

RESEARCH ARTICLE

The annotation of repetitive elements in the genome of channel catfish (*Ictalurus punctatus*)

Zihao Yuan¹, Tao Zhou¹, Lisui Bao¹, Shikai Liu¹, Huitong Shi¹, Yujia Yang¹, Dongya Gao¹, Rex Dunham¹, Geoff Waldbieser², Zhanjiang Liu^{3*}

1 School of Fisheries, Aquaculture and Aquatic Sciences, Auburn University, Auburn, Alabama, United States of America, **2** USDA-ARS Warmwater Aquaculture Research Unit, Stoneville, Mississippi, United States of America, **3** Department of Biology, Syracuse University, Syracuse, New York, United States of America

* johnliu@syr.edu



OPEN ACCESS

Citation: Yuan Z, Zhou T, Bao L, Liu S, Shi H, Yang Y, et al. (2018) The annotation of repetitive elements in the genome of channel catfish (*Ictalurus punctatus*). PLoS ONE 13(5): e0197371. <https://doi.org/10.1371/journal.pone.0197371>

Editor: Hanping Wang, Ohio State University, UNITED STATES

Received: December 17, 2017

Accepted: May 1, 2018

Published: May 15, 2018

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This project was supported by a USDA grant (grant no. 2017-67015-26295) from the Agriculture and Food Research Initiative (AFRI) under the Animal Breeding, Genetics and Genomics Program. Zihao Yuan was supported by China Scholarship Council (201406330063).

Competing interests: The authors have declared that no competing interests exist.

Abstract

Channel catfish (*Ictalurus punctatus*) is a highly adaptive species and has been used as a research model for comparative immunology, physiology, and toxicology among ectothermic vertebrates. It is also economically important for aquaculture. As such, its reference genome was generated and annotated with protein coding genes. However, the repetitive elements in the catfish genome are less well understood. In this study, over 417.8 Megabase (MB) of repetitive elements were identified and characterized in the channel catfish genome. Among them, the DNA/TcMar-Tc1 transposons are the most abundant type, making up ~20% of the total repetitive elements, followed by the microsatellites (14%). The prevalence of repetitive elements, especially the mobile elements, may have provided a driving force for the evolution of the catfish genome. A number of catfish-specific repetitive elements were identified including the previously reported *Xba* elements whose divergence rate was relatively low, slower than that in untranslated regions of genes but faster than the protein coding sequences, suggesting its evolutionary restrictions.

Introduction

Eukaryotic genomes contain significant amount of repetitive DNA sequences, and the collective of the repeated sequences in an organism is known as the repeatome of the organism [1]. Such repetitive sequences were once thought to be junk DNA [2], but recent studies have indicated that they play important roles in propelling genome evolution and adaptation to environments [3–9]. The repeatomes of higher vertebrates, especially those of mammals, have been well studied, but their studies are limited for aquatic species.

Repetitive sequences can be generally divided into three major categories: the dispersed repeats such as transposable elements or transposons, tandem repeats, and high copy number genes [1]. Transposons are dispersed across genomes and their proportion are highly variable among genomes, ranging from 3% to 85% in terms of physical size [10–11]. For instance, the genome of *Utricularia gibba* contains only 3% of repetitive sequences [12–13], while the

genome of maize contains over 85% transposable elements [14–15]. Based on their mechanisms of proliferation, transposons can be further classified into RNA-mediated Class I transposons and RNA-independent Class II DNA transposons. Class I transposons contain three main subclasses: short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs), and transposons with long terminal repeats (LTRs). The Class II transposons can be further divided into two classes based on their transposition mechanisms. The TIR-based Subclass-I elements such as piggyBacs, hATs, which are proliferated through the cut-and-paste mechanism; As well as non-TIR based Sub-class II DNA transposons such as Helitron, and Maverick; which are mobilize by rolling- circle replication via a single stranded DNA intermediate [16–17]. Transposons are capable of moving in the genome, and therefore, they are believed to be a major driving force for genome evolution [18–21].

Tandem repeats are individual repeats of DNA located adjacent to one another comprising variable numbers of nucleotides within each repeat sequence, and variable numbers of repeats such as microsatellites and satellites [22–24]. Tandem repeats are mostly presented in the centromeric, telomeric, and subtelomeric regions of chromosomes. In some cases, the tandem repeats can also make up large fractions of the genome [25–26]. The amplification and/or mutations of tandem repeats may also affect the genome by changing the genome structures or genome sizes [27–29], thereby affecting recombination of genomes, gene expressions, gene conversions, and chromosomal organizations [30–34].

High copy number genes such as ribosomal RNA (rRNA) genes or immunoglobulins also make up significant fractions of the repeatome. For instance, the copy numbers of rRNA genes can be as high as 4,000 copies, such as in the genome of pea (*Pisum sativum*) [35]. In *Saccharomyces cerevisiae*, a single cluster of rRNA can cover about 60% of the chromosome XII [36], thus it has been considered as the “king of the housekeeping genes” in terms of function and quantity [37]. Similarly, immunoglobulin genes have been found to be highly repetitive. For instance, the catfish IgH locus contains at least 200 variable (V) region genes, three diversity (D) and 11 joining (JH) genes for recombination [1, 38–39].

Channel catfish (*Ictalurus punctatus*) is a freshwater fish species distributed in lower Canada and the eastern and northern United States, as well as parts of northern Mexico. Its high tolerance and adaptability to harsh environments made it one of the most popular aquaculture species. In the United States, it is the primary aquaculture species, accounting for over 60% of all U.S. aquaculture production [40–41]. Its reference genome sequence with annotation of protein coding genes was published [42], but its repeatome was not fully characterized. Previous works reported the presence of an A/T-rich tandem Xba elements on the centromere regions but not in closely related species [43–44], and presence of dispersed SINE elements [45] and DNA transposons [46] in the channel catfish genome was also reported. Genomic sequencing surveys provided additional information about other repetitive sequences such as microsatellites, and transposons [47–52]. Here we annotated and characterized the repeatome of the channel catfish genome from sequences generated for whole genome sequencing.

Material and methods

Annotation of repetitive elements in channel catfish genome

The identification and annotation of the repetitive elements in the channel catfish genome assembly along with the degenerate sequences [42] were conducted using the RepeatModeler 1.0.8 package (<http://www.repeatmasker.org/RepeatModeler.html>) containing RECON [53] and RepeatScout [54]. The identified channel catfish repetitive sequences were searched against curated libraries and repetitive DNA sequence database Repbase [55] and Dfam [56] derived from RepeatMasker 4.0.6 package (<http://www.repeatmasker.org/>). To further

determine characters of the repetitive elements classified as “Unknown” by the RepeatModeler, those sequences were first clustered by self-alignments via CD-HIT, with sequence identity cut-off set as 50% [57–58]. Then, the clustered sequences were searched against the entire NCBI Nucleotide collection database (nt) using BLASTN: 2.2.28+ with a relatively relaxed E-value ($<10^{-5}$) to annotate the sequence with the best hit.

The distribution and density of repetitive elements

The distribution frequency of the repetitive elements of DNA/TcMar-Tc1 as well as microsatellites and satellites sequences on the chromosomes were subtotaled and calculated by the location information and abundance information reported by the RepeatMasker. Their density on the chromosomes was presented as bp/MB. The heat map was plotted using the Heml1.0 [59].

Divergence time of channel catfish and blue catfish

The divergence time and their 95% credibility intervals of channel catfish and blue catfish (*Ictalurus furcatus*) were calculated based on the divergence of cytochrome b genes with the calibration of fossil records. The substitution rate of cytochrome b was determined as normal distribution with mean of 1.05% and a standard deviation of 0.0105% [60]. In addition to channel catfish and blue catfish, we also used the cytochrome b sequences of blind cave fish (*Astyanax mexicanus*), common carp (*Cyprinus carpio*), and zebrafish (*Danio rerio*) for phylogenetic analysis (S1 File). The analysis was performed using the BEAST v.1.8.0 package [61]. Two independent runs were performed with 1,000 generations sampled from every 10 million generations for each dataset using MCMC chains [62]. The input files were constructed in BEAUTi, and the best substitution model was selected by Prottest 3.2.1 according to the alignments [63]. Model parameters consisted of a GTR+I+G model with a lognormal relaxed clock [64–65], the speciation birth-death process, and random starting tree were also applied in the phylogenetic analysis.

