Genetica (2007) 131:81–90
DOI 10.1007/s10709-006-9115-4

ORIGINAL RESEARCH

Repeat structure of the catfish genome: a genomic and transcriptomic assessment of *Tc1*-like transposon elements in channel catfish (*Ictalurus punctatus*)

Samiran Nandi · Eric Peatman · Peng Xu · Shaolin Wang · Ping Li · Zhanjiang Liu

Received: 9 December 2005 / Accepted: 2 October 2006 / Published online: 8 November 2006

Abstract We have assessed the distribution and diversity of members of the *Tc1/mariner* superfamily of transposable elements in the channel catfish (*Ictalurus punctatus*) genome as well as evaluating the extent of transcription of *Tc1* transposases in the species. Through use of PCR amplification and sequencing, assessment of random BAC end sequences (BES) equivalent to 1.2% genome coverage, and screening of over 45,000 catfish ESTs, a significant proportion of *Tc1*-like elements and their associated transcripts were captured. Up to 4.2% of the catfish genome in base pairs appears to be composed of *Tc1*-like transposon-related sequences and a significant fraction of the catfish cellular mRNA, approximately 0.6%, was transcribed from transposon-related sequences in both sense and antisense orientations. Based on results of repeat-masking, as much as 10% of BAC end sequences from catfish, which is a random survey of the genome, contain some remnant of *Tc1* elements, suggesting that these elements are present in the catfish genome as numerous, small remnants of the transposons. Phylogenetic analysis allowed comparison of catfish *Tc1* transposase types with those found in other vertebrate and invertebrate species. In spite of the existence of many types of *Tc1*-like sequences that are not yet able to be placed in clades with strong statistical support, it is clear that multiple families of *Tc1*-like elements exist in channel catfish.

Keywords Fish · Catfish · Genome · *Tc1* transposon · Repeat · Transposase · Evolution

Abbreviations

BES	BAC end sequences
Bp	base pair
dNTP	Deoxynucleotide triphosphate
E	Expectation value
EST	Expressed sequence tag
IR	Inverted repeat
ORF	Open reading frame
PCR	Polymerase chain reaction

Introduction

Transposable elements can be divided into retrotransposons and DNA transposons based on their mode of transposition (Finnegan 1989). DNA transposons harbor short terminal inverted repeats (IR) required *in* cis, and encode transposases required *in* trans for transposition. *Tc1*-like elements belong to a *Tc1/mariner* superfamily of DNA transposons that use a cut and paste mechanism with the non-transferred strand being initially cleaved (Dawson and Finnegan, 2003).

Since the discovery of the first DNA transposons in vertebrates (Henikoff 1992), *Tc1*-like elements have been found in a wide range of vertebrate animals

S. Nandi · E. Peatman · P. Xu · S. Wang · P. Li · Z. Liu (✉)
Department of Fisheries and Allied Aquacultures, The Fish Molecular Genetics and Biotechnology Laboratory, Program of Cell and Molecular Biosciences, Auburn University, Auburn, AL 36849, USA
e-mail: zliu@acesag.auburn.edu

S. Nandi
Central Institute of Freshwater Aquaculture, ICAR, Bhubaneswar, India

including poikilotherm vertebrates such as actinopterygians and lissamphibians (Heierhorst et al. 1992; Goodier and Davidson 1994; Radice et al. 1994; Avancini et al. 1996; Lam et al. 1996; Gottgens et al. 1999; Leaver 2001; Miskey et al. 2003; Krasnov et al. 2005; Sinzelle et al. 2005). However, until recently *Tc1*-like elements were all identified as non-autonomous from vertebrate genomes, resulting from extensive insertions/deletions, and in-frame termination codons in their transposase genes (Lam et al. 1996a; Ivics et al. 1996, 1997). Kawakami and his colleagues identified the first endogenous autonomous transposon, an *Ac*-like *Tol2* element from the Japanese medaka fish that can transpose both transiently in fertilized zebrafish eggs and in the zebrafish germ lineage (Kawakami et al. 2000; Koga and Hori 2000, 2001). Several additional transposases with intact open reading frames (ORFs) were recently identified from the amphibian *Xenopus tropicalis* and may still be involved in active transposition (Sinzelle et al. 2005).

In contrast to *P* elements (Castro and Carareto 2004), *Tc1*-like transposases appear to be less demanding on cellular factors for transposition, (Loukeris et al. 1995; Plasterk 1996; Gueiros-Filho and Beverley 1997) and, therefore, are expected to function in a wide host range. This feature of *Tc1* elements favors their practical application as genetic tools. Several reconstructed transposases, such as in *Sleeping Beauty* and *Frog Prince*, were demonstrated to function as transposases in vertebrates (e.g. Ivics et al. 1997; Fadool et al. 1998; Davidson et al. 2003; Miskey et al. 2003; Dupuy et al. 2005, for reviews, see Ivics and Izsvak 2004; Ivics et al. 2004; Wadman et al. 2005; Miskey et al. 2005).

Effective genome research demands an understanding of the repeat structure of the organism under study. Genome-based research has been underway in channel catfish (*Ictalurus punctatus*), the primary aquaculture species in the United States, for almost a decade. However, the repeat structures of the channel catfish genome have not been well studied, particularly the interspersed types of repeats. Ongoing projects in the areas of linkage and physical mapping (Liu et al. 2003) and genomic and expressed sequence tag (EST) sequencing (Cao et al. 2001; Kocabas et al. 2002) will benefit from an assessment of the prevalent *Tc1* transposable elements in catfish, allowing for accurate processing, annotation, and assessment of sequences and accurate assembly of genetic maps. Therefore, here we assessed the distribution and diversity of *Tc1* elements in the channel catfish genome as well as evaluating the extent of transcription of *Tc1* transposases. We also conducted phylogenetic analysis in order to compare catfish *Tc1* transposase types with those found in other vertebrate and invertebrate species. Here we report that: a significant proportion, up to 4.2% of the catfish genome in base pairs, appears to consist of *Tc1*-like transposon-related sequences; a significant fraction of the catfish cellular mRNA, approximately 0.6%, was transcribed from transposon-related sequences in both sense and antisense orientations; and multiple families of *Tc1*-like elements are present in channel catfish.

