Biochemical and Biophysical Research Communications **289,** 317–324 (2001)
doi:10.1006/bbrc.2001.5978, available online at http://www.idealibrary.com on IDEAL®

Microsatellite-Containing Genes from the Channel Catfish Brain: Evidence of Trinucleotide Repeat Expansion in the Coding Region of Nucleotide Excision Repair Gene RAD23B

Zhanjiang Liu,[1] Ping Li, Arif Kocabas, Attila Karsi, and Zhenlin Ju

The Fish Molecular Genetics and Biotechnology Laboratory, Department of Fisheries and Allied Aquacultures and Program of Cell and Molecular Biosciences, Auburn University, Auburn, Alabama 36849

Received October 25, 2001

Expansion of trinucleotide repeats within genes is well known to cause pathological conditions in humans. Here we report a large number of genes containing simple sequence repeats (SSR) from the brain of channel catfish, of which a homologue of the RAD23B gene was found to include (CCA) trinucleotide repeats within its coding region. Because of the importance of the RAD23B gene in the nucleotide excision repair (NER) system, the catfish RAD23B locus was further characterized. The (ACC) repeats encode a polythreonine (T) tract within the catfish RAD23B gene that is absent from the previously cloned human and mouse genes. A survey of the allele variation at the locus indicated the existence of variable microsatellite repeats in the NER RAD23B gene, suggesting that the trinucleotide repeats are expanding or shrinking. The majority of individuals harbor 10 (ACC) repeats within the RAD23B gene, but alleles with 8 and 11 repeats were also detected. The (ACC) repeats are limited to only channel catfish and the closely related blue catfish, but are absent from flathead catfish and the cloned human and mouse genes, suggesting that the microsatellite invasion into the RAD23B gene is a recent event in evolution.

Key Words: **microsatellite; DNA repair; RAD23; fish; marker; mapping; mutation.**

Microsatellites are important markers for development of genetic linkage maps because of their high polymorphism, abundance, codominance, and small locus size. However, they are type II markers for which no known functions are established (1). Type I markers are associated with genes of known functions and thus are of great importance as they can be used in comparative genomics studies to gain understanding on genome evolution. One approach to develop type I markers is to conduct sequencing analysis of expressed sequence tags in searching for microsatellite-containing cDNAs.

Presence of microsatellites within genes may fundamentally affect the outcome and the function of genes. It is well known that trinucleotide repeat expansion within certain genes causes neurological disorders in human (2–6). Expansions in dinucleotide microsatellites would cause frame shift of open reading frames and, therefore, abolish the gene products. To maintain the integrity of the genetic material, cells possess multiple pathways to repair various types of mutational events. Nucleotide excision repair (NER) is the principal pathway in guarding the genetic information (7). A broad spectrum of structurally unrelated lesions can occur in DNA such as ultraviolet (UV)-induced cyclobutane pyrimidine dimmers and (6–4) photoproducts, bulky chemical adducts, and crosslinks (8, 9). Such lesions are principally fixed by NER pathway involving five steps: damage recognition, incision of the damaged strand on both sides of the lesion, excision of the lesion-containing oligonucleotide, synthesis of new DNA using the undamaged strand as template, and ligation (10–13).

Defects in NER cause serious consequences to organisms such as high levels of mutagenesis and/or carcinogenesis. This notion has been demonstrated with xeroderma pigmentosum (XP) disease in humans. XP is an autosomal recessive disease associated with a high incidence of sunlight-induced skin abnormalities (14). XP has been classified into eight complementation groups, XP-A to XP-G, and XP variant (XP-V) (15–17). XP patients show extreme sun sensitivity, pronounced pigmentation abnormalities in UV-exposed areas of the skin, and often accelerated neurodegeneration. The

[1] To whom correspondence and reprint requests should be addressed. Fax: (334)-8449208. E-mail: zliu@acesag.auburn.edu.

0006-291X/01 $35.00

TABLE 1

Microsatellite-Containing Clones from the Brain cDNA Library of Channel Catfish (*Ictalurus punctatus*)

