Mar. Biotechnol. 6, 364–377, 2004
DOI: 10.1007/s10126-003-0039-z

MARINE BIOTECHNOLOGY

Bioinformatic Mining of Type I Microsatellites from Expressed Sequence Tags of Channel Catfish (*Ictalurus punctatus*)

Jerry Serapion,[1] Huseyin Kucuktas,[1] Jinian Feng,[1,2] and Zhanjiang Liu[1]

[1]*The Fish Molecular Genetics and Biotechnology Laboratory, Department of Fisheries and Allied Aquacultures Program of Cell and Molecular Biosciences, Aquatic Genomics Unit, Auburn University, Auburn, AL 36849, USA*
[2]*Northwest Sci-Tech University of Agriculture and Forestry, Yangling, Shaanxi, China*

Abstract: Gene-derived markers are pivotal to the analysis of genome structure, organization, and evolution and necessary for comparative genomics. However, gene-derived markers are relatively difficult to develop. This project utilized the genomic resources of channel catfish expressed sequence tags (ESTs) to identify simple sequence repeats (SSRs), or microsatellites. It took the advantage of ESTs for the establishment of gene identities, and of microsatellites for the acquisition of high polymorphism. When microsatellites are tagged to genes, the microsatellites can then be used as gene markers. A bioinformatic analysis of 43,033 ESTs identified 4855 ESTs containing microsatellites. Cluster analysis indicated that 1312 of these ESTs fell into 569 contigs, and the remaining 3534 ESTs were singletons. A total of 4103 unique microsatellite-containing genes were identified. The dinucleotide CA/TG and GA/TC pairs were the most abundant microsatellites. AT-rich microsatellite types were predominant among trinucleotide and tetranucleotide microsatellites, consistent with our earlier estimation that the catfish genome is highly AT-rich. Our preliminary results indicated that the majority of the identified microsatellites were polymorphic and, therefore, useful for genetic linkage mapping of catfish. Mapping of these gene-derived markers is under way, which will set the foundation for comparative genome analysis in catfish.

Key words: data mining, microsatellite, fish, marker, EST, bioinformatics.

INTRODUCTION

The key issues of structural genomics include analysis of genome structure, organization, and evolution (O'Brien et al., 1999), which are addressed by linkage and physical mapping, genome sequencing, and comparative genome analysis. Development of proper molecular markers is pivotal to structural genomics. Linkage maps have been constructed in several important aquaculture species, including Atlantic salmon (Woram et al., 2003), rainbow trout (Young et al., 1998; Sakamoto et al., 2000; Nichols et al., 2003), tilapia (Kocher et al., 1998; Albertson et al., 2003), catfish (Waldbieser et al., 2001; Liu et al., 2003), eastern and Pacific oysters (Yu and Guo, 2003), and shrimps (Li et al., 2003). However, most markers mapped in these aquaculture species were developed from anony-

Received November 7, 2003; accepted November 10, 2003; online publication May 13, 2004.
Corresponding author: Zhanjiang Liu; e-mail: zliu@acesag.auburn.edu

mous genomic sequences (type II markers, see below), prohibiting comparative genomic analysis.

While variation in DNA sequences is the basis for genetic analysis among different genotypes within a species, genetic colinearity or synteny is the basis for comparative mapping among species. The basic assumption for comparative mapping is that genomes or genome segments are conserved among different species during evolution (Carver and Stubbs, 1997). Within a syntenic group, genes are conserved, though the exact order of genes and their orientations may differ. Under this assumption, the closer the evolutionary relationship between species, the more similar their genomes are expected to be. Consequently, genomes can be mapped in a comparative way, allowing exploitation of research progress from model species (Barbazuk et al., 2000; Postlethwait et al., 2000). For instance, the complete genome maps of human and mouse can serve as a guide for comparative genome analysis of other mammalian species. This saves time, effort, and resources in genome analysis of less-studied species. From the genome information for a model species, one can infer what genes may be involved in the conserved genome segments or chromosomal blocks, transferring genome information from a map-rich species to a map-poor species. Comparative genome analysis, however, requires mapping of a common set of molecular markers from the species under consideration. Such a requirement can only be satisfied by mapping known genes that allow comparisons across species borders of a wide evolutionary spectrum.

Development of markers from known genes has been more difficult. Based on their utilities for comparative mapping, O'Brien 1991 divided molecular markers into type I markers associated with genes of known functions and type II markers associated with anonymous genomic sequences. Only type I markers are useful for comparative gene mapping because they serve as anchorage points for genomic segments for which genomic syntenies are analyzed. A common set of type II markers developed from anonymous sequences can occasionally be identified among very closely related species, allowing comparative genome analysis of these species. Most often, however, such comparative genome analysis is not useful because the genomes involved may not include a map-rich species. In contrast, type I markers are associated with genes that are conserved in a wide spectrum of species through evolution, allowing comparative genome analysis and study of genome evolution. Despite their high utility, type I markers are more difficult to develop (Liu et al., 1999c). Because the gene sequences tend to be less polymorphic due to functional restraints. While nongene sequences are free to mutate, causing higher levels of polymorphism, sequences within protein-coding regions generally show lower levels of polymorphism because of functional selection pressure.

Lack of type I markers in catfish, and in many aquaculture species as well, hinders major progress in genomics and genetic studies (Liu, 2003). Microsatellites and AFLP markers have been used efficiently in construction of the framework linkage map in catfish (Waldbieser et al., 2001; Liu et al., 2003), but genomic locations of these type II markers cannot be compared to other species. Therefore, the lack of type I markers prevents those mapping the catfish genome from taking advantage of progress made in other species such as human, rat, zebrafish, and pufferfish, for which genome sequences will be completed in the near future. We have found that catfish expressed sequence tags (ESTs) are rich resources for the identification of gene-tagged microsatellites (Karsi et al., 2002; Kocabas et al., 2002). The objective of this project was to conduct bioinformatic analysis of ESTs to mine type I microsatellite markers.

ESTs are single-pass sequences of randomly picked complementary DNA clones (Adams et al., 1991). These sequences represent transcribed sequences of the genome or genes. The identity of many ESTs can be revealed by sequence similarity analysis. Therefore, many of the ESTs in the dbEST database (available at http://www.ncbi.nlm.nih.gov) represent known genes. Microsatellites, or simple sequence repeats (SSRs), are short tandem repeats of 1 to 6 bp in length (Tautz et al., 1986). The uniqueness and the value of microsatellites as markers arise from their multiallelic nature, codominant inheritance, relative abundance, even distribution in the genomes, and small locus size, facilitating genotyping using polymerase chain reaction (PCR) (Liu et al., 1999a). When SSRs reside within the transcribed sequences in the exons, they can be identified from ESTs (Liu et al., 1999b, 2001). While ESTs provide means for the identification of genes, microsatellites provide high levels of polymorphism. Tagging of microsatellite sequences to known genes would provide putative type I markers. Here we report the identification of a large number of gene-associated microsatellites for linkage and comparative genome analysis. Development and mapping of large numbers of type I markers will greatly enhance the existing linkage maps, bringing application of genomics for genetic improvement of fish brood through analysis of quantitative trait loci and marker-assisted selection.