Methods and materials

Amplification of a subset of *Tc1*-like elements using the *Tss* IR primer

Genomic DNA of channel catfish was isolated as previously described (Liu et al. 1998a). Blood was expelled into lysis buffer quickly to disperse the blood cells. The cell lysates were incubated at 55°C overnight. DNA was extracted twice with phenol and once with chloroform. DNA was precipitated by adding half the original blood volume of 7.5 M ammonium acetate and two volumes of ethanol. DNA was collected by brief centrifugation and washed twice with 70% ethanol, air-dried, and dissolved in water. The concentration was measured with a spectrophotometer for absorption at 260 nm or estimated from intensities of ethidium bromide staining after agarose gel electrophoresis using a 1 kb ladder DNA as reference (Invitrogen, Carlsbad, California).

In order to characterize a subset of *Tc1* elements in its entire length, PCR reactions were conducted using the *Tss* IR as the primer. PCR reactions of 50 μl were carried out in 50 mM KCl, 10 mM Tris (pH 9.0 at 25 °C), 0.1% Triton X-100, 0.25 mM each of deoxynucleotide triphosphate (dNTPs), 1.5 mM $MgCl_2$, 20 μM of the primers, about 500 ng channel catfish genomic DNA, and 2.5 units of *Taq* DNA polymerase. The primer used in the PCR represents an IR sequence thus only one primer was used for PCR. Its sequence reads 5'-TACAGTTGAAGTCGGAAGTTTACATAC-3'. The temperature profiles used in amplification of *Tc1*-elements were 94 °C for 30 s, 55 °C for 1 min, and 72 °C for 2 min, for 35 cycles. PCR products were cloned into *pCR2.1* (Invitrogen, CA). Ligations and bacterial transformations were performed using the standard procedures (Sambrook et al. 1989). Both the mixture of the PCR products and the gel-purified fragments were used in separate ligation reactions. Mini-scale plasmid preparations were utilized to screen for the clones with PCR product inserts. Clones with expected sizes were picked for sequencing analysis.

Sequence analysis

All DNA clones were sequenced by the Sanger dideoxy termination method using the cycle sequencing kit (Applied Biosystems, Inc., CA). Minipreparation plasmid DNA (1 μl, about 200–500 ng) was used for each reaction. The profiles for cycling were: 94 °C for 1 min, 72 °C for 1 min, 55 °C for 1 min for 30 cycles. An initial 2 min of extra denaturation at 94 °C was used. Initial sequencing was accomplished with both the universal and the reverse sequencing primers. After initial sequencing reactions, overlapping sequencing was finished by the primer-walking method. Sequencing primers were ordered from Sigma Genosys (Woodlands, Texas).

Dot blot hybridization

Known amounts of genomic DNA and plasmid DNA containing the cloned *Tc1*-elements *Tip1* or *Tip2* were first serially diluted. Genomic DNA and isolated fragments of *Tc1*-elements were denatured by mixing with an equal volume of 1 M sodium hydroxide and kept at room temperature for 5 min. An equal volume of 2 M ammonium acetate was added to neutralize the samples. The samples were then loaded onto a nylon membrane by using a dot blot apparatus (BRL, Bethesda, MD). One milliliter of 5 × SSC was then added to each sample spot. The filter was removed and air dried before baking at 80 °C for 2 h prior to hybridization as described (Liu et al. 1998b). The hybridization signals were quantified by using the GS-525 Molecular Imaging System (Bio-Rad, California).

Bioinformatic and phylogenetic analysis

Channel catfish BAC end sequences (BES), after trimming of vector sequences and filtering of the bacterial sequences, were processed through the Repeatmasker software (http://www.repeatmasker.org). Repeats were masked by using the zebrafish repeat database, allowing an assessment of the abundance of *Tc1* transposable elements within the catfish BES. In a second approach, batch BLASTX searches with 20,366 BES were conducted against the non-redundant protein database using a server at the University of Illinois. An expectation value (E) of e^{-5} was used as the significance similarity threshold for the comparison. The BLASTX result was parsed into a tab-delimited format and imported into an Excel spreadsheet. The sheet was sorted according to protein hit, and those BES with a transposase top hit were selected for further analysis.

To analyze *Tc1* transposase abundance and diversity at the level of transcription, 45,457 channel catfish ESTs were subjected to BLASTX searches using a server at the University of Illinois and a cutoff value of e^{-5}. Top protein hits and e-values for each EST were recorded and used to sort the ESTs into 28 groups.

Phylogenetic trees were drawn from ClustalW (Thompson et al. 1994) generated multiple sequence alignments of amino acid sequences using the neighbor-joining method (Saitou and Nei 1987) or the minimum evolution method within the Molecular Evolutionary Genetics Analysis [MEGA (3.0)] package (Kumar et al. 2004), and in some cases, the tree was also checked with maximum likelihood algorithm. Data were analyzed using Poisson correction and gaps were removed by complete deletion. The topological stability of the neighbor joining trees was evaluated by 1,000 bootstrapping replications. First, the catfish EST with the highest e-value for each of the 28 top BLASTX hits was translated, introducing judicious frameshifts where necessary. In several cases, the transposase remnants encoded by the ESTs were too short to allow their placement on the phylogenetic tree. Therefore, only 23 of the 28 EST ''types'' were used on the initial tree. In addition, we combined these translated EST sequences with four additional *Tc1* transposases from BAC end sequences from channel catfish. A representative sequence was chosen from each of the seven clades in the resulting catfish tree. These sequences were aligned with *Tc1* transposases from other vertebrate and invertebrate species and a second tree constructed. Comparison of neighbor-joining and minimum evolution trees derived from this alignment showed that the minimum evolution tree had higher bootstrapping support and more closely replicated previously established phylogenetic relationships between the sequences.

For the identification of any additional *Tc1*-related repeats in the catfish genome, all 20,366 BES of catfish were collected and formatted to the nucleotide database. self-BLASTN searches were conducted against themselves. The cut off P-value was set at 10^{-2} and the identify threshold was set to 90%. The BLASTN results were parsed to tab-delimited format to count the redundant queries and other statistics.