For the files of the resulting trees, we used the TreeAnnotator v1.8.0 to discard 10% of samples as burn-in and summarized the information of the remaining samples of trees onto a maximum clade credibility chronogram, and the results were viewed in Figtree with mean divergence times and 95% age credibility intervals.

To calibrate the divergence time of major clades for a better phylogenetic analysis, we have selected three teleost fossil records for calibration, the following node ages were set using log-normal priors:

1. Time of most recent common ancestor of *Ictalurus* (channel catfish and blue catfish), 19 MYR, (lognormal mean of 19 and standard deviation of 1.9), following Blanton and Hardman [66–67].
2. Time of most recent common ancestor of Characiformes (blind cave fish), 94 MYR, (lognormal mean of 94 and standard deviation of 9.4), with the fossil record discovered in Cenomanian [68–69].
3. Time of most recent common ancestor of Cypriniformes (common carp and zebrafish), 50 MYR, (lognormal mean of 50 and standard deviation of 5.0) with fossil discovered in Ypresian [70].

Substitution rate of the Xba elements

The overall evolutionary dynamics can be referred from the average number of substitutions per site (K). The K was estimated from the divergence levels reported by Repeatmasker, using the one-parameter Jukes-Cantor Formula $K = -300/4 \times \ln(1 - D \times 4/300)$ as described in

previous studies [71], where D represents the proportion of sites that differ between the fragmented repeats and the consensus sequence. For channel catfish and blue catfish Xba elements, the nucleotide substitution rate (r) was calculated using the formula $r = K/(2T)$ [72], where T is the divergence time of channel catfish and blue catfish. To calculate the average K of the different types of repetitive elements, K of each element was multiplied by the length of the element, and the sum of all elements was divided by the sum of the total length of the elements.

Results

Annotation of repetitive elements in the channel catfish genome

The major categories of the repetitive elements in the channel catfish genome are shown in Fig 1 and detailed in S1 Table and S2–S5 Files. The channel catfish genome harbored a total of 417.8 Mb of repetitive elements, accounting for 44% of the catfish genome. Of all the repetitive elements, 84.1% were annotated as known repetitive elements, while 15.9% were previously unclassified repetitive elements in the channel catfish genome. The known repetitive elements fell into 70 major categories, with the category of Tc1/mariner transposons accounting for the largest percentage (19.9% in repeatome, 8.8% in genome), followed by microsatellites (14.1% in repeatome, 6.2% in genome), repetitive proteins (7.2% in repeatome, 3.2% in genome), LINE/L2 (4.3% in repeatome, 1.9% in genome), Xba elements (3.5% in repeatome, 1.6% in genome), LTR/Nagro (3.1% in repeatome, 1.4% in genome), hAT/Ac (3.0% in repeatome, 1.3% in genome), unclassified DNA transposons (2.9% in repeatome, 1.3% in genome), LTR/Gypsy (2.3% in repeatome, 1.0% in genome), LTR/DIRS (2.2% in repeatome, 1.0% in genome), CMC-EnSpm (2.1% in repeatome, 0.9% in genome), Ginger (1.8% in repeatome, 0.8% in genome), satellite (1.7% in repeatome, 0.7% in genome), hAT (1.3% in repeatome, 0.6% in genome), SINE/MIR (1.3% in repeatome, 0.6% in genome), low complexity elements (1.1% in repeatome, 0.5% in genome), DNA/hAT-Charlie (1.1% in repeatome, 0.5% in genome), RC/Helitron (1.1% in repeatome, 0.5% in genome), and LINE/Rex-Babar (1.0% in repeatome, 0.5% in genome). All the remaining categories represented less than 1% each of the repetitive elements (S1 Table).

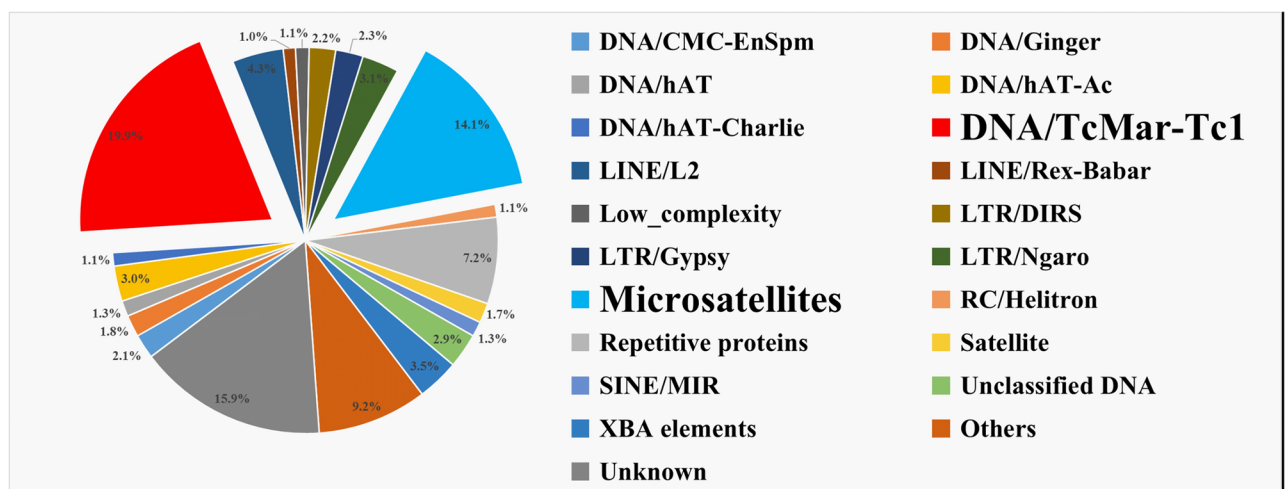


Fig 1. The proportion of major categories of repetitive elements within the channel catfish repeatome.

<https://doi.org/10.1371/journal.pone.0197371.g001>

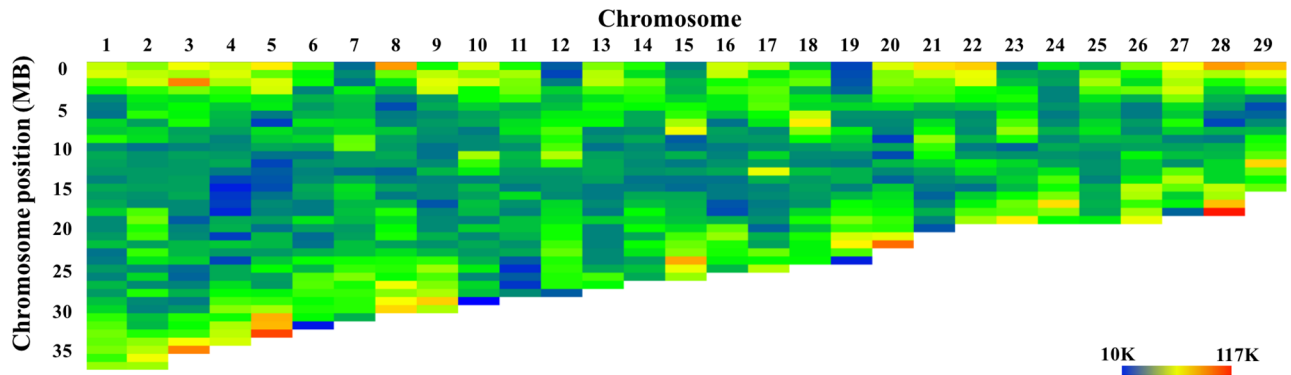


Fig 2. The distribution of Tc1/Mariner transposons cross channel catfish genome. Color key is indicated at the lower right of the figure, with blue color to indicate low and red color to indicate high levels of the transposons in the chromosomal regions. Each color bar represented a physical distance of 1 Mb DNA.

