Original Article

MARINE BIOTECHNOLOGY

Characterization of a BAC Library from Channel Catfish *Ictalurus punctatus*: Indications of High Levels of Chromosomal Reshuffling Among Teleost Genomes

Shaolin Wang,[1] Peng Xu,[1] Jim Thorsen,[2,3] Baoli Zhu,[2] Pieter J de Jong,[2] Geoff Waldbieser,[4] Huseyin Kucuktas,[1] Zhanjiang Liu[1]

[1]The Fish Molecular Genetics and Biotechnology Laboratory, Department of Fisheries and Allied Aquacultures and Program of Cell and Molecular Biosciences, Aquatic Genomics Unit, Auburn University, Auburn, AL 36849, USA
[2]BACPAC Resources, Children's Hospital Oakland Research Institute, Oakland, CA 94609, USA
[3]Department of Basic Sciences and Aquatic Medicine, Section of Genetics, Norwegian School of Veterinary Science, N-0033, Oslo, Norway
[4]USDA, ARS, Catfish Genetics Research Unit, 141 Experiment Station Road, Stoneville, MS 38776, USA

Received: 12 December 2006 / Accepted: 19 April 2007 / Published online: 2 August 2007

Abstract

The CHORI-212 bacterial artificial chromosome (BAC) library was constructed by cloning *Eco*RI/*Eco*RI partially digested DNA into the pTARBAC2.1 vector. The library has an average insert size of 161 kb, and provides 10.6-fold coverage of the channel catfish haploid genome. Screening of 32 genes using overgo or cDNA probes indicated that this library had a good representation of the genome as all tested genes existed in the library. We previously reported sequencing of approximately 25,000 BAC ends that generated 20,366 high-quality BAC end sequences (BES) and identified a large number of sequences similar to known genes using BLASTX searches. In this work, particular attention was given to identification of BAC mate pairs with known genes from both ends. When identified, comparative genome analysis was conducted to determine syntenic regions of the catfish genome with the genomes of zebrafish and *Tetraodon*. Of the 141 mate pairs with known genes from channel catfish, conserved syntenies were identified in 34 (24.1%), with 30 conserved in the zebrafish genome and 14 conserved in the *Tetraodon* genome. Additional analysis of three of the 34 conserved syntenic groups by direct sequencing indicated conserved gene contents in all three species. This indicates that comparative genome analysis may provide shortcuts to genome analysis in catfish, especially for short genomic regions once the conserved syntenies are identified.

Keywords: BAC — catfish — comparative mapping — genome — synteny

Introduction

Large-scale genome research requires a number of genome resources/reagents. These include, but are not limited to, large numbers of polymorphic DNA markers for the construction of genetic linkage maps (Waldbieser et al. 2001; Liu et al. 2003; Serapion et al. 2004a; Xu et al. 2006), normalized cDNA libraries for the analysis of expressed sequence tags (ESTs) (Nonneman and Waldbieser 2005; Liu 2006; Li et al. 2007), a collection of ESTs (Ju et al. 2000; Cao et al. 2001; Karsi et al. 2002; Kocabas et al. 2002; He et al. 2003) supporting the annotation of genes and for the development of cDNA-based microarrays (Ju et al. 2002; Li and Waldbieser 2006; Peatman et al. 2007), and bacterial artificial chromosome (BAC) libraries for the construction of physical maps. BAC libraries are particularly useful not only for the construction of BAC contig-based physical maps, but also for generation of chromosomal markers for fine mapping of regions of interest, for integration of physical and linkage maps, and as the material basis for position-based candidate gene cloning. Because of their high utility, BAC libraries have been developed for many agriculturally important animal species such as cattle (Zhu et al. 1999; Buitkamp et al. 2000; Eggen et al. 2001), swine (Fahrenkrug et al. 2001), and chickens (Zimmer and Verrinder Gibbins 1997; Crooijmans et al. 2000). Recently, a number of BAC libraries have been constructed and

Shaolin Wang and Peng Xu contributed equally to the article.

Correspondence to: Zhanjiang Liu; E-mail: zliu@acesag.auburn.edu

characterized in aquaculture and fish species including salmon, rainbow trout, common carp, tilapia, flounder, oysters, channel catfish, and European sea bass (Katagiri et al. 2000; Katagiri et al. 2001; Quiniou et al. 2003; Katagiri et al. 2005; Thorsen et al. 2005; Cunningham et al. 2006; Whitaker et al. 2006).

For agricultural purposes, the major objective of genome research is to identify genomic regions containing genes controlling performance traits of economic importance. Once the locations of economically important genes are identified, such information can be implemented in marker-assisted selection programs. Initial mapping of the genes of economic importance can be achieved by performing quantitative trait loci (QTL) studies. However, fine mapping of such genes to exact genomic locations can be difficult because of high levels of phenotypic variation, the lack of high-density linkage maps, and the lack of molecular markers in the regions of interest. To circumvent such difficulties, genetic linkage maps containing QTL information is integrated with a well developed physical map. Once the maps are integrated, regional markers can be developed from BAC clones located in the genomic environs of the involved QTLs. A high-quality BAC library resource is crucial for the development of such regional markers (Waldbieser et al. 2003; Rodriguez et al. 2006) leading to the identification and functional characterization of the genes responsible for the QTLs.

Channel catfish is the most important aquaculture species in the United States, representing more than 60% of US aquaculture production. Its genome research is a part of the US National Animal Genome Project NRSP-8. In the area of genome resource development, much progress has been made (for a recent review, see Liu 2003), including the development of a large number of polymorphic markers (Liu et al. 1998b, 1999; Serapion et al. 2004b; Xu et al. 2006), construction of framework genetic linkage map (Waldbieser et al. 2001; Liu et al. 2003), a collection of more than 50,000 ESTs (Li et al. 2007), and generation of more than 20,000 high-quality BAC end sequences (Xu et al. 2006). A first BAC library from brain tissue of gynogenetic catfish (CCBL1/2) was constructed and characterized by Quiniou et al. (2003). A new library from blood of male catfish (CHORI 212) was constructed and the BAC ends sequenced by Xu et al. (2006). The use of alternative restriction enzymes (CCBL1/2: *Hin*dIII; CHORI 212: *Eco*RI) made it possible to clone the refractive fractions (Ng et al. 2005) of the genome that could not be cloned in either. The CHORI 212 library is further characterized in this article. Specifically, the insert size of the BAC library was analyzed; its genomic coverage was characterized; conserved syntenies were identified; and the level of conservation of the identified syntenies and microsyntenies was evaluated to assess the effectiveness of comparative genome analysis of channel catfish using the zebrafish and *Tetraodon* genome sequence resources.

