marker	Accession Number	Forward Primer Sequence (5'→3')	Reverse Primer Sequence (5'→3')
CL3N105	AY326637	TCA GTC AAG GTT CCA CCA GA	GCC AGT CCT CAT AGC CAC AT
CL3N122	AY326638	TGC TCC ACA TCC TAA CAC CA	TGG GCA AAG GAC CTA AAA AG
CL3N97	AY326639	AGG CAG GAG AAT CGT TTG AA	CAG TCT CCC GAA TAG CTG GA
CL3N84	AY326640	TCA TTC CCA GAC GTC CTT TC	AAT CGC TTG AAC CTG GAA GA
CL3N99	AY326641	CTG CCG GTG GGT TTT AAG T	TGC AAG ACG GTT TGA AGA GA
CL3N94	AY326642	TCA GCT GTG TTC ATG CCA TT	AGA GGG AGT GGG TTT GGT TT
CL3N83	AY326643	GTG TGT AAG GGG GAG ACT GG	AAG CCC AAG TGG CAT TCT TA
CL3N95	AY326644	CTG GTC TCG AAC TCC TGA GC	TCT TCT CTA GTT GGC TAT TTC ATT CA
CL3N96	AY326645	CCC ATT CTC TCC CTC TGT CA	TAC CAT GGC TCC CAG AGT TC
CL3N59	AY326646	GCT GGC ACC TTT TAC AGG AA	ATT TGC CAC ATC TTC CCA TC
CL3N58D	AF388525	GCC TAG GGG ACA AAG TGA GA	GGC CTT ATA ACC ATG CAA ATG
CL3N114B	AY326647	TTA ATT GAT TCA AGG ACA TTT GTA GTC T	AAA AAT TAG CTA AGC GTG GTG ACA G
CL3N116	AY326648	AGA TGG GCA AAT GGA AAG TG	GGG CGA GAC TCT GTC TCA AA
CL3N117	AY326649	TCA TCC CAA ACT GAA TCC TCA	CTG GTG ATG CCT TGG AAA AT
CL3N118	AY326650	TCA CAT GCA TTG CCT ACC AT	CAC AGA AAA CTG CAC CCA GA
CL3N119	AY326651	CAA GCA AAT GTT CCC TGA CA	AGG GTC AGA CAG AGC TGG AA
CL3N121	AY326652	AGC CAC TTC ACT CCA GCC TA	GTA CAG CAG GGC CTT GTA GC
CL3N19	AY326653	GAC AGG CCA AAG TGT GTG TG	GGC AAA AAT AAA GCC ACA GC
CL3N23	AY326654	GCC CTT CCC AAC TAT GTA GC	CCA GTC TGG GTG ACA AAG TG

portion of the repeat tract is elongated (Liquori et al. 2001). For several diseases, the loss of sequence interruptions is thought to predispose normal alleles to expansion (Chung et al. 1993; Kunst and Warren 1994; Gunter et al. 1998). Expanded DM2 alleles show unprecedented somatic instability, with significant increases in length over time (e.g., 2,000 bp/3 years) and expanded alleles often appearing as smears by Southern analysis (Liquori et al. 2001; Day et al. 2003).

Linkage disequilibrium played an important role in the identification of the DM2 mutation (Liquori et al. 2001). Several markers within the DM2 critical region were in linkage disequilibrium with the mutation, as shown by transmission/disequilibrium test analysis, including *CL3N58* ( $P < .000001$ ), *CL3N59* ( $P = .0001$ ), *CL3N84* ( $P = .0001$ ), and *CL3N99* ( $P = .0075$ ) (Liquori et al. 2001). These data suggested the possibility that the DM2 expansions may have originated from one or more founder mutations. This would be consistent with DM1, in which expanded CTG alleles in individuals of European ancestry are thought to have arisen from a single 30-kb haplotype that includes nine polymorphic markers (Imbert et al. 1993; Neville et al. 1994; Chakraborty et al. 1996). Several extragenic polymorphisms can extend this associated haplotype up to 160 kb from the expansion, although these polymorphisms are not in complete linkage disequilibrium with expanded alleles.

To further investigate the ancestral origin(s) of the DM2 expansion, we have performed haplotype analyses of 71 families with DM2. In addition, the possibility that the normal repeat tract found in humans has a conserved biological function was investigated by studying the evo-

lutionary conservation of the repeat tract and flanking intron sequence in chimpanzee, gorilla, mouse, and rat.

## Materials and Methods

### Families with DM2

We identified, obtained informed consent from, performed neurological exams of, and collected blood samples from family members of patients with DM2/PROMM. The study was performed on 71 families, consisting of 301 individuals, including affected patients ( $n = 210$ ) and their unaffected relatives. All of the families with DM2 that we have identified are white (i.e., persons of European, North African, or Southwest Asian descent), with 70 families of Northern European descent and 1 from Afghanistan. Genomic DNA was isolated from blood with the Puregene kit D-5000 (Gentra Systems). For evolutionary conservation studies, genomic DNA from three common chimpanzees (*Pan troglodytes*) and two gorillas (*Gorilla gorilla*) was isolated from venous blood of animals at the Como Zoo (St. Paul, MN), and DNA from mouse (FVB/N) was isolated from tail snips with the Puregene kit D-5000 (Gentra Systems).

### Genetic Confirmation of Families with DM2

Affected individuals were determined to be DM2-positive by one or more of the following: (i) PCR, by use of the primers CL3N58D F and CL3N58D R, published elsewhere (GenBank accession number AF388525), which resulted in a blank allele pattern, as described elsewhere (Liquori et al. 2001); (ii) Southern blot analysis,

which resulted in detection of an expanded allele (Liquori et al. 2001; Day et al. 2003); and (iii) repeat assay test, performed as described elsewhere (Day et al. 2003).

*Development of STR Markers in the DM2 Region*

Additional polymorphic markers were developed by designing PCR primers that flank STR sequences identified by analysis of available sequence information from the UCSC Genome Browser and National Center for Biotechnology Information (NCBI) databases. PCR primers and allele frequencies are shown in tables 1 and 2, respectively. Allele frequencies were calculated for each marker by use of the alleles from spouses in all families analyzed. When sufficient data were available, control haplotypes were determined for these same individuals. Marker positions relative to ZNF9 and the DM2 expansion are based on sequence information from UCSC and NCBI (fig. 1).

*Haplotype Analysis*

PCRs for all markers were done with a 5- $\mu$ l reaction (200 mM deoxynucleotide triphosphates [dNTPs], 10 mM tris-HCl [pH 9.0], 50 mM KCl, 0.1% Triton X-100, 0.01% [weight/volume] gelatin, 1/1.5/2 mM MgCl<sub>2</sub>, 0.4 mM each primer, 0.1 U *Taq*) cycled 35 times (94°C for