Clones	Accession No.	Gene identity	Microsatellite repeat type	Location of microsatellites
IpBrn02028	BE574267	Aspartyl-tRNA synthetase	$(CA)_{13}$	3′-NTR
IpBrn01163	BE212676	ATPase, Na^+/K^+ transporting, alpha 3	$(TC)_9$	3′-NTR
IpBrn02134	BE574220	Beta-2 microglobulin precursor	(AT)30	3′-NTR
IpBrn01060	BE212836	CG13621 gene product (*Drosophila melanogaster*)	$(CCT)_6$, $(GT)_{11}$	5′-NTR
IpBrn00173	BE212622	Cytochrome P450 aromatase-like	$(CA)_{11}$	3′-NTR
IpBrn00288	BE574223	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 18 (myc-regulated RNA helicase)	$(TA)_{34}$	3′-NTR
IpBrn01368	BE212977	Hypoxanthine guanine phosphoribosyl transferase	$(GT)_{20}$	3′-NTR
IpBrn00018	BE212593	Immunoglobulin gamma heavy chain	$(GT)_{14}$	5′-NTR
IpBrn01730	BE212939	Nucleotide binding protein 2 (*E. coli* MinD like)	$(TG)_{11}$, $(TG)_{11}$	3′-NTR
IpBrn01166	BE574166	Pancreatic somatostatin-14	$(TAA)_5$	3′-NTR
IpBrn00623	BE212751	Profilin	$(TC)_8(AC)_{16}$	3′-NTR
IpBrn01244	BE574193	Protein inhibitor of neuronal nitric oxide synthase	$(CTT)_7$	3′-NTR
IpBrn01232	BE212687	Protein-tyrosine-phosphatase IF1	$(AC)_{12}$, $(CA)_{27}$	3′-NTR
IpBrn02042	BE213167	RAB7, member RAS oncogene family	$(CA)_{11}$	3′-NTR
IpBrn01330	BE212984	RAD23B homolog	$(CCA)_{11}$	Coding region
IpBrn01331	BE574013	Ribosomal protein S16	$(CA)_3TCA(CT)_8(CA)_4$	3′-NTR
IpBrn02018	BE213163	Synaptosomal-associated protein (snapin)	$(AT)_{12}$	3′-NTR
IpBrn01156	BE212675	Transcription factor IIA	$(AT)_{40}$	3′-NTR
IpBrn01154	BE212674	Translation initiation factor 1A	$(TC)_{30}$	3′-NTR
IpBrn00146	BE212587	Type-1 protein phosphatase catalytic subunit alpha isoenzyme	$(TC)_{14}(CT)_{25}(TC)_{22}(CA)_{17}$	3′-NTR
IpBrn01260	BE212691	Zebrafish LINE DNA	$(TA)_7(GATA)_{12}$	N/A
IpBrn01083	BE212844	Purine nucleoside phosphorylase	$(TG)_5A(TG)_8$	N/D
IpBrn01544	BE213114	Actin-related protein 2/3 complex, subunit 5	$(GACA)_5$	N/D
IpBrn01703	BE212929	NADH–ubiquinone oxidoreductase B8 subunit	$(TG)_8$	3′-NTR
IpBrn00387	BE212550	Similar to mouse RIKEN cDNA 1110018B13	$(TA)_{11}$	N/D
IpBrn00129	BE212504		$(CAAA)_7$, $(TAA)_{11}$, $(AAT)_{10}$	N/D
IpBrn00149	BE212506		$(TTG)_4(TTTG)_4$	N/D
IpBrn00216	BE212525		$(CA)_{14}$	N/D
IpBrn00285	BE212472		$[(GT)_3(GA)_2]_7$	3′-NTR
IpBrn00286	BE212527		$(TAA)_{13}$	N/D
IpBrn00305	BE212534		$(GT)_{36}$	N/D
IpBrn00313	BE212745		$(TTTA)_8$	N/D
IpBrn00330	BE212537		$(AT)_{41}$	N/D
IpBrn00374	BE212545		$(AT)_{13}$	N/D
IpBrn00377	BE212547		$(TC)_{13}$	N/D
IpBrn00401	BE212715		$(TA)_{49}(TG)_7(AT)_5$	N/D
IpBrn00557	BE212649		$(AG)_{12}$	N/D
IpBrn00590	BG381795		$(TG)_{13}$, $(AT)_{10}$	3′-NTR
IpBrn00593	BE212721		$(AT)_{12}(AC)_{11}(AT)_{34}$	3′-NTR
IpBrn00612	BE212729		$(TG)_{20}$	N/D
IpBrn00613	BE212731		$(GT)_{12}(GA)_6$, $(CT)_{23}$	N/D
IpBrn00622	BE212750		$(GT)_{15}$	N/D
IpBrn00803	BE212779		$(AC)_8$	3′-NTR
IpBrn00813	BE212786		$(GT)_{11}$, $(ATGT)_{10}(GT)_8$	3′-NTR
IpBrn00814	BE212787		$(CA)_8$	N/D
IpBrn00859	BE212806		$(GA)_{10}$, $(GAAT)_6(GAGT)_5$, $(ACGA)_7$	N/D
IpBrn00861	BE212808		$(GTT)_9$	3′-NTR
IpBrn01002	BE212699		$(TG)_8$	N/D
IpBrn01019	BE212708		$(AT)_{10}$	3′-NTR
IpBrn01022	BE212710		$(GTCT)_6GTAT(GTCT)_3$	N/D
IpBrn01050	BE212833		$(CA)_8CC(CA)_7$	N/D
IpBrn01071	BE212841		$(CA)_{14}$, $(CTTT)_7$	3′-NTR
IpBrn01085	BE212846		$(ATTT)_5$	N/D
IpBrn01121	BE212864		$(TG)_{20}$	3′-NTR
IpBrn01145	BE212872		$(ATT)_{12}(AGT)_6$, $(ATT)_{18}$	N/D
IpBrn01146	BE212873		$(AC)_{13}T(CA)_6$, $(TA)_{11}$	3′-NTR
IpBrn01148	BE212874		$(GT)_{19}$	3′-NTR
IpBrn01188	BE212886		$(ATT)_{12}$	N/D
IpBrn01189	BE212887		$(AG)_{23}$, $(ATT)_{10}$	N/D
IpBrn01206	BE212892		$(TA)_{16}$	N/D