Materials and Methods

Insert Size Analysis. The CHORI-212 BAC library was constructed by cloning the *Eco*RI/*Eco*RI methylase partially digested high molecular weight DNA prepared from a male channel catfish (USDA103 strain) into the pTARBAC2.1 vector between the *Eco*RI sites and transformed into DH10B (T1-resistant) electro-competent cells (Invitrogen, Carlsbad, CA) (http://bacpac.chori.org/library.php?id=103). A total of 350 clones were analyzed by restriction analysis to determine the average size of the BAC inserts. In brief, BAC clones were inoculated in LB medium containing 20 μg/ml of chloramphenicol for 18 h, and the BAC DNA was purified using the automated plasmid isolation machine AutoGen 960 (AutoGen). BAC DNA was digested with *Not*I and analyzed via pulse filed gel electrophoresis (PFGE) (Osoegawa et al. 1998). Low Range PFG Marker (New England Biolabs, Ipswich, MA) was used as the DNA size marker. The molecular weight determination was achieved using an Alpha Innotech MultiImage digital imager and AlphaEase computer software (Alpha Innotech, San Leandro, CA) as described (Thorsen et al. 2005).

Hybridization Screening. To assess the genome coverage of the channel catfish BAC library, overgo hybridization was conducted to screen the BAC library on the high-density filters. The overgo probes were designed based on EST sequences, and are listed in Table 1. The overgo hybridization method was adapted from a Web protocol (http://www.tree.caltech.edu/) with modifications (Bao et al. 2005; Xu et al. 2005). In briefly, overgo primers of 24 bases were selected following a BLAST search against GenBank to avoid repeated sequences and then purchased from Sigma Genosys (Woodlands, TX). Overgos were labeled with [^{32}P]dATP and [^{32}P]dCTP in overgo labeling buffer, at room temperature for 1 h in a 40-μl reaction containing the following: 0.4 μl of bovine serum albumin, 8 μl of overgo labeling buffer [250 mM Tris (pH 8.0), 25 mM $MgCl_2$, 0.36% 2-mercaptoethanol, 1 mM dTTP, 1 mM dGTP, 1 M HEPES-NaOH (pH 6.6)], 2 μl of overgo primer mix, 1.5 μl of [^{32}P]dATP, 1.5 μl of

Table 1. Primer sequences for overgo probes used for hybridization, and hybridization results

Gene identity or sequence used for the design of overgo probes	*Overgo A primer*	*Overgo B primer*	*Positive clones*
NK-lysin 1 and 2	5′-CCTGTGCAATGCACATGGAATACC-3′	5′-GCAGAGTCAACTCTCAGGTATTCC-3′	
NK-lysin 3	5′-GACAAACTCCCAGTAGTGAAGGAT-3′	5′-CCATTTTCTTACACAAATCCTTCA-3′	34
LEAP-2	5′-AGGAGATCAGAGGTCACTCAAGAG-3′	TGTCATACGGGCCATTCTCTTGAG-3′	23
Hepicidin	5′-CTGCTGCAGGTTCTAATAACGGAC-3′	TGAAAACTTGCATGTGGTCCGTTA-3′	9
Interleukin 1 beta	5′-AATATTCAGTCCACGGAGTTCACC-3′	5′-TGAAAAGCTCCTGGTCGGTGAACT-3′	13
BPI	5′-TATCAGCCTTCACCCTGAACTCAG-3′	5′-TTGTACACGAATCCGGCTGAGTTC-3′	9
SCYA101	5′-GCGTTGCTATTTCGCTGGCAAATC-3′	5′-CACACAGTCTCTCTCTGATTTGCC-3′	13
SCYA102	5′-GTGCTGCTTGCACTTTTTGGATGC-3′	5′-CAGGTGCAGTAGTGATGCATCCAA-3′	11
SCYA103	5′-GTCCTCTGTTTTCTCCTGCTTCTG-3′	5′-TTGGGTACATGCATGCCAGAAGCA-3′	1
SCYA104	5′-CCTGTCTTCAGTCCTTCACAATGG-3′	5′-CCGTTTGCATTCTGTGCCATTGTG-3′	11
SCYA105	5′-ACAAACGTCGTGTGTGTGCAAACC-3′	5′-ACCCACTCATCCTTGGGGTTTGCA-3′	1
SCYA106	5′-AACAGCGGCATCTGATATTGGCAC-3′	5′-CACACGTCCTGTTTCTGTGCCAAT-3′	9
SCYA107	5′-AGGCTTCCACCAAAGAAATCACCG-3′	5′-AATCCTGTGATGGGCACGGTGATT-3′	8
SCYA108	5′-GTAAACACCAGTGTGGAAACGCTG-3′	5′-AGAGGAAAGACCTGAGCAGCGTTT-3′	3
SCYA109	5′-CAACCGTAATGGCAAGAGCAAAGG-3′	5′-GGTCTTTCACTGAGCTCCTTTGCT-3′	9
SCYA110	5′-GAAACAGCACTGTGTGGATCCAAC-3′	5′-GTTGACCCAAACAGCTGTTGGATC-3′	11
SCYA111	5′-GCTCATGTTGTTCCTCCTACTTCC-3′	5′-GGGGGAATTTTCCCATGGAAGTAG-3′	12
SCYA112	5′-CCTCCACAAATGTGTGAACACCTC-3′	5′-GACATAGCCACGTGAAGAGGTGTT-3′	7
SCYA113	5′-CAAAGCCTGGTGGAATCCTACTAC-3′	5′-TCTCTGGAGTCTGAACGTAGTAGG-3′	2
SCYA114	5′-CCATCTGGACTGTAACAGATGCAG-3′	5′-CAGGGGCTCACTTTTTCTGCATCT-3′	7
SCYA115	5′-TTCACTGAAGGGATGCGTTTCACG-3′	5′-AGACGTTTTTGGTGCCCGTGAAAC-3′	15
SCYA116	5′-CATGGCCTTTTTGGACCACAGAGG-3′	5′-ATTCCCTGGTGGCATGCCTCTGTT-3′	5
SCYA117	5′-TCTACTCAGACGCTCAGCCTTTTG-3′	5′-TCAGGATGTGCAGGAGCAAAAGGC-3′	26
SCYA118	5′-TCCTAAGCAAGTCCGTGTGACAAG-3′	5′-CCAGTAGCTCACAATGCTTGTCAC-3′	4
SCYA119	5′-CTGCTCTATCCACTCTTCTTCTGC-3′	5′-AGAGGCAGAACACCATGCAGAAGA-3′	11
SCYA120	cDNA probe	cDNA probe	2
SCYA121	5′-AGATGAATCGTGTGGTTTTGGTCC-3′	5′-ATCAGGAAGAAGCCCAGGACCAAA-3′	23
SCYA122	5′-CAGCAAGGCTTCATTGTTACGACG-3′	5′-GGTTAGGGAACTTAGGCGTCGTAA-3′	13
SCYA123	5′-AACGTAGTGTGTGTGCAAACCCCA-3′	5′-TGCACCCACTTATCCTTGGGGTTT-3′	9
SCYA124	5′-CTCGACCTAACCTCAAACGTGTGT-3′	5′-TGGCCAGAGGATTTAAACACACGT-3′	1
SCYA125	5′-TTGACTCAGAGAGACCTCACCTTG-3′	5′-ACTGAATCGCATGGCTCAAGGTGA-3′	6
SCYA126	5′-CTCGTGCTGCTTATTCGTGGAAAG-3′	5′-TTGGTGCGCACAATCTCTTTCCAC-3′	11

