Biochemical and Biophysical Research Communications **259,** 190–194 (1999)
Article ID bbrc.1999.0751, available online at http://www.idealibrary.com on IDEAL®

Transcribed Dinucleotide Microsatellites and Their Associated Genes from Channel Catfish *Ictalurus punctatus*

Zhanjiang Liu,[1] Guo Tan, Ping Li, and Rex A. Dunham
Fish Molecular Genetics and Biotechnology Laboratory, Department of Fisheries and Allied Aquacultures, Auburn University, Auburn, Alabama 36849

Received April 14, 1999

The presence of trinucleotide microsatellites within genes is a well-known cause for a number of genetic diseases. However, the precise distribution of dinucleotide microsatellites within genes is less well documented. Here we report 15 unique cDNAs containing dinucleotide repeats from the channel catfish *Ictalurus punctatus*. Gene identities of nine of the 15 cDNAs were determined, of which three encode structural genes, and six encode regulatory proteins. Five cDNAs harbored dinucleotide repeats in the 5′ untranslated region (5′-NTR), nine in the 3′-NTR, and one in the coding region. The presence of these transcribed dinucleotide repeats and their potential expansion in size within coding regions could lead to disruption of the original protein and/or formation of new genes by frame shift. The low number of dinucleotide repeats within coding regions suggests that they were strongly selected against. All the transcribed microsatellite loci examined were polymorphic making them useful for gene mapping in catfish.

Key Words: **polymorphism; marker; gene mapping; type 1; repeat, mutation.**

Size expansion of trinucleotide repeats within genes is well known and is the genetic basis for many neurological disorders (1). These disorders include myotonic dystrophy, spinobulbar muscular atrophy, type 1 and type 3 spinocerebellar ataxia, dentatorubro-pallidoluysian atrophy, fragile X, and Huntington's disease. Depending on the position of the repeats, the trinucleotide microsatellites can cause either loss or gain of function of genes harboring such sequences. If the microsatellites are in the upstream non-coding regions, they often reduce gene expression leading to loss of function of the genes. In contrast, if microsatellites exist in the coding regions, they can lead to gain of new functions of the proteins by containing poly-monomeric amino acid tracts, such as the poly-glutamine tract encoded by CAG repeats (2-4).

Microsatellites are evenly distributed in genomes (5). When located within genes, they may have adverse effects on the organism as in the case of human diseases. While expansion of trinucleotide repeats and their relationship with a number of human diseases are well known and much effort has been devoted to characterizing trinucleotide repeat-containing cDNAs (6-13), less is known about transcribed dinucleotide repeats. Size expansion of dinucleotide repeats within genes may have more serious effects than size expansion of trinucleotide repeats because their size expansion causes frame shifting for translation. The frame shifting capacity of dinucleotide repeats within coding regions could lead to loss of function of the original proteins, and/or generation of new proteins with different functions.

Microsatellites are extremely important markers for development of genetic linkage maps because of their high polymorphism, abundance, co-dominance, and their small length, which facilitates genotyping by polymerase chain reaction (PCR). However, they are usually type II markers (14) for which no known function has been established. Type I markers are associated with genes of known function and are more useful for comparative gene mapping to study genomic evolution. Linkage associations between homologous genes of different species may be of functional importance (15). Once a quantitative trait locus (QTL) has been located in one species, synteny analysis of genes may predict possible candidate genes that affect the quantitative trait in another species (16).

Polymorphic type I markers are relatively rare as compared to polymorphic type II markers. One way to develop polymorphic type I markers is to associate type

[1] Correspondence should be sent to this author at the Department of Fisheries and Allied Aquacultures, 203 Swingle Hall, Auburn University, Auburn, AL 36849. Fax: (334)-844-9208. E-mail: zliu@acesag.auburn.edu.