TABLE 1—*Continued*

Clones	Accession No.	Gene identity	Microsatellite repeat type	Location of microsatellites
IpBrn01211	BE212894		$(TG)_{11}$	3′-NTR
IpBrn01216	BE212897		$(TA)_{9}$	3′-NTR
IpBrn01242	BE212903		$(TTA)_{12}$	N/D
IpBrn01276	BE212920		$(ATT)_{14}$	3′-NTR
IpBrn01293	BE212927		$(AAT)_{5}$TAT $(AAT)_{4}$, $(ATT)_{9}$	N/D
IpBrn01332	BE213066		$(AT)_{32}$	N/D
IpBrn01356	BE213071		$(CAA)_{7}$	N/D
IpBrn01381	BE213023		$(CT)_{7}$	3′-NTR
IpBrn01384	BE213076		$(TA)_{30}$	3′-NTR
IpBrn01395	BE213079		$(TA)_{6}A_{2}(TA)_{3}A_{2}TATG(TA)_{41}$	3′-NTR
IpBrn01404	BE213027		$(TTTA)_{7}$	3′-NTR
IpBrn01408	BE213081		$(CA)_{11}GA(CA)_{9}$	5′-NTR
IpBrn01417	BE213030		$(ATT)_{13}$	3′-NTR
IpBrn01430	BE213085		$(AC)_{12}$	3′-NTR
IpBrn01441	BE213089		$(CAA)_{5}(CAAA)_{6}$	3′-NTR
IpBrn01479	BG381796		$(GT)_{10}$	3′-NTR
IpBrn01485	BE213098		$(GT)_{19}$	3′-NTR
IpBrn01517	BE213039		$(GTT)_{6}$	3′-NTR
IpBrn01581	BE213133		$(AAC)_{7}$	N/D
IpBrn01602	BE213137		$(TA)_{32}$	N/D
IpBrn01661	BE213151		$(TA)_{7}TT(TA)_{7}T(TA)_{8}$	N/D
IpBrn01664	BE213152		$(AT)_{12}$	N/D
IpBrn01727	BE212937		$(CT)_{15}$	3′-NTR
IpBrn02066	BE213201		$(TTA)_{10}$	N/D
IpBrn02075	BE213206		$(TAA)_{14}$	N/D
IpBrn02077	BE213207		$(TTG)_{4}(TTA)_{8}$	N/D
IpBrn02114	BE213222		$(CTGT)_{4}$	N/D
IpBrn02123	BE213225		$>(TTA)_{10}$	3′-NTR

disease is associated with over 1000-fold increased risk of skin cancer (18).

Group C is one of the most common forms of XP (18). The human XP-C gene was found to be homologous to the yeast RAD4 gene (19). High susceptibility to UV-induced carcinogenesis was demonstrated in mice lacking the XP-C gene (20). The XP-C protein was found to be tightly complexed with hHR23B, the human homologue of RAD23 from yeast (21, 22). The binding domain of hHR23B has been identified to be responsible for binding the XP-C protein (22, 23). In reconstituted cell-free NER reactions, hHR23 proteins enhance the repair activity of XP-C, suggesting their functional importance in NER (24).

Presence and expansion of trinucleotide repeats within genes can inactivate the gene and/or lead to gaining of new functions of the gene. It is logical to believe that if such trinucleotide repeat expansions occur in important genes such as those involved in NER pathways, the consequences to the organism may be serious with reduced ability to repair its damaged DNA. Here we report a large number of microsatellite-containing cDNA clones from the channel catfish brain. One of these microsatellite-containing clones was identified as a RAD23 homologue which harbors trinucleotide repeat microsatellites in its coding regions, making it prone to be inactivated by trinucleotide repeat expansion.

MATERIALS AND METHODS

Tissue preparation and RNA isolation. All experimental fish were raised in troughs placed inside of the hatchery of the Auburn University Fish Genetics Facility under the same conditions for 4 weeks before the tissues were harvested. Brain tissues were collected from both the young (9–18 months old) and the mature fish (4–5 years old) at different times of the year to include all transcripts, particularly those that may be developmentally or seasonally regulated. Brain tissues were frozen in liquid nitrogen for grinding with a mortar/pestle and then homogenized with a hand-held tissue tearor (Model 985-370, Biospec Products, Inc., WI) in RNA extraction buffer following the guanidium thiocyanate method (25). Poly(A)$^+$ RNA was purified from total cellular RNA using the Poly(A)$^+$ Pure kit (Ambion, Cat. No. 1915) according to the manufacturer's instructions, except that two rounds of purification were performed.

cDNA library construction and EST analysis. Directional cDNA libraries were constructed using the pSPORT-1 SuperScript plasmid cloning system from GIBCO/BRL. The detailed procedures for construction of the channel catfish brain library were previously described (26).

