2005

Isolation and Enrichment of Abundant Microsatellites from a Channel Catfish (*Ictalurus punctatus*) Brain cDNA Library

Dan Nonneman
USDA, ARS, U.S.Meat Animal Research Center, Clay Center, Nebraska

Geoffrey C. Waldbieser
USDA, ARS, Catfish Genetics Research Unit, Thad Cochran National Warm Water Aquaculture Center

Animal Biotechnology, 16: 103–116, 2005

ISSN: 1049-5398 print/1532-2378 online
DOI: 10.1080/10495390500262908

Taylor & Francis
Taylor & Francis Group

ISOLATION AND ENRICHMENT OF ABUNDANT MICROSATELLITES FROM A CHANNEL CATFISH (*ICTALURUS PUNCTATUS*) BRAIN cDNA LIBRARY

Dan Nonneman
USDA, ARS, U.S. Meat Animal Research Center, Clay Center, Nebraska, USA

Geoffrey C. Waldbieser
USDA, ARS, Catfish Genetics Research Unit, Thad Cochran National Warm Water Aquaculture Center, Stoneville, Mississippi, USA

Efforts to construct a genetic linkage map of channel catfish have involved identification of random genomic microsatellite markers, as well as anchored Type I loci (expressed genes) from channel catfish. To identify Type I markers we constructed a directional cDNA library from brain tissue to obtain expressed catfish sequences that could be used for single nucleotide polymorphism (SNP) marker development. These cDNA sequences surprisingly contained a high proportion of microsatellites (about 14%) in noncoding regions of expressed sequence tags (ESTs), many of which were not associated with known sequences. To further identify cDNAs with microsatellites and reduce the number of sequencing reactions needed for marker development, we enriched this library for repeat sequences and sequenced clones from both directions. A total of 1644 clones from seven repeat-enriched captures (CA, GT, CT, GA, MTT, TAG, and TAC) were sequenced from both ends, and 795 nonredundant clones were assembled. Thirty-seven percent of the clones contained microsatellites in the trimmed sequence. After assembly in the TIGR Catfish Gene Index (CfGI), 154 contigs matched known vertebrate genes and 92 contigs contained microsatellites. When BLAST-matched orthologues were available for similarity alignments, 28% of these contigs contained repeats in the 5′-UTR, 72% contained repeats in the 3′-UTR, and 8% contained repeats at both ends. Using biotinylated repeat oligonucleotides coupled with streptavidin-coated magnetic beads, and rapid, single-pass hybridization, we were able to enrich our plasmid library greater than two-fold for repeat sequences and increase the ability to link these ESTs with known sequences greater than six-fold.

Keywords: Channel catfish; Microsatellites; Expressed sequences; Brain

INTRODUCTION

Genetic improvement of livestock requires a dedicated breeding program that relies on the determination of traits that are desirable for future generations. Identification and incorporation of these traits into subsequent generations can be difficult

Mention of trade names or commercial products is solely for the purpose of providing information and does not imply recommendation, endorsement, or exclusion of other suitable products by the U.S. Department of Agriculture.

Address correspondence to Dan Nonneman, USDA, ARS, U.S. Meat Animal Research Center, Spur 18D, P.O. Box 166, Clay Center, NE 68933-0166, USA. E-mail: nonneman@email.marc.usda.gov

and time-consuming based upon phenotype alone. Therefore, marker assisted selection (MAS) of a trait or group of traits is highly desirable and can be achieved through the development of an informative genetic linkage map (1,2). Historically, linkage maps have been developed from random microsatellite markers (Type II markers) that are highly polymorphic and rapidly obtained (3–5), but have little or no comparative value to unrelated species. Anchored loci (Type I markers) are usually then assigned to the map to identify homologous regions of chromosomes between species and further the information obtained with Type II loci (6–9). Ideally, a genetic linkage map would consist of a great number of highly polymorphic, known (Type I) anchored loci (10). However, microsatellite markers in cDNAs of most species are estimated to range from 0.1 to 1% and are therefore not very plentiful for genetic linkage maps (11–13). The degree of polymorphism for Type I and Type II markers varies between species and the reference population from which it is obtained. We constructed a cDNA library from channel catfish brain mRNA and sequenced these clones to identify known genes and find markers that would allow us to place these anchored loci on the catfish genetic linkage map (14).

MATERIALS AND METHODS

Channel Catfish Brain Primary cDNA Library

Brain tissue was obtained from five juvenile USDA103-strain channel catfish at Stoneville, MS, during the month of November, and RNA was extracted using Trizol reagent (Life Technologies, Carlsbad, CA, USA). A channel catfish brain cDNA library was directionally cloned into *Not*I-*Sal*I cloning sites of pSport1 following the protocol for the SuperScript Plasmid cDNA Library kit (Life Technologies, Carlsbad, CA, USA). After transformation into DH5α cells and growth on LB-agar, colonies were grown in 0.1 ml LB media and stored as glycerol stocks at -80°C.