Materials and methods

Downloading Catfish ESTs from the dbEST Database

All channel catfish ESTs from the NCBI dbEST database as of July 2003 were downloaded into VectorNTI software (InforMax Inc.). First, ''channel catfish EST'' was used as a string of keywords to search nucleotide sequences at the NCBI databases (http://www.ncbi.nlm.nih.gov). All matched sequences were displayed in FASTA format and sent to a file in a local desktop computer and saved as a text file. The sequences were downloaded by changing the ''display'' window to FASTA, and the ''send to'' window to FILE. After downloading, the file containing all 15,565 sequences was saved as a text file. A database was then established using VectorNTI by dragging the text file into the window of the database. Another 27,468 ESTs used in this study were produced in an ongoing EST project (Z. Liu, unpublished data).

Identification of Microsatellite-Containing ESTs

After the channel catfish ESTs were downloaded into the local computer, the ''Local VectorNTI database'' was opened. Using the search function, all possible combination of dinucleotide, trinucleotide, and tetranucleotide repeats were searched, and a subbase for each repeat was created. Strings of oligo sequences were used to search microsatellites: 8 repeats were used for dinucleotide microsatellites; 6 repeats were used for trinucleotide microsatellites; and 5 repeats were used for tetranucleotide microsatellites. Single nucleotide repeats were not included because they are not very useful as polymorphic markers. Six types of dinucleotide microsatellites were searched: (TC)8, (GA)8, (TG)8, (GC)8, (TA)8, (CA)8. The VectorNTI software package simultaneously searches both strands, so the searches of CA and TG, and GA and TC, were combined. Similarly, the complementary trinucleotide and tetranucleotide repeat searches were also combined. We searched 20 trinucleotide repeats, (ATC)6, (CTC)6, (GTC)6, (TTC)6, (AGA)6, (CGA)6, GGA(6), (TGA)6, (CTG)6, (GTG)6, (TTG)6, (AGC)6, (CGC)6, (GGC)6, (ATA)6, (CTA)6, (GTA)6, (TTA)6, (ACA)6, (CCA)6, and 79 tetranucleotide repeats, (AATC)5, (CATC)5, (GATC)5, (TATC)5, (ACTC)5, (CCTC)5, (GCTC)5, (AGTC)5, (CGTC)5, (GGTC)5, (TGTC)5, (ATTC)5, (CTTC)5, (GTTC)5, (TGTC)5, (ATTC)5, (CTTC)5, (GTT)5, (TTTC)5, (AAGA)5, (CAGA)5, (GAGA)5, (TAGA)5, (ACGA)5, (CCGA)5, (GCGA)5, (TCGA)5, (AGGA)5, (CGGA)5, (GGGA)5, (TGGA)5, (ATGA)5, (CTGA)5, (GTGA)5, (TTGA)5, (ACTG)5, (CCTG)5, (GCTG)5, (TCTG)5, (AGTG)5, (CGTG)5, (GGTG)5, (ATTG)5, (CTTG)5, (GTTG)5, (TTTG)5, (AAGC)5, (CAGC)5, (GAGC)5, (TAGC)5, (ACGC)5, (CCGC)5, (TCGC)5, (AGGC)5, (CGGC)5, (GGGC)5, (TGGC)5, (AATA)5, (CATA)5, (GATA), (ACTA)5, (CCTA)5, (GCTA)5, (TCTA)5, (AGTA)5, (CGTA)5, (GGTA)5, (TGTA)5, (ATTA)5, (CTTA)5, (GTTA)5, (TTTA)5, (AACA)5, (GACA)5, (TACA)5, (ACCA)5, (CCCA)5, (GCCA)5, (TCCA)5.

The microsatellite-containing EST subbase includes the accession number of the EST-containing microsatellites, EST length, and EST description. After establishing the microsatellite-containing database, the microsatellite-containing ESTs were analyzed by cluster analysis using VectorNTI ContigExpress module. The linear assembly algorithm was used. The criteria for clustering were set at a minimum overlap of 30 bases (default is 20 bases). After the cluster analysis, each cluster was visually inspected to ensure fidelity of alignment to avoid pseudo-clusters caused by repetitive elements or long strings of microsatellite repeats. ESTs belonging to contigs and singletons were recorded. When alignment by VectorNTI was in doubt, a second software package, DNASTAR (DNA Star Inc.), was used for sequence alignment. A unique set of microsatellite-containing genes was obtained by combining the contigs and the singletons.

Determination of Microsatellite Location and Lengths

Fast PCR program (Kalendar, 2003; available at (http://www.biocenter.helsinki.fi/bi/bare-1_html/download.htm) was used to analyze the position and lengths of microsatellites within ESTs. The sequence for each microsatellite-containing EST was obtained by searching GenBank using the accession number. The sequences were copied and pasted into the window of FAST PCR. The sequences were first ''cleaned'' using the ''clean sequences'' function. The ''repeat search'' function was used to search microsatellites. Only ''simple'' was checked under ''type of repeats.'' All other variables were left at default. After each search, microsatellite length and positions within the sequence were recorded. For each microsatellite, ''primer design'' function was used to determine if the sequences had sufficient flanking sequences for primer design. Those clones harboring microsatellites at either the beginning or the end were recorded.

Determination of Gene Identities of Microsatellite-Containing ESTs

BLAST similarity comparison was used to reveal the identities of ESTs containing microsatellites. BLAST searches were conducted by using BLASTX program against the nonredundant (nr) database. When accumulated probability of sequence similarity was less than 1×10^4, the tentative identities were established. BLASTX search results were visually inspected to ensure the sequence similarity was not caused by simple amino acids (stretches of 1 or 2 amino acids).

Estimation of Usefulness of Microsatellites

The usefulness of the identified microsatellites depends on their polymorphism. For genome mapping, their usefulness can be tested in the resource families. We have tested a fraction of the identified microsatellites in one of the resource families, F_1-2 × channel catfish-6. PCR primers were designed by FAST PCR and purchased from Sigma Genosys. PCR products were labeled by using tailed primers (Oetting et al., 1995). The sequence of the IR-700- or IR-800-labeled 19-base primer was 5′-GAGTTTTCCCAGTCACGAC 3′, which was modified from the original M13 sequence to reduce the formation of hairpin structures within the primer. The upper primers (same orientation as the messenger RNA sequence) contained this 19-base extension on the 5′ end. PCR reactions used the following conditions: 1× PCR buffer, 2 mM $MgCl_2$, 0.2 mM each of dNTPs, 4 ng upper PCR primer, 6 ng lower PCR primer, 20 fmol labeled primer, 0.25 units of JumpStart *Taq* polymerase (Sigma), and 20 ng genomic DNA, in a total reaction volume of 5 µl. After an initial incubation at 94°C for 90 seconds, PCR was carried out at 94°C for 30 seconds, 50°C for 30 seconds, 72°C for 45 seconds, for 35 cycles. Upon completion of PCR, the reaction was incubated at 72°C for another 10 minutes. The PCR products were analyzed on 7% sequencing gels using a LI-COR automated DNA sequencer.