Results

Characteristics of a subset of *Tc1*-like elements similar to *Tss*

To characterize a subset of *Tc1*-like elements from channel catfish similar to *Tss*, genomic DNA was

amplified using a single 27-base pair (bp) PCR primer of IR designed from transposon *Tss* (Izsvak et al. 1995). Six bands were produced with sizes of 1.6, 1.3, 1.0, 0.8, 0.6, and 0.5 kb (Fig. 1), suggesting the presence of *Tc1*-like elements with various sizes in the catfish genome. Of the six major bands, the 0.8 and 0.6 kb fragments were not consistently amplified, and the band intensity of the 1.3 kb also varied greatly depending on experiments (Fig. 1), suggesting either the PCR primer did not have a perfect match to the IR, or these bands could have represented heteroduplexes of the PCR products.

With multiple attempts, three consistently amplified bands, the 0.5, 1.0, and 1.6 kb, were cloned. The 0.5 kb product represented a highly abundant (32,000 copies) non-autonomous *Tc1*-like element that was previously characterized (Liu et al. 1999). Attempts were made to clone the 1.3, 0.8, and 0.6 kb PCR products, but were not successful with repeated experiments, suggesting that they could be heteroduplexes formed between different PCR products. Sequence analysis indicated that both the 1.6 and the 1.0 kb elements belong to the *Tc1/mariner* transposon superfamily. Following the common nomenclature of this family, they will be referred to herein as *Tip1* and *Tip2* (for *Tc1*-like elements from *I. punctatus*), for the 1,568 and 1,009 bp elements, respectively. Their complete sequences have been deposited to the GenBank with Accession numbers of DQ318916 and DQ318917.

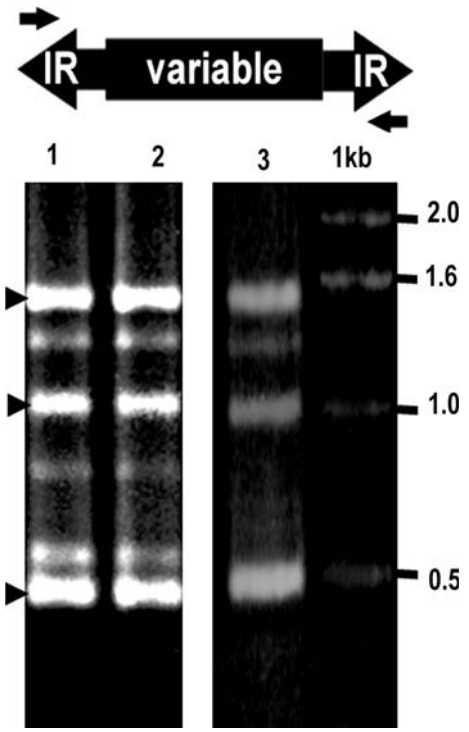

Fig. 1 Mutiple sizes of *Tc1*-like elements from channel catfish as demonstrated by profiles of PCR products using a primer derived from *Tdr1* IR. Lanes 1, 2, and 3 represent profiles of PCR products from different experiments. 1 kb, 1 kb ladder molecular markers with sizes being marked on the right in kilobases. Arrows on the left indicate the cloned bands

The organization of *Tip1* and *Tip2* are similar with the IR and a defective gene homologous to *Tc1*-like transposase. However, lengths and sequence of the IR are drastically different between them. *Tip1* harbors a short IR of 28 bp, essentially with no additional sequences included in the IR beyond the primer sequence. Given the PCR primer used for its amplification, the IR for *Tip1* may differ from the primer sequence. Nonetheless, the *Tip1* sequence is highly similar (Expected value = $2e^{-41}$) to *Tc1*-like element from *Pleuronectes platessa* (Accession number CAB51372). *Tip2* harbors an IR of 176 bp with minor deletions and base substitutions (Fig. 2). *Tip2* was most similar to a *Tc1*-element from *Rana pipiens* (Accession number AAP49009) with an expected value of $1e^{-16}$.

Abundance of *Tip1* and *Tip2* in channel catfish

Tip2 is more abundant than *Tip1*, as determined by dot blot analysis (Fig. 3). To determine the abundance of *Tip1* and *Tip2* in channel catfish, serial dilutions of channel catfish genomic DNA were blotted. The plasmids that harbor respective fragments of the *Tip1* and *Tip2* were used as controls and the corresponding internal segments of inserts were used as probes for dot blots. The radioactivity in each spot on the filter was quantified to determine copy numbers of these elements in the channel catfish genome. *Tip1* exists in the channel catfish genome at approximately 150 copies per haploid genome, and *Tip2* exists at approximately 4,000 copies per haploid genome. In both cases, these elements are less abundant than the previously characterized non-autonomous element *Tipnon*, 32,000 copies of which exist in the channel catfish genome (Liu et al. 1999).

Assessment of *Tc1*-like elements in the channel catfish genome by BAC-end sequencing

The above analysis using PCR was limited to *Tc1*-like transposons that harbor the same IR as the primers used for amplification. Although various IR can be used for PCR amplification of different types of *Tc1*-elements from catfish, it is always limited to the primers used. In addition, given that the vast majority of *Tc1*-elements are likely dispersed in the genome in various forms that may or may not harbor IR sequences, a genomic survey may provide a snapshot