[^{32}P]dCTP, 10 U of Klenow polymerase (Invitrogen), and water to bring the volume to 40 µl. After removal of unincorporated nucleotides using a Sephadex G50 spin column, probes were denatured at 95°C for 10 min and added to the hybridization tubes. Hybridization was performed at 54°C for 18 h in hybridization solution [50 ml of 1% bovine serum albumin (BSA), 1 mM EDTA (pH 8.0), 7% sodium dodecyl sulfate (SDS), 0.5 mM sodium phosphate (pH 7.2)]. Filters were washed and exposed to X-ray film at −80°C for 2 days.

BAC Culture and End Sequencing. BAC ends were sequenced as we previously reported (Xu et al. 2006). In brief, BAC clones from the library were inoculated into 2.2-ml 96-well culturing blocks containing 1.5 ml of 2× YT medium and 12.5 µg/ml of chloramphenicol from 384-well stocking plates using 96-pin replicator (V&P Scientific, Inc., San Diego, CA). Blocks were covered by air permeable seal (Excel Scientific, Wrightwood, CA) and incubated at 37°C for 24 h with shaking at 300 rpm. The blocks were centrifuged at 2000 *g* for 10 min in an Eppendorf 5804R bench top centrifuge to precipitate bacteria. The culture supernatant was decanted and the blocks were inverted and tapped gently on paper towel to remove remaining liquid. BAC DNA was isolated using Perfectprep® BAC kit (Brinkmann Instruments, Westbury, NY) according to the manufacturer's specifications. BAC DNA was collected in 96 plates and stored in –20°C before use.

Dye terminator sequencing reactions were conducted in 96-well semiskirt plates using the following ingredients: 2 µl of 5× sequencing buffer, 2 µl of sequencing primer (3 pmol/µl), 1 µl of BigDye v3.1 Dye Terminator, and 5 µl of BAC DNA. The cycling reactions were conducted with MJ Research Thermal Cyclers under the following conditions: initial 95°C for 5 min; then 100 cycles of 95°C for 30 s, 53°C for 10 s, 60°C for 4 min followed by incubation at 4°C. The standard T7 and SP6 primers were used for sequencing reactions (T7 primer: TAATAC GACTCACTATAGGG; SP6 primer: ATTTAGGT GACACTATAG). After sequencing reactions were completed, 1 µl of 125 mM EDTA and 25 µl of prechilled 100% ethanol were added to each well. After mixing and incubating at room temperature for 10 min, the plate was centrifuged at 2250 *g* at 4°C for 40 min followed by washing in 50 µl of 70% ethanol at 1650 *g* for 15 min. Hi-Di formamide (10 µl) was added to each well to resuspend DNA. The DNA was denatured at 95°C, and then the samples were analyzed with an ABI 3130XL automated capillary sequencer (Perkin Elmer-Applied Biosystems). Specific sequencing primers were designed according to the channel catfish EST sequences for the genes identified to reside within the microsyntenic regions.

Sequence Processing and Bioinformatics. The BAC-end sequences (BES) were trimmed of vector sequences and bacterial sequences, stored in a local Oracle database after base calling and quality assessment using Genome Project Management System (GPMS), a local laboratory information management system for large-scale DNA sequencing projects. Quality assessment was performed using Phred software (Ewing and Green 1998; Ewing et al. 1998) using Q =20 as a cutoff. Repeats were masked using Repeatmasker software (http://www.repeatmasker.org) before BLAST analysis.

BLASTX search of the repeat masked BES was conducted against Non-Redundant Protein database. A cutoff value of e-5 was used as the similarity threshold for the comparison. The BLASTX result was parsed out in a tab-delimited format, which allows the data to be formatted into tables readily using word processing software as well as Excel, and facilitates easy table–text conversions. To anchor the catfish BES to zebrafish and *Tetraodon* genomes, BLASTN searches of the repeat masked catfish BES were conducted against zebrafish and *Tetraodon* genome sequences. The location and chromosome number of each top hit was collected from the results and parsed in tab-delimited format.

Identification and Validation of Conserved Syntenies. Initially, mate paired BES were analyzed by BLASTX searches (cutoff e-5) for the identification of mate pairs with genes on both side of the BAC insert. After the identification, the two mate paired genes in each BES were used as queries to search their chromosomal locations on the zebrafish and *Tetraodon* genomes. We limited the distance between the two genes to 1.2 Mb. Conserved synteny was declared when the mate pair genes also exist within a distance of 1.2 Mb in either zebrafish and/or *Tetraodon*. Once the initial conserved syntenies were identified, further validation for the presence of additional genes found between the two conserved genes of zebrafish or *Tetraodon* within the catfish BACs were deter mined by direct BAC sequencing. First, the zebrafish or *Tetraodon* genes present between the two conserved gene pairs were used to search the catfish EST database to determine if such genes had been identified in catfish. Sequencing primers were then designed based on the catfish EST sequences when present and used for direct BAC sequencing. The generated sequences were

Table 2. Summary of the catfish BAC library CHORI 212

Parameters and descriptions	
Vector	pTARBAC2.1
Restriction enzyme used	*Eco*RI/*Eco*RI methylase
DNA source	Blood
Number of 384-plates	192
Recombinant clones	72,067
Empty wells	1,174 (1.59%)
Nonrecombinant clones	52 (0.07%)
No-insert clones	~435 (0.6%)
Average insert size	161 kb
Genomic coverage	~10.6×

aligned to the catfish EST sequences or subjected to BLASTX searches to determine the putative gene identities of the sequences. The distances and orientations of the conserved genes in catfish were not determined.