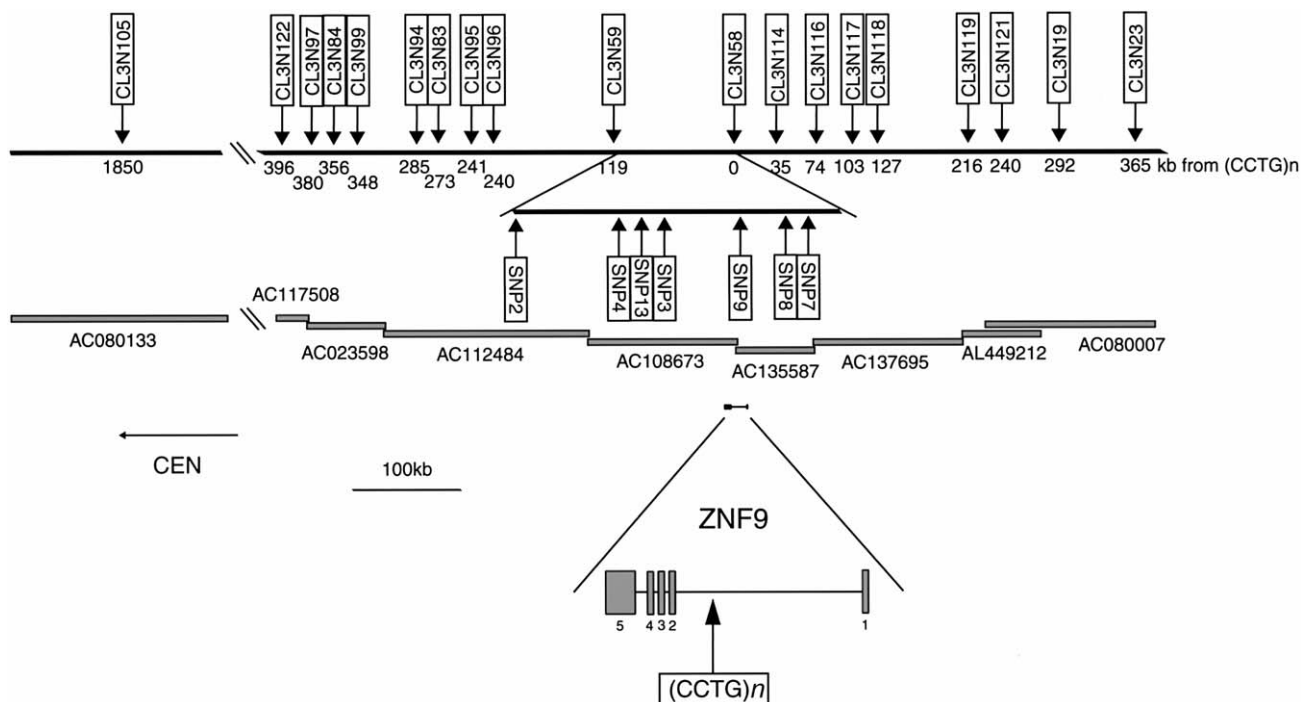
45 s, 51°C/54°C/57°C for 45 s, and 72°C for 60 s). For each reaction, the forward (F) primer was end labeled with 33 $\gamma$ P adenosine triphosphate with T4 polynucleotide kinase (Epicentre). PCR products were run on 4% acrylamide (8M urea) gel at 65 watts.

Affected haplotypes were established by determining which allele cosegregated with the disease in each family. Control haplotypes were constructed by determining the alleles that were passed from parents to their offspring. In the event that the associated allele for a marker could not be unequivocally determined, both alleles are given (fig. 2 [top]).

Haplotypes are given in the order *CL3N105-CL3N122-CL3N97-CL3N84-CL3N99-CL3N94-CL3N83-CL3N95-CL3N96-CL3N59-CL3N58 (DM2-EXP)-CL3N114-CL3N116-CL3N117-CL3N118-CL3N119-CL3N121-CL3N19-CL3N23*, with alleles designated by their size in base pairs and the DM2 expansion designated by “exp” (fig. 2 [top]).

*SNP Analysis*

Potential SNPs were identified on the basis of information from the UCSC Genome Browser and NCBI SNP database. SNPs were analyzed by PCR amplification (as above, except that reactions were performed in 15- $\mu$ l vol-



**Figure 1** Map of newly developed polymorphic markers in DM2 region. We generated a series of novel polymorphic genetic markers containing STRs and SNPs. The markers and relative map positions are indicated with the distance in kilobases from the CCTG repeat given below the marker. The seven SNPs located between markers CL3N59 and CL3N58 are indicated below the solid bar. Nine BACs, spanning >2 Mb, are shown (dark gray horizontal boxes).

**Table 2****Allele Frequencies of STR Markers**

Marker and Alleles in bp <sup>a</sup>	%
CL3N105 ( <i>n</i> = 193):	
240	.5
238	.5
236	1.0
234	4.7
<b>232</b>	<b>13.0</b>
230	15.0
228	20.7
226	17.1
<b>224</b>	<b>16.6</b>
222	10.9
CL3N122 ( <i>n</i> = 80):	
219	1.3
215	5.0
213	5.0
<b>211</b>	<b>25.0</b>
<b>209</b>	<b>46.1</b>
207	15.0
205	1.3
203	1.3
CL3N97 ( <i>n</i> = 128):	
202	.8
198	11.7
<b>194</b>	<b>19.5</b>
<b>190</b>	<b>65.6</b>
186	.8
182	.8
178	.8
CL3N84 ( <i>n</i> = 201):	
<b>157</b>	<b>15.4</b>
155	.5
153	1.5
<b>151</b>	<b>57.8</b>
149	12.4
147	11.9
145	.5
CL3N99 ( <i>n</i> = 209):	
191	.5
185	.5
183	1.0
<b>181</b>	<b>1.4</b>
179	2.9
<b>177</b>	<b>6.2</b>
175	15.3
173	5.3
<b>171</b>	<b>18.6</b>
169	18.2
167	6.2
165	16.7
163	1.4
161	1.9
159	1.9
157	1.0
155	.5
149	.5

*(continued)***Table 2 (continued)**

Marker and Alleles in bp	%
CL3N94 ( <i>n</i> = 202):	
162	1.5
160	6.9
158	6.4
154	3.0
<b>152</b>	<b>50.0</b>
150	2.0
<b>148</b>	<b>29.2</b>
144	1.0
CL3N83 ( <i>n</i> = 213):	
197	1.4
<b>195</b>	<b>.5</b>
193	.9
191	4.7
<b>189</b>	<b>15.5</b>
187	14.6
<b>185</b>	<b>48.7</b>
183	3.3
181	1.9
173	5.2
171	3.3
CL3N95 ( <i>n</i> = 202):	
236	1.0
232	.5
228	.5
224	6.9
<b>222</b>	<b>23.8</b>
220	.5
216	1.5
214	13.9
<b>212</b>	<b>35.0</b>
210	1.5
206	13.4
204	1.5
CL3N96 ( <i>n</i> = 207):	
163	.5
161	9.2
<b>159</b>	<b>23.7</b>
<b>157</b>	<b>51.1</b>
155	15.5
CL3N59 ( <i>n</i> = 202):	
151	.5
<b>149</b>	<b>5.0</b>
147	.5
145	6.4
143	10.9
<b>141</b>	<b>31.7</b>
139	20.3
137	15.3
135	7.9
133	.5
131	1.0
CL3N114 ( <i>n</i> = 82):	
297	1.2
292	1.2
<b>291</b>	<b>7.3</b>

*(continued)*

**Table 2 (continued)**

Marker and Alleles in bp	%
290	2.4
289	8.5
288	69.6
287	9.8
CL3N116 ( <i>n</i> = 68):	
244	2.9
226	1.5
224	4.4
222	89.7
218	1.5
CL3N117 ( <i>n</i> = 78):	
241	2.6
239	82.0
237	2.6
235	12.8
CL3N118 ( <i>n</i> = 74):	
229	9.5
227	78.3
221	12.2
CL3N119 ( <i>n</i> = 78):	
244	1.3
242	1.3
240	1.3
238	1.3
236	3.8
234	11.5
232	28.2
230	35.9
228	3.8
226	10.3
222	1.3
CL3N121 ( <i>n</i> = 80):	
231	13.8
227	44.9
223	41.3
CL3N119 ( <i>n</i> = 72):	
207	1.4
205	4.2
203	5.6
201	9.7
199	18.1
197	15.3
195	22.0
193	4.2
191	4.2
181	12.5
179	2.8
CL3N23 ( <i>n</i> = 70):	
238	7.1
235	1.4
234	8.6
231	1.4
230	12.9
226	24.4
222	27.1
218	5.7
216	1.4
214	8.6
210	1.4

NOTE.—Alleles associated with the affected haplotype are shown in bold italics.