DNA Sequencing and Analysis

Frozen colonies were replicated into 0.35 ml 2XYT media in 0.6 ml deep well plates and cultured overnight in a HiGro shaking incubator (GeneMachines, San Carlos, CA). Bacteria were pelleted by centrifugation in an Eppendorf 5810R centrifuge equipped with an A-2-DWP rotor (Brinkmann Instruments, Inc., Westbury, NY) for 30 min at 1900 × g. Bacterial pellets were resuspended in 75 μl of solution P1 containing 100 μg/ml RNase A (Qiagen, Chatsworth, CA) for 5 min, lysed with 75 μl solution of P2 for 5 min, and neutralized with 105 μl solution of P3. The plate was centrifuged at 1900 × g for 10 min. Two hundred microliters of supernatant were transferred to a 0.45 μm hydrophobic PVDF filter plate (Whatman Inc., Clifton, NJ) with wide bore tips and mixed with 30 μl Procipitate (Ligochem, Fairfield, NJ), which was preloaded onto the filter plate. After a 5 min room temperature incubation, the supernatant was filter-centrifuged into a 0.6 ml deep well receiver plate 20 min at 1900 × g. Two hundred microliters of isopropanol was added to the filtered supernatant, mixed vigorously, and centrifuged 30 min at 1900 × g. Precipitated DNA was washed with 100 μl 70% ethanol, centrifuged 5 min at 1900 × g, inverted to remove the ethanol, and air-dried. The plasmid DNA was resuspended in 40 μl

water, and 400 ng was sequenced using 1/16 dilution (0.5 μl) of BigDye Terminator Cycle Sequencing Ready Reaction Kit or dGTP BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, CA, USA), halfBD Dye Terminator Sequencing Reagent (Sigma, St. Louis, MO, USA), 5% dimethysulfoxide, and 4pmol of T7 or SP6 primer in a total reaction volume of 5 μl. Reactions were precipitated with 20 μl 70% isopropanol, washed with 75% ethanol, resuspended in 10 μl water, and sequences were determined on an ABI PRISM 3700 DNA Analyzer (Applied Biosystems, Foster City, CA).

Base calling was performed using Phred v0.000925.c (15,16). Sequences were processed with Lucy (17) to remove vector and delete sequences with less than 100 Phred20 bases. Before submission to GenBank, mitochondrial and ribosomal RNA transcripts were removed by BlastN (18) screening of the high-quality sequences against a database containing the channel catfish mitochondrial genome and ribosomal sequences. Catfish sequences were downloaded from the TIGR catfish gene index, and microsatellites were identified using Sputnik (www.abajian.net/sputnik) using six repeats as a cutoff score. The TCs were normalized by grouping of clones with identical gene match and/or grouping of clones with separate 5′ and 3′ reads that had different TC designations. The number of microsatellites in expressed sequences of other species was determined by BlastN of the TIGR gene indices (www.tigr.org/tdb/tgi) without a filter, expected threshold of 100, and a word length of 11.

Microsatellite-Enriched Libraries

Microsatellite capture was performed using the GeneTrapper kit (Life Technologies, Carlsbad, CA, USA) and a biotinylated oligonucleotide for hybridization of the total amplified cDNA library. Seven separate enrichments were performed using dCA_9, dGT_9, dCT_9, dGA_9, $dMTT_6$, $dTAG_6$, and $dTAC_6$ oligonucleotides. After capture, repair with T7 primer, and transformation, the clones were sequenced as above using T7 and SP6 primers. Individual colonies of each enriched library were evaluated for presence of the microsatellite repeat by PCR amplification. The reactions used 0.5 μl of a 2 h bacterial culture in 2XYT media, with 5 pmol each of the nonbiotinylated repeat oligonucleotide and the downstream SP6 primer in a total reaction volume of 10 μl. Reaction conditions were 35 cycles of 94°C for 30 sec, 57°C for 45 sec, and 72°C for 1.5 min in a PTC-200 thermalcycler (M.J. Research, Inc., Waltham, MA). Reactions were analyzed by agarose gel electrophoresis and ethidium staining.

Genotype Analysis

Genotyping primers were constructed using GeneRunner (Hastings Software, Hastings-on-Hudson, NY) or Primer3 v.0.9 (19,20) and purchased from Resgen (now Invitrogen, Carlsbad, CA). A 19-bp 5′ extension was added to one primer of the set to facilitate fluorescent labeling of the amplification product (5′-CACGACGTTGTAAAACGAC-3′), and a fluorescent primer with this sequence was synthesized containing 5′ 6-FAM or NED labels (Applied Biosystems). The 15 μl genotyping reactions contained 10 mM Tris-HCl [pH 9.0], 50 mM KCl, 0.1% Triton

X-100, 1 mM $MgCl_2$, 67 μM dNTPs, 0.5 U Taq polymerase (Promega Corporation, Madison, WI) 3 pmoles of the extended primer, 3 pmoles of the fluorescent 19-mer, 6 pmoles of the nonextended primer, and 40 ng genomic DNA. The reaction profile for all loci was 95° for 3 min, 2 cycles of 95° for 1 min and 50° for 30 sec, then 23 cycles of 95° for 30 sec and 50° for 30 sec, then incubation at 72° for 4 min in a PTC-200 thermalcycler (MJ Research, Watertown, MA). One microliter of reaction product was mixed with 9 μl Hi-Dye formamide and 0.2 μl GeneScan-400HD [ROX] size standard (Applied Biosystems). Alleles from two linkage map reference families were determined on the ABI PRISM 3700 DNA Analyzer with GeneScan software (Applied Biosystems).