Results

Searching for ESTs Containing Microsatellites

One of the bottlenecks in catfish genomics is the lack of type I markers (for review, see Liu, 2003). This work aimed to fill some of this gap by bioinformatics mining of microsatellites within catfish ESTs. As of July 2003, a total of 15,565 channel catfish ESTs were available in dbEST database. Another 27,468 ESTs generated from our ongoing EST project were also used for analysis. A complete search of 4 types of dinucleotide microsatellites, 20 trinucleotide microsatellites, and 79 tetranucleotide microsatellites resulted in the identification of 4855 ESTs containing microsatellites, of which 1672 were from ESTs downloaded from dbEST (Table 1). In other words, 12.1% of all catfish ESTs contained microsatellites.

Table 1. Summary of Analysis of ESTs Containing Simple Sequence Repeats

Total number of ESTs analyzed	43,033
Number of ESTs containing SSRs	4,855 (11.2%)
Number of SSR-containing ESTs included in contigs	1,312
Number of SSR-containing EST singletons	3,534
Number of SSR-containing EST contigs	569
Total number of unique genes containing SSRs	4,103

Cluster Analysis of ESTs Containing Microsatellites

Of the 4855 microsatellite-containing ESTs, 1312 clones fell within 569 contigs. These represented genes whose transcripts were sequenced more than once. The remaining 3534 ESTs were singletons; these represented genes whose transcripts were sequenced only once (see Table 1). This study identified a total of 4103 unique genes containing microsatellites. Because our current EST project is ongoing, we attempted to estimate the proportion of genes containing microsatellites using only ESTs downloaded from GenBank. According to the TIGR gene index (available at http://www.tigr.org/tdb/tgi/cfgi/), there were 7764 sequences unique to channel catfish in GenBank as of July 2003, 12.8% of which harbored microsatellites.

Distribution of Microsatellite Types

Dinucleotide repeats were the most abundant within channel catfish ESTs. They accounted for 71.7% of all microsatellite-containing ESTs, followed by 19.2% and 11.5%, respectively, for trinucleotide and tetranucleotide repeats (Table 2). This finding was expected and compatible with the estimate of microsatellites in the catfish gen-

Table 2. Distribution of Simple Sequence Repeat Types Among All SSR-Containing ESTs

SSR Type	No. of ESTs harboring the type of SSR[a]	% of total SSR-harboring ESTs
Dinucleotide SSRs	3481	71.7
Trinucleotide SSRs	932	19.2
Tetranucleotide SSRs	558	11.5
Total	4855[a]	102.4

[a]Note: A total of 116 ESTs (2.4%) harbor more than one type of microsatellite.

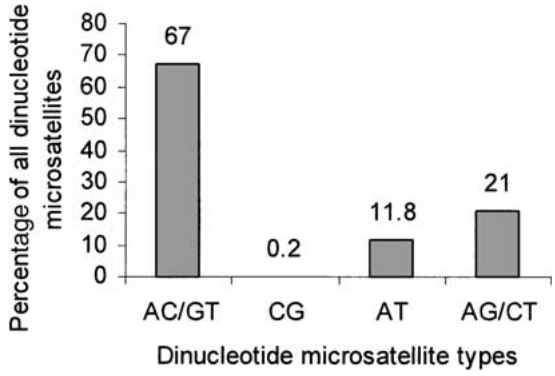

Figure 1. Distribution of dinucleotide microsatellites.

ome (Liu et al., 1999a). A total of 116 ESTs contained more than one type of repeats.

Of the dinucleotide repeats, CA/TG was the most abundant type, accounting for 67% of all dinucleotide repeats found in the catfish ESTs. GA/TC was the second most abundant dinucleotide repeat type, accounting for 21% of all dinucleotide repeats. The AT repeat rate was much lower, at 11.8%, while the CG repeat was rare at a rate of 0.2% (Figure 1). This distribution of dinucleotide repeats is similar to what has been found in other organisms including mammals (Dietrich et al., 1998), arthropods (Toth et al., 2000), and various plant species (Gupta and Varshney, 2000).

Similar to the situation of dinucleotide repeats, trinucleotide repeats were not evenly distributed. The two most abundant types were ATA and TTA, with each being over 25% of all trinucleotide microsatellites found (Figure 2). CAA, ATC, and ATG were each a little over 5%. All 13 other types of trinucleotide repeats were less than 3%. Two types of trinucleotide repeats, CGC and GGC, were not found.

The most abundant tetranucleotide repeats were TTTG (16.3%), AATA (11.1%), and TTTA (9.5%). The intermediately abundant tetranucleotide repeats were AACA (3.9%), AAGA (3.3%), GATA (3.3%), TAGA (3.3%), TATC (3.3%), and TCTA (3.3%). The remaining 70 types of tetranucleotide repeats, when combined, accounted for only 42.8% of all tetranucleotide repeats (Figure 3).

Potential Type I Microsatellites in Catfish

The high proportion of microsatellite-containing ESTs within the channel catfish ESTs offered a unique opportunity for the development of type I markers. As the microsatellite-associated genes are identified and polymorphism is validated, the microsatellites are converted into type I polymorphic markers. To obtain a reasonable estimate of the proportion of known genes among the microsatellite-containing ESTs, we conducted BLASTX searches only on the ESTs downloaded from GenBank. BLASTX searches of the 1023 unique channel catfish microsatellite-containing genes indicated that 302 were from known genes (Table 3).

To determine the usefulness of the microsatellites, the FAST PCR program was used to locate the microsatellite within the ESTs and to determine if primers could be designed from the EST sequences. As summarized in Figure 4, of the 302 known genes containing microsatellites, 241 had enough flanking sequences for primer design, while 61 harbored microsatellite repeats at either the beginning or the end of the EST sequences. The 241 microsatellites should be useful as markers for genetic mapping. We tested 30 randomly selected microsatellites and found that the majority of microsatellites were informative in our resource family of F_1-2 × channel catfish-6 (data not shown). Therefore, the majority of these will prove to be type I markers. However, the 61 ESTs containing microsatellites at their ends are not useful unless additional sequences are produced to generate sufficient flanking sequences for primer design.

Discussion

Several studies have used data mining for the development of microsatellite and single nucleotide markers in different plant species (Temnykh et al., 2001; Kantety et al., 2002; Varshney et al., 2002; Thiel et al., 2003), and several authors have reported microsatellite development from aquaculture

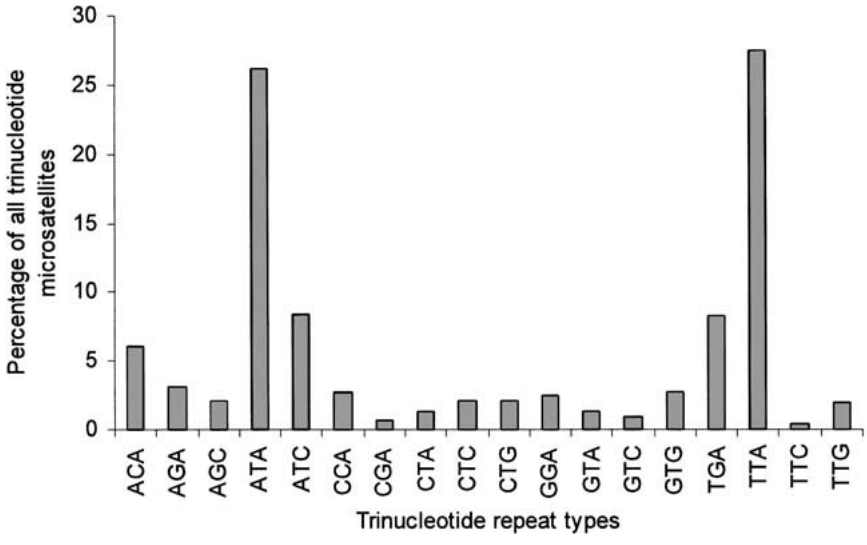

Figure 2. Distribution of trinucleotide microsatellites.