<sup>a</sup> *n* = number of chromosomes analyzed.

umes and without labeling the F primer) of a 400–600-bp region surrounding the SNP of interest, followed by restriction enzyme digestion (1 U/sample) for 3 h at appropriate conditions. The digested products were run out on a 1%–2% agarose gel containing EtBr (1 μl/100 ml gel) at 150 volts.

SNP genotypes are shown in table 3, and PCR primers and restriction-enzyme information are summarized in table 4, with NCBI reference sequence numbers given in parentheses. All restriction enzymes were from New England Biolabs, except for *MaeIII*, which is from Roche Molecular Biochemicals.

#### Sequencing of Normal DM2 Alleles

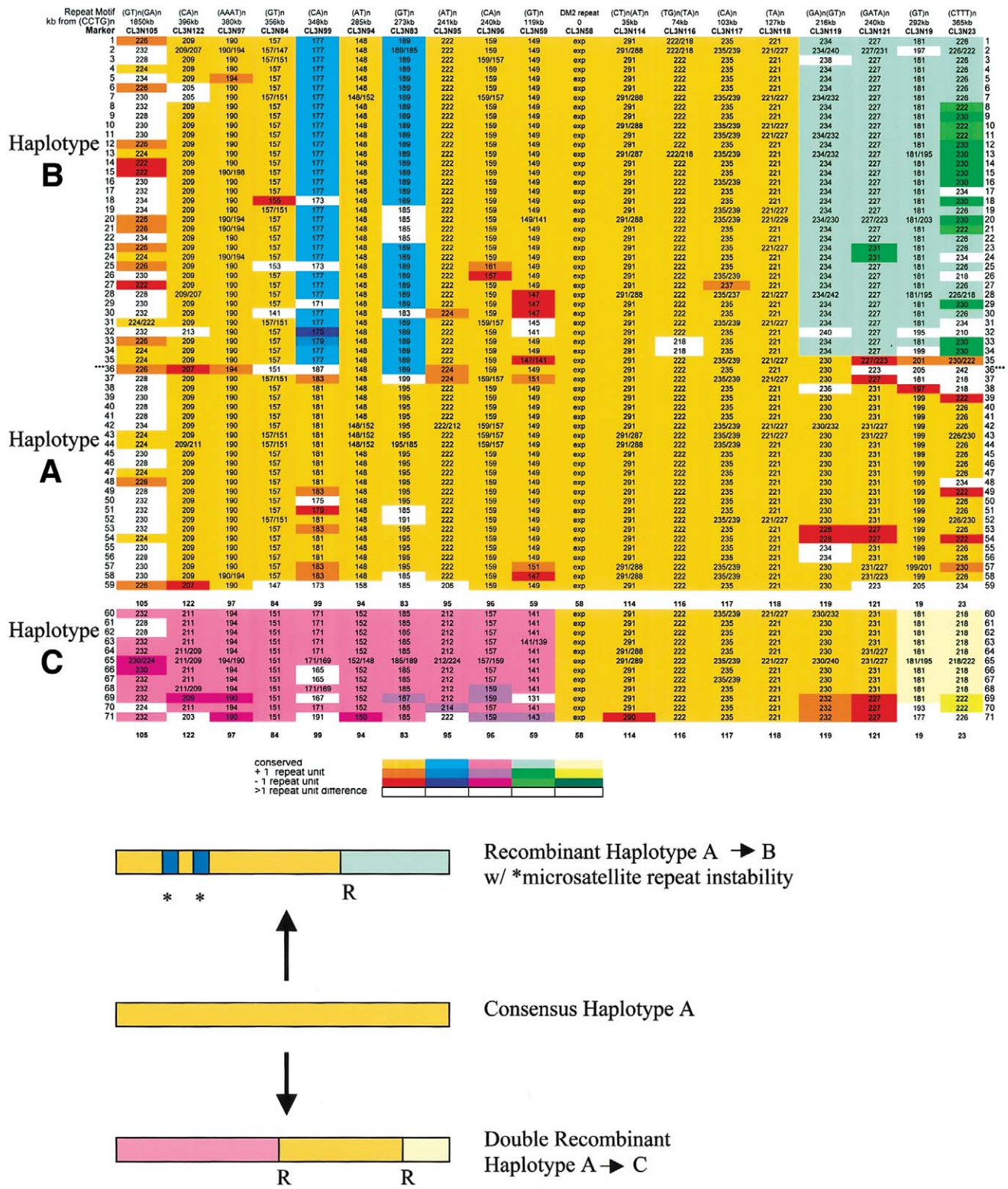
PCR amplification and cloning of normal DM2 (*CL3N58*) alleles was performed as described elsewhere (Liquori et al. 2001). To simplify the sequencing, PCR products were preferentially amplified from affected offspring who inherited the normal haplotype of interest, as only a single allele within the normal range would be amplified. For each allele, the PCR and subsequent sequencing was performed at least twice (fig. 3).

#### Intron Analysis in Chimpanzee and Gorilla

DNA immediately flanking the DM2 repeat was amplified from the genomic DNA of three chimpanzees and two gorillas with primers CL3N58D F and CL3N58D R, as described elsewhere (Liquori et al. 2001). A larger region of flanking sequence (1.4 kb) was amplified by PCR from genomic DNA with primers JJ1 F (5'-CCT TTG CAC ATC TTC CCA TA-3'), JJ1 R (5'-TGG CAG TAA TAC TCA TTC ACT CA-3'), JJ3 F (5'-GCT GGA GTG CAC TGG TAT GA-3'), and JJ3 R (5'-CAG CTA CTT GGG AGG CTG AG-3') in a 15-μl reaction, with an annealing temperature of 54°C and 1.0 mM MgCl<sub>2</sub>. Alleles amplified by PCR were cloned with the TOPO cloning kit (Invitrogen), and at least five independent clones were sequenced for each individual. In some cases, the PCR products were sequenced directly.

#### Mouse CNBP Intron 1 Sequencing

BACs containing murine homolog of *ZNF9* (cellular nucleic acid binding protein [CNBP]) were identified from RPCI-23 mouse BAC high-density membranes (Invitrogen). A probe to mouse exon 5 was generated by PCR with primers mouse E5 F (5'-GGT GAA ACT GGT CAT GTA GCC-3') and mouse E5 R (5'-TGG GAA CAG CCT CTA TCT GC-3'), with an annealing temperature of 54°C and 1.0 mM MgCl<sub>2</sub>, labeled with <sup>32</sup>P and hybridized. BACs were identified as described in the Research Genetics protocol. All positive BACs were grouped by fingerprint maps (NCBI Clone Registry); two BACs were chosen from each group, and each was checked by PCR with the mouse E5 primers. Two clones containing mouse



**Figure 2** *Top*, Haplotypes of 71 families with DM2. The three major affected haplotypes (A, B, and C) found in 71 families with DM2 analyzed are shown. The consensus haplotype A is indicated in gold. Minor deviations in repeat size are indicated by alternative colors, with a color key located below the figure. The markers span 2.2 Mb, and the distance of each marker from the DM2 CCTG expansion is denoted at the top of figure. The STR marker name and the repeat motif associated with each are designated. *Bottom*, Proposed ancestral origin of DM2 haplotypes. Proposed ancestral relationship between the major haplotype variants is shown, with haplotype B and C related to haplotype A, by a small number of ancestral recombination and microsatellite instability events.