RESULTS

Channel Catfish Brain Primary cDNA Library

A cDNA library from channel catfish adult brain was constructed in DH5 cells that contained approximately 500,000 individual clones with an average insert size of 1.8 kb (ranging from 0.2 to 4.7 kb). Direct sequencing of 1632 primary clones resulted in 1219 high-quality trimmed sequences of which 16% of these clones were redundant and 6% contained mitochondrial sequences, resulting in 882 unique clones. More than 80% of the clones sequenced had quality sequence of greater than 500 bp and some contained quality sequence to 1 kb, with an overall average read length of 427 bp (Table 1). Sequencing failures were caused by more than one clone per well, lack of growth, or a sequence compression due to a poly (dG-dC) or microsatellite immediately following the cloning/primer site.

In the primary library, 13% of the 5′-end sequences matched to a known sequence in GenBank (Table 1). Fourteen percent of the catfish sequences contained microsatellites with repeat motifs greater than six dinucleotides, trinucleotides, or tetranucleotides. The most common repeat motif in the microsatellite-containing clones was CA/GT (29%) followed by AAT/ATT (25%). Other repeat motifs identified in the catfish cDNAs were AG/CT, AT/TA, AAC/GTT, ACC/GGT, ACG/CGT, AGG/CCT, ATG/CAT, CAC/GTG, AAAC/GTTT, AAAG/CTTT,

Table 1 Comparison of unique clones from primary and microsatellite-enriched brain cDNA libraries

	Match	No match	Total
Primary library			
Microsatellite	14[a]	112	126
No microsatellite	103[b]	653	756
Total	117	765	882
Enriched library			
Microsatellite	92	202	294
No microsatellite	62	439	501
Total	154	641	795

[a]9 clones common to both libraries.
[b]43 clones common to both libraries.

Table 2 and Table 3 Frequency of dinucleotide repeats in animal ESTs

Species	Sequences[a]	CA_{20} (%)	AT_{20} (%)	CT_{20} (%)	Total (%)
Catfish	7,764	8.15	2.24	3.56	14.02
Salmon	24,783	2.94	1.36	1.34	5.63
Zebrafish	74,236	3.83	0.89	0.89	5.61
Trout	40,628	2.73	0.97	1.02	4.72
Mouse	669,402	1.85	0.32	0.69	2.86
Cattle	89,722	0.98	0.43	0.36	1.77
Pig	58,623	0.85	0.47	0.26	1.58
Chicken	111,742	0.48	0.72	0.30	1.50
Human	843,769	0.74	0.19	0.27	1.20

[a]Totally unique sequences assembled by TIGR (http://www.tigr.org).

CTTT, AAAT/ATTT, and AATC/GATT (Table 2). Therefore, an enrichment of CA, GT, CT, GA, MTT, TAG, and TAC repeats was chosen to select for cDNAs that contained microsatellites, since 400–600 CA/GT repeat microsatellite sequences were estimated to be present in the library.

Microsatellite-Enriched cDNA Library

After enrichment of the brain library for CA/GT repeats using a biotinylated $(CA)_9$ or $(GT)_9$ primer and streptavidin-coated magnetic beads, PCR amplification of transformed clones using the nonbiotinylated $(CA)_9$ primer and SP6 vector primer resulted in 88% positive amplification products of clones, and amplification products ranged from 0.3 to 3.0 kb. The size of these products supported our original observation that the microsatellite sequences resided in both the 5′ and 3′ regions of these cDNAs. A total of 1644 clones from 7 repeat-enriched captures (CA, GT, CT, GA, MTT, TAG, and TAC) were sequenced from both ends using T7 and SP6 primers, and 795 nonredundant clones were assembled with an average read length of 442 bp. The redundancy of the enriched library was 51%. Thirty-seven percent of the unique clones (294) contained microsatellites in the vector trimmed sequence (Table 1). The number of unique clones with microsatellites increased 2.2-fold (126 vs. 294), but the number of unique clones containing microsatellites with a GenBank match increased 6.6-fold (14 vs. 92) with enrichment. The enriched library contained 102 new sequences not found in the primary library; 83 of these contained a microsatellite in the sequenced DNA. After assembly in the TIGR Catfish Gene Index (CfGI), 154 (19%) clones were contained in contigs matching known vertebrate genes or ESTs (Table 3), of which 92 contained microsatellites. Twenty-eight percent of the repeats were identified in the 5′-UTR, 72% in the 3′-UTR, and 8.8% contained repeats at both ends of the contigs (Table 3). This library is listed in the TIGR Catfish Gene Index (CfGI) (http://www.tigr.org/tdb/tgi/cfgi/) as library #9VO. Sixty-four clones contained more than one repeat motif, typically two different dinucleotide repeats, although tri- and tetranucleotide repeats were also associated with dinucleotide repeats in the same clone and could be captured using different probes. For example, one clone selected with a CT probe