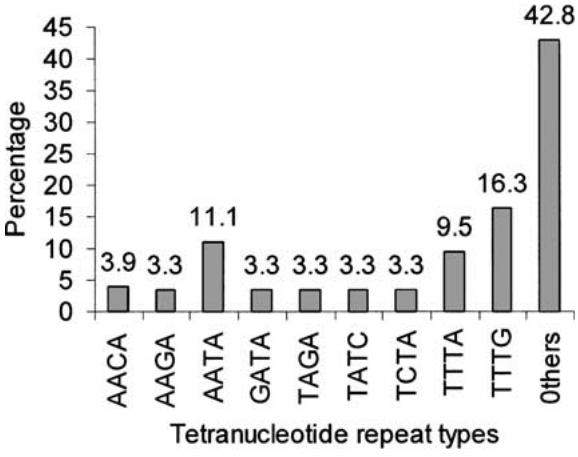

Figure 3. Distribution of tetranucleotide microsatellites. A total of 70 tetranucleotide microsatellites were included in the category of others.

species using traditional approaches (Cagigas et al., 1999; Liu et al., 1999a; Han et al., 2000; Delghandi et al., 2003; Sekino et al., 2003). This study is the first to use bioinformatic analysis for the development of type I markers in an aquaculture species. In addition to the 241 microsatellites identified from known genes, a total of 4103 gene-associated microsatellites were identified from this study, almost 10 times the number of all microsatellites reported from catfish to date. This research suggests that bioinformatic analysis of ESTs is an efficient way to identify polymorphic microsatellite markers, especially type I markers. As large numbers of ESTs have been produced from several aquaculture species (see http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html), the bioinformatic approach is immediately applicable to them, as well as to any species with a large number of ESTs.

Our preliminary testing indicated that the majority of the identified microsatellites are informative in our resource family F_1-2 × channel catfish-6, making them suitable for genome mapping in catfish. This is consistent with a previous report of extremely high polymorphism in catfish microsatellites (Liu et al., 1999a). Mapping a large number of type I markers in the future will allow comparative genomics. In addition to the great value of the type I markers, many of the unknown gene-associated microsatellites are also useful for comparative mapping. One way to do this is to map them by genetic mapping in catfish and also to BLAST search for them in the Zebrafish Genome Sequences (available at http://www.ensembl.org/Danio_rerio/) and the Fugu Genome Sequences (available at http://genome.jgi-psf.org/fugu6/fugu6.home.html) to compare their genomic locations relative to neighboring genes. Despite their unknown identities, the gene sequences are highly conserved through evolution. BLAST searches of some microsatellite-containing ESTs (with unknown gene identity) also have high frequencies of hits on both zebrafish genome sequence contigs and on the fugu genomic scaffolds (data not shown). As the genome sequencing of these two fish species is near completion, the ESTs should serve as frameworks for comparative mapping on the catfish genome. The nearby genes and their order in zebrafish and fugu can be compared with those of catfish. Such comparative genome analysis would make it possible to integrate the genetic linkage maps with the physical maps. The beauty lies in the association of microsatellites with genes. Gene sequences allow mapping of the genes to physical maps, while microsatellites offer polymorphism, allowing them to be mapped to meiotic maps. Mapping of

Table 3. Microsatellite-Containing Genes Identified by Bioinformatic Mining of Channel Catfish Expressed Sequence Tags

Gene Name	Accession No.	Bp	(Repeat type) No.
5S ribosomal RNA	BM425112	474	(gact)6
6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4	BM495608	518	(ca)8,(gt)9
Adenine nucleotide translocase	BM496807	576	(ac)9
ADP-ribosylation factor 1	BM438455	552	multiple
ADP-ribosylation factor binding protein GGA2	CB938743	673	(ca)16
Ah receptor interacting protein	BM497105	468	(gt)13
Alcohol dehydrogenase class III	BM438601	1017	(gt)11
Aldolase C	BM495493	721	multiple
α Actin gene	AF228714	4746	(tg)8,(gatg)6
Aminoacylase	BE469575	860	(gga)9
Aminopeptidase N	BM496665	479	(gt)13
Arp2/3 protein complex subunit p20	BE469169	872	(ag)8
Asialoglycoprotein receptor	BM438603	788	(tg)14
Aspartyl-tRNA synthetase	BE212773	638	(ac)9,(ca)22
ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit	CB938233	208	(tg)15
ATPase, H+ transporting, lysosomal 21 kDa, V0 subunit	BM424733	723	(at)8
ATPase, H+ transporting, lysosomal, V1 subunit H	BM495066	613	(ac)14
ATPase, Na+/K+ transporting, α3	BE212676	709	(ct)9
Basic helix-loop-helix domain containing, class B2	BE468998	487	(ta)8,(ac)8
Bcl2-associated athanogene 2	CB939860	788	(ca)12
β-2 Microglobulin precursor	AF016041	866	(at)27
Bruno-3 CG12478-PA	BM495289	550	multiple
B-type Fas antigen	BM438524	610	(ata)6
C. elegans Y57G11	AI304256	294	(ga)24
CaM-KII inhibitory protein	BM496054	541	(gat)6
NCCR-1	BM028172	987	(tc)9
Cathepsin S	BE469275	732	(ag)8,(ta)8
CCAAT/enhancer binding protein δ	BM438399	593	(tcca)6
CCAAT/enhancer binding protein β	BM438274	950	(gtt)7
CDC-dependent kinase 8	AI304262	295	(ca)21
CG6686-PB	BM496058	401	(tc)11
CGI-142; hepatoma-derived growth factor 2	CB940133	528	(tc)9;(ct)8
CGI-143 protein	BM424544	755	(gt)13, (tcc)6
Choline transporter-like protein 2	BE469164	852	multiple
S-Rex/NSP	BM496435	626	(tg)11
Clone IgD3 IgD secreted form heavy chain	AF363450	10879	(taga)17, (agat)15
Coactosin-like 1	CB937073	744	(ca)20
Coatomer α subunit (α-coat protein)	CB937768	261	(at)26
Collagen, α1 type V/XI	BM029424	607	(tttg)6
Complement component 1q binding protein	CB940728	749	(aat)6
Coronin	CB936368	667	(ag)15
CREB-binding protein	CB939123	472	(at)31
C-terminal binding protein	CB940586	547	(ga)9
CTL2 protein	BM438595	329	(gt)22
Cyclic nucleotide-gated channel modulatory subunit CNGA4b	AF522298	2454	(ca)15
Cyclin B1	AI304252	207	(gt)16,(gt)10