<https://doi.org/10.1371/journal.pone.0197371.g002>

The distribution of repeats cross genome

The Tc1/Mariner transposons are distributed cross the whole genome, with no major differences among chromosomes or among chromosomal regions within chromosomes (Fig 2). Among the annotated microsatellites, the dinucleotide microsatellites are the most abundant type, making up nearly 46% of the total annotated microsatellite sequences followed by tetra- and tri-nucleotide microsatellites, making up 18.6% and 13.6% of the total annotated microsatellites in length, respectively. As shown in Fig 3, the microsatellites and satellites are abundant on both ends of the chromosomes and some of them are distributed on the middle of the chromosomes. This is in consistent with previous results that telomere regions and centromere regions contain large part of short tandem repeats [73–75].

Substitution rates

The analysis of evolutionary rate of the unique Xba elements within catfish is useful to assess their limitations of evolution and assessment of their potential functions. The divergence analysis indicated that the Xba elements have a low mean Jukes-Cantor distance of 3.53, lower

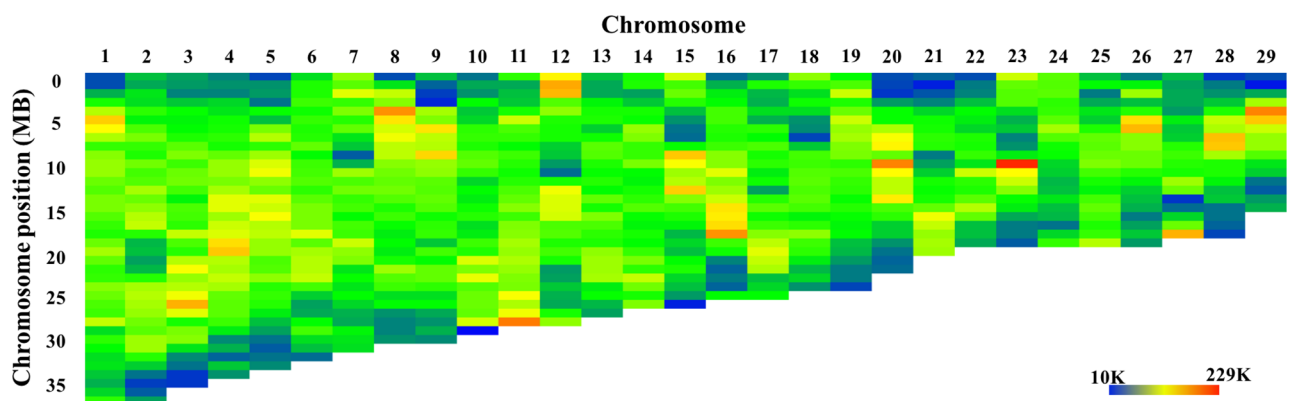


Fig 3. The distribution of microsatellites and satellites along the chromosomes of the channel catfish genome. Color key is indicated at the lower right of the figure, with blue color to indicate low and red color to indicate high levels of the microsatellites and satellites in the chromosomal regions. Each color bar represented a physical distance of 1 Mb DNA.

<https://doi.org/10.1371/journal.pone.0197371.g003>

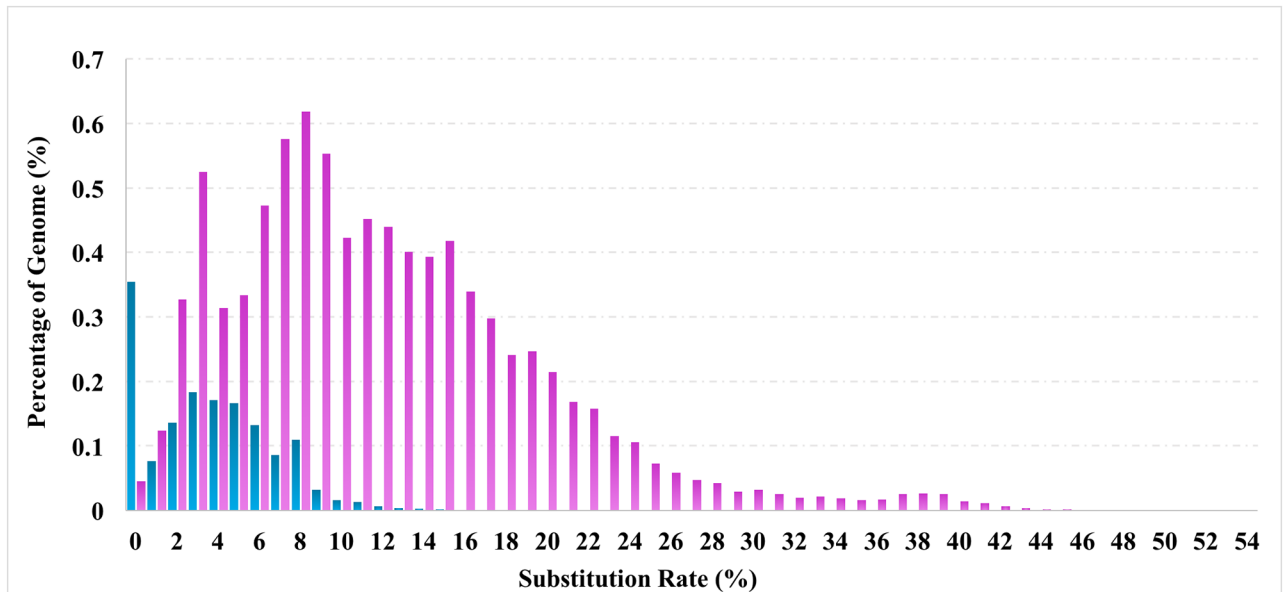


Fig 4. The divergence distribution of channel catfish Xba elements (blue) and DNA/TcMar-Tc1 transposons (pink). The X-axis represents the average number of substitutions per site (%), and the Y-axis represents the percentage sequences that comprise the whole genome (%).

<https://doi.org/10.1371/journal.pone.0197371.g004>

than the average Jukes-Cantor distance of 13.34 of the channel catfish repeatome. Meanwhile, compared with Xba elements, the substitution distribution of the catfish DNA/TcMar-Tc1 transposons, most prevalent in the catfish genome, are characterized not only by a broader distribution of divergence up to more than 50%, but also a larger mean divergence rate of approximately 12% (Fig 4). This indicated a long history of evolution as well as a more active evolutionary dynamics during the evolution of DNA/TcMar-Tc1 transposons in the catfish genomes, and recent acquisition of the Xba elements specific to the *Ictalurus* catfishes.

The inference of divergence time of channel catfish and blue catfish are important for the calculation of the rate of nucleotide substitutions of their unique Xba elements. The maximum clade credibility chronogram analysis indicated that the channel catfish and blue catfish separated approximate 16.6 million years (Myr) ago, with a 95% age credibility intervals of 13.3–19.9 Myr (Fig 5). This is consistent with the earliest fossil record of the channel catfish discovered in Nebraska in the middle Miocene, and agreed with previous analysis of approximate 21 Myr of separation of channel catfish and blue catfish [66]. Based on the average number of substitution per site and the divergence time, the rate of nucleotide substitutions of the Xba elements was calculated as 8.9×10^{-8} to 1.3×10^{-7} substitutions per site per year. Meanwhile, based on the results of the previous research on differences of full length cDNA sequences between channel catfish and blue catfish [76], the rate of nucleotide substitution of Xba elements are higher than those in the open reading frame regions (2.5×10^{-8} to 7.6×10^{-8}), but lower than those in untranslated regions (1.3×10^{-7} to 1.9×10^{-7}).