**Table 3****SNPs Associated with DM2-Affected Haplotypes**

Marker	Accession Number	Distance from (CCTG) <i>n</i> (kb)	Haplotypes		Frequency	
			A/B	C	(%)	Number
CL3N59	AY326646	119.1	149	141		
SNP2	rs762570	109.5	T	T	100	36
SNP4	rs970572	65.1	G	G	99.4	160
SNP13	rs2128342	58.0	G	G	100	36
SNP3	rs1021123	54.3	T	T	100	34
SNP9	rs2062125	29.4	T	T	98.9	178
SNP8	rs1482409	14.0	T	T	96.4	110
SNP7	rs1351596	10.2	C	C	98.8	168
CL3N58	AF388525	0	Expansion	Expansion		

CNBP—RP23-398k17 and RP23-91i11—were identified and grown in 500 ml Luria broth media overnight at 37°C, and DNA was prepared using the QIAGEN large Construct kit (QIAGEN), as described. Sequencing was performed and sequence alignments were done with GeneTool Lite, v. 1.0, (DoubletWist) or BLAST.

## Results

### Haplotype Analysis of 71 Families with DM2

Haplotype analysis was performed on 71 families with DM2. All affected individuals received a diagnosis of DM and tested positive for the DM2 expansion. The families from Germany ( $n = 60$ ) and Minnesota ( $n = 10$ ) are all of European descent, and there was one additional family from Afghanistan (denoted by a triple asterisk [\*\*\*] in fig. 2 [top]). Nineteen STR markers were developed and used in the analysis (table 1). These markers span a region of 2.2 Mb, now with complete coverage by BACs (fig. 1). The CL3N58 marker contains the DM2 CCTG repeat tract, which is expanded in affected individuals.

The majority of families with DM2 share one of three common haplotypes (A, B, or C), each of which spans a core region of ~700 kb between the CL3N122 and CL3N19 markers. In each case, the haplotypes significantly diverge at the flanking markers CL3N105 and CL3N23 (fig. 2). Of the families with DM2, 21 have the conserved haplotype A (shown in gold). Thirty-four families have a second related haplotype (haplotype B), likely derived from haplotype A by an ancestral recombination event involving the four most distal markers and microsatellite repeat slippage at CL3N99 and CL3N83. The large numbers of alleles with differing sizes at the CL3N99 and CL3N83 markers within both the highly related A and B haplotype blocks suggest that these microsatellite markers, (CA)<sub>23-36</sub> and (GT)<sub>14-28</sub>, show higher rates of mutability than neighboring markers. The third major DM2 haplotype, designated “haplotype C,” which is shared by nine families, appears to have had two recombination events, the first between the DM2

repeat tract and the next centromeric marker located 119 kb proximal to the mutation and the second between the CL3N121 and CL3N19 markers on the telomeric side. Our hypothesis that haplotypes B and C arose from the common consensus haplotype A by several ancestral recombination events is shown in figure 2 [bottom]. Six of the seven additional families (35, 36, 37, 59, 69, and 70) have variations of the major haplotypes that can be explained by similar recombination and microsatellite instability events that have occurred in the flanking regions. It is possible that the haplotype found in family 71 is related to that found in families 69 and 70 on the distal side of the repeat tract and that the variation in repeat size for the CL3N114 marker is due to microsatellite instability. Alternatively, the mutation in this family could have arisen independently.

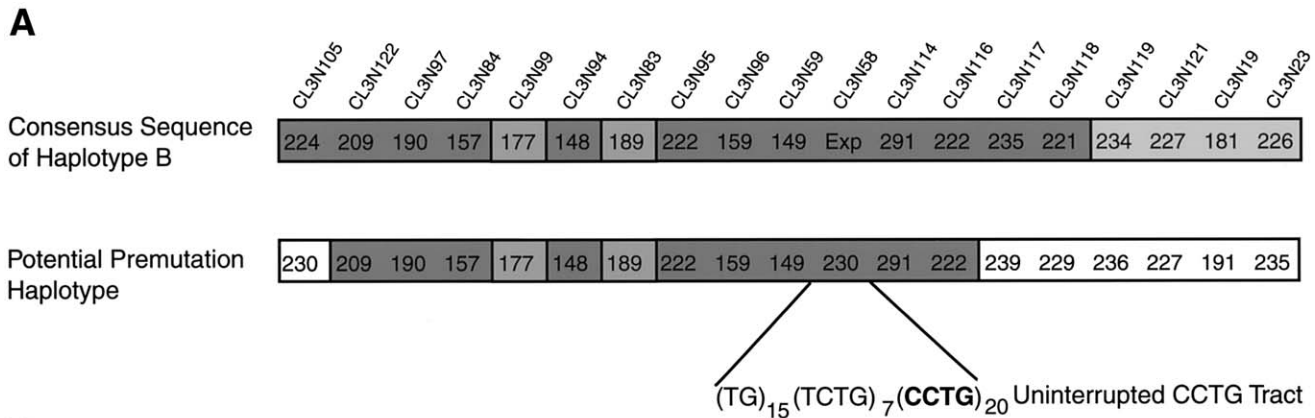
### DM2 Haplotypes Converge, Suggesting a Single Founder

Haplotype A shares seven consecutive STR markers flanking the repeat with haplotype B and six consecutive STR markers adjacent to the telomeric side of the repeat tract with haplotype C. On all three haplotypes, markers CL3N114, CL3N116, CL3N117, and CL3N118 (located 35, 74, 103, and 127 kb telomeric to the repeat tract) are conserved, suggesting that the haplotypes converge near the DM2 mutation.

Haplotype convergence, on the basis of the STR marker data presented in figure 2 (top), is consistent with a group of SNP markers that were typed within 109.5 kb of the DM2 mutation. Although the SNP haplotype was also conserved, it was far less informative and, on the basis of allele frequencies, would be predicted to be found on 94% of chromosomes.

### Possible Premutation Allele Found on Control Chromosome

To determine how common the DM2 expanded allele-associated haplotypes are in the general population, we examined the haplotypes of 228 control chromosomes

**B**

## Sequence Analysis of Control Chromosomes

(TG)<sub>15</sub> (TCTG)<sub>8</sub> (CCTG)<sub>5</sub> GCTG CCTG TCTG (CCTG)<sub>7</sub> *n*=1  
 (TG)<sub>16</sub> (TCTG)<sub>9</sub> (CCTG)<sub>5</sub> GCTG CCTG TCTG (CCTG)<sub>7</sub> *n*=4  
 (TG)<sub>17</sub> (TCTG)<sub>9</sub> (CCTG)<sub>5</sub> GCTG CCTG TCTG (CCTG)<sub>7</sub> *n*=1  
 (TG)<sub>18</sub> (TCTG)<sub>9</sub> (CCTG)<sub>5</sub> GCTG CCTG TCTG (CCTG)<sub>7</sub> *n*=4  
 (TG)<sub>19</sub> (TCTG)<sub>10</sub> (CCTG)<sub>5</sub> GCTG CCTG TCTG (CCTG)<sub>7</sub> *n*=1  
 (TG)<sub>20</sub> (TCTG)<sub>7</sub> (CCTG)<sub>6</sub> GCTG CCTG TCTG (CCTG)<sub>7</sub> *n*=1  
 (TG)<sub>20</sub> (TCTG)<sub>8</sub> (CCTG)<sub>6</sub> GCTG CCTG TCTG (CCTG)<sub>7</sub> *n*=1  
 (TG)<sub>20</sub> (TCTG)<sub>9</sub> (CCTG)<sub>5</sub> GCTG CCTG TCTG (CCTG)<sub>7</sub> *n*=1  
 (TG)<sub>20</sub> (TCTG)<sub>11</sub> (CCTG)<sub>5</sub> GCTG CCTG TCTG (CCTG)<sub>7</sub> *n*=1  
 (TG)<sub>21</sub> (TCTG)<sub>10</sub> (CCTG)<sub>4</sub> GCTG CCTG TCTG (CCTG)<sub>7</sub> *n*=1