Table 3 Channel catfish unique EST TIGR contigs (TCs) with sequence identity to known genes

Contig[1]	Gene match	Repeat type[2]	Location[3]
TC3038	14-3-3 protein	AAAT	3′
TC3802	28 kDa-2 apolipoprotein		
TC3786	42Sp48		
TC3091	Actin-related protein 2/3 complex subunit 4		
TC4193	Adenylosuccinate synthetase 2	CA	3′
TC3878	Aldolase C	GT	3′
TC3545	Alpha-2-macroglobulin-3		
TC4492	Aminopeptidase	CA	3′
TC3400	Aminopeptidase N	CA	3′
TC3144	Amisyn	CCA, GT	3′
TC4400	Amylase		
TC3220	AP-3 complex sigma3B subunit	CT; GT	5′;3′
TC3303	ATP citrate lyase	CA	3′
TC4014	Autophagy protein 5-like	CT	3′
TC4422	BAB15607, unnamed protein product		
TC3268	Beta tubulin	CA	5′
TC2953	Beta-actin		
TC3693	CAC22531. unnamed protein product	CA	3′
TC4175	Calmodulin	GT	3′
TC4660	Casein kinase II, alpha chain	CA	3′
TC3570	Catalase		
TC3951	Cd81	CA	3′
TC4317	Chaperonin-containing T-complex protein 1 eta subunit	CA	3′
TC3245	chS-Rex-s	CA	3′
TC4474	Chymotrypsinogen 2		
TC4376	Clusterin precursor		
TC3869	Cofilin, muscle isoform		
TC3935	Cold-inducible RNA binding protein 2		
TC4413	Complement C4B		
TC4405	Cullin homolog 1	GT, CT	3′
TC4030	Cytosolic malate dehydrogenase thermostable form	CT	3′
TC4084	Diazepam binding inhibitor		
TC4498	dJ347H13.4 (novel protein)	CA	3′
TC4563	DMalpha2b	CA; GT	5′;3′
TC4002	DMgamma1		
TC4458	Dok4	GT	5′
TC4639	Downstream Regulatory Element-Antagonist Modulator	GT, CT	3′
TC3102	Dynein light chain 2	CA; GT	5′;3′
TC4117	Elastase 3 precursor		
TC3786	Elongation factor 1-alpha		
TC3875	Ependymin		
TC3788	Eukaryotic translation initiation factor 3 subunit 6 (48 kD)	CA	3′
TC4023	Fatty acid–binding protein, liver		
TC4138	Fatty acyl elongase	CA	5′
TC4437	Fibrinogen alpha chain E variant		
TC3963	Fibrinogen B-beta subunit precursor	CA	3′

(*Continued*)

Table 3 Continued

Contig[1]	Gene match	Repeat type[2]	Location[3]
TC3842	Fibrinogen gamma polypeptide	CA	3′
TC4423	Fish RyR1 isoform	GA, GT	3′
TC3028	Fructose-bisphosphate aldolase C	CA	3′
TC4040	GABA-A receptor-associated protein	CA, CT	3′
TC3257	GDP-fucose protein O-fucosyltransferase 1 precursor		
TC4225	Gene HMG-T2 protein	GT	3′
TC4092	Glutathione S-transferase	CT	3′
TC2963	Glyceraldehyde 3-phosphate dehydrogenase	GT	3′
TC4457	G-protein gamma 3 subunit	CA	3′
TC4623	Growth arrest-specific 7-cb protein		
TC3972	Growth hormone-releasing hormone/PACAP precursor	CT; CG	5′;3′
TC3150	GTP binding protein Rab1a		
TC3302	GTPase cRhoA	CA	3′
TC3033	Heat shock 90 kD protein 1 alpha		
TC4011	Helix-loop-helix protein	GT	3′
TC2966	Hemoglobin beta chain		
TC4126	High choriolytic enzyme 1 precursor		
TC4662	hnRNP-E2 protein		
TC4479	Homogentisate 1 2-dioxygenase	AT	5′
TC3241	HT028		
TC3221	Hypothetical protein FLJ20625	AT; GT	5′;3′
TC3707	Hypothetical protein FLJ21963	CA	5′
TC3716	IP1		
TC3687	KIAA0472 protein	CA	3′
TC3756	KIAA0659 protein	CA	3′
TC4246	KIAA1089 protein		
TC3359	KIAA1208 protein	GT	3′
TC4417	KIAA1399 protein	GT; GT	5′;3′
TC3605	KIAA1678 protein		
TC3234	Lactate dehydrogenase-A		
TC3090	LanC-like protein 1		
TC3921	LD31834p	CA, CT	3′
TC4103	Legumain precursor	AGT, CT, CA	3′
TC4129	LIM/homeobox protein Lhx6.1	CT	3′
TC3513	Lipoprotein lipase		
TC4236	Low molecular weight protein tyrosine phosphatase isoform A	CA	3′
TC3149	MAP kinase kinase 4	GT	3′
TC3486	Metabotropic glutamate receptor 7 precursor	CA	3′
TC4141	Metalloprotease/disintegrin-like protein	GT, CT	3′
TC3837	MHC class I alpha chain	CA	3′
TC3833	MHC class I alpha chain	GA; CA, TTA	5′;3′
TC3011	MHC class I alpha chain	GA, CA	3′
TC3857	MHC class I antigen	GA	5′
TC4102	MHC class II antigen	GT	5′
TC3679	MHC class II antigen	CA	3′
TC3495	Mitochondrial import stimulation factor L subunit	CT	5′