Plasmid DNA was prepared by alkaline lysis method (27) using Qiagen spin-column miniplasmid kits. Plasmid DNA (1 μl, about 200–500 ng) was used for all sequencing reactions. Sequencing was conducted using the chain termination method (28). Sequencing reactions were performed in a thermocycler using cycleSeq-farOUT polymerase (Display Systems Biotech, Vista, CA). The profiles for cycling were 95°C for 30 s, 55°C for 40 s, 72°C for 45 s for 30 cycles.

```
TGGAGCTCCACCGCGGTGCCTCGCTCTAGAACTAGTGGGTCCCCCGGGCTGCAGGAATTCGGCACGAGAGCAGGCTGTGATGAGCTA
GCGAAAACAGCCTGGACTCACTGTCACACCTTGGGAAAGATCACAGTCTCTCCTCACACACACGCACGGAAACTGCTCAAA ATG C
                                                                                    M
AG ATC ACA CTG AAA ACC CTG CAG CAG CAG ACG TTT AAG ATC GAC ATC GAC GCG GAG GAG ACG GTA
Q  I  T  L  K  T  L  Q  Q  Q  T  F  K  I  D  I  D  A  E  E  T  V
AAG GCG TTA AAG GAG AAA ATT GAG AAT GAG AAA GGG AAG GAG AGT TTT CCT GTC GCT GGA CAG AAG
 K  A  L  K  E  K  I  E  N  E  K  G  K  E  S  F  P  V  A  G  Q  K
TTG ATC TAT GCA GGT AAA ATC CTA AAC GAT GAC ACC GCA CTG AAG GAG TAT AAG ATA GAT GAG AAG
 L  I  Y  A  G  K  I  L  N  D  D  T  A  L  K  E  Y  K  I  D  E  K
AAC TTT GTG GTT GTC ATG GTA GCA AAG[gtaagaaatgaatctttctaatcaaatggtaggctagtttttatattttct
 N  F  V  V  V  M  V  A  K

gtcccgtttttaatgagttcagcaggagattacagctgtgatgttcctcatgcgtaaagccacagatccgaggaaaaagtaaacgaa
atccctttattccaaacacagatggtcagctggggcgtggccactatgtgaagtttttctttagcacaagccagctaacacgtacaa
gctaatctagctaacatgttatgagaatgtggtcacagcacaccagcatcagactctggtaccggccgtaaaagtcttgacgctgga
atcaaaggagtgattcaaatgacagtaatgttggagttttaagtggcttctgctgtattctaaaactgtacacagacagtattcaga
gacgagtgatgtttagcaattctaataatacaaaagtatcagtgcttcaagatgtttctgcaggttaatattccgttcgaattttag
cctaaagcaacatagtagagaggtgtaaatgctgctattaccatagttacgattgttccagatgctctgaagaactgtggcaggttt

ggatttttcagattgtatgaactgattattgtaagtgttttcccag]CCT AAA GCT GCT ACT GCA GCC GCT CAA
                                                P  K  A  A  T  A  A  A  Q
TCT TCA ACT ACT GCA GCA TCA TCG AGT AGT TCC ACC TCT AGC ACT ACA ACG CCC ACA GTC CCC CCT
 S  S  T  T  A  A  S  S  S  S  S  T  S  S  T  T  T  P  T  V  P  P
GTT GCT GCC TCT GCC GCC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACC ACT GAT TCC ACA TCC GAA
 V  A  A  S  A  A  T  T  T  T  T  T  T  T  T  T  T  D  S  T  S  E
AGC AGT GTC ATA GAG GAG AAA GCA GCA GAG GAA AAA CCA CCA TCT TCT ACA CCT GCC AGT TCT GGT
 S  S  V  I  E  E  K  A  A  E  E  K  P  P  S  S  T  P  A  S  S  G
AGT CTG ACG AAC GTC AAC ATA TTC GAG GAG GCG ACT TCT GCT CTG GTG ACA GGA CAG TCG TAT GAG
 S  L  T  N  V  N  I  F  E  E  A  T  S  A  L  V  T  G  Q  S  Y  E
AAC ATG GTG ACT GAG ATC ATG TTA ATG GGC TAT GAG AGA GAG CAA GTG GTG GCA GCG CTG CGT GCC
 N  M  V  T  E  I  M  L  M  G  Y  E  R  E  Q  V  V  A  A  L  R  A
AGT TTC AAC AAC CCA GAC CGA GCT GTT GAG TAC CTG CTC ACG GGA ATT CCA GTG GAG AGT GAG GGG
 S  F  N  N  P  D  R  A  V  E  Y  L  L  T  G  I  P  V  E  S  E  G
AAC GTG GGG GCC TCT GAT CCT GCA GCT CCA GTA GGG GGA GCA CCG GCT GTT ACC ACA GGA CTC TCA
 N  V  G  A  S  D  P  A  A  P  V  G  G  A  P  A  V  T  T  G  L  S
TCA CCA TCC AGC ACA ACA CCT ACA CAG CCC ACT GCT GGA TCA GGA GCT AAT CCT CTG GAG TTT CTG
 S  P  S  S  T  T  P  T  Q  P  T  A  G  S  G  A  N  P  L  E  F  L
AGG AAT CAG CCT CAG TTC CTG CAG ATG AGG CAG ATC ATC CAA CAG AAC CCC TCA CTG CTG CCT GCT
 R  N  Q  P  Q  F  L  Q  M  R  Q  I  I  Q  Q  N  P  S  L  L  P  A
CTA CTG CAA CAG ATC GGC CGA GAG AAC CCC CAG CTA CTG CAG CAA ATC AGC AGT CAC CAG GAG CAG
 L  L  Q  Q  I  G  R  E  N  P  Q  L  L  Q  Q  I  S  S  H  Q  E  Q
TTT ATT CAG ATG CTG AAT GAA CCG GCA CAG GAA GCA GGG CAA GGA GGG GGA GGG GGC GTG TCT GAG
 F  I  Q  M  L  N  E  P  A  Q  E  A  G  Q  G  G  G  G  G  V  S  E
GCT GGT GGT GGG CAC ATG AAC TAC ATA CAG GTC ACA CCC CAG GAA AAG GAA GCA ATT GAG AGG CTA
 A  G  G  G  H  M  N  Y  I  Q  V  T  P  Q  E  K  E  A  I  E  R  L
AAG GCA CTT GGA TTC CCA GAA GGA CTT GTT ATT CAA GCC TAC TTC GCT TGT GAG AAG AAC GAG AAT
 K  A  L  G  F  P  E  G  L  V  I  Q  A  Y  F  A  C  E  K  N  E  N
CTG GCT GCC AAT TTC CTT TTA CAA CAG AAC TTC GAT GAT GAC TAA AAGCCTTTTTTTGTTTTTGTTTTTTGT
 L  A  A  N  F  L  L  Q  Q  N  F  D  D  D  *
TTTGTTTTTGTTCCCTTCCCAATTTACCCGTTTACTTTTTGTTCCTTTTTTTTTCCCTCCAATTAATCATACTGCAGAAAATAGATA
AAATGAAATTCAAACTTAAAAACCCCTCCACAGGCAGTCTTGGTCTATTTTTCTTTACACACTGGATGACTCGTCTGTTTCATCTGC
ATCTTTGTGAAGGCAAGCATGTTCGTTGGGAAATTAAGTAAAACGCATAATCGTGTGAATTCCTCAGATGATGGCCTGGATTATTAT
TTTTGGCTTATATTCATAATTTTACTTTTTGCTGAACACAGGTTTAAAGGGTTCATCGCCATTTTTTTTTCCTCCATCAGTTTGATC
ATTTGTGTCTGAGCTGCACAATCATTTCTGTTCATTGTCAAACTATTTTTCTTTGATCTTTCGCTCGTATCATGCCTTCACAGTTGA
ACGGTATAGGGACCCGAAACAC AATTTTAAA CAACATTAAAAAAGATTTTTAAAGCAAAAAAAAAAAAAAAA
```

