Short Communications

Mammalian Genome

Mammalian Genome 3: 290–292, 1992

© Springer-Verlag New York Inc. 1992

Cross-species conservation of a polymorphic dinucleotide repeat in the dystrophin gene

Andrea J. Maichele and Jeffrey S. Chamberlain

Department of Human Genetics, Box 0618, 3726 Med Sci II, University of Michigan Medical School, Ann Arbor, Michigan 48109-0618, USA

Received January 24, 1992; accepted March 16, 1992

The presence of a polymorphic $[CA]_n$ dinucleotide microsatellite repeat (Weber and May 1989) in the 3′ untranslated region of the human dystrophin gene (Duchenne muscular dystrophy locus, DMD) has been reported (Oudet et al. 1990; Beggs and Kunkel 1990). Sequence analysis of the murine dystrophin cDNA reveals that this dinucleotide repeat is conserved between species and is present at the same location in the mouse and human genes. We have found that the murine dinucleotide repeat, designated UM11, is variable in length among separate mouse species and therefore represents a valuable marker for mapping the X Chromosome (Chr) in interspecific crosses (for example, Chamberlain et al. 1987).

The published repeat in the 3′ untranslated region of the human dystrophin cDNA has the sequence $[CA]_7TA[CA]_8$ (Koenig et al. 1988). The corresponding dinucleotide repeat in DNA from C57BL/10 mice is $[CA]_{17}$ followed closely by a second $[CA]_{13}$. Recently Moore et al. (1991) searched Genbank 66.0 and found 32 dinucleotide repeats conserved among different species of mammals. Only one of these sequences occurred in a coding region; the others were in introns or intergenic regions. It is unusual to have $[CA]_n$ repeats in mRNAs, and the conservation of the dystrophin 3′ $[CA]_n$ repeat through evolutionary time is striking. It has previously been noted that the highest degree of sequence identity between the human and the chicken dystrophin cDNA sequences occurs not in the coding sequences but in the 5′ end of the 3′ untranslated region (LeMaire et al. 1988). This is also true when the murine sequence (Chamberlain et al.: GenBank accession #M68859) is included in the comparison. We have observed that the stretch of greatest similarity between the mouse and human sequences ends immediately after the $[CA]_n$ repeat (Fig. 1), but that this $[CA]_n$ repeat is not present in the chicken sequence (LeMaire et al. 1988). The conserved location of a $[CA]_n$ repeat in the 3′ untranslated region of the mouse and human dystrophin genes is intriguing, and it is tempting to speculate on potential roles for this sequence in regulating mRNA expression or stability. However, $[CA]_n$ repeats are found interspersed throughout the genomes of most eukaryotic species examined, and no functional role for any of these sequences has yet been described.

We used the polymerase chain reaction (PCR) to compare the length of the murine dystrophin 3′ $[CA]_n$ repeat between several wild mouse species and common inbred laboratory strains of mice. A stuttering pattern characteristic of PCR-amplified $[CA]_n$ repeats is observed in each lane of the acrylamide gel (Fig. 2); the dark upper band in each lane corresponds to the full-length allele; fainter bands below are progressively decreased by [CA] units. All 12 of the tested inbred laboratory strains share the 233 bp allele predicted from the cloned C57BL/10 sequence (Figs. 2A and B). In Fig. 2A, the alleles of *Mus musculus castaneus, Mus musculus molossinus,* and *Mus spretus* are, respectively, 14 bp smaller, 6 and 36 bp larger than the 233 bp allele. Figure 2B displays the clear resolution of species-specific alleles in several mice. We have tested four *M. spretus* individuals and observed four different allele sizes; in three animals a single allele was observed (+20, +24, and +36), and a fourth allele (+38) was present with the +36 allele in a heterozygous *M. spretus* female. The expected parental alleles, 233 bp and *spretus* +20, were seen in a CH3 × *M. spretus* female. Allele size differences down to 14 bp can be visualized by ethidium bromide staining of agarose gels without the use of radioisotopes (Fig. 3). Because of the variability in the size of *M. spretus* alleles,

The nucleotide sequence data reported in this paper have been submitted to GenBank and have been assigned the accession number M68859.