0006-291X/99 $30.00

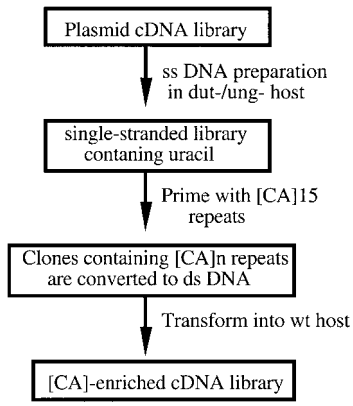

FIG. 1. Strategy for transcribed dinucleotide microsatellite enrichment.

II markers with known genes (17). Microsatellites are good candidates for this because of their even distribution in genomes examined to date (5). The objective of this study was to identify genes that are associated with microsatellites in channel catfish, *Ictalurus punctatus.* Here we report 15 genes that were associated with dinucleotide microsatellites, of which nine were identified by homology comparison. Five genes harbor microsatellites in their 5′ untranslated region (5′-NTR), nine in their 3′-NTR, one in its coding region. All examined clones were polymorphic making them useful for gene mapping and QTL mapping in catfish.

MATERIALS AND METHODS

cDNA library and microsatellite enrichment. The lambda Uni-ZAP II cDNA library constructed from pituitary RNA of channel catfish was utilized (18). The lambda phage library was converted into plasmid library by a mass *in vivo* excision procedure (Stratagene, La Jolla, CA, 19). The resulting plasmid library contained cDNA clones within the plasmid pBluescript SK^-. Conversion into a plasmid library allowed preparation of single-stranded phagemid DNA required for enriching clones containing microsatellites.

Two microsatellite-enriched cDNA libraries were made according to the procedure of Ostrander *et al.* (20) as outlined in Figure 1. Single-stranded phagemid DNA was prepared in *E. coli* CJ236, a dut^-/ung^- strain, and primed by oligonucleotide $[CA]_{15}$ and $[GA]_{15}$ primers, respectively. Microsatellite-containing clones were converted into double-stranded DNA by using *Taq* polymerase. The two microsatellite-enriched cDNA libraries were screened by colony lift hybridization (21) using oligonucleotide primer probes: $[CA]_{15}$, and $[GA]_{15}$ labeled with γ-P^{32} ATP. Positive clones were picked for plasmid preparation and DNA sequencing.

DNA sequencing and sequence analysis. Plasmid DNA was prepared using the Qiagen Miniplasmid Kit (Qiagen, Chatsworth, CA) and sequenced using the 7-deaza thermosequenase cycle sequencing kit (Amersham, Arlington Heights, IL) with T7 and T3 primers (LI-COR, Lincoln, NE), respectively. A single pass sequencing from both strands was sufficient to identify genes and to locate microsatellite sequences using a LI-COR 4200IR automated DNA sequencer. Vector sequences were removed before searching for identity using BLASTN and/or BLASTX servers via the Internet (NCBI, http://www.ncbi.nlm.nih.gov). Matches were considered to be significant only when the smallest sum probability (P) of the BLAST search was less than 0.0001 (22).

PCR amplification and analysis. To determine polymorphism of transcribed microsatellites, DNA from six channel catfish were tested by PCR. Approximately 200 ng of genomic DNA was amplified in PCR reactions of 50 μl containing 50 mM KCl, 10 mM Tris (pH 9.0 at 25°C), 0.1% Triton X-100, 0.25 mM each of deoxynucleotide triphosphate (dNTPs), 1.5 mM $MgCl_2$, 20 μM each of the upper and lower PCR primers, and 2.5 units of *Taq* DNA polymerase. The temperature profiles used for PCR amplification were 94°C for 30 sec, 55°C for 1 min, and 72°C for 2 min, for 35 cycles. Following amplification, samples of 3 μl were mixed with 1 μl loading dye and electrophoresed on 10% acrylamide (19:1 acrylamide:bisacrylamide) gels or on sequencing gels.