## Previously Reported Normal Alleles (Liquori et al. 2001)

(TG)<sub>14-25</sub>(TCTG)<sub>4-9</sub> (CCTG)<sub>5</sub> GCTG CCTG TCTG (CCTG)<sub>4-7</sub> *n*=5  
 (TG)<sub>21</sub> (TCTG)<sub>8</sub> (CCTG)<sub>9</sub> TCTG (CCTG)<sub>3</sub>TCTG (CCTG)<sub>8</sub> *n*=1  
 (TG)<sub>15-17</sub>(TCTG)<sub>7-10</sub>(CCTG)<sub>3-7</sub>TCTG (CCTG)<sub>7</sub> *n*=2

## Expanded Alleles with Uninterrupted CCTG Tract (Liquori et al. 2001)

(TG)<sub>16</sub> (TCTG)<sub>9</sub> (CCTG)<sub>*n*</sub> *n*=3

## Possible Uninterrupted Premutation Allele

(TG)<sub>15</sub> (TCTG)<sub>7</sub> (CCTG)<sub>20</sub> *n*=1

**Figure 3** A, Identification of potential premutation allele in the general population. The consensus between haplotype B and a possible premutation allele in an unaffected control is depicted. Identical coloring indicates the regions where the haplotypes are shared. The sequence of the DM2 expansion region in the possible premutation allele is indicated below the haplotype. B, The schematic diagram showing the repeat configurations of the DM2 expansion region on 24 normal control alleles, 3 expanded affected alleles, and the putative premutation allele.

from unrelated spouses and CEPH family controls. A total of 110 distinct haplotypes was observed, spanning eight markers from CL3N84 to CL3N58 (380 kb). The most common haplotype found in the general population

(*n* = 15; 6.6%) was 151-169-152-185-212-157-141-224, but the majority of haplotypes represented (*n* = 72) were unique.

One control chromosome (0.44%) had a haplotype



**Table 4**  
Summary of SNP Markers

Marker and Primer Sequence (5'→3) <sup>a</sup>	Enzyme	NT <sup>b</sup>	Cut	Product (bp)
SNP2 (rs762570): F: GCC TTC TGA CCT CTC TGG AA R: AAC ATC TGC ACC TCC AAA CC	<i>Bsu36I</i>	C	Yes No	291, 206 497
SNP4 (rs970572): F: CCA AAG TAG CAG GGA ACT GG R: AAA GGT GCA CTG TGG GAA AC	<i>BstUI</i> <i>BssSI</i>	A A G	No Yes Yes No	495 272, 223 276, 219 495
SNP13(rs2128342): F: TCA TGA GAA AAG CGA TGG AA R: TTG GTT TCT CAG CCT AGT TGC	<i>RsaI</i>	A G	Yes No	285, 118, 59 344, 118
SNP3 (rs1021123): F: ACG GGA AGA GAT GCA AAG GT R: GCT TTC CCT GTG ATT CTT TAC AA	<i>MaeIII</i>	C T	Yes No	314, 173 487
SNP9 (rs2062125): F: CAG AGT CTC GCT TCA TCA CC R: TCC GTC GCA AAA ACA AAA AT	<i>HinP1I</i>	C T	Yes No	329, 126 455
SNP8 (rs1482409): F: ACC TCA GCC TCC CCA AGT AT R: CAG GGG TTG TGA GAG TGA CA	<i>SspI</i>	C T	No Yes	502 277, 225
SNP7 (rs1351596): F: GGA ATG ACC CTG CTT GTT TC R: CTA GTG ATG GGC ACC CTA GC	<i>Tsp45I</i> <i>MaeIII</i>	C T	Yes No	267, 133, 104 371, 133

<sup>a</sup> F = forward primer; R = reverse primer.  
<sup>b</sup> NT = nucleotide.

identical to a 470-kb portion of haplotype B, spanning the DM2 repeat tract from markers CL3N122 to CL3N116, with a probable recombination event and a distinct haplotype beginning ~103 kb telomeric to the DM2 repeat tract (fig. 3A). Two others had partial matches to haplotype B, with minor deviations. Similarly, one control chromosome (0.44%) had a haplotype identical to the DM2 haplotype C, spanning the DM2 repeat from CL3N122 to CL3N116, with similar divergence. Thirteen other chromosomes had partial matches to haplotype C. To determine if any of these haplotypes could be related to their DM2 counterparts, we sequenced the normal DM2 alleles associated with these 17 control haplotypes (fig. 3B).

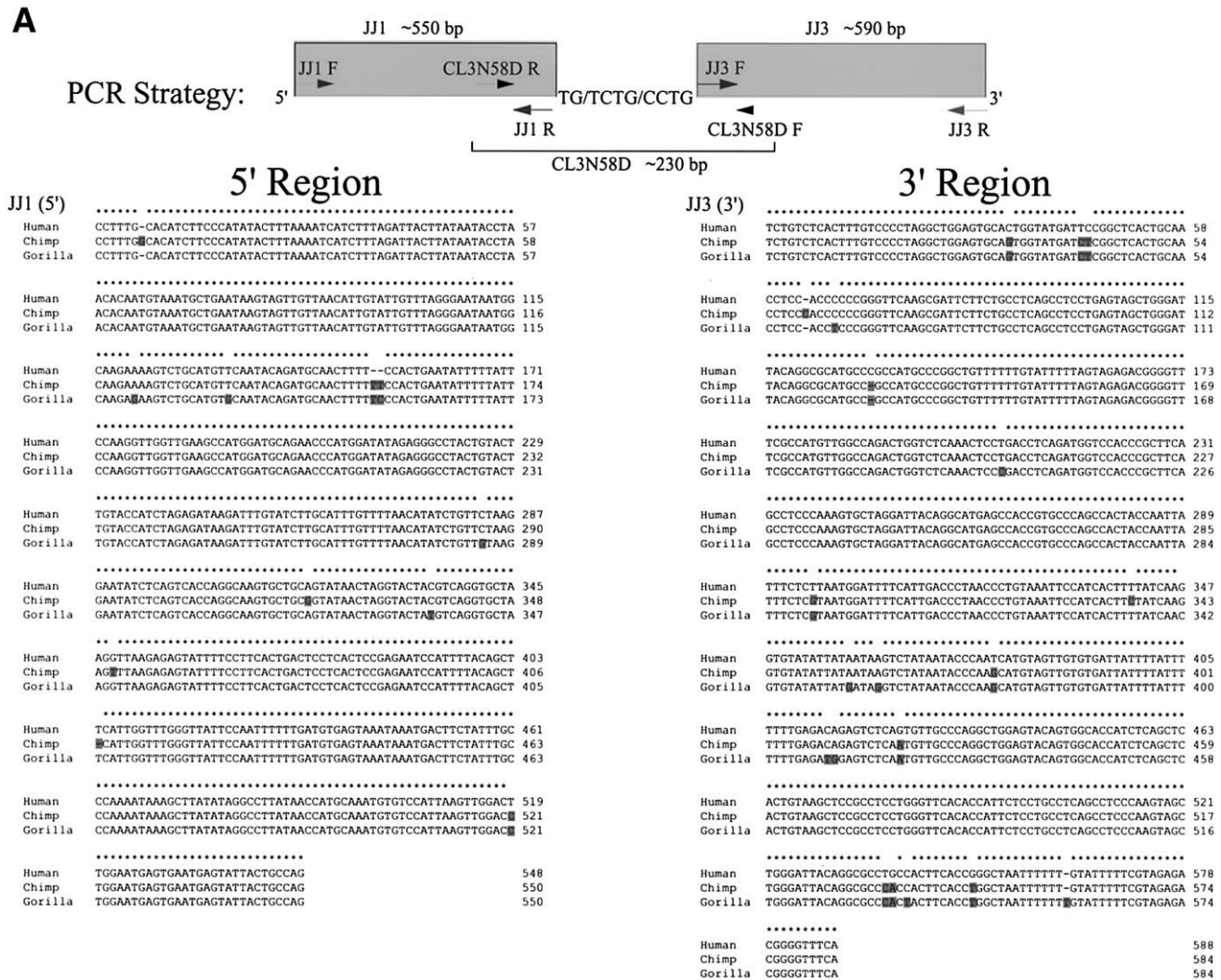
The CL3N58 alleles on the control haplotypes that were partial matches to either DM2 haplotype B or C

all had CCTG repeat tracts that resembled other normal alleles, including the presence of interruptions within the CCTG portion of the repeat tract. Thirteen control chromosomes with partial matches to haplotype B or C possess repeat tracts with the configuration (CCTG)<sub>5</sub>GCTG CCTG TCTG (CCTG)<sub>7</sub>. More size variation in both the TG and TCTG portions of the repeat tracts was observed, indicating that these two motifs are generally more polymorphic than the normally interrupted CCTG portion of the DM2 repeat tract (fig. 3B).