Novel repetitive elements in the catfish genome

Among the repetitive elements in channel catfish, there are still about ~16% of the repetitive sequences which cannot be annotated from neither the repetitive element databases nor the known non-redundant nucleotide database. Those sequences are rich in A/T (58%), the grouping of those sequences with more than 50% in similarity by CD-hit had grouped them into 215 categories (S2 Table). The top categories with over 500 Kb in length and their representative

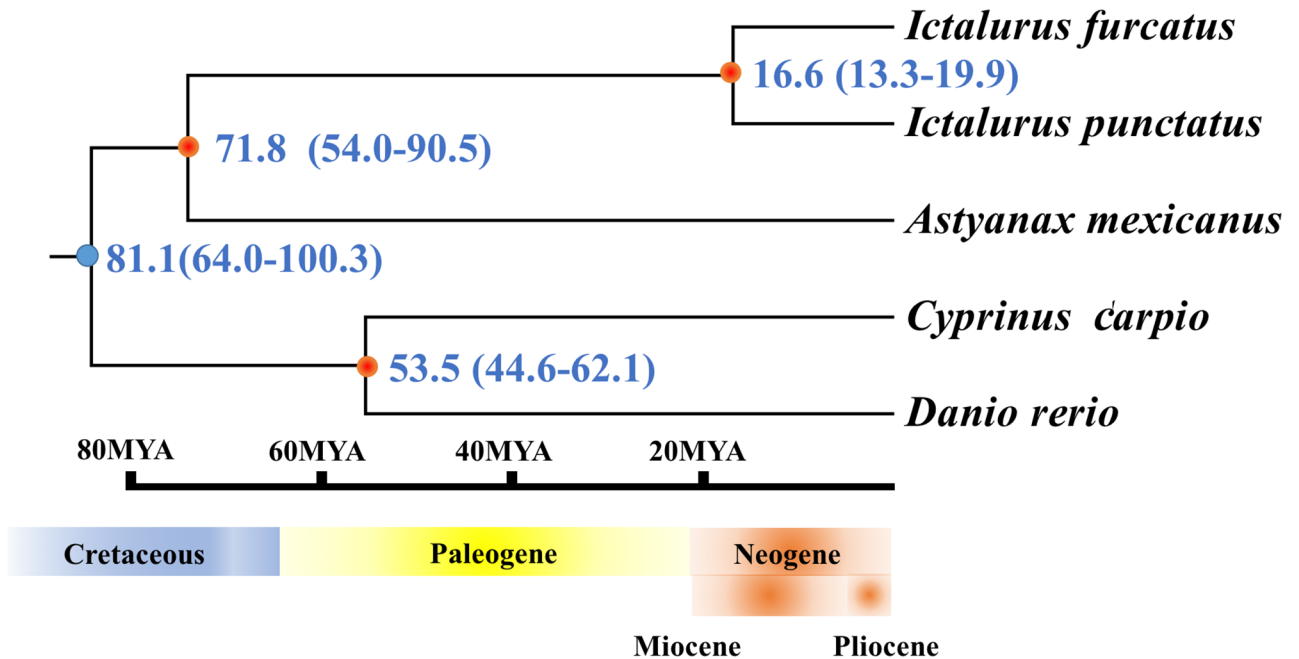


Fig 5. The divergence time of the channel catfish and blue catfish with fossil calibrations (orange nodes) based on the mitochondrial cytochrome b sequences.

<https://doi.org/10.1371/journal.pone.0197371.g005>

sequences on the genome are listed in Table 1. Those categories contain more than 15 Mb of the novel repetitive elements in length and most of them are also A/T enriched. Although there were no previous annotations of those repetitive elements, they may still have potential functions in the genome evolutions or biological processes regulations. Our work provides a

Table 1. The major novel repetitive elements and their characteristics in the channel catfish repeatome.

Size	AT content	Representative scaffold	Representative scf_start	Representative scf_end
1,417K	64.1%	IpCoco_scf00172	16,372,640	16,373,425
1,316K	45.2%	lcl IpCoco_scf00610	3,078	6,525
1,081K	46.5%	IpCoco_scf00517	9,296,133	9,298,172
977K	46.0%	IpCoco_scf00474	141,399	143,674
963K	64.8%	IpCoco_scf00563_1077_652_655	1,427,889	1,431,117
924K	60.8%	IpCoco_scf00369	2,537,431	2,538,007
899K	50.1%	IpCoco_scf00203	3,637,116	3,638,889
881K	64.7%	IpCoco_scf00789	5,607	11,243
870K	64.9%	IpCoco_scf00540	509	4,548
842K	59.5%	IpCoco_scf00570_502_500	2,136,706	2,137,523
793K	44.5%	IpCoco_scf00419	7,439,807	7,440,803
690K	49.2%	IpCoco_scf00021	1,307,023	1,308,497
682K	47.2%	IpCoco_scf00077_78	260,739	262,006
630K	65.4%	IpCoco_scf00799	9,869	11,749
589K	53.3%	IpCoco_scf00354_353	643,250	644,319
546K	60.0%	IpCoco_scf00563_1077_652_655	1,372,133	1,373,248
522K	43.9%	IpCoco_scf00077_78	1,199,900	1,201,056
518K	66.6%	IpCoco_scf00172	2,251,598	2,252,801

<https://doi.org/10.1371/journal.pone.0197371.t001>

brief classification of those repetitive elements (S2 Table). However, whether those sequences are generated internally or are “molecular parasites” from external environments, as well as the more detailed identifications and annotations of the functions of those novel repetitive elements still deserve further studies especially experiment demonstrations.

Discussion

Repetitive elements in channel catfish

Using repetitive element databases combined with the nucleotide (nt) database, we identified, annotated, and characterized the repetitive elements in the channel catfish genome. Channel catfish harbors a large variety of repetitive elements in its genome, accounting for about 44% of its genome. The DNA transposons are the most abundant group of repetitive elements in the channel catfish genome, accounting for 15.9% of the catfish genome. These numbers are in line with our previous observations through genome sequence surveys [48, 52], but the data were analyzed from the whole genome and therefore is more complete.

The DNA/TcMar-Tc1 transposon sequences make up the highest percentage among Class II transposons in channel catfish genome, accounting for about ~20% of the total repetitive elements and are interspersed on the genome. The DNA-TcMar/Tc1 is a typical ‘cut-paste’ transposon ([77]), which is prevalent in nature and can be transferred not only vertically but also horizontally cross species during evolution [78]. It is this character that allows DNA-TcMar/Tc1 transposons to escape from the vertical extinction and being so abundant in nature [79–81]. Channel catfish is a freshwater benthopelagic species that inhabits in rapid fluctuating environments such as muddy ponds and rivers exposing to various biologic agents such as bacteria and viruses. Large amount of DNA/TcMar-Tc1 transposon footprints in channel catfish genome may indicate an external origin of parasitic transposable elements invasion to the genome during evolution [82]. As “parasitic” mobile elements, DNA transposons are known to be potent sources of mutation, and the long-time effective population shrinking in channel catfish can contribute to the evolution of more complex genomes such as more mobile elements or larger genome sizes [42, 83–84]. It is believed that the large amount of mobile transposons such as the DNA/TcMar-Tc1 can in turn contribute to the generation of novel genes and consequently facilitate considerably to species adaptations to novel environments [85–86]. Previous studies indicated that the transposition by a member of the Tc1/mariner family of transposable elements appears to have integrated in the duplicated C μ region of the immunoglobulin [87]. Channel catfish is a quite hardy fish species that can survive in a wide range of environmental conditions [88]. It is also possible that the prevalent of DNA/TcMar-Tc1 sequences, as well as other transposons in channel catfish genomes, play important roles in their adaptations to environments. Currently, there are no specific hotspots of DNA/TcMar-Tc1 on each individual chromosome observed.

Considerable amount of tandem repeats, especially microsatellite sequences, were found in the channel catfish genome. As short tandem DNA repeats of 2–8 nt long are ubiquitous in nearly all eukaryotic genomes [89–91], the expansion of microsatellites is disputable but it is generally considered to be expanded through DNA polymerase slippage [92–94]. High content of microsatellites in catfish genomes compared with other freshwater teleost such as tilapia or medaka [95, 96], indicates a high level of DNA polymerase slippage, may suggest a relationship to the high magnesium concentration (meq/L) in the channel catfish tissue compared with other teleost [97–98]. It was speculated that the magnesium concentration can contribute to DNA polymerase slippage by stabilizing the hairpin structure [99]. However, DNA polymerase slippage is a very complicated process that can be affected by various conditions including the genome structures (such as GC content), DNA repair mechanisms, flanking DNA sequences

(such as SINEs and LINEs), the centromere sequences and proteins involved in various DNA replication processes [100–109]. Whatever the mechanism is, high levels of microsatellites may help modulate the evolutionary mutation rate, thereby serving as a strategy to increase the species' versatility under stressful conditions [110–111]. Our analysis of the distribution of the microsatellites and satellites indicates that those short tandem repeats were mostly presented on the telomere and the centromere regions of the chromosome, consistent with the previous analysis [112–116].