Results and Discussion

Construction of the Channel Catfish BAC Library CHORI 212. The BAC library, CHORI 212, was constructed by cloning large *Eco*RI restriction fragments into the pTARBAC2.1 BAC vector. The library consists of 72,067 recombinant clones arrayed into 192 "384-well" microtiter plates (Table 2). To determine the average insert size and the size distribution of clones in the catfish library, the BAC DNA was digested with *Not*I, and the insert size was analyzed by pulse field electrophoresis. As shown in Figure 1, the vast majority (>96%) of the BAC clones contained insert size greater than 100 kb, with an average insert size of 161 kb. Based on the genome size of catfish being 1.1×10^9 base pairs, this BAC library has a genome coverage of 10.6×.

Assessment of the BAC Library Quality. Several factors are important for the quality of a BAC library, including the average insert size and genome coverage as presented above, and representations of the genome. With the 10.6× genome coverage and an average of 161 kb, the major issue for the CHORI 212 library now is the representation of the genome. To assess the genome representation, we used overgo and cDNA probes to screen the BAC library. As shown in Table 1, all the probes used produced positive clones, suggesting that the BAC library had a good representation of the genome. However, no single BAC library would allow full representation of the genome. In this case, we also integrated known information into consideration. For instance, the highly repetitive elements of catfish were well characterized including the *Xba* elements (Liu et al. 1998a), and the *Tc1* elements (Nandi et al. 2007). As detailed below, BAC end sequencing of more than 10,000 clones from both ends failed to detect any *Xba* elements while more than 4% of the entire catfish genome was composed of *Tc1*-related sequences. This suggested that the interspersed repetitive elements may be well covered in the BAC library, while some of the repetitive elements arranged in tandem arrays may have been excluded from the BA C library depending on whether the elements contained *Eco*R1 restriction sites, which were used

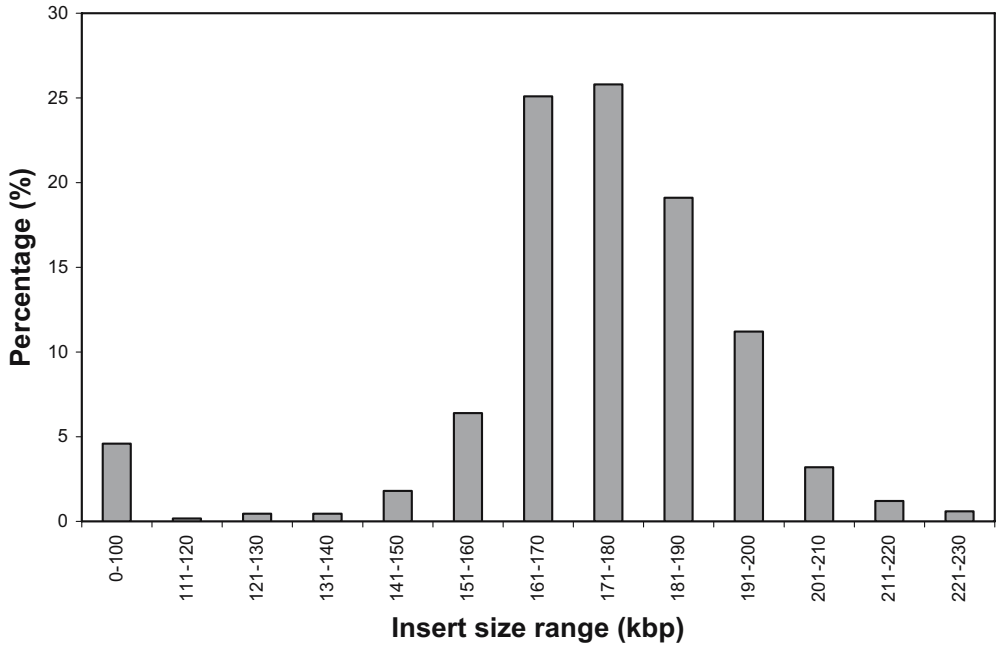

Figure 1. Distribution of insert sizes of the CHORI 212 BAC library.

for the construction of the library. Sequence analysis of *Xba* elements confirmed their lack of *Eco*R1 sites. Therefore, it is clear that complementary BAC libraries would be essential for the whole genome coverage even though the quality of this library was good.

Conserved Syntenies Between Catfish, Zebrafish, and Tetraodon. We previously reported that BAC end sequencing is an effective approach for mapping genes to BACs (Xu et al. 2006). Through the analysis of 20,366 BAC end sequences, we identified a total of 1,877 BAC end sequences that have significant similarities with known gene sequences as revealed by BLAST searches (Table 3). In this study, particular attention was paid to evaluate the level of conservation between the catfish and zebrafish or between the catfish and *Tetraodon* genomes. Of the 20,366 BES, 17,478 BES were mate pair sequences from 8,739 BAC clones. BLASTX searches indicated that 141 sequenced BACs harbor genes on both ends. These paired BAC ends with genes allowed us to compare whether the same sets of genes were located on similar environs in the zebrafish and *Tetraodon* genomes. Zebrafish is the closest to channel catfish in phylogeny with a whole genome sequence, while *Tetraodon* and several other species with whole genome sequences are more distantly related. Of the 141 paired BAC ends with genes, 34 (24.1%) appeared to exhibit a high level of syntenies (Table 4), using approximately one megabase as the cutoff value for conserved syntenies. The level of conserved syntenies was greater between the catfish genome and the zebrafish genome than between the catfish genome and the *Tetraodon* genome. Of the 34 conserved syntenies, 30 were present between the catfish and zebrafish genomes. Of the 34 conserved syntenies among the species, 14 were present in *Tetraodon* and catfish; 15 were absent in *Tetraodon*; and five were unknown because one or both of the two genes involved in the paired BAC ends were not yet designated to specific chromosomes in *Tetraodon*. Although the exact distances between the paired genes sequenced from mate paired BAC ends were unknown, the average insert size of the catfish CHORI 212 BAC library is 161 kb. In most cases, the distances between the sets of two genes were larger in zebrafish than in *Tetraodon* (Table 4), consistent with the more compact genome of *Tetraodon*.