Inverted repeat structure of *Tip1*:

```
TTACAGTTGA AGTCGGAAGT TTACATACGC TTAGTAGGTT GAAGTCATTA AAACTCATTT TTTAACCACT CCACAGATTT CATATTAGCA AACTATAGTT  100
ATGGCAAGTC ATTTATTACA TCTACTTTGT GCGTGACACA AATAATTTTT CCAACAGTTG TTTACAGACA GATTGTTTCA GTTTTAAATG ATTATAACTA  200
TAACTCCAGT GGGTCAGAAG TTTACTTGCA CTAAGTTAAC TGTGCCTTTA AACAGCTTGG AAAATTCCAG GAAATTAATT TCTAGACTTT AGGCAATTCG  300
CTAATTAGCT TCTGCTAGGC TAATTGGCTA ATTTGAGTCA ATTTCTTCCC CTTGGATGTA TTTTAATGGG AAAATCAAAA GAAATCAGCC AAGACCTCAA  400
GAAAAAAAAA TTGTGGACCT CCACAAGTCT GGTTTATCCT TGGGAGCAAT TTCCAAATGC CTGAGGGTAC CATGTTTCAT CTGTACAAAC AATAGTACAC  500
AAGTATAAGC ACCATTGGAC CATGCAGCAA TCATATAGCT CAGGAAGGAG ACGAGAAATA AACGATGCGA AAAGTGCAAA TCAATCACAG AACAACAGCA  600
AAGGACCTTG TGAAGATGCT GGAAGAAACA GCTAAAGTCT CTATATCCAC AGTTAATGAG TCCTGTATCG ACATAACCTG AAAGGCTAGA AACGCCAGAC  700
TAACAGTTTG CACATGGGGA CAAATAGCTT ACCTTTTGGA GAAAAGTCCT CTGGTATGAT GCAACAAATA CTGAACTGTA TGGCCATAAT GACCATCACT  800
ATGTTTGGAG GAAAAAAGGT GAGGCTTGCA AGCTGAATAT CACGAACCTA ATCATGAAGC GTGGGGGTGG TAGCATCATG TTGTGGGGGT GCTTTGCTGC  900
AGGAGGGACC AGTGCACATT ACAAAATAGA TGGCATCATG AGGAAGGAAA ATTATATGGA TATATTGAAG CAACCACTCA AGACATCAGC CAGGAAGTTA 1000
AATCTCGGTC ATAAATGGGT CTTCCAAATG GACAATGACA ACAAGCATAA TTCCAAAGTT GTGGCAAAAT GGCTTAAGGA CAACAAAGTC AAGGTATTGG 1100
AGTGGCTATC ACAAAGCCCT GACCTCAGTT TGATAGAAAA TTTGTGGGCA GAACTGAAAA AGCATGTGCA AGCAAGATGG CCTATAAACC TGACTCAGTT 1200
ACACCAGTTC TGTATGGAGG AATGGGCCAA CATTACAGCA AATTCTGTAA TTGTGAGAAG CCTGTACAAG GCAAAGAAGA AACGTGAAAC AATTCAAAGG 1300
CGATGCTACG AAATACTAAC AAAGTATGTA AACTTCTGAC CACTGGGAAA GTGATGAAAG AAATAAAAGC TGAAATAAAT AATTCTCTCT ACGTTAATTC 1400
TGACAATTCA CATTGTTAAA ATAAAGTAAT GATCCTAACT GACATAAGAC TGGGGATGTT TTCTAAAGAC TGGGGATGTT TTCTACGATT AAATGTCAGG 1500
AATTGTGGAA AAACTGAATT TACATGTATG TGGTTAAGGT GTATGTAAAC TTCCGACTTC AACTGTAA                                    1568
```

Inverted repeat structure of *Tip2*:

```
TACAGTTGAA GTCGGAAGTT TACATACCAT ACACTCATCA TGGACATGAA CGTCATGGCA ATACTGGGCT TCAATAAGTT CTCTGAGCTT ATCTTTTCTG  100
GGAAGAATTA TTGTACAACA TACATCTTTG ACATATATTT TTTAACCAAG AATTTGGTTC ACAGTTTTAA TACATTTGGG GTTTACTGTA ATCAACACTT  200
TTAGAACCCG CTAAAGTTTG CGGCTGACCA TATGAACAAA GAAAAAGCCT TCTGGAGGAA AGTTTTATGG TCAGATGAGA CAAGATTCAG TTGTTTGGCC  300
ACAATGACCA GAGGTATGCT TGGATGAGTA AAGAACACTG TACCAACTGT TAAGCATGGT GGTGGAAGCA CAACGCTCTG GGGCTGTTTT GCTGCCAGTG  400
GAACAGTGGA AGTGCATGGA ATAATGAAGA AGGGCTAACT TAGAATTATT CAGCATAATC TCAAACCATC AGCCAGAAAG TTGAACAGAA GTTTTCCAAC  500
AGGACAATGA CCCCAAACAC ACATCAAAGC TGGTTGTGGA ATGGATAAAG CATGCTAACA TTAAGCTTGT GAAGGTCTCA ACCCTATCAA AAATTTGTGG  600
ACTGTGCTTA AAAGTCGAGT CAGGGCCAGG AAACCAACAA ATGTCATTCA ACTTTACCAA TTCTGCAAGA AGAGTCGTCA AATATCCAAC CAGAATTCTG  700
CCAGAAGCTG GTGTTTGGCT ACCAAAAGCG TCTGCTCAAG GTGAAACTTG CTAAGGACAT TCAACTAAAT ATCGGGTGTG CTGTATGTAT ATTTTTGACC  800
TTGTATGTAT AATTTTGACC TTGTGTCGAC TACAGAAAAT TAAAACATGT GAACCAAATT CTTGTTTTTG TTTTTTATTA AAGATGTATG TTGTGCAATC  900
ATTCTGCCCC AGAAAAAGCA TCACAGTTCA AAGAAATGAT TGAAAGCCCA ATTTTGCCAT GACATTCATG TTCATGACGA GTGTATGTAA ACTTCCGACT 1000
TCAACTGTA                                                                                                    1009
```

Fig. 2 Inverted repeat structure of *Tip1* and *Tip2*. Sequences of *Tip1* and *Tip2* are shown with IR highlighted. Note that the IR for *Tip1* harbored essentially no additional sequences beyond the PCR primer, and therefore, the actual sequence of *Tip1* IR may differ from the sequence shown, but the length and the structure of the IR should be similar to what is shown

look at the status of *Tc1*-elements in the genome. As a part of our catfish genome program, we recently sequenced 25,000 BAC ends and obtained 20,366 BAC end sequences (Xu et al. 2006). These BAC end sequences covered a total of over 11.4 million base pairs, accounting for approximately 1.2% of the catfish genome. This resource made it possible to assess the abundance of *Tc1*-like elements in channel catfish. The assessment was conducted in two ways: first, the catfish