continued

Table 3. Continued

Gene Name	Accession No.	Bp	(Repeat type) No.
Cyclin L ania-6a	CB937680	648	(ca)17
Cytochrome B-245 heavy chain/Flavocytochrome *b*	CB940654	509	(tg)15
Cytochrome P450 CYP11A	AF063836	2550	(gt)10
Cytosolic IMP-GMP specific nucleotidase	BE469287	756	(gag)6
Deacetylase	BE469112	798	(ac)8,(tg)6
DEAD (Asp-Glu-Ala-Asp) box polypeptide 21	CB939649	819	(gaa)7
DMα2b	BM495692	435	(ca)8
DM γ1	BM496602	399	(tg)8
Dopamine receptor	BM028240	479	(ca)35
Downstream of tyrosine kinase 4	BM495681	527	(tg)8,(ctgt)8
DRAL/Slim3/FHL2	BM425105	462	(tgg)8
Dual specificity phosphatase 2	CB939911	695	(aat)19
Dual specificity phosphatase 5	CB939129	763	(aat)9
Dynein light chain 2	BM495288	630	(ac)37,(acat)10
EIG-1/prefoldin subunit 5	BM027925	583	(ac)14 (ac)10
Enolase	AF227804	472	(ac)13
Estrogen receptor type α	AF253507	6756	multiple
Estrogen receptor type β	AF185568	3804	multiple
Ezrin (involed in growth and cell shape)	BE469153	580	(at)12,(at)8
FGF receptor activating protein 1	CB938602	748	(aat)6
Fibrinogen γ	BM438242	927	(ac)13
FLJ20436	CB939722	795	(at)35
FLJ20436-like protein	BE469698	777	(at)23,(tta)13
Formin-like	BE468993	635	(gt)8
Gaba(A) receptor associated protein	BM497096	662	(ag)12, (tg)11
GAP1	BM495595	631	(ag)14
GEF-2	BE469322	741	(tg)15,(gt)7
Gefiltin	BM497054	565	(ctt)6
GH04877p	BM494185	324	(gttt)6
Glucose transporter 1A	BM027884	997	(at)14,(ttta)6
Glutathione S-transferase	BM497044	618	(ag)26
Glyceraldehyde 3-phosphate dehydrogenase	BM496827	664	(ca)8
Groucho homologue2 ZFGRO2	BM438213	955	(ca)6,(ct)8
Growth hormone	S69215	3579	multiple
Growth hormone-releasing hormone	AF321243	1080	multiple
GTP-binding protein	AW351339	252	(aatt)10
H. sapiens clone RP1-34 M23	BM028059	567	(ag)9
H. sapiens hypothetical protein AL365515	BM028055	944	(tc)18
H. sapiens uncharacterized hematopoietic stem/progenitor cells protein MDS032	BM028039	1100	(ac)6,(ca)10
Helicase KIAA0054	CB938117	760	(agc)6
Hematologic and neurologic expressed sequence 1	CB937667	700	(tc)11
Histone H3	CB938298	861	(ac16)
HLA-B associated transcript 8 isoform G9a short	CB940798	693	(gt)16
HMW prekininogen	BM438538	916	(tcc)6
Homolog of secretory protein SEC7; cytoadhesin 1	CB939877	744	(ttg)7
Homolog of yeast mRNA transport regulator 3	CB940490	433	(cag)8
Hypothetical protein FLJ10315	CB940318	641	(ga)11

continued

Table 3. Continued

Gene Name	Accession No.	Bp	(Repeat type) No.
Hypothetical protein FLJ11526	BM495350	498	(ac)25
Hypothetical protein K1AA0692	CB939453	454	(tcc)6
Hypothetical protein MGC28888	CB940693	522	(gt)9
Hypothetical protein T23D8.2	CB938402	717	(at)24
Hypothetical protein XP 136756	BM495047	642	(aca)7
I κB (IκBL) protein	BM424321	568	(ctc)6
Icpu-UA/3 MHC class I antigen gene	AY008848	5104	multiple
Ictacalcin	BM029086	535	(tttg)5
Id1 protein	BM424745	669	(ca)18
IGF-binding protein 7	BM495182	502	(ca)12
Immunoglobulin G light chain LPG5	AY165788	539	(tg)23
Immunoglobulin G light chain LPG16	AY165790	3094	(tg)28
Immunoglobulin γ heavy-chain	BE212593	631	(gt)12
Immunoglobulin heavy chain gene cluster	AF068137	10767	multiple
Immunoglobulin heavy chain gene cluster	BM424715	664	(ca)9, (tc)15
Immunoglobulin μ heavy chain gene	BM425193	661	(gt)28
Jun B transcription factor	BM438319	802	(agc)8
Keratin 12	BM028141	824	(ggt)6
KIAA0569 protein	CB939433	508	(ag)22
KIAA1208 protein	BM496752	633	(ac)22
KIAA2005 protein	CB937469	766	(atc)9
Latent nuclear antigen	BM494163	749	(tc)10,(ca)8
Leukocyte cell-derived chemotaxin 2 (chondromodulin II)	BM438604	596	(ac)14,(ac)8
Leukocyte DNA binding receptor mRNA	IPU97019	2489	(ca)8,(gatg)7
Lysosomal-associated membrane protein 1	CB938401	628	(actg)6
MEGF5	AI304272	288	(ga)17,(gt)14
MHC class II antigen	AF103002	2322	multiple
MHR23B/RAD23b	BE212984	800	(cac)10
MIP1α	CB937269	794	(tca)9
Mitochondrial aconitase	BM495299	597	(ac)12
MKK4	BM496803	720	(ca)17
MO25-like protein	CB939447	530	(ag)10
Moesin	CB937552	663	(at)13; (at)8
MpH312gene	BM424702	631	(tg)10, (gt)20
Mus musculus uterine protein (LOC55978)	BM438248	1016	(tta)7
Myostatin gene	AF396747	5257	multiple
Na,K-ATPase γ subunit	BM438131	887	(gttt)5
NADH dehydrogenase (ubiquinone) 1 α subcomplex, 6	CB940058	526	(tg)16
NADH dehydrogenase (ubiquinone) 1 β subcomplex, 6	CB940056	583	(gagt)7
NADH-ubiquinone oxidoreductase B8 subunit	BM029173	391	(tg)8
NCCR-1 gene	BE468522	879	(ac)13,(tttg)6
N-ethylmaleimide-sensitive factor	BM495757	711	(tg)17
Neuronal protein	CB937285	746	(ac)20
NG22	BM438659	540	(ca)12
Nuclear chloride ion channel protein	CB940677	734	(ac)9
Nucleoporin-like protein RIP	BM496452	423	(gt)17,(att)8
Nucleotide binding protein 2	BE212939	513	multiple
Oct1 transcription factor	IPJ00267	1872	(gca)6
Odorant receptor gene cluster	BE468285	356	(ac)14