The catfish genome also contains a large fraction of repetitive proteins in the repeatome. The main types of repetitive proteins are related to the adaptive immunology and metabolism as previous analysis indicated [38]. This may indicate that the abundance of repetitive genes in the genome is an adaptation that meets the large demand of immune defenses. Remarkably, there are at least 3.8MB of protein coding repetitive domains that are identified to be related to immunoglobulins in the channel catfish genome (Table 2). This may suggest that the expansion of the immunoglobulin family in the channel catfish genome can be one of the mechanisms of its defense against various pathogens.

The divergence of Xba elements sequence in channel catfish

The Xba elements are a group of A/T-rich repetitive sequences that were found in channel catfish and blue catfish centromeres but not in closely related species such as white catfish

Table 2. The major repetitive protein domains characterized from the channel catfish repeatome.

Rank	Total Abundance	Gene Ontology
1	2,990K	Adaptive immunity, Immunity
2	1,046K	Adaptive immunity, Immunity
3	375K	Cellular morphogenesis.
4	353K	Nucleus, cell junction, nucleoplasm
5	255K	Integral component of membrane
6	232K	Unknown
7	214K	Kinase, Transferase
8	204K	DNA replication, mismatch repair
9	203K	Kinase, Transferase
10	174K	Transferase, Ubl conjugation, pathway, zinc binding
11	145K	GTP-binding, Nucleotide-binding
12	132K	Structural molecule activity
13	131K	Osteogenesis, Transcription regulation
14	129K	Cell adhesion
15	124K	Adaptive immunity, Immunity
16	118K	Adaptive immunity, Immunity
17	114K	Guanine-nucleotide releasing factor
18	109K	Integral component of membrane
19	109K	Hydrolase, Ligase, Oxidoreductase, Amino-acid biosynthesis, Histidine biosynthesis, Methionine biosynthesis, One-carbon metabolism, Purine biosynthesis, ATP-binding, NADP, Nucleotide-binding
20	108K	Hormone
21	108K	Zinc binding, Cytosol, Plasma membrane
22	107K	Nucleotide-binding, DNA integration
23	107K	Endopeptidase inhibitor, liver development
24	106K	ATP binding, protein serine/threonine kinase activity

<https://doi.org/10.1371/journal.pone.0197371.t002>

(*Ameiurus catus*) and flathead catfish (*Pylodictus olivaris*) [43–44]. It is conserved among strains with minor changes in sequence identity and length, making it not only potentially important for genetic expression vectors but also of vital importance for the exploration of the channel catfish genome evolutions [43–44]. As the centromeres contain large amounts of DNA and are often packaged into heterochromatin, where the large-scale DNA sequences recombination and rearrangements varies greatly among phylogenetic related species [117–118]. The large amount of conservative *Xba* elements on centromere identified by fluorescence in situ hybridization [44] suggests a unique evolutionary status of the *Ictalurus* catfish. In addition, those centromeric repetitive sequences may be involved in centromere functions, such as kinetochore assembly and chromosome segregation during mitosis or meiosis [119–120], or even some epigenetic regulations [121].

Based on the number of substitutions per site and the divergence time, the rate of nucleotide substitutions of the *Xba* elements is calculated as 8.9×10^{-8} to 1.3×10^{-7} substitutions per site per year. Compared with the rate of nucleotide substitutions of full length cDNA calculated from the divergence level between the channel catfish and blue catfish [76], the rate of nucleotide substitutions of *Xba* elements is higher than that of the sequences in the open reading frames, but lower than those in untranslated regions. Slower rates of evolution suggest functional constraints [72]. The relatively slow evolutionary rate of *Xba* elements in catfish may indicate their potential functions, although unknown at present.

Conclusion

In this study, we identified 417.8 MB of repetitive sequences in the channel catfish genome, among which 84% were annotated. Among the annotated repetitive element, the most prevalent was the DNA/TcMar-Tc1 transposons, making up ~20% of the repeatome, followed by microsatellite (14%). A number of catfish-specific repetitive elements were identified including the previously known *Xba* elements. This work represents the most comprehensive analysis of the repeatome of the channel catfish genome with the best available chromosomal assembly so far, and it should facilitate the annotation of various teleost genomes.

Supporting information

S1 Table. A list of the major categories of repetitive elements in channel catfish and their percentage in the total repeatome.

(DOCX)

S2 Table. A list of the clustering of novel repetitive elements in channel catfish, ranked by the number of contained sequences.

(DOCX)

S1 File. The *cytb* sequences along with the accessions for inferring the phylogenetic tree and divergence time.

(FAS)

S2 File. The GFF files of repeat annotations of channel catfish scaffold-1.

(LZMA)

S3 File. The GFF files of repeat annotations of channel catfish scaffold-2.

(LZMA)

S4 File. The GFF files of repeat annotations of channel catfish degenerate sequences-1.

(LZMA)

S5 File. The GFF files of repeat annotations of channel catfish degenerate sequences-2.
(LZMA)

Acknowledgments

The authors would like to thank the Alabama Supercomputer Center and the Auburn Hopper supercomputer clusters for computational support.

Author Contributions

Conceptualization: Zihao Yuan, Zhanjiang Liu.

Data curation: Geoff Waldbieser, Zhanjiang Liu.

Formal analysis: Zihao Yuan, Tao Zhou.

Funding acquisition: Zhanjiang Liu.

Investigation: Zihao Yuan, Lisui Bao.

Methodology: Zihao Yuan, Shikai Liu.

Project administration: Tao Zhou, Shikai Liu.

Resources: Zhanjiang Liu.

Software: Zihao Yuan, Lisui Bao.

Supervision: Shikai Liu, Zhanjiang Liu.

Validation: Zihao Yuan.

Visualization: Zihao Yuan.

Writing – original draft: Zihao Yuan, Tao Zhou, Shikai Liu, Zhanjiang Liu.

Writing – review & editing: Huitong Shi, Yujia Yang, Dongya Gao, Rex Dunham.

References

1. Maumus F, Quesneville H. Deep investigation of Arabidopsis thaliana junk DNA reveals a continuum between repetitive elements and genomic dark matter. *PLoS One*. 2014; 9(4):e94101. <https://doi.org/10.1371/journal.pone.0094101> PMID: 24709859
2. Ohno S. So much "junk" DNA in our genome. In: Brookhaven symposia in biology. New York: Gordon and Breach; 1972. pp. 366–370. PMID: 5065367
3. Thomas CA Jr. The genetic organization of chromosomes. *Annu Rev Genet*. 1971; 5(1):237–56.
4. Meagher TR, Vassiliadis C. Phenotypic impacts of repetitive DNA in flowering plants. *New Phytol*. 2005; 168(1):71–80. <https://doi.org/10.1111/j.1469-8137.2005.01527.x> PMID: 16159322
5. Schmidt AL, Anderson LM. Repetitive DNA elements as mediators of genomic change in response to environmental cues. *Biol Rev*. 2006; 81(4):531–43. <https://doi.org/10.1017/S146479310600710X> PMID: 16893475
6. Thornburg BG, Gotea V, Makalowski W. Transposable elements as a significant source of transcription regulating signals. *Gene*. 2006; 365:104–10. <https://doi.org/10.1016/j.gene.2005.09.036> PMID: 16376497
7. Sun Y-B, Xiong Z-J, Xiang X-Y, Liu S-P, Zhou W-W, Tu X-L, et al. Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. *P Natl Acad Sci*. 2015; 112(11):E1257–E62.
8. Wang X, Fang X, Yang P, Jiang X, Jiang F, Zhao D, et al. The locust genome provides insight into swarm formation and long-distance flight. *Nat Commun*. 2014; 5:2957. <https://doi.org/10.1038/ncomms3957> PMID: 24423660