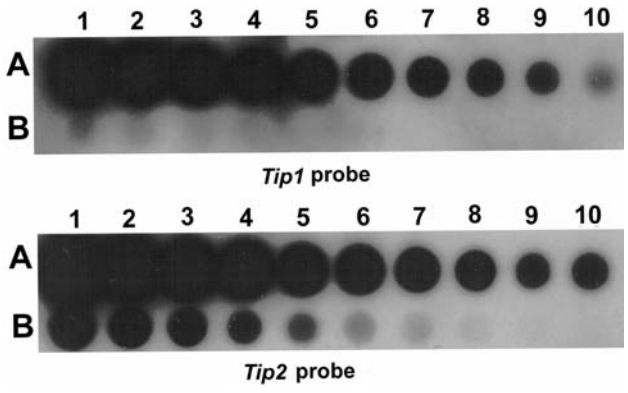

Fig. 3 Dot blot analysis of channel catfish genomic DNA (row B) with control plasmid (Row A) containing *Tip1* (upper) or *Tip2* (lower) to determine copy numbers of *Tip1* and *Tip2*

BAC end sequences were analyzed by using the RepeatMasker program using the zebrafish repeat database that included *Tc1*-transposons. Repeatmasking of *Tc1*-like transposons resulted in masking of 4.22% of bases of the BAC-ends sequences. The masked bases involved 2049 BAC end sequences, suggesting that 10% of the BAC end sequences contained, at least in part, *Tc1*-like sequences. This indicated that the channel catfish genome contained a significantly larger fraction of *Tc1*-like sequences than previously suggested (Liu et al., 1999). In the second approach, we conducted BLASTX searches using the 20,366 BAC end sequences as queries. The BLASTX searches revealed that 1,000 BAC end sequences were similar to *Tc1*-transposases. This indicated that 4.9% of the BAC end sequences contained part of *Tc1*-transposase like sequences. Taken together, these results suggest that at least 4.2% of the catfish genome in base pairs is composed of *Tc1*-like sequences scattered throughout the catfish genome. In consideration of the phylogenetic distance between catfish and zebrafish (even though highly-related), using the zebrafish repeat database to assess the status of *Tc1*-elements in catfish may have underestimated the transposon repeat status in catfish.

In order to determine if a major fraction of *Tc1*-elements was missed in the assessment using zebrafish repeat database, the catfish BES were analyzed by sequence comparison against themselves. A catfish BES database was established and BLASTN was used to search against the catfish BES. After repeatmasking with the zebrafish repeat database, self-BLAST searches were conducted. The BLAST searches resulted in the identification of two major groups of sequences with over 50 (0.25% of the sequences) times of hits. The first group of repeats included 87 BES that had significant hits to one another; and the second group of repeats included 76 BES that had significant hits to one another. Sequence analysis using BLASTX indicated that the first repeat was most similar to rRNA intron-encoded homing endonuclease from chimpanzee (Accession number XP_525925). The second repeat was most similar to an unknown gene sequence from *Schistosoma japonicum* (Accession number AAX30301), and to the cytochrome P450 genes identified from plants (e.g. a cytochrome-like gene from tobacco, Accession number BAA10929). In either case, the number of repeated sequences was not extremely high and the nature of the sequences was not *Tc1*-related. Therefore, we believe that the assessment using the existing zebrafish repeat database provided a reasonable assessment of the status of the *Tc1*-elements in catfish. Given that the entire genome sequences are available only from *Takifugu rubripes*, *Danio rerio*, *Tetraodon nigroviridis*, of which the zebrafish *D. rerio* is evolutionarily the closest to catfish, this genome survey should provide the best assessment of dispersed types repeats in catfish, short of having the entire catfish genome sequences.

Assessment of transcribed *Tc1*-like transposase sequences

ESTs serve as an important resource for the assessment of the expression of *Tc1*-elements (Krasnov et al. 2005). This is particularly true because many *Tc1*-related elements could be transcribed as small remnants, making traditional expression analysis such as Northern-blot analysis or RT-PCR less efficient. In order to assess to what extent the *Tc1*-like transposases were expressed, a total of 45,457 channel catfish ESTs were subjected to BLASTX searches. As we generated the vast majority of these ESTs from non-normalized libraries, a transcriptomic survey of *Tc1*-elements should accurately reflect their proportion in the RNA pool (Peale and Gerritsen 2001; Knox and Skuce 2005; Hoffman 2005). The BLAST searches resulted in a pool of 265 ESTs similar to *Tc1*-like transposases in GenBank (0.6%) with a cutoff value of e^{-5}. We first attempted to cluster these ESTs by nucleotide sequence similarity. However, very few of the transposase transcripts could be placed into contigs due to their having been sequenced from different regions and/or high sequence divergence. We, therefore, sorted the 265 transposase transcripts by their top BLASTX hit and the BLAST e-value, removing redundant top hits. This resulted in a list of 28 transcribed *Tc1*-like transposase sequences. As judged from their top hits using BLASTX searches, it appeared that two groups of transposon-related sequences were expressed at particularly high levels: one was most similar to the *Frog Prince* transposase (Miskey et al. 2003) with 105 ESTs, and the other was most similar to a transposase from the fish *P. platessa* with 51 ESTs. These groups of ESTs accounted for over 58% of the channel catfish transposase-related ESTs.

Phylogenetic analysis of the transcribed *Tc1*-like sequences indicated that at least seven distinct clades of transposase-like ESTs existed in catfish (Fig. 4). The first clade, as represented by CK410194, contained the largest number of ESTs (118). The second largest clade, as represented by CK416451, also contained a large number of ESTs (96), while all other clades contained a smaller number of ESTs.

Both sense and antisense transposase transcripts exist in catfish

With a significant number of transcripts representing *Tc1*-like transposases available, a search was conducted to determine if any of the transcripts contain reading frames that are still open. Although functional transposase genes are very rare among vertebrates, full transposase ORFs were reported from Japanese medaka and from *X. tropicalis* (Koga and Hori 2000; Sinzelle et al. 2005). In order to determine if any of the catfish transposase transcripts harbor a reading frame for transposase that is still open, the 265 transposase-related transcripts were translated in six frames. All the transcripts harbored at least one in-frame termination codon, suggesting that they are all non-functional at present.