RESULTS AND DISCUSSION

Two microsatellite-enriched cDNA libraries containing [CA]n and [GA]n were established. Fifty positive clones were initially isolated, of which 27 contained [GA] repeats, and 23 contained [CA] repeats. The 50 clones were further analyzed by DNA sequencing. All repeated clones and clones that did not contain poly $(A)^+$ sequences were eliminated. Seven unique [GA]-containing, polyadenylated cDNA clones were isolated, as were eight unique [CA]-containing, polyadenylated clones. A large proportion of [GA]-containing clones were redundant clones of several transcripts. For example, growth hormone accounted for 14 of the 27 clones (52%) and ribosomal protein S16 accounted for 4 of the 27 clones (15%). Clearly, their representation in the enriched library is biased toward highly expressed genes in the pituitary (19). This is a shortcoming of this approach for isolating microsatellite-containing cDNAs from microsatellite-enriched cDNA libraries. The polyadenylated [CA]- and [GA]-containing clones were further analyzed for sequence identity by database searches and for polymorphism using PCR.

A variety of cDNAs contained microsatellite sequences (Table 1). The sequences of the expressed sequence tags (ESTs) have been deposited in the dbEST database and their accession numbers are included in Table 1. Among the 15 unique clones, nine are products of known genes as determined by homology comparison. Six of the known cDNAs encode proteins with regulatory functions including cyclin B1 (IpTR006), Y-box protein YB2 (IpTR030), cyclin-dependent kinase 8 (CDK8, IpTR036) (23, 24), growth hormone, epithelial growth factor-like protein, and DNA excision repair protein MHR23B. The other three known cDNAs encode ribosomal protein S16, cytochrome c oxidase subunit 1, and a *C. elegans* homologue Y57G11 (Table 1).

The transcribed dinucleotide microsatellites did not appear to be randomly associated with various types of genes. They were highly associated with genes of regulatory functions. This is consistent with previous findings that trinucleotide microsatellites are highly associated with proteins having regulatory roles such as transcriptional factors (25-27) or tumor suppressor genes (28). In fact, a search of the GenBank database revealed that a large proportion of kinases contained

TABLE 1

Transcribed Microsatellites and Their Associated Genes

Locus names	Accession numbers	Homologous sequences	Probability	Gene identity	Repeat units	Repeat location	Poly $(A)^+$	Number of clones
IpTR002	AI308622 AI308623	S69215	0.0	growth hormone	$[GA]_8$	3′-NTR	Yes	14
IpTR006	AI304251 AI304252	AF036565	2.9e-07	cyclin B1	$[TG]_{23}$	3′-NTR	Yes	2
IpTR007	AI304253 AI304254	M60854	3.4e-15	ribosomal protein S16	$[GT]_4[GA]_9$	3′-NTR	Yes	4
IpTR015	AI304255 AI304256	Z99281	3.4e-07	*C. elegans* Y57G11	$[GT]_4[GA]_{25}$ $(CAT)_3$	3′-NTR	Yes	2
IpTR016	AI304257 AI304258	–	–	unknown	$[CA]_{27}N_7$ $[CA]_{17}$	3′-NTR	Yes	2
IpTR020	AI304259	–	–	unknown	$[TC]_{15}$	3′-NTR	Yes	2
IpTR030	AI304260 AI304261	AA925251	1.0e-06	Y-box protein YB2	$[CA]_{14}$	5′-NTR	Yes	1
IpTR036	AI304262 AI304263	U33015	1.5e-13	Cdk8	$[CA]_{20}$	Coding	Yes	1
IpTR038	AI304264	M91245	2.0e-25	cytochrome c oxydase	$[GA]_{11}$	5′-NTR	Yes	3
IpTR040	AI304265 AI304266	–	–	unknown	$[GT]_{32}$	5′-NTR	Yes	3
IpTR042	AI304267 AI304268	–	–	unknown	$[CA]_{17}$	3′-NTR	Yes	2
IpTR043	AI304269 AI304270	–	–	unknown	$[GA]_{13}CT_{10}$	5′-NTR	Yes	3
IpTR048	AI304271 AI304272	AB011538	1.0e-07	Mouse MEGF5	$[CT]_{16}n[CA]_{14}n$ $[CA]_{16}$	3′-NTR	Yes	2
IpTR049	AI304273 AI304274	–	–	unknown	$[GT]_{30}$	3′-NTR	Yes	2
IpTR093	AI304275 AI304276	X92411	3.0e-18	MHR23B	$[CA]_{12}$ & $[CCA]_{11}$	Coding & 5′-NTR	Yes	1

microsatellite sequences (not shown). Strong association of microsatellites with transcriptional factors (29), kinases (30), and transcription factor binding sites suggest that microsatellite repeats have a role in evolution of genes, regulation of gene expression and consequently species evolution (26).