The syntenic conservation appeared to be low. Part of this low syntenic conservation could have been resulted from random gene loss in the genomes of teleosts that are widely believed to be 3R duplication polyploid lineages. However, even considering total loss of genes (thereby their homologous sequences in the genome), the observed rate of syntenic conservation is still quite low. The predicted rate of syntenic conservation of linked-pair of genes from catfish should be 50% in zebrafish, while the observed syntenic conservation is only 24.1%, suggesting that chromosome fragments are continuously being reshuffled and that this process is more obvious among the less related taxa. A greater level of syntenic conservation was observed between the catfish and zebrafish genomes than between the catfish and *Tetraodon* genomes, consistent with the phylogenetic relationships among the three fish species. Our arbitrary limitation of the 1.2 Mb distance as the cutoff for the linked genes could exclude those conserved syntenies with an inversion involving longer sequence intervals. In addition, the current state of the genome assembly in *Tetraodon* and zebrafish is incomplete, and many genes assigned to unknown chromosomes could reside in close proximity of the known chromosomes harboring some of the genes under study. Nonetheless, only 34 out of 141 (24.1%) gene pairs from catfish had their counterparts arranged in proximity in the zebrafish genome using one mega base pair as the cutoff, which was surprisingly low considering how closely related the catfish and zebrafish are. This finding suggested

Table 3. Mapping of genes to BACs through BAC end sequencing as assessed by BLASTX searches

p-value	*Number of hits*	*Alignment length (amino acids)*	*Average alignment length (amino acids)*	*% Identity*
$<10^{-50}$	58	101–228	167	48–99
10^{-40}–10^{-50}	54	81–207	134	43–97
10^{-30}–10^{-40}	77	66–217	103	40–100
10^{-20}–10^{-30}	253	45–175	75	34–100
10^{-15}–10^{-20}	275	37–199	62	30–100
10^{-10}–10^{-15}	413	30–186	54	31–100
Subtotal	1,130	30–228	73	31–100
10^{-5}–10^{-10}	747	19–193	47	23–100
Total	1,877	19–228	63	23–100

Listed are number of BLASTX hits of genes by BAC end sequences, excluding redundant hits.
The *p*-values, alignment length range, average alignment length, and percentage of identities are provided as indications of the levels of similarities.

Table 4. Summary of conserved syntenies identified by comparison of 141 mate paired genes of channel catfish with genomic locations of those within the *Danio rerio* and *Tetraodon nigroviridis* genomes

			Zebrafish		*Tetraodon*	
Catfish BAC	*Sp6 hits*	*T7 hits*	*Chr*	*Distance*	*Chr*	*distance*
001_L07	CAG01022.1	CAG01025.1			2	130 kb
003_J07	XP_691151.1	Q96Q40	6	920 kb	3	1087 kb
003_H12	AAH65969.1	NP_056346.2	15	380 kb		
004_F18	CAG06494.1	AAG37030.1	15	220 kb		
005_D14	AAP48571.1	CAF89961.1	11	400 kb		
006_P17	CAE51056.1	XP_698666.1	6	180 kb		
007_K16	XP_544816.1	CAG08989.1	15	500 kb	16	100 kb
007_M06	CAG30482.1	CAG08540.1	5	400 kb		
008_I11	NP_001019337.1	NP_001007763.1	14	Cadherin cluster	7	Cadherin cluster
008_I14	XP_545610.2	XP_545544.2	9	710k		
013_P16	XP_691920.1	XP_690925.1	18	740 kb	13	74 kb
014_D16	XP_428910.1	XP_698664.1			Un/19	
018_H11	AAH56818.1	AAH44562.1	13	280 kb	Un/17[a]	
018_I19	CAF99682.1	CAF99686.1	7	160 kb	5	53 kb
020_H13	AAH85663.1	CAF92624.1			3/Un	
020_I23	BAD90503.1	CAF96508.1	9	3 mb	11	91 kb
020_L17	XP_690830.1	AAX46593.1	5	20 kb	Un/12	
020_P11	CAF90170.1	CAF94367.1	6	340 kb		
021_L13	CAF95098.1	XP_688911.1			Un/Un	
022_D09	XP_685117.1	P21359	15	376 kb	16/Un[a]	
022_G21	NP_072140.1	XP_535054.2	17	580 kb		
023_O10	XP_693773.1	XP_685853.1	24	280 kb	6/Un	
025_I21	XP_706772.1	XP_684635.1	11	660 kb		
026_B05	AAQ83456.1	AAB52701.1	3	10 kb		
026_C08	XP_685862.1	NP_956611.1	13	210 kb		
028_J11	AAW38963.1	AAC64076.1	17	Pheromone receptor cluster	16	Pheromone receptor cluster
028_M04	XP_687685.1	XP_695804.1	5	150 kb	16	206 kb
029_C08	CAG14347.1	XP_690272.1	7	310 kb		
029_M13	CAG04452.1	CAG04458.1	12	15 mb	2	158 kb
031_O02	AAL66362.1	XP_696942.1	24	8 mb	6	309 kb
032_F20	XP_691291.1	AAH78367.1	7	270 kb		
033_A11	XP_693134.1	AAH97450.1	17	280 kb	10	57 kb
033_C15	CAF87798.1	XP_686123.1	9	660 kb		
035_O13	XP_683954.1	AAC64076.1	18	30 kb		

The putative identities of the mate paired genes are provided as GenBank accession numbers of their top BLASTX hits. Sp6 hits indicate the gene identities of the BES using the Sp6 sequencing primer, and likewise, T7 hits using the T7 sequencing primer. Chr indicates the chromosome on which the genes are located, and distance indicates the distance found between the two genes in zebrafish or *Tetraodon* as appropriate. Shaded rows are syntenies conserved among all three species of catfish, zebrafish, and *Tetraodon*. For "cadherin cluster" and "pheromone receptor cluster," the distance could not be determined because these genes are arranged in tandem as gene clusters. [a]Indicates partial conserved syntenies. Empty positions are for the lack of conserved syntenies.

that rearrangements could be extensive between the zebrafish and the catfish genomes, posing challenges for genome analysis using comparative approaches. As previously noted, species-specific gene evolution is a widely observed phenomenon in teleost genomes (Peatman and Liu 2006, 2007; Steinke et al. 2006). This is in great contrast to the situation among mammals where high levels of sequence conservation and macro-syntenic regions have been observed and virtual physical and comparative maps can be constructed through sequence similarity comparisons (Larkin et al. 2003). BLASTN searches allowed 29.4% of Cattle BACs to be anchored to the human genome sequences. Of the 1,242 that had both ends matching human genome sequence, 1,011 (81.4%) had ends <300 kb apart on the same human chromosome (Larkin et al. 2003). If this assessment proves to be true, large-scale comparative genome mapping in catfish using zebrafish and *Tetraodon* genome resources would be useful, but caution must be exercised dealing with large genome segments at the chromosome level.