Both sense and antisense transposase transcripts exist among the catfish transcripts. Of the 265 transposase transcripts, 117 transcripts were in a sense strand orientation, while 148 were in an antisense orientation. This may suggest that the majority of the transposase transcripts were derived from transcription using promoters other than their own.

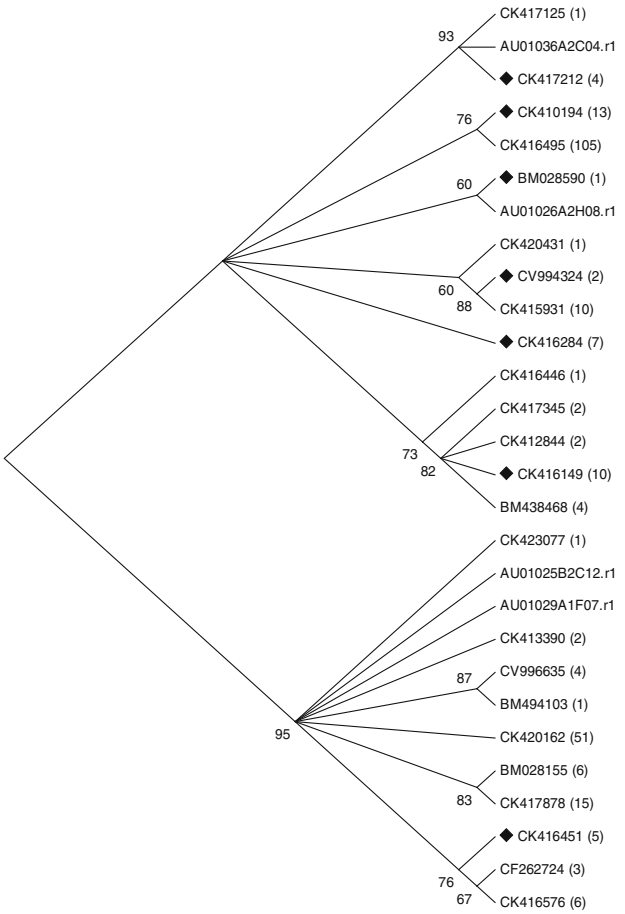

Fig. 4 Phylogenetic analysis of expressed and genomic transposase sequences from catfish. Numbers on branches represent the percentage with which the tree topology is replicated after 1,000 bootstrap replications. Sequences are listed either by their GenBank accession number for ESTs or by a unique plate identifier for BES. Each EST on the tree represents one or more transcripts with a unique transposase BLAST hit. Numbers in parentheses are the number of catfish transcripts that share the same top BLAST hit. Sequences indicated by black diamonds were included in Fig. 5 as representatives of their respective clades

Phylogenetic relationships with other *Tc1*-like elements

To assess the evolutionary relationship between the channel catfish *Tc1*-like elements and those identified from other species, we studied their phylogeny based on sequence similarities within the transposase amino acid sequences. Conceptualized amino acid sequences were produced for vertebrate *Tc1*-like elements as previously used by many researchers (e.g., Leaver, 2001; Turcotte and Bureau 2002; Sinzelle et al. 2005, 2006). For the catfish *Tc1*-like elements, a representative sequence from each of the seven clades of transcribed *Tc1*-like elements was used. In addition, genomic sequences that were not accounted for by the seven clades were also included, if the sequences allowed generation of judicious amino acid sequences long enough for multiple sequence alignment and phylogenetic analysis. As shown in Fig. 5, catfish *Tc1*-like elements fell within a wide spectrum of clades, as well as formed new clades. In spite of low bootstrapping values for many of the clades, the most abundantly expressed types of catfish *Tc1*-elements appeared to be placed into clades with reasonable statistical support. CK410194, the most abundantly expressed type in catfish, was similar to *Tzf* (U51226) from zebrafish; the second most abundantly expressed catfish *Tc1*-element as represented by CK416451 was most similar to *Xeminos*, a recently identified *Minos*-like *Tc1*-element from *X. tropicalis* (Sinzelle et al. 2005; Arca and Savakis 2000). The catfish *Tip1* element was highly related to *Tss* from Atlantic salmon and was placed in a clade also containing *Tdr1* from zebrafish. Several additional catfish *Tc1*-elements exhibited similarities to existing known *Tc1*-elements, but bootstrapping values were low in supporting the clades; these included a catfish *Tc1* element (BM028590) that was similar to *Txz* and *XtTxz* identified from amphibians, a catfish *Tc1*-element (CK417212) that was similar to *Quetzal* (AAB02109), and a catfish *Tc1*-element (CV994324) that was similar to *Maya* elements. Despite the weak support for some clades, it is clear that the channel catfish genome contains multiple families of *Tc1*-like elements, with three of the families placed into clades with reasonably high statistical support.

Discussion

In this work, we assessed the abundance and types of *Tc1*-like elements in the channel catfish genome and their expression. Through a combination of PCR amplification, assessment through BAC end sequencing, and hybridization analysis, we estimated that the catfish genome is rich in *Tc1*-like elements with approximately 4.2% of its nucleotide bases composed of *Tc1*-like elements. The use of PCR amplification allowed us to obtain *Tc1*-like elements most similar to those characterized from teleost species, such as the *Tss* from Atlantic salmon and *Tdr1* from zebrafish, while BAC end sequencing of 1.2% of channel catfish genomic sequences allowed assessment of *Tc1* status in the catfish genome based on the randomly sequenced BAC ends. Interestingly, BAC end sequencing analysis revealed that while *Tc1*-like elements accounted for

Fig. 5 Phylogenetic analysis of catfish transposases combined with transposases from a variety of vertebrate and invertebrate species. Numbers on branches represent the percentage with which the tree topology is replicated after 1,000 bootstrap replications, with only those above 50% shown. Sequences are listed by their species names followed by their GenBank accession number in parentheses, and transposon names when available. Catfish *Tip1* is noted by a black triangle. Sequences indicated by black diamonds were included as representatives of their respective clades, identified in Fig. 4. Numbers in the second parentheses are the number of catfish transcripts from that respective clade

4.2% of the total bases, over 10% of the BAC end sequences contained some level of *Tc1*-like sequences. This result suggests that *Tc1*-like elements are highly dispersed in the catfish genome.