(*Continued*)

Table 3 Continued

Contig[1]	Gene match	Repeat type[2]	Location[3]
TC4199	Myeloid protein-1	CA	3′
TC3062	NAD kinase	CA	5′
TC3048	Nascent polypeptide-associated complex, alpha polypeptide		
TC4054	Neuronal protein NP25	CA	5′
TC3050	Novel protein similar to vertebrate NSF	CA, GT	3′
TC4237	Nucleolar RNA helicase II		
TC3199	Oct2 alpha/beta coding region	ATTC, CA; CT	3′
TC4123	P24B PROTEIN PRECURSOR		
TC3101	Pancreatic carboxypeptidase A1 precursor		
TC4238	Parvalbumin alpha	CT, CA	5′
TC3170	Pentraxin		
TC3070	Phosphofructokinase	GT	3′
TC3756	Pleiotrophin 1	CA	3′
TC3239	Polyadenylate-binding protein, testis-enriched isoform		
TC3538	Procarboxypeptidase B		
TC3906	Protein synthesis factor	CT, GT	3′
TC3117	Ran binding protein 1	GGA; CT	CDS;3′
TC4234	Ras-related protein Rab-18	TTA	3′
TC4182	Ras-related protein Rab-2	CT	3′
TC4107	Ras-related protein Rab-7	CA, CA	3′
TC3779	Ribonucleoprotein		
TC4204	RIKEN cDNA 1110008F13 gene	CT	3′
TC3162	RNA polymerase B transcription factor 3	CA	5′
TC3481	rS-Rex-s	CA, GT	3′
TC3343	Serine protease inhibitor-E2		
TC3349	SH3 domain protein 2 A	CT, GT	5′
TC3852	Solute carrier family 25 member 5 protein	GT	3′
TC3737	Spermine synthase	GA	5′
TC3146	Splicing factor, arginine/serine-rich 1	CA	3′
TC4132	SPPL2a protein		
TC4350	Stathmin 3		
TC3708	Stress-induced-phosphoprotein 1		
TC4219	Succinate dehydrogenase iron-protein subunit B		
TC3136	Superior cervical ganglion-10 protein	CA	3′
TC4269	Synaptotagmin I	AAAT	3′
TC3957	Synuclein	CT; CA	5′;3′
TC4063	TB2		
TC3904	Transcription factor BTF3		
TC3225	Transferrin		
TC3846	Translation elongation factor 2		
TC3016	Translation elongation factor EF-1 gamma		
TC3218	Translation initiation factor 5 A	GT	3′
TC3010	Translationally-controlled tumor protein (TCTP)		
TC3154	Transmembrane 4 superfamily member 2	CTT; CA, TCG	5′;3′
TC3405	Transmembrane protein BRI		
TC4450	Transposase	CA	5′

(*Continued*)

Table 3 Continued

Contig[1]	Gene match	Repeat type[2]	Location[3]
TC3307	Trifunctional enzyme beta subunit, mitochondrial precursor	CA	3′
TC3132	Triglyceride lipase	CA	5′
TC4034	Trypsin III cationic precursor (EC 3.4.21.4)		
TC3928	Trypsinogen 1		
TC3924	Tubulin alpha-3 chain		
TC4223	Tubulin beta-1 chain		
TC3206	Tubulin beta-4 chain	CA	5′
TC3226	Ubiquitin carboxyl-terminal hydrolase	AAAC	5′
TC4674	Ubiquitously expressed transcript		
TC4110	Unnamed protein product gb:BAC04906	GT, GT	5′
TC3548	Vacuolar ATP synthase 16 kDa proteolipid subunit		
TC4003	Voltage-dependent anion-selective channel protein 2		
TC3187	Voltage-gated potassium channel beta-3 subunit	GT	3′
TC2993	Warm-temperature-acclimation-related-65 kDa-protein	CT	5′
TC4120	Wolf-Hirschhorn syndrome candidate 2 protein		
TC4038	WW domain binding protein-2	GT	3′

[1]Contigs and individual clone sequences available at http://www.tigr.org/tdb/tgi/cfgi.
[2]Repeats found in 5′and 3′ end of gene separated by semicolon.
[3]Location with respect to coding region.

(5′:BM496042 and 3′:BM497061) contained a $(CA)_{17}$ motif, a separate $(CA)_{18}$ motif, and an $(ATTT)_6$ motif. Another clone selected with a CA probe (BM495383) contained separate $(CAT)_8$ and $(GT)_{12}$ motifs.