Offprint requests to: J.S. Chamberlain

```
M:11,255 TAGGAAGCCTTTTCCACATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTTTCAGTCATGACAGATGAAGAAGGAGCAGAATAAA
H:11,264 -------T-----------------------------------------------A----------------------------------

M:11,345 TGTTTTACAACTCCTGATTCCCGCATGGTTTTTTATAATATTCGTACAACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAAATCT
H:11,354 ----------------------------------------A-----------------------------------------.------

M:11,435 ATATTTTTGTGAAGGGTAGTGGTACTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAGGTTTTACAC
H:11,443 ------------------------T-------------------------------------------------------------

M:11,525 GTCTATGCAATTGTACAAAAAAGTTAAAAGAAAA...CATGTAAAATCTTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCAT
H:11,533 --------------------------T-------CTA-----------------------------------------------------

M:11,612 TTTGGGTTGTTTAAAAATTTATAACAGTTATAAAGAGAGATTGTAAACTAAAGTGTGCTTTAT.AAAAAAAGTTGTTTATAAAAACCCCT
H:11,623 ------------------------------------A--------------------------A-------------------------

M:11,701 AAACA.AACACACACGCACACACACACACACACACACACACACACACACACGCACACATACATGCA.....CGAACCCACCACA
H:11,713 ---A-C--A---A--A--------------T---------------.......-A--T-TG-G---GCGCATTGTTTT-C-T--.....

M:11,779 CACACACACACACACACACACACTGAGGCAGCACAT.....TGTTTTGCATTACTTTAGCGTGGTATTCATATGGAATTCATG11,857:M
H:11,790 .....T-TT---GTG-T--CCATA--AAA-T---GG----TT-T-TT-T-G-----TA--GAT-A- 11,851:H
```

Fig. 1. Comparison of the mouse and human dystrophin 3′ untranslated regions from the stop codon [base 11,255 in mouse (Chamberlain et al.: GenBank accession #M68859) and 11,264 in human (Koenig et al. 1991)] past the conserved $[CA]_n$ sequence. There are 90 bases per line. The sequences were aligned by Eugene (version 3.2) with the Doolittle alignment program: **dashes** indicate identical bases in the human (**bottom**) and the mouse (**top**) sequences; **dots** indicate gaps. The murine PCR primers are represented by **arrows** above the sequence.

it is necessary to pre-size the parental alleles of any given interspecific cross before deciding whether typing of F_1 animals will require the resolution afforded by acrylamide gels.

$[CA]_n$ polymorphisms are valuable markers because of their variability and ease of typing. The 2.4 megabase human dystrophin gene spans approximately 12 cM of the X Chr, and $[CA]_n$ repeats were recently shown to be highly informative for haplotyping and mutation detection of DMD-affected males and carrier females (Abbs et al. 1990; Clemens et al. 1991). The murine dystrophin gene is of similar physical size to the human locus (J.S. Chamberlain, unpublished); however, few polymorphisms are available for the murine locus (Chamberlain et al. 1987). $[CA]_n$ repeat locus UM11 at the 3′ end of the mouse gene should prove to be a valuable X-linked marker for interspecific mouse crosses.

Acknowledgments. A.J. Maichele was supported by National Institutes of Health (NIH) Service Award #5-T32-GM07544-12 from the National Institute of General Medical Sciences. Supported by a grant from the Muscular Dystrophy Association and by NIH grant 1RO1 AR40864, and by the University of Michigan Multipurpose Arthritis and Musculoskeletal Diseases Center (grant NIH 2 P60 AR20557).