The present study adds to our knowledge of the repeat structures of the channel catfish genome. To date, four classes of repetitive DNA elements have been identified in the channel catfish genome. The most abundant repeat class is an approximately 325 bp A/T-rich *Xba* element with approximately 150,000 copies in the catfish genome (Liu et al. 1998b). The catfish genome also contains approximately 9,000 copies of the 57 bp *Mermaid* and approximately 1,200 copies of the 64 bp *Merman* elements, which are short interspersed repetitive elements found in many bony fish (Kim et al. 2000). We previously reported a highly abundant non-autonomous *Tc1*-element named *Tip-non* that existed in high copy numbers (32,000) accounting for 1.6% of the catfish genome. Here in this work, we assessed the total content of *Tc1*-like elements using random genomic sequences generated from BAC end sequencing. As compared with hybridization approaches, the BAC end sequences should provide a more accurate picture of *Tc1*-like elements in the catfish genome. Unlike the *Xba* elements which are arranged in head-to-tail arrays around the centromere (Liu et al. 1998b; Quiniou et al. 2005), the *Tc1*-transposon like elements are extremely dispersed in the genome. Such information is crucial not only in considerations for entire genome sequencing, but also for comparative genome analysis and the establishment of conserved syntenies. As the *Tc1*-like elements are highly repetitive and may contain pieces of other genes, special caution is needed in orthology and synteny studies involving such sequences. The identified *Tc1*-like sequences in this study should also allow the development of a repeat library for the database of RepeatMasker, thereby simplifying sequence analysis and assembly of major genome projects involving teleost species.

Although our BAC end sequences accounted for only 1.2% of the catfish genome, we believe this

assessment is reliable because the BAC end sequences were sequenced from random BAC clones. The most abundant *Tc1*-like elements in the catfish genome appeared to be those most similar to the *Tzf* elements. At the level of RNA, the largest number of ESTs also fell into the clade containing *Tzf*, suggesting that the transcription of the *Tc1*-like elements was correlated with the abundance of these elements in the catfish genome.

Many catfish *Tc1*-elements exhibited very high levels of nucleotide similarities to *Tc1*-elements from other species. In many cases, the catfish *Tc1*-elements were similar to *Tc1*-elements not only from other closely-related teleosts such as zebrafish or salmonids, but also from distantly related species. For instance, the catfish EST (Accession number BM028155) is most similar to a *Tc1*-element from sea lamprey with 94% nucleotide identity, and to a sequence from the trematode *S. japonicum*, also with 94% nucleotide identity. This may indicate occurrence of recent horizontal gene transfers, as recently reported between salmonids and *Schistosoma* (Melamed et al. 2004).

The catfish transcriptome included approximately 0.6% of transcripts related to *Tc1*-like transposase. This assessment was made with EST expression profiling. We believe this approach is superior to the conventional approaches such as Northern-blot analysis or RT-PCR because the expressed *Tc1*-related sequences may have various transcript sizes as they are present as highly dispersed remnants. Using traditional hybridization-based approaches, cross hybridization would be a major problem for related sequences; also as there are so many different types of transcripts, a genome level analysis would be difficult. Using EST sequencing and profiling, the proportion of any transcript should be represented in the libraries proportional to the abundance of the transcripts, if the ESTs were sequenced from non-normalized cDNA libraries. For catfish ESTs, the vast majority of ESTs were produced in our laboratory and the libraries were not normalized.

The catfish *Tc1*-like elements so far characterized all carried in frame terminations as a result of point mutations or frame shifts. However, despite their current state of inactivity, such a large array of transposable elements, at one point during evolution, must have played a dynamic role in shaping the catfish genome. The fact that large numbers of the elements were in antisense orientation suggest that the transcription of the *Tc1*-like elements in catfish genome, at least at its current stage, was more a consequence of transposition than active expression. Continued genomic sequencing in catfish and an upcoming Joint Genome Institute (JGI) project to sequence 600,000 catfish ESTs will both benefit from our categorization of the *Tc1* transposable elements. Additionally, these projects may in turn reveal the presence of additional, less abundant transposon types and clarify the murky history and future of these elements in teleost genomes.

Acknowledgements This project was supported by National Research Initiative Grant No. # 2006-35616-16685 from the USDA Cooperative State Research, Education, and Extension Service under the Animal Genome Basic Genome Reagents and Tools Program, and in part by a Specific Cooperative Agreement with USDA ARS Aquatic Animal Health Laboratory under the Contract Number 58-6420-5-030. Samiran Nandi was supported by a fellowship from the Department of Biotechnology of the Government of India.

References

Arca B, Savakis C (2000) Distribution of the transposable element *Minos* in the genus *Drosophila*. Genetica 108:263–267

Avancini RM, Walden KK, Robertson HM (1996) The genomes of most animals have multiple members of the *Tc1* family of transposable elements. Genetica 98:131–140

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

Castro JP, Carareto CM (2004) *Drosophila melanogaster* P transposable elements: mechanisms of transposition and regulation. Genetica 121:107–118

Davidson AE, Balciunas D, Mohn D, Shaffer J, Hermanson S, Sivasubbu S, Cliff MP, Hackett PB, Ekker SC (2003) Efficient gene delivery and gene expression in zebrafish using the Sleeping Beauty transposon. Dev Biol 263:191–202

Dawson A, Finnegan DJ (2003) Excision of the Drosophila *mariner* transposon Mos1. Comparison with bacterial transposition and V(D)J recombination. Mol Cell 11:225–235

Dupuy AJ, Akagi K, Largaespada DA, Copeland NG, Jenkins NA (2005) Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. Nature 436:221–226

Fadool JM, Hartl DL, Dowling JE (1998) Transposition of the *mariner* element from *Drosophila mauritiana* in zebrafish. Proc Natl Acad Sci USA 95:5182–5186

Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. Trends Genet 5:103–107

Goodier JL, Davidson WS (1994) *Tc1* transposon-like sequences are widely distributed in salmonids. J Mol Biol 241:26–34

Gottgens B, Barton LM, Grafham D, Vaudin M, Green AR (1999) Tdr2, a new zebrafish transposon of the *Tc1* family. Gene 239:373–379

Gueiros-Filho FJ, Beverley SM (1997) Trans-kingdom transposition of the Drosophila element *mariner* within the protozan Leishmania. Science 276:1716–1719

Heierhorst J, Lederis K, Richter D (1992) Presence of a member of the *Tc1*-like transposon family from nematodes and Drosophila within the vasotocin gene of a primitive vertebrate, the Pacific hagfish Eptatretus stouti. Proc Natl Acad Sci USA 89:6798–802

Henikoff S (1992) Detection of Caenorhabditis transposon homologs in diverse organisms. New Biol 4:382–388

Hoffmann R (2005) Gene expression patterns in human and mouse B cell development. Curr Top Microbiol Immunol 294:19–29

Ivics Z, Izsvak Z (2004) Transposable elements for transgenesis and insertional mutagenesis in vertebrates: a contemporary review of experimental strategies. Methods Mol Biol 260:255–276

Ivics Z, Kaufman CD, Zayed H, Miskey C, Walisko O, Izsvak Z (2004) The Sleeping Beauty transposable element: evolution, regulation and genetic applications. Curr Issues Mol Biol 6:43–55

Ivics Z, Hackett PB, Plasterk RH, Izsvak Z (1997) Molecular reconstruction of *Sleeping Beauty*, a *Tc1*-like transposon from fish, and its transposition in human cells. Cell 91:501–510

Ivics Z, Izsvak Z, Minter A, Hackett PB (1996) Identification of functional domains and evolution of *Tc1*-like transposable elements. Proc Natl Acad Sci USA 93:5008–5013

Izsvak Z, Ivics Z, Hackett PB (1995) Characterization of a *Tc1*-like transposable element in zebrafish (*Danio rerio*). Mol Gen Genet 247:312–322

Kawakami K, Shima A, Kawakami N (2000) Identification of a functional transposase of the Tol2 element, an Ac-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage. Proc Natl Acad Sci USA 97:11403–11408

Kim S, Karsi A, Dunham RA, Liu Z (2000) The skeletal muscle alpha-actin gene of channel catfish (*Ictalurus punctatus*) and its association with piscine specific SINE elements. Gene 252:173–181

Knox DP, Skuce PJ (2005) SAGE and the quantitative analysis of gene expression in parasites. Trends Parasitol 21:322–326

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

Koga A, Hori H (2000) Detection of de novo insertion of the medaka fish transposable element Tol2. Genetics 156:1243–1247

Koga A, Hori H (2001) The Tol2 transposable element of the medaka fish: an active DNA-based element naturally occurring in a vertebrate genome. Genes Genet Syst 76:1–8

Krasnov A, Koskinen H, Afanasyev S, Molsa H (2005) Transcribed *Tc1*-like transposons in salmonid fish. BMC Genomics 6:107

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinfor 5:150–163

Lam WL, Seo P, Robison K, Virk S, Gilbert W (1996) Discovery of amphibian *Tc1*-like transposon families. J Mol Biol 257:359–366

Leaver MJ (2001) A family of *Tc1*-like transposons from the genomes of fishes and frogs: evidence for horizontal transmission. Gene 271:203–214

Liu ZJ, Nichols A, Li P, Dunham RA (1998a) 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 ZJ, 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

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

Liu ZJ, Li P, Kucuktas H, Dunham RA (1999) Characterization of nonautonomous *Tc1*-like transposable elements of channel catfish (*Ictalurus punctatus*). Fish Physiol Biochem 21:65–72

Loukeris TG, Livadaras I, Arca B, Zabalou S, Savakis C (1995) Gene transfer into the medfly, *Ceratitis capitata*, with a *Drosophila hydei* transposable element. Science 270:2002–2005

Melamed P, Chong KL, Johansen MV (2004) Evidence for lateral gene transfer from salmonids to two *Schistosome* species. Nat Genet 36:786–787

Miskey C, Izsvak Z, Kawakami K, Ivics Z (2005) DNA transposons in vertebrate functional genomics. Cell Mol Life Sci 62:629–641

Miskey C, Izsvak Z, Plasterk RH, Ivics Z (2003) The Frog Prince: a reconstructed transposon from *Rana pipiens* with high transpositional activity in vertebrate cells. Nucleic Acids Res 31:6873–6881

Peale FV Jr, Gerritsen ME (2001) Gene profiling techniques and their application in angiogenesis and vascular development. J Pathol 195:7–19

Plasterk RH (1996) The *Tc1*/*mariner* transposon family. Curr Top Microbiol Immunol 204:125–143

Quiniou SM, Wolters WR, Waldbieser GC (2005) Localization of Xba repetitive elements to channel catfish (*Ictalurus punctatus*) centromeres via fluorescence in situ hybridization. Anim Genet 36:353–354

Radice AD, Bugaj B, Fitch DH, Emmons SW (1994) Widespread occurrence of the *Tc1* transposon family: *Tc1*-like transposons from teleost fish. Mol Gen Genet 244:606–612

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sambrook J, Frisch EF, Maniatis T (1989) Molecular Cloning, A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Sinzelle L, Pollet N, Bigot Y, Mazabraud A (2005) Characterization of multiple lineages of *Tc1*-like elements within the genome of the amphibian *Xenopus tropicalis*. Gene 349:187–196

Sinzelle L, Chesneau A, Bigot Y, Mazabraud A, Pollet N (2006) The *mariner* transposons belonging to the irritans subfamily were maintained in chordate genomes by vertical transmission. J Mol Evol 62:53–65

Thompson JD, Higgins DG, Gibson TJ (1994) ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Turcotte K, Bureau T (2002) Phylogenetic analysis reveals stowaway-like elements may represent a fourth family of the IS630-Tc1-*mariner* superfamily. Genome 45:82–90

Wadman SA, Clark KJ, Hackett PB (2005) Fishing for answers with transposons. Mar Biotechnol 7:135–141

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