Ten of the 15 clones contained simple microsatellite sequences while 5 contained composite microsatellites. cDNAs containing [CA] repeats appeared to be more abundant than cDNAs containing [GA] repeats in the pituitary cDNA population. We put greater effort in isolation of the seven unique [GA] containing cDNAs than in isolation of the eight unique [CA]-containing cDNAs. Another factor reducing the number of unique [GA]-containing cDNAs isolated was the complication of the presence of many redundant growth hormone cDNAs containing [GA] repeats in the pituitary cDNA library.

The majority, 9 of 15, of the microsatellites were found in the 3′-NTR. Five clones contained microsatellites in 5′-NTRs. The most interesting clones were IpTR036 and IpTR093, which contained microsatellites in the coding regions. Ip036 contained [CA] dinucleotide repeats in the coding region. Ip093 contained (CCA) trinucleotide repeats in its coding region, additional to presence of [CA] repeats in its 5′-UTR. The dinucleotide [CA] repeats contained in IpTR036 encoded the repeated sequence of two amino acids, histidine (H)-threonine (T). While expansion in size of trinucleotide repeats within genes is known to cause serious human diseases (1), little is known about phenotypes caused by expansion of dinucleotide microsatellite sequences. Obviously repeat number changes of non-integrals of three would cause a frame-shift of the open reading frame downstream from the microsatellite sequence. If such a mutation involved unique genes such as transcription factors, kinases, or other genes of regulatory function, serious consequences may result. Thus identification of such microsatellite containing genes may have clinical importance and allow prediction of potential target genes involved in disease.

One important goal of this work was to develop microsatellite markers for their application to genetic linkage mapping (Table 2). Polymorphisms are required for markers to be used in genetic linkage analysis. To determine if these transcribed dinucleotide microsatellites show similar polymorphism as compared to non-transcribed dinucleotide markers, PCR primers were designed from flanking sequences of

TABLE 2

Determination of Polymorphism of the Transcribed Microsatellite Markers by Polymerase Chain Reaction

Locus names	Upper primer sequences	Lower primer sequences	Polymorphism
IpTR006	ATTCCCCTGTTTACACTCCAG	TTCCATTGTTTTAATGTCAGGAC	yes
IpTR007	AGTCCAAGAAGTTCGGAGGTC	TTCTCCTCAGGTTTATTCTGGTG	yes
IpTR015	TCTAAGCGGAAAAGCCAATCTGC	GCCGCTCATGCCCTACTCTCTT	yes
IpTR016	GCGTTTACTTCGGGTACAAACAC	TTTCATCATTTTCGCTCTTGCTC	yes
IpTR020	TAGTCGAAAGCCAAAGAGGTG	TTTAGTCACTGCATCTGGTGTTG	yes
IpTR042	CTGCTGCCTGCATCTGATCACGA	TGCTTCCCTGTGAGCCAGCTCTG	yes
IpTR049	CAATAAATTAACGTGCTATCTCT	AAAGTGAGTACACCCCTAAGTG	yes

seven loci, randomly selected, and tested in six channel catfish used for production of four Auburn resource families produced for gene and QTL mapping. Use of these parental fish provides a comparison of polymorphisms of transcribed microsatellites with non-transcribed microsatellites in catfish (31, 32). The transcribed microsatellites showed similar levels of polymorphism as compared to non-transcribed microsatellites (31) making these microsatellites useful for construction of catfish genetic linkage and QTL maps. Mapping of these markers will directly map the associated genes to chromosomes. These microsatellites may serve as strong anchor points for comparative gene mapping.