continued

Table 3. Continued

Gene Name	Accession No.	Bp	(Repeat type) No.
Ornithine aminotransferase	CB938959	801	(at)11,(tg)11
Osteopontin-like protein	CB939746	843	(gat)6
Palmitoylated membrane protein 3	BM494396	446	(ac)10
Pancreatic lipase	BM438728	939	(tat)6,(ta)11
Pancreatic somatostatin-14	IPSOM1	784	(aat)7
Parathyroid hormone precursor	BE468334	695	(ca)22
Peroxiredoxin 4	CB941020	571	(ca)8
PHAPI2b protein	CB938827	752	(ag)8
Phosphatidylinositol 3-kinase, P85-α Subunit	CB936847	726	(gt)9
Phosphatidylinositol 3-kinase-related protein kinase	BM425128	739	(gac)6
Phosphatidylinositol-4-phosphate 5-kinase type II	CB938917	786	(ag)16
Phosphofructokinase-P	BM496842	762	(ca)23
Plysia ras-related homolog A2	BM495612	654	(ac)17, (ta)11
Potassium voltage gated channel, shaker related subfamily, member 2	BM496731	535	(ac)14
Preimplantation protein 3	CB939637	776	(ac)13
Prepronerve growth factor homolog	BE469121	722	(aaat)6
Profilin	BM029115	955	(ca)11, (ac)23
Prolactin gene	AF267990	3796	multiple
Proopiomelanocortin (POMC)	AY174051	3022	(aat)19
Protein Htf9C	BM424876	906	(gt)8
Protein L	BM424585	766	(tga)6
Protein phosphatase 2	CB939489	716	(tc)10
Protein tyrosine phosphatase, nonreceptor type 6	BE468514	636	(aat)17
Protein-tyrosine-phosphatase IF1	BM028849	950	(ac)23,(ca)16
PxF protein	CB936502	483	(gt)12
RAB7, member RAS oncogene family	BE213167	773	(ca)11,(ca)9
RalGDS-like protein	CB937529	738	(atg)10,(tc)8
Rap1b	BE468967	756	(ag)10
Ras association domain family 2 isoform 1; rasfadin	CB939349	682	(ac)17
Rasa3 (RAS p21 protein activator 3)	BE469207	798	(ca)8,(tgtt)6
Rac1	BE468617	800	(aat)8,(at)9
Ras-related protein Rab7	BE469205	460	(ac)13,(tg)28
Rho GTPase Cdc42	CB937286	669	(ac)17
Ribonuclease H1	CB940874	510	(ca)12
Ribonuclease inhibitor	BM424646	825	(aat)6
Ribosomal protein L3	BE468958	648	(ag)8
Ribosomal protein S16	BM438251	559	(ct)9
Ribosome associated membrane protein RAMP4	CB938464	727	(cat)11
RIKEN cDNA 1110018B13	BE468706	810	(att)10,(at)11
RIKEN cDNA 1600014C10	CB939303	848	(ac)20 (caa)6
RIKEN cDNA 1600019D15 gene	CB940234	766	(ctc)8
RIKEN cDNA 1810037I17	CB939032	524	(tg)8
RIKEN cDNA 2700027 J02 gene	BM494131	675	(agg)6
RIKEN cDNA 2900070E19	CB939285	455	(tg)16, (tg)9
RIKEN cDNA 4833425P12	BM495657	399	multiple
S100-calcium binding protein A14	BM028229	693	(ct)15
SARA protein	BM027999	902	(tc)21
Secretagogin	CB939225	678	(tc)11
Selenoprotein P	BM027942	733	(ac)25

continued

Table 3. Continued

Gene Name	Accession No.	Bp	(Repeat type) No.
Serine/threonine kinase 17A	BM495747	554	(acc)8
Seven transmembrane helix receptor	BM495683	574	(gt)9
SH2 domain protein 1A	CB938898	732	(gt)15
SH3-domain GRB2-like 2	CB939041	743	(ca)10,(ca)11
Skin cDNA library cDNA 5′, mRNA sequence. Unknown	BM029520	959	(ac)28
SM22 α	BE468618	857	(ac)12
Small inducible cytokine A3	CB939490	632	(cat)14
Small inducible cytokine A8	CB940570	633	(at)18
snRNP core protein D2	CB940539	591	(ac)9
Solute carrier family 25, member 20	CB939628	685	multiple
Solute carrier family 25, member 12	CB939783	643	(tcg)7
Stromal cell derived factor receptor 1	BM494596	368	(ca)25
Stromal interaction molecule 1 (STIM1)	BM425067	346	(tg)19
Small ubiquitin-related protein 1	BM425188	727	(ct)12
Suppressor of cytokine signaling 5	CB939244	637	(aac)10, (agc)7
Synaptosomal-associated protein (snapin)	BE213163	736	(at)12
TAF3 RNA polymerase II, TATA box binding protein	BM497130	656	(ca)14
Tankyrase 1	BM497100	515	(ag)7,(ga)6
T-cell receptor DJC region	AF410785	19570	multiple
Thioredoxin/thioredoxin 2	BM027834	883	(ac)24
Transcription factor IIA	BE212675	708	(ta)41
Transcription factor Sp1	CB940752	740	(agc)6
Transcriptional coactivator (CRSP34)	BE469196	852	(gt)8 (gt)8
Transforming protein bmi1	BE469707	654	(gt)12
Translation initiation factor 1A/4C	BE212674	721	(ga)18
Translation initiation factor 3, subunit 6 interacting protein	CB939205	657	(att)8
Translation initiation factor eIF5A	BE469057	956	(ag)8
NAT-1, N system amino acids transporter	BM438698	935	(ca)8
Tryptophan 2,3-dioxygenase	BM438317	920	(aaac)6,(aac)6
Tumor supressor p53	AF074967	1750	(ac)10
Tumor necrosis factor receptor 2 protein	CB937261	640	(aac)11
Type I phosphatidylinositol-4-phosphate 5-kinase	BM494136	816	(tta)12
Type I rhodopsin gene	BE469135	869	(ttg)7
Type-1 protein phosphatase catalytic subunit isoenzyme	BE212587	801	(tc)10,(tc)21
U6 snRNA-associated Sm-Iike protein LSm2	BE468513	695	(atc)7
Ubiquinol-cytochrome *c* reductase (6.4 kDa) subunit	CB937602	588	(att)6
Ubiquitin specific protease 14	CB940011	687	(ac)10
Ubiquitin-associated protein 2	BE469361	623	(ac)13
Ubiquitin-conjugating enzyme E2D 2	CB937890	576	(tc)8
Ubiquitin-conjugating enzyme E2G 2	CB936804	648	(ac)23, (ac)8
Urokinase receptor	BM425162	470	(ac)25
VAV 1 oncogene	BE469203	953	(ag)10, (ag)20
VHSV-induced protein	BM424658	733	(tc)10
Vitamin-K dependent protein C precursor	BM438559	900	(tta)6
WD-repeat protein 1 (actin interacting protein 1)	CB939294	531	(tc)10
Wiskott-Aldrich syndrome protein interacting protein	CB940207	400	(ac)12, (ct)23
Xpacxe2 protein	BM495706	620	(gt)13
Y box protein YB2	AI304260	388	(ac)14