Twelve clones that were PCR-positive for CA microsatellites, but contained a repeat too far within the clone to be detected by sequencing with vector primers, were resequenced using $(CA)_9$CBN and $(GT)_9$GVN primers. Half of these clones yielded readable sequence from both reactions so genotyping primers could be constructed to amplify across the repeat region. The addition of 5% DMSO to the sequencing reaction and the use of the dGTP sequencing kit facilitated longer and clearer reads through GC-rich repeat regions.

Repeat Polymorphism

Sixty-one percent of the initial primer sets tested (13/33) provided amplification products. The success rate increased to 90% (53/59) for subsequent cDNAs when primer design was weighted toward amplification fragments less than 200 bp. Sample genotyping revealed 86% (20/22) of these markers were polymorphic in catfish linkage map reference families (Table 4). A minimum number of five repeats was tested and found to be polymorphic.

Table 4 Sample of cDNA clones with microsatellite sequences that are polymorphic within two CGRU reference families[a]

Tiger Contig	GenBank ID	Repeat unit	Primers	Alleles (bp)	Polymorphic in reference families
TC2963	BM495658	$(GT)_7(GC)_2(GT)_3GC(GT)_5$	CACGACGTTGTAAAACGACCCACAAA CTCTTTACAAGACC		
			GGTAACACCTACAGCACAACG	188, 191, 193, 200	1 and 2
TC3006	BM496991	$CATA(CA)_{16}$	GTTTGTATACTGGCACGTCACACT		
			CACGACGTTGTAAAACGACAATCCCCA CGTTGTGTCC	215, 217, 224	1 and 2
TC3165	BM496917	$(GA)_{30}$	GTTTACCAATCACTCAGCAAAAC		
			CACGACGTTGTAAAACGACACA CTGTTTACTAATCTTTTTACA	197, 216, 224	1 and 2
TC3210	BM495898	$(CT)_2G(CTGT)_2(CT)_5$	CACGACGTTGTAAAACGACTCTCC GTTTAGTTGTGTG		
			GTTTCAGATAAACAGATGAATTGATG	129, 133, 138	1 and 2
TC3210	BM495924	$(CTGT)_2(CT)_6GTCA$ $(CT)_2ATCT$	CACGACGTTGTAAAACGACGTCTCTCA CTCTGTCTGTTTAC		
			GTTTGCTACTCGGTCACACTACTTTG	171	None
TC3233	BM496906	$(GA)_{31}$	CACGACGTTGTAAAACGACAAAGACTGA CCAATGAACGA		
			GTTTCGTCAGGTGCGTCCATCAGA	263, 267, 269	1 and 2
TC3685	BM496961	$(GA)_{22}CC(GA)_5CC(GA)_6$	CACGACGTTGTAAAACGACTGACAAA CACACCCATAA		
			GTTTAGTCCCCGCATCTG	202, 219, 226	1 and 2
TC3732	BM495911	$(GT)_3GAGC(GT)_7$	CACGACGTTGTAAAACGACGAATATTACT CTTTGCATG		
			GTTTGCAGGAGTAATTGACCAGAG	368	None

TC3751	BM495887	$(TA)_7TTAA$	CACGACGTTGTAAAACGACGGGGG TATTACAGTAAGC		
			GTTTATAGAAATCCCGCTACAAGA	130, 132	1 and 2
TC3963	BM496876	$GA(GT)_7$	GTTTCAGGTTTCACACAGGACGAG		
			CACGACGTTGTAAAACGACGTGGTGT GGATGAACTGGAA	144, 146	1 and 2
TC4011	BM496944	$(AC)_{18}$	CACGACGTTGTAAAACGACACTGTGTA GGACTTCA		
			GTTTGCCGCTGAAATCCACGAATC	134, 142, 145, 149	1 and 2
TC4036	BM496970	$(AC)_3GAGT(GC)_2(AC)_6$	CACGACGTTGTAAAA CGACAGAAGATATTAGACATTGAAA		
			GTTTATCAGTGCCAGTTTAGTTATACAAG	152, 155, 157	1 and 2
TC4046	BM496979	$(CA_4CG)_4(CA)_7GA(CA)_5$ TA$(CA)_4$	GTTTACACAATGACATAGCACC		
			CACGACGTTGTAAAACGACTGCAAATTT CTAACACACAGTTC	158, 160, 162	1 and 2
TC4097	BM496922	$(CT)_4CA(CT)_6TTCT$	CACGACGTTGTAAAACGACGATCTGAT GAGTTACTGTGAG		
			GTTTATATTAGGTAGTGTAGCAAAGG	255	None
TC4143	BM496931	$(CA)_{22}$	CACGACGTTGTAAAACGACGATCA CAACACTGGAAGGGGTC		
			GTTTATTCTGTTATTCGGTGAAAAGTGTG	217, 219, 223, 227	1 and 2
TC4320	BM495903	$(ATG)_7TTGATG$	GTTTCAGGATCAGGATGTTCACAC		
			CACGACGTTGTAAAACGACGAGTACA GTGACAGTTTCA	195, 201, 204	1 and 2
Singleton	BM495774	$(TG)_7CGCATGCG(TG)_4$ CA$(TG)_3$	GTTTATTTCACAACCTGCTAACCTCA		
			CACGACGTTGTAAAACGACGAACTATAGC CACTGCCAGG	201, 203	1 and 2
Singleton	BM495998	$(GT)_{16}$	CACGACGTTGTAAAACGACAGACTCGG ATCAATGTGTC		
			GTTTACAGGAAGTTAAGCAATCTG	142, 161, 164	1 only

[a]Families are F1 offspring of USDA103 strain × Norris strain mating and reciprocal mating.