9. Yu L, Wang G-D, Ruan J, Chen Y-B, Yang C-P, Cao X, et al. Genomic analysis of snub-nosed monkeys (*Rhinopithecus*) identifies genes and processes related to high-altitude adaptation. *Nat Genet.* 2016.
10. SanMiguel P, Tikhonov A, Jin Y-K, Motchoulskaia N, Zakharov D, Melake-Berhan A, et al. Nested retrotransposons in the intergenic regions of the maize genome. *Science.* 1996; 274(5288):765–8. PMID: [8864112](#)
11. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet.* 2007; 8(12):973–82. <https://doi.org/10.1038/nrg2165> PMID: [17984973](#)
12. Ibarra-Laclette E, Lyons E, Hernández-Guzmán G, Pérez-Torres CA, Carretero-Paulet L, Chang T-H, et al. Architecture and evolution of a minute plant genome. *Nature.* 2013; 498(7452):94–8. <https://doi.org/10.1038/nature12132> PMID: [23665961](#)
13. Lee S-I, Kim N-S. Transposable elements and genome size variations in plants. *Genomics Inform.* 2014; 12(3):87–97. <https://doi.org/10.5808/GI.2014.12.3.87> PMID: [25317107](#)
14. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 maize genome: complexity, diversity, and dynamics. *science.* 2009; 326(5956):1112–5. <https://doi.org/10.1126/science.1178534> PMID: [19965430](#)
15. Vicient CM. Transcriptional activity of transposable elements in maize. *BMC Genomics.* 2010; 11(1):601.
16. Mccullers TJ, Steiniger M. Transposable elements in *Drosophila*. *Mobile Genetic Elements.* 2017; 7(3):1.
17. Kapitonov VV, Jurka J. Helitrons on a roll: eukaryotic rolling-circle transposons. *Trends in Genetics Tig.* 2007; 23(10):521. <https://doi.org/10.1016/j.tig.2007.08.004> PMID: [17850916](#)
18. Hurst GD, Schilthuizen M. Selfish genetic elements and speciation. *Heredity.* 1998; 80(1):2–8.
19. Hurst GD, Werren JH. The role of selfish genetic elements in eukaryotic evolution. *Nat Rev Genet.* 2001; 2(8):597–606. <https://doi.org/10.1038/35084545> PMID: [11483984](#)
20. Kazazian HH. An estimated frequency of endogenous insertional mutations in humans. *Nat Genet.* 1999; 22(2):130.
21. Kazazian HH. Mobile elements: drivers of genome evolution. *Science.* 2004; 303(5664):1626–32. <https://doi.org/10.1126/science.1089670> PMID: [15016989](#)
22. Kubis S, Schmidt T, Heslop-Harrison JSP. Repetitive DNA elements as a major component of plant genomes. *Ann Bot.* 1998; 82(suppl 1):45–55.
23. Tóth G, Gáspári Z, Jurka J. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* 2000; 10(7):967–81. PMID: [10899146](#)
24. Ugarković Đ, Plohl M. Variation in satellite DNA profiles—causes and effects. *EMBO J.* 2002; 21(22):5955–9. <https://doi.org/10.1093/emboj/cdf612> PMID: [12426367](#)
25. Hacch F, Mazrimas J. Fractionation and characterization of satellite DNAs of the kangaroo rat (*Dipodomys ordii*). *Nucleic Acids Res.* 1974; 1(4):559–76. PMID: [10793740](#)
26. Petitpierre E, Juan C, Pons J, Plohl M, Ugarkovic D. Satellite DNA and constitutive heterochromatin in tenebrionid beetles. In: *Kew Chromosome Conference IV.* London: Royal Botanic Gardens; 1995. pp. 351–362.
27. Charlesworth B, Sniegowski P, Stephan W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature.* 1994; 371:215–220. <https://doi.org/10.1038/371215a0> PMID: [8078581](#)
28. Lindahl T. DNA repair: DNA surveillance defect in cancer cells. *Curr Biol.* 1994; 4(3):249–51. PMID: [7922329](#)
29. Strand M, Prolla TA, Liskay RM, Petes TD. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature.* 1993; 365(6443):274–6. <https://doi.org/10.1038/365274a0> PMID: [8371783](#)
30. Balaresque P, King TE, Parkin EJ, Heyer E, Carvalho-Silva D, Kraaijenbrink T, et al. Gene Conversion Violates the Stepwise Mutation Model for Microsatellites in Y-Chromosomal Palindromic Repeats. *Hum Mutat.* 2014; 35(5):609–17. <https://doi.org/10.1002/humu.22542> PMID: [24610746](#)
31. Martin P, Makepeace K, Hill SA, Hood DW, Moxon ER. Microsatellite instability regulates transcription factor binding and gene expression. *P Natl Acad Sci USA.* 2005; 102(10):3800–4.
32. Moxon ER, Rainey PB, Nowak MA, Lenski RE. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Biol.* 1994; 4(1):24–33. PMID: [7922307](#)
33. Pardue M, Lowenhaupt K, Rich A, Nordheim A. (dC-dA) n.(dG-dT) n sequences have evolutionarily conserved chromosomal locations in *Drosophila* with implications for roles in chromosome structure and function. *EMBO J.* 1987; 6(6):1781–9. PMID: [3111846](#)

34. Richard GF, Pâques F. Mini-and microsatellite expansions: the recombination connection. *EMBO Rep.* 2000; 1(2):122–6. <https://doi.org/10.1093/embo-reports/kvd031> PMID: 11265750
35. Ingle J, Timmis JN, Sinclair J. The relationship between satellite deoxyribonucleic acid, ribosomal ribonucleic acid gene redundancy, and genome size in plants. *Plant Physiol.* 1975; 55(3):496–501. PMID: 16659109
36. Kobayashi T, Heck DJ, Nomura M, Horiuchi T. Expansion and contraction of ribosomal DNA repeats in *Saccharomyces cerevisiae*: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. *Gene Dev.* 1998; 12(24):3821–30. PMID: 9869636
37. Kobayashi T. Regulation of ribosomal RNA gene copy number and its role in modulating genome integrity and evolutionary adaptability in yeast. *Cell Mol Life Sci.* 2011; 68(8):1395–403. <https://doi.org/10.1007/s00018-010-0613-2> PMID: 21207101
38. Bengtén E, Clem LW, Miller NW, Warr GW, Wilson M. Channel catfish immunoglobulins: repertoire and expression. *Dev Comp Immunol.* 2006; 30(1):77–92.
39. Nagarajan N, Navajas-Pérez R, Pop M, Alam M, Ming R, Paterson AH, et al. Genome-wide analysis of repetitive elements in papaya. *Trop Plant Biol.* 2008; 1(3–4):191–201.
40. USDA-NASS (United States Department of Agriculture, National Agriculture Statistic Service). Census of Aquaculture, Volume 3, Special Studies, Part 2. Washington, DC: United States Department of Agriculture; 2005.
41. United States Department of Agriculture. Census of Aquaculture, Volume 1, Geographic Area Series, Part 51. Washington, DC: United States Department of Agriculture; 2014.
42. Liu Z, Liu S, Yao J, Bao L, Zhang J, Li Y, et al. The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. *Nat Commun.* 2016; 7:11757. <https://doi.org/10.1038/ncomms11757> PMID: 27249958
43. Liu Z, Li P, Dunham RA. Characterization of an A/T-rich family of sequences from channel catfish (*Ictalurus punctatus*). *Mol Mar Biol Biotech.* 1998; 7:232–9.
44. Quiniou S, Wolters W, Waldbieser G. Localization of Xba repetitive elements to channel catfish (*Ictalurus punctatus*) centromeres via fluorescence in situ hybridization. *Anim Genet.* 2005; 36(4):353–4. <https://doi.org/10.1111/j.1365-2052.2005.01304.x> PMID: 16026349
45. Kim S, Karsi A, Dunham RA, Liu Z. The skeletal muscle α -actin gene of channel catfish (*Ictalurus punctatus*) and its association with piscine specific SINE elements. *Gene.* 2000; 252(1):173–81.
46. Liu Z, Li P, Kucuktas H, Dunham R. Characterization of nonautonomous Tc1-like transposable elements of channel catfish (*Ictalurus punctatus*). *Fish Physiol Biochem.* 1999; 21(1):65–72.
47. Chen X, Zhong L, Bian C, Xu P, Qiu Y, You X, et al. High-quality genome assembly of channel catfish, *Ictalurus punctatus*. *GigaScience.* 2016; 5(1):39. <https://doi.org/10.1186/s13742-016-0142-5> PMID: 27549802
48. Xu P, Wang S, Liu L, Peatman E, Somridhivej B, Thimmapuram J, et al. Channel catfish BAC-end sequences for marker development and assessment of syntenic conservation with other fish species. *Anim Genet.* 2006; 37(4):321–6. <https://doi.org/10.1111/j.1365-2052.2006.01453.x> PMID: 16879340
49. Xu P, Wang S, Liu L, Thorsen J, Kucuktas H, Liu Z. A BAC-based physical map of the channel catfish genome. *Genomics.* 2007; 90(3):380–8. <https://doi.org/10.1016/j.ygeno.2007.05.008> PMID: 17582737
50. Serapion J, Kucuktas H, Feng J, Liu Z. Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). *Mar Biotechnol.* 2004; 6(4):364–77. <https://doi.org/10.1007/s10126-003-0039-z> PMID: 15136916
51. Nandi S, Peatman E, Xu P, Wang S, Li P, Liu Z. Repeat structure of the catfish genome: a genomic and transcriptomic assessment of Tc1-like transposon elements in channel catfish (*Ictalurus punctatus*). *Genetica.* 2007; 131(1):81–90. <https://doi.org/10.1007/s10709-006-9115-4> PMID: 17091335
52. Liu Z. Development of genomic resources in support of sequencing, assembly, and annotation of the catfish genome. *Comp Biochem Phys D.* 2011; 6(1):11–7.
53. Bao Z, Eddy SR. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* 2002; 12(8):1269–76. <https://doi.org/10.1101/gr.88502> PMID: 12176934
54. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. *Bioinformatics.* 2005; 21(suppl 1):i351–i8.
55. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA.* 2015; 6(1):1.
56. Wheeler TJ, Clements J, Eddy SR, Hubley R, Jones TA, Jurka J, et al. Dfam: a database of repetitive DNA based on profile hidden Markov models. *Nucleic Acids Res.* 2013; 41(D1):D70–D82.

57. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*. 2010; 26(5):680–2. <https://doi.org/10.1093/bioinformatics/btq003> PMID: 20053844
58. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006; 22(13):1658–9. <https://doi.org/10.1093/bioinformatics/btl158> PMID: 16731699
59. Deng W, Wang Y, Liu Z, Cheng H, Xue Y. HemI: a toolkit for illustrating heatmaps. *PLoS One*. 2014; 9(11):e111988. <https://doi.org/10.1371/journal.pone.0111988> PMID: 25372567
60. Yang J-Q, Tang W-Q, Liao T-Y, Sun Y, Zhou Z-C, Han C-C, et al. Phylogeographical analysis on *Squalidus argentatus* recapitulates historical landscapes and drainage evolution on the island of Taiwan and mainland China. *Int J Mol Sci*. 2012; 13(2):1405–25. <https://doi.org/10.3390/ijms13021405> PMID: 22408398
61. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*. 2012; 29(8):1969–73. <https://doi.org/10.1093/molbev/mss075> PMID: 22367748
62. Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics*. 2002; 161(3):1307–20. PMID: 12136032
63. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*. 2011; 27(8):1164–5. <https://doi.org/10.1093/bioinformatics/btr088> PMID: 21335321
64. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 2006; 4(5):e88. <https://doi.org/10.1371/journal.pbio.0040088> PMID: 16683862
65. Drummond AJ, Suchard MA. Bayesian random local clocks, or one rate to rule them all. *BMC Biol*. 2010; 8(1):114.
66. Blanton RE, Page LM, Hilber SA. Timing of clade divergence and discordant estimates of genetic and morphological diversity in the Slender Madtom, *Noturus exilis* (Ictaluridae). *Mol Phylogenet Evol*. 2013; 66(3):679–93. <https://doi.org/10.1016/j.ympev.2012.10.022> PMID: 23149097
67. Hardman M, Hardman LM. The relative importance of body size and paleoclimatic change as explanatory variables influencing lineage diversification rate: an evolutionary analysis of bullhead catfishes (Siluriformes: Ictaluridae). *Syst Biol*. 2008; 57(1):116–30. <https://doi.org/10.1080/10635150801902193> PMID: 18288621
68. Wang Q, Xu C, Xu C, Wang R. Complete mitochondrial genome of the Southern catfish (*Silurus meridionalis* Chen) and Chinese catfish (*S. asotus* Linnaeus): Structure, phylogeny, and intraspecific variation. *Genet Mol Res*. 2015; 14(4):18198–209. <https://doi.org/10.4238/2015.December.23.7> PMID: 26782467
69. Werner C. Die kontinentale Wirbeltierfauna aus der unteren Oberkreide des Sudan (Wadi Milk Formation). *Berliner Geowissenschaftliche Abhandlungen E (B Krebs-Festschrift)* 1994; 13:221–249.
70. Benton MJ. *The fossil record 2*. Chapman and Hall; 1993.
71. Chinwalla AT, Cook LL, Delehaunty KD, Fewell GA, Fulton LA, Fulton RS, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature*. 2002; 420(6915):520–62. <https://doi.org/10.1038/nature01262> PMID: 12466850
72. Li W, Graur D, Sunderland M. *Fundamental of Molecular Evolution*. Sunderland: Sinauer Associates; 1992. pp. 67–98.
73. Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, et al. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol*. 2013; 14(1):R10. <https://doi.org/10.1186/gb-2013-14-1-r10> PMID: 23363705
74. Wicky C, Villeneuve AM, Lauper N, Codourey L, Tobler H, Müller F. Telomeric repeats (TTAGGC) n are sufficient for chromosome capping function in *Caenorhabditis elegans*. *P Natl Acad Sci*. 1996; 93(17):8983–8.
75. Kamnert I, Lopez CC, Rosen M, Edstrom J-E. Telomeres terminating with long complex tandem repeats. *Hereditas*. 1997; 127(3):175–80. PMID: 9474901
76. Chen F, Lee Y, Jiang Y, Wang S, Peatman E, Abernathy J, et al. Identification and characterization of full-length cDNAs in channel catfish (*Ictalurus punctatus*) and blue catfish (*Ictalurus furcatus*). *PLoS One*. 2010; 5(7):e11546. <https://doi.org/10.1371/journal.pone.0011546> PMID: 20634964
77. Finnegan D. Transposable elements in eukaryotes. *Int Rev Cytol*. 1985; 93:281–326. PMID: 2989205
78. Aziz RK, Breitbart M, Edwards RA. Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res*. 2010; gkq140.