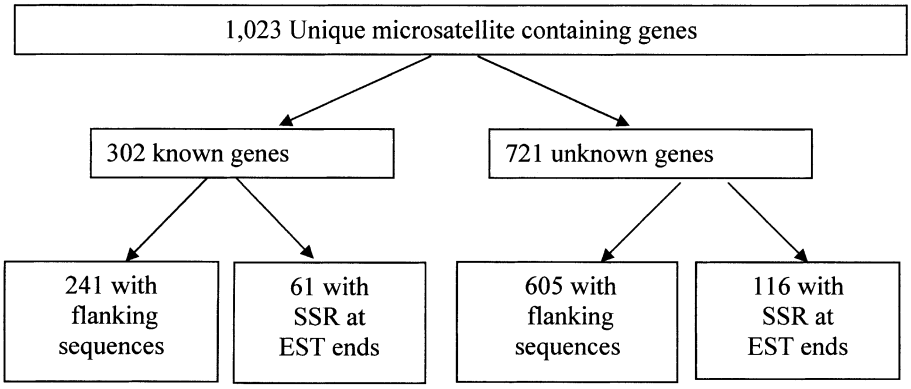

Figure 4. Estimation of the proportion of type I microsatellites and useful microsatellites among the microsatellites identified from ESTs downloaded from GenBank.

a common set of markers on both meiotic and physical maps should lay ground for map integration.

The proportion of microsatellite-containing ESTs in catfish is as high as 11.2%. Considering the unique number of genes containing microsatellites, the proportion in channel catfish is even higher. When only the subfraction of ESTs downloaded from GenBank was used, 1023 unique microsatellite-containing genes were identified from 7764 existing unique gene sequences (TIGR Gene Index, cfgi release 3.0, http://www.tigr.org/tdb/tgi/cfgi/). This indicated that 12.4% of all unique catfish sequences harbor microsatellites. The proportion of microsatellite-containing ESTs in catfish is 2 times that in zebrafish and 7 times that in mammals (Liu, 2003), and is higher than those in several plant species (e.g., barley, wheat, maize, sorghum, and rice), in which microsatellite-containing ESTs were found to range from 1.5% to 4.7% (Kantety et al., 2002).

Part of the high proportion of microsatellite-containing ESTs can be accounted for by high levels of redundant sequencing of some highly expressed microsatellite-containing transcripts. For instance, some of the highly expressed genes such as profilin and MHC type I antigen contained microsatellites. However, the 1312 ESTs formed 569 contigs, with a redundancy number (the number of clones in the contigs divided by the number of contigs) of 2.3, comparable to the redundancy number within contigs of all other catfish ESTs (Cao et al., 2001; Ju et al., 2001; Karsi et al., 2002; Kocabas et al., 2002), indicating that sequencing of highly abundant ESTs containing microsatellites was not the major reason for the high levels of microsatellite-containing ESTs.

Although the exact proportion of microsatellites in the catfish genome is unknown at present, the distribution of microsatellite types, especially the trinucleotide and tetranucleotide repeats, indicates that AT-rich repeat types are the most abundant. This is compatible with the estimation that the catfish genome is AT-rich (Liu et al., 1998). Clearly, GC-rich microsatellite repeats are rare among the catfish ESTs. Further research is needed to determine if this pattern holds true for the entire catfish genome. Toth et al., (2000) observed that ACG and ACT repeats were underrepresented in most taxa. Similar observations were made from this study in catfish. They observed that CCG constituted the most frequent repeat type in exons in primates and rodents, yet in this work no CCG repeats were found. This could be explained by the AT-richness of the catfish genome.

Many ESTs harbor microsatellites at their ends; therefore, additional sequence data are necessary to make them useful. In this study we found 3 types of ESTs harboring microsatellites at their ends. The first type harbors microsatellite at the most 5′ end of the cDNA clone. Microsatellites of this type are not immediately useful. Their flanking sequences are needed for designing PCR primers for genotyping, but cannot be obtained unless the corresponding genomic sequences are available. ESTs of the second type harbor microsatellites immediately upstream of poly(A)$^+$ sequences. The flanking sequences of these microsatellites are also difficult to obtain unless corresponding genomic clones are characterized. ESTs of the third type harbor microsatellites at the end of the EST sequences, but obviously the EST clone can be resequenced to obtain the flanking sequences. ESTs of this last type can be readily resequenced to make them useful.

In spite of the availability of many software packages, suitability of a specific program for bioinformatic analysis of the involved tasks should be tested. Here we used the Windows-based VectorNTI ContigExpress module for contig assembly. The ContigExpress module is an intuitive fragment assembly program for medium-size projects. The linear assembly algorithm was used because of the large number of sequences involved. The advantage of Vector-

NTI lies in its user-friendliness, but its capacity is limited to about 5000 sequences for contig assembly. For contig assembly involving larger data sets, Unix-based programs such as CAP3 and Phrap should be considered. Although many programs such as OLIGO, Primer3, and DNASTAR can be used for PCR primer design, FASTPCR is extremely useful for localization of the microsatellites within sequences as well as for PCR primer design.

Acknowledgments

We thank all the laboratories that deposited channel catfish EST sequences to the dbEST database at NCBI, which made this project possible. This project was supported by a grant from USDA NRI Animal Genome Basic Genome Reagents and Tools Program (USDA/NRICGP 2003-35205-12827). We appreciate the support of Auburn University AAES Foundation Grant and Biogrant from the Office of the Vice President for Research.

References

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., and Moreno, R.F., et al. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252:651–656.

Albertson, R.C., Streelman, J.T., and Kocher, T.D. (2003). Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proc Natl Acad Sci USA* 100:5252–5257.

Barbazuk, W.B., Korf, I., Kadavi, C., Heyen, J., Tate, S., Wun, E., Bedell, J.A., McPherson, J.D., and Johnson, S.L. (2000). The syntenic relationship of the zebrafish and human genomes. *Genome Res* 10:1351–1358.

Cagigas, M.E., Vazquez, E., Blanco, G., and Sanchez, J.A. (1999). Combined assessment of genetic variability in populations of brown trout (*Salmo trutta* L.) based on allozymes, microsatellites, and rapid markers. *Mar Biotechnol* 1:286–296.

Cao, D., Kocabas, A., Ju, Z., Karsi, A., Li, P., Patterson, A., and Liu, Z.J. (2001). Transcriptome of channel catfish (*Ictalurus punctatus*): initial analysis of genes and expression profiles from the head kidney. *Anim Genet* 32:169–188.

Carver, E.A., and Stubbs, L. (1997). Zooming in on the human-mouse comparative map: genome conservation re-examined on a high-resolution scale. *Genome Res* 7:1123–1137.

Delghandi, M., Mortensen, A., and Westgaard, J.I. (2003). Simultaneous analysis of six microsatellite markers in Atlantic cod (*Gadus morhua*): a novel multiplex assay system for use in selective breeding studies. *Mar Biotechnol* 5:141–148.

Dietrich, W.F., Weber, J.L., Nickerson, D.A., and Kwok, P. (1998). Identification and analysis of DNA polymorphisms. In: *Genome Analysis: A Laboratory Manual, Vol. 2: Detecting Genes*, Birren, B., Green, E.D., Klapholz, S., Myers, R.M., and Roskams, J. (eds). Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Gupta, P.K., and Varshney, R.K. (2000). The development and use of microsatellite markers for genetic analyis and plant breeding with emphasis on bread wheat. *Euphytica* 113:163–185.