Recent studies indicated the functional importance of microsatellites, especially when they were adjacent to genes. Reduced expression of the rat androgen receptor gene was observed with CAG repeat expansion in its first exon (33). Attenuation of expression of the rat hepatic polymeric immunoglobulin receptor gene was also associated with the tri-nucleotide repeats in the terminal exon (34). However, CGG repeat expansion in the fragile X gene FMR-1 only affected translation, but did not affect transcription (35). Many of the frequently characterized trinucleotide microsatellites in coding regions encode transcriptional factors, or serve as transcription factor binding sites. Although little is known about functional importance of transcribed dinucleotide repeats, the high mutational potential by frame shifting after size expansion of dinucleotide microsatellites justifies further investigation.

A total of 15 transcribed dinucleotide microsatellite markers were identified. The 15 microsatellites were polymorphic, and can be used in the ongoing, national effort (36) of genetic linkage and QTL mapping in catfish. The association of microsatellites to known genes converts these type II markers into type I markers. Mapping of these markers will allow assignment of their associated genes to chromosomes. Localization of transcribed dinucleotide repeats in coding regions also provides emerging evidence of the importance of transcribed dinucleotide repeats. Their expansion or shrinking in length has great mutational potential by frame shifting, which could lead to disruption of the original genes and/or to formation of new genes. A recent report indicated that over 11% of the gene-associated microsatellites were distributed within the open reading frames of the Japanese pufferfish, *Fugu rubripes* (26) and that many of the involved genes have regulatory roles such as transcription factors (27). In this study, the two catfish genes harboring microsatellites in their coding regions are cdc2-dependent kinase 8 and the DNA excision repair protein MHR23B, both of which are important for normal cellular functions. The nature and possibility of size expansion of microsatellites in these genes are being studied in our laboratory.

ACKNOWLEDGMENTS

This project was supported by a USDA National Research Initiative Competitive Grant Program (USDA NRICGP # 98-35205-6738) to Z.J.L. and R.A.D. We thank Drs. Curtis Bird, Huseyin Kucuktas, Michael Maceina, Len Lovshin, and Ronald Phelps for their critical review of the paper. This paper is journal number 8-996009 of the Alabama Agricultural Experiment Station.

REFERENCES

1. Willems, P. J. (1994) *Nature Genet.* **8,** 213–215.
2. Krahe, R., Ashizawa, T., Abbruzzese, C., Roeder, E., Carango, P., Giacanelli, M., Funanage, V. L., Siciliano, M. J. (1995) *Genomics* **28,** 1–14.
3. Maeda, M., Taft, C. S., Bush, E. W., Holder, E., Bailey, W. M., Neville, H., Perryman, M. B., and Bies, R. D. (1995) *J. Biol. Chem.* **270,** 20246–20249.
4. Wang, J., Pegoraro, E., Menegazzo, E., Gennarelli, M., Hoop, R. C., Angelini, C., and Hoffman, E. P. (1995) *Hum. Mol. Genet.* **4,** 599–606.
5. Tautz, D., and Renz, M. (1984) *Nucl. Acids Res.* **12,** 4127–4137.
6. Davis, B. M., McCurrach, M. E., Taneja, K. L., Singer, R. H., and Housman, D. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 7388–7393.
7. Joseph, D. R. (1998) *Steroids* **63,** 2–4.
8. Koch, K. S., Gleiberman, A. S., Aoki, T., Leffert, H. L., Feren, A., Jones, A. L., Fodor, E. J. (1995) *Nucl. Acids Res.* **23,** 1098–1112.
9. Jiang, J. X., Deprez, R. H., Zwarthoff, E. C., Riegman, P. H. (1995) *Genomics* **30,** 91–93.
10. Li, S., McInnis, M. G., Margolis, R. L., Antonarakis, S. E., and Ross, C. A. (1993) *Genomics* **16,** 572–579.
11. Neri, C., Albanese, V., Lebre, A. S., Holbert, S., Saada, C., Bougueleret, L., Meier-Ewert, S., Le Gall, I., Millasseau, P., Bui,

H., Giudicelli, C., Massart, C., Guillou, S., Gervy, P., Poullier, E., Rigault, P., Weissenbach, J., Lennon, G., Chumakov, I., Dausset, J., Lehrach, H., Cohen, D., Cann, H. M. (1996) *Hum. Mol. Genet.* **5,** 1001–1009.