FIG. 1. Nucleotide sequence of RAD23B cDNA with one intron (in lowercase letters). The (ACC) repeats and their encoded polythreonine tract are shadowed. The polyadenylation signals AATTTTAAA are underlined.

An initial 2-min denaturation at 96°C and a 5-min extension at 72°C were always used. All sequences were analyzed on an automatic LI-COR DNA Sequencer Long ReadIR 4200 or LI-COR DNA Analyzer Gene ReadIR 4200. Gene identities were identified by BLAST searches using NCBI's nonredundant nucleotide database.

Identification of microsatellite-containing clones. Clones containing SSR were identified by determination of a minimal number of repeats in the microsatellite sequences: dinucleotide, eight repeats; trinucleotide, six repeats; and tetranucleotide, five repeats. Single nucleotide repeats were not included since they are not very useful for polymorphic markers. Some cDNA clones contain more than one type of repeats, in which case these clones were categorized according to the longest repeats.

PCR amplification and Southern blot analysis. For the analysis of the RAD23B locus, genomic DNA was used to amplify the trinucleotide repeat region within the RAD23B gene with an upper primer (5′-CGCTGGACAGAAGTTGATCTATGC-3′) and a lower primer (5′-CTTTCTCCTCTATGACACTGCTTTC-3′). The PCR products were electrophoresed on a 1.0% agarose gel and transferred to Zetabind nylon membrane (Whatman, Wisconsin) by capillary transfer with

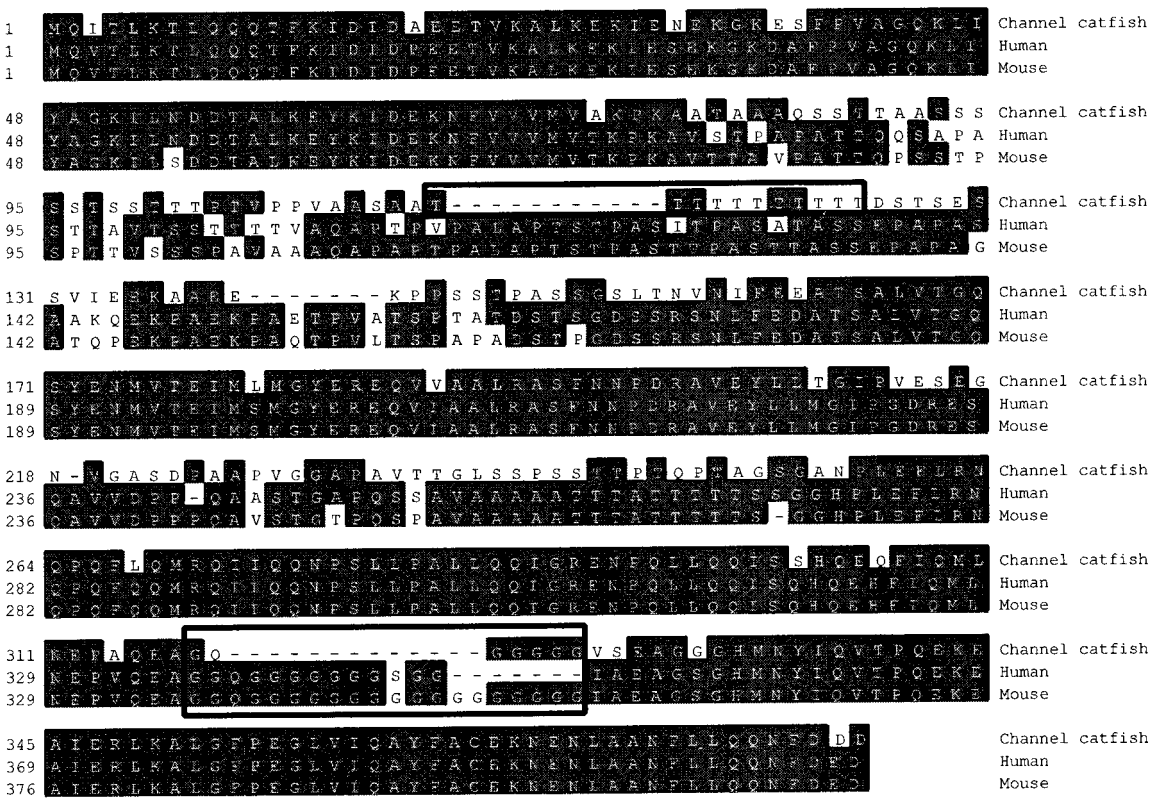

FIG. 2. Similarity comparison of the RAD23B amino acid sequences from channel catfish, human, and mouse. The polythreonine (T) tract and the polyglycine (G) tract are indicated by a black box.