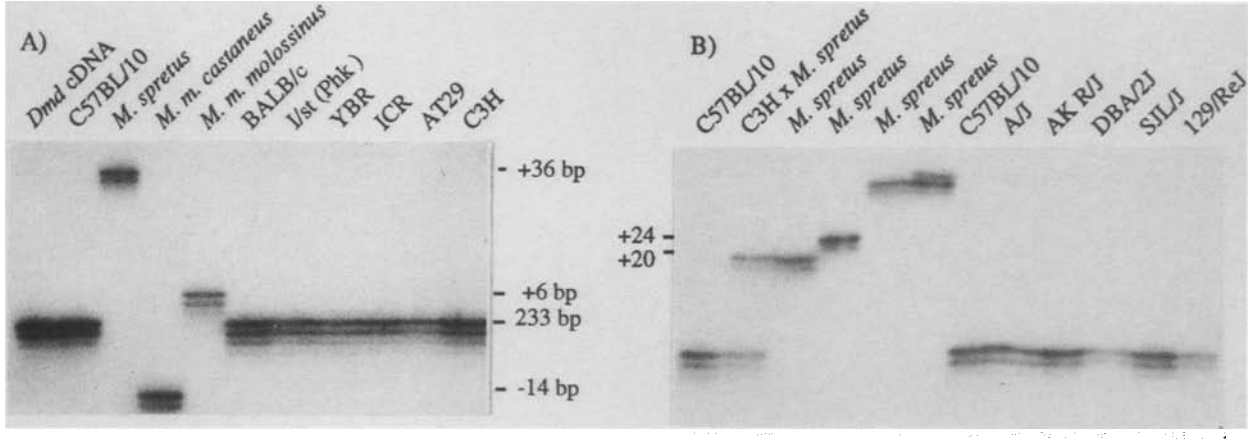

Fig. 2. Amplification of the dinucleotide repeat from cloned C57BL/10 dystrophin cDNA (*Dmd* cDNA), and genomic DNA from the indicated species and laboratory strains of mice. The CA strand was labeled by incorporation of a ^{32}P-end-labeled PCR primer (Weber and May 1989); PCR conditions were 30 s at 92°C, 30 s at 57°C, and 30 s at 72°C for 30 cycles (20 cycles for cloned DNA). Analysis was on a 6% denaturing polyacrylamide sequencing gel with overnight autoradiography. **(A)** The results obtained with DNA from seven inbred laboratory mouse strains and three wild mouse species. To the **right** is indicated the size of each PCR product relative to the 233 bp allele. **(B)** In a separate experiment five additional inbred mouse strains also displayed the 233 bp allele found in the C57BL/10 DNA. Four unrelated *M. spretus* individuals displayed four separate alleles (left to right: +20 bps, +24 bps, +36 bps, and a +36/+38 bp heterozygote). Also shown is the resolution of alleles from a C3H × *M. spretus* F_1 female DNA sample.

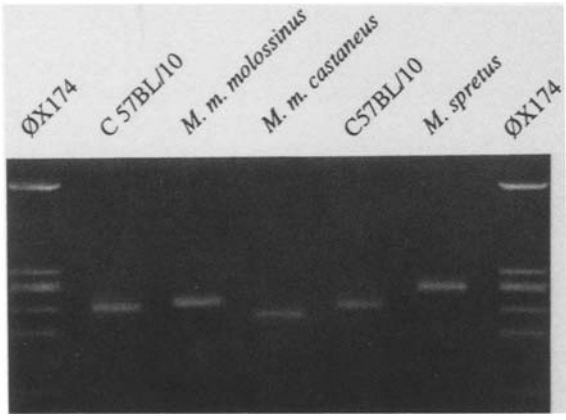

Fig. 3. Resolution of unlabeled PCR products on an ethidium bromide-stained gel (2% NuSieve, 1% LE agarose; FMC Bioproducts, Rockland, Me.). From left to right, the *M.m. molossinus, M.m. castaneus,* and *M. spretus* alleles are 6 bp larger, 14 bp smaller, and 36 bp larger than the 233 bp C57BL/10 allele. With this type of gel the *M.m castaneus* and *M. spretus* alleles are resolved from the C57BL/10 allele.

References

Abbs, S., Roberts, R.G., Mathew, C.G., Bentley, D.R., and Bobrow, M.: Accurate assessment of the intragenic recombination frequency within the Duchenne muscular dystrophy gene. *Genomics 7:* 602–606, 1990.

Beggs, A.H. and Kunkel, L.M.: A polymorphic CACA repeat in the 3′ untranslated region of dystrophin. *Nucleic Acids Res 18:* 1931, 1990.

Chamberlain, J.S., Grant, S., Reeves, A.A., Mullins, L., Stephenson, D., Hoffman, E.P., Monaco, A.P., Kunkel, L.M., Caskey, C.T., and Chapman, V.M.: Regional localization of the murine Duchenne muscular dystrophy gene on the mouse X chromosome. *Somat Cell Mol Genet 13:* 671–678, 1987.

Clemens, P.R., Fenwick, R.G., Chamberlain, J.S., Gibbs, R.A., de Andrade, M., Chakraborty, R., and Caskey, C.T.: Carrier detection and prenatal diagnosis in Duchenne and Becker muscular dystrophy families, using dinucleotide repeat polymorphisms. *Am J Hum Genet 49:* 951–960, 1991.

Koenig, M., Monaco, A.P., and Kunkel, L.M.: The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. *Cell 53:* 219–226, 1988.

LeMaire, C., Heilig, R., and Mandel, J.L.: The chicken dystrophin cDNA: striking conservation of the C terminal coding and 3′ untranslated regions between man and chicken. *EMBO J 7:* 4157–4162, 1988.

Moore, S.S., Sargent, L.L., King, T.J., Mattick, J.S., Georges, M., and Hetzel, D.J.S.: The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. *Genomics 10:* 654–660, 1991.

Oudet, C., Heilig, R., and Mandel, J.L.: An informative polymorphism detectable by polymerase chain reaction at the 3′ end of the dystrophin gene. *Hum Genet 84:* 283–285, 1990.

Weber, J.L. and May, P.E.: Abundant class of human DNA polymorphisms that can be typed using the polymerase chain reaction. *Am J Hum Genet 44:* 388–396, 1989.