12. Takeuchi, T., Chen, B. K., Qiu, Y., Sonobe, H., Ohtsuki, Y. (1997) *Gene* **204,** 71–77.
13. Reddy, P. H., Stockburger, E., Gillevet, P., Tagle, D. A. (1997) *Genomics* **46,** 174–182.
14. O'Brien, S. J. (1991) *Curr. Opin. Genet. Dev.* **1,** 105–111.
15. O'Brien, S. J., Womack, J. E., Lyons, L. A., Moore, K. J., Jenkins, N. A., and Copeland, N. G. (1993) *Nature Genet.* **3,** 103–112.
16. Soller, M. (1994) *Animal Biotechnol.* **5,** 193–207.
17. Ruyter-Spira, C. P., Crooijmans, R. P. M. A., Dijkhof, R. J. M., van Oers, P. A. M., Strijk, J. A., van der Poel, J. J., and Groenen, M. A. M. (1996) *Animal Genet.* **27,** 229–234.
18. Liu, Z. J., Li, P., Argue, B., and Dunham, R. A. (1997) *Mol. Mar. Biol. Biotechnol.* **6,** 221–231.
19. Karsi, A., Li, P., Dunham, R., and Liu, Z. J. (1998) *J. Mol. Endocrinol.* **21,** 121–129.
20. Ostrander, E. A., Jong, P. M., Rine, J., and Duyk, G. (1992). *Proc. Natl. Acad. Sci. USA* **89,** 3419–3423.
21. Sambrook, J., Frisch, E. F., and Maniatis, T. (1989). Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
22. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
23. Tassan, J. P., Jaquenoud, M., Leopold, P., Schultz, S. J., and Nigg, E. A. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 8871–8875.
24. Leclerc, V., Tassan, J. P., O'Farrell, P. H., Nigg, E. A., and Leopold, P. (1996) *Mol. Biol. Cell* **7,** 505–513.
25. Wolford, J. K., and Prochazka, M. (1998) *Gene* **212,** 103–109.
26. Edwards, Y. J. K., Elgar, G., Clark, M. S., and Bishop, M. J. (1998) *J. Mol. Biol.* **278,** 843–854.
27. Kashi, Y., King, D., and Soller, M. (1997) *Trends Genet.* **13,** 74–78.
28. Seifert, M., Seib, T., Engel, M., Dooley, S., and Welter, C. (1995) *Biochem. Biophys. Res. Comm.* **215,** 910–914.
29. Wolford, J. K., Bogardus, C., and Prochazka, M. (1998) *Biochem. Biophys. Res. Comm.* **246,** 624–626.
30. Theodosiou, A. M., Rodrigues, N. R., Nesbit, M. A., Ambrose, H. J., Paterson, H., McLellan-Arnold, E., Boyd, Y., Leversha, M. A., Owen, N., Blake, D. J., Ashworth, A., and Davies, K. E. (1996) *Hum. Mol. Genet.* **5,** 675–684.
31. Liu, Z., Tan, G., Kucuktas, H., Li, P., Karsi, A., Yant, D. R., Dunham, R. A. (1999) *J. Heredity*, in press.
32. Waldbieser, G. C., and Bosworth, B. G. (1997) *Animal Genet.* **28,** 295–298.
33. Choong, C. S., Kemppainen, J. A., Zhou, Z., and Wilson, E. M. (1996) *Mol. Endocrinol.* **10,** 1527–1535.
34. Aoki, T., Koch, K. S., and Leffert, H. L. (1997) *J. Mol. Biol.* **267,** 229–236.
35. Sandberg, G., and Schalling, M. (1997) *Nucl. Acids Res.* **25,** 2883–2887.
36. Liu, Z., Nichols, A., Li, P., and Dunham, R. A. (1998) *Mol. Gen. Genet.* **258,** 260–268.