0.4 M NaOH overnight (29). DNA was cross-linked to the membrane by a UV Stratalinker 2400 (Stratagene, CA) using the "auto cross link" function. The membrane was washed in 0.5% SDS (w/v) at 65°C for 15 min and then prehybridized in 5× SSC, 0.1% SDS (w/v), 5× Denhardt's and 100 μg/ml sonicated and denatured salmon sperm DNA overnight at 42°C (27). Hybridizations were conducted overnight at 45°C in a Hy-Blot 12S hybridization oven (Phoenix Research Products, CA) in hybridization solution that is the same solution as in pre-hybridization, but did not include Denhardt's solution. The probe was an oligonucleotide $(CCA)_{10}$ representing the trinucleotide repeats which was labeled by treatment of the oligonucleotide with polynucleotide kinase (Life Technologies, MD) in presence of [γ-^{32}P]ATP (Amersham). This probe was used to detect if the amplified RAD23B loci from various organisms also contain CCA repeats. After hybridization, Zetabind membranes were washed first in 500 ml of 2× SSC for 5 min, followed by three washes in 0.2× SSC with SDS at 0.2% (w/v) at room temperature for two minutes each. The membranes were then wrapped by Saran wrap and exposed to Kodak BioMax MS film overnight.

RESULTS AND DISCUSSION

Microsatellite-Containing cDNAs from the Brain

A total of 1201 random cDNA clones were sequenced from a channel catfish brain cDNA library. The genes and their expression profiles was previously reported (26). Of the 1201 ESTs, 88 unique ESTs contained microsatellite sequences, excluding single nucleotide repeats (Table 1). Of the microsatellite-containing cDNAs, 25 were orthologs of known genes and 63 were unknown genes. These microsatellites should be useful for genomic studies, particularly those within the known genes.

Among the 88 microsatellites, 12 (13.6%) clones included tetranucleotide repeats; 20 (22.7%) clones included trinucleotide repeats; and the remaining 56 clones included dinucleotide repeats. Most microsatellites existed in the 3′ of the EST clones. The RAD 23B gene harbored microsatellites in its coding region. Because of the importance of the RAD23B gene and its function in nucleotide excision repair, the microsatellite-containing RAD23B cDNA was further characterized as described below.

Presence of Trinucleotide Repeats in the Coding Region of RAD 23B Gene

The complete cDNA sequences with one of the introns are shown in Fig. 1. The cDNA has a 5′ untranslated region of 168 bp and a 3′ untranslated region of 518 bp. A putative poly(A)$^+$ signal AATTTTAAA exists 34 bp upstream of poly(A) tails. The open reading frame of the cDNA encode a protein of 385 amino acids, shorter than the mouse Rad23B ORF (416 amino acids) and the human Rad23B ORF (409 amino acids). A trinucleotide microsatellite repeat $(CCA)_{11}$ existed within the coding region of the channel catfish Rad23B gene encoding a polythreonine tract (Fig. 1).

The channel catfish Rad23B coding sequences are highly conserved through evolution compared to its counterparts in human and mouse. Overall, the channel catfish Rad23B gene showed a higher similarity to the human gene (69.9% identity) than to the mouse gene (68.8% identity). While the human gene and the mouse gene are highly similar in the whole amino acid sequences (89.2% identity), three dissimilar regions of sequences in the channel catfish gene apparently "segmented" the protein into four highly conserved blocks of amino acid sequences (Fig. 2). Block 2 and block 4 have been implicated as the ubiquitin associated domains (21). The high levels of evolutionary conservation in block 1 and block 3 may suggest their functional importance as well although further studies are required to dissect the functionality of these conserved blocks.

One striking feature of the Rad23B genes is the presence of two poly-single-amino-acid tracts in the protein sequences (Fig. 2). A polythreonine (T) tract existed only in the channel catfish gene, but not in the human and mouse genes. The polythreonine tract in the catfish gene is encoded by the ACC repeats as a result of the microsatellite sequences. Direct sequence analysis indicated that the microsatellites could specifically invaded in a specific evolutionary group of organisms which likely do not include mammals. As discussed below, it might be devastating for the catfish gene if "run-off" expansion of these trinucleotide repeats occurs.