DISCUSSION

Microsatellite repeats in EST sequences provide a rich source of potentially highly informative markers that can be placed on comparative maps since these sequences can be linked to known loci. The catfish cDNA sequences are unique in that they appear to contain a large proportion of microsatellite-containing EST's compared to other species (21–23). Liu et al. (22) and Karsi et al. (21) reported finding microsatellites in 7% and 9% of brain and skin ESTs, respectively. This is much higher than that found in mammalian ESTs and even more than twice the rate in zebrafish ESTs (Table 2). Many of these catfish ESTs contained repeat regions in mammalian orthologues. Since many of the EST sequences of most species are generated primarily from the 5′-end, the catfish orthologues may provide targets for full-length or 3′ sequencing projects to identify microsatellites in other species as well. Sequencing from the 3′ end of the catfish clones was worthwhile in providing additional markers that would not have been discovered using 5′ sequencing alone. Some of the finished sequences did not contain repeats probably because the microsatellites are too deep in the transcript to reach with a single pass, the sequencing failed at repetitive sequences or through very long poly(A) tails, the hybridization probes bound to repeats other than microsatellites, or contamination from abundant clones that were not eliminated from the library. The average insert size of the primary library was 1.8 kb and would not be covered by 600 bp reads from both ends. The addition of more EST sequencing reads could assemble these reads into complete contigs that contain repeats and provide identification of repeat-containing clones that currently do not have significant Blast hits. As reported for human brain (24) it is possible that cDNA libraries constructed from catfish brain mRNA contain clones corresponding to 5′-truncated transcripts or clones with very long 3′-untranslated regions that would contain microsatellites and not provide a significant Blast hit.

Initial characterization of these clones demonstrated 86% of the cDNAs could be placed on the catfish genetic linkage map (14). Polymorphism was not entirely dependent on repeat length, since repeats of five dinucleotides were polymorphic while some of six or seven were not. Library enrichment sometimes provided better potential markers because neighboring repeats were longer than the targeted repeat. The presence of introns between the primer locations may have caused some failures for initial PCR testing, but sequencing of longer genomic amplification products, or sequencing of BAC clones, should allow the design of appropriate primers for further marker development. Considering alternative methods for marker development of known loci, this approach is still an efficient means of mapping Type I loci. Genomic sequencing will also allow the discovery of SNPs (25), which can be used to place these genes on the genetic linkage maps and provide specific amplification products for physical mapping. Because these repeats reside in transcribed regions that are primarily single-copy, potential problems of marker development within repetitive sequences in intergenic regions may be reduced.

Many microsatellite sequences began immediately following the 5′ cloning site, which leads to speculation as to whether these microsatellite sequences are associated with transcription start sites or are just a cloning artifact from reverse transcription. Microsatellite polymorphisms in the promoter region of the bovine growth hormone receptor gene and tilapia prolactin gene have been associated with differential

expression (26,27). However, very little data are available from other species, which contain few microsatellites in expressed cDNAs. Further analysis of these cDNAs could address the potential role of repeat regions in the expression of the protein products of these genes, since repeat sequences in the 5′ UTR can decrease protein expression without affecting levels of mRNA (28).

The enrichment reduced by one-sixth the number of clones required to identify the same number of microsatellite markers in known genes and doubled the number of unique clones with microsatellites. Enrichment did increase the redundancy of the sequenced clones three-fold; however, deeper sequencing of the library would probably produce even more microsatellites in unique clones before redundancy would reduce efficiency. Enrichment increased the probability of finding more markers, and sequencing multiple transcripts from both ends provided more contigs and identified more anchored markers than sequencing the primary library alone. Addition of polymorphic Type I markers to the catfish genetic linkage map will provide comparative map information with other vertebrate species and aid identification of genomic regions controlling economically important production traits.

ACKNOWLEDGMENTS

The authors gratefully thank the expert assistance of Mary Duke, Debra Harris, Marjorie Jennings, and Mona Kirby; Sherry Kluver for manuscript preparation, and Brent Page, Lanie Bilodeau, and Gary A. Rohrer for critical review of the manuscript. The sequences described here have been submitted to GenBank under the accession numbers BM495166–BM497150.