The CCTG tract on the haplotype identical to DM2 haplotype C was similar to the other control alleles and also contained an interrupted CCTG repeat with the configuration (CCTG)<sub>n</sub>GCTG CCTG TCTG (CCTG)<sub>n</sub>. In contrast, the CCTG tract on the haplotype identical to DM2 haplotype B lacked interruptions and is the longest



**Figure 4** Evolutionary comparisons of the DM2 repeat tract. Schematic diagram of DM2 repeat region shows consensus sequence configurations found in human, chimpanzee, gorilla, mouse, and rat.

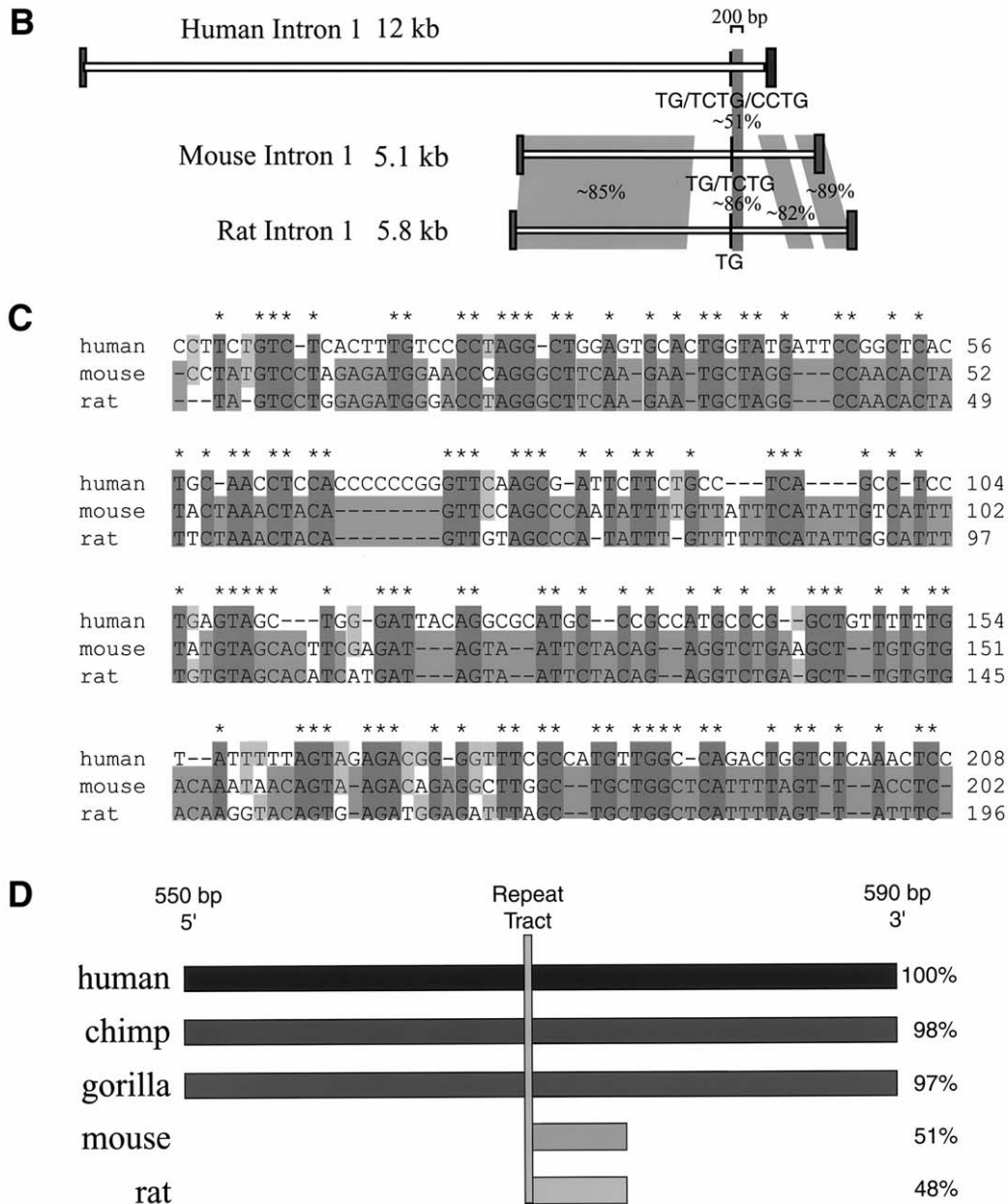


**Figure 5** Conservation of ZNF9 intron 1 regions. *A*, PCR strategy and alignment of the human, chimpanzee, and gorilla sequences 550-bp 5' and 590-bp 3' of the DM2 expansion regions. An asterisk (\*) indicates that the nucleotides at the position are identical in the three sequences. Gray shading indicates base pair differences from the human sequence. *B*, Schematic diagram showing the homology among the human, mouse, and rat intron 1. Gray shading represents regions of homology among human, mouse, and rat sequences, with percentage of homology indicated in each region. Dark gray boxes represent the flanking exons, and the gray vertical line represents the repeat tract. *C*, Alignment of human, mouse, and rat sequence 200-bp 3' of the respective repeats. Dark gray shading indicates that the nucleotides at the position are identical in the three sequences. Medium gray shading indicates nucleotides that are conserved between mouse and rat but differ from the human. Light gray shading indicates nucleotides that are identical between human and either the mouse or the rat but not both. *D*, Summary of regions of intron 1 homology among human, chimpanzee, gorilla, mouse, and rat.

known pure CCTG tract appearing within the normal population at (CCTG)<sub>20</sub> (fig. 3A and 3B). The haplotype conservation of 11 highly polymorphic STR markers >470 kb and the fact that this is the only control allele we have seen with an uninterrupted CCTG repeat suggests that this allele represents a pool of premutation alleles for DM2, which, because they lack sequence interruptions, are predisposed to further expansion.

*Conservation of the Repeat Tract and Intron 1*

We performed PCR on the genomic DNA from chimpanzee and gorilla using the primers for amplifying the human DM2 repeat (CL3N58D F and R) (Liquori et al. 2001). Sequencing of the CL3N58D products revealed that both primates had complex repeat motifs that were similar but not identical to the human DM2 repeat tract



(fig. 4). Chimpanzee has a  $(TG)_nTCTG (CCTG)_n$ , with interruptions in the TG rather than the CCTG tract (fig. 4). Gorilla has a  $(TG)_n(CCTG)_nTCTG CCTG (TCTG)_n$  and a  $(TCTG)_n(CCTG)_nTCTG CCTG TCTG$ , separated by a 38-nt stretch with homology to an *Alu* repeat (fig. 4). To determine if homology in the primate *ZNF9* introns continues beyond the repeat region, PCR was done on two regions flanking the repeat (~550 and 590 bp), again with primers designed for the human sequence (JJ1 F and R; JJ3 F and R). Both the 5' flanking region (chimpanzee 98.9%; gorilla 98.7%) and the 3' flanking region (chimpanzee 98.0%; gorilla 96.8%) are highly conserved (fig. 5A).