Difficulties in comparative genome analysis, however, could be overcome if genome sequence of a closely related species becomes available. For instance, in a recent study, Stemshorn et al. (2005) were able to anchor the genetic map of *Cottus gobio* to the physical map of *Tetraodon* through similarity comparisons using microsatellite flanking sequences. The key question remains as to how closely

two species have to be related for effective comparative genome analysis, and answers to such a question are of great interest to many biologists working on non-model species.

Validation of Conserved Syntenies. To evaluate the extent of the conserved syntenies, we attempted to validate the observed microsyntenies by additional experiments. Three BACs with paired genes on their ends, 018_H11, 022_D09, and 028_M04, were further evaluated by direct BAC sequencing. As shown in Figure 2, the orders of the genes located between the two genes of the mate paired BAC ends were searched in zebrafish and *Tetraodon*. The catfish EST database was searched for the presence of corresponding genes between the two genes identified on both ends of the BACs. When

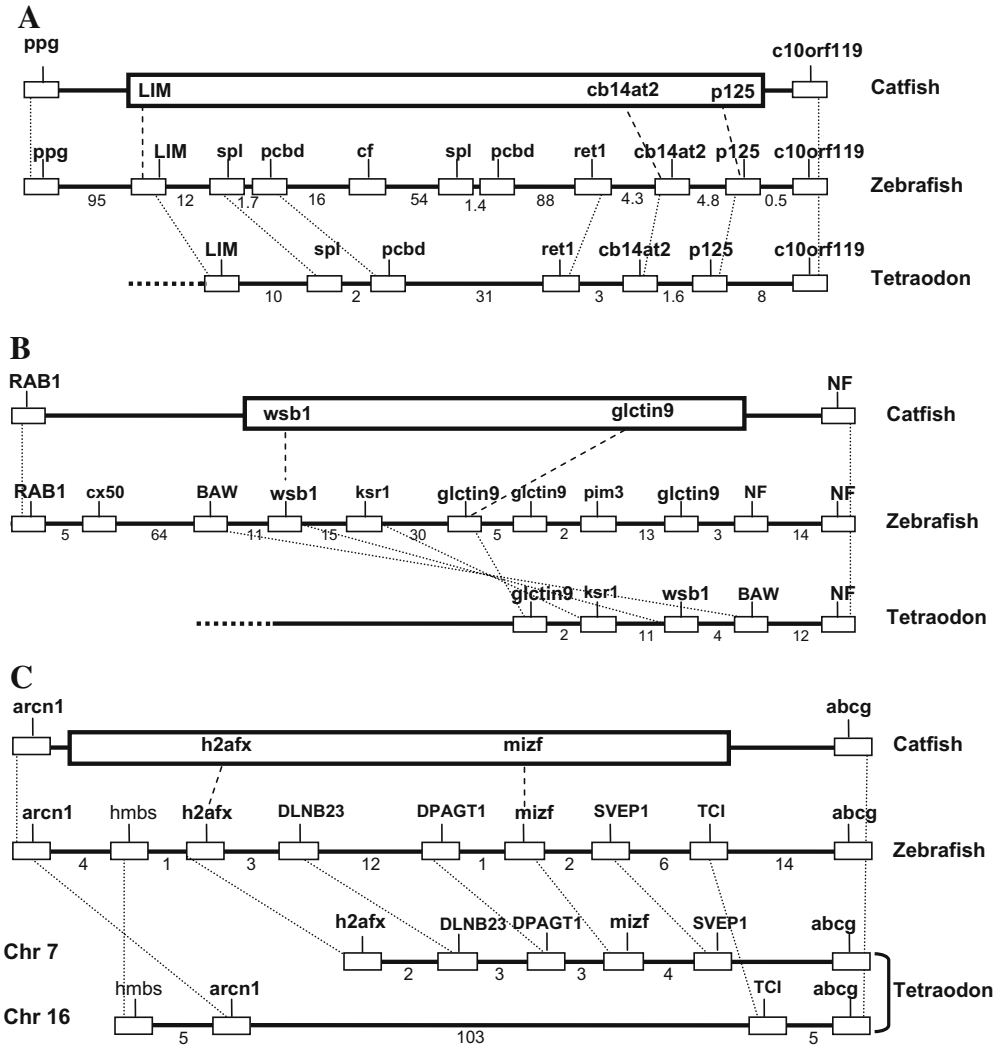

Figure 2. Examples of conserved syntenies extended from the mate paired genes from both ends of the channel catfish BAC end sequences (BES) by comparative analysis of the genes in the catfish genome with those from the zebrafish and *Tetraodon* genomes. Exact gene order, distance, and orientation of the catfish genes internal to the mate paired genes were not determined. Three syntenies are shown from BAC 018_H11 (A), 022_D09 (B), and 028_M04 (C). Abbreviations, in A: ppg, Poly(A) polymerase ã; LIM, LIM protein; spl, sphingosine phosphate lyase 1; pcbd, 6-pyruvoyl-tetrahydropterin synthase/dimerization cofactor of hepatocyte nuclear; cf, cathepsin F; ret1, receptor tyrosine kinase; cb14at2, chondroitin â1,4 *N*-acetylgalactosaminyltransferase 2; p125, SEC23 interacting protein; c10orf119, chromosome 10 open reading frame 119 (*H. sapiens*). B: RAB1, RAB1; CX50, gap junction α-8 protein (lens fiber protein MP70); BAW, BAW protein; wsb1, SOCS box-containing WD protein SWiP-1; ksr1, similar to kinase suppressor of ras-1; glctin9, lectin, galactoside-binding, soluble, 9; pim, similar to serine/threonine-protein kinase; NF, similar to neurofibromatosis type 1. C: arcn1, archain 1; hmbs, hydroxymethylbilane synthase; h2afx, H2A histone family, member X; DLNB23, transmembrane protein 24; DPAGT1, dolichyl-phosphate (UDP-N-acetylglucosamine) *N*-acetylglucosaminephosphotransferase 1 (eGlcNAc-1-P transferase); mizf, MBD2 (methyl-CpG-binding protein)-interacting zinc finger protein; SVEP1, sushi, von Willebrand factor type A, EGF and pentraxin domain containing 1; TC1, transcobalamin I precursor; abcg, ATP-binding cassette, subfamily G, member 4.