Han, K., Li, L., Leclerc, G.M., Hays, A.M., and Ely, B. (2000). Isolation and characterization of microsatellite loci for striped bass (*Morone saxatilis*). *Mar Biotechnol* 2:405–408.

Ju, Z., Karsi, A., Kocabas, A., Patterson, A., Li, P., Cao, D., Dunham, R., and Liu, Z.J. (2001). Transcriptome analysis of channel catfish (*Ictalurus punctatus*): genes and expression profile from the brain. *Gene* 261:373–382.

Kalendar, R. (2003). Fast PCR © 1998–2003 Version 2.4 ©: PCR primer design program, repeats search, DNA and protein tools for Windows. Institute of Biotechnology. Available at http:/www.biocenter.helsinki.fi/bi/bare-1_html/download.htm.

Kantety, R.V., La Rota, M., Matthews, D.E., and Sorrells, M.E. (2002). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol* 48:501–510.

Karsi, A., Cao, D., Li, P., Patterson, A., Kocabas, A., Feng, J., Ju, Z., Mickett, K., and Liu, Z.J. (2002). Transcriptome analysis of channel catfish (*Ictalurus punctatus*): initial analysis of gene expression and microsatellite-containing cDNAs in the skin. *Gene* 285:157–168.

Kocabas, A., Li, P., Cao, D., Karsi, A., He, C., Patterson, A., Ju, Z., Dunham, R., and Liu, Z.J. (2002). Expression profile of the channel catfish spleen: analysis of genes involved in immune functions. *Mar Biotechnol* 4:526–536.

Kocher, T.D., Lee, W.J., Sobolewska, H., Penman, D., and McAndrew, B. (1998). A genetic linkage map of a cichlid fish, the tilapia (*Oreochromis niloticus*). *Genetics* 148:1225–1232.

Li, Y., Byrne, K., Miggiano, E., Whan, V., Moore, S., Keys, S., Crocos, P., Preston, N., and Lehnert, S. (2003). Genetic mapping of the kuruma prawn *Penaeus japonicus* using AFLP markers. *Aquaculture* 219:143–156.

Liu, Z. (2003). A review of catfish genomics: progress and perspectives. *Comp Funct Genomics* 4:259–265.

Liu, Z.J., Li, P., and Dunham, R. (1998). Characterization of an A/T-rich family of sequences from the channel catfish (*Ictalurus punctatus*). *Mol Mar Biol Biotechnol* 7:232–239.

Liu, Z.J., Tan, G., Kucuktas, H., Li, P., Karsi, A., Yant, D.R., and Dunham, R.A. (1999a). High levels of conservation at microsatellite loci among Ictalurid catfishes. *J Hered* 90:307–312.

Liu, Z.J., Tan, G., Li, P., and Dunham, R.A. (1999b). Transcribed dinucleotide microsatellites and their associated genes from channel catfish, *Ictalurus punctatus*. *Biochem Biophys Res Commun* 259:190–194.

Liu, Z.J., Karsi, A., and Dunham, R.A. (1999c). Development of polymorphic EST markers suitable for genetic linkage mapping of catfish. *Mar Biotechnol* 1:437–447.

Liu, Z.J., Li, P., Kocabas, A., Ju, Z., Karsi, A., Cao, D., and Patterson, A. (2001). 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 Commun* 289:317–324.

Liu, Z.J., Karsi, A., Li, P., Cao, D., and Dunham, R. (2003). An AFLP-based genetic linkage map of channel catfish (*Ictalurus punctatus*) constructed by using an interspecific hybrid resource family. *Genetics* 165:687–694.

Nichols, K.M., Young, W.P., Danzmann, R.G., Robison, B.D., Rexroad, C., Noakes, M., Phillips, R.B., Bentzen, P., Spies, I., Knudsen, K., Allendorf, F.W., Cunningham, B.M., Brunelli, J., Zhang, H., Ristow, S., Drew, R., Brown, K.H., Wheeler, P.A., and Thorgaard, G.H. (2003). A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*). *Anim Genet* 34:102–115.

O'Brien, S.J. (1991). Molecular genome mapping lessons and prospects. *Curr Opin Genet Dev* 1:105–111.

O'Brien, S.J., Menotti-Raymond, M., Murphy, W.J., Nash, W.G., Wienberg, J., Stanyon, R., Copeland, N.G., Jenkins, N.A., Womack, J.E., and Marshall Graves, J.A. (1999). The promise of comparative genomics in mammals. *Science* 286:458–462479–481.

Oetting, W.S., Lee, H.K., Flanders, D.J., Wiesner, G.L., Sellers, T.A., and King, R.A. (1995). Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. *Genomics* 30:450–458.

Postlethwait, J.H., Woods, I.G., Ngo-Hazelett, P., Yan, Y.L., Kelly, P.D., Chu, F., Huang, H., Hill-Force, A., and Talbot, W.S. (2000). Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res* 10:1890–1902.

Sakamoto, T., Danzmann, R.G., Gharbi, K., Howard, P., Ozaki, A., Khoo, S.K., Woram, R.A., Okamoto, N., Ferguson, M.M., Holm, L.E., Guyomard, R., and Hoyheim, B. (2000). A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genetics* 155:1331–1345.

Sekino, M., Hamaguchi, M., Aranishi, F., and Okoshi, K. (2003). Development of novel microsatellite DNA markers from the Pacific oyster *Crassostrea gigas*. *Mar Biotechnol* 5:227–233.

Tautz, D., Trick, M., and Dover, G.A. (1986). Cryptic simplicity in DNA is a major source of genetic variation. *Nature* 322:652–656.

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., and McCouch, S. (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res* 11:1441–1452.

Thiel, T., Michalek, W., Varshney, K., and Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 106:411–422.

Toth, G., Gaspari, Z., and Jurka, J. (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res* 10:967–981.

Varshney, R.K., Thiel, T., Stein, N., Langridge, P., and Graner, A. (2002). *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell Mol Biol Lett* 7:537–546.

Waldbieser, G.C., Bosworth, B.G., Nonneman, D.J., and Wolters, W.R. (2001). A microsatellite-based genetic linkage map for channel catfish, *Ictalurus punctatus*. *Genetics* 158:727–734.

Woram, R.A., Gharbi, K., Sakamoto, T., Hoyheim, B., Holm, L.E., Naish, K., McGowan, C., Ferguson, M.M., Phillips, R.B., Stein, J., Guyomard, R., Cairney, M., Taggart, J.B., Powell, R., Davidson, W., and Danzmann, R.G. (2003). Comparative genome analysis of the primary sex-determining locus in salmonid fishes. *Genome Res* 13:272–280.

Young, W.P., Wheeler, P.A., Coryell, V.H., Keim, P., and Thorgaard, G.H. (1998). A detailed linkage map of rainbow trout produced using doubled haploids. *Genetics* 148:839–850.

Yu, Z., and Guo, X. (2003). Genetic linkage map of the eastern oyster *Crassostrea virginica* Gmelin. *Biol Bull* 204:327–338.