We further investigated conservation of the repeat tract and intron 1 by comparing the mouse and rat homologs of the *ZNF9* gene, known as “the cellular nucleic acid binding protein” (*CNBP*) gene, with the human homolog. For mouse, we identified two BAC clones containing the homolog of *ZNF9* (*CNBP*) by screening the BAC high-density membranes by hybridization. Both clones map to mouse chromosome 6, which is syntenic to the human DM2 region. Intron 1 of mouse *CNBP* was sequenced from the BAC clones (AY329620).

The sequence of intron 1 from the rat was determined by comparing rat mRNA sequence (D45254) with the rat genomic sequence (AC097129) by BLAST analysis. The

**Table 5****Intron 1 Sequence Information**

Species	Sequence Description	Accession Number
Human	Complete intron 1	AY329614
Mouse	Complete intron 1	AY329615
Rat	Complete intron 1	AY329616
Chimp 5'	JJ1	AY329608
Chimp 3'	JJ3	AY329609
Gorilla 5'	JJ1	AY329610
Gorilla 3'	JJ3	AY329611
Chimp 5'→3'	JJ1–JJ3, incl 58	AY329612
Gorilla 5'→3'	JJ1–JJ3, incl 58	AY329613
Human conserved		AY329617
Chimp conserved		AY329618
Gorilla conserved		AY329619
Mouse conserved		AY329620
Rat conserved		AY329621

size of intron 1 in both the mouse (5.1 kb) and the rat (5.8 kb) is much smaller than in the human gene (12 kb), with the mouse and rat introns conserved throughout their length (~85%) (fig. 5B). In contrast, only a small region, containing versions of the DM2 repeat tract and the sequence immediately 3' of the repeat tract, is shared among human, mouse, and rat (fig. 5B–5D). The 200-bp sequence downstream of the conserved repeat tracts is 51% conserved between human and mouse and 48% conserved between human and rat (fig. 5D). In mouse, the repeat tract is composed of two of the three sequence motifs, (TG)<sub>5</sub>TCTG TG (TCTG)<sub>2</sub>, found in humans, whereas the rat contains only a simple (TG)<sub>23</sub> repeat at the homologous intronic site (fig. 4). Accession numbers for the newly developed STR markers, as well as sequence information, are summarized in tables 1, 5, and 6.

## Discussion

Haplotype analysis of 71 families with DM2 was performed to assess the possible origins of the DM2 expansion. The majority of families analyzed are of German or European descent, with one family from Afghanistan. Several general predictions can be made from our data. All of the families appear to have variations of the consensus haplotype A, in which haplotypes B and C appear to be derived from haplotype A by a limited number of ancestral recombination and microsatellite instability events (fig. 2 [top]). Microsatellite instability is fairly common at the population level and is the basis for the high rates of heterozygosity of microsatellite repeat markers (Weber and Wong 1993). A diagram depicting how these major haplotype variants could have been derived from a single common haplotype is presented in figure 2 (bottom). Seven additional and slightly variant DM2 chromosomes, all of which contain a core region of conservation, appear to be related to the more common haplotypes by similar recombination and/or mi-

cro-satellite instability events. Similar to DM1, haplotype conservation in the families we have studied suggests that DM2 arose from a single, or perhaps a few, common founders, at least in patients of European and Afghan ancestry.

The affected DM2 haplotype in the native Afghan family (Tajiks from Kabul) is strikingly similar to the core region of the consensus haplotype, with seven markers shared over a 456-kb interval. Although estimates of haplotype age are difficult because rare ancestral recombination events are unlikely to be evenly distributed from generation to generation and because microsatellite mutation rates vary from marker to marker, the conservation of haplotype A in the Afghan family allows us to speculate that DM2 was introduced into the Afghan gene pool sometime between 2000 and 1000 B.C., when the ancient Aryan tribes of Indo-European extraction settled Aryana (ancient Afghanistan).

Our linkage disequilibrium (Liquori et al. 2001) and haplotype data suggest that DM2, like DM1, arose from a common founder chromosome. For DM1, haplotype analysis of control chromosomes has suggested that the DM1 expansion arose through a series of small stepwise expansions that occurred on chromosomes with a single haplotype in the general population and that the longer repeat tracts, which were generally stable within single families, served as a pool of pre-mutation alleles and a reservoir of new mutations. In SCA1, SCA2, and FMR1, the presence of interruptions within the normal CAG and CCG repeats are thought to confer stability on the normal repeat (Chung et al. 1993; Kunst and Warren 1994; Pulst et al. 1996; Gunter et al. 1998), with the loss of these interruptions leading to increased instability and further expansions. Sequence interruptions of the DM2 CCTG repeat tract may serve a similar purpose and may confer stability on the CCTG tract, with the loss of interruptions predisposing the allele to further expansions that could eventually expand into the pathogenic range.

Although our analysis of 228 control chromosomes indicates that perfect matches to the major affected haplotypes are relatively rare, we identified a single control chromosome with a haplotype identical to a large portion of haplotype B spanning the DM2 repeat tract from markers CL3N122 to CL3N116. This chromosome, schematically represented in figure 3A, has an unusual uninterrupted (CCTG)<sub>20</sub> repeat tract. Within the normal population, the largest CCTG tract we have observed contains 26 repeats with two interruptions and the overall repeat motif (CCTG)<sub>12</sub>GCTG CCTG TCTG (CCTG)<sub>11</sub>. The fact that the (CCTG)<sub>20</sub> repeat motif has a significantly longer stretch of uninterrupted repeats than have been observed on other control alleles and is found on a haplotype identical to the haplotype B in a 470-kb region flanking the repeat suggests that this nor-

**Table 6**

Complete Genomic Sequence	
Species	Accession Number
Human	AY329622
Mouse	AY329623
Rat	AY329624

mal allele may represent an allele that is less stable than interrupted alleles in the general population and could eventually be predisposed to further pathogenic expansion. Follow-up studies to determine if uninterrupted alleles are commonly found on control chromosomes with the DM2 haplotype would further support this initial observation.

Linkage disequilibrium with the DM2 mutation was apparent only through the use of STR markers. Although seven putative SNP markers were identified from the NCBI database, none were useful in distinguishing the DM2 haplotypes from haplotypes of the general population, even though they were located extremely close to the mutation (10.2–109.5 kb). All of the SNP alleles associated with the affected DM2 haplotype have frequencies >96% (minor allele frequency <4%) (table 3). A useful threshold for using SNPs in haplotype analysis is generally considered to be a minor allele frequency >20%; thus, none of the SNPs in the region were considered optimal.

Each of the four nonhuman mammals we examined contained repeats that were similar but not identical to the human DM2 repeat. The TG portion of the repeat tract is found in human, chimpanzee, gorilla, mouse, and rat. In contrast to the human repeat, in the chimpanzee, the TG tract is interrupted by one or two TT motifs. The TCTG portion of the repeat tract is found in human, chimpanzee, gorilla, and mouse, although chimpanzee has only a single TCTG. The CCTG portion is found in human, chimpanzee, and gorilla. The least-conserved CCTG portion of the repeat tract is interrupted in normal human and gorilla (although the interruptions in gorilla are more complicated) and is uninterrupted in chimpanzee. A 200-bp sequence just 3' of the respective repeats is conserved among human, chimpanzee, gorilla, mouse, and rat. The evolutionary conservation of the repeat and 3' flanking sequence suggests a biological function. Because both TG dinucleotide and UCAY tetranucleotide repeats have been shown to be involved in the regulation of alternative splicing (Jensen et al. 2000; Gabellini 2001), it is possible that the conserved sequence in *ZNF9* intron 1 regulates splicing or some other aspect of gene regulation.