present, sequencing primers were designed based on the catfish EST sequences and used to directly sequence the relevant catfish BAC clone. The generated sequences were then analyzed by BLASTX searches or by sequence alignment with the ESTs. As shown in Figure 2, the syntenies were well conserved though the order of the genes in catfish was not determined. For BAC 018_H11, the genes on the left and right ends were poly(A) polymerase γ and the chromosome ORF119 gene; three sequencing primers designed for LIM, cb14at2 (chondroitin β1,4 *N*-acetylgalactosaminyltransferase 2), and p125 (SEC23 interacting protein p125) all generated correct sequences by direct BAC sequencing, confirming the presence of these genes within the BAC clone. Similarly, four genes [*RAB1*, *wsb1* (SOCS box-containing WD protein SWiP-1), *Galectin 9*, and neurofibromatosis type 1] were confirmed to be present within the BAC clone 022_D09; and four genes [*archain 1*, *h2afx* (H2A histone family member X), *mizf* (methyl-CpG-binding protein-interacting zinc finger protein), and *abcg* (ATP-binding cassette, subfamily G, member 4)] were confirmed to be present in the BAC clone 028_M04, by direct BAC sequencing. A brief examination of the conserved syntenies also suggested a higher level of genome conservation between the catfish and zebrafish genomes than between the catfish and the *Tetraodon* genomes. In the second and third conserved syntenies as shown in Figure 2, more gene rearrangements were detected in the *Tetraodon* genome as compared to the zebrafish and the catfish genomes. Because of the lack of the sequence information for the design of sequence or polymerase chain reaction (PCR) primers, many genes within the conserved syntenies were not confirmed in catfish, but the demonstrated extension of conserved syntenies suggests a high level of genome conservation. This is in good contrast to the situation of larger genome environs, as discussed in the preceding text. On the microsyntenic scale, gene order and organization appeared to be highly conserved among fishes. This suggested that once the genomic region is located using genetic linkage mapping, comparative mapping should be a powerful tool for the identification of candidate genes.

Acknowledgments

This project was supported by a grant from USDA NRI Animal Genome Basic Genome Reagents and Tools Program (USDA/NRICGP award no. 2006-35616-16685). We appreciate the support of Alabama Agricultural Experiment Station (AAES) for conducting part of the BAC end sequencing.

References

Bao B, Peatman E, Li P, He C, Liu Z (2005) Catfish hepcidin gene is expressed in a wide range of tissues and exhibits tissue-specific upregulation after bacterial infection. Dev Comp Immunol 29, 939–950

Buitkamp J, Kollers S, Durstewitz G, Welzel K, Schafer K, Kellermann A, Lehrach H, Fries R (2000) Construction and characterization of a gridded cattle BAC library. Anim Genet 31, 347–351

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

Crooijmans RP, Vrebalov J, Dijkhof RJ, van der Poel JJ, Groenen MA (2000) Two-dimensional screening of the Wageningen chicken BAC library. Mamm Genome 11, 360–363

Cunningham C, Hikima J, Jenny MJ, Chapman RW, Fang GC, Saski C, Lundqvist ML, Wing RA, Cupit PM, Gross PS, Warr GW, Tomkins JP (2006) New Resources for Marine Genomics: Bacterial Artificial Chromosome Libraries for the Eastern and Pacific Oysters (*Crassostrea virginica* and *C. gigas*). Mar Biotechnol 8, 521–533

Eggen A, Gautier M, Billaut A, Petit E, Hayes H, Laurent P, Urban C, Pfister-Genskow M, Eilertsen K, Bishop MD (2001) Construction and characterization of a bovine BAC library with four genome-equivalent coverage. Genet Sel Evol 33, 543–548

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8, 186–194

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8, 175–185

Fahrenkrug SC, Rohrer GA, Freking BA, Smith TP, Osoegawa K, Shu CL, Catanese JJ, de Jong PJ (2001) A porcine BAC library with tenfold genome coverage: a resource for physical and genetic map integration. Mamm Genome 12, 472–474

He C, Chen L, Simmons M, Li P, Kim S, Liu ZJ (2003) Putative SNP discovery in interspecific hybrids of catfish by comparative EST analysis. Anim Genet 34, 445–448

Ju Z, Karsi A, Kocabas A, Patterson A, Li P, Cao D, Dunham R, Liu Z (2000) Transcriptome analysis of channel catfish (*Ictalurus punctatus*): genes and expression profile from the brain. Gene 261, 373–382

Ju Z, Dunham RA, Liu Z (2002) Differential gene expression in the brain of channel catfish (*Ictalurus punctatus*) in response to cold acclimation. Mol Genet Genomics 268, 87–95

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

Katagiri T, Asakawa S, Hirono I, Aoki T, Shimizu N (2000) Genomic bacterial artificial chromosome library of the Japanese flounder *Paralichthys olivaceus*. Mar Biotechnol 2, 571–576

Katagiri T, Asakawa S, Minagawa S, Shimizu N, Hirono I, Aoki T (2001) Construction and characterization of BAC libraries for three fish species; rainbow trout, carp and tilapia. Anim Genet 32, 200–204

Katagiri T, Kidd C, Tomasino E, Davis JT, Wishon C, Stern JE, Carleton KL, Howe AE, Kocher TD (2005) A BAC-based physical map of the Nile tilapia genome. BMC Genomics 6, 89

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

Larkin DM, Everts-van der Wind A, Rebeiz M, Schweitzer PA, Bachman S, Green C, Wright CL, Campos EJ, Benson LD, Edwards J, Liu L, Osoegawa K, Womack JE, de Jong PJ, Lewin HA (2003) A cattle-human comparative map built with cattle BAC-ends and human genome sequence. Genome Res 13, 1966–1972