79. Lawrence JG, Hartl D. Inference of horizontal genetic transfer from molecular data: an approach using the bootstrap. *Genetics*. 1992; 131(3):753–60. PMID: [1628816](#)
80. Lohe AR, Moriyama EN, Lidholm D-A, Hartl DL. Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. *Mol Biol Evol*. 1995; 12(1):62–72. <https://doi.org/10.1093/oxfordjournals.molbev.a040191> PMID: [7877497](#)
81. Maruyama K, Hartl DL. Evidence for interspecific transfer of the transposable element mariner between *Drosophila* and *Zaprionus*. *J Mol Evol*. 1991; 33(6):514–24. PMID: [1664000](#)
82. Capy P, Gasperi G, Biémont C, Bazin C. Stress and transposable elements: co-evolution or useful parasites? *Heredity*. 2000; 85(2):101.
83. Lynch M, Conery JS. The origins of genome complexity. *Science*. 2003; 302(5649):1401–4. <https://doi.org/10.1126/science.1089370> PMID: [14631042](#)
84. Yi S, Streebman JT. Genome size is negatively correlated with effective population size in ray-finned fish. *Trends Genet*. 2005; 21(12):643–6. <https://doi.org/10.1016/j.tig.2005.09.003> PMID: [16213058](#)
85. González J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA. High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. *PLoS Biol*. 2008; 6(10):e251. <https://doi.org/10.1371/journal.pbio.0060251> PMID: [18942889](#)
86. González J, Macpherson JM, Petrov DA. A recent adaptive transposable element insertion near highly conserved developmental loci in *Drosophila melanogaster*. *Mol Biol Evol*. 2009; 26(9):1949–61. <https://doi.org/10.1093/molbev/msp107> PMID: [19458110](#)
87. Ventura-Holman T, Lobb CJ. Structural organization of the immunoglobulin heavy chain locus in the channel catfish: the IgH locus represents a composite of two gene clusters. *Mol Immunol*. 2002; 38(7):557–64. PMID: [11750657](#)
88. Wellborn TL. Channel catfish: life history and biology. Stoneville: SRAC publication; 1988.
89. Tautz D, Renz M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res*. 1984; 12(10):4127–38. PMID: [6328411](#)
90. Tautz D, Trick M, Dover GA. Cryptic simplicity in DNA is a major source of genetic variation. *Nature*. 1985; 322(6080):652–6.
91. Weber JL. Informativeness of human (dC-dA) n-(dG-dT) n polymorphisms. *Genomics*. 1990; 7(4):524–30. PMID: [1974878](#)
92. Levinson G, Gutman GA. High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res*. 1987a; 15(13):5323–38.
93. Levinson G, Gutman GA. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol*. 1987b; 4(3):203–21.
94. Schlötterer C, Tautz D. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res*. 1992; 20(2):211–5. PMID: [1741246](#)
95. Conte MA, Gammerdinger WJ, Bartie KL, Penman DJ, Kocher TD. A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC genomics*. 2017; 18(1):341. <https://doi.org/10.1186/s12864-017-3723-5> PMID: [28464822](#)
96. Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, et al. The medaka draft genome and insights into vertebrate genome evolution. *Nature*. 2007; 447(7145):714. <https://doi.org/10.1038/nature05846> PMID: [17554307](#)
97. Chen C-Y, Wooster GA, Getchell RG, Bowser PR, Timmons MB. Blood chemistry of healthy, nephrocalcinosis-affected and ozone-treated tilapia in a recirculation system, with application of discriminant analysis. *Aquaculture*. 2003; 218(1):89–102.
98. Wedemeyer G. *Physiology of fish in intensive culture systems*. Boston: Springer; 1996. pp. 54.
99. Castillo-Lizardo M, Henneke G, Viguera E. Replication slippage of the thermophilic DNA polymerases B and D from the Euryarchaeota *Pyrococcus abyssi*. *Front Microbiol*. 2014; 5(403). <https://doi.org/10.3389/fmicb.2014.00403> PMID: [25177316](#)
100. Bachtrog D, Weiss S, Zangerl B, Brem G, Schlötterer C. Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. *Mol Biol Evol*. 1999; 16(5):602–10. <https://doi.org/10.1093/oxfordjournals.molbev.a026142> PMID: [10335653](#)
101. Castoe TA, Hall KT, Mboulas MLG, Gu W, de Koning AJ, Fox SE, et al. Discovery of highly divergent repeat landscapes in snake genomes using high-throughput sequencing. *Genome Biol Evol*. 2011; 3:641–53. <https://doi.org/10.1093/gbe/evr043> PMID: [21572095](#)
102. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nat Rev Genet*. 2009; 10(10):691–703. <https://doi.org/10.1038/nrg2640> PMID: [19763152](#)
103. Glenn TC, Stephan W, Dessauer HC, Braun MJ. Allelic diversity in alligator microsatellite loci is negatively correlated with GC content of flanking sequences and evolutionary conservation of PCR

- amplifiability. *Mol Biol Evol.* 1996; 13(8):1151–4. <https://doi.org/10.1093/oxfordjournals.molbev.a025678> PMID: 8865669
104. Grady DL, Ratliff RL, Robinson DL, McCanlies EC, Meyne J, Moyzis RK. Highly conserved repetitive DNA sequences are present at human centromeres. *P Natl Acad Sci.* 1992; 89(5):1695–9.
 105. Janes DE, Organ CL, Fujita MK, Shedlock AM, Edwards SV. Genome evolution in Reptilia, the sister group of mammals. *Annu Rev Genom Hum G.* 2010; 11:239–64.
 106. Mellon I, Rajpal DK, Koi M, Boland CR, Champe GN. Transcription-coupled repair deficiency and mutations in human mismatch repair genes. *Science.* 1996; 272(5261):557. PMID: 8614807
 107. Plohl M, Meštrović N, Mravinac B. Centromere identity from the DNA point of view. *Chromosoma.* 2014; 123(4):313–25. <https://doi.org/10.1007/s00412-014-0462-0> PMID: 24763964
 108. Primmer CR, Raudsepp T, Chowdhary BP, Møller AP, Ellegren H. Low frequency of microsatellites in the avian genome. *Genome Res.* 1997; 7(5):471–82. PMID: 9149943
 109. Ellegren H. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet.* 2004; 5(6):435–45. <https://doi.org/10.1038/nrg1348> PMID: 15153996
 110. Chang DK, Metzgar D, Wills C, Boland CR. Microsatellites in the eukaryotic DNA mismatch repair genes as modulators of evolutionary mutation rate. *Genome Res.* 2001; 11(7):1145–6. <https://doi.org/10.1101/gr.186301> PMID: 11435395
 111. Rocha EP, Matic I, Taddei F. Over-representation of repeats in stress response genes: a strategy to increase versatility under stressful conditions? *Nucleic Acids Res.* 2002; 30(9):1886–94. PMID: 11972324
 112. Areshchenkova T, Ganai MW. Long tomato microsatellites are predominantly associated with centromeric regions. *Genome.* 1999; 42(3):536–44. PMID: 10382301
 113. Vidaurreta M, Maestro M, Rafael S, Veganzones S, Sanz-Casla M, Cerdán J, et al. Telomerase activity in colorectal cancer, prognostic factor and implications in the microsatellite instability pathway. *World J Gastroentero.* 2007; 13(28):3868.
 114. Wicky C, Villeneuve AM, Lauper N, Codourey L, Tobler H, Müller F. Telomeric repeats (TTAGGC) n are sufficient for chromosome capping function in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences.* 1996; 93(17):8983–8.
 115. Blackburn EH. Structure and function of telomeres. *Nature.* 1991; 350(6319):569. <https://doi.org/10.1038/350569a0> PMID: 1708110
 116. Rosenberg M, Hui L, Ma J, Nusbaum HC, Clark K, Robinson L, et al. Characterization of short tandem repeats from thirty-one human telomeres. *Genome research.* 1997; 7(9):917–23. PMID: 9314497
 117. Mehta GD, Agarwal MP, Ghosh SK. Centromere identity: a challenge to be faced. *Mol Genet Genomics.* 2010; 284(2):75–94. <https://doi.org/10.1007/s00438-010-0553-4> PMID: 20585957
 118. Wang G, Zhang X, Jin W. An overview of plant centromeres. *J Genet Genomics.* 2009; 36(9):529–37. [https://doi.org/10.1016/S1673-8527\(08\)60144-7](https://doi.org/10.1016/S1673-8527(08)60144-7) PMID: 19782954
 119. Pidoux AL, Allshire RC. The role of heterochromatin in centromere function. *Philos T R Soc B.* 2005; 360(1455):569–79.
 120. Westhorpe FG, Straight AF. Functions of the centromere and kinetochore in chromosome segregation. *Curr Opin Cell Biol.* 2013; 25(3):334–40. <https://doi.org/10.1016/j.ceb.2013.02.001> PMID: 23490282
 121. Karpen GH, Allshire RC. The case for epigenetic effects on centromere identity and function. *Trends Genet.* 1997; 13(12):489–96. PMID: 9433139