In humans, the  $\alpha$ -*ZNF9* isoform encodes a 177-amino acid protein, whereas  $\beta$ -*ZNF9* encodes a 170-amino acid protein lacking the last 7 amino acids of the first zinc finger motif. Both donor sites at the 3' end of exon 2

have the same consensus sequence, and both seem to be recognized with a similar frequency (Flink and Morkin 1995). Although *ZNF9* is alternatively spliced and both TG and UCAY repeats have been shown to be involved in the regulation of alternative splicing, it seems unlikely that the normal role of the repeat tract in intron 1, placed well before the splicing branch site, would have a role in regulating alternative use of the splice donor site in exon 2. The fact that the TG portion of the repeat tract is the most highly conserved suggests that the TG sequence may provide the most important clues about the normal function of the repeat tract and the reasons for its conservation from human to mouse. Substantial conservation of noncoding sequences throughout the mouse and human genomes (Mural et al. 2002) suggests that we have much to learn about the normal function of noncoding DNA.

## Acknowledgments

We thank members of families with DM2, for their participation, and Jennifer F. Jacobsen, Katherine A. Dick, and Laura J. Rasmussen, for technical assistance. Funding from the Muscular Dystrophy Association and National Institutes of Health grant NS35870 is gratefully acknowledged. DNA samples from two families included in our study were also used in a separate study by R. Krahe.

## Electronic-Database Information

Accession number and URLs for data presented herein are as follows:

BLAST, <http://www.ncbi.nlm.nih.gov/BLAST/>  
 GenBank, <http://www.ncbi.nlm.nih.gov/Genbank/index.html>  
 (for *ZNF9* [accession number AF388525])  
 NCBI, <http://www.ncbi.nlm.nih.gov/>  
 NCBI Clone Registry, <http://www.ncbi.nlm.nih.gov/genome/clone/>  
 NCBI SNP Database, <http://www.ncbi.nlm.nih.gov/SNP/>  
 Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim/> (for DM1, DM2, and PROMM)  
 UCSC Genome Browser, <http://www.genome.ucsc.edu/>

## References

- Brook JD, McCurrah ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion J-P, Hudson T, Sohn R, Zemelman B, Snell RG, Rundle SA, Crow S, Davies J, Shelbourne P, Buxton J, Jones C, Juvonen V, Johnson K, Harper PS, Shaw DJ, Housman DE (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* 68:799–808
- Buxton J, Shelbourne P, Davies J, Jones C, Van Tongeren T, Aslanidis C, de Jong P, Jansen G, Anvret M, Riley B, Williamson R, Johnson K (1992) Detection of an unstable frag-

- ment of DNA specific to individuals with myotonic dystrophy. *Nature* 355:547–548
- Chakraborty R, Stivers DN, Deka R, Yu LM, Shriver MD, Ferrell RE (1996) Segregation distortion of the CTG repeats at the myotonic dystrophy locus. *Am J Hum Genet* 59:109–118
- Chung MY, Ranum LPW, Duvick LA, Servadio A, Zoghbi HY, Orr HT (1993) Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type 1. *Nat Genet* 5:254–258
- Day JW, Ricker K, Jacobsen JF, Rasmussen LJ, Dick KA, Kress W, Schneider C, Koch MC, Bielman GJ, Harrison AR, Dalton JC, Ranum LPW (2003) Myotonic dystrophy type 2: molecular and clinical spectrum. *Neurology* 60:657–664
- Day JW, Roelofs R, Leroy B, Pech I, Benzow K, Ranum LP (1999) Clinical and genetic characteristics of a five-generation family with a novel form of myotonic dystrophy (DM2). *Neuromuscul Disord* 9:19–27
- Flink IL, Morkin E (1995) Organization of the gene encoding cellular nucleic acid-binding protein. *Gene* 163:279–282
- Fu YH, Pizzuti A, Fenwick RG Jr, King J, Rajnarayan S, Dunne PW, Dubel J, Nasser GA, Ashizawa T, de Jong P, Wieringa B, Korneluk R, Perryman MB, Epstein HF, Caskey CT (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* 255:1256–1258
- Gabellini N (2001) A polymorphic GT repeat from the human cardiac Na<sup>+</sup>Ca<sup>2+</sup> exchanger intron 2 activates splicing. *Eur J Biochemistry* 268:1076–1083
- Gunter C, Paradee W, Crawford DC, Meadows KA, Newman J, Kunst CB, Nelson DL, Schwartz C, Murray A, Macpherson JN, Sherman SL, Warren ST (1998) Re-examination of factors associated with expansion of CGG repeats using a single nucleotide polymorphism in FMR1. *Hum Mol Genet* 7:1935–1946
- Harley HG, Brook JD, Rundle SA, Crow S, Reardon W, Buckler AJ, Harper PS, Housman DE, Shaw DJ (1992) Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature* 355:545–546
- Harper P (2001) *Myotonic dystrophy*, 5th ed. W. B. Saunders, London
- Imbert G, Kretz C, Johnson K, Mandel JL (1993) Origin of the expansion mutation in myotonic dystrophy. *Nat Genet* 4:72–76
- Jensen KB, Musunuru K, Lewis HA, Burley SK, Darnell RB (2000) The tetranucleotide UCAY directs the specific recognition of RNA by the Nova K-homology 3 domain. *Proc Natl Acad Sci USA* 23:5740–5745
- Kunst CB, Warren ST (1994) Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell* 77:853–861
- Liquori C, Ricker K, Moseley ML, Jacobsen JF, Kress W, Naylor S, Day JW, Ranum LPW (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* 293:864–867
- Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, Neville C, Narang M, Barcelo J, O'Hoy K, Leblond S, Earle-MacDonald J, De Jong PJ, Wieringa B, Korneluk RG (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* 255:1253–1255
- Mural R, Adams M, Myers E, Smith H, Miklos G, Wides R, Halpern A, et al (2002) A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. *Science* 296:1661–1671
- Neville C, Mahadevan M, Barcelo J, Korneluk RG (1994) High resolution genetic analysis suggests one ancestral predisposing haplotype for the origin of the myotonic dystrophy mutation. *Hum Mol Genet* 3:45–51
- Pulst S-M, Nechiporuk A, Nechiporuk T, Gispert S, Chen X-N, Lopes-Cendes I, Pearlman S, Starkman S, Orozco-Diaz G, Lunkes A, DeJong P, Rouleau GA, Auburger G, Korenberg JR, Figueroa C, Sahba S (1996) Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat Genet* 14:269–276
- Ranum L, Rasmussen P, Benzow K, Koob M, Day J (1998) Genetic mapping of a second myotonic dystrophy locus. *Nat Genetics* 19:196–198
- Ranum LPW, Day JW (2002) Myotonic dystrophy: clinical and molecular parallels between myotonic dystrophy type 1 and type 2. *Curr Neurol Neurosci Rep* 2:465–470
- Ricker K (1999) Myotonic dystrophy and proximal myotonic myopathy. *J Neurol* 246:334–338
- Ricker K, Koch MC, Lehmann-Horn F, Pongratz D, Otto M, Heine R, Moxley III RT (1994) Proximal myotonic myopathy: a new dominant disorder with myotonia, muscle weakness, and cataracts. *Neurology* 44:1448–1452
- Thornton CA, Griggs RC, Moxley RT (1994) Myotonic dystrophy with no trinucleotide repeat expansion. *Ann Neurol* 35:269–272
- Udd B, Krahe R, Wallgren-Pettersson C, Falck B, Kalimo H (1997) Proximal myotonic dystrophy—a family with autosomal dominant muscular dystrophy, cataracts, hearing loss and hypogonadism: heterogeneity of proximal myotonic syndromes? *Neuromuscul Disord* 7:217–228
- Weber JL, Wong C (1993) Mutation of human short tandem repeats. *Hum Mol Genet* 2:1123–1128