REFERENCES

1. Kappes SM. Utilization of gene mapping information in livestock animals. Theriogenol 1999; 51:135–147.
2. Georges M. Towards marker assisted selection in livestock. Reprod Nutri Develop 1999; 39:555–561.
3. Groenen MA, Crooijmans RP, Veenendaal A, Cheng HH, Siwek M, van der Poel JJ. A comprehensive microsatellite linkage map of the chicken genome. Genomics 1998; 49: 265–274.
4. Kappes SM, Keele JW, Stone RT, McGraw RA, Sonstegard TS, Smith TP, Lopez-Corrales NL, Beattie CW. A second-generation linkage map of the bovine genome. Genome Res 1997; 7:235–249.
5. Rohrer GA, Alexander LJ, Hu Z, Smith TP, Keele JW, Beattie CW. A comprehensive map of the porcine genome. Genome Res 1996; 6:371–391.
6. Ponsuksili S, Wimmers K, Yerle M, Schellander K. Mapping of 93 porcine ESTs preferentially expressed in liver. Mamm Genome 2001; 12:869–872.
7. Rink A, Santschi EM, Eyer KM, Roelofs B, Hess M, Godfrey M, Karajusuf EK, Yerle M, Milan D, Beattie CW. A first-generation EST RH comparative map of the porcine and human genome. Mamm Genome 2002; 13:578–587.
8. Ruyter-Spira CP, de Koning DJ, van der Poel JJ, Crooijmans RP, Dijkhof RJ, Groenen MA. Developing microsatellite markers from cDNA: A tool for adding expressed sequence tags to the genetic linkage map of the chicken. Anim Genet 1998; 29:85–90.

9. Stone RT, Grosse WM, Casas E, Smith TP, Keele JW, Bennett GL. Use of bovine EST data and human genomic sequences to map 100 gene-specific bovine markers. Mamm Genome 2002; 13:211–215.
10. Moran C. Microsatellite repeats in pig (*Sus domestica*) and chicken (*Gallus domesticus*) genomes. J Hered 1993; 84:274–280.
11. Grosse WM, Kappes SM, McGraw RA. Linkage mapping and comparative analysis of bovine expressed sequence tags (ESTs). Anim Genet 2000; 31:171–177.
12. Rohrer GA, Fahrenkrug SC, Nonneman D, Tao N, Warren WC. Mapping microsatellite markers identified in porcine EST sequences. Anim Genet 2002; 33:372–376.
13. Ruyter-Spira CP, Crooijmans RP, Dijkhof RJ, van Oers PA, Strijk JA, van der Poel JJ, Groenen MA. Development and mapping of polymorphic microsatellite markers derived from a chicken brain cDNA library. Anim Genet 1996; 27:229–234.
14. Waldbieser GC, Bosworth BG, Nonneman DJ, Wolters WR. A microsatellite-based genetic linkage map for channel catfish, *Ictalurus punctatus*. Genetics 2001; 158:727–734.
15. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 1998; 8:175–185.
16. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 1998; 8:186–194.
17. Chou H-H, Holmes MH. DNA sequence quality trimming and vector removal. Bioinformatics 2001; 17:1093–1104.
18. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Molec Biol 1990; 215:403–410.
19. Rozen S, Skaletsky HJ. Primer3. 1998. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html.
20. Rozen S, Skaletsky HJ. Primer3 on the WWW for general users and for biologist programmers. Methods Molec Biol 2000; 132:365–386.
21. Karsi A, Cao D, Li P, Patterson A, Kocabas A, Feng J, Ju Z, Mickett KD, Liu Z. Transcriptome analysis of channel catfish (*Ictalurus punctatus*): Initial analysis of gene expression and microsatellite-containing cDNAs in the skin. Gene 2002; 285:157–168.
22. Liu Z, Tan G, Li P, Dunham RA. Transcribed dinucleotide microsatellites and their associated genes from channel catfish *Ictalurus punctatus*. Biochem Biophys Res Comm 1999; 259:190–194.
23. Liu Z, Li P, Kocabas A, Karsi A, Ju Z. Microsatellite-containing genes from the channel catfish brain: Evidence of trinucleotide repeat expansion in the coding region of nucleotide excision repair gene RAD23B. Biochem Biophys Res Comm 2001; 289:317–324.
24. Seki N, Ohira M, Nagase T, Ishikawa K, Miyajima N, Nakajima D, Nomura N, Ohara O. Characterization of cDNA clones in size-fractionated cDNA libraries from human brain. DNA Res 1997; 4:345–349.
25. Fahrenkrug SC, Freking BA, Smith TP, Rohrer GA, Keele JW. Single nucleotide polymorphism (SNP) discovery in porcine expressed genes. Anim Genet 2002; 33:186–195.
26. Hale CS, Herring WO, Shibuya H, Lucy MC, Lubahn DB, Keisler DH, Johnson GS. Decreased growth in angus steers with a short TG-microsatellite allele in the P1 promoter of the growth hormone receptor gene. J Anim Sci 2000; 78:2099–2104.
27. Streelman JT, Kocher TD. Microsatellite variation associated with prolactin expression and growth of salt-challenged tilapia. Physiol Genom 2002; 9:1–4.
28. Amirhaeri S, Wohlrab F, Wells RD. Differential effects of simple repeating DNA sequences on gene expression from the SV40 early promoter. J Biol Chem 1995; 270:3313–3319.