Although a polyglycine (G) tract is not evident in the channel catfish gene, sequence alignments clearly indicated presence of a polyglycine tract in all three genes (Fig. 2). The mouse gene had the longest polyglycine tract involving 19 glycines with 17 continuous glycine residues. The human gene had a polyglycine tract involving 11 glycines, while the catfish gene had the shortest polyglycine tract with only five continuous glycine residues. Examination of the DNA sequences, however, did not reveal presence of simple microsatellite repeats. The polyglycine tract is encoded by GGN, where N was represented by all bases except that G was not used in the human and mouse genes. It is interesting to imagine expansion of the polyglycine tract through polymerase slippage as in the cases of simple sequence repeats (microsatellites). Because of the fundamental importance of the nucleotide excision repair system for guarding the integrity of the genetic materials, studies are needed to investigate if such expansion may exist in natural populations of human and mice, and if so, how such expansion may affect the functions of the Rad23B protein. It is logical that such expansion may lead to malfunction of the Rad23B protein, thereby adversely affect the ability of NER to fix mutations.

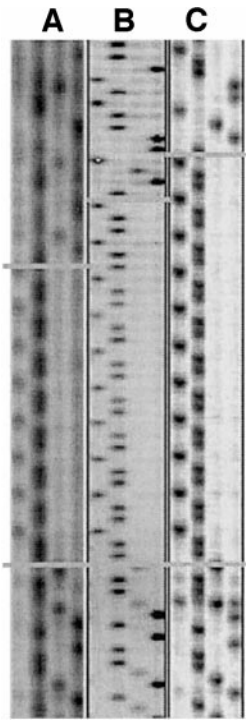

FIG. 3. Allele variations within the (ACC) repeat region of the channel catfish RAD23B locus. Sequencing gels of PCR products from three individuals, A, B, and C, are shown. The (ACC) repeat region is bordered by lines.

Evidence for Trinucleotide Repeat Expansion in the Channel Catfish

The presence of a polythreonine tract in the channel catfish Rad23B coding region raised an important question: do these microsatellite sequences in the coding region expand or shrink as in general cases of microsatellites due to slippage of DNA polymerase (30–32)? This question can be answered by sequence analysis of this locus in the population. To address this question, we collected DNA samples from 20 individuals of channel catfish and amplified their RAD23B locus by PCR. Sequence analysis indicated variation in microsatellite repeat numbers. Although the majority of individuals had 10 (CCA) repeats (18 of 20 or 90%), one individual was homozygous with only 8 (CCA) repeats while the other individual was heterozygous with one allele having 10 (CCA) repeats and the other allele having 11 (CCA) repeats (Fig. 3). Clearly, the microsatellites are not stable in the RAD23B locus. The biological significance of the microsatellite expansion in the RAD23B locus needs to be studied.

Microsatellite Invasion into the RAD23B Gene Is a Recent Event in Evolution

To determine the distribution of microsatellite sequences in various species, genomic DNA was isolated

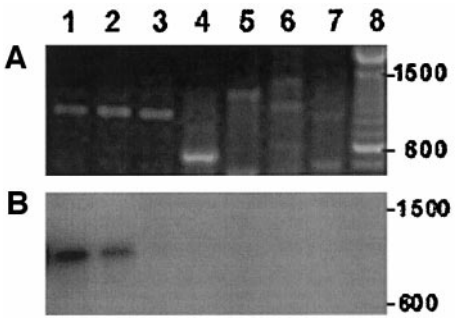

FIG. 4. Distribution of (CCA) repeats in various species as analyzed by PCR. PCR products were analyzed by electrophoresis (A) and by Southern blot (B) using (CCA)10 as probes. Lane 1, channel catfish; lane 2, blue catfish; lane 3, flathead catfish; lane 4, carp; lane 5, tilapia; lane 6, chicken; lane 7, human; lane 8, 100-bp ladder molecular weight markers.

from channel catfish, blue catfish, flathead catfish, carp, tilapia, chicken, and human. The RAD23B segments containing the presumed (CCA) repeats were amplified by PCR and analyzed by Southern blot using $(CCA)_{10}$ oligonucleotide as probes. If the allelic fragments contain CCA repeats, a positive signal should result. As shown in Fig. 4A, fragments of expected sizes were amplified from channel catfish, blue catfish, and flathead catfish, but not from any other species tested. Obviously, we are using primers designed from the channel catfish gene, which may not match the RAD23B sequences in other species. This is particularly true if any sequence variations existed at the 3′ end of the primers. Hybridization with $(CCA)_{10}$ probes indicated that the microsatellite sequences also existed in blue catfish, but not in flathead catfish. This result may be interpreted as either the flathead catfish RAD23B gene does not contain the CCA repeats or that the PCR amplified fragment was non-allelic even though the size was approximately as expected. To confirm the allelic amplification of the flathead catfish RAD23B fragment, the PCR products of flathead catfish were subjected to sequence analysis. The results revealed that the PCR amplified fragment of flathead catfish was indeed the RAD23B fragment, but the flathead catfish gene did not contain continuous CCA repeats. Instead, the CCA repeats were interrupted by CCG and CCT sequences that resulted in failure of hybridization to $(CCA)_{10}$ probes. Flathead catfish belongs to the same family as channel catfish and blue catfish, all in *Ictaluridae.* Although the PCRs failed to produce expected products from all species tested, the lack of continuous CCA repeats in flathead catfish and the fact that there are no CCA repeats in the cloned human and mouse genes indicated that the CCA microsatellite invasion into the RAD23B gene may be a recent event in evolution.

ACKNOWLEDGMENTS

This project was supported by a USDA NRICGP grant to Z.L. and R.D. (98-35205-6738), by the Auburn University Competitive BioGrant (BioGrant J. Liu 99), and partially by a specific cooperative agreement from the USDA ARS Catfish Genetics Unit at Stoneville, Mississippi. We appreciate the support of Auburn University Department of Fisheries and Allied Aquacultures, College of Agriculture, and the Vice President for Research for their matching funds to USDA NRI Equipment Grants to Z.L. (98-35208-6540, 99-35208-8512).