Li P, Peatman E, Wang S, He C, Baoprasertkul P, Xu P, Kucuktas H, Nandi S, Somridhivej B, Serapion J, Simmons M, Liu L, Muir W, Brady Y, Grizzle J, Dunham R, Liu Z (2007) Towards the catfish transcriptome: development of molecular tools from 31,215 catfish ESTs. BMC Genomics 8, 177

Li RW, Waldbieser GC (2006) Production and utilization of a high-density oligonucleotide microarray in channel catfish, *Ictalurus punctatus*. BMC Genomics 7, 134

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

Liu Z (2006) Transcriptome characterization through the generation and analysis of expressed sequence tags: factors to consider for a successful EST project. Israel J Aquaculture-Bamidgeh 58, 328–341

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

Liu Z, Nichols A, Li P, Dunham RA (1998b) Inheritance and usefulness of AFLP markers in channel catfish (*Ictalurus punctatus*), blue catfish (*I. furcatus*), and their F1, F2, and backcross hybrids. Mol Gen Genet 258, 260–268

Liu Z, Karsi A, Dunham RA (1999) Development of polymorphic EST markers suitable for genetic linkage mapping of catfish. Mar Biotechnol 1, 437–447

Liu Z, Karsi A, Li P, Cao D, 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

Nandi S, Peatman E, Xu P, Wang S, Li P, Liu Z (2007) Repeat structure of the catfish genome: a genomic and transcriptomic assessment of Tc1-like transposon elements in channel catfish (*Ictalurus punctatus*). Genetica (in press)

Ng SH, Artieri CG, Bosdet IE, Chiu R, Danzmann RG, Davidson WS, Ferguson MM, Fjell CD, Hoyheim B, Jones SJ, de Jong PJ, Koop BF, Krzywinski MI, Lubieniecki K, Marra MA, Mitchell LA, Mathewson C, Osoegawa K, Parisotto SE, Phillips RB, Rise ML, von Schalburg KR, Schein JE, Shin H, Siddiqui A, Thorsen J, Wye N, Yang G, Zhu B (2005) A physical map of the genome of Atlantic salmon, *Salmo salar*. Genomics 86, 396–404

Nonneman D, Waldbieser GC (2005) Isolation and enrichment of abundant microsatellites from a channel catfish (*Ictalurus punctatus*) brain cDNA library. Anim Biotechnol 16, 103–116

Osoegawa K, Woon PY, Zhao B, Frengen E, Tateno M, Catanese JJ, de Jong PJ (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 52, 1–8

Peatman E, Liu Z (2006) CC chemokines in zebrafish: evidence for extensive intrachromosomal gene duplications. Genomics 88, 381–385

Peatman E, Liu Z (2007) Evolution of CC chemokines in teleost fish: a case study in gene duplication and implications for immune diversity. Immunogenetics 59, in press

Peatman E, Baoprasertkul P, Terhune J, Xu P, Nandi S, Kucuktas H, Li P, Wang S, Somridhivej B, Dunham R, Liu Z (2007) Microarray-base gene profiling of the acute phase response in channel catfish (*Ictalurus punctatus*) after infection with a Gram negative bacterium. Dev Comp Immunol, in press

Quiniou SM, Katagiri T, Miller NW, Wilson M, Wolters WR, Waldbieser GC (2003) Construction and characterization of a BAC library from a gynogenetic channel catfish *Ictalurus punctatus*. Genet Sel Evol 35, 673–683

Rodriguez MF, Gahr SA, Rexroad CE 3rd, Palti Y (2006) A polymerase chain reaction screening method for rapid detection of microsatellites in bacterial artificial chromosomes. Mar Biotechnol 8, 346–350

Serapion J, Kucuktas H, Feng J, Liu Z (2004a) Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). Mar Biotechnol 6, 364–377

Serapion J, Waldbieser GC, Wolters W, Liu ZJ (2004b) Development of type I markers in channel catfish through intron sequencing. Anim Genet 35, 463–466

Steinke D, Salzburger W, Braasch I, Meyer A (2006) Many genes in fish have species-specific asymmetric rates of molecular evolution. BMC Genomics 7, 20

Stemshorn KC, Nolte AW, Tautz D (2005) A genetic map of *Cottus gobio* (Pisces, Teleostei) based on microsatellites can be linked to the physical map of Tetraodon nigroviridis. J Evol Biol 18, 1619–1624

Thorsen J, Zhu B, Frengen E, Osoegawa K, de Jong PJ, Koop BF, Davidson WS, Hoyheim B (2005) A highly redundant BAC library of Atlantic salmon (*Salmo salar*): an important tool for salmon projects. BMC Genomics 6, 50

Waldbieser GC, Bosworth BG, Nonneman DJ, Wolters WR (2001) A microsatellite-based genetic linkage map for channel catfish, *Ictalurus punctatus*. Genetics 158, 727–734

Waldbieser GC, Quiniou SMA, Karsi A (2003) Rapid development of gene-tagged microsatellite markers from bacterial artificial chromosome clones using anchored TAA repeat primers. Biotechniques 35, 976–979

Whitaker HA, McAndrew BJ, Taggart JB (2006) Construction and characterization of a BAC library for the European sea bass *Dicentrarchus labrax*. Anim Genet 37, 526

Xu P, Bao B, He Q, Peatman E, He C, Liu Z (2005) Characterization and expression analysis of bactericidal permeability-increasing protein (BPI) antimicrobial peptide gene from channel catfish *Ictalurus punctatus*. Dev Comp Immunol 29, 865–878

Xu P, Wang S, Liu L, Peatman E, Somridhivej B, Thimmapuram J, Gong G, Liu Z (2006) Channel catfish BAC-end sequences for marker development and assessment of syntenic conservation with other fish species. Anim Genet 37, 321–326

Zhu B, Smith JA, Tracey SM, Konfortov BA, Welzel K, Schalkwyk LC, Lehrach H, Kollers S, Masabanda J, Buitkamp J, Fries R, Williams JL, Miller JR (1999) A 5x genome coverage bovine BAC library: production, characterization, and distribution. Mamm Genome 10, 706–709

Zimmer R, Verrinder Gibbins AM (1997) Construction and characterization of a large-fragment chicken bacterial artificial chromosome library. Genomics 42, 217–226