REFERENCES

1. O'Brien, S. J. (1991) *Curr. Opin. Genet. Dev.* **1,** 105–111.
2. Goldberg, Y. P., Telenius, H., and Hayden, M. R. (1994) *Curr. Opin. Neurol.* **7,** 325–332.
3. Ashley, C. T., Jr., and Warren, S. T. (1995) *Annu. Rev. Genet.* **29,** 703–728.
4. Timchenko, L. T., and Caskey, C. T. (1996) *FASEB J.* **10,** 1589–1597.
5. de Vries, B. B., Halley, D. J., Oostra, B. A., and Niermeijer, M. F. (1998) *J. Med. Genet.* **35,** 579–589.
6. Tapscott, S. J., Klesert, T. R., Widrow, R. J., Stoger, R., and Laird, C. D. (1998) *Curr. Opin. Genet. Dev.* **8,** 245–253.
7. Van der Spek, P. J., Visser, C. E., Hanaoka, F., Smit, B., Hagemeuer, A. Bootsma, D., and Hoeijmakers, H. J. (1996) *Genomics* **31,** 20–27.
8. Friedberg, E. C. (1985) DNA Repair, Freeman, New York.
9. Friedberg, E. C., Walker, G. C., and Siede, W. (1995) DNA Repair and Mutagenesis, Am. Soc. Microbiol. Press, Washington, DC.
10. Van Houten, B. (1990) *Microbiol. Rev.* **54,** 18–51.
11. Hoeijmakers, J. H. J. (1993) *Trends Genet.* **9,** 173–177.
12. Huang, J.-C., Svoboda, D. L., Reardon, J. T., and Sancar, A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 3664–3668.
13. Sancar, A., and Hearst, J. E. (1993) *Science* **259,** 1415–1420.
14. Cleaver, J. E., and Kraemer, K. H. (1989) *In* Xeroderma Pigmentosum. The Metabolic Basis of Inherited Disease (Scriver, C. R., Beaudet, A. L., Sly, W. A., and Valle, D., Eds.), Vol. II, pp. 2949–2971, McGraw-Hill Book Co., New York.
15. Masutani, C., Araki, M., Yamada, A., Kusumoto, R., Nogimori, T., Maekawa, T., Iwai, S., and Hanaoka, F. (1999) *EMBO J.* **18,** 3491–3501.
16. Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F. (1999) *Nature* **399,** 700–704.
17. Yamada, A., Masutani, C., Iwai, S., and Hanaoka, F. (2000) *Nucleic Acids Res.* **28,** 2473–2480.
18. Cleaver, J. E., and Kraemer, K. H. (1994) *In* Xeroderma Pigmentosum. The Metabolic Basis of Inherited Disease (Scriver, C. R., Beaudet, A. L., Sly, W. A., and Valle, D., Eds.), Vol. II, 7th ed., pp. 2949–2971, McGraw-Hill Book Co., New York.
19. Legerski, R., and Peterson, C. (1992) *Nature* **359,** 70–73.
20. Sands, A. T., Abuin, A., Sanchez, A., Conti, C. J., and Bradley, A. (1995) *Nature* **377,** 162–165.
21. Masutani, C., Sugasawa, K., Yanagisawa, J., Sonoyama, T., Ui, M., Enemoto, T., Takio, K., Tanaka, K., van der Spek, P. J., Bootsma, D., Hoeijmakers, J. H. J., and Hanaoka, F. (1994) *EMBO J.* **13,** 1831–1843.
22. van der Spek, P. J., Eker, A., Rademakers, S., Visser, C., Sugasawa, K., Masutani, C., Hanaoka, F., Bootsma, D., and Hoeijmakers, J. H. (1996) *Nucleic Acids Res.* **24,** 2551–2559.
23. Masutani, C., Araki, M., Sugasawa, K., van der Spek, P. J., Yamada, A., Uchida, A., Maekawa, T., Bootsma, D., Hoeijmakers, J. H., and Hanaoka, F. (1997) *Mol. Cell. Biol.* **17,** 6915–6923.

24. Sugasawa, K., Masutani, C., Uchida, A., Maekawa, T., van der Spek, P. J., Bootsma, D., Hoeijmakers, J. H., and Hanaoka, F. (1996) *Mol. Cell. Biol.* **16,** 4852–4861.
25. Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159.
26. Ju, Z., Karsi, A., Kocabas, A., Patterson, A., Li, P., Cao, D., Dunham, R., and Liu, Z. J. (2000) *Gene* **261,** 373–382.
27. Sambrook, J., Frisch, E. F., and Maniatis, T. (1989) Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
28. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
29. Reed, K. C., and Mann, D. A. (1985) *Nucleic Acids Res.* **13,** 7207–7221.
30. Pearson, C. E., Wang, Y. H., Griffith, J. D., and Sinden, R. R. (1998) *Nucleic Acids Res.* **26,** 816–823.
31. Petruska, J., Hartenstine, M. J., and Goodman, M. F. (1998) *J. Biol. Chem.* **273,** 5204–5210.
32. Kruglyak, S., Durrett, R., Schug, M. D., and Aquadro, C. F. (2000) *Mol. Biol. Evol.* **17,** 1210–1219.